

**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 asymptotically 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

1. Introduction	2
1.1 <i>Identification and characterisation of Streptococcus pneumoniae</i>	2
1.1.1 A brief history	2
1.1.2 How pneumococci are characterised ?	3
1.2 <i>The pneumococci have a highly recombinogenic nature</i>	12
1.2.1 Mechanism of recombination.....	12
1.2.2 Early observations of pneumococcal recombination	15
1.2.3 A higher resolution of pneumococcal recombination from whole genome sequencing.....	16
1.3 <i>The pneumococci in carriage</i>	19
1.3.1 Prevalence and duration of carriage.....	19
1.3.2 Interactions between pneumococci and other bacterial species in carriage..	20
1.4 <i>The pneumococci in disease</i>	22
1.4.1 Morbidity and mortality	22
1.4.2 Bacterial progression from carriage to disease	23
1.4.3 Factors influencing the transformation to diseases.....	24
1.4.4 Limited genetic interactions in diseases compared to carriage.....	26
1.5 <i>Natural and clinical mechanisms for pneumococcal elimination and how the pneumococcus evolves to evade them</i>	27
1.5.1 Clearance through natural host immune systems	27
1.5.2 Clinical interventions.....	28
1.5 <i>Project aims and objectives</i>	32
2. Materials and methods	35
2.1 <i>Pneumococcal collections</i>	35
2.1.1 Maela whole genome sequencing collection	35
2.1.2 PMEN14 whole genome sequencing collection.....	36
2.1.3 Other global MLST collections	36
2.2 <i>Whole-genome sequencing</i>	37
2.3 <i>Control for sample mix up through determination of serotype and sequence type</i>	38
2.4 <i>Sequence assembly</i>	38

2.5 Sequence mapping	39
2.6 Visualisation of phylogenetic trees	42
2.7 Statistical analyses.....	42
3. The Maela and global pneumococcal population structure.....	44
3.1 Introduction and aims	44
3.2 Methods.....	46
3.2.1 Estimating Maela population population structure	46
3.2.2 Estimating global pneumococcal population structure.....	47
3.2.3 Serotype switches	47
3.3 Results	48
3.3.1 Maela pneumococcal population structure.....	48
3.3.2 A snapshot of global pneumococcal population structure	55
3.3.3 Multiple introductions of a globally spread lineage to a local community.....	61
3.4 Conclusion	65
4. Maela pneumococcal evolution and population-wide sequence exchange.....	67
4.1 Introduction and aims	67
4.2 Methods.....	68
4.2.1 Estimating lineage-specific evolutionary parameters.....	68
4.2.2 Tracing genetic exchanges through homologous recombination	71
4.3 Results	72
4.3.1 Estimating evolutionary rates within the population	72
4.3.2 Population-wide sequence exchange.....	88
4.4 Conclusion	103
5. Recombination allows rapid adaptation in response to local selective pressure	105
5.1 Introduction and aims	105
5.2 Methods.....	106
5.2.1 Preparation of nucleotide sequences for penicillin-binding proteins (<i>pbps</i>), dihydrofolate reductase (<i>dhfR</i>) and dihydropterpate synthase (<i>folP</i>).....	106
5.2.2 Phylogenetic analyses	106
5.2.3 Statistical tests.....	106
5.3 Results	107
5.3.1 Biological relevance of sequences that have undergone recombination.....	107
5.3.2 Changes in recombination trends reflect changes in selection pressure.....	115

