Norwegian University
of Life Sciences

**Master's Thesis 2022 60 ECTS**
Faculty of Chemistry, Biotechnology and Food Science

# Cost of resistance - stress tolerance and gene expression in wild type and mutant strains of *Staphylococcus haemolyticus* LMGT4071

Kristin Kleivan Solberg
Master in Bioinformatics

# Abstract

Several studies have shown that a bacteria's development of resistance to an antimicrobial substance often comes with a cost related to their general fitness and adaptation to environmental changes. Further research is needed to increase the understanding on how bacteria with mutations in specific genes associated with stress tolerance, regulate their gene expression in response to stress.

The excessive use of antibiotics for treatment of pathogenic infections in patients have led to an increasing number of bacteria resistant to antibiotics. As it is crucial that we find alternative treatment methods, researchers are looking into the possibility of utilizing bacterial produced bacteriocins in combination with other antimicrobials. In the search for appropriate medicine, we need a broader knowledge of the bacteria's response mechanisms related to the treatments.

Through this master's thesis a mutant's loss of fitness due to a resistance mutation will be examined. The study will attempt to give new inside on how a mutant of a bacterium regulates its gene expression when exposed to stress compared to the wild type within the same strain. Specific for this study, mutants within five bacterial strains with mutations connected to the function of the stress-response related gene *rseP*, are inspected in the laboratory. The bacteria were exposed to the bacteriocin Garvicin KS which do not use RseP as a receptor. RNA from samples of *S. haemolyticus* LMGT4071 will be isolated and sequenced to analyse the gene expression in wild type and mutant, with and without exposure to GarKS.

The experiments with *S. haemolyticus* in the laboratory show that the mutant tolerates the exposure to GarKS poorer than the WT, illustrating that the development of resistance comes with a fitness cost. The transcriptome analysis show that the mutant responds to the stress by altering its gene expression to a large extent, in contrast to the WT which turns on only a few genes.

## Sammendrag

Flere studier har vist at en bakteries utvikling av resistens mot et antimikrobielt stoff ofte kommer med en kostnad knyttet til deres generelle fitness og tilpasning til miljøendringer. Det er behov for videre forskning for å øke forståelsen av hvordan bakterier med mutasjoner i spesifikke gener knyttet til stresstoleranse regulerer genuttrykket sitt i respons på stress.

Overdreven bruk av antibiotika for behandling av patogene infeksjoner hos pasienter har ført til at et økende antall bakterier er resistente mot antibiotika. Siden det er avgjørende at vi finner alternative behandlingsmetoder, ser forskerne på muligheten for å utnytte bakteriell produserte bacteriociner i kombinasjon med andre antimikrobielle midler. I jakten på passende medisin trenger vi bredere kunnskap om bakterienes responsmekanismer knyttet til behandlingene.

Gjennom denne masteroppgaven vil en mutants tap av fitness grunnet en resistensmutasjon undersøkes. Studien vil prøve å belyse hvordan en mutant av en bakterie regulerer genuttrykket sitt når det utsettes for stress sammenlignet med villtypen innenfor samme stamme. Spesifikt for denne studien inspiseres mutanter innenfor fem bakteriestammer med mutasjoner knyttet til funksjonen til det stress-responsrelaterte genet *rseP* i laboratoriet. Bakteriene ble eksponert for bacteriocinet Garvicin KS som ikke bruker RseP som reseptor. RNA fra prøver av *S. haemolyticus* LMGT4071 vil bli isolert og sekvensert for å analysere genuttrykket i villtype og mutant, med og uten eksponering av GarKS.

Forsøkene med *S. haemolyticus* i laboratoriet viser at mutanten tåler eksponeringen av GarKS dårligere enn villtypen, noe som illustrerer at utvikling av resistens kommer med en fitnesskostnad. Transkriptomanalysen viser at mutanten reagerer på stresset ved å endre genuttrykket sitt i stor grad, i motsetning til villtypen som kun slår på noen få gener.

## Acknowledgements

This master's thesis is a part of my master's degree in bioinformatics at Norwegian University of Life Sciences with Lars-Gustav Snipen as supervisor, and in collaboration with Dzung Bao Diep at Laboratory of Microbial Gene Technology.

I want to thank the group at Laboratory of Microbial Gene Technology and my supervisor, Dzung Bao Diep, for their warm welcome. Thank you Dzung Bao Diep for giving me the opportunity to gain experience in the laboratory, for giving me helpful feedback and guiding me with your positive and enthusiastic attitude. Thank you Zuzana Rosenbergova for all your guidance and for teaching me in the laboratory. Thank you Thomas Oftedal for taking time to help me with all kinds of things regarding the laboratory work.

A special thanks to my main supervisor Lars-Gustav Snipen (Bioinformatics & Applied Statistics) for all the time he has set aside to help me. Thanks for your good advice and explanations, interesting discussions, and useful feedback regarding my master's thesis. Thank you for your solid guidance, for being so patient, and giving fast response to all my questions.

These five years at Norwegian University of Life Sciences has been a challenging and educational experience, especially the work related to this master's thesis.

I want to thank my parents for always believing in me, and for all their help and support throughout my education. I also want to thank my partner and friends inside and outside the university for their support and great moments.

Thank you all for your tolerance to my stress!

# Table of Contents

# 1 INTRODUCTION

The diversity of prokaryotes is known to be large, estimated to consist of billions of species (1), and can be found everywhere on earth, and outside and inside all eukaryote organisms. Although the laboratory techniques and sequencing technology has improved a lot over the years, the majority of bacteria species are difficult to cultivate (2). Because of the huge impact bacteria have on our lives when it comes to our health, food production and other industries, it is important to broaden our knowledge and understanding of these organisms.

The history of humans preventing bacteria related diseases with antibiotics goes back thousands of years. The Eber's papyrus from 1550 BC cited in an article in COMICR (3) includes treatment of open wounds with mouldy bread. The utilization of antibiotics in a more conscious manner began in the early 20th century, with the most well-known discovery of the antibiotic penicillin by Alexander Fleming in 1928. This resulted in an enormous medical step forward in the fight against infections caused by bacteria. The increased research in this field has led to development of many antibiotic related drugs. As a consequence of the excessive use of antibiotics, and incorrect treatment of patients, an increasing number of bacteria are becoming resistant to antibiotics (3). Therefore, it is crucial that we find alternative treatment methods.

One solution can be found in studying the bacteria themselves. Bacteria lives in an ecosystem, often overcrowded with a lot of different species, causing a constant competition to gain nutrition and space. To survive and reproduce, bacteria have developed a range of weapons to combat each other. Bacteria produce a substance called bacteriocin to inhibit or kill other bacteria (4). However, as bacteria can be resistant to antibiotics, they can also develop resistance against bacteriocin.

Researchers are looking into multiple treatments as a solution to the resistance problem. Several studies have found that the bacteria's development of resistance often comes with a cost related to their general fitness and adaptation to environmental changes (5,6). For instance, one study showed that a bacterium resistant to one bacteriocin was more sensitive to change in temperature compared to the bacterium within the same bacterial strain, which did not have this resistance (5).

This master's thesis examines the divergent stress response in wild type (WT) and mutant. WT is a bacterium that appears as a "normal" phenotype related to its species. The mutant is

a bacterium somewhat different from the WT within the same bacterial strain, due to a mutation in their genome. As a starting point for this study, the mutant is believed to have a mutation that effects the function of the intramembrane protease RseP. RseP is related to the bacteria's stress response (5). The mutant is generated by selecting the bacteria that survives exposure to a bacteriocin which uses RseP as a receptor. To examine the stress response in the WT and mutant, the bacteria will be exposed to another bacteriocin which do not attack RseP.

## 1.1 Aim of study

The main aim of this thesis is to shed light on how a mutant of a bacterium regulates its gene expression when exposed to stress compared to the WT within the same strain. For this purpose, mutants of five bacterial strains were generated by Laboratory of Microbial Gene Technology (LMG): *Staphylococcus haemolyticus* LMGT4071, *Lactococcus lactis* IL 1403, *Enterococcus faecalis* LMGT3560 and two strains of *Enterococcus Faecium* (LMGT7422 and LMGT3160). This collection of mutants is presumed to have mutations in specific genes related to stress-tolerance, specifically in the *rseP* gene, making the mutant resistant to bacteriocins which uses the *rseP* gene product as receptor. To get a better understanding of how the *rseP* mutant and the WT handles stress, they are exposed to Garvicin KS (GarKS), which do not use the RseP protein as a receptor, and their gene expression is inspected.

This study consists of two parts; laboratory experiments in the wetlab and bioinformatic work in the drylab. Performing the experiments in the laboratory prior to the bioinformatic work contributes to a broader understanding on how the data is generated.

In the wetlab the sub goals are:

- To find the type of mutation in all the *rseP* mutants and to confirm whether or not the WT and mutants are resistant to the bacteriocins which uses RseP as a receptor.
- To find the sub-lethal quantity of the bacteriocin GarKS for the WT and mutant of *S. haemolyticus*, *L. lactis* and *E. faecalis*. While exposed to GarKS, the growth of these bacteria will be measured.
- To isolate the RNA from samples of *S. haemolyticus*, WT and mutant, with and without exposure to GarKS.

The drylab is about the bioinformatic work to analyse the gene expression of *S. haemolyticus* LMGT4071, WT and mutant, and the sub goals are:

- To assemble the genome of *S. haemolyticus* LMGT4071 WT and annotate the genes in this genome.
- Mapping the RNA-seq reads to the predicted gene sequences.
- To perform differential expression analysis of WT and mutant samples, with and without exposure of GarKS.
- Clustering of genes in the attempt to predict the bacteria's operon structure.

According to previous studies a mutant appears more sensitive to stress than the WT within the same strain (5,6). Hence, one hypothesis is that the mutant will not tolerate the sub-lethal stress as well as the WT. This means that the WT will be observed to grow faster than the mutant, when exposed to GarKS (in absence of the bacteriocin which the mutant is resistant to). In addition, one hypothesis regarding the transcriptomic inspection is that the differential gene expression analysis will reveal differential expressed gene sets associated with the *rseP* gene.

# 2 THEORY

This chapter will provide information of bacteria in general, the bacteria used in this master's thesis, the *rseP* and the bacteriocins that uses the RseP protein as a receptor and the bacteriocin GarKS. Theory of the methods used in this study is in section 2.4 wetlab and 2.5 drylab.

## 2.1 Bacteria

The bacteria's genetic material is organized in a structure called the nucleoid, consisting of one circular, double-stranded chromosome, located in the cytoplasm not enclosed by any membrane. As a consequence, the transcription and translation processes take place in the same cellular compartment and occur simultaneously (7). In the prokaryotic genome the adjacent genes are often grouped together in regulatory systems called operons, where the genes are transcribed together from one common promotor, resulting in one mRNA product consisting of multiple gene sequences. Consequently, it is expected that the genes in the same operon, show similar gene expression (8).

Due to the bacteria's relatively short generation time, mutations frequently occur in their DNA. A mutation is any change in the DNA sequence of an organism. As previously mentioned, the bacteria can excrete bacteriocins to kill their opponents. However, exposure to bacteriocins can lead to selected growth of bacteria with a mutation in the genome which makes it resistant to the bacteriocin (5).

In the search for the difference in stress tolerance between WT and mutant, this study examines five strains of four different gram-positive bacterial stains.

The *S. haemolyticus* species, similar to other bacteria from the staphylococci genus, exist on the human skin. *S. haemolyticus* is a pathogen that can cause infections, and is known to easily develop resistance to antibiotics, challenging the option of treatment (9). This bacterium is a member of the coagulase-negative staphylococci which was not first considered to be a pathogen (9). As *S. haemolyticus* constitute a large fraction of staphylococcal infections, this bacterium has been given more attention in pathogen research.

*L. lactis* has been part of many studies as the bacterium is used in the food industry related to fermentation of dairy products. Multiple studies address the stress tolerance of these bacteria, for instance tolerance to heat and salt concentration changes (10).

Enterococci is a part of the microbiome in the intestinal flora in animals and humans. The most common Enterococci is *E. faecalis* and *E. faecium*. The bacteria have become a concern when it comes to infections that require hospital treatment. Enterococci are known for their broad range of resistance genes which is easily exchanged between the bacterial species. Treatments with broad-spectrum antibiotics create a challenge as they can lead to bacteria with multiple resistance (11).

## 2.2 The *rseP* gene

The *rseP* gene encodes RseP (Regulator of sigma-E protease) which is an intramembrane protease. RseP is a part of the Regulated Intramembrane Proteolysis (RIP) pathway which is related to the stress response in bacteria, making it a target for antimicrobial substances (12). The active site in RseP is located in the first part of the protein and contains a conserved amino acid (aa) sequence, HExxH (x denotes any aa). It is found that this region of the RseP is important regarding sensitivity to bacteriocin (5).

The gene product of *rseP* is reported to consist of 428 aa in *S. haemolyticus* and *L. lactis* (NCBI accession version numbers: WP_046464779.1, AUS70525.1), while for *E. faecalis* and *E. faecium* RseP consists of 422 aa (NCBI Accession version number: WP_016634144.1, WP_010778111.1).

## 2.3 Bacteriocin

Bacteriocins are ribosomal synthesized antimicrobial peptides that is produced by bacteria with the purpose to outcompete close related or unrelated bacterial species. The majority of bacteria species are found to produce at least one type of bacteriocin. Bacteriocins have a diversity in terms of structure and function, and have become important in research related to treatment of pathogenic bacteria (13).

The Enteriocin K1 (EntK1), Enteriocin EJ97 (EntEJ97) and LsbB are a part of a group of leaderless baceriocins which is known to use RseP as a receptor. LsbB is produced by *L. lactis* and kills bacteria within the lactococci genus. EntK1 and EntEJ97 are produced by Enterococci and has a broader range of inhibition than LsbB (5). EntK1 EJ97 is a fusion of EntK1 and Ent EJ97, where the N-terminal part of the peptide is from EntK1, and the C-terminal part form Ent EJ97. This bacteriocin is reported to have an antimicrobial effect on *S. haemolyticus* (14).

GarKS is a bacteriocin produced by the bacterial strain *Lactococcus garvieae* KS1546. This bacteriocin is reported to have a broad inhibition spectrum with inhabitation of both gram positive and gram negative bacteria. GarKS is a multi-peptide leaderless bacteriocin consisting of three peptides (15).

## 2.4 Theory - Wetlab methods

Polymerase chain reactions

Polymerase chain reaction (PCR) is a method for amplifying a sequence of DNA or RNA and is useful for a broad range of applications associated with genomic analysis. In the PCR machine the process for amplifying a DNA sequence runs as follows: the sample with DNA is heated up to denature the doble stranded DNA, breaking the hydrogen bonds and generating single stranded DNA. The sample is then cooled down, allowing the annealing phase where the added primers bind to each single strand. The temperature is raised in the elongation phase as a thermostable polymerase catalyses the DNA synthesis from the primers, adding deoxynucleotide triphosphate (dNTPs) from the five prime (5')-end to the three prime (3')-end (16). This process is repeated several times depending on the desired number of copies.

Soft-agar overlay assay

The Soft-agar overlay technique is a semiquantitative method common for studying microorganisms' growth and response when exposed to stress, for instance antibiotics or antimicrobial peptides such as bacteriocins (17). For preparing this experiment a petri dish is first filled with agar and left to solidify. The bacteria culture is diluted and mixed with soft-agar (preheated and maintained at 55-60°C) and poured into the petri dish, creating an even distribution of the cells. The selected antimicrobial substances are applied on the petri dish,

either by adding a solution or a disc containing the substance. Once the bacteria are incubated long enough to show growth, any inhibition of the growth can be seen as clearing zones (17). The size and intensity of the clearing zone reflects the strength of the inhibition.

Minimum inhibitory concentration

Minimum inhibitory concentration (MIC) describes the lowest concentration of an antimicrobial needed to inhibit bacterial growth after incubation overnight (ON) (18). This is a quantitative method, in contrast to the Soft-agar overlay assay. The microtiter plate used in the MIC-test contains 96 wells which can be utilized to test a higher number of concentrations of an antimicrobial.

In the microtiter plate the wells in a row is added decreasing concentrations of one antimicrobial. The MIC value is calculated based on the well with the lowest concentration and no bacterial growth after incubation of the microtiter plates. If all wells in a row show bacterial growth, the MIC value is expected to be higher than the concentration of the bacteriocin in the first well. If there is no bacterial growth in any of the wells, the MIC value is expected to be lower than the concentrations tested in the dilution series.

Measuring of growth rate

Measuring the growth of bacteria in liquid medium over a given time is a common method for studying various growth rates. The bacterial growth curve is a characteristic pattern of growth across different bacterial species and can be obtained when a small number of bacteria cells are inoculated in a fresh liquid medium and measured regularly over a period of time (19).

The growth curve can be divided into four phases based on the growth rate, which reflects how the bacteria behave: the lag phase, the log phase, the stationary phase and cell death. After inoculation the bacteria cells are adapting to the new environment in the lag phase. This is observed as slow growth in the growth curve. The log phase, also called the exponential phase, is seen as exponential growth of the bacteria. In this phase the bacteria divide at optimal growth rate. The bacterial growth stagnates in the stationary phase were both cell death and cell division occur, balancing the number of cells. As the number of cells surpasses the available amount of nutrition and toxic components build up, the cells die. Finally, the

phase of cell death occurs when more cells are dying than dividing. This appears as a decreasing line in the growth curve (19).

DNA and RNA isolation

Before sequencing either DNA or RNA, the genetic material must be isolated from the other cell components. The main goal is to optimize the yield, purity, integrity and quality of the molecules. The isolation process, whether it is column-based or solution-based, requires lysis. In the lysis the cell wall breaks down and the cell components are released. DNase (DNA isolation) or RNase (RNA isolation) must be inactivated, and the genetic material must be separation from proteins and other cell components. In addition, it is important to work effectively and precise, and to avoid contamination of the samples This is especially true for isolating RNA, which is a less stable molecule than DNA (20).

## 2.5 Theory - Drylab methods

Read coverage and K-mer coverage

The read coverage, or the sequencing depth, is the average number of reads that cover one position in the genome. The K-mer coverage is somewhat more complicated; a K-mer is a sub-sequence with length k, starting from each position in a sequence except position L – k + 1, where "L" is the sequence length. For instance, if k = 3 the sequence "ATGCA" have three overlapping K-mers: "**ATG**", "TGC" and "**GCA**". In this example the canonical K-mers are in bold letters. A canonical K-mer includes a K-mer and its reverse-compliment. For this sequence, the reverse-complement is "TGCAT" and its K-mers is "TGC", "**GCA**" and "CAT". Whether it is the K-mer or its reverse-compliment that describes the canonical K-mer, is decided based on alphabetical order. K-mer coverage is the average number of canonical K-mers starting at a position in the genome (based on information from Lars-Gustav Snipen's lecture on K-mers in BIN310, NMBU).

$$\textbf{Read coverage} = \frac{\textbf{read length x number of reads}}{\textbf{number of bp in the genome}}$$

$$\textbf{K-mer coverage} = \textbf{read coverage x } \frac{\textbf{read length – k + 1}}{\textbf{read length}}$$

Genome assembly

Genome assembly is the process where sequenced DNA reads are put together in an attempt to achieve longer sequences, called contigs. In case there is already an assembly of a similar genome, the assembly can be simplified by mapping the reads to this existing genome. However, if the genome has not been assembled before, the assembly must be built from scratch (de novo assembly).

Sequencing technologies, like Illumina, provides short reads, up to 300 base pairs (bp). The limitation of read length is due to less compliance in the fluorescent signal, as the sequencing of equal fragments get out of sync (21). To overcome the challenges related to the computational problem that comes with assembling short reads, many assembly algorithms utilize the De Bruijn graph (DBG) approach (22). DBG in genome assembly use K-mer information to merge reads. The DBG is a directed graph consisting of nodes connected by edges. There is one node for each unique K-mer in all reads, and one edge between two nodes observed to follow each other. This means that there can be several edges from one node, and in case two nodes are observed to follow each other several times, this information is stored. When the DBG is built, the order of the K-mers is detected by walking through each edge once (Euclidean walk) (23).

Gene annotation

For transcription studies it is important to uncover the gene content of the genome. Several software tools are developed for predicting genes, by detecting the gene sequence and assign it to a known gene. Gene sequences can be detected either by matching the sequence to some database (24) or by searching for specific patterns in the DNA, that are known to be connected to genes.

The diversity of bacterial genomes makes the gene annotation problematic. Non-coding genes, which encodes for instance rRNA and tRNA species, are highly conserved across organisms (25). Therefore, the non-coding genes are often easily predicted because similar sequences are already known. However, the content and structure of coding genes in different bacteria varies.

Pan genome analysis can be used to compare the gene content of bacteria within a species. The pan-genome describes a species by its core genes (genes common between the strains

within a species), genes shared between almost all strains and genes which are specific to one or few of the strains (26).

Read-mapping by pseudoalignment

Read-mapping is the process where sequenced reads, DNA or RNA, are matched or aligned to a reference sequence. In transcriptomics, RNA reads are mapped to gene sequences to obtain information of the number of RNA reads derived from each gene's transcript. This is used to obtain the read count for each gene for further analysis of the gene expression (27).

Pseudoalignment is K-mer based and uses the DBG for read-mapping. With this algorithm the full length of the reads is not aligned, which is often time consuming. In the DBG, a node represents a K-mer from one or several transcripts (gene sequences) that contain the same K-mer. The K-mers for each read is matched with the K-mers from the transcripts. The reads are then matched to one of the transcripts (or several transcripts) based on the order and number of matching K-mer nodes (28).

Differential gene expression analysis

Differential gene expression analysis can be used to find genes that are significantly upregulated and downregulated in one experimental group compared to another, often utilized to inspect the gene expression of bacteria under different treatments or conditions (29).

Although it is favourable in experimental designs to have multiple biological replicates, the cost sets a limit to the potential number of replicates. It is recommended to use at least three biological replicates to be able to measure the biological variability and to get a more robust statistical test related to the differential expression analysis (30).

Another important element in differential expression analysis is that the read counts must be normalized so that the samples are comparable. The number of reads mapped to each gene will not only depend on the actual expression of the gene, but also on the expression of other genes, the sequencing depth and the length of the transcribed fragments. The goal of normalization is to remove systematic variation and reduce noise in the data as much as possible in order to capture the biological difference (31).

Operons prediction

The beginning of the theory chapter addresses organization of bacterial genes and describes how these are organized into operons. As the genes in the same operon are transcribed together, their gene expression is expected to be similar, which means that all the genes are equally upregulated or downregulated. Genes in the same operon are often found to be functionally related or connected to a common pathway (8).

Predicting which genes that belong to an operon is not strait forward. One study has attempted to cluster the genes based on for instance the distance between adjacent genes and functional relationships (32). Predicting operons based on pair-end read information has also been attempted. Using pair-end RNA sequencing, some of the sequenced fragments will consist of the end of one gene and start of the next in an operon. Aligning the reads directly to the gene sequences should result in some instances where a read pair maps to two different genes. Two studies utilized this information, and clustered genes together into operons when read pairs mapped to two different genes more than the selected threshold (33-34).

# 3 MATERIALS AND METHODS

## 3.1 Wetlab

In this chapter the approach and methods used in the laboratory will be presented. PCR for sequencing of the *rseP* gene, Soft-agar overlay assay and MIC-test were performed for all five bacteria strains. The growth rate of *S. haemolyticus*, *L. lactis* and *E. faecalis* while exposed to GarKS was measured. Additional growth rate measurements and isolation of RNA were performed for *S. haemolyticus*.

### 3.1.1 The bacteria strains; storage, medium and cultivation

Frozen stocks of WT and mutant strains of *S. haemolyticus* LMGT4071, *L. lactis* IL 1403, *E. faecalis* LMGT3560, *E. faecium* LMGT7422 and *E. faecium* LMGT3160 were stored at - 80°C. All strains from the frozen stocks were streaked out on agar plates, of either Brain Heart Infusion (BHI) (OXOID CM1135) or GM17 medium. The plates were incubated overnight in either 30°C or 37°C before stored at 4°C. Unless something else is stated, cells from single colonies on the plates were collected to make overnight (ON) cultures in 5 mL medium for all experiments. Apart from DNA and RNA isolations, all work with and without bacteria was done in a sterile hood.

### 3.1.2 Polymerase chain reaction and sequencing of the *rseP* gene

To confirm if the frozen stocks contained the expected bacterial strains, and to find the type of mutations in the *rseP* gene for each strain, polymerase chain reactions (PCR) was preformed to amplify the gene sequence containing the *rseP* gene before isolating the DNA. For the PCR reactions, cells were collected directly from the frozen stocks.

For the first PCR, only samples for WT and mutant of *S. haemolyticus*, *L. lactis* and *E. faecalis* were prepared. A sterile toothpick was used to collect cells from the frozen stocks into the PCR tubes. The tubes were placed in the microwave oven for 90 seconds to break down the cell membrane (lysis). To avoid direct exposure of radiation, the tubes were placed on one side in the oven. 2.5 µl of forward and reverse primer with a concentration of 10 µmol/L were added to the corresponding PCR tubes. Information about the primers used in the PCR reaction is found in the appendix, **Table A.1**. For the six reactions, master mix for seven reactions was prepared, see **Table 3.1.2.1** below. Q5 High-Fidelity DNA polymerase

was used for all reactions. To avoid early reaction, the polymerase was added at the end. The protocol PCR Using Q5® High-Fidelity DNA Polymerase (M0491) (New England BioLabs (NEB)) was used.

