Identification of non-enterococci microbes carrying the transposon Tn*1549* using CRISPR-Cas9 enrichment for selective long-read sequencing

Siri Bildøy

This thesis is submitted in partial fulfillment of the requirements for the degree of Master of science

Department of Biological Science
Faculty of Mathematics and Natural Science
University of Bergen

June 2022

Acknowledgments

The work in this project was carried out at the Department of Microbiology, Haukeland University Hospital, Bergen. The Department of Microbiology has been very helpful in facilitating working hours and shifts throughout my studies and deserves a big thank you.

Further, I would like to thank my supervisors Randi M. Nygaard and Øyvind Kommedal who had the idea for this thesis and engaged me to dive into the world of *vanB* and anaerobe bacteria. You have been a great inspiration since we started, with endless discussions, thorough and valuable feedback on my writing, and an amazing enthusiasm every step of the way. You both have taught me a lot, and I thank you for your patience and encouraging words throughout this past year, and for giving me the opportunity to work on such a rewarding project.

Secondly, I have to thank my amazing co-workers at the Department of Microbiology. You have been a valuable support for the last two years, both academically and socially. A special thanks to Audun Sivertsen for his amazing help with bioinformatics, and for guiding me into an unknown world of coding and bioinformatic tools. Also, a big thanks to Torbjørn Bruvold for your help and enthusiasm in the initial phase of the project.

Also, I must thank my fellow student and friend Camilla Nundal for your encouragement and for sharing my emotional ups and downs over endless conversations and diners. Even though we have worked on different projects, we did this together!

Finally, I must thank my friends and family, especially my partner Håkon for always encouraging me. You always put me and my needs before yours, and I am grateful for you putting up with me and my study-related stressful mood. I really appreciate the effort you and everyone else have made to make the most of my spare time in between studying the past years.

Bergen, June 2022
Siri Bildøy

TABLE OF CONTENTS

SELECTED ABBREVIATIONS

BLAST	Basic local alignment search tool
bp	Base pair
CRISPR	Clustered regulatory interspaced palindromic repeats
Ct	Cycle threshold
HDR	Homology directed repair
HGT	Horizontal gene transfer
HUS	Haukeland University Hospital
IDT	Integrated DNA technologies
MIA	Department of microbiology
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining repair
nusG	N-utilization substance, protein G
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
PAM	Protospacer adjacent motif
RNP	Ribonucleoprotein
rpoB	RNA polymerase, β subunit
VB	Vancomycin broth
VRE	Vancomycin resistant enterococci

SUMMARY

Vancomycin-resistant enterococci (VRE) have during the last decades emerged and become a public health threat. At Haukeland University Hospital (HUS) in Bergen (Norway), VRE outbreaks have been closely linked to the *vanB*-carrying transposon Tn*1549*. The transposon has also been observed in other intestinal microbes like *Clostridium symbiosum*, and transfer of Tn*1549* to enterococcal species has been demonstrated. These leads to the hypothesis that the formation of VRE may occur in patients' own intestines, where the *vanB* gene possibly is transferred to enterococci from multiple unknown *vanB*-carrying bacteria.

We aimed to develop a culture independent method for identification of Tn1549 harboring species. Many intestinal bacteria that might be carrying the *vanB* gene have strict anaerobic requirements for growth, and are hard to isolate and culture. As an alternative, long-read sequencing technology from Oxford Nanopore Technologies in combination with targeted CRISPR-Cas9 enrichment represents an attractive approach. By use of guide-RNAs, the CRISPR-Cas9 method performs targeted digestion to enrich genomes of interest prior to sequencing. We designed crRNAs targeting both ends of the transposon Tn*1549*, allowing enrichment of the sequences flanking the transposon with information about the host bacteria. Prior to CRISPR-Cas9 enrichment we also performed an enhanced culture-based enrichment of vancomycin-resistant bacteria as an attempt to up concentrate *vanB*-carrying bacteria.

The method was established and evaluated using DNA extracted form cultivated *Enterococcus faecium* carrying the transposon Tn*1549*. We found that pooling of the crRNAs targeting the 5´ and 3´ target ends on the transposon, compared to separate crRNA-directed Cas9 targeting reactions resulted in increased sequencing output by 2-fold to a 63.1% on-target yield. When the method was established, samples from the VRE-surveillance program at HUS with strongly positive vanB-PCR results were collected. Two clinical samples with Ct values below 20 were chosen for targeted CRISPR-Cas9 enrichment and sequencing. The samples obtained an on-target yield of 4.25% and 1.95% from a total of 47 and 154 reads respectively. Despite low yield, we managed to identify two non-enterococcal Tn*1549 vanB* positive bacterial species; *Flavonifractor plautii* and *Enterocloster clostridioformis*. *Flavonifractor plautii* has not previously been described as carrying the vancomycin resistance gene and the result was verified using metagenomic sequencing.

1.0 Introduction

1.1 The emergence of vancomycin-resistant enterococci

Over the past decades, the world has seen a significant increase in the occurrence of antimicrobial resistance in a variety of bacterial species. In Europe, geographical patterns have been observed as to what type of resistance and which drug-resistant bacteria are dominant in different areas of the continent. However, vancomycin-resistant *Enterococcus* stands out as a species not having any particular geographical distribution; (ECDC, 2020). The emergence of vancomycin-resistant enterococci (VRE) is described as a public health threat causing infections in European and Norwegian healthcare facilities (FHI, 2019). Consequently, understanding and learning about their ability to adapt and spread in these environments are of great importance.

Enterococcal species represent less than one percent of the intestinal microbiota in humans and are known to be opportunistic hospital pathogens that cause infection, especially in long-term admitted patients (Van Tyne & Gilmore, 2014). Both *Enterococcus faecalis* and *Enterococcus faecium* are known to have decreased sensibility to several drugs where many of the properties for tolerance are acquired (Clewell, 1990; Hodges et al., 1992). Intestinal enterococci live in a complex microbial community, with many potential sources of genetic material that can change its properties (Gill et al., 2006). This, combined with long-lasting antibiotic exposure, creates an environment for the selection of the enterococci as well as other bacteria tolerant to antibiotics, through both mutation and genetic transfer (Clewell, 1990).

At Haukeland University Hospital (HUS) (Bergen, Norway), there have been several outbreaks of VRE where whole-genome sequencing of cultivated bacteria has shown that the vancomycin resistance has emerged due to the *vanB*-gene containing transposon Tn*1549*. The *vanB*-gene cluster causes vancomycin resistance by providing an alternative biosynthetic pathway for the production of cell wall precursors that binds vancomycin poorly. The *vanB*-gene has been found in other intestinal microbes in hospitals around the world (Ballard et al., 2005; Domingo et al., 2005) and transfer of a *vanB* containing transposon from *Clostridium symbosium* to both *E. faecium* and *E. faecalis* has been demonstrated in the gut of gnotobiotic mice (Launay et al., 2006). This has led to the hypothesis that VRE-carriage in hospitalized patients might not necessarily be caused by the acquisition of a VRE from the hospital environment but that it can

also occur de novo in the patient's own intestine by *vanB* gene uptake by previously vancomycin susceptible enterococci. As of today, there is no clear answer to what stimulates the transmission of *vanB* or Tn*1549* (Patel, 2000). Also, there is little knowledge about the spectrum of non-enterococci *vanB* carrying bacteria, and further mapping of this potential *vanB*-gene reservoir is of great interest. At HUS, VRE screening of patients is part of the laboratory routine analyses. The samples are first enriched in a vancomycin-containing broth overnight, prior to analysis by a VRE real-time PCR targeting both *vanA* and *vanB*. Only samples with a positive result below a certain PCR cycle threshold are further subcultured on a solid vancomycin-containing agar to look for growth of VRE colonies (Palladino et al., 2003). This method allows for the identification of *vanB* positive samples, but not the identification of *vanB* harboring non-enterococci (Mak et al., 2009) due to the strict anaerobic and often highly specific conditions many intestinal microbes require for growth. Therefore, to look for intestinal non-enterococcal *vanB*-gene positive bacteria, new methods like DNA-sequencing with targeted enrichment may prove to be more applicable.

1.2 Passing of genes

The classic way of genetic transfer is by inheritance from mother to daughter cells. From early on it was believed that the passing of DNA occurred only from mother cell to daughter cell, a theory we now know cannot be true as there are tremendous amounts of proof that genomes of one species may contain gens with close homology to both close and distant relatives (Cassier-Chauvat et al., 1997; Jain et al., 2002). Genetic transfer due to mechanisms other than classical inheritance is called horizontal gene transfer (HGT). One mechanism of HGT is conjugation, also referred to as bacterial sex, a contact-based transfer. Bacteria connect through a pilus (a tube-like structure) to transfer conjugative genetic elements like plasmids from the donor cell to the recipient cell (Thomas & Nielsen, 2005). Another mechanism of HGT is transformation, a process in which a cell with competence uptake exogenous genetic material directly from its surroundings. For a bacterial cell to be competent it needs to have the genes necessary to regulate the permeability of the cell wall, enabling free DNA into the cell (Blokesch, 2016). The third mechanism of HGT is transduction, a process by which a bacteriophage (a virus that infects bacteria) enters the bacterial cell and introduces foreign DNA. Non-viral foreign DNA is picked up by bacteriophages when they occupy a bacterial cell and is released into the next bacterium they enter (Figure 1) (Yosef et al., 2017).

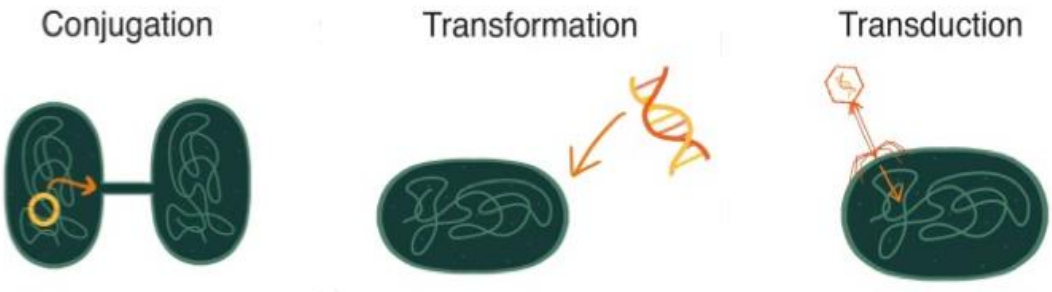


Figure 1: Illustration of horizontal gene transfer by the event of conjugation, transformation, and transduction. Conjugation: a plasmid is transferred from donor to recipient through pilus. Transformation: foreign DNA is taken up by a competent cell from its surroundings. Transduction: bacteriophage containing intrudes a bacteria and leaves foreign DNA (Moralez et al., 2021).

Bacterial DNA exchange is crucial for the ability to obtain tolerance to the many antimicrobial drugs used today. Transfer of phage DNA, transposons, plasmid, and chromosomal DNA are important contributors to the exchange of genes that encode antimicrobial resistance. Plasmids are genetic elements that are not a part of the chromosome but contain the genetic information necessary for self-replication. Transposable elements, like transposons, do not contain this information and therefore cannot act on their own as independent molecules. In order to replicate, the transposon must be integrated into a host genome like a plasmid, chromosomal DNA, or a virus (Clark & Pazdernik, 2016). Several studies have shown that plasmids and phages have an important role in the transmission of transposons harboring antibiotic and other drug resistance genes through horizontal transfer (Colomer-Lluch et al., 2011; Frost et al., 2005). The most common and well-studied transposons are the ones transferred by conjugation, hence the name conjugative transposable elements.

Transposons are divided into the retrotransposons (class I) and the DNA transposons (class II). Bacterial transposons belong to class II and are DNA transposons in the Tn family (Babakhani & Oloomi, 2018). Transposons are classified based on their mechanism of movement. To move, a transposon is dependent on the enzyme transposase which catalyzes DNA movement by a process called transposition – the movement of genes from one location to another (Lambert, 2007). The enzyme recognizes specific repetitive sequences near each end of the transposon, where it excises either the entire transposon or just a single strand of it (Egner & Berg, 1981). The enzyme also recognizes a target sequence on the recipient molecule, where it inserts the transposon. In the process of insertion, the target sequence on the host is replicated, resulting

in one target sequence on each side of the inserted transposon (Clark & Pazdernik, 2016). Depending on the method of excision and insertion the transposition of the transposon is either replicative or conservative (Figure 2). In replicative transposition, the transposase enzyme makes single-stranded nicks at both ends of the transposon and inserts the single strand to the target sequence of the new host genome. Now both the donor and the recipient each have a single-stranded copy of the transposon. This leads to a response from the host repair system to make a complementary strand, thus there are now two copies of the transposon (Derbyshire & Grindley, 1986). The conservative transposition is also called the cut-and-paste transposition. Here, the transposase enzyme recognizes the repetitive sequences and excises the entire transposon, leaving the donor host DNA with a break. The enzyme then inserts the entire transposon into the target sequence of the new host genome (Clark & Pazdernik, 2016).

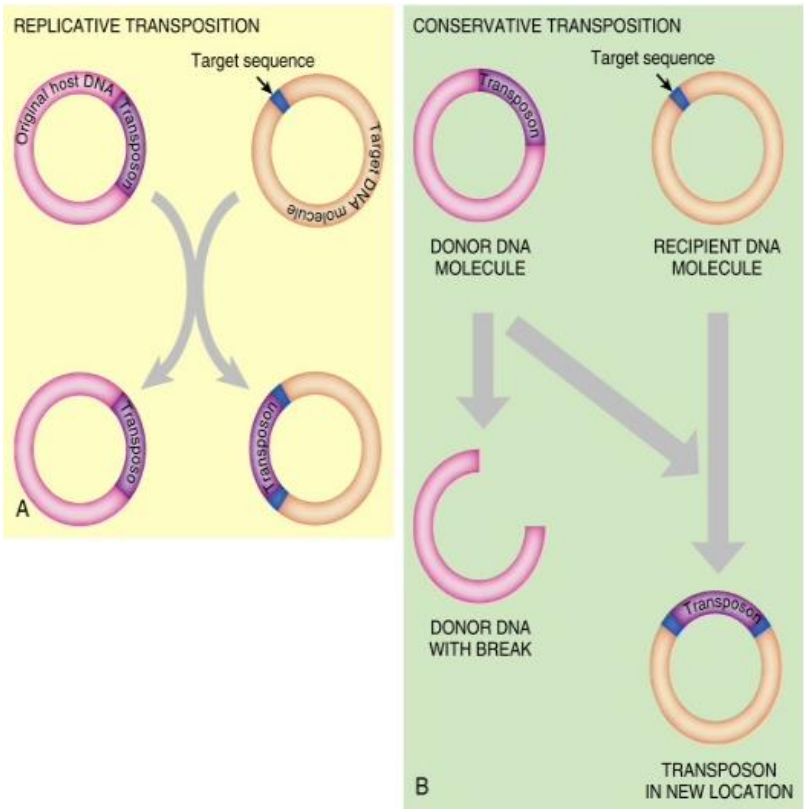


Figure 2: Replicative (A) and conservative (B) transposition of the transposon. A: A single strand of the transposon is nicked from the donor DNA and inserted into the target sequence of the recipient DNA. The host makes the complementary sequences resulting in two copies of the transposon. B: The entire transposon is cut out from the donor DNA, and inserted into the recipient, leaving the donor DNA with a break. (Clark & Pazdernik, 2016).

1.3 The *vanB* gene and Transposon Tn*1549*

Many enterococci have acquired resistance to several drugs used in the treatment of Gram-positive microbes, e.g. high-level resistance against beta-lactam antibiotics and aminoglycosides in *E. faecium* (Sood et al., 2008). This leaves glycopeptides like vancomycin as one of the few effective treatments, making the spread of VRE a great concern. Different classes of *van* genes have been characterized during the last decades, with *vanA* and *vanB* seemingly being the biggest threats (Sadowy, 2021). The mechanism behind resistance is the synthesis of enzymes that produce alternate peptidoglycan precursors where the normal dipeptide termini D-alanine- D-alanine has been replaced with the depsipeptide D-alanine-D-lactate. Changing the peptide sequence prevents vancomycin from binding to the peptide chains (Bugg et al., 1991). Today, we are aware of three different variants of *vanB* (*vanB1*, *2*, and 3) where *vanB2* seems to be the most prevalent. One reason for this is thought to be the increasing number of mobile elements harboring the *vanB2* resistance genes. These elements disseminate both by clonal spread as well as by horizontal gene transfer, one of which is the transposon Tn*1549* (Gold et al., 1993; Zhou et al., 2018).

Tn*1549* is a 34kb *vanB2*-type vancomycin-resistance conjugative transposon first discovered in enterococci. The transposon was carried by the conjugative plasmid pAD1 where it was inserted into the *traE1* gene, a gene that activates its own transcription as well as the transcription of other genes – all involved in the regulation of a pheromone response stimulating conjugation (Garnier et al., 2000; Pontius & Clewell, 1992).

Tn*1549* contains three functional modules; One is encoding *vanB* type vancomycin resistance genes and is very similar to the *vanB* genes of another transposon; Tn*5382* (Carias et al., 1998). Another region is encoding the enzymes necessary for conjugation and mobilization, and the last region is implicated in the process of excision and insertion (Figure 3) (Garnier et al., 2000). Like the transposon Tn*5383*, Tn*1549* can excise directly and form a circular intermediate by recombination of the two transposon ends before insertion into a new host. The transposons do not seem to have specific requirements for the insertion site as the insertions seem to occur at random loci in the host genome (Dahl et al., 2003). The regions containing genes for excision and insertion have shown to share significant homology with the corresponding genes of the Tn*916* family. Tn*1549* also has structural parallels to this transposon family for the open reading frames associated with conjugation (Garnier et al., 2000). Sequence comparisons show that the differences between the two are essentially the region responsible for *vanB* resistance

in Tn*1549* which has replaced the tetracycline resistance protein tet(M) that is found in Tn*916* (Garnier et al., 2000; Sadowy, 2021).