5.4 Conclusion	124
6. Genome-wide association study identifies single nucleotide polymorphic changes associated with beta-lactam resistance	127
6.1 Introduction and aims	127
6.2 Methods.....	129
6.2.1 Subject populations	129
6.2.2 Genotype callings and quality control.....	129
6.2.3 Phenotype information.....	129
6.2.4 Determining the cut-off threshold	130
6.2.5 Case-control analysis	130
6.2.6 Linkage analysis.....	130
6.2.7 Estimation of percentage of resistance in the population explained by candidate loci	131
6.2.8 Specificity to different classes of beta-lactams.....	131
6.2.9 Prevalence of candidate loci in the population	131
6.3 Results	132
6.3.1 Identification of loci associated with beta-lactam non-susceptibility.....	132
6.3.2 Biological relevance of candidate loci	146
6.3.3 Beta-lactam specificity of resistance mutations	149
6.3.4 Distribution of candidate alleles in the Maela and Massachusetts populations	152
6.4 Conclusion	154
7. Conclusions and future directions.....	157
7.1 Biological summary	158
7.1.1 Views from Maela data.....	158
7.1.2 Applications of views from Maela to other global collections	160
7.2 Methodological summary.....	163
7.2.1 Divide and conquer approach.....	163
7.2.2 Genome-wide association study	164
7.3 Future directions.....	164
7.3.1 Pneumococcal transmission	164
7.3.2 Bacterial-host interactions.....	165
7.4 Publications resulting from this thesis.....	165
8. References	167

9. Appendices.....181

List of Figures

Figure 1.1 Effect of recombination on pneumococcal typing

Figure 3.1 Maela pneumococcal population structure

Figure 3.2 Proportion of pneumococcal population commonly observed in multiple locations

Figure 3.3 Pairwise comparisons of similarities and differences in pneumococci detected between different locations

Figure 3.4 Phylogenetic analysis of Maela pneumococci in comparison to global PMEN-14

Figure 4.1 Nucleotide substitution based phylogeny and the clusters from which the nucleotide substitution rates were estimated

Figure 4.2 Demonstration that clock-like signals can be detected from the subclades but not from the whole population

Figure 4.3 Clock-like signals from Path-O-Gen in the subclades where substitution rates were estimated

Figure 4.4 Recombinations per mutation (r/m) of each cluster calculated by linear regression

Figure 4.5 Comparison of evolutionary parameters estimated in dominant clusters

Figure 4.6 Comparison of two recombination detection methods

Figure 4.7 Query length and search specificity

Figure 4.8 Multiple potential donors for a single recipient

Figure 4.9 Trends in genetic exchange

Figure 5.1 Recombination hotspots

Figure 5.2 Association between recombining *pbp* genes and resistance phenotypes

Figure 5.4 Association between recombining *fol* genes and resistant phenotypes.

Figure 6.1 Randomised control for intrinsic noise based on genetic variation alone

Figure 6.2 Summary of the genome-wide association study conducted in two separate datasets

Figure 6.3 Summary of single nucleotide polymorphisms (SNPs) associated with beta-lactam non-susceptibility

Figure 6.4 Linkage analysis for SNPs co-detected in two separate datasets

Figure 6.5 Summary of physical linkage structure in two separate datasets

Figure 6.6 Percentage of the non-susceptible phenotype explained by co-detected loci in the Maela and Massachusetts populations

Figure 6.7 Specificity of association signals for co-detected candidate loci with different classes of beta-lactam antibiotics

Figure 6.8 Frequency of putative resistance alleles from candidate loci in the Maela and Massachusetts data

List of Tables

Table 2.1 Other pneumococcal carriage collections used in the studies

Table 2.2 References used for mapping and mapping coverage generated for each dominant cluster

Table 3.1 Distribution of non-typable serotype (NT) in Maela

Table 3.2 Diversity captured through MLST in each sampling collection

Table 4.1 Nucleotide substitution rates estimated by BEAST

Table 4.2 Recombination per mutation (r/m) calculated from linear regression and arithmetic mean

Table 4.3 Numbers of recombination events used for the search

Table 4.4 Comparison of two recombination detection methods as given by Figure 4.6

Table 4.5 Distribution of length of recipient blocks described in Figure 4.7 a

Table 4.6 Potential donors for each recombinant fragment detected in isolate SMRU1452

Table 5.1 Recombination signals have been refined through time

Table 5.2 Trend in antibiotic consumption based on the Burmese border guidelines (1994-2010)

Table 5.3 Association between recombination, resistance phenotypes and temporal changes in recombination from seven dominant clusters

Table 6.1 Co-occurrence of co-trimoxazole and beta-lactam resistance phenotypes

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