**Table 3.1.2.1** Components of the master mix

| Component | Volume 7 reactions | Volume 5 reactions | Final concentration |
|---|---|---|---|
| 5x Q5 Reaction Buffer | 70 µl | 50 µl | 1x |
| 10 mM dNTP | 7 µl | 5 µl | 0.2 mM |
| Q5 High-Fidelity DNA polymerase | 3.5 µl | 2.5 µl | 0.02 U/µl |
| $dH_2O$ | 234.5 µl | 167.5 µl | |
| 10 µmol/L Forward primer | - | 12.5 µl | 0.5 µmol/L |
| 10 µmol/L Reverse primer | - | 12.5 µl | 0.5 µmol/L |

45 µl of the master mix was added to each PCR tube containing the DNA and primers, quickly mixed by pipetting and placed on ice before they were transferred to the PCR machine. Settings for the different steps of the PCR reaction is listed in **Table 3.1.2.2**.

**Table 3.1.2.2** Termocycling Conditions; settings for running PCR

| Steps | | Temperature | Time |
|---|---|---|---|
| Initial Denaturation | | 98.0°C | 30 seconds |
| 35x Cycles | Denaturation | 98.0°C | 10 seconds |
| | Annealing | 63.0°C | 30 seconds |
| | Elongation | 72.0°C | 1 minute |
| Final Extension | | 72.0°C | 2 minutes |
| Hold | | 4.0°C | |

For the second PCR, with the WT and mutant strains of *E. faecium* LMGT7422 and *E. faecium* LMGT3160, a master mix was made for five reactions with the same forward and reverse primers (see **Table 3.1.2.1**). 50 µl of the master mix was added to each sample of DNA. Besides the annealing temperature, which was 57.0°C, the same settings for PCR in the **Table 3.1.2.2** were used.

To confirm that the amplified fragments corresponded to the expected lengths, the PCR products were run on a 1% agarose gel with 1xTris-acetate-EDTA (TAE) buffer. To make the gel, 50 mL 1xTAE-buffer was added to 0.5 g agarose (UltraPure agarose; Invitrogen), heated in the microwave to dissolve, and cooled down to about 60°C before adding 2 µl peqGREEN DNA/RNA dye (peqlab). For each sample, 5 µl was mixed with 1 µl Gel Loading Dye Purple (6x) (NEB) before applying 4 µl to the well. The gel electrophoresis was run with 75 Volt for the first six samples and 90 Volt for the last four.

The DNA from the rest of the PCR products were isolated using NucleoSpin® Gel and PCR clean-up kit and protocol from Macherey-Nagel. The samples with isolated DNA were prepared for sequencing. Forward and reverse primers was diluted in 2.5 µl $dH_2O$ to get a concentration of the primers equal to 5 µmol/L. If the concentration of the template was 20-120 ng/µl, 5 µl of the template were added to the primer. For templates with higher concentration 2.5 µl of the template and 2.5 µl $dH_2O$ were added. A total number of 24 samples containing 10 µl compound were sent to Eurofins genomics for sequencing (see **Table 3.1.2.3**).

**Table 3.1.2.3** Primer names and sequencing labels for the samples sent for sequencing. All labels start with the initials "ID". The names of forward and reverse primers are above the sequencing labels of each bacteria species, forward primers in the *Primer 1* column and reverse primer in *Primer 2* column. For *E. faecium* an additional forward primer was used (*Primer 3*).

| Bacterial strain | Type | Primer 1 | Primer 2 | Primer 3 |
|---|---|---|---|---|
| | | ShrRseP_F | ShrRseP_R | |
| *S. haemolyticus* LMGT4071 | WT | ID DMX071 | ID DMX079 | |
| | mutant | ID DMX070 | ID DMX078 | |
| | | yvjB_F | yvjB_R | |
| *L. lactis* IL1403 | WT | ID DMX083 | ID DMX069 | |
| | mutant | ID DMX082 | ID DMX074 | |
| | | OE_pair2_F | OE_pair2_R | |
| *E. faecalis* LMGT3560 | WT | ID DMX073 | ID DMX086 | |
| | mutant | ID DMX072 | ID DMX085 | |
| | | Ing_rseP_fm_OFA_B | Ing_resP_fm_ORA_B | Ing_rseP_fm_MF_B |
| *E. faecium* LMGT7422 | WT | ID DMX077 | ID DMX084 | ID DMX092 |
| | mutant | ID DMX076 | ID DMX089 | ID DMX091 |
| *E. faecium* LMGT3160 | WT | ID DMX075 | ID DMX088 | ID DMX090 |
| | mutant | ID DMX080 | ID DMX087 | ID DMX081 |

To detect the mutations in the *rseP* gene of the mutants, the sequenced gene fragments were analysed by using the SnapGene® software (SnapGene® Viewer 5.3.2 from Insightful

Science; available at snapgene.com), Unipro UGENE (35), and CAP3 Sequence Assembly Program from PRABI (36). With SnapGene, the start and the end of each WT sequence was trimmed to only keep the bases with high quality. The sequences from forward and reverse reads were assembled by the CAP3 Sequence Assembly Program and the contigs were downloaded. The reverse compliment of the sequence was copied for the fragments derived from samples with reverse primers. In UGENE, the mutation was found by comparing the gene sequence of the mutant with the WT; mutant reads were mapped to the assembled WT contig.

### 3.1.3 Soft-agar overlay assay

With the purpose of confirming whether the mutants and WT are resistant to the various RseP targeting bacteriocins, Soft-agar overlay assay was performed for WT and mutant of the five strains; *S. haemolyticus* LMGT4071, *L. lactis* IL 1403, *E. faecalis* LMGT3560, *E. faecium* LMGT7422 and *E. faecium* LMGT3160. For each of these bacteria, 50 µl ON culture was added to 5 mL preheated liquid BHI soft agar, mixed by vortexing and poured in separate petri dishes containing BHI agar. When solidified, 3 µl with three different concentrations of the bacteriocins LsbB, EntK1, EntEJ97, EntK1 EJ97 and GarKS were applied to each agar plate: 1 mg/mL, 0.1 mg/mL and 0.01 mg/mL. All plates were incubated at 37°C ON.

### 3.1.4 Minimum inhibitory concentration

To find the lowest concentration of the bacteriocins that prevent growth of the bacteria, and to confirm the results from the Soft-agar overlay assay, the Minimum inhibitory concentration method was used. To find the MIC value for all selected combinations of the various bacteriocins and bacteria strains, WT and mutant, microtiter plates with 96 wells were used. Combination of bacteriocins and bacteria strains are listed in **Table 3.1.4.1**.

**Table 3.1.4.1.** Overview of the bacterial strains exposed to different bacteriocins in the MIC test. A plus sign denotes the exposure of bacteriocin.

| | **Bacteriocin** | | | | |
|---|---|---|---|---|---|
| **Bacterial strain** | **LsbB** | **EntK1** | **EntEJ97** | **EntK1 EJ97** | **GarKS** |
| *S. haemolyticus* LMGT4071 | | | | + | + |
| *L. lactis* IL1403 | + | + | + | + | + |
| *E. faecalis* LMGT3560 | | | + | + | + |
| *E. faecium* LMGT3160 | | + | + | + | + |
| *E. faecium* LMGT7422 | | + | + | + | + |

For this first MIC-test, the microtiter plates were used with 12 rows and 8 columns. 100 µl of BHI medium was added to each well in the plates. Because there were two different BHI medium batches utilised in the experiment, one row of wells for each batch was additionally added 100 µl BHI medium and served as negative controls. As positive controls, two rows for each strain (WT and mutant) were left without added bacteriocin. The first well in every other row, not used as controls, was first added 10 µl of one bacteriocin and 90 µl BHI medium. To make a dilution series through the wells in the rows, a multichannel pipette was used to mix and extract 100 µl from the first wells, transferring this volume to the next well, and so on (see illustration in **Figure 3.1.4.1**). Hence, from one well to the next, the concentration of bacteriocin is diluted 2-fold.

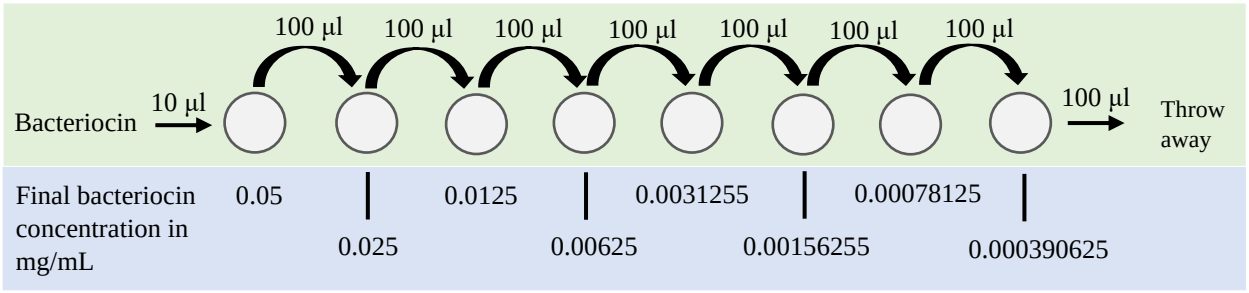


**Figure 3.1.4.1.** Illustration of the dilution series

ON cultures of the bacteria strains were diluted 50 times, before 100 µl of the diluted cultures were added to each well in the rows marked with the different bacteria, resulting in all wells containing 200 µl altogether. Each of the bacteriocins had a start concentration of 1 mg/ml, therefore the concentration in the first well was 0.05 mg/ml.

The MIC value was calculated based on the well with the lowest concentration and no bacterial growth after incubating the microtiter plates at 37°C ON. In addition to visual observation of possible growth on the microtiter plates, the plates were measured with OD600 on a microplate spectrophotometer (SPECTROstar Nano, BMG LABTECH) after four and 24 hours.

### 3.1.5 Additional MIC-tests and measuring bacterial growth in microtiter plates for *S. haemolyticus*, *L. lactis* and *E. faecalis*

With the intention of determine if there is a difference in how efficiently the wild type and the mutant of *S. haemolyticus*, *L. lactis* and *E. faecalis* handles sub-lethal amount of GarKS,

several MIC tests were performed, followed by measuring the growth of the bacteria in different quantities of GarKS in the microtiter plates. In addition, a higher temperature was added as a stress factor together with bacteriocin to examine the effect of exposing the bacteria to two stress factors simultaneously.

To avoid unnecessary sources of error, a new batch of 4 mg/mL GarKS, enough for all remaining experiments, was made and stored at -4°C. The batch was made by adding $dH_2O$ to the three peptides, GakA, GakB and GakC. The peptides were serialized separately by syringe filters before they were combined. The ON cultures were inoculated close to 15.00 hours the day before use. GM17 medium was used when making ON cultures, in the MIC-tests and measuring of growth of *L. lactis* and *E. faecalis*, while BHI medium was used for *S. haemolyticus*. The ON cultures were incubated at the optimal temperature for the bacteria (*S. haemolyticus* and *E. faecalis*: 37°C, *L. lactis*: 30 °C) (37-39)). The ON bacteria cultures were diluted 51 times before transferring 100 µl to the wells in the microtiter plate. For each of the experiments, one microtiter plate was used per bacteria strain. In that way, several rows could be utilized to obtain replicates of the samples. A summary of all these completed experiments is given below.

*S. haemolyticus*: All MIC tests and measuring of the growth rate performed with [GarKS] = 4 mg/mL

- Four MIC tests with optimal temperature
- One measuring of growth rate, optimal temperature
- One measuring of growth rate, 42 °C

*L. lactis*:

- One MIC test, optimal temperature and [GarKS] = 4 mg/mL
- Three MIC tests, optimal temperature and [GarKS] = 0.1 mg/mL
- Two MIC tests, 42°C and [GarKS] = 0.1 mg/mL
- One MIC test, 37°C and [GarKS] = 0.1 mg/mL
- One measuring growth rate, optimal temperature, [GarKS] = 0.1 mg/mL
- One measuring growth rate, 37°C, [GarKS] = 0.1 mg/mL

*E. faecalis*:

- One MIC test, optimal temperature and [GarKS] = 4 mg/mL
- One measuring growth rate, optimal temperature and [GarKS] = 0.25 mg/mL
- One MIC test, 42°C and [GarKS] = 4 mg/mL
- One measuring growth rate, 42°C and [GarKS] = 1 mg/mL

In general, the 12th column of the microtiter plate was only added medium and used as negative control. The 11th column was used as positive control and therefore only containing medium and bacteria. This leaves 10 columns with dilutions series of GarKS. It was desirable to get the concentration of GarKS, corresponding to the MIC value, not far from column five. This to achieve observation of the growth rate for the bacteria exposed to the amount of GarKS less and higher than the MIC value.

The growth rate was determined by measuring the OD600 on a microplate spectrophotometer (SPECTROstar Nano, BMG LABTECH) every 15 minutes, for approximately 24 hours (97 circles).

### 3.1.6 Growth rate of WT and mutant strains of *S. haemolyticus* while exposed to stress in culture tubes

To obtain a sufficient amount of RNA from the *S. haemolyticus* samples, it is required to isolate RNA from a higher number of bacteria cells. A change in medium volume can affect the growth rate measured on the microtiter plates (40). Thus, several concentrations of GarKS and exposure to higher temperatures were additionally tested for WT and mutant of *S. haemolyticus* in a total volume of 5.1 mL. To be sure that the growth patterns are reproducible, the growth of the bacteria when exposed to the selected concentration was repeated several times before the RNA isolation.

ON cultures of *S. haemolyticus*, WT and mutant, were vortexed before 100 µl of each culture were added to different culture tubes containing 5 mL BHI medium. The culture tubes were mixed by vortexing once more before the top of the culture tubes were flame sterilized. The samples were incubated at 37°C. To monitor the cell growth, the spectrophotometer Ultrospec 10 Cell Density Meter (AMERSHAM BIOSCIENCES) was used to measure OD600 each 30-45 minutes. When the samples reached exponential phase, an OD600 around 0.15, the stress was induced for all samples not serving as controls. Samples exposed to higher temperatures, either 42°C or 45 °C, was placed in an incubator. Control samples and the samples exposed to bacteriocin were incubated further at 37°C. The bacteria growth for all samples was observed for 8-10 hours, measuring the OD600 each 30-45 minutes. **Table 3.1.6.1** lists all different concentrations of GarKS and higher temperatures tested.

To avoid batch effect, a new batch of BHI medium was made to be used both for the last 14 samples tested in this experiment and for the cell harvest in the RNA isolation process. The last 14 samples were six WTs and eight mutants; control samples and samples exposed to

[GarKS] equal to 0.0015 mg/mL. In addition, some of these samples were merged in pairs (pairs of WT and pairs of mutants) into the same culture tube before they later were split equally into two different culture tubes. The idea behind this was to get an even growth rate for the samples before inducing stress. The bacteriocin were then added only to one of these culture tubes. The cultures were thoroughly mixed by pipetting before density measuring and before splitting the samples in two.

**Table 3.1.6.1.** Number of tests for the different stress factors (higher temperature and GarKS with different concentrations).

| | Stress factor | Number of samples tested |
|---|---|---|
| WT | control | 6 |
| | 45°C | 2 |
| | 42°C | 1 |
| | [GarKS] = 0.0125 mg/mL | 1 |
| | [GarKS] = 0.00625 mg/mL | 1 |
| | [GarKS] = 0.003125 mg/mL | 1 |
| | [GarKS] = 0.0031 mg/mL | 2 |
| | [GarKS] = 0.0015 mg/mL | 5 |
| | [GarKS] = 0.00075 mg/mL | 2 |
| | [GarKS] = 0.00037 mg/mL | 2 |
| Mutant | control | 7 |
| | 45°C | 2 |
| | 42°C | 1 |
| | [GarKS] = 0.0125 mg/mL | 1 |
| | [GarKS] = 0.00625 mg/mL | 1 |
| | [GarKS] = 0.003125 mg/mL | 1 |
| | [GarKS] = 0.0031 mg/mL | 2 |
| | [GarKS] = 0.0015 mg/mL | 6 |
| | [GarKS] = 0.00075 mg/mL | 2 |
| | [GarKS] = 0.00037 mg/mL | 2 |

### 3.1.7 Isolation of RNA from control and bacteriocin samples; WT and mutant strains of *S. haemolyticus*

In the intent to analyse the gene expression and detect how the WT and mutant of *S. haemolyticus* handles stress due to exposure to the bacteriocin GarKS, the RNA was isolated and sequenced. RNA from two replicates (A and B) of the samples was isolated: WT and mutant without added bacteriocin, and WT and mutant with added bacteriocin. Eight samples in total were sent in for RNA sequencing. A summary of the process is given in **Figure 3.1.7.1**.

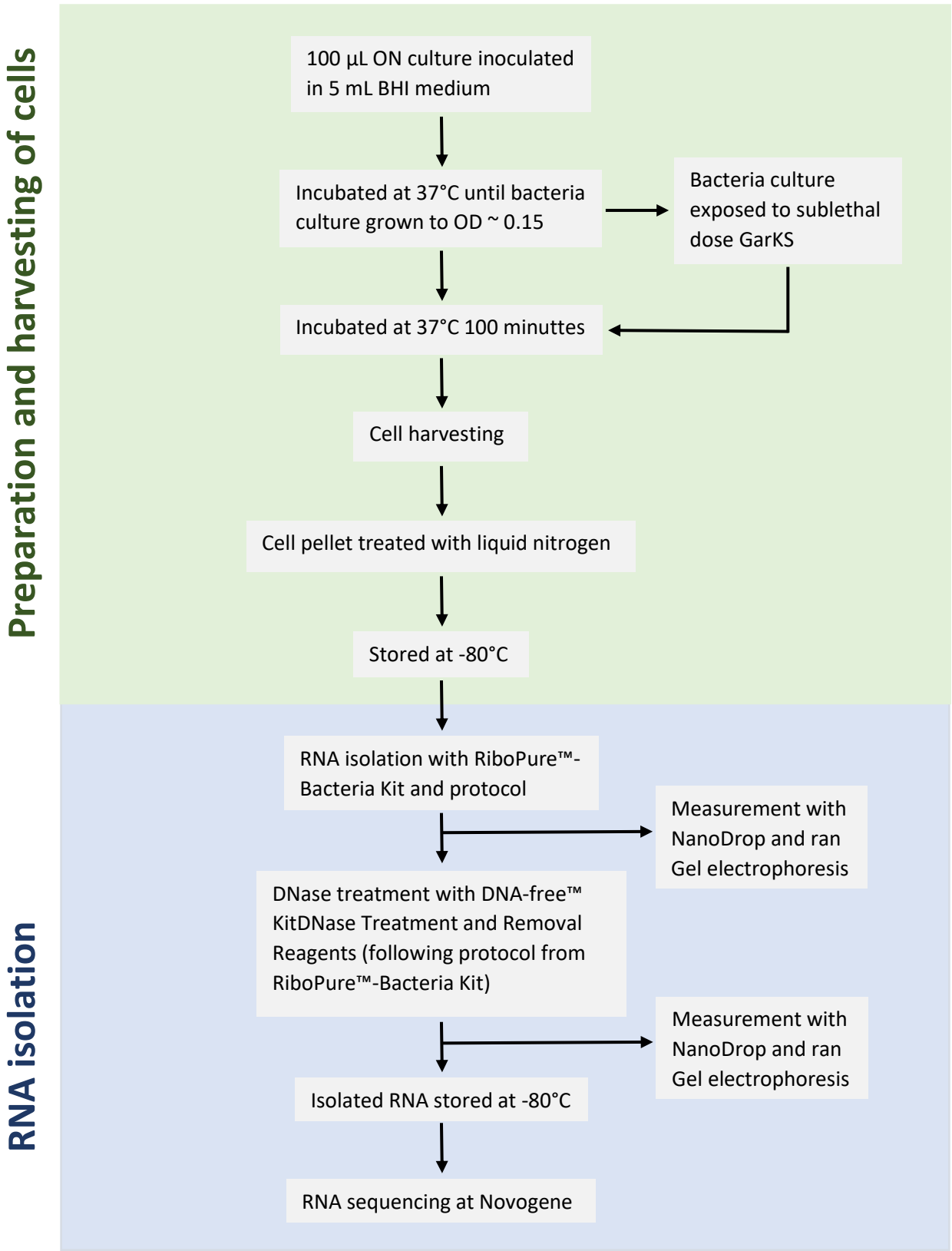


**Figure 1.3.7.1.** Outline of the process of preparing and harvesting of the cells, and RNA isolation.

The four different samples were prepared for the RNA isolation as follows:

- 200 µl ON bacteria culture of the WT was added to 10 mL BHI medium.
- 200 µl ON bacteria culture of the mutant was added to 10 mL BHI medium.
- Additionally, 100 µl ON bacteria culture of the mutant was added to 5 mL BHI medium (extra sample of the mutant).
- The cultures were thoroughly mixed with the medium by pipetting.
- All cell culture tubes were incubated at 37°C until the OD600 value reached 0.15.
- The culture tubes containing 10.2 µl were distributed equally into two culture tubes before adding 7.66 µl 1 mg/mL GarKS to one of each (WT and mutant).
- 7.66 µl GarKS was also added to the last culture tube with the extra mutant.
- All culture tubes were incubated further for 100 minutes at 37°C.
- After incubation, all samples were immediately placed on ice to stop the cell processes.
- The OD600 was measured for all cell culture tubes to calculate the volume needed from each sample to harvest approximately the same number of cells, see **Table 3.1.7.1**.

**Table 3.1.7.1.** Volume of cell culture taken from the samples to harvest cells.

| Sample | OD600 value | Amount of harvested bacteria culture |
|---|---|---|
| WT control (replicate A) | 0.77 | 2 mL |
| Mutant control (replicate A) | 0.57 | 2.5 mL |
| WT bacteriocin (replicate A) | 0.55 | 3 mL |
| Mutant bacteriocin (replicate A) | 0.28 | 5.5 mL |
| WT control (replicate B) | 0.74 | 2 mL |
| Mutant control (replicate B) | 0.73 | 2 mL |
| WT bacteriocin (replicate B) | 0.61 | 2.5 mL |
| Mutant bacteriocin (replicate B) | 0.26 | 6 mL |

**Important precautions related to RNA isolation**

When working with the preparation of cells and RNA isolation, it is critical to avoid contamination, especially RNase contamination, to ensure that the RNA stays intact. Gloves should be worn at all times and changed frequently. To speed up the processes, all equipment should be prepared, and tubes marked in advance. All equipment, components and the working area were washed first with 70% ethanol and sprayed with a generous amount of RNase away (RNase AWAY ™ Surface Decontaminant; Thermo Scientific™) before and throughout the harvesting of cells and the RNA isolation. Pipette tips, Eppendorf tubes, and bottles with $dH_2O$, and 1xTris-Borate-EDTA (TBE) buffer were autoclaved twice. Tubes containing cells and components associated to the isolation process were kept on ice when not in use. For self-protection, safety gloves, safety goggles and a long tweezer were used when handling liquid nitrogen.

Continuing - The four different samples were prepared for the RNA isolation as follows:

- To harvest the cells, the culture tubes were vortexed right before transferring the calculated volume in pre marked Eppendorf tubes. The Eppendorf tubes were balanced and run one minute in the centrifuge at -4°C, 13 000 X g. The supernantant was carefully removed by a pipette without disturbing the cell pellet.
- In a sterile hood the Eppendorf tubes with cell pellet were treated with liquid nitrogen for five minutes for further storage and easier lysis of the cells in the RNA isolation process. The tubes were stored at -80°C.

For the RNA isolation the RiboPure™-Bacteria Kit and protocol (AM1925, Life Technologies) were used. Components from DNase treatment with DNA-free™ Kit DNase Treatment and Removal Reagents (AM1906, Life Technologies) were used to remove remaining DNA molecules. However, the DNA removal was performed with the protocol linked to the RiboPure™-Bacteria Kit. Before and after the DNase treatment, up to 10 µl from each sample of isolated RNA was set aside for running Gel electrophoresis and to measure the concentration, protein and salt contamination (OD260/OD280 and OD260/OD230) on NanoDrop spectrophotometer (NanoDrop 2000; Thermo Scientific).

## 3.2 Sequencing of RNA by Novogene

The eight RNA samples were sent to Novogene Europe (UK) for sequencing. The samples went through quality checks to examine RNA purity and integrity. Since the quality checks from Novogene revealed contamination of DNA in the samples, Novogene performed purification of the samples before the library preparation and the sequencing using the Illumina NovaSeq platform; Illumina NovaSeq 6000.

Novogene used strand-specific RNA library construction with the dUTP method. The read 1 (R1) was sequenced on the original strand and the read 2 (R2) was sequenced on the synthesized strand. Hence, the R1 is the reverse complement of the original fragment, while the R2 corresponds to the original fragment.

Results from Novogene's quality control report of the samples after purification is shown in **Table 3.2.1**. Due to the purification and supplementary gel electrophoresis, a smaller amount of RNA than expected was sequenced for each sample. The integrity value indicates that RNA was still intact even after the RNA isolation process and the processes performed by Novogene.

