**Title:**

Contrasting global and local genomic epidemiology of multidrug-resistant pneumococcal lineages evolving in response to co-trimoxazole prophylaxis in South Africa

**Authors:**

Nicholas J. Croucher1, Anne von Gottberg2, Simon R. Harris3, Lesley McGee4, Mark van der Linden5, Jae-Hoon Song6, Kwan Soo Ko7, Paul Turner8,9, Raquel Sa-Leão10, Herminia de Lencastre10,11, Bernard Beall4, Julian Parkhill3, Keith P. Klugman12,13, Stephen D. Bentley3

**Affiliations:**

1 Department of Infectious Disease Epidemiology, St. Mary’s Campus, Imperial College London, London, W2 1PG, UK

2 Respiratory and Meningeal Pathogens Research Unit, National Institute for Communicable Diseases of the National Health Laboratory Service and University of Witwatersrand, Johannesburg 2000, South Africa

3 Pathogen Genomics, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

4 Respiratory Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

5 Institute for Medical Microbiology, National Reference Center for Streptococci, University Hospital, RWTH Aachen, Pauwelsstrasse 30, 52074 Aachen, Germany

6 Samsung Medical Centre, Sungkyunkwan University School of Medicine and Asia Pacific Foundation for Infectious Disease, Seoul, South Korea

7 Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon 440-746, South Korea

8 Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia

9 Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7LJ, UK

10 Laboratory of Molecular Genetics, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2784-505 Oeiras, Portugal

11 Laboratory of Microbiology, The Rockefeller University, New York, New York 10065, USA

12 Hubert Department of Global Health, Rollins School of Public Health. and Division of Infectious Diseases, School of Medicine, Emory University, Atlanta, GA 30322, USA

13 Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, Gauteng, South Africa

**Abstract**

Pneumococcal disease has long been a major problem in South Africa, exacerbated by the AIDS epidemic in recent decades. In 2000, UNAIDS recommended HIV-infected individuals in sub-Saharan Africa be offered co-trimoxazole prophylaxis against opportunistic infections. The subsequent increase in pneumococci non-susceptible to co-trimoxazole therefore represents a major public health concern. Analysis of 139 genome sequences of multidrug-resistant (MDR) serotype 23F pneumococci from South Africa indicated that the co-trimoxazole resistant PMEN1 (Spain23F-1) lineage has caused sporadic infections since it entered the country in the 1980s. The PMEN15 (Taiwan23F-15) lineage entered the country some years later in a co-trimoxazole sensitive form. However, a clade that acquired intermediate co-trimoxazole resistance, through importing a sulphonamide resistance mutation through transformation, underwent a rapid population expansion contemporaneously with the instigation of co-trimoxazole prophylaxis. These bacteria then repeatedly developed full co-trimoxazole resistance through base substitutions in *dyr* reducing susceptibility to trimethoprim. In contrast to PMEN1, a global collection of PMEN15 representatives revealed strong phylogeographic clustering of isolates, with phylodynamic and accessory genome analyses indicating the MDR phenotype emerged multiple times within this lineage. These data also show how a locally-emerged MDR lineage within South Africa rapidly became successful even in the presence of a globally disseminated strain with the same serotype and an initially more extensive resistance profile. This suggests other aspects of strain genotype influence the structuring of pneumococcal populations.

Infections caused by *Streptococcus pneumoniae* (the pneumococcus) represent both a major historical, and contemporary, problem in South Africa. Worldwide clinical interest in South African pneumococcal disease began when the gold and diamond mines were opened in the late 19th century. Local and migrant mine workers suffered from pneumonia at rates estimated as high as eight cases per thousand workers per month, stimulating interest from local, UK and US experts in methods of preventing transmission of the bacterium (1, 2). Clinical trials in the early 20th century led to the introduction of an octavalent whole cell vaccine in 1918; this continued to be used, in modified forms, in some mines for decades after (1, 2). While effective against the ‘vaccine type’ serotypes included in the formulations, overall pneumonia rates remained high in miners. This was partly a consequence of infections by ‘non-vaccine type’ serotypes; there are at least 95 pneumococcal serotypes, each directly relating to a distinct polysaccharide capsule typically synthesised by genes encoded at the capsule polysaccharide synthesis (*cps*) locus (3). Hence there was interest in developing higher valency vaccines. Acellular capsule polysaccharide vaccines were demonstrated to be efficacious in South African miners by trials in the 1970s, and a 13-valent formula was shown to be around 80% effective at preventing pneumococcal pneumonia (1, 2).

The high burden of pneumonia in South African miners also provided an opportunity to test the effectiveness of early antibiotics. In 1935, a trial in South African miners found sulphapyridine to be an effective treatment for pneumonia (4). This drug was a sulphonamide, an antibiotic class that inhbits bacterial folic acid synthesis through binding dihydropteroate synthase (5). The use of this class of antibiotics has remained high in South Africa in the form of the combination drug co-trimoxazole; this also includes trimethoprim, a drug that inhibits a later step of folic acid synthesis through binding dihydrofolate reductase (6). Following a World Health Organisation and UNAIDS recommendation in 2000 (7), co-trimoxazole prophylaxis has been offered as protection against infection for many individuals living with HIV/AIDS in South Africa (8). As around 17% of the adult South African population are HIV positive (9), and therefore are at high risk for pneumococcal disease (10), the problem of multidrug resistant (MDR) *S. pneumoniae* is particularly severe in this environment.

The very first MDR pneumococcus to be identified was found in a Durban hospital in 1977 (11), and the following year a second report emerged from Johannesburg (12). In recent decades, there has been a marked increase in the prevalence of co-trimoxazole nonsusceptibility, resulting in it becoming the most common pneumococcal resistance phenotype in South Africa (13). Surveys of isolates from invasive pneumococcal disease have found that individuals on co-trimoxazole prophylaxis are significantly more likely to be infected by bacteria nonsusceptible to the antibiotic than a control group (8), suggesting this large-scale drug administration has contributed to the rise in resistance. Epidemiological surveys of the bacterial population have found this increase in resistance has been driven by the rising prevalence of MDR strains commonly found in children (13).

