

From the Department of Microbiology, Tumor- and Cell Biology  
Karolinska Institutet, Stockholm, Sweden and the Swedish  
Institute for Communicable Disease Control, Department of  
Preparedness, Solna, Sweden

**PCR AND SEQUENCING  
BASED METHODS FOR  
DETECTION AND TYPING OF  
PATHOGENIC  
MICROORGANISMS**



**Karolinska  
Institutet**

**Smi**

**SMITTSKYDDSinSTITUTET**

Solna 2011

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Tara Wahab, 2011

ISBN 978-91-7457-594-1

Printed by



[www.reprint.se](http://www.reprint.se)

Gårdsvägen 4, 169 70 Solna

To my wonderful family, the best family in the world



## ABSTRACT

The potential use of bacteria and viruses as biological terror weapons makes certain highly pathogenic microorganisms a worldwide public health threat. In an outbreak investigation involving a possible deliberate spread of a biological warfare agent there is a need for a fast and reliable diagnostic method. Ideally, the same method should be usable for both bacteria and viruses. To distinguish between natural and deliberate spread, information on the prevalence/incidence of the organism/disease, ecology and natural mechanisms for spread is needed. In addition, knowledge of national and international subtypes reflecting both micro- and macro-evolution of the organisms is required for the analysis.

The field of molecular epidemiology has evolved during the last decades with the introduction of several different methods to type and characterize bacteria. There is now a plethora of different molecular techniques available to identify and discriminate between different strains of bacteria. In this thesis, rapid protocols based on pyrosequencing technology were developed and used for discrimination of *Bacillus anthracis* from closely related bacillus species. The technique was also used for identification and subtyping of *Francisella* species, including the human pathogen *Francisella tularensis* and its subspecies. Also a rapid generic protocol for detection and genotyping of infectious agents including bacteria, parasites, and viruses was developed.

Tularemia epidemics occur in limited geographical areas, and at variable intervals. The reasons for these irregularities are still unknown. There may be ecological niches in the affected areas, which harbor the bacteria between epidemics, or the bacterium may be reintroduced into an ecological system that permits its amplification. The establishment of a relevant typing system indentifying individual strains is of great importance .One of the most promising methods, MLVA was evaluated as a tool for practical use in epidemiological investigations of tularemia in Sweden. PFGE analysis was used as reference method. Typing data were combined with geomapping (GIS), in order to predict sources of infection and determine possible reservoirs for the different MLVA types. The results indicate that MLVA has the capacity to be used as a standalone typing method in outbreak investigations. It was possible to use this method directly on clinical specimens without isolating the bacterium by cultivation. In most cases, a certain MLVA type was not correlated to a strict geographic location, indicating that subtyping might be of limited use in surveillance and outbreak investigations of tularemia in endemic countries.

## LIST OF PUBLICATIONS

- I. **Wahab T**, Hjalmarsson S, Wollin R and Engstrand L. Pyrosequencing *Bacillus anthracis*. *Emerg Infect Dis*. 2005 Oct;11(10):1527-31.
- II. **Wahab T**, Ankarklev J, Lebbad M, Glavas S, Svärd S, Palm D. Real-time polymerase chain reaction followed by fast sequencing allows rapid genotyping of microbial pathogens. *Scand J Infect Dis*. 2011 Feb;43(2):95-9.
- III. Jacob D\*, **Wahab T\***, Edvinsson B, Peterzon A, Boskani T, Farhadi L, Barduhn A, Grunow R and Sandtröm G. Identification and subtyping of *Francisella* by pyrosequencing and signature matching of 16S rDNA. *Letters in Applied Microbiology*. 2011 (ahead of print).
- IV. **Wahab T**, Hjertqvist M, Hedenström I and Löfdahl S. Evaluation of MLVA for epidemiological typing of *Francisella tularensis*-experiences from an endemic country. In manuscript.

\*Both authors contributed equally.

# CONTENTS

1	Introduction.....	1
1.1	Historical perspectives.....	1
1.2	General background.....	2
1.3	Genome variability and population structure.....	3
	of bacteria .....	3
1.4	Genomic typing methods.....	3
2	Methods .....	6
2.1	PCR/real-time PCR.....	6
2.2	DNA sequencing technologies .....	6
	2.2.1 Sanger sequencing.....	6
	2.2.2 Pyrosequencing .....	7
2.3	16S rRNA and <i>rpoB</i> analysis .....	8
2.4	PFGE analysis.....	9
2.5	VNTR.....	10
3	Aims.....	11
3.1	Ethical Considerations.....	11
4	Results and Discussion.....	12
4.1	Pyrosequencing technology as a rapid method for identification of <i>Bacillus anthracis</i> .....	12
4.2	Real-Time PCR followed by fast sequencing: an optimized, scaleable diagnostics pipeline for screening and characterization of polymorphic microbial pathogens .....	15
4.3	Rapid identification of <i>Francisella</i> by Pyrosequencing of 16S rDNA .....	18
4.4	Evaluation of MLVA for epidemiological typing of <i>Francisella tularensis</i> - experiences from an endemic country .....	21
5	Concluding remarks .....	24
6	Acknowledgements .....	26
7	References.....	29

## LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
mRNA	Messenger RNA
RNA	Ribonucleic acid
SARS	Severe acute respiratory syndrome
US	United states
H1N1	Influenza A (H1N1) virus
<i>B. anthracis</i>	<i>Bacillus anthracis</i>
<i>F. tularensis</i>	<i>Francisella tularensis</i>
SNP	Single-nucleotide polymorphism
HRM	High resolution melting
MLVA	Multiple-locus variable-number tandem repeat analysis
PFGE	Pulsed-field gel electrophoresis
MLST	Multilocus sequence typing
VNTRs	Variable number of tandem repeats
PCR	Polymerase chain reaction
dNTP	Deoxynucleotide-triphosphate
Ct	Cycle threshold
PPi	Pyrophosphate
ATP	Adenosine-5'-triphosphate
CCD	Charge coupled device
HIV	Human immunodeficiency virus
rRNA	DNA encoding rRNA gene
<i>G. lamblia</i>	<i>Giardia lamblia</i>
BSL3	Bio-safety level 3
BLAST	Basic local alignment tool
SLV	Single locus variants
DLV	Double locus variants
GIS	Geographical Information System
CDC	Center for disease control and prevention

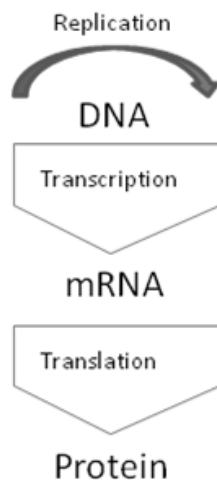


# 1 INTRODUCTION

## 1.1 HISTORICAL PERSPECTIVES

A German pathologist, named Friedrich Henle, declared in 1840 that microorganisms could cause human disease, and by doing so launched the “germ theory”. This theory was confirmed by Koch and Pasteur in the 1870s and 1880s by proving that microorganisms were responsible for causing anthrax, plague and tuberculosis. Since then there has been a great deal of effort directed toward diagnosing microorganisms; first by culturing and later by using microbial DNA as a target for identification.

The central dogma of molecular biology describes how genes encoded by DNA sequences are transcribed to mRNA, which in turn is translated into functional protein (Fig. 1). This was first proposed by Francis Crick in 1957 and became the basis of the hypothesis that the DNA sequence of a gene corresponds to the amino acid sequence of a protein [1]. By 1966, the entire genetic code was independently determined by the groups of Khorana and Nirenberg, enabling prediction of protein sequences by translation of DNA sequences. Ten years later, robust techniques for rapid DNA sequencing were introduced [2-3], allowing for sequencing of large DNA molecules like the 16.5 kb human mitochondrial genome [4]. Since then the sequencing techniques and especially the enzymatic chain termination method of Sanger, have been further developed and adapted to different kinds of automation. A dramatic increase in sequence throughput has been accomplished, and complete sequencing of large genomes has become achievable. Nucleic acids have been used for pathogen detection, as every species of pathogen carries a unique DNA or RNA signature that differentiates it from other organisms. The challenge is to establish this signature for each microorganism of interest for rapid and specific detection, and for that purpose reliable and accurate tools are needed. In this thesis strategies for DNA based methods for detection and typing of different pathogens will be described. The basic techniques as well as general procedures for sequence production and typing analysis will be discussed.



**Figure 1.** The central dogma of genetic information.

## 1.2 GENERAL BACKGROUND

The public and scientific view of the threat from infectious diseases has increased during the last decade. The SARS outbreak, the bio-terror attacks in the US, the avian flu, the H1N1 pandemic, the threats of re-emerging diseases due to climate changes etc, have shown that a new focus on infectious diseases is needed. Increasing rates of antibiotic resistance and the changing epidemiology of different bacterial strains dictate the need to develop new methods to study, control, and prevent infections. Rapid and sensitive diagnostic methods are important in order to provide correct treatment, but are also crucial on the population level for the detection and prevention of new epidemics.

During the last ten years there has been a rapid development of molecular methods for detection and typing of infectious agents. This has resulted in better understanding of many infectious diseases. Molecular methods are of special value in the work regarding highly virulent and contagious agents like *B. anthracis* and *F. tularensis*, as viable preparations should be avoided. Molecular identification offers an attractive alternative method to identify slow growing bacteria, such as *mycobacteria*, because it is rapid and a single technique can be used to identify a variety of different species. Sequencing is an even more powerful molecular identification method, since it answers the question “What is the unknown isolate?”.

Studies combining molecular methods with epidemiological data can provide important information regarding different infections and the results can be used in outbreak investigations and surveillance. Molecular typing techniques allow investigators to determine whether the strain causing disease in one patient is identical to that causing disease in another patient, in order to demonstrate whether transmission has taken place or not in a clinical setting.

In the genomic era, there has been an explosion in the development and utilization of molecular typing for discrimination of bacterial strains in forensics, strain tracking and outbreak investigations as well as in population structure investigations [5]. Molecular typing methods have also been used to assess bacterial population structures in conjunction with epidemiological parameters. Differences in clinical presentation, manifestation, and outcome, as well as geographic distribution have been correlated with these subpopulations [6-8]. Molecular typing methods differentiate among bacterial subpopulations by utilizing the genetic diversity generated by mutations, recombination, and gene transfer. The discriminatory power of a typing method is mathematically defined as the probability that two randomly chosen strains from a population of unrelated strains will be distinguished by the typing method [9]. The collection of strains used in evaluation of a typing system is of utmost importance. A test population, suitably identified at the species level, should reflect as much diversity as possible within the species to determine the discriminatory power of the method.

During the last decades an increasing number of DNA-based techniques fulfilling requirements for rapid detection and accurate identification of different infectious agents have been developed. The work included in this thesis describes molecular techniques for diagnostics and typing, as well as evaluation of some typing methods.

### **1.3 GENOME VARIABILITY AND POPULATION STRUCTURE OF BACTERIA**

Mutations are responsible for all genetic variations within a single replicating genome [10]. Mutations occur as substitutions, duplications, insertions or deletions of nucleotides or as larger deletions and other reorganisations of the genome resulting in loss of genetic functions or entire genes. Duplication of genes can result in the evolution of new functions. Bacteria can also evolve by gaining new genetic material. In many species, chromosomal DNA can be rapidly exchanged between closely related bacteria by homologous recombination, and this event may affect the population structure. Homologous recombination appears to be closely related to the degree of similarity of the recombining DNA, but the recombination rate is inversely proportional to the sequence divergence [11]. The need for homology in order to establish new genetic material can be circumvented by the transfer of plasmid DNA. This DNA is independent of recombination and is maintained in the recipient. Moreover, it has been shown that genetic material of distant foreign origin can also be integrated into the bacterial chromosome, as in the case of transposons and other genomic islands. These changes may affect pathogenicity, antibiotic resistance, increased fitness, or virulence [12].

