Norwegian University of Life Sciences

Master's Thesis 2022 30 ECTS
Faculty of Chemistry, Biotechnology, and Food science

An investigation of the potential role of bacteriocins in the development of the infant gut microbiota

Frida Josefine Pettersson
MSc Chemistry and Biotechnology

An investigation of the potential role of bacteriocins in the development of the infant gut microbiota

Norwegian University of Life Sciences (NMBU),

Faculty of Chemistry, Biotechnology, and Food Science

Acknowledgement

This master's thesis was performed at the Faculty of Chemistry, Biotechnology, and Food Science, at the Norwegian University of Life Sciences, under the supervision of Professor Knut Rudi and PhD. Ida Ormaasen.

First, I want to thank my main supervisor Knut Rudi. Thank you for always helping me with challenges both in the lab and in the writing process and for always having your door open. Your positive attitude and scientific enthusiasm have truly been an inspiration for me during this working process.

I would also like to thank my secondary supervisor Ida Ormaasen for your eagerness to help in the lab and with the writing and for always answering my questions about anything. Your supportive and cheerful outlook has inspired and encouraged me throughout this working process.

Furthermore, I would like to thank the whole MiDiv group for creating such an inclusive working environment and for always sharing your knowledge. Especially thanks to PhD. Morten Nilsen and Laboratory Engineers Tonje Nilsen, Karen Sivertsen Utheim, and Inga Leena Angell for good conversations and always being so supportive and helpful.

Also, I would like to thank senior engineer Else Marie Aasen for your help with the cultivation of bacteria and PhD. Thomas Oftedal for providing bacterial strains and good advice.

Finally, thanks to all my friends and family, for always supporting and believing in me.

Ås, May 2022

Frida Josefine Pettersson

Abstract

The colonization of the infant gut microbiota is important for health and wellbeing later in life. This colonization is dependent on multiple factors, among others the competition between bacteria. Different strategies have been developed to gain an advantage over the other bacteria in the competition for space and nutrition. One such strategy is to synthesize bacteriocins, which are small ribosomally produced peptides that can inhibit the growth of or kill the competing bacteria. The genus *Streptococcus* is a known bacteriocin producer and an important part of the infant gut microbiota. This genus is among the early colonizers of the gut and has been found to dominate this microbiota during the first few days of the infant's life. Since little is known about the role of bacteriocins in the development of the infant gut microbiota, the main aim of this thesis was to investigate the associations between bacteriocins and *Streptococcus* in stool samples.

Seven bacteriocins were chosen for this study based on their significant presence in the gut of infants compared to adults and they were synthesized in vitro before their ability to inhibit bacteria belonging to the genera *Streptococcus, Enterococcus, Listeria,* and *Lactococcctus* was investigated. In addition, qPCR with designed primers was used to identify bacteriocin genes and 16S rRNA sequencing to identify the bacteria present in stool samples provided by the PreventADALL study. A total of 168 stool samples were studied, where 56 samples belonged to each of the three groups: infants at 6 months, infants at 12 months, and mothers.

In conclusion, Sanger sequencing and microscopy confirmed that the cultivation of the streptococcal bacteria was successful. The two bacteriocin activity assays showed that none of the bacteriocins inhibited the growth of the tested bacteria. This could most likely be caused by an unsuccessful bacteriocin synthesis or a classification error of the bacteriocins. The genus *Streptococcus* was significantly enriched in samples from infants compared to mothers, as expected from previous studies. The only significant association between the presence of bacteriocins and the number of streptococcal reads was found in samples from infants at 12 months of age. Where the presence of bacteriocins gave significantly more reads from the streptococcal OTU nr.2, but this association was not identified in samples from infants at 6 months of age or mothers. More associations could possibly be identified in samples taken from an earlier timepoint in the infant's life, and this should be considered in further studies.

Sammendrag

Koloniseringen av spedbarns tarmfloraen er viktig for helse og velvære senere i livet. Denne koloniseringen er avhengig av flere faktorer, blant annet konkurransen mellom bakterier. Ulike strategier er utviklet for å få en fordel i denne konkurransen om næring og plass. En strategi går ut på å produsere bakteriosiner som er små ribosomalt produserte peptider. Disse peptidene kan hemme veksten av eller drepe andre bakterier. Slekten *Streptococcus* er en kjent bakteriosin produsent og utgjør en viktig del av tarmfloraen til spedbarn. Denne slekten er blant de tidlige kolonisatorene og dominerer tarmmikrobiotaen de første dagene av spedbarnets liv. Lite er fortsatt kjent om rollen til bakteriosiner i utvikling av tarmfloraen og hovedmålet til denne oppgaven var dermed å undersøke om det var noen sammenheng mellom bakteriosiner og streptokokker i fekal prøver.

Syv bakteriosiner ble valgt ut til denne studien basert på deres signifikante tilstedeværelse i tarmfloraen til spedbarn sammenlignet med tarmfloraen til voksne og de ble syntetisert in vitro før testing av deres evne til å hemme veksten av bakterier som tilhørte slektene *Streptococcus, Enterococcus, Listeria* og *Lactococcus* ble undersøkt. I tillegg ble qPCR brukt til å identifisere bakteriosin gener og 16S rRNA sekvensering til å identifisere bakterier til stede i fekal prøver fra PreventADALL studien. Totalt ble 168 prøver studert og 56 prøver tilhørte hver av de tre gruppene: spedbarn ved 6 måneder, spedbarn ved 12 måneder og mødre.

For å konkludere, Sanger sekvensering og mikroskopering bekreftet at dyrkingen av streptokokkene var en suksess. I de to testene for bakteriosin aktivitet kom det fram at veksten til de testede bakteriene ikke ble hemmet av bakteriosinene. Dette kunne mest sannsynlig skyldes feil i bakteriosinsyntesen eller en klassifiseringsfeil av bakteriosinene. Videre ble det funnet at slekten *Streptococcus* var signifikant anriket i de samlede prøvene fra spedbarna sammenlignet med prøvene fra mødrene. Den eneste signifikante sammenhengen mellom tilstedeværelse av bakteriosiner og streptokokker ble funnet i prøver tatt fra 12 måneder gamle barn. Her ble det funnet at streptokokk OTU nr.2 var signifikant anriket i prøver hvor det var bakteriosiner til stede, men dette var ikke tilfellet i prøver tatt fra 6 måneder gamle barn eller mor. Flere assosiasjoner kan muligens identifiseres i prøver tatt ved tidligere tidspunkt av barnets liv og dette bør derfor vurderes i videre studier.

Abbreviations

AMPs	Antimicrobial peptides
BCGs	Biosynthetic gene cluster
BHI	Brain hearth infusion
BLAST	Basic Local Alignment Search Tool
CFPS	Cell-free protein synthesis
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
ddNTPs	Dideoxynucleotide triphosphates
dsDNA	Double stranded DNA
GRAS	Generally regarded as safe
HGT	Horizontal gene transfer
HRM	High resolution melting
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
LAB	Lactic acid bacteria
NGS	Next generation sequencing
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PreventADALL	Preventing Atopic Dermatitis and ALLergies
RMS	Reduced metagenome sequencing
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SCFAs	Short chain fatty acids
T_a	Annealing temperature
T_m	Melting temperature
TSYE	Tryptic soy yeast extract
qPCR	Quantitative PCR
QPS	Qualified presumption of safety

Table of contents

1. Introduction

1.1 The human gut microbiota

The human gut microbiota consists of the especially dense populations of microbes located in the human gut and together with all the microbes located in and on the human body, they outnumber our own cells by a factor of ten (Turnbaugh et al., 2007). The diversity in the gut is however quite low compared to for example a soil sample and it can therefore be seen as a quite competitive environment where the bacteria compete for both nutrition and space (Bäckhed et al., 2005). These microbes co-exist with their host in symbiotic, parasitic, or commensal ways. The commensal microbiota in the gut is occupying space that otherwise could be colonized by pathogens, they interact with the human immune system, metabolize nutrients that we otherwise could not digest, produce vitamins, and might even affect our behaviour (Milani et al., 2017, Lozupone et al., 2012, Turnbaugh et al., 2007). In other words, the composition of the human gut microbiota is important for our health.

1.1.1 Development of the infant gut microbiota

There is still a debate going on concerning the colonization of the infant gut microbiota. For decades "the sterile womb" paradigm has been used to describe the colonization, but other hypotheses have now gained momentum. One of the proposed hypotheses is that the colonization might already start with the rupture of the amniotic membrane in utero (Rehbinder et al., 2018). The next few days of the infant's life will consist of rapid colonization of the gut and after two to three years a stable adult-like microbiota will be reached (Koenig et al., 2011). Once the stable conformation is reached, it is harder to influence it. The colonization of the gut microbiota during the first two-three years of the infant's life is therefore seen as a window of opportunity to influence the colonization and thereby also influence the health of the infant throughout life (Cox et al., 2014).

Stool samples taken from infants at different time points in their lives show this development of gut microbiota from a habitat with few inhabitants to a more adult composition with higher alpha diversity, which means that the diversity within the individual increases (Sanidad and Zeng, 2020). The first samples can be taken from meconium and these samples show that the earliest gut microbiota is dominated by the genus *Streptococcus* and these bacteria colonize the gut within 24 hours after birth (Gosalbes et al., 2013, Solís et al., 2010). Stool samples

taken after 10, 30, and 90 days are dominated by *Bifidobacterium*, but the genera *Streptococcus*, *Lactobacillus*, and *Enterococcus* follow straight behind (Solís et al., 2010).

The development and establishment of the infant gut microbiota are influenced by many distinct factors and one of the possible most crucial factors is the mode of delivery. Infants delivered vaginally will come into contact with the mother's vaginal and faecal microbes and be colonized by these, while infants delivered by caesarean section will come into contact with skin microbes and microbes from the hospital environment and be colonized by these (Avershina et al., 2016, Milani et al., 2017). It will therefore be possible to detect differences in the gut microbiota in infants delivered vaginally and infants delivered by caesarean section both in the composition of strains and in the abundance of the different strains of bacteria. Some of the differences that have been observed in several studies are that the genus *Bacteroides* seems to be underrepresented, while the genus *Enterococcus* seems to be overrepresented in infants delivered by caesarean section compared to infants delivered vaginally (Ravi et al., 2018, Jakobsson et al., 2014). As well as leading to the colonization of other bacteria, the caesarean section mode of delivery has also been associated with a greater risk for diseases such as diabetes, celiac disease, asthma, and allergies (Cho and Norman, 2013, Decker et al., 2011, Thavagnanam et al., 2008).

The colonization of the gut is also dependent on the gestational age of birth and it is found that children that are born preterm, which is defined as before 37 weeks of gestation, often have serious health issues (Ruiz et al., 2016). Several studies have found that the colonization of commensal microbes such as *Bifidobacterium* and *Bacteroides* often are delayed in infants delivered preterm ((Hill et al., 2017, Arboleya et al., 2012). This delay gives opportunistic bacteria such as *Enterococcus* and *Enterobacteria* the possibility to grow and establish themselves (Arboleya et al., 2012, Arboleya et al., 2015).

Another major factor that affects gut colonization is the infant feeding mode. The feeding mode will both affect the establishment of the gut microbiota and the gastrointestinal functions (Dominguez-Bello et al., 2011). Antimicrobial agents, promicrobial agents, and a range of different nutrients are present in breast milk and it has been shown that infants fed with breast milk have increased levels of bifidobacteria (O'Sullivan et al., 2015). Formula-fed infants are exposed to a different set of nutrients, and this results in a different colonization pattern in their gut. These infants have a gut microbiota that is more associated with bacteria such as *Staphylococci*, *Bacteroides*, and *Clostridia* (Penders et al., 2006). Factors such as the

introduction of solid food, geographical location, and family situation will also affect the gut microbiota later in life.

1.1.2 The adult gut microbiota

The adult gut microbiota has a relatively stable configuration, but it can still be affected by antibiotic use or dramatic changes in diet among other things (Faith et al., 2013, Thursby and Juge, 2017). Common for healthy individuals is that the gut microbiota is highly variable and especially the two bacterial phyla *Firmicutes* and *Bacteroidetes* are highly represented, but also bacteria from the phyla *Actinobacteria, Proteobacteria, Verrucomicrobia,* and *Fusobacteria* are represented in many healthy gut microbiotas (Lozupone et al., 2012, Eckburg et al., 2005).

There are metabolic, trophic, and protective benefits of having a healthy gut microbiota. The microbes in the gut harbour a high gene diversity that provides the host with various biochemical pathways and enzymes that can be used to digest and utilize diet in a way that would not else be possible for the host (Guarner and Malagelada, 2003). In the colon, the microorganisms are also responsible for a lot of the absorption of iron, magnesium, and calcium and the production of vitamins, such as vitamin K (Younes et al., 2001, Miyazawa et al., 1996, Guarner and Malagelada, 2003, Hill et al., 2017). The trophic effect on the intestinal epithelium is connected to the production of short-chain fatty acids (SCFAs) as a by-product of the digestion by the microorganisms (Guarner and Malagelada, 2003). SCFAs have been shown to stimulate the epithelial cells to differentiation and proliferation both in the small and the large bowel (Frankel et al., 1994). The third major benefit of the gut microbiota is its ability to protect the host against the colonization of pathogens by occupying space, competing for nutrients, and producing antimicrobial substances (Guarner and Malagelada, 2003).

The composition of the gut microbiota is highly divergent between individuals and can be seen as a “fingerprint” for each individual (Lozupone et al., 2012). More similarities in the composition are however found between mother-daughter pairs and between twins, than between unrelated individuals (Dicksved et al., 2008, Turnbaugh et al., 2009). There have also been found similarities between people with conditions such as obesity. Obese people seem to harbour fewer microbes in the gut and have a different abundance of functional genes and taxa compared to lean people (Ley et al., 2006b, Turnbaugh et al., 2006). The low microbial diversity has also been observed in people with Crohn’s disease compared to healthy individuals (Dicksved et al., 2008). Diseases such as diabetes, inflammatory bowel disease

(IBD) and irritable bowel disease (IBS) have also been associated with disturbances in the gut microbiota (Casén et al., 2015).

1.1.3 Ecological interactions in the adult gut microbiota

Even though the gut is tightly colonized by microorganisms, it is also exclusive and the selection pressure upon its residents is strong (Bäckhed et al., 2005). To be able to thrive in the gut, the microorganisms must among other things be able to utilize the available nutrients and they must be able to settle in the right habitat (Ley et al., 2006a). In other words, they must both compete with each other and avoid getting killed by the immune system. This results in selection pressure on two levels. One level can be seen as the "top-down" selection where the host's immune system favours a stable microbial community with functional redundancy. The second level can be seen as the "bottom-up" selection where the microbes compete with each other and this favours more specialized functions (Ley et al., 2006a). Which of these two levels that are the most important for shaping the microbial community in the gut is still unknown, but the "bottom-up" selection is probably a very important factor during the early stages of life since the immune system is not yet fully developed (Trosvik et al., 2010).

1.1.4 The role of the bacterial competition in the establishment of the gut microbiota

The colonization and establishment of the gut microbiota from a sterile environment to a highly complex and densely packed community happens in a relatively short amount of time. During this period, the microorganisms will interact with each other in different ways, and this will in turn give a microbiota that can be associated with health or disease and that can influence the development of the infant's immune system and metabolism (Dogra et al., 2015, Collado et al., 2012, Kerr et al., 2015). The two dominant modes of interaction in the gut are cooperation and competition (Faust and Raes, 2012). To cooperate with other microorganisms might give a more effective utilization of resources, but it can also make the species dependent on each other and if one of the species disappears, the other will likely follow. Cooperation can therefore give unstable communities, and the host might benefit from competition between species if this destabilizes the network of cooperation (Coyte et al., 2015).

The different microbial species in the gut will likely compete for the few niches available and microbes with similar preferences in niches will likely follow Gause's law of competitive exclusion and exclude each other (Gause, 2019). The competition can be against species within the same phylum or species belonging to other phyla. One phylum that has been

observed to have quite a lot of intra-phylum competition is Firmicutes and it is suggested that some of the founder bacteria in the gut belong to this phylum (Trosvik et al., 2010). The high initial growth rate of the founders enables them to swiftly colonize vacant niches, but they are less likely to sustain high competitive pressure in dense populations (Trosvik et al., 2010).

Microorganisms in the gut have developed different strategies to gain an advantage in this competition. One strategy is to utilize the host's immune system to inhibit other microbes (Coyte and Rakoff-Nahoum, 2019) . Another strategy is to use a type IV secretion system where a toxic payload is transferred to other bacteria. This strategy has been observed in many of the Bacteroidales strains from human microbiota samples and analysis has shown that these genes have an important role in the interspecies competition in the human gut (Coyne et al., 2016, Wexler et al., 2016, Chatzidaki-Livanis et al., 2016). A third and widely used strategy amongst bacteria is to produce antimicrobial peptides called bacteriocins. These peptides usually work against closely related species and this can be an advantage since bacteria in the same phyla can be expected to compete because of similar function mechanisms and similar preferences regarding nutrients and habitats (Coyte and Rakoff-Nahoum, 2019).

1.2 Bacteriocins

In 1925, Andre Gratia discovered an unknown substance produced by *Escherichia coli* that killed closely related bacteria. This substance was named Colicin V and was the first bacteriocin to be discovered (Gratia, 2000). Bacteriocins are defined as antimicrobial peptides (AMPs) produced ribosomally by bacteria and some Archaea (Riley and Wertz, 2002). Both Gram-negative and Gram-positive bacteria produce bacteriocins. Bacteriocins produced by Gram-negative bacteria are often found to be relatively large proteins, while bacteriocins produced by Gram-positive bacteria often contain less than 70 amino acids (Zheng et al., 2015).

1.2.1 Classification of bacteriocins

The family of bacteriocins is a diverse family where the members have different microbial targets, size, immunity actions, and modes of action (Riley and Wertz, 2002). The Gram-positive bacteria that produce bacteriocins are mostly in the group called Lactic acid bacteria (LAB). The classification of bacteriocins varies somewhat in the literature, but the following classes are proposed by Alvarez-Sieiro et al. : class I (lanthionine-containing), class II (non-lanthionine containing) and class III (Alvarez-Sieiro et al., 2016). Class I contains

bacteriocins that are less than 10 kDa and are enzymatically modified, which impacts their properties (Alvarez-Sieiro et al., 2016). Class II bacteriocins are also under 10 kDa, they are not enzymatically modified and do therefore not require other enzymes than a transporter and/or a leader peptidase for maturation (Alvarez-Sieiro et al., 2016). Class III bacteriocins are unmodified peptides over 10 kDa (Alvarez-Sieiro et al., 2016).

Class I can be divided into six subgroups (a-f) that all have different modifications, for instance, use of unusual amino acids or glycosylated residues, which gives them different properties (Alvarez-Sieiro et al., 2016). Class II can be divided into four subgroups (a-d), whereas class IIa consists of peptides that have a flexible hinge that binds the two regions of the peptide together and can be divided further into eight subgroups (Alvarez-Sieiro et al., 2016). Class IIb are bacteriocins that consist of two different peptides and only achieve full activity in the presence of both peptides in equal amounts (Nissen-Meyer et al., 2010). Class IIc is leaderless and class IId is a heterogeneous group of unrelated single linear peptide bacteriocin with differences in structure, function, and mechanisms (Alvarez-Sieiro et al., 2016).

1.2.2 Nature and function of bacteriocins

For the production of bacteriocins to be beneficial, it needs to have a greater fitness advantage than the metabolic cost of production. It also needs to spare mutualistic partner strains and avoid resistance developing in the target cells (Heilbronner et al., 2021). The genes responsible for bacteriocin production are gathered in complex and variable biosynthetic gene clusters (BCGs). These BCGs not only encode the bacteriocin, but also biosynthetic enzymes, export proteins, production of immunity mechanisms, and sometimes even regulators for the production of bacteriocins (Alanjary et al., 2017). The BCGs are rapidly evolving and are frequently exposed to horizontal gene transfer (HGT) (Heilbronner et al., 2021). For instance, a phage-mediated manner of HGT has been observed in bacteriocin encoding genetic islands in *Vibrio cholerae* (Thomas et al., 2017). Many of the BCGs are associated with mobile genetic elements such as plasmids or genetic islands with a GC content that can be distinguished from the average of the rest of the genome (Zipperer et al., 2016).

Bacteriocins are seen as quite stable molecules and they use highly diverse mechanisms for their antimicrobial activity. Some bacteriocins work by degrading essential components, while others inhibit specific molecular targets for example molecules necessary for bacterial cell wall synthesis (Heilbronner et al., 2021). Most bacteriocins show antimicrobial activity against closely related species and bacteriocins like microcin and colicins are shown to have a

small target spectrum because of the dependence of cellular receptors on the targets (Heilbronner et al., 2021, Baquero et al., 2019). Small bacteriocins produced by Gram-positive bacteria seem to have a broader spectrum and to be able to affect unrelated bacteria (Sánchez-Hidalgo et al., 2011, Liu et al., 2020). This kind of cross-phyla activity of bacteriocins produced by Gram-positive bacteria has been observed in *Enterococcus, Blautia, Lactobacillus,* and *Staphylococcus* isolates (Simons et al., 2020, Kim et al., 2019, Corr et al., 2007, Janek et al., 2016).

The production of bacteriocin comes at a high metabolic cost and it can affect the growth rate of the producer (Maldonado-Barragán and West, 2020). As well as the metabolic cost tied directly to the production, there is also a cost associated with the maintenance of the BCGs, the transfer of BCGs to the next generation of bacteria, and the self-resistant mechanisms (Heilbronner et al., 2021). The production of bacteriocins is therefore tightly regulated and a system called quorum sensing is often used for this purpose. This system gives the bacteria the possibility to sense the growth of competitors compared to their growth and to communicate and cooperate (Zheng et al., 2015, Nes et al., 1996). The bacteriocins can also be used to interact with the host and it is observed that the cytokine production of dendritic cells can be modulated by plantaricin, which is a bacteriocin produced by *Lactobacillus plantarum* (Meijerink et al., 2010, Zheng et al., 2015).

It is generally acknowledged that bacteria produce bacteriocins to gain an advantage against their competitors (Ghoul and Mitri, 2016). The advantage can be tied to the competition for nutrition, like carbon, phosphorus, and so on, or it can be tied to the competition for space for example in the human body. The killing of the competition will also result in the release of cellular components to the environment and both *Streptococcus pneumonia* and *Vibrio cholerae* are shown to take up DNA from the lysed target bacteria (Shanker and Federle, 2017, Borgeaud et al., 2015). It is also possible that the bacteriocins can be used to kill target bacteria for the release of essential nutrients (Brugger et al., 2020).

1.2.3 Application of bacteriocins

The number of characterized bacteriocins is increasing rapidly and it is estimated that more than 99 % of bacteria can produce at least one bacteriocin (Riley and Wertz, 2002). In addition, many LAB bacteria are generally regarded as safe (GRAS) and possess the status of Qualified Presumption of Safety (QPS), and their bacteriocins are therefore also regarded as safe to use (Alvarez-Sieiro et al., 2016). The number of new antibiotics to be introduced is however low and the use of broad-spectrum antibiotics will also affect the commensal

bacteria and cause unbalance (Gabant and Borrero, 2019, Villarreal et al., 2012). Since bacteriocins often attack specific closely related bacteria of the producer, they have been seen as less intrusive than broad-spectrum antibiotics (Mills et al., 2017). These high potency and low toxicity peptides can be bioengineered and produced in situ by probiotics (Cotter et al., 2013).

Since bacteriocins are seen as a natural product, they are already used to increase the shelf life of different foods and both nisin and pediocin PA-1 are used as food additives (Alvarez-Sieiro et al., 2016). Bacteriocin can also control food-born pathogenic bacteria by being added to packaging films. This prevents the degrading of the bacteriocin by the food components, by gradually releasing the bacteriocin into the food (Guerra et al., 2005). As well as having an antimicrobial effect, some bacteriocins have also shown an effect on cancer cells (Yang et al., 2014) and a study on zebrafish showed that the bacteriocin microcin E492 reduced the size of the tumours in zebrafish larvae (Varas et al., 2020).