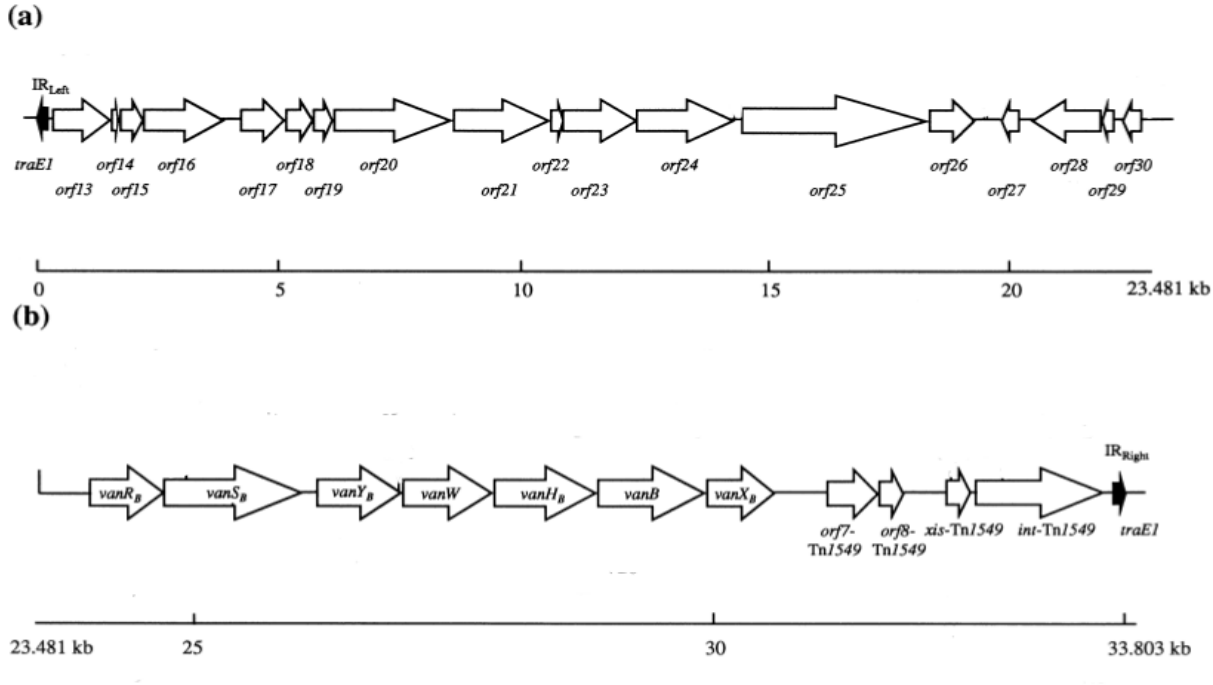


Figure 3: Schematic illustration of the functional modules of transposon Tn1549. The arrows indicate open reading frames (ORFs) as well as the direction of transcription. The left part **(a)** represents genes responsible for mobilization and conjugation. The right part **(b)** represent the genes of the *vanB* operon (*vanRB, vanSB, vanYB, vanW*, *vanHB, vanB* and *vanXB*), followed by the excisionase (xis) and integrase (int) genes. Inverted repeats (IR) at each end of the transposon are also illustrated. (Adapted from Garnier et al., 2000).

After Tn*1549* had been completely sequenced from a clinical isolate of *E. faecalis* in the UK, findings of the transposon were quickly reported from not only other countries but also from other continents (Ballard et al., 2005; Lu et al., 2005; Sivertsen et al., 2014). Most findings were from *E. faecium* isolates and showed that the transposon consistently carried the *vanB2*-subtype. This led to studies of linkage between other gram-positives carrying the *vanB2*-subtype, where the transposon was confirmed in several other microbes like *Clostridium* spp. and *Eggerthella lenta* (Ballard et al., 2005). Furthermore, the transposon has been found integrated into both chromosomal and plasmid DNA but one has still not determined if this transposon family has the ability to move independently and actively between bacteria, or if the propagation is passive (Lambertsen et al., 2018).

1.4 CRISPR-Cas9

The CRISPR-Cas9 system is an adaptive immune system of prokaryotes. CRISPR is an acronym for clustered regulatory interspaced short palindromic repeats – a family of repetitive DNA sequences found in the prokaryote genome. The CRISPR site was first described in *Escherichia coli,* where they observed highly-homologous sequences arranged as repeats with spacers containing approximately 32 nucleotides in between. (Ishino et al., 1987). Sequence analyses of *Yersinia pestis* have reviled that more than two-thirds of the nucleotide spacers found between the homologous sequences were unique in the genome and had a resemblance to bacteriophages and conjugative plasmids (Pourcel et al., 2005). Another important discovery was the genes encoding a protein family that is always found in the proximity to a repeat locus in CRISPR and are called the CRISPR-associated proteins (Cas). The Cas family genes form conserved clusters in prokaryotes, where the different subtypes in the protein family share specific roles in the CRISPR-Cas system (Haft et al., 2005). The proteins encoded have putative nuclease and helix dominance, which in combination with the discovery of CRISPR-loci being transcribed (Tang et al., 2002) substantiated the theory of the CRISPR-Cas system being an adaptive defense system using complementary RNAs to recognize the genomes of past intruders (Haft et al., 2005).

The function of this adaptive immunity mechanism is divided into three main stages; (i) the insertion of the protospacer (a short sequence of foreign intruder DNA) to the CRISPR loci, (ii) transcription of precursor CRISPR-RNA (crRNA) consisting of a complementary to the intruder target sequence, and a repeat fragment, and (iii) crRNA-directed targeted cleavage of the intruding genetic element, executed by Cas proteins at crRNA complementary sites (Doudna & Charpentier, 2014). Depending on the mechanisms used for nucleic acid recognition and cleavage, the CRISPR-Cas systems are divided into types I, II, and III. CRISPR-Cas systems I and III both require a large complex of the Cas proteins, while type II only depends on one large Cas protein (Makarova et al., 2011). This makes the type II systems more suitable for commercial use in comparison to type I and III.

The CRISPR-Cas9 system is a type II system that depends on a protospacer adjacent motif (PAM) and trans-activating crRNA (tracrRNA) for the Cas nuclease to recognize and cut its target. The PAM sequence, often 3-6 base pairs long, is located downstream from the protospacer target DNA (Shah et al., 2013) and interacts with the Cas9 protein with the strand non-complementary to the crRNA (Szczelkun Mark et al., 2014). The tracrRNA is a small trans-

RNA encoded upstream of the cas operon in type II CRISPR-cas systems. It has a partial complementarity to the repeat region of crRNA and is crucial for pre-crRNA processing and maturation. TracrRNA binds to pre-crRNA and with this directs a maturation process by ribonuclease III protein and Cas9. The maturated tracrRNA:crRNA duplex direct the Cas9 endonuclease towards targeted DNA cleavage (Figure 4) (Deltcheva et al., 2011). Cas9 has two known nuclease domains; the HNH and RuvC-like nuclease residues. They have catalytic sites on different DNA strands of the target, and must both be active to achieve a double-stranded break (Sapranauskas et al., 2011).

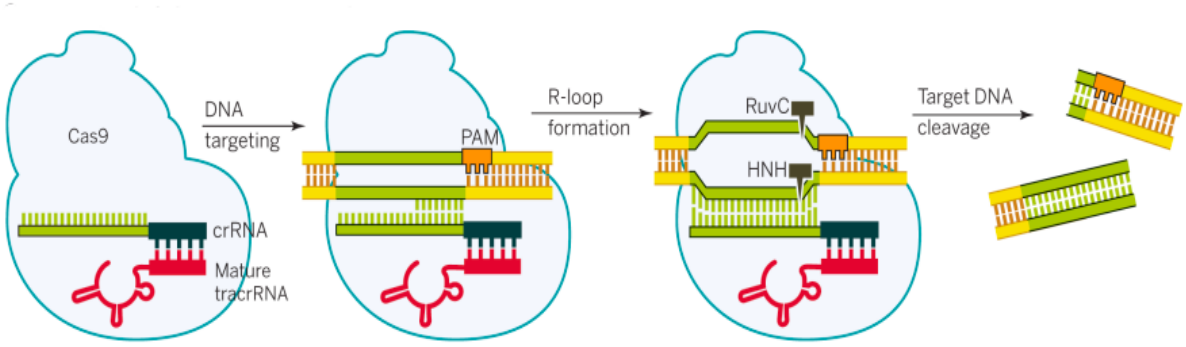


Figure 4: The CRISPR-Cas9 system. The duplex tracrRNA:crRNA guides the Cas9 protein towards the DNA target. The complex is attached to the complementary target sequence upstream of the PAM sequence. PAM sequence interaction with Cas9 leads to R-loop formation (local strand separation) and cleavage by the HNH and RuvC domains. Cas9 complex detaches after DNA cleavage (Doudna & Charpentier, 2014).

Applications of CRISPR-Cas9 have evolved quickly, where modified versions of the system have been used for genome editing both in vivo and in vitro (Latella et al., 2016). Tracr-RNA and custom-designed crRNA, or custom-designed single guide RNA (sgRNA - the engineered version of tracrRNA:crRNA), together with the Cas9 protein have advanced the field of genome engineering. Introducing double-stranded brakes activates the natural DNA repair mechanisms in cells which leads to either nonhomologous end joining repair (NHEJ) or gene replacement by homology directed repair (HDR). NHEJ is prone to errors as it introduces more or less random insertion or deletion to the DNA break, making it a suitable tool for introducing random mutations. HDR on the other hand can introduce precise gene correction or insertion by assisted recombination, making it suitable for site-specific mutation (Doudna & Charpentier, 2014; Maruyama et al., 2015).

1.5 DNA Sequencing

DNA sequencing is the biochemical process in which the order of nucleotides in a DNA molecule is determined. Ever since the double-stranded helix was discovered the methods used to decode genomes have evolved and contributed to the understanding of biological life and diseases (Kchouk et al., 2017). The first generation of sequencing often referred to as "the Sanger sequencing method", can solve one DNA fragment at a time using fluorescent nucleotides to separate bases (Smith et al., 1986). The second generation of sequencing known as massive parallel sequencing or next-generation sequencing (NGS) uses some similar principles as the first generation but has a big advantage when it comes to sequencing volume. NGS is performed with three main steps; library preparation, cluster generation, and DNA sequencing. The hallmark of NGS is the high throughput of short fragmented DNA which is sequenced in parallel using amplification and fluorescence marked nucleotides (Metzker, 2010).

Third-generation sequencing is a long-read technology that can process large DNA inputs without fragmentation. Oxford Nanopore Technologies is one of the leading developers of this single molecular and real-time sequencing technology. The latest versions of the Nanopore chemistry (R9) promise a raw read accuracy of 98.3% with a consensus accuracy of Q50 at 100X. The technology has decreased the library preparation time, has high single-molecule consensus accuracy, and can achieve more than 20Gb of throughput in the smallest of their sequencing instruments (Heikema et al., 2020; Wang et al., 2021). The principle for decoding nucleotides in a sequence is based on changes in current. A Flow Cell or Flongle placed in one of the ONT sequencing machines contains an electro-resistant membrane with hundreds of nanopores. The nanopores are tiny holes connected to both a channel and a chip through electrodes. When sequencing starts, the sensors constantly measure the current flowing through the nanopore and register current disruptions that occur when a molecule passes through. Each nucleotide creates a distinctive current disruption and thus can be identified (Figure 5). In order for a nucleotide sequence to travel through the nanopore, a pure high-quality DNA sample must first undergo an end-fixing step. Here, all DNA ends are enzymatically treated so that the 5´ end is phosphorylated and the 3´end is dA-tailed. An adapter sequence with a DNA helicase "motor" protein is ligated to the phosphorylated 5´ end of each DNA strand and is required for the DNA strand to be sequenced. The adapter-ligated DNA library is loaded to a Flow Cell or Flongle depending on the amount of data one needs to obtain. By applying voltage an ionic current is produced, pulling the DNA strands towards the nanopores. The adapters allow

connection to the nanopores by covalent binding, where the motor protein controls the speed at which the DNA strand flows through the pore (Jain et al., 2016; Wang et al., 2021).

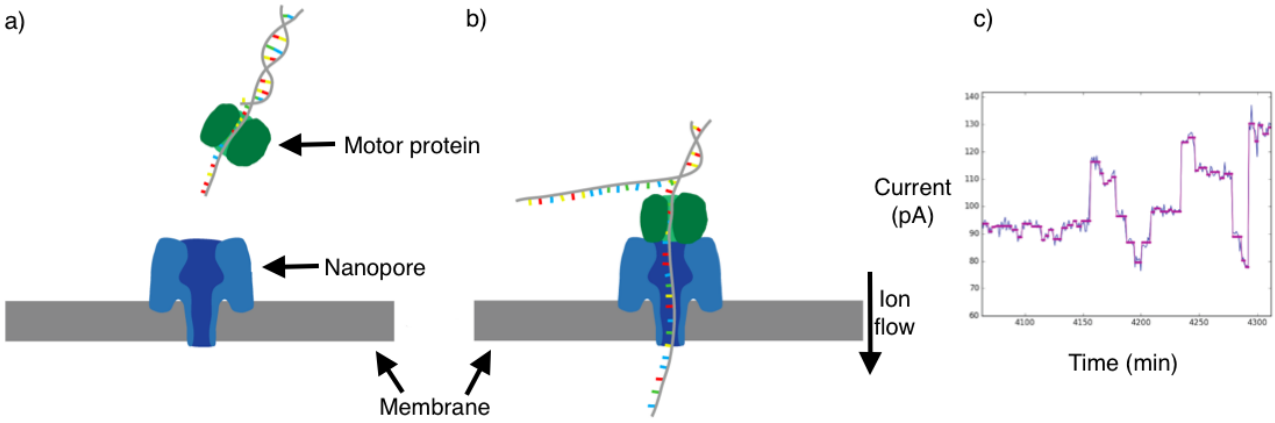


Figure 5: Illustration of sequencing mechanism in an ONT Flow Cell or Flongle. a) A potential across the membrane gives an ion current that leads the double-stranded DNA helix towards the nanopore. b) The motor protein attached to the adapter connects to the nanopore and starts unwinding the strand so it can flow through the pore. c) Each nucleotide molecule creates a disruption in the ion flow that is registered and translated into bases using algorithms (Leggett & Clark, 2017).

1.6 Sequencing with CRISPR-Cas9 enrichment

Nanopore sequencing today allows us to gain high coverage of a genome with long-reads in a relatively rapid analysis workflow. Many of the sequencing methods rely on library amplification by using long-range PCR as an enrichment of the target regions prior to sequencing. (Leija-Salazar et al., 2019). The PCR method depends on at least two known sequences so that both forward and reverse primers can bind to start replication. When the genomic target of interest in a sequencing library is unknown, or only one sequence is known, a PCR enrichment will not be possible. In such cases, the common choice for analysis is whole-genome sequencing which gives a large amount of data that takes time to analyze and is non-selective. A rather new method to ease these areas of research is PCR-free sequencing with CRISPR-Cas9 enrichment which only depends on one known target sequence (Gilpatrick et al., 2020; Watson et al., 2020). This method allows us to target genes and genomes of interest, thus mitigating the post-sequencing data analysis (Figure 6). The CRISPR-Cas9 complex can target any sequence in a genome by selective design of guide-RNAs. The CRISPR-Cas9 enrichment decreases the amount of sequence data that is not of interest and increases output and coverage of the target sequences (López-Girona et al., 2020). This method has mostly been used in tiling

approaches where the Cas9 enzyme targets known sequences on both sides of a region of interest (ROI) followed by sequencing of only the ROI. Another approach to the CRISPR-Cas9 method is to sequence unknown regions of a genome by first targeting a start sequence of interest. A good example is targeting known sequences of a transposon where the carrier host is unknown. By targeting the 5´ and the 3´-end of a transposon it is possible to sequences outwards in the unknown genome. In this way, one can identify both the host and the transposon location in the host genome without analyzing non-specific genomic data, like in whole-genome sequencing (Fiol et al., 2022; Technologies, 2019).

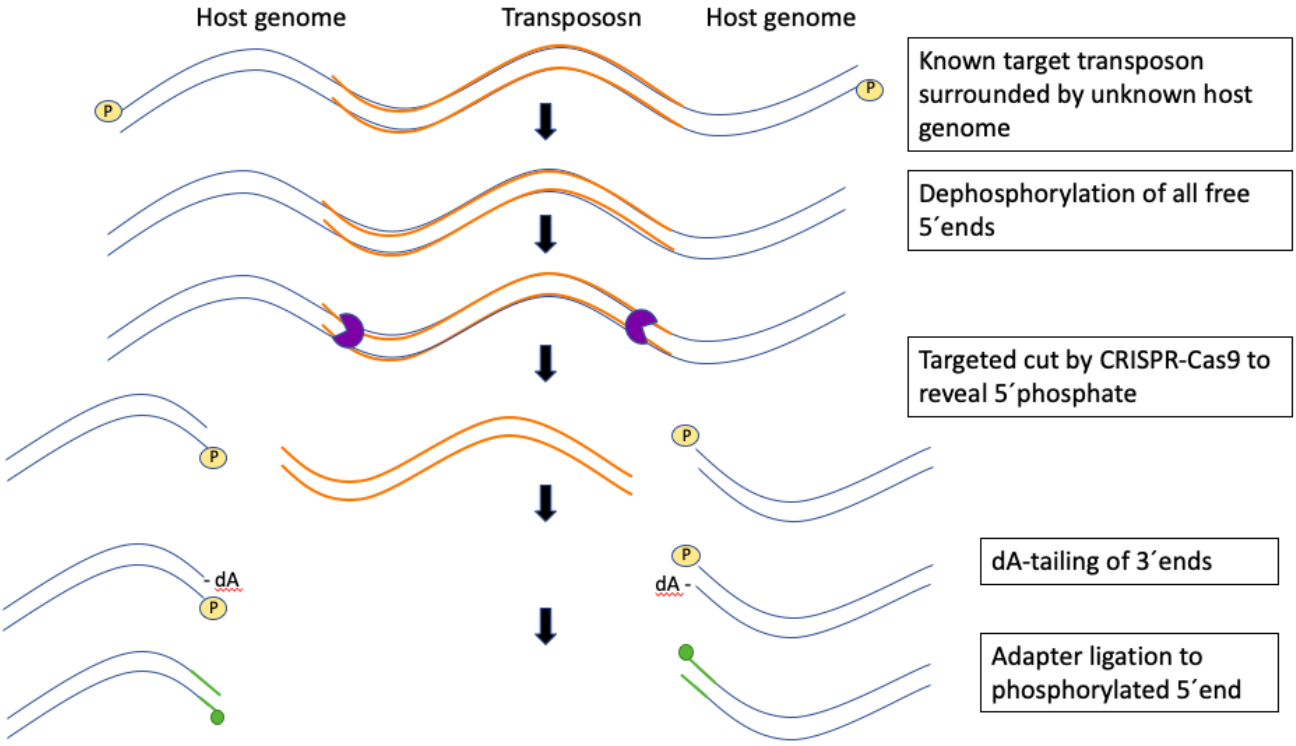


Figure 6: Illustration of CRISPR-Cas9 targeted enrichment prior to sequencing. A host genome with the known transposon is dephosphorylated before being targeted by sgRNA-guided CRISPR-Cas9. The Cas9 enzyme cuts the dsDNA to reveal the 5´phosphate. 3´ends are dA-tailed, and adapters are ligated to the phosphorylated ends to enable Nanopore sequencing.

