Evolution of *Streptococcus pneumoniae* during carriage

Kamolchanok Claire Chewapreecha

Magdalene College

University of Cambridge

August 2014

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

I hereby declare that this dissertation is my own work and contains nothing that is the outcome of work done in collaboration with others, except where specifically indicated at the beginning of each chapter.

All sampling, population survey and microbiology work, which contributed to the metadata of this study were performed by Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University in Thailand through a collaboration with Dr Paul and Claudia Turner. The sequence data used in this thesis was generated at the Wellcome Trust Sanger Institute by Research Development and Sequencing production teams.

None of the work presented here has previously submitted for the purpose of obtaining another degree. This dissertation does not exceed 60,000 words in length, as required by the School of Biological Sciences.

Kamolchanok Claire Chewapreecha

August 2014

Acknowledgements

I am very grateful for Julian Parkhill and Stephen Bentley for trusting and giving me an opportunity to perform this work. I would like to thank both Julian and Ste for their strong support throughout my PhD, offering insightful and invaluable advice, and keeping me going when times were tough. Also, I sincerely thank Sharon Peacock and Matthew Holden, who also help guide me from day one.

I would like to extend my gratitude to my collaborators Paul and Claudia Turner, physicians who spent their PhD setting up a clinic in a remote Maela refugee camp to provide medical services to local people. I am in debt of their hard works and the resources they collected, which constitute almost this entire thesis.

Much of my bioinformatics work would not be possible without guidance and patience of Simon Harris and Nicholas Croucher. I also thank Jukka Corrander, Pekka Martinen, and their team for welcoming the computational challenges and helping develop analysing tools implemented on this study; David Aanensen for his help with graphical visualisation; Alison Mather for her statistical advice; and Susannah Salters for her microbiological expertise. This work could not have been completed without the informatics, systems, sequencing and library making teams of the Sanger institute. Also, I am thankful for all members of team 81, especially E212 residents for their general supports, understanding, and continuous cake supplies.

I thank Joy, all my friends and tutors for being with me through good and bad times during these nine years at Cambridge and yet making it so memorable. A thank to P'Yod whom the burden of mentoring me during difficult times has fallen most heavily. A final thanks to my family mum, dad, grandad, and late grandma for being so supportive and understanding throughout my time studying abroad, and always welcoming me back home with Thai food and warm hugs.

My PhD was funded by a Royal Thai Government Scholarship and a Wellcome Trust PhD Studentship.

Abstract

Streptococcus pneumoniae is a commensal bacterium asymptomatically carried in the nasopharynx of healthy individuals. However, if the bacterium escapes from its natural habitat to other anatomical loci, it can cause a range of invasive pneumococcal diseases, which make it a killer of over one million children annually. Despite high casualties, both treatment and prevention through vaccines have become more difficult as the bacteria rapidly develop antibiotic resistance and vaccine escape serotypes. To understand how this happens, one needs to look at evolution during carriage, a phase where exchange of genetic determinants for antibiotic resistance, virulence, and vaccine escape occurs *via* the process called "recombination".

This thesis summarises findings from a collection of 3,085 genome sequences of pneumococcal isolates from a rural community in Thailand called "Maela". This highly dense sampling gave an opportunity to investigate patterns of recombination and gene flows within the population, as well as changes in evolutionary patterns according to changes in selection pressure, especially the use of antibiotics over time. The non-encapsulated isolates, which are less invasive and unaffected by currently licensed vaccines, have a higher rate of both acceptance and donation of DNA *via* homologous recombination than encapsulated pneumococci. Highly exchanged genes include those associated with antibiotic resistance, implying that the non-encapsulates may act as a reservoir of resistance that can be passed to pathogenic strains and thus enhance the threat posed by antibiotic resistance.