**Table 3.2.1.** Results from Novogene's quality control report after purification of the eight RNA samples sent in for sequencing (Report Date:2022-03-03). A* and B* samples are the two different replicate groups. The Integrity value presents the RNA integrity number (RIN) which ranges from 1 (totally degraded RNA) to 10 (intact RNA).

| Sample Name | Sample description | Concentration (ng/μl) | Volume (μl) | Total amount (μg) | Integrity value |
|---|---|---|---|---|---|
| A1 | WT control | 38.000 | 32.00 | 1.21600 | 9.70 |
| A2 | mutant control | 31.000 | 32.00 | 0.99200 | 9.80 |
| A3 | WT bacteriocin | 36.000 | 32.00 | 1.15200 | 9.80 |
| A4 | mutant bacteriocin | 61.000 | 32.00 | 1.95200 | 10.00 |
| B1 | WT control | 28.000 | 32.00 | 0.89600 | 10.00 |
| B2 | mutant control | 17.000 | 32.00 | 0.54400 | 9.50 |
| B3 | WT bacteriocin | 19.000 | 32.00 | 0.60800 | 9.50 |
| B4 | mutant bacteriocin | 47.000 | 32.00 | 1.50400 | 10.00 |

## 3.3 Drylab

This chapter will present the methods used in the bioinformatic work to inspect the gene expression in *S. haemolyticus*, WT and mutant.

An overview of version numbers and container information of the software used in this thesis can be found in the appendix (**Table A.2**). R-code used, and necessary data will be uploaded to the following GitHub repository:
https://github.com/djevelpumbakyllingen/R_code_Master_thesis

### 3.3.1 Genome data: *S. haemolyticus* LMGT4071 WT

For the differential gene expression analysis, the RNA-seq reads must be mapped to the genes in order to obtain the read counts for each gene. Hence, the genes of the *S. haemolyticus* LMGT4071 strain need to be annotated from an assembly of the sequenced genome. Assembled genomes were downloaded from National Center for Biotechnology Information (NCBI) Reference Sequence Database (RefSeq). This was done to evaluate and strengthen the gene prediction for the LMGT4071 WT. A pan-genome analysis was performed to inspect the variation of gene content between the LMGT4071 strain and the RefSeq genomes.

The genome assembly of LMGT4071

The genome of *S. haemolyticus* LMGT4071 WT has been sequenced with DNBSEQ at Beijing Genomics Institute, directed by LMG at NMBU. To get an overview of the raw data, RStudio (41) was used to compute the number of reads, read length, the read coverage and K-mer coverage (k = 15). There was 5 576 956 read pairs in the FastQ-files, all reads with length 150 bp. The read coverage was 635 and the K-mer coverage 576. The genome size used for this computation is 2 632 932 bp and was chosen based on another sequenced *S. haemolyticus* genome from NCBI (Genbank: CP025031.1).

For checking the read quality and any presence of overrepresented sequences or remaining adapter sequences, read 1 (R1) and read 2 (R2) were analysed by the FastQC software (42). FastQC reported a GC content of 33% and an average sequenced quality of 36, which is based on the quality scores provided in the FastQ-files for each base in each read. The reads

were filtered, and adapter sequences removed by BBduk from the BBmap software (43) before FastQC was run once more on the filtered reads to examine these results.

```
singularity exec /cvmfs/singularity.galaxyproject.org/b/b/bbmap:38.86--h1296035_0
bbduk.sh t=$threads in1=$r1_WT in2=$r2_WT ref=$adapters out1=$r1_WT_out
out2=$r2_WT_out outs=$unpaired_WT_out ktrim=r k=23 mink=11 hdist=1 tpe tbo
```

The filtered FastQ-files were run through SPAdes (22) to assemble the genome. The contigs generated from SPAdes were filtered in RStudio, keeping only the contigs longer than 500 bases and with a K-mer coverage over 50.

```
singularity exec /cvmfs/singularity.galaxyproject.org/s/p/spades:3.15.3--
h95f258a_0 spades.py -o $out_dir_WT --tmp-dir $TMPDIR --memory 99 --threads
$threads --isolate --pe-1 1 $r1_WT --pe-2 1 $r2_WT
```

Both the total number of assembled contigs and the remaining contigs after filtering out short contigs and the ones with low K-mer coverage was evaluated by the QUAST software (44), with and without a reference genome. The commands from running QUAST on all contigs, with and without a reference, are presented:

```
singularity exec /cvmfs/singularity.galaxyproject.org/q/u/quast:5.0.2--
py37pl5262h190e900_4 quast.py --threads $threads --output-dir $out_dir/WT
$contigs_WT
```

```
singularity exec /cvmfs/singularity.galaxyproject.org/q/u/quast:5.0.2--
py37pl5262h190e900_4 quast.py --threads $threads -r $fna_file --output-dir
$out_dir/$fna_file_name\_$line $contigs_WT
```

Gene prediction of the LMGT4071 strain and the RefSeq genomes

Prior to the differential gene expression analysis, the genes in the LMGT4071 strain must be annotated. As the goal is to uncover as much of the gene content as possible, it is more beneficial to have a long list of predicted genes with many false positives than a shorter list with less true positives. Thus, 19 Fasta-files of complete genomes of *S. haemolyticus* from RefSeq were downloaded from NCBI (See RESULTS page 54 for RefSeq and Assembly accession numbers) to increase the possibility of covering all genes in the LMGT4071 WT genome. The genes from these RefSeq genomes were already annotated. However, their genes were predicted together with the LMGT4071 strain genes, attaining all annotated genes through the same process. An overview of the gene prediction process is illustrated in **Figure 3.2.1.1**.

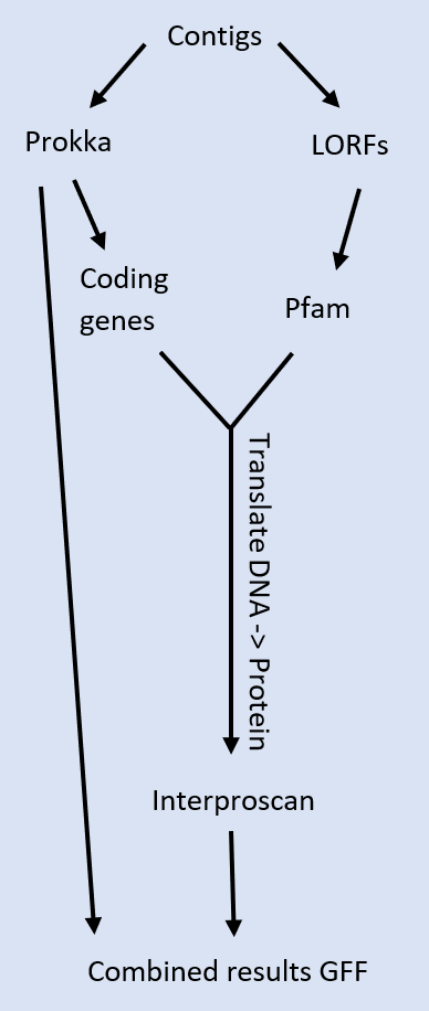


**Figure 3.2.1.1** Illustration of the gene annotation process.

For annotating the genes in the 20 genomes, Prokka was used, a software tool for annotating prokaryotic and viral genomes (24). Prokka is utilizing the software Prodigal, which predicts the coding genes, and four other tools for detecting non-coding genes.

```
singularity exec /cvmfs/singularity.galaxyproject.org/p/r/prokka:1.14.6--
pl5262hdfd78af_1 prokka --force --cpus $threads --outdir $out_dir --prefix
$genome_id $genome_file
```

Although it is the coding genes that are the primary interest in this study, the prediction of non-coding genes will be useful for distinguishing non-coding genes from other non-mapping reads, or to prevent these reads from mapping to the wrong genes resulting in incorrect read count for the gene.

In addition to predicting genes by Prokka, the function *findOrfs()* in the R package microseq (v 2.1.5) was used to collect all Open Reading Frames (ORFs) from the 20 genomes, and keeping only the longest ORFs (LORFs) with the *lorfs()* function. The collected LORFs were

further translated to protein and filtered by a minimum length of 50 aa. The LORFs that included both start and stop (not truncated) were scanned against the Pfam (45) database through the HMMER software (46). The LORFs without any hit against the Pfam database, and genes already predicted from Prokka, were filtered out.

```
singularity exec /cvmfs/singularity.galaxyproject.org/h/m/hmmer:3.3.2--h1b792b2_1
hmmscan --cpu $threads --domtblout $out_file\_$SLURM_ARRAY_TASK_ID.txt $dbase
$genome_file
```

At last, the list of all coding genes predicted with Prokka and the LORFs with hits against Pfam, were translated to protein and run through the InterProScan software (47) with the chosen databases CDD, Hamap, Pfam, TIGRFAM. InterProScan gives information of protein function, domains and families.

```
singularity run
--bind $work_dir:/data \
--bind $db_dir:/opt/interproscan/data \
interproscan_5.45-80.0_1.sif -i proteins.faa -a $apps -d $out_dir -c -C $threads -
f GFF3 -l -g -p
```

The resulting GFF tables from Prokka and InterProScan were merged, keeping all information of the gene annotations from all the 20 genomes. As none of the genes were annotated specifically as "*rseP*", the gene that corresponded to *rseP* was detected through a BLAST search (online version, 48) to ensure the annotation for this gene sequence. In addition, the predicted genes not categorized as either coding or non-coding genes were search through the BLAST database in the attempt to annotate these genes.

Pan genome analysis

In the interest of finding out if the annotated genes in the LMGT4071 genome corresponds to what is expected based on the gene content in other strains within the same species, a pan genome analysis was performed for this genome and the 19 RefSeq genomes. This was done by running Roary (49) to cluster the predicted genes from Prokka from all genomes.

```
singularity exec /cvmfs/singularity.galaxyproject.org/r/o/roary:3.9.1--pl5.22.0_0
roary -p $threads -f $out_dir -e -n -i 90 -r -v $input_gff_files
```

The ape (v 5.6.2), ggtree (v 3.4.0) and micropan (v 2.1) R packages were used to plot a pan-genome tree from manhattan distances of the pan-matrix to uncover how the WT genome relate to the other *S. haemolyticus* RefSeq genomes.

### 3.3.2 RNA-seq data, Kallisto index and read-mapping

In order to analyse the data from the sequencing of the RNA samples, it is essential to detect which gene each of the reads originates from. Kallisto (27) is a software used for this purpose; to map reads to an index built from sequences of the genes in the genome.

Index building with Kallisto

It is crucial that Kallisto builds the index from unique gene sequences to prevent reads originated from the same gene mapping to multiple sequences, giving the wrong picture of read counts. The merged GFF table from the gene annotations contains duplicated sequences due to both the information from Prokka and InterProScan, and from the 20 similar genomes. Therefore, the GFF table were filtered, keeping only unique genes primarily from the LMGT4071 strain. BLAST (48) was used with the LMGT4071 genes as database and all predicted genes from the RefSeq genomes as query to find additional gene sequences.

```
singularity exec /cvmfs/singularity.galaxyproject.org/b/l/blast:2.9.0--
pl526he19e7b1_7 makeblastdb -dbtype nucl -in $LMGT4071_genes -out $dbase
```

```
singularity exec /cvmfs/singularity.galaxyproject.org/b/l/blast:2.9.0--
pl526he19e7b1_7 blastn -num_threads $threads -query $quary_genes -db $dbase -out
$out file -outfmt 6
```

The read mapping was tested both on the Kallisto index based only on the LMGT4071 strain and the index based on the LMGT4071 strain together with the genes from the other genomes with no match in the BLAST search.

```
singularity exec /cvmfs/singularity.galaxyproject.org/k/a/kallisto:0.46.2--
h60f4f9f 2 kallisto index -i index1 LMGT4071 270322.idx $index1
```

Trimming and filtering of RNA-seq reads

The quality of the raw data from the eight sequenced RNA samples were checked with FastQC (42) and the results summarised with MultiQC (50). 11 bases were trimmed of the left side of all reads in addition to filter and remove adapter sequences with the BBmap

software. Read quality was examined once more with FastQC and MultiQC. Short reads produced by the trimming were not removed.

Read mapping

The read mapping with Kallisto (27) was performed first on one RNA-seq sample (WT control) with combinations of raw reads or processed reads, and the index built from only the LMGT4071 genes or the index also containing the additional genes from the reference genomes. This was done to examine the consistency of the mapping results, and to find the combination which resulted in highest percentage of uniquely mapped reads. The combination of processed reads and the index built from only the LMGT4071 genes was selected for further analyses, and therefore performed for all the RNA-seq samples.

```
singularity exec /cvmfs/singularity.galaxyproject.org/k/a/kallisto:0.46.2--
h60f4f9f_2 kallisto quant -i $index1 -o $outdir -t $threads $fastq_r1 $fastq_r2
```

### 3.3.3 Analysis of RNA-seq data

The differential expression analysis was performed to assess the observed difference in phenotype between WT and mutant on a transcript-level. The genes within the same operon are expected to show similar expression as they are transcribed together. Based on this, the genes were attempted to be grouped together to inspect the differential expression within operons.

Related to the analysis of the RNA-seq data, the four different sequenced RNA samples will be referred to as follows:

- WT bacteriocin: samples of WT exposed to GarKS
- WT control: samples of WT not exposed to GarKS
- Mutant bacteriocin: samples of mutant exposed to GarKS
- Mutant control: samples of mutant not exposed to GarKS

Differential gene expression analysis

The *DESeq* function from the DESeq2 package from Bioconductor (v1.36.0) was utilized for the differential gene expression analysis.

After the read-mapping with Kallisto, the raw read counts were inspected to evaluate the read-mapping. To check which RNA molecules the mapped reads corresponded to, the

number of coding genes and non-coding genes were found. Genes with zero mapped reads were filtered out before the differential expression analysis.

To see which samples that cluster together based on the information in the read counts, and to check for batch effect, a Principal Component Analysis (PCA) plot and a dendrogram was made. The PCA plot was generated from the function *plotPCA* from DESeq2, and all coding genes were used. The dendrogram with hierarchical clustering of the samples was plotted using the functions *dist* and *hclust* (stat package in R) with Euclidean distance and complete-linkage. Normalized read counts obtained by the DESeq2 function *varianceStabilizingTransformation* were used to generate both the PCA plot and the dendrogram.

The *DESeq* function requires a specific data object (DESeqDataSet object) containing the raw read counts and metadata of samples and genes. This object was created by using the *tximport* function from the R package tximport (v1.24.0) and the *DESeqDataSetFromTximport* function from DESeq2 with multi-factor design including an interaction term (design = ~ Genotype + Treatment + Genotype:Treatment). With this design, differential expression analysis can account for the effect of the mutant without exposure of GarKS (Genotype) and the effect of the bacteriocin GarKS (Treatment) separately or the two effects combined (29). The contrasts, comparison of samples, were specified with the reference level being either WT control, mutant control or both.

With use of the DESeq function, the raw read counts are used as input and are normalized by taking the median of the rations of the read counts. This normalization of the data accounts for sequencing depth and RNA composition in the samples (51).

Cluster genes in operons

To be able to inspect the gene expression in relation to how the genes are organized into operons, an attempt to uncover these operons were performed. Since the genes in the same operon are transcribed together resulting in one transcript, the RNA-seq reads were assembled with the intent of recover each multi-gene transcript. The sequences of the predicted genes were then aligned to these assembled transcripts.

For the assembly, RNA-seq reads from four samples were used: WT and mutant, control and with bacteriocin. Four different modes of SPAdes (22) were used to assemble the reads: SPAdes with and without the *--careful* option, rnaSPAdes with specifying the read orientation for the paired-end sequencing (rf: “first read in pair corresponds to reverse gene

strand"(52)) and metaSPAdes. For the latter, the reads from one WT control and one mutant with bacteriocin were used.

```
singularity exec /cvmfs/singularity.galaxyproject.org/s/p/spades:3.15.3--
h95f258a_0 rnaspades.py -o $out_dir --tmp-dir $TMPDIR --memory 99 --threads
$threads --ss-rf \
--pe-1 1 $r1_B1 --pe-2 1 $r2_B1 --pe-1 2 $r1_B2 --pe-2 2 $r2_B2 \
--pe-1 3 $r1_B3 --pe-2 3 $r2_B3 --pe-1 4 $r1_B4 --pe-2 4 $r2_B4
```

The contigs were filtered on K-mer coverage and length in RStudio, keeping the contigs longer than 119-149 bases and a K-mer coverage above 4-5. The assembly by rnaSPAdes yield different isoforms of the transcripts. Only the longest isoform variant for each transcript were used for further analyses with BLAST (48).
The command makeblastdb was used to make individual BLAST databases from the contigs from each of the assembly methods. The predicted genes were used as query to align with both blastn (aligning the nucleotide sequences) and tblastx (alignment of database and query both translated to protein).

```
singularity exec /cvmfs/singularity.galaxyproject.org/b/l/blast:2.9.0--
pl526he19e7b1_7 blastn -num_threads $threads -query $query_genes -db $dbase -out
$out_file_blastn -outfmt "6 qseqid sseqid pident length mismatch gapopen qstart
qend sstart send evalue bitscore qlen"
```

The results from BLAST were read into RStudio and filtered as follows: keeping the match with highest bitscore, percentage of the query length aligned and percent identity. In addition, the matches with a percent identity less than 80 and a percentage of the query length aligned less than 50, were filtered out. For the genes that still matched equally good to more than one transcript, information of all assigned groups was kept. The results from the different methods were compared with the *adj.rand.index* function from the fossil R package.

The differential expressed genes and *rseP*

The results from the differential expression analysis were inspected further in the attempt of revealing how the WT handles the stress caused by exposure to GarKS which the mutant cannot manage. In addition, it is noteworthy to examine the genes that appeared to be within a reach of the *rseP* gene.

Of particular interest are the genes found as differentially expressed by comparing WT bacteriocin and WT control, while not differentially expressed by comparing mutant bacteriocin and mutant control. To substantiate that these genes are connected to how the WT handles the stress, the genes were also considered in the contrast between WT and mutant control samples. The gene sequences were search through the BLAST database (online version) to verify their annotations from Prokka and InterProScan.

The genes adjacent to *rseP* were investigated in relation to their log2 Fold Change value from four of the contrasts, assigned clusters and annotation.

# 4 RESULTS

## 4.1 Wetlab

### 4.1.1 The mutations uncovered from sequencing of the *rseP* gene

The PCR and sequencing of the *rseP* gene was done to detect the mutation in the gene for each bacterial strain. The PCR products were run on gel to confirm that the amplified fragments corresponded to their expected lengths: the length of the *rseP* gene in addition to various number of bases upstream and downstream the gene. The amplified fragments are shown in **Figure 4.1.1.1**. Expected lengths of the PCR products in *S. haemolyticus* and *L. lactis* is 1511 and 1510 bp respectively. All four bands belonging to these bacteria are at 1.5 kilo bases (kb). The PCR products in *E. faecalis* have an expected length of 1894 bp, while 1858 bp for the *E. faecium* strains. These bands appear just below 2kb.

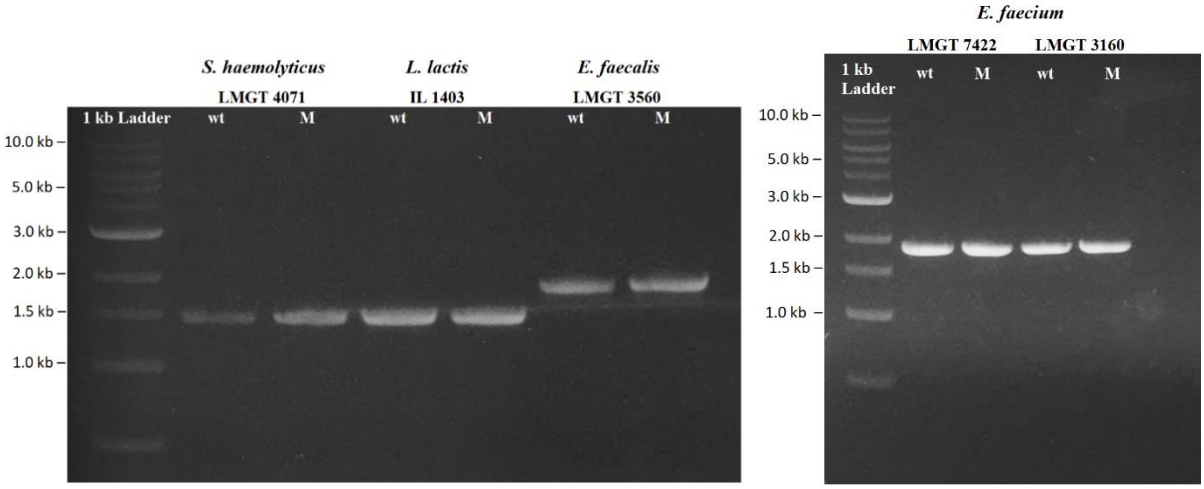


**Figure 4.1.1.1.** Gel images; PCR amplification of the *rseP* gene. The positions of the band from left to right; *S. haemolyticus* LMGT4071 WT and mutant at 1500 bases, *L. lactis* IL1403 WT and mutant at 1500 bases, *E. faecalis* LMGT3560 WT and mutant just below 2000 bases, *E. faecium* LMGT7422 and LMGT3160 WT and mutant just below 2000 bases. The ladder to the left in each picture is a 1 kb DNA Ladder.

The mutations' location, type and the consequences of the different mutations are listed in **Table 4.1.1.1**. The gene product of *rseP* is 428 for *L. lactis* and *S. haemolyticus*, and 422 aa for *E. faecalis* and *E. faecium*.

**Table 4.1.1.1.** Information of the mutations found in the *rseP* gene in the mutant of the five bacterial strains. *Location of the mutation* is described in number of bases from the start of the gene. *Consequence of the mutation* gives the length of the translated product in number of amino acids (aa).

| Bacteria strain | Location of the mutation | Type of mutation | Consequence of the mutation |
|---|---|---|---|
| *S. haemolyticus* LMGT4071 | Position 69 downstream | Substitution T → G | TAT → TAG (stop codon) Prematurely termination resulting in 22 aa |
| *L. lactis* IL1403 | Position 7 downstream | Substitution G → T | GAA → TAA (stop codon) Prematurely termination resulting in length of 2 aa |
| *E. faecalis* LMGT3560 | After position 699 downstream | Multiple insertions | Frame shift, changing aa Prematurely termination resulting in 256 aa |
| *E. faecium* LMGT7422 | After position 520 downstream | Deletion of base C | Frame shift, changing aa Prematurely termination resulting in 179 aa |
| *E. faecium* LMGT3160 | - | - | - |

The sequencing of *S. haemolyticus* LMGT4071 mutant showed a substitution of thymine to guanine at the 69$^{th}$ base from the start of the *rseP* gene. This substitution leads to altering of the original codon TAT, coding for tyrosine, to become the codon TAG which signals termination in the translation process. Hence, the prematurely termination results in a peptide containing 22 aa.

The mutation in the *L. lactis* IL1403 mutant is also a base substitution. A thymine base at position 7 downstream replaced a guanine base, causing a stop codon, TAA, instead of the codon GAA which codes for glutamic acid. The peptide should result in a length of 2 aa due to premature termination.

The *E. faecalis* LMGT3560 mutant had insertions of the bases "CAAAAAAT" somewhere near and downstream of the 699$^{th}$ base. The insertion of 8 bases leads to a frame shift, changing multiple codons downstream of the mutation, generating a stop-codon, TAA, at the 257$^{th}$ codon. The premature termination produces 256 aa.

Deletion of one base after the 520 $^{th}$ base downstream, that occurs as a cytosine in the WT, was found in the *E. faecium* LMGT7422 mutant's *rseP* gene. The frame shift caused by deletion of one base produces a stop codon, TGA, at codon 180, producing a peptide of 179 aa. The mutation in the *E. faecium* LMGT3160 mutant was not detected by sequencing of the gene.

### 4.1.2 Soft-agar overlay assay

The Soft-agar overlay assay was used to test whether or not the WT and mutant of each bacterial strain were resistant against the bacteriocins LsbB, EntK1 EJ97, EntK1, EntEJ97 and GarKS. The results of this experiment can be seen in **Figure 4.1.2.1**.