Variability in bacteria also depends on natural selection and population size. In a large population mutations will occur more frequently than in a small population, and if the mutation is advantageous for the bacteria, the new variant will accumulate within the population by the process of natural selection. Population structures may vary between different bacterial species; some of them display constant horizontal transfer of genetic material and others display inherited gene alterations [13]. Understanding the population structure of a given pathogen is important since it determines which typing method is the most optimal for a specific pathogen.

Microevolution in bacteria is defined as a change in allele frequencies due to mutation or selection. One way to define macroevolution is the combined effects of microevolutionary processes in a larger population with separate gene pools spread over time and space (geographical spread). Stable markers, like housekeeping genes such as *rpoB* and ribosomal genes 16S and 23S, have proven valuable in studies of long term evolution. The mutation rates of these markers are not sufficient to study the microevolution. Tandem repeats of single or oligo nucleotides on the other hand are often associated with a rather high degree of variability and are consequently appropriate for discrimination of individual bacterial isolates within a bacterial species, with little sequence variation between isolates. Tandem repeats are notable hot spots for variability as a consequence of differences in the number of repeats. The mechanism for this has been proposed to be associated with the replication process [14]. For most types of mutations, different parts of a genome generally display different mutation rates, as exemplified by the fact that there is more variability in non-coding regions than in coding regions of genomes.

### **1.4 GENOMIC TYPING METHODS**

At present, there is a growing need to discriminate bacteria at the strain level. This is necessary for a correct epidemiological investigation of a communicable disease, i.e. to elucidate the transmission chain of an infectious disease. In addition, increasing

concerns regarding the emergence and spread of bacterial resistance to antibiotics also requires methods for tracking of individual bacterial strains. Different methods, such as biotyping and serotyping, have been used to address this. Development of molecular typing methods for analysis of bacterial genomes has allowed for assessment of the relationships among different species, as well as among individual isolates.

Culturing of pathogens, a classical bacteriological method, is routinely used in diagnostic laboratories. Some bacterial species can be identified by their fermentation reactions or growth on a certain media, or by phenotypic characteristics, which have been traditionally used for identification purposes. Others methods, making use the genetic differences among strains have, however, been developed.

To determine the relatedness of bacteria or to trace the source in outbreaks, many different typing methods are available. Various techniques have been used to achieve DNA based typing of bacteria such as SNP analysis including e.g. DNA sequencing and high resolution melting (HRM) [15], and multiple-locus variable-number tandem repeat analysis (MLVA) including e.g. DNA sequencing or fragment analysis. Within the last ten years there has been an ongoing revolution in molecular typing of different pathogens with a move away from gel-based methods, such as restriction fragment length polymorphism and pulsed-field gel electrophoresis (PFGE). In certain cases it has been possible to substitute PFGE with novel non gel-based methods, while in other cases these have been a complement to existing PFGE. The choice of typing methods depends on the purpose for which the generated data will be used. Some of the questions may be answered by a classical phenotypic assessment, such as biotype and serotype, while others may require a level of resolution not provided by classical methods, such as the further epidemiological tracking of a strain. A decisive method should be simple, rapid, cheap, safe, reproducible, and give easily interpreted data, as well as accurate estimates of relationships between both distantly related microorganisms as well as those closely related. Today different existing methods often characterize compromises between these criteria.

The favored approaches for typing are based on MLVA and multi-locus sequence typing, MLST. Data generated by such methods are easily stored, and used to develop international databases that are easily accessible for comparisons between laboratories. The use of multiple loci avoids the dangers of inaccurate conclusions being drawn from single loci, the evolution of which may not necessarily reflect that of the genome as whole. In addition, the molecular clocks of the various elements used as markers in these methods differ from each other, and thus provide epidemiological information at different levels. In general, multi-locus sequence approaches use markers with slow molecular clocks, in order to monitor evolution over many thousands of years, while multi-locus tandem repeat markers can have a fast molecular clock more suited to local epidemiological methods, such as outbreak investigation or identification of reactivation. Tandem repeat sequences display allelic hypervariability related to variation in the number of repeats caused by slipped strand mispairing or replication slippage, or DNA recombination between homologous repeat sequences. These markers are known as variable number of tandem repeats (VNTRs), and are considered high speed molecular clocks. They are useful for the discrimination of bacterial species such as *B. anthracis* [16] *Brucella* [17], *F. tularensis* [18], *Mycobacterium tuberculosis*

[19] and *Yersinia pestis* [20]. While PFGE typing usually detects a broader range of the genome as compared to VNTR and standard MLVA it does not necessarily provide a relevant perspective of the entire genome. The disadvantage of all these methods is that they are limited to a number of known markers, which cannot reflect all the events in the genome. An MLVA typing system for *F.tularensis* with 25 VNTR was recently developed by Johansson et al. [18].

The best method to determine the identity of a bacterium is to sequence its entire genome. This is, however, an unrealistic method for typing as it is still too expensive. Instead, commonly used methods targeting genomic variability such as MLST, MLVA and PFGE, are frequently used for molecular epidemiology in different laboratories as well as in this thesis.

## 2 METHODS

### 2.1 PCR/REAL-TIME PCR

Polymerase chain reaction (PCR) is a very powerful technique [21], used to amplify a segment of DNA (template) by using two oligonucleotides (flanking primers) in a series of synthetic reactions (PCR cycles) catalyzed by a DNA-polymerase in the presence of a mixture of the four dNTPs [22]. This technique was initially very labor intensive, since the enzyme was not heat stable, and consequently degenerated by the heat needed to degenerate the DNA in between each cycle of PCR. When the heat stable DNA polymerase was isolated from the bacterium *Thermus aquaticus*, this technique was drastically simplified, as the need to replace the polymerase between cycles was eliminated [22]. Perkin Elmer commercialized the first thermal cycler. This technique has its limitations, namely poor quantification abilities, as scientist quantified the PCR products in agarose gels by measuring the intensity of the ethidium bromide stained gel using UV sensitive cameras. Another limitation was the inability to calculate the efficiency of the PCR reactions. These problems were addressed by the development of real-time PCR.

Higuchi et al further developed PCR to Real-Time PCR by monitoring the reaction so that the efficiency of the reaction could be calculated [23]. The capacity to monitor PCRs in real time has revolutionized the way PCR is used in the clinical microbiology field. This technique is used to amplify and concurrently quantify a targeted DNA molecule and enables both detection and quantification of DNA. Real-time PCR needs a fluorescent reporter that binds to the formed product and reports its presence by fluorescence. There are a number of dyes and probes available. The reporter generates a fluorescent signal that reflects the amount of formed product. In the beginning of the process the signal is weak and cannot be distinguished from the background, but when the product accumulates, the signal increases exponentially. Finally, the signal levels off and is saturated, due to the reaction running out critical components such as primers, reporters, or dNTPs. The number of cycles required to reach this threshold is called the CT value. The amplification response curves are expected to be parallel in the growth phase of the reaction, and the setting of the threshold level should therefore not be critical. Different instrument software use different methods and algorithms to select the threshold.

We have used either conventional or real-time PCR techniques for DNA amplification in all of our studies in order to detect infectious agents and/or to make templates for sequencing.

### 2.2 DNA SEQUENCING TECHNOLOGIES

#### 2.2.1 Sanger sequencing

Sequencing is the process to determine the order of the nucleotide bases along a DNA strand. The most important techniques for DNA sequencing are the enzymatic chain termination method [3], which generate a nested set of single stranded DNA fragments which are separated by size on acrylamide gels. The Sanger method (enzymatic chain termination) is based on the principle that single-stranded DNA molecules differ in

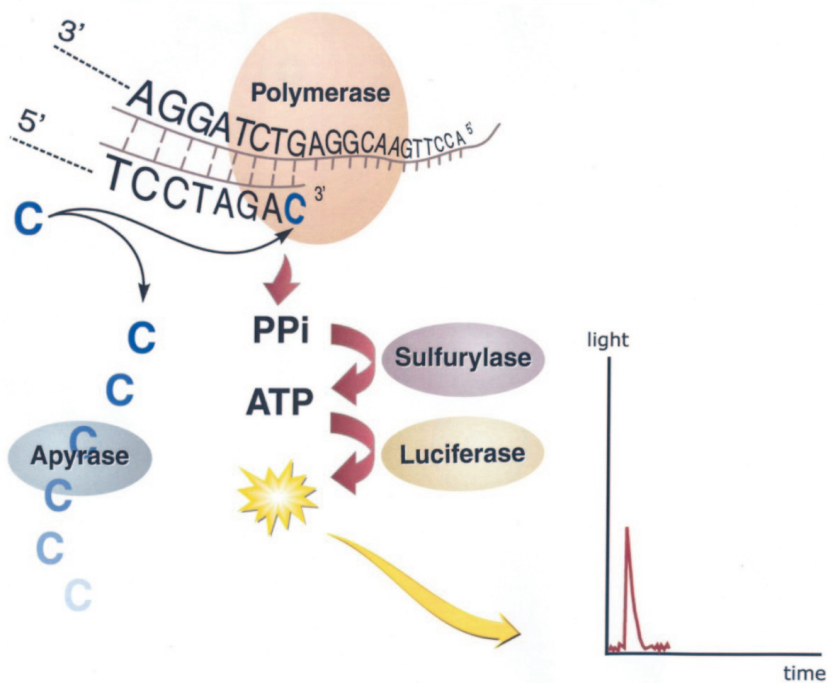
length by just a single nucleotide, and can be separated from one another using polyacrylamide gel electrophoresis. The DNA template is first prepared as a single-stranded DNA, and then a short oligonucleotide is annealed to the same position on each template strand. The oligonucleotide acts as a primer for the synthesis of a new DNA strand that will be complementary to the template DNA. After the first deoxynucleotide is added to the growing complementary sequence, DNA polymerase moves along the template and continues to add base after base. The strand synthesis reaction continues until a dideoxynucleotide is added, blocking further elongation and the reaction terminates. Therefore, the sequencing result is a set of new chains, all of different lengths. This method gives a high yield of detailed data, but has its disadvantages, mainly by being costly, both in terms of apparatus and reagents, but also by needing trained personnel for the laboratory work and maintenance. In 1998, a new sequencing method was reported by a Swedish group at the Royal Technical Institute in Stockholm [24]. The method was named pyrosequencing, as one of the essential products in the reaction was pyrophosphate. Instead of separating the products on gels, this method detects each nucleotide as it is inserted in the elongated DNA chain, giving a signal output much earlier than Sanger sequencing.

### **2.2.2 Pyrosequencing**

Pyrosequencing (Fig. 2) is a sequencing technique based on luminometric real time monitoring of DNA synthesis. In this method, the released pyrophosphate (PPi) generated by nucleotide incorporation of DNA polymerase is converted to ATP by the enzyme sulfurylase. Thereafter, ATP serves as substrate for the light-harvesting enzyme luciferase. The nucleotides are thus added to the DNA enzyme mixture one at the time, iteratively, in a cyclic manner. In order to remove unincorporated nucleotides and ATP between each cycle of nucleotide addition, the nucleotide-degrading enzyme apyrase is included in the system. A CCD camera finally detects the light produced as the end product of the three enzyme cascade (Klenow polymerase, ATP sulfurylase and luciferase) [24]. So far, the method has been used for applications where sequence primers can be designed to produce a read length sufficient for a correct result. Examples include single nucleotide polymorphism (SNP) analysis [25-29], identification and subtyping of bacteria [25], mutation detection in disease genes [30-31], and quantification of and mutation detection in HIV-1 [32-33]. The pyrosequencing technology has been further developed and is today also available as a tool for large-scale sequencing (genome Sequencer FLX, 454 Life Science/Roche). Other full genome sequencing technologies have recently become available, such as the Illumina Genome Analyzer System (Solexa/Illumina) and SOLID (Applied biosystems). Commonly, the laboratory part of all the different methods constitute a minor part, as data handling, follow up sequencing and bioinformatics are still issues to be solved for the application outside the research lab. The costs of sequencing methods remain too high for them to be implemented in routine diagnostics, despite drastically reduced prices. We have developed this method further in paper I and applied it in paper III.