The principle of this method is to utilize the CRISPR-cas9 system's ability to target specific DNA sequences. As explained in the previous section, the ONT long-read sequencing system relies on adapter ligation to the free phosphorylated 5´ DNA ends in order for the nucleotides to be sequenced. When enriching with CRISPR-cas9 all DNA in the sample is first dephosphorylated, thus prohibiting the adapters to ligate. Using the regions of interest in the genome of your sample, one designs sgRNAs, or crRNAs that make the tracrRNA:crRNA

duplex. The guiding RNA is designed using similar rules as when designing PCR primers, except the target region of the guiding RNA must always contain a PAM sequence in order for the Cas9 enzyme to ligate. The regions cut by the sgRNA-guided CRISPR-Cas9 will in theory be the only free DNA ends phosphorylated, thus the only ends able to ligate to adapters. The method also opens for multiple ROIs being targeted by Cas9 proteins in the same reaction and has no limit regarding the size of the ROI (Gilpatrick et al., 2020; Schultzhaus et al., 2021). In principle, the DNA strands enriched by CRISPR-cas9 should be the only nucleotides sequenced, resulting in less sequencing data with a targeted cover sufficient for characterization.

1.7 Aims of the thesis

Vancomycin-resistant enterococci are an emerging threat to healthcare facilities with outbreaks being reported more and more often. Due to the enterococci's inherited and acquired resistance to multiple other antibiotics, vancomycin is today a cornerstone in the treatment of these pathogens.

One of the most important reasons for the fast spread of *vanB* vancomycin resistance genes is transposable mobile elements. Enterococci have the ability to incorporate and utilize these elements, one of which is the *vanB*-carrying transposon Tn*1549*. It has been suggested that new VRE strains may rapidly arise in the intestines of hospital patients due to uptake of *vanB*-carrying transposons by previously vancomycin susceptible enterococci. Hence, screening patients not only for VRE but also for potential donors of *vanB* containing transposons might be of importance.

Many intestinal microbes with the ability to carry and deliver transposons are difficult to culture because of their strict anaerobe requirements. This complicates the search for *vanB*-carrying intestinal organisms. The combination of long-read sequencing with CRISPR-Cas9 enrichment could represent an attractive approach, with the possibility to opt for genomes carrying Tn*1549*.

This thesis aims to use CRISPR-Cas9 enriched long-read sequencing to find intestinal bacterial species other than enterococci which contain the transposon Tn*1549*. To achieve this, we will:

- Collect fecal samples with a strong positive vanB-PCR but without subsequent growth of vancomycin-resistant enterococci
- Attempt to further enrich the growth of vancomycin-resistant anaerobes prior to DNA extraction
- Establish a selective CRISPR-Cas9 enrichment of Tn*1549*, the *vanB*-carrying transposon involved in VRE-outbreaks at Haukeland university hospital
- Use long-read sequencing to obtain sequence data from the genomic DNA flanking the Tn*1549* transposon and use this information to identify the respective carrier bacteria.

2.0 Materials

2.1 Samples

The samples analyzed in this project were all rectal swab samples from hospitalized patients at HUS. A total of 1,130 samples were screened for the presence of VRE in the period from August 2021 to December 2021. The PCR results were monitored and samples with a high probability of containing high concentrations of non-enterococcal *vanB* positive bacteria were included in the study based on a criterion described in detail in chapter 3.3.

2.2 Listed materials

Table 2.1: Controls and bacterial strain

Name	**Batch#**	**Material**	**Supplier**
vanB-positive control	*E. faecalis* ATCC 51299	Purified DNA from cultivated bacteria, diluted in elution buffer.	MIA, HUS
E. faecium containing Tn1549	KogU III (14.09.21)	Cultivated bacteria, clinical isolate	Department of Microbiology, HUS

Table 2.2: Enzymes

Name	**Catalog number**	**Supplier**
Light Cycler® 480 Probes Master, 2x conc.	04887301001	Roche Diagnostics
TB Green Premix Ex Taq (Tli RNase H Plus)	RR420	Takara Bio Inc.
Streptococcus pyogenes HiFi Cas9 Nuclease V3, 62µM	229673812	Integrated DNA Technologies Inc.
S.pyogenes Cas9 tracrRNA, 100µM	229673811	Integrated DNA Technologies Inc.
NEBNext® Ultra™ II End Repair /dA-Tailing Module	(E7546S) E7646AVIAL	New England BioLabs® Inc.
NEBNext Quick T4 DNA ligase	(E6056S) E6057AVIAL	New England BioLabs® Inc.

Table 2.3: PCR Primers*

Name	**Forward sequence 5´-3´**	**Reverse sequence 5´-3´**	**Supplier**
vanB	5´-ctgcttgtcatgaaagaa- 3´	5´-gggataccagacaataca-3´	TIB Molbiol
Enterococci (*E. faecium* and *E. faecalis*)	Efm: 5´-agctcctgagtcatgc-3´ Efc: 5`-agcacctgagtcatgg-3´	Efmfc: 5´-gttggagtcatcgttttc-3´	TIB Molbiol
Tn*1549*	5´-gtgagaagatggatgctatc-3´	5´-aagagcttagtgcaacttca-3´	TIB Molbiol
Cut-confirmation 3´ side (Reverse)	5´-acccaacctgatgcaatgct-3´	5´-catcctgccggtttccagaa-3´	TIB Molbiol
Cut-confirmation 5´ side (Forward)	5´-ttaatcctgcactgtgttctgg-3´	5´acacagcttgttcatgtaaccg-3´	TIB Molbiol
Enterocloster clostridioformis rpoB	5'-gtaactgtattgattcgcgc-3´	5'-acaaggaatctgccatgaa-3´	TIB Molbiol
Enterocloster clostridioformis nusG	5'-tccctgtggttgagcttaag-3´	5'-cctgtggttgagcttaag-3´	TIB Molbiol
Flavonifractor Plautii rpoB	5´-agctctccttcatcgactact-3	5´-tgaagatctcctgctccttga-3´	TIB Molbiol
Flavonifractor Plautii nusG	5'-acgtgctggtgaagatggt-3´	5-'tgcggtccaggtcgatct-3´	TIB Molbiol
Tn*1549*-end 3´	5´-taccgccaagacgctttc-3´	5´-gtcgatgtaaggcgctatg-3´	TIB Molbiol
Tn*1549*-end 5´	5´-tgcctcttacaccagcaca-3´	5´-tgaacggctttgcacctt-3´	TIB Molbiol

*All primers dissolved to a 10µM concentration

Table 2.4: crRNAs *

Name	**Sequence**	**Supplier**
crRNA TN1549 5´/ revers	5´-TCTGTCACCACAAACGAAGT-3´ (PAM=CGG)	Integrated DNA Technologies Inc.
crRNA TN1549 3´/ forward	5´-AAGGCGTTGATGCCGAAGGG-3´ (PAM=AGG)	Integrated DNA Technologies Inc.

*All crRNAs dissolved to a 100µM concentration

Table 2.5: PCR-probes *

Name	**Sequence 5´- 3´**	**Supplier**
vanB-Probe	TEX - cctgtatcgcaccatcctcc- BHQ2	TIB Molbiol
Enterococci (*E. faecium* and *E. faecalis*)	Efm: Cy5-atggtgcttgaggattgatcaac -BHQ2 Efc: Cy5-acggagagcgtggttgaattaac -BHQ2	TIB Molbiol
Tn*1549* Probe	6FAM - ctcgttacgcagtaccgtcttgcaa - BBQ	TIB Molbiol
Enterocloster clostridioformis rpoB probe	6FAM - ccagcttcggaaaggacacctc - BBQ	TIB Molbiol
Enterocloster clostridioformis nusG probe	6FAM - ccacaaagcccgtgcaccg - BBQ	TIB Molbiol
Flavonifractor plautii rpoB probe	6FAM - ccacgtacgccgcgcccatca - BBQ	TIB Molbiol
Flavonifractor plautii nusG probe	6FAM - cctgatagccgaccaccacctcg - BBQ	TIB Molbiol
Tn*1549*-end 3´	6FAM-ccaagcctgcacctgtttgacctg-BBQ	TIB Molbiol
Tn*1549*-end 5´	6FAM-ccgtcagtccgcacagttccgaa-BBQ	TIB Molbiol

*All probes dissolved to a 10µM concentration

Table 2.6: Chemicals, solutions and buffers

Name	**Compound**	**Supplier**
Ethanol 95% AG	95% ethanol diluted in water	Solveco AB
Vancomycin broth	60mg/L aztreonam, 6mg/L vancomycin, RO-water	Department of Microbiology, HUS

Table 2.7: Commercial reagents

Name	**Catalog number**	**Supplier**
Nuclease free Duplex buffer	11-01-03-01	Integrated DNA Technologies Inc.
MagNA Pure Bacteria Lysis buffer	04659180001	Roche Diagnostics
Agencourt AMPure XP Beads	A63881	Beckman Coulter
NEBNext® Ultra™ II End-prep reaction buffer	(E7546S) E7647AVIAL	New England BioLabs® Inc.

Table 2.8: Consumables

Name	**Catalog number**	**Brand**
0.15mL thin-walled PCR tubes, transparent	781318	BRAND GMBH +CO KG
2mL Microcentrifuge tubes, natural	10025-738	VWR® International
DNA LoBind Tube 1.5mL	022431021	Eppendorf AG
Lysing Matrix E, 2ml Tubes	6914-100	Rotaprep Inc.
Qubit® assay tubes – thin-walled, clear, 0.5mL	Q32856	Invitrogen™ Thermo Fisher Scientific
MagNA Pure 96 Processing Cartridge	06241603001	Roche Diagnostics
MagNA Pure 96 Output Plate	06241611001	Roche Diagnostics
MagNA Pure Tip 1000µL	06241620001	Roche Diagnostics
LightCycler® 480 Multiwell Plate 96, white	04729692001	Roche Diagnostics
Flow Cell	Flow-MIN112 R9 version	Oxford Nanopore Technologies
Flongle Flow Cell	FLO-FLG001 R9.4.1	Oxford Nanopore Technologies
SmartCycler Reaction Tube 25µL	900-0003	Cepheid

Table 2.9: Commercial kits and reagents

Name	**Catalogue number**	**Supplier**
Cas9 Sequencing Kit	SQK-CS9109	Oxford Nanopore Technologies
Flow Cell Priming Kit	EXP-FLP002	Oxford Nanopore Technologies
Ligation Sequencing Kit	SQK-LSK109	Oxford Nanopore Technologies
MagNA Pure Compact Nucleic Acid Isolation Kit	03730964001	Roche Diagnostics
MagNA Pure 96 DNA and Viral NA Large Volume Kit	06374891001	Roche Diagnostics
Invitrogen™ Qubit™ dsDNA HS and BR Assay Kit	10616763	Invitrogen™ Thermo Fisher Scientific

Table 2.10: Consumables

Name	**Catalog number**	**Brand**
0.15mL thin-walled PCR tubes, transparent	781318	BRAND GMBH +CO KG
2mL Microcentrifuge tubes, natural	10025-738	VWR® International
DNA LoBind Tube 1.5mL	022431021	Eppendorf AG
Lysing Matrix E, 2ml Tubes	6914-100	Rotaprep Inc.
Qubit® assay tubes – thin-walled, clear, 0.5mL	Q32856	Invitrogen™ Thermo Fisher Scientific
MagNA Pure 96 Processing Cartridge	06241603001	Roche Diagnostics
MagNA Pure 96 Output Plate	06241611001	Roche Diagnostics
MagNA Pure Tip 1000µL	06241620001	Roche Diagnostics
LightCycler® 480 Multiwell Plate 96, white	04729692001	Roche Diagnostics
Flow Cell	Flow-MIN112 R9 version	Oxford Nanopore Technologies
Flongle Flow Cell	FLO-FLG001 R9.4.1	Oxford Nanopore Technologies
SmartCycler Reaction Tube 25µL	900-0003	Cepheid

Table 2.11: Instruments and equipment

Name	**Manufacturer**
Anoxomat® Anaerobic Cultivation System	Advanced Instruments
Anoxomat® Ergonomic Anaerobic Culture Jar AJ9049	Advanced Instruments
Hettich MIKRO 185 Centrifuge (1203)	Hettich
Micocentrifuge, MiniStar (521-2161)	VWR® International
Spectrafuge™ Mini centrifuge (C1301)	Labnet International
Bio-Express Mini centrifuge Cepheid 700-2378 (Labnet# C1301-CEP-230V)	Labnet International/Cepheid
Reax control Vortex Mixer	Heidolph Instruments
MagJET Separation Rack, 2x1.5mL Tube (MR01)	Thermo Fisher Scientific
MagNA Lyser Instrument	Roche Diagnostics
MagNA Pure Compact – Nucleic acid Purification Instrument	Roche Diagnostics
MagNA Pure 96 – Nucleic acid Purification Instrument	Roche Diagnostics
LightCycler® 480 Real-Time PCR System	Roche Diagnostics
Smart Cycler® Real-Time PCR System	Cepheid
ProFlex 3x32 well PCR System	ThermoFisher Scientific Inc.
GridION X5 Long-Read Sequencer	Oxford Nanopore Technologies
Qubit® 3.0 Fluorometer	ThermoFisher Scientific Inc.

Table 2.12: Software

Name	**Brand**
MinKnow v21.22.6	Oxford Nanopore Technologies
Geneious R9.1	Biomatters Inc.
Nucleotide BLAST® (The Basic Local Alignment Search Tool)	National Center for Biotechnology Information (NCBI) U.S. Library of Medicine
Epi2ME v3.4.2	Oxford Nanopore Technologies
CHOPCHOP v3	CHOPCHOP web tool
Custom Alt-R® CRISPR-Cas9 guide RNA Tool	Integrated DNA Technologies Inc.

3.0 Methods

3.1 Establishing the CRISPR-Cas9 selective sequencing method

A clinical isolate of *Enterococcus faecium* isolate carrying the transposon Tn*1549* was used to establish and optimize the CRISPR-Cas9 based method for targeted enrichment prior to sequencing. The workflow of the targeted CRISPR-Cas9 sequencing is illustrated in Figure 7.

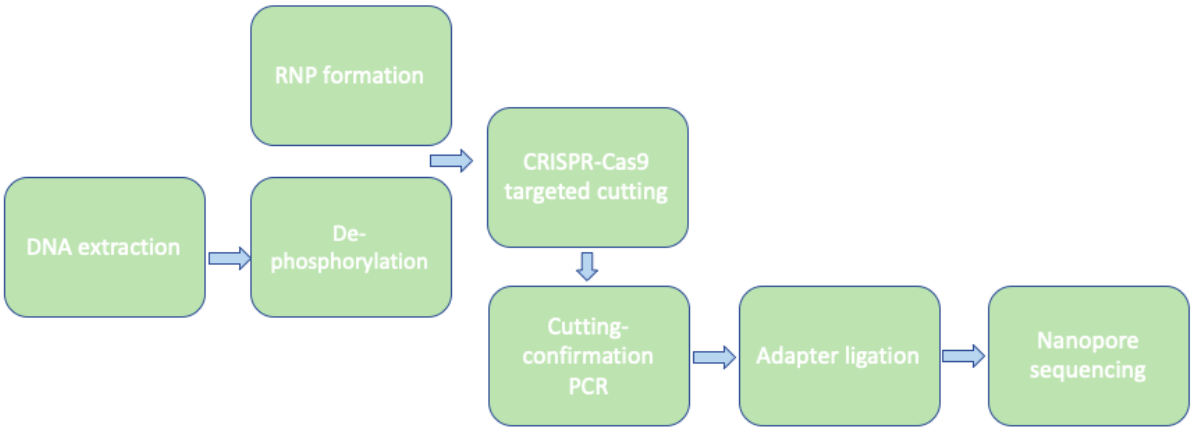


Figure 7: Sequencing with CRISPR-Cas9 workflow. Extracted DNA is dephosphorylated to prevent the binding of Nanopore sequencing adapters. Cas9 ribonucleoprotein complexes (RNPs) are made by annealing tracrRNA:crRNA duplex with the Cas9 enzyme. Targeted cutting is performed on the dephosphorylated DNA, and the cutting is confirmed with a PCR. The targeted cutting will result in novel phosphorylated DNA fragments that can be ligated to the sequencing adapters and sequenced.

3.1.1 DNA extraction from *E. faecium* culture

A swoop of colonies was dissolved in 500µL of nuclease-free water and 500µL of MagNA Pure Bacterial Lysis buffer. The suspension was transferred to a 2mL tube with Matrix E lysing beads. The tube was placed in a MagNA Lyser instrument for cell disruption and homogenization and run for 90 seconds at 1800G. The matrix and cell debris were spun down using a Hettich centrifuge for 3 minutes at 215G. The DNA in the supernatant was purified using a MagNA Pure Compact automated extractor according to the manufacturer's manual using the "MagNA Pure Compact Nucleic Isolation Kit I", with the "Bacterial_v3" protocol in the MAgNA Pure Compact software.

3.1.2 DNA quantification

To quantify DNA concentration the Qubit® 3.0 Fluorometer and the Qubit® dsDNA HS and BR Assay Kit were used. All samples and standards were prepared in Qubit® assay tubes and according to the manufacturer's manual. Two standard solutions (I and II from Qubit® Kit) were used to calibrate the instrument. In the Qubit® 3.0 software one chose DNA>dsDNA>High Sensitivity>Read standards, followed by calibrating the instrument by first reading standard I and then standard II. Before reading the samples, the sample volume used was registered and concentration units were set to ng/µL.

3.1.3 Selection of CRISPR-Cas9 cutting sites

The targeted CRISPR-Cas9 cutting was performed using the Oxford Nanopore Technology (ONT) CRISPR-Cas9 sequencing Kit. Reaction reagents used in chapters 3.1.3 – 3.1.8 are part of this kit unless stated otherwise. The Cas9 enzyme, the tracrRNA, the crRNAs, and the Nuclease-free Duplex buffer were acquired from Integrated DNA Technologies (IDT). The crRNAs were designed using ChopChop and IDTs own software tool for designing custom crRNAs. Since the interesting sequences were the chromosomal DNA flanking the transposon, suitable sequences were searched for on the outer edges of the transposon Tn*1549* with consideration to possible primer dimers, hairpin formation, GC percentage, and possible off-target ligations to both prokaryotic and human DNA. Two crRNAs were selected; one on the 5´-end of the transposon approximately 2750bp from the transposon start (reverse target), and one on the 3´-end of the transposon approximately 2300bp from the end of the transposon (forward target).

3.1.4 Targeted CRISPR-Cas9 cutting

The first step was to prepare the Cas9 ribonucleoprotein complexes (RNPs). Two sets of RNPs were prepared; one with crRNA targeting the 5´-end of the transposon, and one targeting the 3´-end. In a 0.15mL tube, 8µL of duplex buffer was mixed with 1µL tracrRNA and 1µl of crRNA to a total volume of 10µL. The tube was spun down using Spectrafuge™ Mini centrifuge and incubated at 95° C for five minutes using a Pro-Flex thermal cycler instrument to anneal the tracrRNA:crRNA duplex. After annealing, the duplex aliquot was cooled to room temperature and spun down before being transferred to a 1.5mL tube. In the 1.5mL tube, the duplex was mixed with 79.2µL nuclease-free water, 10µL reaction buffer, and 0.8µL of cas9 enzyme to a total volume of 100µL. The tube was incubated for 30 minutes at room temperature

to anneal the Cas9 enzymes and the duplexes. The 5´-targeting and the 3´-targeting RNPs were first tested as separate reactions and in separate sequencing analyses. For the rest of the experiments, these were mixed so that targeted cutting on both ends of the transposon occurred in the same reaction. This was done by mixing 5µL of each crRNA in a 0.15mL tube, where 1µL of the mixed crRNAs were added to the tracrRNA:crRNA duplex annealing. The mixed duplexes were then used for RNP formation, making the tube contain RNPs targeting both the 5´-end and the 3´-end of the transposon.