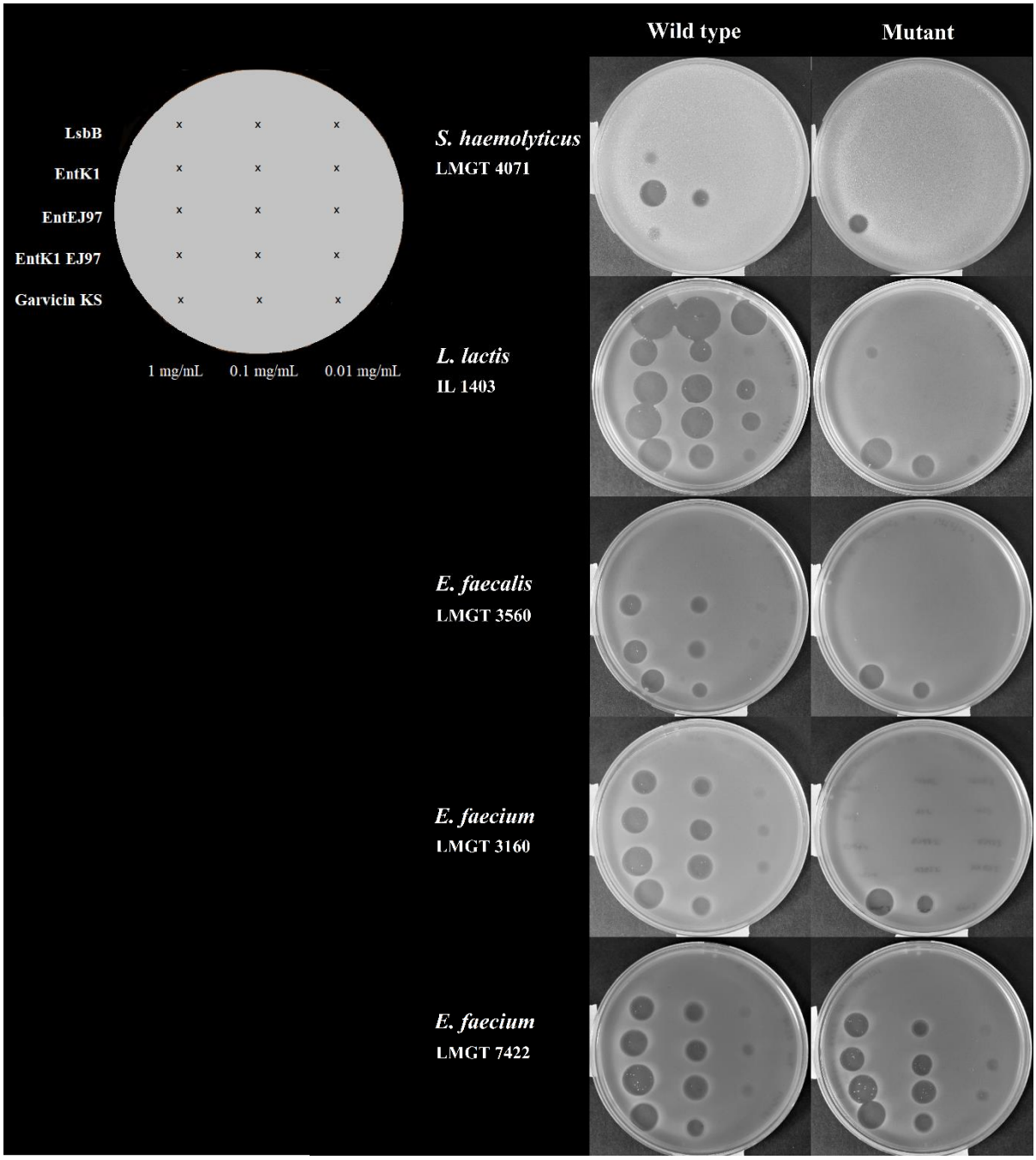


**Figure 4.1.2.1.** Images of the petri dishes in the Soft-agar overlay assay for WT and mutant strains of *S. haemolyticus* LMGT4071, *L. lactis* IL1403 and *E. faecalis* LMGT 3560. The bacteriocin LsbB, EntK1, EntEJ97, EntK1EJ97 and GarKS were applied with the concentrations 1 mg/mL, 0.1 mg/mL and 0.01 mg/mL, and inhibition of bacteria growth appears as dark zones in the petri dish. A legend at the upper left corner indicates were the different bacteriocins with the various concentrations were applied.

The *S. haemolyticus* LMGT401 WT showed inhibition of growth when exposed to 1 mg/ml EntEJ97, EntK1 EJ97 and GarKS and 0.1 mg/ml EntK1 EJ97. The inhibition zones for both concentrations of EntK1 EJ97 were most clear. The petri dish with the mutant show resistance to all bacteriocins, except the highest concentration of GarKS, which produces a larger inhibition zone than for the WT.

The *L. lactis* IL1403 WT appears not resistant to any of the bacteriocins and has smaller inhibition zones for 0.01 mg/ml EntK1 and GarKS. The mutant has a small inhibition zone at 1 mg/ml EntK1. Where GarKS is applied, the size of the inhibition zone decreases with the concentration of bacteriocin. The size of the inhibition zones at GarKS seems equal between WT and mutant.

The *E. faecalis* LMGT3560 WT shows inhibition zones when exposed to all concentrations of EntEJ97 and EntK1 EJ97, smaller zones for the less concentrated solutions. The two highest concentrations of GarKS produces similar inhibition zones both for the WT and the mutant.

Both the *E. faecium* LMGT3160 WT and *E. faecium* LMGT7422 WT show inhibition zones for all concentrations of bacteriocins except for the LsbB and the smallest concentration of GarKS. The mutant of the LMGT3160 strain has only inhibition zones for 1 mg/ml and 0.1 mg/ml GarKS, while the mutant of the LMGT7422 strain show similar inhibition zones as the WT.

### 4.1.3 Tolerance to bacteriocin in microtiter plates

Minimum inhibitory concentration was performed to further assess results from the Soft-agar overlay assay and find the MIC value. Each bacteriocin that produced inhibition zones either on the WT or mutant in the Soft-agar overlay assay were tested on the strain.

**Table 4.1.3.1** Results from MIC test for WT and mutant of *S. haemolyticus* LMGT4071 exposed to EntK1 EJ97 and GarKS. The OD600 measured after four and 24 hours. Final MIC - value is obtained after 24 hours.

***S. haemolyticus* LMGT4071**

| WT/mutant | Bacteriocin | MIC 4 hours | MIC 24 hours |
|---|---|---|---|
| WT | EntK1 EJ97 | 0.00625 mg/mL | 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | GarKS | 0.05 mg/mL | > 0.05 mg/mL |
| mutant | | 0.025 mg/mL | 0.05 mg/mL |

The hybrid bacteriocin EntK1 EJ97 and GarKS were tested on *S. haemolyticus*, see **Table 4.1.3.1**. The results indicates that the WT is more sensitive to EntK1 EJ97 than the mutant, whilst exposed to GarKS the mutant is the most sensitive of the two.

In **Table 4.1.3.2** is the results from exposing *L. lactis* to all five bacteriocins. The WT showed more sensitivity against the bacteriocins LsbB, EntK1, EntEJ97 and EntK1 EJ97 than the mutant, which had growth in all wells after 24 hours. Exposed to GarKS, the WT appears more resistant than the mutant after four hours, but equal sensitive after 24 hours.

**Table 4.1.3.2** Results from MIC test for WT and mutant of *L. lactis* IL1403 exposed to LsbB, EntK1, EJ97, EntK1 EJ97 and GarKS. The OD600 measured after four and 24 hours. Final MIC - value is obtained after 24 hours.

***L. lactis* IL1403**

| WT/mutant | Bacteriocin | MIC 4 hours | MIC 24 hours |
|---|---|---|---|
| WT | LsbB | < 0.000390625 mg/mL | 0.0015625 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntK1 | 0.003125 mg/mL | 0.025 mg/mL |
| mutant | | 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntEJ97 | 0.003125 mg/mL | 0.0125 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntK1 EJ97 | < 0.000390625 mg/mL | 0.0125 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | GarKS | 0.003125 mg/mL | 0.003125 mg/mL |
| mutant | | < 0.000390625 mg/mL | 0.003125 mg/mL |

*E. faecalis* was exposed to EntEJ97, EntK1 EJ97 and GarKS, the results are shown in **Table 4.1.3.3**. After 24 hours, both WT and mutant exposed to either EntEJ97 or EntK1 EJ97 had growth in all wells. From the measurement at four hours, the WT appears less resistant than the mutant. When exposed to GarKS, WT and mutant showed equal sensitivity against the bacteriocin after 24 hours, while the mutant had growth in one more well than the WT after 4 hours.

**Table 4.1.3.3** Results from MIC test for WT and mutant of *E. faecalis* LMGT3560 exposed to EJ97, EntK1 EJ97 and GarKS. The OD600 measured after four and 24 hours. Final MIC - value is obtained after 24 hours

***E. faecalis* LMGT3560**

| WT/mutant | Bacteriocin | MIC 4 hours | MIC 24 hours |
|---|---|---|---|
| WT | EntEJ97 | 0.0125 mg/mL | > 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntK1 EJ97 | 0.05 mg/mL | > 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | GarKS | 0.00625 mg/mL | 0.0125 mg/mL |
| mutant | | 0.0125 mg/mL | 0.0125 mg/mL |

MIC-values for *E. faecium* LMGT7422 exposed to EntK1, EntEJ97, EntK1 EJ97 and GarKS are shown in **Table 4.1.3.4**. After 24 hours both WT and mutant showed no sensitivity to the first three bacteriocins. After four hours, the WT appears less sensitive against EntK1 and more sensitive against Ent EJ97 than the mutant. When exposed to GarKS, the mutant is slightly more sensitive after 24 hours.

**Table 4.1.3.4** Results from MIC test for WT and mutant of *E. faecium* LMGT7422 exposed to EntK1, EJ97, EntK1 EJ97 and GarKS. The OD600 measured after four and 24 hours. Final MIC - value is obtained after 24 hours

***E. faecium* LMGT7422**

| WT/mutant | Bacteriocin | MIC 4 hours | MIC 24 hours |
|---|---|---|---|
| WT | EntK1 | 0.0125 mg/mL | > 0.05 mg/mL |
| mutant | | 0.00625 mg/mL | > 0.05 mg/mL |
| WT | EntEJ97 | 0.0125 mg/mL | > 0.05 mg/mL |
| mutant | | 0.025 mg/mL | > 0.05 mg/mL |
| WT | EntK1 EJ97 | 0.0125 mg/mL | > 0.05 mg/mL |
| mutant | | 0.0125 mg/mL | > 0.05 mg/mL |
| WT | GarKS | 0.00625 mg/mL | 0.025 mg/mL |
| mutant | | 0.00625 mg/mL | 0.0125 mg/mL |

WT and mutant of the LMGT3160 strain were also exposed to EntK1, EntEJ97, EntK1 EJ97 and GarKS, see **Table 4.1.3.5**. As for the other *E. faecium* strain, both WT and mutant showed no sensitivity against the highest concentration of EntK1, EntEJ97 and EntK1 EJ97 after 24 hours. The measurement at four hours indicates that the WT is more sensitive to EntK1 and EntK1 EJ97. After 24 hours of exposure to GarKS the mutant had growth in two more wells than the WT, which indicates the WT to be most sensitive.

**Table 4.1.3.5** Results from MIC test for WT and mutant of *E. faecium* LMGT3160 exposed to EntK1, EJ97, EntK1 EJ97 and GarKS. The OD600 measured after four and 24 hours. Final MIC - value is obtained after 24 hours

***E. faecium* LMGT3160**

| WT/mutant | Bacteriocin | MIC 4 hours | MIC 24 hours |
|---|---|---|---|
| WT | EntK1 | 0.00625 mg/mL | > 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntEJ97 | > 0.05 mg/mL | > 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | EntK1 EJ97 | 0.0125 mg/mL | > 0.05 mg/mL |
| mutant | | > 0.05 mg/mL | > 0.05 mg/mL |
| WT | GarKS | 0.00625 mg/mL | 0.00625 mg/mL |
| mutant | | 0.00625 mg/mL | 0.025 mg/mL |

Due to the absence of mutation in the *rseP* gene for LMGT3160 and the vague difference in the LMGT7422 WT and mutant's phenotype, either of *E. faecium* strains were studied in further experiments.

### 4.1.4 Growth rate and tolerance to bacteriocin and temperature in microtiter plates

The growth rate of *S. haemolyticus*, *L. lactis* and *E. faecalis* exposed to GarKS and higher temperatures in microtiter plates was measured to examine and assess any difference between the WT and mutant within the same strain. Results of selected wells from the microtiter plates are shown in **Figure 4.1.4.1**.

The *S. haemolyticus* LMGT4071 WT shows less sensitivity against GarKS than the mutant, both at 37°C and 42°C. Growth of the mutant are inhibited with the bacteriocin concentration of 0.025 mg/ml, while the WT shows growth after some time under the same conditions. The higher temperature in addition to bacteriocin seems to affect the mutant more than the WT. At 0.0125 mg/mL the growth rate for the WT is still affected, while the lower concentrations of bacteriocin have less effect on the growth. First at the lowest concentration of GarKS, the mutant's growth rate is similar to the WT.

From the growth rate measurements of *L. lactis* IL1403, the mutant appears slightly less sensitive to the various concentrations of GarKS than the WT. This is the case when incubated in both 30°C and 37°C. Note that the difference in bacteriocin concentration across the wells are smaller for the experiment of *L. lactis* than for *S. haemolyticus*.

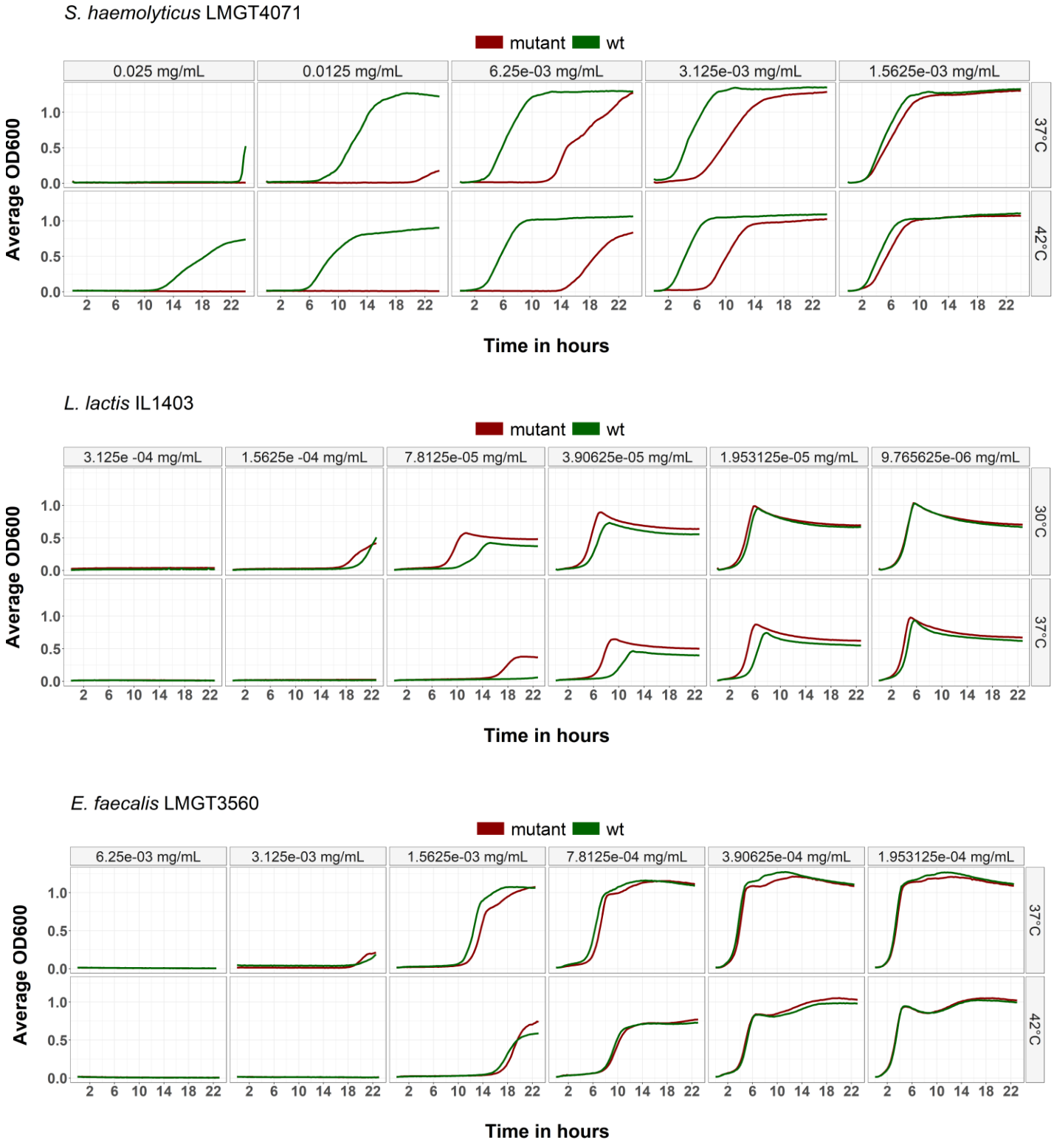


**Figure 4.1.4.1.** Growth curves for WT and mutant strains of *S. haemolyticus* (first plot), *L. lactis* (second plot), and *E. faecalis* (third plot). The rows in each plot correspond to the incubation temperatures, 37°C in the first, and 42°C in the last row for the first and third plot, and 30°C in the first, and 37°C in the last row for the second plot. Columns correspond to the different concentrations of GarKS the bacteria were exposed to. The lines in each cell are drawn from the average OD600 measured of four samples, WT with green lines and mutant with red lines. The plots are made using the R package ggplot2.

There are minor differences in growth rate between the WT and mutant *E. faecalis* LMGT3560 when exposed to GarKS alone and to bacteriocin and the higher temperature combined. Most of the growth is inhibited even in 0.003 mg/mL GarKS. If any difference, the WT appears less sensitive in the combination of bacteriocin and a temperature of 37°C, while more sensitive with bacteriocin and a temperature of 42°C, than the mutant.

Further experiments were performed only for *S. haemolyticus* and did not include either *L. lactis* or *E. faecalis*. This because only the WT and mutant of *S. haemolyticus* LMGT4071 showed difference in growth when exposed to GarKS in these growth measurements.

### 4.1.5 Growth rate and tolerance to bacteriocin and temperature in culture tubes

Measurements of the growth for *S. haemolyticus* LMGT4071, WT and mutant, exposed to higher temperatures and various concentrations of GarKS in culture tubes was done to examine the growth curves for the different conditions, and to find the appropriate condition for further analysis. The resulting growth curves from nine of the conditions tested are shown in **Figure 4.1.5.1**.

The plots in **A** show how the WT and mutant grow in "normal" conditions, incubated at 37°C and not exposed to any stress factors. These serve as control samples. In the plots with growth curves from more than one sample, the measured OD600 at a given time varies between the samples within the same condition. This variation is a result of both a divergent number of cells from the beginning of the experiment and the variation in growth between the samples. However, the growth rate appears approximately equal, both across the samples within WT and mutant separately, and across the WT control and mutant control samples. This pattern shows slow growth in the beginning (the lag phase) exponential growth in the log phase (OD600~0.20-1.5) and a stagnation of growth reaching the beginning of the stationary phase.

The plots in **B** show the growth for WT and mutant incubated at 45°C after they reached OD600~0.15. These two growth curves resemble the growth rate for the control samples and appear not to be different between the WT and the mutant.

The **C** and **D** plots show growth curves for the WT and mutant exposed to the two highest concentrations of GarKS. All the four growth curves are flat and indicate inhibition of growth for both WT and mutant. In plot **D** for the WT exposed to 0.0125 mg/mL GarKS there seems to be beginning of growth. The **E** plots, bacteria exposed to 0.003125 mg/mL GarKS, demonstrate that the mutant are more sensitive than the WT to the bacteriocin. The same can

be seen in the plots **F** and **G**. The last four plots in **H** and **I** show the growth curves for the bacteria samples exposed to the two lowest concentrations of GarKS. In these cases, both the WT and the mutant appears susceptible to the amount of GarKS.

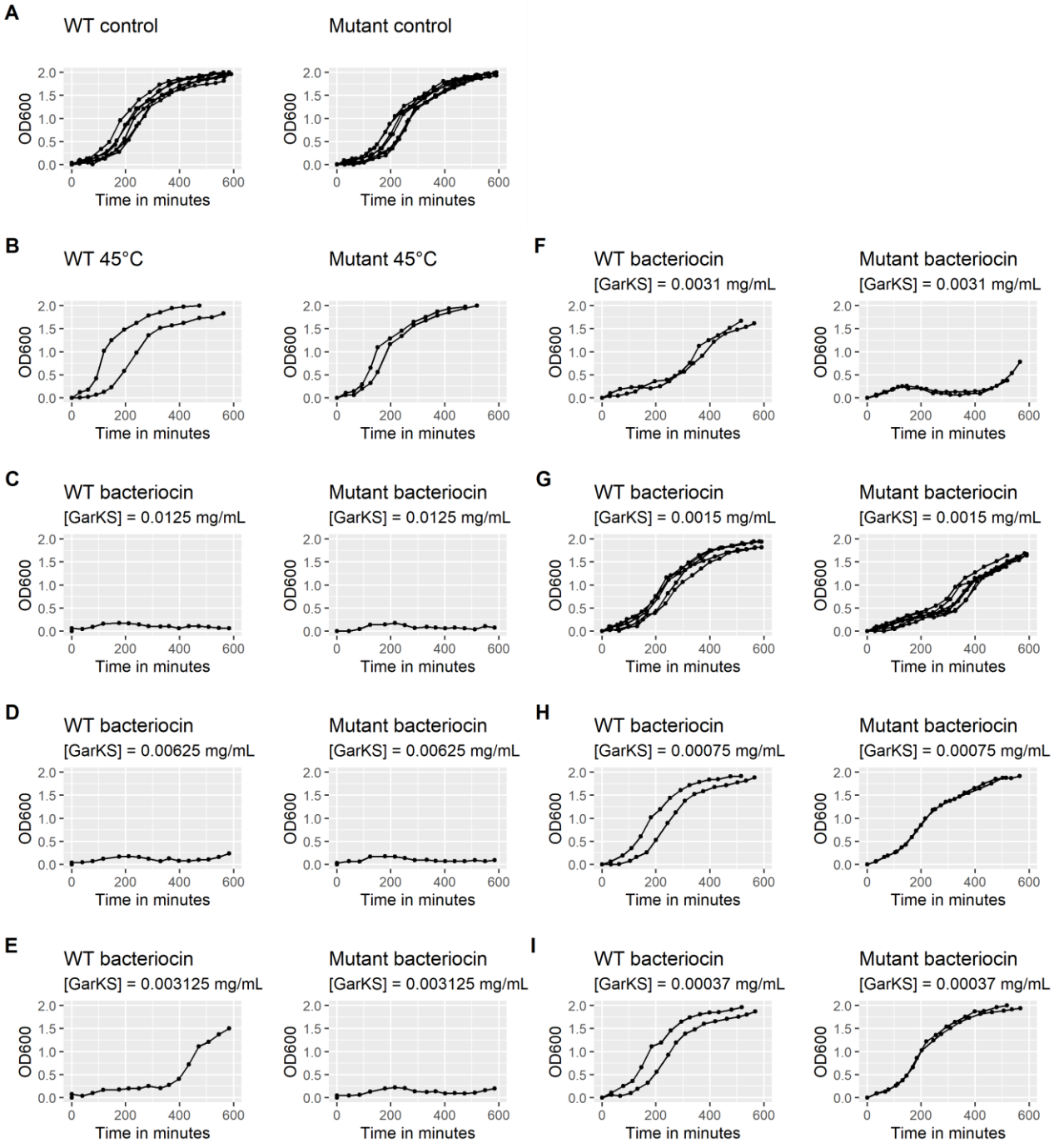


**Figure 4.1.5.1**. Bacterial growth curves from *S. haemolyticus* LMGT4071 WT and mutant grown in various conditions. The character label represents samples under the same condition. **A**: WT and mutant incubated at 37°C without bacteriocin. They serve as controls. **B**: WT and mutant incubated in 45°C after reached OD ~ 0.15. **C-I**: WT and mutant exposed to GarKS, with decreasing concentrations. There are different numbers of samples contributing to each plot, each line in a plot representing either WT or mutant under the same conditions. The points on the lines show the OD600-value measured at a given time. The first point at 0 minutes and OD600 =0 is sat as a starting point and are not an observation. The plots are made using the R packages ggplot2 and cowplot.

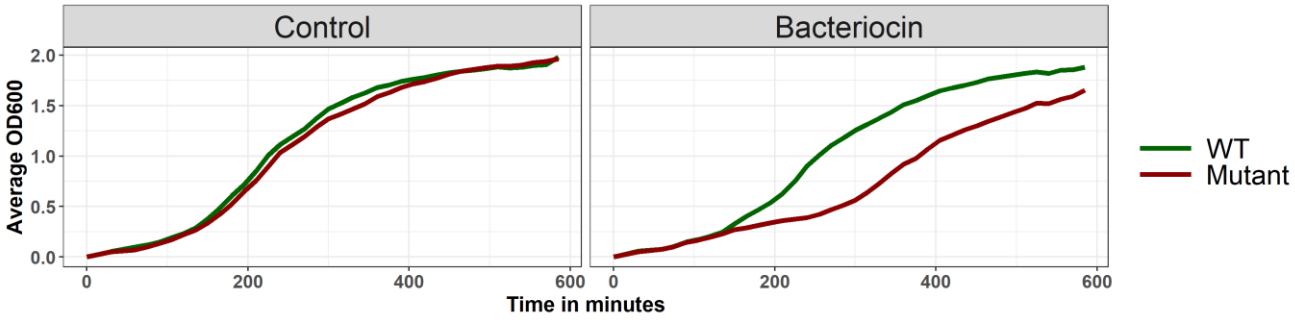


**Figure 4.1.5.2** Growth rate, *S. haemolyticus* LMGT4071 WT and mutant, control and bacteriocin ([GarKS] = 0.0015 mg/mL) samples. For each of the sample group, the growth curves are driven from the average OD600 of four samples. These plots illustrates a comparison in growth rate between WT and mutant. Interpolation between observations was done to compute the mean (the line). The plots are made using the R package ggplot2.

