Towards understanding the genomic epidemiology of bacterial infections in West Africa

Madikay Senghore, BSc, MSc

University of Warwick

This thesis is submitted for the degree of Doctor of Philosophy

December 2016

DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- The bacterial culture, identification and characterisation of isolates by MLST and/or antimicrobial susceptibility testing was carried out by colleagues at the MRC Unit The Gambia:
    - Jacob Otu performed the viability testing and antimicrobial susceptibility testing for all *Mycobacterium tuberculosis* isolates
    - Ebenezer Foster-Nyarko performed the genotyping and antimicrobial susceptibility testing of the invasive disease *Staphylococcus aureus* strains
    - Chinelo Ebruke performed the genotyping and antimicrobial susceptibility testing of the human carriage *S. aureus* strains
    - Jainaba Manneh performed the genotyping and antimicrobial susceptibility testing of the monkey *S. aureus* strains
    - Ebenezer Foster-Nyarko and Jacob Otu performed the culture and antimicrobial susceptibility testing on the outbreak *Streptococcus pneumoniae* strains
    - Catherine Okoi performed the qPCR for detecting pathogens from cerebrospinal fluid specimens from the meningitis outbreak and performed serotyping on *S. pneumoniae* strains

Working with collaborators provided a learning experience for me. I spent ten days at the University of Bath in July 2015 analysing the *S. aureus* dataset with Professor Edward Feil, Dr Sion Bayliss and Harry Thorpe. I also spent

three days at St George's University of London in July 2015 analysing the Nigerian MTBC dataset with Dr Adam Witney.

- Collaborators at the University of Bath performed the following aspects of the *S. aureus* genome sequence analysis:
    - The phylogenetic analysis on the *S. aureus* strains using a custom pipeline developed by Dr Sion Bayliss and Harry Thorpe
    - The proportion of shared accessory genome analysis was performed with Dr Sion Bayliss
    - The heatmap for Figure3.4 and the BLAST ring for Figure3.5 were plotted by Dr Sion Bayliss and annotated by the author
    - Professor Edward Feil proposed the mutation rate used for phylogenetic dating based on published mutation rates
    - Professor Edward Feil identified genes that were conserved in monkeys but absent in humans that were likely to be associated with host adaptation
- The phylogenetic analysis of the *Mycobacterium tuberculosis* complex (MTBC) strains isolated from Nigeria was performed using a custom pipeline developed by Dr Adam Witney at St George's University of London
- The clonal diversity indexes computed for the serotype 1 *Streptococcus pneumoniae* strains was performed by a colleague at the MRC Unit The Gambia, Archibald Worwui

PUBLICATIONS

Parts of this thesis have been published by the author:

1. **Senghore M**, Bayliss SC, Kwambana-Adams BA, Foster-Nyarko E, Manneh J, Dione M, Badji H, Ebruke C, Doughty EL, Thorpe HA, Jasinska AJ, Schmitt CA, Cramer JD, Turner TR, Weinstock G, Freimer NB, Pallen MJ, Feil EJ, Antonio M. Transmission of Staphylococcus aureus from Humans to Green Monkeys in The Gambia as Revealed by Whole-Genome Sequencing. Appl Environ Microbiol. 2016 Sep 16;82(19):5910-7. doi: 10.1128/AEM.01496-16. PubMed PMID: 27474712; PubMed Central PMCID: PMC5038045.
2. Kwambana-Adams BA, Asiedu-Bekoe F, Sarkodie B, Afreh OK, Kuma GK, Owusu-Okyere G, Foster-Nyarko E, Ohene SA, Okot C, Worwui AK, Okoi C, **Senghore M**, Otu JK, Ebruke C, Bannerman R, Amponsa-Achiano K, Opare D, Kay G, Letsa T, Kaluwa O, Appiah-Denkyira E, Bampoe V, Zaman SM, Pallen MJ, D'Alessandro U, Mwenda JM, Antonio M. An outbreak of pneumococcal meningitis among older children (≥5 years)and adults after the implementation of an infant vaccination programme with the13-valent pneumococcal conjugate vaccine in Ghana. BMC Infect Dis. 2016 Oct18;16(1):575. PubMed PMID: 27756235; PubMed Central PMCID: PMC5070171.
3. Gehre F, Otu J, Kendall L, Forson A, Kwara A, Kudzawu S, Kehinde AO, Adebiyi O, Salako K, Baldeh I, Jallow A, Jallow M, Dagnra A, Dissé K, Kadanga EA, Idigbe EO, Onubogu C, Onyejepu N, Gaye-Diallo A, Ba-Diallo A, Rabna P, Mane M, Sanogo M, Diarra B, Dezemon Z, Sanou A, **Senghore M**, Kwambana-Adams BA, Demba E, Faal-Jawara T, Kumar S, Tientcheu LD, Jallow A, Ceesay S, Adetifa I, Jaye A, Pallen MJ, D'Alessandro U, Kampmann B, Adegbola RA, Mboup S, Corrah T, de Jong BC, Antonio M. The emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis research and drug resistance surveillance. BMC Med. 2016 Nov 3;14(1):160. PubMed PMID: 27806714; PubMed Central PMCID: PMC5094099.

Submitted manuscripts:

1. **Senghore, M**., Otu, J., Witney, A., Gehre, F., Doughty, E.L., Kay, G.L., Butcher, P., Salako, K., Kehinde, A., Onyejepu, N., Idigbe, E., Corrah, T., de Jong, B., Pallen, M.J., and Antonio, M. (Submitted Dec 2016). Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. *Microbial Genomics*

Other Publications:

1. Cornick JE, Chaguza C, Harris SR, Yalcin F, Senghore M, Kiran AM, Govindpershad S, Ousmane S, Du Plessis M, Pluschke M, Ebruke C, McGee L, Sigaùque B, Collard JM, Antonio M, von Gottberg A, French N, Klugman KP, Heyderman RS, Bentley SD, Everett DB, for the PAGe Consortium (2015). Region-specific diversification of the highly virulent serotype 1 Streptococcus pneumoniae. Microbial Genomics, 1(2). http://doi.org/10.1099/mgen.0.000027
2. Chaguza C, Cornick JE, Harris SR, Andam CP, Bricio-Moreno L, Yang M, Yalcin F, Ousmane S, Govindpersad S, Senghore M, Ebruke C, Du Plessis M, Kiran AM, Pluschke G, Sigauque B, McGee L, Klugman KP, Turner P, Corander J, Parkhill J, Collard JM, Antonio M, von Gottberg A, Heyderman RS, French N, Kadioglu A, Hanage WP, Everett DB, Bentley SD; PAGe Consortium.. Understanding pneumococcal serotype 1 biology through population genomic analysis. BMC Infect Dis. 2016 Nov 8;16(1):649. PubMed PMID: 27821148; PubMed Central PMCID: PMC5100261.

ACKNOWLEDGEMENTS

In the name of Allah I begin, for all praise is due to Allah and I ask Him to send peace and blessings on His messenger.

I owe a huge debt of gratitude to my supervisors, Professor Mark Pallen and Professor Martin Antonio. Their kind support, continued guidance and consistent supervision were instrumental in completing this thesis. Thank you so much for investing time, effort and resources in helping me complete this thesis. I would like to extend my gratitude to all members of the Microbiology and Infection Unit at the University of Warwick. In particular, Emma Doughty and Gemma Kay; they have contributed considerably to the sequencing of strains that are presented in this thesis.

This September marked ten years since I joined the MRC Unit The Gambia. Words cannot express how much I appreciate my family at the MRC. I wish to thank; my mentors Professor Tumani Corrah and Dr Assan Jaye for encouraging and motivating me to pursue a career in medical research; Dr Brenda Kwambana-Adams for being a role model and a source of inspiration; Dr Bouke De Jong, Dr Florian Gehre and Jacob Otu for sharing their expertise in Mycobacteriology; Dr Thushan Da Silva and Chinelo Ebruke for reviewing parts of this thesis; and Archibald for his assistance in bioinformatics. To my colleagues from the Vaccines and Immunity theme, Molecular Microbiology team, Molecular Microbiology Focus Group and Tuberculosis Diagnostic Laboratory, thank you for your unrelenting support.

This work has benefitted immensely from external collaborators. I would like to thank Professor Edward Feil, Dr Sion Bayliss and Harry Thorpe for hosting me at the University of Bath and teaching me a lot about *Staphylococcus aureus* genomics. Equally I would like to thank Professor Phillip Butcher and Dr Adam Witney for welcoming me at St George's University of London and sharing their expertise in tuberculosis genomics.

A special thanks goes to our long-term collaborators, the members of the Pneumococcal African Genomics consortium who gave me my first training in pathogen genomics and contributed to the analysis of the pneumococcal meningitis outbreak. I want to particularly thank Dr Jennifer Cornick for her

advice on my analysis of the pneumococcal meningitis outbreak dataset. We appreciate the collaborators on the West African Nodes of Excellence for Tuberculosis AIDS and Malaria for their contributions to the tuberculosis drug resistance survey in West Africa. We thank the WHO African regional office and the Ghanaian Ministry of Health for their collaboration during the meningitis outbreak in the Brong-Ahafo region of Ghana in 2016. We also thank the International Vervet consortium for collecting nasopharyngeal swabs from African green monkeys in The Gambia.

Finally, I would like to thank my family for their support. My mother and father raised me upon intellectual values and always pushed me to bring the best out of me. I consider myself very fortunate to have such amazing parents. I wish to thank my wife for putting up with my constant late nights and working odd hours. My brothers Momodou and Abdoukarim, thank you for being there for me when I needed you the most. My dear sister Sukai, thank you for being a good listener and pen pal.

I wish to dedicate this thesis to my late grandfathers Alhagi Momodou Babucarr Njie and Alhagi Abdoukarim Senghore. I know they would have been proud of this achievement.

ABSTRACT

Bacterial infection is a major cause of morbidity and mortality in sub-Saharan Africa especially among young children. Despite the high burden of disease caused by bacterial infection in Africa, there remains a significant paucity of data on the molecular epidemiology of most pathogens in the sub-region. Healthcare facilities are generally underfunded in West Africa and most facilities lack the basic capacity to perform standard microbiological identification of bacterial pathogens.

Understanding the biology and epidemiology of pathogens is fundamental to a successful intervention strategy. Genomics offers unprecedented insights into the epidemiology and biology of infectious diseases, which dominate the public health agenda in West Africa. Here, I introduce a case study of three important pathogens in West Africa. I describe a unique scenario associated with each pathogen and present WGS as a solution to the problem.

Firstly, whole genome sequencing has provided insights into the evolutionary origin of *Staphylococcus aureus* in monkeys from The Gambia and established that monkeys in The Gambia do not pose a threat of serving as reservoirs of highly virulent *S. aureus* that can infect humans. Secondly, genomics has unravelled the evolutionary mechanisms that led to the emergence of a novel clone of serotype 1 *Streptococcus pneumoniae,* which caused an outbreak of meningitis in Ghana following the introduction of the 13-valent pneumococcal vaccine, PCV-13. Thirdly, through genomics we are beginning to build a deeper understanding of the epidemiology of *Mycobacterium tuberculosis* complex in West Africa. Genomics is unravelling the evolutionary mechanisms that are driving the emergence of multidrug resistant tuberculosis. Importantly, genomics has shown that lineages of MTBC that are endemic to West Africa are the principal proponents of multidrug resistance in this sub-region.

The time has come for West Africa to embrace the genomics era and exploit the full potential of microbial genomics. I hope that my work will inspire West African scientists to embrace whole genome sequencing in the fight against infectious bacterial diseases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
AIDS	Acquired Immune deficiency virus
BRIG	BLAST Ring Image generator
CC	Clonal complex
CLIMB	Cloud infrastructure for microbial bioinformatics
CPS	Capsular polysaccharide
CTAB	Cetyl trimethylammonium bromide
DOTS	Directly Observed Treatment Short-Course
ECOWAS	Economic community of West African states
GBS	Group B Streptococcus
GI	Genomic Island
HGT	Horizontal gene transfer
HIV	Human Immunodeficiency virus
IA	Index of association
IEC	Immune evasion cluster
kB	Kilobyte
MAF1	*Mycobacterium africanum* West Africa 1
MAF2	*Mycobacterium africanum* West Africa 2
MB	Megabyte
MDR-TB	Multidrug resistant tuberculosis
MGE	Mobile genetic element
MIRU-VNTR	Mycobacterial interspersed repetitive unit variable number tandem repeat
MLST	Multilocus sequence typing
MRCA	Most recent common ancestor
MRCG	Medical research council unit The Gambia
MRSA	Methicillin resistant *Staphylococcus aureus*
MTBC	*Mycobacterium tuberculosis* complex
PAGe	Pneumococcal African Genomics consortium
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PCV	Pneumococcal conjugate vaccine
PCV-13	13-valent pneumococcal conjugate vaccine
PCV-7	7-valent pneumococcal conjugate vaccine
PreXDR-TB	Pre-extensively drug resistant tuberculosis

rÂ¯ d	Standardized index of association
SaPI	*Staphylococcus aureus* pathogenicity Island
SCCmec	Staphylococcal chromosomal cassette mec
SNV	Single nucleotide variant
ST	Sequence type
TB	Tuberculosis
vSa	*Staphylococcus aureus* genomic Island
WANETAM	West African nodes of excellence for tuberculosis AIDS and malaria
WGS	Whole genome sequencing
WHO	World Health Organisation
XDR-TB	Extensively drug resistant tuberculosis

1 CHAPTER ONE: INTRODUCTION

1.1 Bacterial Infection

1.1.1 The nature and discovery of bacteria

The discovery of bacteria by Antonie van Leeuwenhoek is one of the most important discoveries in the field of microbiology. Armed with a simple microscope, Leeuwenhoek observed microscopic living organisms in specimens of water [1]. In his 1667 letter to the Royal Society, Leeuwenhoek describes these microorganisms as "animalcules" [2].

Bacteria are unicellular microorganisms that are characterized by the absence of a nucleus and membrane-bound cell organelles [3]. Bacteria form a diverse group of organisms that are highly adaptable and occupy niches in all environments. Most bacteria can be considered harmless to humans or beneficial components of the human microbiome [4].

In 1850, Ignaz Semmelweis delivered a lecture to the Vienna Medical Society detailing a link between puerperal fever (also known as childbed fever) and the re-absorption of noxious agents in decaying animal-organic matter [5]. Semmelweis' observation that medical interns who performed autopsies were transmitting the disease to patients in the maternity ward contradicted the accepted scientific opinions of his time. Nonetheless, his observation led him to impose chlorine hand washing on medical personnel, which led to a decrease in puerperal fever mortality [6].

The experiments of Louis Pasteur and Robert Koch were instrumental in establishing the proof that microorganisms can cause disease in healthy hosts. This proof led to the widespread acceptance of the germ theory that infectious diseases are caused by transmission of an infective organism from one host to another [7]. Koch formulated what are now known as Koch's postulates, which were fundamental to the germ theory and establishing the

causative relationship between germs and diseases [7]. Koch's postulates state that in order to define a microorganism as the causal agent of infection it must:

1. Be present in all diseased hosts but not be present in healthy hosts
2. Be isolated from the diseased host and grown in a pure culture
3. Be able to cause disease when the cultured microorganism is introduced into a healthy host
4. Be re-isolated from the diseased experimental host and be identified as identical to the original specific causal agent

It is important to note that Koch's postulates are limited when it comes to defining a causal relation between disease and opportunistic bacteria, uncultivable bacteria or bacteria that elicit toxin-dependent pathogenesis [8]. Bacterial infection involves a complex host-pathogen interaction, whose outcome is not solely dependent on the microbe. A number of factors contribute to disease outcome including but not restricted to [9]:

- A conducive environment for microbial replication
- The microbe's ability to cause disease
- The immune status of the host
- The ability of the microbe to penetrate the host's protective barriers and evade its innate defence mechanisms

1.1.2 The global burden of bacterial infection

Despite advances in sanitation, hygiene, medicine, healthcare and biotechnology, infectious diseases continue to be a major public health burden in all regions of the world [10]. In 2012, the WHO ranked lower respiratory tract infections and diarrhoeal disease as the fourth and seventh most common causes of death globally respectively [11]. Although the 2015 global disease burden estimate reported a decline in the mortality due to communicable disease, diarrhoeal disease, lower respiratory tract infections and other infectious diseases continue to be a major cause of mortality [12].

Diarrhoea and pneumonia are the leading causes of death due to infection among children [13]. The vast majority of these deaths occur in the first two years of life [13]. Africa and South-East Asia bear the highest burden of severe pneumonia and severe diarrhoea globally [13]. The Global Enteric Multisite Study reported that four species caused the majority of diarrhoeal infection in Africa and South-East Asia: *Escherichia coli, Shigella, Cryptosporidium* and rotavirus [14]. Prior to the introduction of pneumococcal conjugate vaccines the bacterial species *Streptococcus pneumoniae* and *Haemophilus influenzae* were the leading causes of pneumonia, but the influenza virus and respiratory syncytial virus were also important causes of pulmonary disease [11].

The global burden of bacterial meningitis has remained relatively steady over the past decade [12] despite vaccine interventions in the African meningitis belt [15]. Healthcare-associated infections are a major public health concern globally and the highest burden exists in developing countries [16]. It is worrying that healthcare-associated multidrug-resistant bacteria continue to emerge in different regions and spread across the globe [17].

The WHO declared tuberculosis a global emergency in 1993 [18]. Mortality due to tuberculosis has been halved since 1990, but it remains one of the leading causes of death globally [19].Tuberculosis is the leading cause of death due to infection globally after HIV, despite the availability of effective treatment for both [12]. HIV-infected individuals are at risk of developing bacterial sepsis, particularly from *S. pneumoniae* and non-typhoidal *Salmonella* [20].

1.1.2.1 Infection is a disease of poverty

In low-income countries, infectious diseases remain the most common cause of death: twenty species — principally viruses and bacteria — caused two-thirds of the deaths from infectious diseases in low-income countries in 2010 [10]. According to the 2010 global disease-burden study, lower-respiratory-tract infection was the leading cause of death in low- to middle-income countries [11]. The symbiotic relationship between poverty and infection breeds a vicious

cycle; the WHO estimates that infection forces 100 million people below the poverty line annually [21].

Under-nutrition is a direct consequence of poverty [22]. Children suffering from protein-calorie malnutrition are more susceptible to infection because they have a weakened innate immune response [23]. Worryingly, the causal links run in both directions: malnutrition is a risk factor for diarrhoea and lower respiratory tract infection, but these infections also perpetuate malnutrition through decreased nutrient intake and energy loss [23,24]. In adults, under-nutrition leads to a weakened immune system, which increases the risk of progression from latent tuberculosis to active disease [25].

Poor housing and suboptimal hygiene conditions have a negative impact on health. Poor housing leads to overcrowding and inadequate ventilation, which increases the risk of airborne disease such as tuberculosis [23]. Lack of a clean water supply and poor sanitation are known risk factors for diarrhoea [26]. People living in the poorest regions of the world have limited access to healthcare. Those who are fortunate to make it hospital often cannot afford to pay for treatment and suffer from lost income [27].

Infectious diseases disproportionately infect the most impoverished regions of the world, particularly sub-Saharan Africa [28,29]. In the last decade mortality due to some neglected tropical diseases has risen: examples include Chagas disease, leishmaniasis, dengue and Ebola virus infection (mainly through the West African Ebola outbreak) [12]. These infections (excluding Ebola) are largely treatable with existing drugs, but delivery is not always forthcoming [28]. Intervention policies need to tackle the socio-economic causes and impacts of infectious diseases.

1.1.3 Sources of bacterial infections

The mechanisms for transmission of bacterial infection are well understood. Infectious disease are transmitted between individuals mainly though inoculation (coming into contact with the body fluids of an infected individual), airborne or waterborne transmission [30]. Non-human vectors also play an important role in the life cycle and transmission of some pathogenic bacteria

[31]. Effective strategies exist for managing the spread of infectious diseases. These include clinical interventions such as vaccination and antibiotic therapy, as well as general measures like sanitation, chemical disinfection, hand washing and vector control [30]. Despite these advances outbreaks of bacterial pathogens occur recurrently in the community and within healthcare facilities.

Outbreaks of infectious disease are marked by a sudden increase in the incidence of a certain disease in a given locality. As little as two epidemiologically linked cases of a rare infectious disease may be considered as an outbreak [32]. During an outbreak, individuals may have acquired the infection from a common source (e.g. a contaminated water source) or the outbreak may be propagated by inter-person transmission. When outbreaks spread to a large proportion of the population, they are referred to as epidemics. For example, the epidemic threshold for meningitis outbreaks is when more than 15 cases of meningitis per hundred thousand population are recorded averaged over two weeks [33].

1.1.4 Zoonosis and emerging disease

Zoonotic infections are infections that humans acquire from animals. Emerging infectious diseases can be described as the onset of a novel infection within a population or a rapid increase in the prevalence and/or geographical range of an existing infection [34]. Emerging infectious diseases represent a major threat to public health globally [8]. Unlike viruses that emerge mainly through rapid evolution, emerging bacterial infections are rarely due to novel pathogenic species/strains [8].

The rise of virulent and antibiotic-resistant forms of known human-adapted bacterial pathogens may be considered as a source of emerging infectious diseases. Two of the best-known examples are methicillin-resistant *Staphylococcus aureus* (MRSA) and multidrug-resistant tuberculosis (MDR-TB). Rapid dissemination of these strains within the community and among hospitalised patients justifies their consideration as emerging infectious

disease [8]. Other notable examples are the vancomycin-resistant *Enterococcus* and extended-beta-lactamase-producing *E. coli* [8].

Zoonoses are an important cause of emerging infections (Table 1.1) [8,34], with zoonotic transmission of bacterial infections to humans via routes including [35]:

- Direct contact with infected animal material e.g. body fluids or faeces
- Being bitten or scratched by an infected animal
- Transmission through an arthropod vector
- Eating contaminated animal products

Table 1.1 Important zoonotic bacterial pathogens that are causes of emerging infection in humans. The list was adapted from [8] citing [36-43].

Species	Animal host	Disease in humans
***Campylobacter* spp**	Poultry and cattle	Diarrhoea
Streptococcus bovis	Sheep, cattle and ruminants	Endocarditis and colorectal cancer
Capnocytophaga canimorsus	Dogs and cats	Sepsis, septic shock and meningitis
***Escherichia coli* 0157:H7**	Cattle and ruminants mainly	Haemorrhagic colitis
Neoehrlichia mikurensis	Ticks and rats	Neoehrlichiosis : Systemic inflammatory response; Vascular and thromboembolic events
Anaplasma phagocytophilum	Wild and domesticated mammals and in ticks	Tick borne fever
Spotted fever group *Rickettsia* spp.	Ticks	Spotted fever rickettsiosis
Borrelia burgdorferi	White-footed mouse and ticks	Lyme disease

Improved diagnostic techniques, both culture- and sequence-based, have been instrumental in the discovery of emerging bacterial infectious diseases [44]. To highlight a few examples:

- Prolongation of culture duration led to the discovery of mycobacteria in blood [8]

- Performing culture on media inoculated with antibiotics made it possible to grow *Campylobacter* and *Helicobacter* species [45,46]
- Culture with non-mammalian cell lines is a promising approach for culturing otherwise uncultivable bacteria [8]
- PCR is a useful tool for detecting emerging infections from sterile sites
- 16S RNA sequencing has been useful in taxonomic classification of emerging infections

1.1.4.1 Ecological factors

The steady growth in human population and the consequent increase in population density aid the spread of bacterial infections among humans. Travelling has become easier over the years and international trade and tourism mean more and more people are travelling around the globe, which further aids the dissemination of infectious agents. Expanding communities continue to encroach deeper into natural habitats increasing the potential for zoonotic transmission of emerging infections. In North America, the re-emergence of Lyme disease has been associated with deforestation and loss of habitat of its animal host [43].

The deadliest emerging infection of our time is HIV/AIDS, with consequent bacterial infection as an emergent threat among HIV positive individuals [47]. Ironically, advances in medicine have contributed to the emergence of new bacterial infections in vulnerable patients. When immunosuppression is employed due to organ transplants or autoimmune disease it renders patients vulnerable to infection. Patients being treated for chronic infections such as cancer, diabetes and renal insufficiencies are also prone to impaired immune systems and are vulnerable to infection. Urinary catheters have contributed to the emergence of bacterial urinary tract infections like *Actinobaculum schaalii* and *Alloscardovia omnicolens* [48].

1.2 Management of bacterial infection

1.2.1 Prevention of bacterial infection: vaccination

Vaccination is an important output from microbiology that embraces synergies between molecular biology, immunology, public health and biotechnology [49]. The aim of vaccination is to induce long-term protection by priming the human immune system with a component of a pathogen [50]. The general principle is that introduction of the vaccine leads to the production of an immune response against the antigenic component of the vaccine, which is protective when the body encounters the pathogenic microorganism. Classical vaccines are mostly comprised of attenuated organisms, inactivated whole-cell preparations or purified cellular components.

The discovery of the smallpox vaccine by the physician Edward Jenner in the 18th century paved the way for the eradication of smallpox [51]. In the 19th century Pasteur discovered that the causative agent of chicken cholera *Pasteurella multocida* could be attenuated into less virulent forms [50]. Although the true rationale behind attenuation would not become apparent until the latter part of the century, it formed the basis for Pasteur's work on rabies and anthrax vaccines [50]. The start of the 20th Century was marked by the discovery of the attenuated *Mycobacterium bovis* Bacille Calmette-Guérin vaccine, which is still the only licensed vaccine for tuberculosis [52]. During this period, inactivated whole-cell vaccines were also produced against typhoid, cholera and the plague [50].

Some bacteria secrete toxins that serve as important virulence factors [53]. The discovery that sub-lethal doses of inactivated bacterial toxins (toxoids) could induce production of antitoxins in human serum, and thus confer protection, led to the development of the tetanus, pertussis and diphtheria vaccines [50]. More recently, vaccines against *Haemophilus influenzae* type B, *Neisseria meningitidis* and *Streptococcus pneumoniae* have been developed by conjugating their capsular polysaccharide to protein molecules such as the diphtheria toxoid [54-56].

1.2.2 Diagnosis of bacterial infection

Identifying the aetiological agent of infection facilitates the appropriate delivery of antibiotic therapy, which increases the chances of treatment success [57]. Phenotypic methods for species identification, including microscopy, culture and biochemical assays, are still widely used as diagnostic tools. Classical taxonomy defined bacterial species as clusters of organisms sharing phenotypic traits (e.g. cell morphology, growth conditions and metabolic characteristics). This approach relies on being able to grow the bacteria in the lab, usually a pure culture, and therefore precludes the classification of uncultivable bacteria [58]. Technological advances have made DNA-based molecular techniques for pathogen detection more accessible and affordable [57,59].

1.2.2.1 Molecular diagnostic approaches

Over time, species classification based on phenotypic traits was complemented with molecular indicators such as GC content and whole-genome DNA-DNA hybridisation, where a cut-off of 70% DNA-DNA relatedness was recommended for classifying organisms into the same species ([60] cited in [61]).

DNA-DNA hybridisation has now been replaced in most labs by the less arduous 16S ribosomal RNA sequencing, leading to an updated species definition which requires ≥97% sequence identity in the 16S rRNA gene [62]. 16S rRNA sequencing is a suitable marker for phylogenetic analysis and was instrumental in distinguishing *Archaea* as an independent domain of life [63]. However, despite its usefulness, 16S rRNA fails to discriminate between some closely related species [63], even at a more stringent cut-off of 99% sequence similarity (e.g. species within the genus *Acinetobacter)* [61].

PCR amplification of single gene loci is now the most widely used molecular technique for pathogen detection in research and diagnostics laboratories [57]. The advantages of PCR-based detection include avoiding the onerous need for culture and the ability to detect specific pathogens at low concentrations

and even from non-viable cells [64,65]. In addition, multiplex PCR assays can detect multiple pathogens in a single reaction [66,67]. However, the chief disadvantage of PCR-based detection is that only known or expected species can be targeted.

1.2.2.2 Rise of genomics and metagenomics

Whole-genome sequencing (WGS) offers an opportunity to study the entire DNA of an organism. This has immense value beyond simply identifying the species—in addition, genomic data can be used for phylogenetic analysis and genotypic characterisation of strains [68].

In the genome era, pairwise genomic comparisons have provided new insights into bacterial diversity, prompting a review of the boundaries of taxonomic classifications [69]. In particular, measuring average nucleotide identity through pairwise comparison of shared sequences in the genome offers improved resolution at the species and sub-species level [70]. An average nucleotide identity of 95% or more corresponds to 70% or more DNA-DNA re-association and strains with >94% average nucleotide identity show consistent phenotypes in the context of species designation [69]. A similar metric based on amino acid identity is a useful measure for comparing distantly related genomes [70].

Culture-free techniques like unbiased metagenomic sequencing are opening new frontiers in clinical microbiology [71], offering potential for detecting and characterising bacterial pathogens directly from clinical samples [72-74]. Metagenomics also identifies uncultivable potentially pathogenic bacteria and offers a holistic snapshot of the microbiome [71].

1.2.3 Treatment of bacterial infection: antibiotics

The antibiotics era was ushered in by the discovery of naturally occurring compounds – penicillin, streptomycin, chloramphenicol and tetracycline – that have bactericidal properties [75]. Antibiotics elicit bactericidal activity by inhibiting important cellular functions such as synthesis of the cell wall, of proteins, of DNA and of RNA [76]. Despite the success of antibiotics as

therapeutic agents the antibiotics era has been marred by the widespread emergence of antibiotic resistance [77].

Most antibiotics operate by binding to their target molecule with a high affinity. Antibiotic resistance can be conferred by a number of mechanisms such as [78]:

1. Preventing the antibiotic from reaching its target through:
    a. Decreased bacterial cell permeability by down-regulation of porin channels [79]
    b. Increased removal of antibiotics from the bacterial cell through efflux pumps
2. Modifying the drug target through:
    a. Point mutations in the gene encoding the drug target
    b. Homologous recombination with genes encoding insensitive drug targets e.g. penicillin binding proteins
    c. Modification of the target protein by methylation or binding of proteins
3. Inactivating or modifying the antibiotic through:
    a. Enzyme-catalysed hydrolysis of the antibiotic e.g. β-lactamases
    b. Chemical modification of the antibiotic e.g. aminoglycoside-modifying-enzymes [80]

Although a crisis looms, the development of new antibiotics is no longer deemed economically viable [81]. Instead pharmaceutical companies have turned their attention the development of more lucrative therapeutic agents for chronic non-communicable disorders [82]. Antibiotic resistance is recognised as a global emergency [83]. Scientists are exploring novel targets such as bacterial cell-division proteins as potential targets for future antibiotics [84]. Bacteriophage therapy is also re-emerging as a promising alternative to antibiotics [85,86]. Regulation, however, is a double-edged sword, because although it may be beneficial in controlling the rise of existing resistance, it can be an obstacle in pursuing the discovery of novel antibiotics [82,87].

1.2.4 Epidemiological typing to monitor spread of bacterial infection

Microbial typing allows us to track the emergence and spread of bacterial pathogens and in some cases can be used to retrace evolutionary history [88]. Typing methods that discriminate between closely related strains are essential in identifying person-to-person transmission in an outbreak [89].

1.2.4.1 Phenotypic typing approaches

Phenotypic typing methods discriminate between strains on the basis of phenotypic traits. Serotyping is perhaps the most widespread phenotypic typing method. Serotyping exploits intra-species strain-specific differences in the expression of cell-surface antigens. For example, variation in the capsular polysaccharide of *S. pneumoniae* is used to classify strains into over 90 serotypes. The gold standard for this is the Quellung reaction, which detects a serotype-specific interaction between an antibody and the capsular antigen. However, molecular techniques are now available for serotyping *S. pneumoniae* based on detection of capsular genes [90,91].

Salmonella isolates are serotyped according to the immunological profiles of their somatic (O) and flagellar (H) antigens. Serotyping has been the cornerstone of epidemiological studies of *Salmonella*. However, keeping stock of the antisera for all the antigenic variants is demanding. Molecular approaches for serotyping *Salmonella* are gradually replacing conventional techniques [92,93].

1.2.4.2 Molecular typing: MLST

The shortfall of the earliest typing methods was that they lacked a standardized nomenclature for clonal lineages and they showed poor reproducibility in unrelated datasets [94]. Multilocus sequence typing (MLST) succeeded in providing a standardised format for documenting gene diversity to facilitate ease of inter-laboratory data comparison [95]. MLST assigns a

sequence type (ST) based on the allelic combination of 7 housekeeping genes that are conserved in a given species [95].

The nucleotide sequence for each of the seven housekeeping genes is compared to known allele sequences that are stored in a curated database. Each time a new sequence for a given gene is reported, it is designated a new allele number. Each unique combination of alleles is designated as a unique ST. STs are clustered into clonal complexes based on the sharing of at least 5 allelic loci with at least one other ST in the clonal complex [96]. By convention the most abundant ST is designated as the founder of the clonal complexes.

MLST is a robust tool for assessing population-based microbial diversity [57], as shown by a few key examples:

- MLST has been instrumental in understanding the global population structure of *Staphylococcus aureus* [97,98]
- Through MLST, the impact of vaccine introduction on the genetic diversity of *Streptococcus pneumoniae* has been studied [99]
- MLST has been implemented in tracking the emergence and spread of antibiotic-resistant strains of *Neisseria meningitidis* [100]
- Associations between bacterial species and animal reservoirs have been revealed through MLST [101,102]

The ease of use and reproducibility of MLST make it the gold standard for typing most common bacterial pathogens [103,104]. However, MLST does not offer enough discriminatory power to make it applicable for:

- Detailed outbreak investigation in most species [89]
- Typing of genetically monomorphic bacteria like *Mycobacterium tuberculosis* (*Mtb*) [88]

1.2.4.3 Molecular typing for Mycobacterium tuberculosis

A wide range of molecular typing methods have been applied to *Mtb*, including:

- **IS6110-RFLP typing**, which involves cleaving genomic DNA at specific sites and using electrophoresis to separate the IS6110 fragments to yield a specific banding pattern
- **Spoligotyping**, which characterizes strains based on the presence or absence of spacer units in the direct repeat locus
- Mycobacterial interspersed repetitive unit variable number tandem repeat (**MIRU-VNTR**) analysis, which relies on the rapid rate of change in repetitive DNA elements, to distinguish between isolates with high-discriminatory power

Although spoligotyping can distinguish between strains at the sub-species level it is limited by the fact that it uses less than 0.1% of the *M. tuberculosis* genome. MIRU-VNTR offers higher discriminatory power than spoligo typing but it lacks standardised laboratory protocols. IS6110-RFLP also offers high discriminatory power between strains, but it requires an expert technician and sophisticated hardware [105].

1.2.4.4 Genomic epidemiology

WGS using next-generation sequencing (see following section) has emerged as a useful tool for the detection and characterisation of outbreaks, as well as informing outbreak management [106,107]. In fact, next-generation sequencing is now ready to contribute routine pathogen surveillance [108-111]. This kind of digital surveillance twinned with effective data sharing will facilitate real time analysis of outbreaks in the context of the global phylogeny [108]. However, whole-genome sequencing needs to be made accessible in low resource settings, where the burden of infection is highest.

The problem with most existing genotyping methods is that they probe only a very small proportion of the genome [112]. WGS has emerged as a superior technique for typing bacteria, since it can detect single nucleotide changes at

all genomic loci [113]. WGS offers high-resolution inter-strain comparison, which has improved our understanding of geographic distribution of species and their evolutionary origins [114,115]. Evolution within a species can be investigated by reconstructing a phylogeny from variant sites in the core genome [114,116].

To establish WGS as the gold standard for microbial typing, standardised schemes for nomenclature and lineage assignment need to be devised. For some species like *Mtb,* a robust typing scheme exists for characterising strains based on a barcode of genome-wide single nucleotide variants (SNVs) [117]. In other bacteria, like *Staphylococcus aureus,* nomenclature of clonal lineages inferred from MLST are still used to annotate whole-genome phylogenies [118]. A ribosomal MLST scheme that probes 53 ribosomal proteins present in almost all bacteria has been proposed to ease the transition from working with MLST to working with whole genomes [119].

1.3 DNA Sequencing

1.3.1 Sanger sequencing

The development of the DNA sequencing technique, now commonly known as Sanger sequencing, by Frederick Sanger and his colleagues in 1977 remains a cornerstone of molecular biology [120]. Sanger sequencing is based on a DNA synthesis polymerase reaction in the presence of dideoxynucleotide chain terminators [121]. Initially, the reaction was run in quadruplicate with different base terminators (ddA, ddC, ddT or ddG) followed by gel separation in separate lanes. The user would then examine the gel to reconstruct the nucleotide sequence at each position based on the terminator base and fragment length [121].

The advent of fluorescently labelled dideoxynucleotide chain terminators increased the throughput of Sanger sequencing and enabled automation. This meant that the sequencing could be performed in a single reaction and the strands could be separated through capillary electrophoresis [122]. A

fluorescent detector connected to a computer was able to detect the nucleotides and reconstruct the sequence in real time [123]. Sanger sequencing can generate reads up to a thousand base pairs long [124].

1.3.2 Whole-genome sequencing

The first bacterial genome sequences were published just over two decades ago [125,126]. Since then, there has been an exponential increase in the number of bacterial genomes sequenced and our understanding of the bacterial genome has improved significantly. This development stems from technological advances in high-throughput sequencing, the decreased cost of sequencing and improved means of data sharing [127].

Bacterial genome sequencing has evolved through three phases [128]:

1. The whole-genome shotgun sequencing era
2. The next-generation sequencing era
3. The single-molecule sequencing era

The advent of shotgun cloning in WGS was the first breakthrough [128]. In this three-step process, genomic DNA is randomly sheared, the fragments are cloned into *E. coli* and sequenced *en masse* by Sanger sequencing [128]. Bioinformatics tools were developed to assemble the short reads into contigs (contiguous strings of bases). Shotgun cloning and subsequent whole-genome sequencing provided novel insights into the biology of the deadly human pathogens *Mycobacterium tuberculosis* [129] and *Yersinia pestis* [130] and identified a myriad of novel genes in the model organism *E. coli* [131]. Genome sequencing that relied on shotgun cloning, however, had its drawbacks: most notably that only clonable regions of the genome could be sequenced [128].

The introduction of next-generation sequencing to bacterial genomics marked the second breakthrough. Next-generation sequencing involves preparing templates by tagging genomic fragments with chemical adapters, which are then amplified by PCR and undergo sequencing-by-synthesis on a solid surface [132]. This technology ushered in the era of bench top sequencers like the Illumina MiSeq and the Roche 454 Jr [128]. Bench top sequencers are

now commonplace in most research groups and for now Illumina dominate the field with short-read sequencing [109].

Next-generation sequencing brought a massive improvement in throughput and volume of data but at the cost of significantly shorter read length [132]. This meant that only draft assemblies could be generated, rather than complete genome sequences. Inadvertently, this influenced bacterial genomics by changing the emphasis from in depth analysis of single genomes to comparative genomics of large dataset of closely related strains [128].

Single-molecule sequencing marks the third breakthrough in bacterial genome sequencing, while also reviving the study of complete genome sequences through long read lengths. Pacific Biosciences launched the first widely used single-molecule sequencing technology, the PacBio. The PacBio can generate high-quality finished genomes in combination with Illumina short reads, or even on its own [133,134]. This makes PacBio ideal for structural genomics and plasmid analysis e.g. in multidrug resistant bacteria [135]. Unfortunately the PacBio is mostly restricted to major sequencing centres due to the size of the instruments and its hefty price tag [128].

Oxford Nanopore sequencing remains a work-in-progress, despite proof-of-principle that it can generate reasonably good draft genomes [128]. Palmtop sequencing with Oxford Nanopore's Minion instrument has made on-site sequencing of samples possible in outbreak surveillance even in remote regions [136,137].

1.3.3 Bacterial genome analysis

Although most bacterial genomes consist of a single circular chromosome, a few species have evolved to harbour multiple chromosomes [138,139]. Bacterial genomes are usually smaller than eukaryotic genomes and vary in size from 112 kB to 14 MB [140,141]. The nucleotide content of bacterial genomes varies markedly, e.g. the GC content of finished assemblies ranges from 15% to 85%. Bacteria from complex environmental niches tend to have larger genomes and higher GC contents than host-associated bacteria [127].

1.3.3.1 Bacterial genome assembly and annotation

Over the years, technological advances in sequencing have been paralleled by computational advances geared towards analysing sequence data. Genome assembly involves stitching together millions of short sequence fragments in an attempt to reconstruct the original genome. Reference-based assembly methods map short reads to a reference genome and overlapping fragments are assembled into contigs. However, reference-based mapping is limited by the fact the regions that do not map to the reference are excluded.

De novo assemblers attempt to reconstruct the genome from scratch without the aid of external data. This eliminates the bias towards the reference, but the assemblies are generally more fragmented [124]. *De novo* assembly is recommended for new pathogens or novel strains of well-studied pathogens [142]. Hybrid approaches that combine *de novo* assembly of long reads from single-molecule sequencing and error correction with high accuracy short reads have produced almost-complete genome assemblies [133,143]. Contigs generated from *de novo* assembly with short reads can be improved by joining them into scaffolds using a reference guided approach [144,145].

Bacterial genomes show a good correlation between genome size and number of genes and have low proportions of non-coding DNA [146]. Automatic annotation pipelines are often employed for annotating new sequences of bacteria. The generic workflow of automated pipelines for annotating bacterial genomes includes the following steps [147]:

1. Predicting open reading frames from a FASTA sequence
    a. GLIMMER is the most common tool for this [148]
2. Comparing the predicted sequence to a reference genome or online databases e.g. Uniprot
3. Search for a homologue
    a. If there is a match, assign annotation from the reference
    b. If there is no match, then annotate the coding sequence (CDS) as a hypothetical protein
4. Predict functional domains for the protein and update annotation
5. Predict other non-coding features e.g. non-coding RNAs

The exponential increase in the number of prokaryotic genomes available provides an amazing resource for comparative genomics. However, one must proceed with caution when transferring annotation from one genome to the next, as there are inconsistencies in the annotation protocols and pipelines implemented by different research groups [147]. In addition, programs like Microsoft Excel have been shown to corrupt gene names [149], while spelling mistakes and use of inconsistent product names for the same gene compounds this problem [147]. For all these reasons, to achieve high-quality annotation, the predicted annotation must be manually curated.

1.3.3.2 Insights into pathogen biology

Bacterial genomics has been instrumental in shaping our understanding of host-pathogen interactions. Knowing the sequence for all genes in the genome has given crucial insights into the biology of bacterial pathogens. Through the genome, we are able to study the full profile of virulence genes that mediate an organism's pathogenesis [150]. WGS has been instrumental in the detection and characterisation of novel variants of important virulence-associated genes e.g. WGS has characterised diversity in the *cagA* gene, the most important virulence factor in *Helicobacter pylori* infection [151].

Genome-wide association studies can be used to dissect the genetic basis for phenotypic traits in bacteria [152]. A genome-wide association study of *Campylobacter* identified a seven-gene vitamin B5 biosynthesis region as a host-specificity factor [153]. In the *Campylobacter jejuni* genome, the presence of imported genes has been linked with evolutionary relationships with specific host species [154,155]. In *Mycobacterium tuberculosis* genome-wide association studies have been employed to detect antibiotic resistance mutations based on convergent evolution [156]. Genome-wide associations in bacteria are confounded by the strong forces of natural selection that are exerted on bacterial populations [152].

Advances in next generation sequencing have made it possible to sequence the entire bacterial transcriptome of a strain [157]. The transcriptome is key to understanding how bacteria respond to changes in their environment.

Genomics has also provided insights into the role of regulatory non-coding RNA. Regulatory non-coding RNA's can affect the physiology of bacteria by binding to messenger RNA in order to reduce translation efficiency [158].

1.3.3.3 Discovery of novel targets

Advances in genomics have led to the discovery of novel targets for diagnostics, antibiotic and vaccines [159]. Comparative genomics can identify diagnostic targets that are conserved within a species, but are not found in other species. These targets can then be incorporated into real-time PCR diagnostic assays [160].

In the genomics era, scientists have looked increasingly towards genomics solutions to the rising antibiotic resistance problem [161,162]. To take just one example: in 2014, Mobegi *et al* published a genomic approach for systematically discovering novel antibiotic targets in three leading respiratory pathogens; *Streptococcus pneumoniae, Haemophilus influenzae and Moraxella catarrhalis* [163]. Their approach exploited high-density transposon mutagenesis, high-throughput sequencing, and integrative genomics to identify essential genes that were conserved, but unlikely to interfere with host function or the microbiome. As proof of concept they identified four novel targets that were experimentally validated [163]. Their approach overcame the shortfalls in specificity and throughput that held back previous attempts to discover novel targets from genomic data [164,165]. As genomics screening for antibiotic targets circumvents the need for traditional laboratory screening, this can fast-track the discovery of novel antibiotics [163].

The genomics era has also revolutionised vaccine development through "reverse vaccinology", which mines bacterial genomes for proteins that can potentially elicit protective immunity as candidates for subunit vaccines [166-168]. Comparative genomics of multiple strains can predict the suitability of a candidate protein by determining how conserved it is across diverse strains [169]. Reverse vaccinology has the potential to lead to vaccines for bacteria where the conventional approaches have failed [167,169].

1.4 Bacterial genome evolution

Bacterial evolution is driven by the accumulation of genetic variation in the bacterial genome. WGS has enhanced our understanding of the evolutionary mechanisms that shape major human pathogens, documenting unprecedented genomic diversity between and within bacterial species [127,170]. Understanding microbial evolutionary mechanisms through the bacterial genome is central to refining intervention measures to counteract pathogenic bacteria.

1.4.1 Processes shaping bacterial evolution

1.4.1.1 Mutation, insertion, deletion

Point mutations are the smallest units of evolution [171]. Point mutations result from the substitution of a single nucleotide (referred to as SNV) or from the insertion or deletion of a single nucleotide. The convergence of point mutations in multiple independent lineages is an indication that the mutation is under positive natural selection [171]. Point mutations may confer a selective advantage that leads them to become fixed in the population.

SNVs that occur in protein-coding genes can yield synonymous mutations that code for the same amino acid or non-synonymous mutations that result in a change in the amino acid, potentially altering *protein function*. However, point mutations in non-coding regions can also affect *gene function,* especially if they occur in promoter regions [171]. Deletions and insertions can introduce a frameshift mutation that leads to downstream sequences being translated in a different reading frame, with major effects on protein sequence, structure and function. Similarly, nonsense mutations result in premature stop codons that truncate proteins with severe consequences for function.

1.4.1.2 Rearrangement and homologous recombination

Genome rearrangements are structural variations involving large chunks of DNA that can result from cellular process such as replication, DNA repair and DNA recombination [172]. The impact of genome rearrangements on the evolution of eukaryotes has been widely studied and there is growing evidence that genome rearrangements are important evolutionary mechanisms in bacteria [172]. Genome rearrangements can alter the order of genes as well as their orientation and copy number. These changes can affect the phenotype of an organism. Evidence shows the rearrangements can also influence gene expression [173,174].

The bacterial genome is highly flexible and can undergo large sequence variation through insertions, deletions, inversions, translocations and duplications [171].

1. **Deletions** can occur in intergenic regulatory regions and lead to gene inactivation. Deletions can also lead to loss of functional domains of proteins and altered gene function [172]
2. **Duplications** can have functional implications depending on the variation in the duplicated segment and how the segment is inserted [172]
3. **Insertions** can introduce novel segments into the genome. The most well studied mechanism of insertion is horizontal gene transfer (HGT) [175] . Mobile genetic elements (MGEs) can mediate the insertion of genomic segments from other loci within the genome. The functional implications of insertions are also governed by the genetic nature of the insert and context of the insertion loci [172]
4. **Translocations** involve the exchange of DNA segments within the same genome. They are important contributors to genome plasticity. Translocations can affect chromosome position, gene orientation and the distance between a target gene and its transcription factors [176]
5. **Inversions** flip the orientation of genomic segments

1.4.1.3 Recombination and horizontal gene transfer

Plasticity within the bacterial genome allows bacteria to adapt to selective pressures from the environment and their hosts [177]. Understanding the evolutionary mechanisms that confer genome plasticity in bacteria can yield valuable insights into their pathogenesis and adaptive mechanisms. Bacteria are able to acquire foreign DNA through HGT [175]. HGT is the largest source of variation in the genome. Uptake of foreign DNA by bacteria is mediated mainly through three processes:

1. **Transformation** involves the incorporation of exogenous DNA from the surroundings into the host
2. **Conjugation** results from direct cell-to-cell contact between cells and transfer of DNA via a sex pilus
3. **Transduction** is mediated by bacteriophages – viruses that infect bacterial cells

Recombination is a DNA repair process in bacteria that facilitates the incorporation of foreign DNA into the bacterial genome [178]. Recombination maintains genetic diversity in bacterial populations while preventing the build-up of unfavourable changes [171]. Homologous recombination classically involves the interaction of two strands of DNA with high sequence similarity. However, recombination can also account for HGT wherein highly divergent sequences are incorporated into the genome [171]. Moreover, recombination can mediate the incorporation of DNA acquired through phage mediated transduction as well as conjugation [171].

1.4.1.4 Mobile Genetic Elements

MGEs shape the evolution of bacterial pathogens by mediating movement of DNA within the genome and between species [179]. MGEs introduce diversity in the accessory genome mainly through plasmids, bacteriophages and genomic islands [180]. Plasmids are large extra chromosomal strands of DNA, linear or circular, that replicate independently and carry a wide range of genes [180]. Plasmids can carry a large number of genes that introduce complex phenotypic traits and lead to quantum leaps in evolution [181,182].

Plasmids can be transferred, mainly by conjugation, between distantly related species and even between bacteria and some eukaryotes [183].

