

# **Preparation of a *Mycobacterium avium* Transposon Library for Mutant Identification by Illumina Sequencing and Study of Genes Potentially Involved in Mycobacterial Iron Uptake**

Maja Helland

Master's thesis in Molecular Medicine  
Trondheim, June 2015

Supervisor: Marte Singsås Dragset

Norwegian University of Science and Technology  
Faculty of Medicine  
Department of Cancer Research and Molecular Medicine  
Centre for Molecular Inflammation Research



**NTNU – Trondheim**  
Norwegian University of  
Science and Technology







## Abstract

*Mycobacterium tuberculosis* is a serious health problem worldwide, and cannot be conquered without an efficient vaccination strategy. Today there are no efficient vaccines available, and it is thereby urgent to develop new ones. *M. tuberculosis* is also acquiring resistance to an increasing number of antibiotics, which has even lead to a *M. tuberculosis* strain that is totally drug-resistant. Consequently, it is crucial to develop new drugs to deal with this threat. To develop new vaccines and drugs it is important to identify and validate new potential mycobacterial drug targets.

*Mycobacterium avium* shares many properties with its relative *M. tuberculosis*, such as ability to establish infection within macrophages. *M. avium* can also cause severe infections, particularly in immunocompromised individuals. Although *M. avium* can serve as a model organism for studying tuberculosis, it is therefore also a bacterium that absolutely needs research attention by its own.

To contribute towards the development of new and effective drugs and vaccines against pathogenic mycobacteria, the specific aim of this study was first of all to organise an *M. avium* transposon mutant library of around 10 000 mutants, and prepare it for Illumina sequencing, in a way that mutants with specific insertion sites could be easily identified in a future application of the library. Such an organised library would make it easy to pick out interesting mutants for further investigation and also to conduct screenings of many thousand mutants at the same time and study their phenotype under different conditions.

Secondly, this study aims to investigate selected genes potentially involved to mycobacterial iron uptake. In drug development it would be of interest to target parts of the mycobacterial iron uptake system, which is essential for the mycobacteria to survive. For this aim, growth curve analysis of non-pathogenic model organism *Mycobacterium smegmatis* with gene potentially involved in iron uptake knocked out were performed in media containing various iron concentrations.

Mutants organised and prepared for Illumina sequencing were sent for Illumina sequencing at Texas A&M University, and sequencing results are expected to ready at the end of June 2015.

From the growth curve analysis performed, the result is that the genes *msmeg\_4898* and *msmeg\_4900* do not seem to be important for the mycobactin pathway in *M. smegmatis*. However, this does not mean that they are not important for this purpose in other mycobacteria.



## Acknowledgements

This thesis project was carried out at the Department of Cancer Research and Molecular Medicine at the Faculty of Medicine at the Norwegian University of Science and Technology (NTNU) and was given from the Centre of Molecular Inflammation Research, (CEMIR). The master's thesis is a part of the Master of Science degree in Molecular Medicine.

First of all, I would like to give a special thanks my supervisor Dr. Marte Singsås Dragset, for helping and supporting me throughout the work with this master's thesis. Thank you for investing a lot of time in me for questions, discussions and conversations, and for all the encouragement you have given. Your scientific experience and knowledge has been really educative and essential for this thesis.

I also want to thank the rest of the Mycobacteria research group, especially Prof. Trude Helen Flo, who is also the leader of the group, for sharing of their knowledge and experience.

During the time working on my master's thesis I have mostly been working at the laboratories at CEMIR, but I also used lab facilities at SINTEF, the Department of Cancer Research and Molecular Medicine and the Genomics Core Facility at NTNU. Thanks to Geir Klinkenberg and Guro Kruge Urdal at SINTEF for using your really cool robots. Thanks to Ruth Haaland Krokan at Department of Cancer Research and Molecular Medicine for helping with Sanger sequencing. Thanks to Sten Even Erlandsen for helping with shearing DNA with Covaris and analysing the DNA fragment sizes at BioAnalyzer.

Thanks a lot to Tom Ioerger at Texas A&M University for doing the Illumina sequencing and analysing the data. That is a huge and valuable job!

Especially, thanks to my fiancé and husband to be, Simon Loevenich, for supporting and encouraging me during this period. Love you!

Trondheim, May 2015

Maja Helland

## Table of contents

Abstract .....	II
Acknowledgements .....	IV
Table of contents .....	V
List of figures .....	VIII
List of tables .....	IX
1. Introduction .....	1
1.1 The genus <i>Mycobacterium</i> .....	1
1.1.1 <i>Mycobacterium tuberculosis</i> .....	1
1.1.2 <i>Mycobacterium avium</i> .....	2
1.1.3 <i>Mycobacterium smegmatis</i> .....	2
1.2 Mycobacterial iron uptake .....	2
1.2.1 Mycobacterial siderophores .....	3
1.2.2 Putative iron uptake genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	4
1.3 Mutagenesis as genetic tools for studying gene functions .....	5
1.3.1 Insertional mutagenesis.....	6
1.3.2 Homologous recombination.....	8
1.4 Illumina Sequencing .....	8
1.4.1 The principle of Illumina sequencing - sequencing by synthesis .....	9
1.4.2 Transposon-sequencing (Tn-seq).....	9
1.4.3 Sample multiplexing and barcoding .....	11
2. Aims of the study .....	12
3. Materials and methods .....	14
3.1 Tables of materials.....	14
3.1.1 Growth media.....	14
3.1.2 Buffers and solutions .....	15
3.1.3 Primers .....	15
3.1.4 Adapters .....	18
3.1.5 Ladders.....	19
3.1.6 PCR kits .....	19
3.1.7 Cloning kits.....	20
3.2 Dilution series to find the amount of bacteria in the transposon library .....	21
3.3 Purification of DNA .....	21
3.3.1 Bead beating of bacteria.....	21
3.3.2 MasterPure DNA Purification Kit from Epicentre .....	22
3.3.3 QIAquick PCR Purification Kit.....	22

3.3.4 QIAGEN DyeEx 2.0 SpinKit.....	23
3.3.5 PureYield Plasmid Miniprep System.....	23
3.3.6 QIAquick Gel Extraction Kit.....	24
3.4 Spectrophotometry.....	25
3.5 Robots used at SINTEF.....	25
3.5.1 Colony picking robot.....	25
3.5.2 Pipetting robots.....	26
3.6 PCR.....	26
3.6.1 PCR for in-house Sanger sequencing of genomic DNA.....	27
3.6.2 Confirming of the presence of the kanamycin resistance gene.....	27
3.6.3 Small scale PCR after adapter ligation.....	28
3.6.4 PCR for addition of Illumina attachment and read homology.....	28
3.6.5 Confirming knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	29
3.7 Gel electrophoresis.....	29
3.8 Preparation of samples for external analysis.....	30
3.8.1 GATC-sequencing.....	30
3.8.2 Preparation Covaris shearing:.....	30
3.8.3 Preparing for BioAnalyzer.....	30
3.9 Analysing data from Sanger sequencing.....	30
3.10 Transformation.....	30
3.10.1 Transformation in competent <i>E. coli DH5α</i> .....	30
3.10.2 Transformation in <i>M. smegmatis</i> .....	31
3.11 Preparation of transposon library mutants for Illumina sequencing.....	31
3.11.1 DNA shearing.....	31
3.11.2 End-repair.....	32
3.11.3 A-tailing.....	32
3.11.4 Making adapter mix.....	33
3.11.5 Adapter ligation.....	33
3.12 Homologous recombination of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	34
3.12.1 Growth curves preparations.....	35
3.13 Freeze down of culture.....	35
4. Results.....	36
4.1 Quality assessment of the <i>M. avium</i> transposon mutant library.....	36
4.1.1 Verification of transposon insertions.....	36
4.1.2 Small scale identification of transposon insertion sites.....	37
4.2 Organising the <i>M. avium</i> library: Growing and pooling the mutants.....	38
4.2.1 Preparation of the stock transposon insertion library for organisation.....	38

4.2.2 Colony picking robot .....	39
4.2.3 Pooling the plates .....	40
4.3 Preparing the <i>M. avium</i> library pools for identification of transposon insertion sites by Illumina sequencing.....	41
4.3.1 Shearing genomic DNA .....	41
4.3.2 End repair, A-tailing and adapter ligation.....	42
4.3.3 Adapter ligation and coding the DNA pools.....	43
4.3.4 PCR of adapter ligated plate, column and row DNA pools.....	47
4.4 Creating knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> by homologous recombination .....	51
4.4.1 Selecting knock-out mutants for growth curve analysis .....	52
4.4.2 Growth of <i>msmeg_4898</i> and <i>msmeg_4900</i> knock-outs under low iron conditions .	54
5. Discussion .....	57
5.1 Transposon mutagenesis versus homologous recombination.....	57
5.2 Identification of transposon insertion site by Sanger sequencing .....	58
5.3 Preparation of the transposon library mutant pools for Illumina sequencing.....	59
5.4 Potential results from Illumina sequencing and their reliability.....	61
5.5 The putative role of <i>msmeg_4898</i> and <i>msmeg_4900</i> in mycobacterial iron uptake.....	62
5.6 Summary/ Conclusion .....	63
5.7 Further perspectives.....	64
6. References .....	65
7. Appendices .....	70
Appendix 1: Gel electrophoresis that checks the fragment length of the column pools after being sheared with Covaris.....	70
Appendix 2: Gel electrophoresis that checks the fragment lengths of the row pools after being sheared with Covaris.....	70
Appendix 3: Gel electrophoresis of PCR-products where all the pools in each category plate, column and row are pooled.....	71
Appendix 4: Gel from Figure 20, after cutting out bands with approximately length of 200 to 400 base pairs. ....	71
Appendix 5: Gel electrophoresis of products from PCR with primer pair 2, 3 and 4 to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	72
Appendix 6: Gel electrophoresis of WT PCR-products from PCR with primer pair 1, 2, 3 and 4 to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> . ....	72
Appendix 7: The second replicate of growth curve analysis.....	73
Appendix 8: The third replicate of growth curve analysis. ....	74

## List of figures

Figure 1: Model of mycobactin (MBT)/carboxymycobactin (cMBT) transport over the cell wall in <i>M. tuberculosis</i> . .....	4
Figure 2: The genes <i>lprD</i> and <i>rv1342c</i> in <i>M. tuberculosis</i> . .....	5
Figure 3: The genes <i>msmeg_4898</i> and <i>msmeg_4900</i> in <i>M. smegmatis</i> . .....	5
Figure 4: Insertions of class I and class II transposons into target DNA sequences. ....	6
Figure 5: An overview of the transposon in $\phi$ MycoMarT7. ....	7
Figure 6: Sample preparation for identification of transposon insertion sites using Illumina sequencing. ....	10
Figure 7: A scheme over the preparation workflow of transposon library mutants for Illumina sequencing. ....	13
Figure 8: The structure of the transposon primers. ....	16
Figure 9: The structure of the adapter primers, Sol_AP1. ....	16
Figure 10: Ladders used in this thesis. ....	19
Figure 11: Checking gene for kanamycin resistance. ....	37
Figure 12: Agar plate showing a colony density suitable for robotic colony picking. ....	39
Figure 13: Grinding the library in plate, column and row pools. ....	40
Figure 14: Optimisation of the activity time for fragmentase. ....	42
Figure 15: Strategy for coding the mutants. ....	44
Figure 16: Gel electrophoresis of transposon junctions of all the plate pools amplified by small scale PCR. ....	45
Figure 17: Gel electrophoresis of transposon junctions of all the column pools amplified by small scale PCR. ....	45
Figure 18: Gel electrophoresis of transposon junctions of all the row pools amplified by small scale PCR. ....	46
Figure 19: Gel electrophoresis of PCR-products where all the pools in each category plate, column and row are pooled. ....	48
Figure 20: Gel from figure 19, after cutting out bands with approximately length of 200 to 400 base pairs. ....	48
Figure 21: Gel electrophoresis of the final pools of plates (P), columns (C) and rows (R). ....	49
Figure 22: Results from BioAnalyzer, showing smears from shearing the final plate, column and row pool with Covaris. ....	50
Figure 23: Electropherogram from BioAnalyzer, showing the distribution of DNA in the final column, row and plate pool. ....	50
Figure 24: Four set of primer pairs were used to confirm knock out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> . ....	52
Figure 25: Primer pairs used to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> . ....	52

Figure 26: Gel electrophoresis of products from PCR with primer pair 1 to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	53
Figure 27: Gel electrophoresis of products from PCR with primer pair 2, 3 and 4 to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> . ....	54
Figure 28: The first replicate of growth curve analysis. ....	56

## List of tables

Table 1: Growth media used in this study.....	14
Table 2: Buffers and solutions used in this study.....	15
Table 3: Primers used in this study to confirm the presence of the kanamycin resistance gene in the transposon library mutants. ....	15
Table 4: The transposon binding primers used in the preparation step for Illumina sequencing. ....	16
Table 5: The adapter binding primers used in the preparation step for Illumina sequencing. .	17
Table 6: Primers used in this study to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> .....	17
Table 7: Sequencing primers used in this study.....	17
Table 8: The barcodes of the adapters number 2 used in this study.....	18
Table 9: PCR kits used in this thesis. ....	19
Table 10: Cloning kits used in this thesis.....	20
Table 11: The set up for the ligation reaction when using InsTAclone PCR Cloning Kit.....	20
Table 12: The set up for the ligation reaction when using CloneJET PCR Cloning Kit. ....	20
Table 13: Sequencing reaction mix for PCR for in-house Sanger sequencing of genomic DNA to identify transposon insertion sites. ....	27
Table 14: Cycling conditions for PCR for in-house Sanger sequencing of genomic DNA to identify transposon insertion sites. ....	27
Table 15: Reaction mix for PCR used to confirm the existence of the kanamycin resistance gene in the transposon library mutants.....	27
Table 16: Cycling conditions for PCR used to confirm the existence of the kanamycin resistance gene in the transposon library mutants. ....	27
Table 17: Reaction mix for small scale PCR after adapter ligation of the transposon library mutants. ....	28
Table 18: Cycling conditions for small scale PCR after adapter ligation of the transposon library mutants.....	28
Table 19: Reaction mix for PCR for addition of Illumina attachment and read homology. ....	28
Table 20: Cycling conditions for PCR for addition of Illumina attachment and read homology. ....	28

Table 21: Reaction mix used for confirming knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> and thereby choosing mutants to growth curve analysis.....	29
Table 22: Cycling conditions used for confirming knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> and thereby choosing mutants to growth curve analysis.....	29
Table 23: Reaction mix used for fragmentase shearing. ....	32
Table 24: Checking that the transposons are inserted in different genes. ....	37
Table 25: Results from Sanger sequencing. ....	47
Table 26: Expected results of successful knock-out mutants at gel electrophoresis after PCR. .....	52
Table 27: The primer pairs used to confirm knock-out of the genes <i>msmeg_4898</i> and <i>msmeg_4900</i> and the length of their PCR-products.....	53

## 1. Introduction

### 1.1 The genus *Mycobacterium*

The bacterial genus *Mycobacterium* consists of aerobic, acid-fast, non-spore-forming rod shaped bacteria with a complex and lipid-rich cell wall. They are 0.2 to 0.6  $\mu\text{m}$  wide and 1 to 10  $\mu\text{m}$  long [1]. Mycobacteria are also non-motile, except for the motile species *Mycobacterium marinum*, which can propel by actin-based motility [2]. In addition most of the mycobacteria are slow-growing, dividing every 12 to 24 hours [1].

#### 1.1.1 *Mycobacterium tuberculosis*

*Mycobacterium tuberculosis*, which causes tuberculosis (TB), was responsible for 1.5 million deaths and an estimate of 9 million new TB cases in 2013 [3]. This makes tuberculosis a serious health problem worldwide. Individuals with HIV currently have TB as the main cause of death [4]. Mycobacteria can establish infections within the human macrophage cells, which are the first line of defence in the immune system [5]. *M. tuberculosis*, which is spread through infectious aerosols, mostly affect the lungs, but can also infect other parts of the body [6]. In some cases the infection is latent, in which cause the infected TB-patients do not show symptoms [7].

The genomic DNA of *M. tuberculosis* encodes around 4000 genes [8]. It is also surrounded by a wax-rich and thick cell wall [6]. This unique cell wall prevents the bacteria from being attacked by medical drugs [9]. It also impair nutrient entry into the bacteria, which explains the slow growth [10].

Tuberculosis cannot be conquered without an efficient vaccination strategy [6]. Today there are no efficient vaccines, and the need to develop new vaccines is urgent. The vaccine that is used today, *Mycobacterium bovis* Bacillus Calmette-Guérin (BCG), gives good protection against severe forms of childhood tuberculosis. However, it does not give protection against adult pulmonary tuberculosis, which is the most common form of the disease [11].

Also the treatment of tuberculosis faces major challenges. *M. tuberculosis* is acquiring resistance to an increasing number of antibiotics, which has even lead to a totally drug-resistant strain of the bacterium [12]. Consequently, it is enormously important to develop new drugs to deal with this threat. To manage this, new potential drug targets have to be identified.

## INTRODUCTION

### 1.1.2 *Mycobacterium avium*

*Mycobacterium avium* is a bacterium with around 5100 genes, where approximately 2900 of these are homologues to genes in *M. tuberculosis*. *M. avium* shares many properties with its relative, such as ability to establish infection within macrophages and a slow growth rate [13]. *M. avium* is an opportunistic human pathogen, less pathogenic than its relative *M. tuberculosis*, and working with this species requires less stringent biosafety laboratory-levels [14].

Infection with *M. avium* is primarily acquired through ingestion of contaminated food or water. Immunocompromised individuals, particularly people living with AIDS, have the greatest risk for disease caused by *M. avium* and an infection can be a severe threat for such patients. This is because the infection is typically disseminated if left untreated, with virtually no organ saved [1]. Although *M. avium* can cause severe infections that can be hard to treat, and thus deserves research attention by its own, the species have many features in common with *M. tuberculosis* and can serve as a model organism for studying TB [14].

### 1.1.3 *Mycobacterium smegmatis*

*Mycobacterium smegmatis* is an often used model organism in *M. tuberculosis* research, and is used to study basic mycobacterial mechanisms. It is saprophytic [15], and contains of around 7000 genes, where 2700 of these genes are homologues to genes in *M. tuberculosis* [16]. The advantages of *M. smegmatis* as a model organism is that it is fast-growing and non-pathogenic [17].

## 1.2 Mycobacterial iron uptake

One of the ways the *M. tuberculosis* survives intercellularly is by employing mechanisms that take up iron at extremely low iron conditions [18].

There is a fight for iron between the human cells and the mycobacteria during infection. The human body will withhold iron from the mycobacteria. This is done by keeping the level of soluble iron low. The extracellular iron is mostly bound to the glycoprotein transferrin, and the intracellular iron is bound to ferritin or in a complex with heme, e.g. haemoglobin [19]. Normally transferrin binds only an average of 0.4 atoms of iron per molecule, even though it can bind up to 2 atoms per molecule. This leaves room for further iron restriction, so upon infections the soluble iron level becomes even lower [20].

The mycobacteria on their side have evolved strategies to take up iron even under conditions with extremely low levels of iron. These uptake strategies can be divided into three groups; using receptors for acquisition of iron bound to transferrin and lactoferrin, heme acquisition systems and siderophore-based iron uptake systems [19].

### 1.2.1 Mycobacterial siderophores

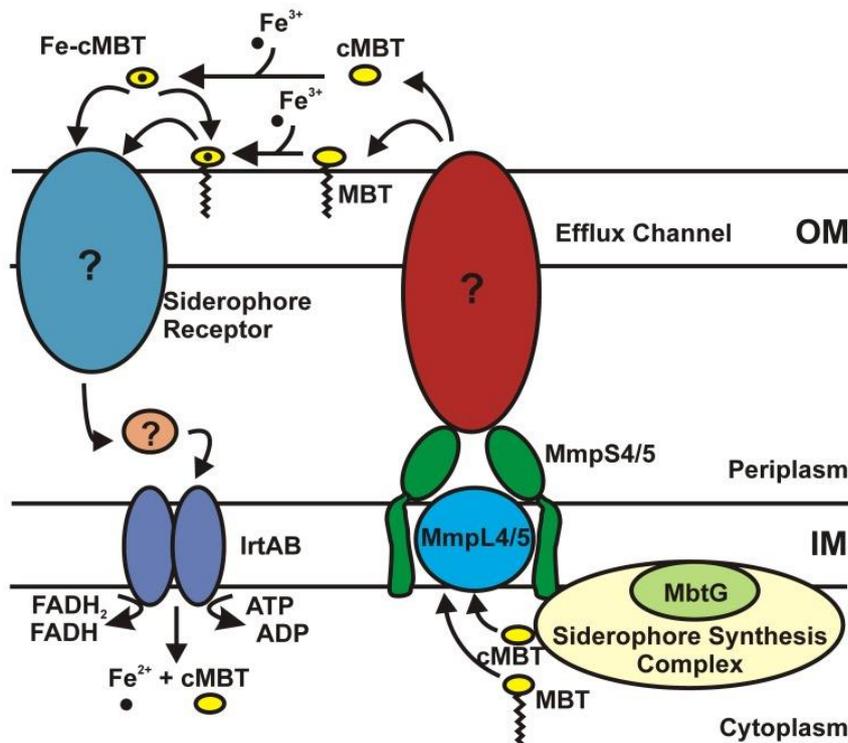
Siderophores are iron-binding molecules which transfer iron from ferritin, lactoferrin and transferrin to the mycobacterial cell wall [21, 22]. Siderophores are synthesised in the cytoplasm transported out of the bacterial cell where they can find and bind iron, before transported back into the cell for usage [23]. However, the siderophores are too large to pass porins in the membrane. Therefore a special transport system is required for importing and exporting siderophores over the outer membrane [24].

Mycobactin and carboxymycobactin are two structurally related siderophores produced by most species of mycobacteria [25]. The only difference between mycobactin and carboxymycobactin is that mycobactin has a long monocarboxyl fatty acid tail, while carboxymycobactin has a shorter dicarboxyl fatty acid tail [26]. Due to the long fatty acid tail, mycobactin is lipophilic, which promotes adherence to the lipidic surface of the mycobacteria. Carboxymycobactin on the other hand is hydrophilic because of the short dicarboxyl tail [27]. Mycobactin is proposed to be an iron storage molecule, that efficiently extracts intracellular iron from macrophages [28].

In addition to mycobactins, some mycobacteria produce a water-soluble siderophore called exochelin. *M. smegmatis* is one of the mycobacteria producing this siderophore [29]. The fact that *M. smegmatis* has two unrelated siderophore-mediated iron uptake mechanisms can be used to find out if a gene is involved in uptake of iron in the mycobactin or exochelin siderophore pathway. To determine if a *M. smegmatis* gene is involved in the mycobactin pathway can potentially be translatable to iron uptake in pathogenic mycobacteria. In fact, the gene cluster *esx-3* was found to be required for mycobactin-mediated iron uptake using the *M. smegmatis* approach [30], however why it is required is still not known. The gene cluster *esx-3* was found to be required for mycobactin-mediated iron uptake by combining mutations in the two different iron uptake pathways. The bacteria should fail to grow under low iron conditions if both siderophore pathways are knocked out. By combining mutations in the same iron uptake system, the bacteria should still be able to grow under low iron conditions, as it would be rescued by the intact siderophore. If a mutant strain where exochelin and another gene of

## INTRODUCTION

interest are knocked out fails to grow at low iron conditions, this indicates that the gene under investigation is important for the mycobactin pathway. Actually, this example is of real interest. Today it is not known how mycobactin is transported over the outer membrane (Figure 1) [23].

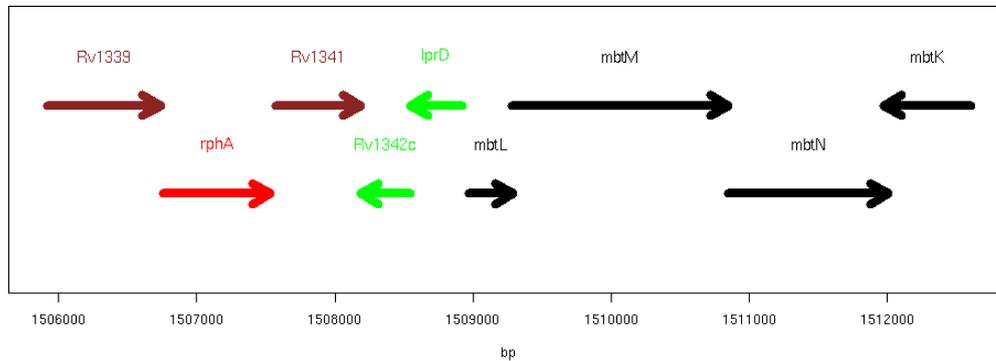


**Figure 1: Model of mycobactin (MBT)/carboxymycobactin (cMBT) transport over the cell wall in *M. tuberculosis*.** To export mycobactin and carboxymycobactin across the cell wall a yet undiscovered protein in the outer membrane is required. To transport the siderophores with bound iron across the outer membrane a siderophore receptor in the outer membrane is required. The proteins required for mycobactin/carboxymycobactin transport over the outer membrane in the cell wall are not known [23].