Pyrosequencing has been applied to organism detection by combining short stretch DNA sequencing with signature matching in a suitable characterized phylogenetic

target, 16S rRNA, in addition to a variety of target genes in bacteria [25, 34-35]. Even though pyrosequencing yields limited amounts of DNA sequence information, pathogen detection could still be made [25]. The utility of pyrosequencing to rapidly identify bacterial species by analysis of variable regions in 16S and 23S rDNA has been demonstrated in previous studies [34, 36]. Speed, cost-effectiveness, robustness and a user-friendly design makes the technique an attractive choice in a routine microbiological laboratory. A drawback for some applications as compared to Sanger sequencing, is, however, its relatively short read-length, which is usually limited to approximately 70 nucleotides [37].



**Figure 2.** The nucleotides A, C, G and T are added sequentially, one at a time, to the primer-template complex. The incorporation of a nucleotide results in the production of one molecule of pyrophosphate (PPi) which is converted to a light flash through a cascade of enzyme reactions. Any non-incorporated nucleotides are broken down by the enzyme apyrase.

### 2.3 16S rRNA AND *rpoB* ANALYSIS

The 16S rRNA gene is highly conserved among bacteria. Its function is essential for the existence of the bacteria, and it is a structural part of the 30S ribosomal small subunit [37]. Therefore, sequence analysis of this gene is widely utilized for identification and phylogenetic classification of bacterial species. However, the 16S rRNA gene is not



useful for typing at the strain level, as the DNA sequence is too conserved, except among a few species, such as *N. meningitidis* [38]. Other genes, with less conserved sequences, have been utilized for both species identification and strain typing. The choice of appropriate genes may vary according to the species. The gene *rpoB* which encodes the  $\beta$  subunit of RNA polymerase, has been used for species identification and subtyping [39]. 16S rDNA analysis was used in paper III for species and subspecies determination. *rpoB* analysis was used for species determination in paper I.

## 2.4 PFGE ANALYSIS

PFGE involves embedding organisms in agarose, lysing the organisms, and digesting the chromosomal DNA with restriction endonucleases that cleave infrequently in the gel [38]. Slices of agarose containing the chromosomal DNA fragments are inserted into the wells of an agarose gel, and the fragments are resolved into a pattern of discrete bands in the gel by an apparatus that switches the direction of current in a predetermined manner. The DNA restriction patterns of the isolates are then compared with one another to determine their relatedness. PFGE represents a gold standard for typing of many microbes such as *Salmonella enteritidis* [39]. This method has been standardized and became a prioritized tool for use in large networks of laboratories in both the USA (PulseNet) and Europe (PulseNet Europe), and this method is used in paper IV. This technique was developed by Schwartz and Cantor [40] by altering the voltage gradient used in normal gel electrophoresis to improve the resolution of larger molecules, due to the size of the fragments being resolved, and the fact that the DNA does not move in a straight line through the gel. Small fragments can find their way through the gel matrix more easily than large DNA fragments. Nevertheless, with a periodic change of field direction, the various lengths of DNA react to the change at differing rates, i.e. larger fragments of DNA will be slower to realign their charge when field direction is changed, while smaller fragments will be quicker. Over the course of time, and with the consistent changing of directions, each band will begin to separate more and more even at very large lengths. Thus a separation of very large DNA fragments by use of PFGE is possible. Instead of constantly running the voltage in one direction, the voltage is periodically switched in three directions: one that runs through the central axis of the gel and two that run at an angle of 120 degrees either side. The pulse times are equal for each direction resulting in a net forward migration of the DNA.

PFGE typing has produced stable and highly reproducible profiles of *F. tularensis* isolates in our laboratory. This method has been used to provide laboratory data for epidemiological characterization and tracing of *F. tularensis* isolates. PFGE analysis is however time-consuming and has a very low resolution. Every gel holds only 30 lanes of which at least three are occupied by a DNA size marker. Processing a gel requires laboratory personnel to work a minimum of five days from the making of bacterial suspensions until the digitalized band patterns are obtained. In our case it creates a risk of laboratory acquired infections due to the extremely contagious nature of *F. tularensis*. PFGE may be used for genotyping or genetic fingerprinting. It is commonly considered a gold standard in epidemiological studies of pathogenic organisms. The choice of typing method depends on both the aim of the assay and the

organism, a discriminatory power of methods varies widely between bacterial species with limited sequence differences, such as *F. tularensis*.

## 2.5 VNTR

Prokaryotic genomes contain a wide array of repetitive DNA elements ranging from single-nucleotide repeats to large, complicated repeats of dozens of nucleotides. VNTRs are repeats that are found in tandem and demonstrate interstrain variability. Multi-locus VNTR analysis (MLVA) has become a reliable way to establish genetic relatedness for epidemiological typing of organisms such as *B. anthracis* [16], *Escherichia coli* O157:H7 [41], *F. tularensis* [42], *Salmonella enteritidis* serovar Typhimurium [43].

VNTR typing is a PCR-based method which amplifies the repetitive region using primers localized in the flanking conserved sequences. The PCR products of the tandem repeat loci are either sequenced or size determined by fragment analysis in order to reveal the number of tandem repeats. In paper IV fragment analysis was performed by using fluorescein labeled primers and capillary electrophoresis to fractionate the PCR amplicons. The actual number of tandem repeats within each PCR amplicon was determined by using a computer program, Gene Mapper® (Applied Biosystem).

For *F. tularensis* MLVA with 25 VNTR loci has enabled discrimination between individual isolates [18]. Five of these VNTR loci were used in paper IV.

### **3 AIMS**

The general aim of this thesis has been the development and/or evaluation of methods for detection and typing of serious pathogens.

The specific aims were:

- I. To discriminate between *B. anthracis* and closely related bacillus species by use of pyrosequencing technology.
- II. To develop a rapid generic protocol for detection and genotyping of infectious agents. The aim was to develop a protocol applicable for bacteria, parasites, and viruses.
- III. To evaluate a rapid method based on 16S PCR and pyrosequencing for the identification and subtyping of *Francisella*, including the human pathogen *F. tularensis* and its subspecies.
- IV. To evaluate MLVA as a tool for practical use in epidemiological investigations of *F. tularensis* infections in Sweden, and combine the results with geomapping (GIS), in order to predict sources of infection and determine reservoirs of the bacteria.

#### **3.1 ETHICAL CONSIDERATIONS**

##### Study II

The study was approved by the Ethics Committee of the Medical Faculty, Karolinska Institutet, and by the Regional Ethical Review Board in Stockholm (2006/911-31/3).

##### Study IV

The study was approved by the Ethics Committee of the Medical Faculty, Karolinska Institutet, and by the Regional Ethical Review Board in Stockholm (2008/1020-31/2).

## 4 RESULTS AND DISCUSSION

### 4.1 PYROSEQUENCING TECHNOLOGY AS A RAPID METHOD FOR IDENTIFICATION OF *BACILLUS ANTHRACIS*

*Bacillus anthracis*, the causative agent of anthrax, is a Gram-positive, aerobic or facultatively anaerobic, endospore-forming and rod-shaped bacterium. It can cause acute infection in both animals and humans. Anthrax is primarily a disease of herbivores, which acquire the infection after coming into contact with soil borne spores. The distribution of anthrax is worldwide, and the disease can be transmitted to humans when spores of *B. anthracis* are introduced by ingestion, inhalation or skin contact. *B. anthracis* is a potential bio-terror agent. To identify *B. anthracis*, both conventional and molecular methods have been applied. Presumptive identification is based on demonstrating a lack of  $\beta$  hemolysis on sheep horse blood agar plates and the organism's lack of motility. *B. anthracis* has two plasmids: the toxin encoding pXO1 (182 kb) and the capsule encoding pXO2 (95 kb) [44-45]. Both plasmids are needed for virulence, and the lack of either of the plasmids results in attenuation of the bacteria. The pXO1 plasmid contains genes *lef*, *cya*, and *pag*, which encode the toxin lethal factor, edema factor, and protective antigen, respectively [46-48]. The pXO2 plasmid contains the genes *capA*, *capB* and *capC*, necessary for capsule formation [49]. These genes have been used as markers to identify *B. anthracis* both in environmental and clinical samples by PCR [50-51].

In the autumn of 2001[52] anthrax was utilized for bioterrorism in a series of attacks where spores from the bacteria were packaged into conventional letters and sent to embassies and governmental institutions worldwide. The practice still occurs occasionally. Some of these letters did indeed contain anthrax spores and 22 persons fell ill. Five patients inhaled the spores and succumbed to the most deadly form of the disease. Many other letters was sent but turned out to contain harmless white powder and all other investigated suspected materials were negative for *B. anthracis*. This occurrence stimulated an international bio-terror scare and increased many countries awareness of preparedness. The interest for *B. anthracis* increased, which can be seen in the increased numbers of scientific publications on this microorganism.

This bio-terror event resulted in efforts to establish rapid protocols for *B. anthracis* detection in clinical and environmental samples. In order to trace infectious agents used for bioterrorism it is important to identify and characterizes the organism with high accuracy. Speed, sensitivity and accuracy are desirable, but often conflicting, properties of diagnostic methods. One such example is the use of separate methods for detection and genotyping of infectious agents. Primary detection is often performed quickly by molecular techniques or microscopy, but genotyping requires more time and resources. A rapid method for genotyping is clinically important since different subtypes of microbes have been shown to be associated with different symptoms and differences in susceptibility to antibiotics [53-57]. Genotyping is also fundamental for the epidemiological understanding of transmission routes and the diversity of infectious agents [58]. Thus, development of new detection methods that are fast, reproducible, inexpensive, user friendly and possible to combine with subsequent genotyping is of great importance.

Many of the available methods for identification of *B. anthracis* are either time consuming or have limited specificity. For this reason a new technical assay that enables rapid and reliable species identification is important and essential as a part of the preparedness against bioterrorism [51, 59]. In this study, we investigated the possibility to use the pyrosequencing technology as a tool for identification of *B. anthracis*. This method has previously been shown to be efficient for the analysis of a wide range of microbial markers [60-61]. One of the main benefits of the pyrosequencing technology is the capacity to perform rapid analyses of multiple samples: up to 96 samples can be analysed in 30 minutes. This technology provides a rapid and accurate way of adding validity to PCR-based assays by qualitatively verifying a positive PCR. The risk of false positives is thereby minimized, which is very important for the treatment of anthrax patients. The *rpoB* gene was used as a chromosomal marker for *B. anthracis* in this study. The *rpoB* gene is common target for real-time PCR analysis for *B. anthracis* [62]-[63] as well as for other bacterial species [64] and was chosen in order to make discrimination between *B. anthracis* and other closely related species of the *Bacillus* genus possible.