Bacteriophages are viruses that infect bacteria. Bacteriophages can replicate within a bacterial cell and kill the cell within minutes of contact when the lytic cycle is activated [184]. However, in the lysogenic cycle the lytic cycle is not activated and the phage genome can be integrated into the host chromosome [184] to become a prophage [171]. Prophages can remain dormant in the chromosome until they are reactivated. In the pseudolysogenic cycle, the phage genome remains in the cytoplasm as an episomal plasmid [184]. Bacteriophage genomes may encode genes that can greatly enhance virulence and transmissibility such as exotoxins [185]. In some cases phages are involved in the transfer of toxins as well as their activation [186]. Phages can alter the microbiome by killing off competitors [187].

Genomic islands are regions of the genomes that are believed to have evolved through the interchange of plasmids and bacteriophages [180]. Genomic islands are present in some strains but are not conserved across a species. Genomic islands can carry genes that mediate genetic mobility but also frequently carry virulence-associated genes [180]. Pathogenicity islands are genomic islands that carry one or more virulence genes and are present in the genomes of pathogenic strains but absent in non-pathogenic strains from the same or closely related species [188]. Pathogenicity islands play an important role in mediating the virulence of bacterial pathogens in a range of host organisms [188].

1.4.1.5 Genome degradation and deletion bias

Genome degradation is believed to be an important evolutionary mechanism that sometimes contributes to the evolution of bacterial pathogenicity [189,190]. Comparing the genomes of some ancient human pathogens (*Yersinia pestis, Salmonella enterica* serovar Typhi CT18 and *Mycobacterium leprae)* with genomes from closely related species reveals extensive genome degradation and gene loss [191-193]. The observed genome degradation in these species is believed to mark a “change in lifestyle” for these organisms towards

occupying a simpler host niche [190]. Gene loss may also contribute to diminished virulence: the *Salmonella* species *S. bongori* has multiple deletion regions compared to its more virulent close relative *S. enterica* [194]. Conversely the loss of the antivirulence *cadA* gene in *Shigella flexeneri* increases the invasiveness of the species [195]. *cadA* deactivates the enzyme lysine decarboxylase, whose product inhibits the organisms plasma-encoded virulence factors [195].

Comparative genomics has revealed that some species show an evolutionary bias towards gene deletion [196]. Studying the dynamics of genome degradation can provide insights into the mechanisms bacterial pathogens employ in host specialisation. The smallest bacterial genomes occur in intracellular pathogens and symbionts [197]. The constant supply of nutrients from the host allows these bacteria to survive with a shrunken genome and reduced functionality [197].

1.4.2 Pan-genomes

All of the genes within a bacterial species make up its *pan-genome*, which is comprised of a conserved core genome and a flexible accessory genome [198,199]. The bacterial core genome largely consists of housekeeping genes that are conserved across the entire species [177]. Accessory genome elements mediate important evolutionary mechanisms such as host adaptation, virulence and antibiotic resistance [200]. Bacterial species that live in highly specialised niches tend to have closed pan genomes that are finite [201]. More generalist species possess open pan-genomes that increase with the addition of more strains, and are characterised by high rates of HGT [201].

The unprecedentedly high rates of HGT in bacterial genomes have led scientists to question the existence of bacterial species [62], particularly as one might suppose that it would not be possible for a bacterial species to maintain its identity in such a genetically promiscuous environment. However, the existence of phenotypically stable clusters despite the observed genome fluidity is explained by the core genome hypothesis. The core genome hypothesis attributes the stability of phenotypic clusters to a

repository of core genes (the core genome) that maintain the identity of a species, despite rampant changes in the accessory/flexible genome [62].

The core genome hypothesis predicts that core genes are primarily under purifying selection as a mechanism to maintain existing function [62]. Core genomes are believed, on average, to express a neutral rate of evolution while a variety of selective pressures are levied on accessory genes [62]. The core genome hypothesis is, however, challenged by the existence of open pan genomes in some species and the high rates of homologous recombination between closely related bacteria [202].

1.4.3 Bacterial evolution in monomorphic lineages

Genetically monomorphic lineages are bacterial taxa that are characterised by very low sequence diversity to such an extent that sequencing housekeeping genes from divergent strains yields few if any SNVs [88]. This trait precludes the use of standard genotyping techniques such as MLST to discriminate between strains from monomorphic lineages, and, thus, the evolutionary history of monomorphic lineages remained unclear until the widespread application of whole-genome sequencing [88].

Some of the deadliest human pathogens form genetically monomorphic lineages that have emerged from species with greater diversity. But why do monomorphic lineages occur? One explanation is that this kind of evolution intensifies disease transmission [203]. Often the population dynamics of genetically monomorphic lineages are distinct from those of the supra-species [88] e.g.:

- *Y. pestis* (causes plague, including the Black Death), is a recently emerged clone of the species *Yersinia pseudotuberculosis* [204]
- *Mycobacterium tuberculosis* (causes tuberculosis, the white plague) is part of a much more diverse *Mycobacterium tuberculosis* complex [205]
- *Bacillus anthracis* (causes anthrax) is part of the *Bacillus cereus* group of soil-dwelling saprophytes [206]

Similarly, highly virulent monomorphic clades can exist within genetically diverse species e.g.:

- The O157:H7 clade of *E. coli,* which infects cattle and can sometimes infect humans inducing haemorrhagic colitis and the haemolytic–uremic syndrome [207]
- The serovar Typhi of *Salmonella enterica* the cause of typhoid fever [208]
- The O1 or O139 clade of *Vibrio cholerae,* the cause of pandemic cholera [209]

There is evidence to suggest that Darwinian selection is a leading contributor to the evolution of bacterial pathogens [210]. Darwinian selection has been demonstrated in laboratory experiments [211] as well as in nature. For example, the widespread use of fluoroquinolones over more than twenty years has resulted in decreased susceptibility in all lineages of serovar Typhi due to accumulation of mutations in the *gyrA* gene [208]. However, Darwinian selection for a single trait sometimes leads to only transient selection, because the selected variants may prove less fit than their progenitors in the absence of the selective pressure [212]. This might explain why susceptible clones of serovar Typhi have co-evolved with the resistant clones.

Comparative genomics suggests that genetically monomorphic lineages are under neutral selection, with mild purifying selection [203]. Comparative genomics shows that synonymous and non-synonymous mutations are accumulated at similar rates in these lineages [115,205,213]. Moreover there is limited evidence of convergent evolution in these lineages since they exhibit almost negligible levels of homoplasy, with the exception of antibiotic-resistant genes [203].

Although monomorphic lineages are devoid of traces of recent homologous recombination, genomic evidence suggests the occurrence of recombination events that predated the emergence of some lineages [214,215]. The absence of HGT makes genomic deletions useful phylogenetic markers in monomorphic lineages like *M. tuberculosis* [216].

1.4.4 Phylogenetic dating

In 1962, Linus Pauling and Emile Zuckerkandl noted a linear correlation between the number of amino acid differences in haemoglobin and time, as determined by fossil evidence [217]. This formed the basis of the molecular clock theory, which states that the rate of evolutionary change in a given protein is constant over time and across different lineages [218].

The assumption that mutation rates are constant in lineages has been known to obscure phylogenetic inference. To overcome this barrier statistical methods like maximum likelihood and Bayesian inference have been employed. These methods use a “relaxed molecular clock” that assumes an independent mutation rate in each lineage [219].

In order to infer dates using a molecular clock, the mutation rate must be calibrated with independent dating evidence from sources such as fossils or ancient DNA [220]. Once a mutation rate is established for a species, it can be employed to estimate the time since divergence of closely related or distant strains.

1.4.5 Population genomics and evolution

1.4.5.1 Within-host evolution and transmission history

Classical genotyping techniques can identify the existence of mixed infection due to different lineages [221,222]. However, they lack the resolution to trace the minute evolutionary changes that occur within a single lineage during colonisation. Advances in WGS continue to shed light on the evolution of single bacterial strains within a host [223]. Microevolution is driven mainly by point mutations and in a number of species the short-term evolution rates are much higher than the long-term evolution rates [224-226]. Clouds of diversity can result from microevolution; within these clouds, variants with different phenotypic traits can potentially co-exist [227]. Changes in the host immune system or antibiotic treatment introduce selection bias that selects for the most favourable trait [227].

Most transmission studies rely on single genomes isolated from patients and do not take into consideration the effect of microevolution, which is a problem, as cloud diversity potentially hinders the ability to accurately infer transmission networks [228]. To understand cloud diversity within a host, multiple genomes need to be sampled—at different time points and/or at a single time point but from multiple colonies [229]. Moreover, heterogeneous transmission from one host to another may occur whereby multiple genotypes are transmitted [228,229]. Co-colonisation of multiple species that are part of the normal flora can lead to recombination which further complicates efforts to detect transmission chains [229].

1.4.5.2 Selection for antibiotic resistance

As noted previously, the world is going through a crisis in antibiotic resistance driven largely by overuse and inappropriate prescription of antibiotics [82]. The natural environment is replete with antibiotic resistance genes, which when transferred to pathogenic bacteria, render them resistant [230]. On top of that, bacteria can spontaneously accumulate mutations that confer antibiotic resistance. Resistant variants typically form minority populations because they are often less fit than their susceptible counterparts. However, in the presence of antibiotic treatment, a selection sweep occurs within the host-associated bacterial population so that antibiotic-resistant variants become dominant [229]. WGS has documented microevolution within a single host leading to the emergence of drug-resistant bacteria [231-233].

The large-scale use of antibiotics has rendered antibiotic resistance a selective advantage leading to the emergence of almost impossible to treat superbugs [234]. Relaxed stewardship of antibiotics in hospitals means that vulnerable patients are at an increased risk of nosocomial antibiotic-resistant infection [81]. Widespread use of antibiotics in agriculture and animal husbandry, particularly in developed countries, has also contributed to the rise in antibiotic resistance [81,235].

1.4.5.3 Vaccine-induced selective pressure

Vaccines have proven to be an effective means of eradicating or substantially reducing the prevalence of infectious diseases. However, occasionally some infections persist despite the availability of a vaccine and some vaccinated individuals remain susceptible. This may arise due to the selection of adaptive features that counteract the impact of the vaccine [221]. Adaptive features include evasion and virulence evolution [221]. Evasion arises through the emergence of novel mutants that can escape the immune response induced by the vaccine [221]. An example of this is the serotype replacement phenomenon that occurs in *S. pneumoniae* in response to conjugate vaccines [236,237].

Bacteria can alter their virulence mechanisms as an evolutionary mechanism to escape the effect of a vaccine. A classical example is the diphtheria antitoxin vaccine, which primes the immune system against the diphtheria toxin, the main virulence agent of *Corynebacterium diphtheriae* [238]. However, despite vaccination, a resurgence of diphtheria has been observed in Russia and surrounding territories. Genomic analysis showed that this resurgence was driven by the persistent circulation of non-toxinogenic strains [239]. Worryingly there is evidence that these strains are evolving into toxinogenic forms and driving a re-emergence [239]. In the UK, vaccine pressure led to an increase in the prevalence of diphtheria due to the closely related species *C. ulcerans* [240].

1.5 Bacterial infection in West Africa: three case studies

1.5.1 West Africa: geography and demographics

Africa is comprised of five regions: Northern Africa, Eastern Africa, Southern Africa, Central Africa and Western Africa. Despite being replete with natural resources, Africa is home to some of the poorest communities in the world

[241]. The WHO reports that children born in Africa are more likely to die within their first year of life than children born elsewhere. Pneumonia, diarrhoea and malaria are the principal causes of death among children under 5 years of age [242].

West Africa is comprised of 16 countries with a landmass of approximately 6.1 million square kilometres (Figure 1.1). The estimated population of West Africa in 2015 was 340 million, 55% of whom are estimated to live in rural areas with limited access to good quality healthcare [243].

Figure 1.1 A map of West Africa highlighting the 15 countries on the mainland. The geographic position of West Africa is highlighted in the thumbnail map of Africa.

West African countries, with the exception of Mauritania, form the Economic Community of West African States (ECOWAS), an economic partnership that mandates free movement of people and goods between member states [244]. This union has obvious economic benefits for the region, but it presents a unique public health challenge: the mass movement of people and goods can facilitate the spread of human infectious diseases. Cross-border cattle trading also presents an opportunity for the spread of zoonotic pathogens across the region [245].

Traditionally, government spending on healthcare has been low in West Africa. The recent Ebola outbreak in West Africa highlighted the fragility of healthcare infrastructure in the region [246]. Neighbouring countries could do little more than close their borders and watch as this deadly disease ravaged through parts of Sierra Leone, Liberia and Guinea.

Understanding the biology and epidemiology of pathogens is fundamental to a successful intervention strategy. Here, I introduce a case study of three important pathogens in West Africa. I describe a unique scenario associated with each pathogen and present WGS as a solution to the problem.

1.5.2 Staphylococcus aureus

1.5.2.1 Clinical features and pathogenesis

S. aureus is a Gram-positive commensal that can cause infection in humans. It is estimated that 50-60% of the human population are intermittently or persistently colonized with *S. aureus* [247]. The most common site of colonisation is the anterior nares, but other sites such as the skin, perineum and pharynx are also commonly colonised by *S. aureus* [248]. Colonisation with *S. aureus* places humans at risk of subsequent infection with the colonizing strain [249,250]. *S. aureus* can cause superficial skin and soft tissue infections, as well as more invasive life threatening infections such as sepsis, endocarditis, meningitis and pneumonia [251,252]. *S. aureus* is a leading cause of bacteraemia and skin and soft tissues infection globally [252-255].

Neutrophils are the host's first line of defence against *S. aureus* infection [247]. Once engulfed in neutrophils, *S. aureus* can avert apoptosis through the release of leukotoxins such as Panton-Valentine leukocidin, which promote the direct lysis of neutrophils by permeating the plasma membrane [256]. *S. aureus* employs other mechanisms to evade the host immune system and invade host tissue, which include [251]:

- **Surface proteins,** which mediate the adhesion to and invasion of host cells as well as biofilm formation. Variation of the surface proteins allows *S. aureus* to evade the host immune system
- **α-Hemolysin,** which causes host cell cytolysis and induces neutrophil recruitment. α-Hemolysin plays an important role in the pathogenesis of *S. aureus* in skin and soft tissue infection
- **Coagulases,** which interfere with the host defences by promoting coagulation

Worryingly, drug resistance has emerged for virtually all antibiotics used to treat *S. aureus* including the "last-line" methicillin and vancomycin [257]. Hospital-associated MRSA levies a significant economic burden on health care services and can lead to loss of life [258]. The rise in morbidity and mortality due to community-associated MRSA is a growing public health problem [259].

1.5.2.2 The burden of staphylococcal infection in West Africa

Population-based studies, mainly in developed countries, estimate the burden of *S. aureus* bacteraemia to be approximately 15–40 per 100,000 population per year [260]. *S. aureus* is the fourth commonest cause of invasive disease in sub-Saharan Africa and, in The Gambia, it is the second commonest pathogen isolated from patients with bacteraemia [261-263]. Despite the importance of this pathogen, serious knowledge gaps still exist regarding the epidemiology of *S. aureus* in sub-Saharan Africa. Studies in Ghana, Senegal and Nigeria show that a higher proportion of urinary tract infections are due to *S. aureus* compared to Europe [264,265]. It is likely that the burden of *S. aureus* in West Africa is underestimated since the majority of cases are

likely to go undetected or treated empirically without laboratory confirmation. Moreover in West Africa research efforts are more concentrated on other infectious agents such as HIV, tuberculosis and malaria that are greater threats to public health.

1.5.2.3 Potential for animal reservoirs in West Africa

S. aureus colonises a wide range of hosts including domestic animals [266] and wild animals [267]. The epidemiology of *S. aureus* in non-human primates has been studied in remote regions of West Africa using MLST [268]. Although these studies describe the population structure of *S. aureus* in non-human primates, they do not investigate the molecular mechanisms that govern host adaptation. Moreover, these genotyping techniques cannot ascertain whether the non-human primates can serve as a reservoir for highly virulent *S. aureus* that can infect humans.

Humans come into frequent contact with Green Vervet Monkeys in The Gambia. This interaction provides opportunities for the exchange of pathogenic bacteria between humans and monkeys. In 2011 The Vervet Consortium partnered with the MRC Unit The Gambia to investigate the epidemiology of Simian Immunodeficiency Virus in Green Vervet monkeys in The Gambia. Speculatively, nasopharyngeal swabs were collected from the monkeys and ~30% of the monkeys were asymptomatic carriers of *S. aureus.*

My task was to employ WGS to infer the evolutionary origin of *S. aureus* in monkeys in The Gambia. This involved investigating the existence of inter-species transmission of *S. aureus* and attempting to elucidate the mechanisms that govern host adaptation. Through this analysis, I sought to establish whether monkeys posed a threat as a reservoir of highly virulent *S. aureus* in The Gambia.

1.5.3 Streptococcus pneumoniae

1.5.3.1 Clinical features and pathogenesis

S. pneumoniae has a propensity to colonise the young, the elderly and the immunocompromised [269]. Successful colonisation by *S. pneumoniae* is a prerequisite to infection, which usually occurs within a month of colonisation [270]. The carriage rates of *S. pneumoniae* are higher in children under 5 compared to older children and adolescents [270]. Carriage of *S. pneumoniae* can lead to transmission through respiratory droplets and children are the main source of transmission [270].

From the upper respiratory tract, *S. pneumoniae* can invade the lungs through aspiration to cause pneumonia (by invading the alveoli) or empyema (by invading the pleura or the pericardium) [269]. *S. pneumoniae* can also enter the bloodstream and cause bacteraemia/sepsis [271]. Bloodstream invasion by *S. pneumoniae* can lead to this pathogen crossing the blood-brain barrier to cause meningitis; pneumococcal meningitis is associated with a high case-fatality rate and long-term neurological damage [271]. *S. pneumoniae* can also spread locally from the upper respiratory tract to cause otitis media and sinusitis [269].

The pneumococcal capsule, which is the main virulence factor of *S. pneumoniae,* is important for nasal colonisation [272]. Adhesion to epithelial cells is a pre-requisite for colonisation. Adhesion is mediated through the binding of pneumococcal surface proteins such as pneumococcal surface adhesin A and choline-binding protein A to cell surface carbohydrates on epithelial cells [273]. The enzyme neuraminidase reduces the viscosity of the mucosal layer making the epithelial surface more accessible for adhesion [274].

1.5.3.2 The burden of pneumococcal infection in sub-Saharan Africa

In sub-Saharan Africa, rates of pneumococcal carriage are generally high, but there is a paucity of data from most countries [275]. Rates of invasive pneumococcal disease are higher among younger children and immunocompromised individuals, due to higher rates of carriage [276]. Other risk factors for invasive disease include respiratory viral co-infection, overcrowding and lack of vaccination [270]. *S. pneumoniae* is the most common cause of pneumoniae among children and is one of the major killers among HIV infected individuals in sub-Saharan Africa [277]. In West Africa, lethal outbreaks of pneumococcal meningitis have been reported across the so-called African meningitis belt [278,279].

1.5.3.3 Vaccination and genomic outbreak surveillance

Pneumococcal conjugate vaccines are efficient intervention tools for preventing pneumococcal infection. They reduce nasopharyngeal carriage in both vaccinated and unvaccinated individuals and reduce the burden of invasive disease due to vaccine serotypes [280-282]. However, throughout most of West Africa the impact of vaccination on the molecular epidemiology of *S. pneumoniae* is unknown due to the lack of carriage surveillance or extensive disease surveillance.

The pneumococcal meningitis outbreak that broke out in the Brong-Ahafo region of Ghana in December 2015 was a reminder that vaccination is not always the end of the story. This outbreak occurred two years after the introduction of the 13-valent pneumococcal conjugate vaccine, PCV-13.

During the outbreak sample integrity was a major shortfall and the majority of samples were not viable for culture. The cause of the meningitis outbreak in Ghana would thus have remained unknown without the use of the quantitative PCR assay for detection of bacterial pathogens from human cerebrospinal fluid [283]. Quantitative PCR identified serotype 1 *S. pneumoniae,* a vaccine serotype, as the main etiological agent from the

outbreak [283]. But how could this happen in a country where PCV-13 had been administered to over 80% of the children under five years of age?

Fortunately, a handful of viable isolates of *S. pneumoniae* were salvaged from the outbreak and as described in a subsequent chapter, were analysed using WGS. Previously sequenced serotype 1 genomes from West Africa served as a baseline for comparing the outbreak strains. This made it possible to identify evolutionary changes that distinguished the outbreak strains. This analysis demonstrated the power of genomics in outbreak surveillance and reiterated the fact that to characterise an outbreak properly, it must be analysed in the context of the global phylogeny [108].

1.5.4 Mycobacterium tuberculosis

1.5.4.1 Clinical features and pathogenesis

Tuberculosis is caused by a group of closely related ecotypes that form the *Mycobacterium tuberculosis* complex [284]. Mycobacteria are slow growing, intracellular, non-motile, non-sporulating, acid-fast bacilli. Mycobacteria have a very thick cell wall, uniquely multi-layered, with varied hydrophobicity. The inner layer consists of hydrophilic peptidoglycan sacculus and arabinogalactan layers, covalently linked to a hydrophobic outer layer of mycolic acids that form a waxy mesh of long chain fatty acids [285,286].

A hallmark of the cell-mediated immune response to tuberculosis infection is the formation of granulomas [287]. Granulomas can suppress the spread of bacteria within tissues by trapping the bacilli within macrophages in a mesh of immune cells [288]. However, mycobacteria can survive in an encapsulated form in lung granulomas for several years in a state of latency [288]. In a tenth of cases, the tubercle bacilli will escape from the granuloma and spread through the body to cause active disease.

The most common site of tuberculous infection is the lungs, but other extra-pulmonary sites including the bones, brain and kidneys can also become infected [289]. Tuberculosis is transmitted by aerosol from patients with active disease through coughing, sneezing, laughing or even talking [289]. Symptoms

of tuberculosis may include persistent cough, productive cough, malaise, night sweats, weight loss, loss of appetite and fever [289].

Treatment of tuberculosis is complicated by the emergence of resistance to anti-TB drugs. Efforts to eradicate tuberculosis have been set back by the emergence of MDR-TB and extensively drug resistant (XDR-TB) forms of tuberculosis. MDR-TB is defined as resistance to at least isoniazid and rifampicin, the two most powerful first-line anti-TB drugs [290]. XDR-TB are MDR-TB that are also resistant to a fluoroquinolone and one of the second-line injectable drugs (amikacin, kanamycin or capreomycin) [291].

1.5.4.2 The burden of mycobacterial infection in West Africa

Tuberculosis is a major cause of death globally and the highest-burden regions are Africa and Southeast Asia. In 2015, over 1.3 million cases of tuberculosis were notified in Africa [292]. However, among the notified cases of pulmonary tuberculosis, only 64% were bacteriologically confirmed [292]. Most health facilities in the continent lack the facilities for bacteriological confirmation of tuberculosis and rely on expert opinion based on clinical symptoms. This highlights an urgent need to strengthen diagnostic capacities in Africa.

In 2015, there were an estimated 480,000 new cases of MDR-TB [292]. Estimating the burden of MDR-TB in Africa is challenging due to lack of accurate surveillance data. In 2015, the reported treatment success rates among MDR-TB and XDR-TB patients were 54% and 24% respectively [292]. Poor adherence due to lack of awareness among patients and shortages of medication are a major factor is treatment failure. However it is worth considering the impact of misdiagnosis at primary health care facilities on the treatment outcomes of MDR-TB and XDR-TB. Most primary health facilities will not be able to perform drug susceptibility testing and are likely to place patients with MDR-TB on an inadequate treatment regimen before they are referred to the tertiary referral centres for retreatment.

Between 2009 and 2013, the West African Nodes of Excellence for Tuberculosis AIDS and Malaria (WANETAM) carried out surveys in eight

West African countries to estimate the true burden of drug-resistant tuberculosis [293]. At the time, only Senegal, Gambia and Nigeria had performed drug-resistance surveys in the sub-region. The estimates of drug resistance from WANETAM were higher than WHO estimates, suggesting that the burden of MDR-TB in West Africa could be underestimated [293]. In particular, the prevalence of MDR-TB among retreatment cases in Mali (59%) and Nigeria, Ibadan and Lagos (39 % and 66 %), was significantly higher than WHO estimate for prevalence of MDR-TB among retreatment cases in the African region (20%) [293].

1.5.4.3 Knowledge gaps for tuberculosis in West Africa

Molecular epidemiology studies that employ classical genotyping techniques have been carried out across West Africa. These studies involve a range sampling strategies and isolated communities, making it challenging to compare the epidemiology across countries [294]. MTBC in West Africa is highly diverse, with six of the seven global lineages present in the region [294]. Despite this rich diversity very little is known about the evolutionary history of MTBC in West Africa, especially of the lineages and genotypes that are exclusively endemic to this region. Moreover, classical genotyping techniques do not have the resolution to distinguish between closely related strains and to infer on going transmission.

The WANETAM dataset of over 900 archived MTBC isolates presents an ideal platform through which we can begin to understand the evolution of MTBC in this region. WGS holds the potential to offer novel insights into the phylogeography of MTBC in West Africa. WGS has the power to unearth on going chains of transmission and even detect microevolution of MTBC within the host over the duration of treatment. With the phenotypic drug-resistance profiles already known, the genome can be studied to understand the molecular mechanisms that are driving drug resistance in this setting.

1.6 Aims and objectives

The time has come for West Africa to embrace the genomics era and exploit the full potential of microbial genomics. Genomics offers unprecedented insights into the epidemiology and biology of infectious diseases, which dominate the public health agenda in West Africa. In this thesis, I aim to exploit the sequencing and analysis of whole bacterial genomes to understand the evolution and spread of bacterial infection and antimicrobial resistance in West Africa. My work focuses on three important human pathogens in different contexts.

1. **In Chapter 3**, I will evaluate the use of whole genomic sequencing to study the molecular epidemiology of *S. aureus* from humans and monkeys in The Gambia. I aim to determine whether transmission has occurred between hosts and, if it has, I aim to identify the underlying evolutionary mechanisms that accompany host adaptation. Through this work, I hope to demonstrate the applicability of genomics to investigate animal reservoirs for deadly human pathogens in West Africa
2. **In Chapter 4**, I will assess the value of WGS in unravelling the genomic epidemiology of a pneumococcal meningitis outbreak that occurred in Ghana after the introduction of the pneumococcal conjugate vaccine. Here, I aim to identify the factors that led to the outbreak and to demonstrate the benefits of genomics in outbreak surveillance
3. **In Chapter 5**, I will explore the genomic epidemiology of tuberculosis in West Africa. I employ pilot datasets from two West African countries, Mali and Nigeria. Through genomics, I investigate the phylogenetic relationships of strains causing tuberculosis in West Africa and the molecular mechanisms that are driving drug resistance in this setting. Through this work I hope to showcase the applicability of genome-scale monitoring to the evolution and the spread of antibiotic resistance in a major pathogen in West Africa.

2 CHAPTER TWO: METHODS

2.1 Bacterial cultures

Colleagues at the MRC Unit The Gambia (MRCG) performed identification, characterisation and antimicrobial susceptibility testing for *S. aureus, S. pneumoniae* and MTBC.

2.1.1 Staphylococcal culture and identification

S. aureus strains were isolated from archived clinical specimens (venous blood pleural fluid, cerebrospinal fluid, joint aspirate, lung aspirate, ascitic fluid, and lymph node aspirates) and from nasopharyngeal and oropharyngeal swabs collected from healthy humans and monkeys using sterile calcium alginate swabs. Swabs were inoculated into chilled skim-milk tryptone glucose glycerol transport medium immediately after collection. All swabs were vortexed for fifteen seconds and stored at -70ºC within eight hours of collection. Specimens were plated on Mannitol Salt Agar and 5% sheep blood agar for 48 hours and 24 hours respectively under aerobic conditions (Oxoid, Basingstoke, UK). Suspected *S. aureus* colonies were sub-cultured on blood agar and confirmed using the Slidex Staph kit (Biomerieux, Basingstoke, Hampshire, UK).

S. aureus isolates were characterised by MLST, targeting seven housekeeping genes, *aroE, ptA, glp, arcC, gmK, tpi,* and *yqiL* as described previously [97]. Sequence types (STs) were grouped into clonal complexes (CCs) by performing an eBURST analysis [295] of all STs in the MLST database (http://saureus.mlst.net) (Appendix 2.1). Occasionally eBURST grouped two or more previously described CCs into one eBURST group. In such cases, the known CC designations were retained in order to maintain consistency with the existing literature.

2.1.2 Pneumococcal culture and identification

Cerebrospinal fluid specimens were collected from suspected meningitis cases in the Brong-Ahafo region by performing a lumbar puncture. Specimens were plated on Columbia blood agar and chocolate agar for overnight incubation. Suspected pneumococcal colonies were confirmed based on growth characteristics, Gram stain and the optochin test (5µg optochin disks; Oxoid Basingstoke, UK). Serotyping using latex agglutination was performed on all confirmed pneumococcus isolates as previously described [296]. Nucleic acid extraction was performed using the Qiagen DNA Mini kit (Qiagen, UK) following manufacturer's instructions. Serotyping was confirmed by qPCR at the World Health Organization Regional Reference Laboratory for Invasive Bacterial Diseases hosted at the MRCG.

2.1.3 Mycobacterial culture and identification

Mycobacterial isolates on Löwenstein–Jensen media were sent to the MRCG from WANETAM collaboration sites [293]. Isolates were sub-cultured in the MGIT 960 system (Becton Dickinson, Oxford Science Park, Oxford, UK) to maintain viability. Positive cultures were tested for purity by inoculation on blood agar, and followed up only if there was no growth after 24 hours incubation. The presence of MTBC was confirmed by Ziehl–Neelsen staining.

In order to extract whole-genome DNA I sub-cultured isolates from the MGIT tubes into Middlebrooks 7H9 liquid medium (Sigma Aldrich, Gillingham, Dorset, UK). Middlebrook 7H9 medium was prepared according to manufacturer's instructions:

1. 4.9g of broth stock dried medium was dissolved in 900ml of distilled water with stirring at room temperature
2. The media was autoclaved
3. 2ml of glycerol was added to the hot media before cooling to 5ºC
4. 100ml of Oleic Albumin Dextrose Catalase enrichment medium (Sigma Aldrich, Gilling, Dorset, UK) was added in a safety cabinet and mixed by pipetting up and down with a 25 ml pipette

5. 10-15ml of the enriched 7H9 medium was aliquoted into 25ml tissue culture flasks (ThermoFisher Scientific, MA, USA)
6. The top of the flasks were sealed with tape to prevent the media from drying
7. Flasks were incubated at 37ºC for 48 hours to check for purity (contamination would lead to turbidity)
8. Flasks were stored at 5ºC

Isolates were grown on 7H9 for 3-8 weeks and the presence of cells at the bottom of the culture flask confirmed growth. MTBC was confirmed by the rapid test Capilia™ TB-Neo (Sigma Aldrich, Gillingham, Dorset, UK) and BD MGIT™ TBc (Becton, Dickinson and Company, Oxford Science Park, Oxford, UK). To test for purity the culture medium was streaked onto blood agar and incubated at 37ºC for 48 hours. Blood agar plates were checked for growth and referred to the senior laboratory microbiologists for confirmation. Contaminated samples were decontaminated by:

1. Transferring the sample to a 50 ml falcon tube and adding an equal amount N-Acetyl-L-Cysteine Sodium Hydroxide solution and vortexing
2. Incubating at room temperature for 15 minutes
3. Adding phosphate buffer up to the 50 ml mark
4. Centrifuging at 3000 x g for 15 minutes
5. Decanting the supernatant in 5% Hycolin
6. Re-suspending pellet in 1ml phosphate buffer
7. Inoculating decontaminated isolates into a fresh 7H9 culture flask

2.1.4 Antimicrobial susceptibility testing

S. aureus isolates were screened for antimicrobial resistance by disc diffusion with penicillin, cotrimoxazole, tetracycline, chloramphenicol, gentamicin, cloxacillin, erythromycin and cefoxitin (Oxoid, Basingstoke, UK) according to Clinical and Laboratory Standards Institute guidelines [297]. For pneumococcal isolates, antimicrobial susceptibility testing was done for clindamycin, vancomycin, trimethoprim-sulfamethoxazole, erythromycin,

chloramphenicol and rifampicin using disc and E-test diffusion testing following CSLI guidelines.

For MTBC, susceptibility testing for the first-line drugs streptomycin (STR, 1 µg/mL), isoniazid (INH, 0.1 µg/mL), rifampicin (RIF, 1µg/mL), and ethambutol (EMB, 4.5 µg/mL) was performed on the MGIT 960 system (Becton Dickinson, Oxford Science Park, Oxford, UK) according to manufacturer's instructions [298]. Multidrug-resistant strains (resistant to isoniazid and rifampicin) were further tested for susceptibility to the second-line drugs capreomycin (CAP, 2.5 µg/mL), ofloxacin (OFX, 2 µg/mL), and ethionamide (ETH, 5 µg/mL) (Sigma-Aldrich, St. Louis, Mo, USA) [293].

2.2 Sequencing

2.2.1 Attribution of effort

Colleagues at the MRCG performed DNA extractions for *S. aureus* and *S. pneumoniae* strains. I performed the DNA extractions for all MTBC strains. All the *S. aureus* and MTBC strains as well as eight of the outbreak *S. pneumoniae* strains were sequenced at the University of Warwick on the Illumina MiSeq following Nextera XT library preparation. Seven outbreak strains were sent to the Crick Institute, UK for sequencing after failing to sequence in Warwick. At the Crick the libraries were prepared with the TruSeq library preparation kit and sequenced on the Illumina MiSeq. Contributions to library preparation and sequencing were as follows:

- ***S. aureus:*** Madikay Senghore (50%) and Emma Doughty (50%)
- ***S. pneumoniae:*** Madikay Senghore (30%), Gemma Kay (25%) and Crick Institute (45%)
- **MTBC (Nigeria):** Madikay Senghore (70%) and Emma Doughty (30%)
- **MTBC (Mali):** Madikay Senghore (80%) and Gemma Kay (20%)

2.2.2 DNA extraction

For *S. pneumoniae* and *S. aureus* strains, genomic DNA was extracted from fresh overnight cultures using the Qiagen DNA Mini kit and Qiagen Genomic DNA extraction Kit (Qiagen, UK) respectively, following manufacturer's instructions. Briefly, following lysis, the lysate suspension was aliquoted into the Qiagen separation column. The lysate was washed with manufacturer's wash buffers before being eluted as pure genomic DNA.

For the MTBC, I performed genomic DNA extraction from 3-8 weeks cultures using the cetyl trimethylammonium bromide (CTAB) method [299]. Briefly:

1. The culture flasks were slanted against the safety cabinet to concentrate cells at the bottom of the flask
2. 1.5 ml of cell suspension was transferred into a 2mL Safe-Lock Eppendorf tube
3. 1.5 ml of cell suspension was mixed with 1.5 ml Tryptone soya broth and aliquoted into two barcode labelled 2 ml cryovials (ThermoFisher) for storage at -70ºC
4. The cells were heat killed at 80ºC for at least 20 minutes
5. The suspension was spun for 5 min at 9000rpm and the supernatant was decanted in a 5% Hycolin waste bucket
6. Cells were suspended in 450µL Tris/EDTA (TE) buffer (Sigma Aldrich, Gillingham, Dorset, UK)
7. Cells were lysed by adding 50µL of 10mg/ml lysozyme and shaking overnight at 37ºC
8. Proteins were denatured by adding 100µL of 20% Sodium dodecyl sulphate and 10µL proteinase K then incubating at 60ºC for 40 minutes
9. The DNA was released by adding 100µL of 5M NaCl (Sigma Aldrich, Gillingham, Dorset, UK) and then adding 100µL of NaCl/CTAB solution and vortexing
10. The DNA was separated by adding chloroform/Isoamyl alcohol (24:1) (Sigma Aldrich, Gillingham, Dorset, UK) and spinning for 8 minutes at 11000 rpm

11. The aqueous layer containing the DNA was carefully transferred to a fresh 1.5 mL Eppendorf tube using a P200 pipette
12. Finally, the DNA was stained with Glycoblue (ThermoFisher Scientific, MA, USA) and resuspended by adding cold Isopropanol stored at -20ºC (Sigma Aldrich, Gillingham, Dorset, UK)
13. The DNA was washed with cold 70% ethanol and resuspended in 200µL TE buffer for storage at -70ºC

2.2.3 Library preparation and sequencing

Genomic DNA libraries were prepared for whole genome sequencing using the Nextera XT library prep kit according to manufacturer's instructions (Illumina, Little Chesterford, UK). Briefly, input DNA was fragmented using a tagmentation step at 55°C. The fragmented DNA was tagged with adapters and indexed for multiplex sequencing. The tagged DNA fragments were then amplified by PCR and the PCR products were purified using Ampure beads. Purified PCR products were quantified using the Qubit® fluorometer (ThermoFisher, MA, USA). Libraries were typically diluted to 4ng/µl. To ensure an equal representation of all samples, a volume representing 40ng of each PCR product (typically 10µl of a 4ng/µl dilution) was added to the pool. Samples were sequenced on an Illumina MiSeq following the manufacturer's instructions (Illumina, Little Chesterford, UK).

2.3 Sequence analysis

The bioinformatics and sequence analysis was performed on a LINUX server at the University of Warwick and on an Ubuntu instance running the Genomics Virtual Lab on the Cloud Infrastructure for Microbial Genomics (CLIMB) [300]. Andrew Millard administers the Pallen group server. The CLIMB instance came preloaded with numerous bioinformatics tools including the Nullarbor pipeline for clinical microbiologists. My collaborators at the University of Bath and at St George's University of London performed analyses on their own servers.

2.3.1 Genome assembly and annotation

For each strain *de novo* contigs were generated from the paired-end sequencing reads using SPAdes (kmers: 21, 33, 55, 77, 99 and 127) [301]. Draft assemblies were annotated with prokka [302]. *S. aureus* genomes were filtered to remove contigs that were likely to be contaminants. This was achieved by removing short contigs (< 300 bp) that had a low coverage (kmer coverage < 2). This information was retrieved from the headers of each contig generated by SPAdes.

2.3.2 MLST and antibiotic resistance prediction from whole genome for *S. aureus*

The MLST profile of each *S. aureus* strain was inferred from the genome sequence. Initially, I used an assembly approach whereby each draft assembly was compared to all alleles in the MLST database (http://saureus.mlst.net) using BLAST with default parameters. Alleles were called based on 100% sequence coverage and 100% nucleotide identity. Some strains had unresolved profiles either due to partial coverage of an allele or less than 100% nucleotide identity with a given allele (Appendix 2.2).

My collaborators at the University of Bath repeated this analysis using the SRST2 tool to map raw sequence reads from each strain to all MLST alleles and call alleles based on the best match [303] (Appendix 2.3). The results from the mapping approach were compared to the results from the BLAST approach and they agreed in most instances. The alleles that could not be resolved by the BLAST comparison were assigned based on the results of the mapping approach. Genomic MLST profiles were compared to MLST profiles reported from the laboratory as a quality control. Eleven isolates that showed discordance between genomic and laboratory MLST profiles were excluded from further analysis.

The antibiotic resistance profiles for *S. aureus* strains were predicted by uploading sequencing reads onto the Mykrobe predictor tool [304]. Mykrobe

predicted antibiotic resistance to the main antibiotics: Beta-lactams (methicillin, penicillin), quinolones (ciprofloxacin), macrolides/lincosamides (erythromycin, clindamycin), tetracycline, aminoglycosides (gentamicin), glycopeptides (vancomycin), rifampicin, mupirocin, fusidic acid and trimethoprim based on the presence of whole genes as well as point mutations.

2.3.3 In silico serotyping for S. pneumoniae

Sequencing reads from each isolate were mapped to a multiFASTA file containing the sequences of the capsular region for 95 pneumococcal serotypes [305] using SMALT. Each capsular type was concatenated into a single contig in the multiFASTA file. Samtools version 0.1.18 was used to convert the SAM file to a BAM file and to sort the BAM file. The depth utility on samtools calculated the depth coverage for each site from the sorted bam file. I wrote a Perl script to calculate the average coverage depth and the proportion of the sequence covered by at least one read for each capsular type (averagecoverage_cps.pl). The average coverage was calculated as the sum of depth coverage at all sites divided by the number of sites. Serotype was assigned based on the capsular type with the highest proportion of its sequence covered by at least one read.

2.3.4 Analysis using Nullarbor

S. pneumoniae strains were analysed using the Nullarbor pipeline [306] , which performed a number of steps:

- Sequencing reads were trimmed and filtered for low quality bases and reads by trimmomatic [307]
- The species of each isolate was determined by a kmer analysis of a database of known genomes using kraken version 0.10.5-beta [308]
- *De novo* assemblies were generated using MEGA-HIT [309] and annotated with prokka [302]. Although Nullarbor has an option to

generate assemblies using SPAdes, this option proved too memory-intensive and caused the program to crash

- MLST profiles [310] were inferred from the assemblies using Seeman's MLST program [311]
- Resistance genes were detected from the assembled genomes using abricate, which BLASTs contigs against a database of known antimicrobial resistance genes [312]

2.3.5 Lineage assignment and drug resistance prediction for MTBC

Three tools were used to infer phylogenetic lineage and antibiotic resistance profiles from MTBC genomes:

- Kvarq run locally, which predicted drug resistance based on canonical mutations [313]. However, kvarq predicted resistance in only 66% of the phenotypically isoniazid resistant isolates. Kvarq placed strains into the seven global lineages, but did not offer phylogenetic placement of strains into the sub-lineages
- PhyResSe and TBProfiler were web-based applications that also predicted drug resistance and placed strains into phylogenetic lineages [314,315]. PhyResSe and TBProfiler provided improved resistance prediction for isoniazid resistance, as they scanned through a wider repository of resistance mutations. Both tools placed strains into the sub-lineages of the Euro-American super-lineage

2.4 Phylogenetic analyses

2.4.1 Staphylococcus aureus

SNVs in the core genome of *S. aureus* were identified using a customized pipeline built by our collaborators at the University of Bath [316]. Briefly, paired-end reads from each genome were mapped against the well-annotated

EMRSA15 reference genome (Accession number: HE681097) using SMALT [317]. Known repeat regions, insertion sequences and mobile genetic elements (identified by our collaborators at Bath) were excluded and SNVs were called using Samtools (version 0.1.18), the GATK toolbox and custom Perl scripts [318,319]. A maximum-likelihood phylogeny was reconstructed using FastTree on default parameters [320].

In line with published estimates, the intra-clonal mutation rate for *S. aureus* was assumed to be ~ 2 x 10^{-6} per site per year, with a core genome size of ~2.5 Mb, and was assumed to be constant across lineages and over time [321]. This equates to a short-term mutation rate of 5 SNVs per year, therefore dividing the number of SNVs between two strains by ten gives an estimate of the number of years since divergence from a most recent common ancestor (MRCA).

$$Time\ since\ MRCA = \frac{Number\ of\ SNVs}{Mutation\ rate \times number\ of\ sites \times 2}$$

2.4.2 Streptococcus pneumoniae

Sequencing reads of *S. pneumoniae* strains from the meningitis outbreak in Ghana were mapped onto the spn1041 reference genome (Accession number: PRJNA50807) using Snippy and Snippy-Core within the Nullarbor suite to create a multiple sequence alignment of SNVs in the core genome [322]. Nullarbor then created a phylogenetic tree using FastTree. This phylogenetic tree was visualised in the context of the geographic origin of the strains. The Nullarbor suite also calculated a distance matrix based on the pairwise SNV difference between strains.

In a subsequent analysis, draft assemblies were generated from *S. pneumoniae* serotype 1 isolates using MEGA-HIT. A quality check was performed by self-mapping sequencing reads from each strain to its draft assembly to ensure that each base in the assembly was supported by at least four reads. These assemblies, together with 108 archived draft genomes from West African serotype 1 isolates (68 from ST303 and 44 from ST217), were mapped against the spn1041 reference genome and SNVs in

the core genome were called using core.ope, an assembly-based SNV-calling pipeline created by Dr Zhemin Zhou (University of Warwick) [323].

The core.ope method initially aligned each assembly to the spn1041 reference genome using the nucmer function in MUMmer [324] and placed all aligned query genomes in a multiple sequence alignment with the reference genome. Custom Perl scripts drawing on a user-generated table with the start and stop positions of all insertion sequences in the reference genome were used to remove insertions. The mobile genetic elements in the spn1041 reference genome were listed in the annotation provided by the Sanger institute (Appendix 2.4). I searched the CRISPR database for CRISPRs in the spn1041 reference genome but none were listed. The core genome was assigned as the regions that were conserved in all genomes that were not insertion sequences. (A folder with the core.ope source code is available in the Appendix folder)

Variants were called from the core genome and recorded in a SNV table listing variant calls in each strain at every polymorphic site in the core genome. The SNV table was converted into a multiple sequence alignment, which was then loaded into RAxML: Randomized Axelerated Maximum Likelihood to reconstruct a maximum likelihood phylogeny using a general time reversible model for substitution [325].

The multiple sequence alignment of the outbreak serotype 1 strains and 108 West African serotype 1 strains was inputted into Gubbins and recombination events were inferred with default parameters [326]. The drawing tools in Gubbins were used to present recombination events as a heatmap adjacent to a phylogenetic tree reconstructed by Gubbins.

Genotypic diversity indices for serotype 1 strains were computed using R (3.3.1) statistics tools and R packages (*poppr 2.2, adegenet 2.1*). The SNV table inferred from the core genome of the outbreak serotype 1 strains and 108 West African serotype 1 strains was inputted into R. Strains were clustered into three groups:

- West African ST303 clade
- Novel ST303 clade

- West African ST217 clade.

Allele frequencies and the Simpson diversity index were calculated for each variable site. As a quality control, uninformative sites that did not distinguish between the three clusters and sites that had a Simpson's index < 0.1 were not considered. The Shannon diversity index for each cluster was calculated to determine the level of diversity within the cluster [327]. The index of association (IA) and the standardized index of association, rÂ¯d were computed for each cluster to determine the level of clonality within each cluster [328].

2.4.3 Mycobacterium tuberculosis complex

The phylogeny of MTBC isolates from Nigeria was reconstructed using a custom pipeline built by Dr Adam Witney (St George's University of London) [329]. Sequencing reads were mapped on to the H37Rv reference genome and variants were called using Samtools and custom Perl scripts. Variants were called from sites that were covered by at least two reads in either direction, were supported by at least 75% of reads and had a quality score of at least 30. An alignment of variants in the core genome from all strains was generated and the maximum likelihood phylogeny was reconstructed using RAxML with a general time reversible model for substitution.

The core.ope pipeline was employed to reconstruct the phylogeny of MTBC isolates from Mali. Draft assemblies generated with SPAdes were mapped to the H37Rv reference genome and SNVs were called from the core genome as described above. The repeat regions in the H37Rv genome, which were excluded from the core genome were identified by Dr Martin Sergeant (University of Warwick) as part of his work on tuberculosis in ancient mummies [330]. A distance matrix was generated by uploading the multiFASTA alignment of core genome SNVs onto MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 [331], which computed the number of differences between strains. The maximum likelihood phylogeny with RAxML using a general time-reversible model of substitution was reconstructed.

2.4.4 Tree visualisation and annotation

Phylogenetic trees were visualised with their associated metadata using FigTree 1.4.2 and were annotated on inkscape. The following context-specific colouring approaches were used:

- For *S. aureus,* tips were coloured based on the source of the isolate (i.e. invasive disease, carriage, reference or monkey) and the branches were coloured to represent the major clades in the phylogeny
- Outbreak *S. pneumoniae* strains were placed into three groups: Novel ST303, West African ST303 and non-serotype 1. Unique tip colours were assigned to each group
- For the broader serotype 1 phylogeny, branches were coloured to show three major clades (1) West African ST303 clade, (2) Novel ST303 clade and (3) West African ST217 clade and tips were coloured based on the country of origin of the strain. The novel clade was highlighted on the phylogenetic tree
- For MTBC, the branches were coloured according to major lineage (lineage 1-7) and the tips were coloured according to sub-lineage

2.5 Accessory genome analysis

An accessory genome analysis for the *S. aureus* dataset was performed using SPINE and AGEnt [332]. To infer a core genome from our dataset, a representative genome was selected from each clade or clonal complex in our phylogeny. The best quality assemblies were identified according to assembly size, number of contigs and n50. Where possible, a published closed genome was used to represent major clonal complexes. The representative dataset included 35 genomes. SPINE used Mummer to align the genomes and then extracted the core genome by identifying genes that

were conserved in all 35 genomes. AGEnt inferred the accessory genome of each query genome by removing the representative set of core genes.

A Bio Perl script authored by Dr Sion Bayliss (University of Bath) was used iteratively to identify orthologous groups of coding sequences (CDSs) between the accessory genomes (FindSharedGeneContent.pl). The scripts iteratively took each CDS in the accessory genome of each strain and searched the accessory genome of all other strains for a homolog using BLAST. To prevent redundancy, CDSs that already had a match were not compared in subsequent queries. For each CDS, presence or absence of an orthologue was scored based on 85% sequence identity at the protein level, while CDSs generating multiple hits in any genome were ignored. A pairwise matrix was generated showing the proportion of shared accessory genome (PercentID.pl). An R script authored by Dr Sion Bayliss plotted the proportion of shared accessory genome as a heatmap linked to the phylogenetic tree (Plot_Gambia.R).