### 1.2.2 Putative iron uptake genes *msmeg\_4898* and *msmeg\_4900*

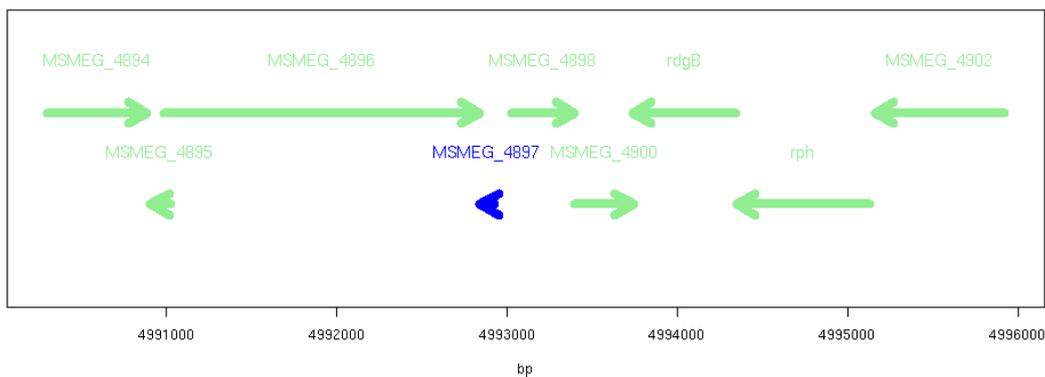
The *M. smegmatis* genes *msmeg\_4898* and *msmeg\_4900* are homologous genes of *lprD* and *rv1342c* respectively in *M. tuberculosis* (Figure 2 and 3). The functions of *msmeg\_4898* and *msmeg\_4900* is so far unknown. The genes *lprD* and *Rv1342c* are identified as encoding membrane proteins involved in the cell wall and in cell processes, but their function is not further clarified yet [31]. Right upstream of *lprD* and *Rv1342c* are the genes *mbtM*, *mbtL*, *mbtN* and *mbtK* (Figure 2), which are known to be involved in mycobactin synthesis in *M. tuberculosis* [32]. The upstream DNA sequence of the gene *lprD* is also predicted to be a binding site for IdeR (iron-dependent regulator), which controls transcription of genes involved in iron uptake [33, 34], and it has been shown that *lprD* is repressed by iron and IdeR [35].

Because of this the genes *msmeg\_4898* and *msmeg\_4900*, which are probably transcribed together, are of interest to study; they can be candidates that can answer the question of how mycobactin is transported over the outer membrane.



**Figure 2: The genes *lprD* and *rv1342c* in *M. tuberculosis*.**

The *M. smegmatis* genes *msmeg\_4898* and *msmeg\_4900* are homologous genes of *lprD* and *rv1342c* respectively in *M. tuberculosis*. Right upstream of *lprD* and *Rv1342c* are the genes *mbtM*, *mbtL*, *mbtN* and *mbtK*, which are known to be involved in mycobactin synthesis in *M. tuberculosis* [36].



**Figure 3: The genes *msmeg\_4898* and *msmeg\_4900* in *M. smegmatis*.**

The *M. smegmatis* genes *msmeg\_4898* and *msmeg\_4900* are homologous genes of *lprD* and *rv1342c* respectively in *M. tuberculosis*. The genes *msmeg\_4898* and *msmeg\_4900* have both an unknown function [37].

### 1.3 Mutagenesis as genetic tools for studying gene functions

In the laboratory, mutagenesis is an important technique to produce mutations in genes. It is possible to manipulate different elements in a gene, such as its coding sequence or its control elements. By mutating a gene in a way that causes it to lose its original function, the function of a gene or protein can be explored. The two common mutagenesis tools used in this project were homologous recombination and insertional mutagenesis [38]. Homologous recombination

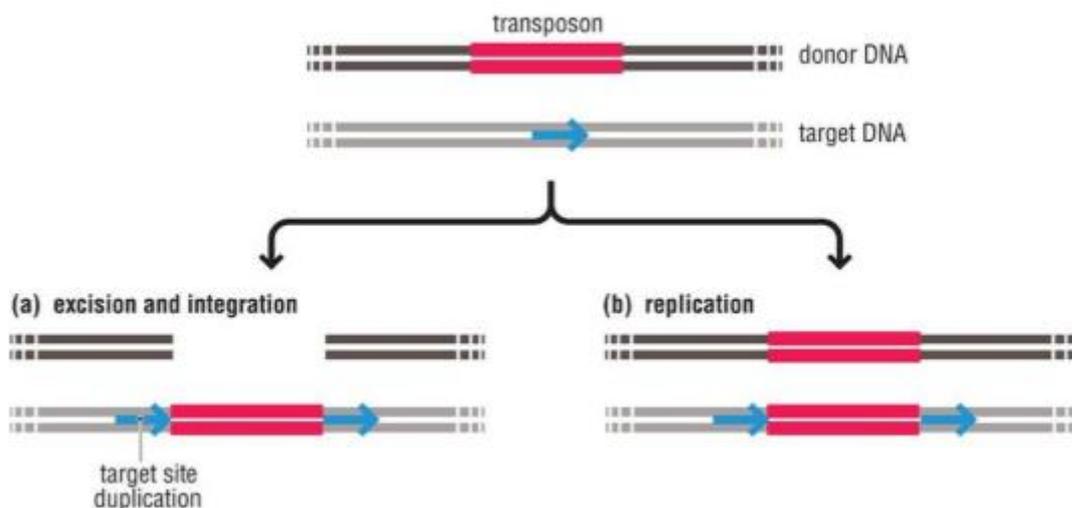
## INTRODUCTION

was used to knock-out genes potentially involved in mycobacterial iron uptake (section 4.4) and insertional mutagenesis were used to generate a transposon mutant library (section 4.1).

### 1.3.1 Insertional mutagenesis

Insertional mutation is when a DNA sequence, such as a transposon, is inserted in the gene to be mutated. This will generally lead to a gene disruption, thus suppressing expression of the affected gene.

Transposons are DNA sequences which can change position within a genome, and can be categorised into two main classes. Class I transposons move by reverse transcription of an RNA intermediate. Class II are transposons that move directly from one DNA- sequence to another. An overview of how the two classes of transposons are inserted into target DNA is shown in Figure 4. Transposons usually have inverted repeats at each end and encode a transposase, which is an enzyme that cuts out the transposon sequence and inserts into a new target site [39, 40]. For most class I transposons insertions occur by replication of a non-moving transposon into a different site (Figure 4b). Insertions of class II transposons mostly occur by excision and moving to a new site (Figure 4a). With the transposase acting on a specific DNA sequence at each end of the transposon, the class II transposon will insert into a new target DNA position. This can allow new genes to be transferred to the host's chromosome [41].



**Figure 4: Insertions of class I and class II transposons into target DNA sequences.**

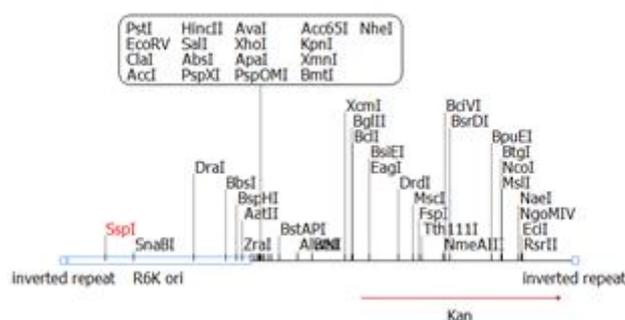
Transposon insertions occur either by a) excision and movement of the transposon to a new site (mostly class II elements), or by b) replication of a non-moving transposon into an additional site (mostly class I elements) [39].

Delivery of a transposon to a cell can be achieved by transformation with plasmid DNA or by using a virus. Disruption of a genetic sequence by a transposon can thus be used as a genetic tool for analysis of protein and gene function, by using them as vectors for genome manipulations [42]. The insertion of a transposon can suppress the natural function of a gene, by altering gene expression or regulation [43]. It is then possible to investigate how this loss of function affects the organism. This is called transposon mutagenesis and is an effective method for creating random mutations at high frequency in target DNA [44].

### 1.3.1.1 Transposon mutagenesis of mycobacteria

Because of the slow growth of mycobacteria and the low availability of refined genetic tools, it has generally been difficult to study and identify conditionally essential genes in mycobacteria [45]. Transposon mutagenesis is a very important technique for finding such genes, because very complex mutant libraries can easily be made [43].

To insert transposons randomly into mycobacterial genome, the phagemid  $\phi$ MycoMarT7 has been constructed by Sassetti *et al.* and shown to be effective for creating mycobacterial insertion mutants [45]. This phagemid contains a class II transposon encoding a kanamycin resistance gene as well as a highly active Himar1 transposase. At both ends of the transposon there are 29 base pairs long inverted repeats. Like other Himar1-based transposons, this transposon inserts only at dinucleotide TA [45]. An overview of the transposon in  $\phi$ MycoMar T7 is shown in Figure 5.



**Figure 5: An overview of the transposon in  $\phi$ MycoMarT7.**

Transposon  $\phi$ MycoMarT7 contains a kanamycin resistance gene and each end has inverted repeats.

## INTRODUCTION

### 1.3.2 Homologous recombination

Homologous recombination makes it possible to exchange gene sequences between two DNA molecules, and is another valuable mutagenesis technique often used in laboratory DNA manipulations. Homologous recombination occurs at low frequencies in human cells, where it increases genetic variation during meiosis [46], and can repair double stranded breaks in the DNA [47]. Some bacteria use homologous recombination to exchange genetic material between strains [48].

In homologous recombination two DNA-molecules exchange parts of their sequence. To make this happen the DNA sequences to be swapped must be similar to each other, which can be done by matching flanking areas. A recombinase, an enzyme that catalyse exchange of a DNA, recognises and binds the flanking areas [49]. Knock-out mutants can be made by using homologous recombination, because the swapping can be used to cut a gene out of a genome, and replace it with another DNA sequence, that e.g. could be coding for an antibiotic resistance gene.

#### 1.3.2.1 *Mycobacterial recombineering plasmid pJV53*

A plasmid used for homologous recombination in mycobacteria is the plasmid pJV53, constructed by van Kessel and Hatfull. The plasmid, encodes the recombinase proteins gp60 and gp61. They are under control of an acetamide-inducible promoter. Gp60 is a 5'-3' double stranded DNA-dependent exonuclease. Gp61 is a single stranded DNA-binding protein which promotes annealing of complementary DNA strands, strand exchange and strand invasion [50].

### 1.4 Illumina Sequencing

Next-generation sequencing (NGS) technologies became available at the onset of the 21<sup>st</sup> century. NGS sequencing platforms are fast and very efficient. Today NGS technologies are used in many fields within biology and they are continually becoming even faster, more efficient and cheaper [51]. One of these NGS platforms is Illumina sequencing, which makes it possible to sequence a whole mutant library in one run. Also by uniquely labelling the mutants it is possible to link the sequencing results to the respective bacteria culture. This is explained in more details in section 1.4.3.

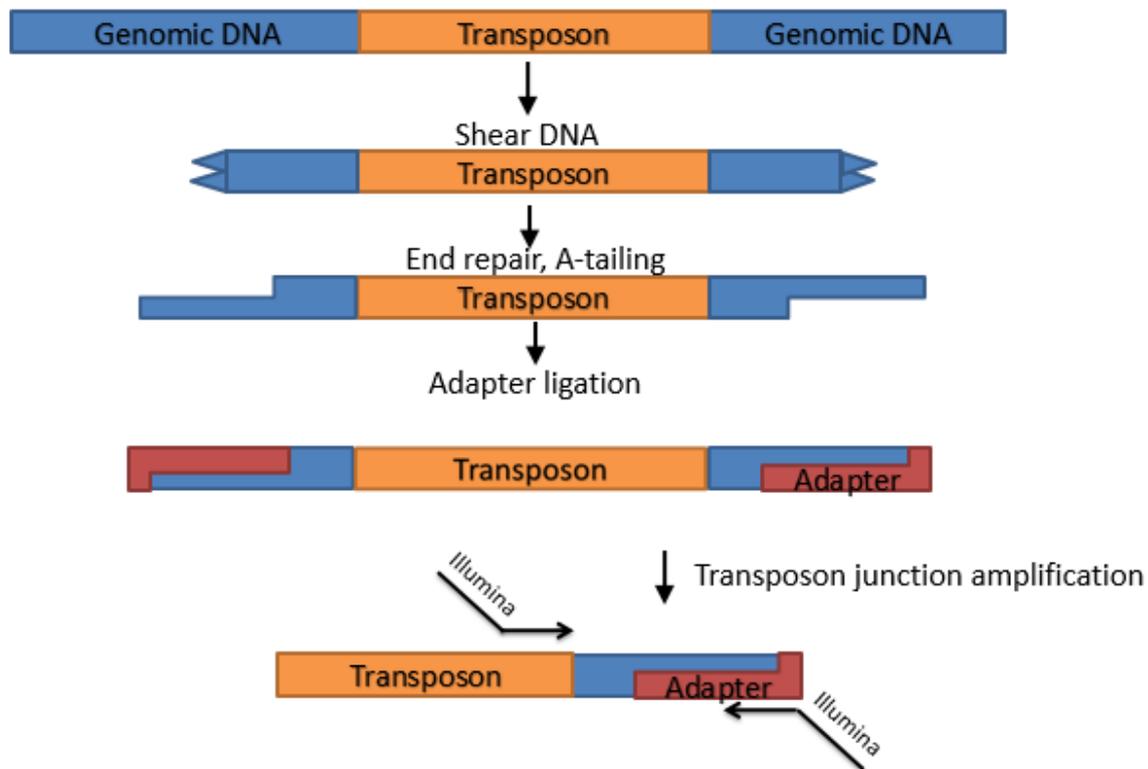
#### 1.4.1 The principle of Illumina sequencing - sequencing by synthesis

Illumina technology makes it possible to sequence a high number of samples in parallel. The flow cells in the Illumina sequencer are coated with two types of oligonucleotides. The DNA-fragments in the sample to be sequenced have adapters ligated to both ends of the strands. When the sample is loaded onto the flow cell, single-stranded fragments bind randomly to the first of the two types of oligonucleotides in the flow cell via a complementary adapter region. The bound fragments are then made double-stranded by adding polymerase and nucleotides. The double-stranded molecules are then denaturated, and the original template is washed away. The single-stranded molecule left behind bends over and the adapter at the other side of the molecule binds to the second type of oligonucleotides. This molecule now form a bridge, and the 2<sup>nd</sup> strand is then synthesised. This is called bridge amplification. The molecule is denaturated again resulting in two single-stranded copies of the molecule. This is simultaneously repeated several times for all the fragments, so that millions of clusters of double stranded DNA will be generated. Each cluster contains the amplification products from the one original single-stranded DNA-fragment. Right before sequencing, the reverse strands are cleaved and washed away, so that there are only forward strands left. The 3'end of these strands are blocked to prevent additional priming to oligonucleotides. A sequencing primer binds to the single-stranded molecule and is extended by fluorescently tagged nucleotides, which are competing to bind. The complementary nucleotide will bind. After the incorporation of each nucleotide the clusters are excited by a laser, and the characteristic fluorescence of the added nucleotide is emitted and captured for each cluster. This is repeated several times and this way one nucleotide after another in the template is identified. The sequencing results can then be compared to a reference genome from a databank [52].

#### 1.4.2 Transposon-sequencing (Tn-seq)

Illumina sequencing can be used to identify where transposons are inserted in a genome by a technique called transposon sequencing [53]. This technique can for example be used to identify which gene has been knocked out by a transposon insertion. To do this, it is essential to prepare the transposon-containing genome before sequencing. The general principle is to use one primer that binds to the end of the transposon, and one primer that binds to a ligated adapter at the end of the sheared genomic fragment. By PCR, using a high-fidelity polymerase, the genome sequence flanking the inserted transposon can thus be specifically amplified before Illumina sequencing. The workflow for the preparation of genomic DNA for transposon sequencing is presented in Figure 6.

## INTRODUCTION



**Figure 6: Sample preparation for identification of transposon insertion sites using Illumina sequencing.**

To prepare a transposon mutant library for Illumina sequencing the genomic DNA from the mutants must first be sheared. Afterwards adapters have to be ligated to the fragments. In this workflow the adapters are ligated by first doing end-repair and A-tailing. The adapters are used to facilitate PCR amplification of the fragments containing transposon sequences.

The first step after isolation and purification of the DNA is to shear it into smaller fragments. This can be done by using enzymatic or mechanical shearing. To prepare the fragments for adaptor attachment, end-repair and A-tailing is carried out (Figure 6). End-repair means the repair of the double strand ends after shearing so that all the fragments have blunt ends. The following A-tailing adds an extra A to the ends of the fragment. The adaptor has a T-overhang, and will then attach to the A-tailed DNA fragment to make an A-T complementary binding between the adaptor and the fragments. The adaptors are then used to facilitate PCR amplification of the fragments containing transposon sequences. After this preparation, massively parallel sequencing is used to determine the location of the transposons in the genome studied. At the same time it is possible to determine the relative abundance of mutants containing a transposon at particular insertion site [54]. Transposon sequencing has previously been performed in *M. tuberculosis* to identify regions required for growth [55].

### 1.4.3 Sample multiplexing and barcoding

To separate different pools of samples, the primers used to amplify DNA before Illumina sequencing can have DNA indexes encoded. By adding index reads to the Illumina run parameters, these indexes will ensure that the various pools can be separated from each other after the end of the run. This strategy is called Illumina multiplexing [56]. Different pools of samples can then be analysed in one single run, without much increase in time and cost [57]. Furthermore, DNA barcodes can also be incorporated into the adapters, and increase the number of pools that can be sequenced and analysed in one single run even further. In this study, to identify the transposon insertion sites of around 10 000 *M. avium* transposon library mutants in one single Illumina sequencing run, a dual approach with primer indexing *and* adapter barcoding will be used.

A newly published article in Nature Communications, describes a similar approach as used in this thesis to organise a transposon mutant library of the *M. bovis* BCG vaccine strain by sample multiplexing and barcoding of the mutants before Illumina sequencing [58].

The *M. avium* transposon library mutants in this thesis were grown in 384-well plates, and then these wells were pooled into plate, row and column pools. The primer indexes referred to if the mutant was in a plate, row or column pool, while the adapter barcodes were used to define the number of the plate, row and column pools. In this way it was possible to identify which well the mutant was grown in originally, although all the pools were analysed in one single run.

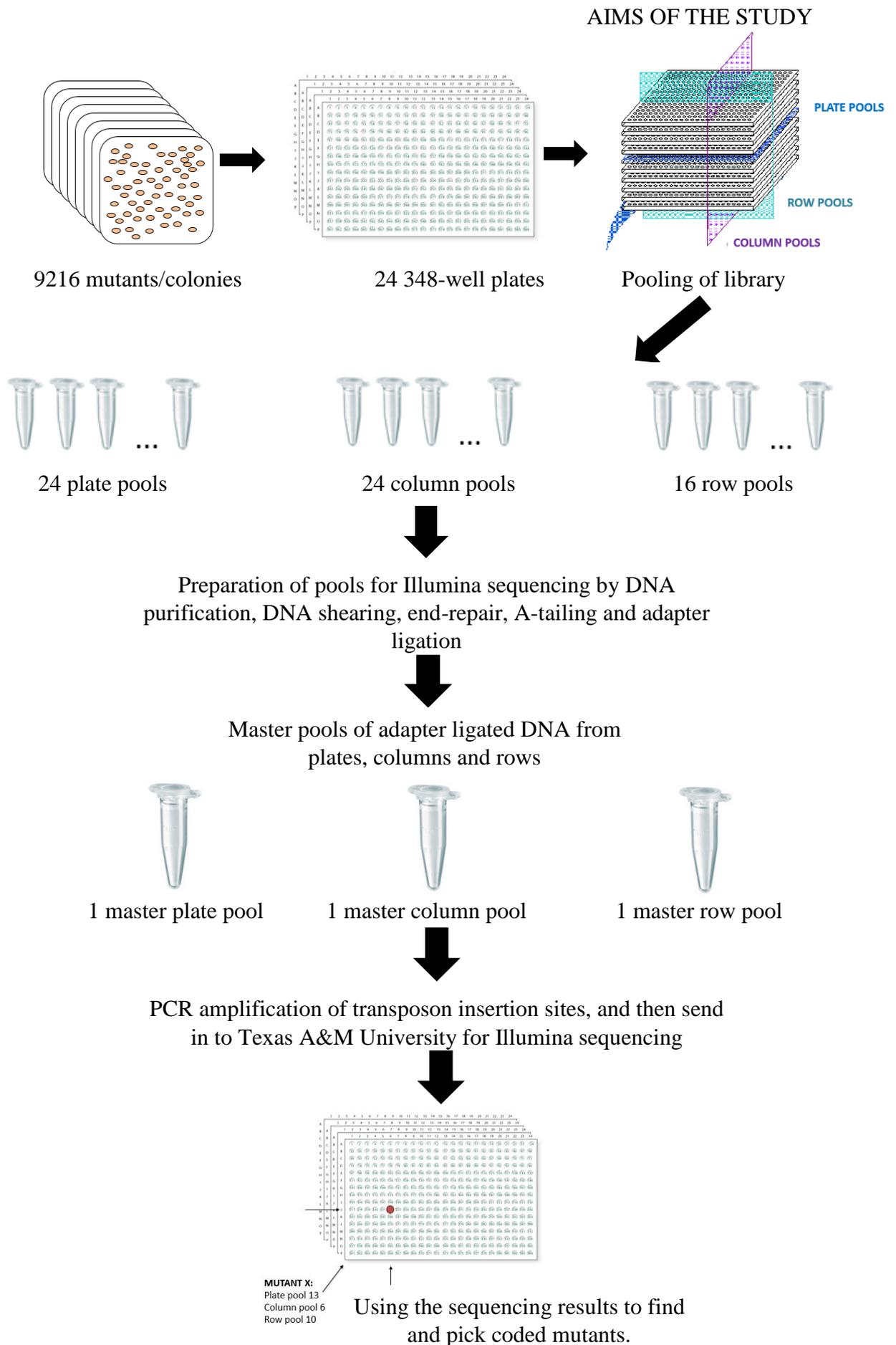
With a transposon library organised in this manner it is possible to perform large-scale screenings and study the phenotype of all or several mutants in the library at the same time. It is also possible to pick out a few genes of interest, e.g. genes that might be related to iron uptake. After the desired mutant has been isolated, its phenotype can be studied, e.g. to investigate how the mutant will grow in media with low and high content of iron.

### 2. Aims of the study

The overall aim with this study was to contribute to identifying and validating new potential mycobacterial drug targets, and by this contribute towards the future development of new and effective drugs against pathogenic mycobacteria.

The specific aims of this study was first of all to organise an *M. avium* transposon mutant library of around 10 000 mutants, and prepare it for Illumina sequencing, in a such way that mutants with specific insertion sites could be easily identified in a future application of the library. The library will then be a tool for investigating the role of the mutated genes have during infection, including the mutants' ability to survive under different environmental conditions. More specifically, such a library could be used e.g. to explore how gene disruptions in *M. avium* affect the bacteria's ability to survive intracellularly in human macrophages. This kind of research will be much easier with an organised mutant library of *M. avium* available, where it will be easy to pick out interesting mutants for further investigation. The library will also make it possible to conduct screenings of many thousand mutants at the same time and study their phenotype under different conditions. By the fact that *M. avium* is not particularly susceptible for homologues recombination to create targeted knockout mutants, a mutant transposon library is also of great value. A scheme over the preparation workflow of transposon library mutant for Illumina sequencing, is shown in Figure 7.

Secondly, this thesis aims to investigate selected genes potentially involved to mycobacterial iron uptake, which is essential for the mycobacteria to survive. For this aim, the non-pathogenic model organism *M. smegmatis*, which is susceptible to homologous recombination, will be used. Thus, the specific genes of interest can be directly targeted for gene mutagenesis. To assess these genes involvement in mycobacterial iron uptake, growth curve analysis of recombined *M. smegmatis* in media containing various iron concentrations will be performed.



**Figure 7: A scheme over the preparation workflow of transposon library mutants for Illumina sequencing.**

### 3. Materials and methods

#### 3.1 Tables of materials

##### 3.1.1 Growth media

The growth media used in this study are described in Table 1.