1.2.4 Bacteriocins produced by the genus *Streptococcus*

Bacteria from the genus *Streptococcus*, which belong to the phyla Firmicutes, are facultative anaerobic and usually occur in chains or pairs (Davey et al., 2016). Streptococci can frequently be found in the mucosa that covers the human body and some of them are harmless and regarded as safe, while others are pathogenic to humans and animals (García-Curiel et al., 2021). This genus is often studied for its competence and for its ability to produce antimicrobial peptides, such as bacteriocins to kill competition (García-Curiel et al., 2021).

The members of the genus *Streptococcus* can produce bacteriocins belonging to all the three classes previously described, but it is the class I and class II bacteriocins that dominate (García-Curiel et al., 2021). The *blp* gene cluster consists of at least two operons, encodes class II bacteriocins, and is responsible for the bacteriocin production in both *S.pneumoniae* and *Streptococcus thermophilus (*García-Curiel et al., 2021*).* The first operon is called *blpABC* and it is responsible for the ABC transporter protein with associated ATPases and BlpC, which is a peptide for induction of quorum detection (Renye et al., 2019). The second operon is called *blpHR* and encodes BlpH, which is the response regulator, and BlpR, which is a histidine kinase (Wang and Dawid, 2018).

The streptococci produce bacteriocins for different purposes. For instance, *Streptococcus mutans* produces bacteriocins for defence, while *Streptococcus salivarius* uses bacteriocins for predation (Lemos et al., 2019, Mignolet et al., 2018). The bacteriocins produced for

predation will destabilize the target cell by generating perforations in the membrane and this results in the death of the target cell. Cellular material will then be released to the surroundings and can be taken up by *S.salivarius*, which then can acquire new DNA (Talagas et al., 2016).

1.3 Approaches to study the human gut microbiota

Even though microorganisms are abundant and can be found everywhere, the mechanical knowledge of the many key roles they play is lacking (Biteen et al., 2016). Only a small fraction of the gut microbiota in humans have successfully been isolated and studied as pure cultures (Biteen et al., 2016, Rajilić-Stojanović and De Vos, 2014). Therefore, methods like metagenomics, metatranscriptomics, and metaproteomics have become increasingly important for identifying the microbes in complex samples and for understanding their functional roles.

1.3.1 Cultivation of bacteria

Different cultivation techniques can be used to grow and maintain stable populations of bacteria under laboratory conditions. To be successful with the cultivation, general knowledge of the bacteria of interest must first be obtained. Bacterial species have different preferences when it comes to the presence of oxygen, nutrients, optimal pH, and optimal temperature. The different oxygen preferences can be obtained by using atmosphere generation systems, which makes it possible to grow bacteria under aerobic, anaerobic, and microaerophilic conditions.

To meet the different nutritional needs of the bacteria, diverse types of media can be used. Media that are selective can promote or suppress the growth of certain species, while non-selective media can promote the growth of a wide range of species (Bonnet et al., 2020). The chosen medium can be used in a solid form if agar is added to the formula or in a liquid form as a broth culture. Broth cultures have been found to obtain a greater mass of bacteria because the bacteria have better access to nutrients and waste products are diluted compared to in a solid medium (Bonnet et al., 2020). To be able to store bacteria over longer amounts of time, they may be grown in a liquid media and then added to glycerol which permits storage at -80 degrees

1.3.2 Metagenome analysis

A standard method for characterizing complex microbial communities is by using high-throughput sequencing on the 16S ribosomal RNA (rRNA) gene (Nguyen et al., 2016). All bacteria express the functionally consistent and experimentally manageable 16S rRNA gene and it can therefore be used to amplify all bacteria present in a sample (Winand et al., 2019,

Angell et al., 2020). The focus of the 16S rRNA analysis is on one gene, while a shotgun metagenome sequencing approach potentially will sequence all DNA fragments present in the sample. This approach has been found to gather substantially more information and to give rise to less bias than the 16S rRNA sequencing approach (Milani et al., 2017). The bottleneck of this method is however that it requires substantial processing power to be able to interpret all the data obtained and it is too expensive in many cases. Methods such as Reduced Metagenome Sequencing (RMS) that sequence only parts of the metagenome, have therefore been developed (Ravi et al., 2018).

In 1977, Fredrick Sanger and his colleagues published a method called Sanger sequencing which was a part of the first-generation sequencing (Sanger et al., 1977). The method is based on the use of monomers of DNA strands called deoxynucleotide triphosphates (dNTPs) and radiolabelled dideoxynucleotide triphosphates (ddNTPs) specific for each of the four DNA bases (Sanger et al., 1977). The second generation of sequencing is also referred to as next-generation sequencing (NGS) and by parallelizing many reactions, these methods made the analysis of DNA much more efficient (Heather and Chain, 2016). Illumina sequencing is an example of a widely used NGS technique that gives many accurate and short reads (Illumina, 2022). The third generation of sequencing or long-read sequencing made it possible to directly sequence single DNA molecules and therefore skip the amplification step (Heather and Chain, 2016). A well-known technique in this group is called Oxford Nanopore sequencing and it generates long reads that can be up to 2 Mb long. This technique is based on nanopores attached to a membrane, which have ionic currents running through them (Magi et al., 2018). The goal of Nanopore is “To enable the analysis of anything, by anyone, anywhere” (Oxford Nanopore, 2021).

1.3.3 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was developed in the early 1980s by Kary Mullins and has since then become a common technique used in laboratories all over the world for different purposes (Bartlett and Stirling, 2003). Even though the technique itself was not established to solve a problem, it has become a solution to many problems as they appear (Bartlett and Stirling, 2003). PCR can be used to make millions of copies of a specific part of the DNA by using three steps: denaturation, annealing, and elongation.

Today, different PCR techniques have been developed to fit the many different requirements in molecular biology. One of these techniques is called Qualitative PCR and makes it possible to confirm that it is the sequence of interest that has been amplified. The specificity of the

method can be confirmed by using an end-point analysis like a gel-electrophoresis (Broeders et al., 2014). Another technique is the Real-time quantitative polymerase chain reaction (Rt-qPCR or qPCR) which uses fluorescent probes to measure the amplification of PCR product during each thermal cycle. For each cycle, the amount of PCR product will increase, and the fluorescent signal will increase correspondingly (Kubista et al., 2006). The C_q-value shows how many cycles are required to reach a chosen threshold and can be used to quantify the target copy number with the help of a standard curve (Bustin et al., 2005). It is therefore possible to gain information about the yield of DNA and to investigate different fragments by using qPCR.

To be able to amplify the correct gene sequence, primers must be designed and optimized. The length of the primer is important for the specificity and a good primer should be around 18-27 bases long with a GC-content of 50 % (Dieffenbach et al., 1993, Geneious, 2022). The primers should have annealing temperatures (T_a) in the range of 50 to 65 °C and it is important that the difference in the T_a between the forward and reverse primer does not differ with more than 4 °C. It is also important to avoid poly-N regions with repetitions and secondary structures, such as hairpin loops (Geneious, 2022). To be able to amplify the gene from different bacterial species, it is important to place the primers in conserved areas of the sequence of interest (Dieffenbach et al., 1993). These areas can be found by aligning the sequences and searching for conserved motifs.

1.3.4 Cell-free protein synthesis

The most abundant biomolecules in living systems are peptides and proteins, and they are critical for structural and functional roles in the cell (Kulkarni et al., 2018). In modern biotechnology, there is a high demand for pure and functional proteins, but these requirements are not often met by natural protein sources and other recombinant protein factories are therefore needed (Sørensen and Mortensen, 2005b). Both in vivo and in vitro techniques have been developed for this purpose. The in vivo techniques require a host organism to express the protein and both *Escherichia coli* and *Lactococcus lactis* are commonly used (Sørensen and Mortensen, 2005b, Zhou et al., 2006). In addition to a host, there is also a need for a plasmid with the gene encoding the protein of interest (Sørensen and Mortensen, 2005a).

Cell-free protein synthesis is an example of an in vitro technique and has been used to increase the understanding of the translation of mRNA into proteins and to search for new antibiotic drugs (Katzen et al., 2005). *E.coli* is the most used source for the preparation of cell-free protein expression, and it can give a yield of up to several milligrams per millilitre

reaction (Kigawa et al., 1999). The bacteria with the gene encoding the target protein are first cultivated and will then be lysed by for instance mechanical disruptions. DNA can then be isolated and purified before the use of a protein synthesis kit. Different protein synthesis kits have been developed and one example is the PURExpress system, which can be used on purified components from *E.coli* (New England Biolabs, 2022). This system will preserve the integrity of the template DNA and RNA and produce proteins that are free of degradation and modification. A one-step reaction is used for both the transcription and translation and in a few hours the results will be available (New England Biolabs, 2022). Since bacteriocins are peptides produced ribosomally, they are well suited for cell-free protein synthesis (CFPS) (Gabant and Borrero, 2019). With the use of CFPS, the bacteriocins are produced outside the cell and it is, therefore, no need for signal peptides to transport them out of the cell. The bacteriocin library PARAGEN 1.0 was built by Gabant and Borrero based on amino acid sequence of the bacteriocins (Gabant and Borrero, 2019). This library can be used to search for the bacteriocin sequence and then further for CFPS.

1.4 PreventADALL cohort

The samples analysed in this thesis were retrieved from the study "Preventing Atopic Dermatitis and Allergies in children" (PreventADALL), which is a multinational population-based mother-child birth cohort study. The main purpose of PreventADALL is to develop strategies to prevent allergic diseases at a low cost and to evaluate if early life factors and exposures, such as xenobiotics, microbiota, and intrauterine environment, have an impact on the development of allergic diseases, asthma, diabetes, obesity, and cardiovascular diseases (Lødrup Carlsen et al., 2018). The study calls attention to the rapid increase of allergic diseases and other immune-related diseases in the western world that cannot be explained by changes in the genotype (Lødrup Carlsen et al., 2018, RodrÍguez et al., 2015).

During a period from December 2014 to October 2016, 2697 mothers at 18-weeks pregnancy were recruited from Oslo University Hospital, Østfold Hospital, and Karolinska Institute in Stockholm. A non-selected, general population was ensured by giving all pregnant women attending routinely ultrasound screening at these three hospitals the opportunity to join the study, as long as they had sufficient language skills (Lødrup Carlsen et al., 2018). Detailed information about the mother's life was collected in electronic questionnaires and biological samples such as blood, urine, skin swabs, and stool samples were collected. The infants enrolled in the study the day they were born and were controlled routinely at 3-, 6-, 12-, 24-, and 36 months (Lødrup Carlsen et al., 2018).

1.5 Aim of thesis

A previous study identified seven bacteriocins produced by the genus *Streptococcus* that were significantly enriched in the gut microbiota of infants compared to adults (Ormaasen, unpublished results). The streptococcal bacteria are among the early colonisers of the infant gut and are these seven bacteriocins could therefore be an important part of the competition between the bacterial species during the establishment of the gut microbiota. Little is known about the function and inhibition spectrum of these bacteriocins, and the main aim of this thesis was therefore to study the associations between these bacteriocin encoding genes and streptococcal species in the infant gut microbiota. To investigate this main goal, several sub-goals were set, and they are listed below.

- Synthesize bacteriocins in vitro
- Test the activity of bacteriocins against chosen species of *Streptococcus*
- Identify the amount of bacteriocin genes in the infant gut, using material from the PreventADALL study
- Associate the amount of bacteriocin genes with the amount of *Streptococcus*

2. Materials and methods

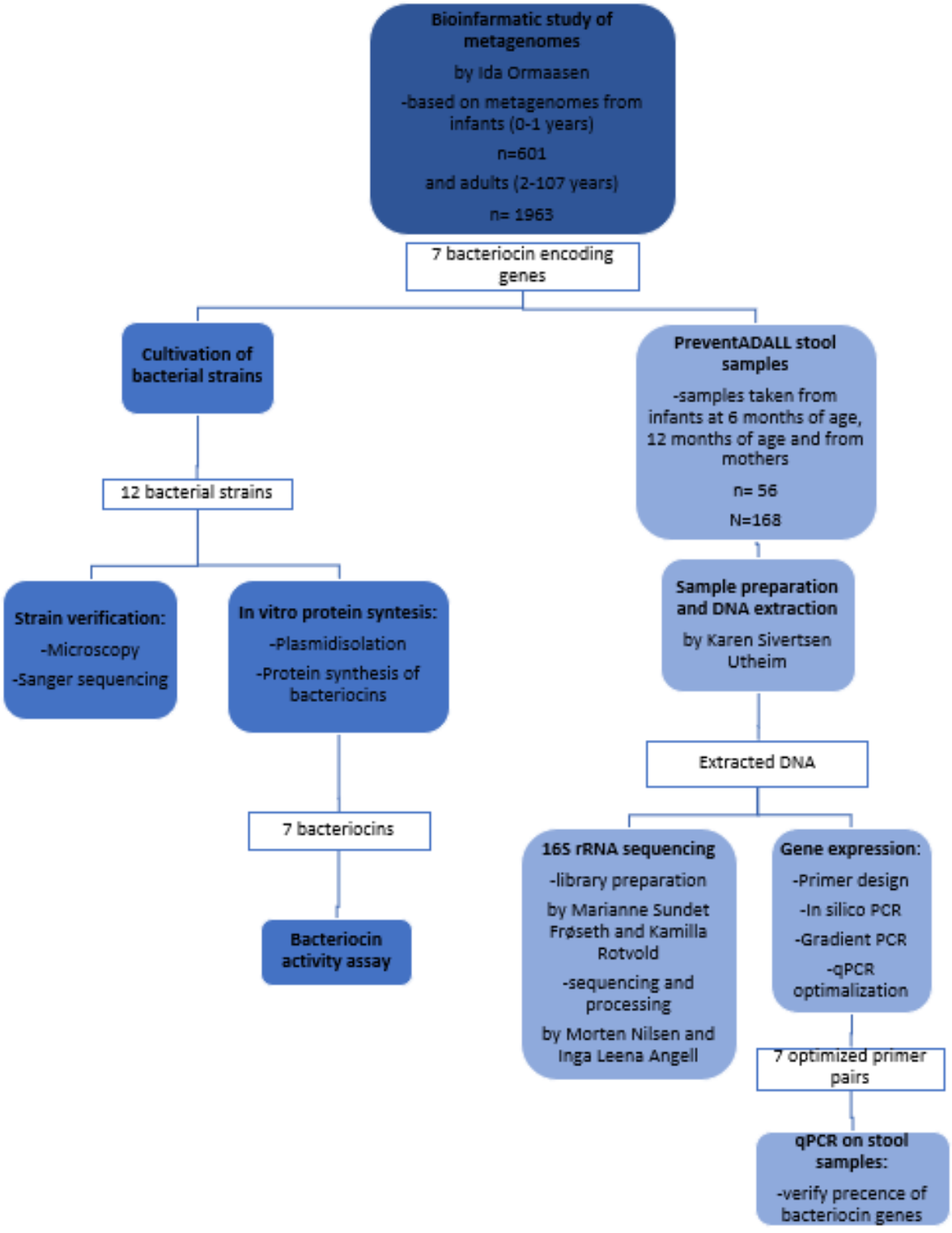

Figure 2.1: Flowchart over the workflow in this thesis. This thesis was based on a bioinformatic study performed by Ida Ormaasen. The metagenomes from this study were used to find bacteriocins that were significantly enriched in samples taken from infants (0-1 years) compared to adults (2-107 years). The first part of the thesis was the cultivation of bacterial strains, in vitro synthesis of bacteriocins, and testing of activity. The second part was the search for bacteriocin genes in stool samples taken from infants at 6 months of age, 12 months of age, and from their mothers. N gives the total number of samples used, while n gives the number of samples in each of the groups.

2.1 Background information and overview of bacterial strains and samples

2.1.1 Background information

This thesis was based on the results of a bioinformatic study by Ida Ormaasen. The metagenomes used in the bioinformatic study were found on NCBI/SRA and chosen if they were taken from healthy gut microbiotas and marked with the age of the donor. 601 samples represented the group of infants (0-1 years), and 1963 samples represented the adults (2-107 years). Then BAGEL4 and BACTIBASE were used to obtain the protein sequence to bacteriocins before the alignment of these protein sequences with the nucleotide sequences of the metagenomes with diamond Blastx. A Wilcoxon test was then used to find the bacteriocins that were significantly enriched in infants compared to adults. The streptococcal bacteriocins used in this thesis were chosen among the resulting bacteriocins from the Wilcoxon test (appendix A).

2.1.2 Samples

The samples used in this thesis were provided by the PreventADALL study and included stool samples taken from infants at 6 months and 12 months of age, and stool samples taken from their mothers. The samples were delivered to the Norwegian University of Life Science, and they were stored at -80 °C until used for DNA analysis.

2.1.3 Bacterial strains

The streptococcal strains cultivated and analysed in this thesis were provided by Leibniz-institute DSMZ- Deutsche Sammlung von Mikroorganismen und Zelkulturen GmbH. The ampoules were provided to NMBU and stored cold and dark in the fridge until use. The control strain of *L.lactis*, the bacteria to the second bacteriocin activity assay, and the control bacteriocin UbeK was provided by Thomas Oftedal, Laboratory for Microbial Gene Technology at NMBU.

2.2 Cultivation of bacteria

The growth media used in this thesis was Tryptic Soy Yeast Extract (TSYE) and Brain Heart Infusion (BHI). These media were used as broth, as soft agar with 0.8 % agar powder, and as solid growth medium with 1.5 % agar powder. DSMZ's and VWR's recipes were followed for the preparation of the growth media followed by autoclaving and storage dark, at room temperature until use (DSZM GmbH, 2012). See Appendix B for reagents used in TSYE medium and BHI medium.

The streptococcal strains were received in ampules and were transferred over to TSYE broth for incubation at 37 °C overnight. Then glycerol stocks were made for each strain for long-term storage. Bacteria were spread out on plates with solid TSYE using a plate streak technique before incubation at 37 °C overnight. Samples from both the overnight cultures and the overnight plates were prepared for microscopy and Sanger sequencing to confirm that the right strains were cultivated. Overnight cultures were also used for the bacteriocin activity test. The bacteria for the second bacteriocin activity assay were cultivated in BHI medium overnight by Thomas Oftedal.

2.3 Sample preparation and DNA extraction

2.3.1 Sample preparation and DNA extraction from cultivated bacteria

Overnight broth cultures of bacterial strains were spun down to make pellets. The pellets were washed with PBS buffer and frozen at -20°C until use. Two different protocols were followed for the DNA extraction. The first protocol was used on the samples that were going to be sequenced with Sanger sequencing. A piece of each pellet and samples taken from overnight plates were scraped on the inside of a PCR strip and put in the microwave on full effect for two minutes to extract DNA. This protocol was suggested by Daniel Straume, Laboratory of molecular microbiology at NMBU. The qubit fluorometer (Invitrogen, USA) was used to ensure that the extraction was successful.

The other DNA extraction protocol was used on the streptococcal samples for the qPCR reactions to look for bacteriocin genes. The pellets were dissolved in 200 µL STAR buffer and transferred to FastPrep tubes together with 0.2 g acid-washed glass beads (Sigma-Aldrich, <106 µm), 0.2 g acid-washed glass beads (Sigma-Aldrich, 425-600 µm), and 0.2 g acid-washed glass beads (Sigma-Aldrich, 2 mm glass beads). The cells were lysed by using FastPrep®-24 (MP biomedicals, France) in two rounds of 4.5 m/s in 40 s. The Mag midi DNA extraction kit (LGC Biosearch Technologies, UK) was used on the mechanically lysed cells by following the manufacturer's protocol. The samples were mixed with lysis buffer, protease, and ethanol before the mag-particle suspension BLm was added. The samples were washed with Wash buffer Blm 1 and 2 before DNA was eluted with Elution buffer BLm.

2.3.2 Sample preparation and DNA extraction from PreventADALL samples

Extraction of DNA from stool samples from infants at 6 months, 12 months, and from mothers was performed by Karen Sivertsen Utheim by using the MagPure Stool DNA LQ kit (Magen, China) and following the instructions from the manufacturer. The stool samples were

transferred to bead tubes and mixed with Buffer ATL/PV and Buffer PCI. Mechanical lysis was performed by using Fastprep 96 (MP biomedicals, USA) and then incubation at 65°C. The KingFisherFlex robot (Thermo Fisher Scientific, USA) was used for the extraction of DNA from the stool samples. Supernatant from the mechanically lysed cells was mixed with RNase A to degrade RNA present in the samples. The KingFisherFlex plates were prepared with Proteinase K, Buffer MLE, MagPure Particles N, Buffer GW1, ethanol, and elution buffer. The program "MagPureStoolDNALQv4" was used for the extraction.

2.4 Polymerase chain reaction

2.4.1 Primer design

Primers for the bacteriocin encoding genes were designed using Geneious Prime 2022.1. The primers were designed based on the amino acid sequence of the seven bacteriocins identified in previous studies by Ida Ormaasen. These sequences were translated into nucleotide sequences by using tblastn. The sequences were aligned, downloaded, and used to find the conserved areas of the bacteriocin sequences with the help of meme-script.org. Default choices were used, except for the number of conserved motifs, which was set to ten. Primers were designed to bind these conserved motifs.

The first sequence of the alignment for each gene was used to design the primer pairs and the preferable characteristics for the primers were a GC-content around 50 %, a length of 18-27 bp, and an annealing temperature between 55-65 °C with no more than two degrees difference between forward and reverse primer. The length of the amplicon should be 80-150 bp (Bustin and Huggett, 2017). These characteristics were set in the primer design functions of Geneious and the primers that matched these characteristics best were tested against sequences from other species. If the primers passed this test, they were ordered from Invitrogen by Thermo Fisher Scientific and used in the experiments.

Primers were also tested in silico against reference genomes of different strains of *Streptococcus* on the website http://insilico.ehu.eus/ (San Millán et al., 2013). All the primer pairs except the Bov255 primers had reference genomes available. The primers were also tested against contigs provided by Ida Ormaasen in the program SnapGene. These contigs came from the metagenomes that were used to identify the different bacteriocins in the infant stool samples.

2.4.2 Cover all PCR

The reaction mix consisted of 1x HOT FIRE pol polymerase Ready to Load (Solis BioDyne, Estonia), 0.5 µL of Mangala F-1, 0.5 µL 16S 1015U R, nuclease-free water, and 2 µL DNA template. The final volume of each reaction was 25 µL. 2720 Thermal cycler (Applied Biosystems, USA) was used to perform the amplification after the following program: 15 minutes at 95 °C for initial denaturation followed by 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 55 °C for 30 seconds, and elongation at 72 °C for 1 minute and 20 seconds before the samples were cooled to 10 °C. The cover all PCR products were used for Sanger sequencing.

2.4.3 Amplicon PCR

The reaction mix consisted of 1x HOT FIRE pol Blend Master Mix Ready to Load (Solis BioDyne, Estonia), 0.5 µL of PRK341F, 0.5 µL PRK806R, nuclease-free water, and 2 µL DNA template. The final volume of each reaction was 25 µL. 2720 Thermal cycler (Applied Biosystems, USA) was used to perform the amplification after the following program: 15 minutes at 95 °C for initial denaturation followed by 25 cycles of denaturation at 95 °C for 30 seconds, annealing at 55 °C for 30 seconds, and elongation at 72 °C for 45 seconds before 72 °C for 7 minutes and then cooling to 10 °C.