3 CHAPTER THREE: HUMAN TO MONKEY TRANSMISSION OF *STAPHYLOCOCCUS AUREUS* IN THE GAMBIA

3.1 Introduction

3.1.1 Genome evolution in *S. aureus*

The *Staphylococcus aureus* genome consists of a circular chromosome of approximately 2.8 Mb [333] together with, in some strains, one or more extra-chromosomal plasmids that can carry virulence and antibiotic resistance genes [334]. *S. aureus* is characterised by high rates of horizontal gene transfer (HGT). This leads to marked strain-to-strain diversity in the repertoire of mobile genetic elements (MGEs) and, consequently, in repertoires of virulence and resistance genes [335,336]. MGE content can vary remarkably even between strains within the same clonal lineage [337]. MGEs make up 15-20% of the *S. aureus* genome [337,338] and include:

- Transposons
- Bacteriophages
- Plasmids
- *S. aureus* pathogenicity islands (SaPIs)
- The staphylococcal cassette chromosome mec (SCCmec), which encodes methicillin-resistance

The *S. aureus* core genome makes up ~75% of the genome content in any particular strain [335,339]. There is marked diversity at the core genome level between the independently evolving clonal lineages of *S. aureus* [340]. Diversity within the core genome is driven mainly by accumulation of point mutations [341]. Core genome transfer is also a source of diversity in the core genome [340]. However, although rates of core genome transfer between closely related strains are low [97], for basal branches of the species tree, there

is strong evidence for high rates of recombination between clonal complexes [340].

3.1.2 Genomic Islands

Genomic islands (GIs) are regions of the genome, 10–200 kb in length, that have an atypical nucleotide composition and contain clusters of genes acquired through HGT [342]. GIs can carry a wide variety of genes that often confer mechanisms for microbial adaptation to the host and environment [342]. Most *S. aureus* strains carry a diverse array of GIs, including three highly polymorphic genomic islands (vSaα, vSaβ and vSaγ) that carry genes for superantigens, lipoproteins and proteases [339,343]. Diversity of vSas is driven by gene loss, high recombination rates and transferability by transducing phage particles [339,344].

3.1.3 Evolution of drug resistance

Shortly after the introduction of penicillin in the early 1940s, penicillin-resistant *S. aureus* emerged [345]. Early penicillin resistance was conferred through acquisition of the *blaZ* gene. *blaZ* encodes for the production of β-lactamase, which hydrolyses the beta-lactam ring [346]. Since then, drug resistance has emerged for virtually all antibiotics used to treat *S. aureus,* even including drugs such as methicillin and vancomycin [257].

Methicillin resistance is conferred by the *mecA* gene, which is carried by the Staphylococcal chromosome cassette mec (SCCmec). *mecA* encodes an altered penicillin binding protein that has a low affinity for all beta-lactam antimicrobials and their derivatives [347,348]. Methicillin-resistant *S. aureus* (MRSA) spread mainly through clonal inheritance, but the SCCmec can also be transferred between strains [349]. MRSA was originally associated with hospital infection, but over the past three decades a remarkable increase has been observed in the prevalence of community-associated MRSA [345].

The unprecedented rise of antibiotic resistance in *S. aureus* has led to increased use of vancomycin for treatment [350]. The selective pressure due to

widespread vancomycin usage has led to the emergence of vancomycin resistant strains of *S. aureus* [351].

3.1.4 Population structure of human *S. aureus*

Staphylococcus aureus is a largely clonal organism, which means that strains cluster into several widespread clonal complexes (CCs) [97,352]. Aside from similar sequence alleles among core genes, each clonal complex is defined by possession of a common set of lineage-specific variable genes [335,337]. However, some variable genes are shared between lineages through vertical transfer and subsequent sequence divergence [335].

Globally ten dominant lineages (CC1, CC5, CC8, CC12, CC15, CC22, CC25, CC30, CC45 and CC51) are believed to cause the majority of human infection [97]. Occasionally localised lineages of *S. aureus* emerge that are predominant in particular regions. For example, CC152 is uncommon in Europe, but is dominant in Mali, in Ghana and probably throughout much of sub-Saharan Africa [118,353,354]. In Japan, the predominant community-associated methicillin resistant lineage CC89 is believed to have emerged locally and evolved under strong selective pressure from antibiotics [355].

3.1.5 Genomic epidemiology of *S. aureus* infections

When the first *S. aureus* genome was sequenced 15 years ago it revealed a vast repository of virulence genes and the contribution of HGT to genome diversity [356]. Whole genome sequencing (WGS) offered a new perspective on the population structure and evolution of *S. aureus,* which built on knowledge from phenotypic methods and genotyping techniques [357] (Table 3.1). WGS has enhanced our understanding of the transmission and spread of MRSA both in the community and in hospital settings [266,321,358]. During an outbreak of *S. aureus* chains of transmission can be identified through WGS [359,360]. Longitudinal studies have revealed fluctuations in gene diversity that occur within the host during asymptomatic carriage and subsequent invasive disease [236,361]. Routine WGS of *S. aureus* in Europe has provided insights into the transmission dynamics of *S .aureus* and served as an early warning

for the emergence of highly virulent clones [118]. In sub-Saharan Africa major knowledge gaps exist in the molecular epidemiology of *S. aureus* highlighting the need for genomic surveillance [362].

Table 3.1 Use of genomic epidemiology to track the evolution of *S. aureus* infection. Important citations are listed and a brief summary is provided for each study.

Citation	Region	Major findings
360	Global	Global epidemiology of the hospital associated MRSA lineage ST239. Evidence for intercontinental spread of the lineage. Convergent evolution on antibiotic resistance genes
363	Oxfordshire, UK	Sequential strains sequenced from asymptomatic carriers of *S. aureus.* Evidence of evolution in asymptomatic carriage dominated by purifying selection.
358	South England, UK	Genome analysis used to infer the routes of transmission of MRSA in a hospital setting. Evidence that patient-to-patient transmission is not the main source of infection. Importantly WGS disproved patient-to-patient transmission inferred by inferior genotyping techniques and detected unsuspected cases of transmission.
321	United States	The evolution of the community-acquired MRSA lineage ST8 (USA300) driven by household reservoir. The expansion of the clone is also driven by emergence of fluoroquinolone resistance
359	Thailand	Diversity of ST239 MRSA lineage in a low resource setting hospital. Remarkable level of diversity in a single hospital. Nosocomial spread and microevolution within a single patient.
364	Denmark, Copenhagen	Monitoring MRSA through routine WGS and epidemiological data. MRSA genome database enhances understanding of transmission.
304	-	Mykrobe predictor, a milestone in genomic epidemiology of *S. aureus*. Prediction of antibiotic resistance from sequencing reads within minutes on a laptop. Error rates comparable to gold standard phenotypic test.
365	England and Ireland	Whole genome sequencing of MRSA isolates from long term care facilities. Major lineage ST22. Evidence of transmission of MRSA between patients in the facility.
366	East of England, UK	Genomic analysis of USA300 MRSA strains isolated from patients with sequenced USA300 strains from USA. Evidence of multiple introductions of USA300 in the region and household transmission.
118	Europe	Snapshot of the European invasive *S. aureus* population structure: “high-risk clones can be identified on the basis of population level properties such as clonal relatedness, abundance, and spatial structuring and by inferring virulence and resistance properties on the basis of gene content”.

3.1.6 Evidence for host switching

S. aureus colonises a wide range of non-human hosts including livestock, companion animals [266] and wild animals [267]. Some clonal lineages of *S. aureus* have a strong association with specific hosts [367], but occasionally bacteria will adapt to a new host and spread through a new host population [352]. For example, some livestock-associated lineages of MRSA lineages, most notably CC97 and CC398, have overcome the species barrier to infect humans [368,369]. Similarly MRSA clones from human-associated lineages, such as CC5, CC9 and ST88, have been reported in livestock [370]. A study based on MLST across multiple host species showed that anthroponotic (human-to-animal) transmission of *S. aureus* occurs more frequently than zoonosis (animal-to-human transmission) [101]. Whole-genome sequencing has been used to detect and analyse transmission of *S. aureus* within and between various host species, including anthroponotic and zoonotic events [368,371,372].

3.1.7 *S. aureus* among non-human primates

Recent data shows that human-associated *S. aureus* lineages can infect great apes that come into contact with humans in captivity and in the wild [373,374]. These infections can be fatal and this poses a threat to some endangered great apes species [373,374]. African monkeys in remote regions are colonized by *S. aureus* from uncharacterized clonal complexes that rarely colonize or infect humans [268]. One highly divergent clade of coagulase-positive staphylococci is known to infect monkeys in sub-Saharan Africa, but has never been recovered from humans so far; this clade has recently been classified as a new species, *S. schweitzeri* [268,375]. However, such monkeys are also often colonized by human-associated genotypes of *S. aureus*, which they are likely to have acquired from humans [268]. Previously *S. aureus* in monkeys was only studied through MLST, which does not have the resolution to infer inter-species transmission and cannot elucidate the evolutionary mechanisms that drive host adaptation. Our study was the first

comparative genomics study on *S. aureus* from humans and monkeys in sub-Saharan Africa [316].

3.1.8 Objectives and study rationale

Here, I describe the use whole-genome sequencing of *S. aureus* isolates from humans and monkeys in The Gambia with the following goals:

- **To determine whether transmission of *S. aureus* has occurred between humans and non-human primates in The Gambia** through a phylogenetic analysis of *S. aureus* isolates from monkeys alongside human *S. aureus* from carriage and invasive disease
- **To estimate the timescale of any inter-host transmission events** using a published mutation rate to estimate the time since the most recent common ancestor based on the number of mutations accumulated in the core genome
- **To investigate the genetic changes accompanying host adaptation** by studying the patterns of gene presence and absence in the accessory genome

3.2 Materials and Methods

3.2.1 Bacterial isolates

Prior to this study, staff at the MRC Unit in The Gambia had compiled an initial dataset of MLST profiles from 298 *S. aureus* isolates from The Gambia. This set of isolates included:

1. Monkey isolates cultured from nasopharyngeal and oropharyngeal swabs taken from African Green Monkeys (*Chlorocebus sabaeus*) in 2011 by the International Vervet Research Consortium during a study on Simian Immunodeficiency Virus in The Gambia [376].
2. Human invasive disease isolates (unpublished data), recovered from archived clinical specimens including venous blood, pleural fluid, cerebrospinal fluid, joint aspirate, lung aspirate, ascetic fluid, and lymph node aspirates
3. Human carriage isolates were recovered from nasopharyngeal and oropharyngeal swabs from healthy humans

For this study, a subset of 90 human and monkey isolates of *S. aureus* was selected. These strains had been previously characterised by MLST and antimicrobial susceptibility testing by staff at the MRC Unit in The Gambia (Table 3.2). This subset included 46 human invasive disease, 13 human carriage and 31 monkey pharyngeal carriage isolates (Table 3.2) This subset included at least one strain from each clonal complex present in the full MLST dataset and 20 reference genomes were included to represent as many of the major clonal complexes as possible.

Table 3.2 Metadata for 90 *S. aureus* isolates analysed by WGS.

ID	Epidemiology	ST	CC	Pen	Sxt	Tet	C	Cn	Ob	E	Meth	Date
A2	Invasive	1	CC1	R	S	S	S	S	S	S	S	25/05/02
C11	Invasive	852	CC1	R	S	S	S	S	S	S	S	11/08/07
C12	Invasive	852	CC1	R	S	S	S	S	S	S	S	28/10/09
SA9	Carriage	1*	CC1	S	S	S	S	S	S	S	S	29/03/04
B12	Invasive	101	CC101	S	S	I	S	S	S	S	S	19/09/02
C1	Invasive	101	CC101	R	S	S	S	S	S	S	S	24/11/05
C3	Invasive	121	CC121	R	S	S	S	S	S	S	S	04/04/03
SA32	Invasive	121	CC121	R	S	R	S	S	S	S	S	27/04/06
SA33	Invasive	121	CC121	R	S	R	S	S	S	S	S	24/12/02
D11	Invasive	2433	CC121	R	S	S	S	S	S	S	S	01/03/03
SA2	Carriage	15	CC15	S	S	S	S	R	S	S	S	13/01/04
A8	Invasive	15	CC15	R	S	S	S	S	S	S	S	11/01/08
A9	Invasive	15	CC15	R	S	R	S	S	S	S	S	18/11/05
A10	Invasive	15	CC15	R	S	R	S	S	S	S	S	02/09/04
A11	Invasive	15	CC15	R	S	S	S	S	S	S	S	23/02/09
B1	Invasive	15	CC15	R	S	S	S	S	S	S	S	14/07/09
B3	Invasive	15	CC15	R	S	R	S	S	S	S	S	18/07/02
B4	Invasive	15	CC15	R	S	R	S	S	S	S	S	30/08/06
SA23	Invasive	15	CC15	R	S	R	S	S	S	S	S	04/08/10
SA24	Invasive	15	CC15	R	S	R	S	S	S	S	S	19/05/09
SA26	Invasive	15	CC15	R	S	S	S	S	S	S	S	11/10/05
SA31	Invasive	15	CC15	R	S	R	S	S	S	S	S	07/07/05
SA29	Monkey	15	CC15	S	S	S	S	S	S	S	S	01/04/11
D12	Invasive	2434	CC15	R	S	S	S	S	S	R	S	26/08/08
C4	Invasive	152	CC152	R	S	S	S	S	S	S	S	10/02/09
C6	Invasive	152	CC152	R	S	S	S	S	S	S	S	24/04/08
C7	Invasive	152	CC152	R	S	R	S	S	S	S	S	24/08/09
C8	Invasive	152	CC152	R	S	S	S	S	S	S	S	12/09/05
SA34	Invasive	152	CC152	R	S	R	S	S	S	S	S	25/09/07
F5	Monkey		CC152	S	S	S	S	S	S	I	S	01/04/11
D10	Invasive	2432	CC2432	S	S	S	S	S	S	S	S	20/02/07
B5	Invasive	25	CC25	R	R	R	S	S	S	S	S	31/07/09
B6	Invasive	25	CC25	R	R	S	S	S	S	S	S	04/11/08
F7	Monkey	Novel	CC2531	S	S	S	S	S	S	S	S	01/04/11
F8	Monkey	Novel	CC2531	S	S	S	S	S	S	S	S	01/04/11
SA22	Monkey	Novel	CC2531	S	S	S	S	S	S	S	S	01/04/11
F9	Monkey	2532	CC2532	S	S	S	S	S	S	S	S	01/04/11
G2	Monkey	2532	CC2532	S	S	S	S	S	S	S	S	01/04/11
F11	Monkey	2533	CC2533	S	S	S	S	S	S	S	S	01/04/11
F12	Monkey	2533	CC2533	S	S	S	S	S	S	S	S	01/04/11

SA27	Monkey	2533	CC2533	S	S	S	S	S	S	S	S	01/04/11
H4	Monkey	Novel	CC2565	S	S	S	S	S	S	I	S	01/04/11
H5	Monkey	Novel	CC2565	S	S	S	S	S	S	S	S	01/04/11
H6	Monkey	Novel	CC2565	S	S	S	S	S	S	S	S	01/04/11
H7	Monkey	Novel	CC2565	S	S	S	S	S	S	I	S	01/04/11
H8	Monkey	Novel	CC2565	S	S	S	S	S	S	S	S	01/04/11
G5	Monkey	2566	CC2566	S	S	S	S	S	S	I	S	01/04/11
H9	Monkey	2569	CC2566	S	S	I	S	S	S	I	S	01/04/11
H10	Monkey	2569	CC2566	S	S	S	S	S	S	S	S	01/04/11
G8	Monkey	2567	CC2567	S	R	S	S	R	R	I	R	01/04/11
G9	Monkey	2567	CC2567	S	S	S	S	S	S	S	S	01/04/11
G10	Monkey	2567	CC2567	S	S	S	S	S	S	S	S	01/04/11
G11	Monkey	2567	CC2567	S	S	S	S	S	S	S	S	01/04/11
G12	Monkey	2567	CC2567	S	S	S	S	S	S	S	S	01/04/11
H2	Monkey	2567	CC2567	S	S	S	S	S	S	S	S	01/04/11
SA18	Carriage	30	CC30	S	S	S	S	I	R	S	S	25/02/04
B8	Invasive	30	CC30	R	S	S	S	S	S	S	S	19/02/02
SA35	Invasive	30	CC30	R	S	R	S	S	S	S	S	20/02/07
D1	Invasive	1472	CC30	R	S	S	S	S	S	S	S	19/02/02
D2	Invasive	1472	CC30	R	S	S	S	S	S	S	S	12/08/05
D4	Invasive	1472	CC30	R	S	S	S	S	S	S	S	09/09/08
D5	Invasive	1472	CC30	R	S	S	S	S	S	S	S	08/11/02
D6	Invasive	1472	CC30	R	S	S	S	S	S	S	S	08/11/02
C9	Invasive	508	CC45	S	S	S	S	S	S	S	S	22/04/10
SA21	Carriage	Novel	CC45	S	S	S	S	S	S	S	S	11/03/04
A3	Invasive	5	CC5	R	S	R	S	S	S	S	R	28/01/09
A5	Invasive	5	CC5									
SA17	Invasive	5	CC5	R	S	S	S	S	S	S	S	12/08/05
E1	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
E2	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
E3	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
E4	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
E5	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
E6	Monkey	6	CC5	S	S	S	S	S	S	I	S	01/04/11
E8	Monkey	6	CC5	S	S	S	S	S	S	S	S	01/04/11
SA6	Carriage	72	CC5	S	S	S	S	S	R	S	S	20/01/04
B9	Invasive	72	CC5	R	S	R	S	S	S	S	S	20/08/05
B10	Invasive	72	CC5	R	S	R	S	S	S	S	S	21/09/04
SA3	Carriage	730	CC5	S	S	S	S	S	S	S	S	05/04/04
SA14	Carriage	730	CC5	S	S	S	S	S	S	S	S	23/03/04
SA15	Carriage	730	CC5	S	S	S	S	S	S	S	S	23/03/04
SA16	Carriage	5*	CC5	S								27/03/04
A6	Invasive	8	CC8	R	R	R	S	S	S	S	S	27/02/07
A7	Invasive	8	CC8	R	S	S	S	S	S	S	S	13/03/08
D8	Invasive	239	CC8	R	R	R	S	R	R	R	R	01/06/07
SA7	Carriage	509	CC89	S	S	S	S	S	S	S	S	10/12/03

SA13	Carriage	509	CC89	S	S	S	S	S	S	S	S	16/03/04
SA19	Carriage	509	CC89	S	S	S	S	S	S	S	S	14/01/04
SA20	Carriage	509	CC89	S	S	S	S	S	S	S	S	21/04/04
B11	Invasive	97	CC97	S	S	R	S	S	S	S	S	16/05/07

Footnote: Isolates metadata showing genomic ID, Lab ID, Epidemiology (human carriage, human invasive disease or Monkey carriage), sequence type ST, clonal complex CC, antibiotic resistance (penicillin Pen, trimethroprim Sxt, tetracycline Tet, Chloramphenicol C, Gentamycin Cn, Cloxacillin Ob, Erythromycin E and methicillin Meth) and date of sample collection. Table sorted by CC then ST then Epidemiology to make it more easily intelligible. * In ST represents a single nucleotide variant.

3.2.2 Geography and demographics of study area

The Gambia is a small West African country with a population of approximately 1.8 million. The Gambia is partitioned across its breadth by the river Gambia, which forms an estuary with the Atlantic Ocean. The most developed region of The Gambia is the Western region, the coastal region south of the River Gambia. My dataset combines isolates from three studies that were carried out in the Western region of The Gambia.

The majority of monkeys sampled had been trapped within a 10km radius of the MRC camp in Fajara, either at the Bijilo Forest Park or the Abuko nature reserve [376] (Figure 3.1). The monkeys were habituated to humans as a result of urbanisation and tourism. The Bijilo Forest Park is a tourist attraction where locals and tourists go to visit the monkeys. Visitors get a chance to get up close to the monkeys, taking pictures with them and on some occasions illegally feeding them peanuts. This interaction presents an opportunity for inter-host transmission of potential pathogens.

The human carriage survey was performed in Sibanor, a village ~80km from the MRC camp in Fajara, between Dec-04 and Apr-05 (Table 3.3) [377]. This study was carried out to document the prevalence and diversity of *S. aureus* colonising humans prior to the introduction of pneumococcal conjugate vaccines in The Gambia. The invasive disease isolates were recovered from patients that reported to the MRC clinic in Fajara. Strains from these two datasets were genotyped in order to provide a baseline for comparing the *S. aureus* strains isolated from monkeys and human *S. aureus.*

Figure 3.1 A Google map screen shot showing the spatial distribution of the study sites where the *S. aureus* isolates were sampled. Red stars indicate monkeys sampling sites and blue star indicates the site for the carriage study.

Table 3.3 A summary of the three studies that formed the basis of the *S. aureus* analysis. For each study the distance from the MRC camp to the study sampling site and the period of sampling was listed. The number of samples analysed by MLST and the number of samples analysed by WGS are indicated.

Epidemiology	Study site	Distance from MRC (km)	Start date	End date	MLST analysis (n)	WGS analysis (n)
Invasive disease	MRC Clinic Fajara	0	2002	2010	116	46
Human carriage	Sibanor, Foni	83	Dec-04	Apr-05	100	13
Monkey carriage	Mainly Abuko nature reserve and Bijilo forest park	10 km radius	Apr-11	May-11	82	31

3.2.3 Workflow for bioinformatics and sequence analysis

A detailed explanation of the methods is provided in the methods section (section 2). Below I present a flow chart that summarises the steps that were taken as part of the analysis (Figure 3.2).

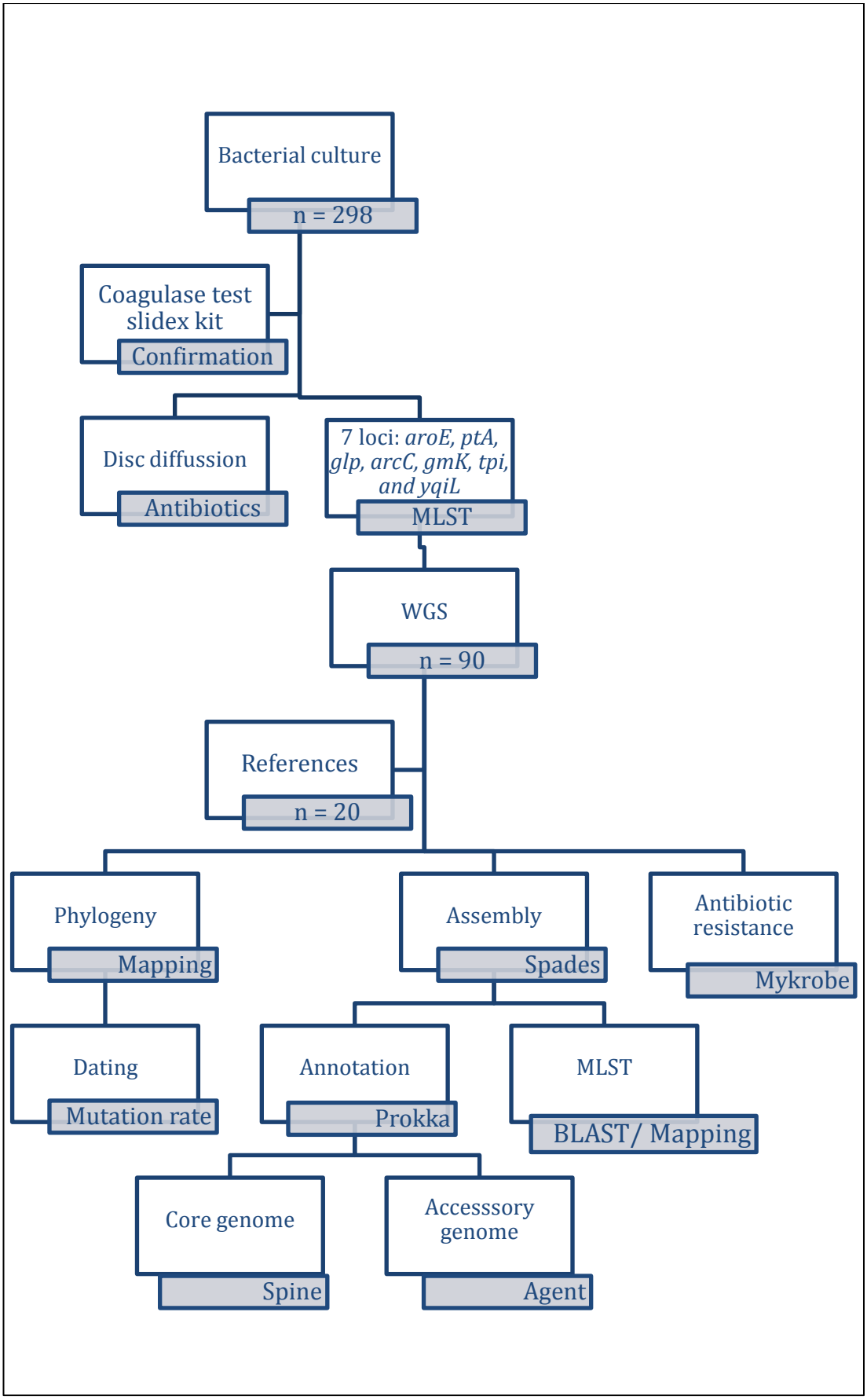


Figure 3.2 A flowchart summarizing the workflow for the bioinformatics analysis of the *S. aureus* dataset.

3.3 Results

3.3.1 Genome sequences

Draft assemblies ranged from 2,701,991 to 3,048,160 nucleotides in length (average 2,793,864.9, median 2,786,149), with an average GC content of 32.7 % (range: 32.5 to 32.9%). The genomes were annotated with an average of 2593 coding sequences (CDSs) (range 2462 to 2921). When mapped to the EMRSA15 reference genome, the mean genome-wide coverage was 61.5 fold (range 12.5 to 126.2).

3.3.2 Population structure of *S. aureus* in The Gambia

Single nucleotide variants (SNVs) were called from 232,276 variable sites in the core genome after excluding all known repeat regions and transposable mobile elements. These SNVs were used to create a maximum-likelihood phylogenetic tree linking isolates with reference strains (See the multiple sequence alignment of core genome SNVs and the core SNV distance matrix in Appendices 3.1 & 3.2 respectively). This phylogenetic tree revealed the clonal population structure of *S. aureus.* Strains belonging to the same clonal complex clustered on monophyletic branches on the phylogenetic tree. The tree resolved five major *S. aureus* clades that were consistent with previous species-wide phylogenetic analyses (Figure 3.3):

- **Clade 1** encompassed pandemic clonal lineages CCs 1, 5, 6, 8, 15 and 25, and closely related genotypes. Clade 1 encompassed both human carriage and human invasive disease strains including two MRSA invasive disease strains from the ST5 and ST339. Within clade 1 there was also a cluster of ST6 isolates from monkeys and a ST15 isolate from a monkey that clustered with the human ST15 isolates
- **Clade 2** represented a monkey-specific clade that contained most of the *S. aureus* isolates recovered from monkeys. This clade was comprised of lineages bearing (at the time) novel STs that were only

isolated from monkeys. Clade 2 was characterised by deep branches that resolved into clusters of unique CCs inferred by eBURST

- **Clade 3** included the EMRASA15 reference genome; human carriage and human invasive disease strains belonging to the globally disseminated lineages CC30 and CC45; and a cluster of ST509 carriage strains, which were single locus variants of the Japanese MRSA lineage ST89
- **Clade 4** housed the pandemic lineage CC121 alongside animal-associated reference genomes. All the CC121 strains were recovered from human invasive disease
- **Clade 5** included the dominant West African lineage ST152 and closely related genotypes. All human strains in clade 5 were recovered from invasive disease specimens. In clade 5 we also recovered a single novel variant from monkeys (F5), which was related to, but still distinct from, ST152

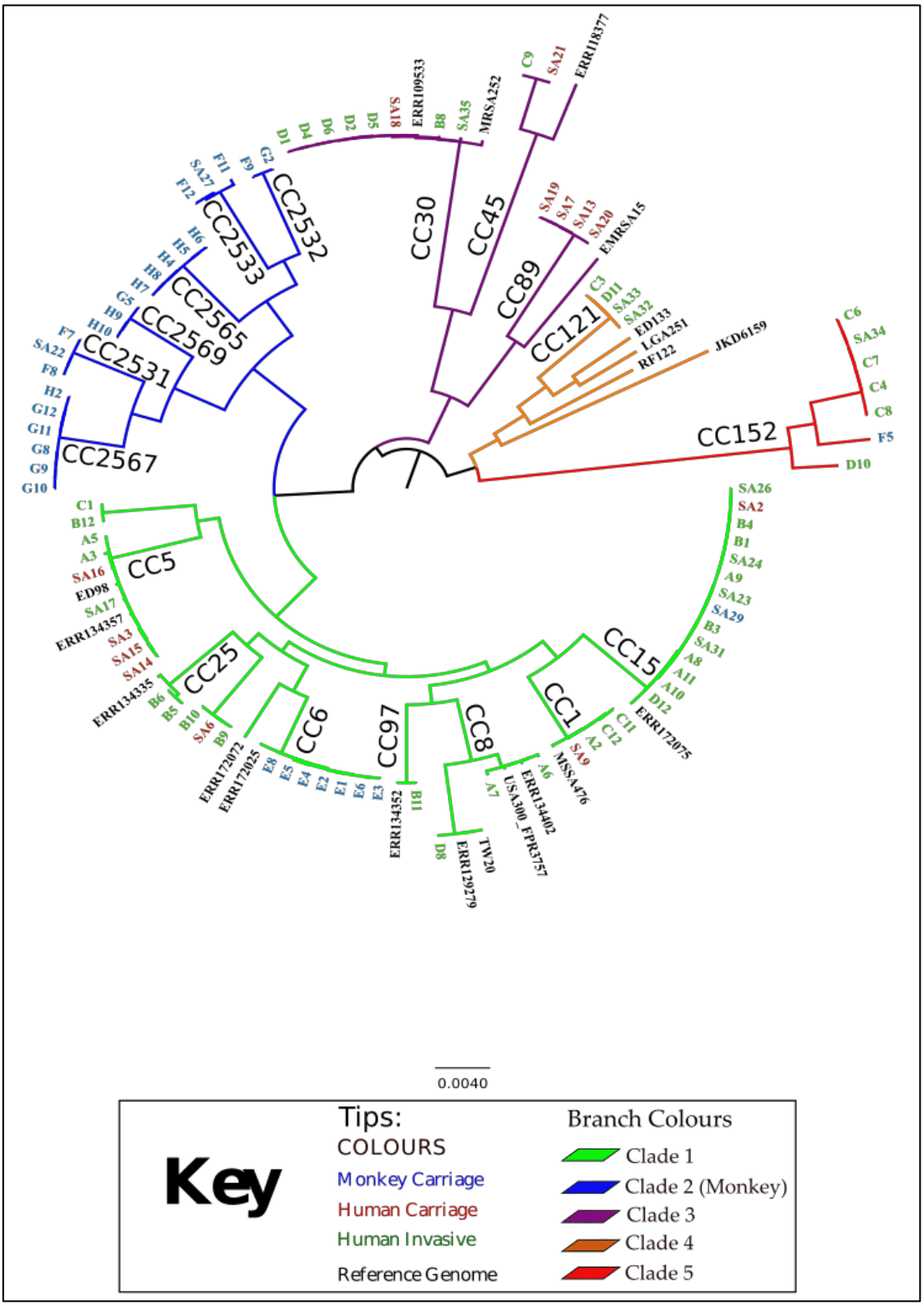


Figure 3.3 Maximum likelihood phylogenetic tree of all *S. aureus* study isolates and reference genomes. Tips coloured by epidemiology and branches coloured based on arbitrary nomenclature (clade 1-5). Branches are also annotated with the major human clonal complexes (CCs) that are described in the literature and CCs for clade 2 strains assigned by eBURST.

3.3.3 Anthroponotic transmission of *S. aureus*

This dataset provided a rare insight into the contemporaneous population structure of *S. aureus* in humans and monkeys. Assuming a constant mutation rate, the timescales for inter-host transmission of *S. aureus* between humans and monkeys in The Gambia were estimated (Table 3.4). Current estimates of intra-clonal mutation rates are consistent in *S. aureus* [321]. Based on the intra-clonal mutation rate for *S. aureus* of ~ 2 x 10^{-6} per site per year [321], and a core genome size of ~2.5 Mb, it was possible to estimate an upper bound date for the transmission events. The short-term mutation rate equates to about 5 SNPs per year. Thus simply dividing the total number of SNPs between any pair of contemporary S. aureus isolates by ten gives the approximate number of years since they shared a common ancestor.

The phylogenetic placement of clade 2 – nested between human-associated clades – suggests this clade has human ancestry. The most divergent pair of clade 2 isolates differed by 26,464 SNVs. Similarly, the closest pairwise distance to an isolate outside clade 2 (The MSSA476 reference genome) was 26,968 SNVs. Thus, clade 2 was likely to have emerged from a host jump that occurred ~2700 years ago. There were no clear examples of zoonotic transmission (from monkeys to humans), as no human-derived isolates fell within the novel monkey-associated clade 2.

The ST6 monkey isolates in clade 1 formed a tight cluster, within which the furthest distance between two ST6 monkey isolates was 270 SNVs. The lowest divergence between a ST6 isolate from a monkey and the reference ST6 isolate from a human in Europe was just 311 SNVs. These observations suggest that ST6 was transmitted from humans to monkeys a little under three decades ago.

The single ST15 isolated from a monkey in clade 1 differed from an invasive disease isolate by 71 SNVs. This most likely represents an anthroponotic host switch that occurred approximately less than 7 years ago. This estimate was not attempted for the monkey isolate in clade 5 due to lack of a closely related human isolate.

Table 3.4 The estimated time-scales for the anthroponotic transmission events.

Monkey Genotype	Closest Human isolates (SNVs)	Estimated time of host jump (years)	Furthest within-clade monkey isolate (SNVs)	Estimated time of divergence (years)
ST15	71	7	-	-
ST6	311	31	270	27
Clade 2	26968	2700	26464	2700

3.3.4 Proportion of shared accessory genome

The core genome was inferred from a representative dataset of 35 genomes, including reference genomes, using SPINE. The accessory genome for each strain was inferred using AGENT, which excluded the representative core genome [332]. Using a custom Perl script, the pairwise proportion of accessory genome shared between strains was calculated based on 85% sequence homology. A heatmap showing pairwise shared proportion of accessory genome was plotted with the phylogenetic tree (Figure 3.4). The gene presence and absence list is presented in Appendix 3.3.

Isolates within the same clonal complexes shared similar accessory genome content. Similarly, the related CCs that formed distinct clades shared more similar accessory genome content than unrelated CCs. For example, in clade 1 CC15/1/8/97/6/25/5 share more accessory genome content with each other that other clades (Figure 3.4). The same trend was observed in the monkey-associated clade 2: CCs within the clade had similar accessory genome content and the accessory genome was highly conserved between strains within the same CCs. The CC152 monkey strain F5 had a unique accessory genome that consisted of 365 CDSs. F5 shared less that 40% of its accessory genome with any other genome. 38% of the accessory genes present in F5 were present in clade5 isolates while 31% were present in CC2566 isolates in clade2. The conservation of accessory genome content within CCs supports the concept of a “core variable” genome [335].

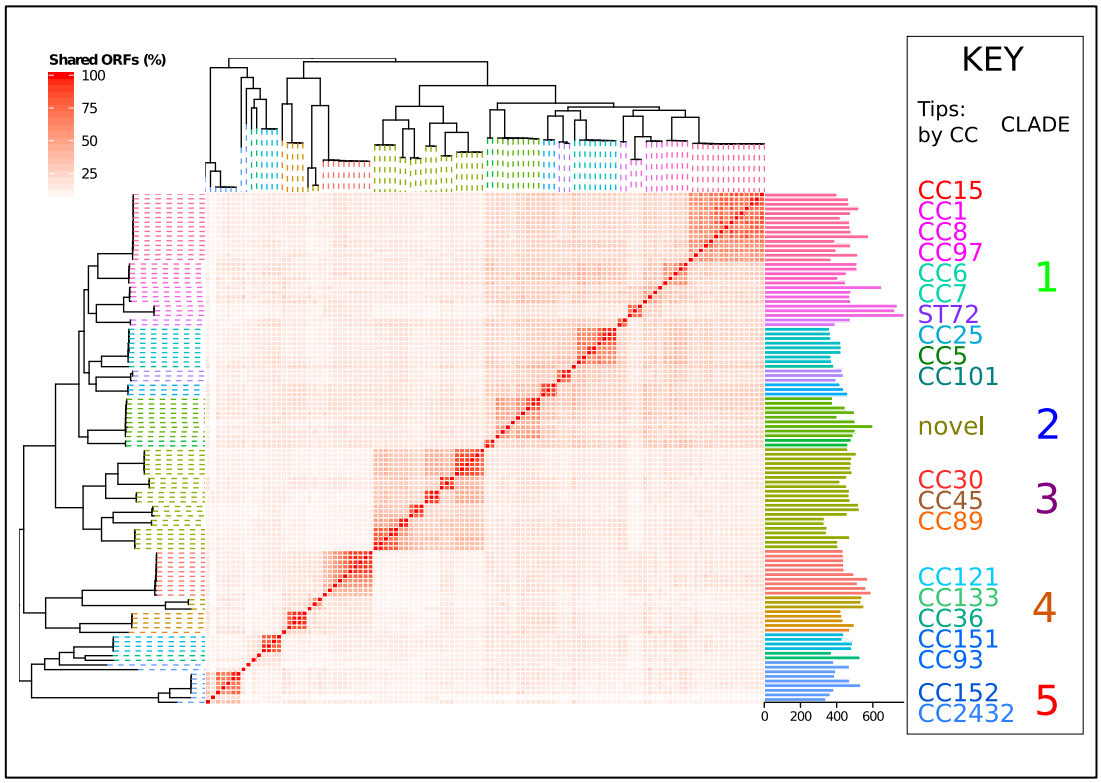


Figure 3.4 A heatmap showing the proportion of accessory genes between *S. aureus* linked to the phylogenetic tree. The nodes on the phylogenetic tree are coloured to represent their respective CCs. The list of tips in the key corresponds to the order of CCs in the vertical tree on the left from top to bottom.

3.3.5 Gene loss in clade 2 and host adaptation

The BLAST Ring Image Generator (BRIG) highlighted genome erosion based on an 85% sequence similarity threshold with the USA300 reference genome. BRIG compared six monkey-derived isolates representing the different CCs within clade 2. Genome erosion was observed among clade 2 isolates on mobile genetic elements such as; the genomic islands vSaα and vSaβ; the SCCmec cassette; the phage element ϕSa2USA; the pathogenicity Island SaPI5; and the Immune evasion cluster (IEC) (Figure 3.5).

The SCCmec cassette and the pathogenicity Island SaPI5 were absent in all of the representative strains from the clade 2 sub-clusters. The genomic islands vSaα and vSaβ were consistently partially eroded in all strains. There was variation within the different clade 2 subgroups, notably isolates H10 (CC2569) and G11 (CC2567) contain a more intact phage φSA2USA than the other clade 2 subgroups. The ϕSa3USA phage element and the IEC were only intact in isolates SA27 (CC2533) and F7 (CC2531) respectively.

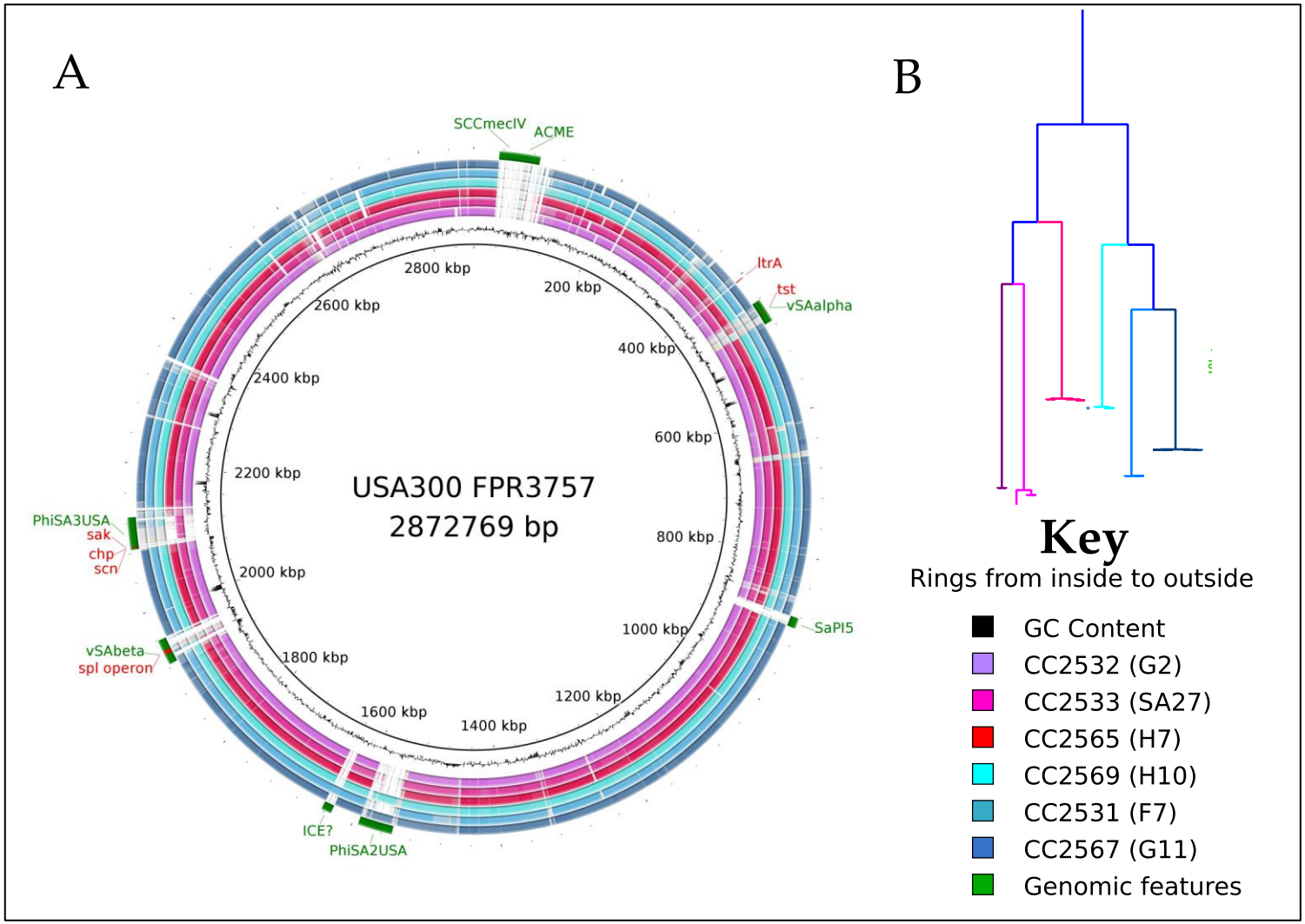


Figure 3.5 Genome-wide BLAST comparison of isolates representing the CCs of clade 2 against the USA300 reference genome showing regions of genome erosion. (A) The inner ring represents the reference genome and each of the outer rings represents a unique strain. Annotations represent mobile genetic elements and Genomic Islands. (B) A schematic of clade 2 wherein each CC is coloured uniquely.

3.3.6 Antibiotic resistance

The Mykrobe predictor was employed to predict antimicrobial resistance from the genome sequence data [304]. Mykrobe predicted antibiotic resistance based on the presence of whole genes as well as point mutations. The Mykrobe predictor also determined the presence of the Panton-Valentine leukocidin toxin genes in the strains.

Based on phenotypic antimicrobial susceptibility testing high levels of penicillin resistance were observed among invasive disease isolates (89%). Penicillin resistance was conferred by presence of the *blaZ* gene in most cases. Among invasive disease isolates 19 were resistant to tetracycline. The *tetL*, *tetM* and/or *tetK* genes mediated tetracycline resistance.

Two invasive disease isolates were confirmed to be methicillin resistant and both isolates bore the *mecA* gene. The MRSA invasive disease strains belonged to the globally disseminated genotypes ST239 and ST5. These represented the first published reports of MRSA causing invasive disease in The Gambia [316].

None of the carriage strains were classified as penicillin or tetracycline resistant based on disc diffusion antimicrobial susceptibility testing. However, based on the genome Mykrobe predicted that 12/13 and 4/13 carriage strains were resistant to penicillin and tetracycline respectively. This indicates that the burden of antimicrobial resistance among *S. aureus* carriage strains in The Gambia is likely to be underestimated by disc diffusion.

This dataset does not confirm drug resistance among *S. aureus* isolates from monkeys in The Gambia. Initially one isolate from a monkey had been characterised as methicillin resistant based on cefotoxin disc diffusion. However, when the antibiotic resistance test was repeated with eTest the isolate was reclassified as methicillin sensitive. Furthermore there was no genetic evidence of MRSA in this isolate based on a Mykrobe prediction [304].

3.4 Discussion

3.4.1 Anthroponotic transmission

We posit that clade 2 represents an ancient human-to-monkey transmission event that occurred a minimum of 2700 years ago. Clade 2 has undergone clonal expansion and diversified into at least six sub-clades within the monkey population. Whilst this can be largely explained in terms of common clonal inheritance (descent with modification), it is also possible that some of the differences in gene content between clade 2 and the other clades in the tree underlie adaptation to the monkey host. The presence of the monkey-associated isolates in both Clade 2 and ST6 provides an opportunity to detect any genes that may have been acquired independently in two unrelated monkey-adapted lineages.

Human-associated genotypes have been previously isolated from monkeys in West Africa: ST1, ST6 and ST15 [268]. We report the presence of ST6 and ST15 in monkeys raising the possibility that these genotypes may be prone to human-to-monkey transmission. The fact that ST15 was only recovered once from monkeys is also consistent with a recent transmission event. The cluster of monkey-derived ST6 isolates is consistent with onward transmission within the monkey population. However, the upper-bound estimate for time since transmission of ST6 to monkeys may be spuriously high since we were unable to sequence the closest human ST6 isolate. Furthermore, the ST6 cluster may have emerged as a result of multiple host switching events. In this case, the within-clade diversity would also represent a high upper-bound estimate for the time since the host switch occurred.

The evolutionary origin of the CC152 monkey isolate F5 is unclear given its unique accessory genome composition and the similarity to ST152 strains at the core genome level. A likely scenario is that a ST152 isolate was transmitted to monkeys then subsequently lost some of its accessory genomes and acquired some accessory content from monkey specific lineages through recombination. ST152 strains are able to adapt to multiple

host niches and have been detected among domestic animals in Africa [268]. Moreover, an overhaul of the accessory genome is not uncommon in *S. aureus*. It has previously been described that ST239 is a product of recombination of a large block of DNA from a CC30 donor into a CC8 background [378].

3.4.2 Gene loss as a contributor to host adaptation

The loss of genes harboured on the genomic islands vSaα and vSaβ in clade 2 isolates may have enhanced adaptation to the monkey host. The *spl operon* (*spl A-F)* is harboured on vSaβ and it encodes serine proteases as well as *lytN*, which is essential for bacterial growth and cell wall synthesis [379,380]. All clade 2 isolates harboured novel variants of *spl* genes that were not present in human isolates or reference genomes highlighting a potential role in host adaptation.

The immune evasion cluster IEC1 harbours the immune evasion genes (*chp, sak and scn)* and enterotoxin genes such as *sea*, *sep*, *sek*, and *seq* [381,382]. These genes elicit human-specific immune modulation using different mechanisms including the blockage of neutrophil chemotaxis [383]. Thus it is not surprising that the IEC1 was lost in the vast majority of isolates recovered from monkeys. Similarly the Panton-Valentine leukocidin genes (lukF-PV and lukS-PV) were absent from all isolates recovered from monkeys. Panton-Valentine leukocidin is a potent human neutrophil cytotoxic factor that does not work on monkeys [384].

Genes carried on MGEs play an important role in host adaptation [343]. However, our data suggests that in addition to gene acquisition, gene loss may play an important role in adaptation to non-human hosts. If gene loss is assumed to occur more readily than gene acquisition, then this relationship might explain why anthroponotic transmission of *S. aureus* is more common than zoonotic transmission. Protein-truncating mutations in *S. aureus* have been previously implicated in the progression from carriage to invasive disease [361].

3.4.3 Risk of monkeys serving as reservoirs for highly virulent *S. aureus*

Our data supports previous suggestions that anthroponotic transmission events may be relatively frequent, while zoonotic transmission is much less common [101]. Although human-to-monkey transmission has been documented in West Africa there is no evidence of zoonotic transmission from monkeys [268]. From a public health perspective, this is reassuring as it suggests that monkeys do not constitute a serious risk of *S. aureus* infection in humans. Related to this, we also note that none of the monkey isolates were MRSA. Furthermore, although there were high levels of resistance to penicillin among invasive disease isolates, the isolates from monkeys were all susceptible to penicillin.

3.4.4 Limitations

The main limitation of this study is that the samples from the three epidemiological classes were collected over different timescales and from different regions. *De novo* sample collection over a standard harmonized timeframe would have enabled us to make more accurate inferences on the direction of transmission. A longitudinal study would also be required to confidently determine the dynamic of the inferred host switching events. Finally, our study lacked precise demographic information that could link each isolates to the location where the monkey bearing it was trapped and sampled. This information would have facilitated a more in depth understanding of the dynamic of *S. aureus* in monkeys in The Gambia.

4 GENOMIC EPIDEMIOLOGY OF A PNEUMOCOCCAL MENINGITIS OUTBREAK IN GHANA, 2016

4.1 Introduction

4.1.1 The African meningitis belt

The African meningitis belt spans a region from Senegal to Ethiopia, a region home to ~300 million people [385,386] (Figure 4.1). Globally, the highest incidence rates of meningitis occur in the African meningitis belt [386]. The main cause of meningitis in the belt is *Neisseria meningitidis*, the meningococcus, but *Streptococcus pneumoniae* or *Haemophilus influenzae* type B are also known to cause localised outbreaks [386,387]. Across the belt, incidence levels peak during the cold dry season, from October to April, causing thousands of deaths annually and large epidemic outbreaks periodically [386,387].

