9 Discussion

9.1 Pneumococcal transformation

9.1.1 Biases in the detection of recombination

The detection of recombination through genetic approaches requires the transfer of polymorphisms from the donor to the recipient; biochemical approaches, such as using isotopically labelled DNA, are required to overcome such limitations. This introduces an inherent degree of uncertainty into demarcating the extent of sequence transfers. At the simplest level, this can be quantified as the FRs in the work in Chapter 5: the position of RSS boundaries can only be measured to a level of precision corresponding to the density of SNPs in the region. This uncertainty is compounded by the mosaic nature of recombinations, alternating between recipient and donor segments: just as it cannot be assumed that the sequence between two neighbouring recipient allele polymorphisms is not affected by a small recombination that does not import any SNPs, within MRs it is not possible to tell whether there has been a reversion to recipient sequence in the distance between two polymorphisms converted to the respective donor alleles.

Nevertheless, the exponential distribution of RSS sizes detected in Chapter 5 suggests that a reasonable approximation of the underlying situation has been observed. Such a distribution of lengths suggests that very small events will be the most common, although there is likely to be a minimum enforced by the requirements for strand exchange on at least one side of the recombination. Unfortunately, the low sensitivity for detecting such short events places limits on the resolution at which this lower bound can be defined. The power to detect short recombinations is further reduced when the sequence of the donor bacterium is not known, as when examining the collection of PMEN1 and ST180 isolates (Chapters 4 and 8). Hence although the distribution of detected homologous recombination sizes is similar to those found in the *in vitro* work, there is a notable under representation of short events less likely to cause a statistically significant increase in SNP density. This leads to the differences in the estimated mean and median sizes of such events in the two studies. More

accurate estimates will require the improved precision achievable through the study of exchanges between more divergent bacteria, which would require good reference sequences to be generated for strains with appropriate selectable markers from other mitis group species.

As indicated by the distribution of RSS lengths spanning the selected locus in Chapter 5, studying a single genetic locus exacerbates the problem of overrepresentation of larger recombination events through 'length biased sampling' (Cox, 1969). Just as bigger events are more likely to contain polymorphisms that make them detectable, they are also more probable to affect any given subset of a chromosome, an issue only overcome by genome-wide analyses.

