Chapter 7: Conclusion and future directions

7. Conclusions and future directions

In summary, this thesis has described diversity in pneumococcal populations, characterised evolutionary rates and genetic exchanges, and identified genetic determinants contributing to antibiotic resistance in isolates from healthy carriage in the human nasopharynx. Carriage is a prerequisite for the development of pneumococcal invasive diseases (Bogaert, De Groot *et al.* 2004). It is also the phase where evolution shapes the wider population structure and, thereby, the prevalence of susceptibility to clinical interventions such as antibiotics and vaccines (O'Brien and Santosham 2004, Dagan and Klugman 2008). Due to the high level of genomic plasticity, the pneumococci have rapidly developed antibiotic resistance. This area has been under intense focus since the late 1960s, when the first resistant pneumococcal isolates were identified (Klugman 1990, Appelbaum 1992, Crook and Spratt 1998, Hanage, Fraser *et al.* 2009, Donkor, Bishop *et al.* 2011). Moreover, the genomic plasticity also led to capsular switching, creating new variants that are not targeted by vaccines. The switches were observed throughout the history of pneumococci (Moore, Gertz *et al.* 2008, Donati, Hiller *et al.* 2010, Wyres, Lambertsen *et al.* 2013, Croucher, Finkelstein *et al.* 2013) and have been under recent study to evaluate vaccine efficacy. These studies have provided greater understanding on pneumococcal evolution and development of antibiotic resistance and vaccine escape serotypes emerging from carriage. I hope this thesis helps add a few pieces to the jigsaw of our current picture of pneumococcal populations.

This thesis presents an unprecedented density of sampling of over 3,000 samples collected over a 3-year period, from a densely populated area of 2.4 km^2 allowed an opportunity to capture the exchange of genetic materials. Moreover, the relatively large sample size (Wyres, Conway *et al.* 2014) allowed robust statistics for measuring lineage-specific evolutionary patterns and performing genome-wide association studies which had previously been difficult in bacteria. This chapter summarises the findings from all chapters, their applications elsewhere and directions for future work.

7.1 Biological summary

7.1.1 Views from Maela data

7.1.1.1 Recombination allows rapid adaptation in response to environmental changes

The role of recombination in mediating sequence exchange and allowing the pneumococci to adapt in response to clinical interventions has long been recognised (Coffey, Dowson *et al.* 1995, Lipsitch 2001, Hanage, Fraser *et al.* 2009, Croucher, Harris *et al.* 2011). This role has been re-emphasised here with genomic evidence given in Chapter 5. Consistent with previous studies, the chapter demonstrated that the most frequently exchanged genes were those associated with antibiotic resistance and immune selection, with the former being sensitive to the levels of antibiotic consumption. This thesis additionally captured bacterial response to temporal changes in the consumptions of two types of antibiotics; beta-lactams and co-trimoxazole. A reduction in consumption of co-trimoxazole over time was manifest in the patterns of recombination, with more recent events having a weaker association with resistance. Vice-versa, a continuing high beta-lactam consumption was consistent with a higher resistance observed in recombining strains in both recent and older recombination events. This demonstrates the role of recombination in enabling the bacteria to adjust to temporal fluctuation in addition to its role in spatial differentiation, as reported previously (Shapiro, Friedman *et al.* 2012).

Importantly, this temporal change in antibiotic consumptions (2002) began several years prior to the Maela pneumococcal carriage study (2007-2010), yet the data was used able to identify evolutionary events from years prior to the sampling time. This highlights the ability to identify temporal changes in selection at loci linked to fluctuating selective pressures, and might allow a prediction of behaviour of this pathogen in response to changes in future. An encouraging prediction based on evidences described here is that a reduction in the use of antibiotic may lead to a decrease of a drug resistant population (in this case for co-trimoxazole), if there is no influx of resistant populations from outside. Evidence given here helps support the models of the relationship between antibiotic consumption and the frequency of

resistance in communities first put forward in 1999 (Austin, Kristinsson and Anderson 1999).