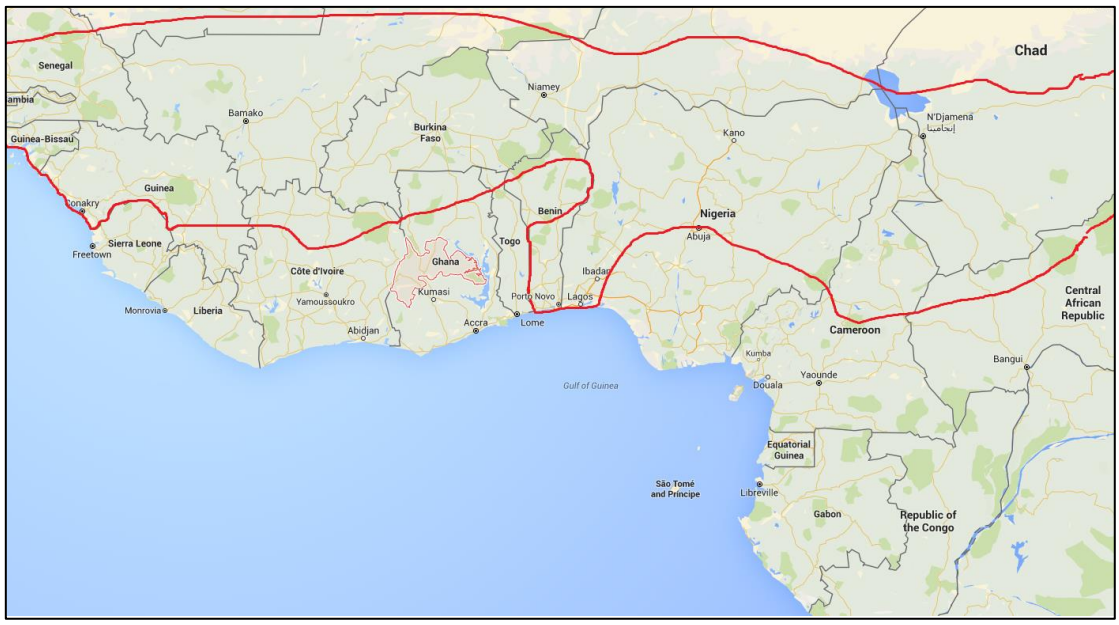


Figure 4.1 A map showing of part of Africa highlighting the African meningitis belt (red line) and the Brong-Ahafo region (red highlight) adapted from [283].

The dynamics behind the seasonality of meningitis outbreaks are unclear, but climate is believed to be the strongest driving force [388,389]. During the dry season, the Harmattan winds blow from the Northeast, carrying high levels of atmospheric mineral dust [390]. The persistent dust and low humidity are believed to damage the mucosal epithelium leading to an increased risk of progression from asymptomatic carriage to invasive disease [386,391]. Similarly, viral respiratory tract infections such as the flu weaken the immune system enhancing transmission and invasion of bacteria [392].

4.1.2 Molecular epidemiology of meningitis in West Africa

Since the 1990s, meningococcal ST5 complex strains belonging to serogroup A have been the most common causes of meningitis outbreaks [393]. In 2000, the serogroup W NmW135 strain emerged as the cause of a major meningitis outbreak during the annual Hajj pilgrimage in Mecca, Saudi Arabia [387]. Shortly afterwards, in 2002, NmW135 was responsible for a major meningitis outbreak in Burkina Faso [394]. NmW135 reached epidemic status due to high incidence rates among children but for unknown reasons the prevalence of NmW135 decreased significantly within a short space of time [395]. The year 2003 saw the emergence of a new serogroup A strain from the ST2859 complex, which demonstrated the potential to cause epidemics in the future [387].

Since the 2000s, lethal community-based pneumococcal meningitis outbreaks have been reported in northern Ghana, Burkina Faso and Togo, affecting all age groups [278,279,396]. *S. pneumoniae* serotype 1 is the leading cause of pneumococcal meningitis in the meningitis belt [397]. Serotype 1 can cause lethal outbreaks of meningitis in high-burden settings [396,398]. Serotype 1 is highly virulent and has a high invasiveness index, since it is rarely detected in asymptomatic carriage [397]. The sequence type (ST) 217 clonal complex (CC) and closely related genotypes ST303 and ST618 have been

implicated in previous outbreaks of pneumococcal meningitis in West Africa [396,398,399].

4.1.3 Vaccine intervention in the African meningitis belt

The MenAfriVac® campaign was launched in 2010 in a bid to eradicate serogroup A. MenAfriVac® has been a success story, since it led to elimination of serogroup A carriage and to herd immunity that protected even unvaccinated individuals [15]. MenAfriVac vaccination in Burkina Faso led to the eradication of serogroup A strains among carriage and diseased individuals. However, following vaccination the serogroup X strains have become the dominant genotype isolated from carriage and invasive disease [400]. Worryingly the serogroup W strains, which have demonstrated the ability to cause lethal outbreaks in the past [394], have re-emerged in Burkina Faso post-vaccination [15,400].

The need for mass vaccination campaigns against *S. pneumoniae,* especially serotype 1, in the meningitis belt cannot be stressed enough. In The Gambia the introduction of the pneumococcal conjugate vaccine PCV-13, which includes serotype 1, has led to a marked decrease in the prevalence of vaccine serotypes causing invasive pneumococcal disease [280]. A study showed that across Burkina Faso, Ghana, Nigeria, Mali, Senegal, and The Gambia 65-87% of pneumococcal isolates had serotypes covered by PCV-13 [401]. Evidence shows that pneumococcal vaccination campaigns should also be extended to cover older children and adults in order to boost efficacy [278].

4.1.4 Evolution of the pneumococcal genome

Variation is introduced into the pneumococcal core genome primarily through homologous recombination (uptake and incorporation of exogenous DNA) and point mutations [178]. Pneumococcal genomes undergo very high rates of recombination; rates are influenced by capsular type and carriage rates

across lineages [402]. Clinical interventions such as vaccination and antibiotic therapy rapidly select for novel pneumococcal genotypes that arise through homologous recombination [116,403]. In addition, *S. pneumoniae* can acquire novel genes that mediate virulence and antimicrobial resistance from other species by horizontal gene transfer (HGT) [403,404].

4.1.5 Evolution of drug resistance

The emergence of drug resistance in *S. pneumoniae* is a major public health concern driven by the emergence of multidrug-resistant strains [405]. Antimicrobial resistance can be spread by clonal dissemination of resistant strains or emergence of resistance in susceptible strains [405]. Pneumococcal strains can acquire antibiotic resistance through target modification, efflux pump and/or enzyme inactivation of the drug [406].

β-lactam resistance is acquired through modification of the penicillin binding protein (PBP). Modifications of PBP occur primarily through point mutations but the mosaic structure of PBP suggests that the PBP may have been modified in other Streptococci before being transformed into *S. pneumoniae* through recombination [178,405]. Fluoroquinolone resistance is driven primarily by spontaneous mutations in the quinolone resistance determinant region [405].

Genes carried on transposable mobile genetic elements confer resistance to a broad range of antibiotics [407]. For example, the Tn916 family transposons carry:

- Macrolide resistance determinants *ermB* and *mef,*
    - *ermB* encodes a protein product that induces resistance by methylating the 23S ribosomal binding site and the
    - *mef* genes, which encode an active efflux pump [408]
- Accessory genes *catQ* and *tetM* that confer chloramphenicol and tetracycline resistance through enzymatic inactivation and target alteration respectively [405,409]

A new antibiotic called teixobactin, which inhibits cell wall biosynthesis has recently been discovered [410]. Teixobactin has shown efficacy in mice infected

with *S. pneumoniae* and a wide range of gram-positive bacteria [410]. Encouragingly it is expected that this compound will avoid the development of resistance that is common to most antibiotics [410]. Perhaps teixobactin and similar compounds that will emerge from this discovery may influence the future of therapeutic intervention for *S. pneumoniae*.

4.1.6 The Brong-Ahafo meningitis outbreak

A meningitis outbreak broke out in the Brong-Ahafo region of Ghana during the peak of the 2015-2016 dry season. The WHO Regional Reference Laboratory hosted at the MRC Unit The Gambia (MRCG) played a crucial role in providing laboratory and data management support to the local health authorities in Brong-Ahafo region in response to the outbreak. A team of scientists and a database manager were dispatched from MRCG to work with the WHO and Ghana Ministry of Health in the Brong-Ahafo region.

There were 886 suspected cases of meningitis in the Brong-Ahafo region between 2nd December 2015 and 26th February 2016. The mortality rate among suspected cases whose outcome was known was 10%. Over 95% of the confirmed cases occurred in children over 5 years and adults. All available cerebrospinal fluid specimens from suspected cases were tested by rapid test, qPCR and/or culture. 135 cases of meningitis were confirmed by identification of a pathogen:

- *Streptococcus pneumoniae* (77%, 104/135)
- *Neisseria meningitidis* (22%, 30/135)
- Group B streptococcus (GBS) (1%, 1/135).

49 pneumococcal strains were successfully serotyped by PCR of which, 40 (78%) were serotype 1 and five (12%) were serotype 12F. Other serotypes detected included 35B, 6A/6B, 7F and 3.

It was surprising that serotype 1 *S. pneumoniae* was the main pathogen identified during the outbreak, as PCV-13 was included in the Ghanaian expanded program of immunization in 2012. Although nationwide coverage of PCV-13 was below 50% in 2012, it rose steadily and by 2014 had reached

81%. PCVs usually induce herd immunity leading to a decrease in the prevalence of vaccine serotypes even in non-vaccinated individuals [281,282,411]. However recent estimates suggests that conjugate vaccines may take up to a decade to induce herd immunity among the unvaccinated population [412].

4.1.7 Objectives and study rationale

To understand the molecular epidemiology of the outbreak we set out to perform whole genome sequencing (WGS) on stored pneumococcal isolates from outbreak patients with the following objectives:

- **To confirm the serotype and determine the MLST profile of the outbreak *S. pneumoniae* strains** by comparing their genome sequence to the sequences of the known pneumococcal serotypes and MLST alleles respectively
- **To reconstruct the phylogeny of the outbreak strains** based on single nucleotide variants in the core genome
- **To study the evolutionary events that led to the emergence of serotype 1 outbreak following the introduction of PCV-13** by gleaning the genome for evidence of capsular switching and changes in the core and accessory genome that could have led to enhanced fitness or virulence
- **To compare the serotype 1 strains from the outbreak with previously sequenced serotype 1 strains from West Africa.** These strains were sequenced as part of a previous collaboration through the Pneumococcal African Genomics Consortium (PAGe)

4.2 Methods

4.2.1 Geography and demographics of the study area

The Brong-Ahafo region lies at heart of Ghana, midway up the North-South axis but below the African meningitis belt (Figure 4.2). The region consists of 27 districts and has a population of 2.4 million. Like much of West Africa, the Brong-Ahafo region has a tropical climate, with a dry season that lasts from November to March and two rainy seasons in between. Prior to the outbreak the WHO did not have standard guidelines for containing a pneumococcal meningitis outbreak. Moreover, such outbreaks were not expected to occur in the Brong-Ahafo region because it lies outside the meningitis belt. As a result, regional health authorities were caught off guard and initially struggled to contain the outbreak. After helping to contain the outbreak the MRCG team made important recommendations to the WHO for dealing with pneumococcal meningitis outbreaks.

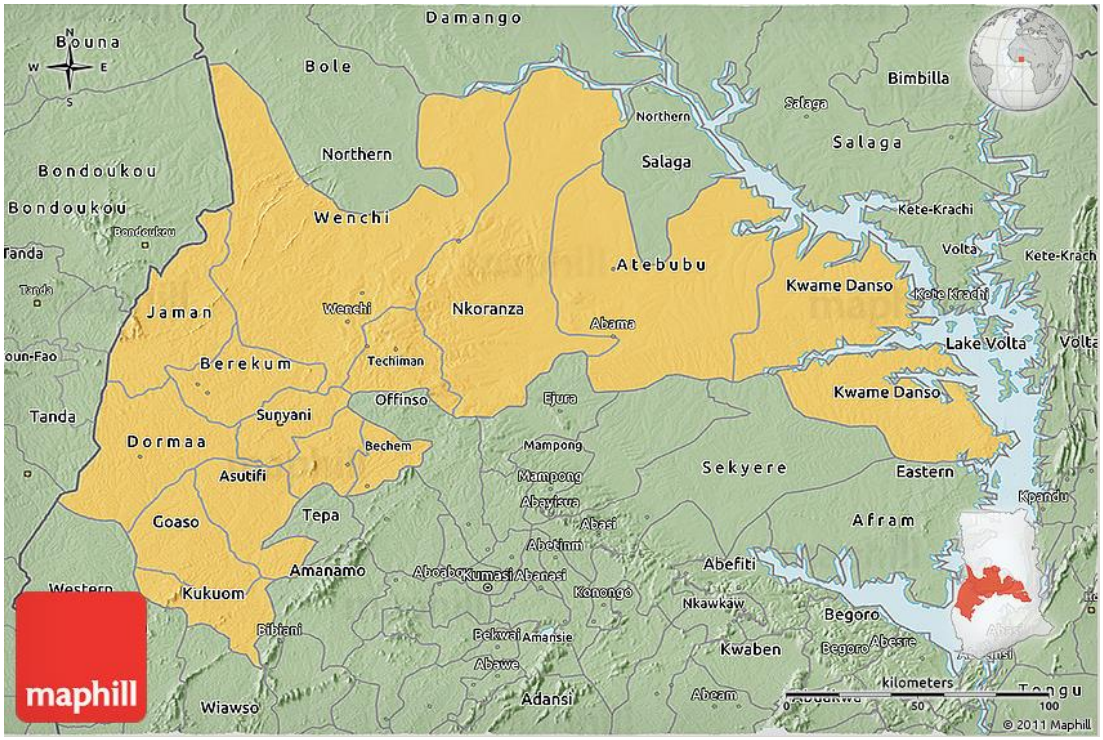


Figure 4.2 A district map of the Brong-Ahafo region of Ghana (source: http://www.maphill.com).

4.2.2 Bacterial isolates

During the outbreak meningitis, the primary method for confirmation of meningitis cases was the Pastorex meningitis kit (Biorad, Hemel Hempstead, Hertfordshire, UK). When the WHO RRL was called in, qPCR assays were performed at the MRCG to identify the aetiological pathogen from cerebrospinal fluid. *S. pneumoniae* serotyping was performed using latex agglutination [296]. Culture facilities were not available at most health centres, so samples had to be sent to the regional hospital in Sunyani for culture. However, sub-optimal storage and transportation conditions meant that most samples were not viable for culture. Consequently out of 104 laboratory-confirmed cases of pneumococcal meningitis, we recovered just sixteen isolates. These isolates were characterised by antibiotic susceptibility testing to Clindamycin, Optochin, Vancomycin, Trimetoprim-sulfamethoxazole, Erythromycin, Chloramphenicol, Rifampin, Ceftriaxone, Penicillin and Tetracycline. WGS was performed on 15 isolates of *S. pneumoniae* that were stored from the outbreak (Figure 4.3). These included 12 serotype 1 and one each of serotype 35B, serotype 7A and serotype 3.

The majority of the strains analysed by WGS were isolated from cerebrospinal fluid specimens that were sampled from patients during the outbreak (14/15). We analysed one strain that was recovered from a throat swab of a healthy carrier in Wenchi district. Unfortunately we were unable to retrieve the patient information for the throat swab and the exact time of sample collection. The majority of isolates were recovered between the 6th and 9th week of 2016 inclusive. Patients' age ranged from 9 to 45 years (mean 21, median 19) and the ratio of male to female was 3:4 (Table 4.1).

The antibiotic susceptibility profiles to commonly used antibiotics are presented in table 4.1 but the antibiogram for the throat swab isolate (BAR_19) was not reported. Staff at the MRCG reported high rates of resistance to tetracycline and trimethoprim-sulfamethoxazole (commonly known as co-trimoxazole) among these isolates. One serotype 1 isolate was also resistant to penicillin in addition to tetracycline and co-trimoxazole

resistance. The serotype 35B strain was susceptible to all antibiotics that were tested.

Table 4.1 The antibiotic susceptibility profiles for the outbreak *S. pneumoniae* strains to commonly used antibiotics in the sub region.

ID	Age	Sex	Outcome	Clindamycin	Optochin	Vancomycin	Trimetoprim-sulfamethoxazole	Erythromycin	Chloramphenicol	Rifampin	Ceftriaxone	Penicillin	Tetracycline
BAR_12	18	F	DEAD	S	S	S	R	S	S	S	S	R	R
BAR_19													
BAR_15	20	M	ALIVE	S	S	S	I	S	S	S	S	S	R
BAR_16	11	M	DEAD	S	S	S	R	S	S	S	S	S	R
BAR_3	45	F	DEAD	S	S	S	S	S	S	S	S	S	S
BAR_11	10	F	ALIVE	S	S	S	I	S	S	S	S	S	R
BAR_10	9	M	ALIVE	S	S	S	R	S	S	S	S	S	R
BAR_17	13	F	ALIVE	S	S	S	I	S	S	S	S	S	S
BAR_7	22	F	ALIVE	S	S	S	I	S	S	S	S	S	R
BAR_5	35	M	DEAD	S	S	S	R	S	S	S	S	S	R
BAR_9	14	F	ALIVE	S	S	S	R	S	S	S	S	S	R
BAR_6	17	F	ALIVE	S	S	S	I	S	S	S	S	S	R
BAR_14	30	F	DEAD	S	S	S	I	S	S	S	S	S	R
BAR_13	24	M	ALIVE	S	S	S	S	S	S	S	S	S	R
BAR_18	35	M	DEAD	S	S	S	R	S	S	S	S	S	R

4.2.3 Outbreak summary and workflow for sequence analysis

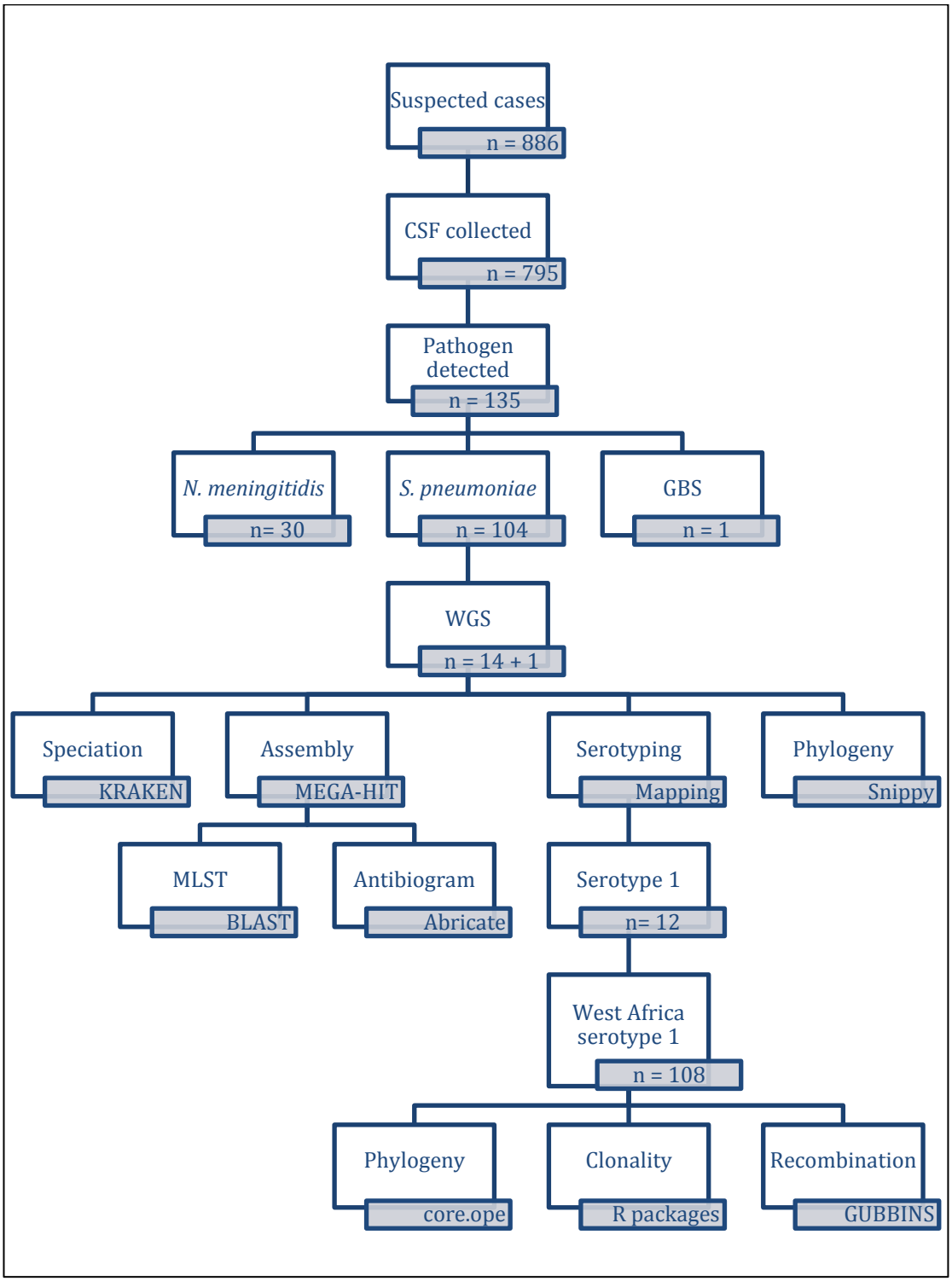


Figure 4.3 A summary of the pneumococcal meningitis outbreak and the workflow for the bioinformatics analysis.

4.3 Results

4.3.1 Genome analysis

The 15 outbreak strains were initially characterised using the nullarbor pipeline (Section 2.3.4). The statistics for read mapping and genome assemblies are summarized in table 4.2. The sequencing reads were mapped to the spn1041 reference genome using snippy. The median depth of coverage was 107 (range 44 to 446). The average GC content for all strains was 40%. Draft assemblies generated by MEGA-HIT ranged in size from 1978923 to 2151432 base pairs. The number of contigs per genome ranged from 90 to 852.

Table 4.2 Summary of statistics for whole genome sequence mapping to spn1041 reference genome and genome assembly.

	Read Mapping			Genome assembly		
BAR_ID	**Reads**	**GeeCee**	**Depth**	**Contigs**	**bp**	**N50**
BAR_12	3171580	39.8	414	742	2088168	4899
BAR_19	2746326	39.8	375	480	2093176	9511
BAR_15	3263496	39.8	428	852	2059204	3736
BAR_16	499400	40.3	56	123	2083942	41179
BAR_3	2608300	39.5	356	429	2151432	11081
BAR_11	343746	40.6	44	90	1978923	54852
BAR_10	451008	40.4	55	128	2084361	35741
BAR_17	701526	40.4	83	116	2080555	43628
BAR_7	471882	40.4	61	125	2118549	41387
BAR_5	387734	40.1	44	124	2055102	39459
BAR_9	3267338	39.8	446	762	2079240	4967
BAR_6	2453604	39.8	336	386	2093558	11772
BAR_14	436534	40.3	55	131	2118859	41265
BAR_13	935610	40.4	114	123	2078016	43265
BAR_18	869786	40.2	107	133	2077163	38302

Sequence data has been submitted to the European Nucleotide Archives under the study accession number PRJEB15437.

4.3.2 Characterisation of strain genotypes

All 15 strains were confirmed to be *S. pneumoniae* by uploading sequencing reads onto kraken through the nullarbor pipeline. By mapping the sequencing reads to the capsular region for 95 pneumococcal serotypes *in silico* serotyping was performed. Serotypes were assigned based on percentage coverage of the capsular region and average coverage depth. In total 12 strains were serotype 1 and the remaining three were serotype 3, 7A and 35B respectively. MLST profiles were inferred by a BLAST comparison of the assemblies to known MLST alleles. Serotype 1 isolates all belonged to ST303 while serotypes 3 7A and 35B strains belonged to the genotypes ST700, ST2833 and ST373 respectively (Table 4.1).

Table 4.3 The genomic characterisation of the *S. pneumoniae* strains by MLST and serotyping.

	***In silico* serotyping**			**MLST**
ID	**Serotype**	**Average coverage**	**Percentage Coverage**	**ST**
BAR_12	1	360.51	99.93%	303
BAR_19	1	342.17	100.00%	303
BAR_15	1	370.68	100.00%	303
BAR_16	1	48.97	100.00%	303
BAR_3	35B	290.75	100.00%	373
BAR_11	7A	17.99	95.51%	2833
BAR_10	1	46.76	100.00%	303
BAR_17	1	64.40	99.98%	303
BAR_7	1	47.26	100.00%	303
BAR_5	3	27.41	54.45%	700
BAR_9	1	413.94	100.00%	303
BAR_6	1	306.04	100.00%	303
BAR_14	1	45.69	100.00%	303
BAR_13	1	85.91	100.00%	303
BAR_18	1	94.98	100.00%	303

4.3.3 Phylogeography of outbreak strains

The sequencing reads from all outbreak strains were mapped to the spn1041 reference genome with snippy and variants were called from the core genome using snippy-core (Alignment in Appendix 4.1). The maximum likelihood was reconstructed using FastTree.

The phylogenetic tree resolved the serotype 1 strains into two clades (Figure 4.4A). Eight serotype 1 strains formed a novel clade that was distinct from the spn1041 reference genome. Isolates within the novel clade were closely related with a within-clade average pairwise distance of 16 core genome single nucleotide variants (SNVs) (Table 4.4). The most similar strains within the novel clade differed at four core genome loci and the most divergent pair differed by 27 SNVs. Four other outbreak serotype 1 strains clustered in a clade with the ST303 reference genome spn1041 (henceforth referred to as reference-type ST303). Strains within the reference-type ST303 clade differed by an average of 91 core genome SNVs (Table 4.4). The strains belonging to other serotypes differed from serotype strains by at least 16,000 SNVs and formed outliers on individual long branches of the phylogenetic tree.

A phylogeography analysis was attempted by linking the annotated phylogenetic tree to a district map showing the incidence rates of confirmed cases in the respective districts of the Brong-Ahafo region. Seven strains in the novel clade were isolated from patients in Techiman, Wenchi or Tain; districts within the epicentre of the outbreak (Figure 4.4B). During the outbreak the incidents rates in the epicentre rose above the epidemic threshold for meningitis outbreaks [283]. The eighth strain from the novel clade was isolated from a patient in Tano South, south of the epicentre. Two reference type ST303 strains were isolated from patients in Nkoranza South and one from Techiman while the origin of the fourth was unknown (on Figure 4.4 this strain was assigned to Sunyani where the regional hospital is located). The non-serotype 1 strains were isolated from the fringe districts, Jaman South, Dorma and Tano South, where the number of confirmed cases was relatively low.

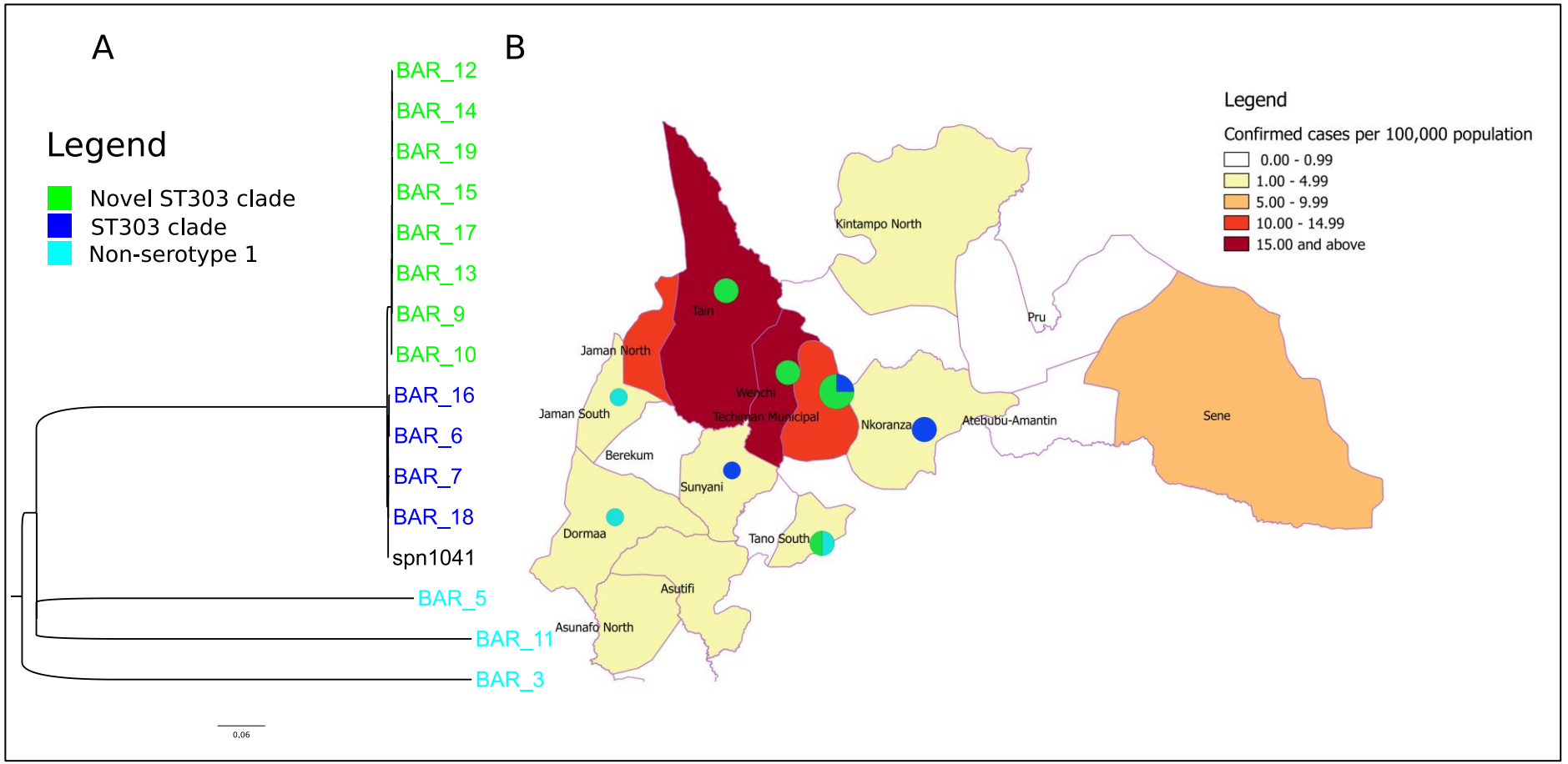


Figure 4.4 A phylogeographic analysis of *S. pneumoniae* strains from the meningitis outbreak. (A) The phylogeny of the outbreak strains was coloured based genotype: novel clade, ST303 and non-serotype 1. (B) The map of Brong-Ahafo shaded to represent the number of confirmed cases per 100,000 in each district during the outbreak. Pie chart shows the prevalence of each genotype in each district.

Table 4.4 A distance matrix showing the number of core SNV differences between pairwise comparisons of the meningitis outbreak *S. pneumoniae* strains. Values are colour coded based on a scale of magnitude: dark blue = 0, white = medium and red = highest value.

ID	BAR_12	BAR_19	BAR_15	BAR_10	BAR_17	BAR_14	BAR_13	BAR_9	Ref	BAR_16	BAR_7	BAR_6	BAR_18	BAR_3	BAR_11	BAR_5
BAR_12	0	6	18	23	11	15	10	18	247	307	298	310	274	17173	16950	16105
BAR_19	6	0	12	17	7	13	4	12	245	305	296	308	272	17171	16948	16103
BAR_15	18	12	0	27	17	23	14	22	257	317	308	320	284	17181	16960	16113
BAR_10	23	17	27	0	22	28	19	23	236	296	287	299	263	17164	16965	16094
BAR_17	11	7	17	22	0	16	7	17	250	310	301	313	277	17176	16953	16106
BAR_14	15	13	23	28	16	0	15	23	250	310	301	313	277	17178	16955	16110
BAR_13	10	4	14	19	7	15	0	14	249	309	300	312	276	17175	16952	16105
BAR_9	18	12	22	23	17	23	14	0	257	317	308	320	284	17183	16960	16115
Reference	247	245	257	236	250	250	249	257	0	80	71	83	47	17138	16900	16050
BAR_16	307	305	317	296	310	310	309	317	80	0	127	73	103	17180	16936	16094
BAR_7	298	296	308	287	301	301	300	308	71	127	0	130	90	17186	16947	16100
BAR_6	310	308	320	299	313	313	312	320	83	73	130	0	106	17183	16939	16091
BAR_18	274	272	284	263	277	277	276	284	47	103	90	106	0	17163	16922	16077
BAR_3	17173	17171	17181	17164	17176	17178	17175	17183	17138	17180	17186	17183	17163	0	17918	17565
BAR_11	16950	16948	16960	16965	16953	16955	16952	16960	16900	16936	16947	16939	16922	17918	0	17048
BAR_5	16105	16103	16113	16094	16106	16110	16105	16115	16050	16094	16100	16091	16077	17565	17048	0

4.3.4 Comparison with West African ST303 and ST217

An assembly based pipeline was employed to reconstruct the phylogeny of the outbreak serotype 1 strains alongside 108 West African serotype 1 archived genomes: 68 from ST303 and 44 from ST217 (full methods in Section 2.4.2). The phylogeny was reconstructed from 9389 variable core genome sites (Alignment in Appendix 4.2). The novel clade from the outbreak was placed on a divergent branch that stemmed from the West African ST303 cluster (Figure 4.5). The ST303 and ST217 clades clustered strains from the same country together. The reference type ST303 strains from the outbreak clustered with Ghanaian ST303 strains within the West African ST303 clade. The phylogenetic placement of ST217 strains highlights that ST217 is the ancestral founder of the ST303 lineage.

Based on the topology of the phylogenetic tree strains were classified into three clusters: West African ST217, West African ST303 and the novel clade. Diversity indexes were computed for each of the clusters using R statistical packages. The Shannon diversity index was computed for each cluster and it confirmed a low level of sequence diversity in the novel clade compared to the West African clades. Similarly, the index of association (IA) and the standardised index of association (rÂ¯d) were computed for each cluster and both showed that the novel clade had a higher degree of clonality than the West African clades (Table 4.5).

Table 4.5 The clonal diversity indexes of West African and outbreak serotype 1 strains.

Cluster name	Number of strains	Shannon-Weiner Index	Index of association (IA)	Standardized index of association (rÂ¯d)
ST303	68	4.22	1157.25	0.19
Novel clade	8	2.08	1582.16	0.35
ST217	44	3.78	314.79	0.11

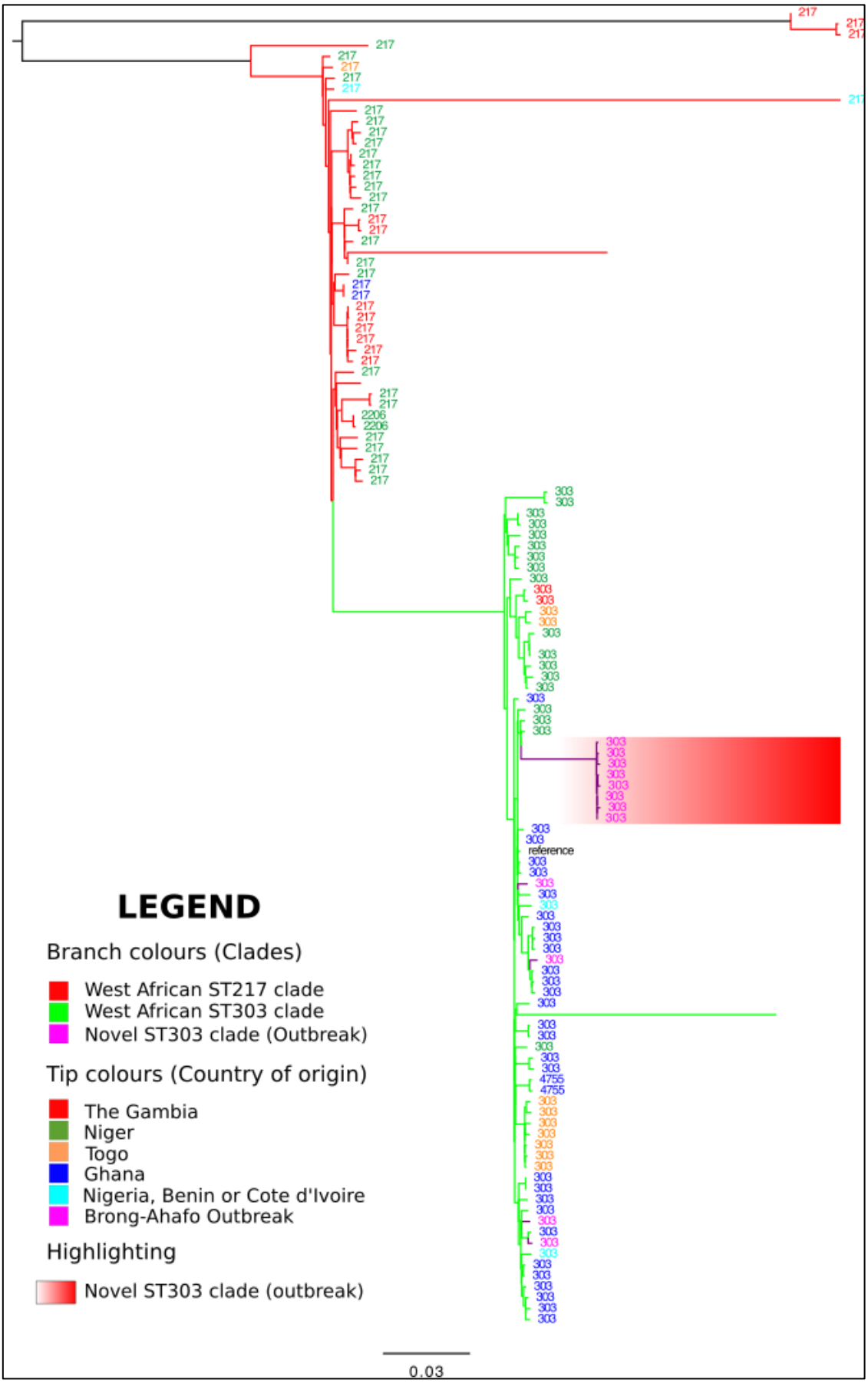


Figure 4.5 The phylogeny of the outbreak serotype 1 strains and the West African serotype 1 genotypes ST303 and ST217. Tips without labels are strains with new STs.

4.3.5 Recombination events that define the novel clade

Recombination regions were predicted from an alignment of 120 serotype 1 strains including the outbreak strains and the West African genotypes ST303 and ST217 using Gubbins (Predicted recombination regions are listed in Appendix 4.3). It appears that the West African ST303 strains evolved from the ST217 ancestral clade through a series of recombination events. Several recombination blocks were conserved in all ST303 strains, including the novel clade, but were absent in ST217 strains (Figure 4.7).

The novel clade was placed on a long branch that had accumulated 271 SNVs: 222 SNVs occurred in recombination regions and 49 SNVs occurred outside of recombination regions. The evolution of the novel clade from the West African ST303 clade was driven by recombination events that were unique to the novel clade. Seven recombination blocks, which varied in size from 349 to 15406 base pairs (median 5091 bp), were conserved in all strains within the novel clade but absent in the West African clades.

The largest recombination block (i) overlapped with a region flanking the fucose metabolism operon (Figure 4.6). This region encompassed numerous regulatory genes most notably the *dltABCD* operon, the zinc uptake regulatory protein *acdA* and a MarR-family regulatory protein (Table 4.6). Other recombination blocks overlapped with numerous metabolism proteins, putative proteins, ATP transporter proteins, the regulatory protein *recX* and a DNA repair system protein C (Table 4.7).

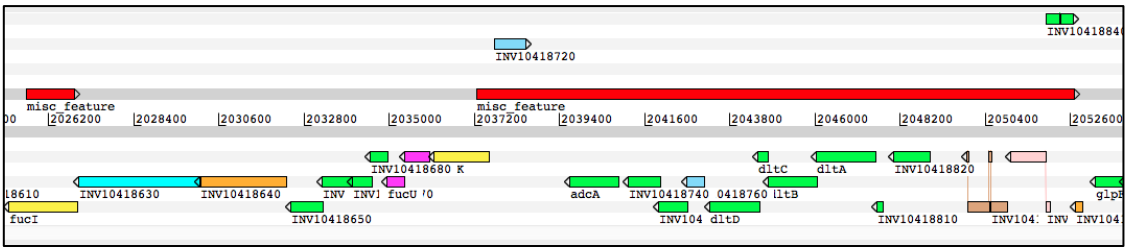


Figure 4.6 An Artemis window snapshot showing the largest recombination region (i) (misc_feature) overlapping with a region flanking the fucose metabolism operon.

Figure 4.7 A heatmap showing the presence/absence of recombination blocks next to the phylogeny of serotype 1 reconstructed by Gubbins. Red blocks are present in multiple strains and blue blocks are present in one strain. The three genotypic clusters are highlighted.

Table 4.6 Table showing the genes that overlap with the recombination regions that characterise the novel clade (blocks I - iii). The start and stop positions show where the recombination starts and stops on the reference genome. The sizes of the recombination blocks are listed.

Region	Start	Stop	Size (b.p.)	Gene name /locus tag	Description
i	2037275	2052681	15406	*fucK*	putative fuculokinase
				INV10418720	putative fucose phosphotransferase system repressor
				adcA	Zinc-binding protein AdcA precursor
				INV10418740	ABC transporter permease protein
				INV10418750	ABC transporter ATP-binding protein
				INV10418760	MarR-family regulatory protein
				dltD	putative D-alanyl-lipoteichoic acid biosynthesis protein
				dltC	D-alanyl carrier protein
				dltB	putative activated D-alanine transport protein
				dltA	D-alanine--poly(phosphoribitol) ligase subunit 1
				INV10418820	major facilitator family transport protein
				INV10418830	putative reverse transcriptase - Group II intron (pseudogene)
				INV10402910	Group II intron maturase
ii	2025618	2026848	1230	*fucI*	putative L-fucose isomerase
iii	1933663	1938754	5091	*gltX*	glutamyl-tRNA synthetase
				pgi	glucose-6-phosphate isomerase
				INV10417860	putative peptidase
				INV10417870	ABC transporter ATP-binding membrane protein

Table 4.7 A continuation of table 4.6 showing the genes that overlap with recombination regions (blocks iv – vii).

Region	Start	Stop	Size (b.p.)	Gene name /locus tag	Description
iv	**208670**	**209752**	**1082**	INV10401540	putative metallopeptidase.
v	**575174**	**581949**	**6775**	*uvrC* INV10405190 *pepV* INV10405210 INV10405180 *brnQ*	UvrABC DNA repair system protein C nitroreductase family protein Oxygen-insensitive NAD(P)H nitroreductase putative Xaa-His dipeptidase uracil DNA glycosylase superfamily protein putative extracellular solute-binding protein putative branched-chain amino acid transport system carrier protein
vi	**1781679**	**1782028**	**349**	*recX*	Regulatory protein recX
vii	**1852685**	**1858834**	**6149**	INV10417140 INV10417170 INV10417180 INV10417190 INV10417200 *asnB*	putative surface anchored protein (pseudogene) putative aminotransferase putative membrane protein putative universal stress protein haloacid dehalogenase-like hydrolase putative L-asparaginase

4.4 Discussion

4.4.1 Novel clade drove the outbreak

Novel lineages of serotype 1 can emerge through clonal expansion and in some cases they can progress to become the dominant clone [413]. It is highly likely that the novel clade was the dominant clone in the meningitis outbreak since it was the most prevalent genotype among the genome-sequenced isolates and it was predominantly isolated from the epicentre of the outbreak. Additionally, the close genetic relationship between the strains suggests that the strains in the novel clade were undergoing clonal expansion and spreading through the community. The other strains isolated from patients during the outbreak such as the reference-type ST303 strains were genetically more diverse, which suggested they were not being transmitted or undergoing clonal expansion. These strains are likely to represent sporadic cases of meningitis that occurred during the time of the outbreak.

4.4.2 Evolution of the novel clade

The evolution of novel clones due to selective pressure from the host is an on going phenomenon for *S. pneumoniae* [414]. However, due to homologous recombination, most novel clonal clusters are transient and get reabsorbed into the main clonal lineages [415]. There is evidence that the novel clade underwent substantial sequence diversity, primarily through homologous recombination, to form a distinct cluster. Recombination blocks overlapped with genes responsible for important cellular functions that could impact on fitness and virulence:

- The *dltABCD* operon mediates D-alanylation of lipoteichoic acids, a mechanism for evasion of host innate immune response through repulsion of cationic antimicrobial peptides [416]. Thus, modification of the *dlt* operon through recombination could have enhanced host immune evasion in the novel clade

- MarR-family proteins regulate expression of virulence factors and responses to antibiotic and oxidative stress [417]. Mutations in MarR-family genes have been associated with increased virulence and invasiveness in a number of bacterial species [418]. In the novel clade, modification of the MarR family protein could have led to increased virulence and invasiveness
- ATP transporter proteins play an important role in bacterial virulence through uptake of nutrients and metal ions as well as in host cell attachment [419]. ATP transporters are highly immunogenic and have been considered as potential vaccine targets for *S. pneumoniae* [419]
- The zinc-uptake regulatory protein *adcA* is highly immunogenic and plays an important role in *S. pneumoniae* infection [417]. Modification of the *adcA* gene by recombination could enhance immune evasion

4.4.3 Vaccine-induced selective pressure

The introduction of vaccines can induce selective pressures that favour serotype replacement through capsular switching or the rise of low prevalence non-vaccine serotypes [116,420]. The very small sample size that we had showed 5 cases due to 12F. This could have represented a substantial transmission network supported by the same conditions that drove the overall outbreak. Alternatively they might all have been sporadic, and just at the usual background level. There was no evidence of capsular switching through the alteration of *cps* genes in the novel clade. However, it cannot be ruled out that vaccine-induced selective pressure may have contributed to the emergence of this novel clone.

Conjugate vaccines alter the nasopharyngeal microbiome by eliminating the most prevalent serotypes [421]. This provides new opportunities for *S. pneumoniae* to acquire novel DNA by recombination within the modified microbial ecology of the nasopharnyx [414]. It is known that vaccine intervention selects for *S. pneumoniae* lineages that have undergone recombination [116,403]. It can be hypothesized that the ancestral clone of the novel clade underwent rapid evolution through recombination and then was

selected for by the introduction of PCV-13. The recombination events are likely to have conferred a selective advantage through enhanced virulence and potential host immune evasion mechanisms.

4.4.4 Potential reasons for outbreak

A likely scenario is that at the time of the outbreak, the strains from the novel clade were already disseminated in the population. The most divergent strains within the novel clade differed by as many as 28 core genome SNVs (without excluding recombination). This is an indication that the novel clade has been undergoing clonal expansion for some time, perhaps a few years. The dust and low humidity that characterises the dry season favoured a switch from carriage to invasiveness, leading to disease. This hypothesis is supported by the emergence of sporadic cases of meningitis due to other serotypes and other genotypes of ST303 serotype 1.

Once the new clade emerged with potentially superior fitness, it could have easily spread among the older unvaccinated population in the community. This is supported by the fact that overall over 95% of confirmed cases occurred among children over 5 and adults. Similarly, the novel clade strains were isolated from patients aged 9 years and above. A strain isolated from a healthy individual was also part of the novel clade, which suggests that strain from the novel clade may have been disseminated in the community prior to the outbreak. Unfortunately this cannot be confirmed since there is no information on when this isolate was collected.

4.4.5 Need for carriage surveillance

A carriage study needs to be carried out in Ghana in order to provide baseline data on the molecular epidemiology of *S. pneumoniae* after the introduction of PCV-13. This is important since the altered dynamics of competition and collaboration between pneumococcal strains and with bacterial species can drive the evolution of novel clones of *S. pneumoniae* [414]. A carriage study would identify the emergence of novel clones and the dissemination of genotypes that have the potential to cause lethal outbreaks.

4.4.6 Limitations

The main limitation of this paper is the small sample size. This highlights the vital need to strengthen laboratory services across West Africa to a level where they can perform bacterial culture and identification. The lack of longitudinal carriage data makes it difficult to predict when the novel clade emerged. A Bayesian estimate phylogenetic dating approach could have shed light on the potential timeline for the emergence of the novel clade. Unfortunately I was unable to attempt phylogenetic dating on the novel clade because the samples were collected within a short time frame. The paucity of patient metadata and small sample size ruled out the feasibility of a risk factor analysis using regression models. The use of wider reference datasets would support more solid inference on the origin of the clade.

5 CHAPTER FIVE: GENOMIC EPIDEMIOLOGY OF TUBERCULOSIS IN WEST AFRICA

5.1 Introduction

5.1.1 Global population structure of *M. tuberculosis*

The *Mycobacterium tuberculosis* complex (MTBC) consists of several ecotypes that are genetically distinct yet closely related. The complex is characterised by identical 16S ribosomal RNA sequences and 99.9% nucleotide identity at the whole genome level [284]. Ecotypes within MTBC are adapted to infect a wide range of hosts. Strains belonging to the same ecotype are differentiated by means of genotyping techniques into lineages and families [422].

The main etiological agents of tuberculosis (TB) in humans are *Mycobacterium tuberculosis sensu stricto* and *Mycobacterium africanum* [284]. *M. tuberculosis* and *M. africanum* form seven distinct phylogenetic lineages that are believed to have co-evolved with humans over millennia [423-425] (Figure 5.1). The global diversity of the MTBC may be explained by recent patterns of migration, trade and conquest [426]. It has been proposed that MTBC originated in East Africa [426] before spreading out of Africa with human migration. Certain lineages (e.g. the Euro-American lineage) were later re-introduced into Africa by travellers, traders and through conquest [423]. The Out-of-Africa scenario estimates on the age of the MTBC is 40,000 – 70,000 years [423]. However studies that relied on MTBC genomes recovered from ancient DNA estimate that the complex is approximately 6000 years old [424,427].