Such strains typically belong to the ‘paediatric serotypes’, which are commonly found colonising the nasopharynx of infants (14). In the early 1980s, paediatric serotypes 6A, 6B and 19A were most commonly associated with penicillin resistance (15). However, in the late 1980s there was a distinct increase in the frequency of serotype 14 and 23F penicillin resistant isolates (15), and by the period of 2003-2008 these two serotypes were most commonly associated with the MDR phenotype (13). Previous whole genome sequencing found many of the serotype 23F clinical isolates from South Africa to be members of the PMEN1, or Spain23F-1, MDR lineage that appeared to have originated in Western Europe around 1970 (16). However, most of the isolates were of a different MDR genotype. Here, these isolates are identified as a distinct subgroup of the PMEN15, or Taiwan23F-15, lineage (17). This genotype was first observed in a Taiwanese hospital (18), and has since been infrequently reported in Taiwan (19, 20), Thailand (21), China (22) and South Korea (23). Outside of South-East Asia, it has also been occasionally observed in Japan (24, 25), the USA (26) and across Europe (27, 28). By generating a reference sequence, and assembling a global collection of isolates, it is possible to show that PMEN15 has developed the MDR phenotype multiple times in parallel around the world, and largely accounts for the increased levels of serotype 23F MDR pneumococcal disease in South Africa since 2000 following strong selection for developing co-trimoxazole resistance.

**Methods**

[to be completed – very similar to previous papers]

**Results**

**The emergence of PMEN15 in South Africa**

All isolates from South Africa were obtained from hospitalised patients, spanning a broad range of ages from infants to the elderly, with pneumococcal infections between 1989 and 2007. This was prior to the introduction of the first polysaccharide conjugate vaccine into South Africa in 2009 [cite http://www.sciencedirect.com/science/article/pii/S0264410X12007955]. The selected bacteria were of serotype 23F and exhibited resistance to penicillin, chloramphenicol and tetracycline. This suggested they were likely to be representatives of the PMEN1 lineage (29). Eighty-four isolates were originally submitted for sequencing, and the resulting data allowed the multilocus sequence type (ST) to be determined for 80 with confidence. Thirty-six of these belonged to clonal complex (CC) 81, indicating that they were members of the PMEN1 lineage (Table S1); this was subsequently confirmed by whole genome sequencing (16). Two isolates were of ST4225, matching the genotype of a serotype 6A isolate from Botswana (30). However, the remaining 42 isolates were of, or related to, ST6279, a single locus variant of ST242, the ST originally associated with the PMEN15 (Taiwan23F-15) multidrug-resistant lineage in the 1990s (18). A further 66 MDR serotype 23F isolates from South Africa were then sequenced, of which 64 could be typed. Analysis with eBURST (31) found 55 of these isolates belonged to the PMEN15 lineage (Fig. 1), while the remaining nine were of PMEN1 (Tables S1 and S2).

Plotting the frequency of these two lineages in South Africa over time (Fig. 1) revealed the incidence of disease caused by PMEN1 rose only slightly between the 1990s and 2000s, a trend that was not significant (Mann-Kendall test for trend, τ = 0.34, *p* = 0.055). However, no PMEN15 isolates were detected prior to 2000, but the frequency at which this lineage caused disease followed a significant increasing trend (Mann-Kendall test for trend, τ = 0.73, *p* = 1.00x10-4), rapidly overtaking PMEN1. Hence, the increase in serotype 23F MDR disease between the 1980s and 2000s (13) seems to have been mainly driven by the emergence of the PMEN15 lineage within South Africa, a trend not previously detected in the absence of genotyping data.

**Genomic differences between PMEN1 and PMEN15**

Genomic DNA from the isolate SA37 was used to generate a reference sequence for PMEN15. The original Illumina sequence reads were combined with 454 data to generate a combined assembly that was annotated (Fig. 2) and compared to the PMEN1 reference genome, *S. pneumoniae* ATCC 700669 (29). As well as sharing similar *cps* loci, the two genomes both harboured an unusual prophage remnant found in only a few pneumococcal lineages (29, 32). However, the emergence of PMEN15 does not appear to represent a case of sequence transfer *en masse* from PMEN1, as inferred from other MDR lineages (33); the PBP1A, PBP2X, and PBP2B proteins of ATCC 700699 and SA37 were only 93%, 97% and 98% identical, respectively. Much variation in the accessory genome related to carbohydrate metabolism, with SA37 possessing gene clusters for the metabolism of 6-phospho-β-glucosides and 6-phospho-β-galactosides, amongst others [here I plan to add a couple of growth curves], but lacking a mannitol utilisation gene cluster present in ATCC 700669. Notable antigenic differences include the presence of the *rrg* island, encoding the type I pneumococcal pilus (34), in SA37, and the presence of *pclA* (35) in ATCC 700669.

The two strains also differed in their complement of mobile genetic elements. While ATCC 700669 carried an ΦMM1-type prophage (29), a novel phage (ΦSA37) is present in SA37. This latter virus is inserted into the *comYC* gene, encoding a component of the competence pilus. Such integrations have previously been observed to inhibit transformation in the laboratory (16) and during evolutionary diversification (36). As with the PMEN1 lineage, SA37’s acquisition of tetracycline and chloramphenicol resistance was the result of the insertion of a Tn*5253*-type integrative and conjugative element. While this element shared a similar structure to ICE*Sp*23FST81, it differed substantially in sequence (Fig. 2).

**Diversity of integrative and conjugative elements**

The sections of ICE*Sp*SA37 encoding antibiotic resistance were similar to those of ICE*Sp*23FST81: the *cat* chloramphenicol acetyltransferase gene was carried on a linearised pC194 plasmid (37) within an Ω(*cat*) element [cite Tn*5253*], and the *tetM* tetracycline resistance gene was carried by a Tn*916*-type element within the larger Tn*5252*-type ICE (29). ICE*Sp*SA37 also carried a Tn*917* element, encoding the *ermB* macrolide resistance gene (38), inserted at the same site within Tn*916*-type element as seen in other PMEN1 representatives (16). The remaining cargo genes were highly distinct; the ICEs each carried highly distinct *uvrD* genes, and different lantibiotic synthesis operons. [will run BAGEL3 on SA37]

To ascertain whether the global spread of PMEN15 followed a similar pattern to that of PMEN1, MDR strains genotyped as being of, or related to, ST242 were collected from multiple countries. This added 49 further isolates from North and South America, Western and Eastern Europe, and South-East Asia to the dataset (Table S1). Searching *de novo* assemblies of these isolates revealed both sequence and structural diversity distinguishing the Tn*5253*-type ICEs causing the MDR phenotype of different PMEN15 representatives (Fig. 2). The number and type of macrolide resistance cassette present, as previously observed in other lineages (16, 36, 39); however, they also differed in more typically conserved loci. Multiple distinct arrangements of the Tn*5252*- and Tn*916*-type elements were observed in the ICEs, and the Ω(*cat*) element and lantibiotic biosynthesis gene clusters of ICE*Sp*SA37 were absent from most other elements found in PMEN15. Each of these ICE variants was observed in a distinct geographical location. This suggested the resistance-bearing components of the Tn*5253*-type ICEs in PMEN15 had been independently acquired in different locations, in contrast to the evolution of PMEN1, in which ICE*Sp*23FST81 was generally conserved throughout the lineage’s global spread, with the exception of macrolide resistance cassette acquisition (16).

