Sonia Soares Appel

Licenciada em Química Aplicada – Biotecnologia

Testimonial of Professional Contributions in the Biotechnology Field (1991-2019)

Pulmonary Infectious Agents *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*

Relatório elaborado nos termos do Despacho n.º 20/2010 para obtenção

do Grau de Mestre em Biotecnologia por Licenciados "Pré-Bolonha"

Orientador

Professor Doutor Carlos Alberto Gomes Salgueiro

Departamento de Química

Júri

Presidente: **Professora Doutora Susana Filipe Barreiros**

Arguente: **Professor Doutor Adriano José Alves de Oliveira Henriques**

Vogal: **Professor Doutor Carlos Alberto Gomes Salgueiro**

Dezembro 2020

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Testimonial of professional contributions in the Biotechnology Field (1991-2019) - Pulmonary Infectious Agents *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*.

Copyright © Sonia Cristina da Veiga Santos Soares Appel, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Dedication

I dedicate this journey to my daughter, Bli.

My father taught me, through his own choices in life, to work honestly for what you want, never stop learning, and never give up.

May we be able to pass it along to future generations.

Enquanto houver estrada p'ra andar

A gente vai continuar...

Enquanto houver ventos e mar

A gente não vai parar...

Jorge Palma

Dedication

Acknowledgements

First and foremost, I am very thankful to Professor **Carlos Salgueiro**, for direction, detailed reading, and especially his words of encouragement during this journey.

I thank Professor **Susana Barreiros**, President of the Chemistry Department, and **Faculdade de Ciências e Tecnologia (FCT)** for the acceptance into this program, and the wonderful opportunity to be their student again. Thank you to **Beatriz do Bem** for the reliable administrative support all along.

I am deeply grateful to my career life mentors **Jeanie Dubnau** and **Joan Haab**. Jeanie was my manager at the PHRI and the best teacher I ever had - an amazing example of a devoted scientist, mother, wife, activist, community leader and friend. Her scientific guidance and emotional support were invaluable to me, and she remains a mentor to this day. Joan was my director at Cepheid; her leadership taught me the kind of leader I wanted to be and changed me forever. I thank Professor **Alexander Tomasz** for the opportunity to work with him at the Rockefeller University. Huge thanks to **Issar Smith** (Smithy) and **Dave Dubnau** for the amazing direction and support at the PHRI – their standards set the tone to multiple generations of young scientists.

I am eternally grateful to my dear friend **Alexandra Sardinha Bernardo**, for her interest in this project, brainstorming sessions on structure and a critical review of the manuscript. I am very thankful to my friend **Kira Foygel** for scientific advice at the beginning of this process. To my sister in life **Adrienne Ho,** and my friends **Maria João Assunção Morgado** and **Catarina Duarte**, a deep thank you for valuable discussions and advice.

To my family, I owe everything. My daughter **Bli**, who made everything in my life make sense, who is growing her wings and blooming into an independent young woman, and whose love and support allowed me to peacefully work through this project. My mother **Aida** has always been there for me through the ups and downs of life; she took me in during the pandemic while I was writing this report - I camped in her living room and she nourished my body and my soul. My father **Oscar**, whose unconditional love, support, and moral compass cross dimensions and boundaries of the universe and guide me every day. My sisters **Catarina** and **Susana**, who helped with the Portuguese language sections, are my best friends, my rocks, my life-rafts; the little time we spend together is the best in the world. I am especially grateful to my bother-in-law **Pedro** for amazing career discussions; his passion for continuing to learn is a life-long inspiration, and without him this journey would not have started.

And most of all, to my darling husband **Jens**, whose love, patience, and infinite support have made this work not only possible, but enjoyable. With you, everything is delightful. Sharing my life with you has been a privilege with no matching words; I love you, and I feel eternally grateful for you every single day.

Preface: Overview of my professional path

In 1991, I graduated from Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa with a pre-Bologna degree of Licenciatura in Applied Chemistry – Biotechnology. I completed my pre-graduation internship from 1990 to 1991 in Instituto de Tecnologia Química e Biológica (ITQB), Oeiras, Portugal, under the direction of Dr. Isabel Sá Nogueira, where I sequenced part of the *Bacillus subtilis* L-arabinose operon.

During my career of twenty-nine years, I spent the first seven in academic research in various institutes and universities. I later shifted to roles in biotechnology industries, specifically in the life sciences and medical devices fields, where I have been working for 22 years, currently focusing on the combination pharmaceutical and medical device industry.

Figure A shows my career timeline and details are covered in the following paragraphs.

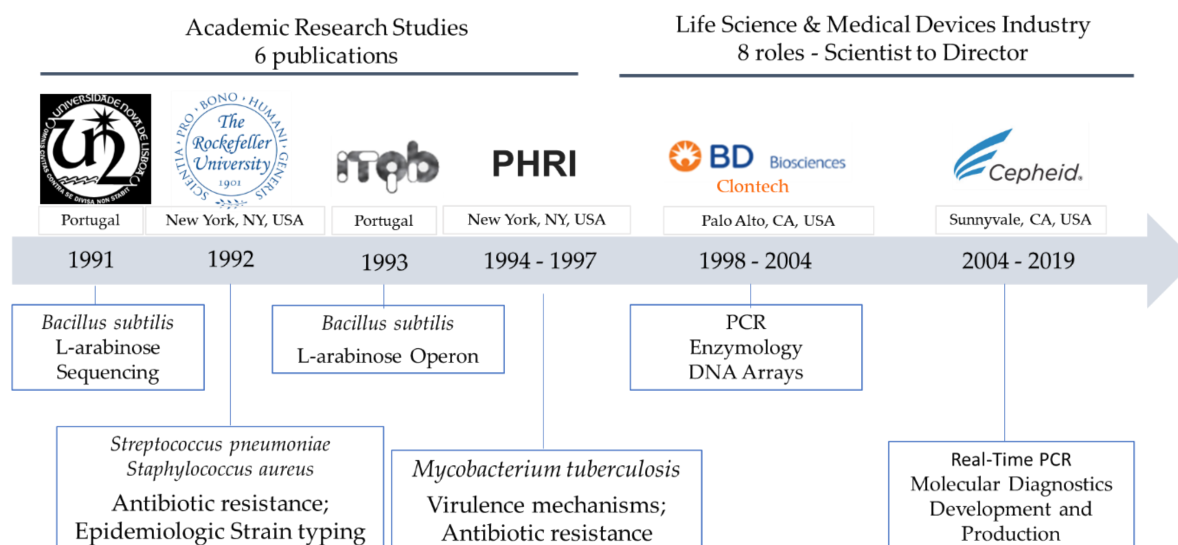

Figure A. Career timeline

In the first seven years of my career, I conducted **Academic Research** studies in the Microbiology field, utilizing mainly Molecular Biology methods to study bacteria. My academic work was included in six scientific publications in peer-reviewed journals (Appendix A - Personal Publications).

From 1991 to 1992, I worked as a Guest Investigator in the Microbiology Laboratory of the **Rockefeller University**, New York, USA, under the direction of Prof. Alexander Tomasz. I performed analysis of penicillin binding proteins of *Streptococcus pneumoniae* and Methicillin Resistant

Staphylococcus aureus (MRSA), and epidemiological strain typing thorough Pulsed Field Gel Electrophoresis. Part of this work is described in Chapter 1.

In 1993 I returned to **ITQB**, Oeiras, Portugal, where I had interned before my graduation – this time as a Ph.D. student under the direction of Dr. Isabel Sá Nogueira, continuing my *Bacillus subtilis* L-arabinose research. The work performed during my internship prior to graduation and during this period is complementary and published together.

From 1994 to 1997, I worked under the direction of Dr. Eugenie Dubnau at the Public Health Research Institute (**PHRI**), New York, USA (at the time associated with New York University and now part of Rutgers University). I was a Research Assistant and studied virulence and antibiotic resistance mechanisms of *Mycobacterium tuberculosis*. This work is reported in Chapter 2.

In the second span of my career, consisting of twenty-two years, I maintained my Molecular Biology focus while switching my activity to roles in the **Biotechnology Industry**. I performed various functions in companies that research, develop and manufacture products for the life science and medical diagnostic fields.

From 1998 to 2004, I performed several roles in the Research and Development and in the Production departments at **Clontech**, Palo Alto, California, USA (acquired by Becton Dickinson Biosciences during my tenure, and now part of TakaraBio). I was a Production Supervisor in the DNA Arrays group, responsible for launching and managing production of the nylon DNA arrays product lines. I later transferred to a Production Scientist role performing investigations in the manufacturing of PCR (Polymerase Chain Reaction) related enzymes. In early 2001, I transferred to a Research Associate role in the R&D Department, where I cloned, modified, and purified novel enzymes, such as DNA-polymerases and Reverse-transcriptases. I also developed and characterized enzymatic systems for PCR, Reverse-transcription, and real-time PCR.

From 2004 to 2019 I worked in the Product Transfer Department at **Cepheid Inc**, Sunnyvale, California, USA. I held a variety of titles from Process Development Scientist to Director of Product Transfer. Under my tenure, Cepheid expanded from 1 to over 25 product lines, each with multiple sub-types. One product line specifically is the XpertMTB Assay line for the detection of *Mycobacterium tuberculosi*s and antibiotic resistance, which public elements are described in Chapter 3.

I opted to focus this report on the detection, virulence, and antibiotic resistance of the pulmonary infectious agents *Streptococcus pneumoniae* and *Mycobacterium tuberculosis,* drawing a unifying line from my work at the Rockefeller University, to the PHRI and to Cepheid, Inc. (Figure B. Unifying line through pulmonary infectious diseases). Some of the work performed by other colleagues is also presented because it is important to justify the conclusions. The work performed by others is identified in the text of the report in each chapter.

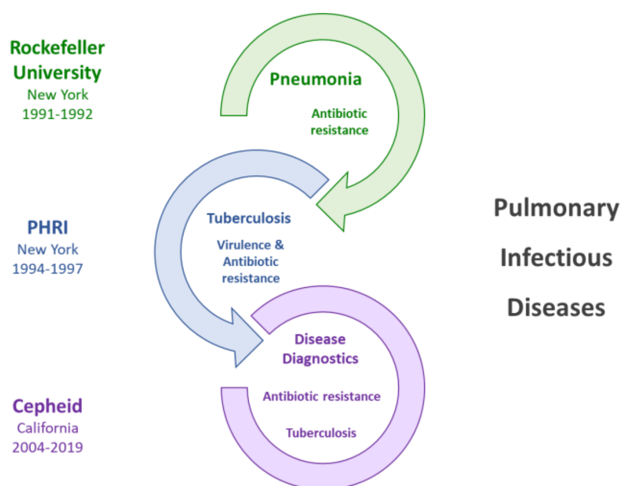

Figure B. Unifying line through pulmonary infectious diseases

Leveraging my experience in the management of development and production projects in the regulated Molecular Diagnostic field, I recently shifted my career to Project Management in the field of Medical Devices/Pharmaceuticals. In January 2020, I completed a Project Manager Professional certification, by the Project Management Institute. I am currently a Project Management Consultant at AstraZeneca, where I lead teams in the development of combination biopharmaceutical / subcutaneous-injection device products, supporting device development and all required verification and validation studies, clinical trials, and transfers to production. The products mainly consist of accessorized pre-filled syringes, auto-injectors and/or pens for home use, which are pre-filled with biopharmaceutical therapeutic products such as monoclonal antibodies, and ultimately will provide methods for safe and effective deliveries of medications to patients.

Abstract

Pulmonary infectious diseases, like pneumonia and tuberculosis, have been pervasive throughout the world for centuries; they kill millions of people annually worldwide and pose an increased threat when associated with antibiotic resistance and viral co-infections such as influenza and, more recently, Covid-19.

This professional activity report summarizes some of my work in the last three decades, using biotechnology to type, characterize or detect two major bacterial infectious agents, *Streptococcus pneumoniae* and *Mycobacterium tuberculosis*.

Strain typing techniques, like Pulsed Field Gel Electrophoresis and Penicillin Binding Protein patterns, were used to identify *Streptococcus pneumoniae* clones from different geographic areas in the early-1990s, showing strong evidence of clonal dissemination locally and across the globe, from Spain to Iceland and to the United States.

In the mid-1990s, an *in vivo* expression technology system based on a promoter-trap was constructed to enable the identification of *Mycobacterium tuberculosis* genes upregulated in human macrophages, through selection for antibiotic resistance. This work led to the identification of eight genes likely relevant for *Mycobacterium tuberculosis* virulence. Several were later confirmed to be required for infection or survival in the host.

A single-step molecular diagnostic real-time PCR assay for the detection of multidrug resistant *Mycobacterium tuberculosis* was commercialized in the late 2000s, with great impact in the detection of tuberculosis, especially in developing countries. The mechanism of the assay is presented here as published by its inventors, as well as with my role in the industry, from the development of processes for production and quality control, to regulation-compliant distribution, troubleshooting and development of improvements.

Keywords: *Streptococcus pneumoniae*, *Mycobacterium tuberculosis,* PFGE, PBP, promoter trap, real-time PCR, molecular diagnostics

Resumo

As doenças infeciosas pulmonares, como a pneumonia e a tuberculose, têm tido, durante séculos uma prevalência mundial significativa, matando milhões de pessoas anualmente e representando uma ameaça crescente quando associadas à resistência a antibióticos ou coinfeções virais, como a gripe e, recentemente, o Covid-19.

Este relatório de atividade profissional resume parte do meu trabalho nas últimas três décadas, usando técnicas de biotecnologia para identificar, caracterizar ou detetar dois importantes agentes infeciosos de origem bacteriana, *Streptococcus pneumoniae* e *Mycobacterium tuberculosis*.

Técnicas de identificação de estirpes, como eletroforese tipo Pulsed Field e padrões de proteínas com afinidade à penicilina, foram usadas para identificar clones de *Streptococcus pneumoniae* de diferentes áreas geográficas no início da década de 90, mostrando forte evidência de disseminação, não só a nível local, mas também através do globo, nomeadamente, de Espanha para a Islândia e Estados Unidos.

Em meados da década de 90, um sistema de expressão *in vivo* foi construído para permitir a identificação de genes de *Mycobacterium tuberculosis* com expressão elevada em macrófagos humanos, através de seleção por resistência a antibióticos. Este trabalho levou à identificação de oito genes com probabilidade de serem relevantes para a virulência de *Mycobacterium tuberculosis*; vários posteriormente confirmados como necessários para a infeção ou sobrevivência no hospedeiro.

Mais recentemente, foi comercializado um teste de real-time PCR para diagnóstico molecular de *Mycobacterium tuberculosis* com multirresistência a antibióticos, com grande impacto na deteção da tuberculose, principalmente em certos países em desenvolvimento. O mecanismo de deteção é apresentado aqui conforme publicado pelos seus inventores, bem como o meu papel na indústria de diagnósticos moleculares, desde o desenvolvimento de processos de produção e controle de qualidade, até a distribuição em conformidade com os respetivos regulamentos, passando por resolução de problemas e desenvolvimento de melhorias.

Palavras-chave: *Streptococcus pneumoniae*, pneumococcus, *Mycobacterium tuberculosis,* Pulsed-Field, PBP, promoter trap, real-time PCR, diagnóstico molecular

Table of Contents

Dedication iii

Acknowledgements v

Preface: Overview of my professional path vii

Abstract xi

Resumo xiii

Table of Contents xv

Abbreviations and Acronyms xvii

Figure Index xix

Table Index xxi

Chapter 1. Molecular Typing of Multi-drug Resistant *Streptococcus pneumoniae* 1

1. Introduction 2
   - 1.1. Relevance of *Streptococcus pneumoniae* 2
   - 1.2. Bacteriology 3
   - 1.3. Virulence factors 4
   - 1.4. Classification of pneumococcus 5
   - 1.5. Genetic diversity 6
   - 1.6. Antibiotic resistance and pneumococcus 6
2. Strain Typing Methods 7
   - 2.1. Multilocus Enzyme Electrophoresis 7
   - 2.2. Penicillin Binding Protein (PBP) patterns 7
   - 2.3. Pulsed-Field Gel Electrophoresis (PFGE) 7
3. Analysis of Serotype 6B Isolates from Iceland 9
   - 3.1. Background 9
   - 3.2. Experiments and results 9
   - 3.3. Summary and analysis 14
4. Serotype 23F Isolates from North Carolina 17
   - 4.1. Background 17
   - 4.2. Experiments and results 17
   - 4.3. Case-study: hypothesized seroconversion 21
5. Conclusions 23
6. Further Developments 24
   - 6.1. Changes in pneumococcal epidemiology 24
   - 6.2. Evolution of typing techniques 25

Chapter 2. Molecular Biology Tools for studies in *Mycobacterium tuberculosis* Virulence 31
1. Introduction 32
1.1. Tuberculosis 32
1.2. History of the tuberculosis disease 32
1.3. Tuberculosis in the world today 34
1.4. The bacillus *Mycobacterium tuberculosis* 35
1.5. The *Mycobacterium tuberculosis* complex 36
2. Tools to Study Mtb Virulence 37
2.1. Background 37
2.2. Construction of an Mtb library in a promoter trap 37
2.3. Verification of performance of controls 39
2.4. Selection of isoniazid resistant Mtb clones during infection of macrophages 40
3. Conclusions 42
4. Further Considerations 43
4.1. Further findings 43
4.2. Gene expression profiling 44
4.3. The Public Health Research Institute 44
Chapter 3. Molecular Diagnostic Assays 47
1. Background 48
2. Introduction 49
2.1. Antibiotic resistance and Mtb 49
2.2. Rifampin resistance and MDR-TB 49
2.3. Real-time PCR assays 50
3. MTB/RIF Assay 52
3.1. Mechanism of detection of Mtb and rifampin resistance 52
3.2. Sample processing and integration into the GeneXpert 52
3.3. Automated result analysis 53
3.4. Assay performance 53
4. Conclusion 54
4.1. My Role in the molecular diagnostics industry 54
4.2. Impact of TB molecular diagnostics in the world 55
General Conclusion 57
References 59
Appendix A - Personal Publications 73

Abbreviations and Acronyms

AIDS	Acquired immune deficiency syndrome
BCG	*Bacillus* Calmette-Guérin
B. globigii	*Bacillus globigii*
bp	Base pairs
Cas	CRISPR-associated protein
CDC	United States Centers for Disease Control and Prevention
CF	Competence factor
CFU	Colony forming units
CRISPR	Clustered regularly interspaced short palindromic repeats
DGH	Distributed-genome hypothesis
E. coli	*Escherichia coli*
ET	Electrophoretic type
HIV	Human immunodeficiency virus
IPD	Invasive Pneumococcal Disease
ITQB	Instituto de Tecnologia Química e Biológica
IVET	*in vivo* expression technology
kb	Kilobases
LOD	Limit of detection
M.	*Mycobacterium* sp.
MDR	Multi-drug resistant
MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MRSA	Methicillin resistant *Staphylococcus aureus*
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
NESp	Non-encapsulated *S. pneumoniae*
NGS	Next-generation sequencing
PBP	Penicillin binding protein
PFGE	Pulsed field gel electrophoresis
PHRI	Public Health Research Institute
PS	Penicillin-susceptible
RFLP	Restriction fragment length polymorphism
RNA-seq	RNA sequencing
RRDR	Rifampicin resistance determining region
RT	Reverse-transcription
S.	*Streptococcus* sp.
SNP	Single nucleotide polymorphism
TA	Toxin-antitoxin
TB	Tuberculosis
UN	United Nations

UV	Ultraviolet
WGS	Whole genome sequencing
WHO	World Health Organization
XDR	Extensively drug-resistant

Figure Index

Figure A. Career timeline vii

Figure B. Unifying line through pulmonary infectious diseases ix

Figure 1.1. Notable scientists in early *Streptococcus* research 2

Figure 1.2. Types of hemolysis 3

Figure 1.3. Optochin disk test 3

Figure 1.4. Microscopical observation of *S. pneumoniae* in patient sputum. 5

Figure 1.5. PBP (Penicillin Binding Protein) profiles of multiresistant isolates 11

Figure 1.6. PFGE patterns of the SmaI fragmented chromosomal DNA 12

Figure 1.7. Prevalent PFGE SmaI pattern in isolates from Iceland, Spain, and England 13

Figure 1.8. Different PFGE SmaI patterns of various isolates. 13

Figure 1.9. PBP pattern of MDR serotype 23F isolates 18

Figure 1.10. PBP pattern of Penicillin-Susceptible isolates 18

Figure 1.11. PFGE of MDR serotype 23F isolates 19

Figure 1.12. PFGE of various isolates 19

Figure 1.13. Schematic timeline of isolate collection from a specific child in the North Carolina daycare 21

Figure 1.14. Example of a Multiplex Real-Time PCR Graph 28

Figure 2.1. Image of the painting The Sick Child 33

Figure 2.2. Graphs showing the prevalence of TB cases in 2018 34

Figure 2.3. Electron micrograph of *Mycobacterium tuberculosis* 35

Figure 2.4. Colony morphology of *Mycobacterium tuberculosis* 35

Figure 2.5. Schematic depiction of the promoter trap constructs 38

Figure 2.6. Growth of Mtb H37RV carrying the promoter trap constructs 40

Figure 3.1. Illustration of the mechanism of Molecular Beacon probes 50

Figure 3.2. Generic overview of an assay production line and lifecycle management 54

Table Index

Table 1.1. Multilocus Enzyme Electrophoresis Allele Types of Isolates by Origin 10

Table 1.2. Epidemiological, phenotypical, and molecular properties of the isolates 15

Table 1.3. Summary of Serotype 23F isolates and PBP/PFGE patterns 20

Table 1.4. Summary of isolates collected from the same child 21

Table 2.1. Genes identified by isoniazid selection in macrophages 41

Chapter 1. Molecular Typing of Multi-drug Resistant *Streptococcus pneumoniae*

1. Introduction

1.1. Relevance of *Streptococcus pneumoniae*

Streptococcus pneumoniae, commonly referred to as pneumococcus, is one of the leading causes of pneumonia. It can also co-infect the host as a complication from influenza, when it is especially dangerous for the elderly, and it is now believed to have caused much of the mortality during the 1918 influenza pandemic [1,2]. According to the World Health Organization, pneumococcus is the most common cause of bacterial pneumonia in children; pneumonia is the single largest infectious cause of death in children worldwide, and in 2017 accounted for 15% of all deaths of children under five years old - over three quarters of a million deaths [3]. Recent reports of superinfection with Covid-19 further increase the concerns [4,5].

Humans are the main reservoir of *S. pneumoniae*; asymptomatic carriage is commonly found in the upper respiratory tract (nasopharynx) of healthy individuals, mainly children. The immunologic mechanism that allows disease to occur in a carrier is not clearly understood, but when a predisposing condition like pulmonary illness exists, immune dysregulation can occur, and potentially lead to infections and diseases such as pneumonia, sepsis, otitis media, conjunctivitis and meningitis [6,7].

S. pneumoniae played a significant role in the history molecular genetics. It was discovered independently by Pasteur and Sternberg in 1881 [2]. In 1928 Frederick Griffith (Figure 1.1) proved that live, harmless *S. pneumoniae* can be transformed into a deadly strain by combination with an extract from heat-killed, virulent *S. pneumoniae* [8], proving that virulence is conferrable. Further, in 1944, Avery, MacLeod, and McCarty (Figure 1.1) showed that the transforming factor in Griffith's experiments was not protein, as popular opinion favored at the time, but DNA [9].

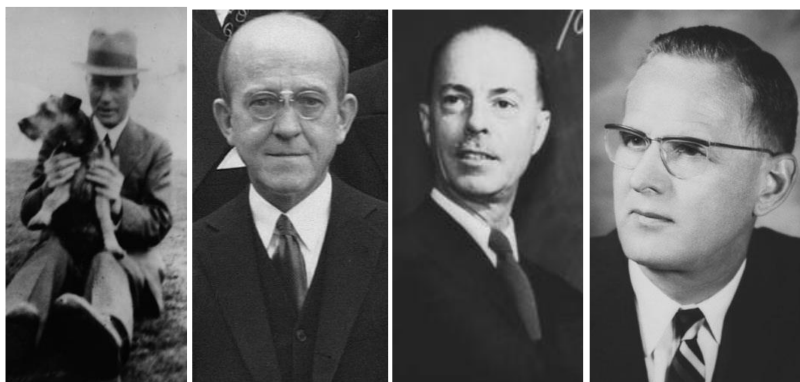

Figure 1.1. Notable scientists in early *Streptococcus* research
Left to right: Frederick Griffith, 1877-1941 [10], Oswald Avery, 1877-1955 [11], Colin MacLeod, 1909-1972 [12] and Maclyn McCarty, 1911-2005 [13].

1.2. Bacteriology

Streptococcus pneumoniae are Gram-positive cocci, often lancet-shaped. They are frequently arranged in pairs (diplococci) but can also be found separately or arranged in short chains. *S. pneumoniae* is a facultative anaerobe; in the laboratory, its growth is best in 5% carbon dioxide and requires a source of catalase (*e.g.,* blood) to neutralize the large amount of hydrogen peroxide produced by the bacteria. In complex media containing blood, the bacterium has a doubling time of 20-30 minutes at 37°C [14]. On blood agar, colonies characteristically produce a green zone due to alpha-hemolysis, which is the partial hemolysis of the red blood cells, as opposed to complete hemolysis, or beta-hemolysis (Figure 1.2). Alpha-hemolysis differentiates *S. pneumoniae* from the Group A (beta-hemolytic) streptococci, but not from *Streptococcus viridans*, another alpha-hemolytic co-inhabitant of the upper respiratory tract. The minimum criteria for traditional identification and distinction of pneumococci from other streptococci are bile solubility or optochin sensitivity (Figure 1.3), Gram-positive staining, and hemolytic activity. [14–16].