Comparative genomics of MTBC has revealed that significant genetic variation exists between different lineages of MTBC, which can impact the pathogenicity of and transmissibility strains [428].

Some lineages of the MTBC are more prevalent in certain geographical regions than others and this is reflected in the nomenclature of the lineages [429] (Figure 5.1):

- The East Asian, East-African-Indian and Euro-American lineages (Lineages 2,3 and 4 respectively) form the so-called “modern clade” of the MTBC, characterised by the absence of the TbD1 region [429]
- Lineages 2 (including Beijing strains) and 4 are more widespread than other MTBC lineages and are more pathogenic and virulent [428]
- The Indo-Oceanic (lineage 1) and the *M. africanum* lineages West Africa 1 and 2 (also called Lineages 5 and 6 or MAF1 and MAF2) are among the ancient lineages of the MTBC. The ancient lineages are more geographically restricted that lineages 2 and 4
- Lineage 7 (Ethiopian) was recently discovered around the horn of Africa: phylogenetically lineage 7 is placed between the modern clade of the MTBC and other lineages [430]

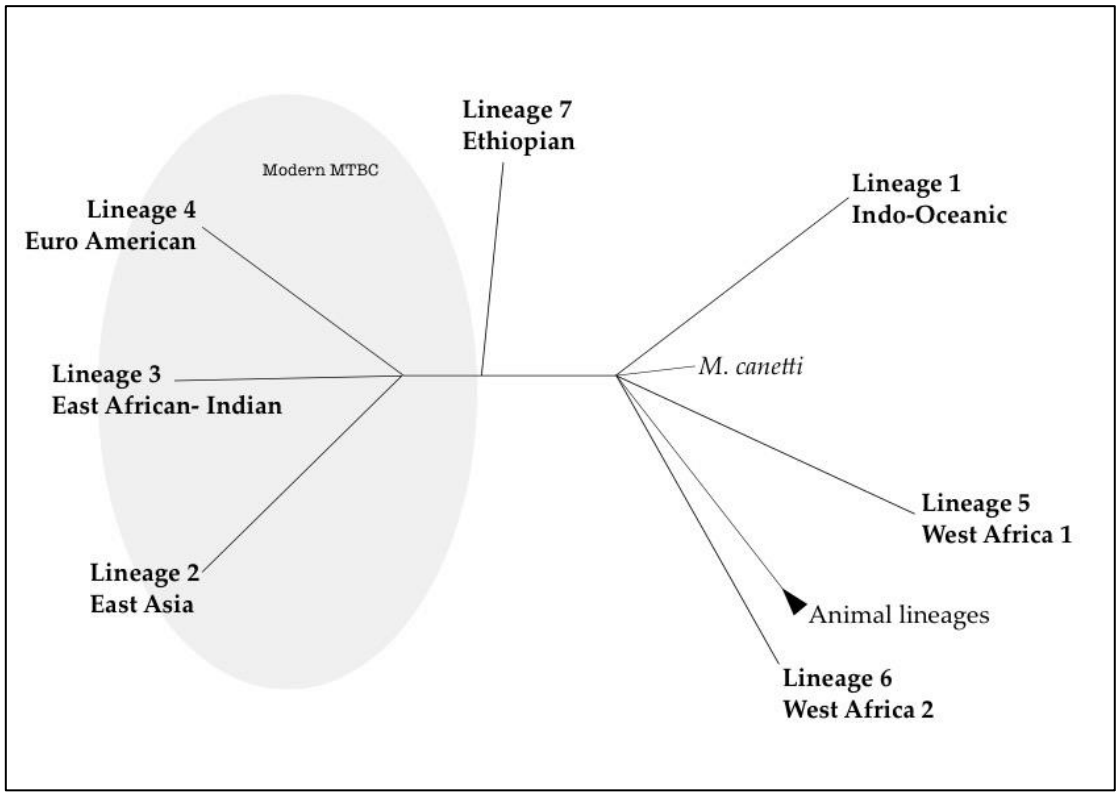


Figure 5.1 A schematic representation summarizing the phylogenetic relationships of the major the MTBC lineages that cause human tuberculosis. Tips are labelled by lineage. Tree topology and nomenclature adapted from [423] and [429] respectively.

5.1.2 Population structure of TB in West Africa

A unique feature of TB in West Africa is the presence of all six major lineages of the MTBC (Lineages 1-6). However, the Euro-American and the two *M. africanum* lineages cause most human pulmonary TB in West Africa [294].

M. africanum is almost exclusive to West Africa and sporadic cases in other regions are generally linked to West African migrants [431]. Geographic clustering is observed within the two *M. africanum* lineages:

- MAF1 is more prevalent around the Gulf of Guinea
- MAF2 is more prevalent in the western regions of West Africa [431]

The restriction of *M. africanum* to West Africa may be driven by an unknown animal reservoir. This hypothesis is supported by the close phylogenetic relationship that exists between *M. africanum* and other animal lineages [428,432]. An alternative hypothesis is that due to its inferior fitness, *M. africanum* cannot compete with other *M. tuberculosis* lineages in other regions [433]. However, this does not explain why *M. africanum* remains an important cause of disease in West Africa, despite the presence of "fitter" lineages in the region [294].

A plausible scenario is that *M. africanum* is adapted to infect the human genotypes that are predominant in West Africa. A recent study has shown an association between lineage 5 (MAF1) and the Ewe ethnic group in Ghana [434]. Similarly, a retrospective study in the US showed reported a high prevalence of *M. africanum* among US-born African Americans [435]. Co-evolution of the MTBC with the human host is supported by striking topological resemblances between the phylogeny of the MTBC and the phylogeny of human mitochondrial genomes [423].

The Cameroon clade has emerged as the dominant genotype in much of West Africa, particularly the eastern parts [436-440]. The reasons for the selection and dissemination of this genotype in West Africa remain unclear, highlighting the need to investigate its evolutionary origins through WGS. In Ghana, the Cameroon clade was most likely introduced through the south of

the country, where a higher proportion of infection is associated with this clade than in the north [441].

The Ghana genotype (or T1 lineage) is disseminated across almost all West African countries [294]. The evolutionary origins of this genotype remain a mystery and its epidemiology is not well described. However, it is the most prevalent genotype in Guinea [442] and evidence from Ghana and Mali suggests that the Ghana genotype is associated with drug resistance [441,443].

The Beijing clade is a globally disseminated lineage that is highly transmissible and is associated with drug resistance [428,444-446]. Although the prevalence of the Beijing genotype in West Africa is generally low, its prevalence is increasing in some countries. For example, Beijing is the most common genotype causing human pulmonary in Benin [447]. Similarly a high prevalence of Beijing strains has been recorded in Senegal [448]: both countries host large international ports.

The animal lineages of the MTBC are associated with a wide range of host species [284]:

- *Mycobacterium* bovis (cattle)
- *Mycobacterium caprae* (sheep and goats)
- *Mycobacterium microti* (voles)
- *Mycobacterium mungi* (mongoose)
- *Mycobacterium pinnipedii* (seals and sealions)
- *Mycobacterium orygis* (antelope)

M. bovis has been found to cause human pulmonary TB in West Africa [449,450]. Patients infected with *M. bovis* need special treatment, because *M. bovis* isolates are intrinsically resistant to pyrazinamide [451], which is part of the recommended treatment regimen for drug-sensitive TB.

5.1.3 Multidrug-resistant tuberculosis (MDR-TB)

Mycobacteria are intrinsically resistant to most commonly prescribed antibiotics due to a thick cell envelope that serves as a barrier and to intracellular mechanisms that evade or deactivate antibiotics [452]. This means that TB is usually treated with drugs specific to management of this infection. However, mycobacteria are able to acquire resistance to anti-TB drugs, through chromosomal mutations in the drug target gene, the drug target promoter region and/or drug activation enzymes [452].

Multidrug resistant tuberculosis (MDR-TB) emerges through the acquisition of resistance in the individual patient during treatment or through transmission of MDR-TB from patients with active pulmonary disease. Drug resistance emerges through spontaneous resistance mutations that are selected for during intermittent exposure to sub-lethal levels of drugs. Such sub-optimal levels arise due to poor adherence, poor drug quality or patient-specific pharmacodynamics [290]. However, in high-burden settings, inter-patient transmission is the main source of MDR-TB infection [453]. WGS can reliably trace the emergence MDR-TB during treatment and can identify on going transmission chains [454,455].

Early diagnosis of MDR-TB is instrumental in ensuring successful treatment. MDR-TB is confirmed by antimicrobial susceptibility testing. Culture-based phenotypic testing is the gold standard, but it is slow and laborious. The WHO now recommends rapid testing, at least for rifampicin resistance, at the time of diagnosis [292]. Molecular based techniques such as Xpert® MTB/RIF (Cepheid Inc.) and the Hain-MTBDR-plus kit provide rapid early diagnosis of MDR-TB. Whole-genome sequencing is emerging as a useful tool for diagnosing and guiding treatment of MDR-TB by offering early detection of resistance to both first- and second-line drugs [456-458].

It is recommended that MDR-TB cases be treated with individualized treatment regimens guided by drug susceptibility testing [459]. Unfortunately the global success rate for treating MDR-TB is only about 50% [460]. Moreover, a large number of cases go undetected or unreported: e.g. in 2012 only 2% of the estimated cases of MDR-TB in Nigeria were reported [460].

5.1.4 Rationale for the WATENAM study

In the 2016 global TB report, the WHO highlights a major shortfall in the annual funding on TB research and development of TB treatment [292]. This shortfall is profound in West Africa where 12 of the 15 countries are among the 42 least developed countries worldwide [461]. Since the year 2000, only The Gambia, Ghana and Nigeria have reported surveillance on the prevalence of TB [292]. There is a paucity of data on the prevalence of drug-resistant TB in West Africa. The vast majority of health facilities in the region lack the facilities for handling *Mycobacterium tuberculosis*, a category 3 pathogen.

The West African Nodes of Excellence for Tuberculosis, AIDS and Malaria (WANETAM) project was set up in 2008, with the aim of tackling these three major public health problems through sub-regional collaboration. The TB nodes of excellence were set up in eight countries (The Gambia, Senegal, Guinea Bissau, Togo, Ghana, Burkina Faso, Mali and Nigeria) with two aims:

- To provide capacity-building through training in mycobacterial culture and identification to prepare the sites for clinical trials
- To estimate the burden of drug resistance through recruitment of patients at tertiary health facilities and antimicrobial susceptibility testing

The WANETAM coordinators have provided numerous training opportunities, including courses in:

- Good clinical and laboratory practice
- Identification, culture, storage and transport of mycobacterial isolates
- Molecular typing and rapid detection of mycobacteria
- Data analysis and publishing
- Introductory bioinformatics, taught by the author

5.1.5 Objectives and study rationale

I wished to exploit whole-genome sequencing to study the epidemiology of TB and genetics of multi-drug resistance among patients with active pulmonary disease in West Africa. Due to limitations in time and money, it was not possible to sequence all the isolates from the WANETAM sites. I therefore undertook a pilot study of isolates from two West African countries:

- **Nigeria**, where the high burden of drug resistance makes it an ideal setting to study the molecular mechanisms of MDR-TB in West Africa
- **Mali**, which has a well-structured hierarchy of TB reference health care facilities in the capital Bamako. The WANETAM dataset presented the opportunity to study the epidemiology of the TB in this city

I used whole genomes sequencing data to:

- **Understand the genomic epidemiology of MTBC in each country** by reconstructing the phylogeny of MTBC and classifying strains into the global phylogenetic lineages and families
- **Investigate the existence of inter-patient transmission** by searching for identical or closely related strains in different patients
- **Understand the molecular mechanisms that are driving the emergence of dug resistance** by scanning the genomes for known resistance inducing mutations
- **Understand the evolution of drug resistance** by studying the patterns of resistance mutations in the context of the phylogeny

5.2 Methods

5.2.1 Study sites

The WANETAM network has coordinated TB drug resistance surveys in nine sites across eight West African countries (Figure 5.2):

1. **Senegal,** Laboratoire Bactériologie Virologie, Le Dantec, Dakar
2. **The Gambia,** National Public Health Laboratory Services, Banjul
3. **Guinea-Bissau,** National Institute for Public Health, Bissau
4. **Mali,** SEREFO (HIV/TB Research and Training Center) Faculté de Médecine, de Pharmacie et d'Odento-Stamologie, University of Bamako, Bamako
5. **Burkina Faso,** Centre Muraz and the National TB Program, Ouagadougou
6. **Ghana**, Korle Bu Teaching Hospital, Accra
7. **Togo,** Laboratoire National de Référence des Mycobactéries, Lome
8. **Nigeria,** Nigerian Institute for Medical Research, Lagos
9. **Nigeria,** College of Medicine, University of Ibadan, Ibadan

Patients with confirmed active pulmonary TB based on WHO guidelines [462] were recruited at the study sites. Mycobacteria were cultured on site and the pure isolates were sent to the MRCG for viability testing and antimicrobial susceptibility testing to first and second line anti-TB drugs. The Joint Gambia Government/MRC Ethics Committee and the relevant ethical bodies in other participating countries approved the study protocols.

As noted, two countries were chosen for investigation in this study: Nigeria and Mali. Nigeria borders Cameroon, Chad, Niger and Benin and has a coastal region on the Gulf of Guinea. Nigeria has implemented WHO recommended DOTS (Directly Observed Treatment Short-Course) since 1996. In recent years, Nigeria has increased use of Xpert® for early detection of rifampicin resistance [463]. Despite these measures, Nigeria still has one of the highest burdens of TB globally and is among the six countries that

account for 60% of the global TB burden [292]. Nigeria also has the second largest population of HIV-positive individuals in the world [464]. The rise of MDR-TB in Nigeria has disrupted efforts to eradicate TB.

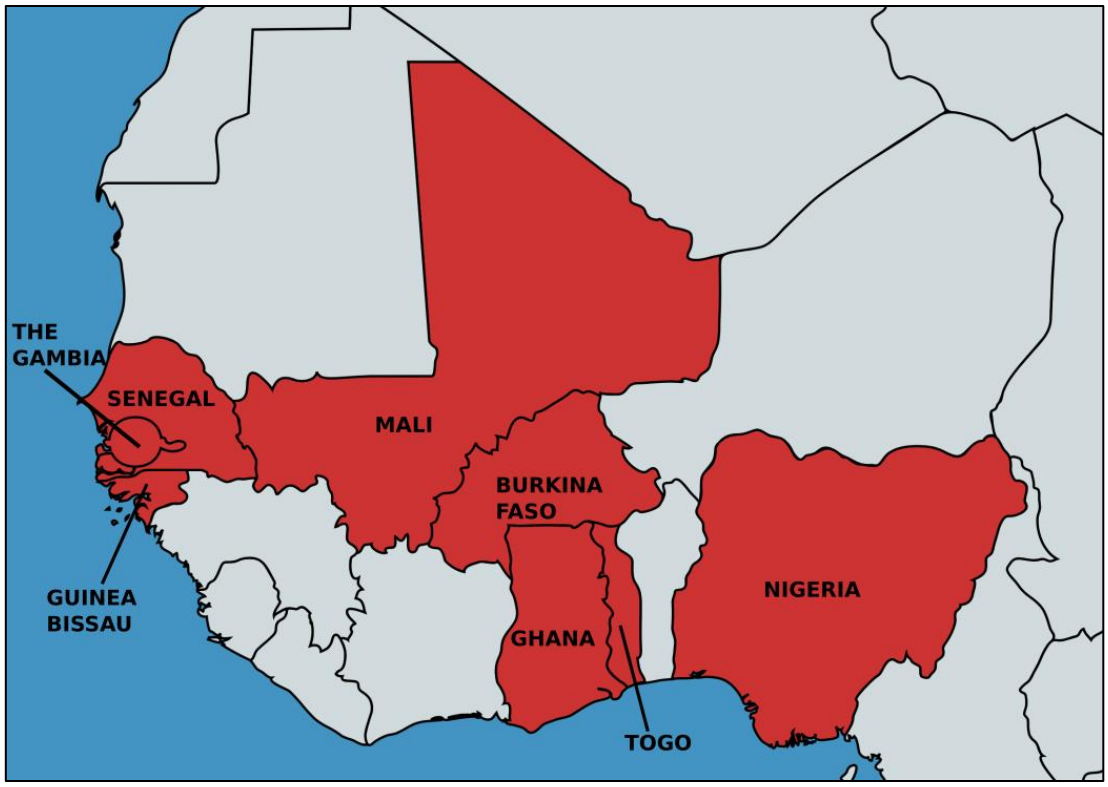


Figure 5.2 Map of West Africa highlighting the countries (in red) that participated in the WANETAM tuberculosis drug resistance survey.

Nigeria was the only country in the WANETAM network with two study sites—Lagos and Ibadan—both in Southwest Nigeria. According to the most recent census figures, Lagos is the most populous city in Africa and Ibadan is the third most populous city in Nigeria [465]. The Nigerian Institute of Medical Research in Lagos is the national TB reference laboratory. The Ibadan University College Hospital is a tertiary health care facility with designated facilities for culturing *M. tuberculosis*.

The WANETAM survey in Nigeria revealed very high rates of MDR-TB (higher than WHO estimates), particularly among retreatment cases. This presented as an opportunity to perform an in-depth study on resistance mutations in this region. Furthermore patients recruited in Lagos were sample at multiple time points over the course of treatment. This gave us a rare opportunity to investigate the microevolution of the MTBC over the course of treatment.

Mali is a large landlocked country that borders seven countries including Senegal, Algeria and Niger. Mali has a population of 15.3 million that consists of people from diverse ethnic backgrounds [466]. The study participants from Mali were recruited in the capital city Bamako. Bamako is highly cosmopolitan and is one of the fastest growing cities in Africa. Bamako has a population of about 2 million people and is divided into six districts. Each district has a TB referral health centre, equipped with diagnosis and treatment facilities. Our isolates were collected at the University Teaching Hospital, which acts as the principal TB referral centre. Retreatment cases from across Mali are referred to the University Teaching Hospital for further investigation and treatment. The Malian Ministry of health reported that in 2014 more than a third of all TB cases in Mali were managed in Bamako [467].

The Mali dataset offered an opportunity to study the molecular epidemiology of the MTBC in Bamako since all the referral health centres report to the central teaching hospital. From the WANETAM survey the prevalence of MDR in Mali was low among new cases, but very high among retreatment cases. Through WGS we set out to study the on going transmission of TB between patients in Bamako and to establish whether MDR-TB among retreatment cases was associated with a particular genotype.

5.2.2 Study isolates and workflow

Mycobacterial isolates were sent to the MRCG laboratories from Mali, Ibadan and Lagos (Table 5.1). Isolates were tested for viability and antimicrobial susceptibility testing to the first-line drugs isoniazid, rifampicin, ethambutol and streptomycin. MDR-TB strains were further tested for susceptibility to the second-line agents fluoroquinolones, kanamycin, capreomycin and ethionamide (Table 5.2). For whole genome sequencing isolates were sub-cultured on Middlebrooks 7H9 liquid medium for 3-8 weeks and genomic DNA was extracted using the cetyl-trimethylammonium-bromide method (Section 2.2.2). Whole-genome sequencing was performed on the Illumina MiSeq using Nextera XT library kits.

The datasets from Mali and Nigeria were analysed separately due to differences in sampling strategy. The datasets comprised of:

- Seventy-five isolates (43 MDR and then 32 non-MDR) that were selected at random from 177 archived Nigerian isolates for whole-genome sequencing. Sequencing failed for two isolates from Nigeria so only 73 isolates were followed up. These isolates had been recovered from 63 patients (21 and 42 from Ibadan and Lagos respectively), with some patients supplying multiple isolates (Tables 5.3 and 5.4).
- Seventy-eight isolates that were selected at random from 208 archived Malian isolates for whole-genome sequencing (Tables 5.3 5.5). All originated from distinct patients. Two sequenced isolates from Mali had no patient metadata or antimicrobial susceptibility profiles.

Detailed methods are provided in Chapter 2 but the workflow for bioinformatics analyses is shown in Figure 5.3. This workflow included the characterisation of strains based on their phylogenetic lineage and genotypically predicted antimicrobial susceptibility profiles. For each dataset a detailed phylogenetic analysis was performed that probed for on going transmission and gave insights into the evolution of drug resistance in tuberculosis infection.

Table 5.1 Summary of patient information for isolates that were collected as part of the WANETAM drug resistance survey in Mali, Ibadan and Lagos.

		Mali		Nigeria Ibadan		Nigeria Lagos	
		No.	%	No.	%	No.	%
	Total	208	-	57	-	120	-
Treatment History	New	150	72	13	23	28	23
	Retreatment	58	28	44	77	88	73
	Unknown	0	0	0	0	4	3
Age	0-14	4	2	3	5	1	1
	15-24	50	24	6	10	16	14
	25-34	78	38	21	37	45	39
	35-44	38	18	15	26	32	28
	45-54	21	10	8	14	12	10
	55-64	11	5	2	4	9	8
	65+	6	3	2	4	1	1
	Unknown	0	-	0	-	4	-
Sex	Female	51	25	23	40	54	46
	Male	157	75	34	60	63	54
	Unknown	0	-	0	-	3	-
HIV	Negative	172	83	42	74	46	38
	Positive	29	14	4	7	27	23
	Unknown	7	3	11	19	47	39

Table 5.2 Prevalence of MDR-TB in Mali and Nigeria (Ibadan and Lagos) from the WANETAM survey stratified by patient treatment history.

Study Site	Treatment history	Total	MDR	
			n	%
Mali	All	**208**	39	19
	New	**150**	5	3
	Retreatment	**58**	34	59
Nigeria Ibadan	All	**57**	17	30
	New	**13**	0	0
	Retreatment	**44**	17	39
Nigeria Lagos	All	**120**	71	59
	New	**28**	9	32
	Retreatment	**88**	58	66
	unknown	**4**	4	100

Table 5.3 Patient data for all genome-sequenced isolates from Mali, Ibadan and Lagos.

		Mali		Ibadan		Lagos	
		MDR	Non-MDR	MDR	Non-MDR	MDR	Non-MDR
Total		12	64	6	15	30	12
HIV Status	Negative	11	51	6	15	21	8
	Positive	1	6	0	0	7	4
	Not Done	0	7	0	0	2	0
Gender	Female	2	19	4	3	11	6
	Male	10	45	2	12	18	6
Treatment History	New	3	58	0	8	4	4
	Retreatment	9	6	6	7	21	1
	Failure	0	0	0	0	2	3
	Relapse	0	0	0	0	2	4
Age (years)	0-15	1	2	0	1	1	1
	16-30	5	30	1	6	10	1
	Above 30	6	32	5	8	18	10

Table 5.4 Metadata for MTBC isolates from Nigeria that were genome-sequenced.

ID	Site	MDR	preXDR	XDR	StudyCase	HIV	Age	Sex	Date
NG46	IBADAN	R	S	S	Retreatment	-ve	35	F	1/8/2012
NG64	IBADAN	R	R	S	Retreatment	-ve	23	F	3/8/2012
NG68	IBADAN	R	S	S	Retreatment	-ve	42	M	7/8/2012
NG70	IBADAN	R	S	S	Retreatment	-ve	47	M	1/8/2012
NG44	IBADAN	R	S	S	Retreatment	-ve	34	F	7/8/2012
NG45	IBADAN	R	S	S	Retreatment	-ve	49	F	1/8/2012
NG88	IBADAN	S			New	-ve	40	M	7/8/2012
NG47	IBADAN	S			New	-ve	38	F	7/8/2012
NG62	IBADAN	S			New	-ve	44	M	7/8/2012
NG81	IBADAN	S			New	-ve	25	F	7/8/2012
NG82	IBADAN	S			New	-ve	7	M	7/8/2012
NG89	IBADAN	S			New	-ve	33	M	7/8/2012
NG84	IBADAN	S			New	-ve	41	M	7/8/2012
NG69	IBADAN	S			New	-ve	77	M	7/8/2012
NG61	IBADAN	S			Retreatment	-ve	51	M	1/8/2012
NG65	IBADAN	S			Retreatment	-ve	29	F	1/8/2012
NG80	IBADAN	S			Retreatment	-ve	47	M	3/8/2012
NG83	IBADAN	S			Retreatment	-ve	17	M	1/8/2012
NG85	IBADAN	S			Retreatment	-ve	27	M	1/8/2012
NG86	IBADAN	S			Retreatment	-ve	28	M	3/8/2012
NG67	IBADAN	S			Retreatment	-ve	30	M	1/8/2012
NG36	LAGOS	R	S	S		NA			3/11/2012
NG55	LAGOS	R	S	S		NA			12/27/2011

NG6	LAGOS	R	S	S	Failure	-ve	28	F	12/27/2011
NG9	LAGOS	R	S	S	Failure	-ve	58	M	12/27/2011
NG12	LAGOS	R	S	S	New	-ve	36	F	12/27/2011
NG13	LAGOS	R	S	S	New	-ve	12	F	12/27/2011
NG25	LAGOS	R	S	S	New	+ve	40	M	3/11/2012
NG2	LAGOS	R	S	S	New	-ve	28	M	3/26/2012
NG10	LAGOS	R	S	S	Relapse	+ve	18	F	12/27/2011
NG16	LAGOS	R	S	S	Relapse	+ve	41	M	12/27/2011
NG42	LAGOS	R	S	S	Relapse	-ve	31	M	12/27/2011
NG1	LAGOS	R	S	S	Retreatment	+ve	41	M	7/17/2013
NG14	LAGOS	R	R	S	Retreatment	-ve	30	F	7/17/2013
NG18	LAGOS	R	S	S	Retreatment	-ve	41	M	7/17/2013
NG22	LAGOS	R	S	S	Retreatment	-ve	30	M	7/17/2013
NG23	LAGOS	R	S	S	Retreatment	-ve	58	M	3/11/2012
NG24	LAGOS	R	S	S	Retreatment	-ve	36	M	3/11/2012
NG26	LAGOS	R	R	S	Retreatment	-ve	35	M	3/11/2012
NG3	LAGOS	R	S	S	Retreatment	-ve	30	M	7/17/2013
NG31	LAGOS	R	S	S	Retreatment	-ve	33	M	7/17/2013
NG32	LAGOS	R	R	S	Retreatment	-ve	26	F	3/11/2012
NG54	LAGOS	R	R	S	Retreatment	-ve	26	M	3/11/2012
NG15	LAGOS	R	S	S	Retreatment	+ve	35	M	3/11/2012
NG30	LAGOS	R	S	S	Retreatment	-ve	57	F	7/17/2013
NG34	LAGOS	R	S	S	Retreatment	-ve	57	F	3/11/2012
NG35	LAGOS	R	S	S	Retreatment	-ve	57	F	3/11/2012
NG21	LAGOS	R	S	S	Retreatment	-ve	30	M	7/17/2013
NG17	LAGOS	R	S	S	Retreatment	-ve	50	F	7/17/2013
NG4	LAGOS	R	S	S	Retreatment	+ve	41	M	7/17/2013
NG43	LAGOS	R	S	S	Retreatment	+ve	41	M	3/11/2012
NG19	LAGOS	R	S	S	Retreatment	NA	32	F	7/17/2013
NG29	LAGOS	R	S	S	Retreatment	-ve	28	M	3/11/2012
NG28	LAGOS	R	S	S	Retreatment	-ve	35	F	3/11/2012
NG33	LAGOS	R	S	S	Retreatment	+ve	45	M	3/11/2012
NG37	LAGOS	R	S	S	Retreatment	+ve	45	M	3/11/2012
NG38	LAGOS	R	S	S	Retreatment	-ve	23	M	3/11/2012
NG5	LAGOS	R	S	S	Retreatment	+ve	50	F	3/11/2012
NG72	LAGOS	S			Failure	-ve	23	M	12/27/2011
NG79	LAGOS	S			Failure	+ve	41	M	12/27/2011
NG39	LAGOS	S			Failure	-ve	15	F	12/27/2011
NG40	LAGOS	S			Failure	-ve	15	F	12/27/2011
NG50	LAGOS	S			New	+ve	46	F	12/27/2011
NG76	LAGOS	S			New	+ve	31	F	12/27/2011
NG51	LAGOS	S			New	-ve	38	M	12/27/2011
NG52	LAGOS	S			New	-ve	32	F	12/27/2011
NG75	LAGOS	S			Relapse	+ve	38	F	12/27/2011
NG77	LAGOS	S			Relapse	+ve	38	F	12/27/2011

NG78	LAGOS	S			Relapse	-ve	36	M	12/27/2011
NG53	LAGOS	S			Relapse	-ve	45	M	12/27/2011
NG71	LAGOS	S			Relapse	-ve	42	F	12/27/2011
NG74	LAGOS	S	S	S	Relapse	-ve	42	F	12/27/2011
NG90	LAGOS	S			Retreatment	-ve	36	M	3/11/2012

Footnote: Table sorted by Site, then MDR status (R = resistant, S = non-MDR) and then Study Case to make the table more intelligible. NA is not applicable or no result available. Date listed is the date that patients were enrolled into the study.

Table 5.5 Metadata for MTBC Isolates from Mali that were genome-sequenced.

ID	MDR	preXDR	XDR	StudyCase	HIV	Age	Sex	Date
MA154								
MA272								
MA209	R	S	S	New	-ve	42	M	02/07/14
MA240	R			New	-ve	72	M	02/07/14
MA026	R	S	S	New	+ve	40	M	05/07/14
MA066	R	S	S	Retreatment	-ve	21	F	05/07/14
MA249	R	S	S	Retreatment	-ve	20	F	02/07/14
MA059	R	S	S	Retreatment	-ve	19	M	05/07/14
MA078	R	S	S	Retreatment	-ve	27	M	05/07/14
MA080	R	R	S	Retreatment	-ve	33	M	05/07/14
MA086	R	S	S	Retreatment	-ve	57	M	05/07/14
MA089	R	S	S	Retreatment	-ve	3	M	05/07/14
MA100	R	S	S	Retreatment	-ve	35	M	05/07/14
MA170	R			Retreatment	-ve	21	M	16/07/14
MA008	S			New	-ve	43	F	05/07/14
MA028	S			New	-ve	24	F	05/07/14
MA057	S			New	-ve	35	F	05/07/14
MA204	S			New	-ve	35	F	02/07/14
MA207	S			New	-ve	45	F	02/07/14
MA210	S			New	-ve	26	F	02/07/14
MA215	S			New	-ve	14	F	02/07/14
MA224	S			New	-ve	23	F	02/07/14
MA225	S			New	-ve	27	F	02/07/14
MA241	S			New	-ve	27	F	02/07/14
MA242	S			New	-ve	22	F	02/07/14
MA245	S			New	-ve	45	F	02/07/14
MA246	S			New	-ve	40	F	02/07/14
MA250	S			New	-ve	4	F	02/07/14
MA003	S			New	-ve	37	M	05/07/14
MA006	S			New	-ve	43	M	05/07/14
MA009	S			New	-ve	34	M	05/07/14
MA010	S			New	-ve	57	M	05/07/14
MA016	S			New	-ve	20	M	05/07/14
MA017	S			New	-ve	26	M	05/07/14

MA018	S			New	-ve	25	M	05/07/14
MA019	S			New	-ve	47	M	05/07/14
MA020	S			New	-ve	25	M	05/07/14
MA036	S			New	-ve	25	M	03/07/14
MA050	S			New	-ve	30	M	05/07/14
MA056	S			New	-ve	28	M	05/07/14
MA129	S			New	-ve	23	M	15/07/14
MA148	S			New	-ve	39	M	15/07/14
MA153	S			New	-ve	34	M	16/07/14
MA202	S			New	-ve	55	M	02/07/14
MA205	S			New	-ve	34	M	02/07/14
MA206	S			New	-ve	20	M	02/07/14
MA208	S			New	-ve	25	M	02/07/14
MA213	S			New	-ve	27	M	02/07/14
MA214	S			New	-ve	19	M	02/07/14
MA218	S			New	-ve	35	M	02/07/14
MA219	S			New	-ve	36	M	02/07/14
MA220	S			New	-ve	35	M	02/07/14
MA221	S			New	-ve	34	M	02/07/14
MA222	S			New	-ve	29	M	02/07/14
MA227	S			New	-ve	52	M	02/07/14
MA228	S			New	-ve	21	M	02/07/14
MA239	S			New	-ve	24	M	02/07/14
MA247	S			New	-ve	25	M	02/07/14
MA248	S			New	-ve	18	M	02/07/14
MA230	S			New	+ve	42	F	02/07/14
MA231	S			New	+ve	26	F	02/07/14
MA243	S			New	+ve	26	F	02/07/14
MA027	S			New	+ve	45	M	05/07/14
MA034	S			New	+ve	36	M	05/07/14
MA048	S			New	+ve	46	M	03/07/14
MA237	S			New	NA	20	F	02/07/14
MA229	S			New	NA	37	M	02/07/14
MA233	S			New	NA	32	M	02/07/14
MA234	S			New	NA	29	M	02/07/14
MA235	S			New	NA	58	M	02/07/14
MA236	S			New	NA	23	M	02/07/14
MA238	S			New	NA	25	M	02/07/14
MA058	S			Retreatment	-ve	34	F	05/07/14
MA067	S			Retreatment	-ve	38	M	05/07/14
MA069	S			Retreatment	-ve	78	M	05/07/14
MA070	S			Retreatment	-ve	31	M	03/07/14
MA087	S			Retreatment	-ve	20	M	05/07/14
MA096	S			Retreatment	-ve	60	M	03/07/14

Footnote: Columns were sorted by MDR status, then Study Case, then HIV status.

5.2.3 Workflow and bioinformatics analysis

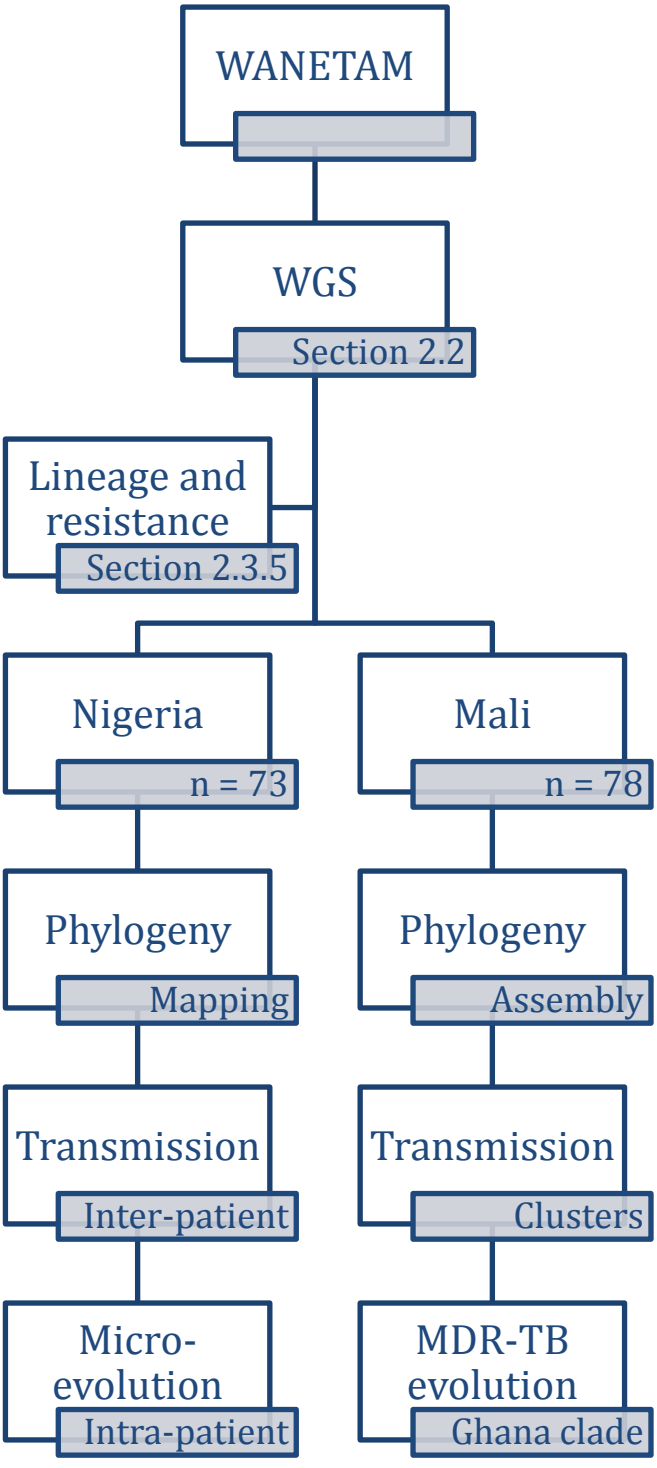


Figure 5.3 A schematic summary of the bioinformatics analysis for the two MTBC datasets from Nigeria and Mali

5.3 Results

5.3.1 Phylogenetic analysis of Nigerian isolates

Sequencing reads were mapped to the H37Rv reference genome with an average genome coverage of 50 X (range 23 to 78). Variants were called from the core genome after excluding known repeat regions (Section 2.4.5.1 and Appendix 5.1). The maximum likelihood was reconstructed from 8430 variable sites in the core genome using RAxML with a general time reversible model for substitution. The phylogenetic lineage of each isolate was inferred by uploading the sequencing reads to kvarq, PhyResSE and TBprofiler. KvarQ only assigned strains into the seven global lineages while TBProfiler and PhyResSE assigned strains to the sub-families and clades within the seven global lineages. There was generally very good agreement in the lineage assignments of PhyResSE and TBprofiler. I noticed only two discrepancies:

1. TBprofiler failed to assign strains to the Ghana genotype
2. PhyResSe failed to assign strains to the Uganda genotype

Consequently, the assignments of Ghana genotype are based on PhyResSE and the Uganda genotype assignments are based on TBprofiler.

Lineage 4 was the dominant lineage in the dataset (56/63 cases, 88%) (Table 5.6 and Appendix 5.2). Within lineage 4 the Cameroon clade was the most common genotype (37/63 cases, 57%). The Cameroon clade strains clustered on a recently diverged clade, with an average pairwise distance within the clade of 97 SNVs in the core-genome (range: 1-195) (Appendix 5.3). Short branch lengths within the Cameroon clade were suggestive of clonal expansion (Figure 5.4).

Within the Cameroon clade, four putative inter-patient transmission events were identified (3 MDR and 1 non-MDR). Inter-patient transmission was defined based on the presence of near-identical genomes (a difference of 3 SNVs or fewer) in two patients. Among the eight patients involved in putative

transmission events, six were retreatment cases, one had succumbed to treatment failure and only one was a new case.

The Haarlem, X-type, Ghana, TUR, LAM, Uganda and H37Rv-like genotypes of lineage 4 were collectively associated with 20 cases of TB (32%). There was no indication that these genotypes were involved in on going transmission. The closest indication of transmission was in two TUR genotype MDR strains from different patients that differed by 18 SNVs.

The *M. africanum* sub-lineage 1 (MAF1) was isolated from six patients: five were MDR and one was non-MDR. Two closely related MAF1 strains that differed by 11 SNVs were isolated from two retreatment cases from Lagos and Ibadan. Although this is not definitive evidence of direct transmission between the two patients, it suggests that MAF1 is involved in transmission of MDR in this setting. The high prevalence of MDR MAF1 is probably because we sampled patients at referral centres where severe cases are managed. One patient was infected with the animal lineage *M. bovis.*

Table 5.6 The prevalence of MTBC lineages and genotypes among MDR and non-MDR cases in Nigeria.

Lineage	Clade	Total	MDR	Non-MDR
M. bovis	Bovis	1		1
Lineage 4	Cameroon	37	19	18
	Uganda	6	2	4
	Ghana	2	2	
	X-type	1	1	
	H37Rv	1	1	
	Haarlem	3	2	1
	LAM	4	2	2
	TUR	2	2	
Lineage 5	West Africa 1	6	5	1
	Grand Total	**63**	**36**	**27**

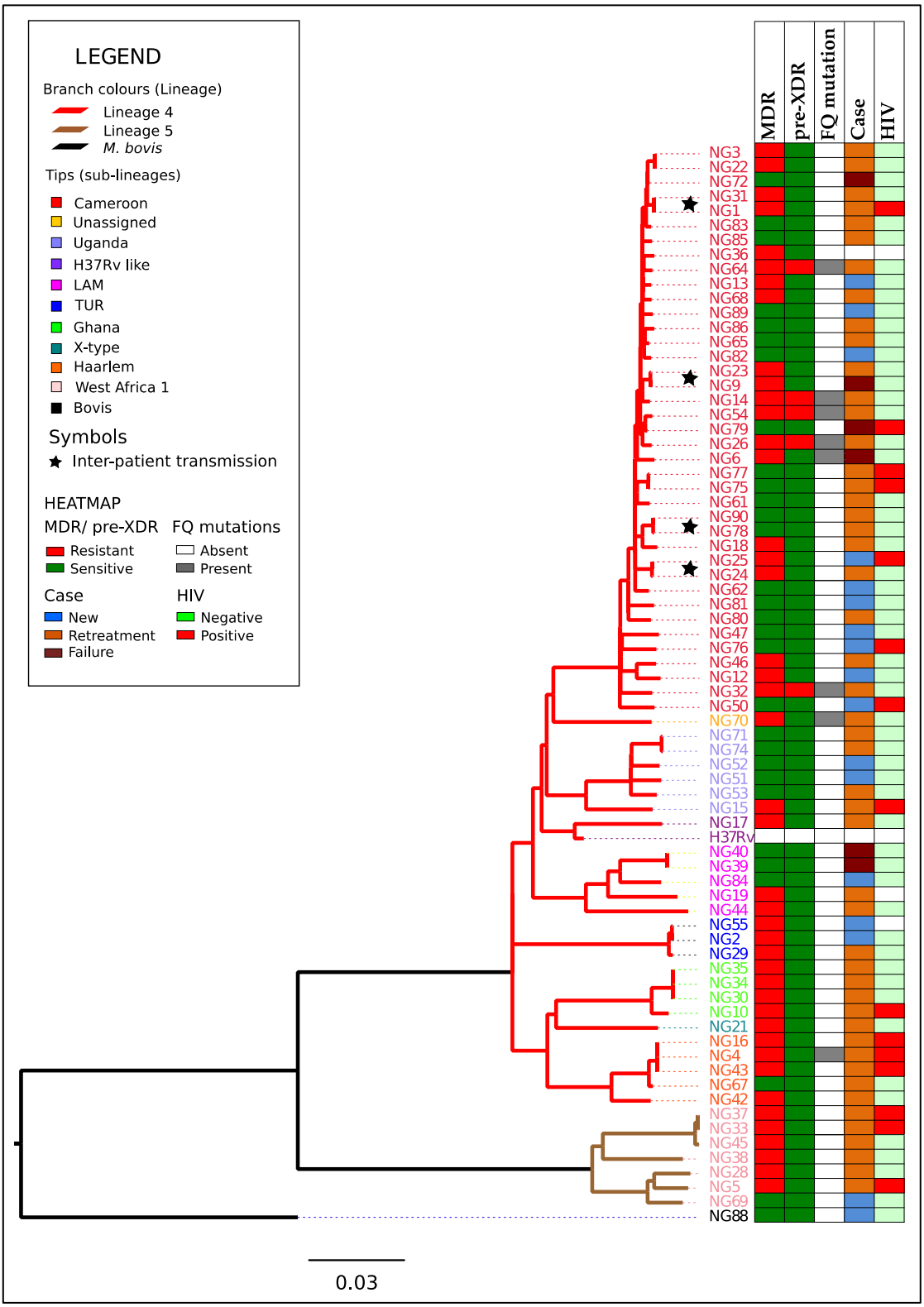


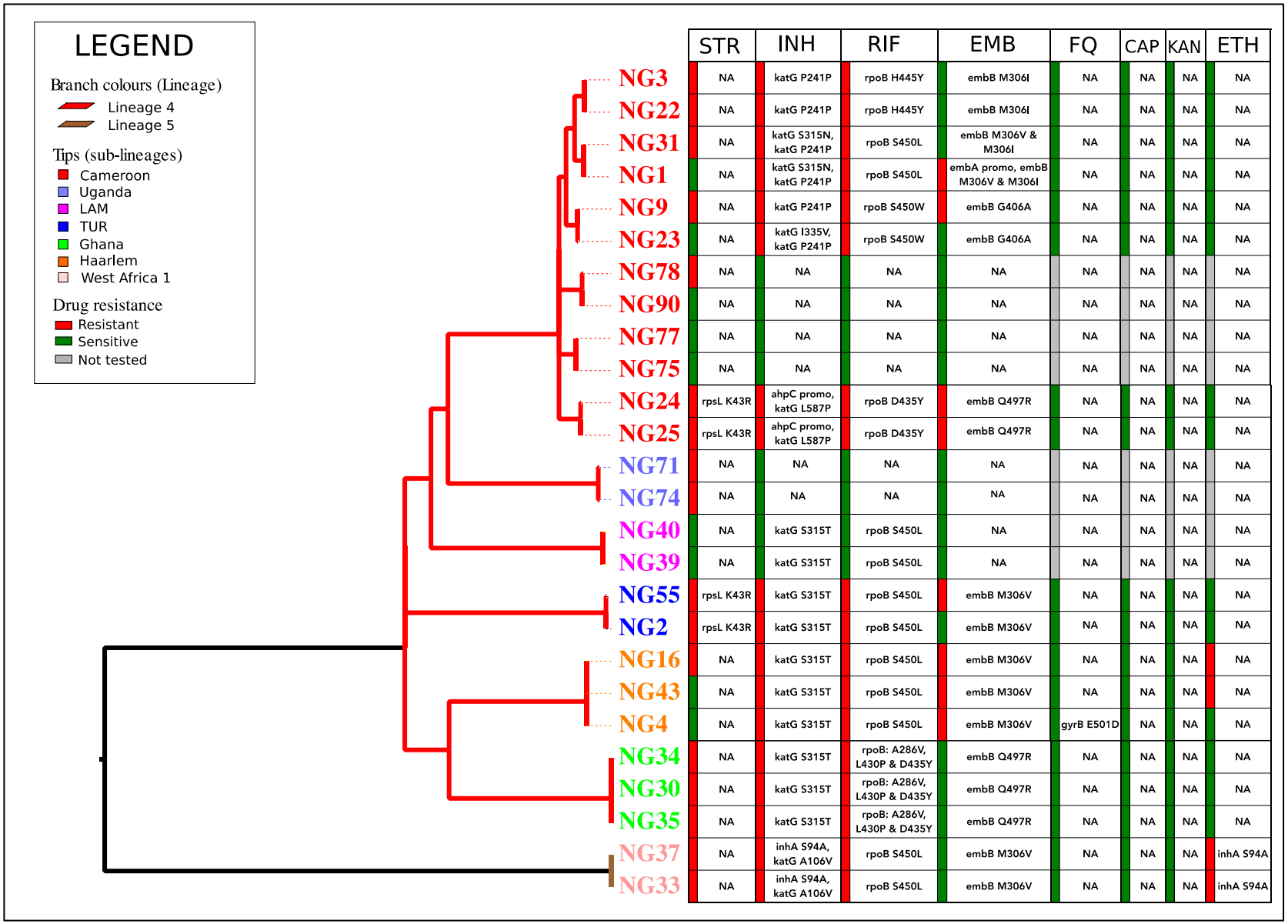
Figure 5.4 Phylogeny of 73 MTBC strains from Nigeria showing patient metadata and inter-patient transmission events. Branches are coloured based by lineage and tips are coloured based on the clade/family. Heatmap shows MDR/pre-XDR status, presence of fluoroquinolone resistance mutations (FQ), study case/treatment status and HIV status.

5.3.2 Microevolution of drug resistance in Nigeria

To study the microevolution of drug resistance we compared resistance profiles of identical/near-identical strains, which formed flat branches in the phylogenetic tree (Figure 5.5). These included genomes recovered from the same patient over the course of treatment as well as those isolated from different patients, which I have cited as putative transmission events. Six patients had isolates from two time points during treatment genome-sequenced and two patients had isolates from three time points genome-sequenced. Three instances were observed where microevolution influenced the susceptibility profiles of strains:

- NG9 and NG23; two patients enrolled in Lagos in 2011 and 2012. Both were MDR strains, but NG23 isolate acquired *katG* mutation L587P
- NG1 and NG31; two patients from Lagos. Both MDR, with canonical mutations in *embB*. NG1 acquired a mutation in the *embA* promoter, so became ethambutol resistant, while NG31 was ethambutol susceptible
- NG16, NG43, NG4 were from the same patient on treatment. All were MDR, but fluoroquinolone-susceptible on phenotypic testing. However the most recent isolate, NG4, acquired a low-level fluoroquinolone-resistance mutation in the *gyrB* gene

To confirm whether the samples were from the same patient or different patients the lab numbers of these strains were sent to our collaborators at the National Institute for Medical Research in Lagos for confirmation. Due to patient confidentiality the patient names are not presented as part of this thesis. NG39 and NG40 were isolated from the same patient at different time points. Both strains were susceptible to all first-line drugs by phenotypic testing. However, both strains bore *katG* and *rpoB* mutations that confer isoniazid and rifampicin respectively. It is unclear why this discrepancy was observed but a sample mix-up cannot be ruled out.