**Parallel emergence of resistance within PMEN15**

A whole genome alignment was created by mapping Illumina sequence reads against the reference genome of SA37, and reconstructing patterns of common ancestry and horizontal sequence transfer (Fig. 3) as described previously (16, 40). Excluding the BM4200 isolate used as an outgroup (16), 47,913 polymorphic sites were identified. Overall, 82,511 base substitutions were predicted to have occurred across these sites by the reconstruction; 79,290 of these were predicted to have been imported by 798 recombinations. Excluding those occurring within putative mobile genetic elements, there were 682 likely transformation events in the reconstruction, resulting in a per site *r*/*m* estimate of 24.6. This robustly inferred five distinct clades (Fig. S1) and predicted a similar pattern of recombinations to an independent analysis with a separate algorithm (Fig. S2). Although these isolates were selected only on the basis of genotype, just one was not of serotype 23F; this was a South Korean isolate that had switched to serotype 6A through a 17.9 kb recombination spanning *wzg* at the 5’ end of the *cps* locus to part of the *rml* operon at the 3’ end. The *pspC* gene, encoding surface protein C, was frequently exchanged by recombination, whereas *pspA*, encoding surface protein A, was more conserved. There was considerable evidence of phage movement over the short term, and multiple recombinations affecting loci determining susceptibility to β–lactams and co-trimoxazole.

The five clades each associated with different geographical regions. All but one of the 94 South African isolates were monophyletic, forming ‘clade SA’. The single outlier was present in a clade of otherwise European isolates (‘clade Europe’); the only European isolate outside of this group was in turn closely related to a set of bacteria from South-East Asia and the USA (‘clade Asia’). The two remaining clades, ‘Brazil’ and ‘Peru’, both consisted entirely of isolates from South America; despite the isolates in each clade coming from a single country, their genetic diversity was comparable to the other internationally-disseminated clades.

Removing base substitutions introduced by recombination from the alignment revealed evidence of a molecular clock, based on the linear correlation between root-to-tip distances and year of isolation (*R*2 = 0.036, *p* = 0.024; Fig. S5). This estimated the last common ancestor of PMEN15 to have existed around 1937, approximately contemporaneous with the first trials of sulphapyridine in South Africa. Using a fitting a Bayesian phylodynamic coalescent model to the data estimated the date of PMEN15’s last common ancestor to have existed in 1938 (95% credibility interval of 1899-1959), with a clock rate of 1.22x10-6 substitutions per site per year (95% credibility interval of 1.00x10-6-1.50x10-6 substitutions per site per year), similar to previous estimates (16, 21, 26, 39, 41). As this predates the introduction of many of the antibiotics to which extant PMEN15 isolates are resistant, this supports the hypothesis that the MDR phenotype has independently emerged several times. Hence it seems likely that PMEN15 began its worldwide spread in a sensitive form prior to different isolates independently acquiring resistance and subsequently spreading locally. This was consistent with the distinct clades’ estimated dates of origin being approximately contemporaneous with, or postdating, the first observations of MDR pneumococci: 1987 for clade SA (95% credibility interval of 1982-1992); 1977 for Clade Asia (95% credibility interval of 1968-1984); 1980 for clade Brazil (95% credibility interval of 1957-1993), and 1985 for clade Europe (95% credibility interval of 1979-1989).

Correspondingly, each of the five clades was stably associated with one of the five ICEs in Fig. 2 (Fig. S8), suggesting these mobile elements had been separately acquired in different regions. There was also telling variation in macrolide resistance cassettes (Fig. S4): all representatives of clade SA carried Tn*917*, while clade Europe all possessed the Mega cassette-encoded Mef efflux pump. Clades Asia, Brazil and Peru each seem to have acquired different cassettes following the acquisition of their respective ICEs, as observed for other lineages [cite PMEN]. Only clade SA appears to have carried Ω(*cat*) since its origin; the only other occurrence of this element is in the South African isolate that is part of clade Europe, suggesting a region-specific selection pressure for retaining this gene.

To test whether resistance had also evolved independently at core genomic loci, the penicillin-binding genes determining β–lactam resistance (*pbpb1a*, *pbpX* and *pbp2b*) were independently aligned and clustered (Fig. 6). This found clades SA and Asia to have distinct allele sets of all three genes from the rest of the population, indicating they had separately acquired β–lactam resistance from different donors. The ongoing diversification of these genes, indicated by their positions coinciding with recombination ‘hotspots’ in the whole genome alignment (Fig. 4), in the other three clades made it difficult to ascertain whether they had originally been separately acquired or diverged from a single common acquisition of resistance. The clades also differed in the amino acid insertions in dihydropteroate synthase causing sulphonamide resistance: clades Europe and Peru both shared the same duplication of arginine and proline at position 59 of dihydropteroate synthase (42); clade Asia had a different insertion around position 66 of the protein (43), while examples of insertions associated with resistance were found at each position in both clade Brazil and SA.

Resistance-associated insertions in *folP* were predicted to have been imported by at least six transformation events, including independent acquisitions of the insertion at position 59 shared by clades Europe and Peru. The only ancestral, sulphonamide-sensitive *folP* alleles in the collection were in Clade SA. A parsimonious interpretation of the data indicated the originally sulphonamide susceptible clade SA imported resistance twice: one isolate acquired an insertion around serine 61, and a large subclade (labelled ‘R’; Fig. 5) imported the duplication of arginine and proline at position 59.