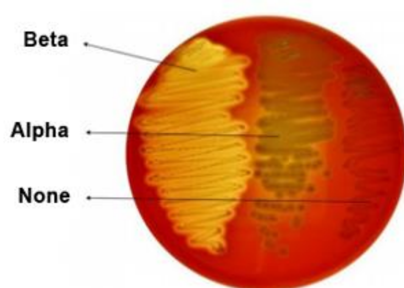

Figure 1.2. Types of hemolysis

The use of a blood agar plate to visually differentiate the extent of hemolysis of the red blood cells. Figure replicated from Garner, Microbiology Nuts & Bolts [17]

Figure 1.3. Optochin disk test

Optochin test for *S. pneumoniae* using optochin disks. Left: *S. mitis*, resistant to optochin with no zone of inhibition; Right: *S. pneumoniae*, susceptible to optochin. Figure replicated from Centers for Disease Control and Prevention [16].

S. pneumoniae possesses autolysin enzymes which disrupt and disintegrate its cells and kill the entire culture when grown to stationary phase. Virtually all clinical isolates of pneumococci harbor this autolysin. In the laboratory, cultures start to undergo lysis between 18 and 24 hours after initiation of growth under optimal conditions. On agar cultures, colonies initially appear with a plateau-type morphology, then start to collapse in the centers when autolysis begins [14]. Autolysins play a major role in pneumococcal virulence [18], which is reviewed below.

1.3. Virulence factors

The major *S. pneumoniae* virulence factors include a polysaccharide capsule that provides protection from the host's immune system, surface proteins with various roles from cell adhesion to metal ion transport mechanisms, autolysins and other lysis proteins with roles in cell growth, and *pili* that enable *S. pneumoniae* to adhere to epithelial cells in the upper respiratory tract [6,18–21].

Pneumococcus cells possess a **capsule of complex polysaccharides** on their outer surface, which acts as a virulence factor for the organism in several ways: by allowing the bacterium to adhere to host cells and cause inflammation/infection, by restricting autolysis and reducing exposure to antibiotics, and by preventing recognition and phagocytosis by the host's immune cells (anti-phagocytosis) [21]. Encapsulated strains of *S. pneumoniae* are found to be 100,000 times more virulent than non-encapsulated strains during invasion of mucosal surfaces [14]. The mechanisms of anti-phagocytosis are not fully understood, but there are studies showing that the capsule may mask antibody recognition of subcapsular antigens localized on the pneumococcal cell wall, as well as prevent the host immunoglobulins that found their way to the pneumococcal cell surface from interacting with their receptors on the surface of phagocytic cells, with the result that the organisms remain extracellular [21,22]. Additionally, the negative charge that the capsule tends to have in many serotypes, may cause electrostatic repulsion, providing a plausible explanation for *S. pneumoniae*'s ability to avoid being trapped by negatively charged mucus layers and phagocytic cells [2,6].

Although the polysaccharide capsule is essential for virulence, and is present in all isolates responsible for invasive pneumococcus disease in otherwise healthy individuals, it was proven recently as non-essential for colonization; non-encapsulated *S. pneumoniae* (NESp) isolates have both been found in up to 19% of asymptomatic carriers, and confirmed to be the cause of non-invasive pneumococcal diseases such as otitis media and infectious conjunctivitis. In rare cases, and mainly in immunodeficient patients, infection with NESp can also lead to invasive pneumococcal disease [23].

In general, **autolysins** degrade the cell wall of bacterial organisms, ultimately leading to cell lysis. These enzymes play roles in a variety of physiological cell functions associated with cell wall growth, cell turnover, and cell separation. The *N*-acetylmuramyl-l-alanine amidase from *S. pneumoniae*, also known as LytA amidase, has been implicated in the pathogenicity of pneumococci by cleaving a specific peptidoglycan bond and breaking the cell wall, leading to cell lysis and to the resulting release of pneumococcal antigens such as pneumolysin, peptidoglycan, and teichoic acids which are all harmful to host cells; these components may help the pneumococci evade the immune system by inhibiting cytokine production and blocking the activation of phagocytes [6].

Pneumolysin (Ply) is a cytoplasmic enzyme that is released due to the action of surface pneumococcal autolysins and has been shown to act as virulence factor through many mechanisms. Pneumolysin binds to membranes containing cholesterol, and forms pores which later lead to host cell lysis. It also plays a role in promoting the formation of biofilms, it reduces mucus clearance of the

bacterium which facilitates the spread of pneumococcal infection, it can interfere with the host's immune system, and can cause DNA damage by dysregulating the production of reactive oxygen species as a result of its pore-forming abilities [6].

Pili are long, thin hair-like structures that extend outside of the polysaccharide capsule. *Pili* aid in the attachment and colonization of the nasopharynx and lungs of hosts by *S. pneumoniae*, and help the bacteria avoid phagocytosis by host immune cells. There are two main types of *pili* found on *S. pneumoniae*: *pilus*-1, which is found in 30% of clinical isolates, and *pilus*-2 which is only found in about 16%. Studies have shown that *pili* are able to stimulate inflammatory responses of the host [6,14,20].

1.4. Classification of pneumococcus

Pneumococcal type-specific antisera were widely used for the treatment of patients in the early 20th century, and pneumococcus responses to the antisera were used to type the isolates, classified as **serotypes**. Although many serotypes were documented and available, antisera were not always effective against all serotypes. The death of Danish Prince Valdemar in 1939, with a novel serotype within serogroup 9 which did not respond to then available antisera, precipitated the development of multiple serotyping studies, simultaneously in Denmark and in America. The Quellung (swelling) reaction, which had been originally developed by Friedrich Neufeld in 1902, was adopted and the Danish system of classification was widely accepted throughout the world [2]. The Quellung reaction is based on an antigen-antibody reaction between antiserum and the highly antigenic polysaccharide capsule, which causes the capsule to appear to swell under microscopical examination (Figure 1.4) [2,14,24].

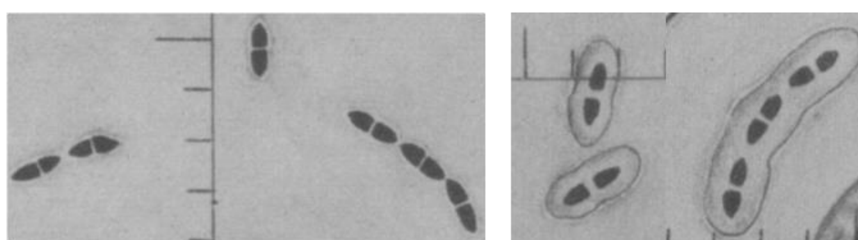

Figure 1.4. Microscopical observation of *S. pneumoniae* in patient sputum.

S. pneumoniae in patient sputum on microscopic slides, mixed with rabbit antiserum, and stained with methylene blue. Type II pneumococcus in sputum is mixed with type I rabbit antiserum on the left panel, and with type II rabbit antiserum on the right panel. The right panel shows the Quellung reaction (swelling of the capsule). Figure replicated from Sabin, 1933 [24].

Over 90 different serotypes of pneumococci have been identified, and they differ in prevalence and extent of drug resistance [6,25]. The virulence of the polysaccharide capsule is enhanced by the pneumococci's ability to undergo recombination; mutations in the capsule polysaccharide synthesis genes (*cps* locus) promote capsular switching, which can generate novel combinations of serotype and genomic backbone, and conferring to pneumococcus the ability to evade the host immune response [25].

1.5. Genetic diversity

Pneumococci are competent and naturally transformable, readily taking up exogenous DNA and incorporating it into the genome by homologous recombination. In laboratory batch cultures, competence typically appears suddenly in most or all cells of the culture at some point during exponential-phase growth, persists for 10 to 20 min, and then decays rapidly. The induction of competence depends on an extracellular protein, named competence factor (CF), which serves as a monitor of population density. The state of competence is accompanied by a gross metabolic shift and a switch in protein synthesis yielding several competence-specific intra- and extra-cellular proteins [26].

S. pneumoniae's virulence thrives because of the bacteria's innate ability to acquire new genetic material via transformation and recombination. Horizontal transfer of genetic material, not only by homologous recombination but also by mobile genetic elements, generally amplifies substantially the heterogeneity of the common gene pool, providing a mechanism for capsular switching as described above, and advantages for the pathogen to circumvent the effects of vaccines and antibiotic treatment through resistance [6,25].

1.6. Antibiotic resistance and pneumococcus

Pneumococci were once among the most highly penicillin-susceptible bacteria, but clinical isolates evolved rapidly since the 1970s to contain a higher incidence of penicillin resistance. Contemporary strains with high levels of penicillin resistance also tend to carry resistance to other antibiotics, creating an additional challenge to treatment [27]. The phenomenon of multi-drug resistance (MDR) is currently observed in many other bacterial pathogens including Methicillin Resistant *Staphylococcus aureus* (MRSA) and *Mycobacterium tuberculosis* [27,28]. According to the United States Center for Disease Control (CDC), at least 2.8 million people are infected each year in the U.S. with antibiotic-resistant bacteria or fungi, and more than 35,000 people die as a result [29].

Penicillin-binding proteins (PBPs) are a group of key proteins with transpeptidase and/or glycosyltransferase functions that are essential for bacterial cell wall synthesis, involved in the assembly of the peptidoglycan after its precursor is transported from the cytoplasm, across the cell membrane, to the outer layers of the cell. PBPs earned their name from their affinity for penicillin and other β-lactam antibiotics [30,31]. Penicillin binds to wild type PBPs, preventing peptidoglycan linking in the last stages of bacterial cell wall synthesis, weakening the cell wall, and causing bacteria to lyse. Penicillin-resistant clones of pneumococci often have altered PBPs with markedly reduced affinity toward the antibiotic molecule, and presumably their natural substrates, and therefore producing peptidoglycans of grossly abnormal muropeptide composition and branching structure [30–32].

2. Strain Typing Methods

In the work that is reported in this chapter, performed in the early 1990s, three techniques were used to type bacterial clones/isolates: Multilocus Enzyme Electrophoresis, Pulsed-Field Gel Electrophoresis (PFGE) and analysis of PBP patterns. They are described below.

2.1. Multilocus Enzyme Electrophoresis

Multilocus Enzyme Electrophoresis is based on the premise that enzyme modifications alter the electrophoretic mobility of proteins. Variations of multiple enzyme loci are assessed by running multiple protein separation gels on culture extracts and assaying each gel for specific enzyme activity using labeling dyes.

The variable mobility patterns of each enzyme are assigned an allele type, and an increased variability of enzyme configurations is associated with genetic diversity. On the contrary, similar patterns are associated with lack of diversity and suggest the possibility of clonality [33,34]. Multiloccus Enzyme Electrophoresis experiments were conducted by collaborators in the laboratory of Dr. James Musser (Baylor College of Medicine, Houston, Texas), and in B. Spratt's laboratory (Brighton, UK) for the Spanish isolates.

2.2. Penicillin Binding Protein (PBP) patterns

PBP patterns are known to change dramatically with resistance to β-lactam antibiotics as a result of the reduced affinity that most resistant mutants show for those antibiotics [27]. Jabes *et al.* showed a direct correlation between the amount of [^{3}H] (tritiated) penicillin needed to detect most PBPs and the minimum inhibitory concentration (MIC), concluding from a relatively large number of isolates that increased penicillin resistance trended with decreased penicillin affinity. They found a high multiplicity of PBP patterns among highly resistant clinical isolates of pneumococci [35].

PBP patterns were assayed as described by Williamson *et al.*, by incubating exponential-phase cultures of each strain with saturating concentrations of tritiated penicillin ([^{3}H]-benzylpenicillin) to radioactively label all PBPs in 10 minutes, followed by lysis, protein separation by SDS-PAGE (Poly-Acrylamide Gel Electrophoresis) and detection by autoradiography [36].

2.3. Pulsed-Field Gel Electrophoresis (PFGE)

PFGE analysis involves the comparison of patterns of discrete fragments of the chromosomal DNA of each isolate after digestion with a specific restriction endonuclease, under the premise that mutations that create or eliminate specific recognition/restriction sites, or large deletions that change the length of

the fragments, should reflect how distant the individual isolates are to each other in their genetic makeover.

The method generally involves the delicate preparation of unbroken purified chromosomal DNA, followed by restriction with a 6 bp recognition site endonuclease, and by separation in agarose gels that can resolve the resulting large fragments. In order to preserve the integrity and length of the specific DNA fragments, a culture of each isolate was mixed with low gelling-temperature agarose and fixed before undergoing lysis, washing, and endonuclease restriction (in our case we used SmaI). The small agarose disks were inserted in the wells of agarose gels and subjected to Pulsed-Field Gel Electrophoresis (PFGE). DNA fragments were identified with Ethidium Bromide and photographed under UV light.

This electrophoretic method is a modification of the traditional gel electrophoresis, where the nucleic acids are subjected to an electric field and migrate towards the positive electrode; in PFGE the direction of the field sequentially alternates between two diagonal axes, inducing the large fragments of nucleic acid to be redirected alternately to the left and to the right of the main direction, and thereby creating a condition where every time the field changes the molecules need to readjust themselves to migrate along the agarose pores. The ability to quickly readjust and start migrating in a different direction is inversely correlated with the size of the molecule, resulting in the discrete separation of large nucleic acid fragments (40-400 kb) that would not be resolved in traditional electrophoresis.

3. Analysis of Serotype 6B Isolates from Iceland

3.1. Background

The work on *S. pneumoniae* described here was performed in 1992 in the laboratory of Dr. Alexander Tomasz at the Rockefeller University in New York.

By the early 1990s, penicillin resistance and MDR pneumococci had emerged in large numbers and at numerous global locations [37]. Some of the work done at the time consisted in developing molecular and genomic based techniques for identifying and typing isolates, and in using them to identify clone migration.

When my work took place in 1992, our collaborators in Iceland were alarmed by the quick spread of penicillin-resistance pneumococci. While sensitive isolates had been consistently collected since 1983, none had been identified with resistance to penicillin until 1988, at which point resistant isolates begun to appear, and rapidly climbed, reaching an incidence of 35 penicillin-resistant strains (more than 25 of them multi-resistant) on the last quarter of 1991. Since there was a predominance of isolates from serotype 6, and the policy for antibiotic use in the general population in Iceland was fairly restrictive, our colleagues postulated that the serotype 6 multi-resistant clone had been introduced from abroad - its spread facilitated by the grouping of children at day-care centers and by the more frequent use of antibiotics in children of that age [38].

Our team collaborated with Dr. James Musser (Baylor College of Medicine, Houston, Texas), and Dr. Karl Kristinsson (National University Hospital in Reykjavik, Iceland) in comparing these isolates to collections of other isolates of serotype 6B from Spain, Alaska, and Texas using molecular-based typing methods.

3.2. Experiments and results

Fifty-seven (57) clinical isolates of *S. pneumoniae* were obtained from the collection of the National Hospital in Reykjavik, representing isolates from Iceland collected from 1989 to 1991. The isolates belonged to serotype 6B, showing virtually identical multi-resistant phenotypes and, as described above, suggesting an epidemic spread of a single bacterial clone. These isolates were studied and compared with a subgroup of multiresistant serotype 6B pneumococci with high incidence in Spain in the prior two decades. Additional pneumococcal serotype 6B isolates from Alaska, Texas and England were used as controls in specific studies, as well as occasional isolates of different serotypes and different clones (as indicated in specific experiments).

Penicillin resistance was initially screened by disk diffusion, plating cultures of each isolate on Oxoid agar plates with 1μg oxacillin and 6% horse blood. Oxacillin, a β-lactam antibiotic of the same class

as penicillin, was first described by Dixon *et al.* in 1977 [39], for use in disk diffusion for screening strains of *S. pneumoniae* as penicillin disks did not reliably detect all resistant strains. Penicillin susceptible strains of pneumococci reliably produce oxacillin zones of 20 mm or greater, but oxacillin zones of ≤ 19 mm can be obtained with resistant, intermediate and with some penicillin-susceptible strains. **MICs (Minimum Inhibitory Concentrations)** were then determined by agar dilutions for strains with zone diameters of ≤ 19 mm.

Variations of 15 enzyme loci were assessed by **Multilocus Enzyme Electrophoresis** by the team in the laboratory of Dr. James Musser, Baylor College of Medicine, Houston, Texas. Extracts of cultures were run through multiple protein separation gels and each assayed for enzymatic activity detected with a labeled dye. Photographs of the protein gels were analyzed, and each variable mobility pattern for each enzyme locus was assigned an allele type. The allele type for each collection of isolates is represented in Table 1.1 with a number, in summarized format due to the large extension of data. The combination of the same alleles over the 15 enzyme loci was assigned a specific Electrophoretic Type (ET), represented in Table 1.1 by a letter on the right-most column.

Table 1.1. Multilocus Enzyme Electrophoresis Allele Types of Isolates by Origin
Numbers in the "Isolate Origin" column indicate the number of isolates that display the allele type indicated on the table over the total number of isolates on the collection. Allele types were arbitrarily assigned a number. For easier visualization of the differences, when a collection of isolates displays a different allele type than the others, that allele type is shown in blue font. Enzyme loci: (NSP) nucleoside phosphorylase, (LAP) leucine aminopeptidase, (ADK) adenylate kinase, (CAK) carbamylate kinase, (PGI) phosphoglucose isomerase, (EST1/EST2) esterase 1/2, (PGM) phosphoglucomutase, (LDH) lactate dehydrogenase, (GP1/GP2) glucose-6-phosphate dehydrogenase 1/2, (6PG) 6-phosphogluconate dehydrogenase, (G3P) glyceraldehyde-3-phosphate dehydrogenase, (GLD) glutamate dehydrogenase, (ADH) alcohol dehydrogenase. (ET) Electrophoretic Type, assigned according to the pattern of allele types at all the enzyme loci. The individual results for each isolate are described in Table 1.2 (page 15).

Isolate Origin	**Allele Type at each Enzyme Locus**															**ET**
	NSP	**LAP**	**ADK**	**CAK**	**PGI**	**EST1**	**EST2**	**PGM**	**LDH**	**GP1**	**GP2**	**6PG**	**G3P**	**GLD**	**ADH**	
Iceland 56/57	**3**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**3**	**5**	**5**	**5**	**5**	**A**
Spain 11/12	**3**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**3**	**5**	**5**	**5**	**5**	**A**
Alaska 4/4	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**5**	**3**	**5**	**E**
Texas 2/2	**3**	**7**	**5**	**5**	**5**	**5**	**3**	**3**	**3**	**5**	**3**	**8**	**5**	**5**	**5**	**F**

The results show that 56 of the 57 Icelandic isolates showed the same ET in all 15 enzyme loci assayed, and they were indistinguishable from 11 of the 12 isolates from Spain. The one Icelandic isolate

that was different (IC45), differed from the others in a single enzyme locus, EST2. The Alaska and Texas isolates differed more extensively from the Spanish-Icelandic ET at 2 to 5 enzyme loci. These results show that most of the isolates from Iceland have the same enzyme electrophoretic profile as the majority of the isolates from Spain, and differ from other serotype 6B isolates collected in Texas and Alaska, substantiating the hypothesis of a single origin between the Icelandic and Spanish isolates.

PBP patterns were assayed as described in section 2.2. As a reference, a penicillin susceptible non-encapsulated laboratory strain with a documented PBP pattern was used (R36A). In addition, a representative of a different clone also from Spain was included. Results are summarized in Figure 1.5, and the PBP pattern observed for each isolate is summarized in Table 1.2 (page 15).

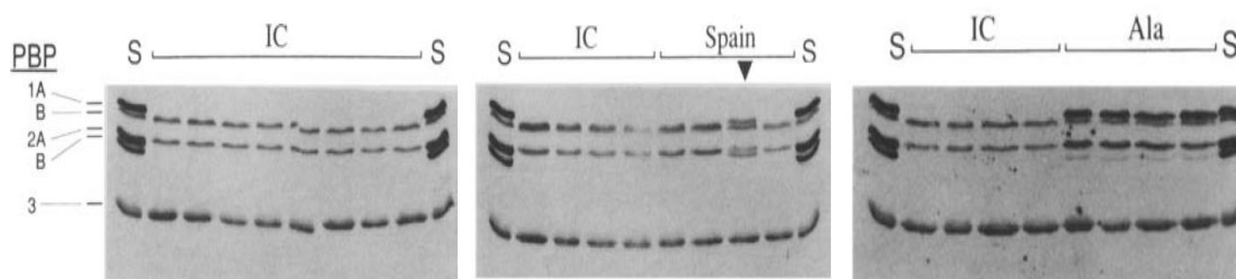

Figure 1.5. PBP (Penicillin Binding Protein) profiles of multiresistant isolates
IC: Iceland, Ala: Alaska, S: Susceptible Strain R36A. Arrow: control isolate from different clone from Spain. It should be noted that PBP 2x is not resolved from 2A as was frequently the case in protein gels at the time of this work [40].

PBP 2X appears unresolved from PBP 2A, as it typically did in protein gels at the time this work was performed; in fact, reference to PBP 2X is largely inconsistent in literature of the late 1980s and early 1990s [40,41].

The results show that the PBP profile is common among all 57 Icelandic isolates, and identical to the 6B isolates from Spain, but distinct from the Alaskan isolates and the control susceptible strain. The arrow indicates a Spanish isolate known to be from a different clone, which also shows a distinct pattern.

The PBPs 1B and 2A of Icelandic and Spanish isolates show somewhat reduced affinity for penicillin, as reflected by the reduced intensity of the bands in the autoradiograph. Additionally, PBPs 1A and 2B of those isolates are either fully altered or have negligible affinity for penicillin, in a way that they are not visible in the autoradiograph. These reduced penicillin-affinities or altered proteins are likely the cause for the low-level penicillin resistance observed in these isolates; in fact, PBPs 1A, 2B and 2X (the latter one being unresolved as previously mentioned) have been identified as the main players in resistance to penicillin, and mutations in PBP 2X and/or PBP 2B have been shown to result in low-level resistance to penicillin [42,43].

We subjected the chromosomal DNA of the pneumococcal isolates to endonuclease restriction with SmaI and to **PFGE** as described in section 2.3, and the results are shown in Figure 1.6, Figure 1.7 and Figure 1.8. Unfortunately, not all photographs of PFGE gels have transferred well to electronic form and some differences are difficult to visualize. Each individual PFGE pattern was assigned a Greek letter and the pattern exhibited by each isolate is summarized in Table 1.2 (page 15).

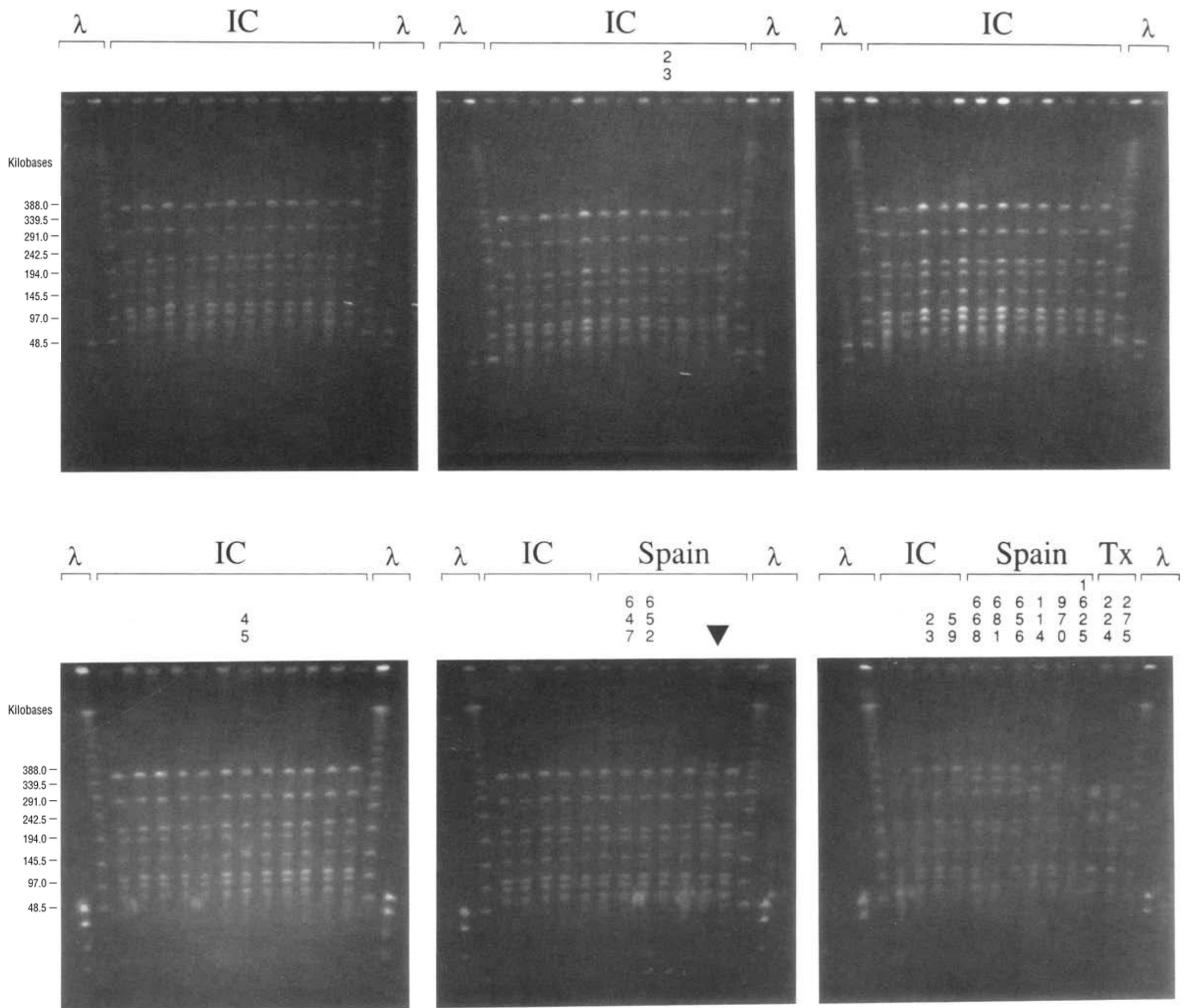

Figure 1.6. PFGE patterns of the SmaI fragmented chromosomal DNA

(IC) Icelandic isolates; (Spain) Spanish isolates; (Tx) Texas isolates; (λ) Molecular size markers λ ladder and low-range PFG (New England Biolabs, Beverly, MA), size indicated in kb. Most Iceland and Spanish isolates display an identical pattern arbitrarily named α, or different patterns that are closely related to α. All the isolates showing a pattern different than α are marked with their isolate number and the arrowhead indicates a strain of penicillin-resistant pneumococcus of a different clone. The pattern displayed by each isolate is summarized in Table 1.2 (page 15).

