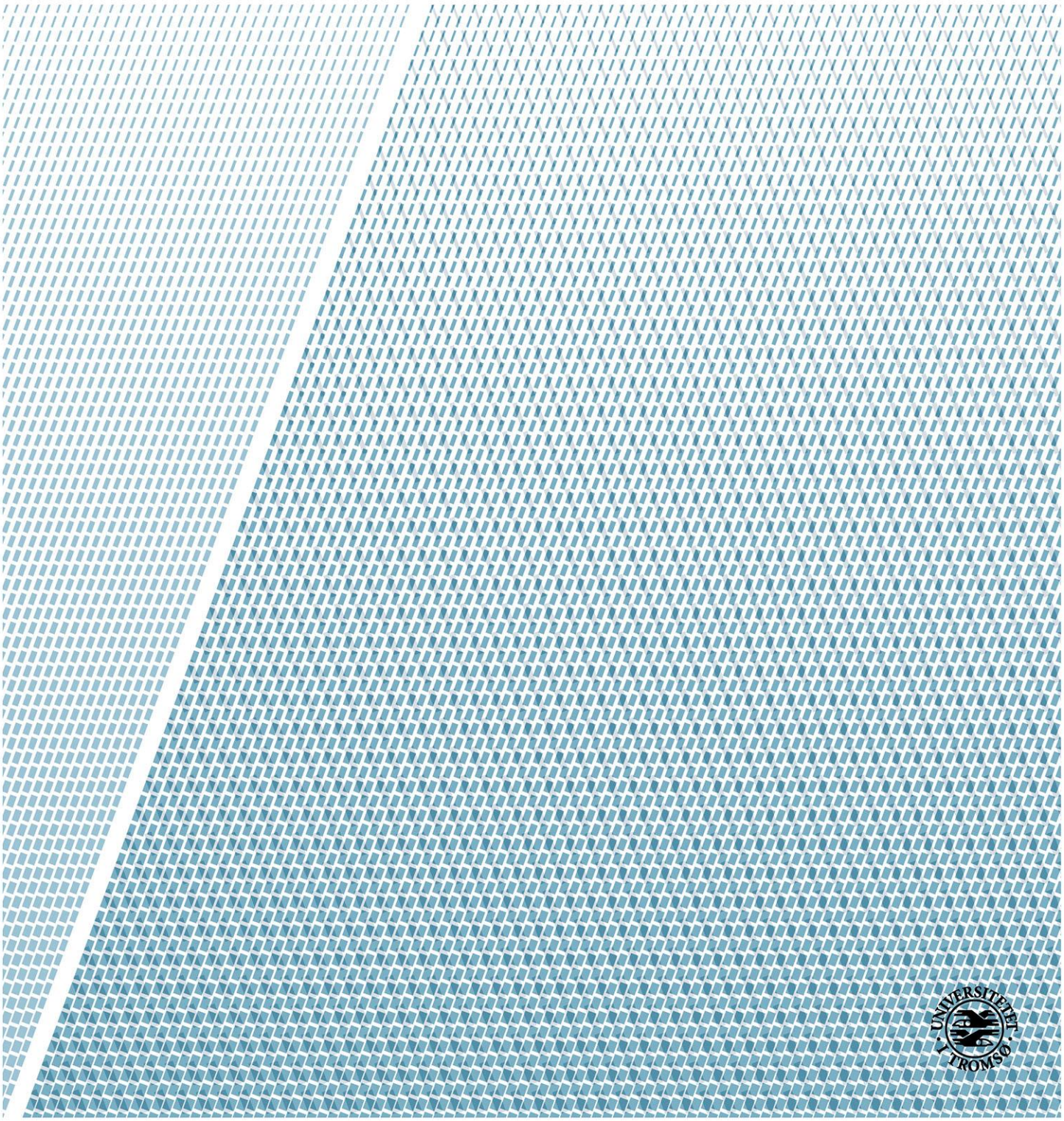
Institute of Pharmacy, the Northern Arctic University of Norway

SCREENING OF GENES ENCODING POTENTIAL VIRULENCE FACTORS IN *ENTEROCOCCUS FAECIUM* AND GAP CLOSURE OF THE rep_{pLG1} REPLICON CLASS PLASMID IN *E. FAECIUM*

Emmanuel Fayia

Master Thesis in Pharmacy, May.20.2014



ACKNOWLEDGEMENTS

This work leading to my Master of Science thesis was done at the department of Medical Microbiology, Centre for Detection of Antimicrobial Resistance (K-res) and Northern Arctic University of Tromsø.

Never forget those who helped you on your way up because when coming down, you might still need them!

First and foremost I want to use this precious opportunity to thank my supervisors, Kristen Hegstad Dahl, Torunn Pedersen and Kaare M. Nielsen for outstanding support council and supervision. I am very grateful for you tireless supervision and for guiding me through this demanding period. These six consecutive months have been a life changing experience. I am delighted by your trust and thoroughly guided by your confidence. I also thank the Reference Centre for Detection of Antimicrobial Resistance (K-res) for given me this privilege.

Also thanks to all staff, especially Theresa Mikalsen for technical guidance and support. Special thanks go to Tracy Munthali Lunde, without whom all would not have been the same. You were always with me when I needed you the most and helped me when I was helpless. For that I must own you love and respect.

To Erik Hjerde, I thank you for the moments I shared with you during my project. It was nice working with you and I will always remember you for your calmness and easy going.

My heartfelt gratitude goes to all friends and families for standing beside me day in and day out. Special thanks to Mathias F. Korfeh for excellent support, encouragement and counselling during my time in school.

Thanks to Rufus Pencimaque for his hospitality and companionship. You have been a driving force in pursue of my dreams and aspirations. I also wish to extend my gratitude to Ernest Tukov for all his prayers and councils.

Above all, I must thank my beloved wife for ignoring all my long hours in school and early departures from home. Finally, I thank my self for never given up. It has been educational and life changing experience.

ABBREVIATIONS

UTI Urinary Tract Infection

Ace Collagen Binding Protein

cylA Haemolysin Activator

efaA Endocarditis Antigen

esp Enterococcal Surface Protien

hyl Hyaluronidase

gelE Gelatinase

hyl (Efm) hyaluronidase *Enterococcus faecium*

UiT University of Tromsø

K-res Reference Centre for Detection of Antimicrobial Resistiance

ORF Open Reading Frames

MGE Mobile Genetic Elements

IS Insertion Sequence

CDS Signal Peptides

CC17 Clonal complex 17

CC2 Clonal complex 2

DNA Deoxyribonucleic acid

ATC Artemis Comparison Tool

PCR Polymerase Chain Reaction

dNTPs Dinucleotide Triphosphates

T4SS TIV Secretion System

mRNA Messenger Ribonucleic Acid

ddNTPs Dideoxynucleotide Triphosphates

SUMMARY

The aims of this study were to obtain a circular map of a pLG1 replicon type plasmid in *E. faecium* (TUH 56-32), derived from the trans-conjugation between *E. faecium* (K60-39), donor and (BM4105-RF) recipient strains and also to detect the presence of and describe the prevalence of selected open reading frames (ORF) from a clinical and non-clinical collection of *E. faecium* and *E. faecalis* strains.

The DNA sequence of *E. faecium* (TUH56-32) plasmid was subjected to gaps closure. Gaps closure was performed by PCR to obtain a circular map of the plasmid sequence. To detect the presence of ORFs total DNA was extracted from a clinical and non-clinical collection of 150 isolates which consisted of *116 E. faecium* and 34 *E. faecalis.* PCR was applied to detect the presence of ORFs.

The gaps closing experiment was not accomplished and therefore did not give a circular DNA. However, the screening experiment was completed and the detection and description of ORFs among *E. faecium* showed that isolates from blood cultures harboured more ORFs than those of other clinical and non-clinical sources. Those of non-clinical cultures however lacked three of the tested ORFs. The prevalence was higher among clinical isolates of all sources compared to non-clinical isolates. Isolates of *E. faecium* detected more prevalently to all ORFs than *E. faecalis*. The isolates derived from blood cultures were highly enriched with all four ORFs compared to isolates from other clinical sources. Blood culture isolates harbouring all ORFs differed significantly to those of other clinical sources. In summary, the tested ORFs were overrepresented in blood culture isolates of *E. faecium* followed by other clinical isolates.

INTRODUCTION

Enterococci

Because of their ability to cause public health hazards, enterococci are gaining more attention in recent years. Growing incidence of hospital associated enterococci has emerged, with *E. faecium* being in the lead (Felmingham et al., 1992). They are involved in antimicrobial resistance and hospital acquired infections (JUREEN, R. 2004). Enterococci belong to the genus enterococcus and were previously composed of 19 species (Monstein et al., 1998). In recent years the number of species has grown significantly as new methods are being implimented (Carvalho Mda et al., 2004, Law-Brown and Meyers, 2003).

General characteristics

The Enterococci are Gran positive bacteria, facultative anaerobic, catalysed negative, occur in pairs, but can also be seen in short chains (Deibel et al., 1963). They are naturally found in the gastrointestinal tract of mammals and are capable of surviving in harsh environments (Arias and Murray, 2012). They are capable of growing in harsh environments such as broth containing high sodium chloride concentration (6, 5%) and broad range pH (4- 9, 6). Growth also occur at 10 ^{0}C and 45 ^{0} C (Deibel et al., 1963). They can also be stored at – 70^{o}C for a long period of time (K-res,UiT).

E. faecium and *E. faecalis*

The genus enterococcus consists of several species including *E. faecium* and *E. faecalis*, which are among the clinically important species causing nosocomial infections in the human population. They are said to have acquired distinctive features allowing them to colonize the human host often causing infection (Vankerckhoven et al., 2004, Seno et al., 2005). Infections of the urinary tract of humans are common examples (Felmingham et al., 1992). Because of genetic diversity, they are regarded as two distinct species. According to Brodal (2011) the pan-genome was shown to be more diverse within *E.*

faecium than *E. faecalis* and larger in *E. faecium* than in *E. faecalis*. Another difference also highlights the presence of larger plasmids that have been detected in *E. faecium* known as the pLG1 replicon type that are not found in *E. faecalis (Rosvoll et al., 2012)* . *E. faecalis* however are shown to harbour pheromone responsive plasmids which are not found in *E. faecium* (Lim et al., 2006). These species are both described of hosting an open pan-genome, which means they are capable of acquiring and integrating foreign genetic elements into their genome (Tettelin et al., 2008). *E. faecalis* and *E. faecium* are the leading cause of hospital acquired infections within their genus, with previous accounts showing 90-95% of cases caused by *E. faecalis* and 5-10% of cases caused by *E. faecium (Arias and Murray, 2012)*. Recent accounts however, showed that *E. faecium* has taken the lead. In 2002, an American study which lasted over a decade showed a 10% increase in clinical isolates within *E. faecium,* while *E. faecalis* were stable throughout the 10 year period (Treitman et al., 2005).

Hospital and community associated lineages

Enterococci can be grouped into hospital associated and community associated clades. They have been discussed in many studies and are said to differ significantly in their genetic property(Kim and Marco, 2013, Brodal, 2011). Furthermore, isolates associated with hospital environments have been grouped into clonal complexes and their subpopulations to differ them from non-clinical isolates, with most common clonal complexes being CC17, *E. faecium* and CC2 *E. faecalis* (Valdezate et al., 2012, Rosvoll et al., 2012, Freitas et al., 2009). It has also being shown that the two groups are genetically far apart, which could explain their choice to populate different niches(Kim and Marco, 2013, Brodal, 2011). Distinguishing those that are disease causing pathogens from those that do not cause disease is important in identifying strains that are high-risk pathogens and those with the potentials to transform from low-risk into high-risk pathogens (de Regt et al., 2012). In the study complete genome sequence of the *E. faecium* strain TX16, Qin et al. (2012) highlighted a significant difference between clinical and non-clinical clades with 3-4% averaged nucleotide sequence difference recorded, which could mean a genetic diversification within *E. faecium*.

Mobile genetic elements (MGEs) of enterococci

The MEGs of Enterococci include plasmids, complex transposons, integrative conjugative plasmids and insertion sequence elements. Recent studies showed that enterococci are capable of absorbing and incorporating foreign elements into their genome as a means of adapting to various environmental pressures (Angulo et al., 2006, Macovei and Zurek, 2007,). However the transfer of mobile genetic elements across genetic lines is an important process and seems to enforce adaptability and persistence in various niches (Santagati et al., 2012). MEGs are by definition simple fragments of DNA consisting of different sizes and genetic properties that are capable of free movement between genomes or within genomes (Santagati et al., 2012). First discovered in the late 1940s, a lot has been understood in terms of how they function, thanks to decade long genome sequence studies. Moreover, they can be crucial in part due to their diverse nature. Studies showed that the acquisition of MEGs could impose unfavourable conditions to the species which they are hosted by leading to fitness problems (Starikova et al., 2013). On the other hand, some encode genes that confer resistance to antibiotics, while others provide key determinants for virulence (Paulsen et al., 2003). Interspecies variation in content of MGEs can also exist among clinical *E. faecium* and non-clinical *E. faecium*. For example previous accounts showed that clinical *E. faecium* are more enriched in MGEs than non-clinical *E. faecium* strains (Kim and Marco, 2013). These differences might also correlate with the fact that there are considerable distances between the two clades on the genome level. Previous accounts also showed that non-clinical *E. faecium* contain smaller genome compared to clinical *E. faecium* strains (Kim and Marco, 2013).

Enterococcal plasmids

Plasmids are described as circular DNA molecules capable of replicating independently of the bacterial chromosome. In enterococci plasmids form an integral part of the cell enabling survival and adaptation in new niches. Genetic studies of enterococci showed different plasmids types, and that plasmids differ greatly in size which also reflects their genetic content (Jensen et al., 2010). Based on their transfer mechanism three main groups of plasmids have been described in enterococci which includes conjugative plasmids, non-conjugative plasmids and pheromone responsive conjugative plasmid (Paulsen et al., 2003). Enterococcal plasmids can be broad host range or narrow host range. Broad host range plasmids exist in more than one species, while narrow host range plasmids are species specific. Example of arrow host range plasmids include pheromone responsive plasmids which are predominantly enriched among *E. faecalis* populations (Clewell, 2011).

Horizontal gene transfer

Horizontal gene transfer highlights the genetic exchange among bacteria and can be described in three ways: 1. Conjugation 2.Transduction 3.Transformation. Since transduction and transformation have not yet been identified in enterococci only conjugation will be discussed here. Bacterial conjugation involves the exchange of genetic elements as a mean of acquiring or spreading antimicrobial resistance genes or virulence factors by direct cell to cell contact (Goessweiner-Mohr et al., 2012), Model of this transfer has been proposed previously in gram-positive bacteria (Abajy et al., 2007). Conjugation provides the basics upon which MGEs such as plasmids and other transposons are exchanged and spread among enterococci (Rosvoll et al., 2012). The mechanism of conjugation involves the recruitment of so called tra-proteins to the type IV secretion-like system (T4SS-like complex) (Abajy et al., 2007). However the T4SS-like system of enterococci have been identified previously and are said to be quite similar to the T4SS system of other gram-positive bacteria (Goessweiner-Mohr et al., 2012).

Pheromone responsive plasmids and transfer mechanism

Enterococcal pheromone responsive plasmid conjugation is based upon the presence of a special protein called sex-pheromones (An and Clewell, 2002). Sex pheromones are, upon their release found on the outside of the cell and are specific for donors conveying conjugative plasmids. These signal proteins are released by the recipient cell and are chromosomally encoded (Palmer et al., 2010). In the presence of sex pheromones donors are triggered to response due to their pheromone sensitivity. In response they produce surface proteins called aggregation substances (ASs) (Kreft et al., 1992). ASs are responsible for linking the two cells into close proximity by binding to proteins called Enterococcal Binding Substances (EBS) that are present at the surface of the recipient cell. This triggers a conformational change which enables the exchange process to take place. Once the transfer is complete the recipient cell then ceases the production of sex pheromone (Clewell, 2011). The mating process shuts down once that plasmid is transferred, but does not hinder the continuing transfer of other plasmid types as pheromones are plasmid specific. However, acting donor cells produce plasmid encoded inhibitors that inhibit aggregation substances preventing further aggregation and hence the transfer of yet the same plasmid type (Clewell, 2011).

Megaplasmids of enterococci

Megaplasmid is an old phenomenon that was used to describe bacteria plasmids of very large sizes. The use of this phenomenon to described bacteria plasmids was first invented in 1981 by Rosenberg and co-workers (Rosenberg et al., 1981). Megaplasmids are also extra chromosomal genetic materials that are capable of self-replicating and therefor do not replicate dependent of the cell chromosomes (Steinbüchel and Schwartz, 2009). No threshold has been adapted for megaplasmids as regards to size; and so the threshold could vary from study to study depending on the kind of species that are studied. Most studies including *E. faecium* consider megaplasmid to be above 100 kilo bases (kb) (Freitas et al., 2010, Laverde Gomez et al., 2011, Panesso et al.,

2010). Laverde Gomez et al. (2011) showed that megaplasmid of *E. faecium* can range from 150 kb to 350 kb. They are said to encode genetic elements with virulence and resistance potentials (Laverde Gomez et al., 2011, Rice et al., 2003) . In their research, Garcia-Migura et al. (2011) showed an *E. faecium* strain that harboured 70 kb plasmid before it was subjected to trans-conjugation. The result was a 120-140 kb size plasmid after the experiment was conducted, suggesting the possibility of genetic recombination of plasmids (Willems and van Schaik, 2009, Willems et al., 2001).

Toxin – Antitoxin systems of enterococci

The toxin-antitoxin system is vital in maintenance of virulence factors, antimicrobial resistance genes and other mobile genetic elements (Grady and Hayes, 2003). It plays an important role in the management of cell growth and prevents premature cell death (Grady and Hayes, 2003). Toxin-antitoxin systems have previously been identified among enterococci. They can be described as a system containing toxins that are neutralized by antitoxins (Grady and Hayes, 2003). However, genes that express toxin and antitoxin are regulated by the transcription of regulators of the TA operon (Boss et al., 2013). Without antitoxins the effect of toxins on mobile genetic materials would make them unstable leading to maintenance problems (Boss et al., 2013). Four TA systems are previously identified in enterococci including axe-txe, mazEF, relB and ω-ε-ζ toxin-antitoxin systems and have been detected on megaplasmids (Rosvoll, 2012). Initially discovered in enterococci was the axe-txe gene (Grady and Hayes, 2003). The toxin-antitoxin pair (axe-txe) is encoded by the axe-txe operon (Halvorsen et al., 2011a). The txe toxin is known to inhibit the growth of *E. coli* cells, but this effect can be highly compromised by the presence of axe antitoxin (Halvorsen et al., 2011b). The Txe toxin inhibits protein synthesis and causes cell death because proteins are the backbone of virtually all cells. The mechanism that inhibits protein synthesis involved the cleavage of an mRNA at the first base just next to the start codon (Halvorsen et al., 2011a). Since antitoxins are co-inhibitors of toxin`s activity in the cell, they could be potential drug targets in the future. The following TA systems, ω-ε-ζ, axe-txe and mazEF have been

detected and are said to induce plasmid stabilization in enterococci (Sletvold et al., 2008, Moritz and Hergenrother, 2007, Grady and Hayes, 2003).

Transposons of enterococci

Transposons are transposable elements that are capable of transferring within genomes or between them (Roberts et al., 2008). Transposons are found in all species including humans and are thought to have a variety of functions (Huang et al., 2012). They are abundantly present in enterococci and are said to involve in the dissemination of virulence and resistance factors (Rosvoll et al., 2012). In enterococci they include insertion sequence elements (ISE), Tn3-like transposons, composite transposons and conjugative transposons (Bonafede et al., 1997, Tomita et al., 2003).

Virulence factors of enterococci

Enterococci were initially recognised as commensals of the human gut flora but are causing serious nosocomial infections in humans. This is in large part due to the emergence of virulent strains. Similar to resistance transfer in enterococci, virulence factors can also be transferred through the exchange of genetic elements (Arias et al., 2009). The virulence of every pathogen is its ability to cause disease. Enterococcal virulence can be described in terms of (1)˝adhering to host tissues (2) invading and forming of abscesses (3) impairment of host inflammatory and immune responses and (3) secretion of toxic materials to enhance biofilm formation.˟ (Salah et al., 2008). Previous study showed a relationship between the accumulation of virulence factors and biofilms formation (the process that leads to disease formation such as urinary tract infection (UTI) and sepsis) (Seno et al.,2005). As indicated by Seno et al. (2005*) E. faecalis* which have accumulated virulence are more likely to form reoccurring biofilms.

Table 1: The most common virulence factors of *Enterococci*

Factors	Found in species:	Function	Referance
Esp	*E. faecium*	Involved in biofilm formation	(Heikens et al., 2007)
AS	*E.faecalis*	Adhesion to soft tissues	(Chow et al., 1993)
Hyl(efm)	*E.faecium*	Involved in glycoside hydrolyses	(Arias et al., 2009)
GelE		Hydrolyses gelatine	(Duprè et al., 2003)
Cytolysin	*E. faecalis, E. faecium*	Lyses erythrocytes	(Chow et al., 1993)

Known virulence genes of enterococci are listed in Table 1 and factors encoded by these genes are said to play an important role in the pathogenesis of *E. faecium* and *E. faecalis*. However, enterococci also differ in terms of virulence factors. For example, while *E. faecalis* harbours aggregation substances (AS), they are entirely absent in *E. faecium (*Hällgren et al. (2009). Aggregations substances are said to be involved in pheromone inducible transfer of plasmids among *E .faecalis* and do not play any role in transfer of plasmids within *E. faecium*. This is because *E. faecium* do not produce sex-pheromones and hence, aggregation substances. The study of potential virulence factors could highlight the prospects of future disease management. Knowledge of enterococcus pathogenesis is therefore important as it would lead to the discovery of new target sites for antimicrobials.

HYPOTHESIS AND AIMS

The hypothesis of this study was that invasive *E. faecium* and *E. faecalis* encode virulence factors with immune invasive and host colonization properties which are secreted into the bloodstream during infection and that those genes encoding such factors are carried by transferable megaplasmids and are therefore more prevalent among isolates of blood cultures than those derived from other sources.

The Primary aims were:

- to obtain a circular map of a pLG1 replicon type plasmid from a blood culture isolate of *E. faecium* (K60-39), transconjugant (TUH56-32).
- to detect the presence of and describe the prevalence of selected open reading frames (ORFs) encoded by megaplasmids in a clinical and non-clinical collection of *E. faecium* and *E. faecalis* strains.

The future aims were to study and identify genetic factors that play an important role for the immune invasive properties of hospital associated *E. faecium*.

MATERIALS

Available contigs and further scaffold formation

Before starting to gap close the K60-39/3D megaplasmid, it had already been DNA sequenced and assembled into what was known as contigs. Contigs are contiguous length of sequences of which the order of bases is known. Gaps are formed where two overlapping contigs end. The process involving contigs formation was done and Illumina DNA sequencing technology and mate pair libraries were applied. The method was not applied in this thesis and would therefore not be discussed. The main aim of this technique can be seen in figure I, which is to obtain what contiguous sequences based on overlapping fragment. Once the material was available for gaps closure, all that needed to be done was based on DNA amplification by a thermal cycler, detection of the amplified DNA on an agarose gel, sequencing of the amplified DNA and using PCR fragments in gaps closure. Different bioinformatics software tools were also needed for gaps closure including those mentioned under bioinformatics analyses of sequencing results.

Strain collection:

The total of 151 isolates was collected including 116 *E.facium* and 34 *E.faecalis*. Isolates were grouped according to source as shown in table 2. In addition transconjugant TUH 56-32 originating from conjugation between donor K60-39 and recipient BM4105-RF was included and was also used in the gap closing project. It was provided to by the Reference Centre for Detection of Antimicrobial Resistance (K-res). Moreover strains were collected in twelve different countries including Norway, Poland, Germany, The UK, Portugal, Australia, The United States, Spain, The Nederland, Denmark, Italy and Sweden. The most recent collection happened in 2013 in Norway and the United States. The first collection however took place in the United States in 1981. This means that isolates included were collected over the period of 30 years.