**Table 1: Growth media used in this study.**

<b>Growth media</b>	<b>Contents</b>	<b>Purpose</b>
<b>LB – Luria Bertoni</b>	5 g tryptone, 2.5 g yeast extract, 5 g NaCl. dH <sub>2</sub> O up to 500 ml. Autoclaved. Optional: Appropriate antibiotics.	Cultivation of <i>E. coli</i> DH5 $\alpha$
<b>LB agar</b>	5g tryptone, 2.5 g yeast extract, 5 g NaCl, 7.5 g agar. dH <sub>2</sub> O up to 500 ml. Autoclaved. Optionally: Appropriate antibiotics. Here 100 $\mu$ g/ml ampicillin.	Cultivation of <i>E. coli</i> DH5 $\alpha$
<b>7H9 Complete</b>	2.35 7H9 broth base, 450 ml dH <sub>2</sub> O, 1.2 ml 85% glycerol. After autoclaving: 50 ml ADC and 1.25 ml 20 % Tween 80. Optional: Appropriate antibiotics.	Cultivation of mycobacteria
<b>7H10</b>	9.5 g 7H10, 450 ml dH <sub>2</sub> O, 3 ml 85% Glycerol. After autoclaving; 500 $\mu$ l 150 mM FeCl <sub>3</sub> , 50 ml ADC. Optional: Appropriate antibiotics. Here 7.5 $\mu$ g/ml gentamicin.	Cultivation of mycobacteria
<b>SOC</b>	20 g bacto tryptone, 5 g bacto yeast extract, 2 ml of 5M NaCl, 2.5 ml of 1M KCl, 10 ml of 1M MgCl <sub>2</sub> , 10 ml of 1M MgSO <sub>4</sub> , 20ml of 1M glucose. dH <sub>2</sub> O up to 1 litre.	Transformation recovery
<b>Chelated Sauton's</b>	70 ml 85 % glycerol, 0.5 g KH <sub>2</sub> PO <sub>4</sub> , 2.2 g citric acid monohydrate, 4g asparagine, 10 g chelex. Adjust pH to 7.4 with NaOH. Stir for in 2-3 days, sterile filter and add 1 g MgSO <sub>4</sub> and 0.05% Tween.	Growth medium for growth curve analyses

## 3.1.2 Buffers and solutions

The buffers and solutions used in this study are described in Table 2.

Table 2: Buffers and solutions used in this study.

<b>Buffers and solutions</b>	<b>Contents / Manufacturer</b>	<b>Purpose</b>
<b>PBS (Phosphate Buffered Saline)</b>	Sigma-Aldrich	Bacteria medium while bead beating
<b>50 x TAE buffer (Tris-Acetate-EDTA buffer)</b>	242 g Tris base, 57.1 ml glacial acetic acid, 37.2 Na <sub>2</sub> EDTA·2H <sub>2</sub> O pH ~8.5, H <sub>2</sub> O up to 1 L	Agarose gel electrophoresis
<b>6x Orange G Loading Dye</b>	40 mg Orange G, 4 g sucrose, 7 ml deionized H <sub>2</sub> O	Agarose gel electrophoresis
<b>Gel Loading Dye, Blue (6X)</b>	New England Biolabs	Agarose gel electrophoresis
<b>EDTA (Ethylenediaminetetraacetic acid)</b>	Sigma	Stopping the activity of fragmentase
<b>Glycerol 85 %</b>	Merck	Freezing of bacterial stock culture. Washing of cells before electroporation.
<b>2'2'-dipyridyl FeCl<sub>3</sub></b>	Alfa Aesar Sigma-Aldrich	Iron chelator Sauton's medium and 7H9 iron supplement
<b>ADC (Albumin dextrose catalase)</b>	4.25 g NaCl, 25 g BSA, 10 g dextrose, 0.015 g catalase. Add dH <sub>2</sub> O up to 500 ml. Sterile filtrate.	7H10 medium supplement
<b>DMSO (Dimethyl sulfoxide)</b>	Sigma-Aldrich	Inhibit formation of secondary structures in DNA during PCR.

## 3.1.3 Primers

All primers used in this study were ordered from Sigma-Aldrich. Primers used to confirm presence of kanamycin resistance gene are shown in Table 3.

Table 3: Primers used in this study to confirm the presence of the kanamycin resistance gene in the transposon library mutants.

<b>Primer</b>	<b>Sequence, 5'-3'</b>
<b>Kan_F</b>	TGGATTGCACGCAGGTTCTC
<b>Kan_R</b>	CGTCAAGAAGGCGATAGAAG

## MATERIALS AND METHODS

Sol\_Mar primers are transposon binding primers, that are used in preparation for Illumina sequencing. Multiple Sol\_Mar primers were mixed to prevent the flow cell image from being overwhelmed by simultaneous fluorescence signals from identical first base pairs. Each of the four Sol\_Mar primers mixed has a different first nucleotide in the variable site and has also a distinct length. See Table 4. The Sol\_Mar primers were mixed 1:1:1:1 to give a final Sol\_Mar mix at 10 mM. The transposon homology site shown in Figure 8 is 5 base pair downstream from the transposon junction. These 5 remaining base pair are “TGTTA”.

(Attachment homology for Illumina chip; Illumina Read 1 primer; Variable site; transposon homology).

Figure 8: The structure of the transposon primers.

Table 4: The transposon binding primers used in the preparation step for Illumina sequencing.

The colour coding of the primers is described in Figure 8.

Primer	Sequence, 5'-3'	Purpose
Sol_Mar_NI	GTCTAGAGACCGGGACTTATC	Confirm that the adapters are inserted in the transposon junction and have the correct barcode.
Sol_Mar	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCTC GGGGACTTATCAGCCAACC	Transposon binding primer
Sol_Mar_1b	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCTT CGGGACTTATCAGCCAACC	Transposon binding primer
Sol_Mar_4b	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCTG ATACGGGGACTTATCAGCCAACC	Transposon binding primer
Sol_Mar_5b	AATGATACGGCGACCACCGAGATCTACAC TCTTTCCCTACACGACGCTCTTCCGATCTA TCTACGGGGACTTATCAGCCAACC	Transposon binding primer

The Sol\_API1-primers bind to the adapters and contain indexes coding for plate, column or row (Figure 9). The adapter binding primers used in this study are shown in Table 5.

(Attachment homology for Illumina chip; Index; Illumina Read 2 primer; adapter homology).

5'CAAGCAGAAGACGGCATAACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGT  
GCTCTTCCGATCTATGATGGCCGGTGGATTTGTG-3'

Figure 9: The structure of the adapter primers, Sol\_API1.

**Table 5: The adapter binding primers used in the preparation step for Illumina sequencing.**  
The colour coding of the primers is described in Figure 9.

<b>Primer</b>	<b>Sequence, 5'-3'</b>	<b>Purpose</b>
<b>Sol_AP1_NI</b>	ATGATGGCCGGTGGATTTGTG	Confirm that the adapters are inserted in the transposon junction and have the correct barcode.
<b>Sol_AP1_tagged_57</b>	CAAGCAGAAGACGGCATAACGAGAT AAGTAGAGGTGACTGGAGTTCAGACGTGT GCTCTCCGATCTATGATGGCCGGTGGATT TGTG	Adapter primer: Code for plate
<b>Sol_AP1_tagged_100</b>	CAAGCAGAAGACGGCATAACGAGAT ACACGATCGTGACTGGAGTTCAGACGTGT GCTCTCCGATCTATGATGGCCGGTGGATT TGTG	Adapter primer: Code for column
<b>Sol_AP1_tagged_473</b>	CAAGCAGAAGACGGCATAACGAGAT CGCGCGGTGTGACTGGAGTTCAGACGTGT GCTCTCCGATCTATGATGGCCGGTGGATT TGTG	Adapter primer: Code for row

The primers used to confirm the knock-out of the genes *msmeg\_4898* and *msmeg\_4900* are shown in Table 6.

**Table 6: Primers used in this study to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.**

<b>Primer</b>	<b>Sequence, 5'-3'</b>
<b>4898_WT_F</b>	CACCTTGGTCAGGCCGAATG
<b>4898_WT_R</b>	TGACGTCGCGGGTGTGAAAGT
<b>4898_WT_UP_F</b>	AGAAAGAGCAGGACCGATAG
<b>4898_WT_UP_R</b>	TGCGACTTTCGAGGAGAACG
<b>GENT_KO_F</b>	CGTAGCCACCTACTCCCAAC
<b>GENT_KO_R</b>	CACGCATCGTCAGGCATTGG

The sequencing primers used for Sanger sequencing are shown in Table 7.

**Table 7: Sequencing primers used in this study.**

<b>Primer</b>	<b>Sequence, 5'-3'</b>	<b>Purpose</b>
<b>Kan_seq2</b>	CTTCCTCGTGCTTTACGG	Sanger sequencing to confirm that the transposon is inserted at different places in the genome.
<b>pJet 1.2 forward sequencing primer</b>	CGACTCACTATAGGGAGAGCGGC	Sanger sequencing to confirm inserts after TA-cloning and blunt end cloning.

## MATERIALS AND METHODS

### 3.1.4 Adapters

The adapter 1 used in this study was ordered from Invitrogen and the number 2 adapters were ordered from Sigma-Aldrich. The barcodes to the number 2 adapters are shown in Table 8.

Purpose: To code each number of plate, column and row.

**Adapter 1:** 5' – TACCACGACCA-NH<sub>2</sub> – 3'

**Adapter 2:** (adapter primer homology; BARCODE ; adapter annealing site)

5' – ATGATGGCCGGTGGATTGTG NNNNNNNN TGGTCGTGGTAT – 3'

Table 8: The barcodes of the adapters number 2 used in this study.

Adapter 2	Barcode
1	ATCACGTT
2	CGATGTTT
3	TTAGGCAT
4	TGACCACT
5	ACAGTGGT
6	GCCAATGT
7	CAGATCTG
8	ACTTGATG
9	GATCAGCG
10	TAGCTTGT
11	GGCTACAG
12	CTTGTA CT
13	TGGTTGTT
14	TCTCGGTT
15	TAAGCGTT
16	TCCGTCTT
17	TGTACCTT
18	TTCTGTGT
19	TCTGCTGT
20	TTGGAGGT
21	TCGAGCGT
22	TGATACGT
23	TGCATAGT
24	TTGACTCT

### 3.1.5 Ladders

The ladders used in this study are shown in Figure 10.

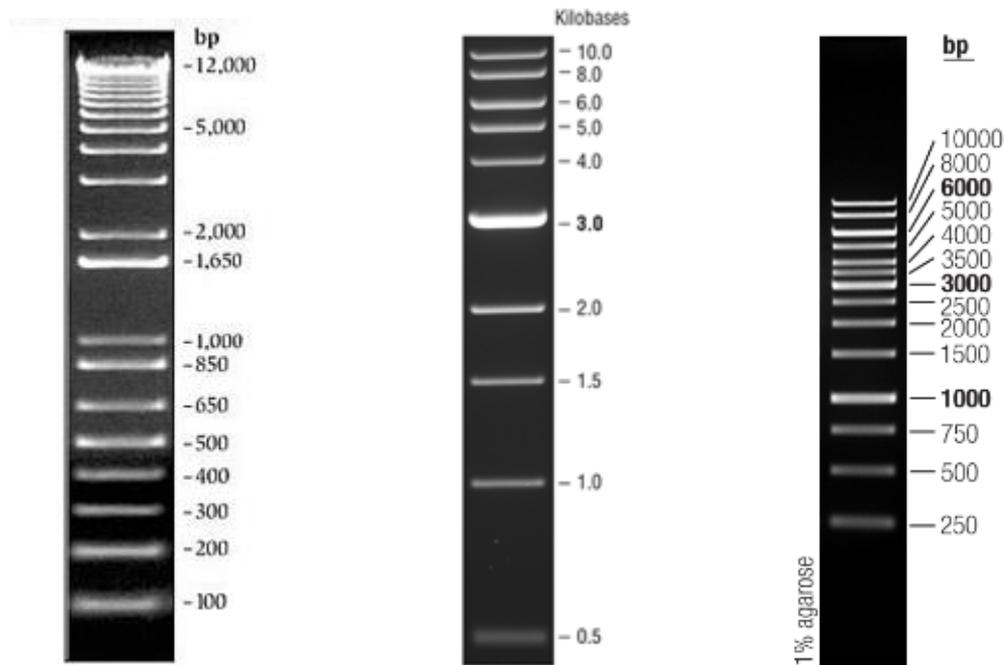


Figure 10: Ladders used in this thesis.

1 kb plus DNA ladder from Life Technologies, 1 kb DNA ladder from NEB (New England Biolabs), Generuler 1kb DNA ladder from Life Technologies; respectively.

### 3.1.6 PCR kits

The PCR kits used in this study are shown in Table 9.

Table 9: PCR kits used in this thesis.

PCR kit	Manufacturer	Purpose
<b>GoTaq Green Master Mix</b>	Promega	Finding the gene for kanamycin resistance. Finding mutants for growth curve analysis.
<b>KOD Xtreme Hot Start DNA Polymerase</b>	Novagen	Illumina sequencing preparations.
<b>BigDye Terminator v1.1 Cycle Sequencing Kit</b>	Applied Biosystems	Preparation for Sanger sequencing

## MATERIALS AND METHODS

### 3.1.7 Cloning kits

The cloning kits used in this study are shown in Table 10.

Table 10: Cloning kits used in this thesis

Cloning kit	Manufacturer	Purpose
<b>InsTAclone PCR Cloning Kit</b>	Thermo Scientific	TA-cloning
<b>CloneJET PCR Cloning Kit</b>	Thermo Scientific	Blunt cloning

The protocols and tables (Table 11 and 12) under are based on the protocols from the manufacturer and adapted to what was done in the laboratory. The ligation mixture was immediately after cloning used for transformation.

#### 3.1.7.1 InsTAclone PCR Cloning Kit

Table 11: The set up for the ligation reaction when using InsTAclone PCR Cloning Kit.

Component	Volume
<b>5x Reaction Buffer</b>	6 $\mu$ l
<b>Non-purified PCR product</b>	1 $\mu$ l
<b>Vector pTZ57R/T</b>	0.5 $\mu$ l
<b>Nuclease free H<sub>2</sub>O</b>	19 $\mu$ l
<b>T4 DNA ligase</b>	1 $\mu$ l
<b>Total volume</b>	30 $\mu$ l

The control was made by adding nuclease free water instead of PCR-product. The ligation reaction was briefly mixed by a vortexer and then centrifuge for 5 seconds. The ligation mixture was then incubated over night at 4 °C.

#### 3.1.7.2 CloneJET PCR Cloning Kit

Table 12: The set up for the ligation reaction when using CloneJET PCR Cloning Kit.

Component	Volume
<b>2x Reaction Buffer</b>	5 $\mu$ l
<b>Non-purified PCR product</b>	0.5 $\mu$ l
<b>pJET 1.2 / blunt cloning vector (50 ng/<math>\mu</math>l)</b>	0.5 $\mu$ l
<b>Nuclease free H<sub>2</sub>O</b>	3.5 $\mu$ l
<b>T4 DNA ligase</b>	0.5 $\mu$ l
<b>Total volume</b>	10 $\mu$ l

The ligation reaction was set up on ice. The control was made by adding nuclease free water instead of PCR-product. The ligation reaction was briefly mixed by a vortexer and then centrifuged for 5 seconds. The ligation mixture was then incubated for 30 minutes at room temperature.

### 3.2 Dilution series to find the amount of bacteria in the transposon library

A culture was made from the library stock and cultured for 2 weeks. A tenfold sequential dilution series reaching from  $10^{-1}$  to  $10^{-10}$  was made by diluting the culture in 7H9 Complete.

The OD<sub>600</sub> was measured for the highest dilution to find the OD for the stock solution. By this OD and the already known CFU at OD = 1 for the library stock, an estimate of how many bacteria would grow on each agar were made. In addition, 100 µl from each dilution were plated on 7H10 agar and the amount of bacteria was determined after an incubation time of 3 weeks. This was done to control the theoretical calculations.

For each dilution 3 times 10 µl were separately plated on the same 7H10-agar plate with 20 µg/ml kanamycin. After three weeks the amount of colonies in each parallel of each agar were manually counted. Therefore, only parallels with clearly distinguishable colonies were used for calculations. Parallels with only 1 or 2 colonies were not seen as reliable because of the low number. The mean value of the colonies per ml was calculated for all the counted parallels together and an estimate of the bacteria concentration in the library stock culture was found.

### 3.3 Purification of DNA

In this study, five different purification kits were used in order to purify DNA from different sources and further analysed in different ways. The protocols presented in this section is rewritten protocols from the manufacturers. The protocols are adapted to describe how the purification was done in the laboratory in this thesis. Before using the first purification kit (MasterPure DNA Purification Kit from Epicentre) the bacteria were lysed by bead beating in order to get to the DNA.

#### 3.3.1 Bead beating of bacteria

For each pool of bacteria the respective tubes were centrifuges, and then solved in 600 µl PBS. Then this was heat inactivated at 95°C in 20 min, so it would not be infectious anymore. The

## MATERIALS AND METHODS

DNA was relieved by 0,1mm glass beads, that first were autoclaved and cooled down. The tubes with the glass beads and the bacteria were bead beaten with the FastPrep-instrument at 6000 rpm for 45 seconds three times.

### 3.3.2 MasterPure DNA Purification Kit from Epicentre

Purpose: To purify DNA from bead beaten bacteria.

The samples were centrifuged quickly and the supernatant collected. This supernatant was then evaporated to a volume of 150  $\mu$ l. Then 1  $\mu$ l Protein K diluted into 150  $\mu$ l of 2XT and C Lysis Solution was added to the 150  $\mu$ l of sample and mixed thoroughly. This mix was incubated at 65°C for 15 minutes. During this incubation the samples were vortexed every 5 minutes. Subsequently, the samples were cooled 37°C and 1 $\mu$ l of 5  $\mu$ g/ $\mu$ l RNase A was added and then everything was mixed. The samples were then placed on ice for 3-5 minutes. Next, 150  $\mu$ l of MPC Protein Precipitation Reagent were added to the lysed sample and vortexed vigorously for 10 seconds. The debris was centrifuged down at 4°C for 10 minutes at 10 000 g. The supernatant was then transferred to a clean Eppendorf tube and the pellet discarded. Then 500  $\mu$ l of isopropanol were added to the recovered supernatant and the tube inverted 30-40 times. The DNA was then centrifuges down at 4°C for 10 minutes at 10 000 g. Afterwards, the isopropanol was carefully poured off without dislodging the DNA pellet. The pellet was rinsed twice with 70% ethanol, still being careful to not dislodge the pellet. Finally, all the residual ethanol was removed and the DNA resuspended in 35 $\mu$ l sterile and deionised water.

### 3.3.3 QIAquick PCR Purification Kit

Purpose: To purify the DNA between each step in the preparation workflow of the mutants before Illumina sequencing.

5 volumes of Buffer PB were added to 1 volume of the PCR sample and mixed. A QIAquick spin column was then placed in a provide 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick column and centrifuged for 1 minute. Afterwards, the flow-through was discarded, and the QIAquick column was placed back into the same tube. 750  $\mu$ l Buffer PE was added to the QIAquick column and centrifuged for 1 minute. Subsequently, the flow-through was discarded and the QIAquick column was placed back in the same tube. Finally, the QIAquick column was placed into a clean 1.5 Eppendorf tube, and the DNA was eluted by

adding 50 µl of sterile and deionised water to the centre of the QIAquick membrane, letting the column stand for 1 minute and then centrifuging for 1 minutes.

### 3.3.4 QIAGEN DyeEx 2.0 SpinKit

Purpose: To remove BigDye from the sequencing reaction before Sanger sequencing.

First, the spin column was vortexed gently to resuspend its resin. Then the cap of the column was loosened a quarter turn to avoid vacuum inside the spin column. The bottom closure of the spin column was broken off and the spin column was then placed in a 2 ml collection tube. This was then centrifuged at 750 g for 3 minutes. Afterwards, the spin column was transferred to a clean Eppendorf tube. The sequencing reaction was carefully applied onto the centre of the slanted gel bed surface in the spin column. Subsequently, the spin column in the Eppendorf tube was centrifuged for 3 minutes at 750 g. The spin column was then discarded, and the eluate in the Eppendorf tube contained the purified DNA.

### 3.3.5 PureYield Plasmid Miniprep System

Purpose: To purify plasmid DNA before transformation.

First, 1.5 ml of bacterial culture was centrifuged for 30 seconds at maximum speed in a microcentrifuge. Afterwards, the supernatant was discarded, and an additional 1.5 ml of bacterial culture was added to the same tube. This was also centrifuged for 30 seconds at maximum speed in a microcentrifuge. 600 µl of water were then added to the cell pellet and this was resuspended completely. Next, 100 µl of Cell Lysis Buffer were added and mixed by inverting the tube 6 times. Then 350 µl of cold (4-8°C) Neutralization Solution were added and mixed thoroughly by inverting. This was centrifuged at maximum speed in a microcentrifuge for 3 minutes. Subsequently, the supernatant was transferred to a PureYield Minicolumn without disturbing the cell debris pellet. The minicolumn was placed into a Collection Tube and centrifuged at maximum speed in a microcentrifuge for 15 seconds. The flow-through was discarded and the minicolumn was placed into the same Collection Tube. Afterwards, 200 µl of Endotoxin Removal Wash (ERB) were added to the minicolumn, and centrifuged at maximum speed in a microcentrifuge for 30 seconds. Then the minicolumn was transferred to a clean 1.5 ml Eppendorf tube, and 30 µl of nuclease-free water were directly added to the minicolumn matrix.

## MATERIALS AND METHODS

This was then standing for 1 minute at room temperature. Finally, the tube with the minicolumn was centrifuged for 15 seconds to elute the plasmid DNA.

### 3.3.6 QIAquick Gel Extraction Kit

Purpose: To purify DNA from agarose gel.

The tubes, used for storing the gel pieces after cutting them out from the agarose gel, were weighted. Then each tube with the gel slice was weighted. By subtracting the weight of the tube the weight of the gel slice was found. 3 volumes of Buffer QG were added to 1 volume of gel (100 mg ~ 100 µl). This was then incubated at 50°C for 10 min or until the gel slice was completely dissolved. To help dissolving the gel, the tube was mixed by vortexing every 2-3 minutes during the incubation. After the gel slice was completely dissolved the colour of the mixture was checked if it was still yellow. Then 1 gel volume of isopropanol was added to the sample and mixed. Next, a QIAquick spin column was placed in a provided 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick column and centrifuged for 1 minute. Afterwards, the flow-through was discarded and the QIAquick column was placed back in the same collection tube. To remove all traces of agarose 0.5 ml of Buffer QG was added to the QIAquick column and centrifuged for 1 minute. To wash, 0,75 ml of Buffer PE was added to the QIAquick column. The column was standing in 5 minutes before it was centrifuged for 1 minute. Subsequently, the flow-through was discarded and the QIAquick column was centrifuged for an additional 1 min at 10 000 g. Finally, the QIAquick column was placed into a clean 1.5 Eppendorf tube, and the DNA was eluted by adding 50 µl of sterile and deionised water to the centre of the QIAquick membrane, letting the column stand for 1 minute and then centrifuging for 1 minutes.

### 3.4 Spectrophotometry

Spectrophotometry was used to determine the concentration of purified DNA, and also in order to determine bacteria concentration in solution. For both instruments, a blank had to be measured first, before measuring the samples. The blank is then the solution the DNA and the bacteria is in when they are measured, and this is done to subtract the absorbance of this solution from what is desired to measure.

To measure concentration of purified DNA, the compact UV-spectrophotometer NanoDrop 1000 from Thermo Fischer Scientific was used. Here only 1-2  $\mu$ l sample is needed for measurement.

DNA concentration is estimated by measuring the absorbance at a wavelength of 260 nm. The absorbance of the sample and its DNA concentration, are proportional values. By measuring the absorption ratio of 260/280 nm, the protein contamination in the sample can be measured, due to proteins maximum absorbance at 280 nm. The 260/280 nm- ratio should generally be over 1.8, to be accepted as “pure” for DNA measurements [59]. The DNA concentrations measured at NanoDrop were given in ng/ $\mu$ l.

To measure bacteria concentration, the compact spectrophotometer Ultrospec 10 Cell density meter from Amersham Bioscience was used. Bacteria was measured at a wavelength of 600 nm.

### 3.5 Robots used at SINTEF

#### 3.5.1 Colony picking robot

Genetix QPixII with Software QSoft XP Picking.

Purpose: Pick colonies from 23 x 23 cm agar plates to 384 well plates

Parameter settings:

Diameter: 10-14

Centre Frame: 116- 178

Roundness: 0.75

Picking head with 384 pickers used.

## MATERIALS AND METHODS

### 3.5.2 Pipetting robots

Biomek NX<sup>P</sup> – Laboratory Automation Workstation from Beckman Coulter with software Biomek

Purpose: To fill each well in the 384-well plates with 50 µl 7H9 complete with 20 µg/ml kanamycin before picking the colonies with the colony picking robot.

Freedom EVO 200 from TECAN with Software EVOsim.

Purpose: Pooling the plates into master plates. 3 µl of the bacteria culture from each well in plates were pooled into a master plate in the same position as in the original plate.

MultiFlo Dispenser from BioTek with the software Liquid Handling Control.

Purpose: Adding glycerol into the well in the 384 well plates that were going to be frozen down, to make a concentration of 15 % glycerol in each well.