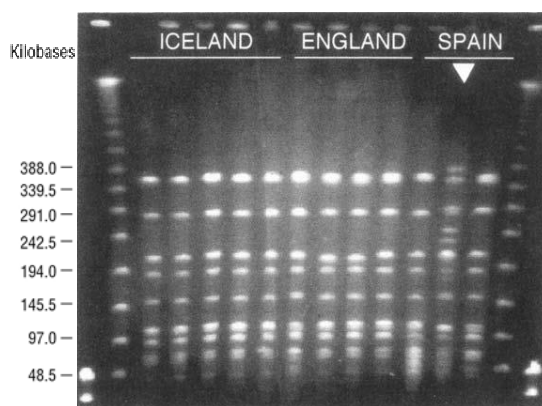

Figure 1.7. Prevalent PFGE SmaI pattern in isolates from Iceland, Spain, and England

Serotype 6B isolates from Spain and Iceland were compared with other serotype 6B isolates from England from the Rockefeller University collection, showing the common pattern α, with minor variations in two England isolates. A multi-resistant isolate of serotype 23F from Spain (marked with arrowhead), belonging to a different clone studied by Muñoz *et al*. [33,44], is included for contrast.

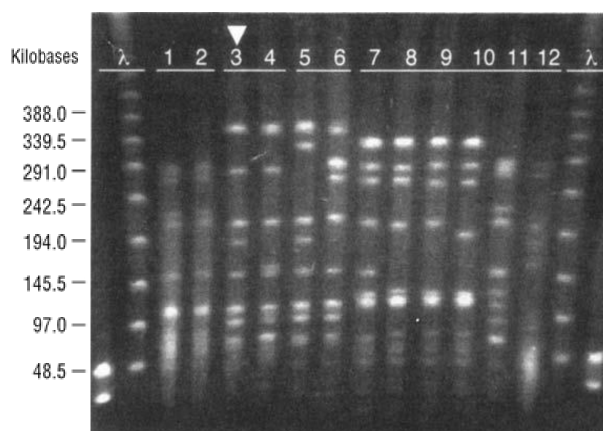

Figure 1.8. Different PFGE SmaI patterns of various isolates.

(1,2) Tex224 and Tex275, multi-resistant isolates from Texas, showing pattern θ (theta) with minor variations, θ1 and θ2. (3) Most common Icelandic isolate pattern α (alpha) represented by IC10. (4) Icelandic isolate IC45 showing pattern γ (gamma). (5,6) Spanish isolates Md970 and Md114 showing patterns σ (sigma) and δ (delta), respectively. (7-12) Penicillin-susceptible serotype 6B isolates collected in Iceland from January 1990 to March 1992, showing multiple patterns, and different from the most common resistant pattern.

The PFGE results show a common pattern in 54 out of the 57 studied isolates, arbitrarily named α, with seven discrete DNA fragments between approx. 90 kb and 388 kb, and only minor differences shown by the remaining three isolates (Figure 1.6 and Figure 1.7). The set of DNA fragments smaller than 97 kb are below the resolution threshold for the PFGE method used; they show non-discrete bands indicating multiple fragments without clear separation and are not consistently resolved in all gels; therefore, they were not considered when analyzing results in this study.

Isolate IC45 (Figure 1.6 and Figure 1.8), when compared to the most common pattern, is missing a band at approximate 194 kb and showing one extra band just above 145.5 kb (clearly visible by comparing lanes 3 and 4 of Figure 1.8). This could be the result of a deletion, or a single point mutation creating a new SmaI restriction site within the 194 kb fragment, therefore breaking it in two: a fragment of approximately 150 kb and a smaller fragment in the neighborhood of 44 kb which would not be clearly visible due to its small size.

Isolates IC23 (Figure 1.6, upper middle panel) and IC59 (data not shown, but summarized in Table 1.2, page 15) showed and even smaller difference from the most common pattern, displaying a very faint band just below the 339.5 kb marker. The pattern, named β1 or β2, is common between these and three isolates from Spain, Md656, Sp668 and Sp681 (Figure 1.6, lower right panel). Pattern β2 displayed by Sp668 and Sp681 shows a slightly more intense band just below the 339.5 kb marker than pattern β1

shared between Md656 and IC23, although difficult to visualize in the figures for IC23. It may be that this is just a visual artifact. However, these small differences do not detract from the main point that both β1 and β2 patterns are almost identical to the most common pattern α, found in most isolates studied.

Interestingly, two serotype 6B isolates from England, drawn from the Rockefeller University collection for comparison in PFGE, but that were not fully part of this study, also show an α-like pattern, matching the common Icelandic and Spanish isolates with only a slightly smaller fragment just above the 194 kb marker. This could be the result of a single deletion without impacting SmaI restriction sites.

It is also relevant to point out that PFGE clearly identifies the genomic variety presented by other *S. pneumoniae* isolates, which is apparent in Figure 1.8.

3.3. Summary and analysis

This work started with the hypothesis that the surge of multi-resistant pneumococcal isolates collected in Iceland from 1988 to 1991 originated from a single clone. The 57 isolates of serotype 6B chosen for this study were analyzed by three molecular typing techniques: PBP patterns by SDS-PAGE, Multiloccus Enzyme Electrophoresis, and PFGE. The Icelandic isolates were compared with 12 isolates from Spain with similar phenotypic characteristics that had a high incidence in Spain a few years before the Iceland spike started. For controls, serotype 6B strains from two different geographic areas in the US were used (Texas and Alaska).

All results collected in this study are listed in Table 1.2 (page 15), in conjunction with epidemiologically relevant Information and antibiotic susceptibility of the isolates. Gel patterns were arbitrarily assigned letters to designate identical or different patterns, with an extra asterisk or a subset 1 or 2, assigned to indicate only a minor difference.

MIC results are also presented in Table 1.2 and show MICs for penicillin between 0.5 and 2 µg/ml, levels that are not very different among the isolates, and which classify at low levels of resistance according to the literature available at the time [37,45]. It may be interesting to note that according later classifications, these MICs were grouped in the category of intermediate or borderline/below resistant [46,47], perhaps representative of the continuing increase in the level and prevalence of resistance, which is widely mentioned in the literature since the 1980s and was carefully studied in Tennessee during a 10-year period [48].

Table 1.2. Epidemiological, phenotypical, and molecular properties of the isolates
Type of Infection: (B) exacerbation of chronic bronchitis, (BL) septicemia, (C) carrier, not infected, (E) eye, conjunctivitis, (M) meningitis, (O) otitis media, (P) pneumonia, (S) sinusitis, (U) undefined. Antibiotic resistance: (S) susceptible, (I) intermediate, (R) resistant. (-) Not determined or not known. Multilocus enzyme electrophoretic types of Spanish Sp isolates were determined in B. Spratt's laboratory (Brighton, UK). Isolates showing small differences from the most common Icelandic patterns are highlighted in blue font; isolates showing substantial differences are highlighted in red font.

Origin	Isolate ID	BACKGROUND				ANTIBIOTIC RESISTANCE				MOLECULAR TYPING		
		Date collected	Patient Age Yrs or (mos)	Type of Infection	Serotype	Tetracyclin	Erythromycin	Chloramphenicol	Penicillin MIC (μg/mL)	PBP Pattern	Multilocus Enzyme Electrophoretic Type	PFGE *Sma* I Pattern
	IC2	1991.09	1	O	6B	R	I	I	1	a	A	α
	IC3	1991.09	3	C	6B	R	I	I	1	a	A	α
	IC4	1991.09	29	P	6B	R	R	I	1	a	A	α
	IC5	1991.09	1	S	6B	R	R	I	1	a	A	α
	IC6	1991.08	17	P	6B	R	R	R	0.5	a	A	α
	IC7	1991.07	2	O	6B	R	R	R	1	a	A	α
	IC8	1991.07	69	B	6B	R	R	R	0.5	a	A	α
	IC9	1991.07	85	P	6B	R	I	R	1	a	A	α
	IC10	1991.06	(8)	O	6B	R	I	I	1	a	A	α
	IC11	1991.06	(10)	O	6B	R	R	R	1	a	A	α
	IC12	1991.05	1	U	6B	R	R	R	0.5	a	A	α
	IC13	1991.05	(4)	C	6B	R	R	I	0.5	a	A	α
	IC14	1991.05	2	U	6B	R	I	I	1	a	A	α
	IC15	1991.05	2	U	6B	R	I	I	1	a	A	α
	IC17	1991.04	(11)	O	6B	R	R	R	2	a	A	α
	IC18	1991.04	18	U	6B	R	R	R	1	a	A	α
	IC19	1991.04	1	O	6B	R	R	I	2	a	A	α
	IC20	1991.04	1	O	6B	R	R	I	2	a	A	α
	IC21	1991.04	1	U	6B	R	R	I	1	a	A	α
	IC22	1991.03	1	O	6B	R	R	I	1	a	A	α
	IC23	1991.03	1	S	6B	R	R	R	2	a	A	β1
	IC24	1991.03	73	P	6B	R	R	R	1	a	A	α
	IC25	1991.03	1	BL	6B	R	R	R	1	a	A	α
	IC26	1991.03	(6)	O	6B	R	R	R	1	a	A	α
	IC27	1991.03	1	O	6B	R	I	I	1	a	A	α
	IC28	1991.03	1	O	6B	R	R	R	1	a	A	α
	IC29	1991.03	(9)	O	6B	R	R	R	2	a	A	α
	IC30	1991.03	1	O	6B	R	R	R	1	a	A	α
Iceland	**IC31**	1991.02	28	S	6B	R	R	R	1	a	A	α
	IC32	1991.02	28	U	6B	R	R	R	1	a	A	α
	IC33	1991.02	31	U	6B	R	R	I	1	a	A	α
	IC34	1991.02	(11)	O	6B	R	R	R	2	a	A	α
	IC35	1991.01	(1)	U	6B	R	I	I	1	a	A	α
	IC36	1991.01	1	O	6B	R	R	I	1	a	A	α
	IC37	1991.01	3	U	6B	R	R	R	1	a	A	α
	IC38	1990.12	67	P	6B	R	R	I	1	a	A	α
	IC39	1990.10	(10)	O	6B	R	R	I	1	a	A	α
	IC40	1990.11	71	P	6B	R	R	I	1	a	A	α
	IC41	1990.10	2	O	6B	R	I	I	2	a	A	α
	IC42	1990.09	79	P	6B	R	R	I	1	a	A	α
	IC43	1990.08	1	O	6B	R	R	R	1	a	A	α
	IC44	1990.08	(4)	O	6B	R	R	R	1	a	A	α
	IC45	1990.06	3	U	6B	R	S	R	1	a	A*	γ
	IC46	1990.05	89	S	6B	R	R	R	1	a	A	α
	IC48	1990.04	72	P	6B	R	R	R	0.5	a	A	α
	IC49	1990.02	33	S	6B	R	R	R	0.5	a	A	α
	IC50	1990.01	12	O	6B	R	R	S	0.5	a	A	α
	IC51	1990.01	(11)	O	6B	R	I	R	1	a	A	α
	IC52	1989.12	4	O	6B	R	R	R	0.5	a	A	α
	IC53	1989.11	1	U	6B	R	R	I	0.5	a	A	α
	IC54	1989.08	40	B	6B	R	R	I	0.5	a	A	α
	IC55	1989.08	1	O	6B	R	R	R	0.5	a	A	α
	IC56	1989.06	(10)	U	6B	R	R	I	1	a	A	α
	IC57	1989.05	1	O	6B	-	R	-	2	a	A	α
	IC58	1989.05	2	O	6B	-	R	-	2	a	A	α
	IC59	1989.04	(6)	O	6B	-	R	I	2	a	A	β1
	IC60	1991.09	2	E	6B	R	R	R	0.5	a	A	α
	Sp522	1987-89	-	-	6B	-	-	-	2	a	A	α
	Md873	1987-89	-	-	6B	R	S	R	2	a	A	α
	Md2067	1987-89	-	-	6B	R	S	R	2	a	A**	α
	Md1238	1987-89	-	-	6B	R	R	R	2	a	A	α
	Sp647	1987-89	-	-	6B	-	-	-	2	a	A	α*
Spain	**Sp652**	1987-89	-	-	6B	-	-	-	2	a	A	α*
	Md656	1987-89	-	-	6B	R	S	R	2	a	A	β1
	Sp668	1987-89	-	-	6B	-	-	-	2	a	A	β2
	Sp681	1987-89	-	-	6B	-	-	-	2	a	A	β2
	Md1625	1987-89	-	-	6B	R	S	R	2	a	A	δ
	Md114	1987-89	-	-	6B	R	S	R	2	a	A	δ
	Md970	1987-89	-	-	6B	R	S	R	2	a	A	σ
Texas,	**Tex224**	1989-91	-	-	6B	S	I	S	4	a*	F	θ1
USA	**Tex275**	1989-91	-	-	6B	S	S	S	4	a*	F	θ2
	Ala1	1982.03	1	P	6B	S	S	S	0.1	e	E	ϕ
Alaska,	**Ala15**	1986.05	1	P	6B	S	S	S	0.3	e	E	ϕ
USA	**Ala31**	1988.07	(10)	-	6B	S	I	S	0.1	e	E	ϕ*
	Ala39	1989.06	(2)	M	6B	S	R	S	0.3	e	E	ϕ*

Our results show that 95% of the Iceland isolates (54/57) show identical patterns in all three molecular typing techniques. Those patterns are shared with 50% of the isolates from Spain (6/12), with minor variants excluded. This evidence supports the hypothesis that most Icelandic isolates originated from a single clone, and that that clone was imported, potentially from Spain. The findings in the England isolates, which show similar PFGE patterns in Figure 1.7, point to a perhaps wider spread of the initial clone. In contrast, the different patterns of several isolates from Spain, Texas, and Alaska show that the techniques can differentiate clones of pneumococci from one another.

It is especially noteworthy that PFGE, an innovative technique at the time, was able to provide a large variety of patterns for the isolates studied and differentiate clones from Spain that would otherwise look identical. PBP and Multilocus Enzyme Electrophoresis both highlight protein differences that impact protein gel mobility and ability to bind substrate. However, PFGE highlights genomic sequence differences at the DNA level, *i.e*., mutations or chromosomal rearrangements. Many of those can be completely silent (not impacting translation), or they may impact protein sequence but not mobility or substrate affinity, or they may also have a significant impact in the structure and function of proteins that are not penicillin-binding, or one of the 15 enzymes selected for Multilocus Enzyme Electrophoresis - in all cases providing possible explanations for the increased discernment of the genomic method. PFGE was one of the first genomic methods used for identification and typing in the early 1990s, leading the way to more sophisticated genomic and mutation analysis techniques which are mentioned later (6.2. Evolution of typing techniques).

This work was published in 1993 in the Journal of Infectious Diseases with the title "Evidence for the Introduction of a Multiresistant Clone of Serotype 6B *S. pneumoniae* from Spain to Iceland in the Late 1980s" [49].

4. Serotype 23F Isolates from North Carolina

4.1. Background

My work with molecular typing, especially PFGE, also helped provide evidence for a study of Serotype 23F pneumococci from North Carolina, USA, where a clonal outbreak was suspected. Children from a day care center in Chapel Hill, North Carolina were under surveillance for nasopharyngeal colonization with *S. pneumoniae* since 1985. In May 1990, an outbreak started, where nasal carriage of MDR resistant serotype 23F was observed in multiple isolates sampled from 14 out of 52 children until December 1991.

Muñoz *et al.* and Reichler *et al*. [33,50] had reported a similar outbreak in a Cleveland, Ohio daycare. Muñoz *et al.* compared it to a collection of Spanish serotype 23F pneumococcus isolates for PBP mobility as well as penicillin binding capacity through PBP autoradiograph patterns, broad enzymatic uniformity through multilocus enzyme electrophoresis, and additionally used a PCR based gene fingerprinting technique to analyze the DNA sequence of PBPs through restriction mapping. They concluded that the isolates collected in Spain in 1989 (where multi-drug resistance had been observed in the prior 5-10 years) were of the same clonal origin as those collected in the Cleveland daycare, likely moving from Spain to Cleveland, Ohio [33].

The body of work reviewed below consisted of using traditional and molecular typing techniques to investigate the correlation of the Chapel Hill isolates to the Cleveland and Spain isolates reported by Muñoz *et al*. [33].

4.2. Experiments and results

A total of 1123 *S. pneumoniae* isolates were collected during the surveillance program in Chapel Hill, with 207 being selected for antibiotic susceptibility testing, each representing a colonization episode of a certain serotype. Serotype 23F was prevalent during 1990 and 1991. A total of 13 MDR and 10 penicillin-susceptible (PS) serotype 23F isolates collected at the time of the North Carolina daycare outbreak were selected for further studies of molecular nature (PBP and PFGE) and compared with several other isolates from Spain, Ohio and Texas.

PBP and PFGE studies were executed as described previously in Section 2. PBP results are shown in Figure 1.9 and Figure 1.10; PFGE results are shown in Figure 1.11 and Figure 1.12 .

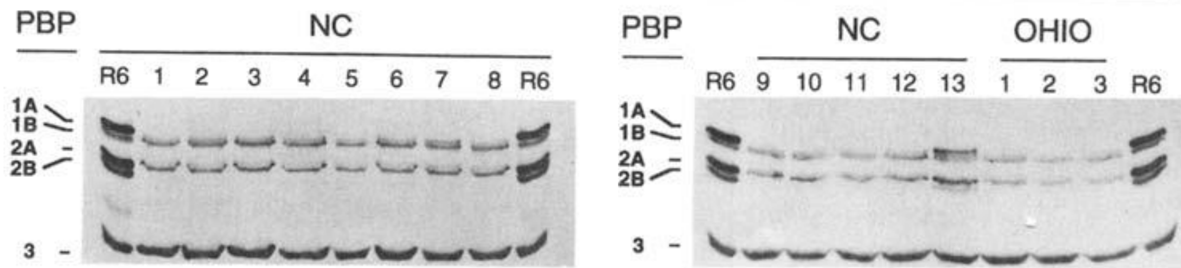

Figure 1.9. PBP pattern of MDR serotype 23F isolates

(NC 1-13) MDR isolates from Chapel Hill, North Carolina. (Ohio 1-3) MDR isolates from Cleveland, Ohio reported by Muñoz *et al*. and Reichler *et al*. [33,50]. (R6) Laboratory standard penicillin-susceptible strain R6. Both collections of isolates show an identical pattern (with only the exception of isolate NC13), with alterations in penicillin-binding proteins 1A and 2B. As previously mentioned, PBP 2X appears unresolved from PBP 2A.

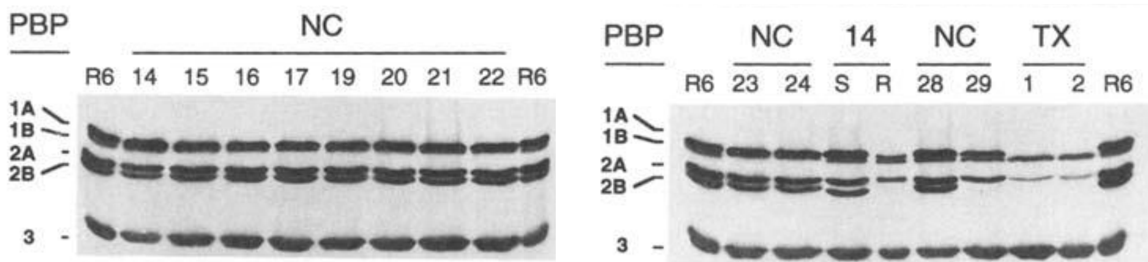

Figure 1.10. PBP pattern of Penicillin-Susceptible isolates

(NC 14-24) PS serotype 23F isolates from Chapel Hill, North Carolina. (14 S-R) Serotype 14 isolates from Chapel Hill daycare collected from the same child about four months apart; S: penicillin-susceptible, R: MDR. (NC 28-29) Serotype 23F isolates from Chapel Hill; 28: penicillin-susceptible, 29: MDR. (TX 1-2) MDR Serotype 6B isolates from Texas. (R6) Laboratory standard penicillin-susceptible strain R6. Penicillin susceptible isolates display PBP patterns similar to the control strain R6 (with perhaps only slight mobility-affecting alterations in PBP 1B or 2B). In contrast, MDR isolates 14R, NC29, TX1 and TX2 of serotypes 14, 23F and 6B, respectively, show PBP patterns with significant alterations in proteins 1A and 2B. Although this pattern is similar across the three serotypes, some differences can be seen, namely in the PBP 1B band and the intensity of the PBP 2A band. As previously mentioned, PBP 2X appears unresolved from PBP 2A.

The PBP autoradiographs show a common pattern, arbitrarily named **a**, between the NC 1-12 MDR isolates of serotype 23F from North Carolina and Ohio, displaying significantly altered PBPs 1A and 2B (Figure 1.9). Isolate NC13 shows a slightly different pattern arbitrarily named **a***. These patterns **a** and **a*** are also displayed by other MDR isolates, namely, 14R, NC29 and two isolates of serotype 14 from Texas, Tx 1-2 (Figure 1.10). Penicillin-susceptible isolates show a pattern named **b** that is common with the reference susceptible strain R6 (Figure 1.10). Results are summarized in Table 1.3 (page 20).

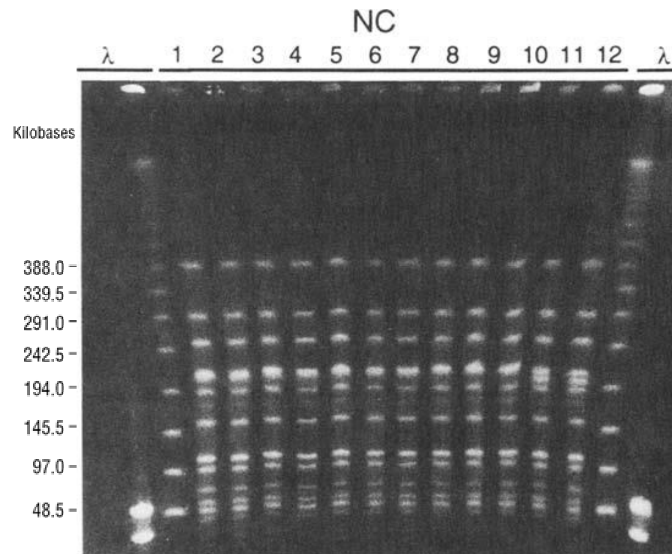

Figure 1.11. PFGE of MDR serotype 23F isolates

(NC 1-12) MDR serotype 23F isolates from Chapel Hill, North Carolina. (λ) Molecular size markers λ ladder and low-range PFG (New England Biolabs, Beverly, MA), size indicated in kb.

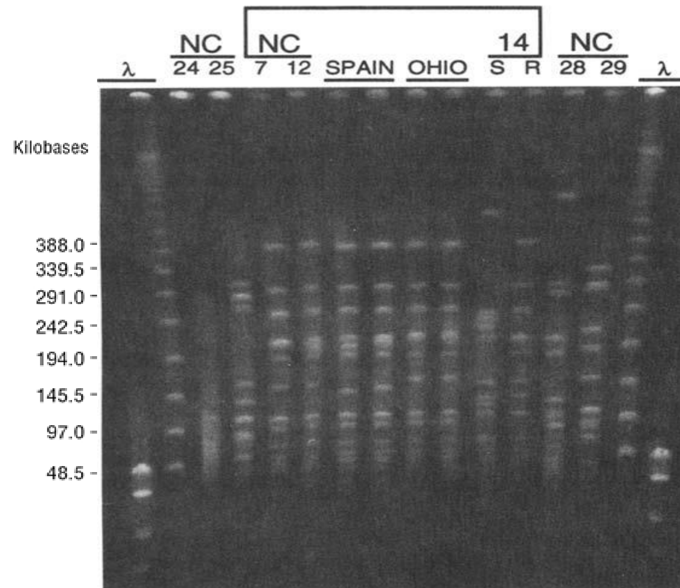

Figure 1.12. PFGE of various isolates

(NC 24-25) Penicillin susceptible serotype 23F isolates from Chapel Hill. NC24's pattern cannot be assessed due to degradation. (NC 7-12) MDR serotype 23F isolates from Chapel Hill. (Spain, Ohio) MDR serotype 23F isolates from Spain and Ohio, respectively. (14 S-R) Serotype 14 isolates from Chapel Hill daycare collected sequentially from the same child; S: penicillin-susceptible, R: MDR. (NC 28-29) Serotype 23F isolates from Chapel Hill; 28: penicillin-susceptible, 29: MDR.