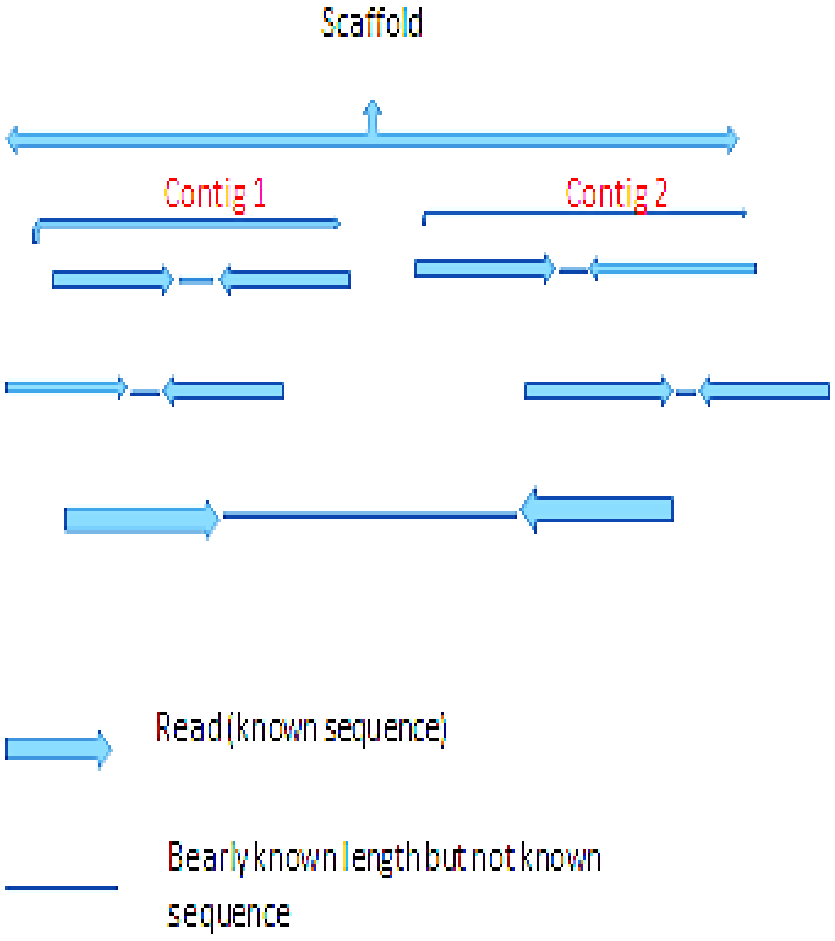


Figure I: Schematic diagram of a scaffold. Contiguous sequences formed contig 1 and contig 2. Between to contiguous sequences are gaps with known length but unknown sequence. Was reprinted from (http://en.wikipedia.org/wiki/Contig) Access date: 16.05.2014

Table 2: The distribution of *E.faecium* and *E.faecalis* strains that were used in this study

Source	*E. faecium*	*E. faecalis*
Human blood	71	27
Other-clinical isolates	30	7
Non-clinical isolates	15	0
Total	116	34

Reference strains:

The reference strains that were used in this study were provided by the Reference Centre for the Detection of Antimicrobial Resistance and included two *E. faecium* strains. One was used as a negative control (TUH 18-06) and the other as a positive control (TUH-56-32). The positive control strain (TUH 56-32) resulted from conjugation between *E. faecium* (K60-39) donor strain and *E. faecium* (*BM4105-RF)* recipient strain and harboured the 3D megaplasmid. Both strains were used as controls for the screening of genes encoding potential virulence factors in the strain collection. Total DNA from TUH 56-32 was also used in the gaps closing project.

METHODS

General bacteria growth condition:

Samples were taken from deep storage at -70 ^{o}C (K-res) and were placed on ice before being transferred on blood agar plates using an inoculation loop. The transfer involved the spreading of small samples on the blood agar plate with an inoculation loop followed by streaking evenly across the entire plate. Once the samples were transferred onto the plates, they were incubated at 37^{0} C for 20 hours. After overnight growth on blood agars, the plates were observed for growth and visual colonies were harvested by using a clean inoculation loop. The colonies were suspended into 2,5 mL of Brain Heart Infusion Broth (BHIB) and incubated at 37 ^{o}C for another 20 hours. The overnight growth mediums were shortly centrifuged at 8000 x g (8000rpm) for 10 min followed by discarding the supernatant. The pelleted bacteria cells were further used for DNA extraction. The extraction involved the use of three different extraction kits according to product manufacturer description.

Total DNA extraction:

Total DNA was extracted from all the strains. The types of extraction Kits that were used included High performance plasmid Midi Kit (Omega, USA), E.Z.N.A Bacterial DNA Kit (Omega, USA) and Qiagen DNA Mini Kit (Qiagen, Germany). The High Performance Plasmid Midi Kit (Hp plasmid Midi Kit, Omega USA) was exclusively used to extract DNA from the trans-conjugant strain used in gap closure and was selected because it has been used to yield clean and quality DNA templates previously. Moreover, total DNA was extracted because it was assumed that most of the strains contained larger plasmids that could not separate from the bacteria chromosomal DNA. The principles for total DNA extraction involved pelleting of overnight cell cultures, lysing of the cell pellet, removal of bacteria proteins and ribonucleic acids (RNA), further washing the DNA template for removal of reagents and final capturing of DNA in a microcentrifuge tube and storage before being used in PCR.

DNA extraction using High Performance plasmid Midi Kit

- To start with 30 mL of overnight cell cultures were transferred in 15 mL falcon tubes and centrifuged at 4000 x g for 10 min and the bacteria pellets were re suspended by vortexing in 2,25 mL of solution I(RNase A) after removal of the supernatant.
- To the suspension 2, 25 mL of solution II and 5 mL of OB protease were added followed by inverting in 7 s and incubation at 5 min.

  3, 2 mL of solution III were so added to the suspension and inverted 10 times
- It was centrifuged at 12000 x g for 10 minutes at 4 ^{0}C to separate the pellet from the supernatant and the supernatant was used further.
- A volume of 1 mL of equilibration buffer was pipetted into a HiBand® DNA Midi column that was pre-inserted into a 50 mL collection tube. It was set to equilibrate in 5 min followed by centrifugation at 3000 x g for 3 min.
- So 3, 75 mL (3 x) of the separate was transferred to the HiBand® DNA Midi column and centrifuged for 5 minutes at 6000 x g. The flow through was discarded.
- The DNA was washed with 3, 5 mL of HB buffer and centrifuged at 6000 x g followed by discarding the flow through.
- The DNA was further washed with DNA wash buffer and 95% ethanol and centrifuged at 6000 x g to discard the flow through.
- The empty HiBand® column was centrifuged for removal of excess ethanol
- The DNA was eluted using 0, 5 mL of DNA elution buffer and centrifugation at 6000 x g to elute the DNA.
- The eluted DNA was stored at -20 ^{o} C until used for PCR

DNA extraction using Qiangen DNA Mini Kit

Protocol D was used for genomic DNA isolation for Gran positive bacteria and included the following steps:

The overnight cultures that are described above were centrifuged for 10min at 8000x g to separate pellet from the supernatant.

- The pellets were re-suspended in 180 µL (20 mg/mL) lysozyme solution and incubated at 37 ^{o} C for 30 min
- A volume 20 µL of proteinase K and 200 µL of Buffer AL were added after incubating as described in product manual
- The mixtures were further incubated at 56^{o} C for 30 min and at 95^{o} for 15 min
- Once incubated they were briefly centrifuged followed by adding 200 µL of 96% ethanol and vortex for few seconds
- The samples were then transferred to a clean QIAamp Spin column, centrifuged for a few seconds again to remove ethanol and further adding 500 µL of buffer AW1 added.
- They were centrifuged at 6000 x g (8000 rpm) to remove buffer AW1 followed by adding another 500 µL of buffer AW2 then centrifuged at 12000 x g (8000rpm) for 6min to remove bufferAW2.
- The QIAamp Spin Columns with bind DNA were inserted into a clean 1,5mL microcentrifuge tube and 200 µL of AE elution buffer were added and set to equilibrate in 3min
- Then columns were centrifuged at 8000 x g (8000 rpm) for 2 min to elute DNA and the eluted DNA was stored at -20^{o} C until used for PCR.

DNA extraction using E.Z.N.A Bacterial DNA Kit

Procedure:
The protocol for bacterial DNA as described by product solution manual was used. Reagents were mixed as recommended on page 4 of this solution manual. The procedure included the following steps:

- To begin with 3 mL of overnight cell culture was centrifuged at 4000 x g for 10 min and aspirated so to discard the supernatant.

- To re-suspend the pellet 100 µL of TE buffer was added and vortex for a few seconds

  10 µL of lysozyme was added and incubated at 30 ^{o} C for 10min

- After incubation 100 µL of BTL buffer and 20 µL of proteinase K were added and vortex thoroughly.

- The samples were incubated at 55^{o} C followed by vortexing after every 25 min

  5 µL of RNase A was added, inverted for a few seconds and incubated at room .temperature for 5 min

- The samples were then centrifuged at 10000 x g for 2 min in order to pellet any undigested materials and transferred to a new 1.5mL microcentrifuge.

  220 µL of DBL buffer were added, vortex and incubated at 65^{o} C for 10 min

- The samples were further transferred into a clean Hibind DNA Mini column with 220 µL 96% ethanol and the column was inserted into a 1, 5 mL microcentrifuge tube and centrifuged at 10000 x g for 2 min in order to discard the flow-through.

- The columns containing band DNA was centrifuged at 10000 x g for another 2 min to remove ethanol and DNA wash buffer.

- The Hibind DNA Mini columns were transferred into 1,5mL microcentrifuge tubes followed by adding 100 mL of pre heated (65^{o} C) elution buffer and set to equilibrate for 3min.

- DNA was eluted following centrifugation at 10000 x g for 2 min and was stored at -20 ^{o} C for use in PCR.

Principles of Polymerase Chain Reaction (PCR)

PCR is an in vitro method used to make copies of a DNA segment and involves an enzymatic reaction to amplify a specific DNA sequence using complimentary oligonucleotide primers, dinucleotide triphosphates (dNTPs) and the DNA polymerase in an optimized buffer. The reaction starts with the binding of primers that hybridize to the opposite single stranded DNA and precedes towards the region of interest (Erlisch, 1989). The DNA polymerase is a thermo stable enzyme that effectively binds at temperatures around 72°C and extends the growing DNA. The reaction volume for PCR varies but sometimes 25µL can be used in PCR. The thermal cycler is a programmable heating block that changes in temperatures for a given number of circles. At 94°C the double stranded DNA denatures into single stranded DNA. This denatured DNA follows the binding of oligonucleotide primers to the single stranded DNA under cooling temperature to about 55^{o} C. The primer binding temperature could also vary depending on the annealing temperature for each primer pair. After primers bind to the single stranded DNA, they are then extended by the work of DNA polymerase which forms a double stranded DNA. The process continues until the number of expected circles is obtained. This process leads to an exponential generation of copies which could result into millions of copies being amplified from a DNA template (Erlisch, 1989).

Some guidelines for a successful PCR reaction are that:

- primers are template specific and can only bind to one specific site on the template DNA,
- Mg^{2+} concentration should be optimized so they do not compromise the binding of primers to the template DNA.
- the right temperature and time should be allocated to each circle in the reaction sequence.

Primer dilution for PCR

The primer stock solution as made by adding appropriate volume of distilled water(K-res, UiT) as recommended by product manufacturer(Eurogentec, Belgium) to obtain a 100 pMol/µL stock solution. The stock solution was further diluted to10 pMol/µL by adding 90 µL of distilled water to 10 µL 100 pMol/L stock solutions. They were stored at - 20 ^{o} C before being use in PCR. 10pMol/L primer solutions were further diluted to 3, 2 pMol/L. To do so, 68 µL of distilled water were added to 32 µL, 10pM/ µL primer solution. The solutions were stored at -20^{o} C for use in cycle sequencing.

PCR for detection of 16SrDNA

The presence and quality of DNA was tested in all the isolates included in the study and involved the use of two specific primers designed to target the 16SrDNA gene. 16SrDNA gene encodes16SrRNA involved in protein synthesis. Therefore the presence of 16SrDNA reflects the fact that bacterial DNA is present in the sample. The presence and quality of DNA was determined by agarose gel electrophoreses. The PCR reaction mixture and program for detection of 16SrDNA are shown in appendix (Table 3, 8) respectively. The PCR was performed in 25 µL of total PCR reaction mix shown in appendix (Table 3).

PCR for screening of genes encoding potential virulence factors

All strains including positive and negative controls were screened for genes encoding potential virulence factors. Conditions for PCR were the same for all the isolates. The master mix and composition for PCR are shown in table 1 under master mix for detection of gene targets. The PCR was performed according to program given in appendix (table 9).

Performing PCR for DNA analyses and gap closure

For gap closure gradient PCR was used for all the samples. The reaction mixture was the same for all the samples as shown in appendix (Table 3) under mix for gap closure. The program for gradient PCR was applied according to appendix (Table 10).

Primer walking for DNA analyses and gap closure

Primer walking was applied in order to primer-walk some specific regions of the K60-39/3D plasmid. The decision was made after suspicion that some regions could have been circular or duplicated. To primer walk three reference primers were designed and were designated by 3D_plasmid-161bF, 3D_plasmid- 160F and 3D_plasmid160F as shown in appendix (Mater table B). To do so 24 µL of highly concentrated DNA (>200 ng/µL) template was used for every PCR reaction. 3 µL of 5 x-buffers, 1µL of primer (50pmol/µL) and 3µL of Big-dye 3.1 were used per PCR reaction.

Long Range PCR for DNA analyses and gap closure

Long range PCRs were run in order to amplify product sizes above in the range between 4 kb and 5kb. This method was used in order amplify unknown product sizes for gap closure. The PCR reaction program is shown in appendix (table 13). Primers are given in appendix (master table B) and the reactions are 3D_49, 3D_46, 3D_42 and 3D_36. Reactions were carried out according to kit description manual and the following reagents and volumes were used:

Reagents

LongRange PCR buffer.	5 µL
dNTPs mix(10mM).	2,5 µL
Primer,A..	2,5 µL
PrimerB	2,5 µL
RNase-free water	24 µL
LongRange PCR enzyme mix	0, 4 µL
Buffer Q	10 µL
Template DNA	2, 5 µL

Principles of DNA sequencing:

The DNA sequencing involved two steps. In the first step direct sequencing of PCR products with BigDye® Terminator v3.1 (Applied Biosystems) was carried out. In the second step however, the precise nucleotide sequences of the DNA fragment were determined by the use of automated DNA sequencing machine. In order to do so, the PCR products were fist purified to remove remaining nucleotides that were added to the growing chain during primer extension. Sequencing reactions were then carried out by PCR (Cycle Sequencing) before the sequences were read and determined by the ABI PRISM 3100 Genetic Analyser (Applied Biosystems). The genetic analyser uses a capillary with liquid polymers which allow smaller molecule to pass rapidly than larger molecules. In each fragment a modified base is incorporated during cycle sequencing and can be detected by a laser which is capable of distinguishing between different fluorescent colours. Each base is assigned its own colour: A green, T red C blue and G black. The detected bases are then registered on the computer which produces a chromatogram showing how the bases are aligned in a sequence as illustrated in Figure II.

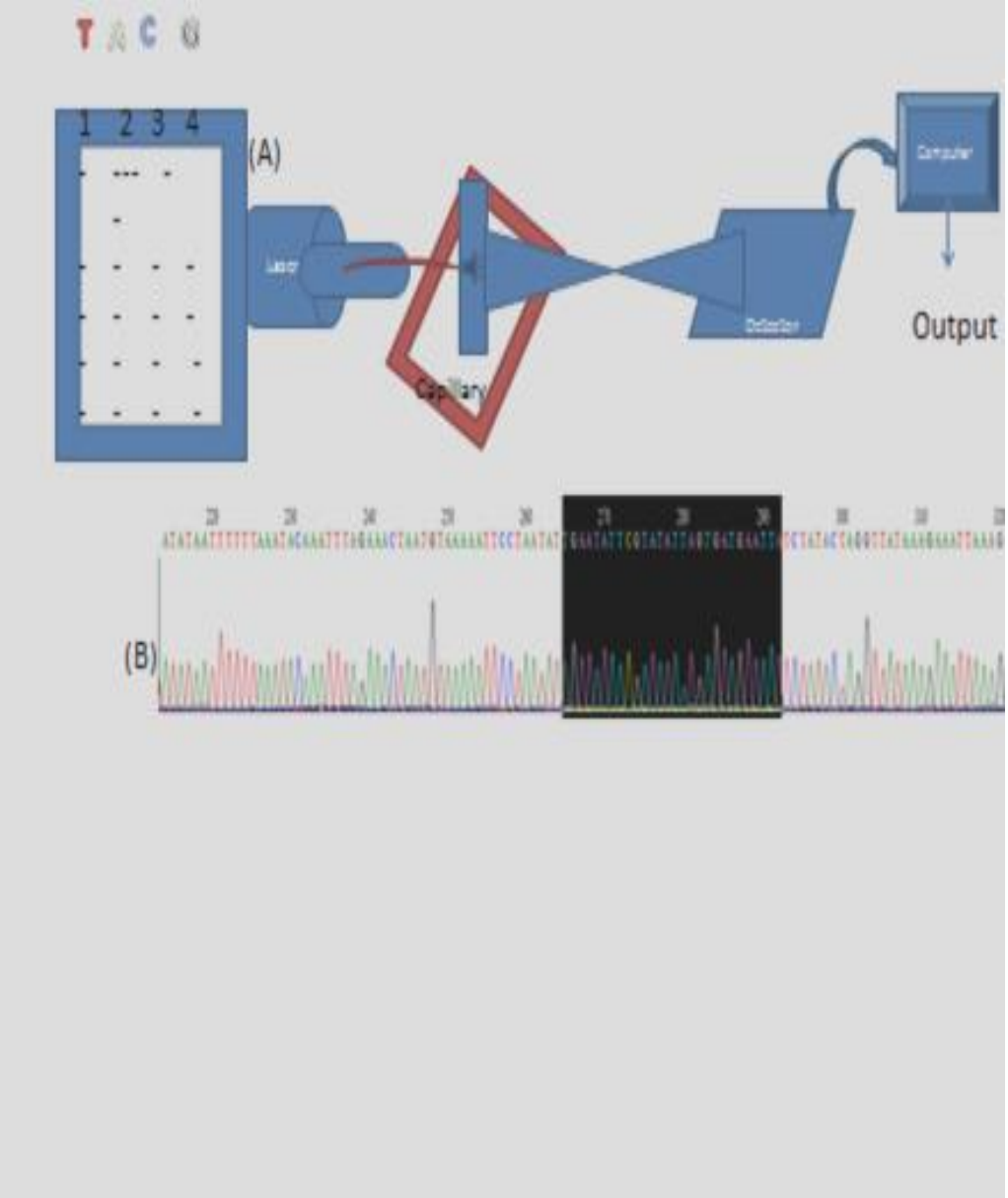


Figure II: The schematic representation of DNA sequencing. In (A the samples are loaded into the capillary and the laser reads the last incorporated base as the DNA fragment migrates through the capillary. In (B the detected bases are recorded by the computer which further displace the information in the form of a chromatogram.

PCR for purifying of PCR products and DNA cycle sequencing

Procedure:

The amplified PCR products were subjected to clean-up for the removal of unused oligonucleotides and dinucleotide triphosphates (dNTPs). To do so the following procedure was applied:

A total volume of 10 µL of PCR product was inserted into PCR reaction tubes for each reaction

In addition 4 µL of Exo-SAP-IT (USB, Cleveland, Ohio) was added to the reaction tubes

The tubes were inserted into the thermal cycler (Applied Biosystems) and run at 37^{o} C for 30 min in order to degrade unused oligonucleotides and dNTPs, then at 80 ^{o} C for another 15 min to deactivate ExoSap-IT.

The clean-up PCR products were amplified in the thermal cycler using one primer per PCR (Applied Biosystem) before being used for sequence detection.

Reaction mixture for sequencing:

13 µL of dH_2O

3 µL of 5x buffer

1 µL of BigDye 3, 1 (table 4)

1 µL (3,2pMol/L), primer

2 µL purified template

The PCR reaction was run according to program given in appendix (Table 11).

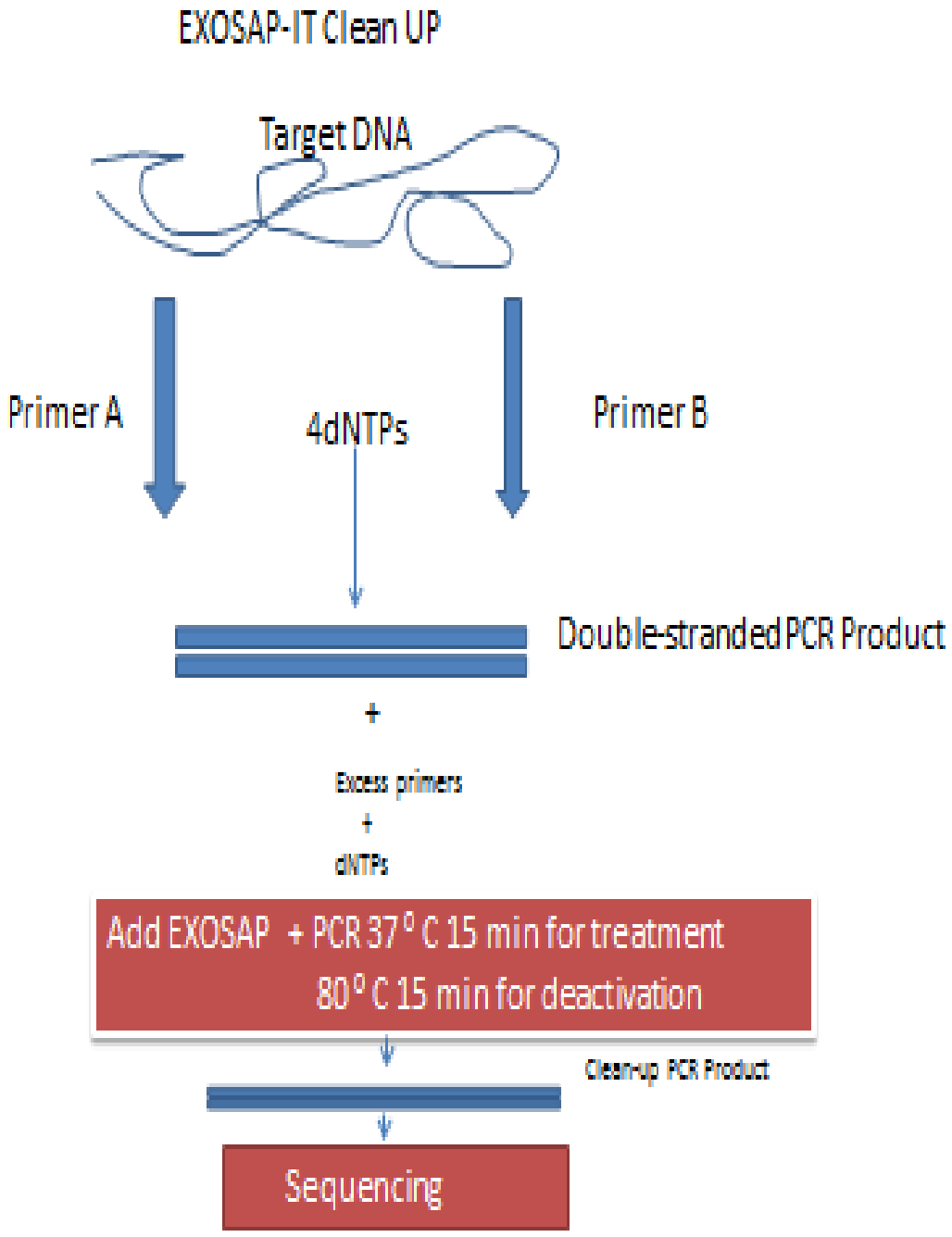


Figure III Schematic diagram of the EXOSAP-IT PCR clean-up method

Principles of agarose gel electrophoreses

The agarose gel electrophoresis separates molecules based on size and is used to separate DNA molecules. In an agarose gel fragments that are larger move slowly while smaller fragments move faster. The movement of molecules is determined by their ability to penetrate the agarose gel, which is made of macromolecules that are tightly packed together forming a network. The smaller the molecule is, the easier it penetrates and larger the molecule is, the slower it penetrates. An agarose gel electrophoresis involves the preparation of agarose gel, the insertion of gel into a gel chamber, insertion of samples and application of electric power to separate the DNA. Because DNA is negatively charged separation follows the movement of DNA from the negatively charged electrode towards the positively charged electrode. This movement in an electric field causes smaller molecules to move faster than larger molecules. Therefore the agarose gel could be used to determine sizes of molecules. A molecular ladder known as DNA ladder is used to compare and determine the size of unknown DNA fragments on the agarose gel. Agarose gel electrophoreses is easy, quicker and highly reproducible and above all time saving.