### 3.6 PCR

PCR is a method generally used to amplify a predetermined DNA sequences in vitro. This is done by specifically designed primers, a thermostable polymerase, deoxynucleotide triphosphates (dNTP's) in addition to the DNA template. These ingredients are mixed together with an appropriate buffer in a PCR tube. The PCR amplification starts by denaturation of the DNA in the sample by heat (> 94°C), to separate the double helix into single strands. Then the primers are hybridised to each strand, by lowering the temperature. This annealing temperature is usually set to 40-65°C. Thereafter, the temperature is raised to the optimum of the DNA polymerase. In this step new DNA-strands is synthesised using the original or newly synthesised DNA-molecule as a template. By repeating these steps in cycles a pre-determined number of times, the amount of DNA sequences amplified is increased exponentially [59].

## 3.6.1 PCR for in-house Sanger sequencing of genomic DNA

The sequencing reaction mixes and cycling conditions for PCR amplification done in this study are shown in Table 13-22.

**Table 13: Sequencing reaction mix for PCR for in-house Sanger sequencing of genomic DNA to identify transposon insertion sites.**

<b>Component</b>	<b>Volume</b>
<b>BigDye v.1.1</b>	2 $\mu$ l
<b>5 x Buffer</b>	3 $\mu$ l
<b>Template</b>	5 $\mu$ g
<b>Primer Kan_seq2</b>	3.2 pmol
<b>dH<sub>2</sub>O</b>	Up to 20 $\mu$ l

**Table 14: Cycling conditions for PCR for in-house Sanger sequencing of genomic DNA to identify transposon insertion sites.**

<b>Step</b>	<b>Temperature and time</b>
<b>1. Polymerase activation</b>	95°C for 4 min
<b>2. Denature</b>	95°C for 30 sec
<b>3. Annealing</b>	52°C for 30 sec
<b>4. Extension</b>	60°C for 4 min
<b>Repeat steps 2-4</b>	60 cycles total
<b>5. Hold</b>	4°C

## 3.6.2 Confirming of the presence of the kanamycin resistance gene

**Table 15: Reaction mix for PCR used to confirm the existence of the kanamycin resistance gene in the transposon library mutants.**

<b>Component</b>	<b>Volume</b>
<b>2 x GoTaq green master mix</b>	5 $\mu$ l
<b>Primer Kan_F</b>	0.5 $\mu$ l
<b>Primer Kan_R</b>	0.5 $\mu$ l
<b>30-80 ng genomic DNA</b>	0.5 $\mu$ l
<b>dH<sub>2</sub>O</b>	3.5 $\mu$ l
<b>Total volume</b>	10 $\mu$ l

**Table 16: Cycling conditions for PCR used to confirm the existence of the kanamycin resistance gene in the transposon library mutants.**

<b>Step</b>	<b>Temperature and time</b>
<b>1. Polymerase activation</b>	95°C for 2 min
<b>2. Denature</b>	95°C for 45 sec
<b>3. Annealing</b>	58°C for 45 sec
<b>4. Extension</b>	72°C for 1 min
<b>Repeat steps 2-4</b>	25 cycles total
<b>5. Final extension</b>	72°C for 5 min
<b>6. Hold</b>	4°C

## MATERIALS AND METHODS

### 3.6.3 Small scale PCR after adapter ligation

**Table 17: Reaction mix for small scale PCR after adapter ligation of the transposon library mutants.**

<b>Component</b>	<b>Volume</b>
<b>2 x Xtreme Buffer</b>	5 $\mu$ l
<b>dNTP's (2 mM each)</b>	2 $\mu$ l
<b>dH<sub>2</sub>O</b>	variable
<b>Sol_Mar_NI</b>	0.3 $\mu$ l
<b>Sol_AP1_NI</b>	0.3 $\mu$ l
<b>Template 20 ng</b>	variable
<b>KOD Xtreme Hot Start DNA Polymerase (1 U/<math>\mu</math>l)</b>	0.2 $\mu$ l
<b>Total volume</b>	10 $\mu$ l

**Table 18: Cycling conditions for small scale PCR after adapter ligation of the transposon library mutants.**

<b>Step</b>	<b>Temperature and time</b>
<b>1. Polymerase activation</b>	94°C for 2 min
<b>2. Denature</b>	98°C for 10 sec
<b>3. Annealing</b>	59°C for 30 sec
<b>4. Extension</b>	68°C for 1 min
<b>Repeat steps 2-4</b>	30 cycles total
<b>5. Final extension</b>	68°C for 7 min
<b>6. Hold</b>	4°C

### 3.6.4 PCR for addition of Illumina attachment and read homology

**Table 19: Reaction mix for PCR for addition of Illumina attachment and read homology.**

<b>Component</b>	<b>Volume</b>
<b>2 x Xtreme Buffer</b>	5 $\mu$ l
<b>dNTP's (2 mM each)</b>	2 $\mu$ l
<b>dH<sub>2</sub>O</b>	variable
<b>Sol_Mar_mix</b>	0.3 $\mu$ l
<b>Sol_AP1_57/100/473</b>	0.3 $\mu$ l
<b>Template 100 ng</b>	variable
<b>KOD Xtreme Hot Start DNA Polymerase (1 U/<math>\mu</math>l)</b>	0.2 $\mu$ l
<b>Total volume</b>	10 $\mu$ l

**Table 20: Cycling conditions for PCR for addition of Illumina attachment and read homology.**

<b>Step</b>	<b>Temperature and time</b>
<b>1. Polymerase activation</b>	94°C for 2 min
<b>2. Denature</b>	98°C for 10 sec
<b>3. Annealing</b>	59°C for 30 sec
<b>4. Extension</b>	68°C for 1 min
<b>Repeat steps 2-4</b>	30 cycles total
<b>5. Final extension</b>	68°C for 7 min
<b>6. Hold</b>	4°C

3.6.5 Confirming knock-out of the genes *msmeg\_4898* and *msmeg\_4900*Table 21: Reaction mix used for confirming knock-out of the genes *msmeg\_4898* and *msmeg\_4900* and thereby choosing mutants to growth curve analysis.

Component	Volume
<b>2 x GoTaq green master mix</b>	5 µl
<b>Forward primer</b>	0.5 µl
<b>Reverse primer</b>	0.5 µl
<b>Bacteria culture</b>	0.5 µl
<b>dH<sub>2</sub>O</b>	3 µl
<b>DMSO</b>	0.5 µL
<b>Total volume</b>	10 µl

Table 22: Cycling conditions used for confirming knock-out of the genes *msmeg\_4898* and *msmeg\_4900* and thereby choosing mutants to growth curve analysis.

Step	Temperature and time
<b>1. Polymerase activation</b>	95°C for 2 min
<b>2. Denature</b>	95°C for 45 sec
<b>3. Annealing</b>	X°C for 45 sec*
<b>4. Extension</b>	72°C for 1 min
<b>Repeat steps 2-4</b>	25 cycles total
<b>5. Final extension</b>	72°C for 5 min
<b>6. Hold</b>	4°C

\*The annealing temperature was set to 5°C below the melting temperature of the primer in the primer pair with the lowest melting temperature.

## 3.7 Gel electrophoresis

DNA can be detected by using gel electrophoresis, which separates DNA-fragments according to size. By supply of voltage, the small fragments are moving fastest, while the bigger fragments are hindered by the polymer network in the gel and are thereby moving slower. By using a ladder can the base pair length of the DNA also be estimated. The separation occurs by the DNA, which has negatively charged phosphates groups, moving toward the positive pole in an electrical field [40]. The DNA-fragments in the gel can be visualised by addition of GelRed, which will bind to the DNA and fluoresce by UV-light [60].

1 % agarose was made by adding 300 µl 1 x TAE buffer (Table 2) to 3 g of SeaKem LE Agarose (Lonza). Gels that were used to cut out slices used for further work, were made 2 %. 2 % agarose was made by adding 300 µl 1 x TAE buffer to 6 g of SeaKem LE Agarose. The mixture was heated in a microwave until boiling three times, to make sure that the agarose was completely

## MATERIALS AND METHODS

dissolved in the buffer. It was stored at around 60°C. 1 µl of GelRed was added and mixed to each 50 µl used in a gel. Then the gel was poured out in a gel form with a comb, to polymerise. After the gel was polymerised the gel was moved to an electrophoresis tray and covered with 1 x TAE buffer. The DNA samples were added to the wells and an electric current of 90 volt was run through the gel. The DNA in the gel was visualised with Gel Logic 212 PRO using the software Carestream Molecular Imaging.

### 3.8 Preparation of samples for external analysis

#### 3.8.1 GATC-sequencing

DNA for Sanger sequencing by GATC was sent in 5 µl of 80-100ng/µl purified plasmid DNA mixed with 5 µl 5 µM of the sequencing primer to the vector used in transformation.

#### 3.8.2 Preparation Covaris shearing:

6 µg of purified DNA were diluted in 100 µl dH<sub>2</sub>O and delivered to the Genomic Core Facility at NTNU for mechanical shearing by the Covaris – Focused ultrasonicator M220.

#### 3.8.3 Preparing for BioAnalyzer

No preparation was needed before delivering to the Genomic Core Facility at NTNU. 1µl sample was needed for analysing at the BioAnalyzer.

### 3.9 Analysing data from Sanger sequencing

Clone-Manager was used to align the sequencing data to a predicted premade result. In this way it was possible to find the TGTTA-sequence in the transposon, the adapter barcode and the adapter primer index. NCBI Nucleotide BLAST was used to find in which gene the transposon was inserted. Since the *M. avium* strain used in the library is not published yet, the strain *M. avium* 104 was used as a reference genome when using BLAST.

### 3.10 Transformation

#### 3.10.1 Transformation in competent *E. coli DH5α*

Transformation was performed to deliver plasmid DNA into the bacteria. This was used to perform blunt-cloning, TA-cloning and as a step before checking sequencing of fragments prepared during the workflow in section 3.12.

50 µl competent *E. coli DH5α* was thawed on ice. 5 µl of 10 to 100 ng DNA were then added to the competent cells, and then leaved on ice for 30 minutes. Next, a heat shock at 42°C in 45 seconds was preformed, and then a rest on ice for 2 minutes. 500 µl pre-heated SOC (Table 1) at 37°C were added, and then incubated at 37°C with shaking for 1 hour. After the last incubation, the cell mixture was centrifuged for 3 minutes at 1.7 g. The most of the supernatant was discarded, and the rest (~200 µl) was resuspended. The cell mixture was plated on pre-heated 37°C plates with antibiotics, and incubated over night at 37°C.

### 3.10.2 Transformation in *M. smegmatis*

It was created a master mix with 1-1.5 µg DNA, in this study PCR-product of a gentamicin resistance gene and flanking regions (section 3.12), per 400 µl of competent cells. In this study it was created 3 master mixes per mutant condition. 400 µl of the master mix were transferred to a pre-cooled 2 mm gap electroporation cuvette and electroporated. The settings for the electroporation were 2.5 kV, 1000 Ω and µF. The time constant after electroporation should ideally be above 20.

Afterwards, the electroporated cells were transferred to tubes containing 2 ml 7H9 complete without antibiotics. The five tubes per mutant category with electroporated cells were incubated for 4 hours, before they were centrifuged and the pellet resuspended in 400 µl 7H9, and plated on 4x5 7H10 plated with 7,5 µg/ml gentamicin and 150 µM FeCl<sub>3</sub>. These plates were incubated for 6 day at 37°C.

## 3.11 Preparation of transposon library mutants for Illumina sequencing

### 3.11.1 DNA shearing

#### **Using an acoustic sonicator; Covaris - Focused ultrasonicator M220**

Parameter setting:

Duty Cycle: 10 %

Intensity: 50

Cycles/ Burst: 200

Times: 110 s

## MATERIALS AND METHODS

### Using NEBNext dsDNA Fragmentase:

First, the NEBNext Fragmentase was vortexed for 3 seconds, span quickly and placed on ice. The reaction mix then mixed is shown in Table 23.

Table 23: Reaction mix used for fragmentase shearing.

Component	Volume
DNA 5 µg	1-16 µl
10X Fragmentase Reaction Buffer v2	2 µl
Sterile water	variable
Final volume	18 µl

Then 2.0 µl dsDNA Fragmentase were added and the mixture vortexed for 3 seconds. Subsequently, the mixture was incubated at 37°C for the optimised time 25 minutes, to generate the desired fragment size. To stop the shearing 5 µl of 0.5 M EDTA was added.

The fragmented DNA was afterwards cleaned up by using QIAquick PCR Purification Kit. Protocol is found in section 3.3.3.

The fragmentation results were checked by using 5 µl of the fragmented DNA at gel electrophoresis. See section 3.7.

#### 3.11.2 End-repair

All the fragmented DNA, ca. 42 µl after gel electrophoresis, was then used for end-repairing. The DNA was mixed with 10 µl of NEBNext End Repair Reaction Buffer (10x), 5 µl of NEBNext End Repair Enzyme Mix and 43 µl of sterile water, so the final volume would be 100 µl. This was then incubated at 20°C for 30 minutes. Afterwards, the DNA was purified by using QIAquick PCR Purification Kit, and eluted in 46 µl sterile water.

#### 3.11.3 A-tailing

All the end-repaired and purified DNA was mixed with 5 µl NEBNext dA-Tailing Reaction Buffer and 3 µl Klenow Fragment (3'→5' exo<sup>-</sup>). The total volume then became 50 µl. This mixture was then incubated at 37°C for 30 minutes. Afterwards, the DNA was purified by using QIAquick PCR Purification Kit. The DNA was eluted in 50 µl water, which was pre-heated to 37°C. Then the tube with the QIAquick column was incubated at 37°C for 10 minutes before spinning down.

#### 3.11.4 Making adapter mix

To make the adapters (Table 8 in section 3.1.4) the following components were mixed together for each type of adapter made:

- 4  $\mu$ l 2mm MgCl<sub>2</sub>
- 48  $\mu$ l adapter 1 100  $\mu$ M
- 48  $\mu$ l barcoded adapter 2 100  $\mu$ M

Annealing conditions:

95°C for 4 minutes

1% descent to 20°C

Hold at 20°C

#### 3.11.5 Adapter ligation

The A-tailed DNA was concentrated by a rotary evaporator so the volume of the DNA was around 11.7  $\mu$ l. Then 0,8  $\mu$ l adapter mix, 1.5  $\mu$ l NEB T4 DNA ligase buffer and 1.0  $\mu$ l NEB T4 DNA ligase were added. The mixture was incubated over night at 16°C. After this incubation, the ligation reaction was spiked with 1  $\mu$ l T4 DNA ligase, 1  $\mu$ l T4 DNA ligase buffer and 8  $\mu$ l water. Subsequently, the ligation mixture was at 16°C for 2 hours. Finally, the DNA was purified by using QIAquick PCR Purification Kit. During this purification, the washing buffer PE was sitting on column for 10 minutes. The DNA was eluted in 50 $\mu$ l 37°C water, and the tube with the QIAquick column was incubated at 37°C for 10 minutes before spinning down.

Small scale PCR (10 $\mu$ l) where the transposon junctions were amplified, was done for the fragments to check that every sample got smears in the right length, and not dimers. Some of the samples were sent for Sanger sequencing, to check the adapters, transposon and insertion in genes.

All the plate pools were pooled together, 42 ng, from each pool. 42 ng was chosen because of 24 samples times 42 ng is around 1000 ng. Then it would be easy to take out volumes of 100 ng, for the next PCR to be done. The same was done for all the column pools and the row pools.

PCR to add the Illumina attachment and read homology was done. The adapter primer annealed directly to the adapter and the transposon primer annealed to the transposon. PCR was done 10 to 11 times for each of the three pools now remaining to ensure that only some fragments were

## MATERIALS AND METHODS

amplified. The PCR products were run at a gel and the fragments from 200 to 400 base pair were cut out of the gel and the DNA purified.

The DNA from the plate pools was then pooled together, and the same was done for the PCR product from the column pool and the row pool. Then it was checked if the pools had a smear with the base pair length wanted. Around 25 ng in 30  $\mu$ l was then sent in from each pool. The three pools were pooled into one tube when arriving Texas, where it is analysed.

### 3.12 Homologous recombination of the genes *msmeg\_4898* and *msmeg\_4900*

Gentamicin-resistance marker was previously cloned into a plasmid vector in between 200 base pair regions flanking the genes *msmeg\_4898* and *msmeg\_4900* by Marte Singsås Dragset. The resistant marker and the flanking regions were amplified with PCR. The template in the PCR tube was digested for 1 hour with DpnI. DpnI is a restriction enzyme cleaving DNA only when its recognition site is methylated [61]. Only the template DNA, and not the newly synthesised DNA during PCR will be methylated. Therefore, only the template DNA will be digested. The PCR product was then purified using QIAquick PCR Purification Kit (See protocol in section 3.3.3).

It was then checked if the PCR product had a concentration of at least 200-250 ng/ $\mu$ l. Four cultures of *M. smegmatis* carrying pJV53 was started in 5 ml 7H9 + 20  $\mu$ g/ml kanamycin. These cultures were wild type, mycobactin knock-out, exochelin knock-out and a double knock-out of mycobactin and exochelin. These were incubated for 3 days until the cultures reached stationary phase. 50  $\mu$ l, 100  $\mu$ l and 150  $\mu$ l from each of the four cultures were inoculated in 25 ml 7H9 + 20  $\mu$ g/ml kanamycin. 17 hours later the OD<sub>600</sub> of these new cultures were measured. And the cultures, one from each mutant type, that were most in the middle on an OD between 0.3 and 0.8, were chosen. These were then induced with acetamide to a final concentration of 0.2 % and incubated for three more hours. The cells were then washed 3 times in 25 ml cold 10 % glycerol to create electrocompetent cells. This was done by first spinning down the cultures at 4700 g in 7 minutes at 4°C. Each culture was then resuspended in 1 ml 10 % glycerol, before 24 ml more were added and everything mixed by inverting the tube. Spin and washing were repeated 3 times. It was made sure that everything was kept cold while preparing the competent cells. After the last wash, each pellet was resuspended in 2 ml 10 % glycerol.

Subsequently, transformation with the homologous recombined plasmid was done, as seen in section 3.10.2. From each mutant category 16 colonies were then plated over on a 7H10-plate

with 7.5 µg/ml gentamicin 150 µM FeCl<sub>3</sub>. 4 days later four of the re-plated colonies in each mutant category that had grown up perfectly, were picked. Each were inoculated in 5 ml 7H9 with 7,5 µg/ml gentamicin and 150 µM FeCl<sub>3</sub>. 4 days later all of them were checked with primer pair number 1, and 1 culture from each mutant category, that the primer pair indicated that the gentamicin resistance gene was inserted, was checked further with the primer pairs 2,3 and 4. For PCR-protocol see section 3.6.5.

### 3.12.1 Growth curves preparations

The cultures chosen by the primer pairs, were cultured again in 5 ml 7H9 with 7.5 µg/ml gentamicin and 150 µM FeCl<sub>3</sub> and incubated in 37°C with shaking. After 3 days, they were centrifuged at 4754 g in 7 minutes. Each pellet was then washed in 5 ml Sauton's chelated medium / ironfree. Then the cultures were span down again and the pellet resuspended in 5 ml Sauton's ironfree. OD<sub>600</sub> were measured for all the cultures. This was done by filling the cuvettes with 100 µl of culture and 900 µl of Sauton's ironfree, and then multiplying the OD by 10 to get the actual OD of the culture.

The three different growth solutions for the growth curve analysis were made:

High iron – Sauton's ironfree with 150 µM FeCl<sub>3</sub>.

Low iron – Sauton's ironfree with 100 µM 2'2 dipyridyl

Low iron 3 µM - Sauton's ironfree with 3 µM FeCl<sub>3</sub>.

Master mixes for each condition and mutant were made, for making three replica for each condition. It was mixed bacteria with the solution so the starting OD<sub>600</sub>, would be around 0.01. 200 µl were inserted in each well in the growth curve plate. The growth curve machine, Bioscreen C MBR, was reading the OD<sub>600</sub> evert second hour in three days.

### 3.13 Freeze down of culture

Freeze culture from the candidates chosen and used in the recombination project were made. By doing this these cultures would be stored and ready for more investigation at a later time point. This was done by freezing 1 ml of each culture grown for 3 days with 15 % glycerol in - 80°C.

## RESULTS

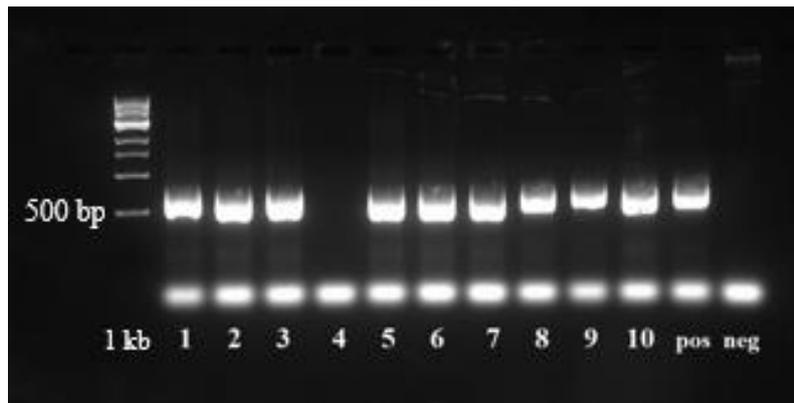
### 4. Results

#### 4.1 Quality assessment of the *M. avium* transposon mutant library

To knock out specific genes in *M. avium* is nearly impossible due to the low transformation efficiency in this bacterium. An organised *M. avium* transposon mutant library, where the insertion sites have been identified, will therefore be an invaluable tool to study mycobacterial gene function. An *M. avium* transposon mutant library containing 170 000 mutants was previously prepared by Marte Singsås Dragset. The library was constructed by transducing *M. avium* with the phagemid  $\phi$ MycoMarT7 (Figure 5), which inserts transposons randomly at TA dinucleotide sites [62]. In order to create a high-density mutant library, an *M. avium* strain which was especially susceptible to transposon mutagenesis by  $\phi$ MycoMarT7 was selected among 15 *M. avium* clinical isolates. Illumina transposon sequencing of the total library pool revealed that over 35 000 of the mutants in this library are unique, that is 35 000 different TA sites had a transposon insertion. The *M. avium* genome contains slightly more than 60 000 TA-sites in the *M. avium* genome, meaning close to 60% of all TA sites were hit by a transposon (unpublished data). The library thus should consist largely of different transposon insertion mutants, and cover most of the non-essential genes in *M. avium*, even when taking into account that some genes contain more than one TA-site.

##### 4.1.1 Verification of transposon insertions

To validate the quality of the replated library, the presence of the transposon was first verified for 10 random colonies by using PCR to check that the bacteria contained the gene for kanamycin resistance (Figure 11). The transposon contains this gene, but the wild type *M. avium* genome does not, so the presence of the kanamycin resistance gene will mean that the transposon is inserted in the bacterial genome. PCR primers for the kanamycin resistance gene were designed by using the software CloneManager. Before the PCR, colonies were picked and incubated for 3 weeks, bead beaten as described in section 3.3.1 and the DNA purified using MasterPure DNA Purification Kit from Epicentre. 10 colonies picked from an agar plate were checked with the primer pair Kan\_F and Kan\_R (Table 3 in section 3.1.3). Sample 4 was analysed twice because of no result first time analysed. When reanalysed sample 4 also got a band (result not shown). Together, this shows that all the 10 randomly picked colonies have the kanamycin resistance gene.



**Figure 11: Checking gene for kanamycin resistance.**

Gel electrophoresis of PCR product of the kanamycin resistance gene from 10 random colonies picked from agar. Ladder: 1 kb DNA ladder from NEB. Positive control:  $\phi$ MycoMarT7 as template. Negative control: water as template.

#### 4.1.2 Small scale identification of transposon insertion sites

Next, to investigate if the transposon was inserted at different sites in the genome for the 9 of the 10 random mutant colonies, the genome sequence flanking the transposon was identified (Table 24).

This was done by in-house Sanger-sequencing, with a primer that binds close to end of the transposon's kanamycin gene. In this way, from the sequence generated, the end of the transposon could also be recognized before the avium genomic sequence starts, to ensure by sequencing as well that the transposon was present in the selected mutants.

**Table 24: Checking that the transposons are inserted in different genes.**

Since the *M. avium* strain used in the library is not published yet, the strain *M. avium* 104 was used as a reference genome when using BLAST. TGTTA are the 5 last base pairs of the transposon, and is here used as an indicator for insertion of the transposon. The TA-site is where the transposon insert in the genome.

Sample nr.	2	3	4	5	6	7	8	9	10
Sequence length	519	654	636	305	463	400	226	277	275
Is there a transposon?	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ends on TGTTA?	yes	yes	yes	yes	yes	yes	yes	yes	yes
Which gene in <i>M. avium</i> 104	Chaperone ClpB	Sugar transport protein	-	2594 bp at 5' side: twin arginine-targeting protein translocase TatC  104 bp at 3' side: conserved hypothetical protein	Hypothetical protein  Regulator of Sig8	Methyltransferase, putative, family protein  Conserved hypothetical protein	Major membrane protein 1  Cysteine desulphurases, SufS	Conserved hypothetical protein	Chaperone ClpB
TA-site	yes	yes	-	yes	yes	yes	yes	yes	yes
Start-end in <i>avium</i> genom	4915391 4915724	3088104 3088544	-	2435036 2434788	4057050 4056932	2795245 2795519	2052101 2052273	5376144 5376332	4915391 4915584

## RESULTS

The transposon sequence was found in each of the samples. Sample 2 and 10 have the transposon inserted in the same gene. The sequence of sample 4 was also good, but it did not match anything in NCBI BLAST.