PFGE results also show a prevalence of two remarkably similar patterns in the North Carolina MDR isolates (Figure 1.11); the pattern arbitrarily named A1 is common between isolates NC 1-10 and identical to the Ohio isolates. Pattern A2 is displayed by isolates NC11 and NC12 and differs from A1 only in the 194-242 kb range. The Spanish isolates also show patterns extraordinarily like A1 and A2 with only one band slightly shifted in the 145 kb area (Figure 1.12). Results are summarized in Table 1.3.

Table 1.3. Summary of Serotype 23F isolates and PBP/PFGE patterns
(MDR) Multidrug Resistant; (MDR*) Multidrug Resistant but susceptible to Erythromycin. PS: Penicillin Susceptible. (-) Not available. PBP and PFGE patterns can be found in Figures 1.9 to 1.12, and explanation is found in the main text.

ID	Origin	Date of Isolation	No. of isolates	Serotype	Antibiotic resistance pattern	PBP pattern	PFGE pattern
NC 1-13	Chapel Hill, NC	1990-91	13	23F	MDR	a, a*	A1, A2
NC 14-24	Chapel Hill, NC	1990-91	10	23F	PS	b	-
NC 25	Chapel Hill, NC	1990-91	1	23F	PS	-	B
NC 28	Chapel Hill, NC	1990-91	1	23F	PS	b	C
NC 29	Chapel Hill, NC	1990-91	1	23F	MDR	a*	D
Ohio 1-3	Cleveland, OH	1990	3	23F	MDR*	a	A1
Spain 1-2	Spain	1989	2	23F	MDR*	-	A1*, A2*
Tx 1-2	Texas	-	2	6B	MDR	a	-

A few point-mutations could be all that is different between these MDR isolates, supporting the hypothesis that they originated from a single *S. pneumoniae* serotype 23F MDR clone that spread across continents from Spain to at least two areas of the United States. A few other PFGE patterns can be observed in individual North Carolina isolates NC25, NC28 (both penicillin-susceptible) and NC29 (MDR) that are not shared with any others, indicating there are other, perhaps less prevalent, clones in the population.

Although the molecular data from these techniques indicate with reasonable certainty that the isolates from North Carolina originate from the same clone as the isolates from Spain and Cleveland, there is a relevant difference that should be mentioned. The isolates from Spain and Cleveland were multidrug resistant but susceptible to Erythromycin, whereas the isolates from Chapel Hill were highly resistant to Erythromycin. Since the Chapel Hill day care surveillance program shows that the last Erythromycin resistant isolate occurred 12 years prior to the MDR 23F outbreak, it is unlikely that resistance was acquired upon arrival in North Carolina; therefore, it is postulated that this clone was introduced into North Carolina through at least one intermediary location, where it would have acquired Erythromycin resistance, and not directly from Cleveland or Spain.

It should be pointed out that although side-by-side comparisons are not available, the PBP pattern of the MDR serotype 23F isolates (Figure 1.9) is very similar to that of the 6B isolates from Iceland reported in Figure 1.5; however the PFGE patterns are quite different between these two collections of isolates (Figure 1.6 and Figure 1.11), confirming earlier conclusions that PFGE provides higher discriminatory power than the PBP pattern in differentiating strains.

4.3. Case-study: hypothesized seroconversion

An interesting hypothesis can be derived from the analysis of a subset of the data presented above. My colleagues studying the serotype and drug resistant patterns of the isolates from the Chapel Hill daycare had noticed that one specific child was colonized with MDR serotype 23F at two sequential time-points: November 1990 and January 1991. That child became colonized with penicillin-susceptible (PS) serotype 14 in February 1991 and then with MDR serotype 14 in June of the same year. A schematic timeline of isolate collection from the child is presented in Figure 1.13.

The PBP and PFGE patterns were shown in Figure 1.9 through Figure 1.12, and are summarized in Table 1.4.

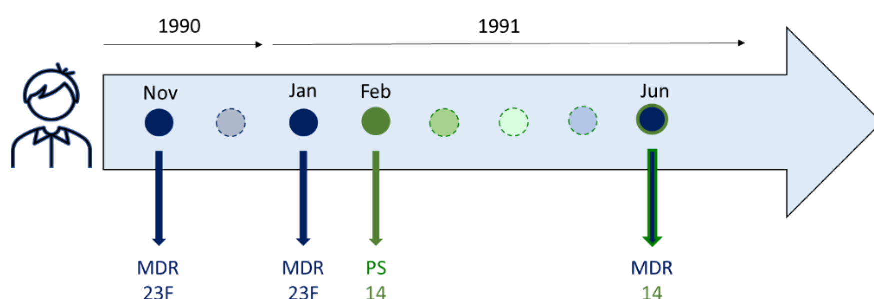

Figure 1.13. Schematic timeline of isolate collection from a specific child in the North Carolina daycare
Time points and isolate characteristics are described in the text. (MDR) Multi Drug Resistant. (PS) Penicillin Susceptible. (23F) Serotype 23F. (14) Serotype 14.

Table 1.4. Summary of isolates collected from the same child
Antibiotic Resistance: (R) Multidrug Resistant. (S) Penicillin Susceptible. PBP Patterns: (a) most common pattern found in MDR serotype 23F isolates. (b*) pattern similar to frequent susceptible pattern b, found in susceptible isolates NC14-24 and in reference strain R6. PFGE Patterns: (A1) Most common pattern found in MDR serotype 23F isolates NC 1-10. (A2) Pattern found in MDR serotype 23F isolates.

Collection Date	Serotype	Antibiotic Resistance	PBP Pattern (Figure 1.9 and Figure 1.10)		PFGE Pattern (Figure 1.11 and Figure 1.12)	
			Isolate No.	Pattern	Isolate No.	Pattern
1990 Nov	23F	R	NC 1 - 12 NC 29	a	NC 1 - 12 NC 29	A1/A2
1991 Jan	23F	R	NC 1 - 12 NC 29	a	NC 1 - 12 NC 29	A1/A2
1991 Feb	14	S	14S	b*	14S	E
1991 Jun	14	R	14R	a	14R	A3

As expected, the patterns of the penicillin susceptible isolate 14S are different from the MDR 23F patterns. However, the subsequently emerged isolate 14R, collected from the same child, shows PBP and PFGE patterns that are remarkably like the most common MDR 23F patterns, which had also been isolated from the same child. This finding leads us to the hypothesis that maybe, instead of undergoing antibiotic resistance conversion from susceptible to resistant, the child might have harbored both serotypes and enabled an exchange in capsular coding material from the new serotype 14 into the original MDR serotype 23F, in such a way that the MDR serotype 23F converted to serotype 14. Capsular switching or serotype conversion is a well described phenomenon in *S. pneumoniae* literature [25] as highlighted in the introduction to this chapter. However, we do not have enough data to confirm this hypothesis.

This work was published in 1995 in the Journal of Infectious Diseases with the title "Transmission of Multidrug-Resistant Serotype 23F *S. pneumoniae* in Group Day Care: Evidence Suggesting Capsular Transformation of the Resistant Strain In Vivo" [51].

5. Conclusions

The findings from the molecular typing work described above support the hypothesis that multidrug resistant clones of *S. pneumoniae* were transferred across borders and propagated in child day care centers.

With respect to the clone of serotype 6B, the conclusions may be summarized as follows: a) 95% of the Iceland isolates show identical patterns and match 50% of the isolates from Spain, with minor variants excluded; b) evidence supports the hypothesis that most Icelandic isolates originated from a single clone; and c) since the clone had been established in Spain for 3 decades and no resistant isolates were found in Iceland until 3 years prior, it is likely that the clone was imported from Spain.

As to the clone of serotype 23F: a) all the North Carolina isolates share two extremely similar patterns, which match the Ohio and Spain isolates; b) evidence supports that the isolates from North Carolina originated from the same clone as the isolates from Spain and Cleveland, and it is likely that the clone was propagated from Spain, since multidrug resistance was prevalent in Spain for decades prior, but not in the Unites States; and c) it is postulated that this clone was introduced into North Carolina through at least one intermediary location, where it would have acquired Erythromycin resistance, because resistance to this antibiotic had not observed in Chapel Hill in the prior 12 years.

The molecular typing of *S. pneumoniae* isolates that I performed was important at a time when the travel of clones across borders was just beginning to be suspected. PFGE was not a prevalent technique at the time, and with my work, I was able to provide reasonable evidence that the travel of multidrug resistant clones from Spain was the most likely reason for the appearance of the 6B clone in Iceland, and that the 23F clone from Spain and Ohio had made it to North Carolina.

The emergence of antibiotic resistance was just beginning to be raised as a problem. From both the analysis of the 6B clone and the 23F clone, there was evidence that children in day care centers were passing on these resistant bacteria and propagating the problem. It was important to raise awareness to the fact that the random and frequent use of antibiotics needed to be reduced – or pathogenic bacteria like *S. pneumoniae* could evolve into full drug resistance at a faster rate than new drugs could be discovered. This battle is by no means over, but today there is a higher level of awareness and increased restraint in the use of antibiotics, by both doctors and patients, than in the early 90s.

PFGE, the technique that I inherited and fine-tuned in Dr. Tomasz's lab continued to be improved and used in epidemiologic work; it is referred today as a "historically important molecular method" [25].

The two clones that I studied went on establish formal pneumococcus lineages and were found in other areas beyond the places that I identified (see section 6.1).

6. Further Developments

6.1. Changes in pneumococcal epidemiology

During their work on the development of a then novel molecular method for pneumococcus epidemiological studies, Multi-Locus Sequence Typing (MLST), in 1998**,** Enright and Spratt included a few samples of the serotype 6B and serotype 23F clones from Spain that I reported in sections 3 and 4, respectively. In their publication they verified with the new technique that each of them likely originated from a single clone and used the data to validate their new method [52].

Aiming at global surveillance of antibiotic-resistant pneumococcus, the Pneumococcal Molecular Epidemiology Network (PMEN) [53] was established in 1997. It brought standardization and structure to the field of *S. pneumoniae* epidemiology, with consistent nomenclature and classification of clones. As of May 2020, there were 43 clones/lineages listed by the PMEN. In the initial release of the clone classification by McGee *et al.* in their 2001 publication, the 23F clone that I reported in section 4 established the PMEN1 lineage. The PMEN2 lineage was established by the serotype 6B clone reported in section 3 [47].

PMEN1, also referred to as the 23F Spanish/USA clone or Spain^{23F}-1, was identified and studied in numerous publications as having been ubiquitous around the globe for decades; it is typically characterized as a low virulence, multi-drug resistant, highly pervasive clone in non-symptomatic carriers [47,54,55].

However, many studies have found that the prevalence of capsular serotypes has shifted after the 1990s for several reasons.

The prevalence of PMEN2, also referred to as or Spain^{6B}-2, the 6B clone prevalent in Iceland reported in section 3, remained the predominant type among penicillin-resistant *S. pneumoniae* through 1996 [56], but declined in Iceland after the mid-1990s [54]. Croucher *et al.* published a study in 2014 with Whole Genome Sequencing analysis of 189 isolates from 12 countries, including 118 from Iceland, collected from the late 1980s to 2009, where they observed that the Icelandic clade of PMEN2 lost antibiotic resistance over time, and lacked the importation of divergent sequences through homologous recombination when compared to other clades of PMEN2 outside Iceland. The decreased selective pressure for multi-drug resistance, achieved by the country's reduction of antibiotic administration to children after 1992, as well as the clone's lack of antigenic diversification resulting from the absence of recombination, might have contributed to the decline [54].

Since the widespread implementation of the **Pneumococcal Conjugate Vaccines** (PCV7 in 2000 and subsequently PCV13 in 2010, which contain polysaccharide antigens for a total of 7 and 13 serotypes respectively), Invasive Pneumococcal Disease (IPD) has significantly decreased. However, *S. pneumoniae* continues to be recovered, frequently from vaccinated children, and the bacterial population has shown

changes. There is multiple evidence of genomic recombination and cases of serotype switches [2,55,57,58]. Interestingly though, serotype replacement occurs mainly due to expansion of non-vaccine types that pre-existed before vaccine introduction; postvaccine capsular switches (from vaccine types to non-vaccine types) did not contribute significantly to the pool of post-vaccine drug resistant isolates [59].

As of May 2020, two new and expanded pneumococcal conjugate vaccines are already on the horizon. PCV15 is in Phase 3 (last stage) clinical trials, with completion expected in May 2021 [60], and PCV20 is in early development. It will be interesting to observe the evolution of the pneumococcal population if these vaccines become widely administered to children, replacing PCV13.

6.2. Evolution of typing techniques

My work on *S. pneumoniae* was performed during a time when molecular techniques of analysis were beginning to be explored for studies of microbial diversity and general epidemiology. Even though the techniques presented here have limited efficacy and very low throughput, they are the base for a set of methods that started brewing at the time and succeeded later with lower costs and higher efficiencies, culminating with genomic sequencing and real-time PCR today.

In the paragraphs below is a modest overview of the evolution of techniques related to my work after its execution in the early 1990s. This does not aim to be a complete scholarly review, but instead to assess how techniques evolved from each other, how they advanced, and how they impacted the field of pneumococcus epidemiology and the science of today.

PFGE continued to be used and was a historically important method of characterization for many microorganisms [25]. It was further validated and standardized for wider use across laboratories with guidelines released in 1995 [61,62]. It is interesting to observe that with wider use and standardized guidelines, the quality of PFGE gels increased remarkably over time, with improved resolution, increased visibility of discrete band patterns, and straighter gel lanes – features that my work did not consistently achieve at the time.

The PFGE method is related to **RFLP (Restriction Fragment Length Polymorphism)**, which was initially developed by Alec Jeffreys starting in the late 1970s and published in 1979 [63]; the original method involved digesting human chromosomal DNA with restriction enzymes, separating the resulting fragments by gel electrophoresis and detecting specific regions by Southern blot hybridization with radioactively labeled probes targeting various types of globin genes. It revealed that specific variants were associated with different individuals and was used to establish the link to maternal heritage [63]. This discovery unleashed several subsequent applications including forensics, paternity/twin testing, and conservation biology [64,65].

The wide use of RFLPs was greatly limited by the need for large quantities of genomic DNA; however, this problem was solved in 1985 with the epic advent of the **Polymerase Chain Reaction (PCR)**

[66,67]. RFLP analysis was combined with PCR amplification, which resolved the large quantity problem and facilitated its wide use, for example as a diagnostic tool for sickle cell anemia [66].

PCR is a well-established and extensively used technology that is based on the selective large-scale amplification of DNA fragments. DNA polymerases that withstand sequential denaturation at high temperatures are allowed repeated and cyclical access to single-stranded and primed DNA template, which they use to construct the complementary strand. The repeated cycles, and the ability to use the product of a previous cycle as the template for the next, provide for a logarithmic amplification power of select DNA fragments.

PCR has a multitude of uses in Molecular Biology, and as it became widely accessible, it quickly started to be incorporated in *S. pneumoniae* typing methods. A few examples are mentioned below.

In 1996, van Belkum [68] published results of a novel DNA fingerprinting technique, **BOX-PCR**, targeting a series of interspersed repetitive DNA sequences that are characteristic of *S. pneumoniae* [69], matching typing resolutions of prior techniques and increasing speed and resolution.

In 2003, Brito *et al*. [70] developed a two-step PCR approach to identify pneumococcal serotypes in a fast and cost-effective way, **PCR-serotyping,** achieving 94.2% agreement with the traditional immunological method (Quellung reaction) on their first attempt. Still today the United States Centers for Disease Control and Prevention (CDC) recommends both conventional and real-time PCR methods on their website for deduction of pneumococcal serotypes and serogroups [71]. Multiple publications followed, with incremental improvements such as Real-time PCR and Whole Genome Sequencing (WGS), which have been fundamental to support studies surrounding the effects of the 7-valent conjugate pneumococcal vaccine (PCV7) released in 2000 and subsequently the 13-valent conjugate pneumococcal vaccine (PCV13) released in 2010 [57,72,73].

PCR was also used to improve the PBP pattern method used in my studies – **PBP Fingerprinting** is referenced in many publications and executed in various ways, but always based on the PCR amplification of the PBP genes, followed by digestion with one or two restriction endonucleases with 4-5 bp recognition sites, and typically detected by ethidium bromide staining [74,75] , although earlier versions included end-labeling with radioactively labeled ^{32}P [76].

RFLP analysis, combined with PCR, was an important early tool in genome mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing [65] but it has since then reached its sunset. Techniques to sequence DNA had been adopted since Sanger's chain-termination technique was published in 1977 [77], but they were very manual, cumbersome and time consuming. However, they were advancing rapidly during the late 20th century: the **Human Genomic Sequencing Project** [78], from 1990 to 2003, unleashed methods of rapid and inexpensive **DNA Sequencing**. In succession, the identification of **Single Nucleotide Polymorphisms (SNPs)** largely replaced RFLPs in prior and novel applications [79–81].

An important molecular method based on sequencing was developed for *S. pneumoniae* epidemiology studies by Enright and Spratt in 1998: **Multi-Locus Sequence Typing (MLST)** [52]. In this method, seven housekeeping genes are sequenced as a sample of genomic variation and are used to define the sequence type. MLST proved to be a powerful new approach to the characterization of pneumococci, providing molecular typing data that is objective and can be electronically portable between laboratories, used to build databases, and establish objective measures of linkage [25,52].

Sequencing has greatly evolved into **Next-Generation Sequencing (NGS)** – a powerful and diverse set of technologies which has drastically reduced the cost and difficulty associated with earlier DNA sequencing [82–85]. These remarkable advances in nucleotide sequencing technology, have revolutionized the field of genomics and created the possibility of **Whole Genome Sequencing (WGS)** in as little as one day. They are widely used in applications such as identification, disease predisposition and human lineage in largely commercialized ancestry kits [86]. NGS also has wide applications in disease management and healthcare. An example is the technology being developed by Natera, San Carlos, CA, USA, combining SNP variability and cell-free DNA detection through NGS to address fundamental questions in the fields of Reproductive Health, Oncology and Organ Transplant [87].

The first pneumococcal genome sequenced was that of a virulent isolate of serotype 4, TIGR4, published in 2001 by Tettelin *et al.*[88]. The ability to sequence multiple whole bacterial genomes through WGS brought an important concept to studies of *S. pneumoniae* and other pathogenic bacteria. The notion that pathogenic bacteria relied on a pool of genomes to draw from and adapt to diverse host environments was first introduced by Ehrlich *et al.* in 2005 as the theory of bacterial plurality [89]. This theory, later referred to as the distributed-genome hypothesis (DGH) was confirmed for *S. pneumoniae* initially by Hiller *et al.* in 2007 [90] and Donatti *et al.* in 2010 [91], and since then by many studies. In May 2020, the genomes of over 14,000 strains of pneumococcus were publicly available and this number is constantly increasing [92]. Data show that *S. pneumoniae* has a **supragenome**, or **pangenome** of typically 5,000 – 7,000 gene clusters, which is much larger than the genome of any individual strain, with typically 500 – 1,100 gene clusters; about three quarters of pneumococcal genes are differentially distributed across strains. The open pangenome guarantees the species a quick and economical response to diverse environments [59,91].

Whole genome sequence analysis confirmed that genetic exchange occurs both within and across the borders of the species. Homologous recombination of short stretches of DNA from related species sharing the same ecological niche is the main mechanism of evolution of *S. pneumoniae*; *Streptococcus mitis*, which also colonizes the human nasopharynx, is the main reservoir of genetic diversity of *S. pneumoniae* outside the species [59,91].

I have observed an interesting change in the tone of conclusions from the early-1990s molecular typing publications to those in the post-sequencing era. What were mere hypothesis and postulates as

to lineages and clone migrations in my publications and those of the time, have now been confirmed and are presented with certainty - with in-depth analyses of phylogenies and distinction from horizontal recombination *versus* vertically acquisition of substitutions. The technological inaccessibility to data led us to publish conclusions with unproven hypotheses in the early 1990s, but WGS has repaired this vacuum. This is very evident in a 2011 publication in the journal Science about pneumococcal evolution, by Croucher *et al.* [58].

Since the beginning of this century, **Real-time PCR** has developed greatly and established itself as the method of choice for identification and characterization of organisms. It has a very quick turnaround time, and the ability to multiplex and detect multiple genes or organisms in a single reaction. When coupled with melting curve analysis and with the right probe design, it can provide discrimination to the level of point mutations. In addition, with the appropriate controls, it can be used to quantitate the starting nucleic acid material, increasing its applications to DNA quantitation. When combined with reverse-transcription (**Real-time RT-PCR**) it can be used in RNA applications and gene expression. An illustration of real time PCR curves is shown in Figure 1.14.

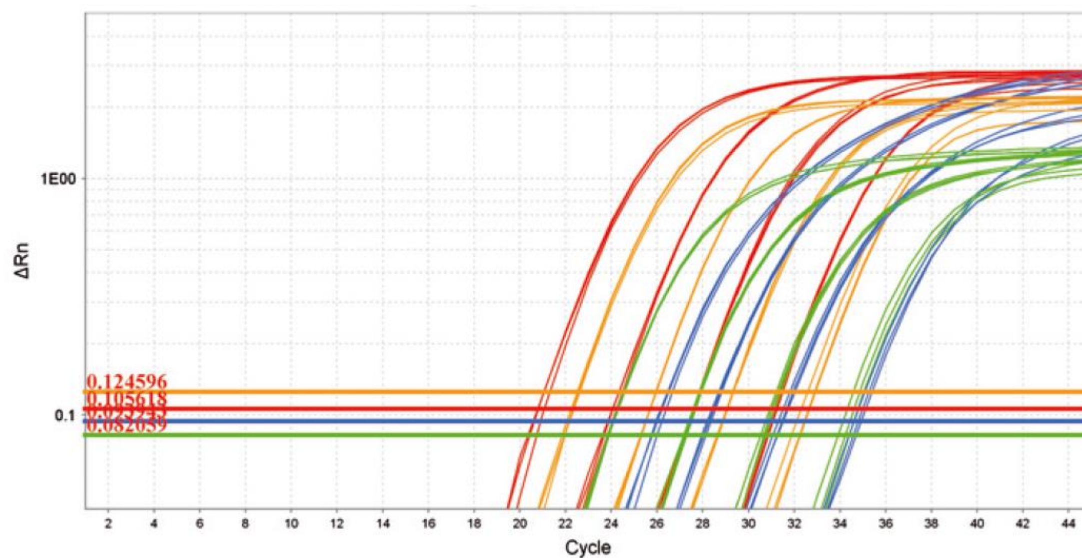

Figure 1.14. Example of a Multiplex Real-Time PCR Graph

Each color represents a probe labeled with a different fluorescence dye to detect a specific target gene. The X-axis represents the PCR cycle and the Y-axis represents the florescence increase. The graph lines portray the logarithmic PCR amplification of the target in real time, as detected by the accumulation of fluorescent probe. Lines of the same color that are separated by approximately 3-4 cycles represent amplifications of tenfold serial dilutions of DNA template. Adapted from Solis Biodyne 5x-HOT-FIREPol®-Multiplex-qPCR-Mix [93].

Real-time PCR is the gold-standard culture-independent assay for ***S. pneumoniae*** **identification**, targeting the major autolysin *lytA*, the pneumococcal surface adhesion *psaA*, and the permease *piaB* genes, among others [94–96]. In 2019, Tavares *et al.* developed a new real-time PCR assay targeting the putative transcriptional regulator gene SP2020 increasing the specificity of the identification by 0.2% to

99.8% with the single new gene, and reducing the misidentifications to none (0/823) when combined with *lytA* [94].

Real-time PCR is also widely used in molecular diagnostic tools outside of *S. pneumoniae*. For example, it is used to identify infection and antibiotic resistance in the assay commercialized by Cepheid Inc., Sunnyvale, CA, USA, for the detection of ***Mycobacterium tuberculosis*** [97–99]; this assay is reviewed in Chapter 3.

While each of the techniques outlined here has brought amazing advances in the molecular biology field, and although it is undeniable that WGS has changed science and the world, I think that the true revolution of the 21st century was brought on by something that is common between many of those techniques: the combination of scientific specialties. The powerful techniques we see today, with their efficiency, high-throughput and accuracy combine multiple platforms of expertise. They rely on amazing combinations of Engineering with Physics, Chemistry, and Biology, and produce complex results that are processed through Computational/Mathematical Algorithms and Statistics. In my early twenties, when I was in university studying for my Chemistry-Biotechnology degree, much was talked about the old rivalry between Physicists and Chemists. In times that we could refer to, one was either one or the other. I did not notice it then, but my degree in Biotechnology already relied heavily in the combination of fields: biotechnology, biochemistry, and enzymology are some examples. Today, there are full degrees in cross-functional subjects, such as Bioengineering and Biostatistics. Those cross combinations of fields are everywhere in science today and have become complex; the truly powerful results tend to come from cross-functional collaborations, diverse combinations of brilliant minds that are not only experts in their field but have the emotional mindset and generosity to be able to truly collaborate and trust others, forming scientific bridges and creating access to cutting-edge technologies and large databases of information. This collaborative approach has brought a paradigm shift to the way we do science in the 21st century, and it is only hindered by the often-limited access to patented technology and for-profit industry breakthroughs.