The concentration of GarKS equal to 0.0015 mg/mL was preferred for further analysis. The growth rate of WT and mutant exposed to this concentration of bacteriocin was distinguishable (see **Figure 4.1.5.2**), while not inhibiting too much of the mutant's growth. RNA was isolated and sequenced for the four sample groups: WT control, WT bacteriocin, mutant control and mutant bacteriocin.

## 4.2 Drylab

### 4.2.1 Genomic data

Assembly

Before assembling the genome, the raw data was inspected. **Figure 4.2.1.1** shows the *Per base sequence quality* plots for R1 and R2 from FastQC. A common pattern to see in these plots is lower scores for the first few positions and the last positions in each read. The reason for the quality decreases for the last positions is that the sequencing will get out of sync longer in the sequencing process. In addition, R2 tend to have generally lower quality scores than R1. This is because the reads from R2 are sequenced in the second round where the flow cell is worn. The quality scores appear evenly good for each position in all reads, both for R1 and R2. As seen in the figure, the quality scores have not decreased from R1 to R2, but decreases slightly for the last positions in both cases.

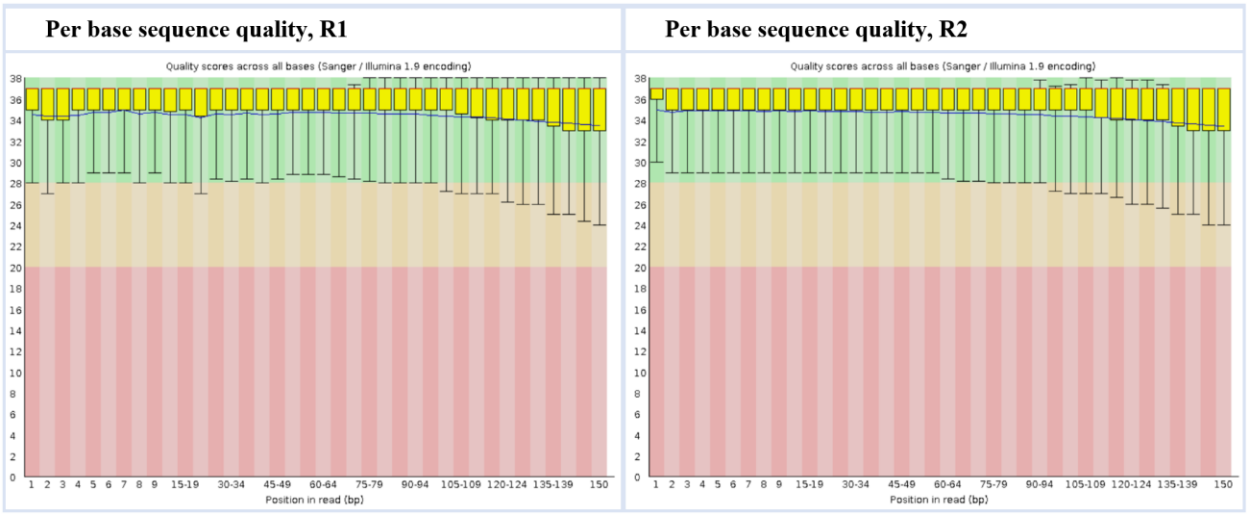


**Figure 4.2.1.1**: Results from FastQC's *Per base sequence quality* – module on the raw data of the sequenced *S. haemolyticus* genome, R1 on the left side and R2 on the right. This plot shows how the quality scores are distributed at each position across all reads, where the given position in the read is on the x-axis and quality scores on the y-axis. The colour coding for the quality scores represents; red = low, yellow = medium, green = high. The red lines are the medium quality score, while the blue is the average. The yellow boxes at each position are the inner-quartile from the 25$^{th}$ to 75$^{th}$ percentile (21).

The log file generated by running BBduk reported 11 153 912 input reads and 11 153 820 output reads. The number of KTrimmed (trimming the matched K-mer and all bases to the right), was 10 218 while 34 140 reads were trimmed by overlap (trimming adapters based on overlap in the read pairs). The number of unpaired reads were found by inspecting the unpaired read file that the software provides. 44 of the 92 reads removed by BBduk were

removed due to loss of their read pair. The filtered FastQ-files were run through FastQC. This report gave approximately the same results as for the raw data. The read lengths ranged from 40-150 bases.

The assembly of the LMGT4071 WT genome by SPAdes resulted in 5117 contigs. After filtering the contigs based on contig length and K-mer coverage, there were 65 contigs left, and these contigs were evaluated through QUAST. The remaining 65 contigs summed to 2379000 bp and the reported GC content was 32.72%. Cumulative length plots from the QUAST report with four of the RefSeq *S. haemolyticus* genomes as references is shown in **Figure 4.2.1.2**. The reason for comparing the assembled LMGT4071 WT genome to these four genomes, was to have reference genomes with diversity in genome size and number of fragments. Information of the references and their alignment with LMGT4071 is in **Table 4.2.1.1**, the order of the RefSeq genomes corresponds to plot A-B.

**Table 4.2.1.1** Information from the QUAST report about the genomes used as references and their alignment to the assembly of LMGT4071 WT. The genome fraction is number of aligned bases in the reference genome.

| RefSeq accession | Genome size (bp) | # Fragments | GC content | Genome fraction | Largest alignment | Total aligned length |
|---|---|---|---|---|---|---|
| GCF_009932885.2 | 2396207 | 6 | 32.69% | 87.511% | 99348 | 2097111 |
| GCF_017355045.1 | 2382565 | 3 | 32.75% | 86.98% | 92050 | 2073628 |
| GCF_019048025.1 | 2568751 | 3 | 32.88% | 87.31 % | 249630 | 2244364 |
| GCF_000972725.1 | 2561368 | 1 | 32.72 % | 88.401% | 178747 | 2265499 |

All RefSeq genomes and LMGT4071 has the same GC content, while the number of bases in the genome and number of fragments varies. The sum of the LMGT4071 contigs length is 3,565 bp less than the smallest of the other genomes.

As seen by the genome fraction and the total aligned length in the table, and in the cumulative length plots, the LMGT4071 contigs covers approximately the same proportion of each genome. The number of contigs needed to cover at least half of the RefSeq genomes seems to be approximately eight.

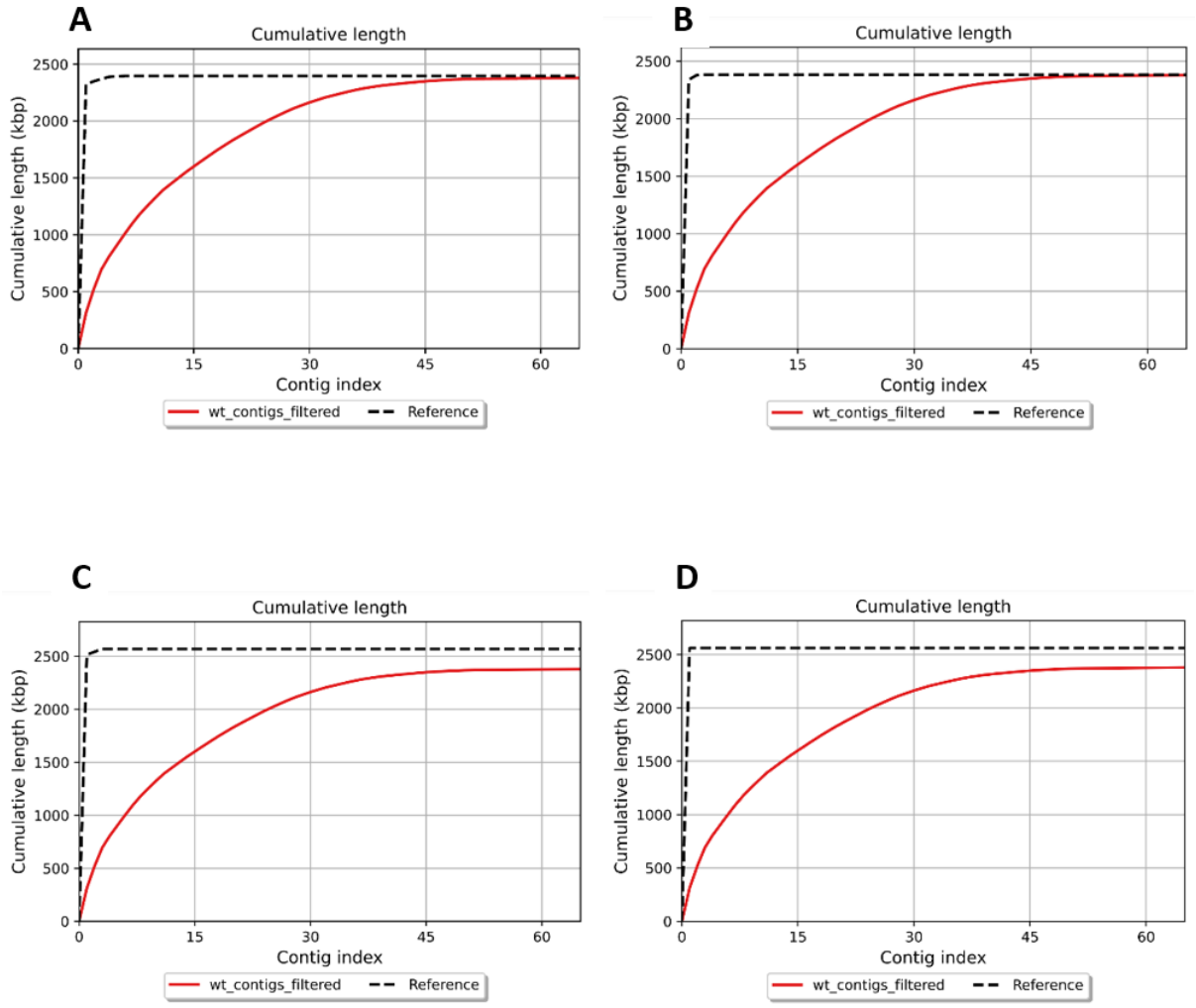


**Figure 4.2.1.2** Cumulative length plot for evaluating the LMGT4071 WT genome against four RefSeq genomes with Quast. **A** genome with RefSeq accession GCF_009932885.2 and assembly accession ASM993288v2. **B** genome with RefSeq accession GCF_017355045.1 and assembly accession ASM1735504v1. **C** genome with RefSeq accession GCF_019048025.1 and assembly accession ASM1904802v1. **D** genome with RefSeq accession GCF_000972725.1 and assembly accession ASM97272v1. These plots show the aligned LMGT4071 WT contigs against the contigs from the RefSeq genomes. In the cumulative length plot, the red line shows the contigs ordered by descending length and indicates number of contigs needed to increasingly cover of the reference genome (dotted line).

A pan genome analysis of the results from Prokka was performed to uncover the similarity of gene content between all the 19 *S. haemolyticus* RefSeq genomes and the LMGT4071 WT genome. As seen in **Figure 4.2.1.3**, LMGT4071 is not separated particularly from the rest of the genomes. Hence, the gene content of LMGT4071 does not appear to deviate from the other genomes. The genome of LMGT4071 is the one with the lowest number of bases with 3565 fewer bases than the 12b strain (**Table 4.2.1.2**).

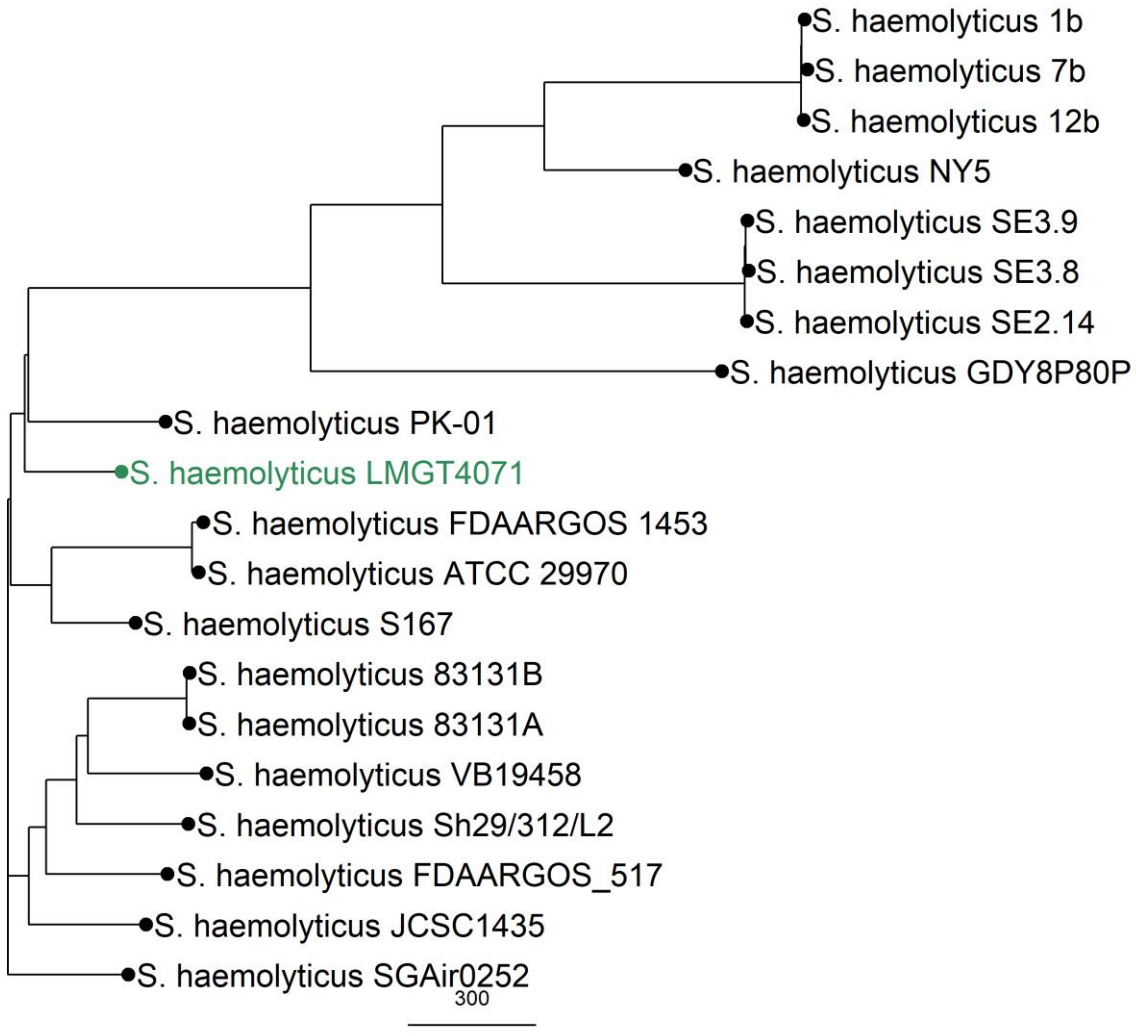


**Figure 4.2.1.3** Tree for the pan genome analysis of the genomes from RefSeq and LMGT4071 WT (marked green). The manhatten distance are used, which gives the number of gene families that differ between the genomes.

**Table 4.2.1.2** Information of RefSeq assembly accession numbers and the genome size including both chromosome and any plasmids (*Number of bases*).

| Strain | RefSeq accession | Assembly name | Number of bases |
|---|---|---|---|
| *S. haemolyticus* JCSC1435 | GCF_000009865.1 | ASM986v1 | 2697861 |
| *S. haemolyticus* Sh29/312/L2 | GCF_000972725.1 | ASM97272v1 | 2561368 |
| *S. haemolyticus* S167 | GCF_001611955.1 | ASM161195v1 | 2560146 |
| *S. haemolyticus* 83131A | GCF_002906595.1 | ASM290659v1 | 2610865 |
| *S. haemolyticus* 83131B | GCF_002906615.1 | ASM290661v1 | 2612666 |
| *S. haemolyticus* SGAir0252 | GCF_002952715.2 | ASM295271v2 | 2674594 |
| *S. haemolyticus* VB19458 | GCF_003596365.3 | ASM359636v3 | 2699210 |
| *S. haemolyticus* FDAARGOS_517 | GCF_003956005.1 | ASM395600v1 | 2669504 |
| *S. haemolyticus* PK-01 | GCF_005706435.1 | ASM570643v1 | 2615683 |
| *S. haemolyticus* ATCC 29970 | GCF_006094395.1 | ASM609439v1 | 2572027 |
| *S. haemolyticus* SE2.14 | GCF_009932715.2 | ASM993271v2 | 2396041 |
| *S. haemolyticus* SE3.8 | GCF_009932885.2 | ASM993288v2 | 2396207 |
| *S. haemolyticus* SE3.9 | GCF_009932935.2 | ASM993293v2 | 2396096 |
| *S. haemolyticus* GDY8P80P | GCF_015169195.1 | ASM1516919v1 | 2771335 |
| *S. haemolyticus* 7b | GCF_017355005.1 | ASM1735500v1 | 2398463 |
| *S. haemolyticus* 1b | GCF_017355025.1 | ASM1735502v1 | 2390652 |
| *S. haemolyticus* 12b | GCF_017355045.1 | ASM1735504v1 | 2382565 |
| *S. haemolyticus* FDAARGOS 1453 | GCF_019048025.1 | ASM1904802v1 | 2568751 |
| *S. haemolyticus* NY5 | GCF_019222685.1 | ASM1922268v1 | 2539247 |
| *S. haemolyticus* LMGT4071 | | | 2379000 |

Annotation

The gene annotation by Prokka resulted in a total of 50 811 predicted genes for all genomes. The number of genes for each of the genomes were computed and are visualized in **Figure 4.2.1.4**. The number of genes in any plasmids from the RefSeq genomes are excluded in this plot. The maximum number of genes in the plasmids is 55. Number of coding genes for each genome ranges from 2220 (NZ_CP084229.1) to 2724 (NZ_CP045187.2). There was 2299 predicted coding genes for LMGT4071. While this genome has fewest number of non-coding genes with 48, the other genomes has twice as many.

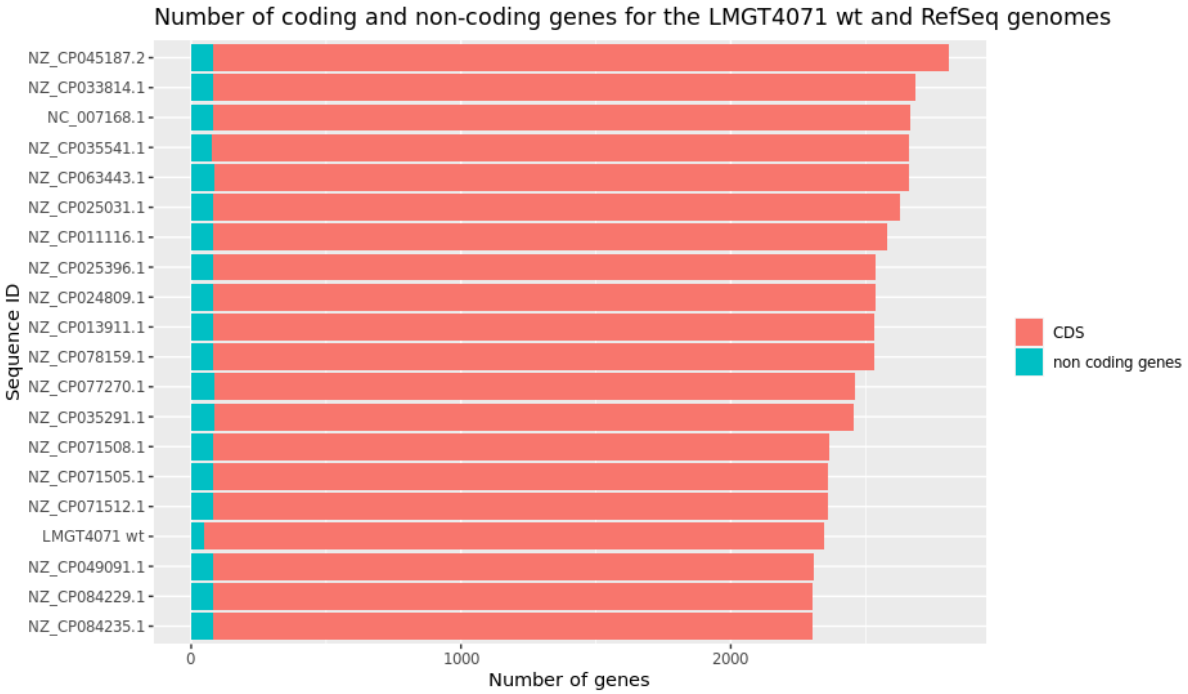


**Figure 4.2.1.4** Number of predicted coding (red) and non-coding (blue) genes by Prokka for LMGT4071 and the RefSeq genomes (excluding predicted genes in any plasmids). The chromosome accession numbers are on the y-axis. The plot is made using the R package ggplot2.

Finding the LORFs in each genome and scanning these against the Pfam database resulted in 256948 hits in total, with 6807 unique descriptions. The scan of the LORFs against the databases in InterProScan resulted in annotations of 661 gene sequences which were not already predicted by Prokka. 57603 of the genes predicted by Prokka and found by the LORFs method, were scanned through InterProScan, resulting in 160570 matches. After merging the GFF tables with annotations from Prokka and InterProScan, the list consisted of 168639 genes, many redundant due to both similar genes across the genomes and several

annotations for each sequence. Number of matches against the InterProScan database for each genome are shown in **Figure 4.2.1.5**. The number of hits for LMGT4071 is around the average.

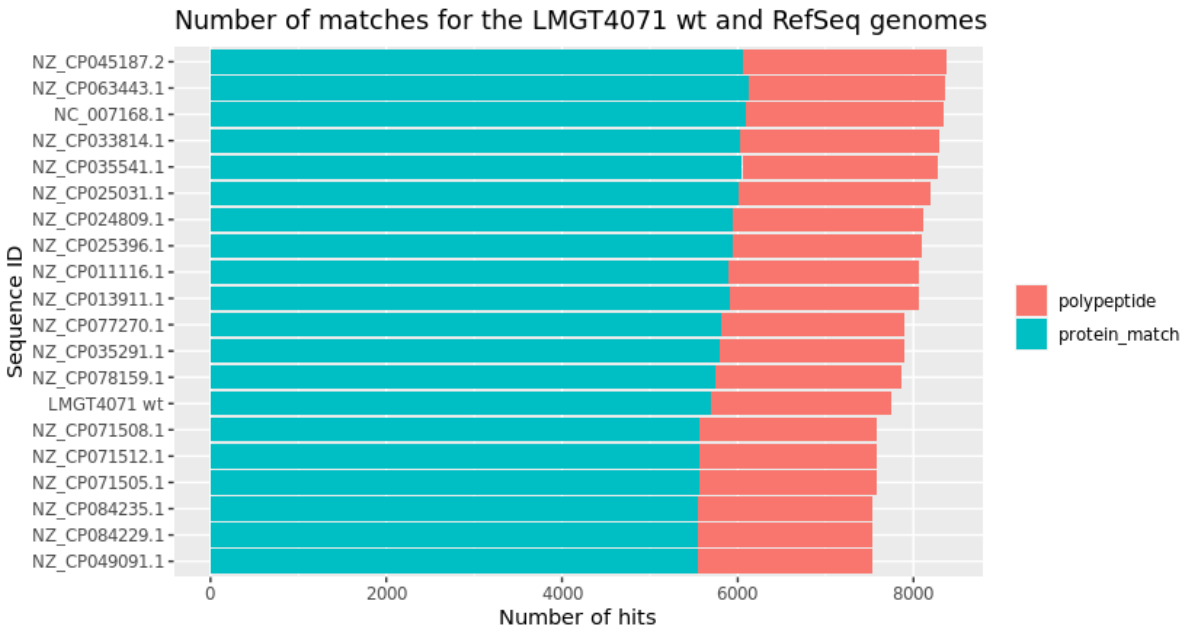


**Figure 4.2.1.5** Number of polypeptide (red) and protein matches (blue) against the InterProScan databases for the predicted genes in LMGT4071 and the RefSeq genomes (without any plasmid genes). Chromosome accession numbers on the y-axis. The plot is made using the R package ggplot2.

### 4.2.2 RNA seq data

Processing RNA-seq reads

The raw data from the sequenced RNA samples and the processed data were evaluated through FastQC. Two of the summary plots from one sample are found in **Figure 4.2.2.1**. The FastQC report of FastQ-files from the other samples is similar to the ones shown in the figure, both R1 and R2. *The Per base sequence quality* after processing reads indicates excellent sequence quality, which was the case also for the raw data. Since it appeared to be non-uniform distribution of A/T and G/C in the raw data, the first 11 bases were trimmed. While the length of the reads from the raw data was 150 bp, their length spanned from 12 to 137 bases after trimming. The number of reads after filtering in each FastQ-file ranged from 9492397-12317376.