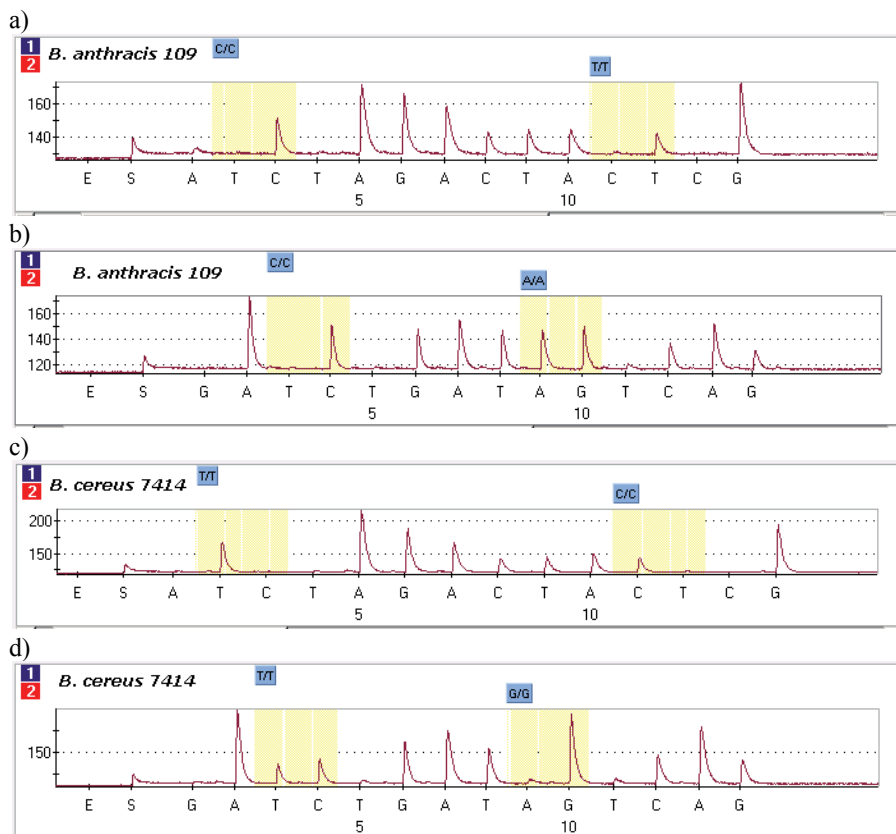
In order to genotype two SNP positions in one single reaction, the *B. anthracis* strains were run in duplicate in the pyrosequencing reaction. SNP genotyping was achieved by sequencing several fragments in one single well using a unique sequencing primer for each desired SNP. The resulting pyrogram was a mix of the overlapping sequences (Fig. 3). The dispensation order had to be designed to generate unique peaks for each fragment in the multiplex genotyping reaction. In order to secure analytic precision, one or two polymorphic peaks from each fragment were analysed without nucleotide incorporation from the other fragment. The advantage of this duplex reaction is that it is time-saving and the cost of the analysis is reduced. The use of one well for all four SNP positions may also further optimise the method. One fragment from each plasmid was sequenced (Fig. 4 and Fig. 5) in order to investigate the virulence of the strains in our collection.

The 16S rRNA gene has also been used for identification of unknown bacteria as well as for phylogenetic analysis. Two fragments from the 16S rRNA gene of *B. anthracis* were amplified and sequenced in this study. The accession number AY138358 was used to design the PCR- and sequencing primers. The positions of interest in the AY138358 sequence were 77, 90, 92, 182, 189, 192, 200, 208, 1015, 1036 and 1045. Unfortunately we were unable to rely solely on the 16S rRNA gene, as these positions are not specific for *B. anthracis*, but our other *Bacillus* species have the same sequence as *B. anthracis* in some of these positions (data not shown).

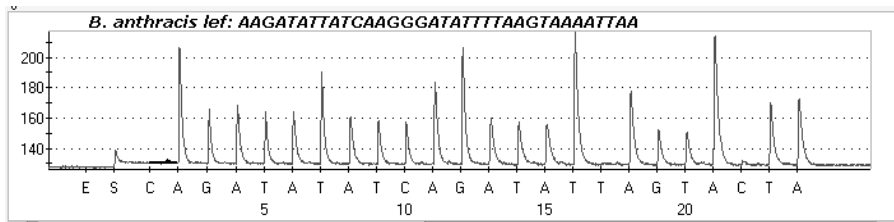
PCR assays combined with rapid and accurate DNA sequencing are valuable methods that have the potential to enhance speed and sensitivity in diagnosis and monitoring of microbial infections. The pyrosequencing technology is powerful for sequence verification of PCR products (19). Although a primer pair may be selected to specifically amplify only a specific target, it is still important to verify that the amplified fragment is of the correct origin, which is especially important when using PCR for diagnostic purposes. Pyrosequencing provides a rapid way of adding confidence to PCR-based assays by qualitatively verifying that a positive PCR result is

not the effect of mis-priming and amplification of unwanted DNA targets. SNPs are principally appropriate markers for subtyping since they are evolutionarily stable and amenable to high-throughput detection methods, such as pyrosequencing and real time PCR [25, 62, 65].

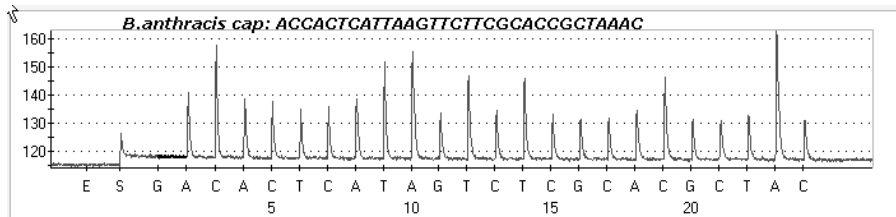
In conclusion, the present study confirms the usefulness of high throughput sequencing of PCR amplicons in diagnostic bacteriology. Volokhov et al., have shown that multiprobe microarray hybridization can be used for identification of *B. anthracis* [66], and here we illustrate for the first time how pyrosequencing technology can be used for identification of the *B. anthracis*. The use of pyrosequencing in diagnostic laboratories has advantages such as rapidness, simplicity, no presence of radioactive substances, relatively low costs, automatization and a high level of accuracy.



**Figure 3.** Pyrograms of duplex SNP determinations of the *rpoB* gene showing *B. anthracis* specific nucleotides 911C, 912T (a) and 913C, 914A (b). The corresponding nucleotides in *B. cereus* are 911T, 912C (c) and 913T, 914G (d).



**Figure 4.** Pyrogram from *B. anthracis* plasmid pXO1. The sequencing fragment is from a region of the *lef* gene located on this plasmid.



**Figure 5.** Pyrogram from *B. anthracis* plasmid pXO2. The sequencing fragment is from a region of the *cap* gene located on this plasmid.

#### 4.2 REAL-TIME PCR FOLLOWED BY FAST SEQUENCING: AN OPTIMIZED, SCALEABLE DIAGNOSTICS PIPELINE FOR SCREENING AND CHARACTERIZATION OF POLYMORPHIC MICROBIAL PATHOGENS

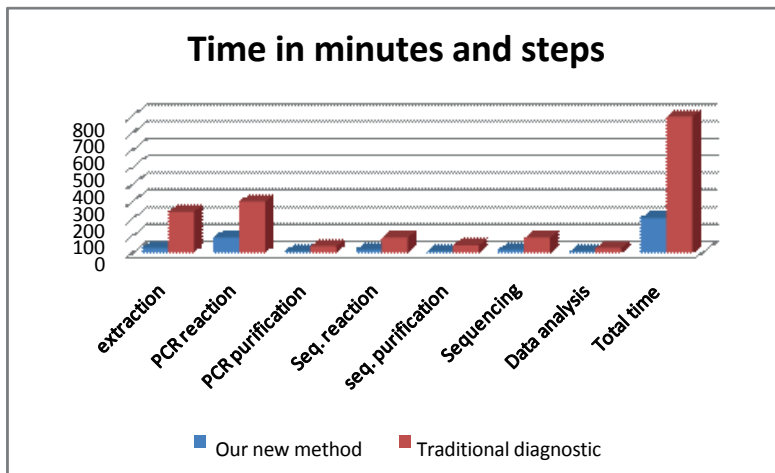
The goal of this study was to produce an easy-to-perform, reproducible clinical diagnostic screening method with a direct sequencing pipeline. The aim was a method that would be open to any organism, any polymorphism, any template, as well as any level of automation, and using only readily available clinical laboratory equipment, as well as meeting the turnaround time demanded by today’s medical community. As a model for this protocol we used *Giardia lamblia*. *G. lamblia* (synonymous with *Giardia intestinalis* and *Giardia duodenalis*) is a flagellated, binucleated, tetraploid protozoan parasite that colonizes and replicates in the small intestine, causing giardiasis. *G. lamblia* infects a wide range of vertebrates including humans, livestock and pets [67]. Currently eight different genotypes or assemblages have been identified [67-68]. Only genotypes A and B have been found in humans [69]. The prevalence of *G. lamblia* in humans varies between 2 to 5% in the industrial world to 20-30% in the developing world. The clinical symptoms vary among patients, ranging from asymptomatic infections to malabsorption and severe chronic diarrhea [67, 70-71]. The parasite is transmitted through ingestion of cysts from sources such as contaminated drinking water, and it has been associated with several large water-borne outbreaks [72]. Primary diagnosis of *G. lamblia* infection is based on microscopy of stool specimens for the presence of *G. lamblia* cysts or trophozoites; however, analysis of single stool specimen has been shown to be effective in only 70% of all cases [73-75]. The sensitivity of microscopy-based diagnostics depends highly on the experience of the microscopist and the number of examined samples. This results in a costly and

time-consuming method lacking the ability to discriminate between different genotypes. Alternative diagnostic methods are needed to increase sensitivity and reduce labour time. PCR-based assays for direct detection of *G. lamblia* in stool samples have been described, with similar or improved sensitivity as compared to microscopy [76-77]. A rapid method for genotyping is clinically important, as mentioned in paper I, since different subtypes of microbes have been shown to be associated with different symptoms and differences in susceptibility to antibiotics [53-55, 57]. Genotyping is also fundamental for the epidemiological understanding of transmission routes and the diversity of infectious agents [58]. Thus, development of new detection methods that are fast, reproducible, user friendly and possible to combine with subsequent genotyping is of great importance.

The *G. lamblia* beta-giardin gene was chosen as target for the PCR and sequencing analyses since it contains conserved regions suitable for screening, as well as characterized polymorphisms that can be used as genetic markers for genotyping [78-79]. Primers were designed using Primer Express 3.0 (ABI), with three deviations from the manufacturer's guidelines: (i) the forward primer was designed with an M13 tail at the 5' end, (ii) a polymorphic region in the reverse primer required the use of two equimolar primers, resulting in amplicons differing by one bp, and (iii) the 286 or 287 bp amplicon is longer than the 150 bp suggested by the manufacturer for optimal Real-Time assays. To investigate if these modifications had any effect on Real-Time PCR efficiency, we ran 10-fold dilution series over five logarithms on a *G. lamblia* control DNA template. A PCR efficiency of 99% was obtained (mean slope = -3.23).