**Comparative phylodynamics of PMEN1 and PMEN15**

The vast majority of South African isolates were monophyletic in both the PMEN15 (clade SA) and PMEN1 (previously identified as clade A (16)) phylogenies; using the detailed dates of isolation against the root-to-tip distances within these clades, it was possible to find evidence of a molecular clock for both (Fig. S5). This allowed a Bayesian phylodynamic coalescent model, including a discrete state phylogeographic model reconstruction of the lineages’ spread, to be fitted to the data (Fig 5.). The analysis of locations is particularly relevant to South Africa, where differences in pneumococcal epidemiology have been observed between communities (15). The PMEN1 lineage appeared to have disseminated throughout the country, with two clades containing multiple isolates from the region around Cape Town. For PMEN15, and particularly subclade R, isolates were generally limited to Johannesburg and the surrounding regions, suggesting a more concentrated outbreak over a shorter timeframe.

The resulting skyline plots allowed the product of the effective population size (*N*e) and the generation time (τ) to be estimated. Based on the *in vitro* growth curves, [here I plan to put in some growth curves], allowing the estimates of *N*e for the two lineages to be compared. Corresponding with the sporadic isolation of PMEN1 isolates since 1989, the phylodynamic reconstruction estimated that the lineage entered South Africa around 1982 (95% credibility interval of 1977-1986), likely causing some of the increase in MDR 23F cases in the late 1990s (15), and increased in prevalence in the late 1990s. The PMEN15 lineage was estimated to have entered South Africa around 1990 (95% credibility interval of 1978-1997), consistent with the overall analysis of the lineage. These dates of entry for the two lineages used similar independent substitution rate estimates: 1.27x10-6 substitutions per site per year for PMEN1 (95% credibility interval of 8.75x10-7-1.74x10-6 substitutions per site per year), and 1.11x10-6 substitutions per site per year for PMEN15 (95% credibility interval of 7.38x10-7-1.48x10-6 substitutions per site per year). However, the *N*e for PMEN15 is predicted to have remained low until a rapid rise in the early 2000s to a similar level to that of PMEN1, clearly mirroring the rise in clinical cases being detected (Fig. 1).

**Strong selection for co-trimoxazole resistance within PMEN15**

The dramatic rise in clinical isolates and *N*e just after to 2000, at the beginning of co-trimoxazole prophylaxis, is associated with the emergence of the sulphonamide-nonsusceptible subclade R. The rapid expansion of this subclade was also reflected by the highly asymmetric tree topology within clade SA; this can be quantified by the Colless index, which was 0.33 for this tree, which ranged between 0.039 and 0.15 for 1,000 randomly generated tree topologies with the same number of isolates. The origin of subclade R is associated with the import of a sulphonamide nonsusceptibility allele by a 7.3 kb long transformation event, but not the acquisition of different penicillin-binding protein alleles (Fig. 5), macrolide resistance cassettes, or ICEs (Fig. S4). This is strong evidence that subclade R rapidly expanded in prevalence as a consequence of selection for sulphonamide nonsusceptibility, which the timing suggests is likely to be linked to the instigation of co-trimoxazole prophylaxis.

However, co-trimoxazole non-susceptibility also requires resistance to trimethoprim. In pneumococci, this is typically the result of an I100L substitution within the dihydrofolate reductase protein (44). The evolutionary reconstruction (Fig. 4) estimates this change to have arisen 24 times across this collection of PMEN15 isolates, and predicts it was imported by transformation events in each case. Recombinations cause the change eleven times within subclade R, but never within the sulphonamide-susceptible clade SA isolates. This highly homoplasic distribution contrasted with that observed in PMEN1, which stably carried the I100L substitution in dihydrofolate reductase throughout its international spread (Fig. S9). However, resistance to rifampicin followed this pattern in both lineages: substitutions at position H499 of the β subunit of RNA polymerase (45) occurred six times within PMEN15, five of which were predicted to have been caused by transformation, and four times in PMEN1, two of which were predicted to have been caused by transformation. Although only one example of the H499 substitutions was observed in PMEN15 outside of clade SA, and similarly only one example was recorded outside of clade A in PMEN1, the mutations did not occur at a sufficiently high frequency to conclude that they were significantly more common in South Africa, where rifampicin is commonly used to treat the high burden of tuberculosis (13), relative to other countries.

**Discussion**

The most recent common ancestor of the PMEN15 lineage appears to have been approximately contemporaneous with the introduction of the first antibiotics. Yet the lineage appears likely to have spread internationally in an entirely antibiotic susceptible form over decades, before subsequently developing the MDR phenotype multiple times independently during the years between the 1970s and 1990s. These timings, inferred through phylodynamic analyses, are consistent with the differences between clades in the sequences underlying the same resistance phenotypes. Particularly unexpected is the appearance of similar ICEs in the same lineage on multiple different occasions. One potential explanation is that the Tn*5252*-type element was present in the progenitor of the lineage, and elements such as the Tn*916*-type sequences and Ω(*cat*) have subsequently inserted into it over subsequent decades. The overall evolutionary pattern has resulted in isolates indistinguishable by multilocus sequence typing (being of ST242; Table S1) sharing similar MDR phenotypes, despite having apparently acquired the traits entirely separately. This contrasts with the global spread of PMEN1, the most recent common ancestor of which was estimated to have existed in the 1970s, having already acquired the MDR phenotype (16).

PMEN1 appears to have reached South Africa in the 1980s, already fully resistant to co-trimoxazole, and has caused sporadic cases of disease over decades across the country. PMEN15 seems to have been imported into South Africa just a few years later, and persisted at a low level while susceptible to both sulphonamides and trimethoprim. Following the introduction of co-trimoxazole prophylaxis in 2000, there was a rapid expansion of the sulphonamide-resistant subclade R, largely due to transmission centred around Johannesburg. This primarily drove the increased frequency of antibiotic-resistant serotype 23F disease by 2003-2008 (13). The acquisition of trimethoprim resistance seems to have occurred through transformation importing the I100L substitution on multiple occasions. However, the gain of this resistance mutation appears to have been inhibited by the repeated infection with ΦSA37, a prophage disrupting the competence system through inserting into *comYC* gene. This is similar to the manner in which prophage ΦIC1, which inserted into the same site, was found to be a likely constraint on the diversification of PMEN2 (36).