7.1.1.2 Behaviour of nontypable pneumococci

The observation of lineage specific variation in rates of recombination, both for donation and receipt of DNA, implies differential rates of response to environmental selection pressures between lineages. An elevated rate of acquisition and donation of recombinant DNA was observed in nontypable cluster BC3-NT (Chapter 4), suggesting a higher capacity for adaptation in nontyable isolates. The observation is consistent with the general idea that capsule could act as a barrier for DNA uptake. Though increased recombination could bring transient benefits, there are potential long-term disadvantages due to increasing genomic instability (Giraud, Matic *et al.* 2001). As noted in Chapter 3, sporadic switches between the NT and encapsulated states were observed. This may serve as a mechanism to modulate the trade-off between benefit and cost of increased recombination rates. Though recombination efficiency is promoted by the non-encapsulated status, other factors should also be considered. Molecular mechanisms limiting the acquisition of foreign DNA, like such as restriction modification systems, haves been described (Johnston, Martin *et al.* 2013). Also the genomic context, including sequence similarity and repeat elements which potentially promote strand exchanges, have been investigated in (Hiller, Ahmed *et al.* 2010) and (Croucher, Harris *et al.* 2012) (see introduction 1.2.1). These factors may also influence the recombinogenic behaviour of each pneumococcal lineage but were not fully investigated here.

7.1.1.3 A potential role of nontypables as a genetic reservoir

The elevated level of receipt and donation of recombinant DNA observed in Maela NT lineages suggests their role as a hub for genetic exchange in the Maela population. Moreover, a heightened level of beta-lactam resistant determinants harboured by BC3-NT, and other NT clusters (Chapter 6) further supports their role as a reservoir of antibiotic resistant genes, which may allows the pneumococcal population the potential to adapt to antibiotics more rapidly.

As many of the vaccine targeted serotypes are associated with antibiotic resistance, it was hoped that the removal of, or reduction in, such serotypes would reduce the pool of resistance alleles and hence decrease pneumococcal resistance to antibiotics (Dagan and Klugman 2008). Although resistance-encoding alleles might be passed from lineages that escape vaccines as a consequence of serotype switching, there is also a possibility that resistant alleles might come from highly recombinogenic lineages, including those that are not currently targeted by available vaccines such as the NT.

As NT lineages are more difficult to detect through conventional serotyping schemes and are less likely to cause invasive pneumococcal diseases, this group of pneumococci has received less attention (Hathaway, Stutzmann Meier *et al.* 2004). However, NTs are common in carriage, making up the majority of isolates collected in Maela camp (Chapter 3). Using MLST typing, (Hanage, Kaijalainen *et al.* 2006) showed that the NT lineages have been transmitted inter-continentally and have been included in recent Pneumococcal Molecular Epidemiology Network analyses as PMEN42 (ST344) and PMEN43 (ST448). This intercontinental spread may reflect an advantage in transmission and these lineages have been associated with outbreaks of conjunctivitis (Hanage, Kaijalainen *et al.* 2006).

This thesis has highlighted several characteristics of NT pneumococci detected in Maela. Given their high recombination tendency in Chapter 4, their predominance in carriage in Chapter 3, the prevalence of genetic determinants for antibiotic resistance detected in these lineages in Chapter 6, the evidence for successful inter-continental transmission and the fact that they are not currently targeted by vaccines, this thesis proposes that the NT could potentially be a genetic reservoir for all pneumococci, allowing harmful traits such as antibiotic resistance to circulate in the population.

7.1.2 Applications of views from Maela to other global collections

The capacity and speed of adaption mediated by homologous recombination seen in Maela may vary between different localities due to distinct population structures and the prevalence of NT pneumococci observed at each location.

7.1.2.1 Distinct population structures may have a distinct adaptive capacity

Chapter 3 presents the comparison of population structures between different locations: UK, US, Kenya, Gambia and Thailand, through MLST profiles collected between 2006-2010. Except for a high similarity between UK and US population structures, the results revealed only a small overlap in genotypes for each population despite the close timeframes of sampling. This small overlap is made up of recognised pneumococcal lineages that have been particularly successful in spreading worldwide. However, the effects of the globally spread clones on local population structure are likely to be buffered by locally distinct lineages, which are present at a higher proportion in each area.

A more in-depth comparison of population structure was performed between the Maela and Massachusetts data (Croucher, Finkelstein *et al.* 2013) where whole genome sequences were available. Consistent with the comparison made by MLST, the Maela and Massachusetts population phylogenies revealed a little overlap in the prevalent genotypes for each population. Each population comprised a large diversity of distinct genotypes manifested as star-like phylogenies, comprised of large clusters of closely related strains interspersed with smaller, looser clusters of divergent isolates.