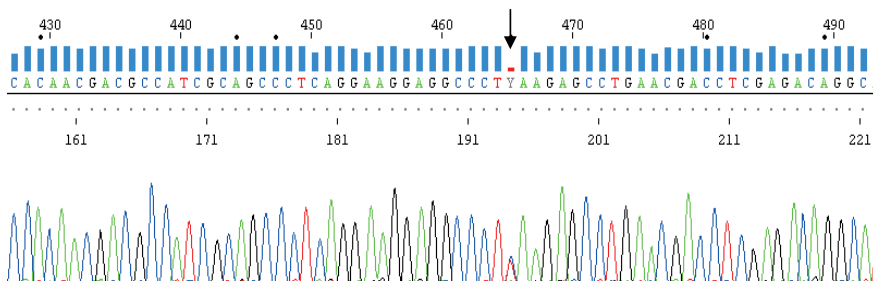
Forty-two clinical samples, all positive for *G. lamblia* by microscopy and nested beta-giardin PCR (except one sample that was negative in nested beta-giardin PCR) [79], were also tested. All samples were positive in Real-Time PCR. Amplicons from positive samples subsequently entered the sequencing pipeline. The total run time from PCR cleanup, Big Dye incorporation, dye terminator clean up, sequencing, through completed data analysis was approximately 70 minutes (Fig 8). All samples were minimally 180 bases ( $QV \geq 20$ ;  $\leq 1\%$  probability of a miscalled base), generally starting 60 bases from the 3'-end of the M13 sequencing primer. The maximum read length was 280 nucleotides. *G. lamblia* genotypes A and B could easily be distinguished and identification of some of the major sub-genotypes was possible. Further, *Giardia* genotypes C to F from infected dogs, sheep, and cats could also be unambiguously identified and discriminated, adding the possibility of implementing the methodology in routine veterinary diagnostic laboratories (data not shown). The sensitivity of our PCR systems, compared to the previously published beta-giardin nested PCR [79], was evaluated, and there were no significant differences between the two methods. All PCRs were specific to *G. lamblia*, and no amplification was detected when testing DNA from a panel of other infectious microorganisms, including *Giardia*-negative faecal samples.





**Figure 8.** Time table for all steps. Blue columns represent the new method and the red columns represent the old method with nested PCR.

By having tailed M13 primers in the front-end PCRs, the sequencing pipeline has no optimization for any assay coming in. A totally universal sequencing step removes human error. The risks that come with a universal primer are the risks of contamination and cross contamination. This is the reason why we favour a cleanup reaction that can be added to the sample, rather than transferring the sample to a new plate, running through filters, or ethanol precipitation. The sample remains in the sequencing plate even as it is loaded and run on the sequencing instrument. An additional benefit is the time saving generated by the highly effective cleanup step as seen in (Fig. 8).



**Figure 9.** *Giardia lamblia* mixed infection. Chromatogram for patient sample D146 with both nucleotides T and C, which indicates double infection with both genotypes A and B of *G. lamblia*.

Some of the chromatograms from clinical samples showed double peaks at one or more positions (Fig. 9). The number and positions of double peaks in *G. lamblia* assemblage B, using our method were in agreement with the results found when using the standard nested beta-giardin PCR and Sanger sequencing. Double peaks present in sequences within assemblage B may be a result of allelic sequence divergence or mixed infections with multiple sub-genotypes [79]. Out of the 42 clinical samples, three indicated multiple double peaks in positions corresponding to double infections with genotypes A and B; previously two of these were characterized as single genotype B infections, and

one was not detected in the nested beta-giardin PCR (sample D146, Fig. 9). These results indicate that this new method is likely to be more sensitive and potentially adds an advantage to screening patient material for mixed infections, as compared to the original nested beta-giardin PCR.

In this study we have developed an innovative protocol for combining Real-Time PCR with a fast sequencing protocol. The proposed methodology allows for increased time savings through further optimization of the Real-Time run. At the present, however, sequencing of amplicons generated using fast Real-Time PCR chemistry has not yielded satisfactory results. The full process from extracted DNA to analyzed sequence is performed within three hours (Fig. 8). The overall benefit of this method, with the inclusion of an M13 tail, allows for the creation of a rapid, universal diagnostic screening and confirmation platform for diagnostic use as well as for subsequent genotyping of any nucleic acid under ~300bp. The use of this method is worth considering, if applied to cases where rapid screening and genotyping is required.

#### **4.3 RAPID IDENTIFICATION OF *FRANCISELLA* BY PYROSEQUENCING OF 16S RDNA**

*F. tularensis* is a small Gram-negative, capsulated, non-motile, aerobic bacterium [80]. This intracellular bacterium is the causative agent of tularemia, a zoonotic disease commonly known as rabbit fever. Tularemia is a severe disease that occurs throughout the Northern hemisphere. It is endemic in Sweden with up to several hundred reported cases annually. The genus *Francisella* comprises two species: *F. tularensis* and *F. philomiragia*, and within *F. tularensis* there are four subspecies, two of which are pathogenic for both humans and animals. *F. tularensis* subspecies *tularensis* (Type A), is highly virulent, and found in North America, whereas *F. tularensis* subspecies *holarctica* (Type B), is moderately virulent and found in Europe, Asia and North America. The less virulent *F. tularensis* subspecies *novicida* has been isolated in North America and Australia and *F. tularensis* subspecies *mediasiatica* in the former Soviet Union. Tularemia can occur in different clinical forms; depending on the route the bacteria enter the body. Inhalation of infected dust leads to the respiratory form of tularemia. Ulceroglandular tularemia is seen after transmission by arthropods or direct contact with infected animals. Oropharyngeal tularemia is seen after oral ingestion of contaminated food or water. The incubation period is usually 3 to 6 days. The initial symptoms after infection with subspecies *tularensis* and *holarctica* are similar, and include chills, fever, muscle pain, and headache. Tularemia caused by *F. tularensis* may rapidly progress to severe disease with septic shock and potentially a fatal outcome. It is a challenge to understand the geographic distribution of type B subpopulations as they occur throughout Northern Hemisphere.

Culture of *F. tularensis* should be attempted only in Bio-safety level 3 (BSL3) laboratories and immunization of the staff should be undertaken as a safety measure. *F. tularensis* is relatively slow growing, and for diagnostic purposes the bacteria require, incubation for 5-8 days in order to exclude growth. That is why the detection of bacterial DNA in patient samples is an attractive alternative to culture, especially since the work can be based on non-viable bacteria and provides rapid analysis in a Bio-safety level 2 laboratory environments. Due to its virulence *F. tularensis* is included

among the top 6 category A pathogens as a potential bioterrorism agent. If deliberate release of the organism is suspected, there is an urgent need to understand the pathogenic potency of an isolate as well as its putative origin. Whenever tularemia appears in an area believed to be free from the agent, characterization of isolates will become urgent. In the case of naturally occurring tularemia the methods for typing and classification of the bacteria are important for epidemiological purposes.

In the present study pyrosequencing was used to analyze V1 16S rDNA sequences of 32 reference strains and 96 Swedish isolates from cases of ulceroglandular tularemia. The 16S rRNA is a structural part of the 30S ribosomal small subunit and its function is essential in bacteria. It consists of 8 conserved and 9 variable regions in between the conserved region V1-V9 [81]. The goal was to develop a method for rapid identification of suspected *Francisella* subspecies isolates using a previously validated set of primers. Although the protocol is generic, and can be used to screen for a large variety of bacterial species, its application here depends on a more rapid approach with dispersion order specifically adapted to the *Francisella* genus sequence. Additional tests were successfully performed to further shorten the time for analysis using a SYBR green qPCR protocol with the ABI7500 FAST platform. The analysis could be completed in less than 1.5 h (data not shown), which is useful not only from a clinical point of view, but also in the context of an outbreak or a suspected bioterrorism event, as the index case may be rapidly identified and treatment initiated at an early stage. The clinical usefulness of the method was demonstrated by analysis of samples from patients with ulceroglandular tularemia collected over a 13-year period. Single nucleotide polymorphism (SNP) analysis, which is able to discriminate between virulent *F. tularensis* subtypes, first used to test reference strains, and subsequently applied to patient samples. All samples were shown to contain markers of human virulence, according to the pyrosequencing results. A 37 nucleotide sequence specifically identifies the *Francisella* genus (Fig. 10) and further subgrouping is possible by SNP analysis. The subgroup, to which all patient isolates were allocated, includes the human virulent subspecies *F. tularensis* subspecies *tularensis*, *F. tularensis* subspecies *holarctica* and *F. tularensis* subspecies *mediasiatica*. A second subgroup includes *F. tularensis* subspecies *novicida* and *F. philomiragia* (Fig. 11).

Human pathogenic isolates can be discriminated from that less significant to human disease, e.g. environmental strains or contaminants in order to determine whether or not an isolate should be processed in a BSL-3 laboratory for further analysis. In conclusion, the method should be considered as a rapid molecular tool to identify *Francisella* and its subspecies with significance to human disease. Further epidemiological investigations require high resolution techniques, such as MLVA or whole genome sequencing.

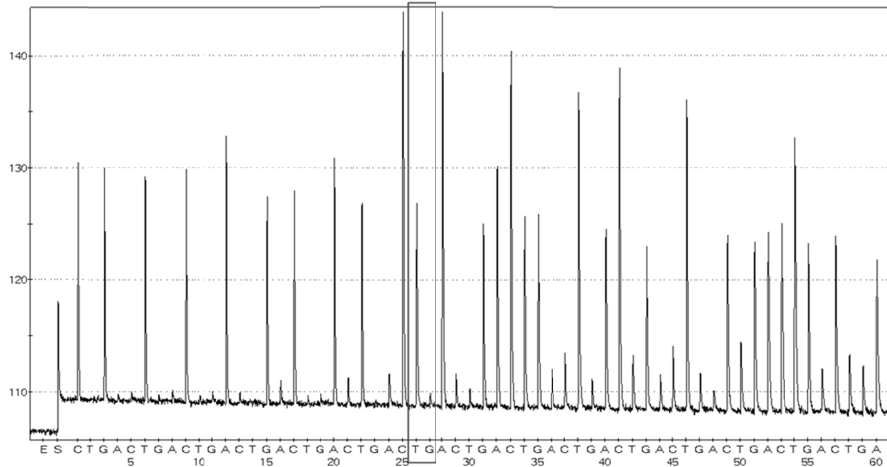
## Pyrogram

Well D1: *F. tularensis* ssp. *holarctica*

Entry: Cyclic\_60\_CTGA

Result: CGTCAGCATCCTAAGACCTGTTACCGTTCGACTTGCA

Quality: Passed (Quality window: 20)



**Figure 10.** Pyrogram of the 37 bases sequence used to discriminate virulent subspecies of *F. tularensis* (*tularensis*, *holarctica* or *mediasiatica*) from less virulent, *novicida*. The SNP T/G used for discrimination is framed.

Sequence Name	< Pos = 1
<input checked="" type="checkbox"/> Consensus	CGTCAGCATCCTAAGACCTGTTACCGTTCGACTTGCA
5 Sequences	
F_tularensis SchuS4_V1.seq	CGTCAGCATCCTAAGACCTGTTACCGTTCGACTTGCA
Ft ssp holarctica ATCC29684 (LVS)_V1	CGTCAGCATCCTAAGACCTGTTACCGTTCGACTTGCA
Ft ssp. mediaasiatica FSC147_V1.seq	CGTCAGCATCCTAAGACCTGTTACCGTTCGACTTGCA
Ft ssp. novicida ATCC15482_V1.seq	CGTCAGCATCCGAAGACCTGTTACCGTTCGACTTGCA
F philomiragia_ATCC25015_V1.seq	CGTCAGCATCCGAAGACCTGTTACCGTTCGACTTGCA

**Figure 11.** The alignment of two groups of *F. tularensis* according to the framed SNP at position 12.

It is obvious that *Francisella* diagnostics using PCR may detect not only human pathogens, but also several closely related species that appear to be of limited clinical relevance. This should be considered when designing new assays for clinical diagnostics or for detection of potentially deliberately released pathogen. Non-virulent *Francisella* in the environment may create misidentification and false alarm signals in either case. The clinical application of the method was demonstrated by analysis of 96 Swedish patient samples collected from patients with ulceroglandular tularemia over a 13 year time period, and all 96 were unambiguously identified by determining the SNP position. Our results show that the method has a clinical application but can also be used to identify less virulent strains, e.g. environmental strains.