	STR	INH	RIF	EMB	FQ	CAP	KAN	ETH
NG3	NA	katG P241P	rpoB H445Y	embB M306I	NA	NA	NA	NA
NG22	NA	katG P241P	rpoB H445Y	embB M306I	NA	NA	NA	NA
NG31	NA	katG S315N, katG P241P	rpoB S450L	embB M306V & M306I	NA	NA	NA	NA
NG1	NA	katG S315N, katG P241P	rpoB S450L	embA promo, embB M306V & M306I	NA	NA	NA	NA
NG9	NA	katG P241P	rpoB S450W	embB G406A	NA	NA	NA	NA
NG23	NA	katG I335V, katG P241P	rpoB S450W	embB G406A	NA	NA	NA	NA
NG78	NA	NA	NA	NA	NA	NA	NA	NA
NG90	NA	NA	NA	NA	NA	NA	NA	NA
NG77	NA	NA	NA	NA	NA	NA	NA	NA
NG75	NA	NA	NA	NA	NA	NA	NA	NA
NG24	rpsL K43R	ahpC promo, katG L587P	rpoB D435Y	embB Q497R	NA	NA	NA	NA
NG25	rpsL K43R	ahpC promo, katG L587P	rpoB D435Y	embB Q497R	NA	NA	NA	NA
NG71	NA	NA	NA	NA	NA	NA	NA	NA
NG74	NA	NA	NA	NA	NA	NA	NA	NA
NG40	NA	katG S315T	rpoB S450L	NA	NA	NA	NA	NA
NG39	NA	katG S315T	rpoB S450L	NA	NA	NA	NA	NA
NG55	rpsL K43R	katG S315T	rpoB S450L	embB M306V	NA	NA	NA	NA
NG2	rpsL K43R	katG S315T	rpoB S450L	embB M306V	NA	NA	NA	NA
NG16	NA	katG S315T	rpoB S450L	embB M306V	NA	NA	NA	NA
NG43	NA	katG S315T	rpoB S450L	embB M306V	NA	NA	NA	NA
NG4	NA	katG S315T	rpoB S450L	embB M306V	gyrB E501D	NA	NA	NA
NG34	NA	katG S315T	rpoB: A286V, L430P & D435Y	embB Q497R	NA	NA	NA	NA
NG30	NA	katG S315T	rpoB: A286V, L430P & D435Y	embB Q497R	NA	NA	NA	NA
NG35	NA	katG S315T	rpoB: A286V, L430P & D435Y	embB Q497R	NA	NA	NA	NA
NG37	NA	inhA S94A, katG A106V	rpoB S450L	embB M306V	NA	NA	NA	inhA S94A
NG33	NA	inhA S94A, katG A106V	rpoB S450L	embB M306V	NA	NA	NA	inhA S94A

Figure 5.5 Microevolution of MTBC strains from Nigeria within patients during treatment and between patients during putative transmission. Phylogeny is shown in the context of antimicrobial susceptibility profiles.

5.3.3 Molecular basis of resistance in Nigeria

Phenotypic testing identified 39 cases as MDR, five of them pre-XDR (Figure 5.4). Almost all isoniazid-resistant strains carried a mutation in *katG* or *inhA*; mutations at codon 315 of *katG* were the commonest cause of isoniazid resistance, but mutations in the *aphC* and *inhA* promoter regions also contributed. All rifampicin-resistant isolates carried at least one resistance mutation in *rpoB*. In two cases, a mutation in *rpoB* was complemented by an apparent epistatic mutation in *rpoC* (Figure 5.6).

A clonally expanding sub-clade within our Cameroon clade was associated with a synonymous SNP at codon 241 of the katG gene (codon change CCC/CCG; katG P241P), which we have included in our analysis of potential resistance-associated mutations [468] (Figure 5.6). However, this mutation was present in both isoniazid-resistant and sensitive strains and, so, although epidemiologically informative, it is unlikely to be causally associated with resistance. However, it is worth noting that some MDR isolates lack any other mutations in katG, suggesting that there may be cryptic genotypic changes causing resistance.

All pre-XDR strains were resistant to fluoroquinolone due to mutations in *gyrA*. Three MDR strains were susceptible to fluoroquinolone by phenotypic testing, but bore mutations in *gyrA* or *gyrB*. One strain had the *ttyI* N236K and the *gyrA* D94H resistance mutations, suggesting XDR-TB. However, this strain was susceptible to both second-line agents on phenotypic testing.

In this dataset, accumulation of mutations in the *embCAB* operon led to phenotypic resistance to ethambutol. However, mutations at codons 306, 497 and 406 in *embB* were detected in both ethambutol-resistant and ethambutol-sensitive isolates. Curiously, only 12 out of 29 (41.4%) cases of streptomycin resistance were associated with a known resistance mutation. Phenotypic resistance profiles are presented alongside predicted genotypic resistance profiles for all strains in Appendix 5.4.

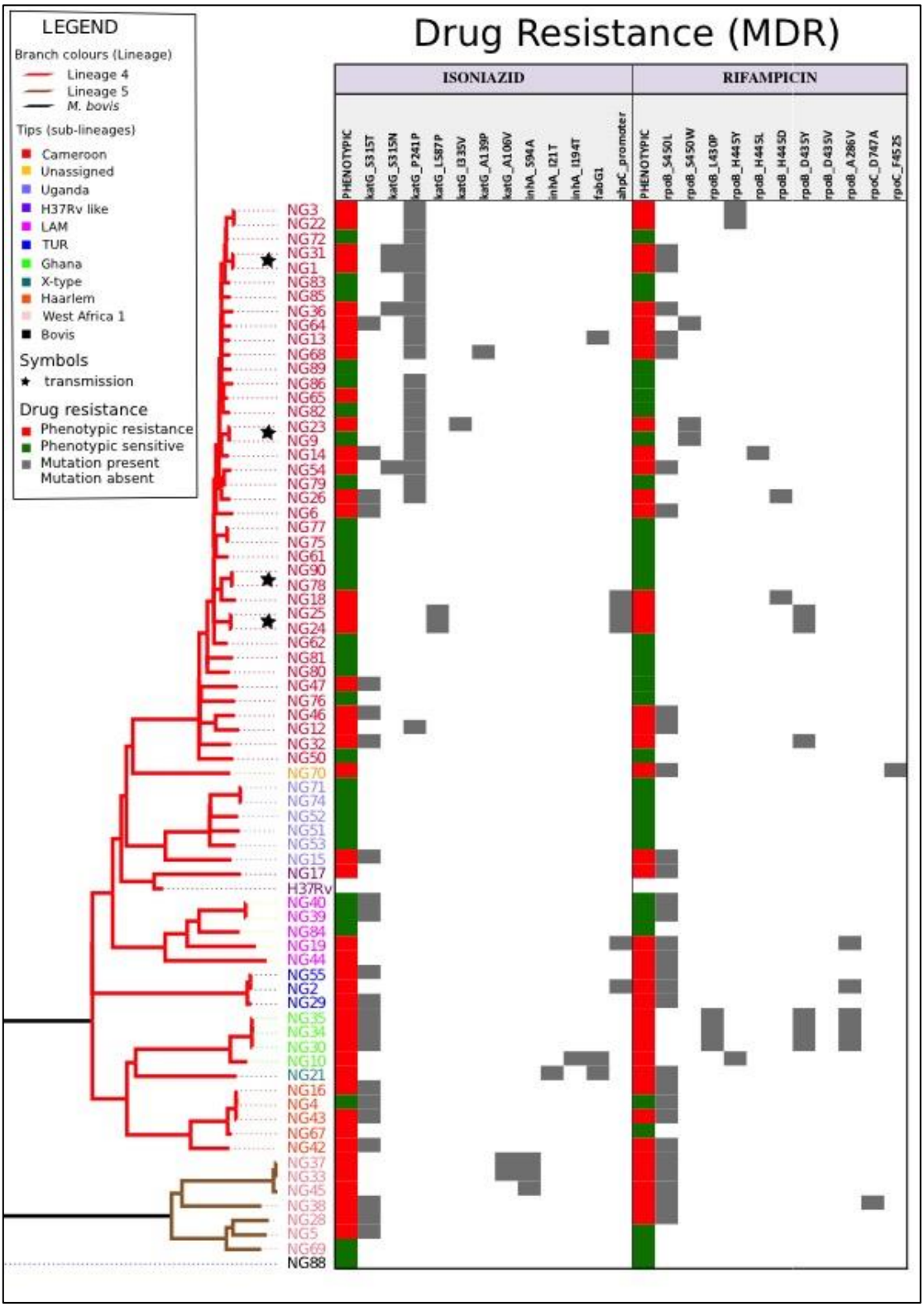


Figure 5.6 The molecular mechanisms of isoniazid and rifampicin resistance among Nigerian MTBC strains in the context of phylogeny. Phylogenetic tree linked to a heatmap showing the phenotypic resistance profiles and the presence/absence of known resistance mutations.

5.3.4 Phylogenetic analysis of Malian isolates

SPAdes assemblies from each strain were mapped against the H37Rv genome and placed in a multiple sequence alignment. Using custom Perl scripts, variants were called from the core genome after excluding repeat regions (Full a alignment Appendix 5.5). The maximum likelihood phylogeny was reconstructed from 8508 variant core genome sites. All the modern human-associated lineages of MTBC (2,3 & 4) were detected among patients from Mali (Table 5.7) (Lineage assignments Appendix 5.6).

The dominant lineage was the Euro-American super lineage, Lineage 4. Within lineage 4, the Cameroon lineage was the most prevalent genotype (25.6%). Strains within the Cameroon clade, including a MDR-TB strain, formed three distinct clusters. Notably, 44% of patients were infected with strains belonging to the related genotypes Ghana, Haarlem and X-type. 7/12 MDR-TB strains belonged to the Ghana genotype. The LAM genotype was recovered from 6 patients but none were MDR-TB.

Six patients were infected with MAF2 strains, but none were MDR. Five patients were infected with strains belonging to the Beijing clade. These strains formed a tight cluster with short branches. Four patients were infected with EAI strains but the long branches separating these strains suggest that the EAI strains were not epidemiologically linked.

Table 5.7 The prevalence of MTBC lineages and genotypes among MDR and non-MDR cases in Mali.

Lineage	Clade	Total	MDR	Non-MDR	Not tested
2	Beijing	5	1	3	1
3	EAI	4	1	3	
4	Cameroon	20	2	18	
	Ghana	16	7	9	
	H37Rv	2	1	1	
	Haarlem	10	0	10	
	LAM	6	0	6	
	Uganda	1	0	1	
	X-type	8	0	8	
6	MAF2	6	0	5	1
	Total	78	12	64	2

5.3.5 Transmission clusters in Mali

In the absence of patient epidemiological data it was not possible to reconstruct chains of transmission. However, it was possible to detect clusters of strains that were likely to be part of a transmission network based on genomic similarity. Transmission clusters were defined as five or more strains that clustered in a clade characterised by short branches and a highly conserved core genome (on average <25 SNVs difference between strains) (Appendix 5.7). The phylogenetic tree suggested that there were three transmission clusters within the dataset (Figure 5.7):

- A sub-clade of 13 Ghana clade strains clustered on a recently divergent branch. On average, strains within this cluster differed by 12 core-genome SNVs. The most divergent pair of strains within this cluster differed by 24 core genome SNVs. Within this transmission cluster there was an MDR sub-cluster and a non-MDR sub-cluster
- Eight non-MDR Cameroon clade strains formed a cluster on a recently divergent branch. The average pairwise distance for strains within this cluster was 7 core genome SNVs and the most divergent strains differed by 13 core genome SNVs. This cluster represented the most conserved transmission cluster in this dataset
- The five Beijing strains formed a monophyletic clade with a within-clade pairwise distance of 21 core-genome SNVs. The pairwise distance between strains within this cluster ranged from 11 to 30 core genome SNVs. The Beijing cluster was less conserved than the Ghana and Cameroon genotype clusters

Two sporadic cases of inter-patient transmission involving non-MDR strains were noted (Figure 5.7):

- Identical LAM genotype strains were isolated from two new cases (MA245 & MA246)
- Identical Cameroon clade strains were recovered from two new cases (MA242 & MA243). They differed from a third strain MA224 by six core genome SNVs

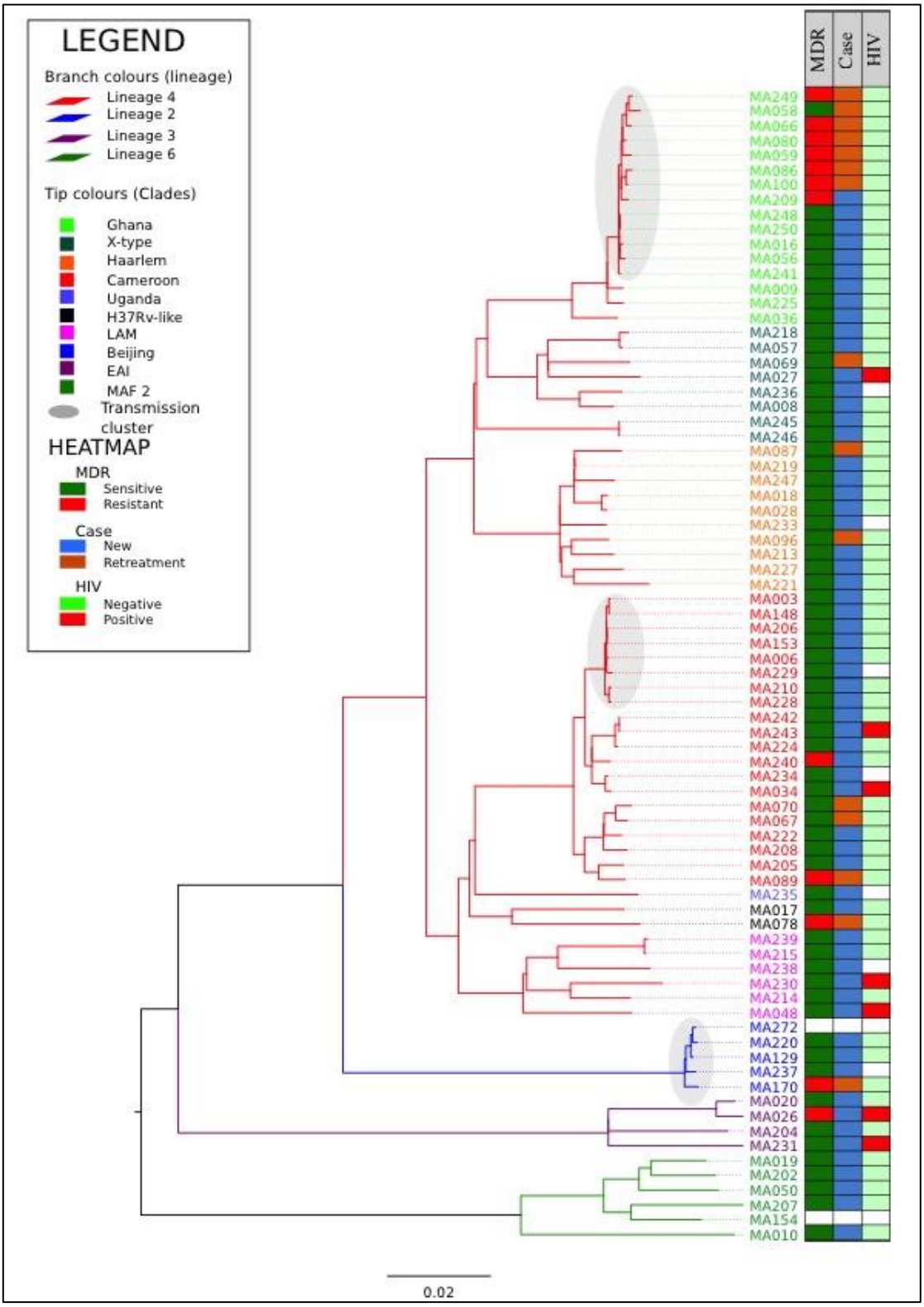


Figure 5.7 Phylogenetic tree for the Malian the MTBC dataset showing patient metadata and highlighting transmission clusters. Tree is linked to a heatmap showing MDR status, study case/treatment status and HIV status. Transmission clusters are highlighted in a grey oval. Branches are coloured by lineage and tips were coloured by clade.

5.3.6 Evolution of MDR in the Ghana clade

To understand the evolution of drug resistance among Ghana genotype strains the phylogeny of the Ghana genotype transmission cluster was visualized in the context of the antimicrobial susceptibility profiles (Figure 5.8). PhyResSE predicted resistance to anti-TB drugs based on the presence of known resistance mutations (Appendix 5.8). The phylogenetic topology of strains within the transmission cluster suggests that the MDR sub-cluster emerged from the non-MDR sub-cluster and then evolved through clonal expansion.

The ancestral clone of the MDR sub-cluster probably acquired the *katG* S315T isoniazid resistance mutation, since it was conserved in all strains within the MDR sub-cluster. Moreover, within the MDR sub-cluster MA058 was resistant to isoniazid but susceptible to rifampicin. Other strains in the MDR sub-cluster developed rifampicin and ethambutol resistance through acquisition of different resistance mutations. Within the non-MDR cluster only MA016 was mono-resistant to rifampicin, the other strains were susceptible to all first-line anti-TB drugs and MA250 & MA248 were identical.

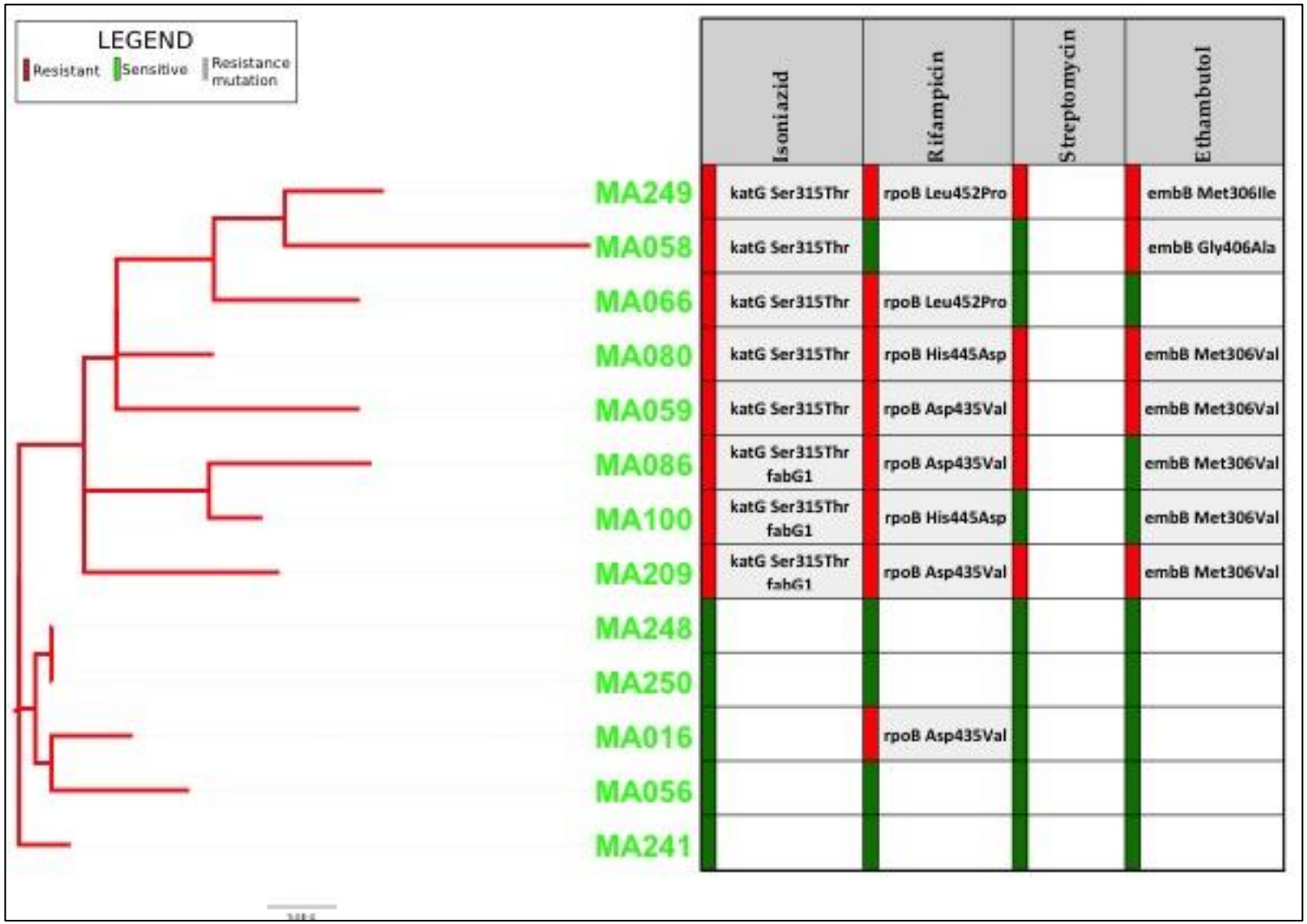


Figure 5.8 The evolution of MDR-TB among Ghana genotype strains in Mail. The phylogeny of the Ghana genotype transmission cluster is shown in the context of antibiotic resistance profiles.

5.3.7 MDR risk factor analysis

In the full Malian WANETAM dataset, 59% of the MDR strains were isolated from retreatment cases. This observation was significantly higher than the prevalence rate estimated by the WHO. In the dataset subjected to whole-genome sequencing, 9 out of 12 MDR strains were recovered from retreatment cases. Seven of the twelve MDR strains belonged to the Ghana genotype; all but one of these were retreatment cases.

To determine whether there was an association between the Ghana genotype and/or retreatment with MDR-TB in Mali, a regression analysis was performed using the Chi Squared test. The test evaluated whether the odds ratio of a strain being MDR if it belonged to the Ghana genotype was significant. The analysis was adjusted for age and study case.

The results show that Ghana genotypes strains were 9 times more likely to be MDR than other genotypes in this setting (Table 5.8). After adjusting for age and study case it was estimated that Ghana genotype strains were still 6 time more likely to be MDR. Retreatment cases were approximately 30 times more likely to be MDR compared to new cases. The number of retreatment cases was low (15/78), which probably explains why the confidence intervals for the retreatment odds ratio are so wide.

Table 5.8 Risk factor analysis for MDR-TB in Mali. Regression analysis showing the relationship between MDR and Ghana genotype strains or retreatment cases by the Chi squared test. P value <0.05 = significant.

	Factor	Observation	Odds ratio	Standard Error	P value	95% CI
Crude	Genotype	Other Ghana	1.00 8.87	6.09	0.001	2.30 - 34.06
	Study case	New Retreatment	1.00 30	23.77	0.00	6.35 - 141.76
Adjusted	Genotype	Other Ghana	1.00 5.77	5.03	0.044	1.05 - 31.86
	Study case	New Retreatment	1.00 22.35	18.70	0.00	4.34 - 115.25

5.4 Discussion

5.4.1 Diversity of TB: Nigeria versus Mali

The Euro-American super-lineage was dominant in all our study sites. In Bamako, two other "modern" lineages, Beijing and EAI, were also detected. The *M. africanum* lineages MAF1 and MAF2 were detected in Southwest Nigeria and Bamako respectively. The geographic restriction of these two lineages to West Africa has been documented [431]. The animal lineage *M. bovis* has been previously reported in both Mali and Nigeria [449,450]. We report one only case of *M bovis* infection that occurred in Ibadan. Prevalence of *M. bovis* is higher in regions where cattle are more widespread. All our study sites were in urban cities where contact with cattle is minimal.

Our data shows that the diversity of MTBC isolates in Bamako, towards the northwest of the West African region, is greater that the diversity in Ibadan and Lagos, towards the Gulf of Guinea. This is in line with the most recent analysis on the phylogeography of the MTBC in West Africa by Gehre *et. al.* [294]. Furthermore, in Southwest Nigeria evidence of on going transmission and clonal expansion was found only in the Cameroon clade. However, in Bamako there was evidence of on going spread not just of Cameroon-clade strains but also of the Ghana genotype and Beijing strains.

As its name suggests, the Cameroon clade is believed to have emerged in Cameroon and has now spread throughout much of West Africa [439]: it has been reported as the most dominant genotype in Nigeria, Ghana, Burkina Faso, Chad and Cameroon [436,438,440,469,470]. Nigeria is probably the largest reservoir of the Cameroon clade genotype [439]. However, our data suggests that the prevalence of the Cameroon clade in Bamako is increasing rapidly. A recent survey based on strains collected between 2006-2010 estimated the prevalence of Cameroon strains to be 10% [443]. This figure has more than doubled based on our dataset. The reasons for the selection and dissemination of the Cameroon clade in West Africa remain unclear, but perhaps will become clearer with additional WGS studies.

Although the Ghana genotype, also known as the T1 clade, has been replaced by the Cameroon clade as the most abundant genotype in Bamako [443], it remains comparable in prevalence. Our data confirms previous reports that the Ghana genotype is associated with MDR-TB [440,443]. Among our isolates, this association appears to have been driven by emergence of a sub-clade of the Ghana genotype, carrying the *katG* S315T isoniazid-resistance mutation. In Lagos, an MDR Ghana-genotype isolate acquired a fluoroquinolone-resistance mutation over the course of treatment. As more strains belonging to this genotype are genome-sequenced, the evolutionary mechanisms driving its association with drug resistance will be unravelled.

There is a paucity of data on the prevalence of Beijing strains in sub-Saharan Africa. Reviews published from the 2000s suggest that the Beijing strains were rarely found in sub-Saharan Africa [471]. In 2008, two cases of TB in Bamako were due to unrelated Beijing strains [472]. Our data shows that Beijing strains are on the rise in Bamako through clonal expansion of drug susceptible strains. Introduction of Beijing strains in Cape Town was followed by a rapid increase in prevalence [473,474]. Similarly, recent evidence shows that the Beijing strain was introduced in Afghanistan during the Soviet invasion in the 1980's and has continued to evolve and spread with human migration since [475]. It is imperative to closely monitor the rise in prevalence of Beijing strains in Bamako and other parts of West Africa, given this clade's propensity for increased virulence, high transmissibility and resistance.

Despite reports of a diminished presence in Cameroon [476] *M. africanum* is still an important cause of TB in West Africa [294,439,440,477]. Although we have not found any recent inter-patient transmission events involving MAF1 in Nigeria, our findings do suggest that multidrug resistant MAF1 isolates have been transmitted in this setting during the recent past. There was no evidence of direct transmission of MAF2 strains isolated from patients in Bamako and none of the strains were MDR. The high prevalence of MDR among MAF1 strains from Nigeria (5/6) is probably because we sampled patients from retreatment centres in a high MDR-TB burden setting.

5.4.2 Impact of MDR transmission on TB control

Globally the burden of MDR-TB is increasing, especially in low-resource settings [292]. This increase is driven largely by the transmission of MDR-TB especially in high-TB-burden regions [453]. Effective TB control strategies deliver prompt identification of cases and adequate treatment [478]. Correct administration of treatment renders the patient non-contagious and completion of treatment results in a high cure rate [478]. It is encouraging to see a decrease in the prevalence of MDR-TB in countries that have implemented effective TB control programs [479]

In Bamako, retreatment cases are referred to the central teaching hospital for antimicrobial susceptibility testing and adequate treatment. However, it is unclear whether the peripheral health facilities are equipped to perform antimicrobial susceptibility testing. Failing to diagnose MDR would lead to treatment failure on the standard regimen for drug-susceptible TB.

In Nigeria, culture facilities are also scarce, in part because of the stringent biosafety requirements. Funding and resources need to be directed to provide adequate diagnostic facilities in low-resource settings where the burden of B is highest. Measures need to be taken to stem the spread of MDR-TB in West Africa. Policy makers are encouraged to consider social intervention such as sensitisation campaigns, improved affordable housing and improved sanitation.

5.4.3 Usefulness of rapid molecular testing for drug resistance in high burden settings

Early diagnosis of MDR-TB can be instrumental in ensuring treatment success. The WHO recommends rapid testing, at least for rifampicin resistance, at the time of diagnosis [292]. The implementation of Xpert® MTB/RIF (Cepheid Inc.) for rapid detection of *Mycobacterium tuberculosis* complex (the MTBC) and rifampicin resistance has been intensified in Nigeria

to good effect [463,469]. However, Xpert® probably needs to be complemented with additional resistance testing [480].

The Hain MTBDR plus kit is a rapid tool for early detection of drug resistance in the MTBC using molecular probes. In Bamako, the Hain kit would be adequate for detecting the MDR Ghana genotype cluster since it has probes for detecting the *katG* S315T isoniazid-resistance mutations. In Nigeria, the Hain kit is unlikely to be sufficient to rule out isoniazid resistance due to the presence of rare isoniazid resistance mutation in the *katG* gene and in gene promoter region. In Nigeria, there appears to be an emerging pre-XDR TB problem driven by fluoroquinolone resistance. The Hain kit also offers early detection of fluoroquinolone.

5.4.4 Need for genomic surveillance of TB across West Africa

West African health authorities may wish to consider a move towards routine *in situ* use of bacterial WGS for characterisation of TB isolates. WGS offers early detection of resistance to both first- and second-line drugs, making it useful for diagnosing and guiding treatment of MDR-TB [457,481,482]. User-friendly web tools such as PhyResSe and TBProfiler are available for inferring the phylogenetic lineages and drug-resistance profiles of the MTBC isolates directly from WGS data [314,315]. WGS boasts superior discriminatory power in delineating human-to-human transmission and it differentiates strains that appear identical by conventional genotyping techniques (e.g. spoligo typing, MIRU VNTR and IS6110 typing) [455,483]. Moreover WGS offers insights into the evolutionary mechanisms that drive important process like transmission and drug resistance.

5.4.5 Limitations

The samples in the WANETAM dataset were collected from patients recruited at tertiary referral centres. This is likely to present a sampling bias since strains belonging to hypervirulent lineages that are prone to resistance are more likely to cause chronic cases of tuberculosis disease that warrant referral. Consequently these samples are likely not to be an accurate representation of the true genomic epidemiology of MTBC in Mali and Nigeria. Due to the lack of patient epidemiology and social network data it was not possible to conclusively infer the routes of inter-patient transmission.

6 CHAPTER SIX: CONCLUSIONS

6.1 The varieties of bacterial evolution

My work has involved three bacterial species that evolve in quite different ways. Through genomics, I have gained valuable insights into the epidemiology and antimicrobial resistance patterns of each pathogen. Some insights have a clear and present translational impact e.g. unravelling the cause of a meningitis outbreak and illuminating the mechanisms of antibiotic resistance in MDR-TB. Others, like the evolutionary history of *S. aureus* in monkeys in The Gambia, satisfy the curiosity, but have a less direct impact on public health.

The mode and tempo of evolution are known to vary between different bacterial species [190] and this applies to the three species studied in this thesis:

- ***Staphylococcus aureus*** readily undergoes horizontal gene transfer, resulting in changes in the accessory genome that contribute significantly to its evolution [337,338]. However, this data demonstrates the distinction between the core genome and accessory genome is somewhat arbitrary, as shown by the conservation of accessory genome components within clonal lineages. Evolution within the core genome of *S. aureus* is driven mainly by point mutations, although there is evidence for recombination between distantly related strains. Recombination events can lead to the emergence of epidemic clones of *S. aureus* e.g. a recombination event between ST8 and ST30 is believed to be the origin of the ST239 “hybrid lineage”, which evolved into a methicillin-resistant lineage [378,484]. However, the full functional implications of recombination in *S. aureus* remain unclear [485]. Point mutations in the core genome were key to reconstructing the phylogenetic tree and in inferring the approximate dates for host

species jumps. These inferences hinged on the assumption of a neutral evolutionary rate and a uniform mutation rate. This approach can be criticised for not considering the effect of variable mutation rates in different lineages. Nonetheless it provides a reasonable estimate for the timescales of the jumps from one host species to another.

- ***Streptococcus pneumoniae*** is also known for its promiscuity in terms of acquiring foreign DNA, but also experiences very high rates of homologous recombination that shape its evolutionary landscape [178]. Homologous recombination proved to be the main driving force behind the emergence of the novel clade that drove the meningitis outbreak described here—over two hundred mutations were introduced into this lineage through recombination. These recombination blocks overlapped with genes that are associated with fitness and virulence. In *S. pneumoniae,* antibiotic resistance is mediated through point mutations as well the acquisition of genes via horizontal gene transfer [406]. Although antibiotic resistance genes were detected in the novel clade, no accessory genes are conserved in this clade yet absent in the West African ST303 and ST217 strains.
- ***Mycobacterium tuberculosis*** is a genetically monomorphic lineage devoid of horizontal gene transfer and homologous recombination [203]. Instead, evolution within this species is driven by point mutations and large genomic deletions. Novel insights have emerged into the phylogeny of the *M. tuberculosis* complex in West Africa, highlighting inter-patient transmission and microevolution of strains within and between patients. Point mutations are the underlying mechanism of resistance to anti-TB drugs. A vast repository of known resistance mutations exists and resistance mutations are confirmed by convergent evolution in different lineages. I characterised antibiotic resistance in my datasets based on the presence/absence of known resistance mutations. This gave valuable insights into the mechanisms behind the emerging MDR-TB and pre-XDR-TB threat in West Africa [293].

6.2 Novel insights

Whole-genome sequencing has identified a novel clade of *S. aureus*, which colonizes Green Vervet monkeys in The Gambia. Analysis of gene repertoires points to gene loss as a major contributor to adaptation of the bacterium to a new host, the monkeys. Genomic epidemiology points towards to humans as the source of *S. aureus* in monkeys in The Gambia with recent anthroponotic transmission of human-associated genotypes to monkeys. The novel clade is believed to well before the advent of recent industrialisation and urbanisation.

Quantitative PCR (qPCR) was instrumental in revealing the aetiological agent of the meningitis outbreak in Ghana. However, WGS went beyond qPCR and standard genotyping to unveil a novel serotype 1/ST303 genotype lineage that showed features of an outbreak strain: high clonality, low sequence divergence and geographic clustering in the epicentre. Comparing the outbreak serotype strains to previously sequenced serotype 1 strains from West Africa confirmed the novelty of the novel clade. This novel clade has probably emerged because of rapid evolution through homologous recombination. Genomic evidence confirmed that strains in the novel clade have undergone capsular switching. A likely scenario is that the novel clade had acquired selective traits that enhanced its dissemination among the unvaccinated older population.

WGS has highlighted the gravity of the MDR-TB burden in Nigeria. MDR strains belonging to almost all the major genotypes of the Euro-America super lineage were detected in Nigeria. This widespread emergence of MDR-TB should be treated as a public health emergency and measures must be taken to prevent Nigeria from becoming a reservoir for MDR-TB. This situation could ominously deteriorate into an XDR-TB problem.

In Mali, the emerging MDR-TB problem appears to be driven by the Ghana genotype. Authorities should work towards enhancing diagnostic capacities across the Bamako area, so that patients are treated appropriately. Currently only retreatment cases are referred to the tertiary tuberculosis referral

hospital in Bamako. It is important that new cases are also referred for antimicrobial susceptibility testing to place patients on the right treatment regimen from the start. Mali also appears to face a rising threat from Beijing strains. This is a concern, given the highly virulent nature of this genotype and its predilection for drug resistance [486].

My results show that MDR transmission is an on going phenomenon in our setting. The confirmation of microevolution of MTBC during treatment should be a wake up call to health-care officials, showing the devastating consequences of sub-optimal treatment: worsening, now-resistant infection that poses an imminent threat to the public.

6.3 Advances in microbial genomics

It is worth pausing to consider the breathless pace of progress in this field that has taken place while the work described in this thesis was on going. Since enrolling for my PhD in October 2013, microbial genomics has become increasingly accessible. The prediction of anti-TB drug resistance from the genome has become increasingly easier. For example, at the time of my upgrade in 2014, I reviewed two major studies that identified drug resistance mutations in MTBC through custom pipelines that detected homoplastic mutations in multiple lineages [156,487]. I put on a brave face at the time, as the prospect of me reproducing a similar feat from The Gambia was farfetched. But today there are several publicly available user-friendly tools that can predict drug resistance from the tuberculosis genome, including kvarq, PhyResSE, TBProfiler and Mykrobe [304,313-315] and such approaches have now become part of routine clinical practice in Public Health England.

The nullarbor pipeline from Torsten Seeman exemplifies the ideal approach for clinical and research microbiologists with basic computational skills. It can characterise a number of bacterial species from the sequencing reads. This data includes speciation, genotyping, resistance predictions, phylogenetic analysis and pan genome analysis. This data can then be fed into downstream analysis pipelines.

Similar considerations apply to the analysis of pan-genomes. I recall building my first pan-genome in 2012, which included 50 genomes of *S. pneumoniae* serotype 1. This was a painstaking experience, as I manually curated comparisons of each strain with the growing pan genome to build the accessory genome. But now ROARY has made pan genome analysis very accessible. All you need to do is feed it annotated genomes and it has tools for visualising output and for comparing sets of sequences within your dataset.

Access to computational facilities has been facilitated by the advent of cloud computing, particularly the environment made available by the Cloud Infrastructure for Microbial Bioinformaticians (CLIMB) project [300]. With CLIMB, you can now fire up a genome virtual lab that is preloaded with the basic bioinformatics tools you need. This circumvents the onerous task of setting up a Linux or Ubuntu workstation from scratch and installing a myriad of programs and dependencies. All you need is a stable Internet connection to log in from anywhere in the world and perform your analysis.

6.4 Future prospects

Our aim at the MRC Unit in The Gambia is to develop our unit into a regional centre of excellence for genomics and bioinformatics. We recently established a bioinformatics core platform at the unit to serve the growing need for big data analysis. Our genomics focus group has successfully lobbied MRCG leadership to invest in an Illumina Miseq and a high performance computer server. By 2017, we aim to perform whole genome sequencing and bioinformatics analysis on site in The Gambia.

The insights gained from analysing 15 outbreak strains of *S. pneumoniae* have opened up new avenues of investigation. We are now collaborating with the WHO to carry out a carriage surveillance of *S. pneumoniae* in Ghana to investigate the effect of PCV-13 on the diversity of *S. pneumoniae.* I will be part of a team that analyses 3000 *S. pneumoniae* genomes from The Gambia and West Africa through the Gates pneumococcal sequencing

project. Through these collaborative efforts we hope to gain insights into the genomic epidemiology of *S. pneumoniae* to guide intervention strategies towards the elimination of this deadly pathogen as a public health problem.

The analysis reported here of MTBC in Mali and Nigeria has paved the way for a more ambitious programme of research. During this PhD, I have created a biobank: I have barcoded and stored 400 MTBC isolates from across West Africa together with associated genomic DNA samples. We aim to create a comprehensive database for these isolates that combines phenotypic drug susceptibility profiles with genomic data. This database will be a useful tool in testing novel diagnostic tools to assess how suited they are for use in West Africa. Moving forward, I am keen to devote effort towards understanding the evolutionary origins of genotypes and lineages that are endemic to West Africa. The Cameroon and Ghana genotypes are on the rise and urgent intervention is needed to curb their rise.

Colleagues at the MRCG continue to work on *S. aureus*, incorporating detailed studies on this dynamic pathogen into their PhD projects, using WGS to assess the impact of azithromycin treatment during labour on antibiotic resistance or to reconstruct perinatal transmission chains for *S. aureus*.

Finally, I hope that I have convinced the reader that West Africa is now poised to embrace microbial genomics as a key weapon in the arms race between man and microbe. We Africans stand ready to turn the tide on infection, harnessing the force of reason and the progress of science against our most fearsome—but far from invincible—microbial adversaries. The struggle continues! *A luta continua!*

APPENDIX

The appendix is presented on a DVD that is attached to this thesis due to the size of the files. If the DVD is not attached or you cannot access the appendix please contact me on msenghore@mrc.gm or madikaysenghore@gmail.com.

BIBLIOGRAPHY

1 Bardell, D (1983) ‘The roles of the sense of taste and clean teeth in the discovery of bacteria by Antoni van Leeuwenhoek’. *Microbiol Rev*, 47(1), pp. 121–126.

2 Porter, J R (1976) ‘Antony van Leeuwenhoek: tercentenary of his discovery of bacteria’. *Bacteriol Rev*, 40(2), pp. 260–269.

3 Baron, S (1996) ‘Introduction to Bacteriology’, in S, B. (ed.), *Medical Microbiology*.

4 Consortium, Microbiome (2012) ‘Structure, function and diversity of the healthy human microbiome’. *Nature*, 486(7402), pp. 207–214.

5 Carter, K C (1985) ‘Ignaz Semmelweis, Carl Mayrhofer, and the rise of germ theory’. *Med Hist*, 29(1), pp. 33–53.

6 Semmelweis, Ignaz Philipp (1981) ‘Childbed Fever’. *Clinical Infectious Diseases*, 3.

7 Stewart, G T (1968) ‘Limitations of the germ theory’. *The Lancet*, 291(7551), pp. 1077–1081.

8 Vouga, M and Greub, G (2016) ‘Emerging bacterial pathogens: the past and beyond’. *Clinical Microbiology and Infection*, 22(1), pp. 12–21.

9 Casadevall, A and Pirofski, L A (1999) ‘Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity’. *Infect Immun*, 67(8), pp. 3703–3713.

10 Dye, C (2014) ‘After 2015: infectious diseases in a new era of health and development’. *Philos Trans R Soc Lond B Biol Sci*, 369(1645).

11 Lozano, Rafael, Naghavi, Mohsen, Foreman, Kyle, Lim, Stephen, et al. (2012) ‘Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010’. *The Lancet*, 380(9859), pp. 2095–2128.

12 Anon (2016) ‘Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015’. *The Lancet*, 388(10053), pp. 1459–1544.

13 Walker, Christa L Fischer, Rudan, Igor, Liu, Li, Nair, Harish, et al. (2013) ‘Global burden of childhood pneumonia and diarrhoea’. *The*

Lancet, 381(9875), pp. 1405–1416.

14 Kotloff, Karen L, Nataro, James P, Blackwelder, William C, Nasrin, Dilruba, et al. (2013) 'Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.'. *Lancet*, 382(9888), pp. 209–222.

15 Kristiansen, Paul A, Diomandé, Fabien, Ba, Absatou Ky, Sanou, Idrissa, et al. (2012) 'Impact of the Serogroup A Meningococcal Conjugate Vaccine, MenAfriVac, on Carriage and Herd Immunity'. *Clinical Infectious Diseases*, 56(3), pp. cis892–363.

16 Allegranzi, Benedetta, Bagheri Nejad, Sepideh, Combescure, Christophe, Graafmans, Wilco, et al. (2011) 'Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis.'. *Lancet*, 377(9761), pp. 228–241.

17 Molton, James S, Tambyah, Paul A, Ang, Brenda S P, Ling, Moi Lin and Fisher, Dale A (2013) 'The global spread of healthcare-associated multidrug-resistant bacteria: a perspective from Asia.' Weinstein, R. A. (ed.). *Clin Infect Dis*, 56(9), pp. 1310–1318.

18 Anon (1993) 'WHO declares tuberculosis a global emergency.'. *Sozial- und Praventivmedizin*, 38(4), pp. 251–252.

19 Eurosurveillance editorial team (2013) 'WHO publishes Global tuberculosis report 2013.'. *Euro Surveill*, 18(43)24 October.

20 Huson, M A M, Stolp, S M, van der Poll, T and Grobusch, M P (2013) 'Community-Acquired Bacterial Bloodstream Infections in HIV-Infected Patients: A Systematic Review'. *Clinical Infectious Diseases*, 58(1), pp. 79–92.

21 World Health Organisation (2008) *World health statistics 2008*, World Health Organization, Geneva Switzerland.

22 Nelson, Michael (2000) 'Childhood nutrition and poverty'. *Proceedings of the Nutrition Society*, 59(2), pp. 307–315.

23 Rodríguez, Leonor, Cervantes, Elsa and Ortiz, Rocío (2011) 'Malnutrition and Gastrointestinal and Respiratory Infections in Children: A Public Health Problem'. *International Journal of Environmental Research and Public Health*, 8(4), pp. 1174–1205.

24 Schaible, Ulrich E and Kaufmann, Stefan H E (2007) 'Malnutrition and Infection: Complex Mechanisms and Global Impacts'. *PLoS medicine*, 4(5), p. e115.

25 Badmus, O and Camorlinga, S (2016) 'Poverty and the emergence of tuberculosis: An agent-based modelling approach'. *2016 IEEE-EMBS ….*

26 Cairncross, Sandy, Hunt, Caroline, Boisson, Sophie, Bostoen, Kristof, et al. (2010) 'Water, sanitation and hygiene for the prevention of diarrhoea.'. *International Journal of Epidemiology*, 39 Suppl 1(suppl 1), pp. i193–205.

27 Tanimura, Tadayuki, Jaramillo, Ernesto, Weil, Diana, Raviglione, Mario and Lönnroth, Knut (2014) 'Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review'. *European Respiratory Journal*, 43(6), pp. 1763–1775.

28 Bhutta, Zulfiqar A, Sommerfeld, Johannes, Lassi, Zohra S, Salam, Rehana A and Das, Jai K (2014) 'Global burden, distribution, and interventions for infectious diseases of poverty'. *Infectious Diseases of Poverty*, 3(1), p. 21.

29 Hotez, Peter J (2010) 'Neglected Infections of Poverty among the Indigenous Peoples of the Arctic'. *PLoS Negl Trop Dis*, 4(1), p. e606.

30 Fauci, Anthony S and Morens, David M (2012) 'The Perpetual Challenge of Infectious Diseases'. *The New England journal of medicine*, 366(5), pp. 454–461.

31 Rosenberg, Ronald and Ben Beard, C (2011) 'Vector-borne Infections'. *Emerg Infect Dis*, 17(5), pp. 769–770.

32 Centers for Disease Control and Prevention (2012) *Principles of epidemiology in public health practice*, … . Self-study course SS1000. 3e éd.

33 Green, Manfred S, Swartz, Tiberio, Mayshar, Elana, Lev, Boaz, et al. (2002) 'When is an epidemic an epidemic?'. *The Israel Medical Association journal : IMAJ*, 4(1), pp. 3–6.

34 Morens, David M, Folkers, Gregory K and Fauci, Anthony S (2004) 'The challenge of emerging and re-emerging infectious diseases'. *Nature*, 430(6996), pp. 242–249.

35 Chikeka, I and Dumler, J S (2015) 'Neglected bacterial zoonoses'. *Clinical Microbiology and Infection*, 21(5), pp. 404–415.

36 Park, Simon F (2002) 'The physiology of Campylobacter species and its relevance to their role as foodborne pathogens'. *Memorial Issue for Gordon Stewart*, 74(3), pp. 177–188.

37 Abdulamir, Ahmed S, Hafidh, Rand R and AbuBakar, Fatimah (2011) 'The association of Streptococcus bovis/gallolyticus with colorectal tumors: The nature and the underlying mechanisms of its etiological role'. *Journal of Experimental & Clinical Cancer Research : CR*, 30(1), pp. 11–24.

38 Butler, T (2015) 'Capnocytophaga canimorsus: an emerging cause of sepsis, meningitis, and post-splenectomy infection after dog bites'. *Eur*

J Clin Microbiol Infect Dis, 34(7), pp. 1271–1280.

39 Ferens, Witold A and Hovde, Carolyn J (2011) 'Escherichia coli O157:H7: Animal Reservoir and Sources of Human Infection'. *Foodborne Pathog Dis*, 8(4), pp. 465–487.

40 Fehr, Jan S, Bloemberg, Guido V, Ritter, Claudia, Hombach, Michael, et al. (2010) 'Septicemia Caused by Tick-borne Bacterial Pathogen Candidatus Neoehrlichia mikurensis'. *Emerging Infectious Disease journal*, 16(7), p. 1127.

41 Stuen, Snorre, Granquist, Erik G and Silaghi, Cornelia (2013) 'Anaplasma phagocytophilum—a widespread multi-host pathogen with highly adaptive strategies'. *Frontiers in Cellular and Infection Microbiology*, 3, p. 31.

42 Socolovschi, Cristina, Mediannikov, Oleg, Raoult, Didier and Parola, Philippe (2009) 'The relationship between spotted fever group Rickettsiae and Ixodid ticks'. *Veterinary Research*, 40(2), p. 34.

43 LoGiudice, Kathleen, Ostfeld, Richard S, Schmidt, Kenneth A and Keesing, Felicia (2003) 'The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk.'. *Proceedings of the National Academy of Sciences*, 100(2), pp. 567–571.

44 Houpikian, Pierre and Raoult, Didier (2002) 'Traditional and Molecular Techniques for the Study of Emerging Bacterial Diseases: One Laboratory's Perspective'. *Emerg Infect Dis*, 8(2), pp. 122–131.

45 Warren, J Robin and Marshall, Barry (1983) 'Unidentified curved baccili on gastric epithelium in active chronic gastritis'. *The Lancet*, 321(8336), pp. 1273–1275.

46 Skirrow, M B (1977) 'Campylobacter enteritis: a "new" disease.'. *British Medical Journal*, 2(6078), pp. 9–11.

47 Berger, B J, Hussain, F and Roistacher, K (1994) 'Bacterial infections in HIV-infected patients.'. *Infectious disease clinics of North America*, 8(2), pp. 449–465.

48 Cattoir, Vincent (2012) 'Actinobaculum schaalii: Review of an emerging uropathogen'. *Journal of Infection*, 64(3), pp. 260–267.

49 Plotkin, Stanley (2014) 'History of vaccination.'. *Proc Natl Acad Sci U S A*, 111(34), pp. 12283–12287.

50 Plotkin, Stanley A and Plotkin, Susan L (2011) 'The development of vaccines: how the past led to the future'. *Nat Rev Microbiol*, 9(12), pp. 889–893.

51 DiMaio, Daniel (2016) 'Thank You, Edward. Merci, Louis.' Spindler, K.

R. (ed.). *PLoS Pathog*, 12(1), p. e1005320.

52 Calmette, A, Guérin, C, Boquet, A and Nègre, L (1927) *La vaccination préventive contre la tuberculose par le‘ BCG,’*,

53 Ivarsson, Mattias E, Leroux, Jean Christophe and Castagner, Bastien (2012) ‘Targeting Bacterial Toxins’. *Angewandte Chemie International Edition*, 51(17), pp. 4024–4045.