The contrasting patterns with which the resistance to sulphonamides and trimethoprim were acquired highlights the difficulty in detecting selection pressures on bacteria. Despite both drugs usually being administered as a combination, the development of sulphonamide resistance appears to have been a comparatively ‘hard’ selective sweep (46), with a resistance allele acquired once by the progenitor of subclade R and subsequently increasing in frequency through proliferation of the resistant descendants. By contrast, the development of trimethoprim resistance looks more like a ‘soft’ selective sweep, with repeated emergence of the same mutation in different bacteria in parallel (39, 46). As both mutations are imported through transformation, rather than arising through point mutation during disease as observed for other base substitutions causing pneumococcal resistance (47), this suggests either a difference in the relative rates at which the two types of mutation are acquired, or a stronger cost associated with the trimethoprim resistance mutation resulting in a higher level of counterselection. The pattern of sulphonamide nonsusceptibility alleles being stably retained across clades, while trimethoprim nonsusceptibility appears sporadically, is repeated across the entire PMEN15 collection. This is in contrast to the stability of trimethoprim resistance in both PMEN1 (16) and PMEN14 (39). These comparisons indicate the pattern of mutation likely reflects a trait of this lineage, suggesting epistasis between resistance mutations and different genetic backgrounds (47).

As this study only considered South Africa isolates of serotype 23F, it is possible that the comparative demographics of these two lineages is distorted by not accounting for isolates that have undergone serotype switching (3). However, when selecting a global collection of isolates based on genotype information, only one non-23F PMEN15 representative was identified (Table S1). A recent epidemiological survey of clinical pneumococcal isolates in South Africa found a minority of PMEN15 isolates to be of serotype 6B, alongside a single non-23F representative of PMEN1 (48). As the global collections indicate both strains were ancestrally of serotype 23F, this indicates that the isolates studied correctly depict serotype 23F MDR disease in South Africa, and very likely accurately reflect the overall dynamics of the PMEN1 and PMEN15 lineages in the country.

What factors underlay the success of PMEN15 in the presence of PMEN1 following the implementation of co-trimoxazole prophylaxis, despite PMEN1’s stable co-trimoxazole resistance, may be clearer if detailed information were available on the changes in the rest of the nasopharyngeal microbiota triggered by this alteration in antibiotic usage. There are some notable differences between the lineages, such as differences in their complement of carbohydrate transporters and subcapsular protein antigens. It seems likely that one or more of these provided a selective advantage following the restructuring of the bacterial population. PMEN15’s success, despite repeated phage infection that deprived the strain of its ability to import variation through transformation, is analogous to the success of PMEN2 in Iceland, where PMEN1 was also present (36). However, in this case the globally disseminated lineage was outcompeted by a locally emerged, serotype-matched MDR strain, highlighting the difficulty of predicting how pneumococcal populations respond to clinical interventions.

**References**

1. **Austrian R**. 1978. The Jeremiah Metzger Lecture: Of gold and pneumococci: a history of pneumococcal vaccines in South Africa. Trans Am Clin Clim Assoc, 1977/01/01 ed. **89**:141–161.

2. **Koornhof H, Madhi S, Feldman C, von Gottberg A, KP K**. 2009. A century of South African battles against the pneumococcus. South Afr J Epidemiol Infect **24**:7–19.

3. **Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, Lipsitch M, Hanage WP**. 2015. Selective and Genetic Constraints on Pneumococcal Serotype Switching. PLoS Genet **11**:e1005095.

4. **Agranat A, Dreosti A, Ordman D**. 1939. Treatment of pneumonia with 2-(p-aminobenzenesulphonamido) pyridine. Lancet **233**:309–317.

5. **Woods D**. 1940. The relation of para-aminobenzoic acid to the mechanism of the action of sulphanilamide. Br J Exp Path **21**:74–90.

6. **Hitchings G, Burchall J**. 1965. folate biosynthesis and function as a basis for chemotherapy. Advan Enzym **27**:417–468.

7. **WHO/UNAIDS**. 2000. Provisional WHO/UNAIDS secretariat recommendations on the use of cotrimoxazole prophylaxis in adults and children living with HIV/AIDS in Africa.

8. **Soeters HM, Von Gottberg A, Cohen C, Quan V, Klugman KP**. 2012. Trimethoprim-sulfamethoxazole prophylaxis and antibiotic nonsusceptibility in invasive pneumococcal disease. Antimicrob Agents Chemother **56**:1602–1605.

9. **UNAIDS**. 2012. Global report: UNAIDS report on the global AIDS epidemic 2012.

10. **Madhi SA, Petersen K, Madhi A, Khoosal M, Klugman KP**. 2000. Increased disease burden and antibiotic resistance of bacteria causing severe community-acquired lower respiratory tract infections in human immunodeficiency virus type 1-infected children. Clin Infect Dis **31**:170–176.

11. **Appelbaum PC, Scragg JN, Bhamjee A, Hallett AF, Bowen AJ, Cooper RC**. 1977. Streptococcus pneumoniae resistant to penicillin and chloramphenicol. Lancet **2**:995–997.

12. **Jacobs MR, Koornhof HJ, Robins-Browne RM, Stevenson CM, Vermaak ZA, Freiman I, Miller GB, Witcomb MA, Isaacson M, Ward JI, Austrian R**. 1978. Emergence of multiply resistant pneumococci. N Engl J Med, 1978/10/05 ed. **299**:735–740.

13. **Crowther-Gibson P, Cohen C, Klugman KP, De Gouveia L, Von Gottberg A**. 2012. Risk factors for multidrug-resistant invasive pneumococcal disease in South Africa, a setting with high HIV prevalence, in the prevaccine era from 2003 to 2008. Antimicrob Agents Chemother **56**:5088–5095.

14. **Austrian R**. 1989. Epidemiology of pneumococcal capsular types causing pediatric infections. Pediatr Infect Dis J **8**:S21–S22.

15. **Koornhof HJ, Wasas A, Klugman K**. 1992. Antimicrobial resistance in Streptococcus pneumoniae: A South African perspective. Clin Infect Dis **15**:84–94.

16. **Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD**. 2011. Rapid pneumococcal evolution in response to clinical interventions. Science (80- ), 2011/01/29 ed. **331**:430–434.

17. **McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefevre JC, Tomasz A, Klugman KP**. 2001. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. J Clin Microbiol, 2001/06/28 ed. **39**:2565–2571.

18. **Shi ZY, Enright MC, Wilkinson P, Griffiths D, Spratt BG**. 1998. Identification of three major clones of multiply antibiotic-resistant Streptococcus pneumoniae in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol, 1998/11/18 ed. **36**:3514–3519.