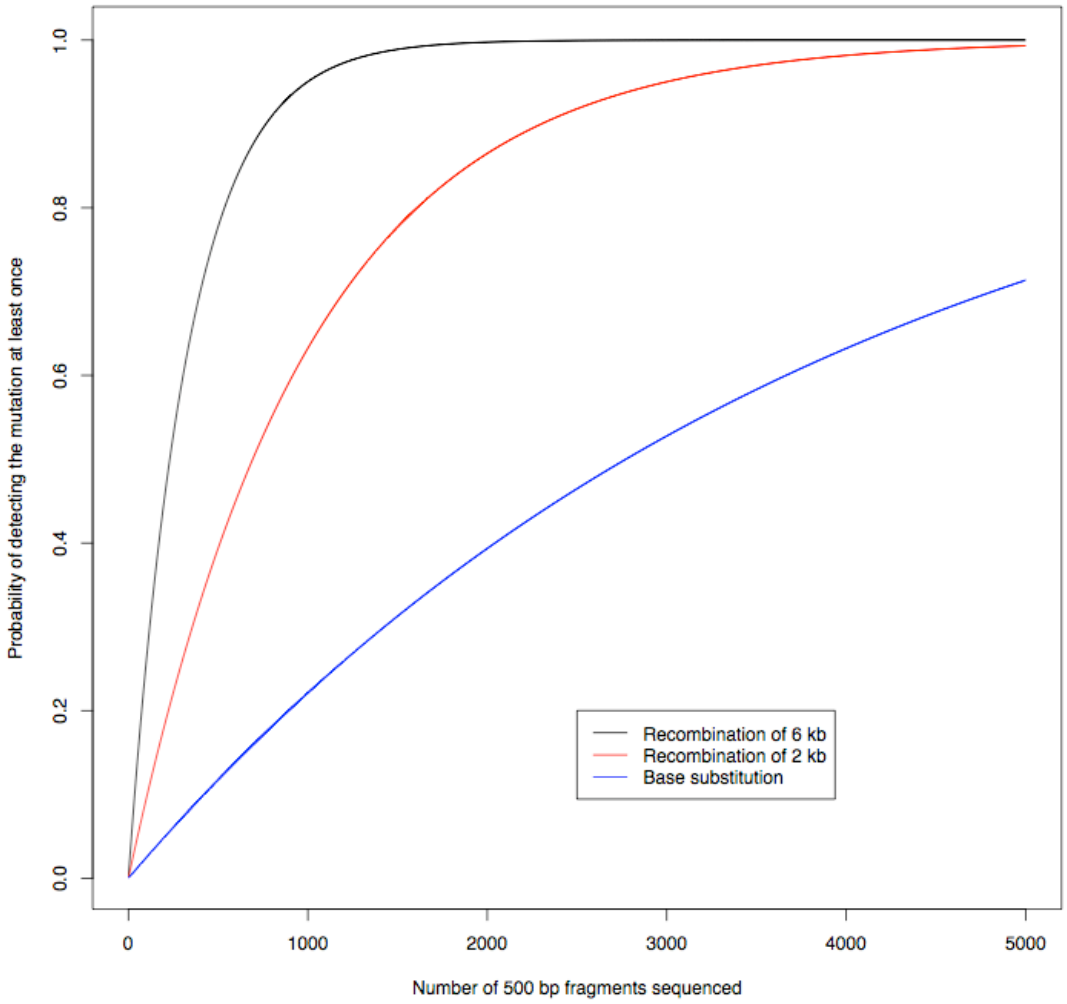

Figure 9.1 Modelling the ability to detect recombinations and mutations through sequencing multiple loci. The probability of detecting a point mutation or recombination at least once through the sequencing of n multiple loci is displayed for two different values of the recombination length, x (Equation 9.1). This shows how, the larger a mutation is, the more likely it is to be detected by incomplete sequencing of a genome. All three lines approach an asymptote of $p = 1$ as a greater proportion of the genome is sequenced.

9.1.2 Impact of biases on *r/m* estimates

This simple principle has an important consequence in resolving the discrepancy between the estimates of *r/m* for *S. pneumoniae* from the PMEN1 and ST180 analyses and the MLST database. By studying seven defined regions of the genome, MLST is more likely to detect larger mutations than smaller ones, because the greater the extent of a mutation, the more likely it is to overlap with one of the studied loci. Hence, given the same sample, the calculated *r/m* value will change as different proportions of the chromosome are sequenced. The effect can be quantified through a simple model. If a single point mutation and recombination of length *x* bp are present in a 2 Mb genome, then when sequencing a 500 bp locus there is a probability of (500 ÷ 2,000,000 =) 2.5×10^{-4} of detecting the point mutation, but a probability of approximately $(2.5 \times 10^{-4})x$ of identifying the recombination, depending on the density of SNPs throughout the event. Using values of *x* corresponding to 2 kb (close to the mean length from the transformation experiment) and 6 kb (close to the mean length estimated from the PMEN1 dataset), the probabilities of detecting a single point mutation (p_{mut}) and single recombination event (p_{rec}) at least once using data from *n* multiple loci can be estimated from a binomial distribution.

$$p_{mut} \sim Binom\left(n, \frac{1}{2000000}\right)$$
$$p_{rec} \sim Binom\left(n, \frac{x}{2000000}\right)$$

Equation 9.1

These outcomes of these models are displayed graphically in Figure 9.1. When only a few loci are sequenced, the bias towards the detection of larger recombinations is greatest. When $n = 7$ and $x = 6$ kb, the ratio *r/m* would be estimated as 11.9 rather than one. This approximates to the observed 9.2-fold and 25.4-fold difference between the *r/m* ratios observed for PMEN1 and ST180, respectively, and the MLST data (Feil *et al*., 2000). Hence the *r/m* statistics derived from whole and partial genome sequences are not easily comparable.