2.4.4 Index PCR

The cleaned products from the amplicon sequencing were used in an Index PCR performed by Kamilla Rotvold and Marianne Sundet Frøseth. 16 forward primers and 36 reverse primers were used so that each sample got a unique combination of primers and could be recognized after Illumina sequencing. A reaction mix was made with 1x HOT FIRE pol polymerase RTL (Solis BioDyne, Estonia), 0.2 µL forward primer, 0.2 µL reverse primer, 2 µL purified PCR product, and nuclease-free water so that the final volume was 25 µL. The samples were amplified with 2720 Thermal cycler (Applied Biosystems, USA) for 5 minutes at 95 °C for initial denaturation followed by 10 cycles of denaturation at 95 °C for 30 seconds, annealing at 55 °C for 1 minute, and elongation at 72 °C for 45 seconds before the samples were held at 72 °C for 7 minutes and then cooled to 10 degrees (Rotvold, 2022, Frøseth, 2022).

2.4.5 Gradient PCR

Gradient PCR was used to find the optimal annealing temperature for the designed primer pairs. The amplicons from which the primers were designed from, were used as a positive control. The amplicon was received as 250 ng dry DNA and was added 50 µL water so that the start concentration was 5 ng/µL. A 10-fold dilution series were made until a 10^{-7} dilution

and water was used as a negative control. The reaction mix consisted of 1x HOT FIREPol® EvaGreen qPCR supermix (Solis BioDyne, Estonia), 0.2 µM of each primer, 2 µL template DNA, and nuclease-free water so that the final volume was 20 µL. Mastercycler gradient (Eppendorf, Germany) was used for the amplification with the following program: 95°C for 15 minutes, followed by 40 cycles of 95 °C for 30 seconds, T_a estimated by Geneious prime for each primer ± 7 °C for 30 seconds, and 72°C for 45 seconds. The products were analysed with gel electrophoresis.

The gradient PCR machine needed service when the other primers were tested, and PCR runs with different temperatures were therefore used to find the optimal annealing temperature for the remaining primers. The same reagents were used with the following program 95°C for 15 minutes, followed by 40 cycles of 95 °C for 30 seconds, T_a for 30 seconds, and 72°C for 45 seconds. T_a was set to 56°C, 60 °C, and 64 °C in the three runs. The products were analysed with gel electrophoresis to estimate the optimal annealing temperature.

2.4.6 Quantitative PCR

The reaction mix contained 1x HOT FIREPol® EvaGreen qPCR supermix (Solis BioDyne, Estonia), 0.2 µM of each primer, 2 µL template DNA, and nuclease-free water so that the final volume was 20 µL. CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA) was used for the amplification with the program: 95°C for 15 minutes, followed by 40 cycles of 95 °C for 30 seconds, 60/67°C for 30 seconds, and 72°C for 45 seconds. A High-Resolution Melting (HRM) analysis was also included. This qPCR was used both to test if there were enough DNA present in the stool samples and to test if there were bacteriocin genes present in the stool samples.

2.4.7 High resolution melting analysis

The qPCR products were analysed with a High-Resolution Melting (HRM) analysis by the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA). This analysis made it possible to verify that the amplicon of interest had been amplified based on its melting point. The melting point of a DNA fragment depends on the length of the fragment and the GC content. CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA) was used to perform the HRM by raising the temperature from 60°C to 95°C and detecting the melting temperatures of the amplicons.

Table 2.1: Primers designed and used in this thesis. Each bacteriocin gene has two primers, one forward and one reverse.

Target gene	Primer	Primer sequence F/R (5' – 3')	Reference
blpU	BlpU_forward	AAAGCAGGTGTTGGAGGAGG	This work
	BlpU_reverse	GTATAGCTCCTCCCACAGCAC	
blpD	BlpD_forward	ATGGCAACTCAAACAATTGAAAACT	This work
	BlpD_reverse	CTCATCACAGGACAACCCACC	
blpK	BlpK_forward	GGAGGAGCGTTTGGAGGTAA	This work
	BlpK_reverse	TGCACCACCTAGAACACCAC	
blpU_2	BlpU_2_forward	TGGATGTAGCTGGGGAGGTT	This work
	BlpU_2_reverse	GACAGCACCTTGCCAAGTAC	
blpK_2	BlpK_2_forward	ACTGAAATGCTTGCTTGCGT	This work
	BlpK_2_reverse	TGCCTAGTTGAAGACCTCTTGC	
blpJ	BlpJ_forward	ATGCTTGCGAAAGTTGAAGGG	This work
	BlpJ_reverse	CTGCTCCAGTTCCACCAGTT	
bov255	Bov255_forward	ACTTTCTACTGTGGAAGGCGG	This work
	Bov255_reverse	ATAACCCCACTCACCATTGCT	
16S rRNA	Mangala F-1	TCCTACGGGAGGCAGCAG	Genetic Analysis
	16S 1015U R	CGGTTACCTTGTTACGACTT	
Region V3-V4 of 16S rRNA gene	PRK341 F	TCCTACGGGAGGCAGCAGT	(Nadkarni et al., 2002)
	PRK806 R	GGACTACCAGGGTATCTAATCCTGT	

2.4.8 Purification of PCR products

The PCR products were purified manually by following the AMPure Clean-up protocol. AMPure XP beads were used to bind the PCR product and isolate it while removing the smaller fragments from the library. The ratio between the PCR products and the AMPure XP beads was 1:1. Freshly made 80 % ethanol was used to wash the samples and nuclease-free water was then used to eluate the cleaned PCR products.

2.5 Qualification of streptococcal strains and quantification of DNA

2.5.1 Microscopy

Light microscopy was used as a quick method to observe the cell morphology of the cultivated bacterial strains and thereby assure that bacteria from the right genus were used in the further analysis. The morphology of bacteria in the genus *Streptococcus* does typically have a coccoid shape and usually occurs in chains or pairs (García-Curiel et al., 2021). If these characteristics were observed, the cultures were analysed further.

2.5.2 Qubit

Qubit® dsDNA HS Assay Kit (Invitrogen, USA) was used to quantify DNA in the different samples. The dye used in this assay is highly selective for double-stranded DNA and emits fluorescence when bound to dsDNA. This fluorescence is then detected by the Qubit Fluorometer (Invitrogen, USA). The manufacturer's recommendations were used on 2 µl sample and 198 µl working solution.

2.5.3 Gel electrophoresis

Gel electrophoresis was used to study the length of DNA from different samples. The gel contained 1 % agarose and 1 x PeqGreen. The gel will consist of a network formed by the added agarose. Because of the applied electrical voltage, the negatively charged nucleic acid fragments will migrate through this network and they will be separated based on their size. By using a 100 bp ladder as a reference, it was possible to determine the size of the different fragments and compare it to what was expected. PCR fragments where 1x HOT FIREPol® EvaGreen qPCR supermix (Solis BioDyne, Estonia) had been used were stained with loading dye. The gel was run for 30 min with 80 V and 400 mA. To visualize the bonds UV lights were used in the MolecularImager® Gel Doc™ XR Imaging system with Quantity One 1-D analysis software (Bio-Rad, USA).

2.5.4 Sanger sequencing

To prepare the samples for Sanger sequencing, 5 µL of cover all PCR products were mixed with 5 µL 5 µM forward primer (Mangala F-1). The samples were stored in the fridge before transport to Eurofins Genomics in Germany for sequencing.

2.5.5 Illumina sequencing

The index PCR samples were pooled together in a library, based on results from the Qubit quantification. This was done with the Biomek 3000 robot (Beckman Coulter, USA) and this normalization ensured that each sample contributed the same amount of nucleotide sequences to the library. The Ampure protocol was used to purify the pooled library before visualization on 1 % agarose gel. This was executed by Kamilla Rotvold and Marianne Sundt Frøseth.

PhD. Morten Nilsen and head engineer Inga Leena Angell performed the Illumina sequencing and sequence processing. KAPA Quantification Kit (Roche Sequencing Solutions, USA) was used on the cleaned index PCR samples to quantify the library by following the manufacturer's recommendations. A 10-fold dilution series of the pooled library was made together with a reaction mix with 2x KAPA SYBR FAST qPCR Master mix, 10x Primer

premix, 2 µL DNA, and PCR grade water so that the final volume was 20 µL. CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, USA) was used for the amplification with the program: 95°C for 5 minutes, followed by 95 °C for 30 seconds, and 60°C for 45 seconds. A High-Resolution Melting (HRM) analysis was also included. The results were quantified with the KAPA Library Quantification Data Analysis Template. The samples were spiked with 15 % PhiX control so that the final concentration was 8 pM. Illumina Miseq (Illumina, USE) was used to perform the sequencing.

2.6 Synthesis of bacteriocins in vitro and screening for bacteriocin activity

2.6.1 Plasmid isolation

The plasmid encoding the control bacteriocin was isolated from an overnight broth culture using the E.Z.N.A.® Plasmid DNA Mini Kit I following the instructions provided by the manufacturer (Omega Bio-Tek, USA). The samples were mixed with solution I which contained RNase, solution II for lysis, and solution III for elution before transfer to E.Z.N.A miniprep columns with collection tubes. Buffer HBC was used for better binding to the column before adding the DNA wash buffer. The washing step was repeated before elution with the elution buffer. The concentrations were then measured using NanoDrop so that the right amount of DNA was used in the next step.

2.6.2 Protein synthesis

The dry plasmids encoding the bacteriocins BlpU, BlpD, BlpK, BlpU_2, BlpK_2, BlpJ, and Bov255 were added 40 µL nuclease-free water so that the final concentrations were 125 ng/µL. The PURExpress® *In vitro* Protein synthesis kit was used together with the PureExress Disulphide Bond Enhancer for the protein synthesis following the instructions provided by the manufacturer (New England Biolabs). The reaction mix contained 10 µL Solution A, 7.5 µL Solution B, 1 µL PURExpress Disulphide Bond Enhancer 1, 1 µL PURExpress Disulphide Bond Enhancer 2, 3.5 µL nuclease-free water, and 2 µL template DNA. The reaction mix for the control bacteriocin Ube K contained 10 µL Solution A, 7.5 µL Solution B, 5.5 µL nuclease-free water, and 2 µL template DNA. All the reactions were incubated at 37 °C for two hours and the reactions were stopped by putting them on ice. The antimicrobial activity was tested shortly after, since the bacteriocins are hydrophobic and therefore quickly precipitates and therefore are considered unstable.

2.6.3 Antimicrobial activity assay

Agar plates with 40 mL solid TYSE agar were prepared beforehand. 200 µL overnight cultures of the bacterial strains were mixed with 10 mL TYSE soft agar and vortexed for 1-2 seconds before spreading on the solid agar plates. The plates were left to dry and then 3 µL of each of the reaction mixes from the protein synthesis were added to marked spots. The plates were incubated overnight at 37 °C in microaerophile conditions, except for *L.lactis* which were incubated in aerobic conditions. Inhibition by the bacteriocin was indicated as clear zones in the soft agar. For the second bacteriocin activity assay, BHI medium was used and the plates were incubated at 30 °C in aerobic conditions.

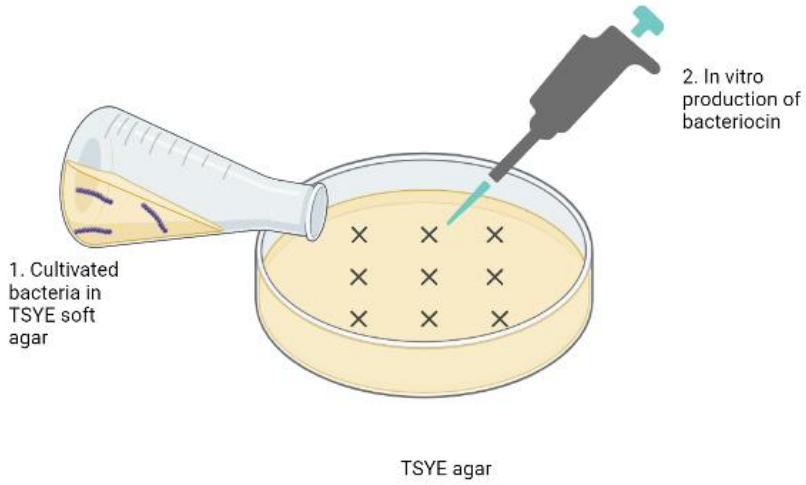

Figure 2.2: Bacteriocin activity assay was used to test if any of the in vitro produced bacteriocins inhibited the growth of the cultivated *Streptococcus*. The figure is made in BioRender.

2.7 Data analysis

2.7.1 16S rRNA data

The 16S rRNA sequencing results were analysed by PhD. Morten Nilsen. Because of the low quality, all the forward reads were removed, and reverse reads were shortened to 220 bp to further optimize the quality. Also, the primer binding sites were removed together with sequences with an expected error mean below 1.0. A threshold of 97 % was set for the clustering of sequences into OTUs. Amplicons containing chimeric sequences were removed before the taxonomic annotation was performed with the database RDP taxonomy 18.

The processed data was used to investigate differences between the three groups: infants at 6 months of age, infants at 12 months of age, and mothers. Data were collected from 168 samples, where 56 samples belonged to each group. The total number of OTUs, OTUs

classified to the genus *Streptococcus* and the number of reads belonging to different OTUs was studied and compared between the three groups of samples.

2.7.2 Kruskal Wallis test

The Kruskal Wallis test is a non-parametric one-way ANOVA test that can be used to test if one variable is significantly different between two or more groups. The test was performed to investigate the associations between *Streptococcus* and age in the three groups of samples. To perform the test, the 16S rRNA sequencing data was uploaded in RStudio and grouped into infants at 6 months, 12 months, and mothers. The test was performed between the groups and the significance level was set to 0.05. The test was also used to see if there were any associations between bacteriocins and *Streptococcus* gut microbiota.

2.7.3 Student t-test

The student t-test can be used to see if the means of two groups are significantly different from one another. The test was performed on groups where the Kruskal Wallis test had shown a significant difference, to test which group contained more streptococcal reads. RStudio was used for the testing with a significance level of 0.05.

3. Results

3.1 Characteristics of cultures and verification of bacterial strains

The initial cultivation of bacteria was checked with light microscopy to verify that it was indeed the streptococcal bacteria growing on the plates and in the broth cultures. The genus *Streptococcus* is typically observed as coccoid cells that occur either in pairs or chains (García-Curiel et al., 2021). These caracteristics were observed for all the six cultivated strains and are shown for *S.salivarius*, *Streptococcus sanguinis*, and *Streptococcus equinius* in figure 3.1.

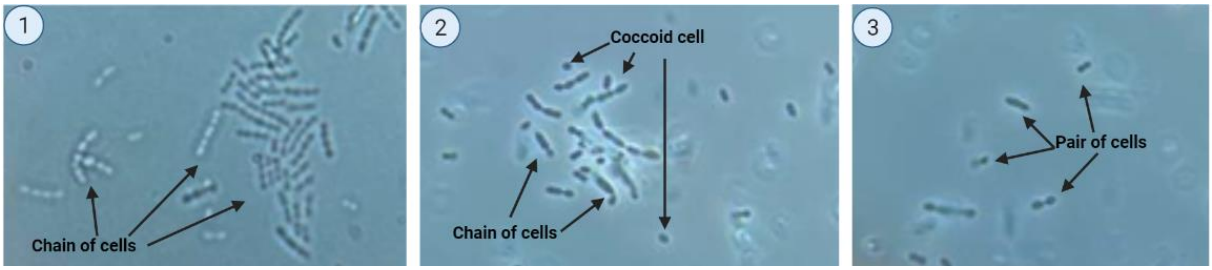

Figure 3.1: Light microscopy of *S.salivarius* (1), *S.sanguinis* (2), and *S.equinius* (3) grown microaerophile at 37 °C. The pictures show cells from overnight cultures of the streptococcal strains cultivated on TSYE plates. The cells show the typical coccoid shape and appear in pairs or longer chains.

Since it was only the genus *Streptococcus* that was confirmed by using the light microscope, Sanger sequencing was performed to ensure that the right strains were cultivated and used for further experiments. All the six different streptococcal strains were identified among the plates, and all showed both coverage and identity above 99 %. They could therefore be used in the bacteriocin activity assay and to identify the presence of bacteriocin genes. Some contamination issues occurred in the early process, but the contamination source was found and removed as described in appendix C.

3.2 Activity of bacteriocins

A spot-on-lawn inhibition assay was performed to test if the bacteriocins inhibited the growth of the streptococcal strains. All the seven bacteriocins were tested against overnight cultures of all the six streptococcal strains. *L.lactis* with the bacteriocin UbeK was used as a positive control since a pilot study had shown that *L.lactis* was inhibited by UbeK (results not included).

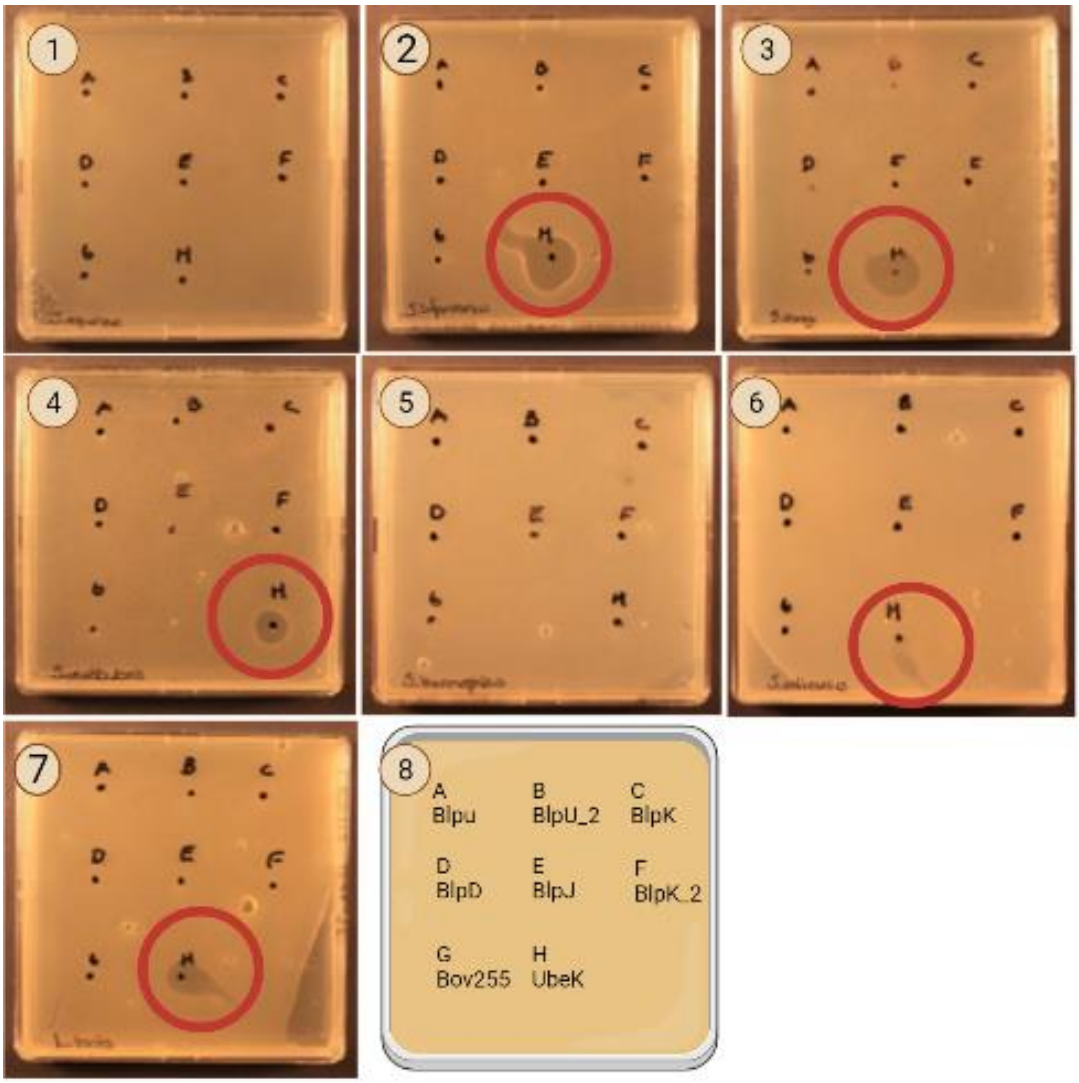

Figure 3.2: Antimicrobial effect of bacteriocins on *S.equinus* (1), *Streptococcus infantarius* (2), *S.sanguinis* (3), *Streptococcus vestibularis* (4), *S.thermophilus* (5), *S.salivarius* (6) and *L.lactis* (7). The figure shows the results of the first bacteriocin activity assay. The marking A-H corresponds to the marking on plate 8 and shows where the different bacteriocins were added to the plates with TSYE medium. The antimicrobial effect of the bacteriocins is observed as small circles where the bacterial growth was inhibited, marked with red circles. The plates were incubated microaerophile at 37 °C overnight.

None of the streptococcal bacteriocins gave inhibition of the tested bacteria, as seen in figure 3.2. Inhibition zones were only observed where the control bacteriocin Ube K had been applied and this bacteriocin was able to inhibit the growth of *S.infantarius, S.sanguinis, S.vestibularis, S.salivarius,* and *L.lactis.* Streptococcal bacteria have previously been shown to inhibit the growth of bacteria belonging to the genera *Enterococcus, Listeria,* and *Lactococcus*. Another bacteriocin activity assay was therefore performed with strains belonging to these genera cultivated with BHI medium.

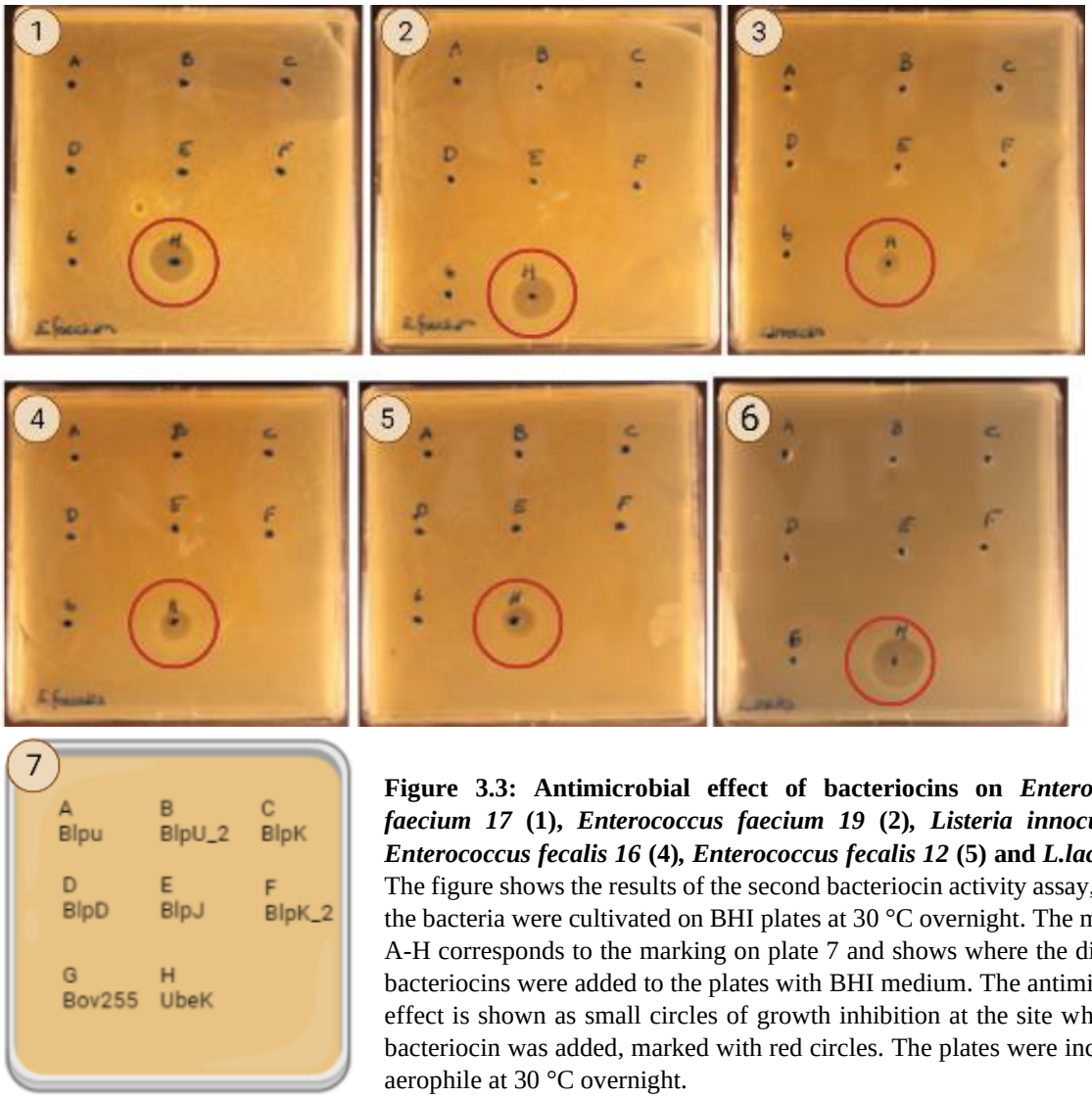

Figure 3.3: Antimicrobial effect of bacteriocins on *Enterococcus faecium 17* (1), *Enterococcus faecium 19* (2), *Listeria innocua* (3), *Enterococcus fecalis 16* (4), *Enterococcus fecalis 12* (5) and *L.lactis* (6). The figure shows the results of the second bacteriocin activity assay, where the bacteria were cultivated on BHI plates at 30 °C overnight. The marking A-H corresponds to the marking on plate 7 and shows where the different bacteriocins were added to the plates with BHI medium. The antimicrobial effect is shown as small circles of growth inhibition at the site where the bacteriocin was added, marked with red circles. The plates were incubated aerophile at 30 °C overnight.