19. **Hsieh Y-C, Chang K-Y, Huang Y-C, Lin H-C, Ho Y-H, Huang L-M, Hsueh P-R**. 2008. Clonal spread of highly beta-lactam-resistant Streptococcus pneumoniae isolates in Taiwan. Antimicrob Agents Chemother **52**:2266–2269.

20. **Ko KS, Song JH**. 2004. Evolution of erythromycin-resistant Streptococcus pneumoniae from Asian countries that contains erm(B) and mef(A) genes. J Infect Dis, 2004/07/24 ed. **190**:739–747.

21. **Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen DM, Mather AE, Page AJ, Salter SJ, Harris D, Nosten F, Goldblatt D, Corander J, Parkhill J, Turner P, Bentley SD**. 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat Genet.

22. **Yang F, Xu XG, Yang MJ, Zhang YY, Klugman KP, McGee L**. 2008. Antimicrobial susceptibility and molecular epidemiology of Streptococcus pneumoniae isolated from Shanghai, China. Int J Antimicrob Agents **32**:386–391.

23. **Waites KB, Jones KE, Kim KH, Moser SA, Johnson CN, Hollingshead SK, Kang ES, Hong KS, Benjamin WH**. 2003. Dissemination of Macrolide-Resistant Streptococcus pneumoniae Isolates Containing Both erm(B) and mef(A) in South Korea. J Clin Microbiol **41**:5787–5791.

24. **Kasahara K, Maeda K, Mikasa K, Uno K, Takahashi K, Konishi M, Yoshimoto E, Murakawa K, Kita E, Kimura H**. 2005. Clonal dissemination of macrolide-resistant and penicillin-susceptible serotype 3 and penicillin-resistant Taiwan 19F-14 and 23F-15 Streptococcus pneumoniae isolates in Japan: A pilot surveillance study. J Clin Microbiol **43**:1640–1645.

25. **Canton R, Morosini M, Enright MC, Morrissey I**. 2003. Worldwide incidence, molecular epidemiology and mutations implicated in fluoroquinolone-resistant Streptococcus pneumoniae: Data from the global PROTEKT surveillance programme. J Antimicrob Chemother **52**:944–952.

26. **Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M**. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet **45**:656–663.

27. **Dicuonzo G, Gherardi G, Gertz RE, D’Ambrosio F, Goglio A, Lorino G, Recchia S, Pantosti A, Beall B**. 2002. Genotypes of invasive pneumococcal isolates recently recovered from Italian patients. J Clin Microbiol **40**:3660–3665.

28. **Sadowy E, Izdebski R, Skoczynska A, Grzesiowski P, Gniadkowski M, Hryniewicz W**. 2007. Phenotypic and molecular analysis of penicillin-nonsusceptible Streptococcus pneumoniae isolates in Poland. Antimicrob Agents Chemother **51**:40–47.

29. **Croucher NJ, Walker D, Romero P, Lennard N, Paterson GK, Bason NC, Mitchell AM, Quail MA, Andrew PW, Parkhill J, Bentley SD, Mitchell TJ**. 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. J Bacteriol, 2008/12/31 ed. **191**:1480–1489.

30. **Aanensen DM, Spratt BG**. 2005. The multilocus sequence typing network: mlst.net. Nucleic Acids Res, 2005/06/28 ed. **33**:W728–33.

31. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG**. 2004. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. J Bacteriol **186**:1518–1530.

32. **Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, Hanage WP**. 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat Commun **5**.

33. **Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, Jacobs MR, Kristinsson KG, Beall BW, Klugman KP**. 2012. The multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. Genome Biol **13**:R103.

34. **Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, Dahlberg S, Fernebro J, Moschioni M, Masignani V, Hultenby K, Taddei AR, Beiter K, Wartha F, von Euler A, Covacci A, Holden DW, Normark S, Rappuoli R, Henriques-Normark B**. 2006. A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci U S A, 2006/02/17 ed. **103**:2857–2862.

35. **Paterson GK, Nieminen L, Jefferies JM, Mitchell TJ**. 2008. PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. FEMS Microbiol Lett, 2008/06/19 ed. **285**:170–176.

36. **Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, Sá-Leão R, Song J-H, Ko KS, Beall B, others**. 2014. Variable recombination dynamics during the emergence, transmission and “disarming” of a multidrug-resistant pneumococcal clone. BMC Biol **12**:49.

37. **Widdowson CA, Adrian P V, Klugman KP**. 2000. Acquisition of chloramphenicol resistance by the linearization and integration of the entire staphylococcal plasmid pC194 into the chromosome of Streptococcus pneumoniae. Antimicrob Agents Chemother, 2000/01/20 ed. **44**:393–395.

38. **Shaw JH, Clewell DB**. 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn917 in Streptococcus faecalis. J Bacteriol, 1985/11/01 ed. **164**:782–796.

39. **Croucher NJ, Chewapreecha C, Hanage WP, Harris SR, McGee L, van der Linden M, Song J-H, Ko KS, de Lencastre H, Turner C, others**. 2014. Evidence for Soft Selective Sweeps in the Evolution of Pneumococcal Multidrug Resistance and Vaccine Escape. Genome Biol Evol **6**:1589–1602.

40. **Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR**. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res **43**:e15.

41. **Croucher NJ, Hanage WP, Harris SR, McGee L, van der Linden M, de Lencastre H, Sá-Leão R, Song J-H, Ko KS, Beall B, others**. 2014. Variable recombination dynamics during the emergence, transmission and’disarming'of a multidrug-resistant pneumococcal clone. BMC Biol **12**:49.

42. **Padayachee T, Klugman KP**. 1999. Novel expansions of the gene encoding dihydropteroate synthase in trimethoprim-sulfamethoxazole-resistant Streptococcus pneumoniae. Antimicrob Agents Chemother **43**:2225–2230.

43. **Lopez P, Espinosa M, Greenberg B, Lacks SA**. 1987. Sulfonamide resistance in Streptococcus pneumoniae: DNA sequence of the gene encoding dihydropteroate synthase and characterization of the enzyme. J Bacteriol **169**:4320–4326.

44. **Adrian P V, Klugman KP**. 1997. Mutations in the dihydrofolate reductase gene of trimethoprim-resistant isolates of Streptococcus pneumoniae. Antimicrob Agents Chemother **41**:2406–2413.