The adapters needed for nanopore sequencing attach to the free phosphorylated 5´-end of a DNA strand. Therefore, all free DNA ends were dephosphorylated to achieve only targeted adapter ligation. When using the ONT Cas9 sequencing kit it is recommended by ONT to use 1-10µg of gDNA as input. 24µL of gDNA from *E. faecium* (0.342µg for the 5´targeting reaction and 0.875µg for the 3´targeting reaction and the mixed reaction) was added to 0.15mL tubes together with 3µL reaction buffer and 3µL of phosphatase with a total volume of 30µL. The tubes were gently flicked and spun down using a Spectrafuge™ Mini centrifuge. The samples were then inserted into the Pro-Flex thermal cycler and incubated at 37° C (phosphatase working temperature) for 10 minutes, then at 80° C (phosphatase inactivation) for 2 minutes, before cooled to room temperature.

The final step was targeted cutting by the RNPs at the outer edges of the transposon where the crRNAs were designed to ligate. In theory, the only free phosphorylated DNA ends should now be the cleaved target cut-sites in the transposon, hence the only strands available for adapter ligation. To the 0.15mL tubes with dephosphorylated gDNA, there was added 10µL of RNPs, 1µL of dATPs, and 1µL of Taq Polymerase, giving a total volume of 42µL. The samples were mixed by gentle inversion, spun down, and inserted into the Pro-Flex thermal cycler. Here, it was incubated first at 37° C (Cas9 working temperature) for 15 minutes and then at 72° C for 5 minutes to inactivate the enzyme. From now on the samples were kept at 4° C. To check the effectiveness of the CRISPR-Cas9 cutting the samples were analyzed and compared with un-cut *E. faecium* DNA using the Cas9 cutting confirmation PCR described in chapter 3.1.5.

3.1.5 PCR-based confirmation of successfully targeted Cas9 cutting

As sequencing equipment is costly, we designed PCRs flanking each of the Cas9 cut sites on the transposon to confirm successful Cas9 cutting prior to sequencing (Figure 8). Successful Cas9 cutting would split the PCR target area and thereby abort PCR amplification. By comparing the PCR Ct values of a sample before and after cutting, it was possible to obtain a semi-quantitative measure of how successful the targeted cutting had been. As there were designed two crRNAs, (see section 3.1.3.) one also designed two sets of primers, one for each target-cut-site. The PCR reaction mix for both sets of primers and probes was prepared according to Table 3.1.

Table 3.1: PCR reaction mix

Component	Amount (µL)
Forward/reverse primer *	2 (1+1)
Nuclease free water	8.5
TB Green RNase	12.5
DNA template	2
Total	25

*Primers and probes dissolved to 10µM

Table 3.2: PCR program

Step	Temperature	Time(min)	Cycles
Activation	95	30	1
Denaturation	95	10	35
Annealing	58	15	
Extension	72	10	

As the difference in Ct value was of interest, a 10-fold dilution series of the *E. faecium* gDNA was made to characterize the PCR efficiency. The dilution series, along with an undiluted sample, were analyzed in triplicates using the Cepheid Smart Cycler® real-time PCR instrument. The reaction tubes were spun down using a Cepheid mini centrifuge, and run with the PCR program shown in Table 3.2. The average Ct values of each triplicate were used to make a linear trendline, where the slope of each line was used to calculate the efficiency number (Eq. 1) and percent efficiency (Eq. 2) of the PCR.

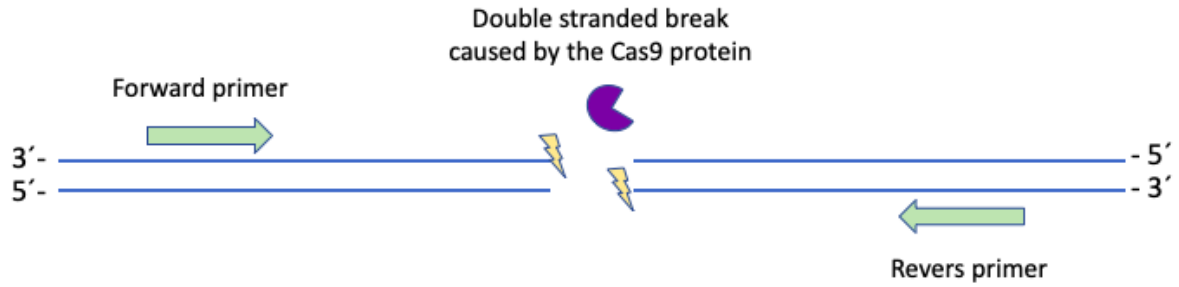


Figure 8: Illustration of primers designed for the cutting confirmation PCR. The primers are targeting each side of the Cas9 cut site, resulting in an incomplete strand extension and no amplification of a new double-stranded template.

Using equations 1 and 2 the PCR efficiency was calculated, where E is the efficiency and the slope is found by plotting Ct values of a dilution series towards the logarithmic copy number of an exponential PCR (Svec et al., 2015).

E=10^(-1/slope) (Eq. 1)

E%=(10^(-1/slope)-1)*100 (Eq. 2)

3.1.6 Adapter ligation

The sample libraries with a total volume of 42µL (from chapter 3.1.4) were transferred to 1.5mL DNA lo bind tubes. In different 1.5mL tubes, an adapter ligation mix was prepared by mixing 20µL Ligation buffer, 3µL nuclease-free water, 10µL T4 DNA ligase, and 5µL adapter, to a total volume of 38µL. The adapter ligation mix was added to the DNA libraries in two parts to avoid protein precipitate and mixed by gentle inversion. The adapter ligation was performed at room temperature with an incubation time of 10 minutes.

3.1.7 Removal of excess adapters

Excess adapters wear out the lifetime of each nanopore during sequencing, therefore it is necessary to remove them from the library. This is done by AMPure XP bead purification, where the beads are positively charged and thus bind to the negatively charged DNA. The 80µL adapter-ligated DNA libraries were first added one volume (80µL) of SPRI dilution buffer giving a total volume of 160µL. Then, a 0.3x volume (48µL) of AMPure XP beads was added to the samples. The beads were properly suspended in the DNA library and kept suspended by gentle inversion during a 10-minute incubation time at room temperature. After incubation, the samples were spun down and put in a MagJET Separation Rack to collect the beads. When all

beads seamed collected to a pellet, the supernatant was removed. The beads were resuspended in 250µL of Long Fragment buffer before returning to the magnet rack to allow the beads to pellet once more. The supernatant was pipetted off and discarded before the washing procedure with the Long Fragment buffer was repeated. The beads were then removed from the magnetic rack and allowed to dry until they were no longer shiny. The beads were resuspended in 13µL of Elution Buffer allowing the DNA to release from the beads. The libraries were incubated at room temperature for 30 minutes, followed by pipetting off the elute, transferring it into new tubes, and discarding the beads.

3.1.8 Nanopore sequencing

The DNA libraries were sequenced using the CRISPR-Cas9 sequencing kit, the Flow cell priming kit, and Flongle Flow cells on the GridION sequencing instrument. The Flongle flow cells were inserted into the Flongle adapter, and the Flongle adapters were inserted into GridION. The number of available nanopores was checked prior to sequencing to ensure sufficient sequencing capacity. Nanopore Technologies recommends a minimum number of 50 active nanopores in the Flongle flow cell, and we decided that all of our experiments should be run with a minimum of 60 active nanopores. The Flongles were prepared by priming the loading port. 117µL of Flush buffer mixed with 3µL of flush tether (both from the Flow cell priming kit) was mixed before being carefully inserted through the port without making air bubbles. Next, the library was prepared to be loaded into the Flongle. In 1.5uL tubes 15µL of sequencing buffer II, 10µL of loading beads, and 5µL of the DNA library were mixed. The library mix was loaded into the loading port without introducing air bubbles. The sequencing was set to run for 48 hours with 180V, with the data output selected to be fastq and fast5 files using Guppy v5.1.12 and the 21.11.6 version of MinKNOW. The program was set to filter out sequences with a mean quality score lower than 7. This allowed for the Guppy software to sort out bases and sequences with low quality and probability of being incorrectly basecalled by the sequencer.

3.1.9 Sequence analysis

The fastq files produced by the GridION sequencer were transferred to the Geneious R9.1 software to study the output sequences. Initially, all sequences were attempted aligned to the outer parts of transposon Tn*1549* that were expected to be found at the beginning of each sequence. The sequencing output from the DNA cut by the 5´-end targeting Cas9 enzymes was aligned with the 2750bp at the start of the transposon and the sequencing output from the DNA

cut by the 3´-end targeting Cas9 enzymes was aligned with the 2300bp at the end of the transposon. From this, it was possible to calculate how many percent of the reads from each output were actually targeted reads. As sequence alignment in Geneious is rather slow, we created four shorter motifs for each of the outer parts of the transposon to make target searches a faster process.

3.2 Method improvements

Due to a disappointingly low yield of confirmed targeted sequences, we tried to increase the output by modification of several of the different steps in the CRISPR-Cas9 enrichment method. A Flongle Flow cell has a lower sequencing capacity than a standard Flow cell. To clarify if a standard Flow cell would give a better yield, one of the two *vanB* positive samples sequenced on a Flongle flow cell (chapter 3.5) was also sequenced using a Flow cell. The sequencing was performed as described in section 3.1.8, where the Flow cell check showed that the Flow cell had more than 1000 nanopores available for sequencing.

We also tested if a longer incubation time with the Cas9 enzyme would increase the target yield. Initially, the activation time was set to 15 minutes as recommended by the manufacturer, but activation for 20, 25, and 30 minutes was also tested and compared using the cutting-confirmation PCR as described above.

To confirm that the target sequences were not lost during any of the sample preparation steps, a PCR was made to check and compare the levels of target DNA before and after the AMPure bead purification of the library. For this we designed primers targeting the very ends of the transposon Tn*1549*, the ends expected to be part of the DNA strands to be sequenced. The eluate ready to be loaded into the Flongle flow cell, which had gone through multiple washing steps with the AMPure XP beads, was analyzed and compared to the Ct value of the DNA input to the dephosphorylation step. The aliquots were analyzed using the Cepheid Smart Cycler® real-time PCR instrument with the PCR program described in Table 3.3, where the PCR amplification mix was made according to Table 3.4 (chapter 3.3).

3.3 Sample collection

We included fecal swabs (Eswab, Copan, Murrieta, CA, USA) from the routine VRE-surveillance program at Haukeland University Hospital. As part of the routine sample treatment, the fecal swabs were transferred to a vancomycin-containing broth and incubated at 35° C overnight for enrichment of vancomycin-resistant bacteria (Figure 9). After incubation, DNA was extracted and purified on a MagNA Pure96 automated extraction instrument using the "MagNA Pure DNA and Viral NA Large volume kit" with an elution volume of 50uL. The purified DNA was screened for the presence of the *vanB* gene using a vanB-PCR. The PCR reaction mix was prepared according to Table 3.4, where a total volume of 20µl was added to a Light Cycler® Multiwell 96 PCR plate with one well for each sample. Using the Light Cycler ® 480 PCR machine the samples were analyzed with real-time PCR with the PCR program shown in Table 3.3.

Table 3.3: PCR program

Step	Temperature °C	Time (min)	Cycles
Enzyme activation	95	10	1
Denaturation	95		45
Annealing	54		
Extension	72		

Table 3.4: PCR reaction mix

Component	Amount (µL)
Primer and probe*	2.5 (1+1+0.5)
Water	2.5
Probes Master	10
DNA template	5
Total	20

*Primers and probes dissolved to 10µM

VanB-gene positive samples with a sigmoid curve and cycle threshold (Ct) value lower than 31 were further cultivated on CHROMagar VRE dishes for 24-48 hours to eventually confirm the presence of VRE.

As part of the study, *vanB*-gene positive samples without growth of enterococci were further analyzed using an enterococci-PCR targeting the *rpoB* gene of both *E. faecalis* and *E. faecium* to ensure that *vanB*-gene positivity was not due to dead VRE. The enterococci-PCR reaction mix was prepared according to Table 3.4, and the samples were analyzed using the PCR multiwell plate, Light Cycler® 480 Instrument, and the PCR program shown in Table 3.3. *VanB*-gene positive samples with a non-corresponding result from both CHROMagar

cultivation and the enterococci-PCR were included as part of the project. Eleven samples fulfilled these criteria and were included in the study.

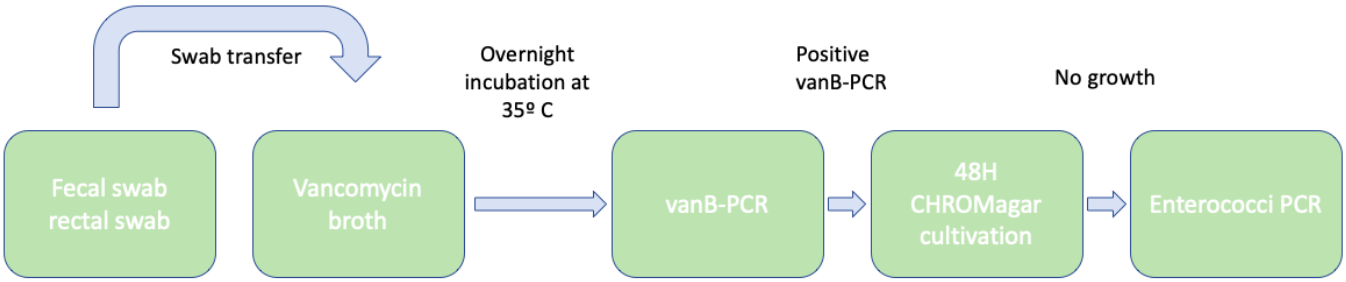


Figure 9: Routine sample workflow in the lab. The fecal swab is transferred to vancomycin broth which is incubated overnight. DNA from the sample in vancomycin broth is extracted and analyzed by vanB-PCR. The sample is cultivated in CHROMagar if the vanB-PCR result is positive. If there is no growth after 48 hours of cultivation the DNA is analyzed again using enterococci-PCR.

3.4 Enhanced culture-based enrichment of vancomycin-resistant bacteria

For the eleven samples, both the primary sample and the aliquot that had already been incubated overnight in a vancomycin broth were used in an additional attempt to better up-concentrate vancomycin-resistant bacteria using anaerobic incubation. One hundred µL from each tube was transferred to a new vancomycin-broth tube, giving two secondary tubes of vancomycin broth, one with a diluted fecal swab sample, and one with a diluted vancomycin broth sample (Figure 10). The secondary tubes with diluted samples were placed in an Anoxomat® anaerobic jar where environmental conditions were adjusted to anaerobic (85% N_2, 10% CO_2, 5% H_2) using the Anoxomat® anaerobic cultivation system. The jar was then incubated at 35° C for 48 hours. After 48 hours bacterial DNA from both the primary tubes and the incubated secondary tubes was extracted using a MagNA Pure 96 instrument as described in section 3.3. The extracted DNA was analyzed by the vanB-PCR to determine if the anaerobe enrichment procedure had increased the growth of vancomycin-resistant bacteria. A PCR specific to Tn*1549* was also performed to compare the Ct values for the transposon to the Ct values for the *vanB* gene. The PCR reaction mix for Tn*1549* was prepared as shown in Table 3.4. The PCR was performed using Light Cycler® 480 Instrument, and the PCR program used is shown in Table 3.3. Samples with strong positive *vanB*-PCR with Ct values lower than 20, and with a corresponding positive Tn*1549* PCR, were chosen for further analysis. Two samples fulfilled these criteria and were further included in the study using targeted sequencing.

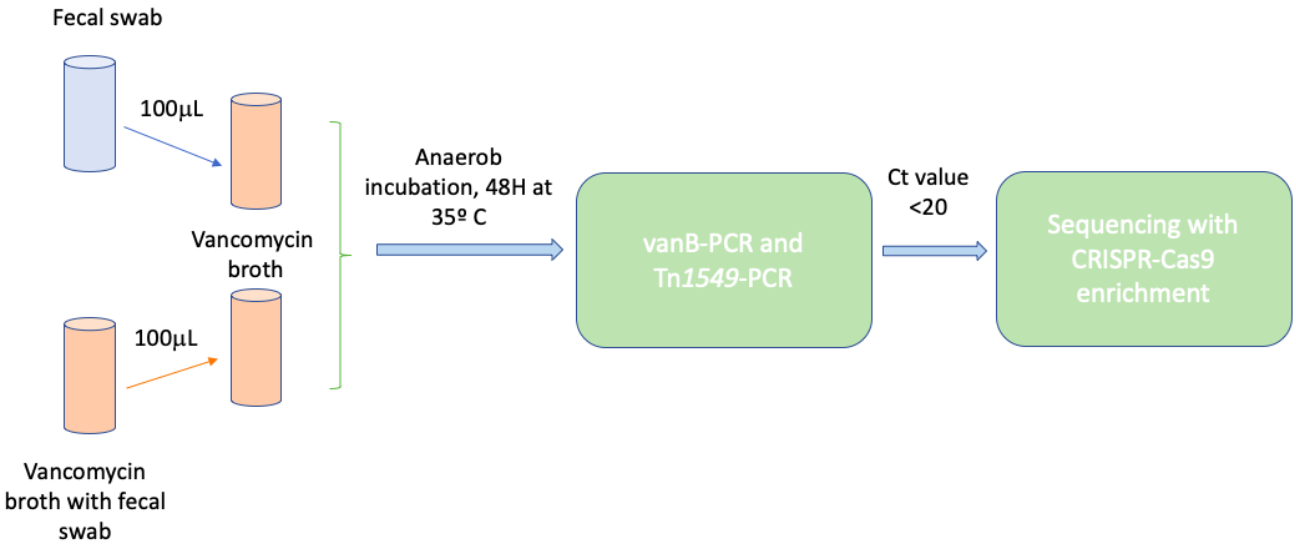


Figure 10: Enhanced sample enrichment workflow. 100µl of sample material from primary fecal swab and vancomycin broth tubes are transferred to new respective vancomycin broth tubes. These tubes are incubated at anaerobic conditions for 48 hours at 35° C. After incubation, DNA is extracted from the samples and then analyzed using both vanB- and Tn1549-PCR. Samples with corresponding Ct values between vanB and Tn1549, and with a Ct value lower than 20, are used to further analysis.