Also, the second bacteriocin activity assay showed inhibition of growth only by UbeK, as shown in figure 3.3.

3.3 Investigation of the presence of bacteriocin genes in stool samples

3.3.1 Primer design and optimization

A primer pair had been designed for each of the seven bacteriocin genes used in this thesis. Three rounds of in silico testing were performed before the primers were used to investigate the presence of bacteriocin genes in stool samples. The initial test was done in Geneious where the primers were tested on aligned sequences from other bacterial species. The second test was performed against reference strains, based on the BAGEL4/BACTIBASE classification of the bacteriocins. The last test was in SnapGene against the contigs used to identify the bacteriocins.

Table 3.1: In silico testing of designed primers. The results from the three rounds of in silico tests are summarized in the table. The first test was performed in Geneious against aligned sequences, the second online against reference genomes based on BAGEL4/BACTIBASE, and the third in SnapGene against contigs (marked Bac). x marks where both forward and reverse primer bound to template and gave amplicons with expected length. F marks where only forward primer bound to template.

Primer pair	Aligned sequence	Reference genome	Bac682	Bac808	Bac809	Bac935
BlpU	x	x	-	-	-	-
BlpD	x	x	F	F	F	F
BlpK	x	x	x	x	-	x
BlpU_2	x	x	-	-	-	-
BlpK_2	x	x	-	-	-	-
BlpJ	x	x	-	-	-	-
Bov255	x	-*	-	-	-	-

*There was no reference genome available for bov255 on the website for in silico PCR

Table 3.1 shows that all the primers were able to bind in the initial rounds of testing against aligned sequences and reference genomes, except the Bov255 primer pair where no reference genome was available. Only the BlpK primer pair was able to bind and amplify segments with the correct length in the test against contigs, but also the BlpD forward primer was able to bind.

Gradient PCR and gel electrophoresis were performed to determine the optimal annealing temperature for the designed primers, see appendix D for gel results. The optimal annealing temperature for BlpU was set to 67 °C, while the annealing temperature for the rest of the primers was set to 60°C. The primers were optimized further by testing if they gave unspecific binding in complex stool samples.

Table 3.2: Testing of primers against complex stool samples. The column marked expected shows the expected melting temperatures if the primers amplify the sequence of interest, from Geneious. The melting temperature of the products from the control samples and the complex stool samples are summarized in this table, see appendix D3 for melting curves. Complex samples with T_m±1°C of control is marked green and are positive for the bacteriocin gene, while red marks the samples that were negative for the bacteriocin gene. – indicates that the primer pair did not bind to the complex samples.

Primer pair	Expected (°C)	Control product (°C)	Complex sample (°C)
BlpU	83.7	83.5	91-92
BlpD	80.1	80	-
BlpK	81.6	80	-
BlpU_2	81.7	80.5	-
BlpK_2	83.7	83.5	-
BlpJ	81.7	80.5	81
Bov255	80.4	79.5	76

The BlpU, BlpJ, and Bov255 primers gave binding in complex stool samples as shown in table 3.2, but it was possible to separate the unspecific and specific binding based on the melting temperature of the amplicons. The binding of BlpU and Bov255 primers were unspecific since the melting temperatures of the products were not ±1°C of the control, while the binding of BlpJ was specific. The primer pairs were also tested on DNA extracted from the streptococcal strains used in the bacteriocin activity assay to see if the bacteriocin genes were present in these strains.

Table 3.3: Testing of primers against DNA extracted from the streptococcal strains. The column marked expected shows the expected melting temperatures if the primers amplify the sequence of interest, from Geneious. The melting temperature of the products from the control samples and the streptococcal strains are summarized in this table, see appendix D3 for melting curves. Samples with T_m±1°C of control is marked green and are positive for the bacteriocin gene, while red marks the samples that were negative for the bacteriocin gene. – indicates that the primer pair did not bind to the streptococcal DNA.

Primer pair	Expected (°C)	Control (°C)	*S.equinus* (°C)	*S.infantarius* (°C)	*S.sanguinis* (°C)	*S.vestibularis* (°C)	*S.thermophilus* (°C)	*S.salivarius* (°C)
BlpU	83.7	83.5	-	-	-	76.5	-	-
BlpD	80.1	80	-	-	-	-	-	-
BlpK	81.6	80	-	-	-	-	80	86
BlpU_2	81.7	80.5	-	-	-	-	-	-
BlpK_2	83.7	83.5	-	-	-	-	82.5	84.5
BlpJ	81.7	80.5	-	-	-	-	-	-
Bov255	80.4	79.5	-	-	-	-	79.5	-

The same criterion for specific binding was used as with the complex samples and specific binding was observed for the BlpK, BlpK_2 and Bov255 primer pair against *S.thermophilus* and for the BlpK_2 primer pair against *S.salivarius*. Unspecific binding was observed to *S.salivarius* with the BlpK primer pair and to *S.vestibularis* with the BlpU primer pair. None of the other primer pairs bound to the streptococcal DNA.

3.3.2 Presence of bacteriocin genes in stool samples

qPCR with 16S rRNA primers confirmed that the amount of DNA in the 100x diluted stool samples was sufficient for further analysis. The Cq-values from these qPCR runs did not differ significantly between infants at 6 months, 12 months, and mothers. The designed primer pairs were used to identify the bacteriocin genes in the samples.

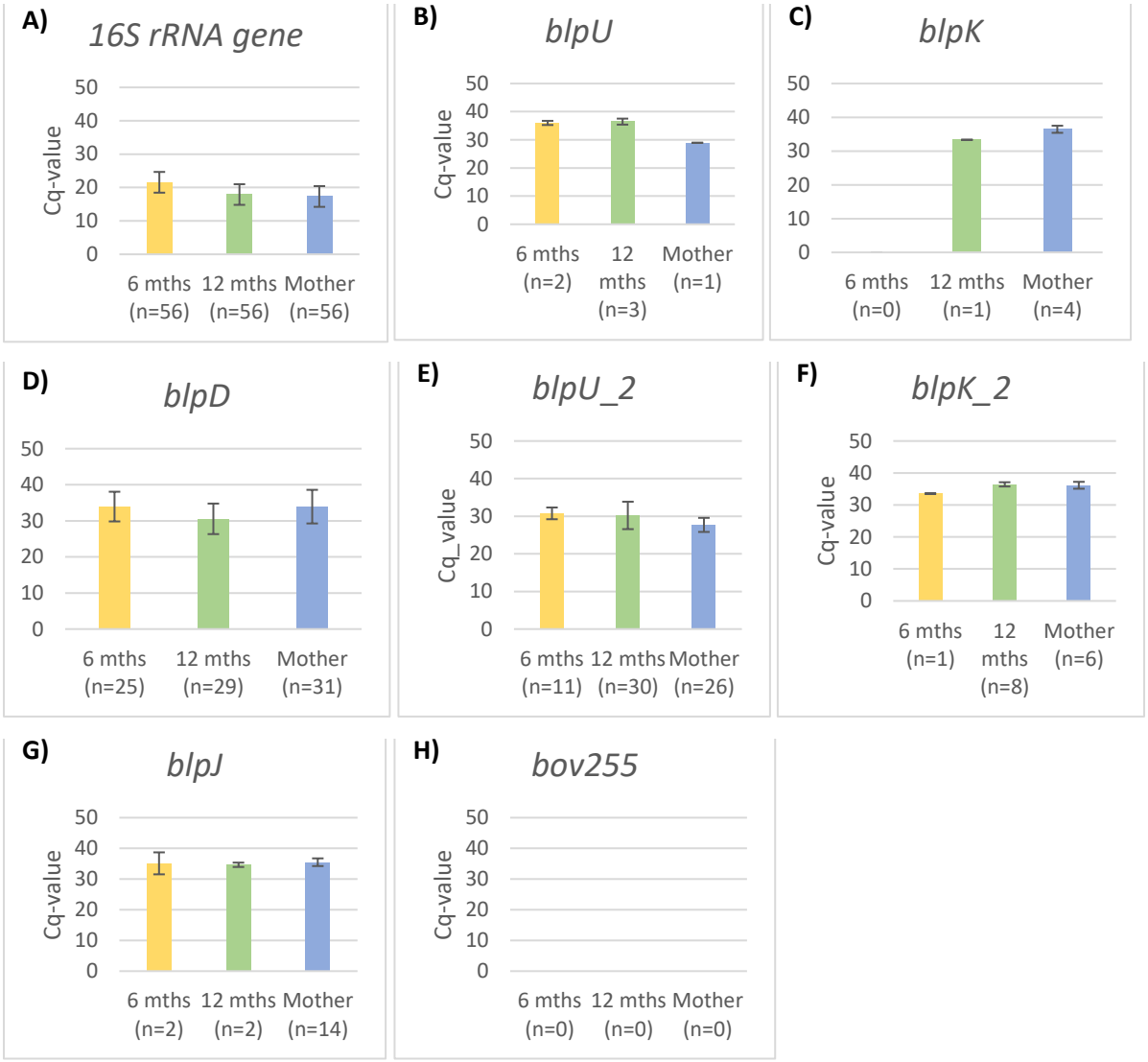

Figure 3.4 A-H: Cq-values from qPCR reactions targeting the *16S rRNA, blpU, blpK, blpD, blpU_2, blpK_2, blpJ* and *bov255 gene.* The Cq-values were estimated as an average of the samples positive for the gene of interest for each of the three groups: infants at 6 months of age, at 12 months of age, and from their mothers. For a stool sample to test positive for a bacteriocin gene it had to meet two criteria. The first criterion was that the amplified PCR product should have T_m of ±1 °C of the positive control and the second criterion was that the amplicon should have a bond length that corresponded to the length of the positive control. No significant difference in Cq-values was found between the three age categories within each gene. No samples tested positive for the *bov255* gene. n gives the number of samples that tested positive for the gene and was used in the calculation of the average Cq-value.

Figure 3.4 shows that the samples amplified with the different primer pairs had quite high Cq-values compared to the samples amplified with the 16S rRNA primer pair. Notice that some of the calculations in the figure are based on only one observation or a few observations. Both specific and unspecific binding of the primer pairs gave high Cq-values, and it was therefore not possible to use Cq-value as a criterion for the presence of bacteriocin. The total number of samples that tested positive for each of the seven bacteriocin genes is shown in table 3.4. A pattern was observed for the *blpD*, *blpK*, and *blpJ* gene, which seemed to increase with age.

Table 3.4: Presence of bacteriocin encoding genes in stool samples from infants at 6 months, 12 months, and mothers. The numbers indicate the number of samples where the criteria for positive results were met. The first criterion was that the melting temperature should be T_m (from the positive control) ± 1°C and the second criterion was that the length of the amplicon should be the same as the positive control. The melting curve and the results from the gel electrophoresis are shown in appendix E.

BACTERIOCIN	INFANTS 6 MONTHS	INFANTS 12 MONTHS	MOTHERS
BLPU	2	3	1
BLPD	25	29	31
BLPK	0	1*	4
BLPU_2	11	30	26
BLPK_2	1	8	6
BLPJ	2	2	14
BOV255	0	0	0

* This sample did not meet the criteria T_m± 1°C but was counted as a positive sample since the BlpK primer pair was observed to bind multiple of the *blp-loci* during the testing in SnapGene

3.3.3 16S rRNA sequencing results

A total of 1309 OTUs were identified and seven OTUs were classified to the genus *Streptococcus*. The average number of OTUs and the percentage of streptococcal reads in each of the three age categories were calculated.

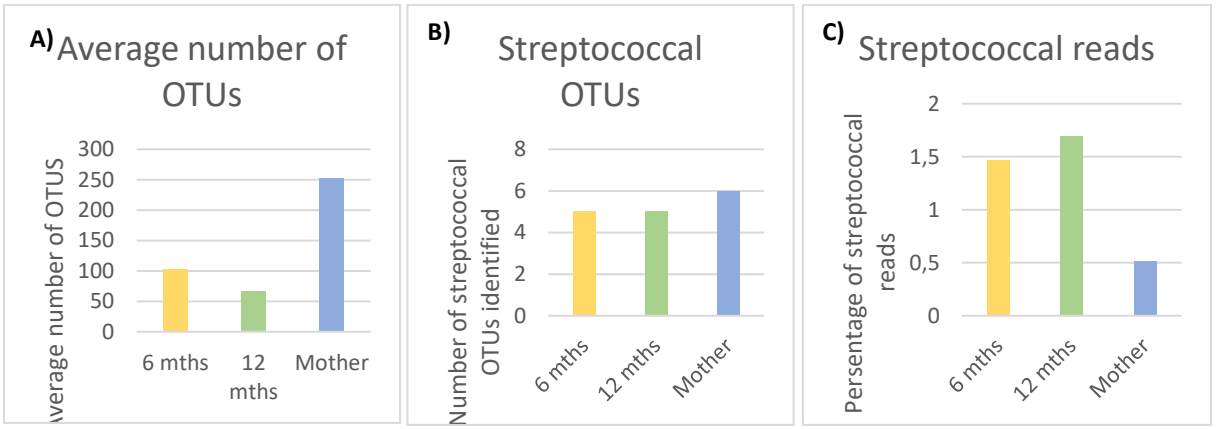

Figure 3.5 A-C: Presence of OTUs, streptococcal OTUs, and streptococcal reads in stool samples from infants at 6 months, 12 months, and from their mothers. A) The average number of OTUs in samples taken from the three age categories. B) The total number of OTUs identified in samples from the three age categories. C) The percentage of streptococcal reads among the total number of reads in each of the three age categories. The number of streptococcal reads was found to be significantly higher in samples from infants (6 months and 12 months) than in samples from mothers. All the 168 samples, 56 for each age category, were used for the calculations in A-C.

Figure 3.5 shows that it was the samples originating from the infants at 12 months of age had the lowest number of OTUs present on average, while mothers clearly had the highest number of OTUs on average. The number of reads classified to the genus *Streptococcus* was highest in infants at 12 months of age with a presence of 1.69 %, while infants at 6 months had 1.46 % streptococcal reads and mothers had 0.52 % streptococcal reads, even though more streptococcal OTUs were identified among the samples taken from mothers. The 16S rRNA

sequencing results was also used to test if there were any associations between age, number of streptococcal reads, and presence of bacteriocin.

3.3.4 Associations between streptococcal reads, presence of bacteriocin, and the three sample groups

Several statistical tests were performed to see if there were any associations between the number of streptococcal reads, presence of bacteriocins, and the three sample groups: infants at 6 months, 12 months, and mothers (see appendix F). A summary of the statistical tests that gave significant results is shown in table 3.5.

Table 3.5: Student t-tests was used to identify associations between the sample groups, streptococcal reads, and the presence of bacteriocins. The table shows the results from two student t-tests with significance level of 0.05 and both tests show significant results. Also, the different median values from each group in both tests are shown. H_0: $\mu_1 \leq \mu_2$, and H_1: $\mu_1 > \mu_2$.

Statistical test	p-value	Median of streptococcal reads
More streptococcal reads in infants (6 months and 12 months) than in mothers	0.045	Infants 6 and 12 months: 32 Mothers: 20
More streptococcal reads from OTU nr.2 in samples (infants 12 months) with bacteriocin than without	0.025	With bacteriocin: 31 Without bacteriocin: 8

As shown in the table above, the number of streptococcal reads was significantly higher in the samples taken from infants (6 and 12 months) than from their mothers. This can also be seen in figure 3.5 where the number of streptococcal reads in percentage of the total number of reads is visualized. Associations between the number of streptococcal reads in OTU nr.2 and the presence of bacteriocins were found in the samples taken from infants at 12 months but could not be found in mothers or infants at 6 months (see appendix F).

4. Discussion

4.1 Bacteriocin activity

The bacteriocin encoding *blp* locus in *S.pneumoniae* and *S.thermophilus* has been known for several years but still, little is known about the activity of the bacteriocins that are encoded by this locus and how this affects their role in the gut microbiota. In this study, six blp-bacteriocins and a bov255 bacteriocin were studied for their ability to inhibit other bacteria. The bacteriocins were synthesised in vitro and evaluated against a set of streptococcal bacteria but only the control bacteriocin Ube K gave inhibition of growth. One possible explanation was that the bacteriocins had small inhibition spectrums and therefore did not inhibit any of the chosen streptococcal strains. This had previously been observed to be the case for the bacteriocin BlpK produced by *S.pneumoniae*, which only inhibited the growth of other pneumococci with deficiencies in their *blp*-locus (Wholey et al., 2019). It would therefore be of interest to test the bacteriocins against other pneumococci, but this was not done in this thesis because of safety regards.

Another possibility that was quickly dismissed was that all the streptococcal strains encoded all the seven bacteriocin genes. To evaluate this hypothesis DNA extracted from the streptococcal strains was included in the qPCR reactions with the designed primer pairs. The qPCR results in table 3.3 show that *S.thermophilus* encoded the *blpK, blpK_2,* and *bov255* gene and *S.salivarius* encoded the *blpk_2* gene. Even though the other strains did not test positive for the bacteriocin genes, they might have acquired the immunity genes. For instance, the immunity mechanism against the Bov255 bacteriocin has previously been found to resemble the immunity mechanisms against Nisin and this is a phenotype that develops quickly among bacteria (Mantovani et al., 2002).

Previous studies have shown that bacteria from the genus *Streptococcus* often inhibit the growth of bacteria belonging to the genera *Enterococcus* and *Listeria (Rossi et al., 2013, Fontaine and Hols, 2008).* Another bacteriocin activity assay was therefore performed to test if this could be the case among these bacteriocins, but no inhibition was observed. As discussed previously regarding the streptococcal strains, these bacteria might be immune to the bacteriocins, or they may be outside the inhibition spectrums of the bacteriocins.

One interesting possibility was that there had been a classification error in the databases BAGEL4 and BACTIBASE, which were used to identify and classify the bacteriocins from the bioinformatic study that was fundamental for this thesis. As previously mentioned, the

classification of bacteriocins varies somewhat in the literature and the classification in BAGEL4/ BACTIBASE was used as the starting point for the planning of this thesis. The bacteriocins were classified to subclass IId, except BlpK from *S.thermophilus* which was classified to subclass IIa and BlpD from *S.thermophilus* which only was classified to class II. Other studies have however placed the blp-bacteriocins in class IIb, which means that they consist of two different peptides and are dependent on both peptides to gain full activity (Fontaine and Hols, 2008, Rossi et al., 2013, Wholey et al., 2019). These studies found that the BlpU bacteriocin were dependent on BlpD to inhibit the growth of bacteria such as *S.thermophilus* and *L.lactis.* Also, the bacteriocin BlpJ was observed to be dependent on another bacteriocin called BlpI. If this is the case, then our bacteriocins will not give any inhibition when added separately. A suggestion for further studies would therefore be to mix the different bacteriocins and see if this gives any inhibition.

4.2 Bacteriocin genes in stool samples

The bioinformatic study that was the foundation for this thesis, observed that the genes encoding the chosen bacteriocins were significantly enriched in metagenomes from infants (0-1 years) compared to adults (2-107 years). The initial hypothesis of this thesis was therefore that the seven bacteriocin genes would be enriched in stool samples from infants, but this did not seem to be the case based on the qPCR results. The presence of all the bacteriocin genes seemed to be reduced in infants at 6 months of age compared to infants at 12 months of age and mothers. It was also surprising that three of the seven genes were present in more samples from mothers than infants. The explanation for this might be that 6 months of age is too late to find enrichment of these streptococcal bacteriocin genes among samples from infants. Previous studies have shown that the gut microbiota is dominated by streptococcal bacteria during the first 10 days of the infant's life and that the amount is more stable in an adult gut microbiota (Gosalbes et al., 2013, Solís et al., 2010). It might be that this stabilization of the amount of *Streptococcus* has occurred already in infants at 6 months of age, and it could therefore be interesting to use samples from the first 10 days of the infant's life.

4.3 Streptococcus in stool samples

The results from the 16S rRNA sequencing identified bacterial genera that typically are found in the human gut, such as *Enterococcus, Bifidobacterium, Lactobacillus,* and *Streptococcus* (Solís et al., 2010). The highest average number of OTUs was present in samples from mothers, which indicated that mothers had the highest alpha diversity and this was consistent with previous knowledge (Derrien et al., 2019). Also, the highest number of OTUs belonging

to the genus *Streptococcus* was identified in samples from mothers. The number of identified streptococcal OTUs was the same in infants at 6 months and 12 months.

The number of streptococcal reads was significantly enriched in samples taken from infants (6 months and 12 months) compared to mothers. This was expected since the genus *Streptococcus* is among the early colonizers of the infant gut microbiota and dominates the microbiota, especially in the first ten days of the infant's life (Gosalbes et al., 2013, Solís et al., 2010). It was therefore a bit surprising that the streptococcal reads made out a bigger share of the total amount of reads in infants at 12 months than at 6 months. The reason for this might be that the samples taken at 6 months of age are taken too late to observe the initial domination of *Streptococcus* and that the amount of *Streptococcus* after the domination period is almost constant but fluctuates slightly. It could therefore be of interest for further studies to look at samples that are taken earlier in the infant's life.

4.4 Associations between bacteriocins and streptococcus in stool samples

Associations between the presence of bacteriocins and streptococcal reads were investigated with Kruskal Wallis tests and student t-tests. All the samples were used in the first test to get a general overview, but no significant association was found. Only when the samples were divided into the three age categories and then further into the seven streptococcal OTUs, an association between bacteriocins and streptococcal reads was found. Among the samples from infants at 12 months of age, the streptococcal OTU nr.2 was observed to be enriched when bacteriocins were present. One possible explanation could be that this streptococcal strain produced these bacteriocins and this possible association was therefore also tested in samples from infants at 6 months of age and mothers. No association was found between the OTU nr.2 and bacteriocins among these samples, which might have been expected if this strain was producing the bacteriocins.

4.5 Methodological considerations

The bacterial strains used in this thesis were chosen from bacterial risk groups 1 and 2 since the laboratory used at NMBU was approved for these types of bacteria. Another criterion was that they should be relatively safe to work with and not be pathogenic towards humans. Therefore, no *S.pneumoniae* strains were tested even though this could be of interest for further studies. Some of the tested bacteria are not usually found in the gut microbiota but were tested since they were relatively safe to work with. In the second round of testing for

bacteriocin activity, the representative for the genus *Listeria* was a *L.innocua* strain since this strain was harmless towards humans.