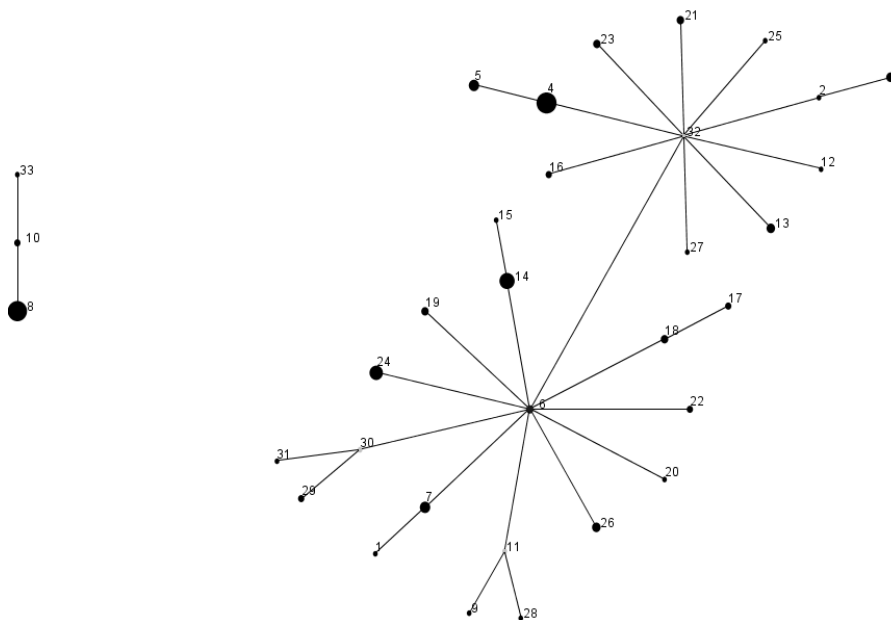
#### 4.4 EVALUATION OF MLVA FOR EPIDEMIOLOGICAL TYPING OF *FRANCISELLA TULARENSIS* - EXPERIENCES FROM AN ENDEMIC COUNTRY

MLVA based on the five published VNTR loci was successfully used in this study for the typing of *F. tularensis* isolates from Swedish patients. Cluster analysis assigned the 127 samples into 33 MLVA types. However, attempts to correlate geographic origin of the isolates to specific MLVA types did not generate a clear cut result. The samples analysed in the present study represent 127 patients from a large geographical area of an endemic country during a time frame of 16 years. Ft-M3, Ft-M6, Ft-M20, Ft-M22 and Ft-M24 displayed repeats sizes of 9, 21, 12, 6 and 21 base pairs, respectively. The most variable allele was observed for the fastest evolving locus, Ft-M3, with 17 different repeat variations, and this locus provided an enormous capacity to discriminate among our isolates. The Ft-M6 and Ft-M20 demonstrated 3 different allele variations, while the Ft-M22 and Ft-M24 showed only 2 variations. Although 25 VNTR loci have been described by Johansson et al [18], we chose 5 VNTR loci based on their published diversity. High mutation rates for VNTR loci can result in homoplasmy, with identical VNTR loci in strains with different genetic backgrounds. Isolates may share mutational changes for reasons other than common ancestry, which could result in an improper isolate link. To exclude homoplasmy our isolates were analyzed by a second typing method, PFGE.

In this work, the restriction enzyme *Pme I* was used for PFGE typing of the patient isolates. Although two other restriction enzymes were also tested, namely *XhoI* and *BamHI*, *Pme I* was chosen due to its higher discrimination power for our isolates. PFGE analysis of the 124 clinical isolates revealed three different but rather similar PFGE profiles (types 1-3). The Bionumerics (Applied Maths, Inc.) version 3.5 software was used to make comparisons of the band patterns. The discriminatory power of typing methods differs, depending on the approach of the method and the possible variability of the studied feature. For PFGE the discriminatory power can vary depending on the used restriction enzyme. For MLVA, on the other hand, the discriminatory power depends on the number of loci and the features of these loci. PFGE is not only very laborious to perform, but also needs viable bacteria. There are also difficulties regarding comparisons of results between laboratories, which is why we set up a real-time-PCR, based on canonical single nucleotide polymorphisms (canSNP) (unpublished data Wahab et al). Our results suggest that this method has the same resolution as PFGE and appears to be a promising tool for the study of the macro-evolution of *F. tularensis*.

The MLVA data was analyzed by eBURST (<http://eburst.mlst.net/>), which uses allele profiles to display the relationships among closely related isolates from a bacterial population [82-83]. It uses an algorithm to classify the founder of the population by classifying the VNTR type that differs from the others by only one (single locus variants, SLV) or two (double locus variants, DLV) loci. eBURST is based on a model of bacterial evolution, whereby a single ancestral founding genotype undergoes diversification to produce a subset of closely related genotypes [83]. A single spot on the eBURST diagram (Fig. 12) represents each individual MLVA type, and the size of the spot is proportional to the number of isolates in the population that share that type. eBURST analysis based on the allelic variants of each VNTR locus

demonstrated two unlinked clonal complexes among the 33 MLVA types found in our study (Fig 12). A large majority of our isolates were related to each other, since they formed one of the clonal complexes, while the remaining three types formed another clonal complex. Out of 127 isolates, just 24 were outliers, which could not be traced back to the founder via single locus variants.



**Figure 12.** Phylogenetic relationship by eBURST. Each spot represents a specific MLVA type. The predicted founder is Type 6. The size of the spot is proportional to the number of isolates in the population that share that MLVA type.

In this study we evaluated MLVA for usefulness in practical epidemiology. MLVA types were compared to detailed epidemiological information, including geomapping of place of exposure. For this purpose the geographical information system (GIS) was used. The results indicate that subtyping might be of limited use in surveillance and outbreak investigations of tularaemia on the national level in endemic countries like Sweden.

Genetic diversity functions as a way for populations to adapt to a changing environment. If there is more variation, it is more likely that certain individuals in a population will have variations of alleles that are suited for a particular environment. Those individuals are more likely to survive and produce offspring carrying that allele. The population will continue to grow for further generations due to the survival of these individuals. VNTR data in this study demonstrated the highest genetic diversity at 0.90 for ft-M3 and the lowest at 0.15 for ft-M20. With an increasing number of laboratories using MLVA, there is a need for standardization of allele nomenclature, so that data can be interpreted and shared effectively between laboratories, Lista et al. have reported on the discrepancies between the relative allele fragment size, as determined by

fragment analysis software for MLVA markers, and the actual sizes determined with complete genome sequencing and direct sequencing of PCR products. This will be a reality if different analysis software, instruments, size standards or fluorescent dyes are used, and the risk is increased when the alleles of the loci contain short repeats, e.g. 2-3 nucleotides [84]. This problem can be overcome by analyzing a set of reference strains of known genotypes and allele sizes or by sequencing the alleles. It is very important that the data can be reliably compared in each laboratory.

## 5 CONCLUDING REMARKS

The last years have been momentous for infectious disease trackers and public health professionals around the globe. The spread of cholera in Haiti following a catastrophic earthquake, the worldwide H1N1 influenza pandemic, the EHEC O104 outbreak and the reemergence of polio in Tajikistan, when we thought this disease had nearly vanished – these are but a few of the challenges we have faced. At the same time, the diagnostic and typing methods for different pathogens have become more sensitive and rapid, which improves the possibility to defeat infections. Neither *B. anthracis* nor *F. tularensis* constitute a part of the human normal bacterial flora, and their presence in clinical specimens therefore indicates infection, regardless of the quantity of the agent. They are easy to disseminate and they have potentially high mortality rates after exposure by inhalation, which is why they are classified as category A bioterrorism agents according to the Center for Disease Control and Prevention (CDC). On the other hand, there is a distinct requirement for a wider range of applications for detection and identification of these two bacteria to be used by veterinary public health services under field conditions. They are zoonotic bacteria, and the first indication of their intentional release might well arise as a disease outbreak among animals. Rapid and accurate assays for microbial identification are essential to ensure proper medical intervention in the case of a suspected intentional release of these two agents. One such protocol was established here in paper II and is based on novel application of DNA sequencing. The analysis, including amplification of target genes, is performed in less than three hours.

Another approach that has been applied on two different pathogens is based on pyrosequencing. Pyrosequencing generates short sequences, and is thus a suitable method for analyzing SNPs within a short gene fragment. As applied in this thesis (papers I and III) it proved to be very useful for the analysis of the *Francisella* 16S rDNA gene for determination of species and subspecies. Likewise, pyrosequencing of the *rpoB* gene was useful for discrimination of *B. anthracis* from closely related bacillus species. The analysis, including amplification of target genes, is performed within two hours.

Protocols for the high resolution molecular typing methods available today are more time consuming than the methods described above. Both MLVA and PFGE, which have been evaluated for molecular epidemiological typing of *F. tularensis*, are very accurate methods for many pathogens. However, for rapid typing they have to be replaced by more rapid and less labour intensive methods.

MLVA studies have shown that *F. tularensis* subspecies *holarctica* isolates from the United States and Sweden grouped together in several of major clades, suggesting a more recent common ancestor for the European and North American *F. tularensis* populations. The minor genetic diversity of subspecies *holarctica* suggests that this subspecies has recently expanded across the Northern hemisphere. Future typing should be based upon methods that result in distinct character data, easily exchanged between laboratories. Investigation of the molecular epidemiology of *F. tularensis* will provide deeper insights into the patterns of spread and be helpful in the establishment new methods and strategies for disease prevention. Tularemia has previously been considered to be a disease of the north (“Norrlandssjuka”) in Sweden. However,



according to the Swedish National Surveillance System, numerous cases occur in central Sweden. This change in the geographical pattern of tularaemia could possibly be attributed to a southward movement of the mosquito vector, but many questions remain unanswered and more research is needed in this area.

## 6 ACKNOWLEDGEMENTS

A PhD thesis is never built up by one person alone, it lays on the foundations of good mentorship and interesting discussions over time. If you are holding this thesis in your hand you have most likely been involved or crossed my scientific road in one way or another. Thank you.

Sven Löfdahl and Lars Engstrand thanks for being my supervisors and teaching me science, Sven you came up with a very good suggestion for how to improve my writing and thanks for teaching me patience, My co-supervisor Professor Lars Engstrand, thank you for believing in me from the beginning and giving me the first project and for all support especially during the last year. Without you two this thesis would never be a reality.

Thanks to my former and present leaders: Anders Tegnell, Johan Struwe, Gunnar Sandström, Ralf Wollin, Åke Lundqvist, Daniel Palm, Öjar Melfors and Andreas Bråve. Andreas I am so happy to have you as a leader, your interest, ambition and appreciation make me work harder and enjoy it even when it is stressful, and hope that there will be less stress in the future.

I want to thank all the co-authors and other people that have contributed to the papers, staff and friends at SMI for the great company in the lunch-room and chats in the morning in the changing room, the chats in the corridors.....

Thanks to Johan Struwe & Kerstin Mannerqvist for being my mentor and a very good friend. Kerstin thanks for our stays on Dalarö. Tuija and Melles for always making sure we were safe in the P3. Many thanks to Monica for arranging everything; you are not just a colleague but even a mum for me. Thanks Andreas Heddini and Karin Tegmark for being my doctor and always give good advice on how to recover from disease and sport injuries without medication. Thanks Andreas for good discussions while you were my roommate, and of course you are always right. Special thanks to all my other wonderful roommates during these years at SMI, Mohammed for all IT-help.