Chapter 2. Molecular Biology Tools for studies in *Mycobacterium tuberculosis* Virulence

1. Introduction

1.1. Tuberculosis

Tuberculosis (TB) is an infectious disease, and one of the oldest recorded human diseases. According to the World Health Organization, TB has recently been one of the top ten causes of death worldwide, and the leading cause of death from a single infectious agent [100]. In 2019, 1.4 million people died from TB [101].

TB is acquired by inhalation of infectious aerosol particles. It is estimated that about a quarter of the world's population is infected with the disease [100]. Most individuals mount an effective response in the lungs leading to a dormant form of the disease - latent tuberculosis; these individuals show no symptoms and do not transmit the disease to others. However, about 5 to 10 percent of the infected individuals eventually progress from latency to an active state. Immunocompromised individuals have a higher risk of developing active TB, and the prevalence of the active form of the disease is strongly associated with poor social and health conditions [100,102,103].

The predominant form of TB affects the lungs and is referred to as pulmonary TB. However, in 5-20% of the cases, infection may develop in other body sites and lead to Pott's disease (spinal and bone deformities), scrofula (swelling of the lymph nodes in the neck), meningitis (central nervous system), or other less prevalent forms of extrapulmonary TB [104].

The German microbiologist Robert Koch isolated the tubercle bacillus in the late 1800s and reproduced the disease by inoculating it into laboratory animals, announcing to the Society of Physiology in Berlin on the 24th of March of 1882 that the bacillus *Mycobacterium tuberculosis* (Mtb) is the cause of tuberculosis. This discovery triggered many developments in the following decades such as the Pirquet and Mantoux tuberculin skin tests, the *Bacillus* Calmette-Guérin (BCG) vaccine, and anti-tuberculosis drugs such as streptomycin. Koch was awarded the Nobel prize in Medicine in 1905 [105,106].

1.2. History of the tuberculosis disease

TB is an ancient disease. Egyptian mummies that are 4,400 years-old show skeletal deformities typical of tuberculosis, and abnormalities like Pott's lesions are clearly illustrated in early Egyptian art [105]. Written descriptions of humans with TB-like symptoms date as far back as 3,300 years ago in writings from India, persisting through recorded ancient history in 700 B.C. clay tablets from the Assyrian Empire, in Hippocrates's detailed accounts of phthisis (an ancient name for tuberculosis) in 500 B.C. Greece, and in 2300 year-old writings from China [104,105].

Recent genomic findings indicate that TB is even older than we thought. Several publications since 2005 have challenged a prior belief that TB originated in animals and was transferred to humans 10,000 years ago, during the Neolithic transition. A study by Gutierrez *et al.*, using DNA sequence data of six housekeeping genes, proposed that an ancient strain, an ancestor of Mtb, would have originated about 3 million years ago in East Africa [107]. Their conclusions have been contested [108] but multiple other evidence summarized in 2012 by Gagneux [109] points to a pre-historic African origin, as well as a host-pathogen coevolution. It is worth mentioning that publications from 2008-2009 discuss a 500,000-year-old fossil of *Homo erectus* discovered in Turkey with bone lesions that may be attributed to TB, albeit among other causes, raising questions around the possibility of TB infections predating modern humans [110,111].

Throughout history, TB remained sporadic up to the 18th century; then it started to increase due to higher population density, unfavorable living conditions and the industrial revolution. By the 19th century, then known as consumption, tuberculosis became epidemic, accounting for nearly 25% of all deaths in London for about two decades in the early 1800s [112,113]. Consumption affected especially infants and young adults, which exhibited the familiar symptoms of chronic cough, fever, weight loss and extreme paleness; tuberculosis is widely mentioned, sometimes romanticized, in the art and literature of the time (example in Figure 2.1), giving rise to tragic heroines such as Mimi in Puccini's La Bohème who dies of consumption in the opera's last scenes.

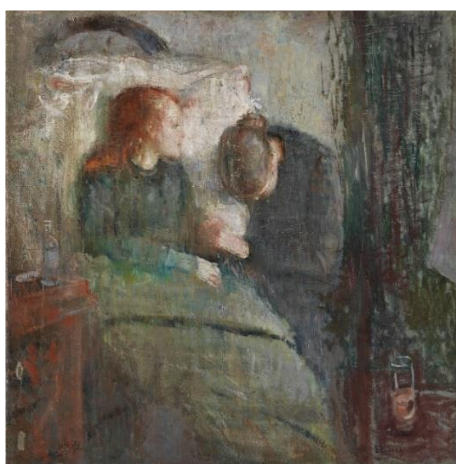

Figure 2.1. Image of the painting The Sick Child
By Edvard Munch, 1885–86 [114].

During the 20th century, the incidence of TB started to decrease rapidly in developed countries thanks to improved nutrition, housing conditions and overall health. The introduction of the BCG vaccine in 1921 and the use of antibiotics, such as streptomycin in 1943, isoniazid in 1952, and rifampicin in 1963, greatly contributed to the decrease [103]. However, in the 1980s, the incidence in TB increased again to

concerning levels due to the HIV pandemic, the deterioration of health conditions in large cities and the appearance of resistance to antibiotics [103].

1.3. Tuberculosis in the world today

TB is a major public health problem worldwide, with a natural reservoir in 25-30% of the world's population with latent forms of the disease, and increasingly sparse treatment options due to the global pervasiveness of multi-drug resistant strains. At a crowded news conference in London on April 23rd, 1993, the WHO declared a global emergency for the first time in history, unleashing a series of media coverage and increasing public awareness for TB and drug resistance [115].

An estimated 10 million people fell ill with TB in 2018, and about 1.2 million died, as reported by the WHO in 2019. Most TB cases occur in developing countries, where conditions are poor, and treatment is difficult. Graphic illustrations of TB prevalence during the year of 2018 are shown in Figure 2.2. Prevalence is highly associated HIV and AIDS - among all TB cases in 2018, 8.6% were people living with HIV [100].

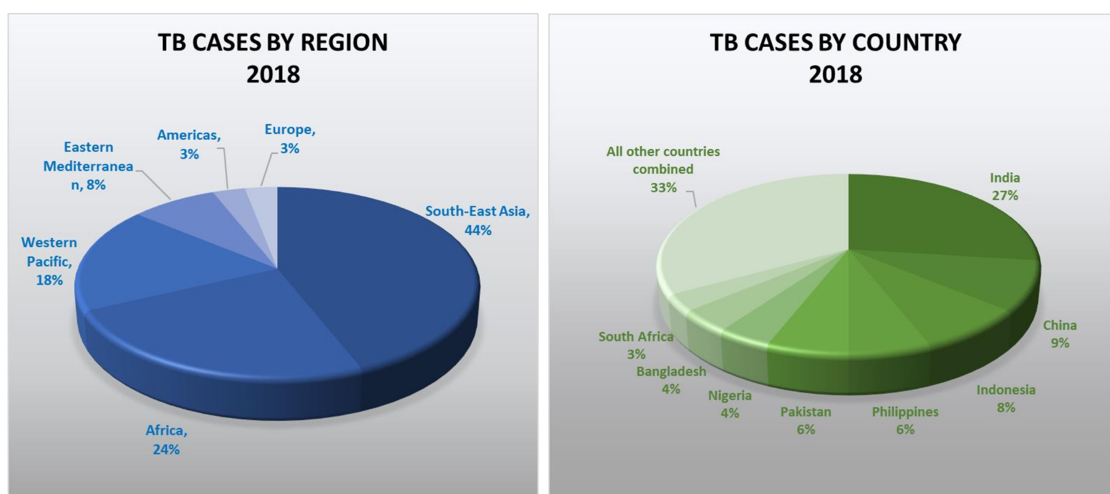

Figure 2.2. Graphs showing the prevalence of TB cases in 2018
Graphs generated with data obtained from the WHO Global TB 2019 Report [100].

Leaders of all member states of the United Nations (UN) have committed to ending the global TB epidemic by 2030. Progress is being made and some countries are on track to reach this goal; however, the worldwide pace towards milestones is slow. Additional resources are needed, and the topic, as of the release of the Global Tuberculosis WHO Report in 2019, was highlighted for discussion at the UN General Assembly in 2020 [100]. The Global Tuberculosis WHO Report of October 2020 again confirms that several milestones have been met, and the annual TB incidence rate and number of deaths are falling globally (9% and 14% respectively in 2019), but they are not declining fast enough to reach the 2015 to 2020 reduction milestones of 20% for incidence rate and 35% for number of deaths [116].

In 2020, TB continues to pose a huge public health risk, potentially made worse with increasing concerns in the context of the increased burden of Covid-19 on hospitals and access to medical care, and potential coinfection with SARS-CoV-2 [116–118].

1.4. The bacillus *Mycobacterium tuberculosis*

Mtb is an obligate aerobic and facultative intracellular bacillus (Figure 2.3). With humans as a single reservoir, Mtb is usually found in large numbers in pulmonary cavities and carried by alveolar macrophages to the host's lymph nodes [119,120].

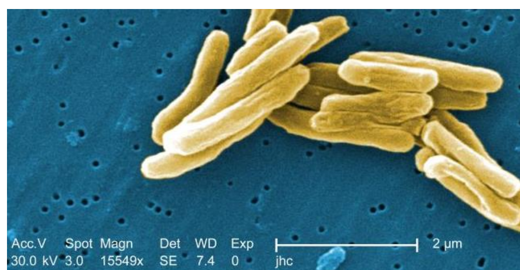

Figure 2.3. Electron micrograph of *Mycobacterium tuberculosis*
Magnification 15549X. Image adapted from Todar, 2015 [14].

Mtb and other mycobacteria have a cell wall which includes peptidoglycan, such as typical Gram-positive bacteria, but it also contains a high content of complex lipids, including mycolic acids, cord factor and wax-D, which are significant determinants of Mtb virulence. These lipids do not stain well with Gram-stain, and therefore mycobacteria have traditionally been identified through Zielh-Neelsen's staining method and classified as acid-fast, instead of Gram. The high concentration of lipids in the cell wall of Mtb have also been associated with impermeability, resistance to certain chemical treatments and survival inside of macrophages [14].

Mtb cultures are slow growing, with generation times of approximately 20 hours in laboratory media and taking 3 to 8 weeks to show visible colonies on agar plates [119,120]. As shown in Figure 2.4, Mtb colonies have a waxy appearance due to the high levels of mycolic acids [14].

Figure 2.4. Colony morphology of *Mycobacterium tuberculosis*
Showing the characteristically rough and waxy colony appearance. Image adapted from Kubica, 1976 [121].

The genome of Mtb has been known for its high GC content. In 1998, S. T. Cole *et al*. [122] published the complete genome sequence of the H37Rv, the strain of choice for laboratory Mtb studies, confirming the GC content as 65.6% and reporting that the genome is comprised of 4,411,529 base pairs and encoding approximately 4,000 genes, many devoted to the production of enzymes involved in lipid metabolism [122].

1.5. The *Mycobacterium tuberculosis* complex

It is now known that tuberculosis affects several mammalian hosts and is caused by a group of phylogenetically closely related bacteria – the *Mycobacterium tuberculosis* complex (MTBC) [109,122].

In humans, TB is caused primarily by *M. tuberculosis* and *M. africanum*, with the latter having a higher prevalence in West Africa. Other members of the MTBC include *M. bovis* (cattle), *M. caprae* (sheep and goats), *M. microti* (voles) and *M. pinnipedii* (seals and sea lions), expanding to include additional species frequently and to varying degrees in the literature. These infectious agents do not transfer disease easily from host-to-host outside of their preferred host species, suggesting a strong adaptation of the pathogen to the host; however, it should be noted that Mtb has been isolated from various animal species and *M. bovis* can cause disease in humans, albeit with no host-to-host transmission documented [109].

MTBC member species display different phenotypic characteristics and affect various mammalian hosts, but they represent one of the most extreme examples of genetic homogeneity, with rare nucleotide variation and no significant trace of genetic exchange among them [107,109,122]. This low level of genetic variation suggests that the entire population may have developed from a common ancestor [107,109].

M. bovis isolated from a sick cow was used by Albert Calmette and Camille Guerin to generate a non-virulent variant, through 230 laboratory passages, yielding the Bacillus Calmette-Guerin (BCG). BCG was used for the first time in humans in 1921 as a TB vaccine and the studies were published in 1931 [123]. Molecular studies confirmed later that significant portions of the *M. bovis* genome are missing in BCG [124].

Because of the risk of infection, the culture of Mtb is classified as a Biosafety-level-3 (BSL-3), requiring special air filtration and extensive gowning and ungowning procedures for laboratory access. In addition, Mtb is a slow growing bacterium, taking 3-8 weeks to yield a visible colony. Therefore, molecular studies of the MTBC have largely used a non-virulent and fast-growing laboratory model as a surrogate: *Mycobacterium smegmatis*. This species shares more than 2000 homologous genes with Mtb, has the same cell wall structure as species in the MTBC, and is easily manipulated with plasmids and mobile genetic elements by transformation [125].

2. Tools to Study Mtb Virulence

2.1. Background

Much of the work carried-on in the PHRI laboratory where I worked from 1994 to 1997 under the direction, mentorship, and extraordinary inspiration of Dr. Eugenie Dubnau, had the ultimate goal of searching for Mtb genes involved in virulence. It typically involved cloning mycobacterial genes from Mtb or from *M. bovis* BCG into mycobacterial plasmid vectors and observing their effect when overexpressed in mycobacterial hosts. The team had expertise in the molecular biology tools that were then relevant, in the culture of mycobacteria, and access to a Biosafety-level 3 laboratory where infectious cultures of Mtb could safely be propagated. My work centered around genetic manipulations, cloning, and generation of genomic libraries, and was enhanced by collaborations with expertise in complementary subjects.

For example, in the work reported in 1997 in the journal Molecular Microbiology [126], I was involved in the isolation and sequence of a cluster of four genes from BCG coding for methyl transferases, and subcloned it in *M. smegmatis*. Collaborators complemented the work by confirming biochemically that the gene cluster was involved in the synthesis of mycolic acids, proposing a model for the pathway of biosynthesis of mycolic acids in BCG.

In another example, I generated a genomic library of Mtb that was used for selection in macrophages [127] and mouse lungs [128] by collaborators, where the identification of promoters specifically upregulated during infection provided clues to genes potentially involved in virulence. The construction of the library in the promoter trap and work in macrophages is described in detail in the sections below.

2.2. Construction of an Mtb library in a promoter trap

I used an *in vivo* expression technology (IVET) system [129–131] in my hunt for Mtb genes involved in virulence, that could later be used as targets for diagnosis and treatment of TB. I created a promoter trap to select for Mtb genes specifically upregulated during growth in human macrophages; selection was based upon increased expression of *inhA*, driven by promoters specifically upregulated during infection.

The gene *inhA* from *M. smegmatis* had been previously cloned and studied [132]. It codes for an enoyl-ACP reductase, which is required for mycolic acid biosynthesis and is a major target for the antibiotic isoniazid in Mtb. Overproduction of this enzyme confers resistance to isoniazid in *M. smegmatis* [132] and, as we suspected and were able to confirm during this study, in Mtb as well.

For the construction of the **promoter trap**, I started by inserting a promoterless PCR-amplified *M. smegmatis inh*A gene [132] into pYUB378, an *Escherichia coli* - *Mycobacterium* shuttle vector conferring

kanamycin resistance in both organisms [133]. The resulting plasmid, pJD32, had a unique BamHI restriction site upstream of the promoterless *inh*A gene which was used for further constructs. The approximate structure of the pJD32 construct is illustrated in Figure 2.5. This plasmid was also used in our studies as a **negative control**.

To create a **positive control**, I cloned the *hsp60* promoter (*Phsp*60) from *M. bovis* BCG into the BamHI site (Figure 2.5). This promoter, driving the production of a highly conserved essential protein in BCG, is expressed constitutively at high levels when present on a multicopy plasmid [134]. The resulting construct, pJD33 is therefore expected to confer isoniazid resistance to its *Mycobacterium* hosts in any condition.

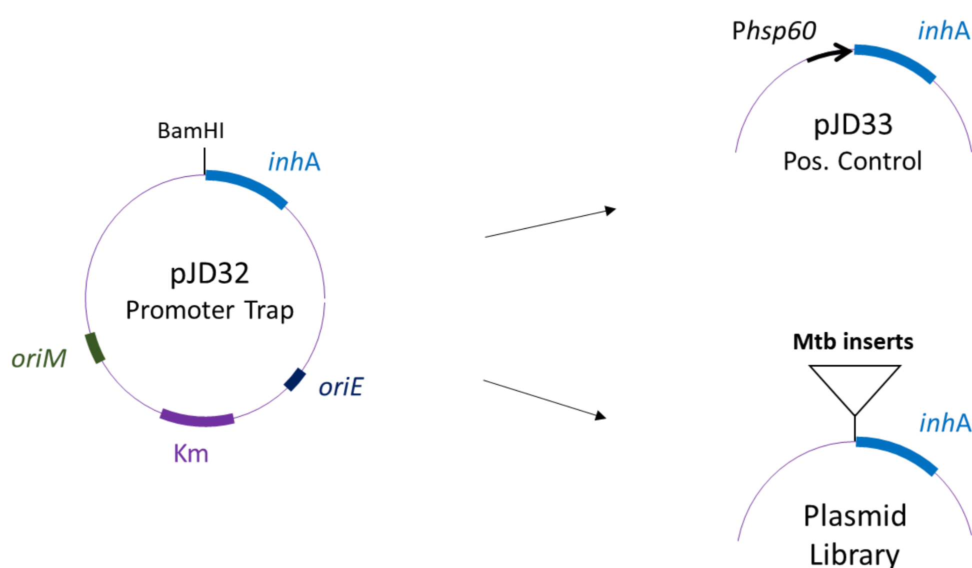

Figure 2.5. Schematic depiction of the promoter trap constructs
Promoter trap constructs were used in the strategy to identify Mtb genes upregulated in human macrophages. BamHI is the unique restriction site selected for cloning. *Inh*A is a gene from *M. smegmatis*. *Phsp*60 is the *hsp*60 constitutive promoter from BCG. Location of the other genetic elements, *ori*E (origin of plasmid replication in *E. coli*), *ori*M (origin of plasmid replication in mycobacteria) and Km (Kanamycin resistance), is shown for illustrative purposes only and may not be accurate. Description of the cloning activities and relevant genetic element functions are provided in the main text.

I built a **library** of Mtb genomic DNA in the pJD32 promoter trap, to later screen and select for virulent genes. I started by culturing the H37Rv strain of Mtb in 7H9 broth (supplemented with albumin, dextrose, NaCl, and Tween 80) and incubating on a rotator at 37°C [135]. H37Rv is a virulent strain, so all H37Rv propagations were carried in the BSL-3 laboratory. I purified genomic DNA from H37Rv and subjected it to a partial digestion with the SauIIIA restriction endonuclease, obtaining an assortment of small DNA fragments. These were then cloned into the BamHI site of pJD32 (Figure 2.5) in *Escherichia coli* (*E. coli*). The resulting plasmid library of 10^6 clones in *E. coli* was estimated by PCR analysis as containing

an average insert size of 300 bp (between 100 and 500 bp), at a minimum insertion rate of 62%. The library was therefore estimated to contain 186 Mb of Mtb genomic DNA. For an organism with a genome of 4.4 Mb, this library in *E. coli* contained a genomic over-representation greater than 42-fold.

The plasmid library, as well as the positive and negative controls, were then used to independently transform Mtb H37Rv via electroporation [135]. Initial selection for the plasmid was done using kanamycin at 10 µg/mL in broth culture. To verify the quality of the library in Mtb, an aliquot of the transformation mixture was plated, and a total of about 9 x 10^4 Mtb clones were obtained. Since the insertion frequency was 62% with an average insertion size of 300 bp, this library was calculated to represent the entire genome of Mtb almost four times (16.7 Mb of the 4.4-Mb genome).

Since the goal was to use the plasmid library to identify Mtb promoters, the H37Rv library was verified for the frequency of promoter inserts by plating with selection for isoniazid at 1 µg/mL (a high concentration, to identify strong promoters). It was expected to be low (due to the need for correct orientation upstream of *inhA*) and determined to be 0.01% - this is presumably the rate of clones containing a strong promoter upstream of *inhA* which was active when Mtb was grown on plates.

2.3. Verification of performance of controls

Strain H37Rv of Mtb, carrying pJD32 (promoterless *inh*A, negative control) or pJD33 (*inh*A driven by P*hsp*60, positive control) was cultured with and without isoniazid, in broth culture and in THP-1 human macrophage cells by my colleagues.

As expected, the culture carrying the negative control (pJD32) remained sensitive to isoniazid, and this culture did not grow in broth or macrophages in the presence of the antibiotic.

The positive control (pJD33) conferred resistance to isoniazid to Mtb as previously reported for *M. smegmatis* [132]. This phenotype was confirmed in both broth and macrophages, and results are shown in Figure 2.6. These results show that as previously reported for *M. smegmatis* [132], overproduction of the InhA protein confers isoniazid resistance to Mtb.

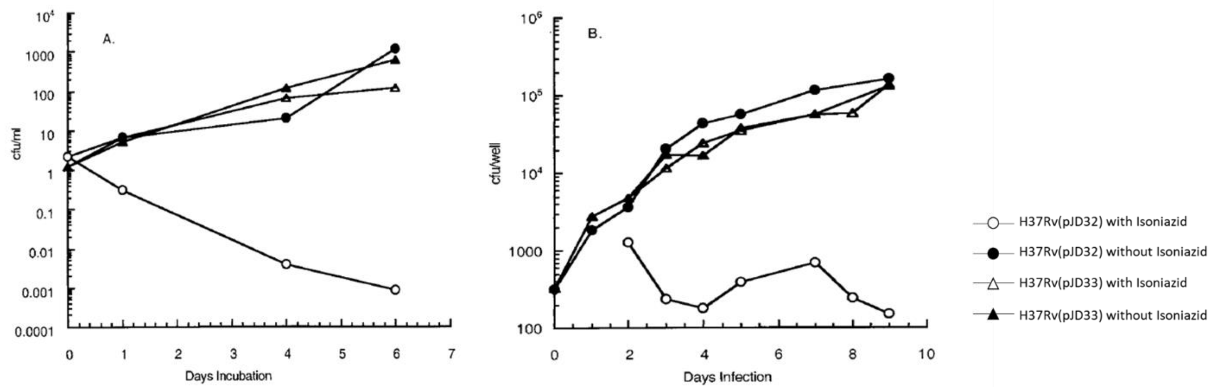

Figure 2.6. Growth of Mtb H37RV carrying the promoter trap constructs

Growth of H37RV carrying the negative control plasmid pJD32 and the positive control plasmid pJD33 in broth (A) and in THP-1 human macrophages (B) as determined by optical density of the culture. As expected, the promoterless *inh*A negative control (pJD32) is sensitive to isoniazid in both growth conditions, and the *inh*A gene preceded by the constitutive promoter (pJD33) confers resistance to isoniazid in both growth conditions. It should be noted that Mtb carrying the plasmid with no promoter driving *inhA* (pJD32) is more sensitive to isoniazid in broth culture (A) than in THP-1 macrophage cells (B). Figure replicated from Dubnau *et al*., 2002 [127].

In conclusion, the promoter trap conferred isoniazid resistance through overexpression of *inhA* only in the presence of a promoter; it was therefore expected to be suitable for the identification of Mtb promoters induced during infection, as intended in later experiments.

2.4. Selection of isoniazid resistant Mtb clones during infection of macrophages

A culture from the library was used to infect THP-1 cells that had been differentiated into macrophage-like cells by collaborators.

Clones that survived a 5-day treatment with isoniazid at 0.5 µg/mL were isolated, diluted into broth, used to reinfect THP-1 cells, and subjected to isoniazid treatment again. Passaging of the library in this manner was repeated for a total of four rounds. In between each round, clones were screened individually for complete sensitivity to isoniazid on agar plates.

Clones that survived isoniazid treatment in macrophages and were isoniazid sensitive on plates, and that were confirmed to contain DNA inserts upstream of *inh*A, as determined by direct PCR analysis, were selected.

The inserts in these plasmids were sequenced and compared to the Mtb H37Rv genome database. After the four rounds of selection for expression in macrophages, approximately two thirds of the 65 clones contained sequences 5' to an open reading frame in the correct orientation for transcription of *inhA*, suggesting the presence of promotors driving the expression of *inh*A during infection of macrophages. A list of the 43 open reading frames found downstream from these putative promoter sequences is shown in Table 2.1.