45. **Ferrandiz MJ, Ardanuy C, Linares J, Garcia-Arenzana JM, Cercenado E, Fleites A, de la Campa AG**. 2005. New mutations and horizontal transfer of rpoB among rifampin-resistant Streptococcus pneumoniae from four Spanish hospitals. Antimicrob Agents Chemother, 2005/05/27 ed. **49**:2237–2245.

46. **Hermisson J, Pennings PS**. 2005. Soft sweeps molecular population genetics of adaptation from standing genetic variation. Genetics **169**:2335–2352.

47. **Croucher NJ, Mitchell AM, Gould KA, Inverarity D, Barquist L, Feltwell T, Fookes MC, Harris SR, Dordel J, Salter SJ, Browall S, Zemlickova H, Parkhill J, Normark S, Henriques-Normark B, Hinds J, Mitchell TJ, Bentley SD**. 2013. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet **9**:e1003868.

48. **Ndlangisa KM, du Plessis M, Wolter N, de Gouveia L, Klugman KP, von Gottberg A, GERMS-SA for**. 2014. Population Snapshot of <italic>Streptococcus pneumoniae</italic> Causing Invasive Disease in South Africa Prior to Introduction of Pneumococcal Conjugate Vaccines. PLoS One **9**:e107666.

**Figure Legends**

**Figure 1** Clonal structure of serotype 23F multidrug-resistant pneumococci in South Africa over time. (a) By extracting multilocus sequence typing MLST data from Illumina sequence reads, all serotype 23F multidrug-resistant pneumococci could be assigned to one of three clonal complexes (CCs) using the criterion of at least five shared loci. Forty-five belonged to CC81, associated with the PMEN1 lineage. Eighty-three belonged to CC6279, which is a member of CC880 in the global MLST dataset; this is associated with the PMEN15 lineage. Two isolates were of ST4225, only previously observed in Botswana. (b) Prevalence of clones over time. PMEN1 isolates were isolated sporadically between the late 1980s and mid-2000s. By contrast, both ST4225 isolates were observed in the mid-1990s, while PMEN15 was not observed pre-2000, but subsequently became the most common serotype 23F multidrug-resistant pneumococcal lineage isolated from disease in South Africa.

**Figure 2** Annotation of the PMEN15 isolate SA37. The outer circle labels the major features of this pneumococcal chromosome. These include genes encoding antigenic surface structures: pneumococcal surface proteins A (*pspA*) and C (*pspC*); the capsule polysaccharide synthesis (*cps*) locus, the *rrg* pilus island and the pneumococcal serine-rich repeat protein (PsrP) island. Also marked are pneumococcal pathogenicity island 1 (PPI-1), the prophage ΦSA37, and the integrative and conjugative element ICE*Sp*SA37. Labels show the positions of *dyr* and *folP*, the sequences of which determine pneumococcal co-trimoxazole resistance, and *pbp2b*, the sequence of which partly determines lactam β-resistance. The next two tracks inwards show the distribution of predicted coding sequences on the forward and reverse strands of the genome, the distribution of which reflect the strong coding bias of a Firmicute genome. The next track inwards shows BLAST matches between SA37 and ATCC 700669, the reference isolate for PMEN1, as red bands; the white gaps indicate regions of difference between the two genomes. One ring further inwards shows the GC content of the genome using a sliding window, and the innermost graph shows the GC skew of the sequence, indicating the position of the origin and terminus of replication.

**Figure 3** Comparison of integrative and conjugative elements (ICEs). The ICE found in the PMEN1 lineage, ICE*Sp*23FST81, is compared to that found in the PMEN15 reference genome, ICE*Sp*SA37. The CDSs predicted within each sequence are represented by pink boxes, except antibiotic resistance genes, which are dark blue. Red bands between the sequences indicate regions of similar sequence, with the intensity of the colour reflecting the strength of the match. In the case of ICE*Sp*SA37, the distinct contigs of the assembly are represented by the alternating orange and brown boxes. Both ICEs have a Tn*5253*-type structure, with a Tn*916*-type element inserted into a Tn*5252*-type element; this arrangement is described by the black bars at the top and bottom of the figure. Underneath ICE*Sp*SA37, four further ICEs assembled from PMEN15 representatives are shown: each comes from a different location, and can be associated with different groups of PMEN15 isolates as indicated by Fig. 4 and Fig. S8.

**Figure 4** Reconstruction of the evolutionary history of PMEN15. (a) Annotation of reference genome of isolate SA37, as displayed in Figure 2. (b) Maximum likelihood phylogeny showing the population structure of PMEN15 based on point mutations. The tree is coloured according to the geographic location of samples’ isolation, as reconstructed by maximum parsimony. This analysis used the unrelated serotype 23F MDR isolate BM4200 as the outgroup on which to root the tree. This allowed five distinct clades to be identified: that associated with South Africa (clade SA), with Western and Eastern Europe (clade Europe), with South-East Asia and the USA (clade Asia), with Peru (clade Peru) and with Brazil (clade Brazil). (c) Putative recombination events occurring in the history of PMEN15. Regions of the genome with an elevated density of base substitutions reconstructed as occurring on a single branch of the tree are marked as likely cases of horizontal sequence transfer. These are displayed as red boxes, where a putative recombination was reconstructed as occurring on an internal branch and is therefore shared by multiple isolates through common descent, or blue boxes, where the putative recombination is reconstructed as occurring on a terminal branch, and is therefore unique to a single isolate. (d) Density of recombination events across the genome. These heatmaps show the frequency of independent recombination events across each base of the reference genome, from unaffected by recombination in any isolate in the collection (blue) to affected by ten or more recombination events (red). The top row shows the summary data for just clade SA, and the bottom row shows the overall data for all five clades.

**Figure 5** Comparative phylodynamic analyses of clade SA in PMEN15 and clade A in PMEN1. (a) Administrative regions of South Africa (b) Phylodynamic reconstruction of PMEN15 clade SA. Tree branches are coloured according to the reconstructed location within South Africa, and scaled according to the likely time of coalescence prior to their isolation. (c) Bayesian skyline plots showing the reconstructed product of the effective population size, *N*e, and generation time, τ. Assuming τ to be constant and the same between the lineages (growth curves!), this shows PMEN1 to have undergone a slight increase in prevalence between 1995 and 2000, whereas the PMEN15 population dramatically increased between 2000 and 2005. Both reached similar *N*e values by the end of the sampling. (d) Phylodynamic reconstruction of PMEN1 clade A, displayed as described for panel (b).