3.5 Sequencing of vanB-PCR strong positive samples with CRISPR-Cas9 targeted enrichment

From the eleven samples chosen for sample enrichment, the two samples with the strongest positivity in the vanB-PCR were selected for CRISPR-Cas9 enrichment and sequencing. Prior to CRISPR-Cas9 enrichment, sample-DNA was quantified using a Qubit 3.0 Fluorometer as described in chapter 3.1.2. The samples were then prepared by CRISPR-Cas9 targeted enrichment as described in chapter 3.1.4, where the crRNAs of the 5´-transposon-cutting and the 3´-transposon-cutting were pooled so that the two targeted cutting occurred in the same reaction. The effectiveness of the cutting was assessed using the cutting-confirmation PCR as described in section 3.1.5. After cutting, adapter ligation, excess adapter disposal, and priming of the Flongle Flow cell loading port (3.1.4. – 3.1.7.), the samples were loaded to separate Flongle Flow cells and sequenced as described in section 3.1.8. The motifs designed in chapter 3.1.9. were used to find the on-target reads after transferring the sequencing output from GridION to Geneious. Using MUSCLE alignment in Geneious, the reads were attempted aligned towards the 2750bp/2300bp at the transposon start/end to find the targeted sequences. Epi2ME was also used – a tool by Nanopore Technology that uses already known bioinformatics tools like Minimap2, Medaka, and Flye to e.g. align sequences or perform De

novo assemble. The first 2750bp/2300bp of the sequence reads should be the outer parts of the transposon, and the sequence flanking this region should be the genome of the host bacteria. From the reads aligning to the outer part of the transposon, the regions flanking the transposon were analyzed using nucleotide BLAST (NCBI, GenBank) to identify the transposon host bacteria using both the nr/nt and wgs databases.

3.6 Metagenomic sequencing

In one of the sequenced samples, the host genome containing the transposon Tn*1549* did not obtain sufficient highly percent identity to conclude on the bacterial species. Therefore, it was decided to perform a long-read metagenomic sequencing analysis of the original sample in an attempt to collect enough sequences for proper identification. Initially, all DNA ends were prepared for adapter ligation. In a 0.15mL tube, 47µL (0.1494µg) of gDNA was mixed with 6.5µL nuclease-free water, 3.5µL Ultra II End-prep reaction buffer, and 3µL of Ultra II End-prep enzyme mix to a total volume of 60µL. The components were mixed by flicking the tube, spun down, and incubated first at 20° C for 5 minutes, then at 65° C for 5 minutes using a Pro-Flex thermal cycler. The sample was transferred to a 1.5mL lo-bind tube and added one volume (60µL) of AMPure XP beads. The beads were kept suspended while incubated for 5 minutes at room temperature to clean up the end-prep enzymes. After incubation, the sample and beads were spun down and put in a MagJET Separation Rack to pellet. While keeping the tube on the magnet rack, the supernatant was pipetted off and discarded before the beads were washed with 200µL of 70% ethanol without disturbing the pellet. The ethanol was pipetted off and the washing was repeated. After the second wash, the beads were removed from the magnet rack and allowed to dry until they were no longer shiny. The beads were then added 61µL of nuclease-free water and incubated for 2 minutes to make the DNA let go of the beads. The tube was put back in the magnet rack to pellet, the sample was pipetted off and the beads discarded.

Using a Ligation kit, adapters were ligated to the end-prepped sample. In a 1.5mL Lo-Bind tube 60µL of the end-prepped DNA was mixed with 25µL Ligation buffer, 5µL Adapter mix, and 10µL NEBNext Quick T4 DNA ligase to a total volume of 100µL. The reagents were incubated at room temperature for 10 minutes. After incubation, an adapter clean-up was performed. The sample was added 40µL AMPure XP beads and incubated for 5 minutes at room temperature while keeping the beads suspended. The sample was then washed and eluted as described in section 3.1.7.

The sample was sequenced using a Ligation kit, a Flow cell, and a Flow cell priming kit. The Flow cell was inserted into the GridION sequencing instrument where the condition of the nanopores was checked and verified to have more than 1000 active nanopores. The Flow cell was then prepared according to the manufacturer's manual "Priming and loading the Flow cell" (ONT). The sequencing was set to run under the same conditions as described in section 3.1.8. The fastq files with sequencing data were analyzed in Epi2ME using the Fastq "What´s in my pot" (WIMP) classification software v3.3.2 to identify the sequencing reads. Further sequencing analysis was performed by a bioinformatician, who performed a De Novo assembly and contig polishing using metaFlye assembler v2.9, and contig classification using Kraken2, and Pavian R package v1.2.0. The program Bandage v0.9.0 was used to visualize the de novo assembled graphs.

3.7 Species confirmation by PCR

Two bacterial species, *Enterocloster Clostridioformis* and *Flavonifractor Plautii*, were found harboring the transposon Tn*1549* in the targeted sequencing analysis. To have a second confirmation of these species, PCRs targeting the *rpoB* and *nusG* genes in both bacteria were made. The PCR reaction mixes were made according to table 3.4 in chapter 3.3, and the PCR program used is described in Table 3.3 in the same chapter, except the annealing temperature used here was 56° C. The samples were analyzed using the Smart Cycler ® instrument. The Ct values from the species-specific PCRs were compared to the Ct values of the vanB-PCR.

3.8 Statistics

Microsoft excel was used to determine if a result was considered statistically significant, where this was assessed using an unpaired two-tailed t-test with Welch's correction. We hypothesized a mean difference of zero, where p values ≤ 0.05 was considered statistically significant.

4.0 Results

4.1 Methods evaluation

4.1.1 Establishing the CRISPR-Cas9 targeted sequencing

To establish and evaluate the CRISPR-Cas9 method, we used DNA from a cultured isolate of *E. faecium* containing the transposon Tn*1549*. After quantification and dephosphorylation, the DNA was treated with CRISPR-Cas9. Following confirmation of effective cutting as described in chapter 3.1.5, the samples were prepared for sequencing. The sequencing data was transferred to Geneious R9 for visualization and analysis of the sequencing output. The sequencing output from the sample treated with the 5´-end targeting Cas9 enzyme was 344 reads after basecalling and filtering of low-quality reads by Guppy, with sequence lengths varying from 207 to 20,113 nucleotides. Out of the 344 reads, we found by multiple alignments that 32,5% were target sequences cut by the 5´-side targeting Cas9 protein. The on-target sequences had varying lengths from 200 to 20,000 nucleotides. The remaining non-target sequences were from random locations in the *E. faecium* genome. The sequencing output from the sample treated with the 3´-end targeting Cas9 enzyme was 1182 reads after filtering by Guppy, with sequence lengths varying from 161 to 39,000 nucleotides. From the 1182 reads, we found by multiple alignments that 38.24% were target sequences cut by the 3´-end targeting Cas9. The on-target sequence lengths varied from 400 to 19,000 nucleotides. Like in the 5´-end targeted sequencing, the remaining reads were sequences from various locations of the *E. faecium* genome not containing any sequences from transposon Tn*1549*. The genomic input to the 3´-end targeting reaction was 2.55-fold larger than the input to the 5´-end targeting reaction. The 3´-end targeting reaction also had a 3.4-fold higher number of reads from the sequencing output than the 5´-end targeting reaction.

After testing the two CRISPR-Cas9 enzymes in separate reactions we wanted to see if they could be combined into a single reaction. This way, the library preparation would be more efficient, and the sample would contain sequences to be read outwards in both directions from the transposon using a single Flongle flow cell. An input of 0.875µg of *E. faecium* DNA was dephosphorylated and treated with a mix of CRISPR-Cas9 enzymes targeting both the start and the end of the transposon Tn*1549*. The Cutting confirmation PCR then confirmed that the cutting in both ends of the transposon had been successful before further sample preparation

and Flongle flow cell loading. After filtering by Guppy, we remained with an output of 1331 sequences with lengths varying from 200 to 23,000 nucleotides. Among these 63.1% represented targeted reads (38.7% from the 5´-end cutting and 24.4% from the 3´-end cutting), indicating that performing the two targeted cuttings in the same reaction might actually increase the target yield in comparison to performing these reactions separately.

4.1.2 Ca9-cutting confirmation PCR

A perfect PCR has an efficiency number of 2, and a percent efficiency of 100%. The PCR targeting the 5´cut site on the transposon obtained an efficiency number of 1.97, corresponding to an efficiency of 97%. The PCR targeting the 3´cut site of the transposon had an efficiency number of 1.95, giving an efficiency of 95%. The average difference between the Ct values of each 10-fold dilution was 3.4 for both PCRs.

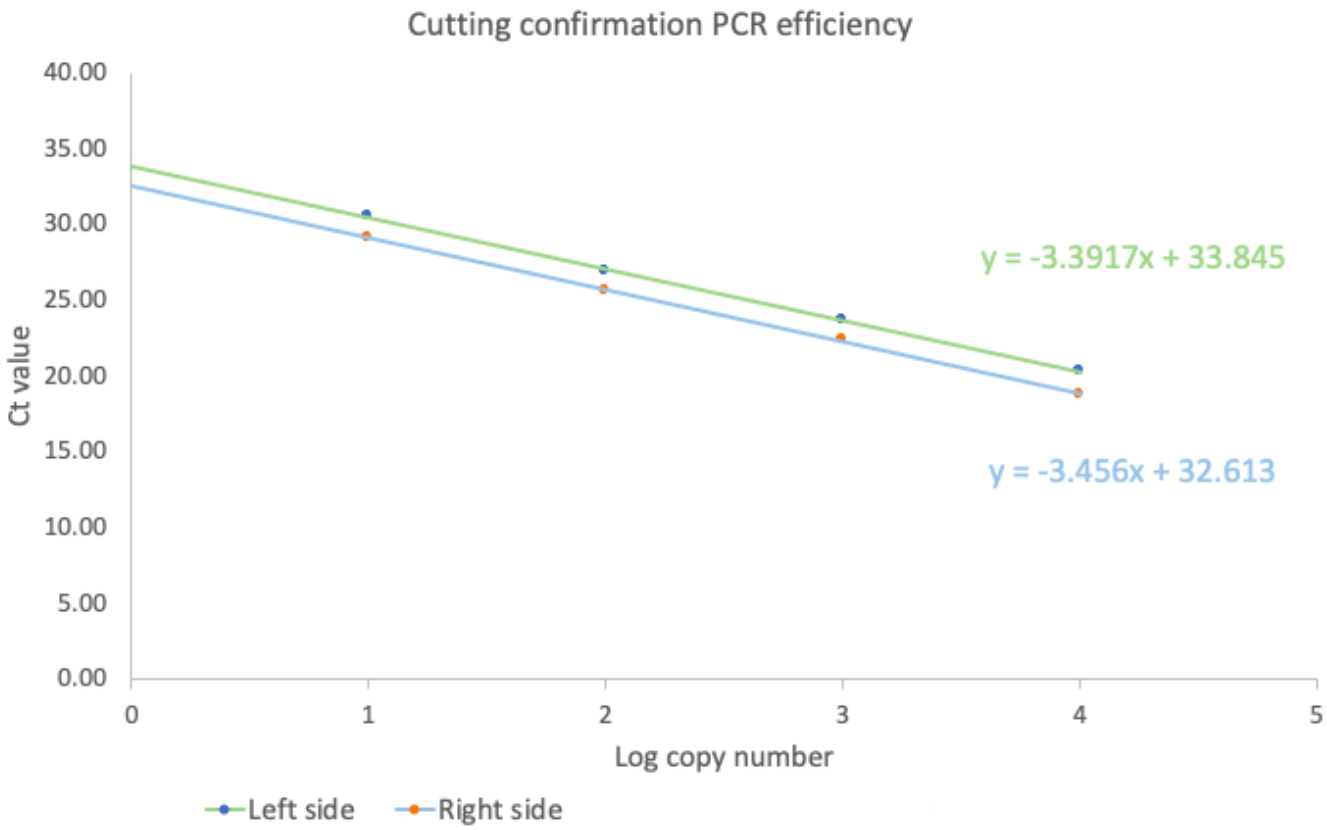


Figure 11: Graphic illustration of the cutting confirmation PCR efficiency. The primary axis shows the logarithmic copy number. The secondary axis shows the Ct value. Both the copy number and the Ct values of each point decrease in correlation to the increasing DNA dilution. The blue line represents the PCR for the left side of the transposon, and the orange line represents the PCR for the right side.

After establishing the PCR efficiencies, the PCRs were used to check if the CRISPR-Cas9 targeted cutting was successful. Extracted DNA from the *E. faecium* culture was treated with the Cas9 enzymes; one aliquot with the enzyme targeting the 5´side and one aliquot with the enzyme targeting the 3´ side of the transposon. Triplicates of the DNA products from the Cas9 reactions, along with an aliquot of DNA untreated by the Cas9 enzyme, were analyzed and

compared. The 5´-end cutting resulted in an average increase of the PCR Ct-value of 4.22 (from 13.24 to 17.46) corresponding to more than a 10-fold dilution, i.e. cutting of more than 90% of the target. The Ct-value for the 3´-end cutting was 3.05 (from 12.44 to 15.49) corresponding to an eight-fold dilution and cutting of more than 80% of the target sequence.

4.2 Sample enrichment

Samples with a Ct value lower than 31 in the hospital's routine vanB-PCR screening, without subsequent growth of VRE and with a negative enterococcus PCR, were selected for anaerobe enrichment. A total of 10 samples qualified for enrichment. The goal was to optimize growth conditions for potential transposon-carrying anaerobes, thus maximizing target input to the CRISPR-Cas9 targeted sequencing. Only one of the ten samples showed an increased vanB-PCR positivity after enrichment with a decrease in Ct value from 25.8 to 18.7 (Table 4.1), indicating selective growth of *vanB*-gene carrying bacteria. An eleventh sample from the VRE screening program were also collected but not enriched, as it already was a strong positive sample with a Ct value of 18 from the vanB-PCR. For both samples the Ct-values of the transposon Tn*1549*-PCR correlated well with the Ct-values in the vanB-PCR, indicating that the *vanB* gene in these samples was located on transposon Tn*1549*.

	1. vanB-PCR		**2. vanB-PCR**	
Sample	VB from VRE screening	Fecal swab	Fecal swab in new VB	VB from VRE screening in new VB
1	27.7	27.08	28.65	31.1
2	28.9	30.29	31.15	33.93
3	28.8	37.04	36.17	33.04
4	30.3	35.32	33.17	31.47
5	29.9	30.03	32.55	35.14
6	30.07	30.64	34.53	31.21
7	28.6	29.06	30.69	31.13
8	28.9	29.53	29.81	32.06
9	25.8	32.8	18.65	30.05
10	28.7	29.75	31.15	32.7

Table 4.1: Overview of Ct values from 10 samples analyzed with vanB-PCR. To the left are the Ct values from the first vanB-PCR used in VRE screening at HUS, including the fecal swab. The light green marked columns to the right are the samples enriched by incubation at anaerobic conditions for 48 hours at 35° C. 100µL of the primary samples (fecal swab and vancomycin broth with fecal swab) were transferred to respective new vancomycin broths (VB). Both the enriched samples were analyzed with a second vanB-PCR to compare Ct value before and after enrichment. Marked in darker green is sample 9, which showed a decrease in Ct value in the enriched VB with the transferred fecal swab.

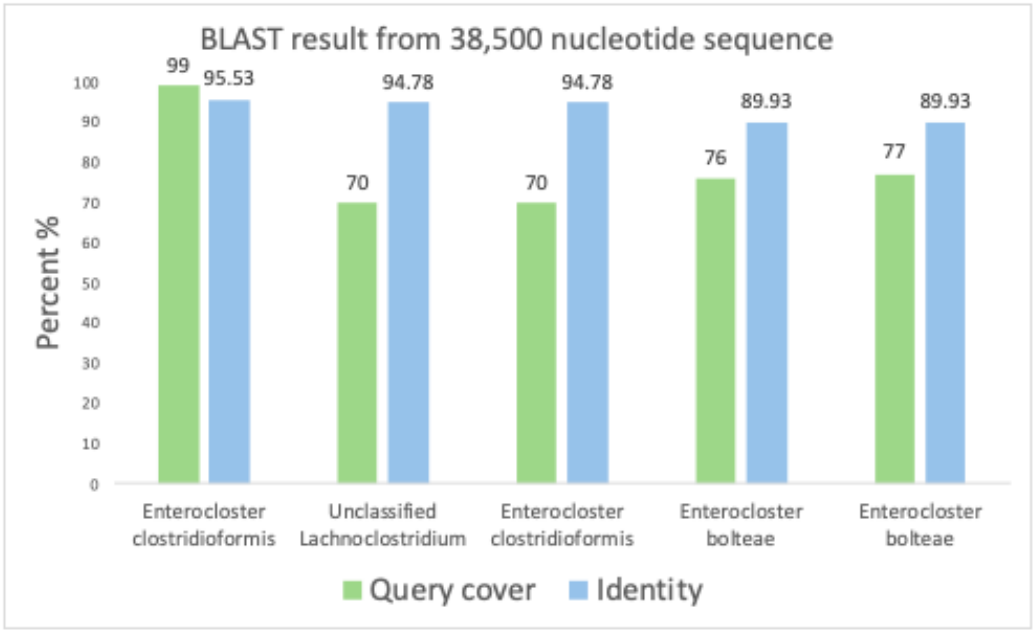
4.3 CRISPR-Cas9 targeted sequencing of strong *vanB*-gene positive samples

The CRISPR-Cas9 targeted sequencing was performed on sample 11 and the enriched sample 9, which both had a Ct value lower than 20 from the vanB-PCR. Sample 9 had a DNA concentration of 3.18ng/L, giving a gDNA input to the library preparation of 0.0763µg. Sample 11 had a concentration of 2.15ng/L, giving an input of 0.06036µg total DNA to the library preparation. The samples were treated with a mix of the two CRISPR-Cas9 enzymes targeting both the 5´- and the 3´-side of the transposon in the same reaction.

Sample 11 obtained a total of 47 reads after filtering by Guppy (minimum q-score of 7) in the range from 400 to 41,000 nucleotides. Only two target sequences were found, corresponding to a target yield of 4.25%. One of the targeted reads had a sequence length of 41,252 nucleotides, and aligned to the 5´-end of the transposon, meaning the sequence contained 38,500 nucleotides with host genome information upstream of the transposon. The second targeted read had a sequence length of 13,160 nucleotides and aligned to the 3´-end of the transposon, giving almost 11,000 nucleotides of host genome information downstream of the transposon. As there were only one sequences covering each side flanking the transposon, construction of consensus sequences was not possible.