Table 2.1. Genes identified by isoniazid selection in macrophages
Table replicated from Dubnau *et al*., 2002 [127]

Category	Gene	Rv no.[a]	Comment
Fatty acid degradation	*aceA*	Rv0467	Isocitrate lyase
	fadA4	Rv1323	Acetyl-coenzyme A acetyltransferase
	echA19	Rv3516	Enoyl-coenzyme A hydratase
	fadA5	Rv3546	Acetyl-coenzyme A acetyltransferase
Possible fatty acid metabolism	*ephF*	Rv0134	Epoxide hydrolase
		Rv0610c	Monooxygenase
		Rv1144	Alcohol dehydrogenase
		Rv1774	Oxidoreductase
Cell envelope		Rv0102	Membrane
		Rv1171	Hydrophobic protein
		Rv2120c	Membrane
	lppM	Rv2171	Lipoprotein
		Rv3237c	Potassium channel
		Rv3524	Membrane, sensor
		Rv3717	*N*-Acetyl-muramoyl-*L*-alanine amidase
Intermediary metabolism	*pckA*	Rv0211	Phosphoenol carboxykinase
	eno	Rv1023	Enolase
PPE/PEPGRS		Rv0977	PE/PGRS
		Rv1361c	PPE
		Rv1840c	PE/PGRS
Putative transcriptional regulators		Rv0549c	Helix-turn-helix motif
		Rv2009	Helix-turn-helix motif
		Rv3321c	Helix-turn-helix motif
Miscellaneous	*fusA2*	Rv0120c	Elongation factor G
	proC	Rv0500	Proline biosynthesis
	uvrC	Rv1420	Exonuclease ABC
	infC	Rv1641	Initiation factor 3
		Rv2224c	Exported protease
	nirA	Rv2391	Nitrite reductase
		Rv2520c	Transmembrane domain
	dut	Rv2697c	Deoxyuridine triphosphatase
		Rv3225c	Aminoglycoside 3′ phosphotransferase
Unknown		Rv0036c, Rv0406c, Rv0811c, Rv1778c, Rv2273, Rv2468c, Rv2632c, Rv2717c, Rv3427c, Rv3493c, Rv3717	

[a] The Rv number is the number assigned to the open reading frame, and the genes are annotated as described by the Pasteur Institute at the TUBERCULIST website (http://genolist.pasteur.fr/TubercuList).

Differential expression between agar culture and macrophages was verified for wild-type Mtb H37Rv through RT-PCR, for 13 of the 43 putative promotors identified by the selection method. Eight of the 13 loci/genes tested were confirmed to be upregulated more than two-fold during growth in macrophage cells relative to growth in broth culture (Rv2224c, Rv3237c, Rv3321c, Rv2520c, *pckA*, *echA19*, *fadA4*, and *aceA*), but the other 5 appeared to be equally expressed or down-regulated in macrophages. These results confirm that the promoter trap system for macrophage selection was an efficient strategy in the identification of candidate Mtb genes involved specifically in macrophage survival, and potentially virulence, but also that further validation methods are necessary to confirm the differential expression.

This work was published in 2002 in Infection and Immunity, under the title "*Mycobacterium tuberculosis* Genes Induced during Infection of Human Macrophages"[127].

3. Conclusions

My contribution to this work was to help engineer and construct a molecular biology tool, the promoter trap, and then build an Mtb genomic library, to enable the selection of potentially virulent genes in macrophage studies performed by my colleagues. The invention of the promoter trap is credited to my phenomenal mentor at the PHRI, Dr. Eugenie Dubnau, who was the principal investigator and fundamentally responsible for the architecture of this strategy.

The promoter trap technology that we used - in a multicopy plasmid - can result in artifacts because it does not necessarily reflect normal gene expression from the chromosome. Other restrictions may also have limited our results, such as the selection of clones that were resistant to isoniazid in macrophages but completely sensitive *in vitro* (in other words, we selected for on/off promoters and not exactly for increased expression). However, using this tool, my colleagues and I were able to identify eight loci/genes, many of which were found to be conserved in infectious species of *Mycobacterium*, beneficial and sometimes essential for infection, or involved in pathways deemed critical for survival in the host (see section 4.1. Further findings).

When witnessed through the lens of the technology resources that are available to researchers today (see section 4.2. Gene expression profiling), the strategy that we developed in the mid-nineties was extremely manual and slow, but I am incredibly grateful for having been a part of it. The use of promotor traps and IVET systems was new at the time. These techniques were early models of bioengineering, examples of applied biotechnology at its best, and the predecessors of amazing architectures in molecular biology that are used today in broader applications, such as adenoviral vector systems engineered to deliver vaccines and genetic therapies that save human lives [136], or CRISPR-Cas systems, tailored from prokaryotic adaptive immune systems into an incredibly powerful toolbox for eukaryotic applications, from genome editing, to RNA targeted diagnostics [137].

4. Further Considerations

4.1. Further findings

The Mtb genomic library that I built went on to support additional research findings. In a posterior study using this promoter trap library, an additional 56 genes were identified from Mtb that are preferentially expressed in live mouse lungs. Some of them were confirmed to be differentially expressed through quantitative real-time PCR [128].

One of the 8 genes identified to be up-regulated in macrophages, ***fadA4***, which codes for Acetyl-CoA acetyltransferase, was also in the group of genes identified in the mouse lung study [128], strengthening the notion that it may be involved in host infection and virulence.

Rv2224c was found to be conserved between *Mycobacterium* species *M. leprae*, *M. bovis* and *M. avium paratuberculosis*, and to encode a probable secreted protease associated with the cell envelope by comparative studies [138]. It was confirmed to be critical for Mtb virulence: disruption of Rv2224c compromised the intracellular survival of Mtb in macrophages, increased its susceptibility to lysozyme, and also led to prolonged survival and highly reduced lung pathology of infected mice [139,140].

Rv3237c encodes a conserved protein with some similarity to putative potassium channel proteins from other species. It was found to be a non-essential gene for *in vitro* growth by analysis of saturated Himar1 transposon mutagenesis [139], but no other information was found in the literature.

The locus **Rv3321c** is now known as gene *vap*B44 [141] encoding antitoxin VapB44. Although not proven to be essential for infection, this protein has been associated with a group of toxin-antitoxin (TA) modules, which are bi-cistronic operons that may act in stress physiology by serving as metabolic regulators of growth. Mtb possesses an unusually large and diverse complement of TA modules [142,143].

Rv2520c is conserved in mycobacterial strains and predicted to encode an integral membrane protein [139].

pckA codes for phosphoenolpyruvate carboxykinase (PEPCK). Deletion of the *pckA* gene of *M. bovis* BCG led to a reduction in the capacity of the bacteria to infect and survive in macrophages, and mice infected with BCG carrying the deletion were able to reduce the bacterial load much more effectively than mice infected with the parental wild-type bacteria [144]. This attenuated virulence of the BCG deletion mutant is in line with the promoter trap Mtb library results, supporting the hypothesis that *pckA* is involved in Mtb virulence.

echA19 codes for a putative enoyl-CoA hydratase, whose absence partially inhibits the process of degradation of cholesterol and β-sitosterol. Studies show that cholesterol may play an important role in the host-pathogen interaction and Mtb pathogenesis [145].

The probable isocitrate lyase encoded by ***aceA*** may play a part in the Glyoxylate cycle. Other isocitrate lyases have been described with roles in this cycle, and found to be essential for host persistence and virulence of Mtb [139,146,147].

In summary, all eight genes show some indication of importance for Mtb, from simply encoding conserved proteins to having been confirmed to be essential for virulence.

4.2. Gene expression profiling

The promoter trap with isoniazid resistance is a tool for gene expression profiling from a time when large data analysis was not readily available to researchers - selection through antibiotic resistance provided the means for downsizing the number of results to only relevant hits, but it carries potential collateral artifacts stemming from overexpression on a multicopy plasmid.

Today, gene expression profiling tools are used to screen differential expression in large-scale. RNA sequencing (RNA-seq) has facilitated genome-wide expression profiling, including the identification of novel and rare transcripts like noncoding RNAs and novel alternative splicing isoforms [148]. Other methods rely on cDNA arrays and real-time quantitative RT-PCR.

These methods are based on previously described technologies (section 6.2. Evolution of typing techniques), but their novelty is that through the aid of computational analysis and biostatistics, differential expression can be determined without the use of *in vivo* artifacts and in large scale – the whole transcriptome can be analyzed in a single setting. The techniques unequivocally underline the perspective I highlighted in the last paragraph of section 6.2 in Chapter 1 – they are built on the bridging of technologies and combination of knowledge across many fields, including Engineering, Computational Sciences, Statistics, Biology, Chemistry and Physics.

4.3. The Public Health Research Institute

The PHRI is now part of the New Jersey Medical school and Rutgers University, and continues to pursue its research mission to help eliminate infectious diseases from spreading across the planet [149]. Standing out from many exceptional discoveries at the PHRI, are the invention of molecular beacon probes, now widely used in multiplex real-time PCR, which were developed while I worked there by cherished colleagues on the same floor [150]. The PHRI and Rutgers University faculty also invented and co-developed the Cepheid Mtb detection assay for commercialization. Both are included in the next chapter of this report.

The four years I worked at the PHRI were fundamental in my development, my scientific approach to problem solving, and ultimately greatly impacted the direction of my career and who I am at my core today. The environment was extraordinary for any young scientist – the floor was brewing with amazing mentors, who provided an umbrella of scientific knowledge, a high standard of ethical principles, friendship, camaraderie, and were models of collaboration for the greater good of science. My work was incredibly interesting and motivating, and it taught me a lot: about biology at a very fundamental level, how to methodically control experiments, how to stay focused on a long-term goal through a day-to-day labyrinth of details and plans, and how to maintain my scientific and ethical principles in the forefront of everything I do. This gem in Manhattan, the PHRI, then housed in the New York City Health Department at 455 First Avenue, provided me with four years that shaped me forever.

Chapter 3. Molecular Diagnostic Assays

1. Background

From 2004 to 2019 I worked in the Product Transfer Department at Cepheid, Sunnyvale, California, USA.

The products that Cepheid develops and produces are individual cartridges that integrate cell lysis, nucleic acid purification, Real-Time PCR, fluidics, software, and data analysis algorithms in order to detect (and sometimes quantify) the presence of specific nucleic acids in human samples. The products are aimed at aiding medical practitioners in the diagnosis of diseases or conditions.

My work involved the transfer of products from Research and Development (R&D) to production and post-launch improvements: establishing the production line processes, designing and validating the quality control methods and specifications, performing hands-on failure investigations, as well as leading cross-functional teams with various skills from molecular biology to software integration and automation line engineering.

The XpertMTB assay line includes a few molecular diagnostic kits options which, in a single step, can detect *Mycobacterium tuberculosi*s and characterize antibiotic resistance patterns in patient sputum samples, aiding doctors in the appropriate and effective selection of treatments.

Cepheid's assays are part of a connected system where the reagents are contained in a cartridge that performs sample preparation as well as real-time PCR detection. This cartridge works when placed in the GeneXpert – a family of instrument systems that is designed to provide real-time PCR at varying scales, from point-of-care detection in doctor's offices to large volume laboratories.

The work that I performed to commercialize and consistently produce the many sub-types of the XpertMTB and other Cepheid assays has many elements that I will not disclose, respecting Cepheid's confidentiality. However, the real-time PCR functionality of the XpertMTB assays was published by their inventors [97–99], and I will use the public elements to describe it and share my role in the company.

2. Introduction

2.1. Antibiotic resistance and Mtb

The Mtb bacillus has developed multiple mechanisms enabling antibiotic resistance since the use of antibiotics became widespread in the mid-20th century, unfortunately a prevalent reality common to most infectious bacterial agents including, as previously mentioned, *S. pneumoniae*.

During the 1990s, multidrug-resistant tuberculosis (MDR-TB) emerged worldwide as a threat to TB control; **MDR-TB** is defined as a form of TB that is resistant to at least the two front line antibiotic TB treatments: isoniazid and rifampicin [151,152].

In the 2000s, extensively drug-resistant tuberculosis (XDR-TB) emerged. **XDR-TB** is a form of TB which is resistant to at least four of the core anti-TB drugs: isoniazid, rifampicin, any of the fluoroquinolones (such as levofloxacin or moxifloxacin) and to at least one of the three injectable second-line drugs (amikacin, capreomycin or kanamycin) [152]. By 2016, XDR-TB had been reported in 123 member states of the World Health Organization, with data suggesting that about 6.2% of MDR-TB cases worldwide are XDR-TB. XDR-TB patients can be cured, but with the current drugs available, the likelihood of success is much smaller than in patients with ordinary TB or even MDR-TB. Cure depends on the extent of the drug resistance, the severity of the disease and whether the patient's immune system is compromised [152].

Methods to rapidly identify Mtb and characterize the antibiotic resistance pattern in each patient are important from an individual health perspective, and critical from a public health standpoint.

2.2. Rifampin resistance and MDR-TB

Rifampicin monoresistance is extremely rare, and usually preceded by development of resistance to isoniazid, so rifampicin resistance is used as the marker for identification of MDR-TB [97,153–156].

It has been shown that 95 to 98% of rifampicin resistant strains have mutations, insertions or deletions in an 81-bp core region of the *rpoB* gene, named the rifampicin resistance determining region (RRDR), encoding a specific 27 amino-acid sequence of the β-subunit of the RNA polymerase [153,156]. These are often point mutations in specific codons [156]. In contrast, drug susceptible strains always display the wild-type sequence in this region [157].

For these reasons, the *rpoB* gene, and specifically the RRDR, is an ideal target for molecular screening of MDR-TB.

2.3. Real-time PCR assays

Real-time PCR is a powerful technology, at present widely used for many applications including the identification of organisms, as detailed in Chapter 1, section 6.2. Evolution of typing techniques.

Real-time PCR gets its name for the ability to monitor the PCR amplification of a targeted DNA template during the reaction, *i.e.*, in real time.

Early methods of real-time PCR included the use of intercalation dyes such as SYBR Green, which became fluorescent upon the non-specific incorporation into any double-stranded DNA molecule. By monitoring the increase in fluorescence throughout the PCR reaction, one can follow the increase in the amount of double-stranded PCR product. This method has some problems in that it relies solely on the specificity of the primers to identify the correct target.

Improved methods more frequently used currently rely on the use of sequence-specific DNA probes, which add to the specificity of the PCR primers by matching the target. These short nucleotide probes typically contain a mechanism that will greatly increase the amount of fluorescence emitted only if DNA amplification of the specific target occurs.

There are numerous types of probes, many including a fluorophore chemical component in one end of the probe and a fluorescence quencher on the other end. In a **molecular beacon** [150], the probe is designed to include complementary nucleotides on each end and form a hairpin that brings the quencher close to the fluorophore, greatly reducing the emission of fluorescence; in the presence of target amplified from the primers, the probe is designed to selectively hybridize to the target, breaking the harpin and therefore separating the quencher from the fluorophore, inducing a dramatic increase in the amount of fluorescence emitted (Figure 3.1).

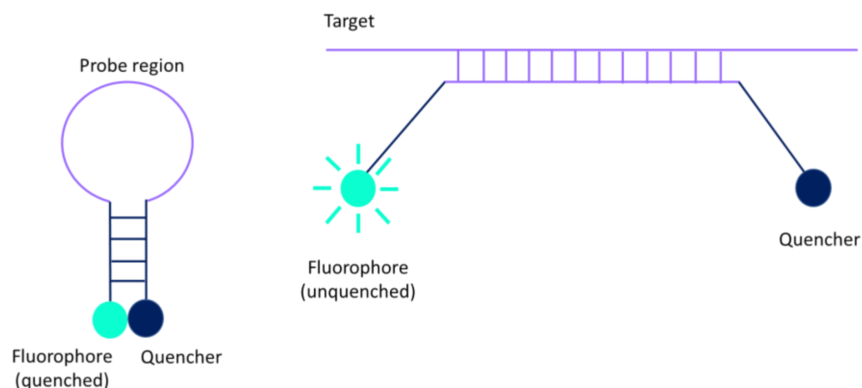

Figure 3.1. Illustration of the mechanism of Molecular Beacon probes
Adapted from Tyagi and Kramer, 1996 [150]

The use of different fluorophores (with separate fluorescence emission wavelengths) attached to different specific probes, lets several different PCR reactions be monitored in the same reaction tube. This powerful technique called multiplexing, allows for the simultaneous detection of different target sequences in a sample, and for the design and inclusion of internal controls in an amplification.

The device needed for a real-time PCR reaction is a PCR thermal cycler equipped with a detection system that can read fluorescence throughout the reaction. The typical unit of measure of real time PCR is Ct, which stands for Cycle threshold, and represents the PCR cycle number at which the fluorescence crossed a nominal threshold level. This threshold level varies; it is assigned to represent the level of fluorescence that is minimally sufficient to distinguish a negative reaction (one where the probe does not detect the target) from a positive reaction (where the probe detects the target).

In a well-designed conventional or real-time PCR system, with balanced melting temperatures of primers and probes, optimum buffer conditions for the DNA polymerase, and optimal temperatures and cycle durations, mathematical perfection can be approached, and amplification of the target is precise. The higher the concentration of target in the reaction at the starting point, the earlier the threshold will be reached. Real-time PCR has the ability to quantitate the initial amount of target material added to the reaction because it inversely correlates to the Ct. Mathematically, a 10-fold increase in the amount of template results in a 3.3 decrease in the Ct (within a serial dilution range for the template). Dose response curves and internal controls are some of the methods used to facilitate quantitative real-time PCR.

The high accuracy, high sensitivity, high specificity, low turn-around-time, and low cost of a real-time PCR reaction make it a perfect candidate for use in multiple applications. Furthermore, real-time PCR systems may be designed to discern point mutations. For these reasons, real-time PCR is widely used in the molecular diagnostics field, and I will describe below an elegant system in use worldwide for the detection of Mtb present in patient sputum samples.

3. MTB/RIF Assay

3.1. Mechanism of detection of Mtb and rifampin resistance

Several publications described assays designed to allelic discriminate mutations in the *rpoB* gene sequence, and were validated in population studies as highly predictable of MDR-TB [154,155,157]. Later on, Piatek *et al*. developed a rapid closed-tube real-time PCR assay using molecular beacons to detect *rpoB* mutations [158,159]. Molecular beacons are able to detect amplicons as they are synthesized during real-time PCR and can discriminate between DNA sequences that differ from one another by as little as a single nucleotide substitution [158,160]. Five allele discriminating molecular beacons were designed against the RRDR of the *rpoB* gene [159].

Each of the five molecular beacon probes contained a 15-20 nucleotide sequence designed to collectively span the entire wild-type *rpoB* RRDR 81-bp region, with overlapping sequences of 1-3 bases. The probes detected the wild-type version of the gene and not a mutant with a single base change [159].

To increase the specificity (minimize cross-amplification of non-tuberculosis *Mycobacterium* species) and to maximize mutation detection, the sequences of the primers and probes were modified by Helb *et al.* [97] to incorporate a heminested PCR - two successive reactions where the product of the first PCR reaction is used as template for the second reaction, and a third primer is designed to match a binding site nested within the first amplicon [97].

Bacillus globigii spores, as well as an assay designed to target their DNA with a sixth molecular beacon probe, are also included in the Xpert MTB/RIF cartridge to serve as an internal control for sample processing and PCR amplification [97,161].

Fluorescent dyes and quenchers were developed to enable all six molecular beacons to be multiplexed within the same reaction [97,162].

3.2. Sample processing and integration into the GeneXpert

The GeneXpert instrument system is a software-driven microfluidics-based cartridge processor and integrated fluorescence-based quantitative thermal cycler that processes biological samples to nucleic acids and performs real-time PCR [161].

The Xpert cartridge consists of multiple chambers that are designed to hold all reagents, buffers, and enzymes, to process the biological sample, and to retain all waste. The cartridge has an attached tube where after sample processing, the PCR reaction occurs with real-time fluorescence detection [161].

3.3. Automated result analysis

At the end of the real-time PCR, the data analysis algorithm integrates the results of the test and the control. Samples that are negative for Mtb and positive for the control *B. globigii* are reported as negative; samples that are negative for both Mtb and *B. globigii* are reported as invalid [97].

A sample is reported as Mtb positive if at least two of the five *rpoB* probes are positive and within 2 Ct of each other. Rifampin resistance is detected by the failure of one or more of the *rpoB* probes to hybridize properly to the *rpoB* amplicon. This is defined by a greater than 3.5 Ct difference between the earliest and latest *rpoB* probe signals, or if one to three *rpoB* probes did not produce any measurable signal at all [97].

3.4. Assay performance

Assay performance was evaluated for sensitivity, specificity, and the ability to detect *rpo*B mutations [97].

Sensitivity was assessed by evaluating the limit of detection (LOD) of the assay for purified Mtb genomic DNA, as well as for Mtb cells spiked in clinical sputum. LOD is defined as the amount of target material required for 95% probability of detection. Results showed an LOD of 4.5 genomes per reaction, and 131 colony forming units (CFU) per mL of sputum [97].

The specificity of the assay was confirmed by testing the ability to detect mutations in the *rpoB* gene associated with rifampin resistance, including DNA samples from an assortment of 6 wildtype (susceptible) and 23 RRDR mutant (resistant) clinical isolates, or artificial full-length double-stranded nucleotides, to test all regions of each molecular beacon probe. The results showed that all wild type and mutant samples were correctly identified by the assay [97].

Additional tests were completed to assess and validate the performance of the assay, showing rare cross-reactivity with other mycobacterial species, negligible impact of freeze-thawing the sample prior to performing the assay, and the ability to inactivate live Mtb to levels compliant with international decontamination standards [97]. Clinical studies showed an overall performance close to traditional culture-based methods of Mtb detection, with the advantage of providing results in a few hours instead of weeks [97].

4. Conclusion

4.1. My Role in the molecular diagnostics industry

Even though I was not part of the team that developed the Xpert MTB/RIF assay, I was involved in its development and launch, and had to understand its mechanism and performance. I was responsible for its ongoing quality after launch, any changes and upgrades to the production line and testing procedures, improvements, and validation.

A production line is a vast network with many layers of complexity (Figure 3.2) across space and time.

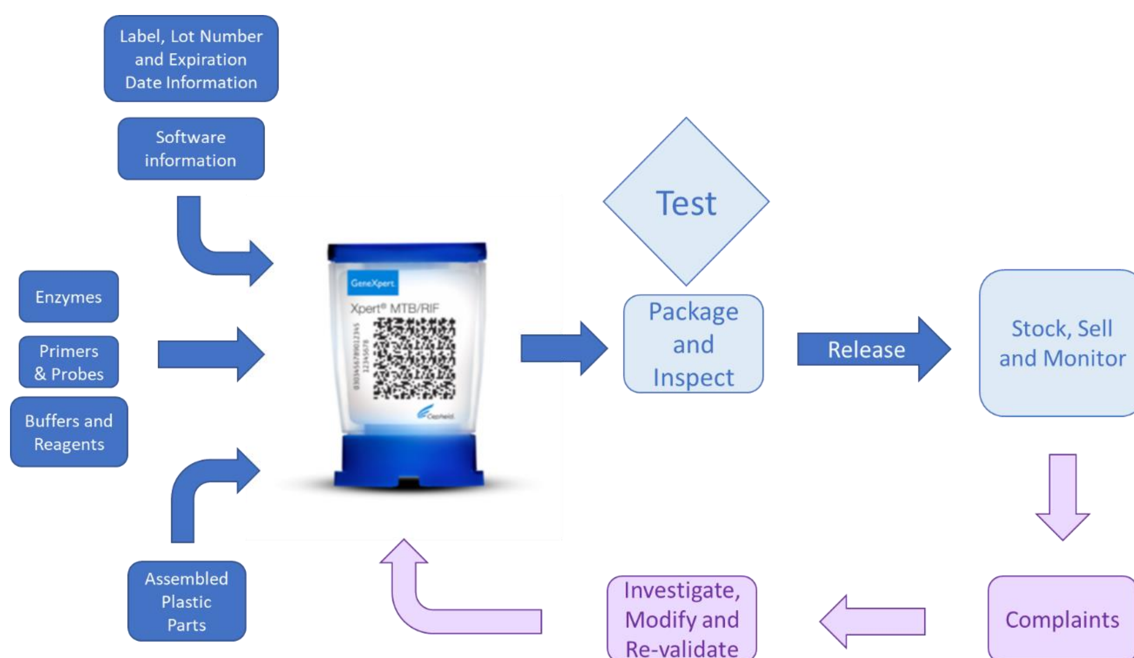

Figure 3.2. Generic overview of an assay production line and lifecycle management

Each lot of a critical raw material, such as an enzyme, needs to be tested when it is received. For that one test only, test methods need to be developed and validated, and materials to be used in those tests need to be ordered and tested as well. The time that the raw material lasts in stock before it is incorporated into the assay needs to be determined, and then monitored. Each time a vendor changes a process or a small part of the raw material, an assessment of risk needs to be made, and re-validation of the assay may be needed.

Processes need to be developed to integrate reagents, cartridge parts, software, and other pieces of information. These processes usually involve multiple departments and handoffs of material or information, which are at a higher risk for errors. Everything needs to be documented rigorously. These processes must be validated for accuracy and reliability and are subject to inspections and audits by regulatory agencies that oversee the quality of products in the medical field. These agencies vary from

country to country, and many have specific and dynamic labeling requirements, which result in strict and complex additional processes associated with labeling and distribution.

Production and testing processes also must be made safe for the operators, as well as efficient and cost effective. Automation of production lines is common; it involves ambitious design and validation projects upfront that reduce manufacturing costs, errors, and safety incidents later.

Many of these processes depart from Biotechnology, but many rely on it and leverage it; so, my team was generally composed of scientists with a deep knowledge of the real-time PCR assay and general molecular biology techniques, balanced with a strong pragmatic inclination for process development. I built and managed a team that was the liaison between all the departments involved with a product's lifecycle and its scientific roots. I was accountable for the development of numerous processes and test methods, countless troubleshooting and improvement projects, and my work was fundamental to the company's strategy and operations for over 15 years - across all of Cepheid's line of over 25 assays, including MTB/RIF.