### 4. 2 Organising the *M. avium* library: Growing and pooling the mutants

To be able to identify which well the mutant was originally grown in after all the mutants are analysed together in the Illumina sequencer, the mutants are pooled into plate, column and row pools. To separate different pools of samples, the different pools were coded.

#### 4.2.1 Preparation of the stock transposon insertion library for organisation

In order to organise the pooled *M. avium* library, and identify transposon insertion sites, the library had to be replated. The *M. avium* genome consists of around 5100 genes. To maximize the chance of obtaining at least one mutant for each non-essential gene, a library size of around twice the size of the number of genes was decided upon for the organised library. Via plating of dilution series it was determined that the stock library suspension contained around  $3.9 \times 10^{11}$  bacterial cells per ml. In order to use the Genetix QPixII colony picker, 2000 colonies per 23 cm x 23 cm agar plate is advisable. The stock suspension was therefore diluted to 2000 bacteria/ml. Plating of 1 ml of the diluted suspension resulted in around 2000 colonies per plate, as seen in Figure 12. Initial testing had shown that at least 50 % of the colonies plated were accepted as pickable by the colony picker robot. 1 ml diluted suspension was therefore plated on each of a total of twelve 7H10 agar plates with 20 µg/ml kanamycin. This would ensure sufficient colonies to pick into 24 well plates consisting of 384 wells each, even if only half of the colonies were be accepted as pickable.



**Figure 12: Agar plate showing a colony density suitable for robotic colony picking.**

For using the colony picker robot Genetix QPixII, it is advisable to have 2000 colonies per 23cm x 23cm agar plate. The goal was reached by diluting the library stock to 2000 bacteria/ml. By plating out 1 ml of this, around 2000 colonies grew up per plate.

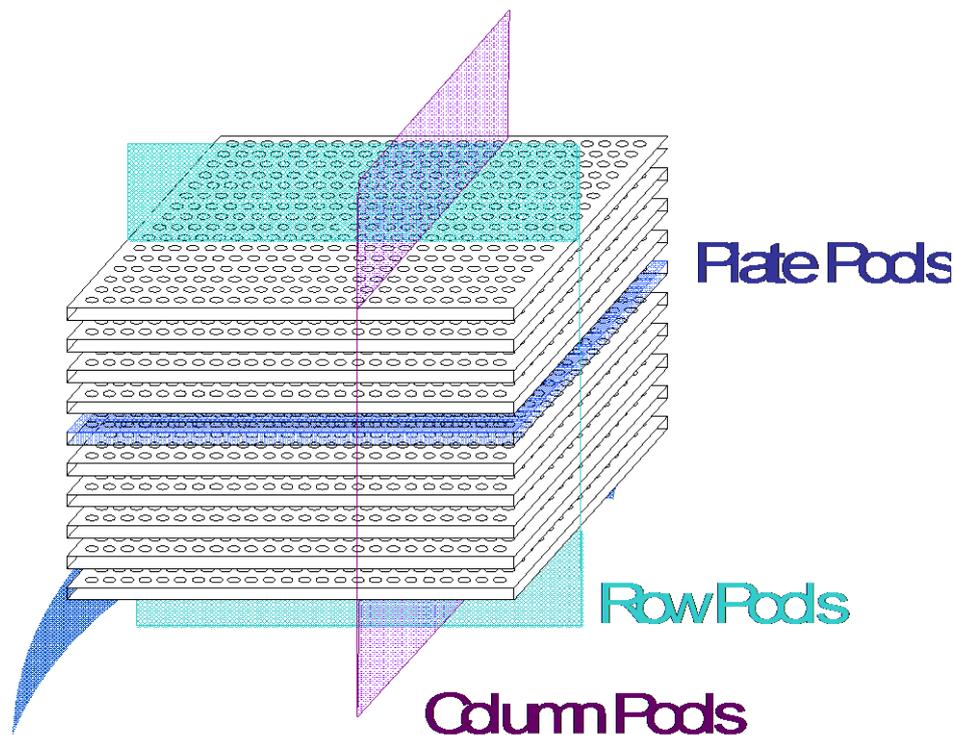
#### 4.2.2 Colony picking robot

Before pooling the plates, the replated colonies were picked into 384-well plates. It was decided to use the colony picking robot Genetix QPixII, to pick the colonies into the well plates. First, a round of test-picking was done into one 384-well plate. This went well, the bacteria were growing in the complete 7H9 media with 20  $\mu\text{g/ml}$  kanamycin in this plate after 3 weeks incubation at 37°C. The plate was covered with a sealing tape in addition to the respective lid of the well plate. The wells did not dry out and there was no sign of contamination. Thereby, 9216 colonies were picked from the agar plates in duplicates into 48 plates with 384 wells each. Each duplicates was then consisting of 24 well plate. The library was picked in duplicates, so one library could be pooled and the other copy of the library could be stored and used to pick out interesting mutants later. The library was incubated at 37°C for 3 weeks, and then stored at 4°C until pooling of the plates was done.

## RESULTS

### 4.2.3 Pooling the plates

The pooling strategy was to pool 3  $\mu\text{l}$  every well at the same positions at every plate together in a master plate, and at the same time make two extra master plates, so that it would be enough material for DNA purification and preparation of DNA for Illumina sequencing. The master plates were pooled with help from the pipetting robot Freedom EVO 200 from TECAN. 10  $\mu\text{l}$  of every well in each single plate were then manually pooled to create the plate pools. To create column pools, 30  $\mu\text{l}$  of every well in each single column in the master plates were manually pooled. The same applies for every row in the master plates, to create row pools. This strategy gave 24 plate pools, 24 column pools and 16 row pools. Each well is then represented three times - in a whole plate pool, a pool consisting of each well in each column and each row, as seen in Figure 13.



**Figure 13: Grinding the library in plate, column and row pools.**

3D pool organisation of 384-well plates. Each well is represented three times by this strategy – in a plate pool, in a row pool and in a column pool [63].

### 4.3 Preparing the *M. avium* library pools for identification of transposon insertion sites by Illumina sequencing

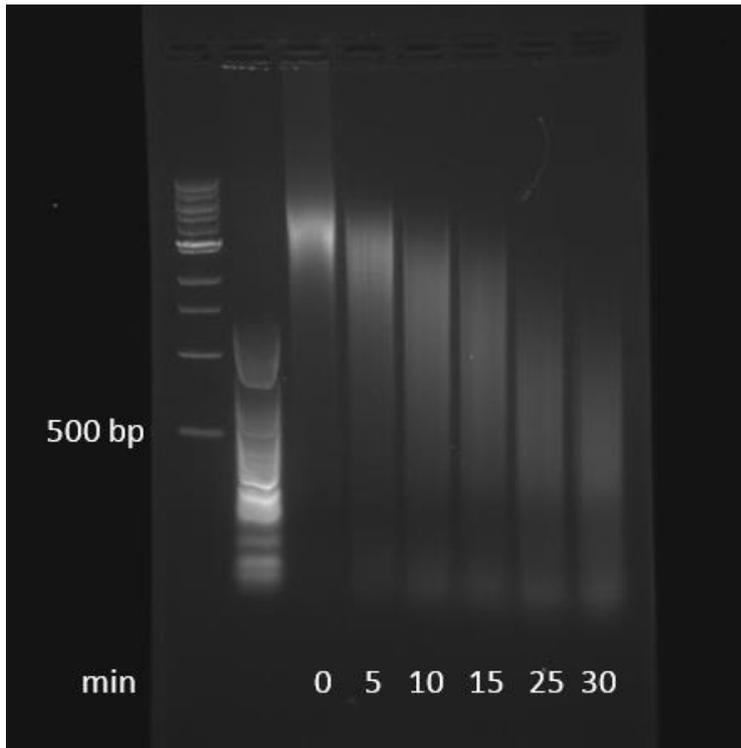
To separate different pools of samples, the primers and adapters used in the pool preparation contained indexes and barcodes, respectively. The adapter barcodes were used to define the number of plate, row and column pool, while the primer indexes referred to if the mutant was in a plate, row or column pool. In this way it was possible to identify which well the mutant was grown in originally, although all the pools are analysed in one single run in the Illumina sequencer.

The genomic DNA from each of the 64 pools were isolated by bead beating and DNA purification with MasterPure DNA Purification Kit from Epicentre. The workflow for the preparation of the mutants consists of shearing, end-repair, A-tailing, adapter ligation and PCR amplification. This is described in section 1.4.2 and 3.11.

#### 4.3.1 Shearing genomic DNA

Most of the samples were sheared enzymatically by fragmentase, which randomly creates double strand breaks in DNA [64]. The activity time of fragmentase was first optimised. It was desired to have fragments that were between 200 and 400 base pair. Therefore, 5 µg genomic DNA were incubated in the presence of fragmentase in 0, 5, 10, 15, 25 and 30 minutes before inactivation of fragmentase by EDTA. Subsequently, the 5 µl of the sample were mixed with 1 µl of Gel Loading Dye, Blue (6X) and the fragments were visualised by gel electrophoresis. As seen in Figure 14, an activity time of 25 minutes seemed to be the best.

## RESULTS



**Figure 14: Optimisation of the activity time for fragmentase.**

The activity time of fragmentase was optimised before it was used to shear the genomic DNA of transposon library mutants into smaller fragments. It was desired to have fragments that were between 200 and 400 base pair. Different time points were tested. From the gel electrophoresis done, an activity time of 25 minutes seemed to be the best.

Due to unexpected lack of fragmentase, nine samples, four row pools and five column pools, were sheared by mechanical shearing by the instrument Covaris (Appendix 1 and 2). All 64 samples were checked if they had the desired base pair length by gel electrophoresis after shearing (results not shown).

### 4.3.2 End repair, A-tailing and adapter ligation

In order to ligate adapters to the ends of the sheared DNA, the ends of the DNA fragments were first enzymatically blunt-ended using the end-repair kit from NEB (section 3.11.2). To check if the fragments were blunt-ended, CloneJET PCR Cloning Kit from Thermo Scientific was used. This kit contains a blunt cloning vector, which is able to insert blunt ended DNA fragments. The vector contains a gene for ampicillin resistance and a gene that is lethal when it is not disrupted by ligation of a DNA insert into the cloning site [65]. By ligating end-repaired DNA fragments into this vector, transform it into bacteria and plate the transformation mix at LB-plates with 100 µg/ml ampicillin, it is possible to see if the DNA fragments were able to be inserted in the vector. At one plate it was plated bacteria that was transformed with vector with

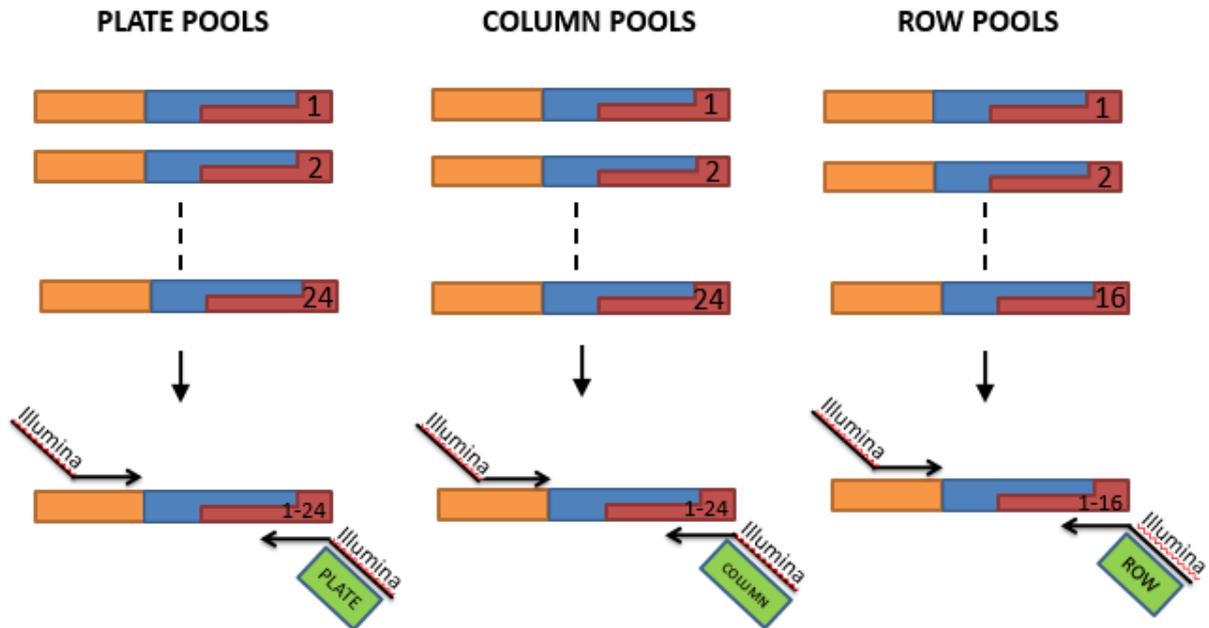
the DNA fragments and at the other plate it was plated bacteria that was transformed with vector without the DNA fragments as a control. The plate with the bacteria transformed with DNA fragments inserted in vectors should have significantly more colonies grown up at the plates than the control plate. The test-plate had 100 colonies grown up and the control-plate had no colonies, which means that after the end-repair step at least a fraction of the fragments should be end-repaired and ready for the next step: A-tailing, which is to add an extra A to the ends of the fragment (further explained in section 1.4.2).

Successful A-tailing was checked by using InsTAclone PCR Cloning Kit from Thermo Scientific. This kit contains a TA-cloning vector, which is able to insert dA-tailed DNA fragments. The vector contains a gene for ampicillin resistance and has a 3' ddT-overhang at each end. The transformed bacteria were plated at LB-plates with 100 µg/ml ampicillin. A control was also used here, which was made the same way as for the blunt end cloning. The test plate had over 1000 colonies grown up and the control plate had around 100 colonies, which means that after the A-tailing step the fragments got an extra adenine added to their ends.

#### 4.3.3 Adapter ligation and coding the DNA pools

To separate the mutants, they were coded with different adapters. Each adapter has a unique barcode. There will be one specific barcode for each number of the plates, columns and rows. In other words, the plate, column and row number 1, will all have the barcode number 1, then the plate, column and row number will all have barcode 2, and so on. Later on, with the PCR that will amplify the transposon junctions (section 4.3.4), the primers binding to the adapters will have indexes that will separate the mutants belonging to a plate, row or column (Sol\_AP1\_57, Sol\_AP1\_473, Sol\_AP1\_100, respectively). This dual coding strategy is shown in Figure 15.

## RESULTS



**Figure 15: Strategy for coding the mutants.**

To separate the mutants, they were coded. First with 24 different adapters. Each adapter has a unique barcode. There will be one specific barcode for each number of the plates, columns and rows. Then the primers used will have barcodes that will tell if the mutant is in a plate, row or column. Colours in figure: Transposon is orange, genomic DNA is blue and adaptor is red.

To check that the mutants were successfully adapter ligated, all the 64 adapter ligated samples were run in a small scale PCR by using the transposon binding primer Sol\_Mar\_NI and the adapter binding primer Sol\_AP1\_NI (Figure 16-18). This was done to check that they all had a smear, and not only formed dimers. Smears were expected because the since the fragments had been randomly sheared, and therefore had variable base pair lengths. Dimers would indicate that the transposon primer and the adapter primer had bond to themselves, and would be visualised as a band with length of around 100 base pair.

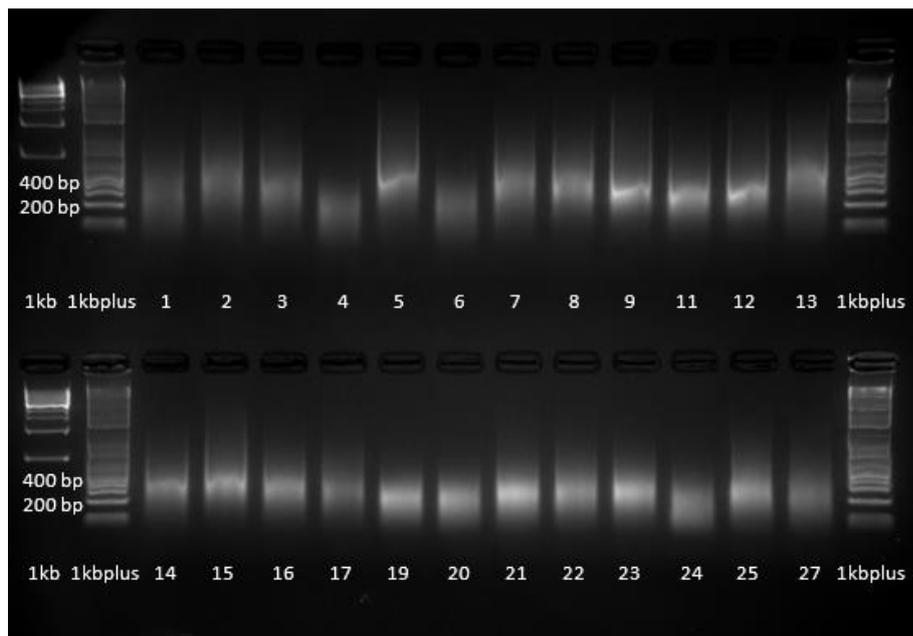


Figure 16: Gel electrophoresis of transposon junctions of all the plate pools amplified by small scale PCR.

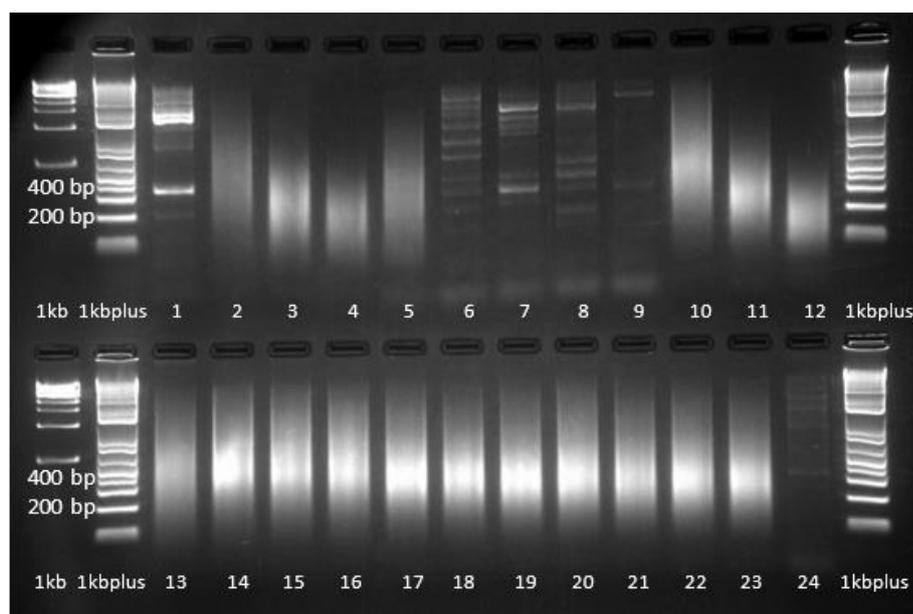


Figure 17: Gel electrophoresis of transposon junctions of all the column pools amplified by small scale PCR.

Six column pools did not show smears as expected after the shearing by fragmentase (Figure 17). These pools were therefore prepared over again, and this time sheared by Covaris. The pools then showed smears in the gel electrophoresis, which were in the desired base length region.

## RESULTS

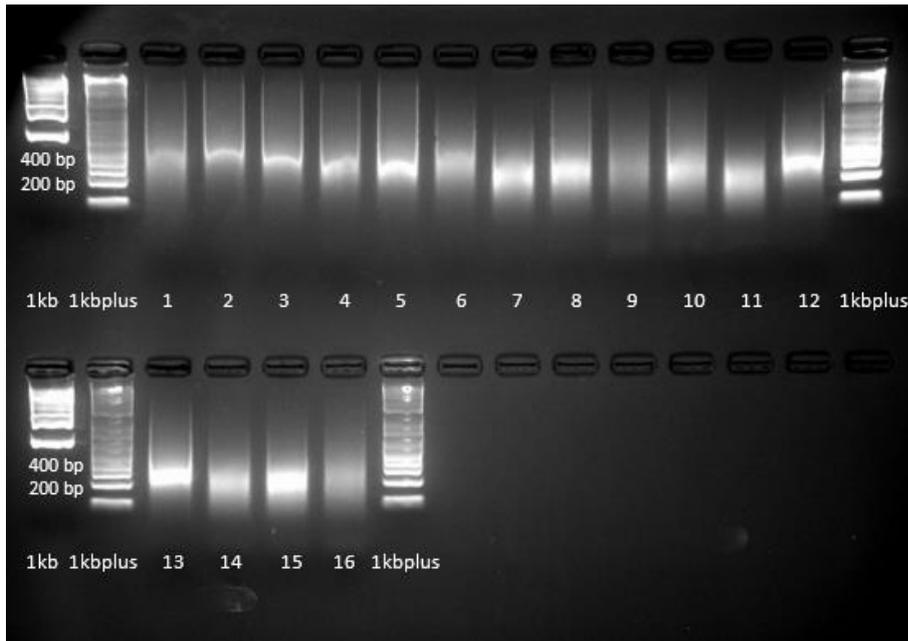


Figure 18: Gel electrophoresis of transposon junctions of all the row pools amplified by small scale PCR.

Some of the samples were Sanger sequenced to confirm that the adapters were inserted in the transposon junction and had the correct barcode (Table 25). Before the samples were ready for Sanger sequencing, the samples had their transposon junctions amplified by PCR by using the transposon binding primer Sol\_Mar\_NI and the adapter binding primer Sol\_API1\_NI. Then the transposon junctions were blunt cloned by CloneJET PCR Cloning Kit from Thermo Scientific. The transposon junction fragments were cloned into the vector pJET1.2 by following the protocol in section 3.1.7.2.

The transposon junctions were then amplified by PCR again by using the primers Sol\_Mar\_NI and Sol\_API1\_NI. The samples that had PCR-product would also be clones that had the transposon fragment inserted. This to ensure that the samples sent for Sanger sequencing most likely had the amplified PCR product inserted. The samples with PCR-product were then cut out of the gel and purified. The cloning vector's respective sequencing primer, pJet1.2 forward sequencing primer, were mixed with the sample DNA before sequencing (section 3.8.1). The sequencing showed that the adapter barcodes corresponded to the sample sequenced, e.g. the samples sequenced from plate number 3 had the adapter barcode number 3. The results are shown in Table 25.

**Table 25: Results from Sanger sequencing.**

Since the *M. avium* strain used in the library is not published yet, the strain *M. avium* 104 was used as a reference genome when using BLAST. TGTTA are the 5 last base pairs of the transposon, and is here used as an indicator for insertion of the transposon.

<b>Pool</b>	<b>TGTTA</b>	<b>Adapter BARCODE</b>	<b>Blast GENE</b>
<b>Plate 9</b>	Yes	9	-
<b>Plate 9</b>	Yes	9	-
<b>Column 3</b>	Yes	3	-
<b>Column 15</b>	Yes	15	Pyrioxamine 5'-phosphate oxidase family protein
<b>Row 9</b>	Yes	9	-
<b>Row 9</b>	Yes	9	Conserved hypothetical protein
<b>Row 15</b>	Yes	15	Anti-anti-sigma factor
<b>Row 15</b>	Yes	15	-

#### 4.3.4 PCR of adapter ligated plate, column and row DNA pools

PCR was done to add the Illumina attachment and read homology. Before the PCR, all the plate pools were pooled together, making a master plate pool. Master pools were also made for all the column pools and the row pools. PCR with the transposon primer mix and the pools respective adapter primers was done 10 to 11 times for each of the three pools now remaining. For protocol see section 3.6.4. The PCR was done several times to ensure that only some fragments were amplified. The PCR products were run at a gel and the fragments from 200 to 400 base pair were cut out of the gel and the DNA purified. See figure 19 and 20, and Appendix 3 and 4.