The BLAST result for both target sequences found in sample 11 showed several matches to bacteria within the *Enterocloster* genus, with the best match to *Enterocloster clostridioformis* (accession number: CP050964.1). The longest sequence upstream of the transposon, *E. clostridioformis* had a query cover of 99% with an identity of 95.53%. The shorter sequence downstream of the transposon, *E. clostridioformis* had a query cover of 49% with an identity of 95.59%. The other matches against the long upstream sequence also had relatively high percent homologies but with a considerably lower query cover. For the shorter downstream sequence, the percent identity was also relatively high for all matches. The query cover on the other hand was under 50% for all, including the *E. clostridioformis* with the best match to both sequences. Nevertheless, the high percent identity in combination with the relatively large coverage considering the 38,500-nucleotide upstream sequence, indicated that the *vanB*-gene containing transposon Tn*1549* was carried by an *E. clostridioformis* with close similarity to the *E. clostridioformis* strain FDAARGOS_739 chromosome.

a)



b)

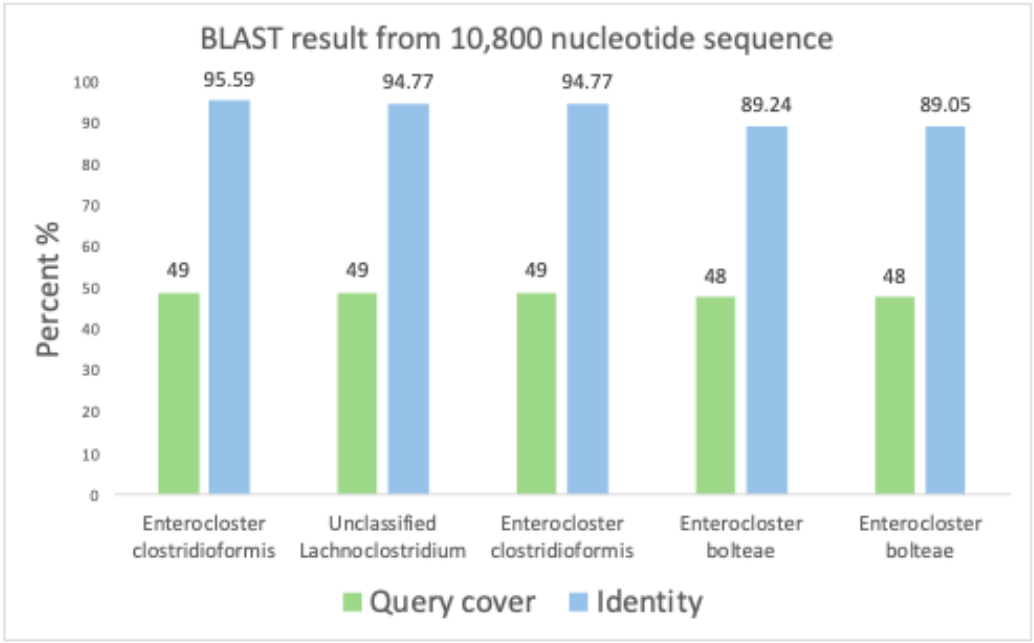


Figure 12: BLAST results of the two target sequences found in sequencing output from sample 11. The diagrams show the five BLAST results with the highest percent identity match to the query sequence. The green columns represent query cover in percent, meaning how much of the query sequence is covered by the matching sequences. The blue column represents percent identity, meaning how many percent of the query is identical to the aligned sequence. **a)** The BLAST results from the target sequence of 41,252 nucleotides flanking the upstream (5´) part of the transposon Tn*1549*. The best match is to *E. clostridioformis* sequence, with a percent identity of 95.53% and query cover of 99%. **b)** The BLAST results from the target sequence of 13,160 nucleotides flanking the downstream (3´) part of the transposon Tn*1549*. The best match is *E. clostridioformis* with 95.59% identity and 49% query cover. For both sequence-BLASTs, the other matches are also in the *Enterocloster* genus, except for the unclassified match of a *Lachnoclostridium* spp.

From sample 9 we remained with 154 reads after filtering by Guppy, with sequence lengths varying from 120 to 20,000 nucleotides. Among these there were three target sequences, giving a yield of 1.95%. All three sequences aligned to the 2750 base pairs at the start of the transposon Tn*1549*, thus they all gave information about their host genome upstream of the transposon. The sequences had lengths of 11,705 nucleotides, 5,059 nucleotides, and 3,911 nucleotides. Using Geneious R9, the three sequences were aligned in an attempt to construct consensus sequences. However, the shortest of the three sequences turned out not to represent the same host genome as the two longer sequences. Consequently, we were also this time prevented from constructing a consensus sequence. The shortest of the three sequences had 97% query cover and 90% identity to more than 15 different *Enterococcus faecium* isolates. The sequence of 5,059 nucleotides had a query cover of 96-97% and 89.25-89.76% identity to four different strains of chromosomal *Flavonifractor plautii,* where the best match was a chromosomal genome with accession number CP095094.1. The longest target sequence of 11,705 nucleotides had a query cover of 96% with 81.98-82.23% identity to the same four different strains of chromosomal *Flavonifractor plautii,* where the best match was to a chromosomal genome with accession number CP015406.2.

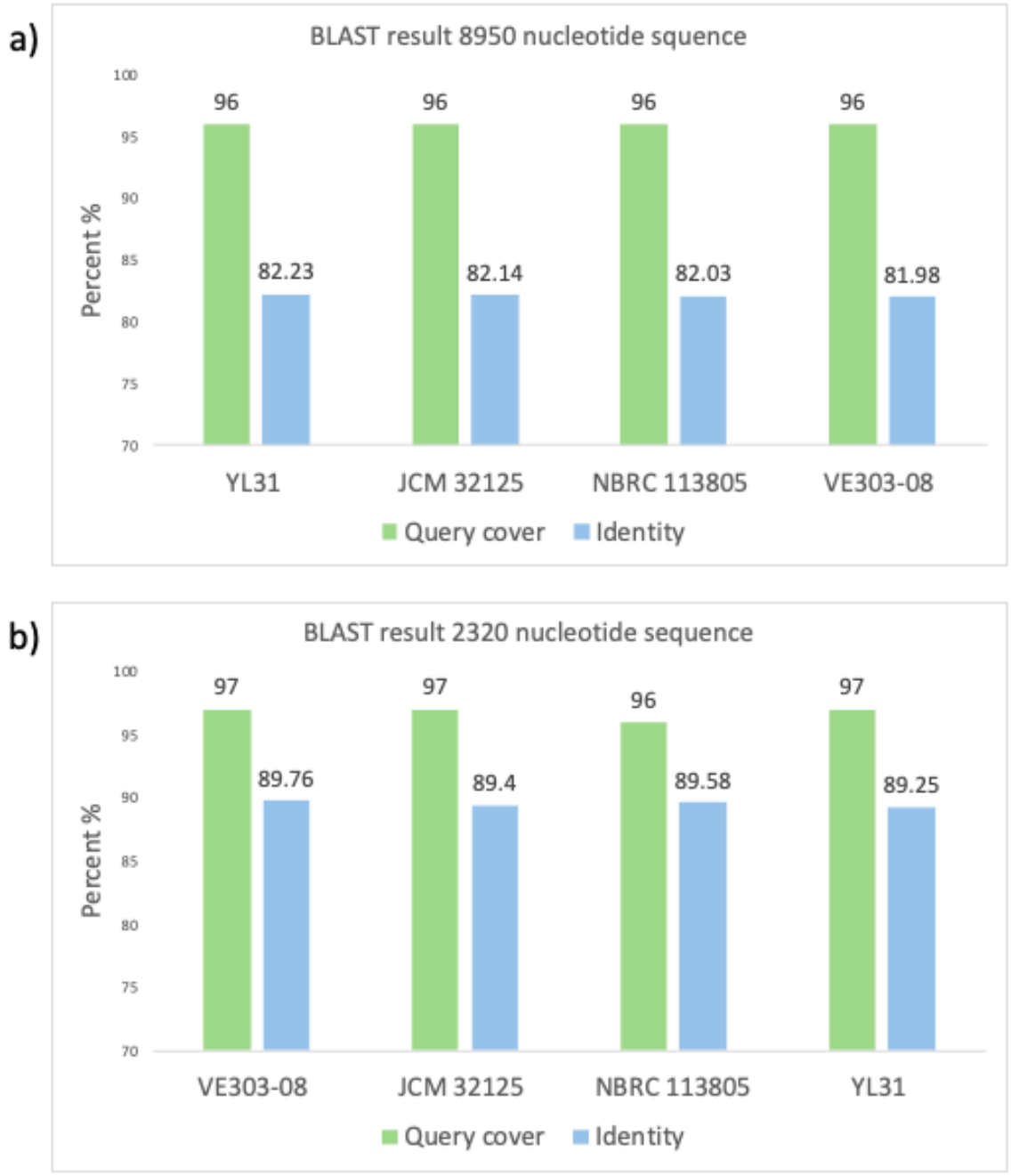


Figure 13: BLAST result from the two target sequences found in sample 9 with matches only to *Flavonifractor plautii* strains. The green columns represent the query cover and the blue columns represent the percent identity between the query and the aligned matches. Both target sequences a and b are located upstream of the transposon Tn*1549* and had matches to different strains of *F. plautii*. **a)** All strains of *F. platuii* had a query cover of 96% to the BLAST input sequence of 8950 nucleotides, with an identity match varying between 82.23 to 81.98. **b)** The *F. plautii* strains had a query cover of 96-96% to the BLAST input of 2320 nucleotides. There was a low variation in percent identity match which varied from 89.76 to 89.25. The identity match to *F. plautii* strains was on average 7.4% higher for the shorter sequence, b, compared to sequence a.

Due to the high sequencing error rate of nanopore sequencing, individual sequences will obtain a relatively low % homology even against the correct reference Therefore, to underpin the findings that the transposon carrying bacteria in samples 9 and 11 were indeed *F. plautii* and *E. clostridioformis*, species-specific PCRs targeting the *rpoB* and *nusG* genes of each bacterium were designed. The Ct values of the species-specific PCRs correlated well with those of the vanB- and Tn1549-PCRs.

4.4 Attempted method improvements

The yield from the Nanopore sequencing following CRISPR-Cas9 enrichment was disappointingly low, both when analyzing *E. faecium* DNA from pure culture during the method establishment, and especially when investigating the strong positive vanB-fecal samples. To ensure that the low yield was not due to too low accessibility of nanopores during sequencing, we re-sequenced sample 9 on a larger standard Flow cell. The sample DNA had a concentration of 3.18ng/µL, giving an input of 0.0763µg for the CRISPR-Cas9 library preparation. After sequencing and filtering by Guppy, we remained with an output of 245 reads with sequence lengths varying from 200 to 21k nucleotides. From the 245 reads, only 2 of the reads aligned with the transposon, giving a target yield of 0.8%. This indicated that using a larger and more expensive standard Flow cell did not improve the sequencing target yield.

We also tested if the Cas9 enzyme would produce a larger amount of target cuts if the enzyme activation time was increased. In parallel, four samples of E. faecium DNA containing the transposon Tn*1549* were treated with the Cas9 enzymes targeting both the 3´and the 5´end of the transposon in the same reaction. Each sample treatment had a different Cas9 activation time; 15, 20, 25, and 30 minutes. The samples were then analyzed in duplicates using the cutting confirmation PCR, where the largest increase in Ct values was found for the samples treated in 15 and 30 minutes for both the 3´and the 5´taregt site. The difference in Ct values for these was compared, and both were found not to be significant as p= 0.080 and 0.072 for the 5´and the 3´ target sites respectively. In conclusion, increasing the enzyme activation time did not seem to be of significance.

As the low yield did not seem to be due to a low amount of available nanopores or output capacity, and the activation time of the Cas9 enzyme seemed to be optimal, we wanted to see if there had been any loss of target during the library preparation procedure. To check this, we used a PCR targeting the very end of both the 5´and 3´end of the transposon. The only steps during the library preparation in which loss of genomic DNA was possible, were during the washing steps after adapter ligation. However, the DNA ready to be sequenced showed a Ct value similar to the DNA analyzed prior to the dephosphorylation step, confirming that there had been no loss of target sequences during the library preparation.

4.5 Metagenomic sequencing of the sample with assumed transposon Tn*1549* positive *F. plautii*

Due to the low yield, there was some uncertainty about the true identity of the Tn*1549* positive bacterium in sample 9. Therefore, we performed metagenomic sequencing of the sample using a Flow cell on the GridION sequencing instrument.

The sequencing output contained 6.4Gb after filtering by Guppy, with sequences varying in lengths from 100 to 70,000 nucleotides. The large data set was first analyzed without any assembling or polishing, using Nanopore Technologies visual bioinformatics tool; Epi2ME. All raw data were uploaded and analyzed using the Fastq "What's in my pot" (WIMP) for abundance classification of the nanopore reads. The Program managed to throughput 1,5 million of the reads and was able to classify roughly 1,1 million of these. The most abundant sequences represented the bacterium *Butyricomonas faecalis* with more than 455 thousand matches to the NCBI (National Center for Biotechnology Information) database. The second most abundant sequence found was *F. plautii*, with more than 283 thousand matches to the NCBI database. Down to the third most abundant sequences, there was a significant jump, where different species of *Bacteroides* had approximately 80k matches to the database.

A second bioinformatics tool, Kraken2 (Wood et al., 2019), was also used to make an abundance classification estimate. Different tools may use different databases and/ or are not up to date, resulting in the same data set getting different abundance classifications. Out of 1,54 million reads, Kraken2 managed to classify 78%, where the most abundant sequences were found to be a *Butyricomonas*, only this time it was classified as a *Butyricomonas virosa* (533k). The second most abundant sequences were classified to be *F. plautii* here as well (280k), giving decent grounds to believe that the sample does in fact contain a large amount of this bacteria. In this classification as well, there was a significant jump down to the third most abundant of sequences, which in this case was classified to be *Butyricomonas faecalis* with 74k matched sequences. The Pavian R package (Breitwieser & Salzberg, 2020) was used to visualize the classification results.

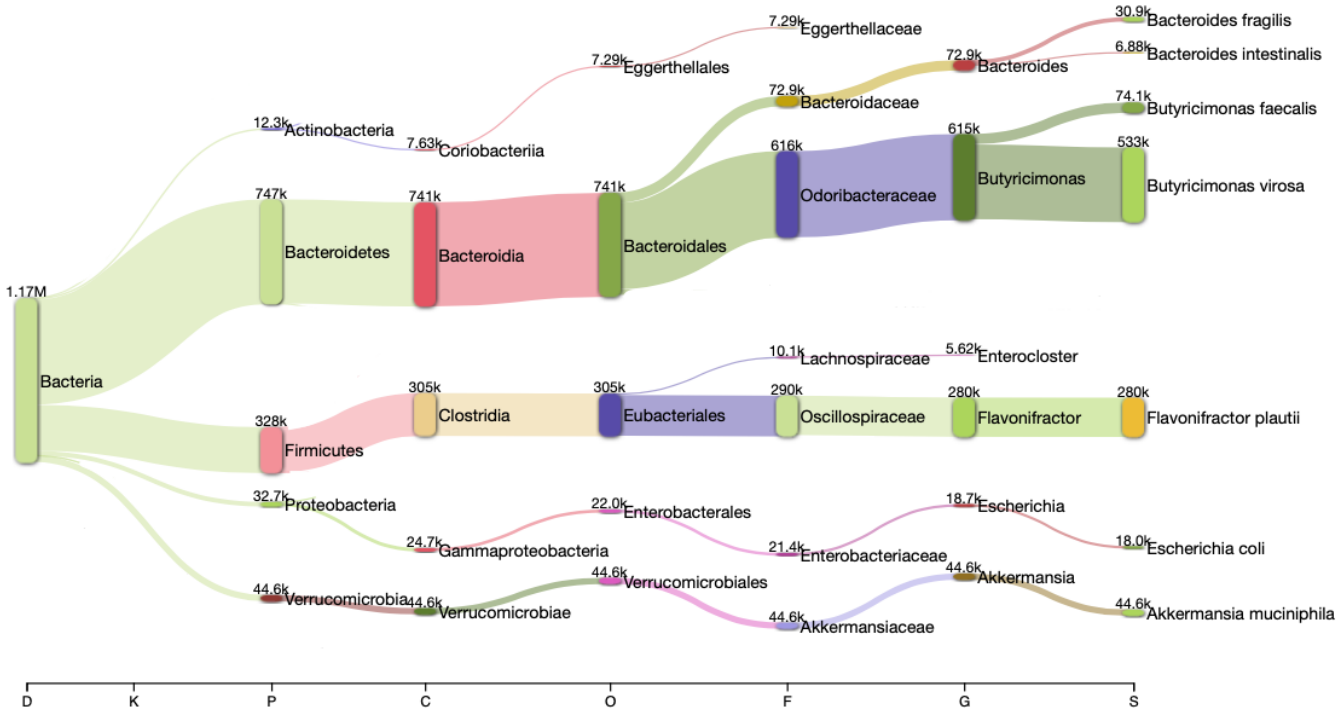


Figure 14: A Sankey visualization of abundance classification of sequence reads from whole-genome sequencing of sample 9, generated by Pavian. The figure illustrates some of the classified reads from Kraken2, where the size of the columns reflects the number of sequences classified within that strain. The amount of sequences classified within each branch is also stated. The largest of the columns lead to the genus *Butyricomonas*, where *B. virosa* is classified as the largest branch and *B. faecalis* is classified as the third largest branch of the total data analyzed. The second-largest column lead to *Flavonifractor plautii* with 280k classified reads.

In order to confirm that the *vanB* containing transposon Tn*1549* was carried by the second most abundant bacteria, *F. plautii*, in sample 9, a De novo assembly was performed from the whole genome sequencing output. The assembly was performed using metaFlye assembler which performs de novo assembly on metagenomic single-molecule sequencing reads. The metaFlye assembler also polished the assembled contig output to remove possible errors from the assembly, like those which typically occur in homopolymer regions. The finished assembly output contained approximately 50Mb of assembled DNA, where the largest assembled contig had a size of 3,7Mb and was classified to be *F. plautii*. Using BLAST, one searched for the transposon Tn*1549* on the contig, resulting in a match where the transposon was found located 91k nucleotides downstream of the start of the contig. The 11,705-long read obtained from the CRISPR-Cas9 targeted sequencing of sample 9 was then mapped against the 3.7Mb contig using Geneious High Sensitivity “Map to reference” tool. The mapping showed that the

placement of the transposon in the *F. plautii* assembled genome corresponded to the targeted sequencing read. Other contigs of *F. plautii* were also found, indicating fragmented contigs of the genome. After the assembly had finished, the metaFlye assembler also generated a Gfa file which to visualize the assembled contigs. The Gfa file was studied using Bandage.