Procedure:

The agarose gel electrophoresis was widely used in order to detect positive PCR products. It was also used in detection of 16SrDNA so as to verify whether DNA was present in all the samples. The agarose gel electrophoreses included the following procedures:

- A total of 1g agarose (Seakem LE agarose, Lenza, Rochland, USA) was dissolved in 100 mL 0, 5* TAE buffer (K-res, UiT), by boiling in the microwave for 2 minutes
- The solution was cooled down for a few minutes followed by adding 5µL of GelRed nucleic acid stain (Biotium, Hayward).
- It was further poured into a gel chamber containing a comb and set to cool for about 20 minutes.

- A volume of 5 µL*2 of 1 Kb+ DNA ladder (Biotium, Hayward) were loaded into the gel wells at both end of the of agarose gel for each reaction and 5 µL of each PCR product was loaded into the gel wells
- The agarose gel electrophoreses was run at 124 V for about 45 minutes and the DNA bands were visualized using GeneSnap (Teble 4)

Bioinformatics analyses of sequencing results.

Sequence analyses were done using various computer software programmes. Artemis was used for visualization of sequence features. Artemis Comparison Tool (ACT) was used to pairwise compare two or more sequences. EBiox was used to align sequence reads against the plasmid. Algorithm-based automatic contiguation of assembled sequence (Abacas) were used for recording sequences and BioEdit sequence alignment editor was used in order to verify sequence identity.

Statistical analyses

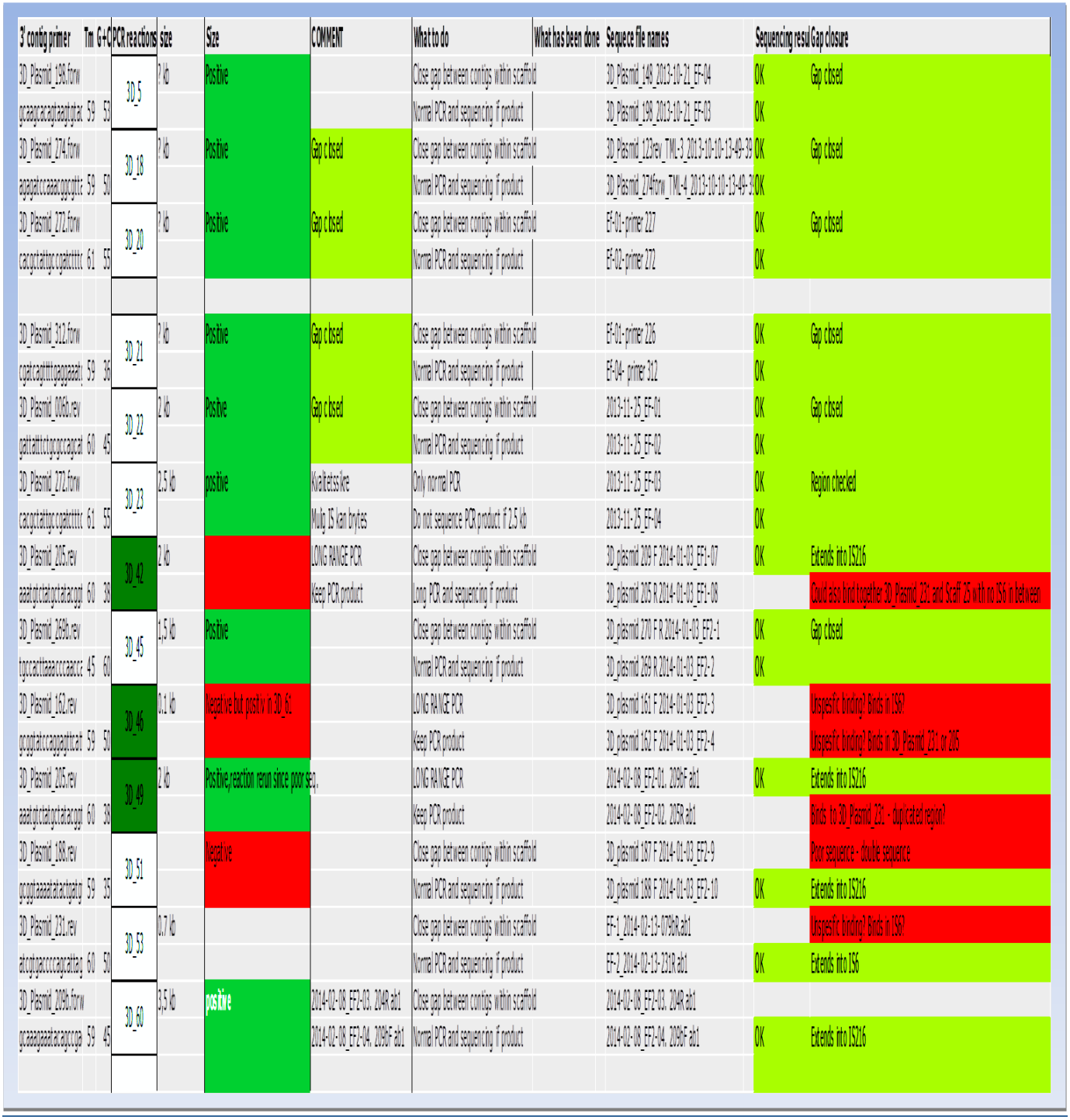
The correlation between the presence of all tested ORFs and the clinical source of isolates among the two populations as well as between them, was determined in a contingency table of Fisher`s exact test. A $p < 0, 05$ was considered significant to show that there is an association between the various sets of variables.

RESULTS

Gap closure of trans-conjugant (TUH56-32) megaplasmid

The DNA sequence of the 300 kb megaplasmid detected in blood culture isolate of *E. faecium* (K60-39), transconjugant (TUH 56-39/3D) had been obtained using next generation sequencing. The selected ORFs that were included in the study were present in this plasmid DNA sequence. To verify the linkage to the plasmid and to enable further studies of the plasmid, gaps closure of the plasmid DNA to obtain a circle DNA molecule was carried out by PCR analyses followed by DNA sequencing of PCR products.
The Illumina DNA sequencing of the TUH 56-39/3D plasmid showed that it is composed of 286 kb, given in 18 scaffolds with 26 gaps and more than 60 IS (Appendix, Table 6). It was primarily selected because it did test positive to all four open reading frames that were screened in the screening project. Therefore it was used as positive control in addition to gap closure.

The 286 kb DNA sequence was subjected to gap closure using 66 specific primer pairs as shown in appendix (Mater table B). All were used in PCR and 13 PCR positive reactions were obtained. In addition positive PCR products were further subjected to automated DNA sequencing and the sequences were again analysed and used in gaps closure. The gaps closing results can be seen in the figures V-VII. Figures were taken from the sequence assembly in Artemis and ACT, software programs that were used in the study for sequence comparison and alignment. Of the 13 PCR positive reactions, 7 were used in gaps closure. The remaining PCR products did not perfectly fit into their respective gaps as Figure VII indicates and were therefore excluded from the gaps closure.



3' contig primer	Tm	G+C	PCR reactions	size	Size	COMMENT	What to do	What has been done	Sequence file names	Sequencing resu	Gap closure
3D_Plasmid_198.forw			3D_5	? kb	Positive		Close gap between contigs within scaffold		3D_Plasmid_148_2013-10-21_EF-04	OK	Gap closed
gcaagcacagtcaatgcta	59	55					Normal PCR and sequencing if product		3D_Plasmid_198_2013-10-21_EF-03	OK	
3D_Plasmid_274.forw			3D_18	? kb	Positive	Gap closed	Close gap between contigs within scaffold		3D_Plasmid_123rev_TML-3_2013-10-10-13-49-39	OK	Gap closed
agagatccaaacagcgttcc	59	50					Normal PCR and sequencing if product		3D_Plasmid_274forw_TML-4_2013-10-10-13-49-3	OK	
3D_Plasmid_272.forw			3D_20	? kb	Positive	Gap closed	Close gap between contigs within scaffold		EF-01- primer 227	OK	Gap closed
cacgctattgccgtatttttc	61	55					Normal PCR and sequencing if product		EF-02- primer 272	OK	
3D_Plasmid_312.forw			3D_21	? kb	Positive	Gap closed	Close gap between contigs within scaffold		EF-01- primer 226	OK	Gap closed
cgatcagttttcgcgaaaac	59	36					Normal PCR and sequencing if product		EF-04- primer 312	OK	
3D_Plasmid_006b.rev			3D_22	? kb	Positive	Gap closed	Close gap between contigs within scaffold		2013-11-15_EF-01	OK	Gap closed
gattatttctgcgccagcat	60	45					Normal PCR and sequencing if product		2013-11-15_EF-02	OK	
3D_Plasmid_272.forw			3D_23	2.5 kb	positive	Kraftessive	Only normal PCR		2013-11-15_EF-03	OK	Region checked
cacgctattgccgtatttttc	61	55				Multp IS Kan bnytes	Do not sequence PCR product if 2.5 kb		2013-11-15_EF-04	OK	
3D_Plasmid_205.rev			3D_42	? kb		LONG RANGE PCR	Close gap between contigs within scaffold		3D_plasmid 209 F 2014-01-03_EF1-07	OK	Extends into IS216
aaatgctcatagctatcagcgg	60	39				Keep PCR product	Long PCR and sequencing if product		3D_plasmid 205 R 2014-01-03_EF1-08		Could also bind together 3D_Plasmid_231 and Scaff 25 with no IS6 in between
3D_Plasmid_269b.rev			3D_45	1.5 kb	Positive		Close gap between contigs within scaffold		3D_plasmid 270 F R 2014-01-03_EF2-1	OK	Gap closed
tgccactttaaaccccacca	45	60					Normal PCR and sequencing if product		3D_plasmid 269 R 2014-01-03_EF2-2	OK	
3D_Plasmid_162.rev			3D_46	0.1 kb	Negative but positiv in 3D_61		LONG RANGE PCR		3D_plasmid 161 F 2014-01-03_EF2-3		Unspesific binding? Binds in IS6?
gcgatatccaggagttcatt	59	50					Keep PCR product		3D_plasmid 162 F 2014-01-03_EF2-4		Unspesific binding? Binds in 3D_Plasmid_231 or 205
3D_Plasmid_205.rev			3D_49	? kb	Positive,reaction rerun since poor seq.		LONG RANGE PCR		2014-02-08_EF2-01. 205bF ab1	OK	Extends into IS216
aaatgctcatagctatcagcgg	60	39					Keep PCR product		2014-02-08_EF2-02. 205R ab1		Binds to 3D_Plasmid_231. - duplicated region?
3D_Plasmid_168.rev			3D_51		Negative		Close gap between contigs within scaffold		3D_plasmid 167 F 2014-01-03_EF2-9		Poor sequence - double sequence
gcgatcaaaatcatcatcatg	59	35					Normal PCR and sequencing if product		3D_plasmid 168 F 2014-01-03_EF2-10	OK	Extends into IS216
3D_Plasmid_231.rev			3D_53	0.7 kb			Close gap between contigs within scaffold		EF-1 2014-02-13-0736R ab1		Unspesific binding? Binds in IS6?
atcatgaaccccagcattacg	60	50					Normal PCR and sequencing if product		EF-2 2014-02-13-2212R ab1	OK	Extends into IS6
3D_Plasmid_209b.forw			3D_60	3.5 kb	positive	2014-02-08_EF2-03. 204R ab1	Close gap between contigs within scaffold		2014-02-08_EF2-03. 204R ab1		
gcaaagaaaatcaagccgg	59	45				2014-02-08_EF2-04. 209bF ab1	Normal PCR and sequencing if product		2014-02-08_EF2-04. 209bF ab1	OK	Extends into IS216

Figure V: **F**igure showed positive PCR reactions that were obtained. Reactions that led to gap closure are shown in green far right. The area below mixed in red and light green shows that the primer combinations were PCR positive but sequences were poor to be used in gaps closure.

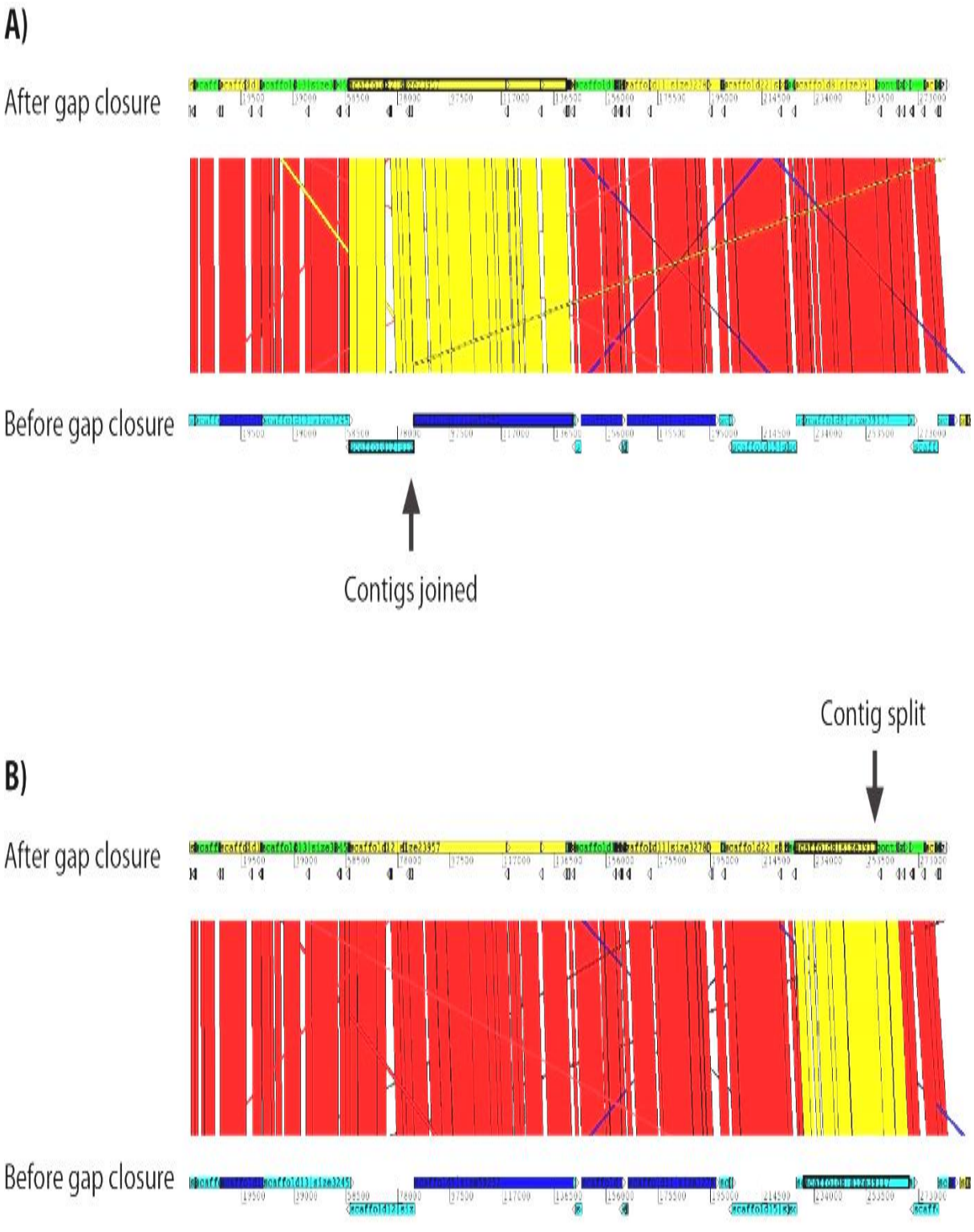


Figure VI: The blue and cyan bars in A and B show contigs of the old sequence before gap closure, while the yellow and green are the contigs of the new sequence after gap closure. The yellow lines between the old and new sequence show that the sequences are similar. In A) the arrows pointing upwards shows where to contigs were joined and the gap closed. In B) the arrow pointing downwards shows where a contig was split because of misassembling.

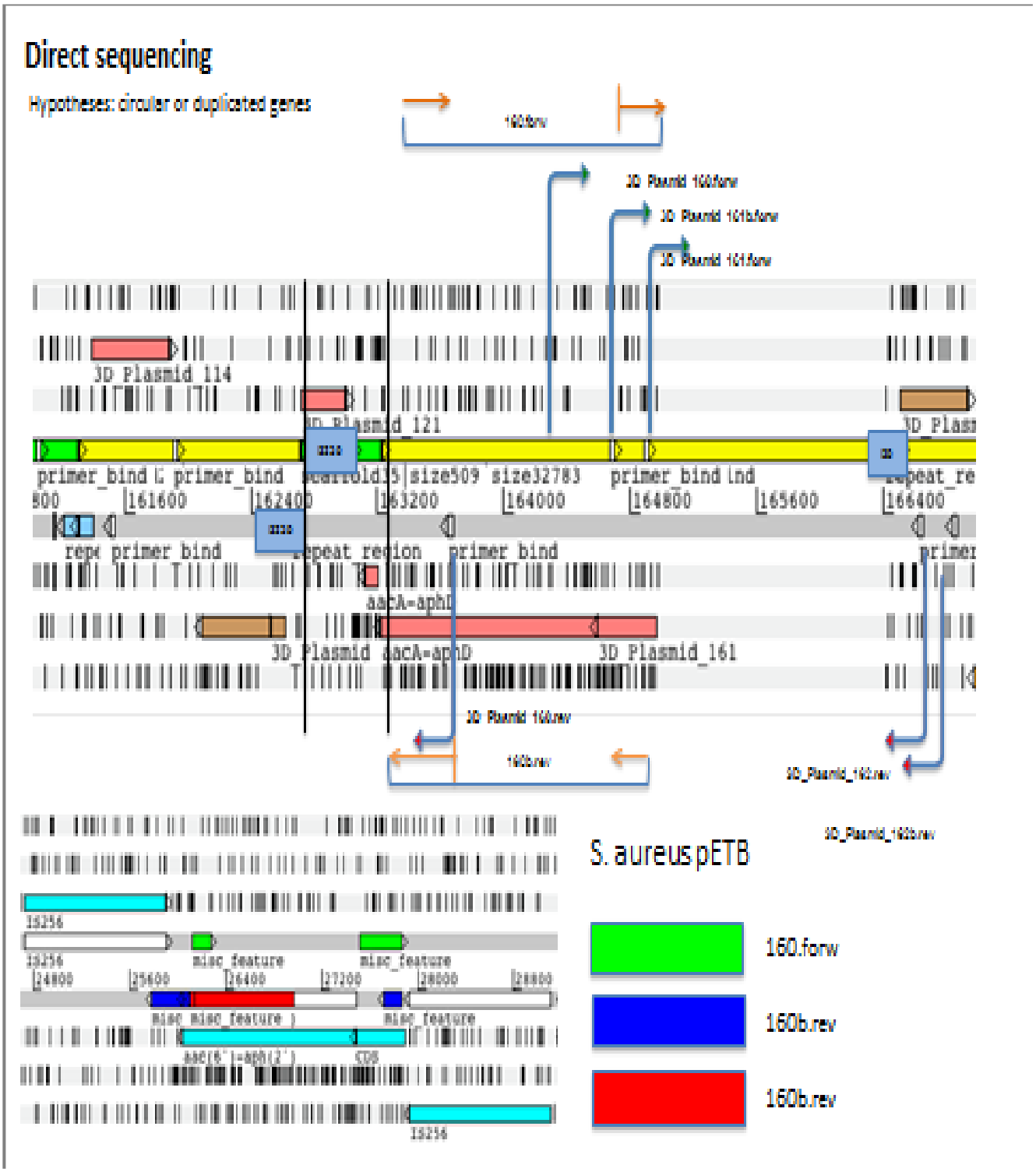


Figure VII: The green, blue and red arrows below indicate targets that are present in *Staphylococcus aureus*, but identical to those found in *E. faecium*. Insertion sequence elements are shown in light blue and light blue arrows are pointing in direction of primers. Shown in salmon are regions covering the aacA-aphD gene and in brown showing opposite genes to taacA-aphD.

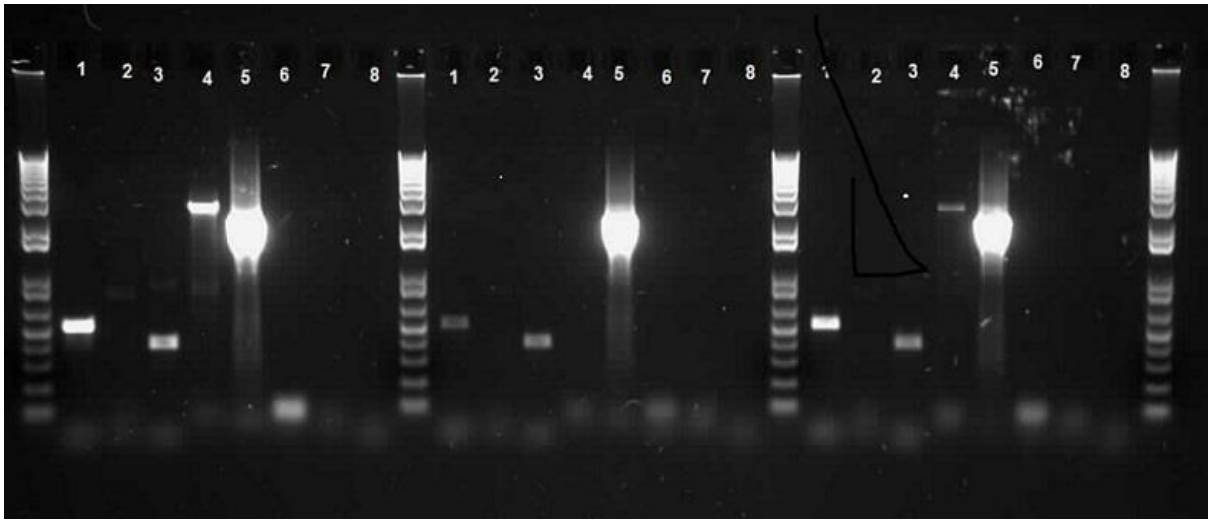


Figure VIII: An agarose gel electrophoreses result of the amplified plasmid DNA in transconjugant TUH56-32. Primers that were used included: 1. 161bF/162R, 3. 187F/188R, 4. 204R/209bF, 5. 209bF/205R. The results 1 through 8 were obtained using gradient PCR.