The results generated by the designed primer pairs indicated that there was some unspecific binding and that the specificity of the primers varied quite a bit (appendix D and E). Table 3.1 showed that it was only the BlpK primer pair that bound and amplified products in the in silico test against different contigs. This indicated that the specificity of the primer pair probably was more aimed at the *blp*-locus than only the *blpK* gene. As seen in appendix E, the BlpK primers was the primers that gave most unspecific binding which could be explained by the observed lower specificity. The only other primer that bound to contigs was the BlpD forward primer and this seems to give rise to less unspecific binding among the other primers (appendix D and E). Common for all the primers was that it was possible to distinguish between specific and unspecific binding based on the melting temperature and the length of the amplified PCR product. To be certain that the primers amplified the sequence of interest, it could be wise to sequence the PCR products. The length of the PCR products in this thesis was optimized for the PCR reaction and was too short for the clean-up with AMPure beads. Therefore, sequencing was not performed and a suggestion for further studies would be to design the primers so that the products become long enough for sequencing.

As seen in figure 3.4 all the primer pairs gave high Cq-values when used in stool samples which means that the number of cycles required to give a detectable signal in the qPCR was quite high. This was the case both for specific and unspecific binding and Cq-value could therefore not be used as a criterion for a positive result. Table 3.1 shows that most of the primers did not bind to the contigs from which they were identified, and it was, therefore, likely that the primers would underestimate the number of bacteriocin genes present in the samples. It was therefore decided that the results would only be presented qualitatively rather than quantitative. Further optimization of the primers could therefore be preferred in further studies.

The study of the bacteriocin activity in this thesis may be limited by the synthesis step and since there was no positive control for the bacteriocins, there was no way of knowing if the synthesis step worked. The method with the PURExpression kit had previously been successfully used on the control bacteriocin Ube K but had not yet been confirmed to work on the bacteriocins tested in this thesis (Oftedal et al., 2021). Previous studies with these bacteriocins have been successful in their synthesis by constructing vectors with the genes of interest and using *L.lactis* for protein expression (Fontaine and Hols, 2008). It was therefore

possible that the lack of inhibition zones in the activity assay was caused by an unsuccessful bacteriocin synthesis. It could be of interest for further studies to find a method to measure if the bacteriocins have been synthesised so that this source of error could be ruled out. Another possibility was that the synthesis resulted in non-functional proteins or proteins with low activity.

Some uncertainties were also tied to the 16S rRNA sequencing data. This dataset was only based on the reverse reads, since the forward reads had too low quality and this gave smaller fragments than if both forward and reverse reads could be used. The analysis was not able to classify further than to the genus level and it could be of interest to see if any of the streptococcal OTUs found matched the bacteriocin producer strains.

5. Conclusion and further research

In this thesis, it was possible to identify seven different OTUs belonging to *Streptococcus,* and these reads were significantly enriched in samples taken from infants compared to mothers which were consistent with previous studies. The identification of bacteriocin genes in stool samples was also successful. The pattern of bacteriocins identified did, however, not seem to fit with the initial hypothesis, since many of the bacteriocins were present in more samples from mothers compared to infants.

The only association between streptococcal reads and bacteriocins was found in samples from infants at 12 months of age. Here it was found that streptococcal reads belonging to OTU nr.2 were enriched in samples with bacteriocins, but this association was not found in samples from infants at 6 months of age or mothers. This could possibly be explained by the fact that the earliest samples were taken from infants at 6 months of age, which might be too late to discover the associations between bacteriocins and streptococcal strains during the colonization process. Further studies should therefore include samples from earlier in the infant's life.

None of the bacteriocins showed any activity during the bacteriocin activity test and this might be caused by a classification error, so it would be of interest for further studies to mix the bacteriocins. Another explanation was that the protein synthesis did not work for the bacteriocins, and it should therefore be tested if there were any bacteriocins produced by the synthesis. It might also be wise to add other bacteria typically found in the gut microbiota and to look for immunity genes amongst the bacteria.

References

ALANJARY, M., KRONMILLER, B., ADAMEK, M., BLIN, K., WEBER, T., HUSON, D., PHILMUS, B. & ZIEMERT, N. 2017. The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic acids research,* 45**,** W42-W48.

ALVAREZ-SIEIRO, P., MONTALBÁN-LÓPEZ, M., MU, D. & KUIPERS, O. P. 2016. Bacteriocins of lactic acid bacteria: extending the family. *Applied microbiology and biotechnology,* 100**,** 2939-2951.

ANGELL, I. L., NILSEN, M., CARLSEN, K. C. L., CARLSEN, K.-H., HEDLIN, G., JONASSEN, C. M., MARSLAND, B., NORDLUND, B., REHBINDER, E. M., SCHINAGL, C. M., SKJERVEN, H. O., STAFF, A. C., SÖDERHÄLL, C., VETTUKATTIL, M. R. & RUDI, K. 2020. De novo species identification using 16S rRNA gene nanopore sequencing.

ARBOLEYA, S., BINETTI, A., SALAZAR, N., FERNÁNDEZ, N., SOLÍS, G., HERNANDEZ-BARRANCO, A., MARGOLLES, A., DE LOS REYES-GAVILÁN, C. G. & GUEIMONDE, M. 2012. Establishment and development of intestinal microbiota in preterm neonates. *FEMS microbiology ecology,* 79**,** 763-772.

ARBOLEYA, S., SÁNCHEZ, B., MILANI, C., DURANTI, S., SOLÍS, G., FERNÁNDEZ, N., CLARA, G., VENTURA, M., MARGOLLES, A. & GUEIMONDE, M. 2015. Intestinal microbiota development in preterm neonates and effect of perinatal antibiotics. *The Journal of pediatrics,* 166**,** 538-544.

AVERSHINA, E., LUNDGÅRD, K., SEKELJA, M., DOTTERUD, C., STORRØ, O., ØIEN, T., JOHNSEN, R. & RUDI, K. 2016. Transition from infant- to adult-like gut microbiota. *Environ Microbiol,* 18**,** 2226-2236.

BAQUERO, F., LANZA, V. F., BAQUERO, M.-R., DEL CAMPO, R. & BRAVO-VÁZQUEZ, D. A. 2019. Microcins in Enterobacteriaceae: peptide antimicrobials in the eco-active intestinal chemosphere. *Frontiers in microbiology,* 10**,** 2261.

BARTLETT, J. M. & STIRLING, D. 2003. A short history of the polymerase chain reaction. *PCR protocols.* Springer.

BITEEN, J. S., BLAINEY, P. C., CARDON, Z. G., CHUN, M., CHURCH, G. M., DORRESTEIN, P. C., FRASER, S. E., GILBERT, J. A., JANSSON, J. K. & KNIGHT, R. 2016. Tools for the microbiome: nano and beyond. *ACS nano,* 10**,** 6-37.

BONNET, M., LAGIER, J. C., RAOULT, D. & KHELAIFIA, S. 2020. Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology. *New microbes and new infections,* 34**,** 100622.

BORGEAUD, S., METZGER, L. C., SCRIGNARI, T. & BLOKESCH, M. 2015. The type VI secretion system of Vibrio cholerae fosters horizontal gene transfer. *Science,* 347**,** 63-67.

BROEDERS, S., HUBER, I., GROHMANN, L., BERBEN, G., TAVERNIERS, I., MAZZARA, M., ROOSENS, N. & MORISSET, D. 2014. Guidelines for validation of qualitative real-time PCR methods. *Trends in Food Science & Technology,* 37**,** 115-126.

BRUGGER, S. D., ESLAMI, S. M., PETTIGREW, M. M., ESCAPA, I. F., HENKE, M. T., KONG, Y. & LEMON, K. P. 2020. Dolosigranulum pigrum cooperation and competition in human nasal microbiota. *Msphere,* 5**,** e00852-20.

BUSTIN, S., BENES, V., NOLAN, T. & PFAFFL, M. 2005. Quantitative real-time RT-PCR–a perspective. *Journal of molecular endocrinology,* 34**,** 597-601.

BUSTIN, S. & HUGGETT, J. 2017. qPCR primer design revisited. *Biomolecular detection and quantification,* 14**,** 19-28.

BÄCKHED, F., LEY, R. E., SONNENBURG, J. L., PETERSON, D. A. & GORDON, J. I. 2005. Host-bacterial mutualism in the human intestine. *science,* 307**,** 1915-1920.

CASÉN, C., VEBØ, H. C., SEKELJA, M., HEGGE, F. T., KARLSSON, M. K., CIEMNIEJEWSKA, E., DZANKOVIC, S., FRØYLAND, C., NESTESTOG, R., ENGSTRAND, L., MUNKHOLM, P., NIELSEN, O. H., ROGLER, G., SIMRÉN, M., ÖHMAN, L., VATN, M. H. & RUDI, K. 2015. Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Alimentary pharmacology & therapeutics,* 42**,** 71-83.

CHATZIDAKI-LIVANIS, M., GEVA-ZATORSKY, N. & COMSTOCK, L. E. 2016. Bacteroides fragilis type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proceedings of the National Academy of Sciences,* 113**,** 3627-3632.

CHO, C. E. B. & NORMAN, M. M. D. P. 2013. Cesarean section and development of the immune system in the offspring. *Am J Obstet Gynecol,* 208**,** 249-254.

COLLADO, M. C., CERNADA, M., BAÜERL, C., VENTO, M. & PÉREZ-MARTÍNEZ, G. 2012. Microbial ecology and host-microbiota interactions during early life stages. *Gut microbes,* 3**,** 352-365.

CORR, S. C., LI, Y., RIEDEL, C. U., O'TOOLE, P. W., HILL, C. & GAHAN, C. G. 2007. Bacteriocin production as a mechanism for the antiinfective activity of Lactobacillus salivarius UCC118. *Proceedings of the National Academy of Sciences,* 104**,** 7617-7621.

COTTER, P. D., ROSS, R. P. & HILL, C. 2013. Bacteriocins—a viable alternative to antibiotics? *Nature Reviews Microbiology,* 11**,** 95-105.

COX, L. M., YAMANISHI, S., SOHN, J., ALEKSEYENKO, A. V., LEUNG, J. M., RODRIGUEZ, Z., RODGERS, A. B., ROBINE, N., LOKE, P. N. & BLASER, M. J. 2014. Alternating the intestinal microbiota during critical developmental window has lasting metabolic consequences. *Cell,* 158**,** 705-721.

COYNE, M. J., ROELOFS, K. G. & COMSTOCK, L. E. 2016. Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC genomics,* 17**,** 1-21.

COYTE, K. Z. & RAKOFF-NAHOUM, S. 2019. Understanding competition and cooperation within the mammalian gut microbiome. *Current Biology,* 29**,** R538-R544.

COYTE, K. Z., SCHLUTER, J. & FOSTER, K. R. 2015. The ecology of the microbiome: networks, competition, and stability. *Science,* 350**,** 663-666.

DAVEY, L., HALPERIN, S. A. & LEE, S. F. 2016. Thiol-disulfide exchange in Gram-positive firmicutes. *Trends in microbiology,* 24**,** 902-915.

DECKER, E., HORNEF, M. & STOCKINGER, S. 2011. Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children. *Gut Microbes,* 2**,** 91-98.

DERRIEN, M., ALVAREZ, A.-S. & DE VOS, W. M. 2019. The gut microbiota in the first decade of life. *Trends in microbiology,* 27**,** 997-1010.

DICKSVED, J., HALFVARSON, J., ROSENQUIST, M., JÄRNEROT, G., TYSK, C., APAJALAHTI, J., ENGSTRAND, L. & JANSSON, J. K. 2008. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *The ISME journal,* 2**,** 716-727.

DIEFFENBACH, C., LOWE, T. & DVEKSLER, G. 1993. General concepts for PCR primer design. *PCR methods appl,* 3**,** S30-S37.

DOGRA, S., SAKWINSKA, O., SOH, S.-E., NGOM-BRU, C., BRÜCK, W. M., BERGER, B., BRÜSSOW, H., KARNANI, N., LEE, Y. S. & YAP, F. 2015. Rate of establishing the gut microbiota in infancy has consequences for future health. *Gut microbes,* 6**,** 321-325.

DOMINGUEZ-BELLO, M. G., BLASER, M. J., LEY, R. E. & KNIGHT, R. 2011. Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. *Gastroenterology,* 140**,** 1713-1719.

DSZM GMBH. 2012. *92. Trypticase Soy Yeast Extract Medium* [Online]. Available: https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium92.pdf [Accessed 07. February 2022].

ECKBURG, P. B., BIK, E. M., BERNSTEIN, C. N., PURDOM, E., DETHLEFSEN, L., SARGENT, M., GILL, S. R., NELSON, K. E. & RELMAN, D. A. 2005. Diversity of the human intestinal microbial flora. *science,* 308**,** 1635-1638.

FAITH, J. J., GURUGE, J. L., CHARBONNEAU, M., SUBRAMANIAN, S., SEEDORF, H., GOODMAN, A. L., CLEMENTE, J. C., KNIGHT, R., HEATH, A. C. & LEIBEL, R. L. 2013. The long-term stability of the human gut microbiota. *Science,* 341.

FAUST, K. & RAES, J. 2012. Microbial interactions: from networks to models. *Nature Reviews Microbiology,* 10**,** 538-550.

FONTAINE, L. & HOLS, P. 2008. The inhibitory spectrum of thermophilin 9 from Streptococcus thermophilus LMD-9 depends on the production of multiple peptides and the activity of BlpGSt, a thiol-disulfide oxidase. *Applied and environmental microbiology,* 74**,** 1102-1110.
FRANKEL, W. L., ZHANG, W., SINGH, A., KLURFELD, D. M., DON, S., SAKATA, T., MODLIN, I. & ROMBEAU, J. L. 1994. Mediation of the trophic effects of short-chain fatty acids on the rat jejunum and colon. *Gastroenterology,* 106**,** 375-380.
FRØSETH, M. S. 2022. *Factors Associated with Host DNA Methylation in Infant Stool.* Master's thesis, Norwegian University Of Life science

GABANT, P. & BORRERO, J. 2019. PARAGEN 1.0: A standardized synthetic gene library for fast cell-free bacteriocin synthesis. *Frontiers in bioengineering and biotechnology,* 7**,** 213.
GARCÍA-CURIEL, L., DEL ROCÍO LÓPEZ-CUELLAR, M., RODRÍGUEZ-HERNÁNDEZ, A. I. & CHAVARRÍA-HERNÁNDEZ, N. 2021. Toward understanding the signals of bacteriocin production by Streptococcus spp. and their importance in current applications. *World Journal of Microbiology and Biotechnology,* 37**,** 1-14.
GAUSE, G. F. 2019. *The struggle for existence: a classic of mathematical biology and ecology*, Courier Dover Publications.
GENEIOUS. 2022. *Primer design* [Online]. [Accessed 21. January 2022].
GHOUL, M. & MITRI, S. 2016. The ecology and evolution of microbial competition. *Trends in microbiology,* 24**,** 833-845.
GOSALBES, M., LLOP, S., VALLES, Y., MOYA, A., BALLESTER, F. & FRANCINO, M. 2013. Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. *Clinical & Experimental Allergy,* 43**,** 198-211.
GRATIA, J.-P. 2000. Andre Gratia: a forerunner in microbial and viral genetics. *Genetics,* 156**,** 471-476.
GUARNER, F. & MALAGELADA, J.-R. 2003. Gut flora in health and disease. *The Lancet,* 361**,** 512-519.
GUERRA, N. P., MACIAS, C. L., AGRASAR, A. T. & CASTRO, L. 2005. Development of a bioactive packaging cellophane using Nisaplin® as biopreservative agent. *Letters in applied microbiology,* 40**,** 106-110.
HEATHER, J. M. & CHAIN, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107**,** 1-8.
HEILBRONNER, S., KRISMER, B., BRÖTZ-OESTERHELT, H. & PESCHEL, A. 2021. The microbiome-shaping roles of bacteriocins. *Nature Reviews Microbiology***,** 1-14.
HILL, C. J., LYNCH, D. B., MURPHY, K., ULASZEWSKA, M., JEFFERY, I. B., O'SHEA, C. A., WATKINS, C., DEMPSEY, E., MATTIVI, F. & TUOHY, K. 2017. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome,* 5**,** 1-18.
ILLUMINA. 2022. *Understanding the NGS workflow* [Online]. Available: https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html [Accessed 14. January 2022].
JAKOBSSON, H. E., ABRAHAMSSON, T. R., JENMALM, M. C., HARRIS, K., QUINCE, C., JERNBERG, C., BJÖRKSTÉN, B., ENGSTRAND, L. & ANDERSSON, A. F. 2014. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut,* 63**,** 559-566.
JANEK, D., ZIPPERER, A., KULIK, A., KRISMER, B. & PESCHEL, A. 2016. High frequency and diversity of antimicrobial activities produced by nasal Staphylococcus strains against bacterial competitors. *PLoS pathogens,* 12**,** e1005812.
KATZEN, F., CHANG, G. & KUDLICKI, W. 2005. The past, present and future of cell-free protein synthesis. *Trends in biotechnology,* 23**,** 150-156.
KERR, C. A., GRICE, D. M., TRAN, C. D., BAUER, D. C., LI, D., HENDRY, P. & HANNAN, G. N. 2015. Early life events influence whole-of-life metabolic health via gut microflora and gut permeability. *Critical reviews in microbiology,* 41**,** 326-340.

KIGAWA, T., YABUKI, T., YOSHIDA, Y., TSUTSUI, M., ITO, Y., SHIBATA, T. & YOKOYAMA, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS letters,* 442**,** 15-19.

KIM, S. G., BECATTINI, S., MOODY, T. U., SHLIAHA, P. V., LITTMANN, E. R., SEOK, R., GJONBALAJ, M., EATON, V., FONTANA, E. & AMORETTI, L. 2019. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature,* 572**,** 665-669.

KOENIG, J. E., SPOR, A., SCALFONE, N., FRICKER, A. D., STOMBAUGH, J., KNIGHT, R., ANGENENT, L. T. & LEY, R. E. 2011. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences,* 108**,** 4578-4585.

KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., FOROOTAN, A., JONÁK, J., LIND, K., SINDELKA, R., SJÖBACK, R., SJÖGREEN, B. & STRÖMBOM, L. 2006. The real-time polymerase chain reaction. *Molecular aspects of medicine,* 27**,** 95-125.

KULKARNI, S. S., SAYERS, J., PREMDJEE, B. & PAYNE, R. J. 2018. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature Reviews Chemistry,* 2**,** 1-17.

LEMOS, J., PALMER, S., ZENG, L., WEN, Z., KAJFASZ, J., FREIRES, I., ABRANCHES, J. & BRADY, L. 2019. The biology of Streptococcus mutans. *Microbiology spectrum,* 7**,** 7.1. 03.

LEY, R. E., PETERSON, D. A. & GORDON, J. I. 2006a. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell,* 124**,** 837-848.

LEY, R. E., TURNBAUGH, P. J., KLEIN, S. & GORDON, J. I. 2006b. Human gut microbes associated with obesity. *nature,* 444**,** 1022-1023.

LIU, Y., LIU, Y., DU, Z., ZHANG, L., CHEN, J., SHEN, Z., LIU, Q., QIN, J., LV, H. & WANG, H. 2020. Skin microbiota analysis-inspired development of novel anti-infectives. *Microbiome,* 8**,** 1-19.

LOZUPONE, C. A., STOMBAUGH, J. I., GORDON, J. I., JANSSON, J. K. & KNIGHT, R. 2012. Diversity, stability and resilience of the human gut microbiota. *Nature,* 489**,** 220-230.

LØDRUP CARLSEN, K. C., REHBINDER, E. M., SKJERVEN, H. O., CARLSEN, M. H., FATNES, T. A., FUGELLI, P., GRANUM, B., HAUGEN, G., HEDLIN, G., JONASSEN, C. M., LANDRØ, L., LUNDE, J., MARSLAND, B. J., NORDLUND, B., RUDI, K., SJØBORG, K., SÖDERHÄLL, C., CATHRINE STAFF, A., VETTUKATTIL, R., CARLSEN, K. H., ASARNOJ, A., AUVINEN, P., STENSBY BAINS, K. E., CARLSEN, O. C., ENDRE, K. A., GRANLUND, P. A., GUDMUNDSDÓTTIR, H. K., HAAHTELA, T., HILDE, K., HOLMSTRØM, H., HÅLAND, G., KREYBERG, I., MÄGI, C. A. O., NORDHAGEN, L. S., NYGAARD, U. C., SCHINAGL, C. M., SKRINDO, I., SJELMO, S., TEDNER, S. G., VÆRNESBRANDEN, M. R. & WIIK, J. 2018. Preventing Atopic Dermatitis and ALLergies in Children—the PreventADALL study. *Allergy,* 73**,** 2063-2070.

MAGI, A., SEMERARO, R., MINGRINO, A., GIUSTI, B. & D'AURIZIO, R. 2018. Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief Bioinform,* 19**,** 1256-1272.

MALDONADO-BARRAGÁN, A. & WEST, S. A. 2020. The cost and benefit of quorum sensing-controlled bacteriocin production in Lactobacillus plantarum. *Journal of evolutionary biology,* 33**,** 101-111.

MANTOVANI, H. C., HU, H., WOROBO, R. W. & RUSSELL, J. B. 2002. Bovicin HC5, a bacteriocin from Streptococcus bovis HC5. *Microbiology,* 148**,** 3347-3352.

MEIJERINK, M., VAN HEMERT, S., TAVERNE, N., WELS, M., DE VOS, P., BRON, P. A., SAVELKOUL, H. F., VAN BILSEN, J., KLEEREBEZEM, M. & WELLS, J. M. 2010. Identification of genetic loci in Lactobacillus plantarum that modulate the immune response of dendritic cells using comparative genome hybridization. *PloS one,* 5**,** e10632.

MIGNOLET, J., FONTAINE, L., SASS, A., NANNAN, C., MAHILLON, J., COENYE, T. & HOLS, P. 2018. Circuitry rewiring directly couples competence to predation in the gut dweller Streptococcus salivarius. *Cell reports,* 22**,** 1627-1638.