9.1.3 The advantage of being transformable

A number of different hypotheses have been proposed as the main function of the competence system in *S. pneumoniae* and other bacterial species. These fall into two categories: the advantages of generating genetic diversity, and the use of exogenous DNA as a source of metabolic substrates.

Hypotheses of the former category were originally formulated for sexually reproducing eukaryotic species. These suggest the main advantage of exchange is the acquisition of beneficial genetic material. The first model of this type, the Fisher-Muller hypothesis (Fisher, 1930; Muller, 1932), proposed that recombining populations evolve more quickly than asexual populations because advantageous alleles at different loci can accumulate into a single genotype more rapidly. However, this explanation was criticised for being formulated in terms of group selection (Maynard Smith, 1978), the validity of which is very doubtful (Maynard Smith, 1964; Williams, 1966). Subsequent descriptions have focussed on demonstrating advantages on the level of selection of individuals (Hill and Robertson, 1966; Felsenstein, 1974; Felsenstein and Yokoyama, 1976).

When applied to bacteria, this category of hypotheses is further subdivided. One set of extensions to this concept has focussed on the increased speed with which populations can respond to diversifying or balancing selection pressures (Michod *et al*., 2008; Vos, 2009). Given the nature of the regulation of the pneumococcal competence system, this has lead to the proposal that *S. pneumoniae* is adapted to diversify under stress, when it is maladapted to its environment (Claverys *et al*., 2006; Prudhomme *et al*., 2006). A second set of hypotheses suggest that transformation allows an improved response to purifying selection pressures: these focus on the possibility of recombination repairing genes afflicted by disruptive mutations (Michod *et al*., 2008; Vos, 2009), and thereby reversing Muller's ratchet (Muller, 1964; Felsenstein, 1974).

The main problem with such explanations is that the competence system, *a priori*, cannot distinguish beneficial and deleterious mutations, hence particular circumstances are required for transformation events to be of an overall selective advantage to an individual cell. For instance, the nature of epistatic interactions

between different loci around the chromosome has a bearing on the fitness of competence (de Visser and Elena, 2007). Positive epistasis, occurring when different advantageous alleles in a genotype adapt to one another, selects against transformation because mutually beneficial interactions are disrupted. Conversely, negative epistasis, involving deleterious loci reinforcing one another's negative impact, favours the evolution of transformation. Hence competence is beneficial when there is strong selection for the repair of deleterious mutations, while there are relatively small fitness costs when adaptive polymorphisms are lost. An alternative explanation is that the system can be biased towards the uptake of adaptive mutations through triggering competence in response to stress (Redfield, 1988). An increase in fitness following the acquisition of DNA is most likely when the host genotype has been damaged, or is maladapted to its environment, through the simple principle of regression to the mean. One criticism levelled at this model is that free DNA in the environment is most likely to have been released by cells that have already lysed, and hence this exogenous gene pool is likely to overrepresent genotypes poorly adapted to the environment (Redfield, 1988). A more direct response to genetic damage is the proposed triggering of transformation events at the site of lesions created by genotoxic stress (Hoelzer and Michod, 1991). However, this is not congruent with the observation that an increase in external DNA concentration can cause a rise in the amount of sequence imported without a commensurate increase in the level of host DNA damage (*e.g.* Chapter 5), suggesting only a small proportion of transformation events, if any, are directly instigated by genomic lesions.

Such couterpoints, and the observation that *H. influenzae* and *B. subtilis* do not regulate competence in response to DNA damage (Redfield, 1993a), lead to an alternative suggestion for the purpose of competence. Rather than a source of genetic information, imported DNA may serve as a source of nucleotides for DNA and RNA synthesis. Exogenous DNA is abundant in the nasopharynx, estimated to be present at a concentration of ~300 mg L^{-1} mucus (Matthews *et al*., 1963) and, in the case of the purine and pyrimidine auxotroph *H. influenzae*, transcription of the competence genes is repressed by an abundance of nucleotides in the growth medium (MacFadyen *et al*., 2001). However, several aspects of the competence machinery imply it is unlikely to have been optimised for the acquisition of nucleotides. Firstly, in the case of the