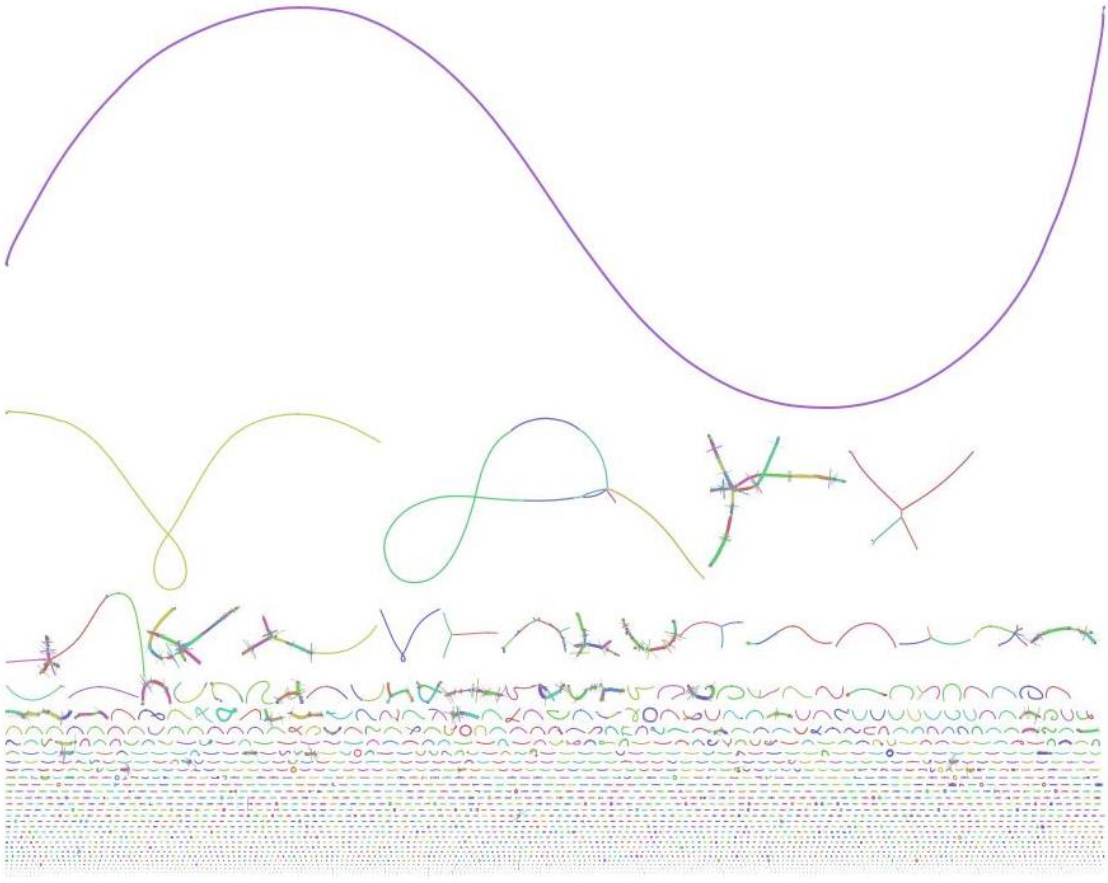


Figure 15: Visualization of assembled contigs in Bandage. The metagenomic output provided by the Nanopore whole genome sequencing of sample 9 was de novo assembled using metaFlye assembler. The assembled contig output was presented as a Gfa file and was visualized using Bandage. In Bandage every contig was assigned a random color, where the largest contig in purple represents the 3,7Mb assembled *F. plautii* containing the transposon Tn*1549*.

5.0 Discussion

It has been suggested that intestinal non-enterococci bacteria carrying transposable vancomycin resistance genes could play a role in the emergence of VRE. However, the spectrum of intestinal bacteria that can potentially harbor such elements is not well studied. At HUS, multiple outbreaks have been caused by enterococci carrying the *vanB*-gene positive transposon Tn*1549*. The main goal of this thesis was to establish a method for finding other intestinal carriers of Tn*1549* based on CRISPR-Cas9 enrichment and long-read sequencing. We have proved that the concept works and have identified two species other than enterococci carrying the Tn*1549* transposon.

5.1 Evaluation of the targeted CRISPR-Cas9 sequencing method

During the establishment and evaluation of the CRISPR-Cas9 method, we found the sequencing output and the on-target yield to be disappointingly low. Therefore, we did various attempts to improve the method. Firstly, we tested if the on-target yield would improve if the sequencing was performed on a larger and more expensive Flow cell as compared to the Flongles we had been using. Secondly, we tested if increasing the activation time and allowing the Cas9 enzyme to work for a longer period would improve the cutting efficiency. None of these attempts improved the output, and the Cas9 cutting-confirmation PCR also allowed us to confirm that the cutting process itself was effective. We therefore speculated that there might have been a loss of target input during the washing steps in the sample preparation. This suspicion was disproved by a real-time PCR which confirmed that there were equal amounts of target DNA both before and after the sample preparation. We therefore concluded that the pre-sequencing part of the protocol has been well optimized by the manufacturer.

5.1.1 Targeted reads obtained from sequencing

Using *E. faecium* DNA extracted from pure culture, we first tested the crRNA-directed Cas9 enzymes in separate reactions that were sequenced on separate Flongle flow cells. Here, the 5´-targeting aliquot had an on-target yield of 32.5%, and the 3´-targeting aliquot had an on-target yield of 38.24%. As the input to the library preparation only contained pure *E. faecium* DNA, we had expected the proportion of targeted reads to be much higher. Interestingly, the total on-target yield increased by approximately 2-fold when the cutting reactions for the 5´- and the 3´-end were combined by pooling the crRNAs for both target sites. It is speculated if the number

of targeted cut sites i.e. the number of pooled guide-RNAs influences the fraction of on-target reads, where a higher number of cleaved DNA fragments with compatible ends for adapter ligation result in a higher fraction of on-target reads. This was the conclusion reached by McDonald et al. who used a high number of target sites for digestion and obtained an on-target value of 44% (McDonald et al., 2021), a conclusion that is also shared by Fiol et al (Fiol et al., 2022). Most reports with descriptions of targeted yield have used pools of guide RNAs targeting several locations in their target sequences at the same time. One of the most recent reports from Fiol et al. describes an on-target yield of 8.6%, which is distinctively higher than other studies like Gilpatrick et al. which describes an on-target yield of 4.61%, and López-Girona et al. with on-target yields of 3.04% and 2.08% (Gilpatrick et al., 2020; López-Girona et al., 2020). The on-target yield obtained from sequencing the *vanB*-positive patient samples with CRISPR-Cas9 enrichment has similar results as to the latter reports with an on-target yield of 1.95% and 4.25% in samples 9 and 11 respectively.

The number of total reads obtained both after sequencing of the *E. faecium* DNA during the initial testing and of the strong positive *vanB* samples was relatively low. A possible reason for this in addition to the use of only two cleavage sites during digestion might be the sequencing input. When testing the method using pure *E. faecium* DNA containing the transposon Tn*1549*, we knew that the total input corresponded to the amount of gDNA containing the transposon. That was not the case for the *vanB* positive patient samples, which contained a mixture of bacteria and where the host genome(s) of the transposon represented an unknown fraction of this input. Most of the earlier studies have been based on DNA from cultured cell lines, which makes up-concentration of high molecular weight DNA a rather simple task compared to up-concentrating the anaerobic bacteria from the fecal swab samples used in this project. Sample 9 had an input of 0.0763µg and sample 11 had an input of 0.06036µg and obtained a total of only 154 and 47 reads respectively. The low input is likely to have impacted the number of reads, whereas it did not seem to change the fraction of on-target reads. López-Girona et al. had a genomic input of 5µg to their library preparation, and Gilpatrick et al. had an input of 3µg, both reporting fractions of on-target reads similar to those obtained here but with a higher number of total reads. Even if our on-target fractions were similar to those obtained in other studies our low total output resulted in only two or three on-target sequences. Nanopore sequencing has a relatively high error rate, meaning that the nanopore approach relies on multiple reads to construct reliable consensus sequences (Chandak et al., 2020). This made the

low number of on-target sequences a problem, as it prevented us from constructing adequate consensus sequences, which again made sequence interpretation more difficult.

One of the main challenges in this thesis therefore seems to be the large amount of input that Nanopore Technologies requires for the sequencing results to be optimal. Even the reports that fulfill the input requirements receive a low fraction of on-target sequences unless using a high number of cutting sites by pooling guide RNAs. When testing the 5´and the 3´trageting reactions separately we found that the 3´targeting reaction had a 2.55-fold larger input than the 5´targeting reaction (chapter 4.1.1), which resulted in a 3.4-fold higher number of reads. This indicates that a larger input to the sequencing library gives a larger sequencing output. The adapters ligated to the target sequences work to concentrate the DNA on the membrane surface near the nanopores to enable a more efficient sequencing (Jain et al., 2016). Because of low yield despite PCRs confirming large amounts of target DNA, we speculate if it might be possible that the excess DNA that does not have ligated adapters is physically blocking this event so that the targeted strands are unable to reach the nanopores (Figure 16). There is also a possibility that when the nanopores are restricted from sequencing, because of physical blockage by excess non-adapter-ligated DNA or simply because there is a low DNA input, this affects the health of the nanopores so that they cannot work optimally. This has also been discussed on ONT´s own community website, where users describe a fast decrease of active nanopores during the first hours of sequencing in experiments with low DNA input. Both circumstances might explain the higher on-target yield in the testing of the CRISPR-Cas9 method with the pure *E. faecium* DNA where both high DNA concentrations could easily be obtained and a larger fraction of the library input was expected to be adapter-ligated as compared to the input from the strongly positive *vanB* samples.

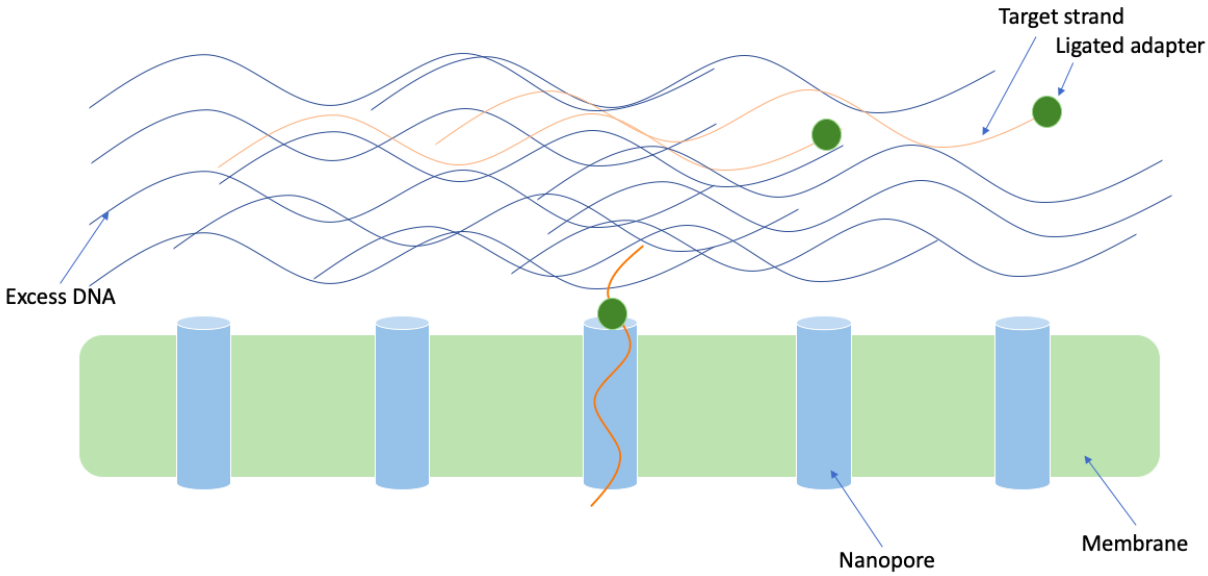


Figure 16: Illustration of excess DNA blocking target DNA sequencing in CRISPR-Cas9 sequencing method. The adapters ligated to the targeted DNA are drawn towards the membrane to bind to the membrane-bound nanopores. Only DNA strands with ligated adapters are sequenced as the adapters must covalently bind to the nanopores for the nucleotide sequence to be allowed into the nanopores. Excess DNA not cut by the Cas9 proteins during the library preparation is not ligated to adapters but remains part of the sequencing library. We speculate if this excess DNA might be blocking adapter-ligated DNA from accessing the nanopores.

After sequencing the two strongly positive *vanB* samples with the lowest Ct values of all our samples, we concluded that it was unlikely that any of the other samples with higher Ct values from the vanB-PCR would provide any relevant results. A single attempt was made to sequence a sample with a Ct value of 25 using the CRISPR-Cas9 enrichment (data not shown), where the sequencing output gave zero on-target reads.

5.1.2 crRNAs and target site

The crRNAs were designed with consideration for off-target ligation to both prokaryotic and human DNA and had a high probability for on-target ligation according to the design tools in ChopChop and IDT. The target sites for both crRNAs were more than 2k nucleotides into the transposon sequence from both ends, as these sites were given the best predictions according to the tools used. These two kilobases of known start sequence at each side of the transposon made the search for on-targets sequences easy but also gave some limitations. According to Nanopore Technologies, the Cas9 targeted sequencing provides the best coverage when the target

sequences are from 5-20k. Other studies report that the Nanopore read-length using both the CRISPR method and whole-genome sequencing gives a mean length of 1000-6000bp, where the tiling approach is not used (Croville et al., 2018; Fiol et al., 2022). This is longer than observed in this project, where the mean length was only 500-700bp (data not shown). The sequences of interest in this thesis are the ones flanking the 2.7kb/2.3kb transposon start area. Because of this, it was expected that a portion of our targeted reads would not provide data from the transposon host genomes, simply because they were too short to reach outside the transposon. Nevertheless, the same articles reporting the average read lengths also reported longer reads with sequence lengths over 30kb, which became useful in mapping and de novo assembly, and in the case of this project, useful for identification of the transposon carrier hosts.

5.2 *VanB* positive samples

Despite the low total output and the disappointing number of on-target reads, the results from the sequencing proved that the concept of using targeted CRISPR-Cas9 for the identification of unknown carriers of the transposon Tn*1549* is a method that works. From samples 9 and 11 we managed to identify two transposon carries other than enterococci.

5.2.1 *Enterocloster clostridioformis*

In sample 11 we obtained two on-target sequences, both with more than 95% identity to *Enterocloster clostridioformis* (formerly known as *Clostridium clostridioforme*) with accession number CP050964. The longest target read contained 38,500 nucleotides of host information genome upstream of the transposon and also had a 99% cover to the aligned *E. clostridioformis*, making the identification relatively certain. The shorter target sequence contained 10,800 nucleotides of host information downstream of the transposon but only half the sequence was aligned to other sequences in nucleotide BLAST. It is possible that the low query cover is due to a highly variable region in the *E. clostridioformis* genome, making alignments uncertain and difficult. This is not unusual for some bacteria and has been observed in different Enterocloster species several times before (Mohan et al., 2006; Song et al., 2003). *E. clostridioformis* has been reported to harbor the transposon Tn*1549* in earlier studies, where assays for antibiotic susceptibility showed vancomycin resistance and teicoplanin susceptibility corresponding to the expression of the *vanB* gene (Dehoux et al., 2016; Marvaud et al., 2011). The report from Marvaud et al. also demonstrated both *in vivo* and *in vitro* the transfer of Tn*1549* from *Clostridium symbiosum* to *Enterococcus* spp., but the transfer from *E. clostridioformis* to *E.*

faecium in vitro was unsuccessful. That being said, *in vitro* studies do not reproduce the complex conditions in the human colon, and a negative *in vitro* experiment does not in any way exclude the possibility of gene transfer. Looking at our BLAST results, *Enterocloster bolteae* was one of the closes matches to our query after *E. clostridioformis*. This might not be surprising as they are closely related, in addition to *Enterocloster bolteae* previously belonging to the *E. clostridioformis* complex. The two species have close to 3% divergence in the *16S* rRNA gene and a 37% association value *(Song et al., 2003)*. Because of this resemblance, it was important to classify this species with more than one method, making the species-specific two-target PCR that we designed to separate *Enterocloster bolteae* and *E. clostridioformis*, a valuable tool that confirmed our BLAST results.

5.2.2. *Flavonifractor plautii*

In sample 9 we obtained two targeted sequences from upstream of the transposon Tn*1549* that did not align with enterococci. The sequences contained less than 9k and 2.5k nucleotides of host information, where the respective nucleotide BLAST results showed 82.23% and 89.76% homology to the bacterium *Flavonifractor plautii* (formerly known as *Eubacterium plautii*). Due to the short sequences and the relatively low homologies with the aligned strains, there were uncertainties regarding the true identity of our transposon carrier. *Flavonifractor plautii* is known to have a close genetic and functional resemblance to for example *Clostridium orbiscindens* (Carlier et al., 2010), where isolates have shown to share up to 99.7% sequence similarity, they can both degrade quercetin and are both known to be susceptible to teicoplanin and have reduced susceptibility to vancomycin. This knowledge, in addition to *F. plautii* never being described as a *vanB*-carrier before, strengthened the basis for further investigation to confirm the true identity of the transposon carrier in sample 9. The metagenomic sequencing analysis of the enrichment broth showed that *F. plautii* was the second most abundant bacterium after a *Butyricimonas* spp. The Butyricomonas genus was first described in 2009 and currently includes six species, whereof *Butyricimonas virosa* is the only one known to cause human infection (Enemchukwu et al., 2016). The de novo assembly of the metagenomic sequencing output resulted in a large contig of 3.7Mb representing the *F. plautii* bacterium and carrying the transposon Tn*1549*. The transposon was located 91k nucleotides downstream of the 5´-end of the contig, which correlated well to the host-sequence information from the two sequences obtained from the targeted CRISPR-Cas9 sequencing. *F. plautii* is a strictly anaerobic and partially gram-positive bacterium that is part of the human normal intestinal flora. It has been

reported that *F. plautii* can cause bloodstream infection (Karpat et al., 2021), and the bacterium has also been isolated from blood cultures earlier at Haukeland University Hospital (data not published). To our knowledge, this is the first time *F. plautii* is observed carrying the *vanB* gene and the transposon Tn*1549*. Assays for antibiotic susceptibility are yet to be performed but due to the significant decrease in PCR Ct values during the anaerobe vancomycin broth enrichment, we assume that the bacterium expresses the *vanB* gene and is indeed tolerant to vancomycin.

5.3 Enrichment and sequencing

At this time, due to the high DNA input requirements of the current Oxford Nanopore sequencing technology, our CRISPR-Cas9-based method is clearly dependent on the culture-based enrichment of *vanB*-positive organisms using vancomycin-containing broth. This is disappointing since culture-based enrichment requires the bacteria to be alive when the sample arrives in the laboratory and to be able to grow in the vancomycin broth under the conditions provided. Since many intestinal bacteria are strictly anaerobe with highly specific growth requirements, this is likely to prevent the discovery of some potential *vanB*-carriers. Initially, we had hoped that targeted CRISPR-Cas9 enrichment could provide us with a culture-independent method that could be applied to DNA extracted directly from the fecal samples.

Metagenomic sequencing worked well from the enrichment broth. In this study, it seems to represent a more robust approach than the CRISPR-Cas9 enrichment although it requires a more expensive standard Flow cell and some bioinformatics experience. In theory, metagenomic sequencing could also be used as a culture-independent approach to DNA extracted directly from the fecal sample without prior enrichment. However, due to the very high number of bacterial species present, and with the *vanB*-positive species probably only constituting a very minor part of the total microbiota, this would require an enormous sequencing depth. In addition, bioinformatics would become very complicated and computationally demanding (Rajan et al., 2019).

5.3 Future perspectives

This thesis aimed to use CRISPR-Cas9 enriched long-read sequencing to find and identify bacterial species other than enterococci which contain the *vanB*-carrying transposon Tn*1549*. Using this method, we managed to identify two Tn*1549 vanB* positive bacterial species, one of which has not previously been described as carrying the vancomycin resistance gene. All the work put down into establishing and evaluating this method, including PCR design and error searches, resulted in a thorough and educational process.