RESULTS

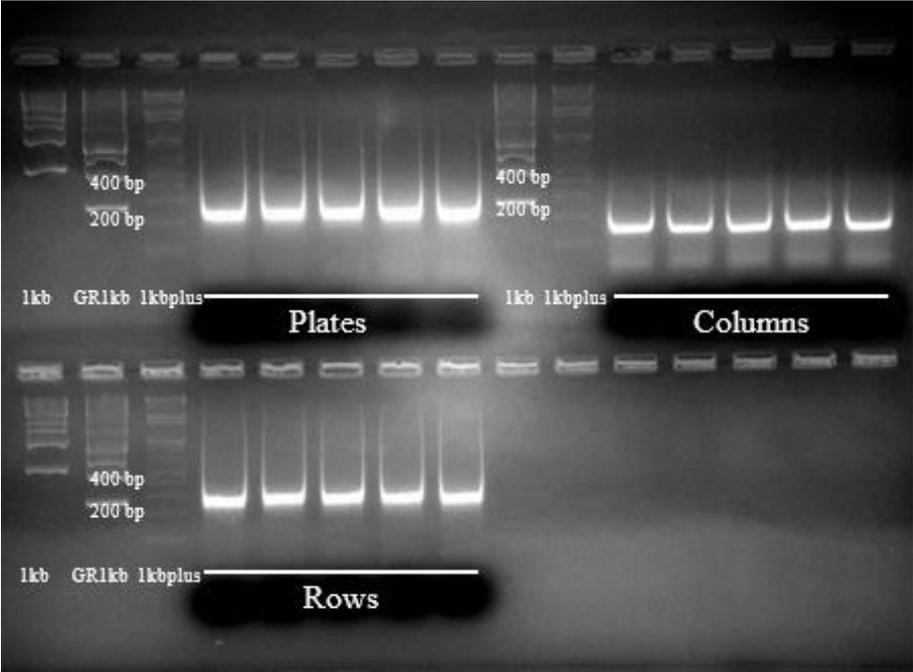


Figure 19: Gel electrophoresis of PCR-products where all the pools in each category plate, column and row are pooled.

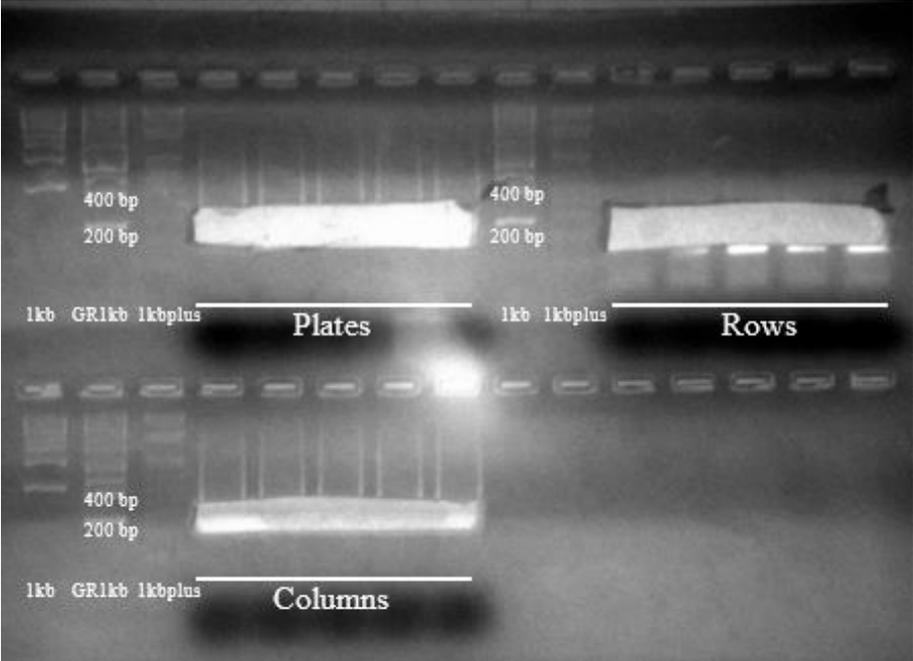
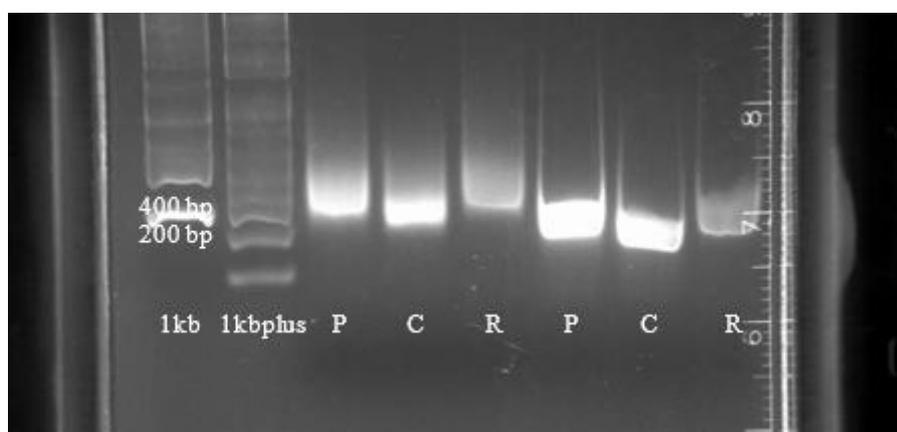


Figure 20: Gel from figure 19, after cutting out bands with approximately length of 200 to 400 base pairs.

## RESULTS

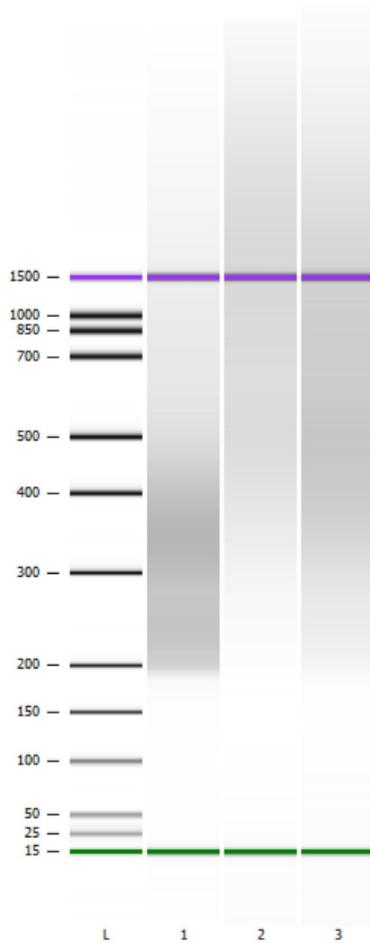
The PCR-products of the master pools were purified and the products that were from the same kind of pool were pooled. The fragments in the three remaining pools were then visualised by gel electrophoresis. Quite sharp bands were seen (Figure 21). The formation of sharp bands had also been seen for some of the samples in the small scale PCR tests in Figure 16-18. Sharp bands on the PCR could indicate contamination, that one specific fragment was amplified instead of a smear of different sized fragments. Therefore, the samples were analysed by a BioAnalyzer also to ensure that samples were not contaminated. The results from the BioAnalyzer showed smears and also acceptable base pair lengths for Illumina sequencing (Figure 22 and 23). Around 25 ng in 30  $\mu$ l was then sent to Texas A&M University for Illumina sequencing from each of the three final pools.



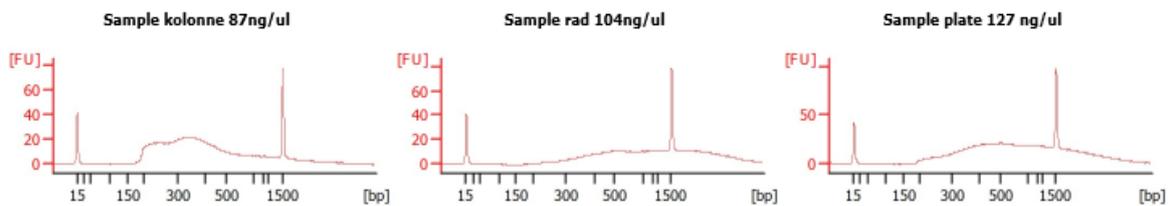
**Figure 21: Gel electrophoresis of the final pools of plates (P), columns (C) and rows (R).**

In the three first sample lines 1  $\mu$ l DNA of the final pools of plates (127 ng/ $\mu$ l), columns (87 ng/ $\mu$ l), and rows (104 ng/ $\mu$ l) was loaded. In the three last sample lines 3  $\mu$ l of the same samples were loaded on the gel.

## RESULTS



**Figure 22: Results from BioAnalyzer, showing smears from shearing the final plate, column and row pool with Covaris.** L-lane is showing the ladder with base pair lengths annotated. Lane 1 is the column pool, lane 2 is the row pool and lane 3 is the plate pool.



**Figure 23: Electropherogram from BioAnalyzer, showing the distribution of DNA in the final column, row and plate pool.**

The first electropherogram is the DNA distribution in the column pool, the second is the row pool and the last one show the DNA distribution of the plate pool. The DNA distribution is shown in relation to the base pair (bp) length of the DNA in the sample, which is represented with the x-axis. The y-axis of the electropherogram represents fluorescence units (FU). The two high peaks in each electropherogram are markers, and not part of the DNA sample.

#### 4.4 Creating knock-out of the genes *msmeg\_4898* and *msmeg\_4900* by homologous recombination

Iron uptake is an important field in mycobacteria research because one of the ways the *M. tuberculosis* survives intercellularly is by having mechanisms to take up iron when there are extremely low iron conditions [18]. *M. smegmatis*, as opposed to *M. avium*, is susceptible to homologous recombination: This method was therefore used to knock out the genes *msmeg\_4898* and *msmeg\_4900* in *M. smegmatis*. These genes were genes potentially related to iron uptake in mycobacteria, because of homology to genes in *M. tuberculosis* laying in a cluster of genes known to be involved in iron uptake [32]. The *M. tuberculosis* homologue to *msmeg\_4898* is in addition predicted to be a binding site for IdeR (iron-dependent regulator, and this gene in *M. tuberculosis* is also shown to be repressed by the presence of iron and IdeR [33-35].

Four bacteria cultures of *M. smegmatis* were started; the wild type (WT), a mycobactin knock-out mutant ( $\Delta mbtD$ ), an exochelin knock-out mutant ( $\Delta fxbA$ ) and a double knock-out mutant, where both mycobactin and exochelin are knocked out ( $\Delta mbtD_{\Delta fxbA}$ ).  $\Delta mbtD$  is a polyketide synthase required for the synthesis of mycobactin and  $\Delta fxbA$  is a formyl transferase involved in the synthesis of exochelin [66, 67]. By knocking down genes involved in synthesis of a siderophore, the mutant is no longer able to produce the siderophore. Knock-out mutants of the genes *msmeg\_4898* and *msmeg\_4900* were made in all of the four different strains. This was done to be able to see how this knock-out mutation would affect the growth of the strains, and be able to identify which siderophore pathway the genes *msmeg\_4898* and *msmeg\_4900* might were a part of. By combining mutations in the two different iron uptake pathways the bacteria should fail to grow under low iron conditions. By combining mutations in the same iron uptake system, the bacteria should still be able to grow under low iron conditions, as it would be rescued by the intact siderophore.

All these bacteria contained the recombination plasmid pJV53. They were induced by actemide and homologously recombined with a construct consisting of a gentamicin resistance gene between 200 base pair flanking the genes *msmeg\_4898* and *msmeg\_4900*. The recombined bacteria were then plated at 7H10 agars with 7.5 $\mu$ g/ml gentamicin and incubated for 3 days.

## RESULTS

### 4.4.1 Selecting knock-out mutants for growth curve analysis

The mutants correctly recombined was found by using four pairs of primer set (Figure 24). The names and where they bind are shown in Figure 25. The expected results of bands in gel electrophoresis after PCR amplification for successful knock-out are seen in Table 26. In Table 27 the primer pair and their respective length of their PCR-product are shown.

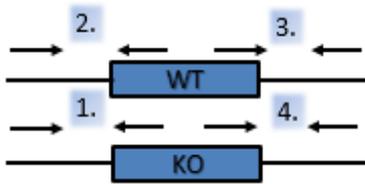


Figure 24: Four set of primer pairs were used to confirm knock out of the genes *msmeg\_4898* and *msmeg\_4900*. WT stand for the wild type, and KO for the knock-out mutant.

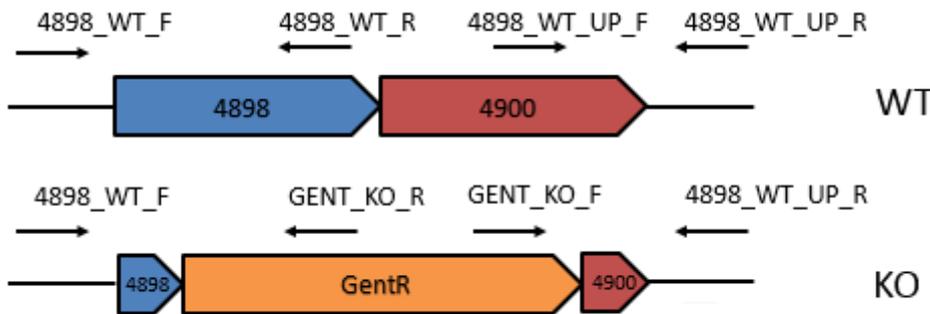


Figure 25: Primer pairs used to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.

A scheme over where the different primer pairs used to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900* and where these primers would bind at the wild type (WT) and the knock-out mutants (KO). This scheme also shows the names of the primers.

Table 26: Expected results of successful knock-out mutants at gel electrophoresis after PCR.

	Primer pair 1	Primer pair 2	Primer pair 3	Primer pair 4
Band at all candidates	Yes	No	No	Yes
Band at control, <i>M. smegmatis</i>	No	Yes	Yes	No

Table 27: The primer pairs used to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900* and the length of their PCR-products.

Primer pair	Base pair length of PCR-product
1. 4898_WT_F + GENT_KO_R	853
2. 4898_WT_F + 4898_WT_R	684
3. 4898_WT_UP_F + 4898_WT_UP_R	824
4. 4898_WT_UP_R + GENT_KO_F	1128

First DNA from four colonies from each category WT,  $\Delta mbtD$ ,  $\Delta fxbA$  and  $\Delta mbtD \Delta fxbA$  were amplified using primer pair 1. A successful mutant would here show a band using this primer pair, indicating that the gentamicin resistance gene is inserted. The WT-category had one with a weak band. All the mycobactin knock-outs checked had bands. Half of the exochelin knock-outs had bands, and all the double knock-outs checked got bands. The control used, genomic DNA from *M. smegmatis* that had not been recombined (100 ng/ $\mu$ l), showed no bands as expected. One candidate with band was chosen from each category for further check with the other primer pairs. The candidates chosen is indicates by an asterisk in Figure 26.

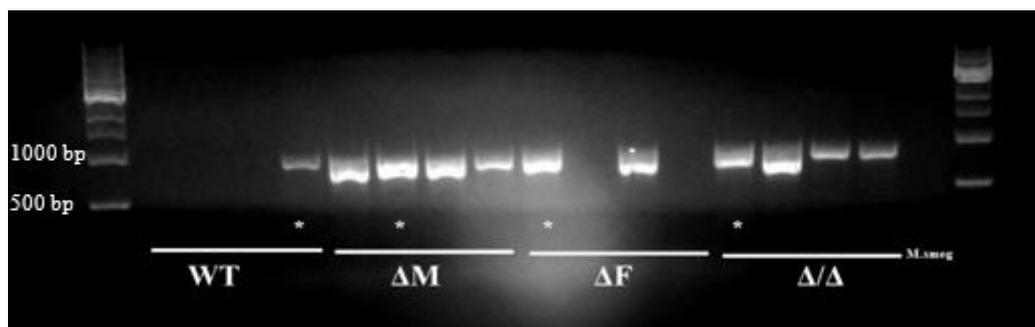


Figure 26: Gel electrophoresis of products from PCR with primer pair 1 to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.

One candidate with band was chosen from each category for further check with the other primer pairs. The candidates chosen is shown with a \*. The control used, *M. smegmatis* that had not been recombined, showed no bands as expected. WT is wild type,  $\Delta M$  is mycobactin knock-out mutant,  $\Delta F$  is exochelin knock-out mutant and  $\Delta/\Delta$  is a mutant with both mycobactin and exochelin knocked out.

The chosen candidates from the run with primer pair 1, were then tested with primer pair 2, primer pair 3 and primer pair 4. As seen in Figure 27, the results for primer pair 2 and 3 were as expected. For primer pair 4 the control was as expected, and the double knock-out also got a clear band. However, only weak bands were shown for  $\Delta mbtD$  and  $\Delta fxbA$ , and there was no band to see for the wild type. PCR with primer pair 4 was thus repeated and correct mutants were identified, although, for the *msmeg\_4898-4900* mutant in the WT background, new colonies had to be analysed before a correct mutant was identified (results shown in Appendix 5 and 6).

## RESULTS

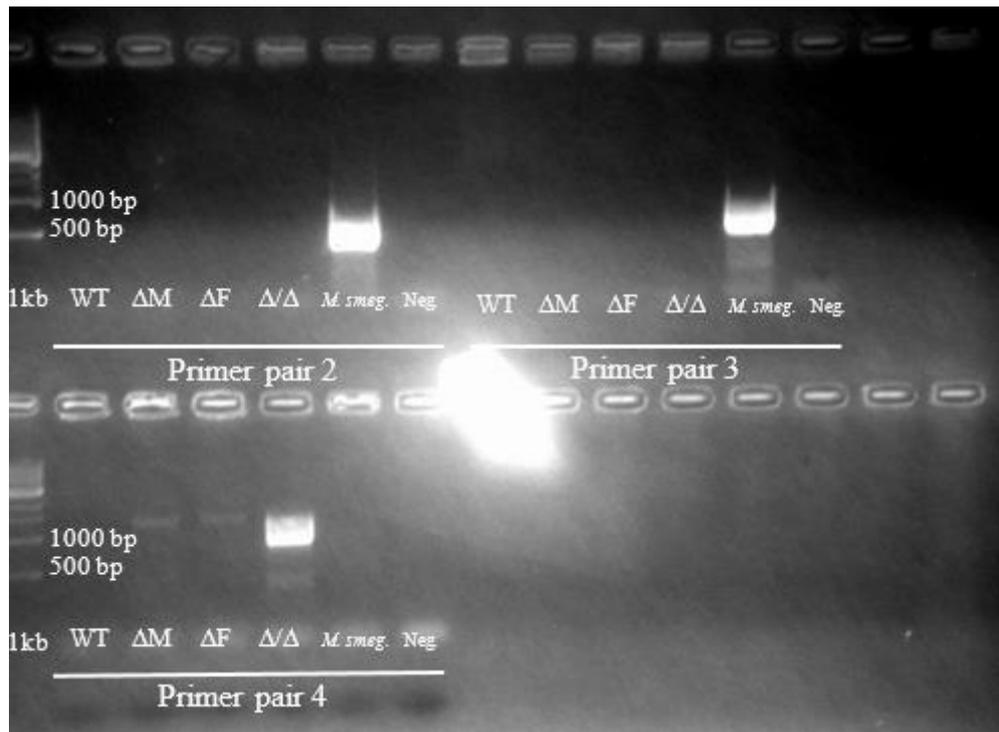


Figure 27: Gel electrophoresis of products from PCR with primer pair 2, 3 and 4 to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.

WT is wild type,  $\Delta M$  is mycobactin knock-out mutant,  $\Delta F$  is exochelin knock-out mutant and  $\Delta/\Delta$  is a mutant with both mycobactin and exochelin knocked out. The control used is *M. smegmatis* that had not been recombined.

### 4.4.2 Growth of *msmeg\_4898* and *msmeg\_4900* knock-outs under low iron conditions

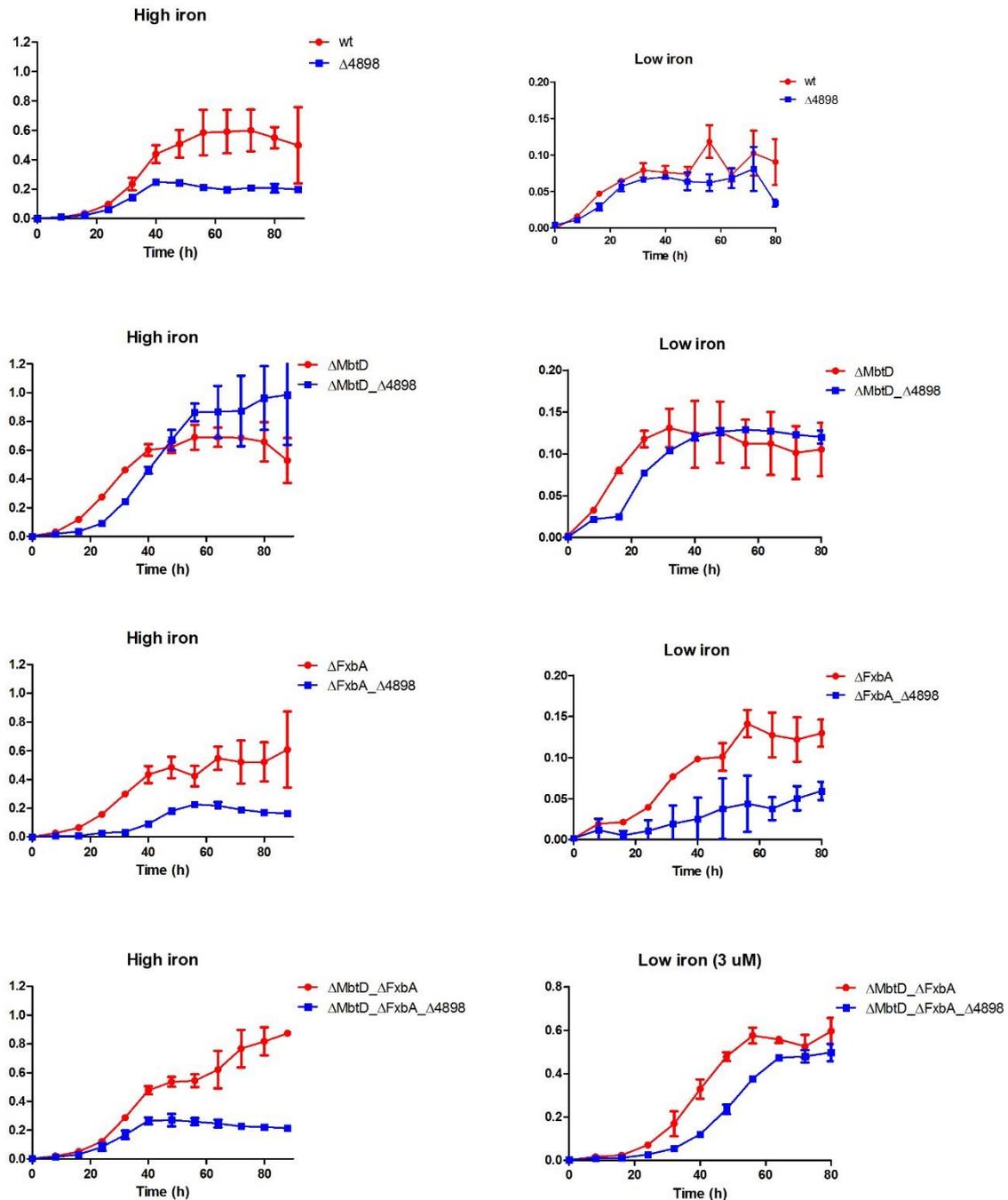
In order to assess the potential function of *msmeg\_4898* and *msmeg\_4900* in mycobactin-mediated iron uptake, the *msmeg\_4898-4900* knockout mutants were prepared for growth curve analysis. There was made high and low iron conditions for all the candidates, and each condition was set up with three technical replicates in the well plate for the growth curve instrument, Bioscreen C MBR. The  $OD_{600}$  was read every second hour in three days by Bioscreen C MBR. This experiment was repeated three times.

If the bacteria do not grow at low iron conditions when exochelin and the genes *msmeg\_4898* and *msmeg\_4900* studied are knocked-out it indicates that the genes studied are important for the mycobactin pathway. With exochelin knocked out, mycobactin is the siderophore to take up iron. If genes that are important for the mycobactin pathway also are knocked-out, the bacteria do not have any functional siderophore iron uptake system, and therefore lose an important way to take up iron from the outside of the cell. Since iron is crucial for the growth of the bacteria, the bacteria will fail to grow under low iron conditions [30].

As seen in Figure 28, it is seen that the bacteria are still growing under low iron conditions, although exochelin and the genes *msmeg\_4898* and *msmeg\_4900* are knocked out in the same

mutant. Similar results are shown in the two other growth analysis done (Appendix 7 and 8). In the growth curves shown in Figure 28 is the *msmeg\_4898* and *msmeg\_4900* knock out called  $\Delta 4898$ . The  $\Delta fxbA_{\Delta 4898}$  is growing less than the  $\Delta fxbA$  at low iron conditions. However the same trend is seen at high iron conditions. The WT and the WT\_ $\Delta 4898$  are growing similarly under low iron conditions, and this is also seen under high iron conditions for WT and the WT\_ $\Delta 4898$ . Similarly are the results for  $\Delta mbtD$  and  $\Delta mbtD_{\Delta 4898}$ . In addition, the  $\Delta mbtD_{\Delta fxbA}$  and the  $\Delta mbtD_{\Delta fxbA_{\Delta 4898}}$  are growing similarly under low iron (3 $\mu$ M iron added) conditions and also under high iron conditions. Thus, it seems like *msmeg\_4898* and *msmeg\_4900* is not required for iron uptake in *M. smegmatis*.