pneumococcal system, only one strand is imported through the pore while the other is degraded, making its constituents available to other cells (Dubnau, 1999; Johnsborg *et al*., 2007). Secondly, once inside the cell, the ssDNA is protected from degradation through the loading of ssDNA binding proteins, the binding and disassociation cycle of which requires ATP hydrolysis. Thirdly, species such as *H. influenzae* and *N. meningitidis* do not take up all DNA available to them, but rather selectively uptake homologous sequences containing distinctive repeats, hence limiting their ability to acquire nucleotides (Dubnau, 1999; Smith *et al*., 1999).

Whatever the primary function of the competence system, it must be sufficiently advantageous to counteract the inherent instability of such a mechanism. If transformation events only rarely prove to be beneficial, then the competence system will become defunct comparatively rapidly relative to purely metabolic genes. This results from an the asymmetric situation in which transformable bacteria are able to horizontally acquire mutations, or genes that inhibit transformation, that render the competence system non-functional, while non-competent cells cannot acquire sequences that cause them to revert to a competent state (Redfield, 1993b).

9.1.4 Inferences from PMEN1 and ST180

The 'hotspots' of recombination in the PMEN1 population, such as the capsule and proteinaceous antigens *pspA* and *pspC*, would appear to support the hypothesis that the competence system benefits the population through allowing it to respond to diversifying selection. However, this perception may be due to the bias of studying a single lineage: both *pspA* and *pspC* are very diverse genes within the species, hence are likely to be identified as an import each time they are transferred, whereas there is less sensitivity to detect transfer of other, more conserved, regions of the genome. Furthermore, it is likely that the acquired alleles of these antigens only appear diverse relative to the PMEN1 background and are actually sequences already frequent in the species. This is because transformation will sample alleles in proportion to their frequency in the population, thereby making such recombination events an inefficient mechanism for responding to balancing selection, which drives an increase in the frequency of rare alleles.

Furthermore, the mean rate of import from other lineages was approximately one event per three years in PMEN1 and one event per 30 years in ST180, although these rates were highly heterogeneous across the phylogenies. Analysis of MLST data from *H. influenzae* suggest that such horizontal transfer of divergent sequence is even rarer in that species (Feil *et al*., 2001). Hence it seems likely that transformation within a clonal population of bacteria, rather than between lineages, is likely to be the main selective pressure maintaining the competence system; this would suggest repair is a more important function than diversification. However, the relatively low recombination rate in ST180 does not appear to result in a particularly elevated rate of pseudogene formation: 6.7% of the CDSs in *S. pneumoniae* ATCC 700669 are pseudogenes, compared to 7.6% in *S. pneumoniae* OXC141, while 0.077 gene disruptions per SNP were observed in the PMEN1 sample, relative to 0.025 per SNP in ST180. However, this dissimilarity is likely to result from the differences in sequence data quality and indel identification, as well as the longer time period over which selection has been able to act to remove deleterious mutations from the ST180 lineage. Furthermore, the most frequent spontaneous mutations observed in *S. pneumoniae*, transitions and small indels, are efficiently repaired by the MMR system when imported in low numbers (Claverys *et al*., 1981; Lacks *et al*., 1982; Claverys *et al*., 1983), making this an ineffective use of the transformation machinery.

9.2 Site-specific and homologous recombination

9.2.1 The characteristics of horizontal sequence transfer

Transduction and conjugation, like transformation, both require the donor and recipient bacteria to co-colonise the same environment. However, the donor cell is usually lysed in order to release genomic DNA or phage particles for transformation and transduction respectively, whereas the donor cell remains intact following conjugation. Another trait of conjugation is that the donor may be distantly related to the recipient, given the wide host range of ICE (Chapter 3). By contrast, heterospecific transfers via phage or transformation are relatively rare (Chapter 5).