**Figure 6** Co-trimoxazole and β–lactam resistance in PMEN15. (a) Phylogeny displayed as in Figure 4. (b) Diversity of genes involved in β–lactam and co-trimoxazole resistance. The first three columns correspond to the genes that determine pneumococcal β–lactam resistance: *pbpX*, *pbp1a*, and *pbp2b*. Each row corresponds to an individual sequence, as defined by the position of isolates in the tree. These were individually analysed with BAPS and BRATNextGen to cluster them and identify any detectable mosaic structure. The overall clustering is shown by the background ‘recipient cluster’ colour; the imported sequences are indicated by the ‘donor cluster’ coloured segments. Diversity of genes involved in co-trimoxazole resistance. The next two columns correspond to *folP* and *dyr*. Small insertions at positions 59 or 61 within the dihydropteroate synthase protein, corresponding to the locus within *folP* indicated by the downwards arrow, are associated with resistance to sulphonamides. Different sequences at these loci are represented by different coloured boxes, with grey boxes corresponding to the ancestral susceptible alleles in each case. The I100L substitution in the dihydrofolate reductase protein, caused by to base substitutions within *dyr* at the position indicated by the downwards arrow, is associated with resistance to trimethoprim. The susceptible allele is again grey.

**Supplementary Figure Legends**

**Figure S1** Details of the maximum likelihood phylogeny shown in Figure 4. This figure shows the isolate names and the bootstrap support for each node, as calculated from 100 bootstrap replicates.

**Figure S2** NextGenBRAT analysis of the PMEN15 isolates. (a) Annotation of reference genome of isolate SA37, as displayed in Figure 2. (b) Maximum likelihood phylogeny as displayed in Figure 4. (c) Clustering and recombinations predicted by NextGenBRAT. The background ‘recipient cluster’ colouring describes the overall clustering, while the regions coloured as belonging to ‘donor clusters’ represent putative instances of horizontal sequence transfer. This independent analysis of the whole genome alignment identified groups consistent with the phylogeny, albeit differing from those annotated, which also drew on geographical and accessory genome information. Patterns of recombination predictions are generally consistent, although the methods sometimes assign recombination events to complementary sets of leaf nodes in cases where there may be some ambiguity in the reconstruction.

**Figure S3** Analysis of the new PMEN1 whole genome alignment. This dataset has been modified from a previously published version through the addition of extra isolates from South Africa, and the exclusion of some samples for which lower levels of sequence data were available, to make it consistent with the analysis of the PMEN15 isolates. The results of the evolutionary history reconstruction are shown as in Figure 4. For the summary heatmaps, the top row summarises just the recombination events within clade A, while the bottom row summarises data across all isolates.

**Figure S4** Distribution of antibiotic resistance-associated sequences. (a) The sequences of five antibiotic resistance-associated accessory genome loci. Tn*916*-type elements carry the *tetM* tetracycline resistance gene; the Omega and Tn*917* elements insert into Tn*916*-type sequences and carry the *ermB* macrolide resistance gene; the Mega cassette carries the *mefE* macrolide resistance gene; and Ω(*cat*) encodes for a chloramphenicol acetyltransferase. (b) Maximum likelihood phylogeny, as displayed in Figure 4. (c) Heatmap showing distribution of antibiotic resistance-associated sequences. Each row corresponds to a single isolate in the phylogeny. The depth of sequence read mapping to each base in the reference sequence is shown by the colour of the heatmap: blue show zero mapping, suggesting the sequence is absent, while red indicates mapping up to a level of 50-fold coverage, indicating the presence of a sequence.

**Figure S5** Evidence of molecular clock signals. Each graph shows the root-to-tip distances extracted from the relevant phylogeny relative to the date of isolation for each isolate, where available. The red line shows the best linear relationship between these data. (a) Shows the absence of any evidence of a molecular clock signal when a tree is constructed from the raw whole genome alignment (n = 142, *R*2 = 0.0066, *p* = 0.34). (b) Shows the significant positive correlation between root-to-tip distance and date of isolation when recombination is removed from the whole genome alignment by the analysis displayed in Figure 4 (n = 142, *R*2 = 0.036, *p* = 0.029). This estimates the last common ancestor of the PMEN15 isolates to have existed around 1937. (c) Shows the evidence for a molecular clock from clade SA of PMEN15. The linear regression estimates a date of origin around 1994 (n = 94, *R*2 = 0.16, *p* = 1.01x10-4). (d) Shows the evidence for a molecular clock from clade A of PMEN1. The linear regression estimates a date of origin around 1976 (n = 34, *R*2 = 0.34, *p* = 1.07x10-4).

**Figure S6** Transformation event lengths. Histograms (a) and (b) show the distribution of transformation event lengths in clade SA and across the full PMEN15 collection, respectively. The red lines show the best-fitting exponential distributions, which have rate parameters of 1.26x10-4 bp-1 and 1.42x10-4 bp-1, respectively.

**Figure S7** Distribution of prophage sequences. (a) Relatedness of 32 pneumococcal prophage. A hierarchical clustering of 32 pneumococcal prophage sequences, of which four were extracted from sequences that were part of this study, was constructed as described previously (see Methods). The sequences themselves are represented by alternating orange and brown bars underneath the tree. (b) Maximum likelihood phylogeny of PMEN15 isolates, displayed as described in Fig. 4. (c) Heatmap showing the mapping of PMEN15 sequence reads to the prophage sequences, as described in Fig. S4. (d) Maximum likelihood phylogeny of PMEN1 isolates, displayed as described in Fig. S3. (e) Heatmap showing the mapping of PMEN1 sequence reads to the prophage sequences, as described in Fig. S4.

**Figure S8** Distribution of integrative and conjugative element sequences. (a) The five ICE sequences shown in Fig. 2. (b) The maximum likelihood phylogeny, displayed as in Figure 4. (c) Heatmap showing sequence read distribution, displayed as described in Fig. S4. This shows the clonal association between particular pneumococci and the integrative and conjugative elements they carry.

**Figure S9** Co-trimoxazole and β–lactam resistance in PMEN15. (a) Phylogeny displayed as in Fig. S3. (b) Analysis of the three genes associated with pneumococcal β–lactam resistance (*pbpX*, *pbp1a*, and *pbp2b*) and the two genes associated with pneumococcal co-trimoxazole resistance (*folP* and *dyr*) as described in Fig. 6.