54 Kelly, Dominic F, Moxon, E Richard and Pollard, Andrew J (2004) ‘Haemophilus influenzae type b conjugate vaccines’. *Immunology*, 113(2), pp. 163–174.

55 ESKOLA, JUHANI and ANTTILA, MERJA (1999) ‘Pneumococcal conjugate vaccines’. *Pediatr Infect Dis J*, 18 N2 -(6).

56 English, M (2000) ‘A randomised, double-blind, controlled trial of the immunogenicity and tolerability of a meningococcal group C conjugate vaccine in young British infants’. *Vaccine*, 19(9-10), pp. 1232–1238.

57 Sibley, C D, Peirano, G and Church, D L (2012) ‘Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology’. *Infect Genet Evol*, 12(3), pp. 505–521.

58 Stewart, E J (2012) ‘Growing unculturable bacteria’. *J Bacteriol*, 194(16), pp. 4151–4160.

59 Foxman, B and Riley, L (2001) ‘Molecular epidemiology: focus on infection’. *Am J Epidemiol*, 153(12), pp. 1135–1141.

60 Wayne, L G, Brenner, D J, Colwell, R R, Grimont, P A D, et al. (1987) ‘Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics’. *Int J Syst Evol Microbiol*, 37(4), pp. 463–464.

61 Chan, J Z, Halachev, M R, Loman, N J, Constantinidou, C and Pallen, M J (2012) ‘Defining bacterial species in the genomic era: insights from the genus Acinetobacter’. *BMC Microbiol*, 12, p. 302.

62 Riley, M A and Lizotte-Waniewski, M (2009) ‘Population genomics and the bacterial species concept’. *Methods Mol Biol*, 532, pp. 367–377.

63 Rossello-Mora, R and Amann, R (2001) ‘The species concept for prokaryotes’. *FEMS Microbiol Rev*, 25(1), pp. 39–67.

64 Mothershed, E A and Whitney, A M (2006) ‘Nucleic acid-based methods for the detection of bacterial pathogens: present and future considerations for the clinical laboratory’. *Clin Chim Acta*, 363(1-2), pp. 206–220.

65 Josephson, K L, Gerba, C P and Pepper, I L (1993) ‘Polymerase chain reaction detection of nonviable bacterial pathogens.’. *Appl Environ*

Microbiol, 59(10), pp. 3513–3515.

66 Chang, S S, Hsieh, W H, Liu, T S, Lee, S H, et al. (2013) 'Multiplex PCR system for rapid detection of pathogens in patients with presumed sepsis - a systemic review and meta-analysis'. *PLoS One*, 8(5), p. e62323.

67 Mengelle, C, Mansuy, J M, Prere, M F, Grouteau, E, et al. (2013) 'Simultaneous detection of gastrointestinal pathogens with a multiplex Luminex-based molecular assay in stool samples from diarrhoeic patients'. *Clin Microbiol Infect*, 19(10), pp. E458–65.

68 Didelot, X, Bowden, R, Wilson, D J, Peto, T E and Crook, D W (2012) 'Transforming clinical microbiology with bacterial genome sequencing'. *Nat Rev Genet*, 13(9), pp. 601–612.

69 Konstantinidis, K T and Tiedje, J M (2005) 'Genomic insights that advance the species definition for prokaryotes'. *Proc Natl Acad Sci U S A*, 102(7), pp. 2567–2572.

70 Rossello-Mora, R (2005) 'Updating prokaryotic taxonomy'. *J Bacteriol*, 187(18), pp. 6255–6257.

71 Pallen, M J (2014) 'Diagnostic metagenomics: potential applications to bacterial, viral and parasitic infections'. *Parasitology*, 141(14), pp. 1856–1862.

72 Nakamura, S, Maeda, N, Miron, I M, Yoh, M, et al. (2008) 'Metagenomic diagnosis of bacterial infections'. *Emerg Infect Dis*, 14(11), pp. 1784–1786.

73 Loman, N J, Constantinidou, C, Christner, M, Rohde, H, et al. (2013) 'A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4'. *JAMA*, 309(14), pp. 1502–1510.

74 Doughty, E L, Sergeant, M J, Adetifa, I, Antonio, M and Pallen, M J (2014) 'Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer'. *PeerJ*, 2, p. e585.

75 Clardy, Jon, Fischbach, Michael A and Currie, Cameron R (2009) 'The natural history of antibiotics'. *Current biology*, 19(11), pp. R437–R441.

76 Kohanski, Michael A, Dwyer, Daniel J and Collins, James J (2010) 'How antibiotics kill bacteria: from targets to networks'. *Nat Rev Microbiol*, 8(6), pp. 423–435.

77 Aminov, Rustam I (2010) 'A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future'. *Frontiers in microbiology*, 1.

78 Blair, Jessica M A, Webber, Mark A, Baylay, Alison J, Ogbolu, David O and Piddock, Laura J V (2015) 'Molecular mechanisms of antibiotic resistance'. *Nat Rev Microbiol*, 13(1), pp. 42–51.

79 Kojima, Seiji and Nikaido, Hiroshi (2013) 'Permeation rates of penicillins indicate that Escherichia coli porins function principally as nonspecific channels.'. *Proc Natl Acad Sci U S A*, 110(28), pp. E2629–34.

80 Qin, Shangshang, Wang, Yang, Zhang, Qijing, Chen, Xia, et al. (2012) 'Identification of a novel genomic island conferring resistance to multiple aminoglycoside antibiotics in Campylobacter coli.'. *Antimicrob Agents Chemother*, 56(10), pp. 5332–5339.

81 Bartlett, John G, Gilbert, David N and Spellberg, Brad (2013) 'Seven ways to preserve the miracle of antibiotics.'. *Clin Infect Dis*, 56(10), pp. 1445–1450.

82 Ventola, C Lee (2015) 'The Antibiotic Resistance Crisis: Part 1: Causes and Threats'. *Pharmacy and Therapeutics*, 40(4), pp. 277–283.

83 Hoffman, Steven J, Outterson, Kevin, Røttingen, John-Arne, Cars, Otto, et al. (2015) 'An international legal framework to address antimicrobial resistance'. *Bulletin of the World Health Organization*, 93(2), pp. 66–66.

84 Sass, Peter and Brötz-Oesterhelt, Heike (2013) 'Bacterial cell division as a target for new antibiotics'. *Curr Opin Microbiol*, 16(5), pp. 522–530.

85 Golkar, Zhabiz, Bagasra, Omar and Pace, Donald Gene (2014) 'Bacteriophage therapy: a potential solution for the antibiotic resistance crisis'. *The Journal of Infection in Developing Countries*, 8(02), pp. 129–136.

86 Wittebole, Xavier, De Roock, Sophie and Opal, Steven M (2013) 'A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens'. *Virulence*, 5(1), pp. 226–235.

87 Piddock, Laura JV (2012) 'The crisis of no new antibiotics—what is the way forward?'. *Lancet Infect Dis*, 12(3), pp. 249–253.

88 Achtman, M (2008) 'Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens'. *Annu Rev Microbiol*, 62, pp. 53–70.

89 Sabat, A J, Budimir, A, Nashev, D, Sa-Leao, R, et al. (2013) 'Overview of molecular typing methods for outbreak detection and epidemiological surveillance'. *Euro Surveill*, 18(4), p. 20380.

90 Habib, Maha, Porter, Barbara D and Satzke, Catherine (2014)

'Capsular Serotyping of Streptococcus pneumoniae Using the Quellung Reaction'. *J Vis Exp*, (84), pp. e51208–e51208.

91 Raymond, Frédéric, Boucher, Nancy, Allary, Robin, Robitaille, Lynda, et al. (2013) 'Serotyping of Streptococcus pneumoniae Based on Capsular Genes Polymorphisms' de Lencastre, H. (ed.). *PLoS One*, 8(9), p. e76197.

92 McQuiston, John R, Waters, R Jordan, Dinsmore, Blake A, Mikoleit, Matthew L and Fields, Patricia I (2011) 'Molecular determination of H antigens of Salmonella by use of a microsphere-based liquid array.'. *J Clin Microbiol*, 49(2), pp. 565–573.

93 Herrera-León, Silvia, McQuiston, John R, Usera, Miguel A, Fields, Patricia I, et al. (2004) 'Multiplex PCR for distinguishing the most common phase-1 flagellar antigens of Salmonella spp.'. *J Clin Microbiol*, 42(6), pp. 2581–2586.

94 Achtman, M (1996) 'A surfeit of YATMs?'. *J Clin Microbiol*, 34(7), p. 1870.

95 Maiden, M C (2006) 'Multilocus sequence typing of bacteria'. *Annu Rev Microbiol*, 60, pp. 561–588.

96 Francisco, A P, Bugalho, M, Ramirez, M and Carrico, J A (2009) 'Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach'. *BMC Bioinformatics*, 10, p. 152.

97 Feil, E J, Cooper, J E, Grundmann, H, Robinson, D A, et al. (2003) 'How clonal is Staphylococcus aureus?'. *J Bacteriol*, 185(11), pp. 3307–3316.

98 Oliveira, Duarte C, Tomasz, Alexander and de Lencastre, Herminia (2002) 'Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant Staphylococcus aureus.'. *Lancet Infect Dis*, 2(3), pp. 180–189.

99 Adetifa, Ifedayo M O, Antonio, Martin, Okoromah, Christy A N, Ebruke, Chinelo, et al. (2012) 'Pre-Vaccination Nasopharyngeal Pneumococcal Carriage in a Nigerian Population: Epidemiology and Population Biology' Wright, L. (ed.). *PLoS One*, 7(1), p. e30548.

100 Ibarz-Pavón, Ana Belén, Morais, Luis, Sigaúque, Betuel, Mandomando, Inacio, et al. (2011) 'Epidemiology, Molecular Characterization and Antibiotic Resistance of Neisseria meningitidis from Patients ≤15 Years in Manhiça, Rural Mozambique' Gagneux, S. (ed.). *PLoS One*, 6(6), p. e19717.

101 Shepheard, M A, Fleming, V M, Connor, T R, Corander, J, et al. (2013) 'Historical zoonoses and other changes in host tropism of Staphylococcus aureus, identified by phylogenetic analysis of a population dataset'. *PLoS One*, 8(5), p. e62369.

102 Sheppard, Samuel K, Colles, Frances, Richardson, Judith, Cody, Alison J, et al. (2010) 'Host association of Campylobacter genotypes transcends geographic variation.'. *Appl Environ Microbiol*, 76(15), pp. 5269–5277.

103 Pérez-Losada, Marcos, Cabezas, Patricia, Castro-Nallar, Eduardo and Crandall, Keith A (2013) 'Pathogen typing in the genomics era: MLST and the future of molecular epidemiology'. *Infection, Genetics and Evolution*, 16, pp. 38–53.

104 Larsen, Mette V, Cosentino, Salvatore, Rasmussen, Simon, Friis, Carsten, et al. (2012) 'Multilocus sequence typing of total-genome-sequenced bacteria.'. *J Clin Microbiol*, 50(4), pp. 1355–1361.

105 Jagielski, T, van Ingen, J, Rastogi, N, Dziadek, J, et al. (2014) 'Current methods in the molecular typing of Mycobacterium tuberculosis and other mycobacteria'. *Biomed Res Int*, 2014, p. 645802.

106 Halachev, Mihail R, Chan, Jacqueline Z-M, Constantinidou, Chrystala I, Cumley, Nicola, et al. (2014) 'Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England'. *Genome medicine*, 6(11), p. 70.

107 Tang, Patrick and Gardy, Jennifer L (2014) 'Stopping outbreaks with real-time genomic epidemiology'. *Genome medicine*, 6(11), p. 104.

108 Gardy, Jennifer, Loman, Nicholas J and Rambaut, Andrew (2015) 'Real-time digital pathogen surveillance — the time is now'. *Genome biology*, 16(1), p. 155.

109 Pallen, Mark J (2016) 'Microbial bioinformatics 2020.'. *Microb Biotechnol*, 9(5), pp. 681–686.

110 Luheshi, Leila M, Raza, Sobia and Peacock, Sharon J (2015) 'Moving pathogen genomics out of the lab and into the clinic: what will it take?'. *Genome medicine*, 7(1), p. 132.

111 Köser, Claudio U, Ellington, Matthew J, Cartwright, Edward J P, Gillespie, Stephen H, et al. (2012) 'Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology' Rall, G. F. (ed.). *PLoS Pathog*, 8(8), p. e1002824.

112 Fournier, Pierre-Edouard, Drancourt, Michel and Raoult, Didier (2007) 'Bacterial genome sequencing and its use in infectious diseases'. *Lancet Infect Dis*, 7(11), pp. 711–723.

113 Croucher, Nicholas J, Harris, Simon R, Grad, Yonatan H and Hanage, William P (2013) 'Bacterial genomes in epidemiology—present and future'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614).

114 Parkhill, Julian and Wren, Brendan W (2011) 'Bacterial epidemiology

and biology - lessons from genome sequencing'. *Genome biology*, 12(10), p. 230.

115 Morelli, Giovanna, Song, Yajun, Mazzoni, Camila J, Eppinger, Mark, et al. (2010) 'Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity'. *Nat Genet*, 42(12), pp. 1140–1143.

116 Croucher, Nicholas J, Harris, Simon R, Fraser, Christophe, Quail, Michael A, et al. (2011) 'Rapid pneumococcal evolution in response to clinical interventions.'. *Science*, 331(6016), pp. 430–434.

117 Coll, F, McNerney, R, Guerra-Assuncao, J A, Glynn, J R, et al. (2014) 'A robust SNP barcode for typing Mycobacterium tuberculosis complex strains'. *Nat Commun*, 5, p. 4812.

118 Aanensen, D M, Feil, E J, Holden, M T, Dordel, J, et al. (2016) 'Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive Staphylococcus aureus in Europe'. *MBio*, 7(3).

119 Jolley, K A, Bliss, C M, Bennett, J S, Bratcher, H B, et al. (2012) 'Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain'. *Microbiology*, 158(Pt 4), pp. 1005–1015.

120 Sanger, F (1988) 'Sequences, sequences, and sequences'. *Annual review of biochemistry*.

121 Sanger, F (1981) 'Determination of nucleotide sequences in DNA'. *Bioscience reports*.

122 LM, Smith, JZ, Sanders, RJ, Kaiser, P, Hughes, et al. (1986) 'Fluorescence detection in automated DNA sequence analysis.'. *Nature*, 321(6071), pp. 674–679.

123 Luckey, John A, Drossman, Howard, Kostichka, Anthony J, Mead, David A, et al. (1990) 'High speed DNA sequencing by capillary electrophoresis'. *Nucleic Acids Res*, 18(15), pp. 4417–4421.

124 Baker, M (2012) 'De novo genome assembly: what every biologist should know'. *Nature Methods*, 9, pp. 333–337.

125 Fleischmann, R D, Adams, M D, White, O, Clayton, R A, et al. (1995) 'Whole-genome random sequencing and assembly of Haemophilus influenzae Rd'. *Science*, 269(5223), pp. 496–512.

126 Fraser, C M, Gocayne, J D, White, O, Adams, M D, et al. (1995) 'The minimal gene complement of Mycoplasma genitalium'. *Science*, 270(5235), pp. 397–403.

127 Land, M, Hauser, L, Jun, S R, Nookaew, I, et al. (2015) 'Insights from 20 years of bacterial genome sequencing'. *Funct Integr Genomics*,

15(2), pp. 141–161.

128 Loman, N J and Pallen, M J (2015) ‘Twenty years of bacterial genome sequencing’. *Nat Rev Microbiol*.

129 Cole, S T, Brosch, R, Parkhill, J, Garnier, T, et al. (1998) ‘Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence’. *Nature*, 393(6685), pp. 537–544.

130 Parkhill, J, Wren, B W, Thomson, N R, Titball, R W, et al. (2001) ‘Genome sequence of Yersinia pestis, the causative agent of plague’. *Nature*, 413(6855), pp. 523–527.

131 Blattner, Frederick R, Plunkett, Guy, Bloch, Craig A, Perna, Nicole T, et al. (1997) ‘The Complete Genome Sequence of Escherichia coli K-12’. *Science*, 277(5331), pp. 1453–1462.

132 Metzker, Michael L (2010) ‘Sequencing technologies |[mdash]| the next generation’. *Nat Rev Genet*, 11(1), pp. 31–46.

133 Bashir, Ali, Klammer, Aaron A, Robins, William P, Chin, Chen-Shan, et al. (2012) ‘A hybrid approach for the automated finishing of bacterial genomes’. *Nature Biotechnology*, 30(7), pp. 701–707.

134 Chin, Chen-Shan, Alexander, David H, Marks, Patrick, Klammer, Aaron A, et al. (2013) ‘Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data’. *Nature Methods*, 10(6), pp. 563–569.

135 Stoesser, N, Sheppard, A E, Shakya, M, Sthapit, B, et al. (2015) ‘Dynamics of MDR Enterobacter cloacae outbreaks in a neonatal unit in Nepal: insights using wider sampling frames and next-generation sequencing.’. *Journal of Antimicrobial Chemotherapy*, 70(4), pp. 1008–1015.

136 Quick, Joshua, Ashton, Philip, Calus, Szymon, Chatt, Carole, et al. (2015) ‘Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella’. *Genome biology*, 16(1), p. 114.

137 Quick, Joshua, Loman, Nicholas J, Duraffour, Sophie, Simpson, Jared T, et al. (2016) ‘Real-time, portable genome sequencing for Ebola surveillance’. *Nature*, 530(7589), pp. 228–232.

138 White, O, Eisen, J A, Heidelberg, J F, Hickey, E K, et al. (1999) ‘Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1’. *Science*, 286(5444), pp. 1571–1577.

139 Mendell, J E, Clements, K D, Choat, J H and Angert, E R (2008) ‘Extreme polyploidy in a large bacterium’. 105(18), pp. 6730–6734.

140 McCutcheon, J P and Dohlen, von, C D (2011) ‘An interdependent metabolic patchwork in the nested symbiosis of mealybugs’. *Curr Biol*,

21(16), pp. 1366–1372.

141 Bennett, G M and Moran, N A (2013) ‘Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect’. *Genome Biol Evol*, 5(9), pp. 1675–1688.

142 Loman, Nicholas J, Constantinidou, Chrystala, Chan, Jacqueline Z-M, Halachev, Mihail, et al. (2012) ‘High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity’. *Nat Rev Microbiol*, 10(9), pp. 599–606.

143 Koren, Sergey, Schatz, Michael C, Walenz, Brian P, Martin, Jeffrey, et al. (2012) ‘Hybrid error correction and de novo assembly of single-molecule sequencing reads’. *Nature Biotechnology*, 30(7), pp. 693–700.

144 Page, Andrew J, De Silva, Nishadi, Hunt, Martin, Quail, Michael A, et al. (2016) ‘Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data’. *Microbial Genomics*, 2(8).

145 Silva, Genivaldo GZ, Dutilh, Bas E, Matthews, T David, Elkins, Keri, et al. (2013) ‘Combining de novo and reference-guided assembly with scaffold_builder’. *Source Code for Biology and Medicine*, 8(1), p. 23.

146 Mira, A, Ochman, H and Moran, N A (2001) ‘Deletional bias and the evolution of bacterial genomes’. *Trends Genet*, 17(10), pp. 589–596.

147 Richardson, Emily J and Watson, Mick (2013) ‘The automatic annotation of bacterial genomes.’. *Brief Bioinform*, 14(1), pp. 1–12.

148 Delcher, A L, Harmon, D, Kasif, S, White, O and Salzberg, S L (1999) ‘Improved microbial gene identification with GLIMMER.’. *Nucleic Acids Res*, 27(23), pp. 4636–4641.

149 Ziemann, Mark, Eren, Yotam and El-Osta, Assam (2016) ‘Gene name errors are widespread in the scientific literature’. *Genome biology*, 17(1), p. 177.

150 Suez, Jotham, Porwollik, Steffen, Dagan, Amir, Marzel, Alex, et al. (2013) ‘Virulence Gene Profiling and Pathogenicity Characterization of Non-Typhoidal Salmonella Accounted for Invasive Disease in Humans’ Chabalgoity, J. A. (ed.). *PLoS One*, 8(3), p. e58449.

151 Iwamoto, Akira, Tanahashi, Toshihito, Okada, Rina, Ogawa, Hirofumi, et al. (2015) ‘696 Whole-Genome Sequencing Detects Novel Virulence Variants of Helicobacter pylori cagA Gene and Its Relationship With Oncogenic cagA’. *Gastroenterology*, 148(4), pp. S–137–147.e9.

152 Chen, Peter E and Shapiro, B Jesse (2015) ‘The advent of genome-wide association studies for bacteria’. *Curr Opin Microbiol*, 25, pp. 17–24.

153 Sheppard, Samuel K, Didelot, Xavier, Meric, Guillaume, Torralbo, Alicia, et al. (2013) ‘Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter.’. *Proc Natl Acad Sci U S A*, 110(29), pp. 11923–11927.

154 McCarthy, Noel (2013) ‘An epidemiological view of microbial genomic data.’. *Lancet Infect Dis*, 13(2), pp. 104–105.

155 McCarthy, Noel D, Colles, Frances M, Dingle, Kate E, Bagnall, Mary C, et al. (2007) ‘Population genetic approaches to assigning the source of human pathogens: host associated genetic import in Campylobacter jejuni.’. *Emerg Infect Dis*, 13(2), pp. 267–272.

156 Farhat, Maha R, Shapiro, B Jesse, Kieser, Karen J, Sultana, Razvan, et al. (2013) ‘Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis’. *Nat Genet*, 45(10), pp. 1183–1189.

157 Croucher, Nicholas J and Thomson, Nicholas R (2010) ‘Studying bacterial transcriptomes using RNA-seq’. *Curr Opin Microbiol*, 13(5), pp. 619–624.

158 Repoila, Francis and Darfeuille, Fabien (2009) ‘Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects’. *Biology of the Cell*, 101(2), pp. 117–131.

159 Medini, Duccio, Serruto, Davide, Parkhill, Julian, Relman, David A, et al. (2008) ‘Microbiology in the post-genomic era’. *Nat Rev Microbiol*, 6(6), pp. 419–430.

160 Behr, Jürgen, Geissler, Andreas J, Schmid, Jonas, Zehe, Anja and Vogel, Rudi F (2016) ‘The Identification of Novel Diagnostic Marker Genes for the Detection of Beer Spoiling Pediococcus damnosus Strains Using the BIAst Diagnostic Gene findEr’ Green, S. J. (ed.). *PLoS One*, 11(3), p. e0152747.

161 Dougherty, T J, Barrett, J F and Pucci, M J (2002) ‘Microbial Genomics and Novel Antibiotic Discovery: New Technology to Search for New Drugs’. *Current Pharmaceutical Design*, 8(13), pp. 1119–1135.

162 Moir, Donald T, Shaw, Karen J, Hare, Roberta S and Vovis, Gerald F (1999) ‘Genomics and Antimicrobial Drug Discovery’. *Antimicrob Agents Chemother*, 43(3), pp. 439–446.

163 Mobegi, Fredrick M, van Hijum, Sacha AFT, Burghout, Peter, Bootsma, Hester J, et al. (2014) ‘From microbial gene essentiality to novel antimicrobial drug targets’. *BMC Genomics*, 15(1), p. 958.

164 Molzen, T E, Burghout, P, Bootsma, H J, Brandt, C T, et al. (2011) ‘Genome-wide identification of Streptococcus pneumoniae genes essential for bacterial replication during experimental meningitis.’.

Infect Immun, 79(1), pp. 288–297.

165 Sassetti, C M, Boyd, D H and Rubin, E J (2001) 'Comprehensive identification of conditionally essential genes in mycobacteria.'. *Proceedings of the National Academy of Sciences*, 98(22), pp. 12712–12717.

166 Sette, Alessandro and Rappuoli, Rino (2010) 'Reverse Vaccinology: Developing Vaccines in the Era of Genomics'. *Immunity*, 33(4), pp. 530–541.

167 Heinson, Ashley I, Woelk, Christopher H and Newell, Marie-Louise (2015) 'The promise of reverse vaccinology.'. *International health*, 7(2), pp. 85–89.

168 Rinaudo, C Daniela, Telford, John L, Rappuoli, Rino and Seib, Kate L (2009) 'Vaccinology in the genome era'. *J Clin Invest*, 119(9), pp. 2515–2525.

169 Maione, D (2005) 'Identification of a Universal Group B Streptococcus Vaccine by Multiple Genome Screen'. *Science*, 309(5731), pp. 148–150.

170 Binnewies, T T, Motro, Y, Hallin, P F, Lund, O, et al. (2006) 'Ten years of bacterial genome sequencing: comparative-genomics-based discoveries'. *Funct Integr Genomics*, 6(3), pp. 165–185.

171 Bryant, J, Chewapreecha, C and Bentley, S D (2012) 'Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences'. *Future Microbiol*, 7(11), pp. 1283–1296.

172 Periwal, Vinita and Scaria, Vinod (2014) 'Insights into structural variations and genome re-arrangements in prokaryotic genomes'. *Bioinformatics*, 31(1), pp. btu600–9.

173 Sousa, Carolina, de Lorenzo, Victor and Cebolla, Angel (1997) 'Modulation of gene expression through chromosomal positioning in Escherichia coli'. *Microbiology*, 143(6), pp. 2071–2078.

174 Couturier, EtienneRocha, Eduardo P. C. (2006) 'Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes'. *Mol Microbiol*, 59(5), pp. 1506–1518.

175 Thomas, C M and Nielsen, K M (2005) 'Mechanisms of, and barriers to, horizontal gene transfer between bacteria'. *Nat Rev Microbiol*, 3(9), pp. 711–721.

176 Block, Dena H S, Hussein, Razika, Liang, Lusha W and Lim, Han N (2012) 'Regulatory consequences of gene translocation in bacteria.'. *Nucleic Acids Res*, 40(18), pp. 8979–8992.

177 Dobrindt, U, Zdziarski, J, Salvador, E and Hacker, J (2010) ‘Bacterial genome plasticity and its impact on adaptation during persistent infection’. *Int J Med Microbiol*, 300(6), pp. 363–366.

178 Chaguza, C, Cornick, J E and Everett, D B (2015) ‘Mechanisms and impact of genetic recombination in the evolution of Streptococcus pneumoniae’. *Comput Struct Biotechnol J*, 13, pp. 241–247.

179 Frost, Laura S, Leplae, Raphael, Summers, Anne O and Toussaint, Ariane (2005) ‘Mobile genetic elements: the agents of open source evolution’. *Nat Rev Microbiol*, 3(9), pp. 722–732.

180 Jackson, R W, Vinatzer, B, Arnold, D L, Dorus, S and Murillo, J (2011) *The influence of the accessory genome on bacterial pathogen evolution. Mob. Genet. Elements 1: 55–65*,

181 Kumwenda, Benjamin, Litthauer, Derek and Reva, Oleg (2014) ‘Analysis of genomic rearrangements, horizontal gene transfer and role of plasmids in the evolution of industrial important Thermus species’. *BMC Genomics*, 15(1), p. 813.

182 Paauw, Armand, Hall, Maurine A Leverstein-van, Verhoef, Jan and Fluit, Ad C (2010) ‘Evolution in Quantum Leaps: Multiple Combinatorial Transfers of HPI and Other Genetic Modules in Enterobacteriaceae’ Dobrindt, U. (ed.). *PLoS One*, 5(1), p. e8662.

183 Zupan, John, Ward, Doyle and Zambryski, Patricia (2002) ‘Inter-kingdom DNA transfer decoded’. *Nature Biotechnology*, 20(2), pp. 129–131.

184 Penadés, José R, Chen, John, Quiles-Puchalt, Nuria, Carpena, Nuria and Novick, Richard P (2015) ‘Bacteriophage-mediated spread of bacterial virulence genes’. *Curr Opin Microbiol*, 23, pp. 171–178.

185 Wagner, Patrick L and Waldor, Matthew K (2002) ‘Bacteriophage control of bacterial virulence.’. *Infect Immun*, 70(8), pp. 3985–3993.

186 Sumby, Paul and Waldor, Matthew K (2003) ‘Transcription of the toxin genes present within the Staphylococcal phage phiSa3ms is intimately linked with the phage's life cycle.’. *J Bacteriol*, 185(23), pp. 6841–6851.

187 Selva, Laura, Viana, David, Regev-Yochay, Gili, Trzcinski, Krzysztof, et al. (2009) ‘Killing niche competitors by remote-control bacteriophage induction.’. *Proc Natl Acad Sci U S A*, 106(4), pp. 1234–1238.

188 Schmidt, H and Hensel, M (2004) ‘Pathogenicity islands in bacterial pathogenesis’. *Clin Microbiol Rev*, 17(1), pp. 14–56.

189 Moran, N A (2002) ‘Microbial minimalism: genome reduction in bacterial pathogens’. *Cell*, 108(5), pp. 583–586.

190 Pallen, M J and Wren, B W (2007) ‘Bacterial pathogenomics’. *Nature*,

449(7164), pp. 835–842.

191 Cole, S T, Eiglmeier, K, Parkhill, J, James, K D, et al. (2001) ‘Massive gene decay in the leprosy bacillus’. *Nature*, 409(6823), pp. 1007–1011.

192 Parkhill, J, Dougan, G, James, K D, Thomson, N R, et al. (2001) ‘Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18’. *Nature*, 413(6858), pp. 848–852.

193 Chain, P S, Carniel, E, Larimer, F W, Lamerdin, J, et al. (2004) ‘Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis’. *Proc Natl Acad Sci U S A*, 101(38), pp. 13826–13831.

194 Fookes, Maria, Schroeder, Gunnar N, Langridge, Gemma C, Blondel, Carlos J, et al. (2011) ‘ Salmonella bongori Provides Insights into the Evolution of the Salmonellae’ Ochman, H. (ed.). *PLoS Pathog*, 7(8), p. e1002191.

195 Maurelli, A T (2007) ‘Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens’. *FEMS Microbiol Lett*, 267(1), pp. 1–8.

196 Moran, Nancy A (2002) ‘Microbial Minimalism’. *Cell*, 108(5), pp. 583–586.

197 Ochman, H (2005) ‘Genomes on the shrink’. *Proc Natl Acad Sci U S A*, 102(34), pp. 11959–11960.

198 Mira, A, Martin-Cuadrado, A B, D'Auria, G and Rodriguez-Valera, F (2010) ‘The bacterial pan-genome:a new paradigm in microbiology’. *Int Microbiol*, 13(2), pp. 45–57.

199 van Tonder, A J, Mistry, S, Bray, J E, Hill, D M, et al. (2014) ‘Defining the estimated core genome of bacterial populations using a Bayesian decision model’. *PLoS computational biology*, 10(8), p. e1003788.

200 Croll, Daniel and McDonald, Bruce A (2012) ‘The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens’ Heitman, J. (ed.). *PLoS Pathog*, 8(4), p. e1002608.

201 Rouli, L, Merhej, V, Fournier, P E and Raoult, D (2015) ‘The bacterial pangenome as a new tool for analysing pathogenic bacteria’. *New Microbes and New Infections*, 7, pp. 72–85.

202 Doolittle, W F and Papke, R T (2006) ‘Genomics and the bacterial species problem’. *Genome biology*, 7(9), p. 116.

203 Achtman, Mark (2012) ‘Insights from genomic comparisons of genetically monomorphic bacterial pathogens’. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1590), pp. 860–867.

204 Achtman, M, Zurth, K, Morelli, G, Torrea, G, et al. (1999) ‘Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis.’. *Proceedings of the National Academy of Sciences*, 96(24), pp. 14043–14048.

205 Comas, Iñaki, Chakravartti, Jaidip, Small, Peter M, Galagan, James, et al. (2010) ‘Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved’. *Nat Genet*, 42(6), pp. 498–503.

206 Didelot, Xavier, Barker, Margaret, Falush, Daniel and Priest, Fergus G (2009) ‘Evolution of pathogenicity in the Bacillus cereus group’. *Systematic and Applied Microbiology*, 32(2), pp. 81–90.

207 Leopold, Shana R, Magrini, Vincent, Holt, Nicholas J, Shaikh, Nurmohammad, et al. (2009) ‘A precise reconstruction of the emergence and constrained radiations of Escherichia coli O157 portrayed by backbone concatenomic analysis.’. *Proc Natl Acad Sci U S A*, 106(21), pp. 8713–8718.

208 Roumagnac, Philippe, Weill, François-Xavier, Dolecek, Christiane, Baker, Stephen, et al. (2006) ‘Evolutionary History of Salmonella Typhi’. *Science*, 314(5803), pp. 1301–1304.

209 Chun, Jongsik, Grim, Christopher J, Hasan, Nur A, Lee, Je Hee, et al. (2009) ‘Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae.’. *Proc Natl Acad Sci U S A*, 106(36), pp. 15442–15447.

210 Torres-Morquecho, Araceli, Giono-Cerezo, Silvia, Camorlinga-Ponce, Margarita, Vargas-Mendoza, Carlos F and Torres, Javier (2010) ‘Evolution of bacterial genes: Evidences of positive Darwinian selection and fixation of base substitutions in virulence genes of Helicobacter pylori’. *Infection, Genetics and Evolution*, 10(6), pp. 764–776.

211 Barrick, Jeffrey E, Yu, Dong Su, Yoon, Sung Ho, Jeong, Haeyoung, et al. (2009) ‘Genome evolution and adaptation in a long-term experiment with Escherichia coli’. *Nature*, 461(7268), pp. 1243–1247.

212 Zhou, Zhemin, McCann, Angela, Weill, François-Xavier, Blin, Camille, et al. (2014) ‘Transient Darwinian selection in Salmonella enterica serovar Paratyphi A during 450 years of global spread of enteric fever.’. *Proc Natl Acad Sci U S A*, 111(33), pp. 12199–12204.

213 Holt, Kathryn E, Parkhill, Julian, Mazzoni, Camila J, Roumagnac, Philippe, et al. (2008) ‘High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi’. *Nat Genet*, 40(8), pp. 987–993.

214 Gutierrez, M Cristina, Brisse, Sylvain, Brosch, Roland, Fabre, Michel, et al. (2005) ‘Ancient Origin and Gene Mosaicism of the Progenitor of Mycobacterium tuberculosis’. *PLoS Pathog*, 1(1), p. e5.

215 Didelot, X, Achtman, M and Parkhill, J (2007) ‘A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination?’. 17(1), pp. 61–68.

216 Hershberg, Ruth, Lipatov, Mikhail, Small, Peter M, Sheffer, Hadar, et al. (2008) ‘High Functional Diversity in Mycobacterium tuberculosis Driven by Genetic Drift and Human Demography’ Blaser, M. J. (ed.). *PLoS Biol*, 6(12), p. e311.

217 Zuckerkandl, E and Pauling L B (1962) ‘Molecular disease, evolution, and genetic heterogeneity’. *Horizons in Biochemistry, Aca-demic Press, New York*, (189), p. 36.

218 Bromham, L and Penny, D (2003) ‘The modern molecular clock’. *Nat Rev Genet*, 4(3), pp. 216–224.

219 Drummond, A J, Ho, S Y, Phillips, M J and Rambaut, A (2006) ‘Relaxed phylogenetics and dating with confidence’. *PLoS Biol*, 4(5), p. e88.

220 Benton, M JDonoghue, P. C. (2007) ‘Paleontological evidence to date the tree of life’. *Mol Biol Evol*, 24(1), pp. 26–53.

221 Mongkolrattanothai, Kanokporn, Gray, Barry M, Mankin, Peggy, Stanfill, Amy B, et al. (2011) ‘Simultaneous carriage of multiple genotypes of Staphylococcus aureus in children’. *Journal of Medical Microbiology*, 60(3), pp. 317–322.

222 Votintseva, A A, Miller, R R, Fung, R, Knox, K, et al. (2014) ‘Multiple-strain colonization in nasal carriers of Staphylococcus aureus.’. *J Clin Microbiol*, 52(4), pp. 1192–1200.

223 Wilson, Daniel J (2012) ‘Insights from Genomics into Bacterial Pathogen Populations’ Chitnis, C. E. (ed.). *PLoS Pathog*, 8(9), p. e1002874.

224 Didelot, Xavier, Eyre, David W, Cule, Madeleine, Ip, Camilla LC, et al. (2012) ‘Microevolutionary analysis of Clostridium difficile genomes to investigate transmission’. *Genome biology*, 13(12), p. R118.

225 Wilson, Daniel J, Gabriel, Edith, Leatherbarrow, Andrew J H, Cheesbrough, John, et al. (2009) ‘Rapid evolution and the importance of recombination to the gastroenteric pathogen Campylobacter jejuni.’. *Mol Biol Evol*, 26(2), pp. 385–397.

226 Morelli, Giovanna, Didelot, Xavier, Kusecek, Barica, Schwarz, Sandra, et al. (2010) ‘Microevolution of Helicobacter pylori during Prolonged Infection of Single Hosts and within Families’ Malik, H. S. (ed.). *PLoS Genet*, 6(7), p. e1001036.

227 Paterson, Gavin K, Harrison, Ewan M, Murray, Gemma G R, Welch,

John J, et al. (2015) ‘Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission’. *Nat Commun*, 6, p. 6560.

228 Worby, Colin J, Lipsitch, Marc and Hanage, William P (2014) ‘Within-Host Bacterial Diversity Hinders Accurate Reconstruction of Transmission Networks from Genomic Distance Data’ Koelle, K. (ed.). *PLoS computational biology*, 10(3), p. e1003549.

229 Didelot, Xavier, Walker, A Sarah, Peto, Tim E, Crook, Derrick W and Wilson, Daniel J (2016) ‘Within-host evolution of bacterial pathogens’. *Nat Rev Microbiol*, 14(3), pp. 150–162.

230 Wright, Gerard D (2010) ‘Antibiotic resistance in the environment: a link to the clinic?’. *Curr Opin Microbiol*, 13(5), pp. 589–594.

231 Mwangi, Michael M, Wu, Shang Wei, Zhou, Yanjiao, Sieradzki, Krzysztof, et al. (2007) ‘Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing.’. *Proceedings of the National Academy of Sciences*, 104(22), pp. 9451–9456.

232 Eldholm, Vegard, Norheim, Gunnstein, Lippe, von der, Bent, Kinander, Wibeke, et al. (2014) ‘Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient’. *Genome biology*, 15(11), p. 490.

233 Koch, Anastasia and Wilkinson, Robert John (2014) ‘The road to drug resistance in Mycobacterium tuberculosis’. *Genome biology*, 15(11), p. 520.

234 Davies, J and Davies, D (2010) ‘Origins and evolution of antibiotic resistance’. *Microbiology and Molecular Biology Reviews*.

235 Marshall, Bonnie M and Levy, Stuart B (2011) ‘Food animals and antimicrobials: impacts on human health.’. *Clin Microbiol Rev*, 24(4), pp. 718–733.

236 Golubchik, T, Brueggemann, A B, Street, T and Gertz, R E, Jr (2012) ‘Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event’. *Nature*.

237 Croucher, Nicholas J, Finkelstein, Jonathan A, Pelton, Stephen I, Mitchell, Patrick K, et al. (2013) ‘Population genomics of post-vaccine changes in pneumococcal epidemiology’. *Nat Genet*, 45(6), pp. 656–663.

238 Robbins, John B, Schneerson, Rachel, Trollfors, Birger, Sato, Hiroko, et al. (2005) ‘The diphtheria and pertussis components of diphtheria-tetanus toxoids-pertussis vaccine should be genetically inactivated mutant toxins.’. *J Infect Dis*, 191(1), pp. 81–88.

239 Mokrousov, I (2009) ‘Corynebacterium diphtheriae: Genome diversity, population structure and genotyping perspectives’. *Infection, Genetics and Evolution*, 9(1), pp. 1–15.

240 Wagner, K S, White, J M, Crowcroft, N S, MARTIN, S DE, et al. (2010) ‘Diphtheria in the United Kingdom, 1986–2008: the increasing role of Corynebacterium ulcerans’. *Epidemiology & Infection*, 138(11), pp. 1519–1530.

241 (null), The World Bank (2016) ‘**Poverty rising in Africa report**’. *Poverty in Africa*. [online] Available from: ttpwww.worldbank.orgenregionafrpublicationpoverty-rising-africa-poverty-report (Accessed 13 November 2016)

242 Liu, Li, Johnson, Hope L, Cousens, Simon, Perin, Jamie, et al. (2012) ‘Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000’. *The Lancet*, 379(9832), pp. 2151–2161.

243 United NationsUN Social Affairs, Population Division (2014) ‘**Urban /rural division of countries for the years 2015 and 2025**’. *Urban/ Rural population*. [online] Available from: http://www.geohive.com/earth/pop_urban.aspx (Accessed 12 November 2016)

244 Bolarinwa, Joshua Olusegun (2015) ‘The ECOWAS Free Movement Protocol: Obstacle or Driver of Regional Integration?’. *Insight on Africa*, 7(2), pp. 154–168.

245 Dean, Anna S, Fournié, Guillaume, Kulo, Abalo E, Boukaya, G Aboudou, et al. (2013) ‘Potential Risk of Regional Disease Spread in West Africa through Cross-Border Cattle Trade’ Kaltenboeck, B. (ed.). *PLoS One*, 8(10), p. e75570.

246 Cenciarelli, Orlando, Pietropaoli, Stefano, Malizia, Andrea, Carestia, Mariachiara, et al. (2015) ‘Ebola Virus Disease 2013-2014 Outbreak in West Africa: An Analysis of the Epidemic Spread and Response’. *International Journal of Microbiology*, 2015(2), pp. 1–12.

247 Kobayashi, Scott D, Malachowa, Natalia and DeLeo, Frank R (2015) ‘Pathogenesis of Staphylococcus aureus Abscesses’. *The American Journal of Pathology*, 185(6), pp. 1518–1527.

248 Wertheim, Heiman FL, Melles, Damian C, Vos, Margreet C, van Leeuwen, Willem, et al. (2005) ‘The role of nasal carriage in Staphylococcus aureus infections’. *Lancet Infect Dis*, 5(12), pp. 751–762.

249 Kooistra-Smid, Mirjam, Nieuwenhuis, Marianne, van Belkum, Alex and Verbrugh, Henri (2009) ‘The role of nasal carriage in Staphylococcus

aureus burn wound colonization'. *FEMS Immunol Med Microbiol*, 57(1), pp. 1–13.

250 Clarke, Simon R, Brummell, Kirsten J, Horsburgh, Malcolm J, McDowell, Philip W, et al. (2006) 'Identification of in vivo-expressed antigens of Staphylococcus aureus and their use in vaccinations for protection against nasal carriage.'. *J Infect Dis*, 193(8), pp. 1098–1108.

251 Foster, Timothy J, Geoghegan, Joan A, Ganesh, Vannakambadi K and Höök, Magnus (2014) 'Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus'. *Nat Rev Microbiol*, 12(1), pp. 49–62.

252 Tong, S Y, Davis, J S, Eichenberger, E, Holland, T L and Fowler, V G Jr (2015) 'Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management'. *Clin Microbiol Rev*, 28(3), pp. 603–661.

253 Naber, Christoph K (2009) 'Staphylococcus aureus bacteremia: epidemiology, pathophysiology, and management strategies.'. *Clin Infect Dis*, 48 Suppl 4(Supplement 4), pp. S231–7.

254 Malachowa, Natalia, Kobayashi, Scott D, Sturdevant, Daniel E, Scott, Dana P and DeLeo, Frank R (2015) 'Insights into the Staphylococcus aureus -Host Interface: Global Changes in Host and Pathogen Gene Expression in a Rabbit Skin Infection Model' Kaufmann, G. F. (ed.). *PLoS One*, 10(2), p. e0117713.

255 Ray, G Thomas, Suaya, Jose A and Baxter, Roger (2013) 'Microbiology of skin and soft tissue infections in the age of community-acquired methicillin-resistant Staphylococcus aureus'. *Diagn Microbiol Infect Dis*, 76(1), pp. 24–30.

256 Spaan, András N, Henry, Thomas, van Rooijen, Willemien J M, Perret, Magali, et al. (2013) 'The Staphylococcal Toxin Panton-Valentine Leukocidin Targets Human C5a Receptors'. *Cell Host & Microbe*, 13(5), pp. 584–594.

257 Stryjewski, Martin E and Corey, G Ralph (2014) 'Methicillin-resistant Staphylococcus aureus: an evolving pathogen.'. *Clin Infect Dis*, 58 Suppl 1(suppl 1), pp. S10–9.

258 Kim, Chung-Jong, Kim, Hong-Bin, Oh, Myoung-don, Kim, Yunhee, et al. (2014) 'The burden of nosocomial staphylococcus aureus bloodstream infection in South Korea: a prospective hospital-based nationwide study'. *BMC Infect Dis*, 14(1), p. 1.

259 Zetola, Nicola, Francis, John S, Nuermberger, Eric L and Bishai, William R (2005) 'Community-acquired meticillin-resistant Staphylococcus aureus: an emerging threat'. *Lancet Infect Dis*, 5(5), pp. 275–286.

260 Laupland, K B, Lyytikäinen, O, Søgaard, M, Kennedy, K J, et al. (2013) 'The changing epidemiology of Staphylococcus aureus bloodstream infection: a multinational population-based surveillance study'. *Clin Microbiol Infect*, 19(5), pp. 465–471.

261 Musiime, V, Cook, A, Bakeera-Kitaka, S, Vhembo, T, et al. (2013) 'Bacteremia, causative agents and antimicrobial susceptibility among HIV-1-infected children on antiretroviral therapy in Uganda and Zimbabwe'. *Pediatr Infect Dis J*, 32(8), pp. 856–862.

262 Hill, P. C., Onyeama, C O, Ikumapayi, U N, Secka, O, et al. (2007) 'Bacteraemia in patients admitted to an urban hospital in West Africa'. *BMC Infect Dis*, 7, p. 2.

263 Mulholland, E K, Ogunlesi, O O, Adegbola, R A, Weber, M, et al. (1999) 'Etiology of serious infections in young Gambian infants'. *Pediatr Infect Dis J*, 18(10 Suppl), pp. S35–41.

264 Dromigny, J A, Nabeth, P and Perrier Gros Claude, J D (2002) 'Distribution and susceptibility of bacterial urinary tract infections in Dakar, Senegal'. *International Journal of Antimicrobial Agents*, 20(5), pp. 339–347.

265 Adjei, Ohene and Opoku, Charles (2004) 'Urinary tract infections in African infants'. *International Journal of Antimicrobial Agents*, 24, pp. 32–34.

266 Harrison, E M, Weinert, L A, Holden, M T, Welch, J J, et al. (2014) 'A shared population of epidemic methicillin-resistant Staphylococcus aureus 15 circulates in humans and companion animals'. *MBio*, 5(3), pp. e00985–13.

267 Porrero, M Concepción, Mentaberre, Gregorio, Sánchez, Sergio, Fernández-Llario, Pedro, et al. (2014) 'Carriage of Staphylococcus aureus by free-living wild animals in Spain.'. *Appl Environ Microbiol*, 80(16), pp. 4865–4870.

268 Schaumburg, F, Pauly, M, Anoh, E, Mossoun, A, et al. (2015) 'Staphylococcus aureus complex from animals and humans in three remote African regions'. *Clin Microbiol Infect*, 21(4), pp. 345 e1–8.

269 Bogaert, D, de Groot, R and Hermans, PWM (2004) 'Streptococcus pneumoniae colonisation: the key to pneumococcal disease'. *Lancet Infect Dis*, 4(3), pp. 144–154.

270 Mehr, Sam and Wood, Nicholas (2012) 'Streptococcus pneumoniae – a review of carriage, infection, serotype replacement and vaccination'. *Paediatric respiratory reviews*, 13(4), pp. 258–264.

271 Mook-Kanamori, Barry B, Geldhoff, Madelijn, van der Poll, Tom and van de Beek, Diederik (2011) 'Pathogenesis and pathophysiology of

pneumococcal meningitis.'. *Clin Microbiol Rev*, 24(3), pp. 557–591.

272 Magee, A D and Yother, J (2001) 'Requirement for capsule in colonization by Streptococcus pneumoniae.'. *Infect Immun*, 69(6), pp. 3755–3761.

273 Swiatlo, Edwin, Champlin, Franklin R, Holman, Steven C, Wilson, W William and Watt, James M (2002) 'Contribution of choline-binding proteins to cell surface properties of Streptococcus pneumoniae.'. *Infect Immun*, 70(1), pp. 412–415.

274 Tong, H H, Blue, L E, James, M A and DeMaria, T F (2000) 'Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model.'. *Infect Immun*, 68(2), pp. 921–924.

275 Usuf, E, Bottomley, C, Adegbola, R A and Hall, A (2014) 'Pneumococcal carriage in sub-Saharan Africa—a systematic review'. *PLoS One*.

276 Centers for Disease Control and Prevention (CDC) (2008) 'Invasive pneumococcal disease in children 5 years after conjugate vaccine introduction--eight states, 1998-2005.'. *MMWR. Morbidity and mortality weekly report*, 57(6), pp. 144–148.

277 Lawn, Stephen D, Harries, Anthony D, Anglaret, Xavier, Myer, Landon and Wood, Robin (2008) 'Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa'. *AIDS (London, England)*, 22(15), pp. 1897–1908.

278 Gessner, Bradford D, Mueller, Judith E and Yaro, Seydou (2010) 'African meningitis belt pneumococcal disease epidemiology indicates a need for an effective serotype 1 containing vaccine, including for older children and adults'. *BMC Infect Dis*, 10(1), p. 1.

279 Traore, Yves, Tameklo, Tsidi Agbeko, Njanpop-Lafourcade, Berthe-Marie, Lourd, Mathilde, et al. (2009) 'Incidence, seasonality, age distribution, and mortality of pneumococcal meningitis in Burkina Faso and Togo.'. *Clin Infect Dis*, 48 Suppl 2(s2), pp. S181–9.