While the main evolutionary advantage of transformation is still debated, conjugation and phage infection are both driven by mobile genetic elements (MGEs), which, unless they carry selectively advantageous cargo, are likely to be detrimental to the host. This necessity of transmitting an entire MGE means that conjugation and phage infection events are large, spanning tens of kilobases, which contrasts with the observed range of transformation events, which have a median length of around 2.3 kb (Chapter 4). These differences suggest a novel hypothesis for the role of transformation in bacterial evolution.

9.2.2 Transformation as bacterial 'gene therapy'

As stated in Chapter 5, given the exponential distribution of RSS sizes and the apparent exclusion of donor insertions from such regions, it is evident that at any given polymorphic locus in a transformable pneumococcal population the smallest allele will transfer most quickly among the population and hence, in the absence of selection, will drift to fixation. Therefore transformation tends to reduce the size of the genome, as opposed to site-specific recombination, which inserts material directly into the chromosome. The insertions that are removed the most quickly will be those that are large and lack any selective advantage, the most obvious example of which are prophage and cryptic ICE. Hence it may be that homologous recombination serves primarily as a mechanism for removing parasites from the genome. Such DNA exchanges would be effective at eliminating any recently acquired parasite that was polymorphic within the population, allowing transformation to act as a post-integrative MGE defence mechanism.

Unlike the import of SNPs and small indels, where deleterious alleles are as likely to be acquired as beneficial ones, the system would be adapted to specifically importing short sequences in order to remove large insertions, and the transfers would be more efficient because long indels are much less frequently repaired through MMR (Claverys *et al*., 1980). Given this hypothesis, it may be expected that selection for transformation events at the prophage integration sites in the pneumococcal genome would lead to them being identified as 'hotspots' in the PMEN1 population. However, more detailed consideration suggests this is unlikely. Transformation will only be effective at removing prophage shortly after infection, when the element is

polymorphic in a clonally-related population; during this period, the co-colonising strain would almost certainly have been the origin of the prophage, hence could not donate non-lysogenic sequence that would lead to its removal. Only sequence donated by non-lysogenic members of the same clonally-related population could be used for repair, and these would not be identified as imports in the analysis outlined in Chapter 4. Once the prophage has been fixed through transmission bottlenecks, its subsequent elimination would be unlikely, as reinfection would be highly probable to undo any repair events.

Is MGE transmission within clonally-related bacteria a sufficiently strong selection pressure to maintain the competence system? The availability of prophage sequences within pneumococcal genomes permitted surveys of clinical isolates to be conducted; these have found that the majority of strains are lysogenic (Ramirez *et al*., 1999; Romero *et al*., 2009). When present at such a high level within the population, phage infections must be more frequent than imports of sequence from other lineage. While the physical distance to be traversed between pneumococcal communities is the same in each case, the phage is adapted for efficient and targeted transmission between pneumococci, protected against DNAse activity and is not competing for uptake with high concentrations of heterologous host DNA. One difference is that several restriction systems are present in the pneumococcal genome to protect against the acquisition of MGE as dsDNA, but will not act on ssDNA entering as a RecA- or DprA-coated nucleoprotein filament. One such system, displayed in Chapter 6, is the *hsdS* hypervariable restriction enzyme locus, which alters in its specificity over very short timescales; this seems likely to be importing in inhibiting the spread of recently acquired MGEs between clonally-related pneumococci, as this model proposes the competence system may do. The rapid alternation between forms means an MGE that was previously integrated into, and methylated as part of, a clonally related bacterium can be recognised and degraded by the cell, which would only be advantageous if the cell itself was not yet hosting the element itself.