Many thanks to KCB folks, starting with my fadder START girl Marie I miss you, Linn, Anna, Sandra, Benjamin, Melinda ( I miss your lovely cakes, since you finished here I just get flowers and candle from Talar and Rebaz for my birthdays, nobody gives me cakes, please come back), Ingela (it is lovely to work with you, nothing is impossible, by trial and error everything is solved in one way or another), Kicki for good discussions about raising boys, Malin S., Shawon, Karin S. (I love your energy), Jonas K., Cecilia, Elba, Jonas H., Ida, Linda T., Ann-Marie, Åsa B., Ewa, Sandor, Hadi, Amir, Henrik, Zhe Hu, Elsie, Linda B., Elenor, Kjell-Olof, Ann Andersson, Helen, Gunnel, Ali, Anette K., Vittja, Maria Wallström, Jolle, Kerstin Falk, Anna-Lena, Sirkka (for English reviewing this), Angerd, Anna P., Mehrdad, Carole. TB group (Pontus, Solomon, Ramona, Alex, Juan Carlos, Anna, Jolanta, Jim, Lisbeth K., Sven), Britt-Marie (I miss you on 2-floor), Matilda. Parasitology folks Jessica, Silvia, Lisbet G., Johan Lindh, Görel, Anette, Caroline, Sabina, Jonas, Elisabet H., Anneli, Marianne Lebbad for being such a good friend and reading my second job. And of course Margareta L., Britta, Cia, Gloria, Sara B., Maria Axelsson, Reza, Marika. Elisabeth S. (for Palasso help), Alma, Sara H., Lena G, Maj and Ann (thank you for lending me so many plates during these years), Mattias Mild, Steve Glavas I miss you so much and hope that our scientific road cross again, Benjamin Turner for a good collaboration, and finally to SMIs queen my Talar (I am not missing my sister Talar when I have you around), and many thanks for taking care of the diagnostic while I was writing, I will

make it up to you after this, I promise and you know I usually keep my promises. Thank you so much for concerning so much not just for me but even for my family.

Thank you Sven Britton for revealing Ethiopia to me, and to all my co-Ethiopian friends: Charlotta, Melissa, Sayma, Kaj, Biborka, Linda, Patricia and my lovely Salma (you changed my view about all Iranian folks with your good heart) as well as Girma and Yitzak for a wonderful trip and of course where we were warmly welcomed and got to discover this beautiful country with its wonderful peoples.

Many thanks to all EQDeBa members for such a nice and friendly atmosphere when we meet and for your generosity to share information and protocols with us. Special thanks to Pietro, Sandra, Mandy and Antonio for making my and Dylan's trip to Italy an unforgettable episode and to Antonios family for taking such good care of Dylan while we were in the field for four days. "All Italy just 5 EUR". And of course a lot thanks to you Daniela, Urshi and Sabina for your friendship. A lot of thanks to FBD members especially to Johanna, Kerstin, Catarina, Anna-Lena and Carin for a nice meeting and good project.

Roger Karlsson and Uno Lindberg for giving me opportunity to work in your lab at Stockholms University. Many thanks to professor Richard W. Padgett for giving me the opportunity to work in your lab at Waksman Institute, Rutgers University, and to Lisa Maduza for helping and learning me all about *C. elegans* and all the other members of a group. It was there I discovered that being in the lab, having lab coat on and pipette in your hand is the best way to kill homesickness and other private problems. Thanks to Hong-Yan, Ulf, Pia, Anne, Bo, Olga, Alistair and Mao for a good collaboration and for having so much fun together at Gene-Tag Neuromics.

Many thanks to the Friskis crew at the reception Susanne, Sandra, Helene, Anna, Gunilla and of course the best leaders Jonas S., Emma J., Sussane J., Rose-Marie A., Anette N., Camilla R., Katarina T. and all other leaders and last Abram Katz (I miss your funny stories), for giving so much energy and making one feel so good after having been there, a perfect place to go to when you feel down because you will always get happy by stepping in there.

Many thanks to my freinds outside the lab for making life more enjoyable Kazaw, Hero&Akram, Paywast&Khalaf, Dlgir, Gllara&Farhad, Farhad&Sawza, Rezan&Ashti, Aram&Bahra, Khadga&Hussien, Taha&Nasrin, mama Mansor, Caroline&Omer (I miss your company, we should find time to spend together again), and Maja (I miss our cinemas and dinners, pity that you moved to Australia), Jwan Ismael, Avin Shawkat, Sara, Nian, Sairan, Hawler group Famida&Addulkadir, Awat&Saedi, Jwan Mhedin, Shko, Darya, Frial, Shler, Sawza, Dlshad, Shawkat and Aras, London folks Dunia&Siwail, Rzgar, Asda, Natali, Kwestan&Jwanro, Niga. Many thanks to Sten for making life much easier and repairing everything between the sky and earth for me; you are the most helpful janitor in the world, my neighbours Erica, Rizgar and Ledy. Thanks to all my wonderful relatives, all my lovely uncles (especially for Omer&Suad and mama Mustafa, your encouragement meant a lot when I was young) and their families and all my cousins including Saman Nagm and of course to you pora Fatma and kaka Ismail gian.

Many many thanks to my family the most lovable and wonderful family in the world. **Dad** you are definitely the best, thank you so much for always being there, never saying no to our requests, giving us so much love, and for making me always feel that I was so special that made

other jealous. **Mum** for having such a big and good heart and for giving so much to everybody without expecting anything in return, please take care a bit more of yourself for our sakes so that we don't get so much worry for your health, **Pasha** for being such a caring adoptive father to Dylan, for your ability to make me in a good mood just by hearing your voice and for always having a solution to all my problems, you are a real Pasha. **Rebin** for being you and always standing up and making me know that I can just tell you what I need and know that by the end of the day I have it regardless what I need. **Talar&Fouad** for just being there and fulfilling my place when I am away. Talar dlakam you are my angel. **Bahar&Alom** for taking care of Dylan so often and so long especially when I was in Africa, you even fixed school for him and he was joining yours and my lovely Arin and Dileen in school, and special thanks to you Bahar gian for being so wise and always bringing me back to the road when I get lost, I am privileged to have you so close to me. Sorry that sometimes it takes a long time to realize that you are right and I know you get frustrated but you are stubborn and don't give up and I appreciate that. **Rebaz** there are no words for you to be thanked because your contribution to me for doing this thesis was so valuable. How can I ever be able to thank you and for what. There are many things, because you are always there for me when I cannot fulfil my duties you do it directly for me without asking why, do you remember when you carried Dylan and swung him for 45 minutes when he was 10 years old, heavy and had itch and pain because of chickenpox and could not sleep. For all the funny time you spend with Dylan and taking him to ice-hockey matches and cinemas, hanker after and mathematic teaching, Dylan is a genius in math guess why? And for taking care of him so often while I am away at conferences, meetings or having fun with my colleges and friends. I am truly very sorry for taking so much time from you which you don't have. Many thanks for the new members in our family **Avin** Bahadin (dada Afin gian) and my flower **Hadil** gol you both are amorous, I once said to my brothers that they will be alone in their lives because the girls who were worth them were not born yet, but what I didn't know was that there were just two very lovely girls there and I'm so happy that they could find you. And last many thanks for all the kids **Baban, Bahast, Arin, Anja** and **Dileen** (my bilbilaichaw) and the little two **Nalin** Nonogian and **Luin** (koka chaw tre rashka) gian for being so lovely kids and making us so proud, hugging you make me fly with joy. Many many thanks goes to my son and sun of my life **Dylan** for giving a meaning to my life and making all my cold days in Sweden a warm summer day like home. When you say good morning my sun goes up and down when I get my hug from you and you say good night, thanks for making me so proud so often. I really mean it when I tell you please take away the moon and stars because I am so tall now of proud they hit my head and I know you think that I am nagging a lot about how proud I am and how much I love you but it is not my fault I have the most wonderful son in the world. You are a miracle and sometimes I wonder if I am worth having such a nice boy like you. You are definitely my best friend and thank you for giving me such a meaningful life I just love you so..... much. Do I need to write it down how much I love you or by now you know my light? It feels like I inherited from my mum to have such a nice boy and hope that you will get as nice kids as me and my mum.

If I by any chance have forgotten to mention you, please forgive me and I want you to know that I am truly sorry for that but I am too old and the memory unfortunately to short.

## 7 REFERENCES

1. Crick, F.H., J.S. Griffith, and L.E. Orgel, *Codes without Commas*. Proc Natl Acad Sci U S A, 1957. **43**(5): p. 416-21.
2. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA*. Proc Natl Acad Sci U S A, 1977. **74**(2): p. 560-4.
3. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
4. Anderson, S., *Shotgun DNA sequencing using cloned DNase I-generated fragments*. Nucleic Acids Res, 1981. **9**(13): p. 3015-27.
5. Achtman, M., *Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens*. Annu Rev Microbiol, 2008. **62**: p. 53-70.
6. Mathema, B., et al., *Lessons from molecular epidemiology and comparative genomics*. Semin Respir Crit Care Med, 2008. **29**(5): p. 467-80.
7. Manning, S.D., et al., *Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks*. Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4868-73.
8. Achtman, M., et al., *Microevolution and history of the plague bacillus, Yersinia pestis*. Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17837-42.
9. Hunter, P.R. and M.A. Gaston, *Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity*. J Clin Microbiol, 1988. **26**(11): p. 2465-6.
10. van Belkum, A., et al., *Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology*. Clin Microbiol Rev, 2001. **14**(3): p. 547-60.
11. Fraser, C., W.P. Hanage, and B.G. Spratt, *Recombination and the nature of bacterial speciation*. Science, 2007. **315**(5811): p. 476-80.
12. Gal-Mor, O. and B.B. Finlay, *Pathogenicity islands: a molecular toolbox for bacterial virulence*. Cell Microbiol, 2006. **8**(11): p. 1707-19.
13. Smith, J.M., et al., *How clonal are bacteria?* Proc Natl Acad Sci U S A, 1993. **90**(10): p. 4384-8.
14. Levinson, G. and G.A. Gutman, *Slipped-strand mispairing: a major mechanism for DNA sequence evolution*. Mol Biol Evol, 1987. **4**(3): p. 203-21.
15. Vossen, R.H., et al., *High-resolution melting analysis (HRMA): more than just sequence variant screening*. Hum Mutat, 2009. **30**(6): p. 860-6.
16. Keim, P., et al., *Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis*. J Bacteriol, 2000. **182**(10): p. 2928-36.
17. Whatmore, A.M., et al., *Identification and characterization of variable-number tandem-repeat markers for typing of Brucella spp.* J Clin Microbiol, 2006. **44**(6): p. 1982-93.
18. Johansson, A., et al., *Worldwide genetic relationships among Francisella tularensis isolates determined by multiple-locus variable-number tandem repeat analysis*. J Bacteriol, 2004. **186**(17): p. 5808-18.
19. Mazars, E., et al., *High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology*. Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1901-6.
20. Klevytska, A.M., et al., *Identification and characterization of variable-number tandem repeats in the Yersinia pestis genome*. J Clin Microbiol, 2001. **39**(9): p. 3179-85.