The difference in population structures observed here might have an impact on differential response to clinical interventions implemented in different locations. What determines this impact remains obscure but detailed comparison of genomic datasets across a wide range of locations might provide clearer insights into the parameters that govern the process.

7.1.2.2 Differences in prevalence of nontypables in different population settings

Based on our previous hypothesis that NT lineages could act as the hub of genetic exchanges, the different proportion of NT pneumococci observed at each location might imply that strong adaptive responses detected in Maela (Chapter 5) may not be the same in other locations. While NT pneumococci appeared to be highly prevalent in Maela, they were observed at a lower frequency in Massachusetts, US; Southampton, UK; The Gambia; and Kilifi, Kenya (Chapter 3). This discrepancy may be due to different laboratory techniques, which might ignore NT due to its atypical morphology (Rolo, A *et al.* 2013) and subsequently lead to different numbers of NT being reported. However, laboratory protocols used in the two cohort studies: Maela, Thailand and Massachusetts, USA (Croucher, Finkelstein *et al.* 2013) were compared and neither showed a bias in NT detection; thereby confirming an actual lower prevalence of NT detected in Massachusetts compared to Maela (Personal communication with Dr Nicholas Croucher).

Although the NT population was confirmed to be low in Massachusetts, a resource-rich state of US, it has a high prevalence in poorer US communities including the Native American communities from Navajo and White Mountain Apache (Millar, O'Brien *et al.* 2009). In these Native American communities, the NT was the third largest serotype group, after 6A and 6B, detected in nasopharyngeal carriage of children < 6 years prior to vaccine introduction (Millar, O'Brien *et al.* 2009), and remained among the top 10 serotypes in all age strata post vaccine (Scott, Millar *et al.* 2012). Five percent (95% confidence interval (CI), 4.2% - 5.7%) of pneumococcal carriage isolates from Navajo and White Mountain Apache children collected between 2006-2008 were NT (Scott, Millar *et al.* 2012). In contrast, only 1.88% of carriage isolates from children aged < 7 years were NT in the study conducted in the Massachusetts during the same sampling time frame (Hanage, Huang *et al.* 2007). Based on multilocus sequence typing, the majority of STs detected in these Native American communities included ST344, ST448, ST1054, ST1186 and ST2011 (Scott, Hinds *et al.* 2012). The first two were members of PMEN clones: PMEN42 ($Norway^{NT}$-42) and PMEN43 (USA^{NT}-43) which have been reported globally, though with different prevalence.

It is unclear why NT pneumococci have a high prevalence in places like Maela refugee camp, Thailand and Native American communities from Navajo and White

Mountain Apache, US, but not in Massachusetts US. As the first two locations reported higher carriage rates than Massachusetts (Scott, Millar *et al.* 2012, Turner, Turner *et al.* 2012, Croucher, Finkelstein *et al.* 2013), it is possible that the non-encapsulated status allows some adaptive advantages in a densely colonised niche. A lack of capsule promotes greater bindings of pneumococcal surface attachment proteins to epithelia (Weiser, Austrian *et al.* 1994); thereby supporting their colonisation of the nasopharynx. Beyond this, little is known about factors governing the prevalence of NT pneumococci in different locations. Without densely sampling done in similar fashion to Maela, it will be difficult to predict the contributions of NT to genetic exchanges in different pneumococcal populations.

7.2 Methodological summary

In addition to pneumococcal biology, this thesis has documented some unprecedented technical challenges including the ability to handle large-scale data, and the application of genome-wide association approaches, which are common in human genetics, onto bacterial genomes.

7.2.1 Divide and conquer approach

With help from Professor Jukka Corander and his team, we showed that a rough population structure could be constructed from a large genomic data set by both the Bayesian clustering method (BAPS) and maximum likelihood phylogeny. However, the preference for large-scale analysis lies with BAPS as it operates much faster, while giving similar results when compared to phylogenetic tree generation (Chapter 3). By partitioning the diverse population into closely related clusters of smaller size, the data can be handled more effectively.

Population stratification is essential for genome-wide association studies (GWAS) (discussed in the next section) to determine whether the detected signals are due to true genetic associations or arise from shared common ancestry. BAPS allowed rapid estimation of the population structure in each the separate Maela and Massachusetts datasets. The tool has been further developed so it is now possible to perform BAPS

on the combined Maela and Massachusetts data and more, to determine their population structures at the same time. This will help reduce a potential bias introduced by running BAPS on the separate datasets, which might lead to more strict control over population stratification in one population than the other (discussed in Chapter 6).