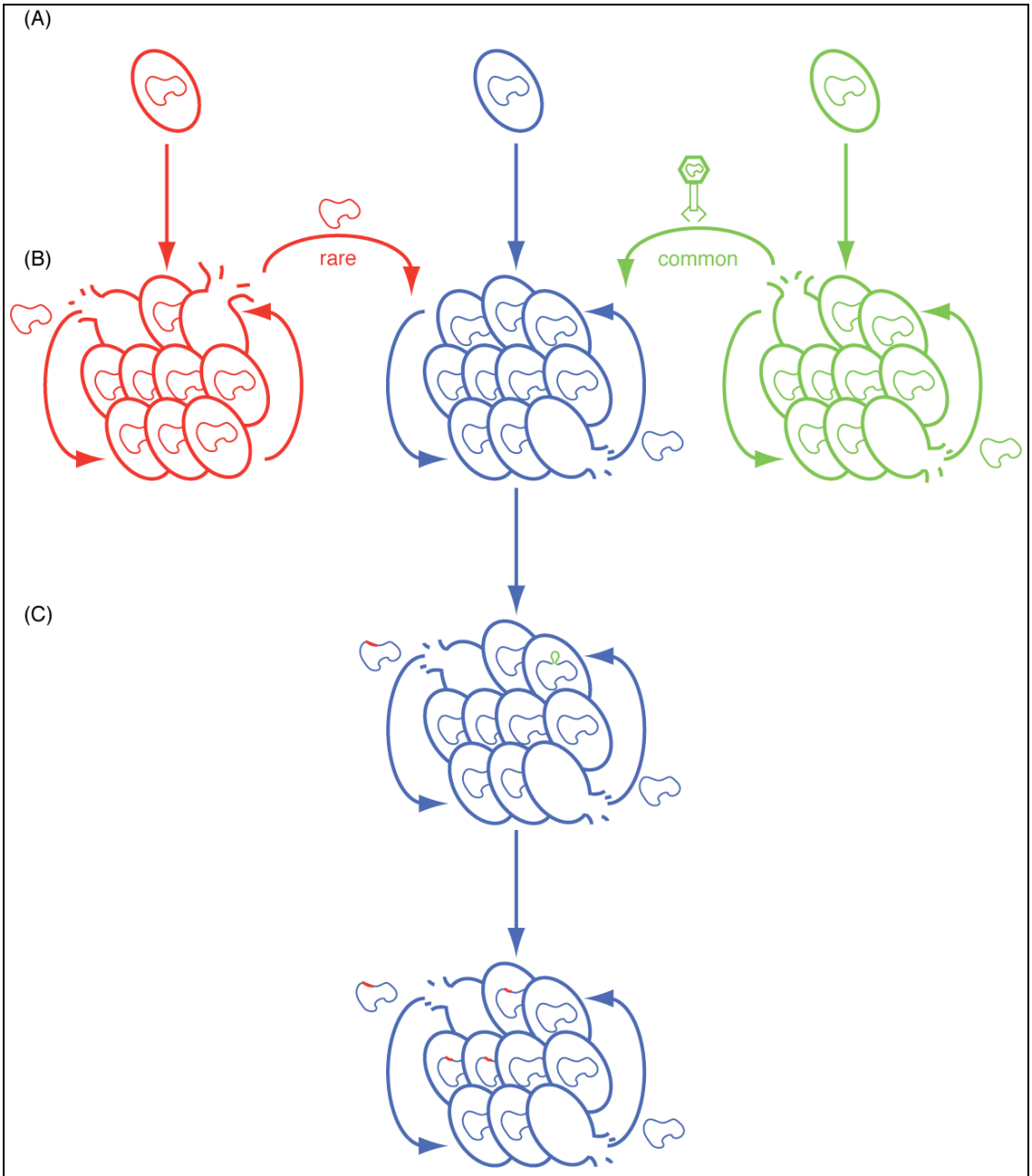

Figure 9.2 Model for the role of transformation in pneumococci. (A) Three different co-colonising *S. pneumoniae* genotypes are represented in red, green and blue, each of which gives rise to a spatially separated clonal population. (B) DNA exchange is rapid within each clonal population. MGEs, adapted to transmission between bacteria, are frequently passed between lineages. The movement of genomic DNA between different strains is likely to be much less common. (C) Once diversity has arisen in the originally clonal population through DNA import and MGE acquisition, its prevalence is determined by exchange within the population. Alleles of equal size are likely to be exchanged at equal rates, whereas the transfer of smaller alleles will be more rapid than that of larger alleles at the same locus, leading to elimination of recently acquired MGEs unless selection acts to retain them.