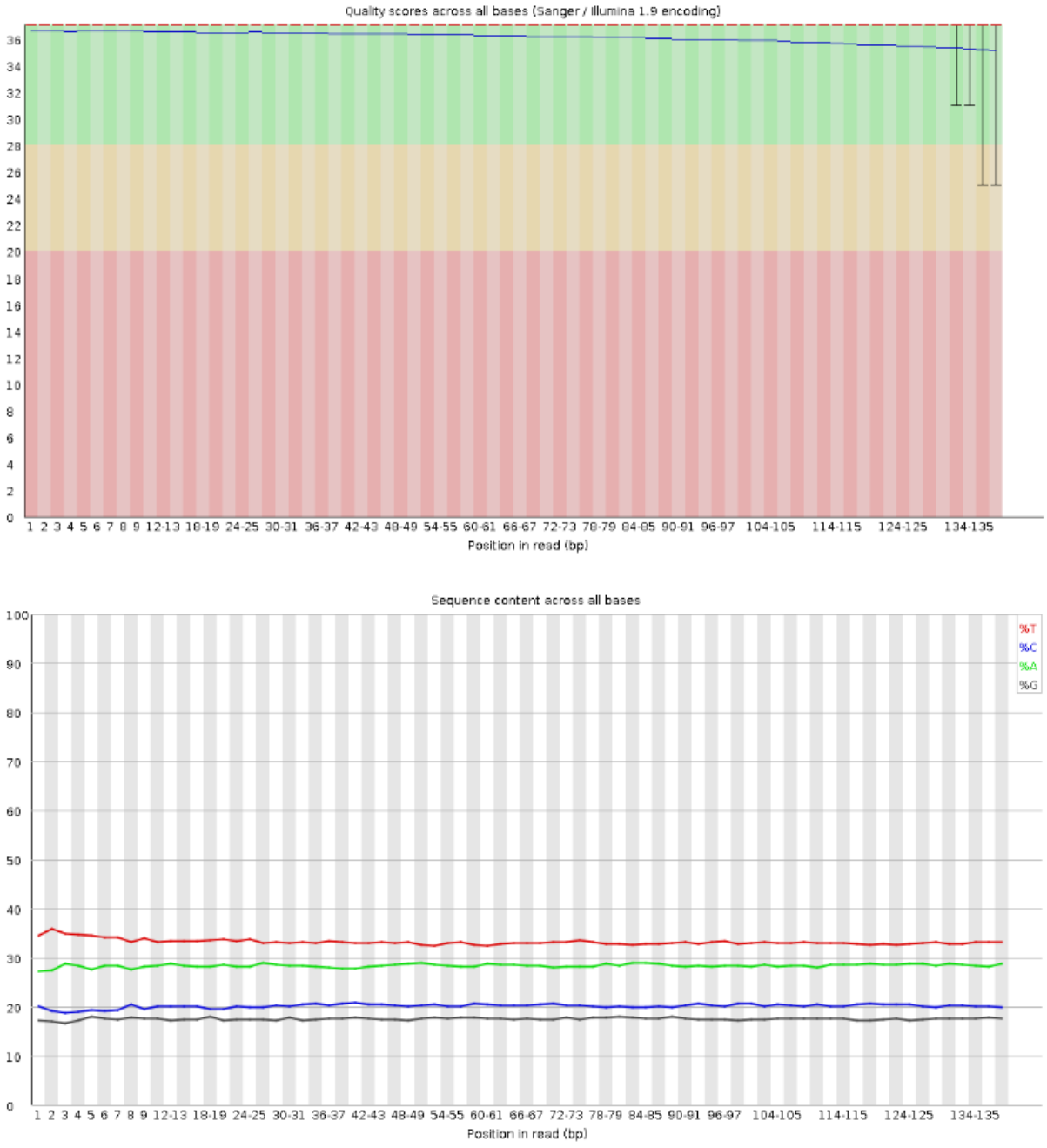


**Figure 4.2.2.1** Results from FastQC report of R1 reads from sequencing of WT control (sequencing name A1) after processing reads with BBduk. *Per base sequence quality* (at the top) shows the distribution of quality scores per position and *Per base sequence content* (at the bottom) the content of each base for every position for all reads.

Read mapping

The reads from all samples were first mapped to an index built from the LMGT4071 predicted genes (idx1) and then to an index built on selected genes from the RefSeq genes as well (idx2). Selected summary of the runs provided by the Kallisto output JSON file are in **Table 4.2.2.1**. Percentage of the reads that mapped in both cases ranged from 72.1 - 84.7%, with 72 - 84.6% of the reads mapping to unique gene sequences. Generally, the total number of pseudoalignments was slightly higher for the mapping using idx2 than idx1. However, for most samples there were more reads which mapped to only one sequence when using idx1. Based on these results the read counts from mapping to idx1 were selected for further work.

**Table 4.2.2.1** Comparison of the reads mapping to the Kallisto index built on LMGT4071 predicted genes (Idx1) and the index built with additional genes from the RefSeq genomes (Idx2). The number of genes is 2348 in Idx1 and 5731 in Idx2. *# Processed* is the number of reads used as input for mapping to the Kallisto index, while *# Pseudoaligned* is the number of reads that was aligned and *# Unique* is the number of reads which mapped to one single gene sequence.

| Sample ID | Sample Description | # Processed | # Pseudoaligned | | # Unique | |
|---|---|---|---|---|---|---|
| | | | Idx1 | Idx2 | Idx1 | Idx2 |
| A1 | WT control | 10327964 | 8066225 | 8066004 | 8065858 | 8056029 |
| A2 | Mutant control | 10954674 | 8835680 | 8837990 | 8835315 | 8827201 |
| A3 | WT bacteriocin | 9888830 | 8277312 | 8281291 | 8276955 | 8270201 |
| A4 | Mutant bacteriocin | 9492397 | 7569360 | 7611283 | 7569067 | 7594193 |
| B1 | WT control | 12317376 | 9130578 | 9126166 | 9130076 | 9116201 |
| B2 | Mutant control | 11998300 | 8649260 | 8645850 | 8648911 | 8636847 |
| B3 | WT bacteriocin | 10784816 | 9128897 | 9129435 | 9128468 | 9118723 |
| B4 | Mutant bacteriocin | 10334965 | 8018147 | 8075729 | 8017828 | 8061553 |

To get a knowledge of how much of the reads that corresponded to non-coding genes (rRNA, tRNA etc.), the read count for each gene type was inspected. **Figure 4.2.2.2** illustrates the variation of proportion of non-coding genes between the eight sequenced samples. In general, the proportion of non-coding genes are larger for all control samples than the samples exposed to bacteriocin. Observing the latter, the mutant samples appear to have slightly larger proportions of non-coding genes than the WT.

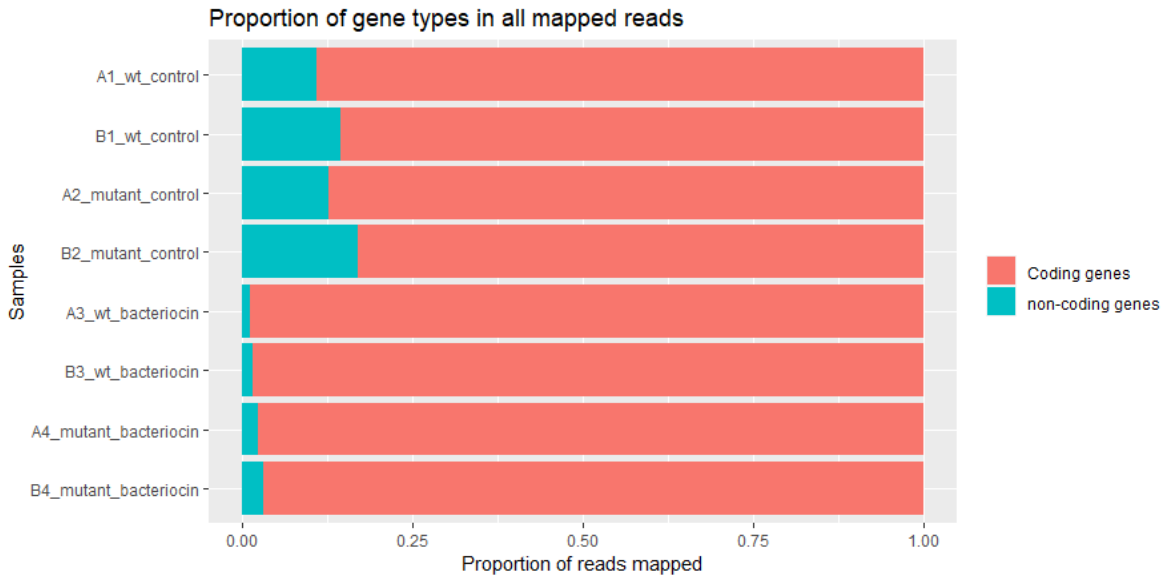


**Figure 4.2.2.2** Proportion of coding and non-coding genes mapped to the predicted LMGT4071 genes for each sample (y-axis), all control samples at the top four and samples exposed to GarKS below. The majority of non-coding genes were rRNA. The predicted genes not categorized as either coding or non-coding genes, eight instances, are excluded. The plot is made using the R package ggplot2.

Looking at the raw read counts for both replicates of WT control the mean was 3254.64 and the median was 905. One gene sequence had zero mapped reads across all samples and three genes had zero mapped reads for at least one sample.

To examine if the read counts for each gene in the samples can reflect any difference between the samples, the counts were used to cluster the samples according to similarity in their gene expression patterns. The clustering of samples was done by generating both a dendrogram from hierarchical clustering of the samples (see **Figure 4.2.2.3)** and a PCA plot through dimension reduction by using the *plotPCA* function from the DESeq R package (see **Figure 4.2.2.4**). In the dendrogram, both the mutant samples exposed to bacteriocin forms a clade separated from all other samples. Except for WT control replicate A1, the WT and mutant control appear in a separate clade.

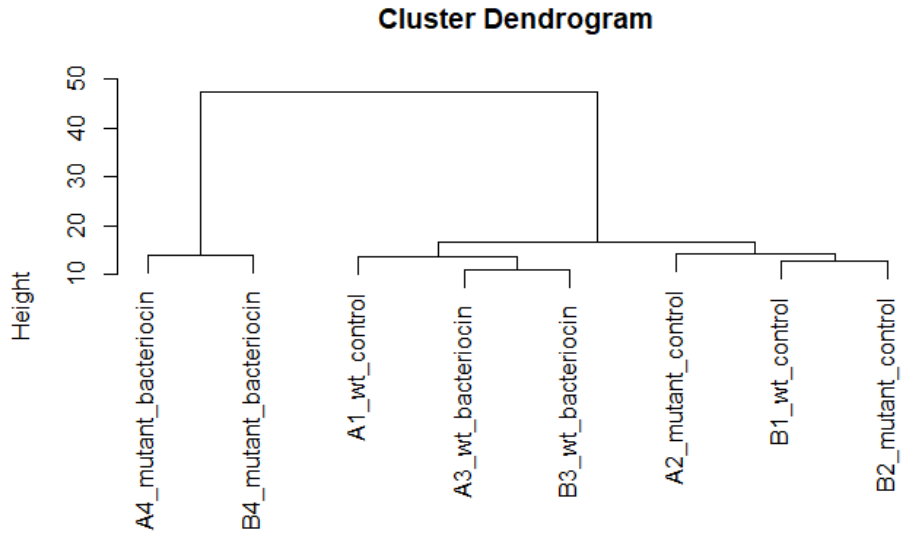


**Figure 4.2.2.3** Hierarchical clustering of the samples based on the read counts normalized through variance Stabilizing Transformation (vst) and complete linkage using Euclidean distances.

The PCA plot shows that the first principal component, that explains most of the variance, separates the mutant bacteriocin (red) samples from the rest. The other replicated samples also cluster together. The second principal component separates the replicates of equal samples, a larger distance within the mutant bacteriocin samples and the WT control samples.

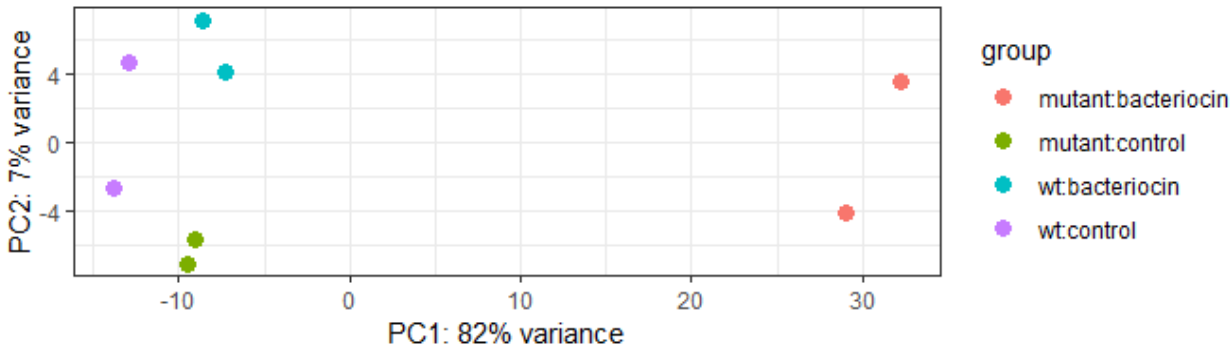


**Figure 4.2.2.4** PCA plot gained from the plotPCA function from DESeq R package. The two principal components which explains most of the variance are plotted, the first explaining 82% (x-axis) and the second explaining 7% (y-axis) of the variation. All 2305 genes used as input in the differential expression analysis by DESeq contribute to the plot. The read counts are transformed by vst.

### 4.2.3 Further analysis of RNA-seq data

Differential expression analysis

After filtering out non-coding genes and genes with zero mapped reads, 2305 genes in total were further analysed. The differential expression analysis by DESeq2 gave results from five contrasts:

- WT bacteriocin vs WT control: the effect of bacteriocin in the WT
- mutant control vs WT control: the effect of bacteriocin in the mutant
- mutant bacteriocin vs mutant control: the general effect of the mutant's genotype
- mutant bacteriocin vs WT bacteriocin: the difference between the genotype when exposed to GarKS
- interaction term: related to how the genotypes respond to the bacteriocin

The number of significant differential expressed genes (DEGs) obtained from each contrast are listed in **Table 4.2.3.1**, and MA-plots from the former four contrasts are in **Figure 4.2.3.1**.

The greatest numbers of DEGs are reported from comparing the mutant exposed to bacteriocin against mutant control and WT bacteriocin. 330 of the DEGs from mutant versus WT bacteriocin is upregulated, while 119 are downregulated in the mutant. 333 DEGs were upregulated and 104 downregulated in mutant bacteriocin compared to mutant control. All 19 DEGs were upregulated in WT bacteriocin when compared to WT control. Nine genes were found as upregulated and five genes as downregulated in the mutant control compared to WT control.

**Table 4.2.3.1** Number of DEGs from each of the comparison groups reported from the analysis through the R package DESeq2. Only the significant results with q-value less than 0.05 are included. Number of significant instances that simultaneously had an absolute log2 Fold Change value above 1 are shown in the column to the right, noted "up" for number of upregulated genes and "down" for downregulated with respect to the first sample specified in the *Comparison* column.

| Comparison | # DE genes (q-value<0.05) | # DE genes (q-value<0.05 and -1 > FC > 1) |
|---|---|---|
| WT bacteriocin vs WT control | 43 | 19 (19 up) |
| Mutant control vs WT control | 60 | 14 (9 up, 5 down) |
| Mutant bacteriocin vs Mutant control | 1293 | 437 (333 up, 104 down) |
| Mutant bacteriocin vs WT bacteriocin | 1293 | 449 (330 up, 119 down) |
| Interaction term | 815 | 341 (274 up, 67 down) |

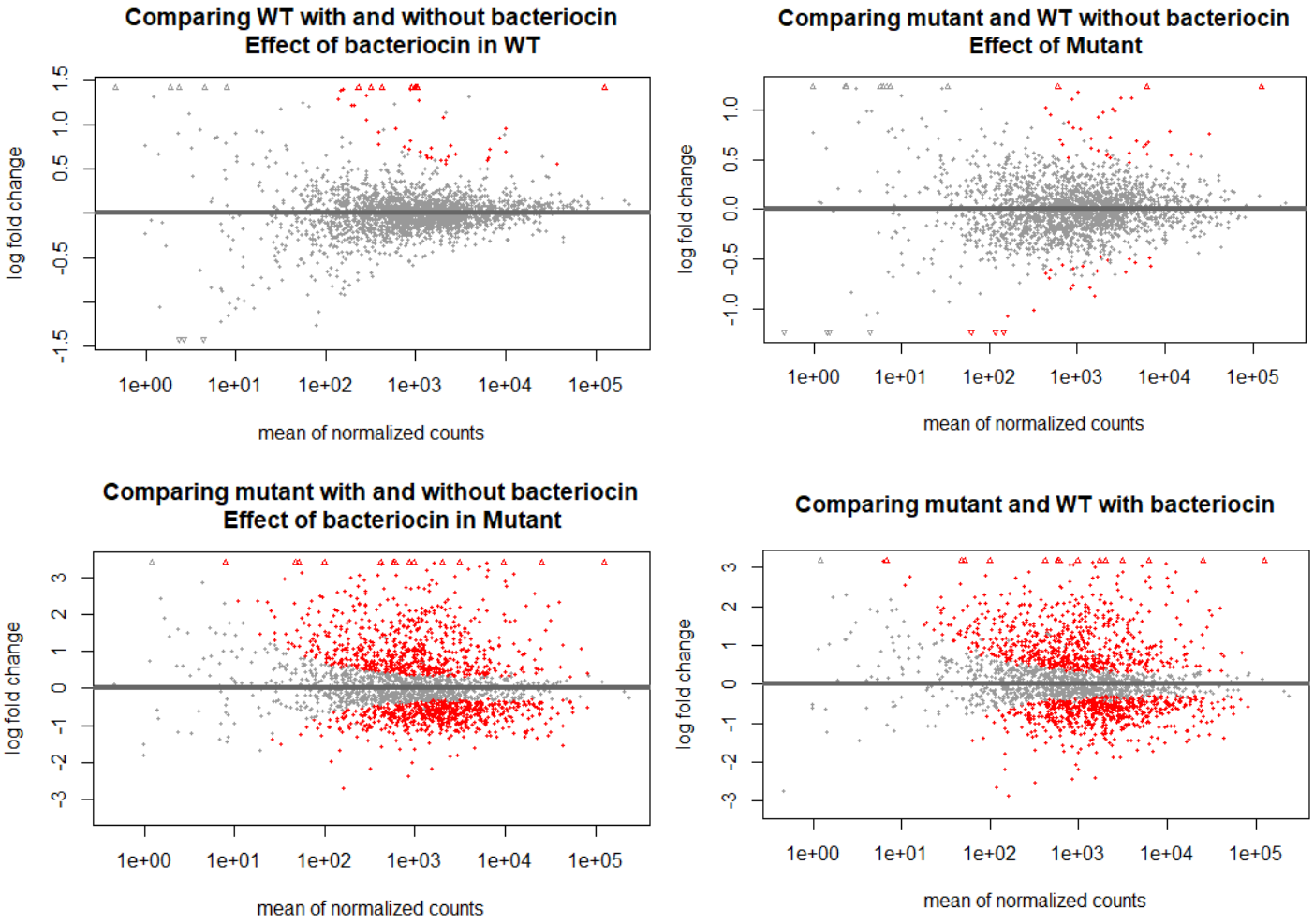


**Figure 4.2.3.1** Plots from the *plotMA* function in the DESeq2 R package. **Top left**: effect of GarKS in the WT. **Top right**: effect of mutant without GarKS. **Bottom left**: effect of GarKS in the mutant. **Bottom right**: effect of mutant with GarKS. From each of the four contrasts, the mean of normalized read counts for all samples (x-axis) are plotted against the log2 Fold Change value in the given contrast. The significant DEGs (q-value < 0.05) are marked red and points outside the limitation of the y-axis are marked with triangles.

The MA-plots in **Figure 4.2.3.1** illustrates how different the outcome of the contrasts is. The significant DEGs, with q-value less than 0.05, are marked red. These results indicate that the gene expression in WT control, WT bacteriocin and mutant control samples are not that different, while the mutant bacteriocin samples deviates from the other samples.

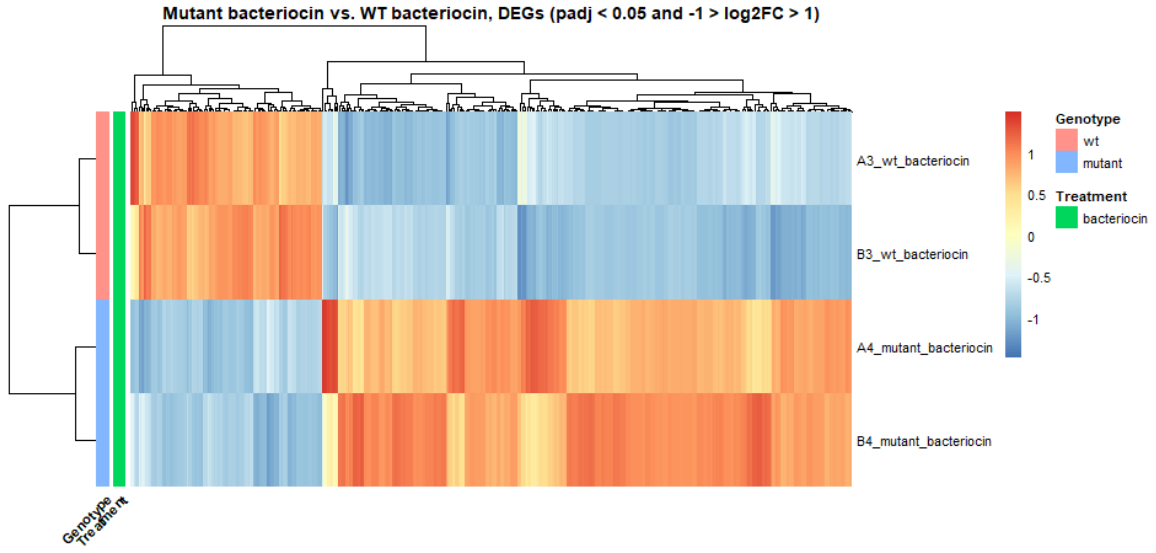


**Figure 4.2.3.2** Heatmap from the *pheatmap* function in the pheatmap R package. Clustering of the samples (rows) and selected DEGs (columns) from comparing mutant bacteriocin against WT bacteriocin. The DEGs that are included, 449, have a q-value less than 0.05 and absolute log2 Fold Change value greater than 1. The intensity of the colours corresponds to higher (red) or lower (blue) value of log2 Fold Change centred and scaled in the columns.

A heatmap with clustering of both samples and genes are shown in **Figure 4.2.3.2**. This heatmap gives another image of the contrast mutant bacteriocin vs WT bacteriocin as in the lower right MA plot in **Figure 4.2.3.1**. The equal samples are clustered together. The gene expression is similar for the replicates within the sample type, only small differences can be observed. The genes are clustered in two main groups: genes upregulated and genes downregulated in the mutant bacteriocin compared to WT bacteriocin. The heatmap of the contrast mutant bacteriocin vs mutant control is not presented here. Although the DEGs was not completely the same in these two contrasts, the heatmap with the mutant bacteriocin vs mutant control contrast did not contribute to any additional information as it illustrated a similar situation as the heatmap above (already seen in **Figure 4.2.3.1**).

Cluster genes in operons

All predicted genes were attempted to be assigned to groups representing operons. This was done with combinations of different SPAdes modes and BLAST programs. To compare the gene clusters across the methods, a dendrogram was made, see **Figure 4.2.3.3**. According to the y-axis, the largest distance is 0.3, indicating small differences in the clustering across all methods. Aligning with blastn and tblastx within the same assembly method gave similar results. Assembly by metaSPAdes led to slightly different clustering than the other methods.

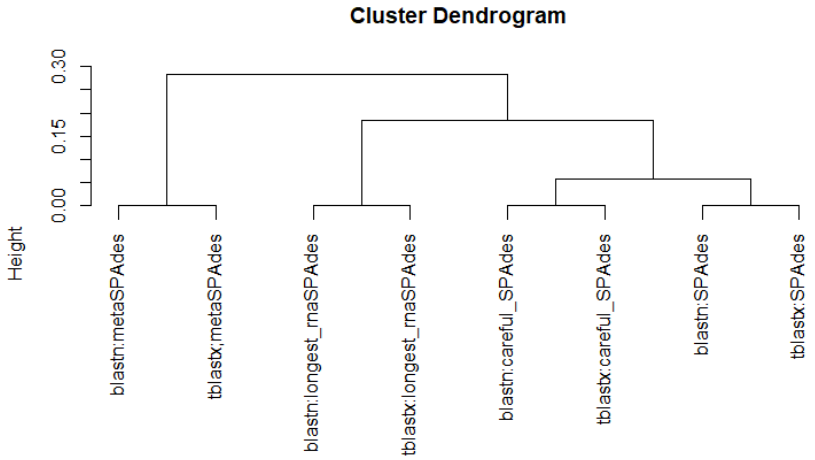


**Figure 4.2.3.3** Hierarchical clustering (complete-linkage) of the results from the attempt to cluster the genes in operons with different methods. The text in each leaf refers to the type of SPAdes mode used (SPAdes, with and without the ---*careful* option, rnaSPAdes and metaSPAdes), and if the predicted genes were aligned with the assembled transcripts through blastn or tblastx. The distances between the grouping methods are derived from the *adj.rand.index* function from the fossil R package.