4.2. Impact of TB molecular diagnostics in the world

The Xpert MTB/RIF assay provided a rapid form of detection of MDR-TB with low-complexity and a sensitivity that approached culture methods, reducing the accurate detection time from weeks to hours [97].

In December 2010, the WHO publicly endorsed the Xpert MTB/RIF rapid-test as "a major milestone for global TB diagnosis and care" [163]. With affordability being a key concern, Cepheid and the Foundation for Innovative and New Diagnostics (FIND) negotiated of a 75% reduction in price for 116 low- and middle-income countries with endemic TB [163].

In 2012 the Bill and Melinda Gates Foundation announced a partnership with three other global humanitarian organizations, jointly providing funds to reduce the costs of the Xpert MTB/RIF test by more than 40% to 145 high-burden and developing countries affected by TB [164].

Diagnosis of TB is on a positive trend. In 2019, 7.1 million people were newly diagnosed with TB globally, up from 7.0 million in 2018 and largely increased from previous years [116]. This is mainly due to intensified efforts around testing. Despite these increases, there is still a large gap (2.9 million in 2019) between the number of people newly diagnosed and the 10 million estimated to have developed TB, due to underreporting and underdiagnosis [116]. Further improvements are needed to close the gap, by reducing underreporting, improving access to diagnosis, and ideally and achieve 100% detection.

Global estimates by the WHO indicate that 10 million people fell ill with TB in 2019, down 2.3% from 2018 [116]. This estimation has been declining consistently in recent years, even if slowly, due to

global efforts that include early and effective detection, and allow patients to start as early as possible on the most effective treatment regimen [116].

Cepheid continues to improve and expand its menu of Mtb diagnostic assays, including improved detection algorithms [98] and expansion to detection of XDR-TB [99], making a significant impact in the positive trends reported by the WHO. By mid-2017, 23 million Cepheid Xpert MTB/RIF tests had been procured in 130 countries [98]. Among countries reporting data to the WHO in 2019, an average of 63% of testing sites used the Xpert MTB/RIF assay, the most commonly used WHO-recommended rapid diagnostic test worldwide [116].

General Conclusion

My great-grandmother, Maria Silvina, died in her early thirties in the countryside of Portugal leaving behind my orphaned grandmother Maria Julia at the age of 11 with little more than a few loving memories. She died in 1918, like many around the word, from the great influenza pandemic, which my grandmother, who lived to see her three granddaughters into adulthood, always called the "*pneumónica*" in Portuguese.

While I was writing this report of my work in the field of infectious lung diseases, another pandemic of viral origin affected the world like the great influenza: Covid-19. Pneumonia and tuberculosis have posed threats to public health for centuries, affecting young and old, rich or poor, and having a penchant for the frail. These bacterial lung infections are still of great relevance today and are of increased concern and when potentially involved in superinfections with viruses such as influenza or, recently, SARS-CoV-2.

My contributions were small. I showed early evidence of the spread of two main clones of *S. pneumoniae* across the globe, identified eight genes potentially involved in virulence of Mtb, and helped assure a steady supply of effective Mtb diagnostics tests, with consistently high quality, and at the lowest possible cost.

But the impact of my work, through the accumulation of other people's efforts across the passing of time, was high. The two pneumococcus clones ended up starting a network for surveillance of pneumococcus spread and a new nomenclature; the eight Mtb genes and their involvement in virulence of TB are relevant for research in global ongoing efforts profoundly needed for treatment; and the Mtb diagnostic tests make a difference every day in developing countries and areas of endemic TB; they enable fast detection and enhance the outcomes of treatment, helping global efforts to eradicate the disease.

Through the stories that I heard so many times from my grandmother, I formed a personal connection with lung diseases and their effects in the world at large and on the life of one crucial 11-year-old girl. My work is a small drop in the fight to improve the lives of human beings around the world. But it may help prevent a child from losing a parent, and that is a source of immense satisfaction for me.

References

1. Morens DM, Taubenberger JK, Fauci AS. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. J Infect Dis [Internet]. 2008 Oct [cited 2020 Jun 1];198(7):962–70. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1086/591708

2. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev [Internet]. 2015 Jul [cited 2020 Jun 1];28(3):871–99. Available from: https://cmr.asm.org/content/28/3/871

3. World Health Organization, Pneumonia Fact Sheet [Internet]. 2019 [cited 2020 Apr 8]. Available from: https://www.who.int/en/news-room/fact-sheets/detail/pneumonia

4. Cucchiari D, Pericàs JM, Riera J, Gumucio R, Md EC, Nicolás D. Pneumococcal superinfection in COVID-19 patients: A series of 5 cases. Med Clínica [Internet]. 2020 Jun [cited 2020 Nov 21];S0025775320303493. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0025775320303493

5. Toombs JM, Van den Abbeele K, Democratis J, Mandal AKJ, Missouris CG. Pneumococcal coinfection in COVID-19 patients. J Med Virol [Internet]. 2020 Jul 17 [cited 2020 Nov 21];jmv.26278. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26278

6. Brooks LRK, Mias GI. *Streptococcus pneumoniae*'s Virulence and Host Immunity: Aging, Diagnostics, and Prevention. Front Immunol [Internet]. 2018 Jun 22 [cited 2020 May 29];9:1366. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2018.01366/full

7. Keller LE, Robinson DA, McDaniel LS. Nonencapsulated *Streptococcus pneumoniae*: Emergence and Pathogenesis. mBio [Internet]. 2016 May 4 [cited 2020 Jun 1];7(2):e01792-15, /mbio/7/2/e01792-15.atom. Available from: https://mbio.asm.org/content/7/2/e01792-15

8. Griffith F. The significance of pneumococcal types. J Hyg (Lond). 1928 Jan;27:113–59.

9. Lederberg J, Gotschlich EC. A Path to Discovery: The Career of Maclyn McCarty. PLoS Biol [Internet]. 2005 Oct 11 [cited 2020 Jun 15];3(10):e341. Available from: https://dx.plos.org/10.1371/journal.pbio.0030341

10. Frederick Griffith [Internet]. Wikipedia. 2020 [cited 2020 Jun 18]. Available from: https://en.wikipedia.org/wiki/Frederick_Griffith

11. Oswald Avery [Internet]. Wikipedia. 2020 [cited 2020 Jun 18]. Available from: https://en.wikipedia.org/wiki/Oswald_Avery

12. Colin Munro MacLeod [Internet]. Wikipedia. 2019 [cited 2020 Jun 18]. Available from: https://en.wikipedia.org/wiki/Colin_Munro_MacLeod

13. Visible Proofs, Forensic Views of the Body, Galleries, Technologies [Internet]. National Library of Medicine, National Institutes of Health. 2014 [cited 2020 Jun 18]. Available from: https://www.nlm.nih.gov/exhibition/visibleproofs/galleries/technologies/dna_image_4.html

14. Todar K. Todar's Online Textbook of Bacteriology [Internet]. 2015 [cited 2020 Apr 15]. Available from: http://textbookofbacteriology.net/index.html

15. Barich D (Editor). MicrobeWiki, *Streptococcus pneumoniae* [Internet]. MicrobeWiki, Kenyon College, Department of Biology. 2016 [cited 2020 Apr 8]. Available from: https://microbewiki.kenyon.edu/index.php/Streptococcus_pneumoniae

16. Identification and Characterization of *Streptococcus pneumoniae* (Chapter 8) [Internet]. Laboratory Methods for the Diagnosis of MeningitisCenters for Disease Control and Prevention. 2016 [cited 2020 Apr 8]. Available from: https://www.cdc.gov/meningitis/lab-manual/chpt08-id-characterization-streppneumo.html

17. Garner DP. Microbiology Nuts & Bolts [Internet]. 2013 [cited 2020 Apr 16]. Available from: http://www.microbiologynutsandbolts.co.uk/the-bug-blog/discombobulated-by-beta-haemolytic-streptococci

18. Jedrzejas MJ. Pneumococcal Virulence Factors: Structure and Function. Microbiol Mol Biol Rev [Internet]. 2001 Jun 1 [cited 2020 Jun 1];65(2):187–207. Available from: https://MMBR.asm.org/content/65/2/187

19. Berry AM, Lock RA, Hansman D, Paton JC. Contribution of autolysin to virulence of *Streptococcus pneumoniae*. Infect Immun [Internet]. 1989 [cited 2020 Jun 17];57(8):2324–30. Available from: https://IAI.asm.org/content/57/8/2324

20. Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, Dahlberg S, Fernebro J, Moschioni M, Masignani V, Hultenby K, Taddei AR, Beiter K, Wartha F, von Euler A, Covacci A, Holden DW, Normark S, Rappuoli R, Henriques-Normark B. A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci [Internet]. 2006 Feb 21 [cited 2020 Jun 1];103(8):2857–62. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0511017103

21. Mitchell AM, Mitchell TJ. *Streptococcus pneumoniae*: virulence factors and variation. Clin Microbiol Infect [Internet]. 2010 May [cited 2020 Jun 1];16(5):411–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1198743X14633115

22. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The Streptococcus pneumoniae Capsule Inhibits Complement Activity and Neutrophil Phagocytosis by Multiple Mechanisms. Infect Immun [Internet]. 2010 Feb [cited 2020 Jun 1];78(2):704–15. Available from: https://IAI.asm.org/content/78/2/704

23. Langereis JD, de Jonge MI. Non-encapsulated *Streptococcus pneumoniae*, vaccination as a measure to interfere with horizontal gene transfer. Virulence [Internet]. 2017 Aug 18 [cited 2020 Jun 1];8(6):637–9. Available from: https://www.tandfonline.com/doi/full/10.1080/21505594.2017.1309492

24. Sabin AB. Immediate Pneumococcus Typing Directly From Sputum By The Neufeld Reaction. JAMA J Am Med Assoc [Internet]. 1933 May 20 [cited 2020 Jun 1];100(20):1584. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1933.02740200018004

25. Andam CP, Hanage WP. Mechanisms of genome evolution of *Streptococcus*. Infect Genet Evol [Internet]. 2015 Jul [cited 2020 Jun 1];33:334–42. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1567134814004109

26. Hui FM, Morrison DA. Genetic transformation in *Streptococcus pneumoniae*: nucleotide sequence analysis shows comA, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. J Bacteriol [Internet]. 1991 [cited 2020 May 29];173(1):372–81. Available from: https://JB.asm.org/content/173/1/372

27. Tomasz A. Antibiotic Resistance in *Streptococcus pneumoniae*. Clin Infect Dis. 1997;(24 (Suppl 1)):S85–8.

28. Dye C, Espinal MA, Watt CJ, Mbiaga C, Williams BG. Worldwide Incidence of Multidrug-Resistant Tuberculosis. J Infect Dis [Internet]. 2002 Apr 15 [cited 2020 Jun 18];185(8):1197–202. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1086/339818

29. Drug Resistance [Internet]. Centers for Disease Control and Prevention. 2020 [cited 2020 Apr 20]. Available from: https://www.cdc.gov/drugresistance/about.html

30. Sauvage E, Terrak M. Glycosyltransferases and Transpeptidases/Penicillin-Binding Proteins: Valuable Targets for New Antibacterials. Antibiotics [Internet]. 2016 Feb 17 [cited 2020 Jun 22];5(1):12. Available from: http://www.mdpi.com/2079-6382/5/1/12

31. Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev [Internet]. 2008 Mar [cited 2020 Jun 22];32(2):234–58. Available from: https://academic.oup.com/femsre/article-lookup/doi/10.1111/j.1574-6976.2008.00105.x

32. Kawai Y, Mickiewicz K, Errington J. Lysozyme Counteracts β-Lactam Antibiotics by Promoting the Emergence of L-Form Bacteria. Cell [Internet]. 2018 Feb [cited 2020 Jun 1];172(5):1038-1049.e10. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867418300588

33. Munoz R, Coffey TJ, Daniels M, Dowson CG, Laible G, Casal J, Hakenbeck R, Jacobs M, Musser JM, Spratt BG, Tomasz A. Intercontinental Spread of a Multiresistant Clone of Serotype 23F *Streptococcus pneumoniae*. J Infect Dis [Internet]. 1991 Aug 1 [cited 2020 May 29];164(2):302–6. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/164.2.302

34. Jacobs MR. Antimicrobial Susceptibility Testing of Pneumococci: Determination of Kirby-Bauer Breakpoints for Penicillin G, Erythromycin, Clindamycin, Tetracycline, Chloramphenicol, and Rifampin. 1979;16:8.

35. Jabes D, Nachman S, Tomasz A. Penicillin-Binding Protein Families: Evidence for the Clonal Nature of Penicillin Resistance in Clinical Isolates of Pneumococci. J Infect Dis [Internet]. 1989;159(1):16–25. Available from: http://www.jstor.org/stable/30136423

36. Williamson R, Hakenbeck R, Tomasz A. In vivo interaction of beta-lactam antibiotics with the penicillin-binding proteins of *Streptococcus pneumoniae*. Antimicrob Agents Chemother [Internet]. 1980 Oct 1;18(4):629. Available from: http://aac.asm.org/content/18/4/629.abstract

37. Appelbaum PC. Antimicrobial Resistance in *Streptococcus pneumoniae*: An Overview. Clin Infect Dis [Internet]. 1992;15(1):77–83. Available from: http://www.jstor.org/stable/4456550

38. Kristinsson KG, Hjálmarsdóttir MA, Steingrímsson O. Increasing penicillin resistance in pneumococci in Iceland. Lancet Lond Engl. 1992 Jun 27;339(8809):1606–7.

39. Dixon JMS, Lipinski, A.E., Graham, M.E.P. Detection and prevalence of pneumococci with increased resistance to penicillin. Can Med Assoc J. 1977 Nov 19;117:1159–61.

40. Zerfaß I, Hakenbeck R, Denapaite D. An Important Site in PBP2x of Penicillin-Resistant Clinical Isolates of *Streptococcus pneumoniae*: Mutational Analysis of Thr338. Antimicrob Agents Chemother [Internet]. 2009 Mar [cited 2020 Jun 23];53(3):1107–15. Available from: https://AAC.asm.org/content/53/3/1107

41. Hakenbeck R, Ellerbrok H, Briese T, Tomasz A. Penicillin-Binding Proteins of Penicillin-Susceptible and -Resistant Pneumococci: Immunological Relatedness of Altered Proteins and Changes in Peptides Carrying the 3-Lactam Binding Site. ANTIMICROB AGENTS CHEMOTHER. 1986;30:6.

42. Hakenbeck R, Brückner R, Denapaite D, Maurer P. Molecular mechanisms of β-lactam resistance in *Streptococcus pneumoniae*. Future Microbiol [Internet]. 2012 Mar [cited 2020 Jun 22];7(3):395–410. Available from: https://www.futuremedicine.com/doi/10.2217/fmb.12.2

43. Hakenbeck R, Denapaite D, Maurer P. Mechanisms of Penicillin Resistance in *Streptococcus pneumoniae*: Targets, Gene Transfer and Mutations. In: Dougherty TJ, Pucci MJ, editors. Antibiotic Discovery and Development [Internet]. Boston, MA: Springer US; 2012 [cited 2020 Jun 22]. p. 593–616. Available from: http://link.springer.com/10.1007/978-1-4614-1400-1_18

44. Munoz R, Musser JM, Crain M, Briles DE, Marton A, Parkinson AJ, Sorensen U, Tomasz A. Geographic Distribution of Penicillin-Resistant Clones of *Streptococcus pneumoniae*: Characterization by Penicillin-Binding Protein Profile, Surface Protein A Typing, and Multilocus Enzyme Analysis. Clin Infect Dis [Internet]. 1992;15(1):112–8. Available from: http://www.jstor.org/stable/4456555

45. Jacobs MR, Koornhof HJ, Robins-Browne RM, Stevenson CM, Vermaak ZA, Freiman I, Miller GB, Witcomb MA, Isaäcson M, Ward JI, Austrian R. Emergence of multiply resistant pneumococci. N Engl J Med. 1978 Oct 5;299(14):735–40.

46. Jacobs MR, Bajaksouzian S, Palavecino-Fasola EL, Holoszyc HM, Appelbaum PC. Determination of Penicillin MICs for *Streptococcus pneumoniae* by Using a Two- or Three-Disk Diffusion Procedure. J Clin Microbiol [Internet]. 1998 [cited 2020 May 29];36(1):179–83. Available from: https://JCM.asm.org/content/36/1/179

47. Mcgee L, Mcdougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W. Nomenclature of Major Antimicrobial-Resistant Clones of *Streptococcus pneumoniae* Defined by the Pneumococcal Molecular Epidemiology Network. J CLIN MICROBIOL. 2001;39:7.

48. Tang Y-W, Li H, Griffin JP, Haas DW, D'Agata EMC. Rapidly increasing prevalence of penicillin-resistant Streptococcus pneumoniae in middle Tennessee: a 10-year clinical and molecular analysis. J Clin Microbiol. 2002 Feb;40(2):395–9.

49. Soares S, Kristinsson KG, Musser JM, Tomasz A. Evidence for the Introduction of a Multiresistant Clone of Serotype 6B *Streptococcus pneumoniae* from Spain to Iceland in the Late 1980s. J Infect Dis [Internet]. 1993 Jul 1 [cited 2020 May 29];168(1):158–63. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/168.1.158

50. Reichler MR, Allphin AA, Breiman RF, Schreiber JR, Arnold JE, McDougal LK, Facklam RR, Boxerbaum B, May D, Walton RO, Jacobs MR. The Spread of Multiply Resistant *Streptococcus pneumoniae* at a Day Care Center in Ohio. J Infect Dis [Internet]. 1992 Dec 1 [cited 2020 May 29];166(6):1346–53. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/166.6.1346

51. Barnes DM, Whittier S, Gilligan PH, Soares S, Tomasz A, Henderson FW. Transmission of Multidrug-Resistant Serotype 23F *Streptococcus pneumoniae* in Group Day Care: Evidence Suggesting Capsular Transformation of the Resistant Strain In Vivo. J Infect Dis [Internet]. 1995 Apr 1 [cited 2020 May 29];171(4):890–6. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/171.4.890

52. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology [Internet]. 1998 Nov 1 [cited 2020 May 29];144(11):3049–60. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-144-11-3049

53. Pneumococcal Molecular Epidemiology Network (PMEN) [Internet]. Pneumococcal Molecular Epidemiology Network (PMEN). [cited 2020 May 10]. Available from: https://www.pneumogen.net/pmen/

54. Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, Sá-Leão R, Song J-H, Ko K, Beall B, Klugman KP, Parkhill J, Tomasz A, Kristinsson KG, Bentley SD. Variable recombination dynamics during the emergence, transmission and ‘disarming’ of a multidrug-resistant pneumococcal clone. BMC Biol [Internet]. 2014 [cited 2020 May 29];12(1):49. Available from: http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-12-49

55. Sá-Leão R, Tomasz A, Sanches IS, Brito-Avô A, Vilhelmsson SE, Kristinsson KG, de Lencastre H. Carriage of Internationally Spread Clones of *Streptococcus pneumoniae* with Unusual Drug Resistance Patterns in Children Attending Day Care Centers in Lisbon, Portugal. J Infect Dis [Internet]. 2000 Oct [cited 2020 May 29];182(4):1153–60. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1086/315813

56. Vilhelmsson SE, Tomasz A, Kristinsson KG. Molecular evolution in a multidrug-resistant lineage of *Streptococcus pneumoniae*: emergence of strains belonging to the serotype 6B Icelandic clone that lost antibiotic resistance traits. J Clin Microbiol. 2000 Apr;38(4):1375–81.

57. Beall B, Chochua S, Gertz RE, Li Y, Li Z, McGee L, Metcalf BJ, Ricaldi J, Tran T, Walker H, Pilishvili T. A Population-Based Descriptive Atlas of Invasive Pneumococcal Strains Recovered Within the U.S. During 2015–2016. Front Microbiol [Internet]. 2018 Nov 19 [cited 2020 May 29];9:2670. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.02670/full

58. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song J-H, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. Rapid Pneumococcal Evolution in Response to Clinical Interventions. SCIENCE. 2011 Jan 28;331:430–4.

59. Hiller NL, Sá-Leão R. Puzzling Over the Pneumococcal Pangenome. Front Microbiol [Internet]. 2018 Oct 30 [cited 2020 May 29];9:2580. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.02580/full

60. US National Library of Medicine, Clinical Trails Identifier NCT03692871 [Internet]. 2019 [cited 2020 May 10]. Available from: https://clinicaltrials.gov/ct2/show/NCT03692871?term=v114&cond=Pneumococcal+Infections&draw=2&rank=4

61. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol [Internet]. 1995 [cited 2020 Jun 24];33(9):2233–9. Available from: https://JCM.asm.org/content/33/9/2233

62. Arbeit R, Arbique JC, Beall, Bernard, Critchley, Ian A., Marsik, Frederic, Michaud, Sophie, Steward, Christine, Tenover, Fred C., Trees, David L. Molecular Methods for Bacterial Strain Typing; Approved Guideline. (Clinical and Laboratory Standards Institute; vols. 27–10).

63. Jeffreys AJ. DNA Sequence Variants in the G-gamma-, A-gamma-, delta- and beta-Globin Genes of Man. Cell. 1979 Sep;18:1–10.

64. Zagorski N. Profile of Alec J. Jeffreys. Proc Natl Acad Sci [Internet]. 2006 Jun 13;103(24):8918. Available from: http://www.pnas.org/content/103/24/8918.abstract

65. Jobling MA. Curiosity in the genes: the DNA fingerprinting story. Investig Genet [Internet]. 2013 Nov 18 [cited 2020 Jun 24];4(1):20. Available from: https://doi.org/10.1186/2041-2223-4-20

66. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic Amplification of ,-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. Science. 1985;230:6.

67. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In: Methods in Enzymology [Internet]. Academic Press; 1987 [cited 2020 Jun 25]. p. 335–50. (Recombinant DNA Part F; vol. 155). Available from: http://www.sciencedirect.com/science/article/pii/0076687987550236

68. van Belkum A, Sluijuter M, de Groot R, Verbrugh H, Hermans PW. Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. J Clin Microbiol [Internet]. 1996 [cited 2020 May 29];34(5):1176–9. Available from: https://JCM.asm.org/content/34/5/1176

69. Koeuth T, Versalovic J, Lupski JR. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. Genome Res [Internet]. 1995 Nov 1 [cited 2020 May 29];5(4):408–18. Available from: http://www.genome.org/cgi/doi/10.1101/gr.5.4.408

70. Brito DA, Ramirez M, de Lencastre H. Serotyping *Streptococcus pneumoniae* by Multiplex PCR. J Clin Microbiol [Internet]. 2003 Jun 1 [cited 2020 May 29];41(6):2378–84. Available from: http://jcm.asm.org/cgi/doi/10.1128/JCM.41.6.2378-2384.2003

71. *Streptococcus* Laboratory, Resources and Protocols [Internet]. Centers for Disease Control and Prevention. 2018 [cited 2020 May 31]. Available from: https://www.cdc.gov/streplab/pneumococcus/resources.html

72. Sá-Leão R, Nunes S, Brito-Avô A, Frazão N, Simões AS, Crisóstomo MI, Paulo ACS, Saldanha J, Santos-Sanches I, de Lencastre H. Changes in pneumococcal serotypes and antibiotypes

carried by vaccinated and unvaccinated day-care centre attendees in Portugal, a country with widespread use of the seven-valent pneumococcal conjugate vaccine. Clin Microbiol Infect [Internet]. 2009 Nov [cited 2020 May 29];15(11):1002–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1198743X14617948

73. Nunes S, Sá-Leão R, Carriço J, Alves CR, Mato R, Avô AB, Saldanha J, Almeida JS, Sanches IS, de Lencastre H. Trends in drug resistance, serotypes, and molecular types of *Streptococcus pneumoniae* colonizing preschool-age children attending day care centers in Lisbon, Portugal: a summary of 4 years of annual surveillance. J Clin Microbiol. 2005 Mar;43(3):1285–93.

74. Gherardi G, Whitney CG, Facklam RR, Beall B. Major Related Sets of Antibiotic-Resistant Pneumococci in the United States as Determined by Pulsed-Field Gel Electrophoresis and pbp1a-pbp2b-pbp2x-dhf Restriction Profiles. J Infect Dis [Internet]. 2000 Jan 1 [cited 2020 Jun 24];181(1):216–29. Available from: https://academic.oup.com/jid/article/181/1/216/893205

75. Bonofiglio L, Gardella, Noella, Mollerach, Marta. Gel Electrophoresis: Advanced Techniques. Application of Molecular Typing Methods to the Study of Medically Relevant Gram-Positive Cocci (Chapter 7). Magdeldin S, editor. BoD – Books on Demand; 2012. 516 p.

76. Coffey TJ, Dowson CG, Daniels M, Zhou J, Martin C, Spratt BG, Musser JM. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. Mol Microbiol [Internet]. 1991 Sep [cited 2020 Jun 24];5(9):2255–60. Available from: http://doi.wiley.com/10.1111/j.1365-2958.1991.tb02155.x

77. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A [Internet]. 1977 Dec [cited 2020 Jun 24];74(12):5463–7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765/

78. The Human Genome Project [Internet]. NIH - National Human Genome Research Institute. 2019 [cited 2020 May 3]. Available from: https://www.genome.gov/human-genome-project

79. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med [Internet]. 2001 Apr 10 [cited 2020 Jun 24];114(4–5):204–10. Available from: http://link.springer.com/10.1007/s004149900117

80. Chanock S. Candidate Genes and Single Nucleotide Polymorphisms (SNPs) in the Study of Human Disease. Dis Markers [Internet]. 2001 [cited 2020 Jun 24];17(2):89–98. Available from: http://www.hindawi.com/journals/dm/2001/858760/abs/

81. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet. 1999 Jul;22(3):239–47.

82. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet [Internet]. 2014 Sep [cited 2020 Jun 24];30(9):418–26. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168952514001127

83. Rhoads A, Au KF. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics [Internet]. 2015 Oct [cited 2020 Jun 24];13(5):278–89. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1672022915001345

84. Illumina, Inc. Illumina Sequencing by Synthesis, Video [Internet]. [cited 2019 Jun 24]. Available from: https://www.youtube.com/watch?time_continue=5&v=fCd6B5HRaZ8&feature=emb_logo

85. Illumina, Inc. An Introduction to Next-Generation Sequencing Technology [Internet]. [cited 2019 Jun 24]. Available from: www.illumina.com/technology/next-generation-sequencing.html

86. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics [Internet]. 2016 Jan [cited 2020 May 29];107(1):1–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0888754315300410

87. Natera [Internet]. Natera, Inc. [cited 2020 May 4]. Available from: https://www.natera.com/our-technology

88. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald... LA. Complete Genome Sequence of a Virulent Isolate of *Streptococcus pneumoniae*. Sci New Ser [Internet]. 2001;293(5529):498–506. Available from: http://www.jstor.org/stable/3084097

89. Ehrlich GD, Hu FZ, Shen K, Stoodley P, Post JC. Bacterial Plurality as a General Mechanism Driving Persistence in Chronic Infections: Clin Orthop [Internet]. 2005 Aug [cited 2020 May 29];NA;(437):20–4. Available from: http://gateway.ovid.com/ovidweb.cgi?T=JS&PAGE=crossref&AN=00003086-200508000-00005

90. Hiller NL, Janto B, Hogg JS, Boissy R, Yu S, Powell E, Keefe R, Ehrlich NE, Shen K, Hayes J, Barbadora K, Klimke W, Dernovoy D, Tatusova T, Parkhill J, Bentley SD, Post JC, Ehrlich GD, Hu FZ. Comparative Genomic Analyses of Seventeen *Streptococcus pneumoniae* Strains: Insights into the Pneumococcal Supragenome. J Bacteriol [Internet]. 2007 Nov 15 [cited 2020 May 29];189(22):8186–95. Available from: https://JB.asm.org/content/189/22/8186

91. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Hotopp JCD, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. 2010;19.

92. Streptococcus pneumoniae MLST Home Page [Internet]. [cited 2020 Jun 24]. Available from: https://pubmlst.org/spneumoniae/

93. Solis BioDyne | HOT FIREPol Multiplex qPCR Mix [Internet]. [cited 2020 Jun 29]. Available from: https://www.solisbiodyne.com/EN/product/name=5x-HOT-FIREPol-Multiplex-qPCR-Mix&catno=08-03-0000S

94. Tavares DA, Handem S, Carvalho RJ, Paulo AC, de Lencastre H, Hinds J, Sá-Leão R. Identification of *Streptococcus pneumoniae* by a real-time PCR assay targeting SP2020. Sci Rep [Internet]. 2019 Dec [cited 2020 May 29];9(1):3285. Available from: http://www.nature.com/articles/s41598-019-39791-1

95. World Health Organization. PCR for Detection and Characterization of Bacterial Meningitis Pathogens: *Neisseria meningitidis*, *Haemophilus influenzae*, and *Streptococcus pneumoniae*

(Chapter 10). In: Laboratory Methods for the Diagnosis of Meningitis [Internet]. 2nd ed. 2011 [cited 2020 Jun 25]. Available from: https://www.cdc.gov/meningitis/lab-manual/chpt10-pcr.html

96. Carvalho M da GS, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, Steigerwalt A, Whaley M, Facklam RR, Fields B, Carlone G, Ades EW, Dagan R, Sampson JS. Evaluation and improvement of real-time PCR assays targeting *lyt*A, *ply*, and *psa*A genes for detection of pneumococcal DNA. J Clin Microbiol. 2007 Aug;45(8):2460–6.

97. Helb D, Jones M, Story E, Boehme C, Wallace E, Ho K, Kop J, Owens MR, Rodgers R, Banada P, Safi H, Blakemore R, Lan NTN, Jones-López EC, Levi M, Burday M, Ayakaka I, Mugerwa RD, McMillan B, Winn-Deen E, Christel L, Dailey P, Perkins MD, Persing DH, Alland D. Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. J Clin Microbiol. 2010 Jan;48(1):229–37.

98. Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, Banada PP, Deshpande S, Shenai S, Gall A, Glass J, Krieswirth B, Schumacher SG, Nabeta P, Tukvadze N, Rodrigues C, Skrahina A, Tagliani E, Cirillo DM, Davidow A, Denkinger CM, Persing D, Kwiatkowski R, Jones M, Alland D. The New Xpert MTB/RIF Ultra: Improving Detection of *Mycobacterium tuberculosis* and Resistance to Rifampin in an Assay Suitable for Point-of-Care Testing. mBio. 2017 29;8(4).

99. Chakravorty S, Roh SS, Glass J, Smith LE, Simmons AM, Lund K, Lokhov S, Liu X, Xu P, Zhang G, Via LE, Shen Q, Ruan X, Yuan X, Zhu HZ, Viazovkina E, Shenai S, Rowneki M, Lee JS, Barry CE, Gao Q, Persing D, Kwiatkawoski R, Jones M, Gall A, Alland D. Detection of Isoniazid-, Fluoroquinolone-, Amikacin-, and Kanamycin-Resistant Tuberculosis in an Automated, Multiplexed 10-Color Assay Suitable for Point-of-Care Use. J Clin Microbiol. 2017;55(1):183–98.

100. World Health Organization. Global tuberculosis report 2019. Geneva: World Health Organization; 2019.

101. WHO. Tuberculosis [Internet]. World Health Organization. 2020 [cited 2020 Nov 21]. Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis

102. Fogel N. Tuberculosis: A disease without boundaries. Tuberculosis [Internet]. 2015 Sep [cited 2020 May 29];95(5):527–31. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1472979214206950

103. Bañuls A-L, Sanou A, Van Anh NT, Godreuil S. *Mycobacterium tuberculosis*: ecology and evolution of a human bacterium. J Med Microbiol [Internet]. 2015 Nov 1 [cited 2020 May 29];64(11):1261–9. Available from: https://www.microbiologyresearch.org/content/journal/jmm/10.1099/jmm.0.000171

104. Smith I. *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. Clin Microbiol Rev [Internet]. 2003 Jul [cited 2020 May 29];16(3):463–96. Available from: https://CMR.asm.org/content/16/3/463

105. Barberis I, Bragazzi NL, Galluzzo L, Martini M. The history of tuberculosis: from the first historical records to the isolation of Koch’s bacillus. 2017;4.

106. Keshavjee S, Farmer PE. Tuberculosis, Drug Resistance, and the History of Modern Medicine. N Engl J Med [Internet]. 2012 Sep 6 [cited 2020 May 29];367(10):931–6. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMra1205429

107. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, Supply P, Vincent V. Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. Ramakrishnan L, editor. PLoS Pathog [Internet]. 2005 Aug 19 [cited 2020 May 29];1(1):e5. Available from: https://dx.plos.org/10.1371/journal.ppat.0010005

108. Smith NH. A Re-Evaluation of M. prototuberculosis. Manchester M, editor. PLoS Pathog [Internet]. 2006 Sep 29 [cited 2020 May 29];2(9):e98. Available from: https://dx.plos.org/10.1371/journal.ppat.0020098

109. Gagneux S. Host–pathogen coevolution in human tuberculosis. Philos Trans R Soc B Biol Sci [Internet]. 2012 Mar 19 [cited 2020 May 29];367(1590):850–9. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0316

110. Roberts CA, Pfister L, Mays S. Letter to the editor: Was tuberculosis present in *Homo erectus* in Turkey? Am J Phys Anthropol [Internet]. 2009 Jul [cited 2020 Sep 4];139(3):442–4. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.21056

111. Kappelman J, Alçiçek MC, Kazancı N, Schultz M, Özkul M, Şen Ş. FirstHomo erectus from Turkey and implications for migrations into temperate Eurasia. Am J Phys Anthropol [Internet]. 2008 Jan [cited 2020 Sep 4];135(1):110–6. Available from: http://doi.wiley.com/10.1002/ajpa.20739

112. Wilson LG. Commentary: Medicine, population, and tuberculosis. Int J Epidemiol [Internet]. 2005 Jun 1 [cited 2020 Sep 4];34(3):521–4. Available from: http://academic.oup.com/ije/article/34/3/521/682217/Commentary-Medicine-population-and-tuberculosis

113. Murray JF. Willing Suspension of Disbelief: New Roles for Violetta and Mimi. Am J Respir Crit Care Med [Internet]. 2013 Mar [cited 2020 May 29];187(5):465–7. Available from: http://www.atsjournals.org/doi/abs/10.1164/rccm.201212-2247OE

114. The Sick Child (Munch) - Wikipedia [Internet]. Wikipedia. Available from: https://en.wikipedia.org/wiki/The_Sick_Child_(Munch)

115. Klaudt K. TB, A Global Emergency. World Health Organization; 1994. Report No.: WHO/TB/94.177.

116. World Health Organization. Global tuberculosis report 2020 [Internet]. Geneva: World Health Organization; 2020. Available from: https://www.who.int/teams/global-tuberculosis-programme/data

117. Crisan-Dabija R, Grigorescu C, Pavel C-A, Artene B, Popa IV, Cernomaz A, Burlacu A. Tuberculosis and COVID-19: Lessons from the Past Viral Outbreaks and Possible Future Outcomes. Can Respir J [Internet]. 2020 Sep 8 [cited 2020 Nov 21];2020:1–10. Available from: https://www.hindawi.com/journals/crj/2020/1401053/

118. Kumar DrR, Bhattacharya DrB, Meena DrV, Soneja DrM, Wig DrN. COVID-19 and TB co-infection - “Finishing touch”’ in perfect recipe to 'severity’ or ‘death’. J Infect [Internet]. 2020

Sep [cited 2020 Nov 21];81(3):e39–40. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0163445320304436

119. Koch A, Mizrahi V. Mycobacterium tuberculosis. Trends Microbiol [Internet]. 2018 Jun [cited 2020 Sep 4];26(6):555–6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0966842X18300489

120. Fitzgerald DW, Sterling TR, Haas DW. Mycobacterium tuberculosis (251). In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases [Internet]. 8th ed. 2015. Available from: https://doi.org/10.1016/B978-1-4557-4801-3.00251-4

121. Kubica G. Public Health Image Library (PHIL), Centers for Disease Control and Prevention [Internet]. 1976. Available from: https://phil.cdc.gov/details.aspx?pid=4428

122. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Iii CEB, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, ..., Whitehead S, Barrel BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998 Jun 11;393:27.

123. Calmette A. Preventive Vaccination against Tuberculosis with BCG. Proc R Soc Med [Internet]. 1931 Sep [cited 2020 Sep 27];24(11):1481–90. Available from: http://journals.sagepub.com/doi/10.1177/003591573102401109

124. Liu J, Tran V, Leung AS, Alexander DC, Zhu B. BCG Vaccines: Their mechanisms of attenuation and impact on safety and protective efficacy. Hum Vaccin [Internet]. 2009 Feb [cited 2020 Sep 27];5(2):70–8. Available from: http://www.tandfonline.com/doi/abs/10.4161/hv.5.2.7210

125. *Mycobacterium smegmatis* - Wikipedia [Internet]. Available from: https://en.wikipedia.org/wiki/Mycobacterium_smegmatis

126. Dubnau E, Lanéelle M, Soares S, Bénichou A, Vaz T, Promé D, Promé J, Daffé M, Quémard A. *Mycobacterium bovis* BCG genes involved in the biosynthesis of cyclopropyl keto- and hydroxy-mycolic acids. Mol Microbiol [Internet]. 1997 Jan [cited 2020 May 29];23(2):313–22. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1997.2301589.x

127. Dubnau E, Fontán P, Manganelli R, Soares-Appel S, Smith I. *Mycobacterium tuberculosis* Genes Induced during Infection of Human Macrophages. Infect Immun [Internet]. 2002 Jun [cited 2020 May 29];70(6):2787–95. Available from: https://IAI.asm.org/content/70/6/2787

128. Dubnau E, Chan J, Mohan VP, Smith I. Responses of Mycobacterium tuberculosis to Growth in the Mouse Lung. Infect Immun [Internet]. 2005 Jun [cited 2020 Nov 8];73(6):3754–7. Available from: https://IAI.asm.org/content/73/6/3754

129. Mahan, Michael J., Slauch, James M., Mekalanos, John J. Selection of bacterial virulence genes that are specifically induced in host tissues. Science. 1993 Jan 29;259.

130. Mahan MJ, Tobias JW, Slauch JM, Hanna PC, Collier RJ, Mekalanos JJ. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. Proc Natl Acad Sci [Internet]. 1995 Jan 31 [cited 2020 Oct 11];92(3):669–73. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.92.3.669

131. Chiang SL, Mekalanos JJ, Holden DW. In Vivo Genetic Analysis of Bacterial Virulence. Annu Rev Microbiol [Internet]. 1999 Oct [cited 2020 Oct 11];53(1):129–54. Available from: http://www.annualreviews.org/doi/10.1146/annurev.micro.53.1.129

132. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um K, Wilson T, Collins D, de Lisle G, Jacobs W. inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science [Internet]. 1994 Jan 14 [cited 2020 Oct 9];263(5144):227–30. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.8284673

133. Dubnau E, Soares S, Huang TJ, Jacobs WR. Overproduction of mycobacterial ribosomal protein S13 induces catalase/peroxidase activity and hypersensitivity to isoniazid in *Mycobacterium smegmatis*. Gene [Internet]. 1996 Apr [cited 2020 May 29];170(1):17–22. Available from: https://linkinghub.elsevier.com/retrieve/pii/0378111995008411

134. Stover, CK, de la Cruz, VF, Fuerst, TR, Burlein, JE, Benson, LA, Bennett, LT, Bansal, GP, Young, JF, Lee, MH, Hatfull, GF, Snapper, SB, Barletta, RG, Jacobs, WR, Bloom, BR. New use of BCG for recombinant vaccines. Nature. 1991 Jun 6;351:456–60.

135. Jacobs WR, Kalpana GV, Cirillo JD, Pascopella L, Snapper SB, Udani RA, Jones W, Barletta RG, Bloom BR. [25] Genetic systems for mycobacteria. In: Methods in Enzymology [Internet]. Elsevier; 1991 [cited 2020 Oct 9]. p. 537–55. Available from: https://linkinghub.elsevier.com/retrieve/pii/007668799104027L

136. Wold W, Toth K. Adenovirus Vectors for Gene Therapy, Vaccination and Cancer Gene Therapy. Curr Gene Ther [Internet]. 2014 Jan 31 [cited 2020 Nov 7];13(6):421–33. Available from: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1566-5232&volume=13&issue=6&spage=421

137. Pickar-Oliver A, Gersbach CA. The next generation of CRISPR–Cas technologies and applications. Nat Rev Mol Cell Biol [Internet]. 2019 Aug [cited 2020 Nov 9];20(8):490–507. Available from: http://www.nature.com/articles/s41580-019-0131-5

138. Ribeiro-Guimarães ML, Pessolani MCV. Comparative genomics of mycobacterial proteases. Microb Pathog [Internet]. 2007 Nov [cited 2020 Nov 8];43(5–6):173–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0882401007000691

139. Kapopoulou A, Lew JM, Cole ST. Mycobrowser [Internet]. The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. Tuberculosis (Edinb). Jan 91(1):8-13 (2011). 2011 [cited 2020 Nov 8]. Available from: https://mycobrowser.epfl.ch/

140. Rengarajan J, Murphy E, Park A, Krone CL, Hett EC, Bloom BR, Glimcher LH, Rubin EJ. Mycobacterium tuberculosis Rv2224c modulates innate immune responses. Proc Natl Acad Sci [Internet]. 2008 Jan 8 [cited 2020 Nov 8];105(1):264–9. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0710601105

141. antitoxin VapB44 [Mycobacterium tuberculosis H37Rv] - Protein - NCBI [Internet]. [cited 2020 Nov 8]. Available from: https://www.ncbi.nlm.nih.gov/protein/NP_217838.1

142. Ramage HR, Connolly LE, Cox JS. Comprehensive Functional Analysis of Mycobacterium tuberculosis Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. Rosenberg SM, editor. PLoS Genet [Internet]. 2009 Dec 11 [cited 2020 Nov 8];5(12):e1000767. Available from: https://dx.plos.org/10.1371/journal.pgen.1000767

143. Ahidjo BA, Kuhnert D, McKenzie JL, Machowski EE, Gordhan BG, Arcus V, Abrahams GL, Mizrahi V. VapC Toxins from Mycobacterium tuberculosis Are Ribonucleases that Differentially Inhibit Growth and Are Neutralized by Cognate VapB Antitoxins. Ahmed N, editor. PLoS ONE [Internet]. 2011 Jun 29 [cited 2020 Nov 8];6(6):e21738. Available from: https://dx.plos.org/10.1371/journal.pone.0021738

144. Liu K, Yu J, Russell DG. pckA-deficient Mycobacterium bovis BCG shows attenuated virulence in mice and in macrophages. Microbiology [Internet]. 2003 Jul 1 [cited 2020 Nov 8];149(7):1829–35. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.26234-0

145. Wrońska N, Brzostek A, Szewczyk R, Soboń A, Dziadek J, Lisowska K. The Role of fadD19 and echA19 in Sterol Side Chain Degradation by Mycobacterium smegmatis. Molecules [Internet]. 2016 May 6 [cited 2020 Nov 8];21(5):598. Available from: http://www.mdpi.com/1420-3049/21/5/598

146. Muñoz-Elías EJ, McKinney JD. M. tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. 2006;16.

147. Gould TA, van de Langemheen H, Munoz-Elias EJ, McKinney JD, Sacchettini JC. Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis. Mol Microbiol [Internet]. 2006 Aug [cited 2020 Nov 8];61(4):940–7. Available from: http://doi.wiley.com/10.1111/j.1365-2958.2006.05297.x

148. Arrigoni A, Ranzani V, Rossetti G, Panzeri I, Abrignani S, Bonnal RJP, Pagani M. Analysis RNA-seq and Noncoding RNA. In: Lanzuolo C, Bodega B, editors. Polycomb Group Proteins: Methods and Protocols [Internet]. New York, NY: Springer; 2016 [cited 2020 Nov 9]. p. 125–35. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-4939-6380-5_11

149. PHRI [Internet]. [cited 2020 Nov 9]. Available from: https://phri.njms.rutgers.edu/

150. Tyagi S, Kramer FR. Molecular Beacons: Probes that Fluoresce upon Hybridization. Nat Biotechnol [Internet]. 1996 Mar [cited 2020 Nov 9];14(3):303–8. Available from: http://www.nature.com/articles/nbt0396-303

151. Centers for Disease Control and Prevention (CDC). Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. MMWR Morb Mortal Wkly Rep [Internet]. 2006 Mar 24;55(11):301–5. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5511a2.htm

152. World Health Organization. Drug-resistant TB: XDR-TB FAQ. Available from: https://www.who.int/tb/areas-of-work/drug-resistant-tb/xdr-tb-faq/en/

153. Mokrousov I, Otten T, Vyshnevskiy B, Narvskaya O. Allele-Specific rpoB PCR Assays for Detection of Rifampin-Resistant Mycobacterium tuberculosis in Sputum Smears. Antimicrob Agents Chemother [Internet]. 2003 Jul [cited 2020 Nov 14];47(7):2231–5. Available from: https://AAC.asm.org/content/47/7/2231

154. Prammananan T, Cheunoy W, Taechamahapun D, Yorsangsukkamol J, Phunpruch S, Phdarat P, Leechawengwong M, Chaiprasert A. Distribution of rpoB mutations among multidrug-resistant Mycobacterium tuberculosis (MDRTB) strains from Thailand and development of a rapid method for mutation detection. Clin Microbiol Infect [Internet]. 2008 May [cited 2020

Nov 14];14(5):446–53. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1198743X1460937X

155. Traore H, Fissette K, Bastian I, Devleeschouwer M, Portaels F. Detection of rifampicin resistance in Mycobacterium tuberculosis isolates from diverse countries by a commercial line probe assay as an initial indicator of multidrug resistance. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2000 May;4(5):481–4.

156. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance inMycobacterium tuberculosis: 1998 update. Tuber Lung Dis [Internet]. 1998 [cited 2020 Nov 15];79(1):3–29. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0962847998900028

157. Telenti A, Imboden P, Marchesi F, Matter L, Schopfer K, Bodmer T, Lowrie D, Colston MJ, Cole S. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. The Lancet [Internet]. 1993 Mar [cited 2020 Nov 15];341(8846):647–51. Available from: https://linkinghub.elsevier.com/retrieve/pii/014067369390417F

158. Piatek AS, Tyagi S, Pol AC, Telenti A, Miller LP, Kramer FR, Alland D. Molecular beacon sequence analysis for detecting drug resistance in Mycobacterium tuberculosis. Nat Biotechnol [Internet]. 1998 Apr [cited 2020 Nov 14];16(4):359–63. Available from: http://www.nature.com/articles/nbt0498-359

159. Piatek AS, Telenti A, Murray MR, El-Hajj H, Jacobs WR, Kramer FR, Alland D. Genotypic Analysis of Mycobacterium tuberculosis in Two Distinct Populations Using Molecular Beacons: Implications for Rapid Susceptibility Testing. Antimicrob Agents Chemother [Internet]. 2000 Jan 1 [cited 2020 Nov 14];44(1):103–10. Available from: https://AAC.asm.org/content/44/1/103

160. Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. Nat Biotechnol [Internet]. 1998 Jan [cited 2020 Nov 15];16(1):49–53. Available from: http://www.nature.com/articles/nbt0198-49

161. Raja S, Ching J, Xi L, Hughes SJ, Chang R, Wong W, McMillan W, Gooding WE, McCarty KS, Chestney M, Luketich JD, Godfrey TE. Technology for Automated, Rapid, and Quantitative PCR or Reverse Transcription-PCR Clinical Testing. Clin Chem [Internet]. 2005 May 1 [cited 2020 Nov 14];51(5):882–90. Available from: https://academic.oup.com/clinchem/article/51/5/882/5629815

162. El-Hajj HH, Marras SAE, Tyagi S, Kramer FR, Alland D. Detection of Rifampin Resistance in Mycobacterium tuberculosis in a Single Tube with Molecular Beacons. J Clin Microbiol [Internet]. 2001 Nov 1 [cited 2020 Nov 15];39(11):4131–7. Available from: http://jcm.asm.org/cgi/doi/10.1128/JCM.39.11.4131-4137.2001

163. WHO. WHO | WHO endorses new rapid tuberculosis test [Internet]. 2010 [cited 2020 Nov 28]. Available from: https://www.who.int/mediacentre/news/releases/2010/tb_test_20101208/en/

164. Gates Foundation. Public-Private Partnership Announces Immediate 40 Percent Cost Reduction for Rapid TB Test | The Bill & Melinda Gates Foundation - Bill & Melinda Gates Foundation [Internet]. 2012 [cited 2020 Nov 28]. Available from: https://www.gatesfoundation.org/Media-Center/Press-Releases/2012/08/PublicPrivate-Partnership-Announces-Immediate-40-Percent-Cost-Reduction-for-Rapid-TB-Test

Appendix A - Personal Publications

Dubnau, E., P. Fontan, R. Manganelli, S. Soares-Appel and I. Smith. 2002. Mycobacterium tuberculosis genes induced during infection of human macrophages. Infection and Immunity 70 (6), 2787-95. http://iai.asm.org/content/70/6/2787.full.pdf+html

Dubnau, E., M. Lanelle, S. Soares, A. Benichou, T. Vaz, D. Prome, J. Prome, M. Daffe and A. Quemard. 1997. Mycobacterium bovis BCG genes involved in the biosynthesis of cyclopropyl keto- and hydroxy-mycolic acids. Molecular Microbiology 23 (2), 313-22. http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2958.1997.2301589.x/epdf

Sa-Nogueira, I., T. V. Nogueira, S. Soares and H. Lencastre. 1997. The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression. Microbiology 143, 957-69. https://doi.org/10.1099/00221287-143-3-957

Dubnau, E., S. Soares, T. J. Huang and W. R. Jacobs, Jr. 1996. Overproduction of mycobacterial ribosomal protein S13 induces catalase/peroxidase activity and hypersensitivity to isoniazid in Mycobacterium smegmatis. Gene 170, 17-22. https://doi.org/10.1016/0378-1119(95)00841-1

Barnes D.M., S. Whittier, P. H. Gilligan, S. Soares, A. Tomasz and F. W. Henderson. 1995. Transmission of multidrug-resistant serotype 23F Streptococcus pneumoniae in group day care: evidence suggesting capsular transformation of the resistant strain in vivo. J. Infectious Diseases 171, 890-6. https://doi.org/10.1093/infdis/171.4.890

Soares S., K. G. Kristinson, J. M. Musser and A. Tomasz. 1993. Evidence for the introduction of a multiresistant clone of serotype 6B Streptococcus pneumoniae from Spain to Iceland in the late 1980s. J. Infectious Diseases 168, 158-63. https://doi.org/10.1093/infdis/168.1.158