9.2.3 The interaction of mobile elements and transformation

A line of evidence that would support the hypothesis that recombination acts to remove mobile elements from the genome is the mechanisms that parasitic entities in

pneumococcal genomes appear to employ to counter this host defence tactic. The simplest mechanism, observed to occur with the group 2b prophage in the PMEN1 population, was complete abrogation of the host's competence system. This resulted from the insertion of the lysogen into the *comYC* gene, encoding a subunit of the competence pseudopilus essential for transformation. Similarly, the non-transformable equine pathogen *S. equi* is distinguished from its progenitor, the transformable streptococcus *S. zooepidemicus*, by the insertion of a prophage into a homologue of *comFA*, essential for competence (Holden *et al*., 2009b). Such targeting of the viral integration event immediately removes the chance of the element being removed by a homologous recombination.

Another putative defence against recombinatorial removal is seen with the *Tn*5252-type ICE common in multidrug-resistant *S. pneumoniae* strains. Notable among their cargo is the presence of *uvrD* or *umuCD* DNA repair genes, the *E. coli* orthologues of which are both upregulated as part of the SOS response to genotoxic stress (Munoz-Najar and Vijayakumar, 1999). The diverse sources of these sequences suggest they are cargo genes that have been acquired in parallel, rather than constituting part of the much more strongly conserved conjugative machinery. In *E. coli*, RecA filaments recruit UmuCD to ssDNA lesions within the genome to trigger repair via translesion synthesis; as part of this function, the UmuCD complex appears to be able to bind and disrupt RecA nucleoprotein filaments (Sommer *et al*., 1993; Rehrauer *et al*., 1998). Similarly, mutant *E. coli* strains found to perform RecA-mediated integration of conjugatively-transferred DNA at an increased rate frequently have a disrupted *uvrD* gene (Arthur and Lloyd, 1980; Feinstein and Low, 1986; Bierne *et al*., 1997), with overexpression of *uvrD* inhibiting RecA-mediated recombination (Petranovic *et al*., 2001). This appears to be a consequence of UvrD's ability to dismantle RecA nucleoprotein filaments (Veaute *et al*., 2005) and inhibit RecA-facilitated pairing of homologous sequence (Morel *et al*., 1993). Hence these ICE-borne genes would be able to reduce the rate of recombination by disrupting RecA nucleoprotein filaments before they invade the genomic DNA duplex. Such an advantage would also apply *in trans* to any other element sharing the same genome; this may explain the observed association of the shorter, independently mobile Tn*916* element with the larger

transposon, and the consequent association of chloramphenicol and tetracycline resistance.

Lastly, the small interspersed repeat sequences so prevalent in *S. pneumoniae* (Chapter 7) seem likely to be exactly the type of parasites that could propagate effectively despite transformation. While transformation seems likely to be effective at removing large, single copy parasites, smaller elements would be more easily horizontally acquired and more slowly removed. Also, their high copy number would mean that they could not be eliminated effectively by a small number of horizontal sequence transfer events.

The high rate of MGE turnover in the population is likely to have the consequence that examples of such elements retained in the host are likely to be both selectively advantageous and actively mobile. While the prophage population in PMEN1 undergoes rapid flux, ICE*Sp*23FST81 remains relatively stable, presumably due to the selection for the antibiotic resistance genes it carries. However, the presence of ICE*Sp*11876 and ICE*Sp*11930 in the population appear to represent cases where ICE*Sp*23FST81 has been lost and alternative ICEs, carrying similar resistance elements, rapidly acquired in its place. In conjunction with the 'scars' observed adjacent to *rplL*, this supports the hypothesis that there is a rapid flux of conjugative elements through the pneumococcal population, which are swiftly eliminated by transformation unless there is a selection pressure to retain them.