MILANI, C., DURANTI, S., BOTTACINI, F., CASEY, E., TURRONI, F., MAHONY, J., BELZER, C., PALACIO, S. D., MONTES, S. A., MANCABELLI, L., LUGLI, G. A., ROIGUEZ, J. M., BODE, L., VOS, D. W., GUEIMONDE, M., MARGOLLES, A., SINDEREN, V. D. & VENTURA, M. 2017. The first microbial

colonizers of the human gut : Composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev,* 81.
MILLS, S., ROSS, R. P. & HILL, C. 2017. Bacteriocins and bacteriophage; a narrow-minded approach to food and gut microbiology. *FEMS microbiology reviews,* 41**,** S129-S153.
MIYAZAWA, E., IWABUCHI, A. & YOSHIDA, T. 1996. Phytate breakdown and apparent absorption of phosphorus, calcium and magnesium in germfree and conventionalized rats. *Nutrition Research,* 16**,** 603-613.
NADKARNI, M. A., MARTIN, F. E., JACQUES, N. A. & HUNTER, N. 2002. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology,* 148**,** 257-266.
NES, I. F., DIEP, D. B., HÅVARSTEIN, L. S., BRURBERG, M. B., EIJSINK, V. & HOLO, H. 1996. Biosynthesis of bacteriocins in lactic acid bacteria. *Antonie van Leeuwenhoek,* 70**,** 113-128.
NEW ENGLAND BIOLABS. 2022. *PURExpress In Vitro Protein Synthesis Kit* [Online]. Available: https://international.neb.com/products/e6800-purexpress-invitro-protein-synthesis-kit# [Accessed 24.January 2022].
NGUYEN, N.-P., WARNOW, T., POP, M. & WHITE, B. 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ biofilms and microbiomes,* 2**,** 1-8.
NISSEN-MEYER, J., OPPEGÅRD, C., ROGNE, P., HAUGEN, H. S. & KRISTIANSEN, P. E. 2010. Structure and mode-of-action of the two-peptide (class-IIb) bacteriocins. *Probiotics and antimicrobial proteins,* 2**,** 52-60.
O'SULLIVAN, A., FARVER, M. & SMILOWITZ, J. T. 2015. Article Commentary: The influence of early infant-feeding practices on the intestinal microbiome and body composition in infants. *Nutrition and metabolic insights,* 8**,** NMI. S29530.
OFTEDAL, T. F., OVCHINNIKOV, K. V., HESTAD, K. A., GOLDBECK, O., PORCELLATO, D., NARVHUS, J., RIEDEL, C. U., KJOS, M. & DIEP, D. B. 2021. Ubericin K, a New Pore-Forming Bacteriocin Targeting mannose-PTS. *Microbiology spectrum,* 9**,** e00299-21.
OXFORD NANOPORE, T. 2021. *To enable the analysis of anything, by anyone, anywhere* [Online]. Available: https://nanoporetech.com/about-us [Accessed 14. January 2022].
PENDERS, J., THIJS, C., VINK, C., STELMA, F. F., SNIJDERS, B., KUMMELING, I., VAN DEN BRANDT, P. A. & STOBBERINGH, E. E. 2006. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics,* 118**,** 511-521.
RAJILIĆ-STOJANOVIĆ, M. & DE VOS, W. M. 2014. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS microbiology reviews,* 38**,** 996-1047.
RAVI, A., AVERSHINA, E., ANGELL, I. L., LUDVIGSEN, J., MANOHAR, P., PADMANABAN, S., NACHIMUTHU, R., SNIPEN, L. & RUDI, K. 2018. Comparison of reduced metagenome and 16S rRNA gene sequencing for determination of genetic diversity and mother-child overlap of the gut associated microbiota. *J Microbiol Methods,* 149**,** 44-52.
REHBINDER, E. M., CARLSEN, K. C. L., STAFF, A. C., ANGELL, I. L., LANDRØ, L., HILDE, K., GAUSTAD, P. & RUDI, K. 2018. Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? *American journal of obstetrics and gynecology,* 219**,** 289. e1-289. e12.
RENYE, J. A., SOMKUTI, G. A. & STEINBERG, D. H. 2019. Thermophilin 109 is a naturally produced broad spectrum bacteriocin encoded within the blp gene cluster of Streptococcus thermophilus. *Biotechnology letters,* 41**,** 283-292.
RILEY, M. A. & WERTZ, J. E. 2002. Bacteriocins: evolution, ecology, and application. *Annual Reviews in Microbiology,* 56**,** 117-137.
RODRÍGUEZ, J. M., MURPHY, K., STANTON, C., ROSS, R. P., KOBER, O. I., JUGE, N., AVERSHINA, E., RUDI, K., NARBAD, A., JENMALM, M. C., MARCHESI, J. R. & COLLADO, M. C. 2015. The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis,* 26**,** 26050-26050.
ROSSI, F., MARZOTTO, M., CREMONESE, S., RIZZOTTI, L. & TORRIANI, S. 2013. Diversity of Streptococcus thermophilus in bacteriocin production; inhibitory spectrum and occurrence of thermophilin genes. *Food microbiology,* 35**,** 27-33.

ROTVOLD, K. 2022. *A study on the potential role of bacteriocins in shaping the infant gut microbiota, through inhibition of bacterial growth by selected bacteriocins and presence of enterocin A encoding enterococci in fecal samples.* Master's thesis, Norwegian University Of Life science.

RUIZ, L., MOLES, L., GUEIMONDE, M. & RODRIGUEZ, J. M. 2016. Perinatal microbiomes' influence on preterm birth and preterms' health: influencing factors and modulation strategies. *Journal of pediatric gastroenterology and nutrition,* 63**,** e193-e203.

SAN MILLÁN, R. M., MARTÍNEZ-BALLESTEROS, I., REMENTERIA, A., GARAIZAR, J. & BIKANDI, J. 2013. Online exercise for the design and simulation of PCR and PCR-RFLP experiments. *BMC Research Notes,* 6**,** 1-4.

SÁNCHEZ-HIDALGO, M., MONTALBÁN-LÓPEZ, M., CEBRIÁN, R., VALDIVIA, E., MARTÍNEZ-BUENO, M. & MAQUEDA, M. 2011. AS-48 bacteriocin: close to perfection. *Cellular and Molecular Life Sciences,* 68**,** 2845-2857.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SANIDAD, K. Z. & ZENG, M. Y. 2020. Neonatal gut microbiome and immunity. *Current opinion in microbiology,* 56**,** 30-37.

SHANKER, E. & FEDERLE, M. J. 2017. Quorum sensing regulation of competence and bacteriocins in Streptococcus pneumoniae and mutans. *Genes,* 8**,** 15.

SIMONS, A., ALHANOUT, K. & DUVAL, R. E. 2020. Bacteriocins, antimicrobial peptides from bacterial origin: overview of their biology and their impact against multidrug-resistant bacteria. *Microorganisms,* 8**,** 639.

SOLÍS, G., DE LOS REYES-GAVILAN, C., FERNÁNDEZ, N., MARGOLLES, A. & GUEIMONDE, M. 2010. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe,* 16**,** 307-310.

SØRENSEN, H. P. & MORTENSEN, K. K. 2005a. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of biotechnology,* 115**,** 113-128.

SØRENSEN, H. P. & MORTENSEN, K. K. 2005b. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial cell factories,* 4**,** 1-8.

TALAGAS, A., FONTAINE, L., LEDESMA-GARCA, L., MIGNOLET, J., LI DE LA SIERRA-GALLAY, I., LAZAR, N., AUMONT-NICAISE, M., FEDERLE, M. J., PREHNA, G. & HOLS, P. 2016. Structural insights into streptococcal competence regulation by the cell-to-cell communication system ComRS. *PLoS pathogens,* 12**,** e1005980.

THAVAGNANAM, S., FLEMING, J., BROMLEY, A., SHIELDS, M. D. & CARDWELL, C. R. 2008. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy,* 38**,** 629-633.

THOMAS, J., WATVE, S. S., RATCLIFF, W. C. & HAMMER, B. K. 2017. Horizontal gene transfer of functional type VI killing genes by natural transformation. *MBio,* 8**,** e00654-17.

THURSBY, E. & JUGE, N. 2017. Introduction to the human gut microbiota. *Biochemical Journal,* 474**,** 1823-1836.

TROSVIK, P., STENSETH, N. C. & RUDI, K. 2010. Convergent temporal dynamics of the human infant gut microbiota. *The ISME journal,* 4**,** 151-158.

TURNBAUGH, P. J., HAMADY, M., YATSUNENKO, T., CANTAREL, B. L., DUNCAN, A., LEY, R. E., SOGIN, M. L., JONES, W. J., ROE, B. A. & AFFOURTIT, J. P. 2009. A core gut microbiome in obese and lean twins. *nature,* 457**,** 480-484.

TURNBAUGH, P. J., LEY, R. E., HAMADY, M., FRASER-LIGGETT, C. M., KNIGHT, R. & GORDON, J. I. 2007. The human microbiome project. *Nature,* 449**,** 804-810.

TURNBAUGH, P. J., LEY, R. E., MAHOWALD, M. A., MAGRINI, V., MARDIS, E. R. & GORDON, J. I. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *nature,* 444**,** 1027-1031.

VARAS, M. A., MUÑOZ-MONTECINOS, C., KALLENS, V., SIMON, V., ALLENDE, M. L., MARCOLETA, A. E. & LAGOS, R. 2020. Exploiting zebrafish xenografts for testing the in vivo antitumorigenic

activity of microcin E492 against human colorectal cancer cells. *Frontiers in microbiology,* 11**,** 405.

VILLARREAL, A. A., ABERGER, F. J., BENRUD, R. & GUNDRUM, J. D. 2012. Use of broad-spectrum antibiotics and the development of irritable bowel syndrome. *Wmj,* 111**,** 17-20.

WANG, C. Y. & DAWID, S. 2018. Mobilization of bacteriocins during competence in streptococci. *Trends in microbiology,* 26**,** 389-391.

WEXLER, A. G., BAO, Y., WHITNEY, J. C., BOBAY, L.-M., XAVIER, J. B., SCHOFIELD, W. B., BARRY, N. A., RUSSELL, A. B., TRAN, B. Q. & GOO, Y. A. 2016. Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proceedings of the National Academy of Sciences,* 113**,** 3639-3644.

WHOLEY, W.-Y., ABU-KHDEIR, M., YU, E. A., SIDDIQUI, S., ESIMAI, O. & DAWID, S. 2019. Characterization of the competitive pneumocin peptides of Streptococcus pneumoniae. *Frontiers in cellular and infection microbiology,* 9**,** 55.

WINAND, R., BOGAERTS, B., HOFFMAN, S., LEFEVRE, L., DELVOYE, M., BRAEKEL, J. V., FU, Q., ROOSENS, N. H., KEERSMAECKER, S. C. D. & VANNESTE, K. 2019. Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *Int J Mol Sci,* 21**,** 298.

YANG, S.-C., LIN, C.-H., SUNG, C. T. & FANG, J.-Y. 2014. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Frontiers in microbiology,* 5**,** 241.

YOUNES, H., COUDRAY, C., BELLANGER, J., DEMIGNÉ, C., RAYSSIGUIER, Y. & RÉMÉSY, C. 2001. Effects of two fermentable carbohydrates (inulin and resistant starch) and their combination on calcium and magnesium balance in rats. *British Journal of Nutrition,* 86**,** 479-485.

ZHENG, J., GÄNZLE, M. G., LIN, X. B., RUAN, L. & SUN, M. 2015. Diversity and dynamics of bacteriocins from human microbiome. *Environmental microbiology,* 17**,** 2133-2143.

ZHOU, X. X., LI, W. F., MA, G. X. & PAN, Y. J. 2006. The nisin-controlled gene expression system: construction, application and improvements. *Biotechnology Advances,* 24**,** 285-295.

ZIPPERER, A., KONNERTH, M. C., LAUX, C., BERSCHEID, A., JANEK, D., WEIDENMAIER, C., BURIAN, M., SCHILLING, N. A., SLAVETINSKY, C. & MARSCHAL, M. 2016. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature,* 535**,** 511-516.

Appendix

Appendix A: Overview of the seven bacteriocins synthesized and used for the activity assay

Table A.1: Overview of the seven bacteriocins synthesized and tested for activity in this thesis. The bacteriocins were chosen based on their significant presence in gut metagenomes from infants compared to mothers in the bioinformatic study by Ida Ormaasen. The classification in the table is based on the results from BAGEL4/BACTIBASE.

Bacteriocin	Bacteriocin class	Producer species	Protein sequence (length)
BlpU	2d	*Streptococcus pneumoniae*	MNTKTMSQFEIMDTEMLACVEGGGCN WGDFAKAGVGGGAARGLQLGIKTRTW QGAATGAVGGAILGGVAYAATCWW (76)
BlpD	2	*Streptococcus thermophilus LMD-9*	MATQTIENFNTLDLETLASVEGGLSCDEG MLAVGGLGAVGGPWGAVGGVLVGAALY CF (58)
BlpK	2a	*Streptococcus thermophilus*	MATQTIENFNTLDLETLASVEGGGCSWRG AGGATVQGAIGGAFGGNVVLPVVGSVPG YLAGGVLGGAGGTVAYGATCWWS (80)
BlpU_2	2d	*Streptococcus thermophilus*	MATQTIENFNTLDLETLASVEGGGCSWGG FAKQGVATGVGNGLRLGIKTRTWQGAVAG AAGGAIVGGVGYGATCWW (76)
BlpK_2	2d	*Streptococcus pneumoniae*	MDTKMMSQFSVMDTEMLACVEGGGCNW GDFAKAGVGGGAARGLQLGIKTGTWQGAA TGAAGGAILGGVAYAATCWW (76)
BlpJ	2d	*Streptococcus pneumoniae TIGR4*	MNTKMLSQLEVMDTEMLAKVEGGYSSTDC QNALITGVTTGIITGGTGAGLATLGVAGLAG AFVGAHIGAIGGGLTCLGGMVGDKLGLSW (89)
Bov255	2d	*Streptococcus sp.*	MNTKTFEQFDVMTDEALSTVEGGGKGYCK PVYYAANGYSCRYSNGEWGYVVTKGAFQA TTDVIANGWVSSLGGGYFGKP (79)

Appendix B: Growth medium

Tryptic soy yeast extract agar/broth (400 mL)

12 g Tryptic soy broth (Sigma-Aldrich)

1.2 g Yeast extract (Merck)

6 g * agar (VWR)

400 mL H_2O

*For TYSE broth, leave out the agar

Brain Hearth Infusion agar/broth (400 mL)

14.8 g Brain hearth infusion broth (VWR, Belgium)

6 g* agar (VWR)

400 mL H_2O

*For BHI broth, leave out the agar

Appendix C: Measurements to find the early source of contamination

The streptococcal strains ordered from DSMZ were grown on the TSYE medium (described in appendix B) at 37 degrees with microaerophile conditions overnight. When the plates were inspected the next day, clear contaminations were observed. To confirm that it was not only *Streptococcus* on the plates, samples taken from the plates were also inspected by light microscopy and Sanger sequencing.

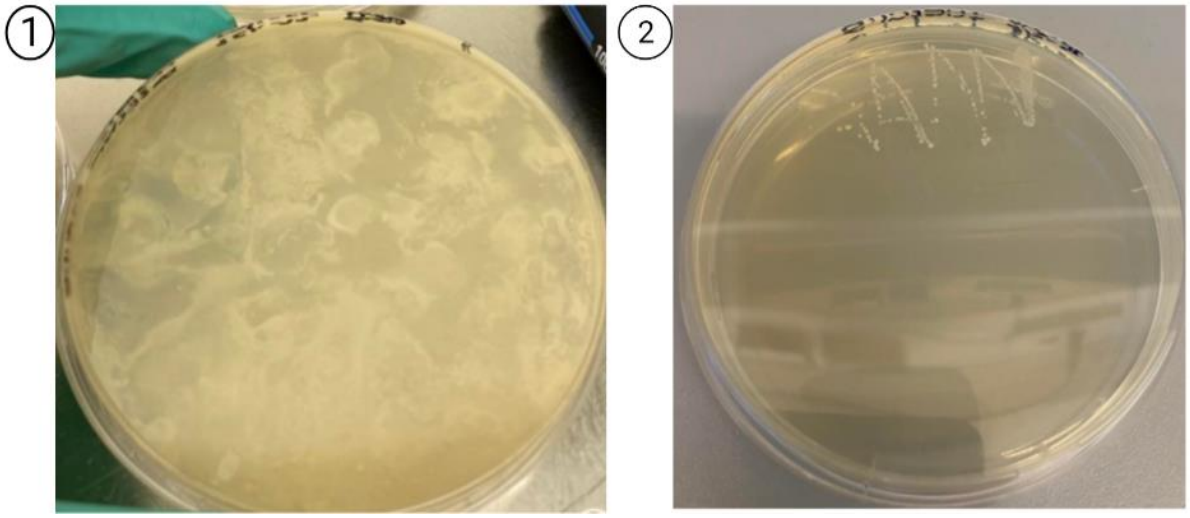

Figure C.1: Plates with (1) and without (2) contamination. *S.infantarius* was spread out using a streak plate technique on plates with TSYE agar and was incubated overnight at 37 degrees in microaerophilic conditions. Plate nr.1 shows clear contaminations on a plate early in the experiment, while plate nr.2 shows how colonies from *S.infantarius* were expected to look. The picture to the right is taken after the source of contamination was found.

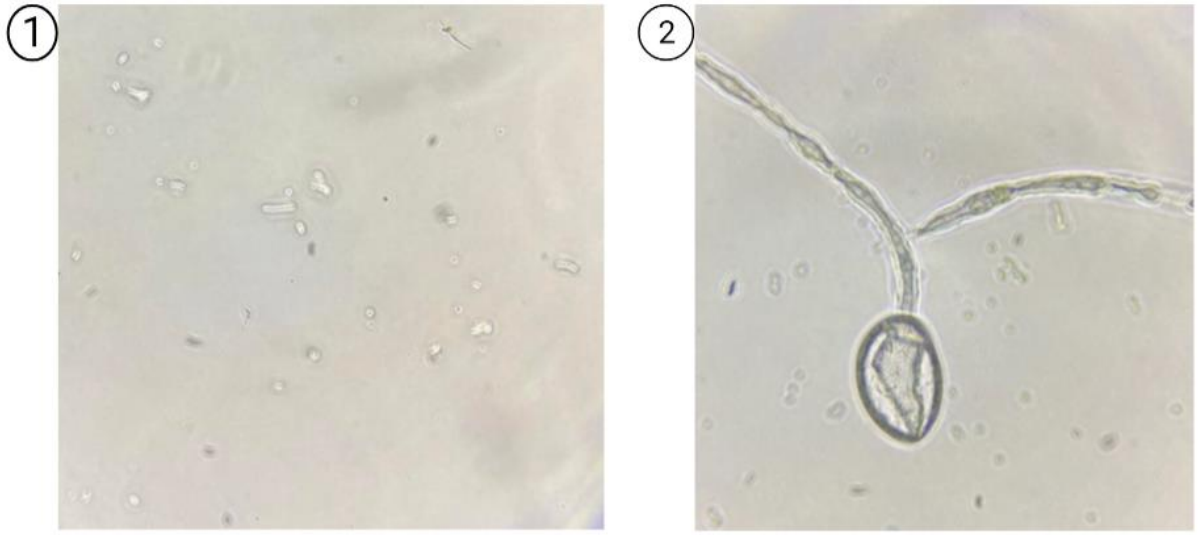

Figure C.2: Light microscopy of plates with contaminations. Picture nr.1 shows bacteria that during the Sanger sequencing were confirmed to be *Bacillus* while the picture nr.2 shows a cell that likely belongs to a fungus, these pictures were taken in the early phases of the experiment.

Table C.1: Bacterial strains identified by Sanger sequencing and BLAST search. The table shows the correspondence between bacterial strains ordered from DSMZ and the BLAST hits based on the results of selected samples.

DSMZ STRAIN	BLAST HIT	QUERY COVER (%)	IDENTITY (%)
S.EQUINIS	Streptococcus equinius	99	99,72
S.INFANTARIUS	Streptococcus infanterius	99	99,71
S.SANGUINIS	Streptococcus sanguinis	99	99,64
S.VESTIBULARIS	Streptococcus vestibularis	99	99,64
S.SALIVARIUS	Streptococcus salivarius	99	99,21
S.THERMOPHILUS	Streptococcus thermophilus	99	99,64
S.SANGUINIS	Lepadogaster lepadogaster voucher *	100	100
S.SANGUINIS	Bacillus licheniformis	100	100
S.EQUINIS	Bacillus licheniformis	100	99,63
S.EQUINIS	Lutra lutra*	100	100
S.THERMOPHILUS	Bacillus hayenesii	99	99,91
S.SANGUINIS	Pseudimonas protegens	89	97,44
S.SANGUINIS	Bacillus anthracis	91	93,4
S.SALIVARIUS	Populus trichocarpa*	100	100
S.VESTIBULARIS	Bacillus licheniformis	98	99,83
S.INFANTRSIUS	Bacillus licheniformis	99	99,51

*Marks the sequences that were shorter than 30 bp

The streptococcal colonies were supposed to be small and white as shown in the right part of figure C.1, but at the start of the experiment, all the plates had contaminations such as the plate in the left part of figure C.1. The source of the contamination could possibly be the sterile bench, the autoclave, the incubator, or the handling of the plates. To find the most likely source of contamination, different measurements were taken. Else Marie Aasen provided KPG-plates that were used to test the different hypotheses.

To exclude the incubator as the contamination source, empty KPG-plates were incubated both aerobic and microaerophile in the incubator at 37 °C overnight and for two days. No growth was observed on these plates. The next source that was excluded was the sterile bench. This was done by putting two KPG-plates with their lids open for 30 sec and 1 min on the sterile bench, before incubation at 37 °C. The plates were checked overnight and after two days. Neither of these plates had any growth. KPG-plates were also used to test if the streak plate technique and the handling of the plates could be the source of contamination, but only colonies belonging to the streptococcal strains were observed.

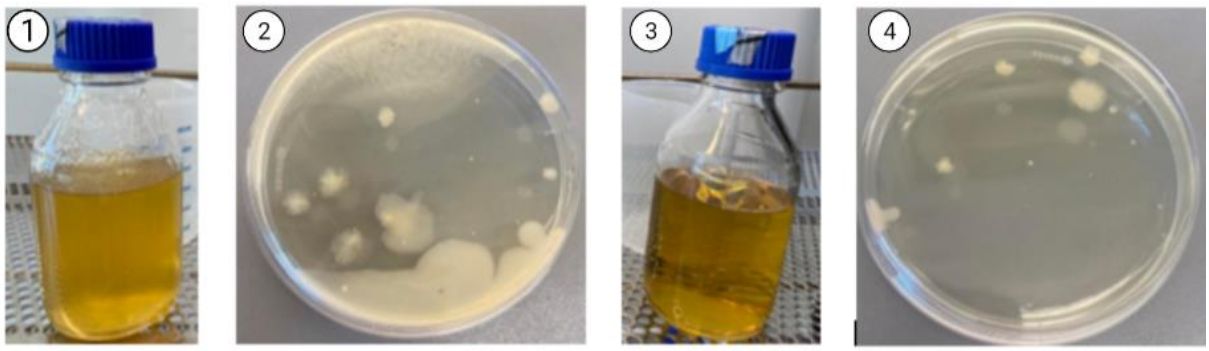

Figure C.3: Bottles with TSYE agar and overnight plates with TSYE agar after incubation. The two bottles contained the same batch of TSYE agar and stood beside each other in the autoclave, but the agar in the left bottle (1) had started to solidify while the agar in the right bottle (3) was still fluent. The plates contain TSYE agar from the bottles and were incubated overnight at room temperature.

The last possible contamination source that was tested was the autoclave. This test was performed by making the TSYE medium as described by the manufacturer and autoclaving it. The batch was divided into two bottles before autoclaving. The two bottles were taken out of the autoclave, but they did not appear to be similar. In one of the bottles, the agar had solidified at the bottom, while the agar in the other bottle was liquid. The bottle with liquid agar was spread on plates at once, while the bottle with solid agar was warmed in the microwave before spreading on plates. All the plates were incubated at room temperature in the dark for two days and growth was observed on all the plates. Therefore, the same steps were used to make a new batch of TSYE agar, but a new smaller autoclave was used before incubation at room temperature in the dark. No growth was observed on these plates after two days and the autoclave was, therefore, the likely source of contamination, and the smaller autoclave was used throughout this thesis to avoid further contamination.

Appendix D: Optimalization of primers and technical issues with the gradient PCR

Appendix D1: Optimalization of annealing temperatures

Gradient PCR and gel electrophoresis were performed to estimate the optimal annealing temperature for the designed primers. The primers were designed based on metagenome results from previous studies performed by Ida Ormaasen and this study also showed that the BlpU accounted for the highest proportion of the bacteriocins identified in the samples. The primer pair designed for BlpU was therefore chosen for the initial testing since it most likely would be present in these samples as well. The results of the gradient PCR are shown in figure D.1 and the different annealing temperatures are shown in table D.1.

Table D.1: Annealing temperatures (T_a) used in gradient PCR. The number of the wells corresponds to the wells on the three gels.