7.2.2 Genome-wide association study

GWAS has been commonly used in human genetics to identify genetic loci associated with a particular trait. However, it has been difficult in bacteria due to the intrinsic clonal population structure, which hinders the differentiation of truly associated genetic variations from hitchhikers. Chapter 6 showed that it was possible to apply the technique on highly recombinogenic bacteria, where recombination has occurred frequently enough to disrupt the clonal structure. Moreover, the large sampling size used in this study: 3,085 isolates from Maela, and 616 isolates from Massachusetts cohorts, has allowed more robust statistics. This led to the identification of genetic determinants of antibiotic resistance down to single polymorphic changes or small loci. Apart from an application in *S. pneumoniae* as documented in this thesis, the method has been recently been applied to *Campylobacter* (Sheppard, Didelot *et al.* 2013) and *Staphylococcus aureus* (Laabei, Recker *et al.* 2014), suggesting GWAS as a promising avenue for identifying bacterial genes or genetic loci associated with traits such as antibiotic resistance, transmission and virulence.

7.3 Future directions

There are two potential areas that can be followed up from the works described here: first, transmission analysis; and second, bacterial components that would elicit host immune responses. The former may help control the spread of pneumococci in carriage and the latter might help identify suitable gene candidates for pneumococcal vaccines.

7.3.1 Pneumococcal transmission

Though it was not documented in this thesis, preliminary investigation showed that donors of recombinant fragments (Chapter 3) were largely observed to be strains colonising individuals living in the same household rather than those colonising individuals living in separate houses (Fisher's exact test, p value 2.2 x 10^{-16}). The observation implies high rates of household transmission, which is consistent with previous reports conducted elsewhere (Shimada, Yamanaka *et al.* 2002, Mosser, Grant *et al.* 2014). Whole genome sequencing is increasingly being used to investigate transmission and disease outbreaks caused by various bacterial pathogens (Bryant, Grogono *et al.* 2013, Peacock 2014, Price, Golubchik *et al.* 2014). It should be possible to adapt the techniques developed previously to track transmission in other bacteria for *S. pneumoniae,* although the high levels of recombination detected in this species will likely pose some challenges.

7.3.2 Bacterial-host interactions

Chapter 5 has captured the signals of selection pressure from host immunity that results in diversifying antigenic proteins, manifested here as recombination hotspots. The interactions between host immune systems and pneumococcal protein antigens are complex, and involve many players. However, given that information on i) whole genome sequences of pneumococcal strains colonising individuals and ii) measured antibody responses to pneumococcal proteins, are both available; one could apply the GWAS approach to identify both generic and lineage specific candidate loci that might elicit host responses. This would allows all signatures of diversifying selection by specific immune responses to be captured in the genomic data, beyond the classical dN/dS measures. Also, the allelic variation of antigens that have signals of diversifying selection could be demonstrated in *vitro* to confirm their selective advantage in the presence of antibodies targeting particular alleles.

7.4 Publications resulting from this thesis

Works described in Chapter 3, Chapter 4 and Chapter 5 form the publication

C. Chewapreecha, S. R. Harris, N. J. Croucher, C. Turner, P. Marttinen, L. Cheng, A. Pessia, D. M. Aanensen, A. E. Mather, A. J. Page, S. J. Salter, D. Harris, F. Nosten, D. Goldblatt, J. Corander, J. Parkhill, P. Turner and S. D. Bentley (2014). "Dense genomic sampling identifies highways of pneumococcal recombination." Nat Genet 46(3): 305-309.

Works described in Chapter 6 were published in

C. Chewapreecha, P. Marttinen, N. J. Croucher, S. J. Salter, S. R. Harris, A. E. Mather, W. P. Hanage, D. Goldblatt, F. H. Nosten, C. Turner, P. Turner, S. D. Bentley, and J. Parkhill (2014). “Comprehensive Identification of Single Nucleotide Polymophisms Associated with Beta-Iactam Resistance within Pneumococcal Mosaic Genes”. PLOS Genetics. DOI:10.1371/journal.pgen.1004547

Thank you so much for reading till the end of this thesis. I hope you enjoyed it.