9.2.4 Criticism of this model

Transformation would appear to be a poor method of responding to diversifying selection, because it would sample the most common alleles at any given locus most frequently. It also appears to be an inefficient way of repairing the most common forms of genome damage, because disruptive mutations would be as frequently acquired as advantageous ones, and also the MMR system is adept at inhibiting such transfers anyway. However, in order for this hypothesis to account for the advantages of transformation, it is necessary that large insertions in the recipient, relative to the donor strand, do not inhibit recombination events. Evidence for this was presented in Chapter 5, but more precise experiments focussing on this question would provide a

more definitive answer. Additionally, it is necessary that exchange within a clonally descended population is sufficiently fast to overcome the spread of MGEs through infection. While competence is upregulated under stressful conditions, when lysogens are most likely to be induced, phage have an intrinsic advantage of targeting a specific locus in the genome, whereas transformation necessarily has to cover the entire genome at random with short DNA fragments, hence must be a very common occurrence in order to compete with prophage integration at any one site. *In vitro* systems that more closely mimic the *in vivo* niche of the pneumococcus will be required to accurately test these relative rates. However, even if the kinetics of transformation are only sufficient to retard the spread of prophage, rather than eliminate them, then this would still lead to a greater chance of a non-lysogenic member of the population being transmitted to another host.

9.2.5 Comparing the evolutionary dynamics of PMEN1 and ST180

The comparison of PMEN1 and ST180 reveals starkly contrasting patterns of variation, even though both appear to have dramatically expanded in their population size over the past few decades. Assuming these differences do not represent a much stronger purifying selection pressure on ST180, all three mechanisms of horizontal DNA transfer seem to be more rapid in PMEN1, hence its more frequent acquisition of antibiotic resistance and different capsule types.

The success of ST180, despite its infrequent imports of divergent sequence, would seem to contradict the hypothesis that transformation increases a bacterium's fitness through enabling it to respond to diversifying or balancing selection. Furthermore, there is no evidence for a significantly increased proportion of pseudogenes arising in the population, implying transformation is unlikely to have a significant role in repair. It could be considered that ST180 may have switched from using imported DNA for genetic purposes to using it as a metabolic substrate instead, which could make sense in the light of the lineage's increased requirement for nucleotide sugars for the biosynthesis of the extensive mucoid capsule. However, given such a pressing need of the cell's biochemistry, it would not be expected to find the observed mutants with disruptions to key competence genes.

Rather, these observations are more in keeping with the lack of selection pressure for the retention of transformability in ST180, in turn congruent with the hypothesis that transformation acts to remove mobile elements. The thick capsular layer is likely to inhibit phage adhesion (Bernheimer and Tiraby, 1976), which would correspond with the very low rate of prophage integration observed. The capsule is likely to prevent conjugation as well, given the rarity of ICE in the sequenced ST180 isolates. With this physical barrier protecting against such parasites, the selection pressure to maintain competence would be weakened. The low rate of transformation in ST180 may account for the resilience of ΦOXC141, despite DNA and RNA sequencing revealing it to be active, and rare homoplasic deletions suggesting its loss may be advantageous. This comparison between PMEN1 and ST180 suggests that the thickness of the capsular layer may link the rates of phage infection, conjugative transfer and transformation, but these lineages are likely to represent extremes of the species. Whether this relationship holds when considering a greater number of serotypes will require the sequencing of further lineages.

9.3 Concluding remarks

Horizontal sequence transfer through transformation, conjugation and viral infection all contribute to the genetic diversity of *S. pneumoniae*. The flux of mobile elements appears to be rapid in lineages such as PMEN1, where site-specific integration leads to rapid acquisition of MGEs, while homologous recombination frequently acts to remove such features from the genome. In ST180, the mucoid capsule inhibits mobile element acquisitions but also slows the rate of transformation correspondingly. This leads to a much more static genotype, hence reducing the incidence of antibiotic resistance gene acquisition and serotype switching. However, the lower level of variation alone is not sufficient to explain the low prevalence of resistance mutations in this lineage, suggesting that selection maintains the universally susceptible phenotype. Improved understanding of the evolutionary dynamics of lineages in between these two extremes of the population should inform our knowledge of how *S. pneumoniae* populations are likely to react to clinical interventions in the future.