21. Mullis, K.B. and F.A. Faloona, *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction*. Methods Enzymol, 1987. **155**: p. 335-50.
22. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*. Science, 1988. **239**(4839): p. 487-91.
23. Higuchi, R., et al., *Kinetic PCR analysis: real-time monitoring of DNA amplification reactions*. Biotechnology (N Y), 1993. **11**(9): p. 1026-30.
24. Ronaghi, M., M. Uhlen, and P. Nyren, *A sequencing method based on real-time pyrophosphate*. Science, 1998. **281**(5375): p. 363, 365.
25. Wahab, T., et al., *Pyrosequencing Bacillus anthracis*. Emerg Infect Dis, 2005. **11**(10): p. 1527-31.
26. Ahmadian, A., et al., *Single-nucleotide polymorphism analysis by pyrosequencing*. Anal Biochem, 2000. **280**(1): p. 103-10.
27. Alderborn, A., A. Kristofferson, and U. Hammerling, *Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing*. Genome Res, 2000. **10**(8): p. 1249-58.
28. Gustafsson, A.C., et al., *HPV-related cancer susceptibility and p53 codon 72 polymorphism*. Acta Derm Venereol, 2001. **81**(2): p. 125-9.
29. Gustafsson, A.C., et al., *Screening and scanning of single nucleotide polymorphisms in the pig melanocortin 1 receptor gene (MC1R) by pyrosequencing*. Anim Biotechnol, 2001. **12**(2): p. 145-53.
30. Ahmadian, A., et al., *Analysis of the p53 tumor suppressor gene by pyrosequencing*. Biotechniques, 2000. **28**(1): p. 140-4, 146-7.
31. Garcia, C.A., et al., *Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene*. Gene, 2000. **253**(2): p. 249-57.
32. Nygren, M., et al., *Quantification of HIV-1 using multiple quantitative polymerase chain reaction standards and bioluminometric detection*. Anal Biochem, 2001. **288**(1): p. 28-38.
33. O'Meara, D., et al., *Monitoring resistance to human immunodeficiency virus type 1 protease inhibitors by pyrosequencing*. J Clin Microbiol, 2001. **39**(2): p. 464-73.
34. Jonasson, J., M. Olofsson, and H.J. Monstein, *Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments*. APMIS, 2002. **110**(3): p. 263-72.
35. Innings, A., et al., *Identification of 43 Streptococcus species by pyrosequencing analysis of the rnpB gene*. J Clin Microbiol, 2005. **43**(12): p. 5983-91.
36. Jordan, J.A., J. Jones-Laughner, and M.B. Durso, *Utility of pyrosequencing in identifying bacteria directly from positive blood culture bottles*. J Clin Microbiol, 2009. **47**(2): p. 368-72.
37. Clarke, S.C., *Pyrosequencing: nucleotide sequencing technology with bacterial genotyping applications*. Expert Rev Mol Diagn, 2005. **5**(6): p. 947-53.
38. Tenover, F.C., et al., *Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing*. J Clin Microbiol, 1995. **33**(9): p. 2233-9.
39. Pang, J.C., et al., *A pulsed field gel electrophoresis (PFGE) study that suggests a major world-wide clone of Salmonella enterica serovar Enteritidis*. Int J Food Microbiol, 2007. **116**(3): p. 305-12.
40. Schwartz, D.C. and C.R. Cantor, *Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis*. Cell, 1984. **37**(1): p. 67-75.
41. Noller, A.C., et al., *Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic Escherichia coli O157:H7 isolates*. J Clin Microbiol, 2003. **41**(12): p. 5389-97.

42. Farlow, J., et al., *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *J Clin Microbiol*, 2001. **39**(9): p. 3186-92.
43. Lindstedt, B.A., et al., *DNA fingerprinting of Salmonella enterica subsp. enterica serovar typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci*. *J Clin Microbiol*, 2003. **41**(4): p. 1469-79.
44. Mikesell, P., et al., *Evidence for plasmid-mediated toxin production in Bacillus anthracis*. *Infect Immun*, 1983. **39**(1): p. 371-6.
45. Uchida, I., et al., *Association of the encapsulation of Bacillus anthracis with a 60 megadalton plasmid*. *J Gen Microbiol*, 1985. **131**(2): p. 363-7.
46. Bragg, T.S. and D.L. Robertson, *Nucleotide sequence and analysis of the lethal factor gene (lef) from Bacillus anthracis*. *Gene*, 1989. **81**(1): p. 45-54.
47. Robertson, D.L., M.T. Tippetts, and S.H. Leppla, *Nucleotide sequence of the Bacillus anthracis edema factor gene (cya): a calmodulin-dependent adenylate cyclase*. *Gene*, 1988. **73**(2): p. 363-71.
48. Welkos, S.L., et al., *Sequence and analysis of the DNA encoding protective antigen of Bacillus anthracis*. *Gene*, 1988. **69**(2): p. 287-300.
49. Makino, S., et al., *Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in Bacillus anthracis*. *J Bacteriol*, 1989. **171**(2): p. 722-30.
50. Makino, S.I., et al., *Direct detection of Bacillus anthracis DNA in animals by polymerase chain reaction*. *J Clin Microbiol*, 1993. **31**(3): p. 547-51.
51. Ramisse, V., et al., *Identification and characterization of Bacillus anthracis by multiplex PCR analysis of sequences on plasmids pXO1 and pXO2 and chromosomal DNA*. *FEMS Microbiol Lett*, 1996. **145**(1): p. 9-16.
52. Jernigan, J.A., et al., *Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States*. *Emerg Infect Dis*, 2001. **7**(6): p. 933-44.
53. Farlow, J., et al., *Francisella tularensis in the United States*. *Emerg Infect Dis*, 2005. **11**(12): p. 1835-41.
54. Staples, J.E., et al., *Epidemiologic and molecular analysis of human tularemia, United States, 1964-2004*. *Emerg Infect Dis*, 2006. **12**(7): p. 1113-8.
55. Escobedo, A.A. and S. Cimerman, *Giardiasis: a pharmacotherapy review*. *Expert Opin Pharmacother*, 2007. **8**(12): p. 1885-902.
56. Sedinova, J., et al., *Use of random amplified polymorphic DNA (RAPD) analysis for the identification of Giardia intestinalis subtypes and phylogenetic tree construction*. *J Eukaryot Microbiol*, 2003. **50**(3): p. 198-203.
57. Marfurt, J., et al., *Identification and differentiation of Leishmania species in clinical samples by PCR amplification of the miniexon sequence and subsequent restriction fragment length polymorphism analysis*. *J Clin Microbiol*, 2003. **41**(7): p. 3147-53.
58. Xiao, L. and R. Fayer, *Molecular characterisation of species and genotypes of Cryptosporidium and Giardia and assessment of zoonotic transmission*. *Int J Parasitol*, 2008. **38**(11): p. 1239-55.
59. Ramisse, V., et al., *The Ba813 chromosomal DNA sequence effectively traces the whole Bacillus anthracis community*. *J Appl Microbiol*, 1999. **87**(2): p. 224-8.
60. Hjalmarsson, S., et al., *Rapid combined characterization of microorganism and host genotypes using a single technology*. *Helicobacter*, 2004. **9**(2): p. 138-45.
61. Sinclair, A., C. Arnold, and N. Woodford, *Rapid detection and estimation by pyrosequencing of 23S rRNA genes with a single nucleotide polymorphism*

- conferring linezolid resistance in Enterococci*. Antimicrob Agents Chemother, 2003. **47**(11): p. 3620-2.
62. Qi, Y., et al., *Utilization of the rpoB gene as a specific chromosomal marker for real-time PCR detection of Bacillus anthracis*. Appl Environ Microbiol, 2001. **67**(8): p. 3720-7.
  63. Klee, S.R., et al., *Evaluation of different methods to discriminate Bacillus anthracis from other bacteria of the Bacillus cereus group*. J Appl Microbiol, 2006. **100**(4): p. 673-81.
  64. Mollet, C., M. Drancourt, and D. Raoult, *rpoB sequence analysis as a novel basis for bacterial identification*. Mol Microbiol, 1997. **26**(5): p. 1005-11.
  65. Keim, P., et al., *Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales*. Infect Genet Evol, 2004. **4**(3): p. 205-13.
  66. Volokhov, D., et al., *Identification of Bacillus anthracis by multiprobe microarray hybridization*. Diagn Microbiol Infect Dis, 2004. **49**(3): p. 163-71.
  67. Thompson, R.C., *Giardiasis as a re-emerging infectious disease and its zoonotic potential*. Int J Parasitol, 2000. **30**(12-13): p. 1259-67.
  68. Lasek-Nesselquist, E., D.M. Welch, and M.L. Sogin, *The identification of a new Giardia duodenalis assemblage in marine vertebrates and a preliminary analysis of G. duodenalis population biology in marine systems*. Int J Parasitol, 2010. **40**(9): p. 1063-74.
  69. Monis, P.T., et al., *Molecular systematics of the parasitic protozoan Giardia intestinalis*. Mol Biol Evol, 1999. **16**(9): p. 1135-44.
  70. Thompson, R.C., R.M. Hopkins, and W.L. Homan, *Nomenclature and genetic groupings of Giardia infecting mammals*. Parasitol Today, 2000. **16**(5): p. 210-3.
  71. Homan, W.L. and T.G. Mank, *Human giardiasis: genotype linked differences in clinical symptomatology*. Int J Parasitol, 2001. **31**(8): p. 822-6.
  72. Levy, D.A., et al., *Surveillance for waterborne-disease outbreaks--United States, 1995-1996*. MMWR CDC Surveill Summ, 1998. **47**(5): p. 1-34.
  73. Burke, J.A., *The clinical and laboratory diagnosis of giardiasis*. CRC Crit Rev Clin Lab Sci, 1977. **7**(4): p. 373-91.
  74. Cartwright, C.P., *Utility of multiple-stool-specimen ova and parasite examinations in a high-prevalence setting*. J Clin Microbiol, 1999. **37**(8): p. 2408-11.
  75. Danciger, M. and M. Lopez, *Numbers of Giardia in the feces of infected children*. Am J Trop Med Hyg, 1975. **24**(2): p. 237-42.
  76. Verweij, J.J., et al., *Real-time PCR for the detection of Giardia lamblia*. Mol Cell Probes, 2003. **17**(5): p. 223-5.
  77. Guy, R.A., C. Xiao, and P.A. Horgen, *Real-time PCR assay for detection and genotype differentiation of Giardia lamblia in stool specimens*. J Clin Microbiol, 2004. **42**(7): p. 3317-20.
  78. Lalle, M., et al., *Genetic heterogeneity at the beta-giardin locus among human and animal isolates of Giardia duodenalis and identification of potentially zoonotic subgenotypes*. Int J Parasitol, 2005. **35**(2): p. 207-13.
  79. Lebbad, M., et al., *Dominance of Giardia assemblage B in Leon, Nicaragua*. Acta Trop, 2008. **106**(1): p. 44-53.
  80. Hopla, C.E., *The ecology of tularemia*. Adv Vet Sci Comp Med, 1974. **18**(0): p. 25-53.
  81. Gray, M.W., D. Sankoff, and R.J. Cedergren, *On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved*



- structural core in small subunit ribosomal RNA*. Nucleic Acids Res, 1984. **12**(14): p. 5837-52.
82. Spratt, B.G., et al., *Displaying the relatedness among isolates of bacterial species -- the eBURST approach*. FEMS Microbiol Lett, 2004. **241**(2): p. 129-34.
  83. Feil, E.J. and M.C. Enright, *Analyses of clonality and the evolution of bacterial pathogens*. Curr Opin Microbiol, 2004. **7**(3): p. 308-13.
  84. Lista, F., et al., *Genotyping of Bacillus anthracis strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis*. BMC Microbiol, 2006. **6**: p. 33.