## RESULTS



**Figure 28: The first replicate of growth curve analysis.**

In order to assess the potential function of *msmeg\_4898* and *msmeg\_4900* in mycobactin-mediated iron uptake, the *msmeg\_4898-4900* knockout mutants ( $\Delta 4898$ ) were prepared for growth curve analysis. There was made high and low iron conditions for all the candidates, and each condition was set up with three technical replicates in the well plate for the growth curve instrument, Bioscreen C MBR. The OD<sub>600</sub> was read every second hour in three days by Bioscreen C MBR. This experiment was repeated three times. The four bacteria candidates of *M. smegmatis* analysed, both with and without the genes *msmeg\_4898* and *msmeg\_4900* knocked out were; the wild type (WT), a mycobactin knock-out mutant ( $\Delta mbtD$ ), an exochelin knock-out mutant ( $\Delta fxbA$ ) and a double knock-out mutant, where both mycobactin and exochelin are knocked out ( $\Delta mbtD_{\Delta fxbA}$ ).

## 5. Discussion

In this study, it is shown how it is possible to identify the mutated genes in a large *M. avium* transposon mutant library by using Illumina sequencing. To have a transposon library of *M. avium* is of special worth, because this species is not that susceptible to mutagenesis by homologous recombination due to low transformation efficiency (Marte Dragset, personal communication). Therefore, the library would be a good resource for pick out mutant to study gene function in *M. avium*, and new discoveries might be transferable to *M. tuberculosis* gene function. *M. smegmatis* is susceptible to mutagenesis by homologous recombination [50]. Due to an efficient transformation efficiency *M. smegmatis* is useful as a model to study mycobacterial basic mechanisms, including uptake of iron [30]. Therefore, to study their proposed role in mycobacterial iron uptake, two *M. smegmatis* genes suggested to be involved in uptake of iron were knocked out by recombineering. Their requirement for mycobacterial growth in low iron was subsequently investigated.

### 5.1 Transposon mutagenesis versus homologous recombination

There are positive and negative sides by using transposon mutagenesis. And there are also advantages and disadvantages by using homologous recombination. Transposon mutagenesis is efficient in the way that it is a not that much work needed to make mutations evenly distributed all over a genome. Due to this, transposon mutagenesis is a great technique for creating mutant libraries. By homologous recombination just a few genes, which then are next to each other, can be mutated in one round of mutagenesis done in the laboratory. Furthermore, by homologous recombination the genes mutated can be deleted from the genome. By transposon mutagenesis, the sequence of a gene will still be in the genome, just disrupted by the transposon. Consequently, depending on where in the gene the transposon is inserted, the function of the gene can still be intact. The insertion of the transposon can also lead to an uncontrolled altered function of the gene disrupted. However, most mutations by transposons lead to a total inactivation of the gene [68].

Which gene that is mutated, is easy to control by homologous recombination. However, it is also possible to do targeted mutagenesis by transposons that bind to defined DNA-binding domains [69]. Both transposons and homologous recombined genes can be designed to carry selectable markers, such as antibiotic resistance. Due to this it is easy to find mutated bacteria,

## DISCUSSION

by using phenotypic screening for example for antibiotic resistance. However, some times the recombination can occur at a not desired place in the genome.

Sometimes also just a part of the gene that is desired to knock-out is recombined. Therefore, it is important to control for appropriated recombination by PCR, as done in this thesis, to ensure that the whole targeted gene sequence actually is deleted in the recombined version of the genome. Transposon mutagenesis can to some extent be controlled. Transposons can be inserted in a desired DNA-binding domain, but can also be inserted randomly in a genome independent of specific nucleotides [40, 69, 70]. The transposon used in this thesis only inserted at TA-sites. Consequently, genes which do not have any TA-site would not be mutated. Due to the even distribution of mutations, essential genes can be easily discovered by transposon mutagenesis. When the transposon is inserted in an essential gene, the bacteria with this insertion would not grow.

### 5.2 Identification of transposon insertion site by Sanger sequencing

During the preparation of the *M. avium* transposon library, Sanger sequencing was used on several occasions to check that transposon mutants. Although transposons were detected in the random mutants investigated, both by PCR and by Sanger sequencing, not all of them matched the *M. avium* 104 reference genome, or any other sequence on NCBI BLAST. One possibility here is that this is a gene that is typical for the strain used in the library, which is sequenced, but not published yet. The transposon can also have been inserted in a mycobacterial plasmid without known sequence.

As shown in Table 24 in section 4.1.2, results from this Sanger sequencing were showing that the transposon was inserted in the same gene in two of the nine mutants investigated. This could mean that some mutants are highly overrepresented in the library, which would result in a library that would not give a desired coverage of the genome.

It statistical unlikely that for two of nine samples investigated the transposon should be inserted in the same gene, since the library consists of 35 000 unique mutants, where all had transposon inserted at different TA-sites. The bacterium mutated has around 5000 genes. This seems to be a coincidence that two of the mutants had the transposon inserted in the same gene, while the gene they were inserted in has not shown up in the other Sanger sequencings done. It could be that by coincidence two mutants for the same gene were picked. Another explanation would be that two samples were mixed or one mutant was picked twice from the agar plate.

The other results from the sequencing to control that the transposon was inserted in different genes in the transposon library, were as it was hoped for. Most of the genes were inserted in different genes and the transposon was identified in all mutants.

In the other Sanger sequencings done, it is sometimes seen that the genome sequence detected is very short. This was the example for the majority of the samples listed in Table 25. In this case, PCR products were ligated into a blunt vector, transformed and prepared for sequencing. The reason for the short sequence could be due to that the vector prefer to insert the short fragments in comparison to the longer fragments. Anyway, the results from the Sanger sequencing, including additional results not shown in this thesis, showed that over 50% of the fragments analysed were as expected after preparation for Illumina based transposon sequencing. The remaining 50 % was fragments that didn't have the transposon, adapter or genomic DNA inserted. It was concluded that this would be sufficient for a successful Illumina sequencing of the library.

### 5.3 Preparation of the transposon library mutant pools for Illumina sequencing

During the preparation of the transposon library mutant pools for Illumina sequencing, there were some problems getting enough DNA by using the MasterPure kit. This might have been better by using another method to isolate the DNA, e.g. by phenol-chloroform extraction. However, using phenol-chloroform extraction would be more time consuming and laborious, and would include the use of an organic solvent and a dangerous corrosive chemical. In the end of the preparation period, some of the original bacteria sample pools started to run empty, due to that the fragmentase step had to be repeated several times for some samples to get the desired fragment length. Therefore a small portion of the bacteria culture from each of these pools was diluted 1:10 in 5 ml 7H9 Complete and incubated in 5 days before the purification of DNA to increase the number of bacteria before isolating. To increase the amount of bacterial in each pool by had a good effect on the DNA amount got after MasterPure purification. Therefore, in a retrospective view, it might have been better to do this for all the samples from the beginning. The reason why not all the sample pools were incubated some days longer after they were pooled and before the purification of DNA from the beginning was because of fear of that a few of the mutants would grow faster than others dominate the pool completely. However, it should not be problematic, because this method do not count on amount of each mutant, but its presence or absence.

## DISCUSSION

When fragmenting the DNA, the activity time of fragmentase, which enzymatically sheared the purified DNA, was optimised to reach the desired fragment lengths. However, even with optimisation, the DNA purified from the various 64 pools often showed too small or too large DNA fragments after shearing compared to the desired, optimised length. This could be due to different purity of DNA in the samples, or that the DNA concentrations measured by NanoDrop was not precise. By mechanical shearing though, a desired length of the fragments was much easier to control. Mechanical shearing was not used from the start due to that this method is more expensive than shearing with fragmentase. However, thinking of all the work with redoing the fragmentase shearing due to incorrect fragment size in many of the DNA pools, it seems to be a better approach to use mechanical shearing to shear the DNA.

During the testing of adapter ligation, it was by several occasions seen sharp bands after amplification of the transposon junction by PCR, when visualised by gel electrophoresis. It was expected to see smears here, which would indicate the presence of several different fragments in the pool. Sharp bands could represent primer-dimers, amplification of just a few specific fragments during PCR amplification or contaminations. Therefore, the final three samples were also analysed on BioAnalyzer before the samples were sent to Texas A&M University for sequencing. The results from the BioAnalyzer showed smears, but with a different fragment size distribution in the three different final pools. From the gel electrophoresis results obtained in the gel, which the DNA from the master pools were purified from, it looked as the samples had around the same fragment size distributions. The fragments in the three final pools, which were also analysed on BioAnalyzer, were also visualised by conventional gel electrophoresis. This time it was loaded less DNA than done in the gel electrophoresis done for cutting out the desired fragment length from the gel, and quite sharp bands were seen, and the bands were sharpest at the highest concentration of DNA. In addition, the fragment size distribution in the three final pool showed the same as the results from the BioAnalyzer. By this it looks as the DNA in gels with GelRed run different in the gel when high concentrations of DNA are loaded at the gel. Therefore, it would be better to colour the gel after gel electrophoresis, than having GelRed in the gel before running DNA samples.

#### 5.4 Potential results from Illumina sequencing and their reliability

All the steps in the preparation of the transposon library to make it ready for Illumina sequencing were checked after each step was done, to ensure the Illumina sequencing to give readable and trustworthy sequencing results. However, it will always be chances for that contamination will give wrong results. In this case, it could be if some bacteria that not belong to a pool, accidentally got into a wrong pool. For example was the sealing tape that was put on the wells before the incubation sometimes difficult to put on and also to take off. Although these sealing tapes were handled with extreme care when taking them on and off the well plates, the difficulties could have led to contamination between the wells. Other steps that could have led to contamination are that DNA not belonging to the library contaminated the library DNA and was then amplified by PCR. It could also be that some residue from the phagemid used to insert the transposon into the bacteria was still left in the library stock, and then sequences from the phagemid can be found in the Illumina sequencing. However, Sanger sequenced test samples have not matched the phagemid in NCBI BLAST.

When the results from the Illumina sequencing in Texas is ready at the end of June 2015, some wells are expected to have no results because of that they went dry during the incubation time. If there is two or more mutants in the library, which have their transposon insertion in the same TA-site in the genome of *M. avium*, several combinations of coding would be available for these identical mutants, which would make it difficult to trace back to the original well position.

The newly published article in Nature Communications by Vandewalle *et al.*, describes a similar approach to organise a transposon mutant library and identify them by Illumina sequencing as done in this study. The article describes how this is done in *M. bovis* [58]. By looking at their results it is therefore possible to say something about what is expected of results from the library organised and prepared in this study.

Vandewalle *et al.* found that in their library 64% of the non-essential *M. tuberculosis* orthologues were disrupted. Over 77% of the 8259 unique TA-sites in *M. bovis* had sufficient sequence coverage for the software used to identify the position of the gene in the organised library. Around 84% of the mutations were in known genes, which means 57% of the genes in the genome of *M. bovis* were mutated. By testing around 100 of 9216 well positions if the mutants in the specific wells matched the mutants given by Illumina sequencing, all the wells checked had matching mutations with the sequence that the position should have according to the Illumina sequencing results. They also assume that if the number of wells with mutants were

## DISCUSSION

doubled, 90% of the genes that is possible to target by transposon mutagenesis would be covered in the library [58].

With an organised transposon library it is possible to perform large-scale screenings and study the phenotype of all or several mutants in the library at the same time. It is also possible to pick out a few genes of interest, e.g. genes that might be related to iron uptake. After the desired mutant has been isolated, its phenotype can be studied, e.g. to investigate how the mutant will grow in media with low and high content of iron.

### 5.5 The putative role of *msmeg\_4898* and *msmeg\_4900* in mycobacterial iron uptake.

The gene cluster *esx-3* was found to be required for mycobactin-mediated iron uptake by combining mutations in the two different iron uptake pathways. It was indicated that the gene cluster *esx-3* is important for the mycobactin pathway by that the mutant strain where exochelin and *esx-3* were knocked out failed to grow at low iron conditions [30]. This pattern would also be expected to see if the gene *msmeg\_4898* and *msmeg\_4900* is a part of the mycobactin pathway in *M. smegmatis*.

From the growth curves seen in Figure 28 and Appendix 7 and 8, it is seen that the WT and the WT\_Δ4898 are growing similarly under low iron condition, and this is also seen under high iron conditions for WT and the WT\_Δ4898. Also for the Δ*mbtD* and the Δ*mbtD*\_Δ4898 it is seen similar growth curves under low iron condition, and the same under high iron conditions. This indicates that the genes *msmeg\_4898* and *msmeg\_4900* are not a part of the exochelin pathway, because the exochelin pathway here saves the Δ*mbtD*\_Δ4898. The Δ*mbtD*\_Δ4898 would not grow if it was a part of the exochelin pathway, due to the fact that both siderophore pathways then would be knocked out.

In contrast to the double knock-out of exochelin and *esx-3*, the double knock-out Δ*fxbA*\_Δ4898 is still growing under low iron conditions. However, the Δ*fxbA*\_Δ4898 is growing less than the Δ*fxbA* at low iron conditions. This could indicate that *msmeg\_4898* and *msmeg\_4900* are important for iron uptake under low iron conditions. Nevertheless, since the same trend is seen at high iron conditions, the conclusion about *msmeg\_4898* and *msmeg\_4900* being important for mycobactin pathway cannot be taken.

The Δ*mbtD*\_Δ*fxbA* and the Δ*mbtD*\_Δ*fxbA*\_Δ4898 are growing similarly under low iron (3 μM iron added) conditions and also under high iron conditions. If the Δ*mbtD*\_Δ*fxbA* had been

growing better than the  $\Delta mbtD_{\Delta fxbA_{\Delta 4898}}$  on low iron levels, it would have indicated that *msmeg\_4898* and *msmeg\_4900* is a part of another iron uptake system that would take over the iron acquisition when the siderophore pathways do not work.

Although, it seems to be no important role for the genes *msmeg\_4898* and *msmeg\_4900* in the mycobactin pathway, the true role of the genes could be masked if the 200 base pairs flanking sites of the genes, if these are enough to maintain the iron uptake function of the genes.

An interesting observation was made when the mutant colonies from the original agar plates were plated on LB-plates without additional iron. The knock-out mutants didn't grow on these plates. However, they grew on 7H10-plates with 7.5  $\mu\text{g/ml}$  gentamicin 150  $\mu\text{M}$   $\text{FeCl}_3$ . This could be due to that the bacteria without the genes *msmeg\_4898* and *msmeg\_4900* are essential to iron. However, it could be that the bacteria for reasons unknown do not like one or several ingredients in LB-agar, or that the agar lacks one or several important ingredients. The assumptions about the LB-agar are here most likely, due to the fact that the mutants were growing under low iron conditions during the growth curve assay.

## 5.6 Summary/ Conclusion

In this study, it is shown how it is possible to identify the mutated genes in a large *M. avium* transposon mutant library by using Illumina sequencing. However, it will not be possible to tell if this approach was successful before the sequencing results are available at the end of June 2015. Since all the quality checks during the preparation have been positive, it is likely that the results from the Illumina sequencing will allow to link mutants and sequences, thereby generating the desired mutant library for future research. The approach will then also be able to be used for generating other mutant libraries.

The genes *msmeg\_4898* and *msmeg\_4900* do not seem to be important for the mycobactin pathway in *M. smegmatis*. This does not mean that they are not important for this purpose in other mycobacteria.

## DISCUSSION

### 5.7 Further perspectives

Depending on the results from the Illumina sequencing, there will be different further perspectives. If the sequencing results do not allow to link the sequencing reads to specific mutants, error searching has to be done. It could also be interesting if samples that were grown more before DNA purification and were sheared mechanically instead of enzymatically would have different outcome shown through the sequencing.

If the sequencing is successful, this library will be a valuable tool in research. Future use of the library would be to use it to pick out mutants of interest to investigate. With this library it will also be possible to conduct screenings with many thousand mutants at the same time and study their phenotype under different conditions.

The homologues of the genes *msmeg\_4898* and *msmeg\_4900* would be of interest to study also in *M. avium* and *M. tuberculosis*, to see the importance of the homologous genes in iron uptake in these mycobacterial species. The homologous genes in *M. avium* and *M. tuberculosis*, could be important although the genes *msmeg\_4898* and *msmeg\_4900* are not important in *M. smegmatis*. For example, *M. smegmatis* could have developed another system for transporting mycobactin over the outer membrane in the cell wall, than other mycobacteria. To study the homologues of the genes *msmeg\_4898* and *msmeg\_4900* in *M. avium*, the organised transposon library would be of great value, due to the fact that *M. avium* is not susceptible for mutagenesis through recombination.

## 6. References

1. Murray, P.R., K.S. Rosenthal, and M.A. Pfaller, *Medical microbiology*. 6th ed. ed. 2009, Philadelphia: Elsevier Mosby: p: 277, 284-85
2. Stevens, J.M., E.E. Galyov, and M.P. Stevens, *Actin-dependent movement of bacterial pathogens*. *Nat Rev Microbiol*, 2006. **4**(2): p. 91-101.
3. World Health Organization. *Tuberculosis factsheet 104*. 2015 March 2015 27.04.15]; Available from: <http://www.who.int/mediacentre/factsheets/fs104/en/>.
4. World Health Organization. *HIV/AIDS*. 2014 November 2014 28.04.15]; Available from: <http://www.who.int/features/qa/71/en/>.
5. Cambier, C.J., S. Falkow, and L. Ramakrishnan, *Host evasion and exploitation schemes of Mycobacterium tuberculosis*. *Cell*, 2014. **159**(7): p. 1497-509.
6. Kaufmann, S.H., *How can immunology contribute to the control of tuberculosis?* *Nat Rev Immunol*, 2001. **1**(1): p. 20-30.
7. Gengenbacher, M. and S.H.E. Kaufmann, *Mycobacterium tuberculosis: Success through dormancy*. *FEMS Microbiol Rev*, 2012. **36**(3): p. 514-32.
8. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence*. *Nature*, 1998. **393**(6685): p. 537-44.
9. Takayama, K., C. Wang, and G.S. Besra, *Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis*. *Clin Microbiol Rev*, 2005. **18**(1): p. 81-101.
10. Kleinnijenhuis, J., et al., *Innate immune recognition of Mycobacterium tuberculosis*. *Clin Dev Immunol*, 2011. **2011**: p. 405310.
11. Gengenbacher, M. and S.H. Kaufmann, *Mycobacterium tuberculosis: success through dormancy*. *FEMS Microbiol Rev*, 2012. **36**(3): p. 514-32.
12. Velayati, A.A., et al., *Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in iran*. *Chest*, 2009. **136**(2): p. 420-5.
13. Agdestein, A., et al., *Intracellular growth of Mycobacterium avium subspecies and global transcriptional responses in human macrophages after infection*. *BMC Genomics*, 2014. **15**: p. 58.
14. Otero, J., W.R. Jacobs, Jr., and M.S. Glickman, *Efficient allelic exchange and transposon mutagenesis in Mycobacterium avium by specialized transduction*. *Appl Environ Microbiol*, 2003. **69**(9): p. 5039-44.
15. Roy, S., et al., *The ftsZ Gene of Mycobacterium smegmatis is expressed Through Multiple Transcripts*. *Open Microbiol J*, 2011. **5**: p. 43-53.

## REFERENCES

16. NCBI Gene Plot. *Pairwise genome comparison of protein homologs (symmetrical best hits)*. 28.05.15]; Available from: <http://www.ncbi.nlm.nih.gov/sutils/geneplot.cgi>.
17. Niederweis, M., *Nutrient acquisition by mycobacteria*. Microbiology, 2008. **154**(Pt 3): p. 679-92.
18. Banerjee, S., et al., *Iron acquisition, assimilation and regulation in mycobacteria*. Infect Genet Evol, 2011. **11**(5): p. 825-38.
19. Skaar, E.P., *The battle for iron between bacterial pathogens and their vertebrate hosts*. PLoS Pathog, 2010. **6**(8): p. e1000949.
20. Ratledge, C., *Iron, mycobacteria and tuberculosis*. Tuberculosis (Edinb), 2004. **84**(1-2): p. 110-30.
21. Gobin, J. and M.A. Horwitz, *Exochelins of Mycobacterium tuberculosis remove iron from human iron-binding proteins and donate iron to mycobactins in the M. tuberculosis cell wall*. J Exp Med, 1996. **183**(4): p. 1527-32.
22. Wandersman, C. and P. Delepelaire, *Bacterial iron sources: from siderophores to hemophores*. Annu Rev Microbiol, 2004. **58**: p. 611-47.
23. Wells, R.M., et al., *Discovery of a siderophore export system essential for virulence of Mycobacterium tuberculosis*. PLoS Pathog, 2013. **9**(1): p. e1003120.
24. Chu, B.C., et al., *Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view*. Biometals, 2010. **23**(4): p. 601-11.
25. Snow, G.A., *Mycobactins: iron-chelating growth factors from mycobacteria*. Bacteriol Rev, 1970. **34**(2): p. 99-125.
26. Madigan, C.A., et al., *Lipidomic discovery of deoxysiderophores reveals a revised mycobactin biosynthesis pathway in Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences, 2012. **109**(4): p. 1257-1262.
27. Madigan, C.A., et al., *Lipidomic Analysis Links Mycobactin Synthase K to Iron Uptake and Virulence in M. tuberculosis*. PLoS Pathog, 2015. **11**(3): p. e1004792.
28. Luo, M., E.A. Fadeev, and J.T. Groves, *Mycobactin-mediated iron acquisition within macrophages*. Nat Chem Biol, 2005. **1**(3): p. 149-53.
29. Macham, L.P. and C. Ratledge, *A new group of water-soluble iron-binding compounds from Mycobacteria: the exochelins*. J Gen Microbiol, 1975. **89**(2): p. 379-82.
30. Siegrist, M.S., et al., *Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition*. Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18792-7.
31. de Souza, G.A., et al., *Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway*. J Proteomics, 2011. **75**(2): p. 502-10.

32. Dragset, M.S. and b. Norges teknisk-naturvitenskapelige universitet Institutt for, *Riding the ferrous wheel : identification and study of genes involved in mycobacterial iron acquisition*, 2013, Norwegian University of Science and Technology, Faculty of Natural Sciences and Technology, Department of Biotechnology: Trondheim: p. 22
33. Gold, B., et al., *The Mycobacterium tuberculosis IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages*. Mol Microbiol, 2001. **42**(3): p. 851-65.
34. Prakash, P., et al., *Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of Mycobacterium tuberculosis open reading frames*. Bioinformatics, 2005. **21**(10): p. 2161-6.
35. Rodriguez, G.M., et al., *ideR, An essential gene in mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response*. Infect Immun, 2002. **70**(7): p. 3371-81.
36. TubercuList. Search term: *lprd*. 2015 05.05.15]; Available from: <http://tuberculist.epfl.ch/quicksearch.php?gene+name=lprd&submit=Search>.
37. SmegaList. Gene: *MSMEG\_4898*. 2015 05.05.15]; Available from: [http://mycobrowser.epfl.ch/smegmasearch.php?gene+name=MSMEG\\_4898&submit=Search](http://mycobrowser.epfl.ch/smegmasearch.php?gene+name=MSMEG_4898&submit=Search).
38. Rubin-Pitel, S.B., et al., *Chapter 3 - Directed Evolution Tools in Bioproduct and Bioprocess Development*, in *Bioprocessing for Value-Added Products from Renewable Resources*, S.-T. Yang, Editor. 2007, Elsevier: Amsterdam. p. 49-72.
39. Craig, N.L., *Molecular biology : principles of genome function*. 2010, Oxford: Oxford University Press: p. 602
40. Alberts, B., *Molecular biology of the cell*. 5th ed. ed. 2008, New York: Garland Science: p. 317-8, 534
41. Finnegan, D.J., *Transposable elements*. Current Opinion in Genetics & Development, 1992. **2**(6): p. 861-867.
42. Ivics, Z., et al., *Transposon-mediated genome manipulation in vertebrates*. Nat Methods, 2009. **6**(6): p. 415-22.
43. Hamer, L., et al., *Recent advances in large-scale transposon mutagenesis*. Curr Opin Chem Biol, 2001. **5**(1): p. 67-73.
44. Xu, T., N. Bharucha, and A. Kumar, *Genome-wide transposon mutagenesis in Saccharomyces cerevisiae and Candida albicans*. Methods Mol Biol, 2011. **765**: p. 207-24.
45. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Comprehensive identification of conditionally essential genes in mycobacteria*. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12712-7.