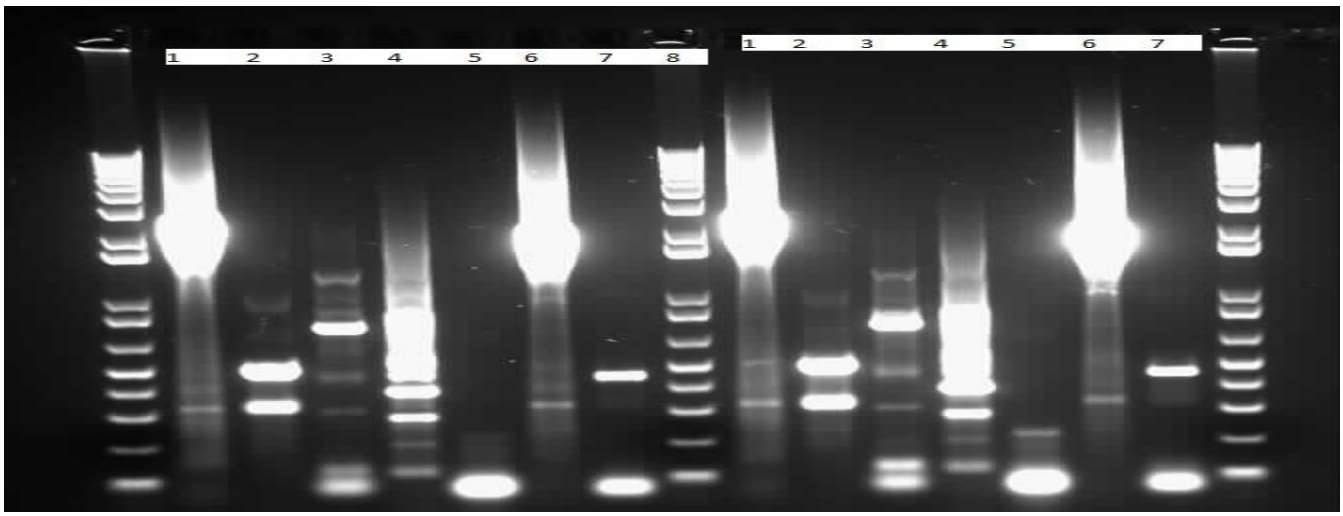


Figure IX: An agarose gel electrophoreses result of the amplified plasmid DNA in transconjugant TUH56-32. Primers that were used included: 1. 161bF/162R, 3. 187F/188R, 5. 209bF/205R. The results 1 through 8 were obtained using gradient PCR.

Screening of the 150 isolates

The ORFs harbored by a megaplasmid from a blood culture isolate of *E. faecium* and encoding hypothetical proteins with a signal peptide for secretion and with unknown function were selected for prevalence analyses in an enterococcal strain collection. The prevalence of four ORFs in the strain collection was investigated by PCR analyses. Screening results for the detection of genes encoding potential virulence factors in blood cultures, other clinical and non-clinical isolates of *E. faecium* and *E. faecalis* are presented in Table A in appendix and in figure 2-14 below.

Distribution and sources of isolates

The strain collection consisted of 150 isolates including 116 *E. faecium* and 34 *E. faecalis (*Figure 1). All were screened for presence of ORFs using specific primers for each ORF as given in appendix (Table 4). Isolates were assigned into three different categories and included isolates of blood cultures, other clinical isolates and non-clinical isolates. Other clinical isolates included those of human wounds, urine and human faeces from hospital outbreak. Non-clinical isolates however were derived from faeces community outbreak of enterococci. As shown in Figure 1 none of *E. faecalis* isolates were derived from non-clinical sources. Blood culture isolates were over represented among *E. faecium* strains.

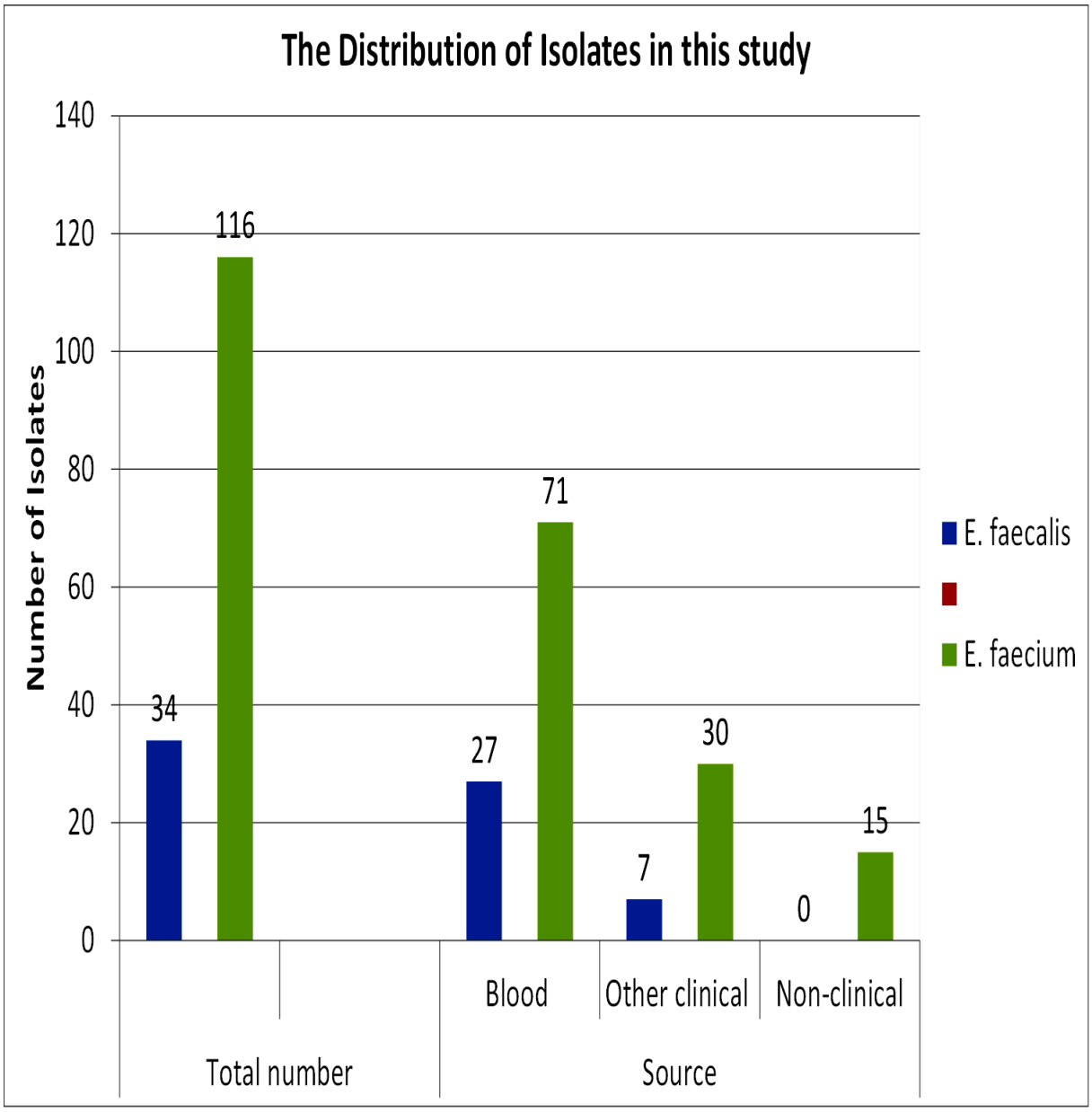


Figure 1: The distribution of isolates that were used in this study.

Prevalence of ORFs in *E. faecium* compared to *E. faecalis*

The distribution of positive and negative test results among *E. faecium* and *E. faecalis* is given in Figure 2. The PCR test results presented show that 23(68%) out of 34 *E. faecalis* isolates were negative to all four ORFs compare to 14(12%) out of all 116 *E. faecium* (p> 0, 0001). In addition 102(87%) out of 116 *E. faecium* isolates tested positive to at least one ORFs, whereas only 11 (32%) out of 34 *E. faecalis* harbored one or more ORFs (P< 0, 0087).

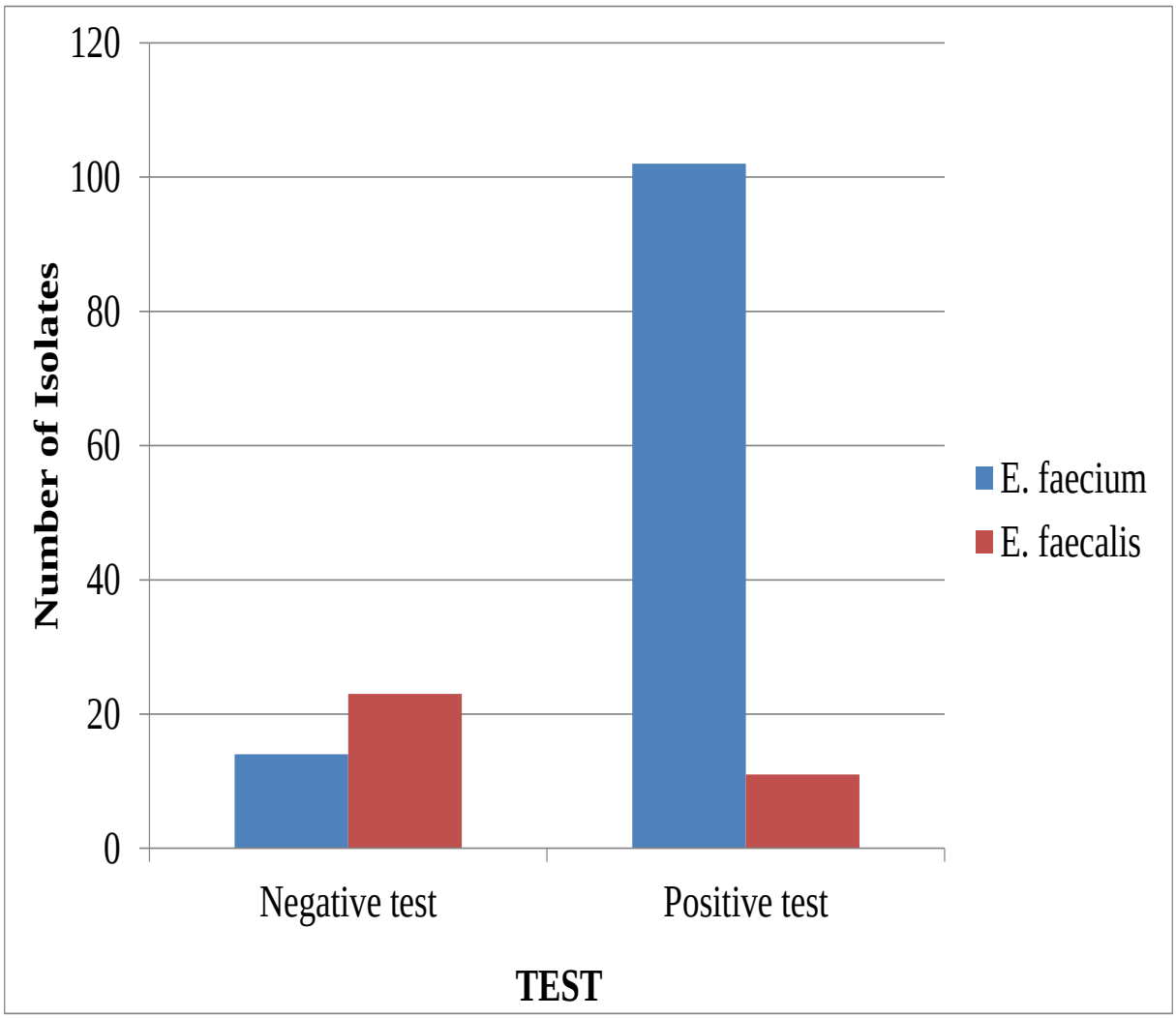


Figure 2: The distribution of positive and negative PCR test results. The negative test here means that the test was negative to all four gene targets. And positive test means that the test was positive to at least one or more targets.

Figure 3 compares the prevalence of ORFs detected among *E. faecium* and *E. faecalis* strains. The test results show that ORFs were detected among *E. faecium* and *E. faecalis* isolates that were included, although the frequency of ORFs were not the same in each group. Compared to *E. faecalis*, ORF147, ORF117 and ORF118 were highly present among *E. faecium* isolates. The two species differed significantly in prevalence to all ORFs as given in Table 14. The most prevalently detected among *E. faecium* was ORF023; while the most possessed ORFs among *E. faecalis* were ORF117 and ORF118. A significantly low prevalence was detected for ORF023 and ORF 147 among *E. faecalis* isolates.

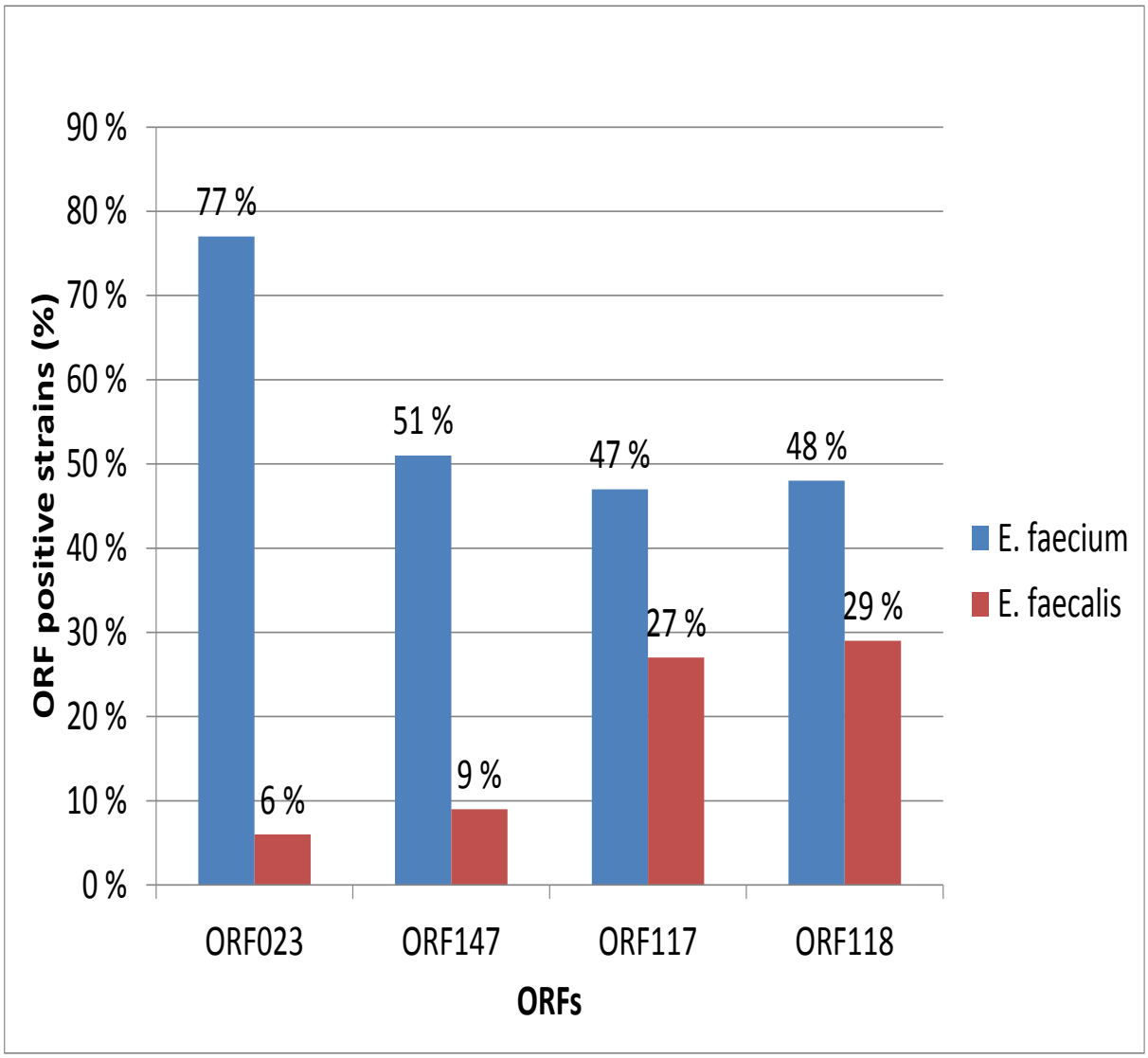


Figure 3 comparison of the different ORFs between *E. faecium* and *E. faecalis*

Table 14: The statistical significance for presence and absence of ORFs between isolates of *E. faecium* and those of *E. faecalis*

ORFs	Positive *E. faecium*	Negative *E. faecium*	Positive *E. faecalis*	Negative *E.faecalis*	P-value
ORF023	89	27	2	32	0,0001
ORF147	59	57	3	31	0,0001
ORF117	54	62	9	25	0,0479
ORF118	56	60	10	24	0,0761

Distribution of ORFs in *E. faecium*

To investigate the distribution of ORFs among *E. faecium* isolates, the number of isolates that were PCR positive was calculated for each ORF. Figure 4 presents the percentage distribution of ORFs among *E. faecium*. None of the isolates from non-clinical cultures were harboring ORF147, ORF117 and ORF118. The tested ORFs were overrepresented among isolates derived from blood cultures, but also highly present in other clinical isolates. Figures 12-13 show the statistical significance between isolates of blood cultures and those of other clinical and non- clinical sources. The most prevalently detected was ORF023 and was highly present in isolates of all sources.

Significant difference was revealed between isolates of blood cultures and those of other clinical isolates (Table 12) with regard to ORF147 (P<0, 0003), ORF117 (0, 0, 0097) and ORF118 (P<0, 0046). However no significant difference in prevalence between isolates of blood cultures and those of other clinical and non-clinical cultures in term of ORF023 was observed. The p values were p: 0, 4403 for isolates of other clinical and p: 1, 0000 for isolates of non-clinical background. A statistical significance was however revealed between isolates of blood culture and those of non-clinical as regard to ORF147, ORF117 and ORF118 as presented in Table 13.

Table 12: The statistical significance for presence and absence of ORFs between isolates of blood cultures and those of other clinical isolates

ORFs	Positive BC	Negative BC	Positive OC	Negative OC	P-value
ORF023	53	18	25	5	0,4403
ORF147	50	21	9	21	0,0003
ORF117	44	27	10	20	0,0097
ORF118	46	25	10	20	0,0046

BC: Isolates of Blood culture, OC: Isolates of other clinicals

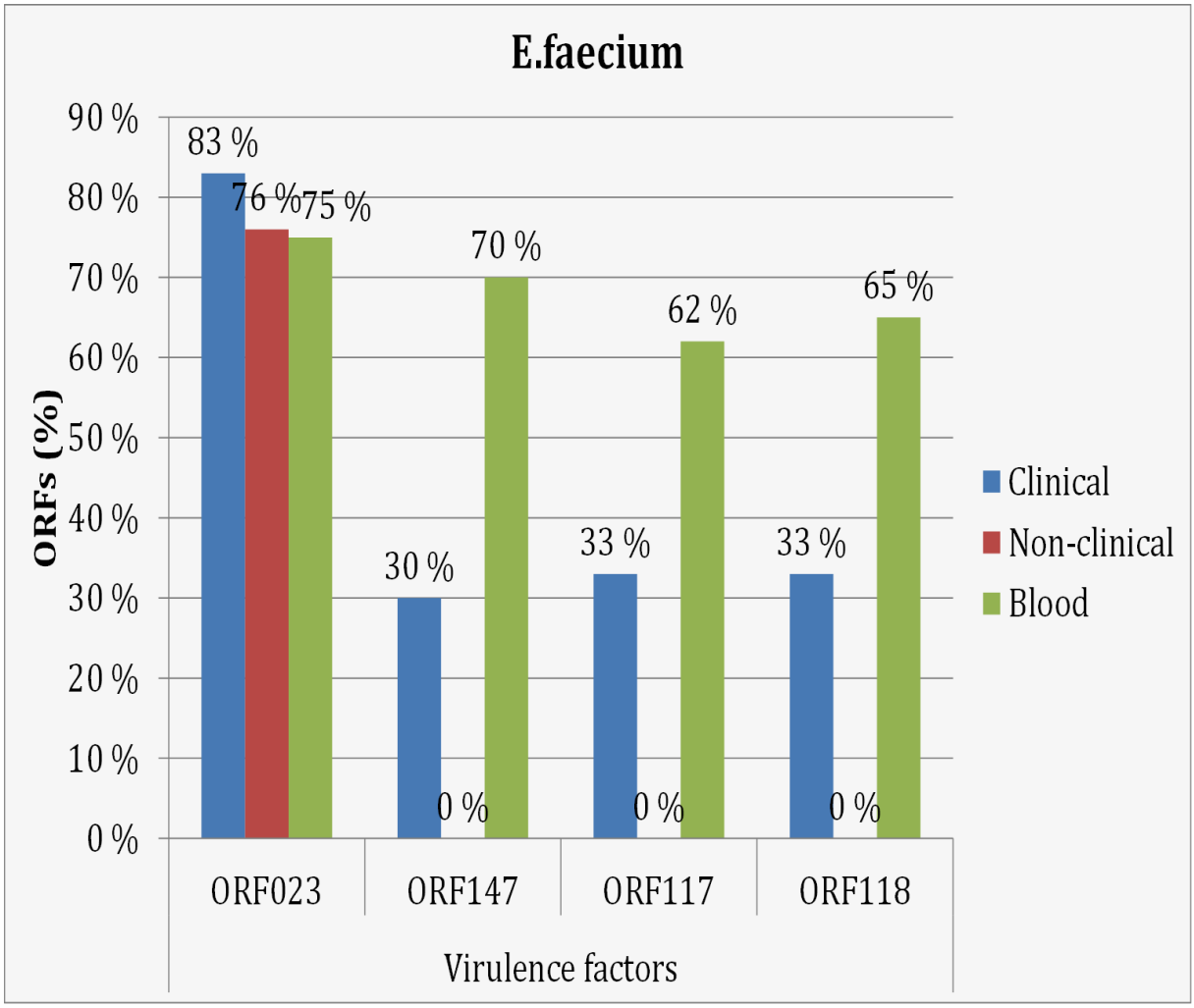


Figure 4: The comparison and prevalence of ORFs among *E. faecium* blood culture isolates (green bars) other clinical (blue bars) and non-clinical isolates (red bars)

Table 13: The statistical significance for presence and absence of ORFs between isolates of blood cultures and those of non- clinical isolates

ORFs	Positive BC	Negative BC	Positive NC	Negative NC	P-value
ORF023	53	18	11	4	1,0000
ORF147	50	21	0	15	0,0001
ORF117	44	27	0	15	0,0001
ORF118	46	25	0	15	0,0001

BC: Isolates of blood cultures: NC: Isolates of non-clinical

Targets distribution among *E. faecium* isolates

To better understand how these genes targets were distributed among *E. faecium* isolates the PCR positive results were analyzed in terms of number of ORFs detected in each isolate and the numbers of ORFs are shown in Figure 5 for the three different groups of *E. faecium*. The results show that 36 (51%) of total 71 blood culture isolates were harboring all four ORFs while only 4 (13%) out of total 30 other clinical isolates tested positive to all four ORFs. ORF023 was detected in 11 (73%) of total 15 non-clinical isolates, but none were harboring ORF147, ORF117 and ORF118. A high number of negative samples to all the tested ORFs were observed among blood culture isolates. Significant difference was revealed between blood cultures isolates and those of other clinical isolates when considering the number of isolates harboring all four ORFs ($p<0,0004$).