The CRISPR-Cas9 enrichment turned out to be less effective than initially expected for the identification of the *vanB* carriers. The main issue is the large genomic input required by Nanopore Technologies, as we obtained only two or three on-target sequences even when using DNA with Ct values lower than 20 in the transposon Tn*1549* PCR. That being said, we were successful in identifying two different *vanB* carriers, including the one that has not previously been acknowledged. Such work might be important in understanding the epidemiology behind VRE. Once one has identified potential *vanB* and transposon Tn*1549* carriers, further work to find and map the occurrence of these bacteria becomes easier as one can for example use species-specific PCRs. When carrier bacteria have been identified, a natural next step is to investigate the frequency of these bacteria in the human gut, as well as the frequency of *vanB*-carriage among these species. After the identification of *vanB* in both *E. clostridioformis* and *F. plautii*, a bachelor thesis for the students of Biomedical laboratory science was designed to continue some of this work. They have investigated the correlation between the *vanB* gene and the occurrence of other bacteria known to carry *vanB*, like *E. clostridioformis* and *F. plautii*, in patients screened for VRE. During their investigation they did find a statistically significant correlation between the occurrence of *vanB* and *F. plautii*, without knowing the causation. Other work to be carried out would be mating experiments to demonstrate if the *vanB*-gene carriers have the ability to transfer these resistance genes and potentially create VRE. As discussed earlier, unsuccessful mating experiments do not exclude the possibility of transfer. Also, a tiling approach to investigate the transposon Tn*1549* would be of interest to visually confirm that the vanB-carriage is actually due to the carriage of Tn*1549*.

In this project, we found the CRISPR-Cas9 enrichment method in itself to be an efficient method. In the future, if long-read sequencing technologies with lower input requirements are being developed it will represent an attractive method to search for unknown carriers of defined genes.

References:

Babakhani, S., & Oloomi, M. (2018). Transposons: the agents of antibiotic resistance in bacteria. *Journal of basic microbiology*, *58*(11), 905-917.

Ballard, S. A., Pertile, K. K., Lim, M., Johnson, P. D. R., & Grayson, M. L. (2005). Molecular characterization of vanB elements in naturally occurring gut anaerobes. *Antimicrobial Agents and Chemotherapy*, *49*(5), 1688-1694.

Blokesch, M. (2016). Natural competence for transformation. *Current Biology*, *26*(21), R1126-R1130. https://doi.org/https://doi.org/10.1016/j.cub.2016.08.058

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics*, *36*(4), 1303-1304. https://doi.org/10.1093/bioinformatics/btz715

Bugg, T. D. H., Wright, G. D., Dutka-Malen, S., Arthur, M., Courvalin, P., & Walsh, C. T. (1991). Molecular basis for vancomycin resistance in Enterococcus faecium BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry*, *30*(43), 10408-10415.

Carias, L. L., Rudin, S. D., Donskey, C. J., & Rice, L. B. (1998). Genetic linkage and cotransfer of a novel, vanB-containing transposon (Tn 5382) and a low-affinity penicillin-binding protein 5 gene in a clinical vancomycin-resistant Enterococcus faecium isolate. *Journal of Bacteriology*, *180*(17), 4426-4434.

Carlier, J.-P., Bedora-Faure, M., K'Ouas, G., Alauzet, C., & Mory, F. (2010). Proposal to unify Clostridium orbiscindens Winter et al. 1991 and Eubacterium plautii (Séguin 1928) Hofstad and Aasjord 1982, with description of Flavonifractor plautii gen. nov., comb. nov., and reassignment of Bacteroides capillosus to Pseudoflavonifractor capillosus gen. nov., comb. nov. *International journal of systematic and evolutionary microbiology*, *60*(3), 585-590.

Cassier-Chauvat, C., Poncelet, M., & Chauvat, F. (1997). Three insertion sequences from the cyanobacterium Synechocystis PCC6803 support the occurrence of horizontal DNA transfer among bacteria. *Gene*, *195*(2), 257-266.

Chandak, S., Neu, J., Tatwawadi, K., Mardia, J., Lau, B., Kubit, M., . . . Ji, H. (2020, 4-8 May 2020). Overcoming High Nanopore Basecaller Error Rates for DNA Storage via Basecaller-Decoder Integration and Convolutional Codes. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Clark, D. P., & Pazdernik, N. J. (2016). Chapter 1 - Basics of Biotechnology. In D. P. Clark & N. J. Pazdernik (Eds.), *Biotechnology (Second Edition)* (pp. 1-31). Academic Cell. https://doi.org/https://doi.org/10.1016/B978-0-12-385015-7.00001-6

Clewell, D. B. (1990). Movable genetic elements and antibiotic resistance in enterococci. *European Journal of Clinical Microbiology and Infectious Diseases*, *9*(2), 90-102. https://doi.org/10.1007/BF01963632

Colomer-Lluch, M., Jofre, J., & Muniesa, M. (2011). Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PloS one*, *6*(3), e17549.

Croville, G., Le Loc'h, G., Zanchetta, C., Manno, M., Camus-Bouclainville, C., Klopp, C., . . . Guérin, J.-L. (2018). Rapid whole-genome based typing and surveillance of avipoxviruses using nanopore sequencing. *Journal of Virological Methods*, *261*, 34-39. https://doi.org/https://doi.org/10.1016/j.jviromet.2018.08.003

Dahl, K. H., Røkenes, T. P., Lundblad, E. W., & Sundsfjord, A. (2003). Nonconjugative transposition of the vanB-containing Tn5382-like element in Enterococcus faecium.

Antimicrobial agents and chemotherapy, *47*(2), 786-789. https://doi.org/10.1128/AAC.47.2.786-789.2003

Dehoux, P., Marvaud, J. C., Abouelleil, A., Earl, A. M., Lambert, T., & Dauga, C. (2016). Comparative genomics of Clostridium bolteae and Clostridium clostridioforme reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. *BMC genomics*, *17*(1), 1-16.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., . . . Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602-607. https://doi.org/10.1038/nature09886

Derbyshire, K. M., & Grindley, N. D. F. (1986). Replicative and conservative transposition in bacteria. *Cell*, *47*(3), 325-327.

Domingo, M. C., Huletsky, A., Giroux, R., Boissinot, K., Picard, F. J., Lebel, P., . . . Bergeron, M. G. (2005). High prevalence of glycopeptide resistance genes vanB, vanD, and vanG not associated with enterococci in human fecal flora. *Antimicrob Agents Chemother*, *49*(11), 4784-4786. https://doi.org/10.1128/aac.49.11.4784-4786.2005

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096.

ECDC, E. C. f. D. P. a. C. (2020). *Antimicrobial resistance in the EU/EEa (EARS-Net). Annual epidemiological report for 2019.* ECDC.

Egner, C., & Berg, D. E. (1981). Excision of transposon Tn5 is dependent on the inverted repeats but not on the transposase function of Tn5. *Proceedings of the National Academy of Sciences*, *78*(1), 459-463.

Enemchukwu, C. U., Ben-Faras, H., Gialanella, P., Szymczak, W. A., Nosanchuk, J. D., & Madaline, T. F. (2016). Butyricimonas virosa bacteraemia and bowel disease: case report and review. *New Microbes and New Infections*, *13*, 34-36.

FHI. (2019). Enterokokkinfeksjon (inkl. vankomycinresistente enterokokker, VRE) - veileder for helsepersonell. https://www.fhi.no/nettpub/smittevernveilederen/sykdommer-a-a/enterokokkinfeksjon-inkl.-vankomyci/

Fiol, A., Jurado-Ruiz, F., López-Girona, E., & Aranzana, M. J. (2022). An efficient CRISPR-Cas9 enrichment sequencing strategy for characterizing complex and highly duplicated genomic regions. A case study in the Prunus salicina LG3-MYB10 genes cluster. *BioRxiv*.

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, *3*(9), 722-732.

Garnier, F., Taourit, S., Glaser, P., Courvalin, P., & Galimand, M. (2000). Characterization of transposon Tn1549, conferring VanB-type resistance in Enterococcus spp. The GenBank accession number for the 33803 bp sequence of Tn1549 is AJ192329. *Microbiology*, *146*(6), 1481-1489.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., . . . Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *science*, *312*(5778), 1355-1359.

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., . . . Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature biotechnology*, *38*(4), 433-438.

Gold, H. S., Unal, S., Cercenado, E., Thauvin-Eliopoulos, C., Eliopoulos, G. M., Wennersten, C. B., & Moellering Jr, R. C. (1993). A gene conferring resistance to vancomycin but not teicoplanin in isolates of Enterococcus faecalis and Enterococcus faecium

demonstrates homology with vanB, vanA, and vanC genes of enterococci. *Antimicrobial Agents and Chemotherapy*, *37*(8), 1604-1609.
Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, *1*(6), e60.
Heikema, A. P., Horst-Kreft, D., Boers, S. A., Jansen, R., Hiltemann, S. D., de Koning, W., . . . Hays, J. P. (2020). Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota. *Genes*, *11*(9). https://doi.org/10.3390/genes11091105
Hodges, T., Zighelboim-Daum, S., Eliopoulos, G., Wennersten, C., & Moellering Jr, R. (1992). Antimicrobial susceptibility changes in Enterococcus faecalis following various penicillin exposure regimens. *Antimicrobial agents and chemotherapy*, *36*(1), 121-125.
Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, *169*(12), 5429-5433.
Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology*, *17*(1), 1-11.
Jain, R., Rivera, M. C., Moore, J. E., & Lake, J. A. (2002). Horizontal gene transfer in microbial genome evolution. *Theoretical population biology*, *61*(4), 489-495.
Karpat, I., Karolyi, M., Pawelka, E., Seitz, T., Thaller, F., & Wenisch, C. (2021). Flavonifractor plautii bloodstream infection in an asplenic patient with infectious colitis. *Wiener klinische Wochenschrift*, *133*(13), 724-726.
Kchouk, M., Gibrat, J.-F., & Elloumi, M. (2017). Generations of sequencing technologies: from first to next generation. *Biology and Medicine*, *9*(3).
Lambert, M. S. (2007). Molecular Biosafety. *Applied Biosafety*, *12*(3), 196-197.
Lambertsen, L., Rubio-Cosials, A., Patil, K. R., & Barabas, O. (2018). Conjugative transposition of the vancomycin resistance carrying Tn1549: enzymatic requirements and target site preferences. *Molecular Microbiology*, *107*(5), 639-658.
Latella, M. C., Di Salvo, M. T., Cocchiarella, F., Benati, D., Grisendi, G., Comitato, A., . . . Recchia, A. (2016). In vivo Editing of the Human Mutant Rhodopsin Gene by Electroporation of Plasmid-based CRISPR/Cas9 in the Mouse Retina. *Molecular Therapy - Nucleic Acids*, *5*, e389. https://doi.org/https://doi.org/10.1038/mtna.2016.92
Launay, A., Ballard, S. A., Johnson, P. D., Grayson, M. L., & Lambert, T. (2006). Transfer of vancomycin resistance transposon Tn1549 from Clostridium symbiosum to Enterococcus spp. in the gut of gnotobiotic mice. *Antimicrobial agents and chemotherapy*, *50*(3), 1054-1062.
Leggett, R. M., & Clark, M. D. (2017). A world of opportunities with nanopore sequencing. *Journal of experimental botany*, *68*(20), 5419-5429.
Leija-Salazar, M., Sedlazeck, F. J., Toffoli, M., Mullin, S., Mokretar, K., Athanasopoulou, M., . . . Schapira, A. H. V. (2019). Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Molecular genetics & genomic medicine*, *7*(3), e564.
Lu, J.-J., Chang, T.-Y., Perng, C.-L., & Lee, S.-Y. (2005). The vanB2 gene cluster of the majority of vancomycin-resistant Enterococcus faecium isolates from Taiwan is associated

with the pbp5 gene and is carried by Tn 5382 containing a novel insertion sequence. *Antimicrobial agents and chemotherapy*, *49*(9), 3937-3939.
López-Girona, E., Davy, M. W., Albert, N. W., Hilario, E., Smart, M. E. M., Kirk, C., . . . Chagné, D. (2020). CRISPR-Cas9 enrichment and long read sequencing for fine mapping in plants. *Plant Methods*, *16*(1), 1-13.
Mak, A., Miller, M. A., Chong, G., & Monczak, Y. (2009). Comparison of PCR and culture for screening of vancomycin-resistant Enterococci: highly disparate results for vanA and vanB. *Journal of clinical microbiology*, *47*(12), 4136-4137. https://doi.org/10.1128/JCM.01547-09
Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., . . . Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, *9*(6), 467-477. https://doi.org/10.1038/nrmicro2577
Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., & Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology*, *33*(5), 538-542.
Marvaud, J.-C., Mory, F., & Lambert, T. (2011). Clostridium clostridioforme and Atopobium minutum clinical isolates with vanB-type resistance in France. *Journal of clinical microbiology*, *49*(9), 3436-3438.
McDonald, T. L., Zhou, W., Castro, C. P., Mumm, C., Switzenberg, J. A., Mills, R. E., & Boyle, A. P. (2021). Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nature Communications*, *12*(1), 3586. https://doi.org/10.1038/s41467-021-23918-y
Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, *11*(1), 31-46.
Mohan, R., Namsolleck, P., Lawson, P. A., Osterhoff, M., Collins, M. D., Alpert, C.-A., & Blaut, M. (2006). Clostridium asparagiforme sp. nov., isolated from a human faecal sample. *Systematic and Applied Microbiology*, *29*(4), 292-299. https://doi.org/https://doi.org/10.1016/j.syapm.2005.11.001
Moralez, J., Szenkiel, K., Hamilton, K., Pruden, A., & Lopatkin, A. J. (2021). Quantitative analysis of horizontal gene transfer in complex systems. *Current Opinion in Microbiology*, *62*, 103-109. https://doi.org/https://doi.org/10.1016/j.mib.2021.05.001
Palladino, S., Kay, I. D., Flexman, J. P., Boehm, I., Costa, A. M. G., Lambert, E. J., & Christiansen, K. J. (2003). Rapid detection of vanA and vanB genes directly from clinical specimens and enrichment broths by real-time multiplex PCR assay. *Journal of clinical microbiology*, *41*(6), 2483-2486. https://doi.org/10.1128/JCM.41.6.2483-2486.2003
Patel, R. (2000). Enterococcal-type glycopeptide resistance genes in non-enterococcal organisms. *FEMS microbiology letters*, *185*(1), 1-7.
Pontius, L. T., & Clewell, D. B. (1992). Conjugative transfer of Enterococcus faecalis plasmid pAD1: nucleotide sequence and transcriptional fusion analysis of a region involved in positive regulation. *Journal of bacteriology*, *174*(10), 3152-3160.
Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.
Rajan, S. K., Lindqvist, M., Brummer, R. J., Schoultz, I., & Repsilber, D. (2019). Phylogenetic microbiota profiling in fecal samples depends on combination of sequencing depth

and choice of NGS analysis method. *PloS one*, *14*(9), e0222171-e0222171. https://doi.org/10.1371/journal.pone.0222171

Sadowy, E. (2021). Mobile genetic elements beyond the VanB-resistance dissemination among hospital-associated enterococci and other Gram-positive bacteria. *Plasmid*, *114*, 102558. https://doi.org/https://doi.org/10.1016/j.plasmid.2021.102558

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research*, *39*(21), 9275-9282. https://doi.org/10.1093/nar/gkr606

Schultzhaus, Z., Wang, Z., & Stenger, D. (2021). CRISPR-based enrichment strategies for targeted sequencing. *Biotechnology Advances*, *46*, 107672. https://doi.org/https://doi.org/10.1016/j.biotechadv.2020.107672

Shah, S. A., Erdmann, S., Mojica, F. J. M., & Garrett, R. A. (2013). Protospacer recognition motifs. *RNA Biology*, *10*(5), 891-899. https://doi.org/10.4161/rna.23764

Sivertsen, A., Billström, H., Melefors, Ö., Liljequist, B. O., Wisell, K. T., Ullberg, M., . . . Hegstad, K. (2014). A multicentre hospital outbreak in Sweden caused by introduction of a vanB2 transposon into a stably maintained pRUM-plasmid in an Enterococcus faecium ST192 clone. *PLoS One*, *9*(8), e103274.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., . . . Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, *321*(6071), 674-679.

Song, Y., Liu, C., Molitoris, D. R., Tomzynski, T. J., Lawson, P. A., Collins, M. D., & Finegold, S. M. (2003). Clostridium bolteae sp. nov., Isolated from Human Sources. *Systematic and Applied Microbiology*, *26*(1), 84-89. https://doi.org/https://doi.org/10.1078/072320203322337353

Sood, S., Malhotra, M., Das, B. K., & Kapil, A. (2008). Enterococcal infections & antimicrobial resistance. *Indian journal of medical research*, *128*(2), 111.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W., & Kubista, M. (2015). How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*, *3*, 9-16. https://doi.org/https://doi.org/10.1016/j.bdq.2015.01.005

Szczelkun Mark, D., Tikhomirova Maria, S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., . . . Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences*, *111*(27), 9798-9803. https://doi.org/10.1073/pnas.1402597111

Tang, T.-H., Bachellerie, J.-P., Rozhdestvensky, T., Bortolin, M.-L., Huber, H., Drungowski, M., . . . Hüttenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences*, *99*(11), 7536-7541.

Technologies, O. N. (2019). *Amplification-free target enrichment for native-strand sequencing using CRISPR-Cas9*. Oxford Nanopore Technologies. https://nanoporetech.com/resource-centre/amplification-free-target-enrichment-native-strand-sequencing-using-crisprcas9

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology*, *3*(9), 711-721.

Van Tyne, D., & Gilmore, M. S. (2014). Friend Turned Foe: Evolution of Enterococcal Virulence and Antibiotic Resistance. *Annual Review of Microbiology*, *68*(1), 337-356. https://doi.org/10.1146/annurev-micro-091213-113003

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348-1365. https://doi.org/10.1038/s41587-021-01108-x

Watson, C. M., Crinnion, L. A., Hewitt, S., Bates, J., Robinson, R., Carr, I. M., . . . Bonthron, D. T. (2020). Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Laboratory Investigation*, *100*(1), 135-146. https://doi.org/10.1038/s41374-019-0283-0

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Yosef, I., Goren, M. G., Globus, R., Molshanski-Mor, S., & Qimron, U. (2017). Extending the Host Range of Bacteriophage Particles for DNA Transduction. *Molecular Cell*, *66*(5), 721-728.e723. https://doi.org/https://doi.org/10.1016/j.molcel.2017.04.025

Zhou, X., Chlebowicz, M. A., Bathoorn, E., Rosema, S., Couto, N., Lokate, M., . . . Rossen, J. W. A. (2018). Elucidating vancomycin-resistant Enterococcus faecium outbreaks: the role of clonal spread and movement of mobile genetic elements. *Journal of Antimicrobial Chemotherapy*, *73*(12), 3259-3267.