## REFERENCES

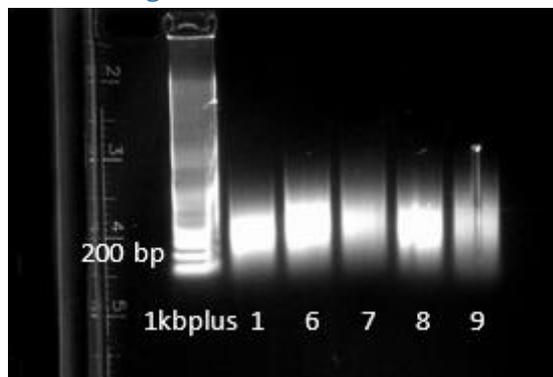
46. Creighton, H.B. and B. McClintock, *A Correlation of Cytological and Genetical Crossing-Over in Zea Mays*. Proc Natl Acad Sci U S A, 1931. **17**(8): p. 492-7.
47. Nakajima, N.I., et al., *Pre-exposure to ionizing radiation stimulates DNA double strand break end resection, promoting the use of homologous recombination repair*. PLoS One, 2015. **10**(3): p. e0122582.
48. Lawrence, J.G. and A.C. Retchless, *The interplay of homologous recombination and horizontal gene transfer in bacterial speciation*. Methods Mol Biol, 2009. **532**: p. 29-53.
49. Kilby, N.J., M.R. Snaith, and J.A. Murray, *Site-specific recombinases: tools for genome engineering*. Trends Genet, 1993. **9**(12): p. 413-21.
50. van Kessel, J.C. and G.F. Hatfull, *Recombineering in Mycobacterium tuberculosis*. Nat Methods, 2007. **4**(2): p. 147-52.
51. Barba, M., H. Czosnek, and A. Hadidi, *Historical perspective, development and applications of next-generation sequencing in plant virology*. Viruses, 2014. **6**(1): p. 106-36.
52. Illumina Inc. *Illumina Sequencing Technology - Highest data accuracy, simple workflow, and a broad range of applications*. 2010 11. October 2010 29.04.15]; Available from: [http://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf).
53. van Opijnen, T., K.L. Bodi, and A. Camilli, *Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms*. Nat Methods, 2009. **6**(10): p. 767-72.
54. van Opijnen, T. and A. Camilli, *Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms*. Nat Rev Microbiol, 2013. **11**(7): p. 435-42.
55. Zhang, Y.J., et al., *Global assessment of genomic regions required for growth in Mycobacterium tuberculosis*. PLoS Pathog, 2012. **8**(9): p. e1002946.
56. Illumina Inc. *Multiplex Sequencing Assay*. 2015 2015 16.05.15]; Available from: <http://www.illumina.com/technology/next-generation-sequencing/multiplexing-sequencing-assay.html>.
57. Pomraning, K.R., et al., *Library preparation and data analysis packages for rapid genome sequencing*. Methods Mol Biol, 2012. **944**: p. 1-22.
58. Vandewalle, K., et al., *Characterization of genome-wide ordered sequence-tagged Mycobacterium mutant libraries by Cartesian Pooling-Coordinate Sequencing*. Nat Commun, 2015. **6**: p. 7106.
59. Pelt-Verkuil, E., et al., *Principles and Technical Aspects of PCR Amplification*, 2008, Springer Science + Business Media B.V: Dordrecht: p. 1, 38

60. Crisafuli, F.A., E.B. Ramos, and M.S. Rocha, *Characterizing the interaction between DNA and GelRed fluorescent stain*. Eur Biophys J, 2015. **44**(1-2): p. 1-7.
61. New England Biolabs. *DpnI*. 2015 04.05.15]; Available from: <https://www.neb.com/products/r0176-dpni>.
62. Rubin, E.J., et al., *In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria*. Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1645-50.
63. CNRGV-INRA. *DNA Pool production*. 2008 08.04.10 28.04.15]; Available from: <http://cnrgv.toulouse.inra.fr/Services/DNA-Pool-production>.
64. New England Biolabs. *NEBNext® dsDNA Fragmentase®* 15.05.2015]; Available from: <https://www.neb.com/products/m0348-nebnext-dsdna-fragmentase>.
65. Thermo Scientific. *PRODUCT INFORMATION: Thermo Scientific CloneJET PCR Cloning Kit* 2015 20.01.15]; Available from: [https://tools.lifetechnologies.com/content/sfs/manuals/MAN0012707\\_CloneJET\\_PCR\\_Cloning\\_UG.pdf](https://tools.lifetechnologies.com/content/sfs/manuals/MAN0012707_CloneJET_PCR_Cloning_UG.pdf).
66. Quadri, L.E., et al., *Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin*. Chem Biol, 1998. **5**(11): p. 631-45.
67. Fiss, E.H., S. Yu, and W.R. Jacobs, Jr., *Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways*. Mol Microbiol, 1994. **14**(3): p. 557-69.
68. Botstein, D. and D. Shortle, *Strategies and applications of in vitro mutagenesis*. Science, 1985. **229**(4719): p. 1193-201.
69. Wu, S.C., et al., *Steps toward targeted insertional mutagenesis with class II transposable elements*. Methods Mol Biol, 2008. **435**: p. 139-51.
70. Aboklaish, A.F., et al., *Random insertion and gene disruption via transposon mutagenesis of Ureaplasma parvum using a mini-transposon plasmid*. Int J Med Microbiol, 2014. **304**(8): p. 1218-25.

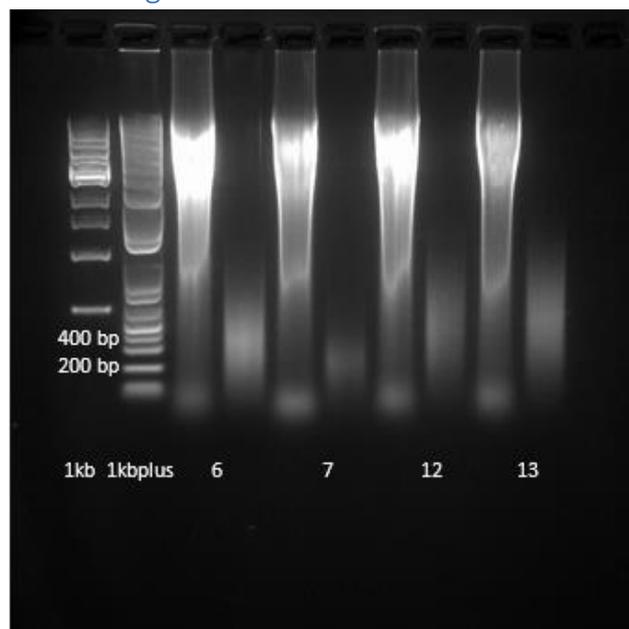
## APPENDICES

### 7. Appendices

Appendix 1: Gel electrophoresis that checks the fragment length of the column pools after being sheared with Covaris.

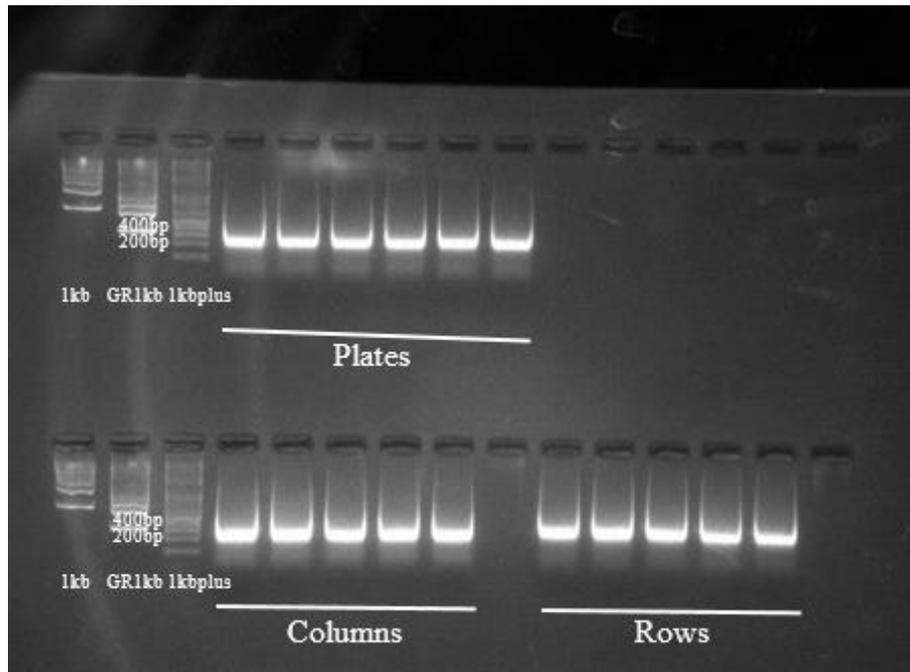


Appendix 2: Gel electrophoresis that checks the fragment lengths of the row pools after being sheared with Covaris.

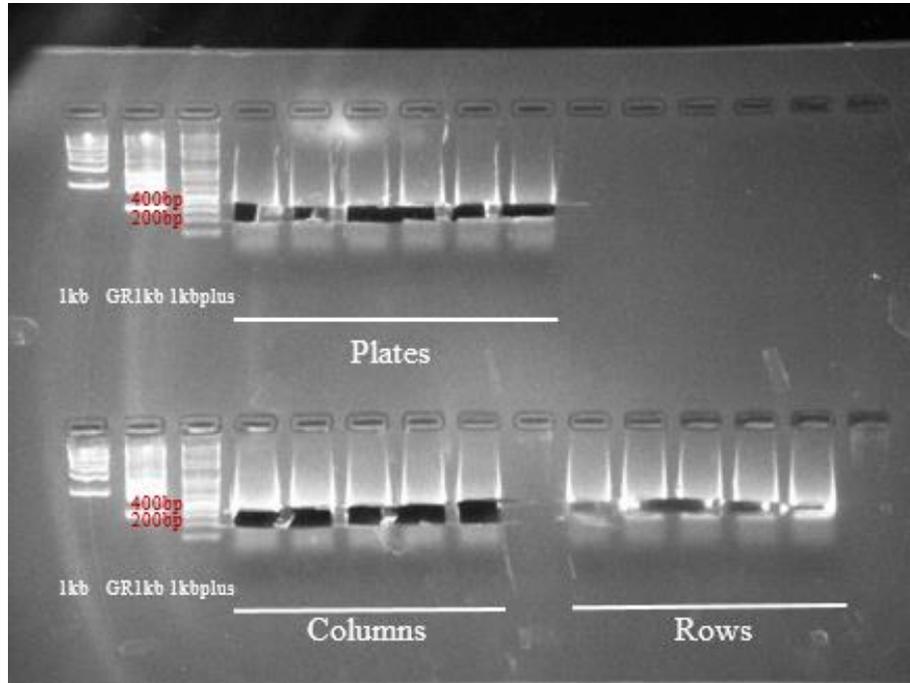


The row pools tested here are number 6, 7, 12 and 13. In front of every lane with the Covaris-sheared fragment, is unfragmented DNA from each respective row pool loaded.

Appendix 3: Gel electrophoresis of PCR-products where all the pools in each category plate, column and row are pooled.

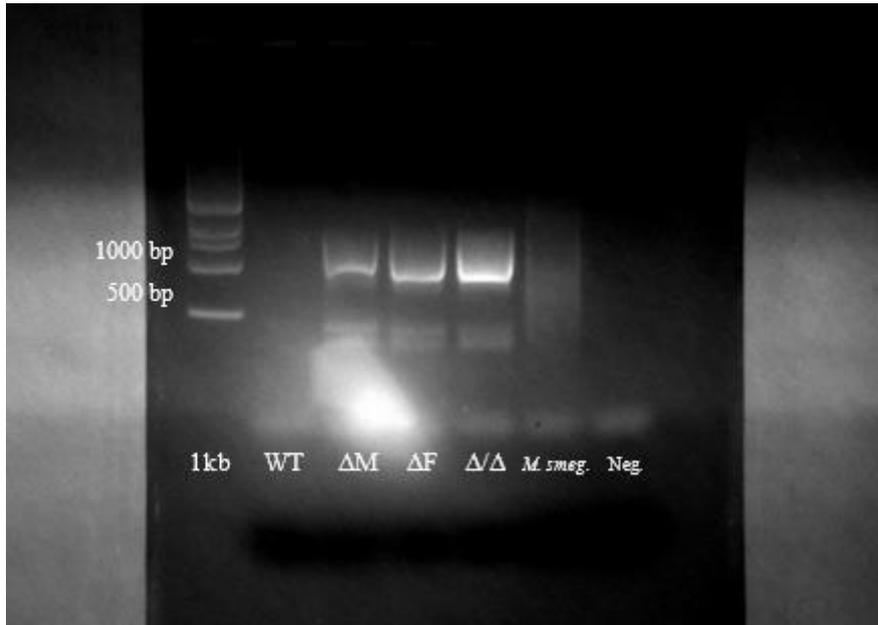


Appendix 4: Gel from Figure 20, after cutting out bands with approximately length of 200 to 400 base pairs.



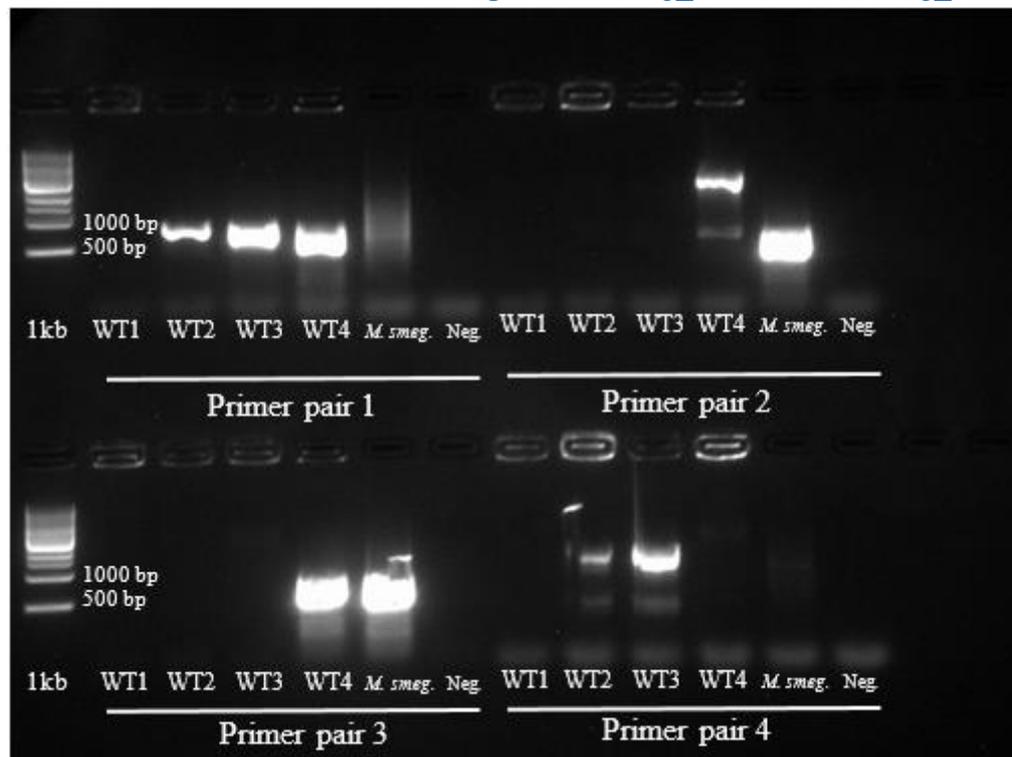
## APPENDICES

Appendix 5: Gel electrophoresis of products from PCR with primer pair 2, 3 and 4 to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.



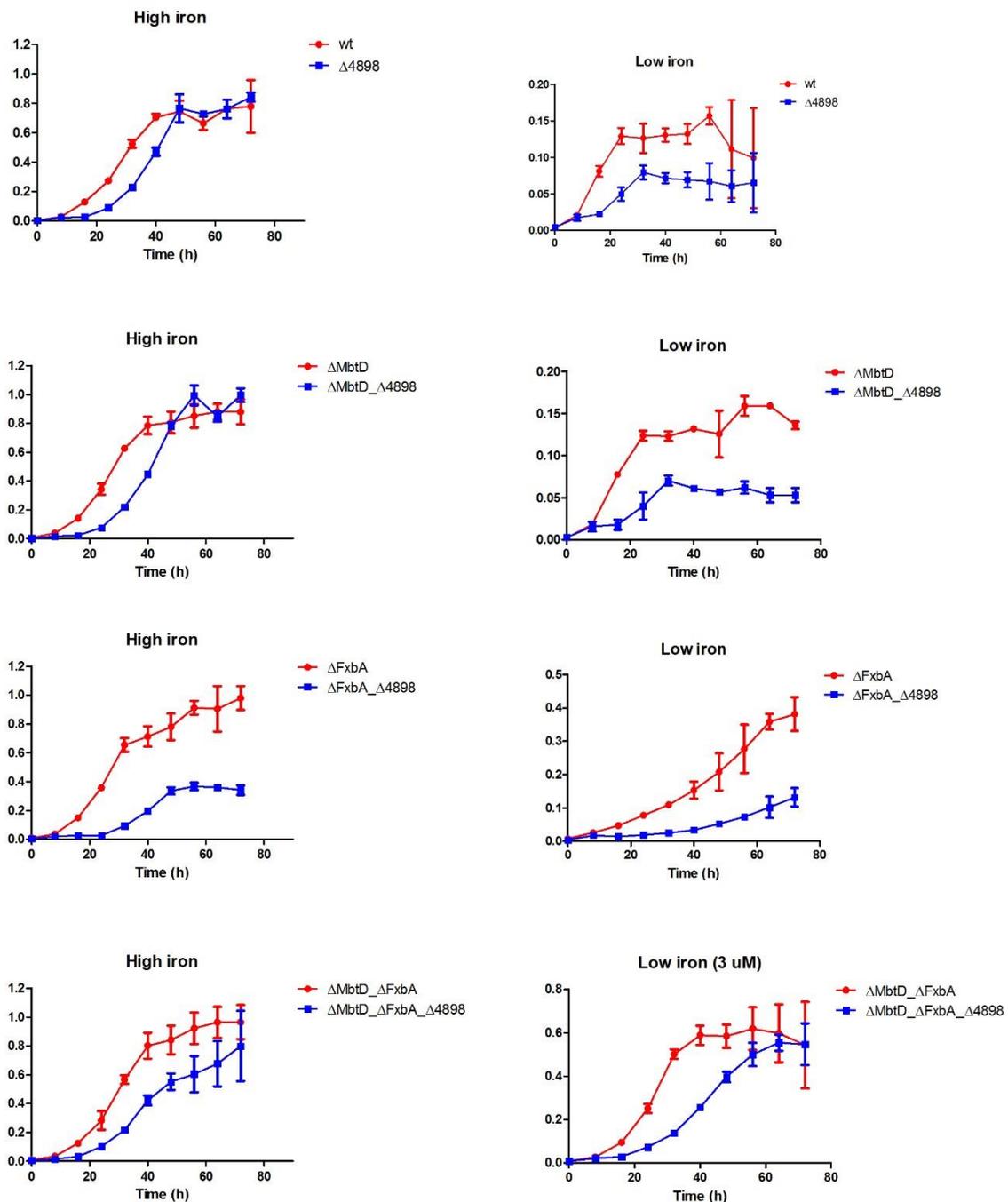
WT is wild type,  $\Delta$ M is mycobactin knock-out mutant,  $\Delta$ F is exochelin knock-out mutant and  $\Delta/\Delta$  is a mutant with both mycobactin and exochelin knocked out. The control used is *M. smegmatis* that had not been recombined.

Appendix 6: Gel electrophoresis of WT PCR-products from PCR with primer pair 1, 2, 3 and 4 to confirm knock-out of the genes *msmeg\_4898* and *msmeg\_4900*.



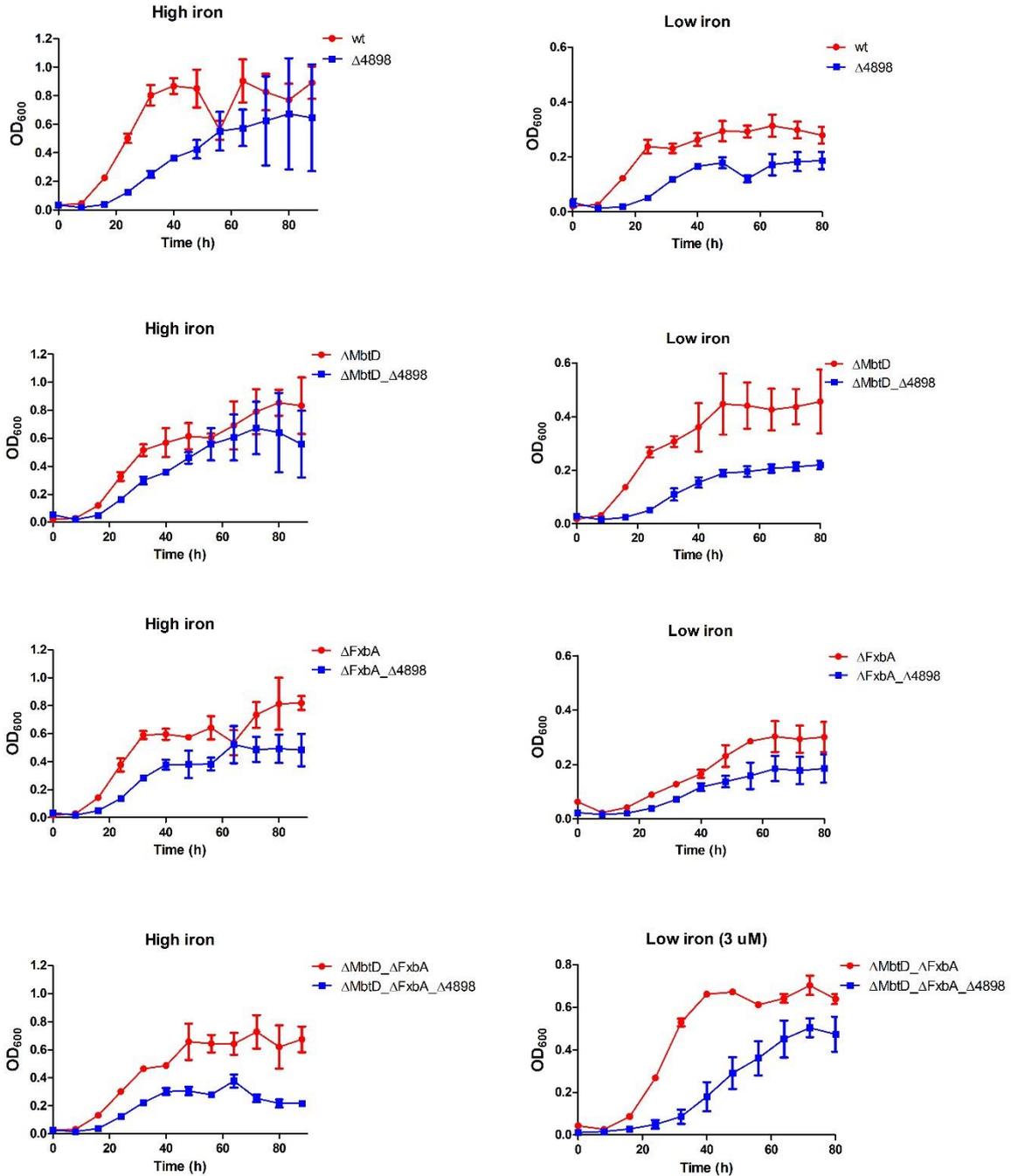
This was done to find a WT-candidate. WT3 was the candidate chosen. WT is wild type,  $\Delta$ M is mycobactin knock-out mutant,  $\Delta$ F is exochelin knock-out mutant and  $\Delta/\Delta$  is a mutant with both mycobactin and exochelin knocked out. The control used is *M. smegmatis* that had not been recombined.

## Appendix 7: The second replicate of growth curve analysis.



In order to assess the potential function of *msmeg\_4898* and *msmeg\_4900* in mycobactin-mediated iron uptake, the *msmeg\_4898-4900* knockout mutants ( $\Delta 4898$ ) were prepared for growth curve analysis. There was made high and low iron conditions for all the candidates, and each condition was set up with three technical replicates in the well plate for the growth curve instrument, Bioscreen C MBR. The OD<sub>600</sub> was read every second hour in three days by Bioscreen C MBR. This experiment was repeated three times. The four bacteria candidates of *M. smegmatis* analysed, both with and without the genes *msmeg\_4898* and *msmeg\_4900* knocked out were; the wild type (WT), a mycobactin knock-out mutant ( $\Delta mbtD$ ), an exochelin knock-out mutant ( $\Delta fxbA$ ) and a double knock-out mutant, where both mycobactin and exochelin are knocked out ( $\Delta mbtD \Delta fxbA$ ).

Appendix 8: The third replicate of growth curve analysis.



In order to assess the potential function of *msmeg\_4898* and *msmeg\_4900* in mycobactin-mediated iron uptake, the *msmeg\_4898-4900* knockout mutants ( $\Delta 4898$ ) were prepared for growth curve analysis. There was made high and low iron conditions for all the candidates, and each condition was set up with three technical replicates in the well plate for the growth curve instrument, Bioscreen C MBR. The OD<sub>600</sub> was read every second hour in three days by Bioscreen C MBR. This experiment was repeated three times. The four bacteria candidates of *M. smegmatis* analysed, both with and without the genes *msmeg\_4898* and *msmeg\_4900* knocked out were; the wild type (WT), a mycobactin knock-out mutant ( $\Delta MbtD$ ), an exochelin knock-out mutant ( $\Delta fxbA$ ) and a double knock-out mutant, where both mycobactin and exochelin are knocked out ( $\Delta MbtD \Delta fxbA$ ).