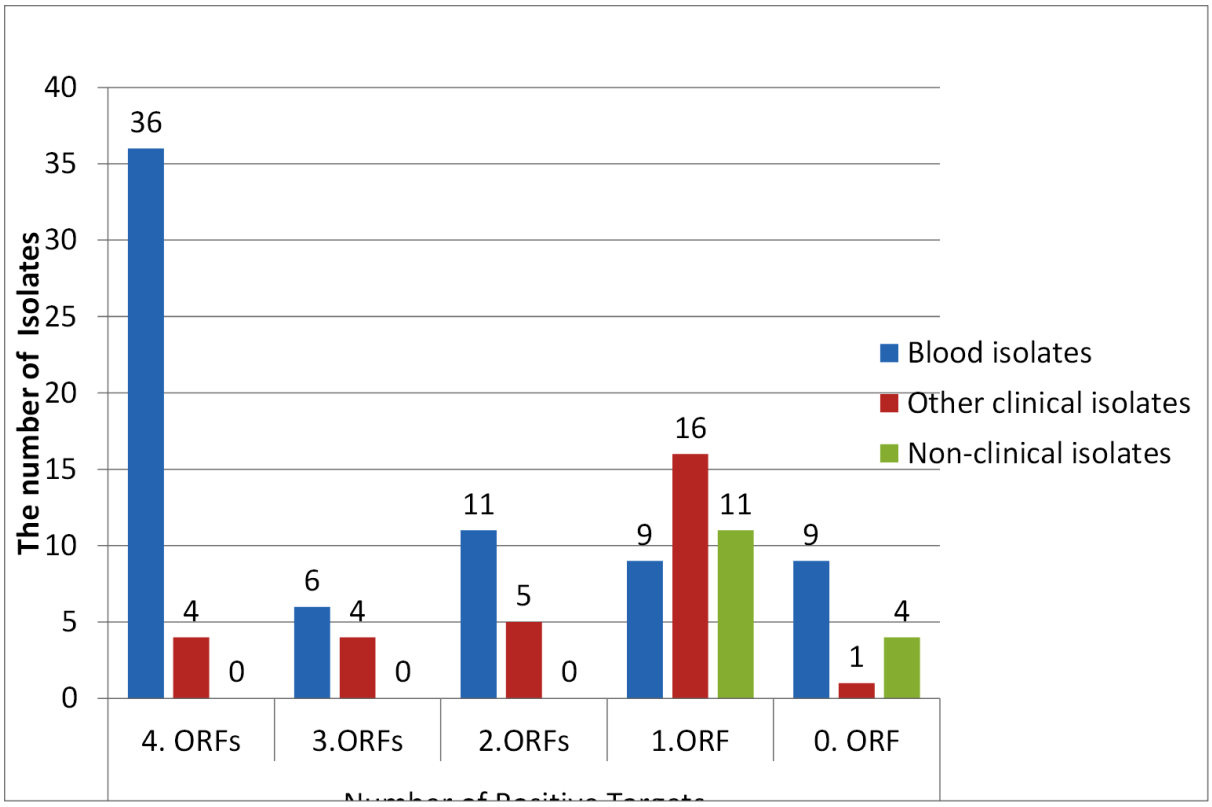


Figure 5: The number of *E. faecium* strains that were harboring 0 ORF, 1 ORF, 2 ORFs 3 ORFs and 4 ORFs**.**

Targets distribution among *E. faecalis* isolates

The presence and number of ORFs per isolate in each group of *E. faecalis* are illustrated in Figure 6. The numbers show that 19 (56%) out of total 27 blood culture isolates did not detect positive to any of the ORFs. The same show that 4 (57%) out of total 7 other clinical isolates were negative to all ORFs. One isolate among other clinical isolates was detected for all four ORFs, but none of the blood culture isolates tested positive to all four targets. Also 2 (29%) out of total 7 other clinical isolates were positive to three of the ORFs.

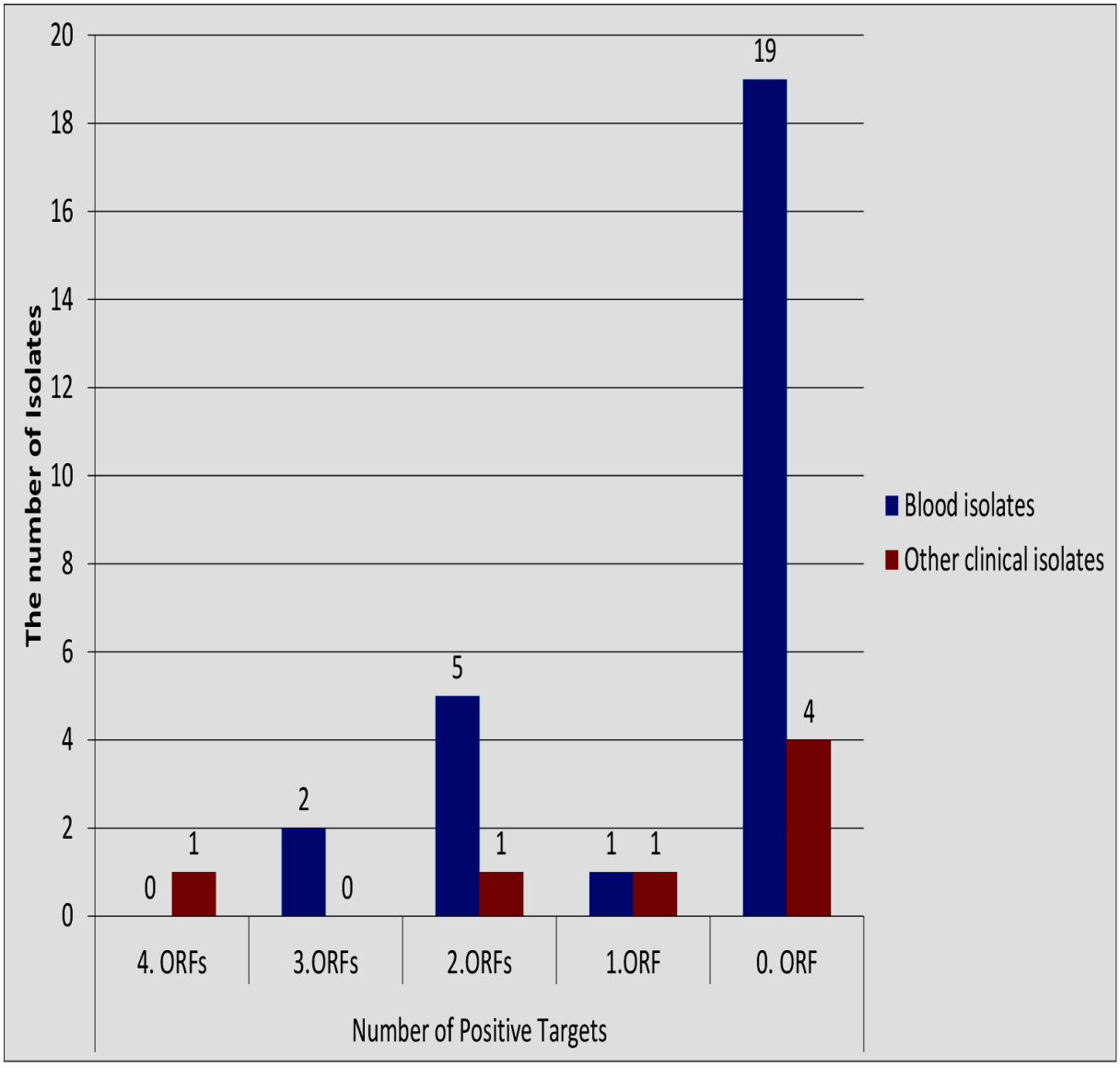


Figure 6: The number of *E. faecalis* strains that were harboring 0 ORF, 1 ORF, 2 ORFs 3 ORFs and 4 ORFs

The agarose gel electrophoreses for detection of ORFs

The agarose gel electrophoreses results are given for a representative collection of isolates that were screened for presence of all four ORFs. Figure 7 shows results over all CC17 and CC 2 strains that were included in the study. It was obtained for the second time to verify whether the previously tested CC 17 and CC 2 strains that were positive actually could test positive in the verification test. The positive control was positive and negative control also negative throughout the test. The fidelity of the control samples was very crucial as it allowed for better judgment as to whether the tested samples were negative or positive. Figures 8-11 were obtained after an agarose gel electrophoreses and showed post PCR reaction products for ORF023, ORF147, ORF118 and ORF117 respectively.

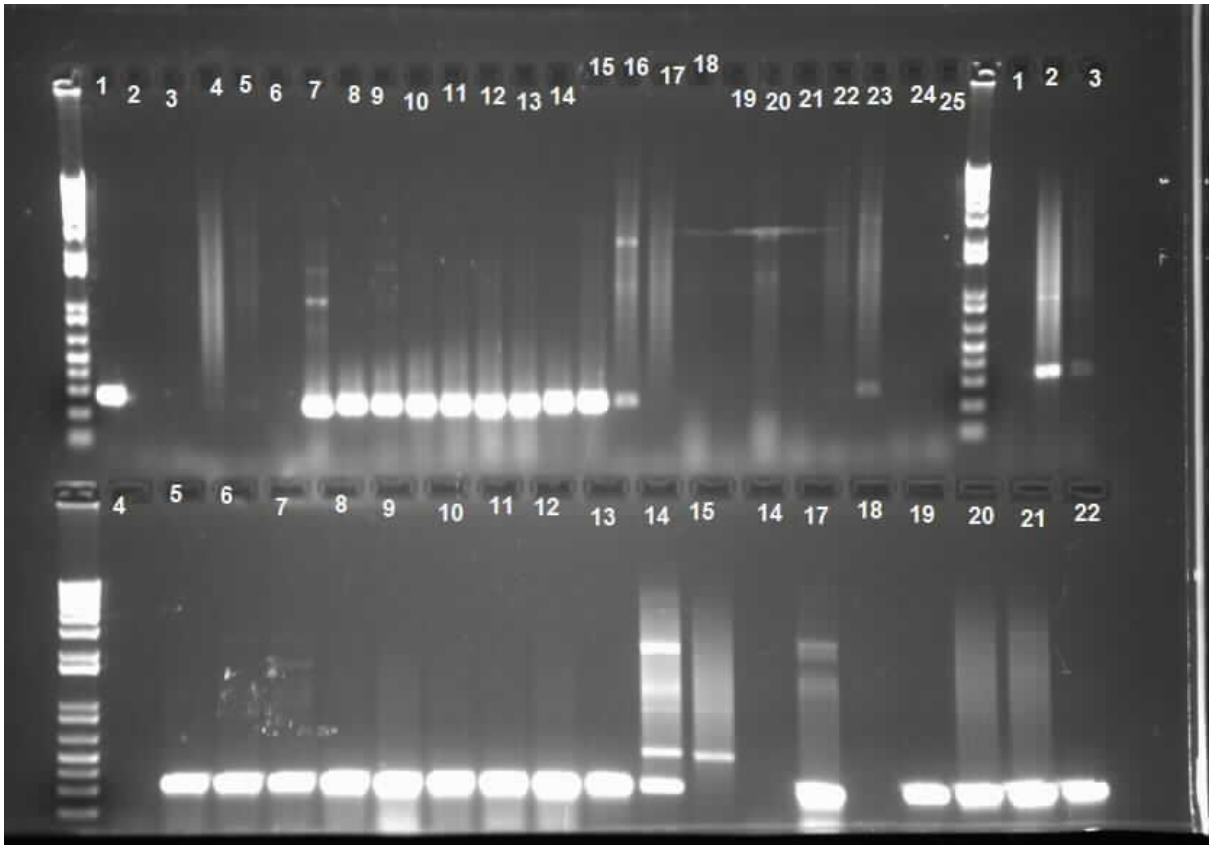


Figure 7: An agarose gel electrophoreses results for genes encoding potential virulence factors in clonal complex 17 and 2(mater table A) strains that were included in this study. Targets were ORF147, above and ORF118 below. Samples 1, 2 above and far left

were positive and negative controls respectively. The agarose gel was re-run in order to verify that previously tested clonal complexes were actually positive or negative for the tested targets

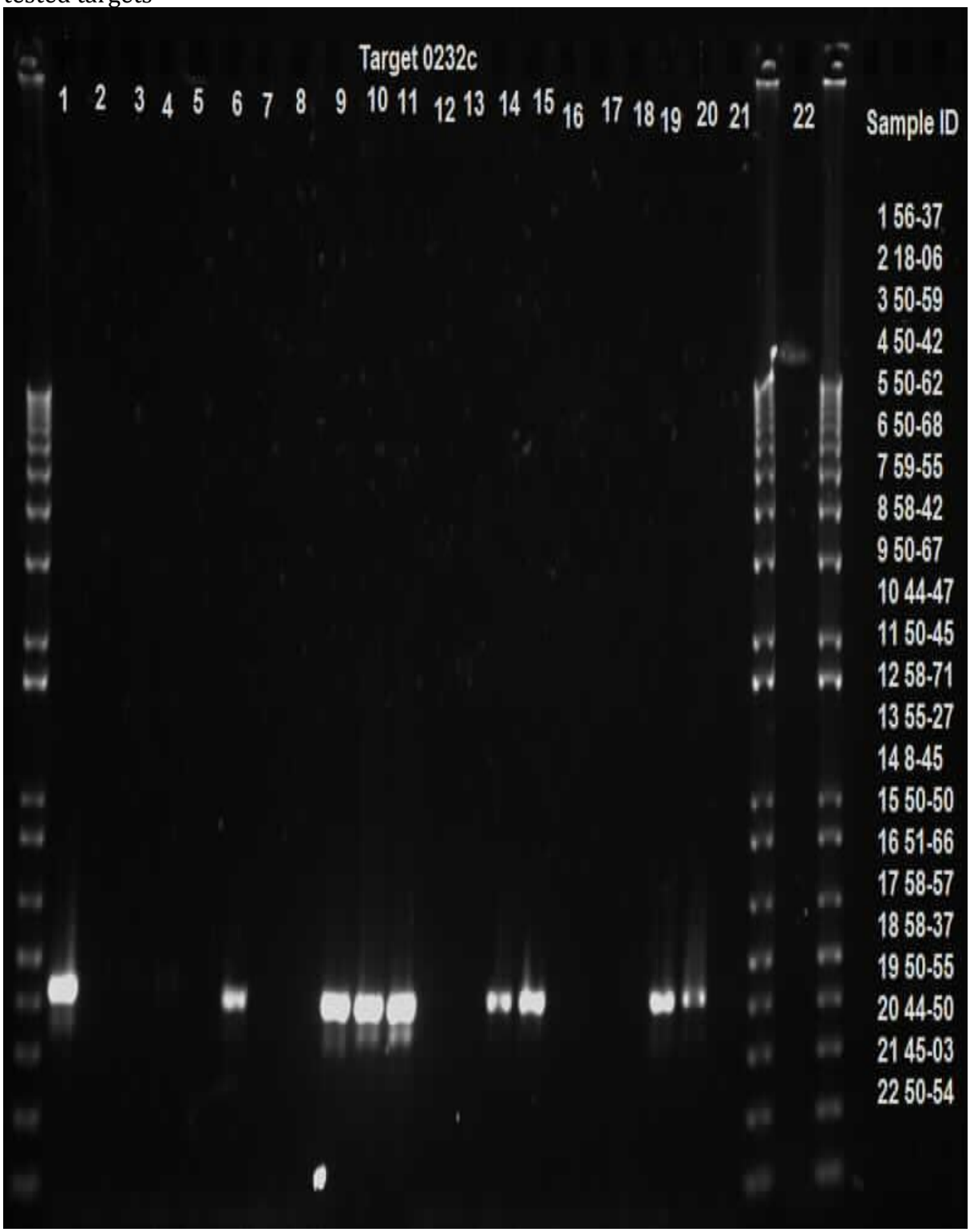


Figure 8: An agarose gel electrophoreses results of ORF023. Sample 1, 2 were positive and negative controls respectively.

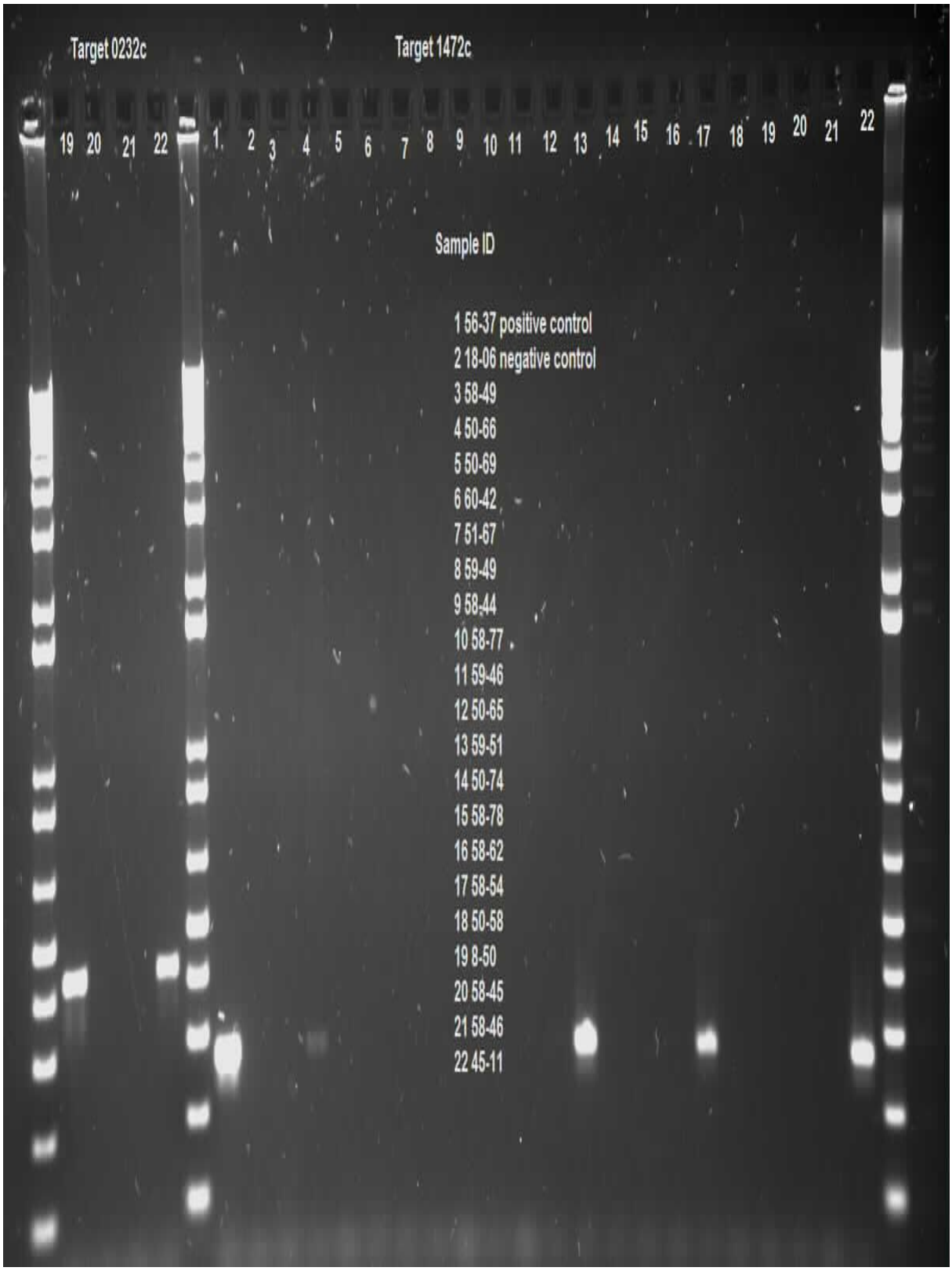


Figure 9: An agarose gel electrophoreses results ORF147. Sample 1, 2 were positive and negative controls respectively

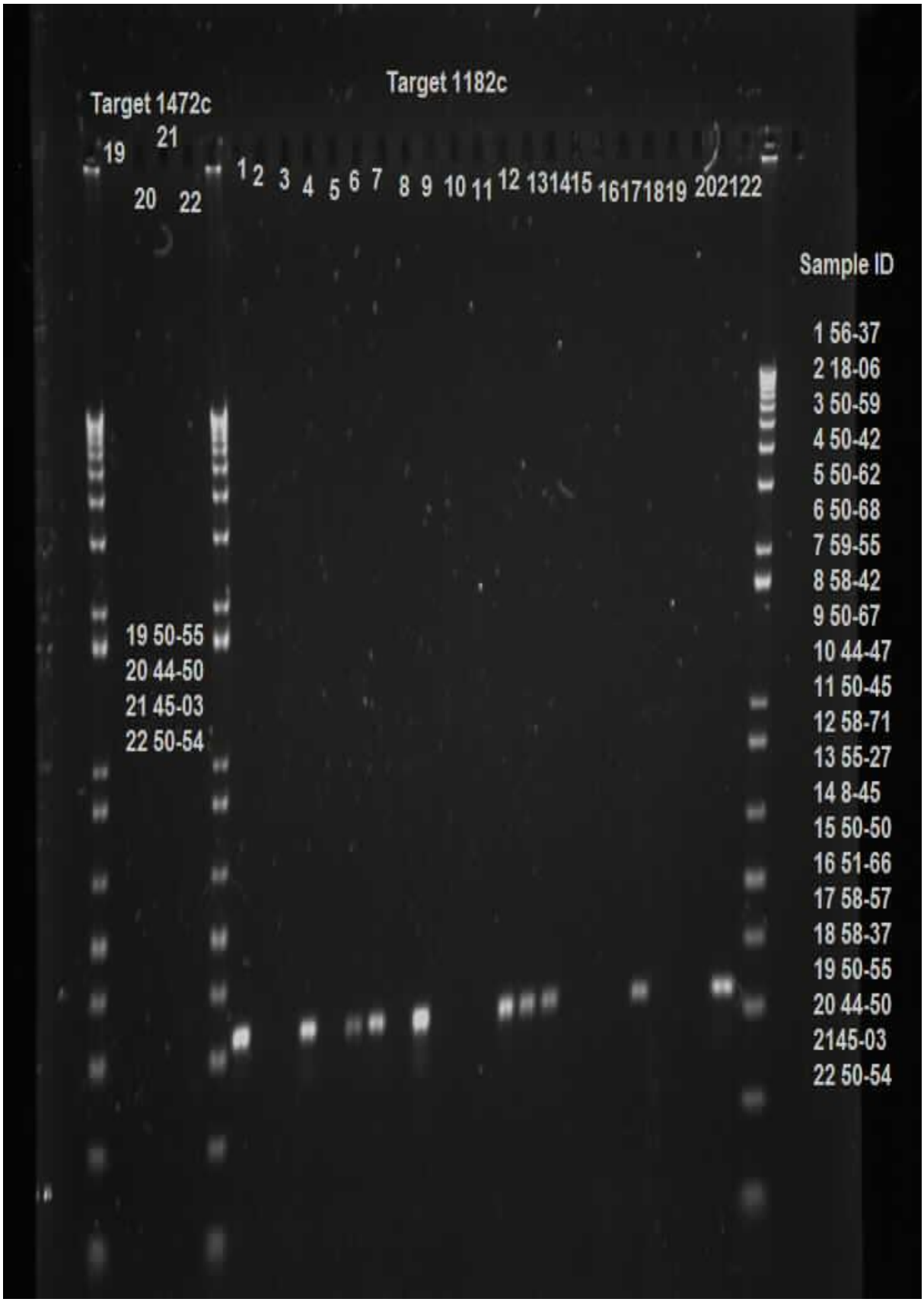


Figure 10: An agarose gel electrophoreses results of ORF118. Sample 1, 2 were positive and negative controls respectively

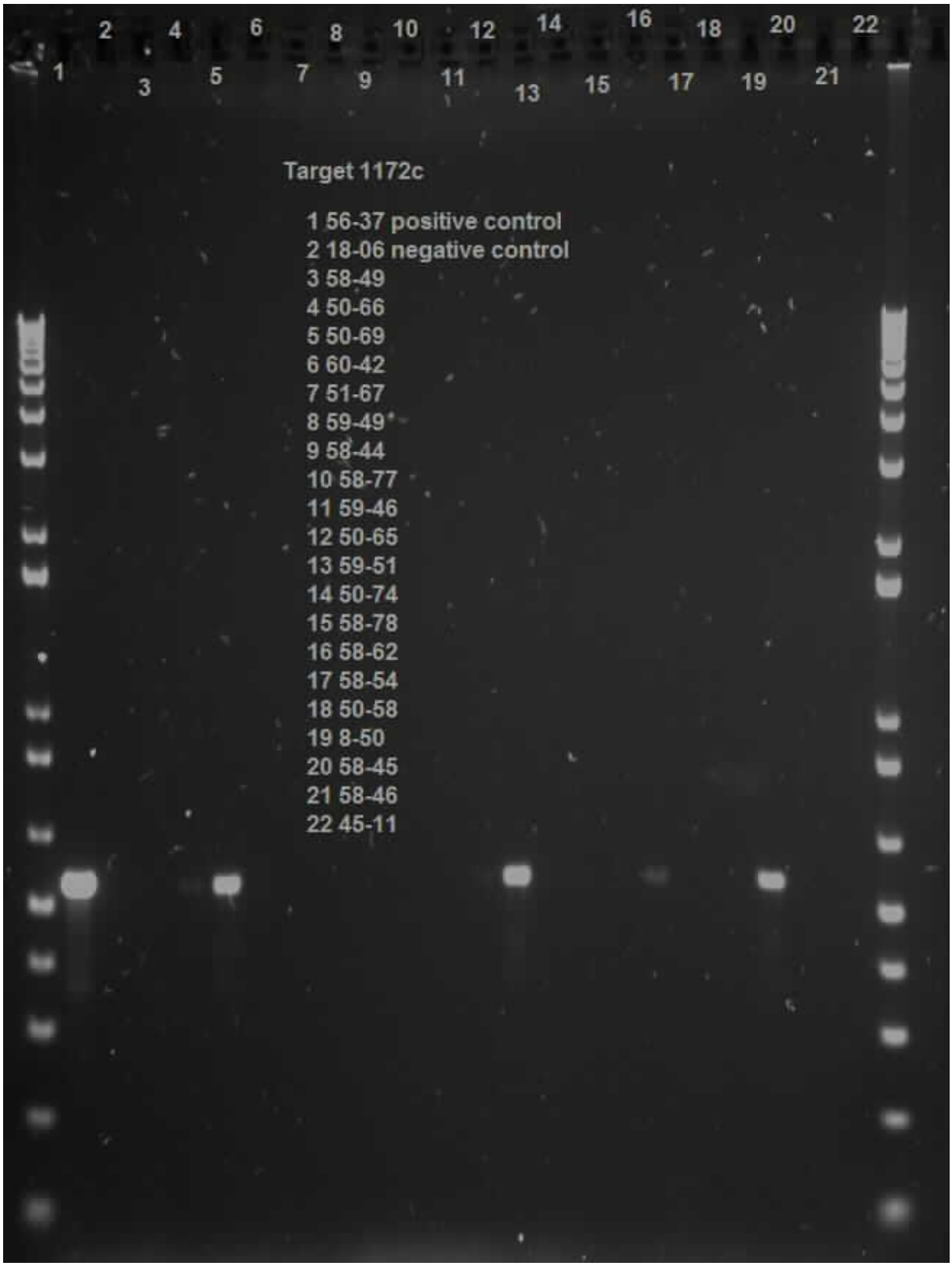


Figure 11: An agarose gel electrophoreses results of ORF117. Sample 1, 2 were positive and negative controls respectively.