In **Figure 4.2.3.4** is the results of the gene clustering in operons by assembling the RNA-seq reads with SPAdes (option --*careful*) and then aligning the predicted genes by blastn. The number of genes in each predicted operon varies greatly. While there were instances of clusters with one single gene, ten of the clusters contained more than 40 genes. The grouped genes tended to be on the same strand, in the same contig and adjacent to each other. However, some clusters were spread across contigs and strand or within the contig.

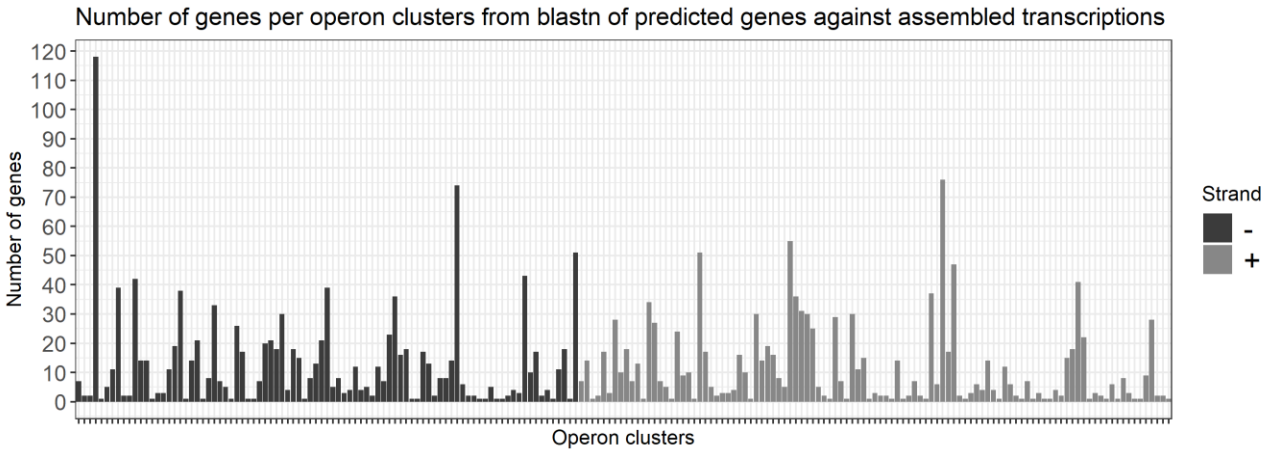


**Figure 4.2.3.4** Overview of number of genes per operon cluster from assembly by SPAdes (--*careful*) and alignment using blastn. The predicted genes are sorted by associated strand and the order of the genes in each contig form the genome assembly (the order of the contigs is not known). The 11 genes that were not assigned to any cluster are excluded.

Inspection of selected differential expressed genes

There was found four genes which were differential expressed in the contrast WT bacteriocin vs WT control that at the same time was not differential expressed in the contrast mutant bacteriocin vs mutant control. The log2 Fold Change for these four genes are presented for different contrasts in **Figure 4.2.3.5**. None of the four genes appear differential expressed when comparing mutant and WT control samples. All genes are downregulated in mutant bacteriocin compared to WT bacteriocin and upregulated in WT bacteriocin compared to WT control. By comparing mutant, with and without bacteriocin exposure, there is no significant difference in gene expression. However, it appears that the gene expression differs more in this contrast compared to the contrast with the control samples.

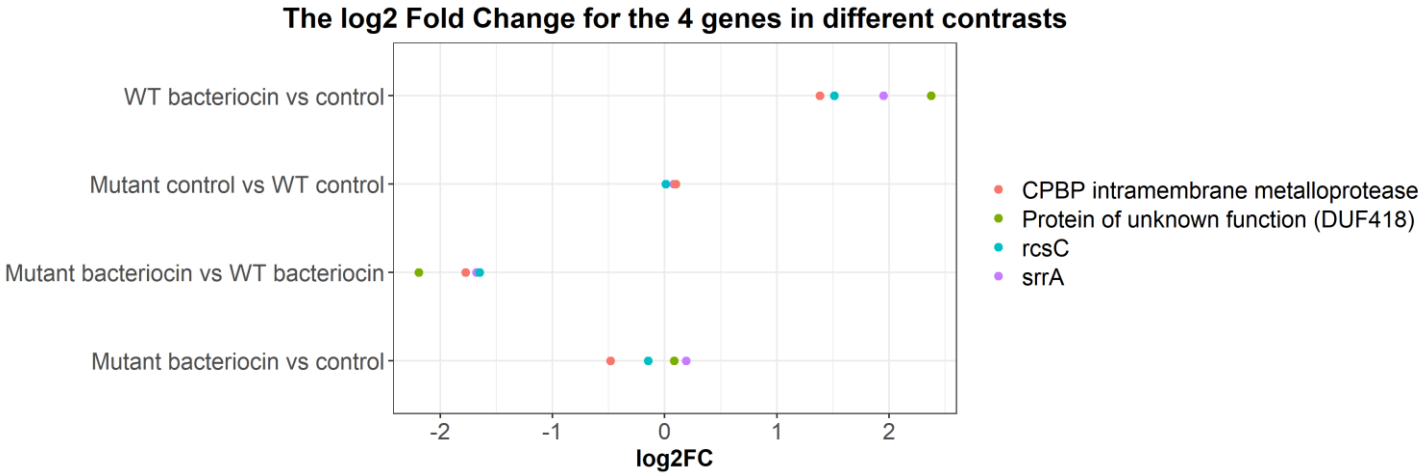


**Figure 4.2.3.5** Log2 Fold Change values (x-axis) for four of the upregulated genes in the contrast: WT bacteriocin vs WT control. Log2 Fold Change values are shown for the four contrasts, from the left: mutant bacteriocin vs to mutant control, mutant bacteriocin vs to WT bacteriocin, mutant control vs to WT control and WT bacteriocin vs to WT control. Annotation of each gene is seen in the legend. The plots are made using the R package ggplot2

The gene predicted as "*srrA"* from Prokka is described as a "Response regulator receiver domain" and "Transcriptional regulatory protein" from InterProScan. This was verified in the BLAST search which matched the gene sequence to a response regulator transcription factor in *S. haemolyticus* (Accession version number: WP_011274925.1).

The gene predicted as *rcsC* from Prokka codes for Sensor histidine kinase. This was verified by the BLAST search (Accession version number: WP_011274924.1). *srrA* and *rcsC* are both predicted to be on the minus strand and were assigned to the same cluster in the operon

prediction. Their predicted distance was -8 and as seen in the figure above, the genes had similar log2 Fold Change values in the four contrasts.

The gene described as “Protein of unknown function (DUF418)” from InterProScan was not predicted by Prokka. No additional information was received from the BLAST search.

The last of the four genes was predicted to be on the plus strand. This gene was not predicted by Prokka. However, the InterProScan description was “CPBP intramembrane metalloprotease”. This was confirmed in the BLAST search (Accession version number: QFR05682.1).

The genes predicted to be located near *rseP* were also inspected. Table with all annotations provided from InterProScan and Prokka for the selected genes are in the appendix (**Table A.3**). **Table 4.2.3.2** shows selected annotations from the gene prediction and log2 Fold Change values from different contrasts for the selected genes.

| Annotation | Distance to next gene | Cluster nr | WT | Mutant | Genotype | Treatment |
|---|---|---|---|---|---|---|
| Cytidylyltransferase family | 210 | 78 | -0.11558 | **-0.49882** | 0.159808 | -0.22343 |
| rseP | 18 | 50 | -0.01136 | -0.0575 | -0.01686 | -0.063 |
| proS | 260 | 50 | 0.090924 | 0.175439 | -0.21386 | -0.12934 |
| polC | 168 | 50 | -0.04769 | **-0.68536** | 0.000716 | **-0.63695** |
| rimP | 20 | 50 | -0.00237 | **1.528688** | -0.25092 | **1.280133** |
| nusA | 25 | 50 | 0.032333 | **0.580363** | 0.030322 | **0.578351** |
| Protein of unknown function (DUF448), YlxR | -4 | 50 | 0.082295 | **0.88679** | -0.03378 | **0.770719** |
| rplGA | 4 | 50 | 0.058315 | **0.875583** | 0.045439 | **0.862708** |
| infB | 161 | 50 | 0.138539 | **0.926438** | -0.00273 | **0.785166** |
| rbfA | 228 | 50 | 0.195096 | **0.960912** | 0.074852 | **0.840667** |
| truB | 14 | 50 | 0.048882 | -0.20295 | -0.01979 | -0.27163 |
| ribF | 113 | 50 | 0.013823 | -0.21011 | -0.04654 | -0.27047 |
| rpsO | 135 | 19 | 0.144547 | **0.519338** | 0.072817 | **0.447608** |

**Table 4.2.3.2** Information of genes predicted to be located near the *rseP* gene. *rseP* is marked green. The log2 Fold Change values for the contrasts are in the four last columns (marked with bold font for genes with q-value < 0.05 from DESeq2). *WT*: WT bacteriocin vs WT control. *Mutant*: mutant bacteriocin vs mutant control, *Genotype*: mutant control vs mutant bacteriocin, *Treatment*: mutant bacteriocin vs WT bacteriocin. *Distance to next gene*: predicted distances between the end of one gene and the start of the next, *Cluster nr:* predicted operon cluster.

None of the genes are significantly differential expressed when comparing WT bacteriocin and WT control or comparing mutant control and WT control. *rseP* is not differential expressed in any of the contrasts. However, *rimP*, which is clustered in the same group as *rseP*, has a q-value < 0.05 and a log2 Fold Change value above 1 in these contrasts. This indicates that this gene is upregulated in the mutant exposed to bacteriocin.

The protein predicted as YlxR and the genes predicted *rplGA*, *infB,* and *rbfA* has a log2 Fold Change near 1 and q-value < 0.05 in the contrasts mutant bacteriocin vs mutant control and mutant bacteriocin vs WT bacteriocin. The three gene's log2 Fold Change are similar within each contrast and they were clustered together in the gene prediction, in the same cluster as *rseP*.

# 5 DISCUSSION

In this master's thesis the phenotype and the genotype of mutant and WT of five bacterial strains have been examined, in addition to inspect the gene expression for *S. haemolyticus* LMGT4071 mutant and WT, with and without exposure to GarKS.

Through this chapter the methods and results from the laboratory experiments will be discussed first, followed by discussion of the assembly, gene annotation process and differential expression analysis for *S. haemolyticus*. The findings in this study for *S. haemolyticus* LMGT4071 will be summarized in the conclusion to reveal if the main aim of this thesis is achieved.

## 5.1 Tolerance to bacteriocin: phenotype and genotype

The laboratory experiments with sequencing of the *rseP* gene, Soft-agar overlay assay and MIC tests were performed for the five bacterial strains *S. haemolyticus* LMGT4071, *L. lactis* IL 1403, *E. faecalis* LMGT3560 and *E. faecium* LMGT7422 and LMGT3160.

The mutation in *E. faecium* LMGT3160 was not revealed by sequencing of the *rseP* gene, while both the results from the Soft-agar overlay assay (**Figure 4.1.2.1**) and MIC test (**Table 4.1.3.5**) indicated that the mutant was more resistant to EntK1 and EntK1 EJ97 than the WT. It is conceivable there could be a mutation located in another gene connected to the function of RseP, which was the case in a previous study performed by LMG on the *rseP* gene in *S. haemolyticus* (14). Here they found that only two of 14 mutants that showed resistance to the fusion of Ent K1 and Ent EJ97 had mutations within the *rseP* gene.

The mutation in the *rseP* gene from the *E. faecium* LMGT7422 mutant resulted in a gene product of 179 aa instead of 422, see **Table 4.1.1.1**. However, as seen in **Figure 4.1.2.1** and **Table 4.1.3.4** this mutant showed similar phenotype as the WT. The active site in RseP is located within the 25 first aa (5). Hence, a possible explanation for the mutant's phenotype is that the functionality of RseP still is sufficient for working as a receptor for EntK1, EntEJ97 and the fusion of these bacteriocins. Nor can the possibility of some mix-up between the two *E. faecium* strains in the laboratory work be eliminated.

As seen in **Table 4.1.1.1**, the mutant of *S. haemolyticus* had one base substitution early in the *rseP* gene that led to prematurely termination and probably a non-functional gene product.

This complies with the Soft-agar overlay assay and MIC-test (**Table 4.1.3.1**) which showed that the mutant was resistant to the leaderless bacteriocins that attacks RseP.

The *L. lactis* mutant had a substitution of one base early in the *rseP* gene, resulting in a prematurely termination. Therefore, it is most likely that the *L. lactis* mutant do not possess any functional RseP. Observing the mutant's phenotype supported this. In the Soft-agar overlay assay, *L. lactis* mutant showed more resistance against the bacteriocins LsbB, EntK1, EntEJ97 and EntK1 EJ97 than the WT which was confirmed in the MIC test (**Table 4.1.3.2**). Surprisingly, there was an inhibition zone for the mutant where the EntK1 with the highest concentration was applied, indicating sensitivity. The experiment should have been repeated to see if this observation can reoccur and is worthy a further inspection.

The insertion of multiple bases in the *rseP* gene of the *E. faecalis* mutant resulted in a gene product of 256 aa. The results from the Soft-agar overlay assay showed that the mutant was resistant to the leaderless bacteriocins that uses RseP as receptor, while the WT was sensitive to EntEJ97 and EntK1 EJ97. Results from the MIC-test, which is presented in **Table 4.1.3.3**, indicates little difference in sensitivity between WT and mutant for EntEJ97, EntK1 EJ97 and GarKS. These results may point towards that the RseP in the mutant is partly functional.

The growth rate for WT and mutants of *S. haemolyticus*, *L. lactis* and *E. faecalis*, both with and without exposure of GarKS, was measured. Results from this are in **Figure 4.1.4.1.** These measurements of growth rate in microtiter plates illustrated three different scenarios. While it was clear that the *S. haemolyticus* WT tolerated GarKS better than the mutant, the growth of *E. faecalis*, WT and mutant, showed no noticeable difference. If any difference between the WT and mutant of *L. lactis*, it pointed to the mutant being less sensitive to GarKS than the WT. As for the experiments mentioned in the previous paragraph, the similarity of growth between WT and mutant of *E. faecalis* substantiates that the RseP in the mutant still is partly functional.

In this study the focus was to examine a mutant's loss of fitness due to a resistance mutation. With this in mind, there were no further experiments that involved *L. lactis* or *E. faecalis*. This because the mutant within each strain did not appear to be more sensitive to GarKS than the WT. Therefore, only the *S. haemolyticus* strain were chosen for further experiments as the mutant showed more sensitivity than the WT to the exposure of GarKS. However, it would have been interesting to also investigate the unexpected results from the growth rate measurements of the *L. lactis*, WT and mutant.

For *S. haemolyticus*, measurements of the growth rate in culture tubes (**Figure 4.1.5.1 and Figure 4.1.5.2**), RNA isolation and RNA sequencing were performed. The goal was to sequence RNA from WT and mutant samples where they were exposed to a concentration of GarKS that could distinguish their growth, while not inhibiting too much of the mutant's growth. Although the growth rate measurements in the microtiter plates signified an appropriate concentration of 0.003125 mg/mL, the bacteria were both more sensitive to this concentration in the culture tubes with higher volume. To make sure the growth pattern was reproducible, measurements of the growth rate in the culture tubes were performed with several concentrations of GarKS.

## 5.2 Genomic data: assembly and annotation

The raw reads from sequencing of *S. haemolyticus* LMGT4071 WT genome were processed using the BBmap software. Only 92 reads were removed by filtering and 44358 were trimmed either by K-mer trimming or trimming adapters based on overlap in the read pairs. As the quality scores were decent in the raw data (see **Figure 4.2.1.1**), the reads were not trimmed based on quality. Although it did not appear that the processing made any major difference, it was useful to have adapter sequences removed. To be even more certain that the reads did not contain any adapter sequences, a certain number of bases from the left could have been trimmed for all reads. It might also have been useful to merge overlapping read pairs before the assembly with SPAdes to save memory.

The 5117 contigs assembled by SPAdes were filtered to remove the contigs with low K-mer coverage. This because low K-mer coverage may indicate that the contig is assembled from reads containing error. As short contigs will most likely not contribute to relevant information in the gene annotation, these were also filtered out. This resulted in 65 contigs and fewer bp than in the genomes from RefSeq. However, the QUAST report comparing the assembled LMGT4071 wt genome against the four RefSeq genomes (**Table 4.2.1.1** and **Figure 4.2.1.2**), and the results from the pan-genome analyses (**Figure 4.2.1.3** and **Table 4.2.1.2**) indicated that our assembly was sufficient for the purpose of this study. To get an even better assembly, it would have been useful if the LMGT4071 mutant had been sequenced as well, allowing assembly based on the reads from both the WT and mutant. In addition, assembling of the WT and mutant genomes separately could have been used to evaluate the assemblies by examine their similarity. This would also have given an opportunity to search for other differences in the mutant's genome in addition to the mutation in the *rseP* gene.

The pan-genome analysis indicates that the annotation of LMGT4071 gave similar results as the annotation of the RefSeq genomes. This reflects that the assembly of LMGT4071 was good. For evaluating the annotation method itself, it may be more appropriate to perform the pan-genome analysis with the annotation information from the NCBI RefSeq database, rather than obtaining the annotations the same way as for the LMGT4071 genome.

The *S. haemolyticus* reference genome from NCBI (Genome assembly accession: ASM161195v1) has 2,523 annotated genes, where 83 of these are non-coding genes. From Prokka, the number of coding genes variated from 2220 to 2724 across the RefSeq genomes, see **Figure 4.2.1.4**. While the number of coding genes in the LMGT4071 was about average, LMGT4071 had only half of the non-coding genes compared to the reference and RefSeq genomes. The reason for this may be that non-coding genes, especially those encoding rRNA, have multiple copies. The reads derived from these copies are often assembled into a single contig in the genome assembly (53).

InterProScan was used to annotate the gene sequences from the LORFs method, and to get additional information for the genes annotated by Prokka. Using these annotation methods combined with annotation of 20 genomes, resulted in a large gene list with both redundant gene sequences and annotations. Most likely, not all these genes were in the LMGT4071 strain. Although the reads were mapped to only the genes found in the LMGT4071 genome, it was valuable to have the additional genes in case the mapping based also on these, gave a better result. The result of the mapping with the additional genes can also be used to evaluate the mapping with the LMGT4071 genes. The similar outcome from the two mapping methods indicates that the mapping with the LMGT4071 genes was applicable.

## 5.3 RNA-seq data – mapping and differential expression analysis

The differential gene expression analysis showed little effect of GarKS in the *S. haemolyticus* WT, while considerable effect in the mutant. As seen in **Table 4.2.3.1** and in the MA-plots in **Figure 4.2.3.1**, the WT only turned on a few genes in response to the exposure to GarKS, while the mutant responded by upregulating a large number of genes (333 genes). This illustrates that the WT has a mechanism that makes it capable to cope with the amount of GarKS efficiently, which the mutant does not have. The heatmap in **Figure 4.2.3.2** illustrates the number of genes that are up-and down regulated in mutant bacteriocin in contrast to WT bacteriocin and conveys a clearer picture of the corresponding MA-plot. The mutant responds to the GarKS exposure mainly by upregulating genes. The genes that are turned on in the WT

bacteriocin and not in the mutant bacteriocin may reveal the mechanism in the WT which is missing in the mutant.

The difference in growth between the WT and mutant samples exposed to GarKS, as seen in **Figure 4.1.5.2**, supports that only the WT have an efficient response. The amount of GarKS inhibit the growth of mutant, but not the WT. However, neither the observation of the gene expression nor the growth measurement transpires what cause the mutant's slow growth. The results from this study does not reveal if the slow growth reflects cell death or if the mutant is too occupied trying to cope with the added stress and therefore is unable to divide as quickly as the WT.

From the PCA plot in **Figure 4.2.2.4** and the dendrogram in **Figure 4.2.2.3** it is evident that the mutant's gene expression deviates greatly from the other samples when exposed to bacteriocin. Under optimal conditions it is expected no significant difference in gene expression between the control samples of WT and mutant, still it was reported 14 DEGs. This can be a result of batch effect in the laboratory or systematic errors from downstream processes, and it is conceivable that the reported DEGs are false positives. As mentioned in the theory chapter, in gene expression analysis it is recommended to have at least three biological replicates to account for the biological variation (30). The samples A1 and B1 (WT control) showed relatively high variation. In the dendrogram with clustering of the eight samples, sample A1 was separated from the other control samples. In the PCA-plot the WT control samples were further apart from each other than the replicates of mutant control and WT bacteriocin. Since there were only two biological replicates in this study, it is difficult to separate the biological variation from systematic variation in the differential expression analysis.

Although the mutant bacteriocin samples were clustered together in the dendrogram, the second principal components in PCA-plot illustrated a larger variation between these replicates. This can be seen as sign of batch effect or that the mutant does not have a specific method of dealing with the stress due to GarKS.

It is important to have in mind that imprecision can occur in each of the processes finally leading to reported DEGs. This is true for all experiments in the laboratory with final RNA isolation, the sequencing of the RNA, how the RNA reads are processed and mapped to the genes and how the gene sequences are obtained.

After the trimming of the reads, none of the short reads were filtered out. As short reads lead to less specific sequences, the chance of reads mapping to the wrong gene increases. It is also possible that a read can be mapped to multiple genes. This can be observed in **Table 4.2.2.1**, by inspecting the difference in the total number of pseudoaligned and the number of reads uniquely pseudoaligned to a gene. By filtering out multimapping reads, which was not done in this study, one could risk losing some information, while at the same time be more confident that the remaining information is more correct. In this case, since the number of multimapping reads is small, it might have been appropriate to filter these out. To avoid too many multimapping reads, the mapping with only the predicted genes from LMGT4071 was chosen. There was little additional information obtained from the mapping based also on the RefSeq genes.

**Table 4.2.2.1** also indicated that some of the reads did not map to any gene sequence. Novogene's quality check of the RNA samples reported contamination of DNA. Although Novogene purified the samples, the chance of some remaining DNA molecules cannot be eliminated. Hence, it is not unlikely that some of the none-mapping reads are reads from DNA. Non-mapping reads can also occur if the gene they originated from was not found through the gene prediction process.

Because the mRNA was not specifically extracted from the RNA samples, the fragments from non-coding genes were also sequenced. **Figure 4.2.2.2** illustrates how much of the sequencing resources that were allocated to these fragments. It may appear as the control samples did express non-coding genes more than the samples with added bacteriocin. However, the figure simply shows the proportions of non-coding and coding genes. Hence, what this figure illustrates is that the expression of coding genes increases in response to bacteriocin. The difference in proportion of non-coding genes between the mutant bacteriocin and WT bacteriocin reflects the high number of upregulated genes in the mutant bacteriocin compared to WT bacteriocin.

Subsequently to the differential gene expression analysis, some of the genes were further inspected. Among these were the genes *srrA* and *rcsC* which were clustered together in the operon prediction. The analysis showed that the WT, and not the mutant, upregulated *srrA* and *rcsC* when exposed to bacteriocin (see **Figure 4.2.3.5**), it may be that the regulation of these genes is somehow connected to the function of RseP.

The *srrA* and *rcsC* gene's similarity in log2 Fold Change across the contrasts and the predicted distance of -8 bases supports the prediction that they are in the same operon. If *srrA* and *rcsC* are organized in the same operon, it is possible that they are functionally related or connected to a common pathway as stated in the theory section about operons. For other bacteria, studies have shown that *srrA* and *rcsC* are related to a Two-Component System. SrrA is part of the SrrAB Two-Component System (54). Research by Nadim Majdalani and Susan Gottesman, cited in "New Insights into the Non-orthodox Two Component Rcs Phosphorelay System" (55) shows that RcsC is part of the Two-Component Rcs phosphorelay System. The Two-Component Systems are related to the bacteria's response to environmental changes (56). Hence, it is sensible that the WT upregulates both these genes in response to the exposure of GarKS.

**Table 4.2.3.2** showed that the *rseP* gene was not found differential expressed in any of the contrasts. This result was not as expected since RseP is found to be related to the stress response in bacteria (5). Some of the genes clustered together with *rseP* in the operon prediction were reported either significantly differential expressed (*rimP*) or nearly significantly differential expressed (*ylxR*, *rplGA*, *infB*, and *rbfA*) in the contrasts mutant bacteriocin vs mutant control and mutant bacteriocin vs wt bacteriocin. These genes were five of many that the mutant upregulated in response to GarKS. Although the operon prediction clustered all five genes together with *rseP*, *rseP* was not upregulated in the mutant exposed to GarKS. This points towards that these genes are not transcribed together.

## 5.4 Operon prediction

In this study, it was attempted to cluster the genes based on assembling the reads and align the gene sequences to the assembled reads. Each of the contigs from the assembly should represent one transcript. In theory, this seemed to be a reasonable approach and as seen in **Figure 4.2.3.3**, the eight clustering methods gave similar results which substantiates the clustering. However, some of the transcripts from SPAdes were much longer than expected (see **Figure 4.2.3.4**). One possible explanation of long transcripts is that several transcripts connect due to reads from DNA fragments if these are present. In addition, the precited distances between the adjacent genes and the log2 Fold Change from DESeq2 did not harmonize with the gene clusters as expected. However, the prediction of the start of the gene sequence is uncertain (57). Hence, it is not given that the calculated distances between the genes are correct. The observed variation in gene expression between genes clustered

together seems to point towards wrong clustering. However, the expression for the genes in the same operon are reported not to be as consistent (58) as expected based on the fact that the genes are transcribed together.