However, the view from the Maela community may not be directly applicable to the population elsewhere, as different population structures may result in a different capacity for adaption. I therefore compared pneumococcal lineages detected in Maela with other contemporaneous carriage collections from the USA, UK, Gambia and Kenya based on multilocus sequence typing. The results showed that while the USA and UK share a lot of common lineages, large proportions of pneumococci detected in Gambia, Kenya and Thailand are unique to each location. Therefore, the propensity for genetic exchange may vary geographically and temporally.

The next part of the thesis identifies genetic determinants of resistance to beta-

lactams, a group of antibiotics frequently prescribed for upper respiratory infections. Here I performed a genome-wide association study - a technique commonly used in human genetics but difficult in bacteria due to their clonal population structure. Nevertheless, the large sample size and highly recombinogenic nature of *S. pneumoniae* allowed me to identify potential sources of resistance with improved resolution from "mosaic" genes described in the literature to several discrete causative sites, some of which are novel. The non-uniform distribution of these alleles in both vaccine-targeted and non-vaccine targeted lineages also highlights the limitations of vaccine in the control of spread of antibiotic resistance.

Together, this snapshot of the evolution of pneumococci and their interactions during carriage highlights the speed at which *S. pneumoniae* can adapt to new challenges, including antibiotics, while informing limitations in current health control policy.

Table of Contents

List of Figures

List of Tables

Abbreviations

ANCOVA	Analysis of Covariance
BAPS	Bayesian Analysis of Population Structure
BC(s)	primary BAPS Cluster(s)
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BLAST	Basic Local Alignment Search Tool
BURST	Based Upon Related Sequence Types
BratNextGen	Bayesian recombination tracker Next Generation
CC(s)	Clonal complex(es)
CLSI	Clinical and Laboratory Standard Institute
cps	capsular polysaccharide synthesis locus
CSP	Competence stimulating peptide
DNA	Deoxyribonucleic acid
DVL(s)	double-locus variant(s)
ENA	European Nucleotide Archive
ICE	Integrative conjugated element
IPD	Invasive pneumococcal diseases
MLEE	Multilocus enzyme electrophoresis
MLST	Multilocus sequence typing
MMR	Mismatch repair system
NT(s)	Nontypable
PCV	Pneumococcal conjugate vaccine
PFGE	Pulse field gel electrophoresis
PMEN	Pneumococcal Molecular Epidemiology Network
sBC	secondary BAPS cluster
SMRU	Shoklo Malaria Research Unit
SNP(s)	Single nucleotide polymorphism(s)
ST(s)	sequence type(s)
SVL(s)	single-locus variant(s)
TVL(s)	triple-locus variant(s)
WGS	Whole genome sequencing

Chapter 1: Introduction

1.1 Identification and characterisation of *Streptococcus pneumoniae*

1.1.1 A brief history

1.1.2 How pneumococci are characterised ?

1.1.2.1 Capsular typing

1.1.2.2 Multi-locus typing

1.1.2.3 Whole genome sequencing

1.1.2.4 Typing methods are affected by recombination

1.2 The pneumococcus has a highly recombinogenic nature

1.2.1 Mechanism of recombination

1.2.2 Early observations of pneumococcal recombination

1.2.3 A higher resolution of pneumococcal recombination from whole genome sequencing

1.3 The pneumococcus in carriage

1.3.1 Prevalence and duration of carriage

1.3.2 Interactions between pneumococci and other bacterial species

1.3.2.1 Interactions between pneumococci

1.3.2.2 Interactions between pneumococcus and other species

1.4 The pneumococcus in disease

1.4.1 Morbidity and mortality

1.4.2 Bacterial progression from carriage to disease

1.4.3 Factors influencing the transformation to disease

1.4.4 Limited genetic interactions in disease compared to carriage

1.5 Natural and clinical mechanisms for pneumococcal elimination and how the pneumococcus evolves to evade them

1.5.1 Clearance through natural host immune systems

1.5.2 Clinical interventions

1.5.2.1 Vaccines

1.5.2.2 Antibiotics

1.6 Project aims and objectives