DISCUSSION

Plasmid sequencing and gap closure

In this section, the pLG1 replicon type plasmid originating from *E. faecium* (K60-39/3D) had been studied. The aim of the study was to obtain a circular map as well as study the genetic content of the pLG1 replicon type megaplasmid from *E. faecium* (K60-39). Primers were designed in order to amplify sequence targets and to assemble two and two contigs into one (closed gap). The assembly was done by amplification carried out on the PCR machine (Table 5) and PCR products were further analysed in terms of sizes and quality based on agarose gel electrophoreses results (Figure VI, VII), before they were purified and sent for sequencing. Sequencing results were then analysed and used to confirm the sequence in gap closure.

The 3D megaplasmid is described as narrow host range plasmid and has been exclusively found among *E. faecium* population *(Laverde Gomez et al., 2011)*. The reason of obtaining a genetic map is that it can be used in future comparative genomics. If obtained the K60-39/3D plasmid could have been used to study genes content and furthermore determine whether all the ORFs detected in this study were intact. However, the gap closing project has been a challenging encounter from the very beginning; and so the circular DNA was not obtained because of several reasons.

Unspecified primer binding

As shown in Figure IX the result of unspecific primer binding led to the acquisition of multiple bands on the agarose gel and for most of these products it was increasingly difficult to obtain a precise DNA of interest. Three methods could have been implemented to obtain the amplified DNA: 1. by purifying DNA from the gel band, 2 by applying gradient PCR and 3 by optimizing PCR reaction conditions. Both gradient PCR and optimization of PCR reaction conditions were implemented. The agarose gel electrophoreses results in Figure VIII was obtained from gradient PCR. The second

method was not used in this study because the sizes of some gaps were unknown and moreover, it was extremely time consuming. For double bands that were present on the agarose gel following agarose gel electrophoreses, optimization of reaction conditions was implemented, but was not effective because the primers used were of different annealing temperatures for each pair. The gradient PCR also did not work because different annealing temperatures were given for most of the primers that were used.

Insertion sequence elements

Moreover, the presence of many insertion sequence elements was also put into question as to whether they may have cause failure in gap closure. For example, the gene encoding aminoglycoside modifying enzymes was shown to have been engulfed by insertion sequence elements, IS 256 (Figure Appendix) on both sides and could have potentially knock out this region to form an extra plasmid. In this case it could make the PCR fragments useless because there will be no gaps to fill them into due to the fact that the entire region has been knocked out.

The aacA-aphD gene

It was also assumed that the gene (aacA-aphD) encoding aminoglycoside modifying enzyme in enterococci was present and may have been duplicated (Figure VII). Considering that was the case it would have meant that two identical fragments would be lying adjacent to each other or far apart. Have they been distance apart from each other it would cause the pasted sequence to move back and forward in the direction of both fragments as was observed in this study (Figure VI). Two different methods were applied in order to fix these problems. One was the use of direct sequencing of the DNA so that sequences could be used to cover regions between identical genes that were thought of being far apart. The second method however was the used of long range PCR for the same purpose as direct sequencing. Because the sequencing machine could not properly read sequences above one-thousand base pairs, both methods were not further used in the study.

Unsuccessful PCRs reaction

Another possible problem was due to the lack of enough positive PCRs reactions to be included in gaps closure. Out of 66 PCR reactions, only 13 were PCR positive and good for use in sequencing. Of these 7 (Figure V) were selected and used in gap closure. It was not clear what caused these negative PCR reactions, but it was speculated that they could have been because of poor primer quality or that during the time primers were designed it was not easy to predict overlapping contigs. In addition the product sizes for most positive PCRs were above one-thousand base pairs, but during the time of sequencing not more than five-hundred base pairs were obtained for most of the sequences. Any reads above five-hundred were of poor quality for most of the sequences. The sequencing machine used could only sequence up to one-thousand base pairs in one reaction. The result of this however, was that several positive PCRs were needed to close gaps that were bigger than one-thousand base pairs. With little time left for laboratory work this was not an option. Furthermore, an increasing number of primers did not bind specifically to their target sites (Figure VII), which lead to unclean PCR products and further into poor DNA sequences.

Detection of ORFs and prevalence determination

Although Enterococci are commensals of the human gut flora; their ability to cause infection cannot be neglected. They are implicated in hospital acquired infections and are ranked Europe`s 4th most common cause of hospital acquired infections (Hidron et al., 2008). That being said, virulence factors are also common among enterococcal species. The ability of pathogens to invade bloodstream requires that they have the mechanism to resist the host immune system. In this study, results were obtained for four selected ORFs with potential immune invasive properties.

Investigation has being made on the prevalence of OFRs with potential virulence properties in enterococci causing infectious diseases and those that are commensal of the human and animal gut flora. The observed prevalence of ORF023 among *E. faecium* isolates was quite high in the strain collection (Figure 4). Compare to other ORFs it was the most detected among *E. faecium* of all sources. There was no statistical significance

in prevalence of ORF023 between isolates of blood cultures and those of other sources. This highlights the fact that ORF023 could have been of little or no clinical significance due to its abundance in non-clinical culture isolates (76%). Interestingly, prevalence did not differ from blood culture isolates, although the result was not expected. Because of these findings it was concluded that blood culture isolates of *E. faecium* could have possessed ORF023 for other unknown purposes then what was previously proposed, which is immune invasion and host colonization. The presence of ORF023 among non-clinical isolates also highlights the possibility of its role in colonization of the gastrointestinal tract, given its frequency in non-clinical isolates of human and animal fecal as shown in appendix (Master table A).

The prevalence of ORF147 was higher among blood culture isolates than among other clinical and non-clinical isolates as the results in Figure 4 indicate. None of the isolates obtained from non-clinical sources detected positive to ORF147. Therefore the lack of ORF147 among non-clinical isolates, but presence in isolates from blood and other clinical sources shows that ORF147 could have been a marker for isolates obtained from blood cultures and other clinical sources, but not isolates of non-clinical origin.

It was shown that ORF117 and ORF118 were most prevalent in isolates of blood cultures compare to other clinical sources. A statistical significance was revealed between isolates of blood cultures and those of other origins. None of these ORFs were detected among isolates of non-clinical origin. Their absences among non-clinical isolates, but presence in isolates of blood cultures and other clinical cultures also highlights the fact that they could have been clinical markers and therefore have potential virulence properties. However, further investigation might be needed in order to conclude whether they do play a role in immune invasion and host colonization.

The prevalence of ORF147, ORF117 and ORF118 differed between isolates from blood cultures and those of other clinical sources and statistical significance was revealed indication that the tested ORFs were more present among blood culture isolates compared to those of other clinical sources (Table 12). Statistical significance was shown for presence of all ORFs detected in each isolate between blood cultures and other clinical isolates ($P< 0, 0004$).

E. faecium versus *E. faecalis*

On the population level *E. faecium* seems to outnumber *E. faecalis* in the clinical settings (Treitman et al., 2005). This transition of *E. faecium* could have global implication in terms of disease management and antimicrobial resistance control. Therefore it is important to study these recent trends and so as to take possible measures in tackling the spread of high-risk pathogens of *E. faecium*. It is assumed that the growing mobilization of *E. faecium* into high-risk pathogens has to do with their ability to acquire and spread mobile genetic elements (Rosvoll et al., 2010). Therefore on the species level *E. faecium* are said to contain more plasmids than *E. faecalis* (Rosvoll et al., 2010). By considering *E. faecium* and *E. faecalis*, *E. faecium* has the largest genome (Tettelin et al., 2008). However larger genome means that the bacterial can have elevated capacity to add on foreign genetic materials. The ability has been tested in many studies (Duprè et al., 2003, Oancea et al., 2004, Rosvoll et al., 2010). The ability to acquire and disseminate genetic elements might have been the driving force behind *E. faecium* popularity in hospital environments in recent years. In this thesis *E. faecium* of blood, other clinical and non-clinical origins had been compared with *E. faecalis* derived from blood and other clinical cultures in order to see if the prevalence of genes encoding potential ORFs with immune invasive and host colonization properties was to be more in *E. faecium* than *E. faecalis.* The comparison between *E. faecium* and *E. faecalis* showed that *E. faecium* were most detected for presence of all four ORFs. The reason for this trend could have been because the selected ORFs were only selected from blood cultures isolates of *E. faecium* and not *E. faecalis.* The results in Table 14 show that *E. faecium* differs significantly in presence of all four ORFs. Based on the assumption that the tested ORFs were not derived from blood cultures isolates of *E.faecalis* the need for further investigation should be inevitable in deciding whether the two species differ in prevalence to the studied ORFs. However, this would need to be elucidated in the near future in order to make any final conclusion. Until then, the current results showed the two species differed significantly.

CONCLUDING REMARKS

In this thesis an attempt has been made to address two important aspects of enterococci with main focus on *E. faecium*. An attempt to obtain a circular map of a pLG1 replicon type plasmid in an *E. faecium* strain failed following several unidentified problems. It was not easy to identify one specific cause of the problems except that the factors discussed above were thought to have been involved. Because of time limit, not more has been done to resolve these problems. For future study test should be done to assure that the primers are working to avoid unsuccessful PCRs and that the sequencing machine is capable of producing good sequences.

For the screening experiment the clinical significance of all ORFs are unknown and would there have to be elucidated in the nearest future. However, it has been revealed that the selected ORFs were significantly more common among isolates of blood cultures of *E. faecium* compared to those derived from other sources. Isolates of non-clinical cultures were not harbouring ORF147, ORF117 and ORF118. Differences in prevalence were significant between non-clinical isolates and isolates from other clinical sources, except for ORF023 which did not differ in prevalence among *E. faecium*. A significant number of Isolates among blood cultures were positive to all four ORFs. In addition ORFs were also more prevalent but and statistically significant different between isolates of *E. faecium* and *E. faecalis*. In conclusion, the tested hypothesis was correct that the selected ORFs were highly prevalent in blood cultures isolates of *E. faecium* compare to isolates of other human cultures. The selected ORFs were derived from blood culture isolates of *E. faecium*. With that being said, further studies might be needed to compare *E. faecium* and *E. faecalis*.

REFERENCES

ABAJY, M. Y., KOPEC, J., SCHIWON, K., BURZYNSKI, M., DORING, M., BOHN, C. & GROHMANN, E. 2007. A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria. J Bacteriol, 189, 2487-96.

ANGULO, F. J., HEUER, O. E., HAMMERUM, A. M., COLLIGNON, P. & WEGENER, H. C. 2006. Human Health Hazard from Antimicrobial-Resistant Enterococci in 3 Animals and Food. Clinical Infectious Diseases, 43, 911-916.

ARIAS, C. A. & MURRAY, B. E. 2012. The rise of the Enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol,* 10**,** 266-78.

ARIAS, C. A., PANESSO, D., SINGH, K. V., RICE, L. B. & MURRAY, B. E. 2009. Cotransfer of antibiotic resistance genes and a hylEfm-containing virulence plasmid in Enterococcus faecium. *Antimicrob Agents Chemother,* 53**,** 4240-6.

ARIAS, C. A., PANESSO, D., SINGH, K. V., RICE, L. B. & MURRAY, B. E. 2009. Cotransfer of Antibiotic Resistance Genes and a hylEfm-Containing Virulence Plasmid in Enterococcus faecium. *Antimicrobial Agents and Chemotherapy,* 53**,** 4240-4246.

BOSS, L., LABUDDA, L., WEGRZYN, G., HAYES, F. & KEDZIERSKA, B. 2013. The Axe-Txe Complex of Enterococcus faecium Presents a Multilayered Mode of Toxin-Antitoxin Gene Expression Regulation. PLoS One, 8, e73569.

BRODAL, H. P. 2011. *Pangenom-studier av genus Enterococcus,* Ås, [H.P. Brodal].

CHOW, J. W. 2000. Aminoglycoside Resistance in Enterococci. Clinical Infectious Diseases, 31, 586-589.

CLEWELL, D. B. 2011. Tales of conjugation and sex pheromones: A plasmid and enterococcal odyssey. Mob Genet Elements, 1, 38-54.

COBURN, P. S., PILLAR, C. M., JETT, B. D., HAAS, W. & GILMORE, M. S. 2004. Enterococcus faecalis Senses Target Cells and in Response Expresses Cytolysin. Science, 306, 2270-2272.

DE REGT, M. J. A., VAN SCHAIK, W., VAN LUIT-ASBROEK, M., DEKKER, H. A. T., VAN DUIJKEREN, E., KONING, C. J. M., BONTEN, M. J. M. & WILLEMS, R. J. L. 2012. Hospital and Community Ampicillin-Resistant Enterococcus faecium Are Evolutionarily
Closely Linked but Have Diversified through Niche Adaptation. *PLoS ONE,* 7**,** e30319.

FISHER, K. & PHILLIPS, C. 2009. The ecology, epidemiology and virulence of Enterococcus. Microbiology, 155, 1749-1757.

FLANNAGAN, S. E., CHOW, J. W., DONABEDIAN, S. M., BROWN, W. J., PERRI, M. B., ZERVOS, M. J., OZAWA, Y. & CLEWELL, D. B. 2003. Plasmid content of a vancomycin-resistant Enterococcus faecalis isolate from a patient also colonized by Staphylococcus aureus with a VanA phenotype. Antimicrob Agents Chemother, 47, 3954-9.

FRANCIA, M. V., WITTE, W. & WERNER, G. 2011. A multiresistance megaplasmid pLG1 bearing a hylEfm genomic island in hospital Enterococcus faecium isolates. Int J Med Microbiol, 301, 165-75.

FREITAS, A. R., COQUE, T. M., NOVAIS, C., HAMMERUM, A. M., LESTER, C. H., ZERVOS, M. J., DONABEDIAN, S., JENSEN, L. B., FRANCIA, M. V., BAQUERO, F. & PEIXE, L. 2011. Human and swine hosts share vancomycin-resistant Enterococcus faecium CC17 and CC5 and Enterococcus faecalis CC2 clonal clusters harboring Tn1546 on indistinguishable plasmids. J Clin Microbiol, 49, 925-31.

GARCIA-MIGURA, L., SANCHEZ-VALENZUELA, A. J. & JENSEN, L. B. 2011. Presence of glycopeptide-encoding plasmids in enterococcal isolates from food and humans in Denmark. Foodborne Pathog Dis, 8, 1191-7.

GOESSWEINER-MOHR, N., ARENDS, K., KELLER, W. & GROHMANN, E. 2013. Conjugative type IV secretion systems in Gram-positive bacteria. Plasmid, 70, 289-302.

GRADY, R. & HAYES, F. 2003. Axe–Txe, a broad-spectrum proteic toxin–antitoxin system specified by a multidrug-resistant, clinical isolate of Enterococcus faecium. Molecular Microbiology, 47, 1419-1432.

GRAHAM, J. C. & GOULD, F. K. 2002. Role of aminoglycosides in the treatment of bacterial endocarditis. Journal of Antimicrobial Chemotherapy, 49, 437-444.

HÄLLGREN, A., CLAESSON, C., SAEEDI, B., MONSTEIN, H.-J., HANBERGER, H. & NILSSON, L. E. 2009. Molecular detection of aggregation substance, enterococcal surface protein, and cytolysin genes and in vitro adhesion to urinary catheters of Enterococcus faecalis and E. faecium of clinical origin. International Journal of Medical Microbiology, 299, 323-332.

HALVORSEN, E. M., WILLIAMS, J. J., BHIMANI, A. J., BILLINGS, E. A. & HERGENROTHER, P. J. 2011a. Txe, an endoribonuclease of the enterococcal Axe-Txe toxin-antitoxin system, cleaves mRNA and inhibits protein synthesis. Microbiology, 157, 387-97.

HEIKENS, E., BONTEN, M. & WILLEMS, R. 2007. Enterococcal surface protein Esp is important for biofilm formation of Enterococcus faecium E1162. *J Bacteriol,* 189**,** 8233 - 8240.
HEIKENS, E., BONTEN, M. & WILLEMS, R. 2007. Enterococcal surface protein Esp is important for biofilm formation of Enterococcus faecium E1162. *J Bacteriol,* 189**,** 8233 - 8240.
HIDRON, A., EDWARDS, J., PATEL, J., HORAN, T., SIEVERT, D., POLLOCK, D. & FRIDKIN, S. 2008. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol,* 29**,** 996 - 1011.

HOLLENBECK, B. L. & RICE, L. B. 2012. Intrinsic and acquired resistance mechanisms in enterococcus. Virulence, 3, 421-33.

HUANG, C. R., BURNS, K. H. & BOEKE, J. D. 2012. Active transposition in genomes. Annu Rev Genet, 46, 651-75.

IKE, Y., HASHIMOTO, H. & CLEWELL, D. B. 1987. High incidence of hemolysin production by Enterococcus (Streptococcus) faecalis strains associated with human parenteral infections. *J Clin Microbiol,* 25**,** 1524-8.

ENSEN, L. B., GARCIA-MIGURA, L., VALENZUELA, A. J., LOHR, M., HASMAN, H. & AARESTRUP, F. M. 2010. A classification system for plasmids from enterococci and other Gram-positive bacteria. J Microbiol Methods, 80, 25-43.

JUREEN, R. 2004. Ampicillin- and high-level gentamicin-resistant enterococci in hospitalised patients in Norway, Bergen, Institute of Medicine, University of Bergen.

LAM, M. M. C., SEEMANN, T., BULACH, D. M., GLADMAN, S. L., CHEN, H., HARING, V., MOORE, R. J., BALLARD, S., GRAYSON, M. L., JOHNSON, P. D. R., HOWDEN, B. P. & STINEAR, T. P. 2012. Comparative Analysis of the First Complete Enterococcus faecium Genome. Journal of Bacteriology, 194, 2334-2341.

LAVERDE GOMEZ, J. A., VAN SCHAIK, W., FREITAS, A. R., COQUE, T. M., WEAVER, K. E., FRANCIA, M. V., WITTE, W. & WERNER, G. 2011. A multiresistance megaplasmid pLG1 bearing a hylEfm genomic island in hospital Enterococcus faecium isolates. Int J Med Microbiol, 301, 165-75.

LEAVIS, H., WILLEMS, R., TOP, J. & BONTEN, M. 2006. High-level ciprofloxacin resistance from point mutations in gyrA and parC confined to global hospital-adapted clonal lineage CC17 of Enterococcus faecium. *J Clin Microbiol,* 44**,** 1059 - 1064.

MONSTEIN, H. J., QUEDNAU, M., SAMUELSSON, A., AHRNE, S., ISAKSSON, B. & JONASSON, J. 1998. Division of the genus Enterococcus into species groups using PCR-based molecular typing methods. *Microbiology,* 144 (Pt 5)**,** 1171-9.

MUNDY, L. M., SAHM, D. F. & GILMORE, M. 2000. Relationships between Enterococcal Virulence and Antimicrobial Resistance. Clinical Microbiology Reviews, 13, 513-522.

NALLAPAREDDY, S., SINGH, K. & MURRAY, B. 2008. Contribution of the collagen adhesin Acm to pathogenesis of Enterococcus faecium in experimental endocarditis. *Infect Immun,* 76**,** 4120 - 4128.

PALAZZO, I. C., PITONDO-SILVA, A., LEVY, C. E. & DA COSTA DARINI, A. L. 2011. Changes in vancomycin-resistant Enterococcus faecium causing outbreaks in Brazil. J Hosp Infect, 79, 70-4.

PALMER, K. L., KOS, V. N. & GILMORE, M. S. 2010. Horizontal gene transfer and the genomics of enterococcal antibiotic resistance. Curr Opin Microbiol, 13, 632-9.

.PAULSEN, I., BANERJEI, L., MYERS, G., NELSON, K., SESHADRI, R., READ, T., FOUTS, D., EISEN, J., GILL, S., HEIDELBERG, J., TETTELIN, H., DODSON, R., UMAYAM, L., BRINKAC, L., BEANAN, M., DAUGHERTY, S., DEBOY, R., DURKIN, S., KOLONAY, J., MADUPU, R., NELSON, W., VAMATHEVAN, J., TRAN, B., UPTON, J., HANSEN, T., SHETTY, J., KHOURI, H., UTTERBACK, T., RADUNE, D. & KETCHUM, K. 2003. Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. *Science,* 299**,** 2071 - 2074.

QIN, X., GALLOWAY-PENA, J. R., SILLANPAA, J., ROH, J. H., NALLAPAREDDY, S. R., CHOWDHURY, S., BOURGOGNE, A., CHOUDHURY, T., MUZNY, D. M., BUHAY, C. J., DING, Y., DUGAN-ROCHA, S., LIU, W., KOVAR, C., SODERGREN, E., HIGHLANDER, S., PETROSINO, J. F., WORLEY, K. C., GIBBS, R. A., WEINSTOCK, G. M. & MURRAY, B. E. 2012. Complete genome sequence of Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus faecium genomes. BMC Microbiol, 12, 135.

QU, T. T., YANG, Q., SHEN, P., WEI, Z. Q. & YU, Y. S. 2012. Novel vancomycin-resistance transposon, plasmid replicon types, and virulence factors of vancomycin-resistant Enterococci in Zhejiang, China. Microb Drug Resist, 18, 183-8.