280 Mackenzie, Grant A, Hill, Philip C, Jeffries, David J, Hossain, Ilias, et al. (2016) 'Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study'. *Lancet Infect Dis*, 16(6), pp. 703–711.

281 Egere, Uzochukwu, Townend, John, Roca, Anna, Akinsanya, Abiodun, et al. (2012) 'Indirect Effect of 7-Valent Pneumococcal Conjugate Vaccine on Pneumococcal Carriage in Newborns in Rural Gambia: A Randomised Controlled Trial' Ellis, R. D. (ed.). *PLoS One*, 7(11), p. e49143.

282 Millar, Eugene V, Watt, James P, Bronsdon, Melinda A, Dallas, Jean, et al. (2008) 'Indirect effect of 7-valent pneumococcal conjugate vaccine on pneumococcal colonization among unvaccinated household members.'. *Clin Infect Dis*, 47(8), pp. 989–996.

283 Kwambana-Adams, Brenda Anna, Asiedu-Bekoe, Franklin, Sarkodie, Badu, Afreh, Osei Kuffour, et al. (2016) 'An outbreak of pneumococcal meningitis among older children (≥5 years) and adults after the implementation of an infant vaccination programme with the 13-valent pneumococcal conjugate vaccine in Ghana'. *BMC Infect Dis*, 16(1), p. 17.

284 Gagneux, Sebastien (2013) 'Genetic Diversity in Mycobacterium tuberculosis', in *Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism*, Current Topics in Microbiology and Immunology, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 1–25.

285 Zuber, Benoît, Chami, Mohamed, Houssin, Christine, Dubochet, Jacques, et al. (2008) 'Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state.'. *J Bacteriol*, 190(16), pp. 5672–5680.

286 Brennan, Patrick J and Nikaido, Hiroshi (2003) 'The Envelope of Mycobacteria'. *Annual Reviews*, 64(1), pp. 29–63.

287 Kahnert, Antje, Höpken, Uta E, Stein, Maik, Bandermann, Silke, et al. (2007) 'Mycobacterium tuberculosis triggers formation of lymphoid structure in murine lungs.'. *J Infect Dis*, 195(1), pp. 46–54.

288 Miranda, M Silva, Breiman, A and Allain, S (2012) 'The tuberculous granuloma: an unsuccessful host defence mechanism providing a safety shelter for the bacteria?'. *Clinical and and Developmental Immunology*.

289 Lawn, Stephen D and Zumla, Alimuddin I (2011) 'Tuberculosis'. *The Lancet*, 378(9785), pp. 57–72.

290 Müller, Borna, Borrell, Sonia, Rose, Graham and Gagneux, Sebastien (2013) 'The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis'. *Trends in Genetics*, 29(3), pp. 160–169.

291 Jassal, Mandeep and Bishai, William R (2009) 'Extensively drug-resistant tuberculosis'. *Lancet Infect Dis*, 9(1), pp. 19–30.

292 World Health Organization (2016) 'WHO | Global tuberculosis report 2016'. *WHO*. [online] Available from: http://www.who.int/tb/publications/global_report/en/

293 Gehre, Florian, Otu, Jacob, Kendall, Lindsay, Forson, Audrey, et al. (2016) 'The emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for large-scale tuberculosis research and drug resistance surveillance'. *BMC Medicine*, 14(1), p.

160.

294 Gehre, Florian, Kumar, Samrat, Kendall, Lindsay, Ejo, Mebrat, et al. (2016) 'A Mycobacterial Perspective on Tuberculosis in West Africa: Significant Geographical Variation of M. africanum and Other M. tuberculosis Complex Lineages.' Burke, M. (ed.). *PLoS Negl Trop Dis*, 10(3), p. e0004408.

295 Feil, E J, Li, B C, Aanensen, D M, Hanage, W P and Spratt, B G (2004) 'eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data'. *J Bacteriol*, 186(5), pp. 1518–1530.

296 Hill, Philip C, Cheung, Yin Bun, Akisanya, Abiodun, Sankareh, Kawsu, et al. (2008) 'Nasopharyngeal carriage of Streptococcus pneumoniae in Gambian infants: a longitudinal study.'. *Clin Infect Dis*, 46(6), pp. 807–814.

297 Clinical Laboartory Standards institute (2014) '**Performance Standards for Antimicrobial Susceptibility Testing. Twenty-Fourth Information Supplement (M100-S24)**'. *CLSI*.

298 Ardito, F, Posteraro, B, Sanguinetti, M, Zanetti, S and Fadda, G (2001) 'Evaluation of BACTEC Mycobacteria Growth Indicator Tube (MGIT 960) automated system for drug susceptibility testing of Mycobacterium tuberculosis'. *J Clin Microbiol*, 39(12), pp. 4440–4444.

299 Hosek, J, Svastova, P, Moravkova, M and Pavlik, I (2006) 'Methods of mycobacterial DNA isolation from different biological material: a review'. *Veterinarni medicina*.

300 Connor, Thomas R, Loman, Nicholas J, Thompson, Simon, Smith, Andy, et al. (2016) 'CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community'. *Microbial Genomics*, 2(9).

301 Bankevich, Anton, Nurk, Sergey, Antipov, Dmitry, Gurevich, Alexey A, et al. (2012) 'SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing'. *dx.doi.org*, 19(5), pp. 455–477.

302 Seemann, Torsten (2014) 'Prokka: rapid prokaryotic genome annotation'. *Bioinformatics*, 30(14), pp. btu153–2069.

303 Inouye, Michael, Dashnow, Harriet, Raven, Lesley-Ann, Schultz, Mark B, et al. (2014) 'SRST2: Rapid genomic surveillance for public health and hospital microbiology labs.'. *Genome medicine*, 6(11), p. 90.

304 Bradley, P, Gordon, N C, Walker, T M, Dunn, L, et al. (2015) 'Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis'. *Nat Commun*, 6, p. 10063.

305 Bentley, Stephen D, Aanensen, David M, Mavroidi, Angeliki, Saunders, David, et al. (2006) 'Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes.'. *PLoS Genet*, 2(3), p. e31.

306 Seemann, Torsten (n.d.) 'Nullarbor'. *Nullarbor*. [online] Available from: httpsgithub.comtseemannnullarbor (Accessed 2016)

307 Bolger, Anthony M, Lohse, Marc and Usadel, Bjoern (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data'. *Bioinformatics*, 30(15), pp. btu170–2120.

308 Wood, Derrick E and Salzberg, Steven L (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments'. *Genome biology*, 15(3), p. R46.

309 Li, Dinghua, Liu, Chi-Man, Luo, Ruibang, Sadakane, Kunihiko and Lam, Tak-Wah (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.'. *Bioinformatics*, 31(10), pp. 1674–1676.

310 Enright, M C and Spratt, B G (1998) 'A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease.'. *Microbiology*, 144 (Pt 11)(11), pp. 3049–3060.

311 Seemann, Torsten (2016) 'MLST'. [online] Available from: https://github.com/tseemann/mlst

312 Seemann, Torsten (n.d.) 'Abricate'. *Abricate*. [online] Available from: https://github.com/tseemann/abricate (Accessed May 2016)

313 Steiner, A, Stucki, D, Coscolla, M, Borrell, S and Gagneux, S (2014) 'KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes'. *BMC Genomics*, 15, p. 881.

314 Feuerriegel, Silke, Schleusener, Viola, Beckert, Patrick, Kohl, Thomas A, et al. (2015) 'PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data' Carroll, K. C. (ed.). *J Clin Microbiol*, 53(6), pp. 1908–1914.

315 Coll, Francesc, McNerney, Ruth, Preston, Mark D, Guerra-Assunção, José Afonso, et al. (2015) 'Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences'. *Genome medicine*, 7(1), p. 51.

316 Senghore, Madikay, Bayliss, Sion C, Kwambana-Adams, Brenda A, Foster-Nyarko, Ebenezer, et al. (2016) 'Transmission of Staphylococcus aureus from Humans to Green Monkeys in The Gambia as Revealed by Whole-Genome Sequencing' Björkroth, J. (ed.). *Appl Environ Microbiol*, 82(19), pp. 5910–5917.

317 Ponstingl, Hannes (n.d.) ‘SMALT’. *SMALT*. [online] Available from: http://www.sanger.ac.uksciencetoolssmalt- (Accessed 18 November 2016)

318 Li, Heng (2011) ‘A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.’. *Bioinformatics*, 27(21), pp. 2987–2993.

319 McKenna, Aaron, Hanna, Matthew, Banks, Eric, Sivachenko, Andrey, et al. (2010) ‘The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.’. *Genome Res*, 20(9), pp. 1297–1303.

320 Price, Morgan N, Dehal, Paramvir S and Arkin, Adam P (2010) ‘FastTree 2--approximately maximum-likelihood trees for large alignments.’ Poon, A. F. Y. (ed.). *PLoS One*, 5(3), p. e9490.

321 Uhlemann, A C, Dordel, J, Knox, J R, Raven, K E, et al. (2014) ‘Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community’. *Proceedings of the National Academy of Science*, 111(18), pp. 6738–6743.

322 Seemann, Torsten (n.d.) ‘Snippy’. *SNIPPY*. [online] Available from: httpsgithub.comtseemannsnippy (Accessed 18 November 2016)

323 Zhou, Zhemin, McCann, Angela, Litrup, Eva, Murphy, Ronan, et al. (2013) ‘Neutral genomic microevolution of a recently emerged pathogen, Salmonella enterica serovar Agona.’ Casadesús, J. (ed.). *PLoS Genet*, 9(4), p. e1003471.

324 Delcher, Arthur L, Salzberg, Steven L and Phillippy, Adam M (2003) *Using MUMmer to Identify Similar Regions in Large Sequence Sets*, Hoboken, NJ, USA, John Wiley & Sons, Inc.

325 Stamatakis, Alexandros (2014) ‘RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.’. *Bioinformatics*, 30(9), pp. 1312–1313.

326 Croucher, Nicholas J, Page, Andrew J, Connor, Thomas R, Delaney, Aidan J, et al. (2014) ‘Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins’. *Nucleic Acids Res*, 43(3), pp. gku1196–e15.

327 Shannon, C E (2001) ‘A mathematical theory of communication’. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), pp. 3–55.

328 Agapow, Paul-Michael and Burt, Austin (2001) ‘Indices of multilocus linkage disequilibrium’. *Molecular Ecology Notes*, 1(1-2), pp. 101–102.

329 Witney, Adam A, Gould, Katherine A, Arnold, Amber, Coleman, David,

et al. (2015) ‘Clinical Application of Whole-Genome Sequencing To Inform Treatment for Multidrug-Resistant Tuberculosis Cases’ Forbes, B. A. (ed.). *J Clin Microbiol*, 53(5), pp. 1473–1483.

330 Chan, Jacqueline Z-M, Sergeant, Martin J, Lee, Oona Y C, Minnikin, David E, et al. (2013) ‘Metagenomic Analysis of Tuberculosis in a Mummy’. *N Engl J Med*, 369(3), pp. 289–290.

331 Kumar, Sudhir, Stecher, Glen and Tamura, Koichiro (2016) ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.’. *Mol Biol Evol*, 33(7), pp. 1870–1874.

332 Ozer, Egon A, Allen, Jonathan P and Hauser, Alan R (2014) ‘Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGEnt.’. *BMC Genomics*, 15(1), p. 737.

333 Lowy, Franklin D (2009) ‘ Staphylococcus aureus Infections’. *dx.doi.org*, 339(8), pp. 520–532.

334 McCarthy, Alex J and Lindsay, Jodi A (2012) ‘The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated’. *BMC Microbiol*, 12(1), p. 1.

335 Lindsay, J A, Moore, C E, Day, N P, Peacock, S J, et al. (2006) ‘Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes’. *J Bacteriol*, 188(2), pp. 669–676.

336 Goerke, Christiane, Pantucek, Roman, Holtfreter, Silva, Schulte, Berit, et al. (2009) ‘Diversity of prophages in dominant Staphylococcus aureus clonal lineages.’. *J Bacteriol*, 191(11), pp. 3462–3468.

337 Lindsay, Jodi A (2014) ‘Staphylococcus aureus genomics and the impact of horizontal gene transfer’. *International Journal of Medical Microbiology*, 304(2), pp. 103–109.

338 McCarthy, Alex J, Loeffler, Anette, Witney, Adam A, Gould, Katherine A, et al. (2014) ‘Extensive horizontal gene transfer during Staphylococcus aureus co-colonization in vivo.’. *Genome Biol Evol*, 6(10), pp. 2697–2708.

339 Lindsay, J A and Holden, M T (2004) ‘Staphylococcus aureus: superbug, super genome?’. *Trends Microbiol*, 12(8), pp. 378–385.

340 Everitt, R G, Didelot, X, Batty, E M, Miller, R R, et al. (2014) ‘Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus’. *Nat Commun*, 5, p. 3956.

341 Uhlemann, A C, Porcella, S F, Trivedi, S, Sullivan, S B, et al. (2012) ‘Identification of a highly transmissible animal-independent Staphylococcus aureus ST398 clone with distinct genomic and cell

adhesion properties'. *MBio*, 3(2).

342 Langille, M G, Hsiao, W W and Brinkman, F S (2010) 'Detecting genomic islands using bioinformatics approaches'. *Nat Rev Microbiol*, 8(5), pp. 373–382.

343 Malachowa, Natalia and DeLeo, Frank R (2010) 'Mobile genetic elements of Staphylococcus aureus'. *Cellular and molecular life sciences*, 67(18), pp. 3057–3071.

344 Moon, B Y, Park, J Y, Hwang, S Y, Robinson, D A, et al. (2015) 'Phage-mediated horizontal transfer of a Staphylococcus aureus virulence-associated genomic island'. *Sci Rep*, 5, p. 9784.

345 Chambers, Henry F and DeLeo, Frank R (2009) 'Waves of resistance: Staphylococcus aureus in the antibiotic era'. *Nat Rev Microbiol*, 7(9), pp. 629–641.

346 Olsen, John Elmerdahl, Christensen, Henrik and Aarestrup, Frank Møller (2006) 'Diversity and evolution of blaZ from Staphylococcus aureus and coagulase-negative staphylococci.'. *Journal of Antimicrobial Chemotherapy*, 57(3), pp. 450–460.

347 Lee, John Hwa (2003) 'Methicillin (Oxacillin)-resistant Staphylococcus aureus strains isolated from major food animals and their potential transmission to humans.'. *Appl Environ Microbiol*, 69(11), pp. 6489–6494.

348 Brumfitt, William and Hamilton-Miller, Jeremy (2010) 'Methicillin-Resistant Staphylococcus aureus'. *dx.doi.org*, 320(18), pp. 1188–1196.

349 Berger-Bächi, Brigitte and Rohrer, Susanne (2002) 'Factors influencing methicillin resistance in staphylococci'. *Archives of Microbiology*, 178(3), pp. 165–171.

350 Appelbaum, Peter C (2007) 'Microbiology of antibiotic resistance in Staphylococcus aureus.'. *Clin Infect Dis*, 45 Suppl 3(Supplement 3), pp. S165–70.

351 Hiramatsu, Keiichi, Aritaka, Nanae, Hanaki, Hideaki, Kawasaki, Shiori, et al. (1997) 'Dissemination in Japanese hospitals of strains of Staphylococcus aureus heterogeneously resistant to vancomycin'. *The Lancet*, 350(9092), pp. 1670–1673.

352 Shambat, S, Nadig, S, Prabhakara, S, Bes, M, et al. (2012) 'Clonal complexes and virulence factors of Staphylococcus aureus from several cities in India'. *BMC Microbiol*, 12, p. 64.

353 Ruimy, Raymond, Maiga, Aminata, Armand-Lefevre, Laurence, Maiga, Ibrahim, et al. (2008) 'The carriage population of Staphylococcus aureus from Mali is composed of a combination of pandemic clones and the divergent Panton-Valentine leukocidin-positive genotype

ST152.'. *J Bacteriol*, 190(11), pp. 3962–3968.

354 Kraef, Christian, Alabi, Abraham S, Peters, Georg, Becker, Karsten, et al. (2015) 'Co-detection of Panton-Valentine leukocidin encoding genes and cotrimoxazole resistance in Staphylococcus aureus in Gabon: implications for HIV-patients' care.'. *Frontiers in microbiology*, 6(455), p. 60.

355 Kikuta, H, Shibata, M, Nakata, S, Yamanaka, T, et al. (2011) 'Predominant Dissemination of PVL-Negative CC89 MRSA with SCCmec Type II in Children with Impetigo in Japan.'. *International journal of pediatrics*, 2011(5), pp. 143872–8.

356 Kuroda, M, Ohta, T, Uchiyama, I, Baba, T, et al. (2001) 'Whole genome sequencing of meticillin-resistant Staphylococcus aureus.'. *Lancet*, 357(9264), pp. 1225–1240.

357 Fitzgerald, J Ross and Holden, Matthew T G (2016) 'Genomics of Natural Populations of Staphylococcus aureus.'. *Annu Rev Microbiol*, 70(1), pp. 459–478.

358 Price, James R, Golubchik, Tanya, Cole, Kevin, Wilson, Daniel J, et al. (2014) 'Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of Staphylococcus aureus in an intensive care unit.'. *Clin Infect Dis*, 58(5), pp. 609–618.

359 Tong, SYC, Holden, MTG and Nickerson, E K (2015) 'Genome sequencing defines phylogeny and spread of methicillin-resistant Staphylococcus aureus in a high transmission setting'. *Genome Res*, 25(3), pp. 111–118.

360 Harris, S R, Feil, E J, Holden, M T G, Quail, M A, et al. (2010) 'Evolution of MRSA During Hospital Transmission and Intercontinental Spread'. *Science*, 327(5964), pp. 469–474.

361 Young, Bernadette C, Golubchik, Tanya, Batty, Elizabeth M, Fung, Rowena, et al. (2012) 'Evolutionary dynamics of Staphylococcus aureus during progression from carriage to disease.'. *Proc Natl Acad Sci U S A*, 109(12), pp. 4550–4555.

362 Schaumburg, F, Alabi, A S, Peters, G and Becker, K (2014) 'New epidemiology of Staphylococcus aureus infection in Africa'. *Clin Microbiol Infect*, 20(7), pp. 589–596.

363 Golubchik, Tanya, Batty, Elizabeth M, Miller, Ruth R, Farr, Helen, et al. (2013) 'Within-host evolution of Staphylococcus aureus during asymptomatic carriage.' Aziz, R. K. (ed.). *PLoS One*, 8(5), p. e61319.

364 Bartels, M D, Larner-Svensson, H and Meiniche, H (2015) 'Monitoring meticillin resistant Staphylococcus aureus and its spread in Copenhagen, Denmark, 2013, through routine whole genome sequencing'. *Eurosurveillance*.

365 Harrison, Ewan M, Ludden, Catherine, Brodrick, Hayley J, Blane, Beth, et al. (2016) 'Transmission of methicillin-resistant Staphylococcus aureus in long-term care facilities and their related healthcare networks'. *Genome medicine*, 8(1), p. 102.

366 Toleman, M S, Reuter, S, Coll, F, Harrison, E M and Blane, B (2016) 'Systematic surveillance detects multiple silent introductions and household transmission of'. *J Infect Dis*.

367 Smyth, Davida S, Feil, Edward J, Meaney, William J, Hartigan, Patrick J, et al. (2009) 'Molecular genetic typing reveals further insights into the diversity of animal-associated Staphylococcus aureus'. *Journal of Medical Microbiology*, 58(10), pp. 1343–1353.

368 Spoor, L E, McAdam, P R, Weinert, L A, Rambaut, A, et al. (2013) 'Livestock origin for a human pandemic clone of community-associated methicillin-resistant Staphylococcus aureus'. *MBio*, 4(4).

369 Lewis, H C, Molbak, K, Reese, C, Aarestrup, F M, et al. (2008) 'Pigs as source of methicillin-resistant Staphylococcus aureus CC398 infections in humans, Denmark'. *Emerg Infect Dis*, 14(9), pp. 1383–1389.

370 Kock, R, Schaumburg, F, Mellmann, A, Koksal, M, et al. (2013) 'Livestock-associated methicillin-resistant Staphylococcus aureus (MRSA) as causes of human infection and colonization in Germany'. *PLoS One*, 8(2), p. e55040.

371 McCarthy, A J, van Wamel, W, Vandendriessche, S, Larsen, J, et al. (2012) 'Staphylococcus aureus CC398 clade associated with human-to-human transmission'. *Appl Environ Microbiol*, 78(24), pp. 8845–8848.

372 McCarthy, A J and Lindsay, J A (2013) 'Staphylococcus aureus innate immune evasion is lineage-specific: a bioinfomatics study'. *Infect Genet Evol*, 19, pp. 7–14.

373 Schaumburg, F, Mugisha, L, Peck, B, Becker, K, et al. (2012) 'Drug-resistant human Staphylococcus aureus in sanctuary apes pose a threat to endangered wild ape populations'. *Am J Primatol*, 74(12), pp. 1071–1075.

374 Nagel, M, Dischinger, J, Turck, M, Verrier, D, et al. (2013) 'Human-associated Staphylococcus aureus strains within great ape populations in Central Africa (Gabon)'. *Clin Microbiol Infect*, 19(11), pp. 1072–1077.

375 Tong, S Y, Schaumburg, F, Ellington, M J, Corander, J, et al. (2015) 'Novel staphylococcal species that form part of a Staphylococcus aureus-related complex: the non-pigmented Staphylococcus argenteus sp. nov. and the non-human primate-associated Staphylococcus schweitzeri sp. nov'. *Int J Syst Evol Microbiol*, 65(Pt 1), pp. 15–22.

376 Ma, Dongzhu, Jasinska, Anna J, Feyertag, Felix, Wijewardana,

Viskam, et al. (2014) ‘Factors associated with siman immunodeficiency virus transmission in a natural African nonhuman primate host in the wild.’. *J Virol*, 88(10), pp. 5687–5705.

377 Ebruke, C, Dione, M M, Walter, B, Worwui, A, et al. (2016) ‘High genetic diversity of Staphylococcus aureus strains colonising the nasopharynx of Gambian villagers before widespread use of pneumococcal conjugate vaccines’. *BMC Microbiol*, 16, p. 38.

378 Robinson, D Ashley and Enright, Mark C (2004) ‘Evolution of Staphylococcus aureus by large chromosomal replacements.’. *J Bacteriol*, 186(4), pp. 1060–1064.

379 Reed, S B, Wesson, C A, Liou, L E, Trumble, W R, et al. (2001) ‘Molecular characterization of a novel Staphylococcus aureus serine protease operon.’. *Infect Immun*, 69(3), pp. 1521–1527.

380 Frankel, Matthew B, Hendrickx, Antoni P A, Missiakas, Dominique M and Schneewind, Olaf (2011) ‘LytN, a murein hydrolase in the cross-wall compartment of Staphylococcus aureus, is involved in proper bacterial growth and envelope assembly.’. *J Biol Chem*, 286(37), pp. 32593–32605.

381 Cuny, Christiane, Abdelbary, Mohamed, Layer, Franziska, Werner, Guido and Witte, Wolfgang (2015) ‘Prevalence of the immune evasion gene cluster in Staphylococcus aureus CC398.’. *Vet Microbiol*, 177(1-2), pp. 219–223.

382 Verkaik, N J, Benard, M, Boelens, H A, de Vogel, C P, et al. (2011) ‘Immune evasion cluster-positive bacteriophages are highly prevalent among human Staphylococcus aureus strains, but they are not essential in the first stages of nasal colonization’. *Clin Microbiol Infect*, 17(3), pp. 343–348.

383 van Wamel, W J, Rooijakkers, S H, Ruyken, M, van Kessel, K P and van Strijp, J A (2006) ‘The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of Staphylococcus aureus are located on beta-hemolysin-converting bacteriophages’. *J Bacteriol*, 188(4), pp. 1310–1315.

384 Loffler, B, Hussain, M, Grundmeier, M, Bruck, M, et al. (2010) ‘Staphylococcus aureus panton-valentine leukocidin is a very potent cytotoxic factor for human neutrophils’. *PLoS Pathog*, 6(1), p. e1000715.

385 Molesworth, Anna M, Thomson, Madeleine C, Connor, Stephen J, Cresswell, Mark P, et al. (2002) ‘Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa’. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 96(3), pp. 242–249.

386 Agier, L, Deroubaix, A, Martiny, N, Yaka, P, et al. (2013) ‘Seasonality

of meningitis in Africa and climate forcing: aerosols stand out'. *Journal of The Royal Society Interface*, 10(79), pp. 20120814–20120814.

387 Teyssou, Rémy and Muros-Le Rouzic, Erwan (2007) 'Meningitis epidemics in Africa: A brief overview'. *Vaccine*, 25, pp. A3–A7.

388 Sultan, Benjamin, Labadi, Karima, Guégan, Jean-François and Janicot, Serge (2005) 'Climate Drives the Meningitis Epidemics Onset in West Africa' Hales, S. (ed.). *PLoS medicine*, 2(1), p. e6.

389 Yaka, Pascal, Sultan, Benjamin, Broutin, Hélène, Janicot, Serge, et al. (2008) 'Relationships between climate and year-to-year variability in meningitis outbreaks: A case study in Burkina Faso and Niger'. *International Journal of Health Geographics*, 7(1), p. 34.

390 Marticorena, B, Chatenet, B, Rajot, J L, Traoré, S, et al. (2010) 'Temporal variability of mineral dust concentrations over West Africa: analyses of a pluriannual monitoring from the AMMA Sahelian Dust Transect'. *Atmospheric Chemistry and Physics Discussions*, 10(3), pp. 8051–8101.

391 Mueller, Judith E and Gessner, Bradford D (2010) 'A hypothetical explanatory model for meningococcal meningitis in the African meningitis belt'. *International Journal of Infectious Diseases*, 14(7), pp. e553–e559.

392 Mueller, Judith E, Yaro, Seydou, Madec, Yoann, Somda, Paulin K, et al. (2008) 'Association of respiratory tract infection symptoms and air humidity with meningococcal carriage in Burkina Faso'. *Tropical Medicine & International Health*, 13(12), pp. 1543–1552.

393 Caugant, Dominique A (1998) 'Population genetics and molecular epidemiology of Neisseria meningitidis'. *Apmis*, 106(1-6), pp. 505–525.

394 Nathan, Nicolas, Rose, Angela M C, Legros, Dominique, Tiendrebeogo, Sylvestre R M, et al. (2007) 'Meningitis serogroup W135 outbreak, Burkina Faso, 2002.'. *Emerg Infect Dis*, 13(6), pp. 920–923.

395 Traore, Yves, Njanpop-Lafourcade, Berthe-Marie, Adjogble, Kokou-Louis-Sewonou, Lourd, Mathilde, et al. (2006) 'The rise and fall of epidemic Neisseria meningitidis serogroup W135 meningitis in Burkina Faso, 2002-2005.'. *Clin Infect Dis*, 43(7), pp. 817–822.

396 Leimkugel, Julia, Adams Forgor, Abudulai, Gagneux, Sebastien, Pflüger, Valentin, et al. (2005) 'An outbreak of serotype 1 Streptococcus pneumoniae meningitis in northern Ghana with features that are characteristic of Neisseria meningitidis meningitis epidemics.'. *J Infect Dis*, 192(2), pp. 192–199.

397 Mueller, Judith E, Yaro, Seydou, Ouedraogo, Macaire S, Levina, Natalia, et al. (2012) 'Pneumococci in the African Meningitis Belt:

Meningitis Incidence and Carriage Prevalence in Children and Adults' Beall, B. (ed.). *PLoS One*, 7(12), p. e52464.

398 Yaro, Seydou, Lourd, Mathilde, Traore, Yves, Njanpop-Lafourcade, Berthe-Marie, et al. (2006) 'Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso.'. *Clin Infect Dis*, 43(6), pp. 693–700.

399 Antonio, Martin, Hakeem, Ishrat, Awine, Timothy, Secka, Ousman, et al. (2008) 'Seasonality and outbreak of a predominant Streptococcus pneumoniae serotype 1 clone from The Gambia: expansion of ST217 hypervirulent clonal complex in West Africa.'. *BMC Microbiol*, 8(1), p. 198.

400 Kristiansen, Paul A, Ba, Absatou Ky, Sanou, Idrissa, Ouédraogo, Abdoul Salam, et al. (2013) 'Phenotypic and genotypic characterization of meningococcal carriage and disease isolates in Burkina Faso after mass vaccination with a serogroup a conjugate vaccine.'. *BMC Infect Dis*, 13(1), p. 363.

401 Donkor, Eric S, Dayie, Nicholas TKD and Badoe, Ebenezer V (2013) 'Vaccination against pneumococcus in West Africa: perspectives and prospects'. *International Journal of General Medicine*, 6, pp. 757–764.

402 Chaguza, Chrispin, Andam, Cheryl P, Harris, Simon R, Cornick, Jennifer E, et al. (2016) 'Recombination in Streptococcus pneumoniae Lineages Increase with Carriage Duration and Size of the Polysaccharide Capsule.'. *MBio*, 7(5), pp. e01053–16.

403 Andam, Cheryl P and Hanage, William P (2015) 'Mechanisms of genome evolution of Streptococcus'. *Infection, Genetics and Evolution*, 33, pp. 334–342.

404 Donati, Claudio, Hiller, N Luisa, Tettelin, Hervé, Muzzi, Alessandro, et al. (2010) 'Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species'. *Genome biology*, 11(10), p. R107.

405 Cornick, J E and Bentley, S D (2012) 'Streptococcus pneumoniae: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides'. *Microbes and Infection*, 14(7-8), pp. 573–583.

406 Charpentier, Emmanuelle and Tuomanen, Elaine (2000) 'Mechanisms of antibiotic resistance and tolerance in Streptococcus pneumoniae'. *Microbes and Infection*, 2(15), pp. 1855–1864.

407 Rice, Louis B (1998) 'Tn916 Family Conjugative Transposons and Dissemination of Antimicrobial Resistance Determinants'. *Antimicrob Agents Chemother*, 42(8), pp. 1871–1877.

408 Okitsu, Naohiro, Kaieda, Satoru, Yano, Hisakazu, Nakano, Ryuichi, et al. (2005) 'Characterization of ermB gene transposition by Tn1545 and

Tn917 in macrolide-resistant Streptococcus pneumoniae isolates.'. *J Clin Microbiol*, 43(1), pp. 168–173.

409 Ayoubi, P, Kilic, A O and Vijayakumar, M N (1991) 'Tn5253, the pneumococcal omega (cat tet) BM6001 element, is a composite structure of two conjugative transposons, Tn5251 and Tn5252.'. *J Bacteriol*, 173(5), pp. 1617–1622.

410 Ling, Losee L, Schneider, Tanja, Peoples, Aaron J, Spoering, Amy L, et al. (2015) 'A new antibiotic kills pathogens without detectable resistance.'. *Nature*, 517(7535), pp. 455–459.

411 O'Brien, K L and Dagan, R (2003) 'The potential indirect effect of conjugate pneumococcal vaccines'. *Vaccine*, 21(17-18), pp. 1815–1825.

412 Shiri, Tinevimbo, Datta, Samik, Madan, Jason, Tsertsvadze, Alexander, et al. (2017) 'Indirect effects of childhood pneumococcal conjugate vaccination on invasive pneumococcal disease: a systematic review and meta-analysis'. *Lancet Glob Health*, 5(1), pp. e51–e59.

413 Cornick, Jennifer E, Chaguza, Chrispin, Harris, Simon R, Yalcin, Feyruz, et al. (2015) 'Region-specific diversification of the highly virulent serotype 1 Streptococcus pneumoniae'. *Microbial Genomics*, 1(2).

414 Henriques-Normark, Birgitta, Blomberg, Christel, Dagerhamn, Jessica, Bättig, Patrick and Normark, Staffan (2008) 'The rise and fall of bacterial clones: Streptococcus pneumoniae'. *Nat Rev Microbiol*, 6(11), pp. 827–837.

415 Hanage, William P, Spratt, Brian G, Turner, Katherine M E and Fraser, Christophe (2006) 'Modelling bacterial speciation'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), pp. 2039–2044.

416 Saar-Dover, Ron, Bitler, Arkadi, Nezer, Ravit, Shmuel-Galia, Liraz, et al. (2012) 'D-alanylation of lipoteichoic acids confers resistance to cationic peptides in group B streptococcus by increasing the cell wall density.' Peschel, A. (ed.). *PLoS Pathog*, 8(9), p. e1002891.

417 Reyes-Caballero, Hermes, Guerra, Alfredo J, Jacobsen, Faith E, Kazmierczak, Krystyna M, et al. (2010) 'The metalloregulatory zinc site in Streptococcus pneumoniae AdcR, a zinc-activated MarR family repressor.'. *J Mol Biol*, 403(2), pp. 197–216.

418 Ellison, Damon W and Miller, Virginia L (2006) 'Regulation of virulence by members of the MarR/SlyA family.'. *Curr Opin Microbiol*, 9(2), pp. 153–159.

419 Garmory, Helen S and Titball, Richard W (2004) 'ATP-binding cassette transporters are targets for the development of antibacterial vaccines

and therapies.'. *Infect Immun*, 72(12), pp. 6757–6763.

420 Wyres, Kelly L, Lambertsen, Lotte M, Croucher, Nicholas J, McGee, Lesley, et al. (2013) 'Pneumococcal capsular switching: a historical perspective.'. *J Infect Dis*, 207(3), pp. 439–449.

421 McDaniel, Larry S and Swiatlo, Edwin (2016) 'Should Pneumococcal Vaccines Eliminate Nasopharyngeal Colonization?'. *MBio*, 7(3), pp. e00545–16.

422 Smith, Noel H, Kremer, Kristin, Inwald, Jacqueline, Dale, James, et al. (2006) 'Ecotypes of the Mycobacterium tuberculosis complex'. *Journal of Theoretical Biology*, 239(2), pp. 220–225.

423 Comas, Iñaki, Coscolla, Mireia, Luo, Tao, Borrell, Sonia, et al. (2013) 'Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans.'. *Nat Genet*, 45(10), pp. 1176–1182.

424 Bos, Kirsten I, Harkins, Kelly M, Herbig, Alexander, Coscolla, Mireia, et al. (2014) 'Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis'. *Nature*, 514(7523), pp. 494–497.

425 Gagneux, S and Small, P M (2007) 'Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development'. *Lancet Infect Dis*, 7(5), pp. 328–337.

426 Niemann, Stefan and Supply, Philip (2014) 'Diversity and Evolution of Mycobacterium tuberculosis: Moving to Whole-Genome-Based Approaches'. *Cold Spring Harbor Perspectives in Medicine*, 4(12), pp. a021188–a021188.

427 Kay, Gemma L, Sergeant, Martin J, Zhou, Zhemin, Chan, Jacqueline Z-M, et al. (2015) 'Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe.'. *Nat Commun*, 6, p. 6717.

428 Coscolla, Mireia and Gagneux, Sebastien (2014) 'Consequences of genomic diversity in Mycobacterium tuberculosis.'. *Semin Immunol*, 26(6), pp. 431–444.

429 Gagneux, Sebastien, DeRiemer, Kathryn, Van, Tran, Kato-Maeda, Midori, et al. (2006) 'Variable host-pathogen compatibility in Mycobacterium tuberculosis.'. *Proc Natl Acad Sci U S A*, 103(8), pp. 2869–2873.

430 Firdessa, Rebuma, Berg, Stefan, Hailu, Elena, Schelling, Esther, et al. (2013) 'Mycobacterial Lineages Causing Pulmonary and Extrapulmonary Tuberculosis, Ethiopia'. *Emerg Infect Dis*, 19(3), pp. 460–463.

431 de Jong, Bouke C, Antonio, Martin and Gagneux, Sebastien (2010)

'Mycobacterium africanum—Review of an Important Cause of Human Tuberculosis in West Africa' Picardeau, M. (ed.). *PLoS Negl Trop Dis*, 4(9), p. e744.

432 Winglee, Kathryn, McGuire, Abigail Manson, Maiga, Mamoudou, Abeel, Thomas, et al. (2016) 'Whole Genome Sequencing of Mycobacterium africanum Strains from Mali Provides Insights into the Mechanisms of Geographic Restriction' Small, P. L. C. (ed.). *PLoS Negl Trop Dis*, 10(1), p. e0004332.

433 Bold, Tyler D, Davis, Daphne C, Penberthy, Kristen K, Cox, Laura M, et al. (2012) 'Impaired fitness of Mycobacterium africanum despite secretion of ESAT-6.'. *J Infect Dis*, 205(6), pp. 984–990.

434 Asante-Poku, A, Yeboah-Manu, D, Otchere, I D, Aboagye, S Y, et al. (2015) 'Mycobacterium africanum is associated with patient ethnicity in Ghana'. *PLoS Negl Trop Dis*, 9(1), p. e3370.

435 Sharma, Aditya, Bloss, Emily, Heilig, Charles M and Click, Eleanor S (2016) 'Tuberculosis Caused by Mycobacterium africanum, United States, 2004–2013'. *Emerg Infect Dis*, 22(3), pp. 396–403.

436 Niobe-Eyangoh, S N, Kuaban, C, Sorlin, P, Cunin, P, et al. (2003) 'Genetic biodiversity of Mycobacterium tuberculosis complex strains from patients with pulmonary tuberculosis in Cameroon'. *J Clin Microbiol*, 41(6), pp. 2547–2553.

437 Thumamo, B P, Asuquo, A E, Abia-Bassey, L N, Lawson, L, et al. (2012) 'Molecular epidemiology and genetic diversity of Mycobacterium tuberculosis complex in the Cross River State, Nigeria'. *Infect Genet Evol*, 12(4), pp. 671–677.

438 Godreuil, S, Torrea, G, Terru, D, Chevenet, F, et al. (2007) 'First molecular epidemiology study of Mycobacterium tuberculosis in Burkina Faso'. *J Clin Microbiol*, 45(3), pp. 921–927.

439 Lawson, L, Zhang, J, Gomgnimbou, M K, Abdurrahman, S T, et al. (2012) 'A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria'. *PLoS One*, 7(6), p. e38409.

440 Yeboah-Manu, D, Asante-Poku, A, Bodmer, T, Stucki, D, et al. (2011) 'Genotypic diversity and drug susceptibility patterns among M. tuberculosis complex isolates from South-Western Ghana'. *PLoS One*, 6(7), p. e21906.

441 Yeboah-Manu, Dorothy, Asare, P, Asante-Poku, A, Otchere, I D, et al. (2016) 'Spatio-Temporal Distribution of Mycobacterium tuberculosis Complex Strains in Ghana.' Sreevatsan, S. (ed.). *PLoS One*, 11(8), p. e0161892.

442 Ejo, Mebrat, Gehre, Florian, Barry, Mamadou Dian, Sow, Oumou, et al.

(2015) ‘First insights into circulating Mycobacterium tuberculosis complex lineages and drug resistance in Guinea.’. *Infect Genet Evol*, 33, pp. 314–319.

443 Traore, B, Diarra, B, Dembele, B P P, Somboro, A M, et al. (2012) ‘Molecular strain typing of <I>Mycobacterium tuberculosis</I> complex in Bamako, Mali’. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 16(7), pp. 911–916.

444 Park, Young Kil, Shin, Sonya, Ryu, Sungweon, Cho, Sang Nae, et al. (2005) ‘Comparison of drug resistance genotypes between Beijing and non-Beijing family strains of Mycobacterium tuberculosis in Korea.’. *Journal of microbiological methods*, 63(2), pp. 165–172.

445 Glynn, Judith R, Whiteley, Jennifer, Bifani, Pablo J, Kremer, Kristin and van Soolingen, Dick (2002) ‘Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review.’. *Emerg Infect Dis*, 8(8), pp. 843–849.

446 Ioerger, Thomas R, Feng, Yicheng, Chen, Xiaohua, Dobos, Karen M, et al. (2010) ‘The non-clonality of drug resistance in Beijing-genotype isolates of Mycobacterium tuberculosis from the Western Cape of South Africa.’. *BMC Genomics*, 11(1), p. 670.

447 Affolabi, D, Anyo, G, Faïhun, F, Sanoussi, N, et al. (2009) ‘First molecular epidemiological study of tuberculosis in Benin.’. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 13(3), pp. 317–322.

448 Diallo, A B, Walbang, O G and Camara, M (2016) ‘Molecular genotypes of Mycobacterium tuberculosis strains circulating in Dakar, Senegal’. *African Journal of ….*

449 Adesokan, H K, Jenkins, A O, van Soolingen, D and Cadmus, S I B (2012) ‘ Mycobacterium bovis infection in livestock workers in Ibadan, Nigeria: evidence of occupational exposure’. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 16(10), pp. 1388–1392.

450 Diallo, Mamadou, Diarra, Bassirou, Sanogo, Moumine, Togo, Antieme C G, et al. (2016) ‘Molecular identification of Mycobacterium bovis from cattle and human host in Mali: expanded genetic diversity’. *BMC Veterinary Research*, 12(1), p. 145.

451 Michel, A L, Muller, B and van Helden, P D (2010) ‘Mycobacterium bovis at the animal-human interface: a problem, or not?’. *Vet Microbiol*, 140(3-4), pp. 371–381.

452 Smith, Tasha, Wolff, Kerstin A and Nguyen, Liem (2012) ‘Molecular

Biology of Drug Resistance in Mycobacterium tuberculosis', in *Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism*, Current Topics in Microbiology and Immunology, Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 53–80.

453 Kendall, Emily A, Fofana, Mariam O and Dowdy, David W (2015) 'Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis'. *The Lancet Respiratory Medicine*, 3(12), pp. 963–972.

454 Clark, Taane G, Mallard, Kim, Coll, Francesc, Preston, Mark, et al. (2013) 'Elucidating Emergence and Transmission of Multidrug-Resistant Tuberculosis in Treatment Experienced Patients by Whole Genome Sequencing' Metcalfe, J. Z. (ed.). *PLoS One*, 8(12), p. e83012.

455 Gardy, Jennifer L, Johnston, James C, Sui, Shannan J Ho, Cook, Victoria J, et al. (2011) 'Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak'. *dx.doi.org*, 364(8), pp. 730–739.

456 Walker, Timothy M, Kohl, Thomas A, Omar, Shaheed V, Hedge, Jessica, et al. (2015) 'Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study'. *Lancet Infect Dis*, 15(10), pp. 1193–1202.

457 Köser, Claudio U, Bryant, Josephine M, Becq, Jennifer, Török, M Estee, et al. (2013) 'Whole-Genome Sequencing for Rapid Susceptibility Testing of M. tuberculosis'. *dx.doi.org*, 369(3), pp. 290–292.

458 Gardy, J L (2015) 'Towards genomic prediction of drug resistance in tuberculosis'. *Lancet Infect Dis*, 15(10), pp. 1124–1125.

459 Ahuja, Shama D, Ashkin, David, Avendano, Monika, Banerjee, Rita, et al. (2012) 'Multidrug Resistant Pulmonary Tuberculosis Treatment Regimens and Patient Outcomes: An Individual Patient Data Meta-analysis of 9,153 Patients' Evans, C. (ed.). *PLoS medicine*, 9(8), p. e1001300.

460 Falzon, Dennis, Mirzayev, Fuad, Wares, Fraser, Baena, Inés Garcia, et al. (2015) 'Multidrug-resistant tuberculosis around the world: what progress has been made?'. *European Respiratory Journal*, 45(1), pp. 150–160.

461 UNDP (2014) *The 2014 Human Development Report—Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience* UNDP (ed.),

462 World Health Organization (2015) *Global tuberculosis report 2015*, World Health Organization.

463 Abdurrahman, Saddiq Tsimiri, Emenyonu, Nnamdi, Obasanya, Olusegun Joshua, Lawson, Lovett, et al. (2014) 'The hidden costs of installing xpert machines in a tuberculosis high-burden country: experiences from Nigeria'. *Pan Afr Med J*, 18.

464 Musa, B M, Musa, B, Muhammed, H, Ibrahim, N and Musa, A G (2015) 'Incidence of tuberculosis and immunological profile of TB/HIV co-infected patients in Nigeria'. *Ann Thorac Med*, 10(3), pp. 185–192.

465 NIGERIA, FEDERAL REPUBLIC OF (2009) *2006 Population Census*,

466 CIA (n.d.) 'The World Factbook AFRICA::MALI'.

467 la santé et de lHygiene Publique, de, Ministère (2014) *Direction nationale de la sante- Rapport d'Activités du Programme National de Lutte contre la Tuberculose, 2013, Mali 2014*, Bamako.

468 Baker, L V, Brown, T J, Maxwell, O, Gibson, A L, et al. (2005) 'Molecular analysis of isoniazid-resistant Mycobacterium tuberculosis isolates from England and Wales reveals the phylogenetic significance of the ahpC -46A polymorphism.'. *Antimicrob Agents Chemother*, 49(4), pp. 1455–1464.

469 Agatha, Ani, Yetunde, Isah, Rosemary, Pwol, Chindak, Lekuk, et al. (2015) 'Detection of Mycobacterium tuberculosis by rapid molecular methods augments acid fast bacilli (AFB) smear microscopy in a non-culture tuberculosis laboratory'. *African Journal of Microbiology Research*, 9(13), pp. 960–964.

470 Diguimbaye, C, Hilty, M, Ngandolo, R, Mahamat, H H, et al. (2006) 'Molecular Characterization and Drug Resistance Testing of Mycobacterium tuberculosis Isolates from Chad'. *J Clin Microbiol*, 44(4), pp. 1575–1577.

471 Glynn, Judith R, Whiteley, Jennifer, Bifani, Pablo J, Kremer, Kristin and van Soolingen, Dick (2002) 'Worldwide Occurrence of Beijing/W Strains of Mycobacterium tuberculosis: A Systematic Review'. *Emerg Infect Dis*, 8(8), pp. 843–849.

472 Diarra, Bassirou, Siddiqui, Sophia, Sogoba, Dramane, Traoré, Bréhima, et al. (2010) 'Mycobacterium tuberculosisBeijing Strain, Bamako, Mali'. *Emerg Infect Dis*, 16(2), pp. 361–363.

473 van der Spuy, G D, Kremer, K, Ndabambi, S L, Beyers, N, et al. (2009) 'Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics.'. *Tuberculosis (Edinb)*, 89(2), pp. 120–125.

474 Cowley, Donna, Govender, Dhirendra, February, Bradley, Wolfe, Mary, et al. (2008) 'Recent and Rapid Emergence of W-Beijing Strains of Mycobacterium tuberculosisin Cape Town, South Africa'. *Clinical*

Infectious Diseases, 47(10), pp. 1252–1259.

475 Eldholm, Vegard, Pettersson, John H-O, Brynildsrud, Ola B, Kitchen, Andrew, et al. (2016) 'Armed conflict and population displacement as drivers of the evolution and dispersal of Mycobacterium tuberculosis.'. *Proc Natl Acad Sci U S A*, p. 201611283.

476 Koro Koro, Francioli, Kamdem Simo, Yannick, Piam, Félix Fotso, Noeske, Jurgen, et al. (2013) 'Population dynamics of tuberculous Bacilli in Cameroon as assessed by spoligotyping.'. *J Clin Microbiol*, 51(1), pp. 299–302.

477 Asante-Poku, A, Yeboah-Manu, D, Otchere, I D, Aboagye, S Y, et al. (2015) 'Mycobacterium africanum is associated with patient ethnicity in Ghana'. *PLoS Negl Trop Dis*, 9(1), p. e3370.

478 Campbell, P J, Morlock, G P, Sikes, R D, Dalton, T L, et al. (2011) 'Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis'. *Antimicrob Agents Chemother*, 55(5), pp. 2032–2041.

479 Cohen, Ted, Jenkins, Helen E, Lu, Chunling, McLaughlin, Megan, et al. (2014) 'On the spread and control of MDR-TB epidemics: An examination of trends in anti-tuberculosis drug resistance surveillance data'. *Drug Resistance Updates*, 17(4), pp. 105–123.

480 Drobniewski, Francis, Nikolayevskyy, Vladyslav, Maxeiner, Horst, Balabanova, Yanina, et al. (2013) 'Rapid diagnostics of tuberculosis and drug resistance in the industrialized world: clinical and public health benefits and barriers to implementation'. *BMC Medicine*, 11(1), p. S15.

481 Peacock, Sharon J and Weinstock, George M (2014) 'Microbial sequencing to improve individual and population health.'. *Genome medicine*, 6(11), p. 103.

482 Walker, Timothy M, Kohl, Thomas A, Omar, Shaheed V, Hedge, Jessica, et al. (2015) 'Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study'. *Lancet Infect Dis*, 15(10), pp. 1193–1202.

483 Roetzer, Andreas, Diel, Roland, Kohl, Thomas A, Rückert, Christian, et al. (2013) 'Whole Genome Sequencing versus Traditional Genotyping for Investigation of a Mycobacterium tuberculosis Outbreak: A Longitudinal Molecular Epidemiological Study' Neyrolles, O. (ed.). *PLoS medicine*, 10(2), p. e1001387.

484 Smyth, Davida S, McDougal, Linda K, Gran, Frode W, Manoharan, Anand, et al. (2010) 'Population Structure of a Hybrid Clonal Group of Methicillin-Resistant Staphylococcus aureus, ST239-MRSA-III' DeLeo,

F. R. (ed.). *PLoS One*, 5(1), p. e8582.

485 Spoor, Laura E, Richardson, Emily, Richards, Amy C, Wilson, Gillian J, et al. (2015) 'Recombination-mediated remodelling of host–pathogen interactions during Staphylococcus aureus niche adaptation'. *Microbial Genomics*, 1(4).

486 European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis (2006) 'Beijing/W Genotype Mycobacterium tuberculosisand Drug Resistance'. *Emerg Infect Dis*, 12(5), pp. 736–743.

487 Casali, Nicola, Nikolayevskyy, Vladyslav, Balabanova, Yanina, Harris, Simon R, et al. (2014) 'Evolution and transmission of drug-resistant tuberculosis in a Russian population.'. *Nat Genet*, 46(3), pp. 279–286.