Well	1	2	3	4	5	6	7	8	9	10	11	12
T_a (°C)	52.9	53.2	54.0	55.2	56.8	58.7	60.6	62.5	64.3	65.8	66.9	67.5

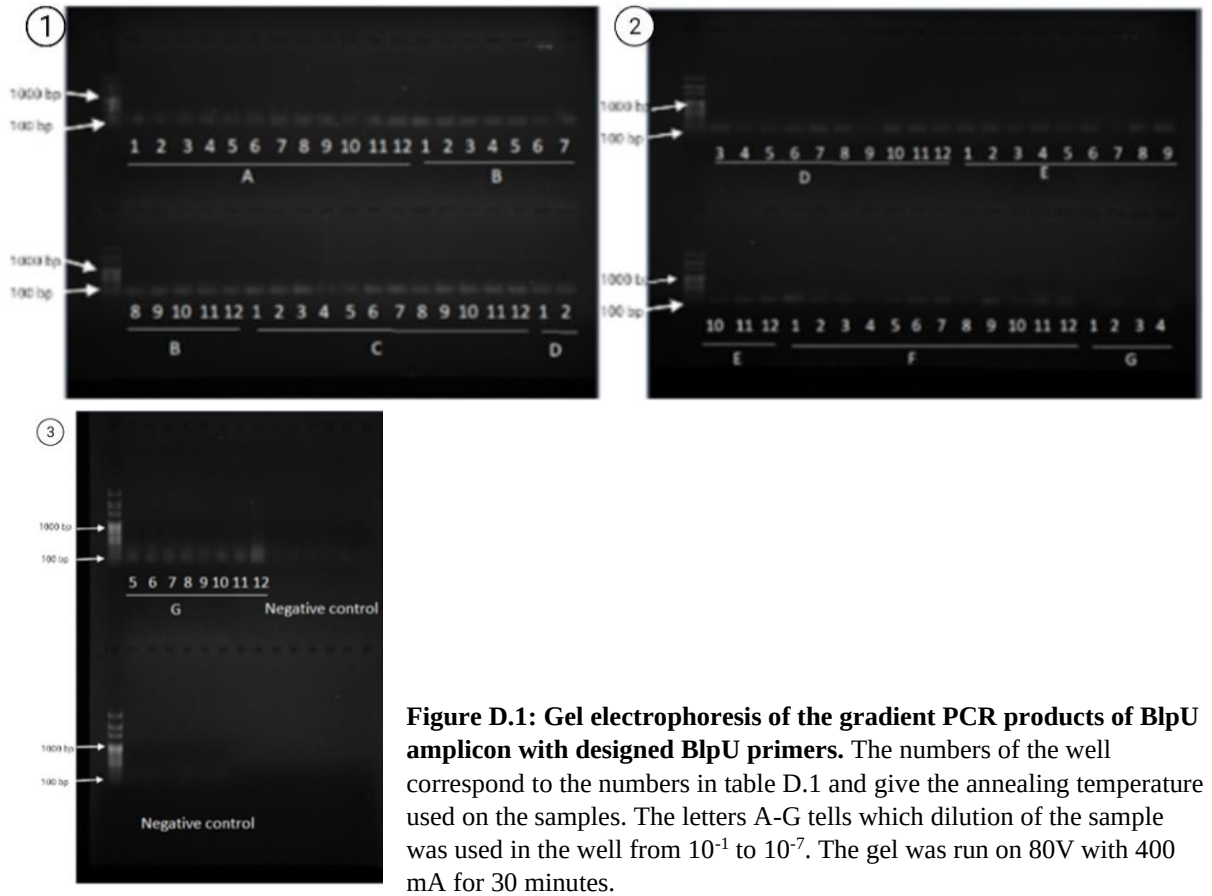

Figure D.1: Gel electrophoresis of the gradient PCR products of BlpU amplicon with designed BlpU primers. The numbers of the well correspond to the numbers in table D.1 and give the annealing temperature used on the samples. The letters A-G tells which dilution of the sample was used in the well from 10^{-1} to 10^{-7}. The gel was run on 80V with 400 mA for 30 minutes.

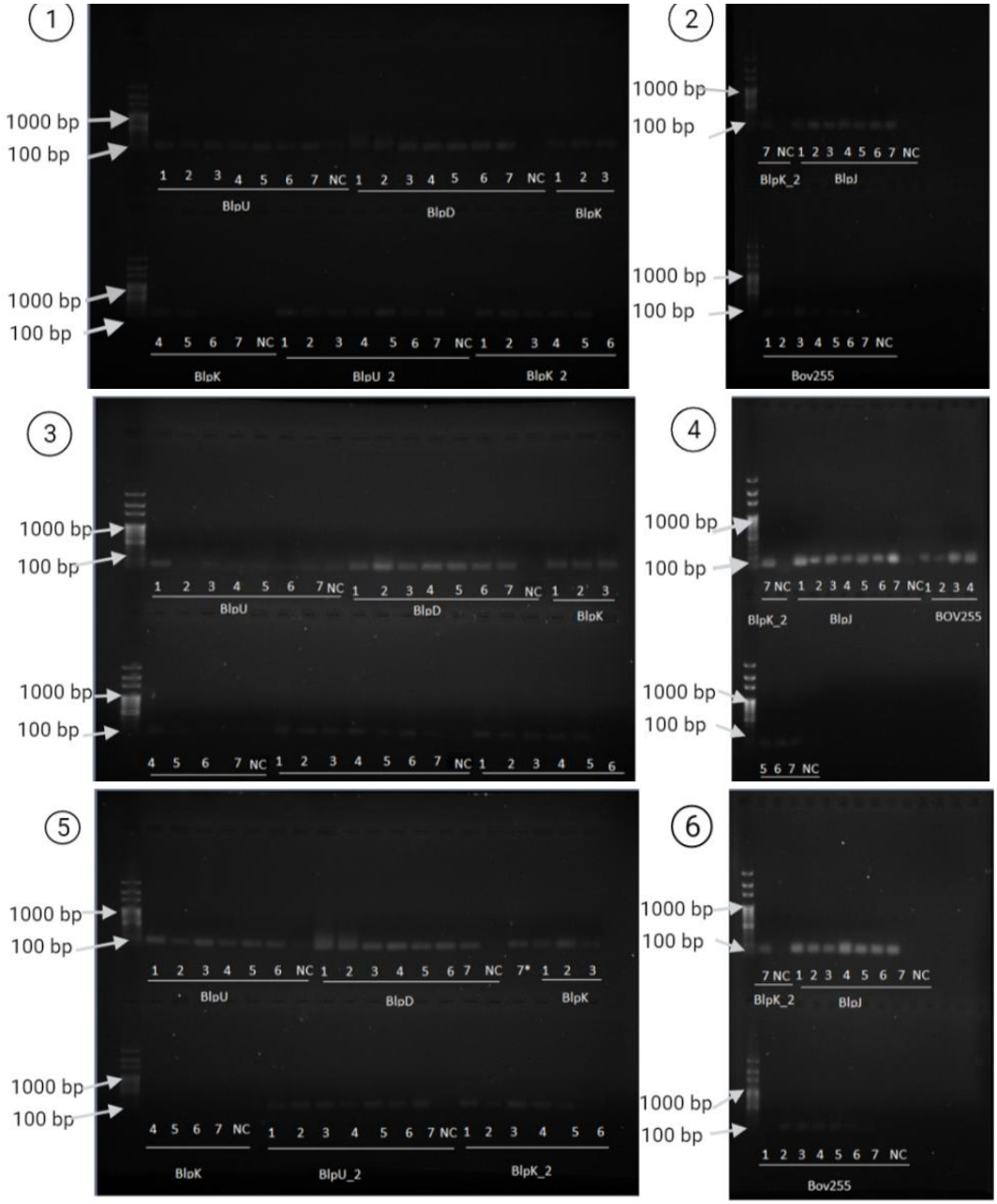

Figure D.2: PCR products of the designed primers with an annealing temperature of 56 °C (1-2), 60 °C (3-4), and 64 °C (5-6). The primers were used on the amplicons which they were designed for, and the numbers correspond with the dilution of the template from 10^{-1} to 10^{-7}. The gel was run on 80V with 400 mA for 30 minutes.

Figure D.2 shows that all the primers gave amplification with all the three different annealing temperatures. Since the BlpK primer pair had bound to multiple contigs during the in silico testing this primer pair was the next to be optimized further. On qPCR run with an annealing temperature of 64°C and one with an annealing temperature of 60°C were run.

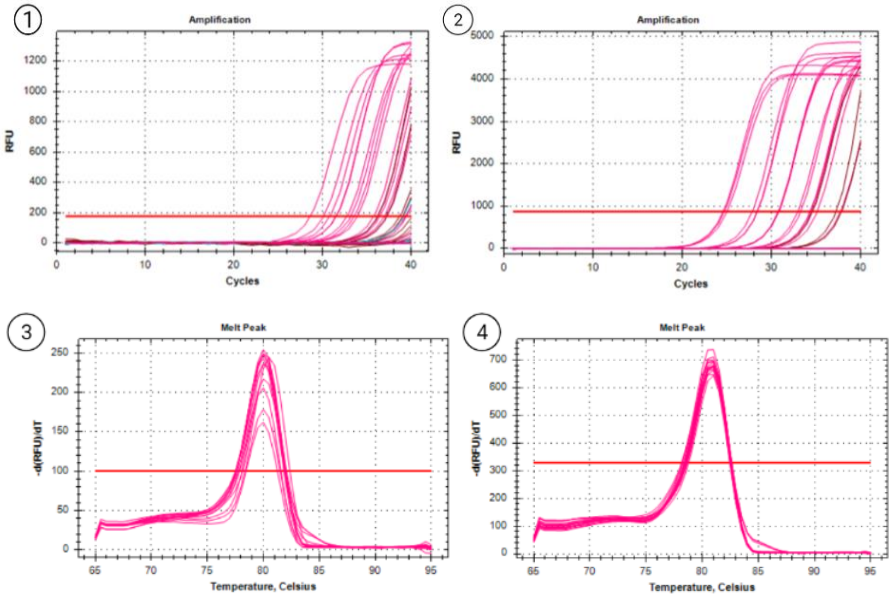

Figure D.3: Amplification and melting curves of positive control with the BlpK primer pair and an annealing temperature of 64 °C (1,3) and 60 °C (2,4). The pink curves show different dilutions of the positive control amplicon,.The threshold (red line) is set above the fluorescence of the background and was used by the software to find valid peaks.

The figure above shows that an annealing temperature of 64 °C gives higher Cq-values than an annealing temperature of 60 °C. The melting curves show that lowering the annealing temperature to 60 °C did not increase the unspecific binding. 60 °C was therefore tested on the remaining primer pairs.

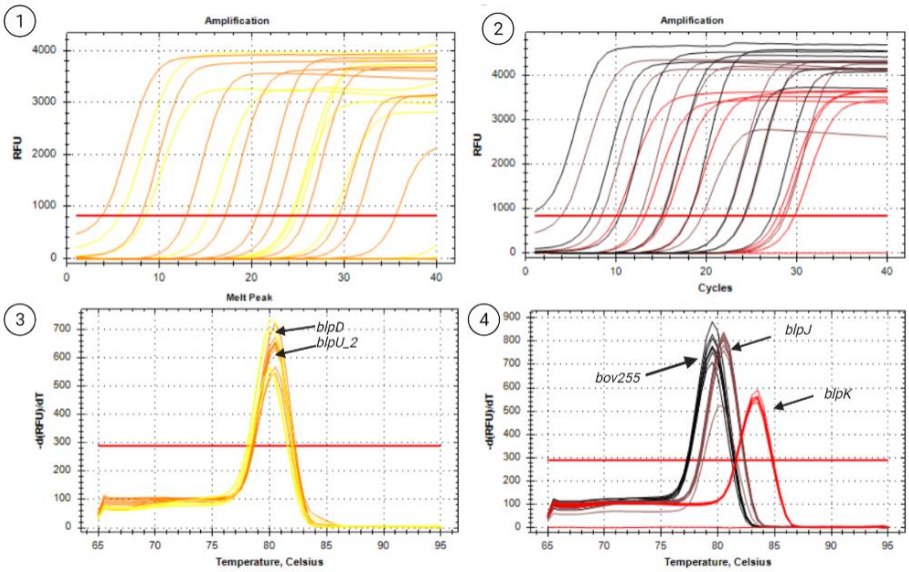

Figure D.4: Amplification (1-2) and melting curves (3-4) of positive controls with the BlpD, BlpU_2, BlpK_2, BlpJ, and Bov255 primer pair and an annealing temperature of 60 °C. The yellow curves belong to BlpD, orange to BlpU_2, red to BlpK_2, brown to BlpJ, and black to Bov255. The threshold (red line) is set above the fluorescence of the background and was used by the software to find valid peaks.

All the remaining primers gave amplification with 60 °C as annealing temperature.

Appendix D.2: Testing of primers on complex stool samples and streptococcal DNA

All the primers were tested on complex stool samples and DNA extracted from the streptococcal strains used in the bacteriocin activity assay.

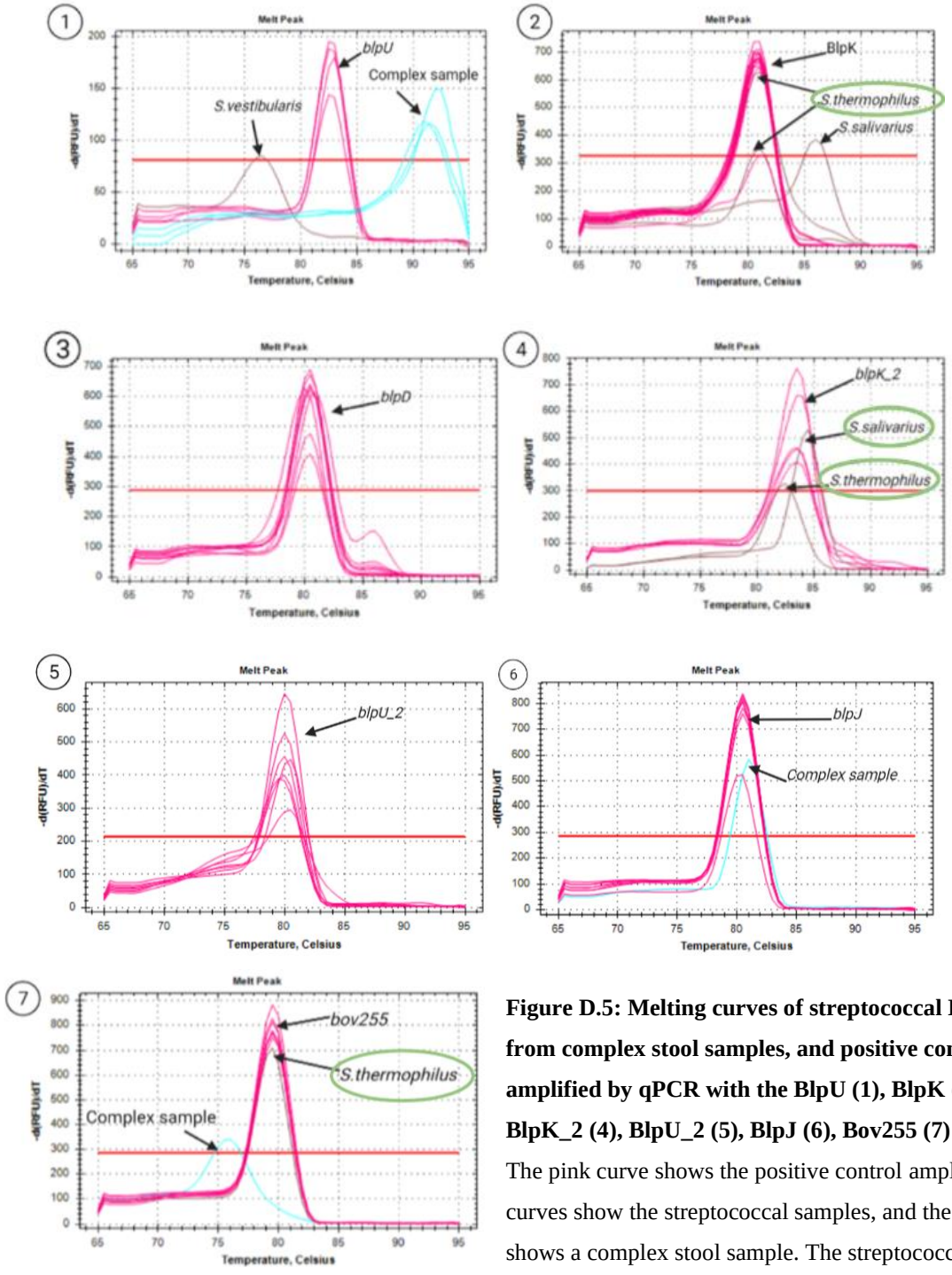

Figure D.5: Melting curves of streptococcal DNA, DNA from complex stool samples, and positive control amplicon amplified by qPCR with the BlpU (1), BlpK (2), BlpD (3), BlpK_2 (4), BlpU_2 (5), BlpJ (6), Bov255 (7) primer pair. The pink curve shows the positive control amplicon, the brown curves show the streptococcal samples, and the blue curve shows a complex stool sample. The streptococcal strains that tested positive for the gene are marked with green circles. The red line shows the threshold set by the software to find the valid peaks.

Appendix D.3: Technical issues with the gradient PCR

Estimation of the optimal annealing temperature of the designed primers was performed by using gradient PCR on samples with control DNA. The control DNA used in this experiment was the double-stranded amplicons for which these primers were designed. Figure D.1 shows that the initial gradient PCR of the BlpU primers gave results that could be used to find the optimal annealing temperature for this primer pair. After a few weeks, the Mastercycler gradient (Eppendorf, Germany) was used on the BlpK primer pair with dilutions from 10^{-1} to 10^{-7} (A-G) and the annealing temperatures are described in table D.2.

Table D.2: Annealing temperatures (T_a) used in gradient PCR. The number of the wells corresponds to the wells on the three gels.

Well	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
T_{a1} (°C)	52.9	53.2	54.0	55.2	56.8	58.7	60.6	62.5	64.3	65.8	66.9	67.5
T_{a2} (°C)	42.9	43.2	44.0	45.2	46.8	48.7	50.6	52.5	54.3	55.8	56.9	57.5

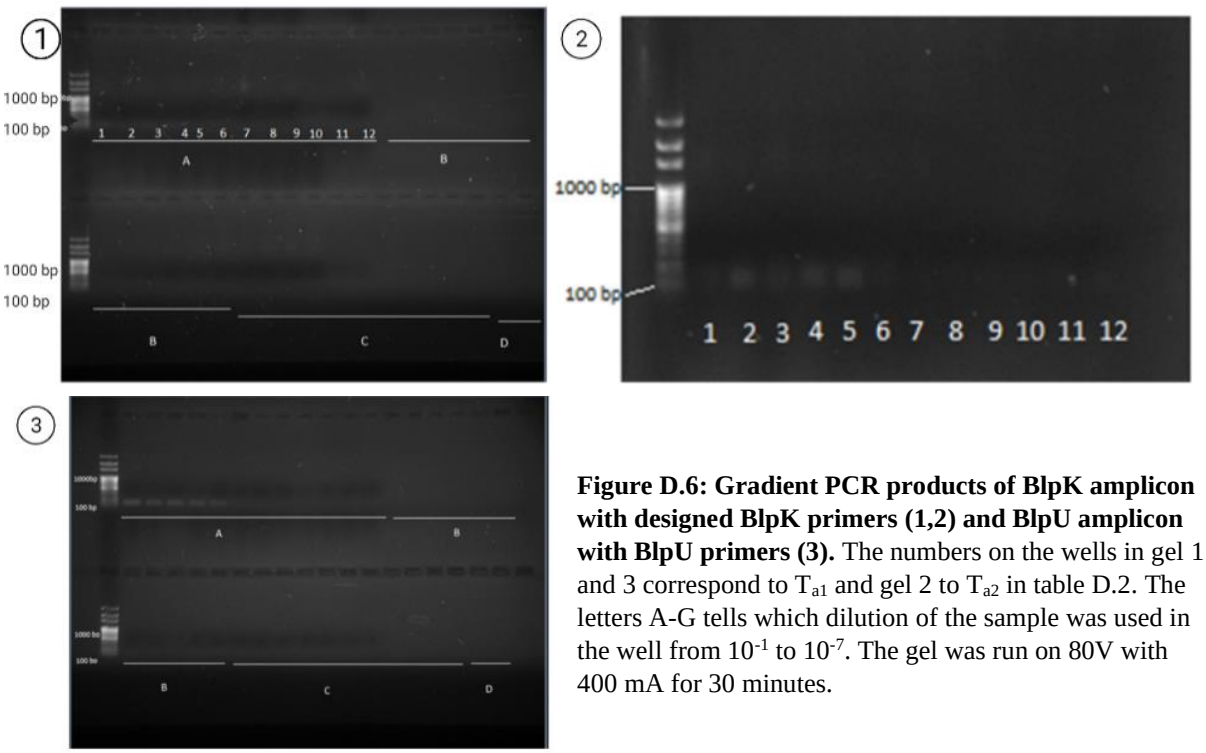

Figure D.6: Gradient PCR products of BlpK amplicon with designed BlpK primers (1,2) and BlpU amplicon with BlpU primers (3). The numbers on the wells in gel 1 and 3 correspond to T_{a1} and gel 2 to T_{a2} in table D.2. The letters A-G tells which dilution of the sample was used in the well from 10^{-1} to 10^{-7}. The gel was run on 80V with 400 mA for 30 minutes.

Figure D.6 (1) shows that it was only the highest dilution of amplicons that were amplified by using the blpK primer pair and that the primers seemed to prefer the lower end of the temperature scale. Another gradient PCR was, therefore, run on the highest dilution of the control DNA, but the annealing temperature was set to 50 ± 7 °C. Also, this reaction gave some weak bonds at the lower end of the temperature scale, but this was not expected based

on the results from figure D.6 (1). In figure D.6 (1) it was possible to identify amplification at annealing temperature 52.92-58.7 °C and it was therefore expected that the second gradient PCR would have bonds in the upper-temperature scale as well. To test if there was something wrong with the Mastercycler gradient (Eppendorf, Germany), another gradient PCR with the BlpU primer pairs was run.

The new gradient PCR run of the BlpU primers on the BlpU amplicon gave rise to amplification at the lower end of the temperature scale and only at the highest dilution of control. This did not correspond with the previous result where bonds were observed for all the dilutions and all the annealing temperatures. It was therefore decided that the Mastercycler gradient probably needed service and the annealing temperatures of the remaining primers were estimated by running three PCR reactions with different annealing temperatures: 56 °C, 60 °C, and 64 °C.