While the clustering of genes performed in this study does not necessarily provide correct information on which genes the operon consists of, it may provide information on which of the genes that do not belong in the same operon.

## Conclusion

The main aim of this thesis was to examine how a mutant of a bacterium regulates its gene expression when exposed to stress, compared to the WT within the same strain. After confirming the mutation in the *rseP* gene in the mutant, the phenotype of *S. haemolyticus* LMGT4071 was studied in the laboratory for both the mutant and the WT. It was found one mutation early in the *rseP* gene in the mutant, and the mutant showed resistance to the leaderless bacteriocins that uses RseP as receptor. The WT showed sensitivity to EntK1 EJ97. While the growth of the mutant did not differ from the growth of the WT in optimal conditions, the slower growth of the mutant exposed to GarKS showed that the mutant was more sensitive than the WT to this bacteriocin. These results point out that the function of RseP is somehow important for the stress response in *S. haemolyticus*.

The inspection of the transcriptome in *S. haemolyticus* LMGT4071 showed that the WT upregulated a few genes in response to the GarKS exposure. *srrA* and *rcsC* were two of the genes upregulated in the WT and not in the mutant. This may indicate that the function of RseP is related to the regulation of these genes. The mutant altered its gene expression considerably in response to the GarKS exposure. Altogether this study shows that the mutation in the mutant which makes it resistant to the EntK1 EJ97, comes with a cost as the mutant lost some mechanism related to its stress response.

## Further work

The results from this master's thesis showed that the *S. haemolyticus* LMGT4071 mutant was more sensitive than the WT to the GarKS exposure which may be related to the mutation in the *rseP* gene in the mutant. To support this result, the genome of the mutant can be sequenced to eliminate the possibility of another mutation in the mutant's genome that could affect the stress response.

The transcriptome analysis in this study was based on a differential gene expression analysis with further inspection of a few genes. Further research with another approach, like for instance a co-expression analysis, can contribute to a broader understanding of how the mutant regulates its gene expression in response to stress.

It was attempted to uncover the operon structure in *S. haemolyticus* LMGT4071 by clustering the genes based on mapping the genes to assembled transcripts from the RNA reads. A more accurate approach for operon prediction and more research related to the gene expression within an operon is desirable. Combining several approaches for operon prediction could be attempted; utilizing the pair-end read information as in the two studies (33-34) mentioned in the theory chapter, mapping the genes to assembled transcripts as attempted in this study, and accounting for the gene function and the distance between adjacent genes (32).

# 6 REFERENCES


1. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. PNAS. 2016;113(21):5970-5975.
2. Tang L. Culturing uncultivated bacteria. Nat Methods. 2019;16:1078.
3. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. COMICR. 2019;51:72-80.
4. Simons A, Alhanout K, Duval RE. Bacteriocins, Antimicrobial Peptides from Bacterial Origin: Overview of Their Biology and Their Impact against Multidrug-Resistant Bacteria. Microorganisms. 2020;8(5):639.
5. Aleksandrzak-Piekarczyk T, Diep DB, Jensen MS, Kristiansen PE, Nes IF, Ovchinnikov KV et al. The Leaderless Bacteriocin Enterocin K1 Is Highly Potent against Enterococcus faecium: A Study on Structure, Target Spectrum and Receptor. Front. Microbiol. 2017;8:774.
6. Dykes GA, Hastings JW. Fitness costs associated with class IIa bacteriocin resistance in Listeria monocytogenesB73. Lett Appl Microbiol. 1998;26(1):5-8.
7. Rodnina MV. Translation in Prokaryotes. Cold Spring Harb Perspect Biol. 2018;10(9).
8. Osbourn AE, Field B. Operons. Cell Mol Life Sci. 2009;66(23):3755-3775.
9. Czekaj T, Ciszewski M, Szewczyk EM. Staphylococcus haemolyticus – an emerging threat in the twilight of the antibiotics age. Microbiology Society. 2015;161(11):2061-2068.
10. Kleerebezem M, Bachmann H, Pelt-KleinJan E, Douwenga S, Smid EJ, Teusink B et al. Lifestyle, metabolism and environmental adaptation in Lactococcus lactis. FEMS Microbiology Reviews. 2020;44(6):804–820.
11. Jabbari Shiadeh SM, Pormohammad A, Hashemi A, Lak P. Global prevalence of antibiotic resistance in blood-isolated Enterococcus faecalis and Enterococcus faecium: a systematic review and meta-analysis. Infect Drug Resist. 2019;2019(12):2713-2735.
12. Kayo K, Maegawa S, Ito K, Akiyama Y. Environment of the Active Site Region of RseP, an Escherichia coli Regulated Intramembrane Proteolysis Protease, Assessed by Site-directed Cysteine Alkylation. JBC. 2007;282(7):4553-4560.
13. Yang S-C, Lin C-H, Sung CT, Fang J-Y. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. Front. Microbiol. 2014;5:241.
14. Kranjec C, Kristensen SS, Bartkiewicz KT, Brønner M, Cavanagh JP, Srikantam A et al. A bacteriocin-based treatment option for Staphylococcus haemolyticus biofilms. Sci Rep. 2021;11(1):13909.
15. Dubey S, Diep DB, Evensen Ø, Munang'andu HM. Garvicin KS, a Broad-Spectrum Bacteriocin Protects Zebrafish Larvae against Lactococcus garvieae Infection. Int J Mol Sci. 2022;23(5):2833.
16. RAMESH R, MUNSHI A, PANDA SK. Polymerase chain reaction. Natl Med J India. 1992;5(3):115-119.
17. Hockett KL, Baltrus DA. Use of the Soft-agar Overlay Technique to Screen for Bacterially Produced Inhibitory Compounds. JoVE. 2017;(119):55064.

18. Andrews JM. Determination of minimum inhibitory concentrations. Journal of Antimicrobial Chemotherapy. 2001;48(1):5-16.
19. Aryal S. Bacterial Growth Curve Protocol [Internet]. The Biology Notes: Sagar Aryal; 2018 [cited 2022 June 17]. Available from: https://thebiologynotes.com/bacterial-growth-curve-protocol/
20. Lundeberg J. DNA, RNA, and Protein Extraction: The Past and The Present. BioMed Research International. 2009;2009(574398):2314-6133.
21. Harvard Chan Bioinformatics Core (HBC). Introduction to RNA-Seq using high-performance computing [Internet]. GitHub Pages: hbctraining; [cited 2022 March 13]. Available from: https://hbctraining.github.io/Intro-to-rnaseq-hpc-salmon/lessons/qc_fastqc_assessment.html
22. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 2012;19(5):455-477.
23. Compeau PE, Pevzner PA, Tesler G. How to apply de Bruijn graphs to genome assembly. Nat Biotechnol. 2011;29(11):987-91.
24. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14): 2068- 2069.
25. Isenbarger TA, Carr CE, Johnson SS, Finney M, Church GM, Gilbert W et al. The Most Conserved Genome Segments for Life Detection on Earth and Other Planets. Orig Life Evol Biosph. 2008;38(6):517-533.
26. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. The microbial pan-genome. Current opinion in genetics & development. 2006;15(6):589-594.
27. Bray N, Pimentel H, Melsted P, Pacther L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–527.
28. Chan FC. Fong Chun Chan's Blog [Internet]. GitHub; Fong Chun Chan. 2015. How "Pseudoalignments" Work in Kallisto; 2015 Sep 2 [cited 2022 July 8]. Available from: https://tinyheero.github.io/2015/09/02/pseudoalignments-kallisto.html
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
30. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V et al. Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment. arXiv. 2015.
31. Zhou Y, Zhu J, Tong T, Wang J, Lin B, Zhang J. A statistical normalization method and differential expression analysis for RNA-seq data between different species. BMC Bioinformatics. 2019;20(1):1471-2105.
32. Taboada B, Estrada K, Ciria R, Merino E. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. Bioinformatics. 2018;34(23):4118–4120.
33. Irla M, Neshat A, Brautaset T, Rückert C, Kalinowski J, Wendisch VF. Transcriptome analysis of thermophilic methylotrophic Bacillus methanolicus MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. BMC Genomics. 2015;16(1):73.
34. Pfeifer-Sancar K, Mentz A, Rückert C. Kalinowski J. Comprehensive analysis of the Corynebacterium glutamicum transcriptome using an improved RNAseq technique. BMC Genomics. 2013;14(888):1471-2164.

35. Fursov M, Golosova O, Okonechnikov K, the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28(8):1166-1167.
36. Huang X, Madan A. CAP3: A DNA sequence assembly program. Genome Res. 1999;9(9)868-877.
37. Dunstan RH, Gottfries J, Onyango LA, Roberts TK, von Eiff C. Effect of low temperature on growth and ultra-structure of Staphylococcus spp. PLoS One. 2012;7(1):e29031.
38. Ji X, Li W, Song F, Wei Y, Xiang Y, Yang X et al. Biological characteristics and whole-genome analysis of the Enterococcus faecalis phage PEf771. Canadian Journal of Microbiology. 2020;66(9):505-520.
39. Dorau R, Chen J, Liu J, Jensen PR, Solem C, Ercolini D. Adaptive Laboratory Evolution as a Means To Generate Lactococcus lactis Strains with Improved Thermotolerance and Ability To Autolyze. Applied and Environmental Microbiology. 2021;87(21):e01035-21.
40. Danial, EN. Al-Zahrani SHM, Al-Mahmoudi ZAM. Enhancement of novel extracellular bacteriocin production by media optimization using LAB isolate from meat. Journal of Applied Pharmaceutical Science. 2016;6(12):020-027
41. RStudio Team. RStudio: Integrated Development for R. RStudio. 2022.2.3.492 [software]. Boston, MA: PBC; 2020 [cited 2022 July 11]. Available from: http://www.rstudio.com/
42. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 0.11.9--0 [software]. 2010 [cited 2022 July 11]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
43. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 38.86--h1296035_0 [software]. Lawrence Berkeley National Laboratory: LBNL; 2014 [cited 2022 July 11]. Available from: https://escholarship.org/uc/item/1h3515gn
44. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142-i150.
45. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(34.0):D222-D230.
46. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7(10): e1002195.
47. Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236-1240.
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
49. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3693.
50. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–3048.
51. Harvard Chan Bioinformatics Core (HBC). Introduction to DGE [Internet]. GitHub Pages: hbctraining; [cited 2022 April 8]. Available from: https://hbctraining.github.io/DGE_workshop/lessons/02_DGE_count_normalization.html

52. Lapidus A, Korobeynikov A, Antipov D, Bushmanova E, Gurevich A, Kunyavskaya O et al. rnaSPAdes manual [Internet]. St.Petersburg State University: CAB SPbU; 2016 [cited 2022 July 11]. Available from: https://cab.spbu.ru/files/release3.11.1/rnaspades_manual.html#sec4
53. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. rrndb: the Ribosomal RNA Operon Copy Number Database. Nucleic Acids Res. 2001;29(1):181-184.
54. Pragman AA, Yarwood JM, Tripp TJ, Schlievert PM. Characterization of Virulence Factor Regulation by SrrAB, a Two-Component System in Staphylococcus aureus. J Bacteriol. 2004;186(8); 2430–2438.
55. Guo X, Sun Y. New Insights into the Non-orthodox Two Component Rcs Phosphorelay System. Front Microbiol. 2017;8:2014.
56. Tierney AR, Rather PN. Roles of two-component regulatory systems in antibiotic resistance. Future Microbiol. 2019;14(6):533-552.
57. Korandla DR, Wozniak JM, Campeau A, Gonzalez DJ, Wright ES, AssessORF: combining evolutionary conservation and proteomics to assess prokaryotic gene predictions. Bioinformatics. 2020;36(4):1022–1029.
58. Lim AHN, Lee AY, Hussein AR. Fundamental relationship between operon organization and gene expression. Proceedings of the National Academy of Sciences. 2011;108(26):10626-10631.

# 7 APPENDIX

**Table A.1**. Information about the primers used in PCR reaction to amplify the *rseP* gene. *Tm* shows the estimated annealing temperature (https://tmcalculator.neb.com/#!/main). *Ref. genome* has the accession number of the genomes and the location of the *rseP* sequence (used to design the primers).

| Bacteria | Primer name | Primer sequence | Tm | GC content | Ref.genome |
|---|---|---|---|---|---|
| L. lactis | yvjB_R | GCTGATAACCTGGGCATCAG | 66°C | 55% | NZ_PKRZ01000001.1, 814800- 816086 |
| | yvjB_F | TGGAATTCTTGACCGCTTTGAC | 65°C | 45% | NZ_PKRZ01000001.1, 814800- 816086 |
| S. haemolyticus | ShsRseP_F | TTGAGGAGTTATTGAGTGCAC | 62°C | 43% | NZ_CP013911.1, 917209- 918495 |
| | ShsRseP_R | GGGACATCTCTCCTAGTTGG | 64°C | 55% | NZ_CP013911.1, 917209- 918495 |
| E. faecium | Ing_rseP_fm_OFA_B | TGCTGTTTGTATTTCCTATC | 57°C | 35% | CP038996, 1619664-1620932 |
| | Ing_resP_fm_ORA_B | AAGGATCATTTTTCGATCAT | 56°C | 30% | CP038996, 1619664-1620932 |
| E. faecalis | OE_pair2_F | AGGTTTCTTCATGCGTTGGG | 66°C | 50% | CP051005, 2109513-2110781 |
| | OE_pair2_R | GCGAAGTGGTCAAGTCCAATG | 66°C | 52% | CP051005, 2109513-2110781 |

**Table A.2** Version number and container information for the software used

| Software | Version number | Singularity container |
|---|---|---|
| SnapGene | 5.3.2 | |
| UGENE | 40 | |
| RStudio/R version | 4.1.2 – 4.2.0 | |
| FastQC | | fastqc:0.11.9--0 |
| BBmap | | bbmap:38.86--h1296035_0 |
| SPAdes | | spades:3.15.3--h95f258a_0 |
| QUAST | | quast:5.0.2--py37pl5262h190e900_4 |
| Roary | | roary:3.9.1--pl5.22.0_0 |
| Prokka | | prokka:1.14.6--pl5262hdfd78af_1 |
| BLAST | | blast:2.9.0--pl526he19e7b1_7 |
| MultiQC | v1.12 | |
| HMMER | | hmmer:3.3.2--h1b792b2_1 |
| InterProScan | | InterProScan-5.45-80.0 |
| Kallisto | | kallisto:0.46.2--h60f4f9f_2 |

**Table A.3**. Annotations from Prokka and InterProScan for genes predicted to be near the *rseP* gene.

| Prokka; Gene name | Prokka; Gene product | InterProScan; Signature description |
|---|---|---|
| NA | hypothetical protein | Cytidylyltransferase family, NA |
| NA | Putative zinc metalloprotease | PDZ domain, PDZ_metalloprotease, Peptidase family M50, S2P-M50_PDZ_RseP-like, TIGR00054: RIP metalloprotease, NA |
| proS | Proline--tRNA ligase | Aminoacyl-tRNA editing domain, ProRS-INS, ProRS_core_prok, tRNA synthetase class II core domain (G, H, P, S and T), ProRS_anticodon_short, Anticodon binding domain, proS_fam_II: proline--tRNA ligase, Proline--tRNA ligase [proS]., NA |
| polC | DNA polymerase III PolC-type | DNA polymerase III polC-type N-terminus I, DNA polymerase III polC-type N-terminus II, polC_OBF, Exonuclease, DEDDh, dnaq: exonuclease, DNA polymerase III, epsilon subunit family, PHP domain, PHP_PolIIIA_POLC, Bacterial DNA polymerase III alpha NTPase domain, Bacterial DNA polymerase III alpha subunit finger domain, Helix-hairpin-helix motif, polC_Gram_pos: DNA polymerase III, alpha subunit, Gram-positive type, NA, DNA polymerase III PolC-type [polC]. |
| rimP | Ribosome maturation factor RimP | RimP N-terminal domain, RimP C-terminal SH3 domain, YlxS_C, NA, Ribosome maturation factor RimP [rimP]. |
| nusA | Transcription termination/antitermination protein NusA | NusA N-terminal domain, S1 RNA binding domain, S1_NusA, NusA-like KH domain, NusA_KH, NusA: transcription termination factor NusA, NA, Transcription termination/antitermination protein NusA [nusA]. |
| NA | hypothetical protein | Protein of unknown function (DUF448), YlxR, NA |
| rplGA | putative ribosomal protein YlxQ | Ribosomal protein L7Ae/L30e/S12e/Gadd45 family, NA |
| infB | Translation initiation factor IF-2 | Translation initiation factor IF-2, N-terminal region, small_GTP: small GTP-binding protein domain, Elongation factor Tu GTP binding domain, IF2_eIF5B, IF2_mtIF2_II, Translation-initiation factor 2, mtIF2_IVc, NA, Translation initiation factor IF-2 [infB]., IF-2: translation initiation factor IF-2 |
| rbfA | Ribosome-binding factor A | Ribosome-binding factor A, Ribosome-binding factor A [rbfA]., rbfA: ribosome-binding factor A, NA |

Alignment of the *rseP* gene in WT and mutant of *S. haemolyticus* LMGT4071 (Clustal Omega)

```
CLUSTAL O(1.2.4) multiple sequence alignment


S_haemolyticus_wt_rseP          GTGAGCTATTTAATCACTATTGTCTCATTTATGATCGTGTTTGGTGTACTTGTTACGGTA
S_haemolyticus_mutant_rseP      GTGAGCTATTTAATCACTATTGTCTCATTTATGATCGTGTTTGGTGTACTTGTTACGGTA
                                ************************************************************

S_haemolyticus_wt_rseP          CATGAATATGGTCACATGTTCTTTGCTAAGCGTGCTGGAATAATGTGTCCTGAATTCGCG
S_haemolyticus_mutant_rseP      CATGAATAGGGTCACATGTTCTTTGCTAAGCGTGCTGGAATAATGTGTCCTGAATTCGCG
                                ******** ***************************************************

S_haemolyticus_wt_rseP          ATAGGTATGGGACCAAAAATATTTAGTTTCCGTAAGAACGAAACTTTGTATACCATCCGT
S_haemolyticus_mutant_rseP      ATAGGTATGGGACCAAAAATATTTAGTTTCCGTAAGAACGAAACTTTGTATACCATCCGT
                                ************************************************************

S_haemolyticus_wt_rseP          TTATTACCTGTAGGTGGTTATGTAAGAATGGCTGGAGACGGTTTAGAAGAGCCACCAGTT
S_haemolyticus_mutant_rseP      TTATTACCTGTAGGTGGTTATGTAAGAATGGCTGGAGACGGTTTAGAAGAGCCACCAGTT
                                ************************************************************

S_haemolyticus_wt_rseP          GAACCAGGTATGAATGTTAAAGTAAAACTTAATGATAAAGATGAGATTACGCACATAATT
S_haemolyticus_mutant_rseP      GAACCAGGTATGAATGTTAAAGTAAAACTTAATGATAAAGATGAGATTACGCACATAATT
                                ************************************************************

S_haemolyticus_wt_rseP          TTAGATGATCAGCATAAATTTCAAAAGATAGAAGCAATTGAAGTTAAACAATGTGACTTC
S_haemolyticus_mutant_rseP      TTAGATGATCAGCATAAATTTCAAAAGATAGAAGCAATTGAAGTTAAACAATGTGACTTC
                                ************************************************************

S_haemolyticus_wt_rseP          AAGGATGATTTGTACATTGAAGGTATTACTTCATATGATAATGAGCGACATCATTTTAAT
S_haemolyticus_mutant_rseP      AAGGATGATTTGTACATTGAAGGTATTACTTCATATGATAATGAGCGACATCATTTTAAT
                                ************************************************************

S_haemolyticus_wt_rseP          ATTGCTGAAAAAGCATATTTTGTTGAGAATGGTAGTTTAATTCAAATCGCTCCGCGACAT
S_haemolyticus_mutant_rseP      ATTGCTGAAAAAGCATATTTTGTTGAGAATGGTAGTTTAATTCAAATCGCTCCGCGACAT
                                ************************************************************

S_haemolyticus_wt_rseP          CGACAATTTGCACATAAGAAACCTTTACCCAAATTTTTAACGTTATTTGCAGGACCATTA
S_haemolyticus_mutant_rseP      CGACAATTTGCACATAAGAAACCTTTACCCAAATTTTTAACGTTATTTGCAGGACCATTA
                                ************************************************************

S_haemolyticus_wt_rseP          TTTAACTTTATATTAGCCTTAATATTATTTATCGCATTAGCTTATTTCCAAGGAACACCT
S_haemolyticus_mutant_rseP      TTTAACTTTATATTAGCCTTAATATTATTTATCGCATTAGCTTATTTCCAAGGAACACCT
                                ************************************************************

S_haemolyticus_wt_rseP          ACTACCTCAGTGGGGCAATTAGCTGATCACTATCCAGCGCAACAAGCAGGATTAAAATCC
S_haemolyticus_mutant_rseP      ACTACCTCAGTGGGGCAATTAGCTGATCACTATCCAGCGCAACAAGCAGGATTAAAATCC
                                ************************************************************

S_haemolyticus_wt_rseP          GGAGATAAAATCGTTCAAGTAGGTCAATATAAAACAAAGAGTTTTGATGACATTCAGTCT
S_haemolyticus_mutant_rseP      GGAGATAAAATCGTTCAAGTAGGTCAATATAAAACAAAGAGTTTTGATGACATTCAGTCT
                                ************************************************************

S_haemolyticus_wt_rseP          GCAGCAAATAAAATTAAAGATAATAAAACAACTATAAAATTTGAAAGAGATAATCAAACA
S_haemolyticus_mutant_rseP      GCAGCAAATAAAATTAAAGATAATAAAACAACTATAAAATTTGAAAGAGATAATCAAACA
                                ************************************************************

S_haemolyticus_wt_rseP          AAGACAGTGGACATAACTCCTAAAAAGCAAGTTATTAAGCAAACTAAATTAAATTCTGAG
S_haemolyticus_mutant_rseP      AAGACAGTGGACATAACTCCTAAAAAGCAAGTTATTAAGCAAACTAAATTAAATTCTGAG
                                ************************************************************

S_haemolyticus_wt_rseP          ACGACGTACATTTTAGGTTTCCAACCAGAGAAAGAACATACTTTAATAAAACCAATTGCG
S_haemolyticus_mutant_rseP      ACGACGTACATTTTAGGTTTCCAACCAGAGAAAGAACATACTTTAATAAAACCAATTGCG
                                ************************************************************

S_haemolyticus_wt_rseP          CTAGGATTTGATCAGTTTGTTAGTGCCAGTACATTAATCTTTAAAGCTGTAGGAACAATG
S_haemolyticus_mutant_rseP      CTAGGATTTGATCAGTTTGTTAGTGCCAGTACATTAATCTTTAAAGCTGTAGGAACAATG
                                ************************************************************

S_haemolyticus_wt_rseP          ATTGCAAGTATATTCACAGGTCAATTCTCATTTGATATGTTAAATGGTCCAGTGGGTATT
S_haemolyticus_mutant_rseP      ATTGCAAGTATATTCACAGGTCAATTCTCATTTGATATGTTAAATGGTCCAGTGGGTATT
                                ************************************************************

S_haemolyticus_wt_rseP          TATCATAATGTTGACTCTGTAGTTAAGCAGGGTATCATTGCTTTAACATACTACACTGCA
S_haemolyticus_mutant_rseP      TATCATAATGTTGACTCTGTAGTTAAGCAGGGTATCATTGCTTTAACATACTACACTGCA
                                ************************************************************

S_haemolyticus_wt_rseP          CTATTAAGTGTTAACTTAGGTATAATGAACTTATTACCAATTCCAGCACTTGATGGTGGT
S_haemolyticus_mutant_rseP      CTATTAAGTGTTAACTTAGGTATAATGAACTTATTACCAATTCCAGCACTTGATGGTGGT
                                ************************************************************
```

```
S_haemolyticus_wt_rseP          CGAATTTTATTTGTTATCTATGAAGCAATTTTCAGAAGACCAGTTAATAAAAAAGCAGAA
S_haemolyticus_mutant_rseP      CGAATTTTATTTGTTATCTATGAAGCAATTTTCAGAAGACCAGTTAATAAAAAAGCAGAA
                                ************************************************************

S_haemolyticus_wt_rseP          ACAATTATAATAGCTGCTGGTGCTATTTTTGTCTTAATTATAATGGTTCTAGTAACTTGG
S_haemolyticus_mutant_rseP      ACAATTATAATAGCTGCTGGTGCTATTTTTGTCTTAATTATAATGGTTCTAGTAACTTGG
                                ************************************************************

S_haemolyticus_wt_rseP          AACGATATACAACGTTATTTCTTGTAA
S_haemolyticus_mutant_rseP      AACGATATACAACGTTATTTCTTGTAA
                                ***************************
```


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

Postboks 5003
NO-1432 Ås
Norway