RICE, L., CARIAS, L., RUDIN, S., VAEL, C., GOOSSENS, H., KONSTABEL, C., KLARE, I., NALLAPAREDDY, S., HUANG, W. & MURRAY, B. 2003. A potential virulence gene, hylEfm, predominates in Enterococcus faecium of clinical origin. *J Infect Dis,* 187**,** 508 - 512.

SALAH, R., DAR-ODEH, N., ABU HAMMAD, O. & SHEHABI, A. A. 2008. Prevalence of putative virulence factors and antimicrobial susceptibility of Enterococcus faecalis isolates from patients with dental Diseases. BMC Oral Health, 8, 17.

SANCHEZ, B., URDACI, M. C. & MARGOLLES, A. 2010. Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa-bacteria interactions. Microbiology, 156, 3232-42.

SANTAGATI, M., CAMPANILE, F. & STEFANI, S. 2012. Genomic diversification of enterococci in hosts: the role of the mobilome. Front Microbiol, 3, 95.

SENO, Y., KARIYAMA, R., MITSUHATA, R., MONDEN, K. & KUMON, H. 2005. Clinical implications of biofilm formation by Enterococcus faecalis in the urinary tract. Acta Med Okayama, 59, 79-87.

SLETVOLD, H., JOHNSEN, P., HAMRE, I., SIMONSEN, G., SUNDSFJORD, A. & NIELSEN, K. 2008. Complete sequence of Enterococcus faecium pVEF3 and the detection of an omega-epsilon-zeta toxin-antitoxin module and an ABC transporter. *Plasmid,* 60**,** 75 - 85.

SONG, X., SUN, J., MIKALSEN, T., ROBERTS, A. P. & SUNDSFJORD, A. 2013. Characterisation of the Plasmidome within Enterococcus faecalis Isolated from Marginal Periodontitis Patients in Norway. PLoS ONE, 8, e62248.

SOOD, S., MALHOTRA, M., DAS, B. K. & KAPIL, A. 2008. Enterococcal infections & antimicrobial resistance. Indian J Med Res, 128, 111-21.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. 2008. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol,* 11**,** 472 - 477.

TREITMAN, A. N., YARNOLD, P. R., WARREN, J. & NOSKIN, G. A. 2005. Emerging Incidence of Enterococcus faecium among Hospital Isolates (1993 to 2002). *Journal of Clinical Microbiology,* 43**,** 462-463.

UBOSS, L., LABUDDA, L., WEGRZYN, G., HAYES, F. & KEDZIERSKA, B. 2013. The Axe-Txe Complex of Enterococcus faecium Presents a Multilayered Mode of Toxin-Antitoxin Gene Expression Regulation. PLoS One, 8, e73569.

VANKERCKHOVEN, V., VAN AUTGAERDEN, T., VAEL, C., LAMMENS, C., CHAPELLE, S., ROSSI, R., JABES, D. & GOOSSENS, H. 2004. Development of a Multiplex PCR for the Detection of asa1, gelE, cylA, esp, and hyl Genes in Enterococci and Survey for Virulence Determinants among European Hospital Isolates of Enterococcus faecium. Journal of Clinical Microbiology, 42, 4473-4479.

VANNUFFEL, P. & COCITO, C. 1996. Mechanism of action of streptogramins and macrolides. Drugs, 51 Suppl 1, 20-30.

WARDAL EWA, S. E., HRYNIEWICY VERLERIA 2010. Complex Nature of Enterococcal Pheromone-Responsive Plasmids. Vol . 59, 87.

WENDELBO, O., JUREEN, R., EIDE, G. E., DIGRANES, A., LANGELAND, N. & HARTHUG, S. 2003. Outbreak of infection with high-level gentamicin-resistant Enterococcus faecalis (HLGRE) in a Norwegian hospital. Clin Microbiol Infect, 9, 662-9.

WENDELBO, O., JUREEN, R., EIDE, G. E., DIGRANES, A., LANGELAND, N. & WERNER, G., COQUE, T. M., FRANZ, C. M. A. P., GROHMANN, E., HEGSTAD, K., JENSEN, L., VAN SCHAIK, W. & WEAVER, K. 2013. Antibiotic resistant enterococci—Tales of a drug resistance gene trafficker. International Journal of Medical Microbiology, 303, 360-379.

WILLEMS, R. & VAN SCHAIK, W. 2009. Transition of Enterococcus faecium from commensal organism to nosocomial pathogen. *Future Microbiol,* 4**,** 1125 - 1135.

WILLEMS, R., HOMAN, W., TOP, J., VAN SANTEN-VERHEUVEL, M., TRIBE, D., MANZIOROS, X., GAILLARD, C., VANDENBROUCKE-GRAULS, C., MASCINI, E., VAN KREGTEN, E., VAN EMBDEN, J. & BONTEN, M. 2001. Variant esp gene as a marker of a distinct genetic lineage of vancomycin-resistant Enterococcus faecium spreading in hospitals. *Lancet,* 357**,** 853 - 855.

ZIRAKZADEH, A. & PATEL, R. 2006. Vancomycin-Resistant Enterococci: Colonization, Infection, Detection, and Treatment. Mayo Clinic proceedings Mayo Clinic, 81, 529-536.

ABAJY, M. Y., KOPEC, J., SCHIWON, K., BURZYNSKI, M., DORING, M., BOHN, C. & GROHMANN, E. 2007. A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria. *J Bacteriol,* 189**,** 2487-96.

AN, F. Y. & CLEWELL, D. B. 2002. Identification of the cAD1 Sex Pheromone Precursor in Enterococcus faecalis. *Journal of Bacteriology,* 184**,** 1880-1887.

ARIAS, C. A. & MURRAY, B. E. 2012. The rise of the Enterococcus: beyond vancomycin resistance. *Nat Rev Microbiol,* 10**,** 266-78.
BONAFEDE, M. E., CARIAS, L. L. & RICE, L. B. 1997. Enterococcal transposon Tn5384: evolution of a composite transposon through cointegration of enterococcal and staphylococcal plasmids. *Antimicrob Agents Chemother,* 41**,** 1854-8.
BRODAL, H. P. 2011. *Pangenom-studier av genus Enterococcus,* Ås, [H.P. Brodal].
CARVALHO MDA, G., STEIGERWALT, A. G., MOREY, R. E., SHEWMAKER, P. L., TEIXEIRA, L. M. & FACKLAM, R. R. 2004. Characterization of three new enterococcal species, Enterococcus sp. nov. CDC PNS-E1, Enterococcus sp. nov. CDC PNS-E2, and Enterococcus sp. nov. CDC PNS-E3, isolated from human clinical specimens. *J Clin Microbiol,* 42**,** 1192-8.
CHOW, J. W., THAL, L. A., PERRI, M. B., VAZQUEZ, J. A., DONABEDIAN, S. M., CLEWELL, D. B. & ZERVOS, M. J. 1993. Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis. *Antimicrob Agents Chemother,* 37**,** 2474-7.
DEIBEL, R. H., LAKE, D. E. & NIVEN, C. F., JR. 1963. PHYSIOLOGY OF THE ENTEROCOCCI AS RELATED TO THEIR TAXONOMY. *J Bacteriol,* 86**,** 1275-82.
DUPRÈ, I., ZANETTI, S., SCHITO, A. M., FADDA, G. & SECHI, L. A. 2003. Incidence of virulence determinants in clinical Enterococcus faecium and Enterococcus faecalis isolates collected in Sardinia (Italy). *Journal of Medical Microbiology,* 52**,** 491-498.
ERLISCH, H. A. 1989. *PCR Technology,* United States and Canada Stockton Press.
FELMINGHAM, D., WILSON, A. P., QUINTANA, A. I. & GRUNEBERG, R. N. 1992. Enterococcus species in urinary tract infection. *Clin Infect Dis,* 15**,** 295-301.
FREITAS, A. R., NOVAIS, C., RUIZ-GARBAJOSA, P., COQUE, T. M. & PEIXE, L. 2009. Clonal expansion within clonal complex 2 and spread of vancomycin-resistant plasmids among different genetic lineages of Enterococcus faecalis from Portugal. *Journal of Antimicrobial Chemotherapy,* 63**,** 1104-1111.
FREITAS, A. R., TEDIM, A. P., NOVAIS, C., RUIZ-GARBAJOSA, P., WERNER, G., LAVERDE-GOMEZ, J. A., CANTON, R., PEIXE, L., BAQUERO, F. & COQUE, T. M. 2010. Global spread of the hyl(Efm) colonization-virulence gene in megaplasmids of the Enterococcus faecium CC17 polyclonal subcluster. *Antimicrob Agents Chemother,* 54**,** 2660-5.
GOESSWEINER-MOHR, N., GRUMET, L., ARENDS, K., PAVKOV-KELLER, T., GRUBER, C. C., GRUBER, K., BIRNER-GRUENBERGER, R., KROPEC-HUEBNER, A., HUEBNER, J., GROHMANN, E. & KELLER, W. 2012. The 2.5 Å Structure of the Enterococcus Conjugation Protein TraM resembles VirB8 Type IV Secretion Proteins. *Journal of Biological Chemistry*.
GRADY, R. & HAYES, F. 2003. Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of Enterococcus faecium. *Mol Microbiol,* 47**,** 1419-32.
KIM, E. B. & MARCO, M. L. 2013. Non-clinical and clinical Enterococcus faecium but not Enterococcus faecalis have distinct structural and functional genomic features. *Applied and Environmental Microbiology*.
KREFT, B., MARRE, R., SCHRAMM, U. & WIRTH, R. 1992. Aggregation substance of Enterococcus faecalis mediates adhesion to cultured renal tubular cells. *Infection and Immunity,* 60**,** 25-30.
LAVERDE GOMEZ, J. A., VAN SCHAIK, W., FREITAS, A. R., COQUE, T. M., WEAVER, K. E., FRANCIA, M. V., WITTE, W. & WERNER, G. 2011. A multiresistance megaplasmid

pLG1 bearing a hylEfm genomic island in hospital Enterococcus faecium isolates. *Int J Med Microbiol,* 301**,** 165-75.

LAW-BROWN, J. & MEYERS, P. R. 2003. Enterococcus phoeniculicola sp. nov., a novel member of the enterococci isolated from the uropygial gland of the Red-billed Woodhoopoe, Phoeniculus purpureus. *Int J Syst Evol Microbiol,* 53**,** 683-5.

LIM, S. K., TANIMOTO, K., TOMITA, H. & IKE, Y. 2006. Pheromone-responsive conjugative vancomycin resistance plasmids in Enterococcus faecalis isolates from humans and chicken feces. *Appl Environ Microbiol,* 72**,** 6544-53.

MORITZ, E. M. & HERGENROTHER, P. J. 2007. Toxin–antitoxin systems are ubiquitous and plasmid-encoded in vancomycin-resistant enterococci. *Proceedings of the National Academy of Sciences,* 104**,** 311-316.

OANCEA, C., KLARE, I., WITTE, W. & WERNER, G. 2004. Conjugative transfer of the virulence gene, esp, among isolates of Enterococcus faecium and Enterococcus faecalis. *J Antimicrob Chemother,* 54**,** 232 - 235.

PANESSO, D., REYES, J., RINCÓN, S., DÍAZ, L., GALLOWAY-PEÑA, J., ZURITA, J., CARRILLO, C., MERENTES, A., GUZMÁN, M., ADACHI, J. A., MURRAY, B. E. & ARIAS, C. A. 2010. Molecular Epidemiology of Vancomycin-Resistant Enterococcus faecium: a Prospective, Multicenter Study in South American Hospitals. *Journal of Clinical Microbiology,* 48**,** 1562-1569.

RICE, L., CARIAS, L., RUDIN, S., VAEL, C., GOOSSENS, H., KONSTABEL, C., KLARE, I., NALLAPAREDDY, S., HUANG, W. & MURRAY, B. 2003. A potential virulence gene, hylEfm, predominates in Enterococcus faecium of clinical origin. *J Infect Dis,* 187**,** 508 - 512.

ROBERTS, A. P., CHANDLER, M., COURVALIN, P., GUÉDON, G., MULLANY, P., PEMBROKE, T., ROOD, J. I., JEFFERY SMITH, C., SUMMERS, A. O., TSUDA, M. & BERG, D. E. 2008. Revised nomenclature for transposable genetic elements. *Plasmid,* 60**,** 167-173.

ROSENBERG, C., BOISTARD, P., DÉNARIÉ, J. & CASSE-DELBART, F. 1981. Genes controlling early and late functions in symbiosis are located on a megaplasmid in Rhizobium meliloti. *Molecular and General Genetics MGG,* 184**,** 326-333.

ROSVOLL, T., PEDERSEN, T., SLETVOLD, H., JOHNSEN, P., SOLLID, J., SIMONSEN, G., JENSEN, L., NIELSEN, K. & SUNDSFJORD, A. 2010. PCR-based plasmid typing in Enterococcus faecium strains reveals widely distributed pRE25-, pRUM-, pIP501- and pHTbeta-related replicons associated with glycopeptide resistance and stabilizing toxin-antitoxin systems. *FEMS Immunol Med Microbiol,* 58**,** 254 - 268.

ROSVOLL, T. C., LINDSTAD, B. L., LUNDE, T. M., HEGSTAD, K., AASNAES, B., HAMMERUM, A. M., LESTER, C. H., SIMONSEN, G. S., SUNDSFJORD, A. & PEDERSEN, T. 2012. Increased high-level gentamicin resistance in invasive Enterococcus faecium is associated with aac(6')Ie-aph(2'')Ia-encoding transferable megaplasmids hosted by major hospital-adapted lineages. *FEMS Immunol Med Microbiol,* 66**,** 166-76.

ROSVOLL, T. C. S. 2012. Plasmids, Resistance and Hospital adaptation in Enterococci : an epidemiological approach

SENO, Y., KARIYAMA, R., MITSUHATA, R., MONDEN, K. & KUMON, H. 2005. Clinical implications of biofilm formation by Enterococcus faecalis in the urinary tract. *Acta Med Okayama,* 59**,** 79-87.

SLETVOLD, H., JOHNSEN, P., HAMRE, I., SIMONSEN, G., SUNDSFJORD, A. & NIELSEN, K. 2008. Complete sequence of Enterococcus faecium pVEF3 and the detection of an omega-epsilon-zeta toxin-antitoxin module and an ABC transporter. *Plasmid,* 60**,** 75 - 85.

STARIKOVA, I., AL-HARONI, M., WERNER, G., ROBERTS, A. P., SORUM, V., NIELSEN, K. M. & JOHNSEN, P. J. 2013. Fitness costs of various mobile genetic elements in Enterococcus faecium and Enterococcus faecalis. *J Antimicrob Chemother,* 68**,** 2755-65.

STEINBÜCHEL, A. & SCHWARTZ, E. 2009. *Microbial Megaplasmids,* Berlin, Heidelberg, Springer Berlin Heidelberg.

TETTELIN, H., RILEY, D., CATTUTO, C. & MEDINI, D. 2008. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol,* 11**,** 472 - 477.

TOMITA, H., TANIMOTO, K., HAYAKAWA, S., MORINAGA, K., EZAKI, K., OSHIMA, H. & IKE, Y. 2003. Highly Conjugative pMG1-Like Plasmids Carrying Tn1546-Like Transposons That Encode Vancomycin Resistance in Enterococcus faecium. *Journal of Bacteriology,* 185**,** 7024-7028.

TREITMAN, A., YARNOLD, P., WARREN, J. & NOSKIN, G. 2005. Emerging incidence of Enterococcus faecium among hospital isolates (1993 to 2002). *J Clin Microbiol,* 43**,** 462 - 463.

VALDEZATE, S., MIRANDA, C., NAVARRO, A., FREITAS, A. R., CABRERA, J. J., CARRASCO, G., COQUE, T. M., JIMÉNEZ-ROMANO, E. & SAÉZ-NIETO, J. A. 2012. Clonal outbreak of ST17 multidrug-resistant Enterococcus faecium harbouring an Inc18-like::Tn1546 plasmid in a haemo-oncology ward of a Spanish hospital. *Journal of Antimicrobial Chemotherapy,* 67**,** 832-836.

VANKERCKHOVEN, V., VAN AUTGAERDEN, T., VAEL, C., LAMMENS, C., CHAPELLE, S., ROSSI, R., JABES, D. & GOOSSENS, H. 2004. Development of a multiplex PCR for the detection of asa1, gelE, cylA, esp, and hyl genes in enterococci and survey for virulence determinants among European hospital isolates of Enterococcus faecium. *J Clin Microbiol,* 42**,** 4473-9.

APPENDIX

Table 3: PCR master mixes used in this study

Master mix for detection of 16SrDNa	**Amount (µL)**
TagRed Ready Mix	12,5
$_dH_2O$	9,0
16S R primer	1,0
16S F primer	1,0
DNA template	1,0
Master mix for detection of genes target	**Amount(µL)**
TagRed Ready Mix	12,5
$_dH_2O$.	9,5
R primer.	0,5
F primer.	0,5
DNA Template.	2.0
Master mix for gap closing	**Amount(µL)**
TagRed Ready Mix	12,5
$_dH_2O$	9,5
R primer.	0,5
F primer.	0,5
DNA template	2.0

Table 4: Shows screening primers, their melting temperature™ expected product sizes and name of ORFs encoding hypothetical proteins

Oligonucleotides	**Primer ID**	**Hypothetical protein**	**Expected PCR product size**	**TM**
5'-TGGGGGCTATTTTCTTTTGA-3' 5'CGAAATAGCGTTGGAATCAGT-3'	118 2C F 118 2C R	3D hypothetical secretory protein	321 bp	60,5 61,4
5'-GGT GTT TCT GTT GGG CTA CTT AT-3' 5'-GTT GGC TTG ATT TAC ATT TCC TAC-3'	117 2C F 117 2C R	168 3D hypothetical secretory protein	519bp	45,3 41,4
5'-CTT TAG ATA GGG ACG GAA TAG C-3' 5'-TTT TCT TTC CAA TTC CGT GAT-3'	147 2C F 147 2C R	257 3D hypothetical secretory protein	187bp	47,9 41,4
5'TCAAGCAATACGAGCCAAAGT3' 5'-TCTTGAAGTTTTTGAATGCTC-3'	023 2C F 023 2C R	281 3D hypotetisk sekretorisk protein	459 bp	45,3 41,4

Based on Signal P analysis, it was projected that the following gene targets encode hypothetical proteins with signal peptides sequences.

Table 6 Illumina DNA sequencing results of the 3D megaplasmid of *E. faecium* (K60-39)

Variables	Data
Strain	K60-39
Plasmid size(kb) Sample Library Plasmid assembled(kb) Number of scaffolds GC content (%) CDS IS elements Number of gaps Number of gaps closed Primers available Number PCRs run	300 3D Mate paired assembly 286 18 34,24 315 60 26 7 71 61

Table 5: Reagents and materials that were used in the study

Reagents	**Amount/PCR**	**Source**
RED Taq DNA Polymerase Ready Mix(2x Master Mix Kit) Comp: 75 mM Tris-HCl pH 8.5 20 mM (NH4)2SO4 0,2 mM of each dNTP 1,5 mM MgCl2 0,01% Tween 20 0,625 units/µl VWR Tag polymerase Inert Red dye and a stabilizer	12,5 µL	VWR, Denmark
Reaction Master Mix Comp: Red Tag DNA polymerase 2x Master Mix dH2O	23 µL	This study
BigDye 3.1 Reaction Kit Comp: Ready Reaction Mix pGEM®-3Zf(+) double-stranded DNA template 21 M12 control primer Big dye Terminator 3.1 sequencing buffer(5X)	1 µL	Applied Biosystems, USA
QIAGEN Long Range PCR kit Comp: Long Range PCR buffer dNTPs(10mM) RNase-free water Long Range PCR Enzyme Mix 5x buffer Q		QIAGEN, Hilden, Germany
ExoSAP- IT Comp: Exonuclease 1and Shrimp Alkaline Phosphatase in specially formulated buffer		USB, Cleveland,Ohio, USA
GelRed Nucleic Acid Stain	5 µL i 100 mL gel	Biotium, Hayward, USA
TAE buffer 1X 40 mM. Tris-HCl (Triza-base) 20 mM. acetic acid 1 mM. EDTA pH 8	-	K-res, UiT
1 Kb+ DNA Ladder	5 µL per reaction	Biotium, Hayward, USA
Seakem LE Agarose	-	Lonza, Rochland, USA
Materials		
GeneSnap Version 7.12	-	Syngene, UK

BioEdit	-	Ibis, Bioscience
PCR Reaction Tubes	-	Thermo Scientific, MIC
Primers for gaps closure	-	Eurogentic S.A, Belgium
ART, Aerosol Resistant tips		San, Diego California, USA
ABI PRISM 3700 DNA Analyser	-	Applied Biosystems
PCR Thermal cycler		BIORAD, Laboratories, Inc, Singapore

Table 7: Shows PCR reaction program for Screening of genes encoding potential virulence factors in *E. faecium* and *E. faecalis*

	Temperature (oC)	**Time**	**Number of cycles**
Initial denature	95	5 min	
Denaturation	95	30 s	25 cycles
Annealing	56	30s	
Extension	72	30s	
Final extension	72	5min	
End of PCR	12	indefinate	

Table 8: Shows PCR reaction program for the detection of 16SrDNA

	Temperature (^oC)	**Time**	**Number of cycles**
Initial denature	95	5 min	
Denaturation	95	1 min	30 cycles
Annealing	58	1min	
Extension	72	1min	
Final extension	72	5min	
End of PCR	12	indefinate	

Table 9: Cycle sequencing protocal for clean-up PCR products

	Temperature (°C)	**Time**	**Number of cycles**
Initial temperature	95	5 min	
Denaturation Tem:	95	1 min	
Annealing Tem:	50	1 min	30 cycles
Extension Tem:	60	2 min	
Final extension	60	5min	
End of PCR	12	indefinate	

Table 10: Gap closing PCR reaction program

	Temperature (°C)	**Time**	**Number of cycles**
Initial temperature	95	5 min	
Denaturation	95	1 min	
Annealing(gradient)	55-60	1min	30 cycles
Extension Tem:	72	1min	
Final extension	72	5min	
End of PCR	12	indefinate	

Table 11: Gap closing Long Range PCR reaction program

	Temperature (°C)	**Time**	**Number of cycles**
Initial temperature	93	3 min	
Denaturation	93	15 s	
Annealing	62	30 s	30 cycles
Extension Tem:	68	1 min/kb	
Final extension	68	5min	
End of PCR	12	indefinate	

Master table A:

ID/fryse stock	Name	Species	ST	Source	Year	Origin	Reference or source	16S rDNA	ORF 023 3D	ORF 147 3D	ORF 117 3D	ORF 118 3D
TUH 51-66	229710	*E.faecalis*	**6 (2)**	Urine	1992	Portugal	Mikalsen et al, 2013	+	**0**	**0**	**1**	**0**
TUH 50-66	UW1833	*E.faecalis*	**40(40)**	Urine	1998	Germany	Mikalsen et al, 2013	+	**0**	**1**	**0**	**1**
11-E-4	50710071	*E. faecalis*	**ND**	Urine	2013	Norway	K-Res	+	**0**	**0**	**0**	**0**
TUH 50-58	1638/07	*E faecalis*	**40(40)**	Urine	2007	Denmark	Mikalsen et al, 2014	+	**0**	**0**	**0**	**0**
TUH 50-59	3992/09	*E.faecalis*	**40(40)**	Urine	2009	Netherlands	Mikalsen et al, 2014	+	**0**	**0**	**0**	**0**
TUH 50-74	UW7782/" 435/96"	*E faecalis*	**40(40)**	Urine	1996	Poland	Mikalsen et al, 2014	+	**0**	**0**	**0**	**0**
TUH 50-57	916/07	*E.faecalis*	**40(40)**	Faeces	2007	Poland	Mikalsen et al, 2013	+	**1**	**1**	**1**	**1**
K58-54	11290601	*E.faecalis*	**NY**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**1**	**1**	**1**
TUH 50-55	1665/07	*E faecalis*	**6 (2)**	Human blood	2007	Denmark	Mikalsen et al, 2013	+	**1**	**0**	**1**	**1**
TUH 50-42	E3450/605 785	*E.faecalis*	**6 (2)**	Human blood	2006	Netherlands	Mikalsen et al, 2013	+	**0**	**0**	**1**	**1**
TUH 50-54	340/07	*E.faecalis*	**6 (2)**	Human blood	2006	Poland	Mikalsen et al, 2013	+	**0**	**0**	**1**	**1**
TUH 50-69	UW7001	*E faecalis*	**6 (2)**	Human blood	2006	Germany	Mikalsen et al, 2013	+	**0**	**0**	**1**	**1**
K58-45	et7	*E.faecalis*	**21**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**1**	**1**
K58-55	92774	*E.faecalis*	**28**	Human blood	2008	Norway	Rosvoll et al 2013	+	0	0	1	1
K58-37	41996	*E faecalis*	**6**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-42	8569	*E faecalis*	**245**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-44	et2	*E.faecalis*	**447**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-46	et8	*E.faecalis*	**SLV av 451**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-48	et25	*E.faecalis*	**448**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**1**
K58-49	10024691	*E.faecalis*	**451**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-57	62883	*E faecalis*	**30**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-62	50095526	*E.faecalis*	**273**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-71	908301544	*E faecalis*	**449**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-72	908302487	*E.faecalis*	**450**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-77	656406	*E faecalis*	**21**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K58-78	658545	*E faecalis*	**16**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-02	75753	*E faecalis*	**268**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**

K59-04	691	*E.faecalis*	**55**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-06	16301	*E faecalis*	**179**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-08	34425	*E.faecalis*	**SVL av 16**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-09	35015	*E.faecalis*	**211**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-12	1410807275	*E faecalis*	**SVL av 30**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
K59-13	1410813435	*E.faecalis*	**28**	Human blood	2008	Norway	Rosvoll et al 2013	+	**0**	**0**	**0**	**0**
TUH 2-06	V583	*E faecalis*	**6(2)**	Human blood	1981	USA		+	**0**	**0**	**0**	**0**
		34							**2**	**3**	**9**	**10**

Strain ID	Name	*Skpecies*	ST	Source	Year	Origin	Reference	16SrRNA	ORF023	ORF147	ORF1172	ORF118
K59-17	18317	E. faecium	22	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K60-42	1410820462	E. faecium	22	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K59-44	43134	E. faecium	32	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K59-49	129766	E. faecium	32	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K59-19	36488	E. faecium	296	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K59-46	64750	E. faecium	533	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K60-09	506120	E. faecium	579	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K60-21	87205	E. faecium	580	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
K60-25	18580	E. faecium	581	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	0	0
TUH 50-45	U0229/O0160/506619	E. faecium	17	Human blood	1995	The Netherlands	Mikalsen et al, 2014	+	1	0	0	1
TUH 50-65	VRE-10	E. faecium	17	Human blood	2004	Denmark	Mikalsen et al, 2014	+	1	0	0	0
TUH 51-67	E1463/A5FLR(vre)	E. faecium	17	Human blood	1998	Spain	Mikalsen et al, 2014	+	1	0	0	0
K59-67	50103605	E. faecium	18	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	0	0
K59-57	10527341	E. faecium	38	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	0
K60-02	908301850	E. faecium	52	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	0
K59-26	et10	E. faecium	94	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	0
K59-48	115920	E. faecium	94	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	0
K59-50	908900774	E. faecium	202	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	0	0
K59-52	10181971	E. Faecium	576	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	0
K45-11	O2T878	E. faecium	17	Human blood	1998	Sweden		+	1	1	0	0
K60-36	1410804500	E. faecium	18	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	0	0
K60-29	31229	E. faecium	19	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	0	0
K59-58	10852801	E. faecium	132	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	0	0
K60-40	1410811268	E. faecium	202	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	1	1
K59-71	50117412	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	1	0
K59-55	10454581	E. faecium	279	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	0	1	1

K59-73	50150755	E. faecium	440	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	0	1
K59-18	21864	E. faecium	574	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	1	0
11-C-4	50708930	E. faecium	ND	Human blood	2012	Norway	Linezolid, Ålesund	+	1	1	0	0
K59-28	20575	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	0	1
K60-27	27719	E. faecium	18	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	1	1
K59-30	22135	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	1	1
K60-22	14473	E. faecium	202	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	0	1	1
K60-26	26495	E. faecium	202	Human blood	2008	Norway	Rosvoll et al, 2012	+	0	1	1	1
K59-37	14944	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	0	1
K59-27	et20	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-63	79932	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-08	657214	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-13	20167	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-20	73323	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-30	39254	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-33	1410803517	E. faecium	17	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
TUH 50-68	UW6900	E. faecium	17	Human blood	2005	Germany	Mikalsen et al, 2014	+	1	1	1	1
K59-51	10146163	E. faecium	18	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-15	40648	E. faecium	18	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-22	29966	E. faecium	78	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-77	50154026	E. faecium	78	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-78	7101	E. faecium	78	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
TUH 50-48	U0262/X2209	E. faecium	78	Human blood	2002	The Netherlands	Mikalsen et al, 2014	+	1	1	1	1
TUH 50-67	UW6847	E. faecium	78	Human	2006	Germany	Mikalsen et al, 2014	+	1	1	1	1

				blood								
K59-34	104499	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-42	36783	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-12	9339	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-14	38459	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-39	1410810681	E. faecium	192	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-05	653294	E. faecium	202	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-20	5977	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-21	20641	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-25	418541	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-29	20972	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-32	54271	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-33	78135	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-40	30697	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-54	10403171	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-04	651578	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-31	1410801937	E. faecium	203	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-62	81163	E. faecium	282	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-16	89332	E. faecium	440	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-36	5282	E. faecium	575	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K59-68	50108366	E. faecium	577	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
K60-07	656443	E. faecium	578	Human blood	2008	Norway	Rosvoll et al, 2012	+	1	1	1	1
		71							**53**	**50**	**44**	**46**

ID	Name	*Species*	ST	Soruce	+	Year	Origin	Reference	16srDNA	ORF023	ORF147	ORF117	ORF118
TUH 12-01	C68	*E.faecium*	16	human clinical specimen	+	1996	USA	(Carias et al., 1998)	+	1	0	0	0
TUH 32-62	E0510	*E. faecium*	17	human clinical specimen	+	1998	Australia	Willems, R. J. L.	+	0	1	0	1
TUH 50-62	VRE-106	*E. faecium*	78	Urine	+	2007	Denmark	Mikalsen et al, 2014	+	0	0	0	0
TUH 07-15		*E. faecium*	16	human clinical specimen	+	not available	USA	(Dahl *et al*., 1999)	+	1	0	0	0
TUH 02-18		*E. faecium*	17	human clinical specimen	+	1996	Norway	(Dahl *et al*., 2003)	+	1	0	0	0
TUH 07-55		*E. faecium*	17	human clinical specimen	+	not available	Germany	(Dahl et al., 1999)	+	1	0	0	0
TUH 32-74	E0300	*E. faecium*	20	human clinical specimen	+	1994	USA	Willems, R. J. L.	+	1	0	0	0
TUH 32-78	E0125	*E.faecium*	21	human clinical specimen	+	1995	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-63	E1626	*E.faecium*	92	human clinical specimen	+	1965	The Netherlands	Willems et al, 2011	+	1	0	0	0
TUH 12-04	A0885	*E.faecium*	308	human clinical specimen	+	1996	USA	(Carias *et al*., 1998)	+	1	0	0	0
TUH 04-65	A0884	*E. faecium*	313	human clinical specimen	+	not available	USA	(Dahl *et al*., 1999)	+	1	0	0	0
TUH 02-13		*E.faecium*	ND	human clinical specimen	+	1996	Norway	(Haarr *et al*., 1998)	+	1	0	0	0
TUH 41-64	U37	*E.faecium*	ND	human clinical specimen	+	1996	USA	(Grady and Hayes, 2003)	+	1	0	0	0
K08-50		*E.faecium*	ND	human clinical specimen	+	2001	Norway	Rosvoll *et al.* 2010	+	1	0	0	0
TUH 32-56	TX0016 (DO)	*E.faecium*	18	human clinical specimen	+	1992	USA	Arduino *et. Al.*, 1994	+	1	1	0	0
TUH 32-64	E0013	*E. faecium*	18	human clinical specimen	+	1992	UK	Willems, R. J. L.	+	0	0	1	1
9-C-2	50660268	*E.faecium*	ND	Urine	+	2012	Norway	K-Res	+	1	1	0	0
K08-45		*E.faecium*	ND	human clinical specimen	+	2003	Norway	Rosvoll *et al.* 2010	+	1	0	1	1
TUH 50-50	NIZP292/02	*E.faecium*	17	Wound	+	2002	Poland	Mikalsen et al, 2014	+	1	0	1	1
10-B-6	50691662	*E.faecium*	ND	Urine	+	2013	Norway	K-Res	+	1	1	1	1
TUH 32-69	E0470	*E.faecium*	16*	human faeces hospital outbreakspital	+	1999	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-72	E0155	*E.faecium*	17*	human faeces hospital outbreakspital	+	1995	USA	Willems, R. J. L.	+	1	0	0	0
TUH 32-75	E0073	*E.faecium*	22	human faeces hospital outbreakspital	+	1995	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-79	E0729	*E.faecium*	5	human faeces hospital outbreakspital	+	2000	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-73	E1652	*E.faecium*	18*	human faeces hospital outbreakspital	+	2002	The Netherlands	Willems, R. J. L.	+	0	0	1	1
K55-27	K55-27	*E.faecium*	78	faeces	+	2008	Sweden	Mikalsen et al, 2014	+	0	1	1	1
11-G-1	50701700	*E.faecium*	ND	Faeces	+	2013	Norway	K-Res	+	1	1	1	0
11-A-3	50705414	*E.faecium*	ND	Faeces		2013	USA	K-Res	+	1	1	1	1
10-C-9	50698098	*E.faecium*	ND	Feaces		2013	Norway	K-Res	+	1	1	1	1

10-E-7	50694469	*E. faecium*	ND	Feaces		2013	Norway	K-Res	+	1	1	1	1
	Total	**30**								**25**	**9**	**10**	**10**
TUH 44-31	A0888	E. faecium	311			1998	Norway	(Johnsen *et al*., 2005)	+	0	0	0	0
TUH 44-39	A0249	E. faecium	241	animal faeces animal faeces		1999	Norway	(Johnsen *et al*., 2005)	+	0	0	0	0
TUH 44-23		E. faecium	8	human faeces community		1998	Norway	(Johnsen *et al*., 2005)	+	0	0	0	0
TUH 45-03	A0889	E. faecium	312	human faeces community		1998	Norway	(Johnsen *et al*., 2005)	+	0	0	0	0
TUH 44-40	A0887	E. faecium	310	animal faeces	1	1999	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 44-50	A0251	E. faecium	195		1	1998	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 45-12	A0886	E. faecium	309		1	1999	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 45-30		E. faecium	247	animal faeces animal faeces human faeces	1	1999	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 32-67	E0403	E. faecium	7	human faeces community	1	1997	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-68	E0996	E. faecium	47	human faeces community	1	1998	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 32-80	E0092	E. faecium	6	human faeces community	1	1997	The Netherlands	Willems, R. J. L.	+	1	0	0	0
TUH 41-67		E. faecium	18*	human faeces community	1	2002	Italy	(Biavasco *et al*., 2007)	+	1	0	0	0
TUH 44-47		E. faecium	246	human faeces community	1	1999	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 45-05		E. faecium	48	human faeces community	1	1999	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
TUH 45-25		E. faecium	60	human faeces community	1	1998	Norway	(Johnsen *et al*., 2005)	+	1	0	0	0
	Total	**15**								**11**	**0**	**0**	**0**

Master table B: Primers used in gap PCR

5' contig primer	**Tm**	**G+C**	**3' contig primer**	**Tm 2**	**G+C 3**	**PCR reactions:**
3D_Plasmid_006.rev			3D_Plasmid_202.rev			3D_1
atggtggatctttagctggaa	59	43	aaggcgaaaagacaaataagga	59	37	
3D_Plasmid_079.rev			3D_Plasmid_231.rev			3D_2
tcacatctatgcattgacactca	59	39	atcgtgaccccagcattaga	60	50	
3D_Plasmid_112.forw			3D_Plasmid_266.forw			3D_3
gaatgccaactgcgatatga	60	45	tttcgacccaccgtttattc	60	45	
3D_Plasmid_147.forw			3D_Plasmid_225.forw			3D_4
gctcgaatagcagggaacg	61	58	cgccatcgataaacgattca	62	45	
3D_Plasmid_148.rev			3D_Plasmid_198.forw			3D_5
ggctacacaaagaatttcaacaga	60	38	gcaagcacagtaagtgtacgg	59	53	
3D_Plasmid_114.forw			3D_Plasmid_112.forw			3D_6
cctgttgttccttggcgtat	60	55	gaatgccaactgcgatatga	60	45	
3D_Plasmid_114.forw			3D_Plasmid_266.forw			3D_7
cctgttgttccttggcgtat	60	55	tttcgacccaccgtttattc	60	45	
3D_Plasmid_002.forw			3D_Plasmid_155.rev			3D_8
attggatacgtggcttttgg	60	45	gaacgaaatactgaaccccaag	60	45	
3D_Plasmid_002.forw			3D_Plasmid_159.forw			3D_9
attggatacgtggcttttgg	60	45	tctgacggcttattcggagt	60	50	
3D_Plasmid_001.rev			3D_Plasmid_155.rev			3D_10
tggctatcagttaactcacgatt	58	39	gaacgaaatatgaaccccaag	60	45	
3D_Plasmid_001.rev			3D_Plasmid_159.forw			3D_11
tggctatcagttaactcacgatt	58	39	tctgacggcttattcggagt	60	50	
3D_Plasmid_083.rev			3D_Plasmid_114.forw			3D_12
ccagcacacacgatgaaact	60	45	cctgttgttccttggcgtat	60	55	

3D_Plasmid_083.rev			3D_Plasmid_244.forw			3D_13
ccagcacacacgatgaaact	60	45	catttcgatgccagatttga	60	40	
3D_Plasmid_083.rev			3D_Plasmid_266.forw			3D_14
ccagcacacacgatgaaact	60	45	tttcgacccaccgtttattc	60	45	
3D_Plasmid_312.forw			3D_Plasmid_006.rev			3D_15
cgatcagttttgaggaaatgaa	59	36	atggtggatctttagctggaa	59	43	
3D_Plasmid_274.rev			3D_Plasmid_277.rev			3D_16
tgttacgatctccccaccat	60	50	agcggagtttgcgtcaatta	61	45	
3D_Plasmid_272.forw			3D_Plasmid_312.forw			3D_17
cacgctattgccgatctttc	61	55	cgatcagttttgaggaaatgaa	59	36	
3D_Plasmid_123.rev			3D_Plasmid_274.forw			3D_18
tgcttagggtcaccgatctt	60	50	agagatccaaacggcgttac	59	50	
3D_Plasmid_227.forw			3D_Plasmid_245.rev			3D_19
gccttagttgtcggtcttcg	60	55	cttcaaggataggttgccatt	58	43	
gccttagttgtcggtcttcg	60	55	cacgctattgccgatctttc	61	55	3D_20

Master table B: Primers used in gap PCR

5' contig primer	**TM**	**G+C**	**3' contig primer**	**TM2**	**G+C2**	**PCR reactions:**
3D_Plasmid_226.rev			3D_Plasmid_312.forw			3D_21
gaactacggggatagcttgg	59	55	cgatcagttttgaggaaatgaa	59	36	
3D_Plasmid_202b.rev			3D_Plasmid_006b.rev			3D_22
ttgtgcttctcccgaacttt	60	45	gattatttctgcgccagcat	60	45	
3D_Plasmid_312.forw			3D_Plasmid_272.forw			3D_23
cgatcagttttgaggaaatgaa	59	36	cacgctattgccgatctttc	61	55	
3D_Plasmid_270.forw			3D_Plasmid_269.rev			3D_24
gaaatagtacagatgtcgtaatga	51	33	cgaatagctggtccaaaactatg	44	53	
3D_Plasmid_109.rev			3D_Plasmid_108.forw			3D_25

tggtgccattacacataaaagg	60	41	ctacacaactccgcaaatcg	59	50	
3D_Plasmid_147.forw			3D_Plasmid_153.forw			3D_26
gctcgaatagcagggaacg	61	58	tgccatatctactccccctct	60	53	
3D_Plasmid_161.forw			3D_Plasmid_162.rev			3D_27
gctataaacccaaatgaactgg	58	41	gcggtatccaggagttcatt	59	50	
3D_Plasmid_160.rev			3D_Plasmid_114.rev			3D_28
aagacaaatgcacggtttaga	57	38	ccatagggctgtgtcaagt	60	55	
3D_Plasmid_114.forw			3D_Plasmid_266.forw			3D_29
cctgttgttccttggcgtat	60	55	tttcgacccaccgtttattc	60	45	
3D_Plasmid_114.forw			3D_Plasmid_244.forw			3D_30
cctgttgttccttggcgtat	60	55	catttcgatgccagatttga	60	40	
3D_Plasmid_114.forw			3D_Plasmid_244.forw			3D_31
cctgttgttccttggcgtat	60	55	catttcgatgccagatttga	60	40	
3D_Plasmid_199.forw			3D_Plasmid_244.forw			3D_32
tcatttcattcgccttttcc	60	40	catttcgatgccagatttga	60	40	
3D_Plasmid_241.rev			3D_Plasmid_240.forw			3D_33
attagggatgctgcaaatgg	60	45	caggaaacactgcttgtgga	60	50	
3D_Plasmid_231.rev			3D_Plasmid_079.rev			3D_34
atcgtgaccccagcattaga	60	50	tcacatctatgcattgacactca	59	39	
3D_Plasmid_155.rev			3D_Plasmid_159.forw			3D_35
gaacgaaatactgaaccccaag	60	45	tctgacggcttattcggagt	60	50	
3D_Plasmid_095.forw			3D_Plasmid_096.rev			3D_36
tggcattactcacctttttgg	60	42	atcggttggtttctgctttt	59	40	
3D_Plasmid_076.forw			3D_Plasmid_066b.rev			3D_37
aaggatcatgcttaagaataga caga	59	35	cctactatttccaatatcaatcaa tca	59	30	
3D_Plasmid_199.rev			3D_Plasmid_065.forw			3D_38
tgtgtaaatgagaaactgccaa a	60	35	tatgcgtgcgatttcatcat	60	40	
3D_Plasmid_225.forw			3D_Plasmid_076.forw			3D_39
cgccatcgataaacgattca	62	45	aaggatcatgcttaagaataga caga	59	35	

3D_Plasmid_225.forw			3D_Plasmid_066b.rev			3D_40
cgccatcgataaacgattca	62	45	cctactatttccaatatcaatcaatca	59	30	
3D_Plasmid_209.forw			3D_Plasmid_208.rev			3D_41
caagcaaccgaaagaaaaca	59	40	gcgggtttacctttttctgg	61	50	
3D_Plasmid_209.forw			3D_Plasmid_205.rev			3D_42
caagcaaccgaaagaaaaca	59	40	aaatgtctatgctatacggtgcaa	60	38	

Master table B: Primers used in gap PCR

5' contig primer	**Tm**	**G+C**	**3' contig primer**	**Tm2**	**G+C2**	**PCR reactions:**
3D_Plasmid_041.forw			3D_Plasmid_043.rev			3D_43
tgcgaataatttgcttctgttg	60	36	gcaaaccacccaatgaactc	60	50	
3D_Plasmid_054.forw			3D_Plasmid_055.rev			3D_44
ccagaagttgaggcgaagtg	61	55	tcaactaagccactccttttg	56	41	
3D_Plasmid_270.forw			3D_Plasmid_269b.rev			3D_45
gaaatagtacagatgtcgtaatga	51	33	tgccacttaaacccaaccat	45	60	
3D_Plasmid_161.forw			3D_Plasmid_162.rev			3D_46
gctataaacccaaatgaactgg	58	41	gcggtatccaggagttcatt	59	50	
3D_Plasmid_108b.forw			3D_Plasmid_109b.rev			3D_47
tcaagtggacattcattaagctg	60	39	ccacctgtaccagtcaattcaa	60	46	
3D_Plasmid_209b.forw			3D_Plasmid_208.rev			3D_48
gcaaagaaatacagccgaca	59	45	gcgggtttacctttttctgg	61	50	
3D_Plasmid_209b.for			3D_Plasmid_205.rev			3D_49

w						
gcaaagaaatacagccgaca	59	45	aaatgtctatgctatacggtgcaa	60	38	
3D_Plasmid_199.forw			3D_Plasmid_114.forw			3D_50
tcatttcattcgccttttcc	60	40	cctgttgttccttggcgtat	60	55	
3D_Plasmid_187.forw			3D_Plasmid_188.rev			3D_51
gcggcaaaagcttatcaaag	60	45	gcggtaaaatatactgatgtaaatgg	59	35	
3D_Plasmid_187.forw			3D_Plasmid_274.rev			3D_52
gcggcaaaagcttatcaaag	60	45	tgttacgatctccccaccat	60	50	
3D_Plasmid_079b.rev			3D_Plasmid_231.rev			3D_53
taaatgagctgccgaaaagc	60	45	atcgtgaccccagcattaga	60	50	
3D_Plasmid_153.forw			3D_Plasmid_108.forw			3D_54
tgccatatctactccccctct	60	53	ctacacaactccgcaaatcg	59	50	
3D_Plasmid_147.forw			3D_Plasmid_109.rev			3D_55
gctcgaatagcagggaacg	61	58	tggtgccattacacataaaagg	60	41	
3D_Plasmid_147.forw			3D_Plasmid_270.forw			3D_56
gctcgaatagcagggaacg	61	58	gaaatagtacagatgtcgtaatga	51	33	
3D_Plasmid_108.forw			3D_Plasmid_270.forw			3D_57
ctacacaactccgcaaatcg	59	50	gaaatagtacagatgtcgtaatga	51	33	
3D_Plasmid_232.rev			3D_Plasmid_079b.rev			3D_58
cttttgctggcagacagtga	60	50	taaatgagctgccgaaaagc	60	45	
3D_Plasmid_204.rev			3D_Plasmid_079b.rev			3D_59
ggcgatattcttggagcatt	59	45	taaatgagctgccgaaaagc	60	45	
3D_Plasmid_204.rev			3D_Plasmid_209b.forw			3D_60
ggcgatattcttggagcatt	59	45	gcaaagaaatacagccgaca	59	45	
3D_Plasmid_161b.forw			3D_Plasmid_162b.rev			3D_61
tggcaagctctaggattacctt	59	45	caactggaatgccttgcttt	60	45	
3D_Plasmid_160.forw			3D_Plasmid_160b.rev			3D_62

tcatcttcccaaggctctgt	60	50	tgaattacacgagggcaaaa	59	40	
3D_Plasmid_160.forw			3D_Plasmid_162b.rev			3D_63
tcatcttcccaaggctctgt	60	50	caactggaatgccttgcttt	60	45	
3D_Plasmid_161b.forw			3D_Plasmid_162b.rev			3D_64
tggcaagctctaggattacctt	59	45	caactggaatgccttgcttt	60	45	
3D_Plasmid_161.forw			3D_Plasmid_162.rev			3D_65
gctataaacccaaatgaactgg	58	41	gcggtatccaggagttcatt	59	50	
2C_Plasmid_229.forw			3D_Plasmid_160b.rev			3D_66
tcccgaatgaacaacctacc	60	50	tgaattacacgagggcaaaa	59	40	
3D_Plasmid_161b.forw						
tggcaagctctaggattacctt	59	45				
3D_Plasmid_160.forw						
tcatcttcccaaggctctgt	60	50				
3D_Plasmid_160b.rev						
tgaattacacgagggcaaaa	59	40				