Appendix E: High resolution melting analysis of qPCR amplified samples and gel electrophoresis to confirm amplicon length

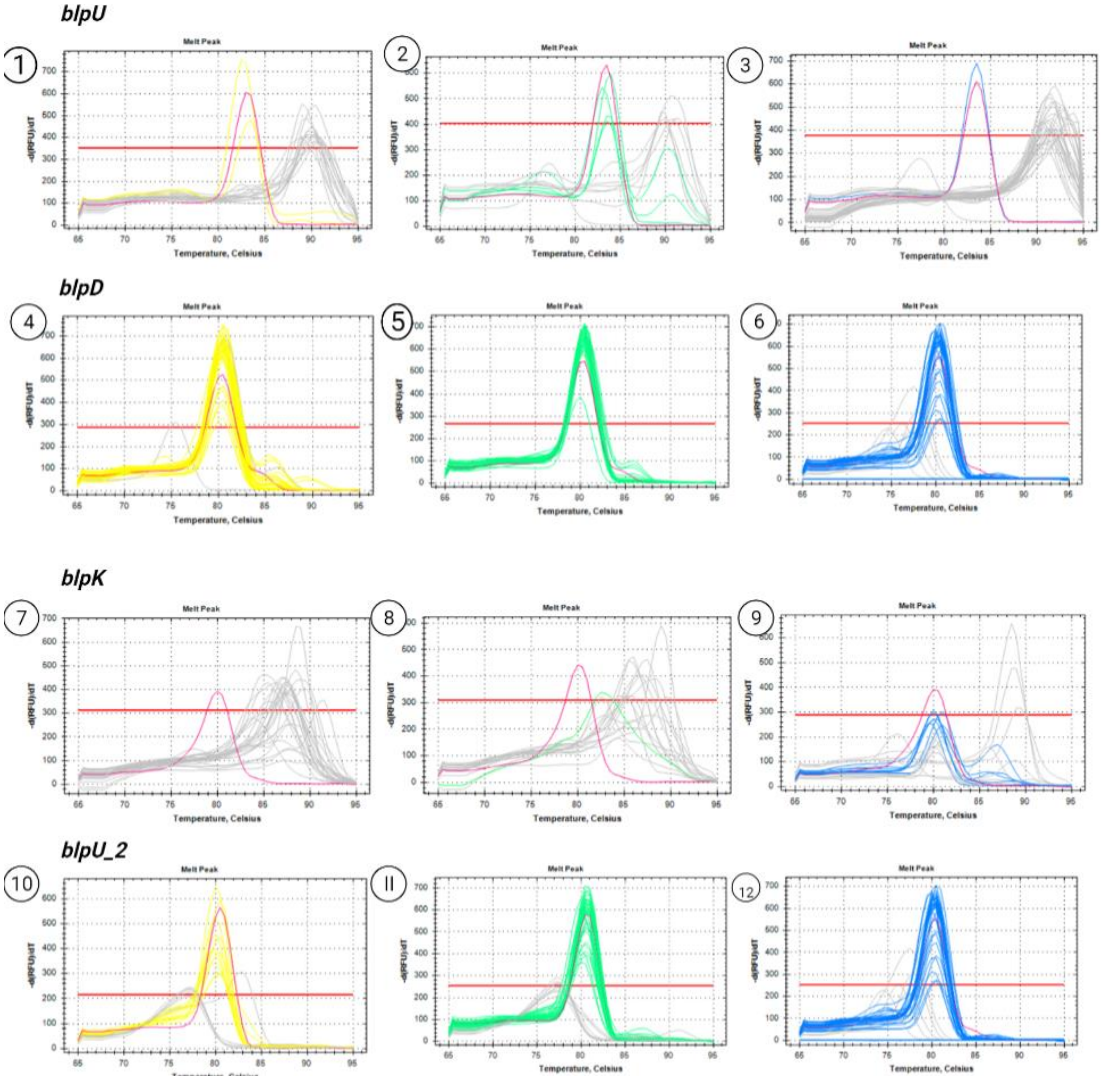

Figure E.1: The melting point graphs of samples from infants at 6 months of age (1,4,7,10), 12 months (2,5,8,11) of age, and from mothers (3,6,9,12) amplified with the BlpU, BlpD, BlpK, and BlpU_2 primer pair. The pink graph shows the positive control, while the yellow shows the positive samples from infants at 6 months, green at 12 months, and blue from mothers. The threshold (red line) was used by the software to find valid peaks.

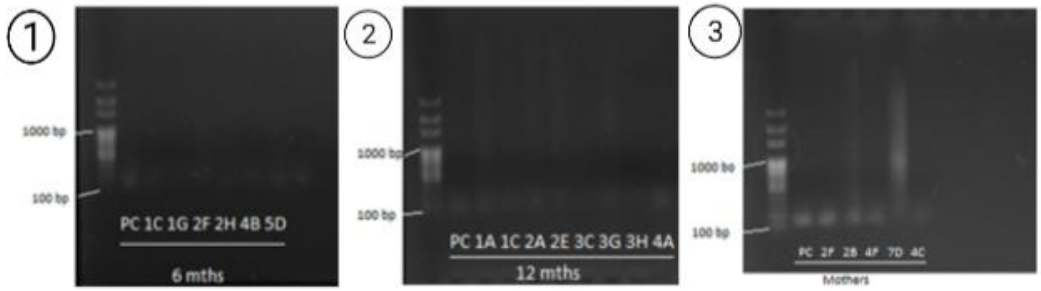

Figure E.2: Gel electrophoresis of samples taken from infants at 6 months (1), 12 months (2), and mothers (3), amplified with the BlpU primer pair. The gels were run for 30 min with 400 mA and 80 V.

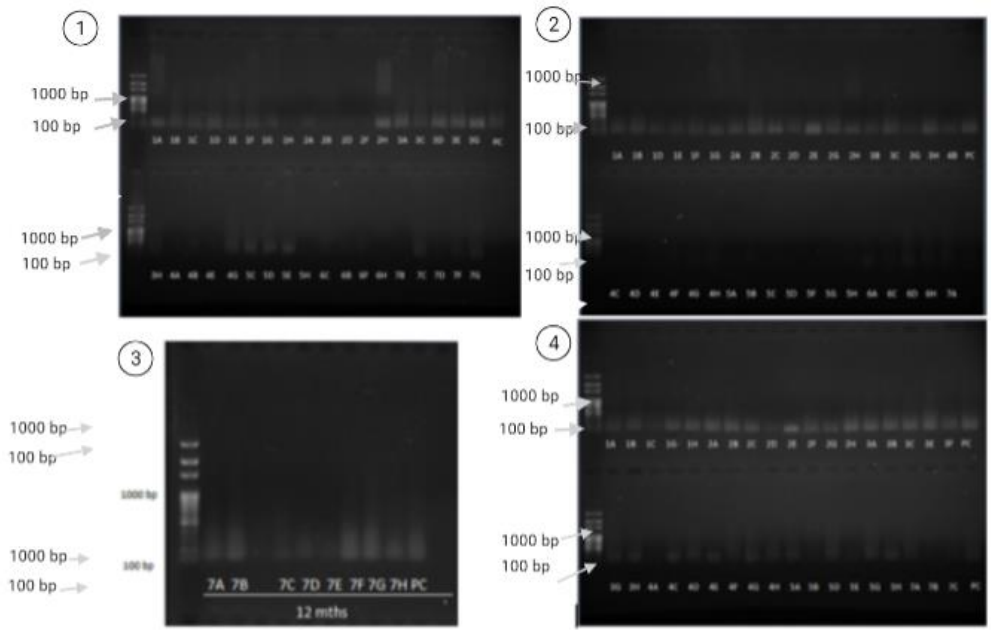

Figure E.3: Gel electrophoresis of samples taken from infants at 6 months (1), 12 months (2,3), and mothers (4), amplified with the BlpD primer pair. The gels were run for 30 min with 400 mA and 80 V.

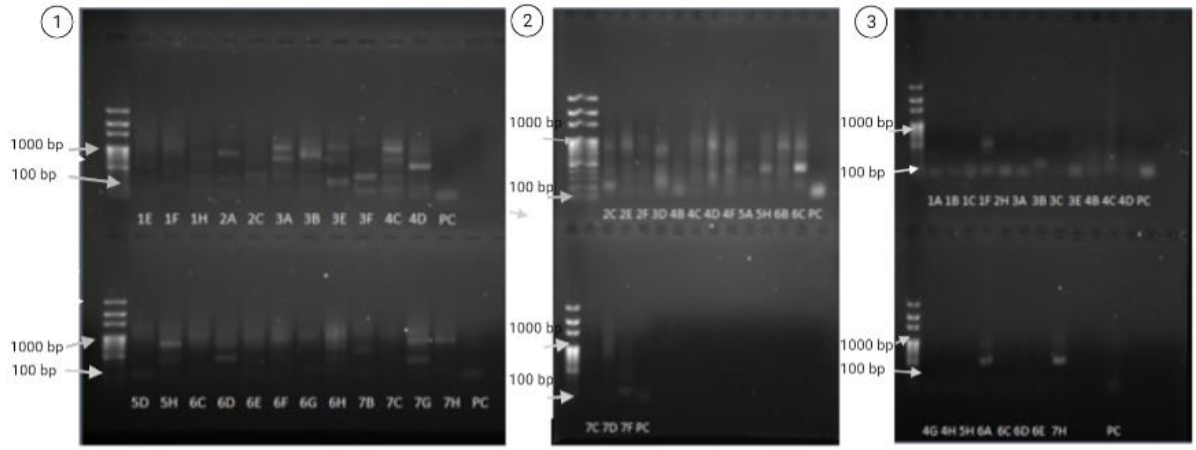

Figure E.4: Gel electrophoresis of samples taken from infants at 6 months(1), 12 months (2), and mothers (3), amplified with the BlpK primer pair. The gels were run for 30 min with 400 mA and 80 V.

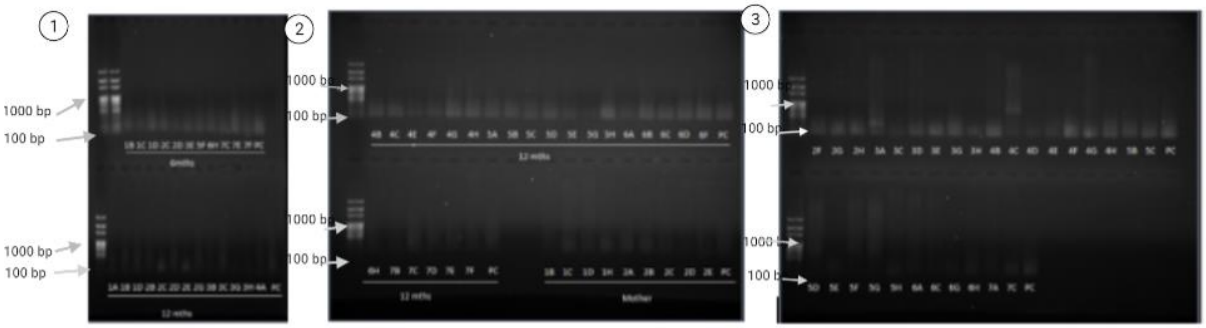

Figure E.5: Gel electrophoresis of samples taken from infants at 6 months (1), 12 months (1,2), and mothers (2,3), amplified with the BlpU_2 primer pair. The panel to the left shows the samples taken from infants at 6 and 12 months, the middle shows the samples from infants at 12 months and mothers. The right panel shows the samples taken from mothers. The gels were run for 30 min with 400 mA and 80 V.

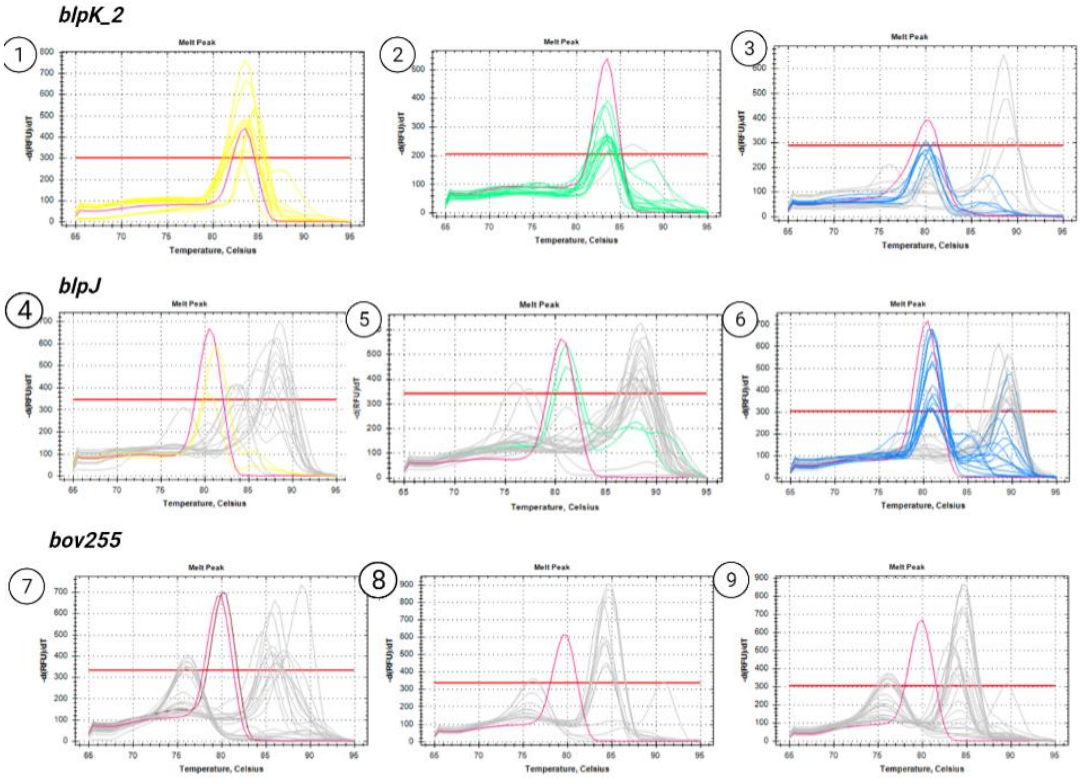

Figure E.6: The melting point graphs of samples from infants at 6 months of age (1,4,7), 12 months (2,5,8) of age, and from mothers (3,6,9) amplified with the BlpK_2, BlpJ, and Bov255 primer pair. The pink graph shows the positive control, while the yellow shows the positive samples from infants at 6 months, green at 12 months, and blue from mothers. The threshold (red line) was used by the software to find valid peaks.

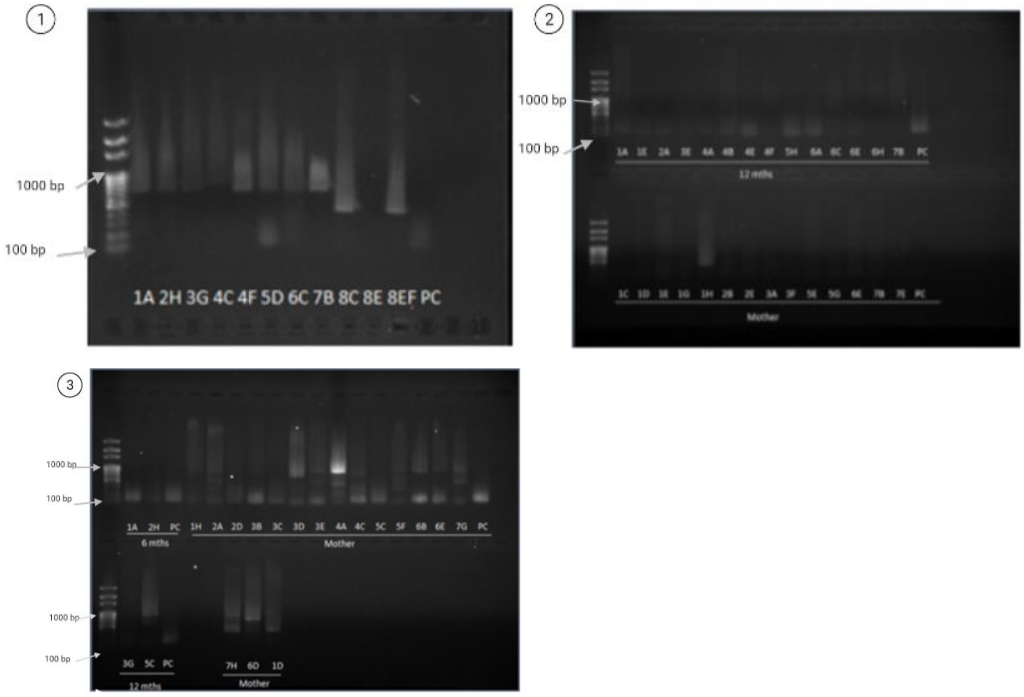

Figure E.7: Gel electrophoresis of samples taken from infants at 6 months, 12 months, and mothers, amplified with the BlpK_2 (1,2) and BlpJ (3) primer pair. The panel to the left shows the samples taken from infants at 6months, and the right shows samples from infants at 12 months and mothers. The gels were run for 30 min with 400 mA and 80 V.

Appendix F: Statistical tests

Several statistical tests were used to identify any associations between the number of streptococcal reads, the presence of bacteriocins, and the sample group. The first test looked at infants at 6 months, 12 months, and mothers in three groups and compared the number of streptococcal reads between them. The next looked at infants (6 months and 12 months) against mothers.

To find associations between streptococcal reads and the presence of bacteriocins, all the samples were divided into two groups: with or without bacteriocin. No associations were found, but a couple of outliers were observed as seen in figure F.1. Further testing was therefore performed by dividing the samples into three groups based on age and then further into the streptococcal OTUs found in the 16S rRNA sequencing. The argument with outliers was used here as well to go deeper into the data. The results of the tests are shown in table F.1.

Table F.1: Summary of the statistical tests performed to see if there were any associations between the number of streptococcal reads, age, and presence of bacteriocins. The statistical test used was Kruskal Wallis test with a significance level of 0.05. The null hypothesis was that there were no differences between the groups tested, while the alternative hypothesis was that there was a difference between the groups. The age categories refer to the samples taken from infants at 6 months, 12 months, and from their mothers.

Test for an association between	Origin of samples	p-value	Median of streptococcal reads
Age categories and number of streptococcal reds	All samples	0.1339	Infant 6 mths: 0 Infant 12 mths: 0 Mother: 0
Streptococcal reads in infants (6mths and 12mths) and mothers	All samples	0.045	With bacteriocin: 32 Without bacteriocin: 20
***More streptococcal reads in infants (6 mths and 12 mths) than in mothers**	All samples	0.0019	With bacteriocin: 32 Without bacteriocin: 20
Number of streptococcal reads and presence of bacteriocins	All samples	0.47	With bacteriocin: 31 Without bacteriocin: 25.5
Number of streptococcal reads and presence of bacteriocins	Infants 6 months	0.93	With bacteriocin: 0 Without bacteriocin: 0
Number of streptococcal reads and presence of bacteriocins	Infants 12 months	0.11	With bacteriocin: 0 Without bacteriocin: 0
Number of streptococcal reads and presence of bacteriocins	Mothers	0.97	With bacteriocin: 0 Without bacteriocin: 0
Number of streptococcal reads from OTU nr.2 and presence of bacteriocins	Infants 6 months	0.62	With bacteriocin: 18 Without bacteriocin: 20

Number of streptococcal reads from OTU nr.4 and presence of bacteriocins	Infants 6 months	0.41	With bacteriocin: 7 Without bacteriocin: 8
Number of streptococcal reads from OTU nr.1 and presence of bacteriocins	Infants 12 months	0.73	With bacteriocin: 0 Without bacteriocin: 0
Number of streptococcal reads from OTU nr.2 and presence of bacteriocins	Infants 12 months	0.019	With bacteriocin: 32 Without bacteriocin: 8
***More streptococcal reads from OTU nr.2 in samples with bacteriocin**	Infants 12 months	0.025	With bacteriocin: 32 Without bacteriocin: 8
Number of streptococcal reads from OTU nr.2 and presence of bacteriocins	Mothers	0.60	With bacteriocin: 0 Without bacteriocin: 0
Number of streptococcal reads from OTU nr. 3 and presence of bacteriocins	Mothers	0.98	With bacteriocin: 0 Without bacteriocin: 21.5
Number of streptococcal reads from OTU nr.4 and presence of bacteriocins	Mothers	0.61	With bacteriocin: 2.5 Without bacteriocin: 6.75

* These tests were performed using an unpaired, one-sided student t-test with H_0: equal or less streptococcal reads belonging to OUT nr.2 in samples with bacteriocin and H_1: more streptococcal reads in samples with bacteriocins. The significance level was set to 0.05.

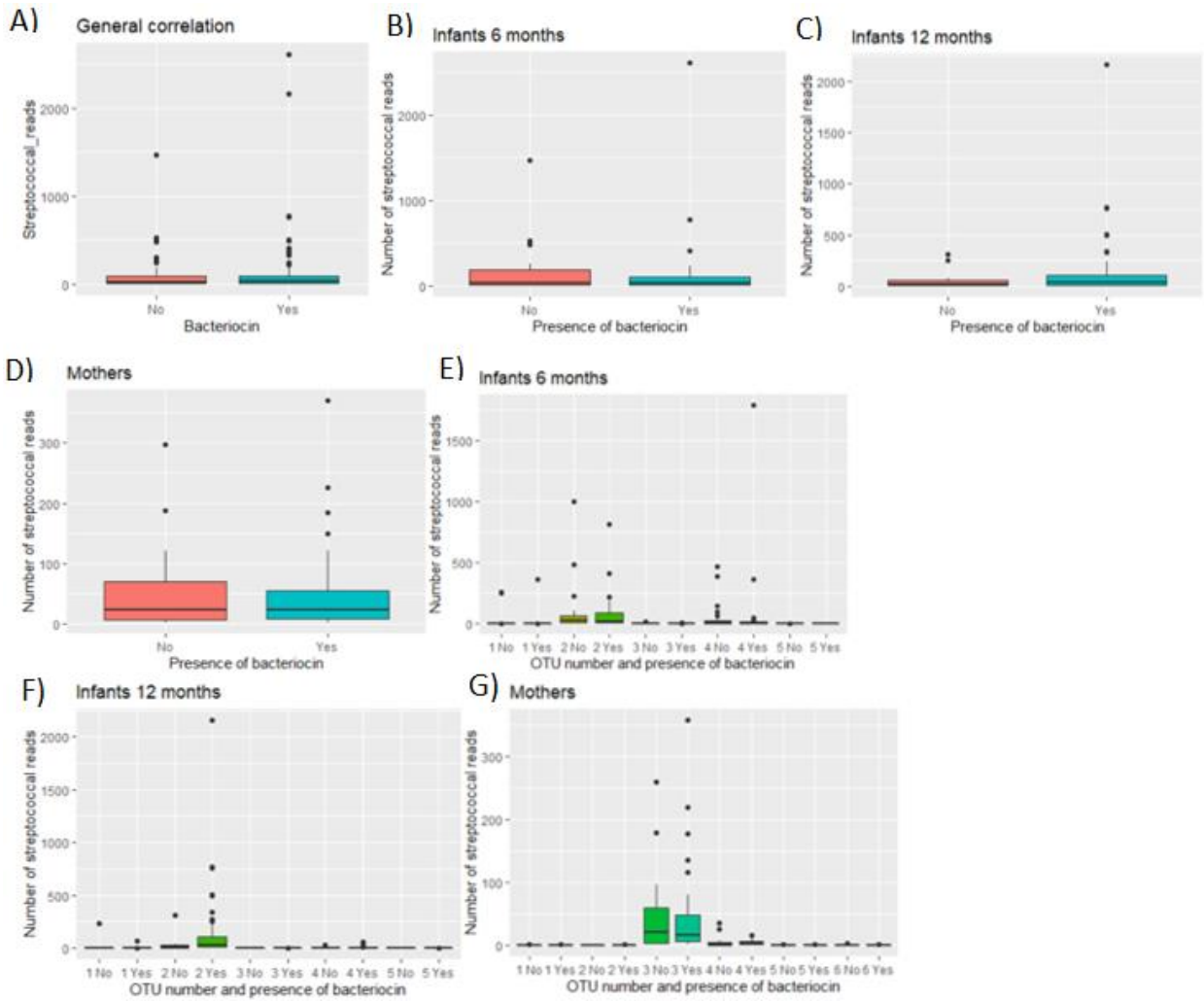

Figure F.1: Boxplot displaying the number of streptococcal reads in samples with and without bacteriocin. A) The distribution of streptococcal reads amongst all samples with or without the presence of bacteriocin. B) The distribution of streptococcal reads amongst samples from infants at 6 months of age with or without the presence of bacteriocin. C) The distribution of streptococcal reads amongst samples from infants at 12 months of age with or without the presence of bacteriocin. D) The distribution of streptococcal reads amongst samples from mothers with or without the presence of bacteriocin. E) The distribution of streptococcal reads among the different OTUs in samples from infants at 6 months with respect to the presence of bacteriocin. F) The distribution of streptococcal reads among the different OTUs in samples from infants at 12 months with respect to the presence of bacteriocin. G) The distribution of streptococcal reads among the different OTUs in samples from mothers with respect to the presence of bacteriocin.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway