Expansion of pneumococcal serotype 23F and 14 lineages with genotypic changes in capsule polysaccharide locus and virulence gene profiles post introduction of pneumococcal conjugate vaccine in Blantyre, Malawi

Author names

Rory Cave[1], Akuzike Kalizang'oma[1,2], Crispin Chaguza[3,4], Thandie S. Mwalukomo[5], Arox Kamng'ona[5], Comfort Brown[2], Jacquline Msefula[5], Farouck Bonomali[2], Roseline Nyirenda[2], Todd D. Swarthout[1,2,6], Brenda Kwambana-Adams[1,2], Neil French[7], Robert S. Heyderman[1,2]

Affiliation(s)

1.Mucosal Pathogens Research Group, Research Department of Infection, Division of Infection & Immunity, University College London, London, United Kingdom

2.Malawi Liverpool Wellcome Programme, Blantyre, Malawi

3.Parasites and Microbes, Wellcome Sanger Institute, Cambridge, UK

4.Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT, USA

5.Kamuzu University of Health Sciences, Blantyre, Malawi

6.Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands

7.Clinical Infection, Microbiology and Immunology, Institute of Infection Veterinary & Ecological Science, University of Liverpool, Liverpool, United Kingdom

Corresponding author and email address

Rory Cave rory.cave@ucl.ac.uk
Robert Heyderman r.heyderman@ucl.ac.uk

Keywords

Streptococcus pneumoniae, Capsule polysaccharide *locus,* PCV13*,* vaccine escape, Antimicrobial resistance, Africa

1. Abstract

Since the introduction of the 13-valent pneumococcal conjugate vaccine (PCV13) in Malawi in 2011, there has been persistent carriage of vaccine serotype (VT) *Streptococcus pneumoniae*, despite high vaccine coverage. To determine if there has been a genetic change within the VT capsule polysaccharide (cps) loci since the vaccine's introduction, we compared 1,022 whole-genome-sequenced VT isolates from 1998 to 2019. We identified the clonal expansion of a multidrug-resistant, penicillin non-susceptible serotype 23F GPSC14-ST2059 lineage, a serotype 14 GPSC9-ST782 lineage and a novel serotype 14 sequence type GPSC9-ST18728 lineage. Serotype 23F GPSC14-ST2059 had an I253T mutation within the capsule oligosaccharide repeat unit polymerase Wzy protein, which is predicted *in silico* to alter the protein pocket cavity. Moreover, serotype 23F GPSC14-ST2059 had SNPs in the DNA binding

sites for the cps transcriptional repressors CspR and SpxR. Serotype 14 GPSC9-ST782 harbour a non-truncated version of the large repetitive protein (Lrp), containing a Cna protein B-type domain which is also present in proteins associated with infection and colonisation. These emergent lineages also harboured genes associated with antibiotic resistance, and the promotion of colonisation and infection which were absent in other lineages of the same serotype. Together these data suggest that in addition to serotype replacement, modifications of the capsule locus associated with changes in virulence factor expression and antibiotic resistance may promote vaccine escape. In summary, the study highlights that the persistence of vaccine serotype carriage despite high vaccine coverage in Malawi may be partly caused by expansion of VT lineages post PCV13 rollout.

Impact Statement

Our findings highlight the potential for clonal expansion of multidrug-resistant, penicillin-non-susceptible vaccine serotype lineages with capsule locus modifications, within a high carriage and disease burden population. This shift has occurred among young children where there has been high vaccine coverage, posing challenges for effective vaccine scheduling and design. Furthermore, this study emphasises the importance of ongoing *Streptococcus pneumoniae* genomic surveillance as new or modified pneumococcal vaccines are implemented.

2. Data summary

Whole genome sequencing assemblies for the PCVPA survey have been deposited in the BioProject PRJNA1011974.

3. Introduction

Pneumonia, meningitis and sepsis caused by *Streptococcus pneumoniae* (the pneumococcus) is a major global public health concern, with an estimated 300,000 global deaths due to invasive disease reported among children aged under 5 of which 57% occur within resource poor settings (1,2). Asymptomatic nasopharyngeal carriage of the pneumococcus is a prerequisite for pneumococcal disease, transmission, and the development of natural immunity (3). Since the introduction of pneumococcal conjugate vaccines (PCVs) in childhood vaccination programs worldwide, there has been a considerable reduction in invasive disease (4).

The outer capsule polysaccharide (cps) is a major pneumococcal virulence factor, protecting the pneumococcus against desiccation, complement-mediated opsonophagocytic and other host antimicrobial pathways (5–7). The cps biosynthesis genes are found on a single locus controlled by a single promoter region for most serotypes (8). Cps consists of diverse sugar structures that vary among isolates, serving as the basis for classifying *S. pneumoniae* serotypes, with more than 100 immunologically-distinct serotypes identified to-date (9). Pneumococcal conjugate vaccines (PCVs) are formulated with a select array of serotype-specific capsule polysaccharides, chosen to target the most commonly occurring invasive serotypes, with a particular focus on those that cause the most severe diseases or are associated with AMR (10).

Since the 2011 PCV13 rollout in Blantyre, Malawi, we have shown a reduction in vaccine serotype (VT) invasive pneumococcal disease (IPD) with the incidence of post-PCV13 VT IPD 74% lower among children aged 1–4 years, and 79% lower among children aged 5–14 years from 2006 to 2018 (11). However, among PCV13 age ineligible populations, we noted only 38% lower VT IPD among infants and 47% lower among adolescents and adults. We have also shown that VT IPD has persisted amongst infants <90 (12). Alongside this, we have shown that in contrast to high income settings, there is considerable residual VT carriage, 7 years after PCV13 introduction despite high vaccine uptake (13). We have also shown that this imperfect direct and indirect control of pneumococcal carriage and disease is associated with waning of protective vaccine-induced anti-pneumococcal immunity in the first year of life (14).

The pneumococcus is highly transformable such that the *cps* locus, a known recombination hotspot, often acquires changes and can facilitate pneumococcal vaccine escape(15). Serotype switching from VT to non-vaccine serotypes (NVT), gene deletions or mutations resulting in pseudogenes lead to capsule loss (16,17). Genetic changes within the cps locus which do not lead to capsule switch or loss could also enable VT serotype persistence post PCV rollout (18). Here, we investigated the hypothesis that the residual VT IPD and persistent VT carriage observed following PCV13 introduction in Malawi, is at least in part due to the clonal expansion of VT lineages that have acquired changes in their *cps* locus while maintaining their serotype. We further postulate that these capsule locus variants have also acquired genetic traits that together could promote a competitive advantage in colonisation and transmission, with the potential to result in vaccine escape.

4. Methods

Whole genome sequences

We obtained the whole genome sequences of PCV13 VT *S. pneumoniae* isolates from Blantyre, Malawi through the Global Pneumococcal Sequencing Project (GPS) dataset (n=329). These were derived from isolates collected between 1998-2015, including both carriage and disease isolates. We also used the Pneumococcal Conjugative Vaccine Prospective Analysis (PCVPA) dataset (n=693) from serologically typed carriage isolates collected from vaccinated, unvaccinated children, and adults living with HIV between 2015 to 2019 (13,19,20). Additionally, for the purpose of genomic comparison between Blantyre VT isolates with those isolated in other countries we incorporated publicly available isolates from Pathogenwatch which contains sequences from carriage and disease from early 1900's to 2021 (https://pathogen.watch/, Accessed February 2023).

Genome genetic typing and annotation

Genetic typing and antimicrobial resistance genotype of isolates was also conducted using Pathogenwatch. Lineages were defined by Pathogenwatch using the Global Pneumococcal Sequence Cluster (GPSC) nomenclature employing the PopPUNK framework, and the Multi-locus Sequence Type (MLST) system based on the pneumococcal scheme (19,21,22). Full genome annotation was conducted using Bakta v1.9.1 (23).

127 ***cps locus* comparison**
128 We used parsnp v1.7.4 (https://github.com/marbl/parsnp) to extract and identify single
129 nucleotide polymorphisms (SNPs) within the Blantyre pneumococcal *cps* locus by aligning
130 them against reference serotype specific *cps* locus described by Bentley *et al*. (8). These SNPs
131 were then annotated to distinguish synonymous from non-synonymous mutations using the
132 vcf-annotator tool v0.5 (https://github.com/rpetit3/vcf-annotator). A phylogenetic tree from
133 the alignment of the *cps* locus was constructed using IQ-TREE v2.1.2 with the best model for
134 each alignment selected by ModelFinder (24,25). Phylogenetic trees and SNPS within the *cps*
135 locus were visualised using the R package ggtree v3.18 (26). BLASTP v2.14.1 was used to
136 determine if similar mutations were found in the amino acid sequence of isolates from other
137 countries (27).
138
139 We performed sequence alignment for the *cps* locus using BLASTN v2.14.1 to enhance the
140 detection of indels within the *cps* locus that are segmented into multiple contigs. The
141 alignment coverage against the reference was then visualised using the R package gggenomes
142 v0.9.12 (https://github.com/thackl/gggenomes).
143
144 To determine genetic synteny within the intergenic regions of the *cps locus*, we extracted the
145 DNA sequences located between *dexB* and *wzg* for each individual genome. Synteny was
146 established by aligning these sequences with one another using minimap2, and the resulting
147 synteny patterns were visualised using the R package gggenomes (28). We investigated
148 alterations within the 37-CE region located upstream of the cps promoter, where the
149 transcriptional factors SpxR and CpsR are known to bind to suppress expression of the cps
150 locus (29). To locate and extract the 37-CE sequence, we developed a custom Python script
151 (https://github.com/rorycave/37-CE_finder) that identified the position of the sequence
152 pattern "TTGAAAC," which is typically conserved in the 37-CE region across various serotypes
153 within the *cps* locus intergenic region. Subsequently, the 37-CE sequences were extracted
154 using bedtools v2.28 'getfasta' commands based on their sequence positions, including 156
155 nucleotides upstream and 14 nucleotides downstream of the sequence (30). The extracted
156 sequences were then manually checked by aligning them to the reference 37-CE sequence
157 using Clustal Omega v1.2.4 (29,31).
158
159 **The impact of synonymous SNPs on protein structure within the *cps* locus**
160 To determine the impact of nonsynonymous SNPs on a protein structure within the *cps* locus,
161 we first used SWISS-MODEL, to find homologous protein models that had a Global Model
162 Quality Estimate (GMQE) >0.95 (32). Furthermore, within that GMQE range we chose X-ray
163 diffraction models if present over models predicted by in silico methods. For transmembrane
164 proteins that only have in silico predicted models, TMBed was used to identify if mutations
165 occurred in the transmembrane cytoplasmic or extracellular domain of the protein (33).
166 Additionally, we used PrankWeb 3 web server which runs P2RANK to determine if the
167 mutation occurred in a protein pocket, and Missesne3D to assess whether mutations would
168 cause structural damage to the protein's 3D structure (34–36).
169
170

171 **Phylogenetic and accessory genome analysis**
172 A core SNP maximum-likelihood (ML) phylogenetic tree was constructed by aligning
173 assemblies to a serotype specific complete reference genome sequence using Snippy v4.6.0
174 (https://github.com/tseemann/snippy). Recombination within aligned sequences was then
175 filtered out with Gubbins v3.3.1 (37). A phylogenetic tree was then constructed from
176 recombinant-free alignment using IQ-TREE v2.1.2 with the best model for each alignment
177 using selected by ModelFinder and set ultrafast bootstrap replication to 1,000 (24,25). The
178 phylogenetic tree was visualised and annotated in Microreact (38).
179
180 To find differences in the accessory genome between lineages that have the same serotype,
181 a pangenome form the annotated genome for each serotype was then constructed using
182 Panaroo v1.3.4 with the merge paralogs setting (39). Scoary v1.6.16 was then used to identify
183 genes that belong to certain lineages that were absent in others (40).

184 ## 5. Results

185 **Expansion of cps locus variant lineages of serotype 23F and 14 following PCV13**
186 **introduction**
187 To determine if the predominate VT cps locus genotypes changed after PCV13 introduction
188 in Malawi, we compared the PCVPA dataset (2015 to 2019) to earlier pneumococcal
189 sequences (1998 to 2014) in Blantyre, Malawi. Serotype 23F and 14 lineages were found to
190 have expanded with changes in their *cps* locus genotypes.
191
192 The dominant serotype 23F lineage shifted from GPSC20 (n=23, 69.7%) pre-PCV13 rollout to
193 a GPSC14 lineage that dominant from 2014 onwards (n=57, 78.03%). This includes the
194 emergence of a GPSC14 ST2059 (n=42, 57.5%) that became dominant among all 23F serotype
195 genotypes collected in the PCVPA dataset.
196
197 For serotype 14 isolates (n=77, collected between 2000 and 2019), the dominant GPSC9
198 remained unchanged after the PCV13 introduction (Figure 1B). However, there was a shift in
199 the most prevalent STs, transitioning from GPSC9 ST63 pre-vaccine (n=13, 100%) to GPSC9
200 ST782 (n=20, 35.1% of PCVPA serotype 14 isolates) between 2015 and 2019. We also
201 observed the clonal expansion of a novel serotype 14 sequence type ST18728 (n=13, 26.3%
202 of PCVPA serotype 14 isolates), which was dominant from 2018 to 2019.
203
204 **Genetic drift within 23F GPSC14 *cps* locus predicted to impact on capsule expression and**
205 **phenotype**
206 To explore the hypothesis that the clonal expansion of 23F GPSC14 ST2059 post-PCV13
207 introduction and the decline in 23F GPSC20 may have been due a competitive advantage, we
208 first looked for genetic differences within the *cps* locus. Alignment with the reference 23F *cps*
209 locus (GenBank accession: CR931685) revealed a higher SNP density with the emergent 23F
210 GPSC14 lineage (mean = 11.5 SNPs/Kbp, range 7.7 - 29.2 SNP/Kbps per isolate) when
211 compared to 23F GPSC20 (mean = 5.7 SNPs/Kbp, range 5.7 to 5.8 SNPs/Kbp per isolate).
212 Furthermore, the 23F GPSC14 lineages exhibited four non-synonymous SNPs, resulting in
213 amino acid changes: W54S, L56F, T237I, and I253T, in the oligosaccharide repeat unit
214 polymerase gene, *wzy*. These same mutations were also found in a single serotype 23F
215 GPSC455 isolate (Figure 2).

Additionally, from the cps locus alignment, we identified a subpopulation within the GPSC14 lineage composed of ST2059 (n=49), ST12347 (n=1), and isolates (n=7) that belong to five novel sequence types (ST18717, n=1; ST18707, n=2; ST18727, n=1; ST18722, n=1; ST18723, n=2). These were single allele variants of ST2059 that displayed recombination within the rhamnose synthesis locus (*rmlACDB*). This resulted in three non-synonymous SNPs (leading to amino acid changes: L214V, R258M, and S272P) in the *rmlA* gene and seven non-synonymous SNPs (leading to amino acid changes: T2S, A12V, E13I, L46E, E57A, D89G, and STOP198E) in the *rmlC* gene.

To evaluate the impact of non-synonymous SNPs in 23F GPSC14 on protein structure, we utilised an AlphaFold model for *S. pneumoniae* Serotype 23F Wzy (UniProt ID: Q9R925), an Alphafold model of RmlA (UniProt ID: A0A2I1UG67) from *Streptococcus oralis subsp dentisani* and a X-ray diffraction model for RmlC (SMTL ID: 1ker.1.) of *Streptococcus suis* (32). From these three, the I253T mutation in Wzy was predicted to be part of a protein pocket (confidence score of 0.970), and the alteration of the amino acid residue was predicted to induce structural changes in the protein by reducing the cavity volume by 126.576 $Å^3$ (Figure S1).

We then investigated the intergenic regions of the serotype 23F *cps* locus, situated between *dexA* and the *wzg* gene where transcriptional regulators bind to alter capsule expression (27). We observed structural variations among several different lineages (Figure 3A). GPSC14 and GPSC455 exhibited a longer intergenic region (1,401bp) compared to GPSC20, 228, 5, and 22 (554bp) due to the truncation of insertion sequences. Conversely, GPSC176, 272, 328, 116, 22, and 40 displayed a shift in the repeat unit pneumococcal (RUP) preceding the insertion sequences. These structural variations in the intergenic regions of the cps locus may lead to different level of gene expression in the cps locus.

Additionally, we identified differences between lineages in DNA binding site sequences for SpxR and CpsR in the 37-CE, which are known to suppress capsule expression (8) (figure 3B). Notably, GPSC14 and GPSC455 showed a higher similarity (92%) to the reference strain (D39 Serotype 4) in the SpxR1 and CpsR binding sites described by Glanvielle *et al.*, compared to GPSC20, 228, 22, 857, 882, 10, and 116 (62% similarity). Moreover, within GPSC5, 176, 272, 328, and 116 there was 72% similarity compared to the reference sequence. The primary distinction between the GPSC20 SpxR1 and CpsR binding sites and the reference sequence lies in the alterations involving A to T or T to A base changes at positions 1, 3, 4, and 5, along with a T-to-G substitution at position 10. Conversely, for GPSC14 and GPSC455, the sole variation from the reference occurs at position 10, where a T is replaced by a G, whereas GPSC5, 176, 272, 328, and 116 also had two additional changes at positions 7 and 9, both being A to G. Ultimately these changes may alter the binding affinity of the SpxR1 and CpsR proteins, leading to changes in cps locus gene expression.

Emergent serotype 14 isolates with the non-truncated version of the large repetitive protein gene within the *cps* locus.

Through a comparative analysis of the serotype 14 *cps* locus over time, we did not observe the emergence of non-synonymous SNPs within cps genes or changes in cps transcriptional binding sites, that we had seen in the 23F lineage. Instead, by BLAST sequence analysis using a reference serotype 14 sequences (GenBank accession: CR931662) we identified isolates harbouring large repetitive protein (*lrp*) gene that was not truncated, containing the Cna protein B-type domain a feature absent from serotype 14 isolates *lrp* gene collected before 2015 (figure 4). The emergence of isolates with the complete version of the *lrp* gene post-PCV13 introduction may suggest that the bacteria have a fitness advantage, potentially aiding them in evading vaccine-induced antibodies through alterations in the bacteria's immunogenicity.

Global comparative analysis of serotype 23F Wzy, SpxR1 and CpsR binding motif, and serotype 14 Lrp changes

To further understand whether the mutations related to the serotype 23F Wzy amino acid structure and the SpxR1 and CpsR binding sites in 37-CE are specific to certain lineages, and if they are confined to Malawi, we conducted genomic comparisons with 1,385 isolates worldwide. Notably, mutations in Wzy protein sequences and the genetic structures of the 37-CE were consistent across the majority of GPSC14 isolates globally (Figure S2). Moreover, the Wzy protein in serotype 23F exhibits a high level of sequence conservation, with just two distinct versions of the protein found in more than one isolate within the available dataset. This suggests, that emergent of a mutated Wzy protein in GPSC14 isolates may confer a fitness benefit.

In our core maximum likelihood phylogenetic analysis of serotype 23F isolates, constructed by aligning against the complete genome of serotype 23F *S. pneumoniae* ATCC 700669 (GenBank accession: FM211187), we found that the 37-CE sequences are primarily associated with specific lineages (Figure S2). This implies that different lineages may be linked to varying levels of cps expression. The phylogenetic tree also suggests a close genetic relationship between the newly emergent GPSC14 ST2059 in Blantyre and GPSC14 ST2059 isolates from South Africa, with the closest South African Malawian isolate differing by only 32 SNPs. This implies possible country-to-country transmission, most likely from South Africa to Malawi, as the ST2059 lineage was prominent in South Africa pre and post -PCV vaccine introduction but not in Malawi.

We also conducted a comparative analysis of serotype 14 isolates' *lrp* gene. Utilising data from 1,607 isolates collected globally, including Malawian isolates from various regions, our aim was to determine whether the non-truncated version of the *lrp* gene is associated with specific lineages. Interestingly, we observed that, aside from GPSC9, other lineages mainly had the complete or truncated version of the gene, indicating that the versions of the gene evolved separately from each other, and recombination is less likely to occur in the *lrp* gene.(table S1). Our phylogenetic analysis of GPSC9 serotype 14 isolates, aligned against the serotype 14 *S. pneumoniae* G54 complete genome (GenBank accession: CP001015), indicates a genetic divergence based on the completeness of the *lrp* gene. We observe one cluster containing all truncated versions of the *lrp* gene, and the other comprises a mixture of

complete and incomplete *lrp* genes, suggesting that variations in the *lrp* gene could potentially play a role in the differences observed among these isolates (Figure S4).

Emergent serotype 23F and 14 lineages are associated with virulence genes and AMR

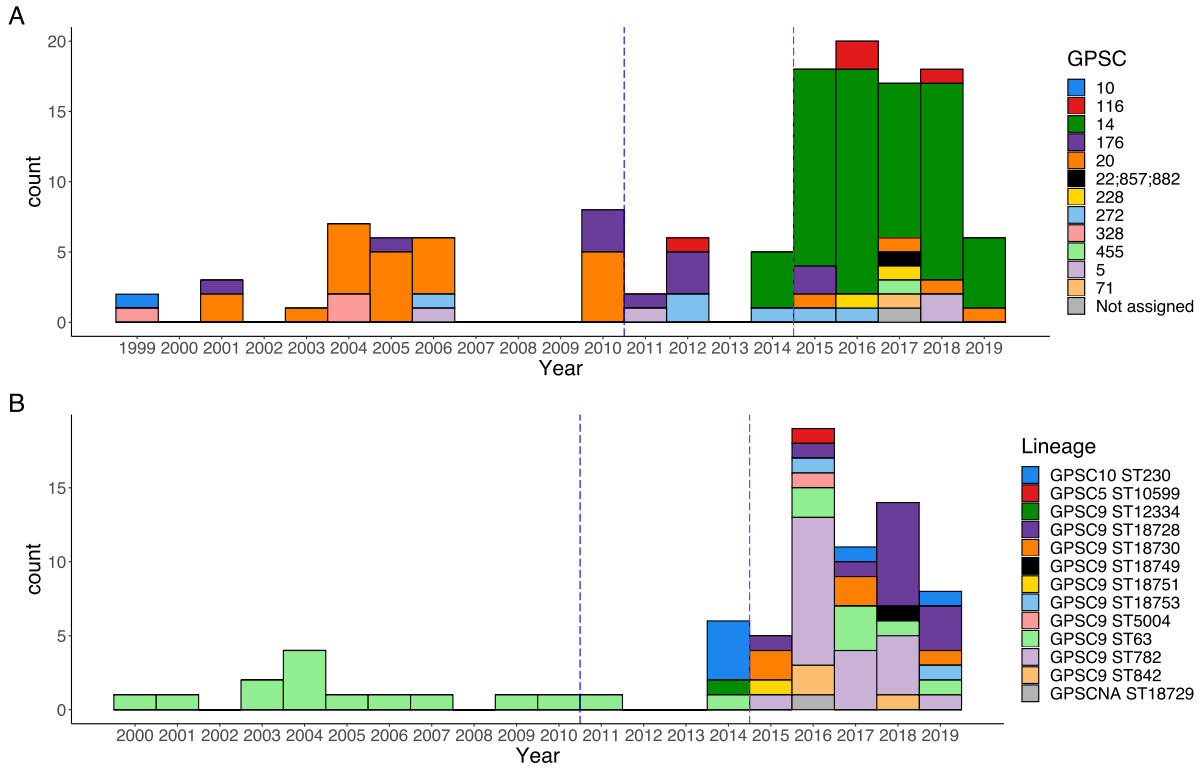
To explore the hypothesis that the emerging serotype 23F and 14 lineages had additional virulence and AMR characteristics that conveyed advantage, we first compared the accessory gene content of the different lineages (20) (Table S2). A toxin-antitoxin system gene associated bacterial stress response, cell growth, and biofilm formation, and the pneumococcal serine-rich repeat protein (PsrP)-accessory Sec system (secY2A2) pathogenicity island witch involved in biofilm formation, adhesion to epithelial cells, and the export of contact-dependent pneumolysin toxin were consistently present in the Serotype 23F GPSC14 isolates but absent in the 23F GPSC20 isolates (41–47). Furthermore, all 23F ST2059 isolates carried the macrolide resistance gene *mefA*, which was only present in three 23F isolates outside of GPSC14. Additionally, they harboured Thiazolylpeptide-type bacteriocin and Lantibiotic resistance genes associated with bacterial colonisation (48).

Regarding serotype 14 GPSC9 lineage, the Choline-binding protein *pcpA* virulence gene which mediates pneumococcal adhesion was present in all but one isolate in cluster C, which contains the post-PCV13 emergent lineage on the core phylogenetic tree (Figure 5)(49). ST18728 also carried the Zinc metalloprotease virulence gene associated with inflammation in the lower respiratory tract, found in only one additional isolate indicating the emergent isolates may have an increased virulence potential enabling the lineage to the dominate lineages pre-PCV13 introduction (50).

To assess AMR among the emergent serotype 23F and 14 lineages, we compared them to the lineage that dominated before PCV 13 introduction (Figure 6). For 23F ST2059 harboured *mefA* (macrolide resistance), *tetM* (tetracycline resistance) and the mutations folP_aa_insert_57-70 and folA_I100L (resistance to sulfamethoxazole and trimethoprim, respectively). Additionally, 23F ST2059 isolates were penicillin non-susceptible (MIC 0.25 to 0.5 µg/µL), unlike GPSC20 ST82, which was penicillin-susceptible (MIC 0.035 µg/µL) (Figure 5). There were three 23F GPSC20 ST9530 isolates recovered before PCV13's introduction also exhibited penicillin non-susceptibility (MIC 0.25 µg/µL).

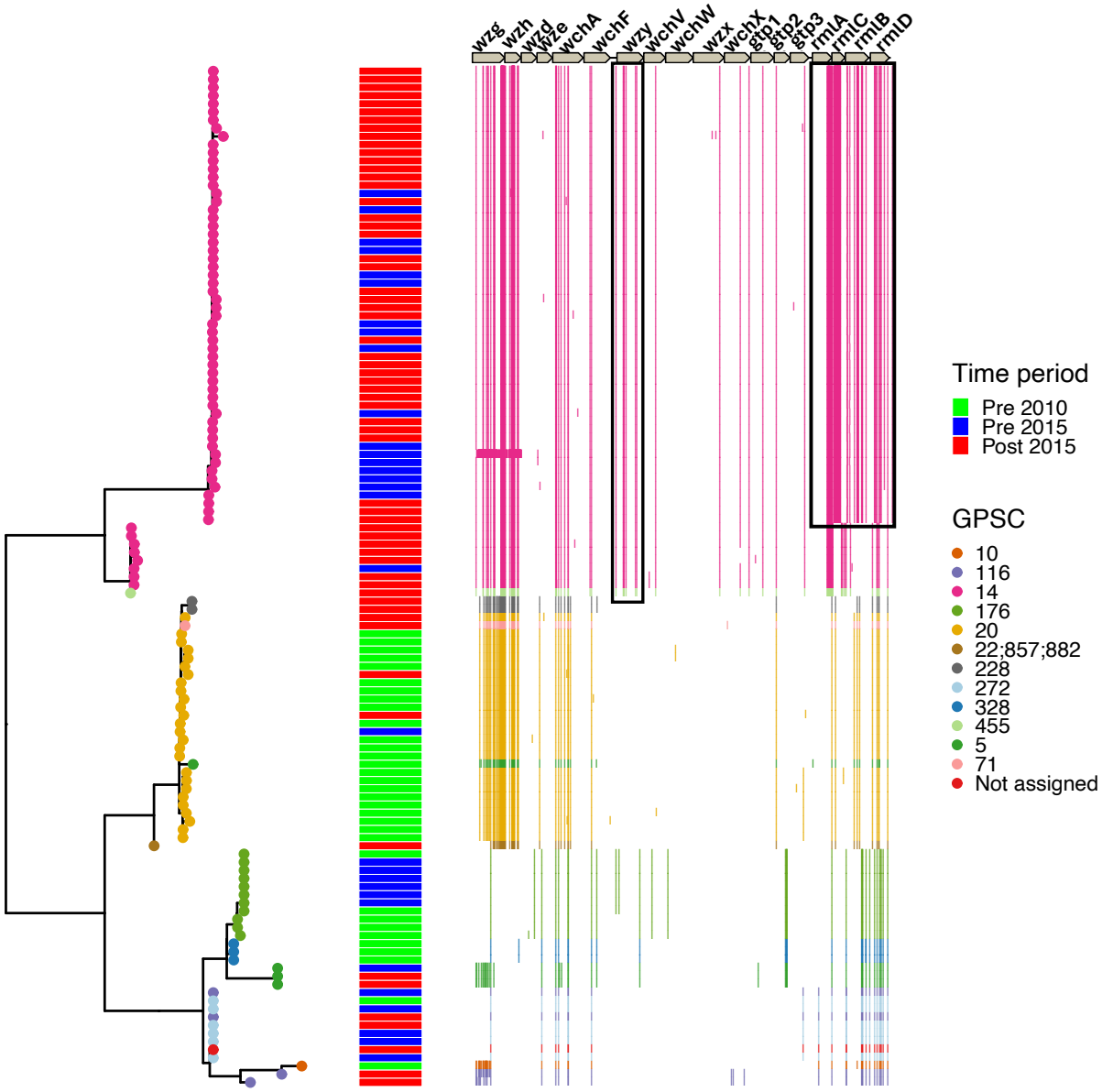
For Serotype 14 isolates there were no additional gene mutations (folP_aa_insert_57-70 and folA_I100L) or acquisitions (*tetM*) conferring AMR. All serotype 14 lineages were penicillin non-susceptible. There was no additional increase in penicillin MIC amongst the emerging GPSC9 lineage.

342 ## 6. Figures and tables



343
344 Figure 1: Change in genetic lineage/strain overtime among A) Serotype 23F and B) Serotype
345 14 isolates. Blue dash line indicates the time that PCV13 was introduced in Blantyre, Malawi.
346 Purple line indicates the start of PCVPA survey.

347



348
349 Figure 2: Maximum likelihood phylogenetic tree and position of SNPS within serotype 23F
350 *cps* locus shows that change in genotype in *wzy* and *rmlACBD* in the emergent GPSC14 *S.*
351 *pneumoniae* isolates in Blantyre, Malawi based on phenotype clustering.

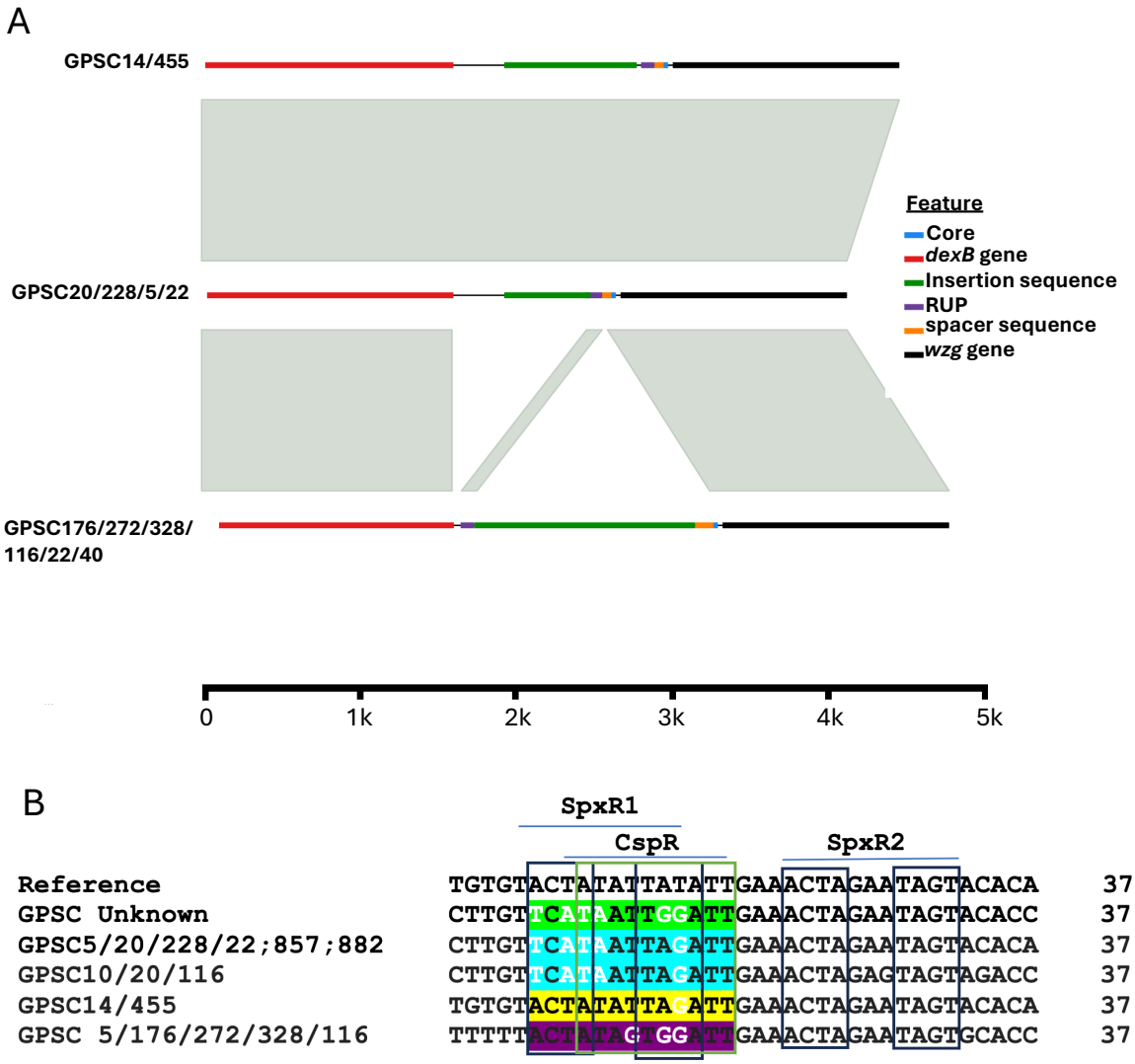


Figure 3: Differences in the *cps* locus intergenic region in the emergent 23F GPSC14 isolates in Blantyre, Malawi. A) Genetic synteny of intergenic regions between isolates. B) Changes in the CpsR and SpxR1 nucleotide binding site in the 37-CE.

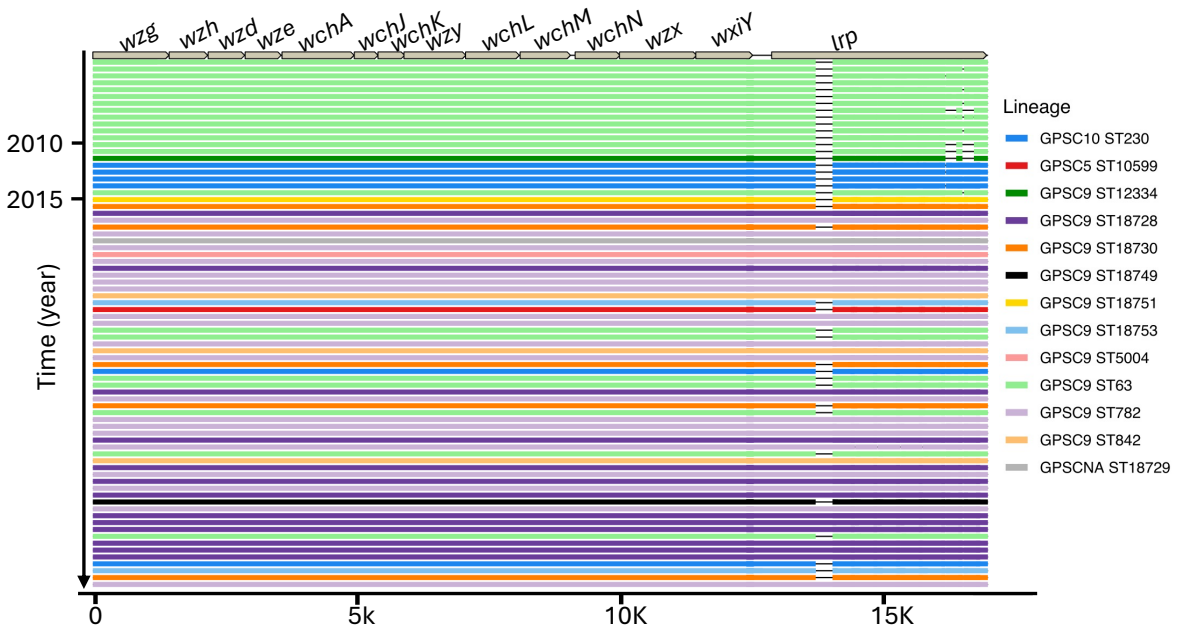


Figure 4: BLASTN analysis to identify large genomic changes in the serotype 14 cps sequences showing the emergence of sequence types within GPSC9 lineage that have complete version of *lrp* gene. GPSCNA= non-classified GPSC.

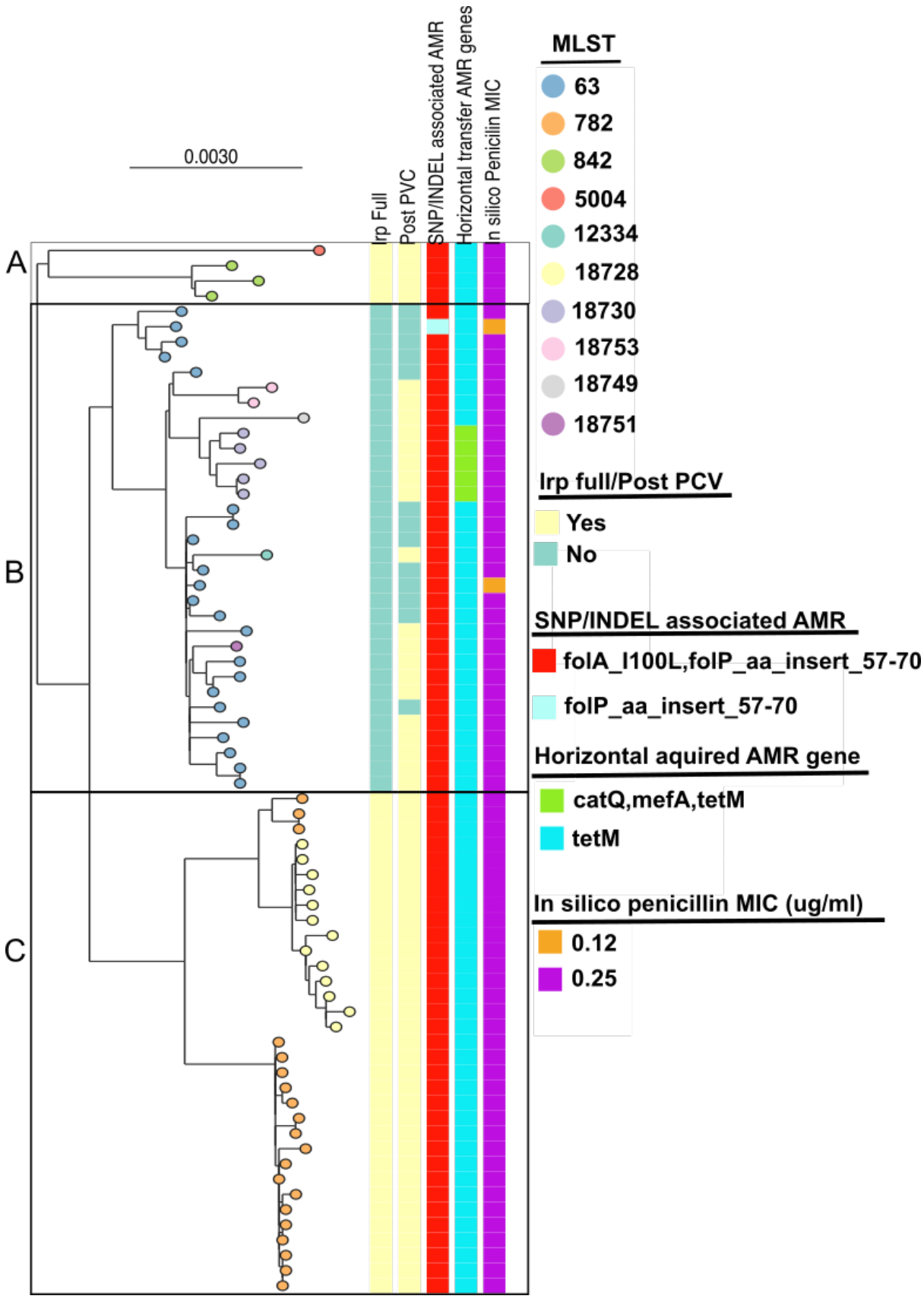


Figure 5: Phylogenetic tree of serotype 14 GPSC9 isolates from Blantyre, Malawi, and the clustering used to define different sub-lineages of GPSC9 for accessory gene content comparison.

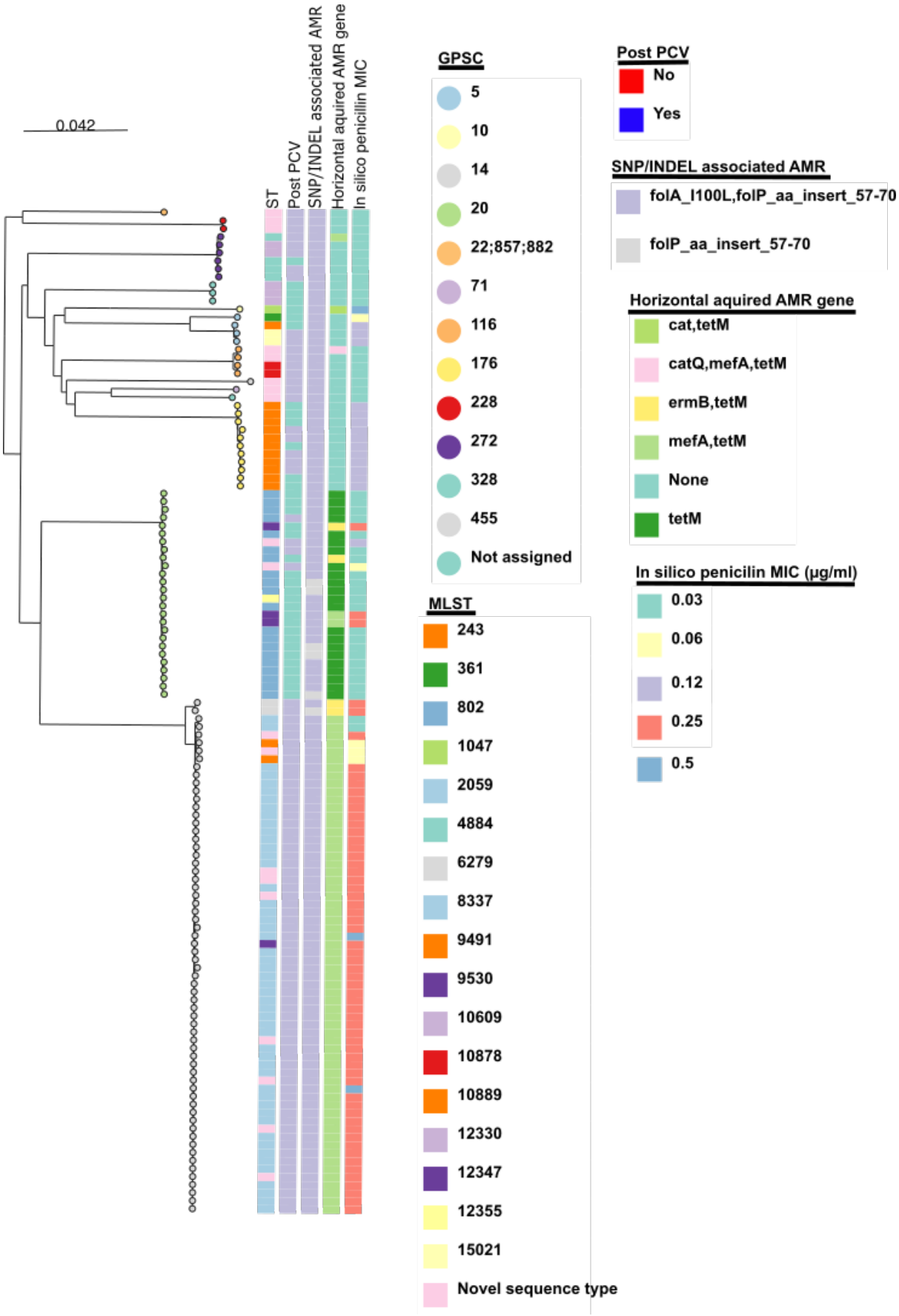


Figure 6: Phylogenetic tree of serotype 23F isolates from Blantyre, Malawi showing the

366 emergent GPSC14 ST2059 isolates having resistance to more antibiotic classes from
367 genotype data over other lineages.
368

369 7. Discussion

370 In the context of persistent *S. pneumoniae* VT carriage and high PCV13 uptake in Malawi, we
371 have identified serotype 23F and 14 lineages, serotypes that have been prominent causes of
372 IPD, with potentially functionally important genetic changes in their *cps* locus (13,51). These
373 lineages harbour additional virulence attributes and AMR, which may further enhance their
374 potential for vaccine escape. Indeed, together with a recently described serotype 3 GPSC
375 ST700–GPSC10, clonal expansion of these lineages may explain the residual circulation of some
376 VT serotypes after vaccination in Blantyre, Malawi (13). The detection of these lineages was
377 only possible through continuous genomic surveillance at the population level, highlighting
378 the importance of such surveys following PCV introduction.
379
380 The predominant serotype 23F lineage before PCV13 introduction in Malawi was GPSC20
381 ST802, a lineage found on multiple continents (52,53). However, post-PCV13 introduction,
382 there was a decrease in ST802, and an expansion of GPSC14 ST2059, which, to date, has only
383 been reported from neighbouring South Africa and Mozambique (52,54). Intriguingly, data
384 from South Africa suggests that 23F ST2059 has a more invasive phenotype (19). Whether
385 vaccine pressure was directly responsible for this shift is uncertain, as genotype replacement
386 of 23F isolates has been reported in China (GPSC24 ST342 to multidrug-resistant GPSC16 ST81
387 also known as the PMEN1 clone lineage) which does not have a routine PCV programme (53).
388 However, once established in Malawi, the genotypic attributes of the 23F GPSC14 ST2059
389 lineage may explain its persistence, particularly in the context of vaccine-induced serotype-
390 specific immunity that has waned below the correlates of protection for both carriage and
391 disease in the first year of life (14).
392
393 We have identified two mutations within the *cps locus* of the emergent 23F GPSC14 lineage
394 that were absent in other 23F genetic lineages. These are predicted to have an impact on
395 capsule expression and production(29,55). These mutations included Wzy, an oligosaccharide
396 repeat unit polymerase responsible for linking sugar chains outside the cell wall as part of the
397 capsule. Wzy shows significant diversity between serotypes but is highly conserved within
398 serotypes (18,56). It was therefore unexpected to find that the 23F GPSC14 lineage has
399 multiple non-synonymous SNPs, including a SNP I253T that reduces the cavity volume found
400 within a protein pocket of Wzy. Mutagenesis studies of Wzy involved in lipopolysaccharide
401 synthesis in gram-negative bacteria such as *Pseudomonas aeruginosa* and *Shigella flexneri*,
402 affect the O-antigen chain length and distribution affect capsule thickness (57,58). Moreover,
403 differences in the intergenic region structure, especially the SNP's in the SpxR1 and CpsR DNA
404 binding sites in the 37-CE have previously been shown to alter the level of capsule production
405 affecting the strain's virulence (29,55). Furthermore, recent phenotypic studies have revealed
406 a 6.6-fold variation in cps production among 23F isolates, though the lineage and *cps locus*
407 genotype was not explored (59). We therefore hypothesise that mutations in Wzy and 37-CE
408 will most likely affect capsule function and expression respectfully, conferring a competitive
409 advantage.

410 The prevalent serotype 14 lineage before PCV13 introduction in Malawi was GPSC9 ST63, a
411 globally distributed sequence type (60). Post-PCV introduction, *S. pneumoniae* GPSC9 isolates
412 from the USA, Israel, South Africa, and Cambodia have been identified that have switched
413 serotype from 14 to nonvaccine type 15A (60). This has not been seen in Malawi, instead
414 GPSC9 ST63 has been overtaken by an expanding serotype 14 GPSC9 ST782, another global
415 sequence type (61,62). Moreover, there was also the expansion of phylogenetically closely
416 related novel sequence type, ST18728, which to date has only been reported in Malawi. This
417 expanding serotype 14 GPSC9 are associated with genetic alteration in the *lrp* gene, where
418 we identified the emergent strain GPSC9 ST782 and a novel sequence type closely related to
419 Blantyre's ST782, which harbours the complete 1359-amino acid-long version of the Lrp
420 protein containing the Cna protein B-type collagen-binding domain. Variants in the lrp gene
421 have been previously reported, particularly the serotype 14-like isolates from Papua New
422 Guinea (63,64). However, the Papua New Guinea isolates differed from Blantyre isolates, not
423 reacting to serotype 14 antibodies, and lacking key genes. The Lrp protein's function is
424 unknown, but it's hypothesised to be a dominant antigen, overriding serological similarities
425 between serotype 14 and 15 (8). The Lrp protein's collagen-binding Cna B domain, also found
426 in the virulent factor RrgB protein, could affect serotype 14's colonisation ability and
427 immunogenicity (59,60).

428 In addition to changes in the *cps* locus, possible factors contributing to the emergence and
429 spread of these 23F and 14 lineages include genes associated with AMR, colonisation, and
430 virulence that would provide bacteria with an advantage over lineages from the same setting.
431 The question is whether these lineages that seem to have a competitive advantage will
432 disseminate more widely. It is noteworthy that a different serotype 23F lineage, PMEN1,
433 identified before pneumococcal vaccines were widely introduced, became transmitted across
434 the globe, causing invasive disease (65–67). The emergence and spread of PMEN1 were
435 attributed to the acquisition of antibiotic resistance, transmission, and virulence genes absent
436 in closely related ancestor strains (68). Together this shows the importance of the identifying
437 emergent strains with genetic adaption and assessing their potential to transmit locally and
438 globally.

439 The main limitation of this study is the limited number of genomes from Malawi prior to and
440 immediately after the introduction of PCV13 and that these historical genomes were biased
441 towards invasive rather than carriage isolates. However, we speculate that, based on closely
442 related isolates from neighbouring countries, such as the ST2059 isolates in South Africa, that
443 these expanding lineages are likely to cause invasive disease in Malawi (69,70).

444 In conclusion, following the introduction of PCV13 in Malawi, there has been clonal expansion
445 of 23F and 14 lineages, characterised by genotypic *cps* locus changes that affect capsule
446 expression and production, and potentially interaction with host immune system. These
447 lineages also have genetic features that confer a competitive advantage in terms of
448 colonisation, transmission and AMR. These findings underscore the value of robust *S.*
449 *pneumoniae* genomic surveillance to inform vaccine programmes in high burden settings
450 where, in the face of a high force of infection, sufficient control of pneumococcal colonisation
451 and disease has not yet been achieved (11,71,72).
452

Author statements

8. Author contributions

[Mandatory for Access Microbiology, encouraged for all other journals. A section describing each author's contribution to the research, using the CRediT taxonomy from CASRAI: https://casrai.org/credit/]

9. Conflicts of interest

The author(s) declare that there are no conflicts of interest.

10. Funding information

This work received funding for the initial sample collection for the PCVPA survey from Bill & Melinda Gates Foundation, Wellcome Trust, and National Institute for Health Research.

11. Ethical approval

The original study protocol for the PCVPA dataset *was* proved by the College of Medicine Research and Ethics Committee, University of Malawi (P.02/15/1677) and the Liverpool School of Tropical Medicine Research Ethics Committee (14.056). Written informed consent was obtained from adult participants and parents/guardians of child participants. Additionally, children aged 8–10 years provided informed assent. The consent encompassed permission for publication.

12. Acknowledgements

[This section is optional for all journals. However, if materials and results were obtained from outside the authors' laboratories (e.g. production of antibodies, properties of strains), this must be acknowledged.] This section should also be used to list the individuals who are part of a listed Group Author/Consortium].

13. References

1. Global Pneumococcal Disease and Vaccination | CDC [Internet]. 2023 [cited 2024 Feb 26]. Available from: https://www.cdc.gov/pneumococcal/global.html

2. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, et al. Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. Lancet Glob Health. 2018 Jun 13;6(7):e744–57.

3. Rodgers GL, Whitney CG, Klugman KP. Triumph of Pneumococcal Conjugate Vaccines: Overcoming a Common Foe. J Infect Dis. 2021 Sep 30;224(Suppl 4):S352–9.

4. Berman-Rosa M, O'Donnell S, Barker M, Quach C. Efficacy and Effectiveness of the PCV-10 and PCV-13 Vaccines Against Invasive Pneumococcal Disease. Pediatrics. 2020 Apr 1;145(4):e20190377.

491 5. Wartha F, Beiter K, Albiger B, Fernebro J, Zychlinsky A, Normark S, et al. Capsule and D-
492 alanylated lipoteichoic acids protect Streptococcus pneumoniae against neutrophil
493 extracellular traps. Cell Microbiol. 2007 May;9(5):1162–71.

494 6. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. The Streptococcus pneumoniae
495 Capsule Inhibits Complement Activity and Neutrophil Phagocytosis by Multiple
496 Mechanisms. Infect Immun. 2010 Feb;78(2):704–15.

497 7. Abeyta M, Hardy GG, Yother J. Genetic Alteration of Capsule Type but Not PspA Type
498 Affects Accessibility of Surface-Bound Complement and Surface Antigens of
499 Streptococcus pneumoniae. Infect Immun. 2003 Jan;71(1):218–25.

500 8. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al.
501 Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal
502 Serotypes. PLOS Genet. 2006 Mar 10;2(3):e31.

503 9. GPS :: Global Pneumococcal Sequencing Project [Internet]. [cited 2023 Sep 28]. Available
504 from: https://www.pneumogen.net/gps/serotypes.html

505 10. Flasche S, Hoek AJV, Sheasby E, Waight P, Andrews N, Sheppard C, et al. Effect of
506 Pneumococcal Conjugate Vaccination on Serotype-Specific Carriage and Invasive Disease
507 in England: A Cross-Sectional Study. PLOS Med. 2011 Apr 5;8(4):e1001017.

508 11. Bar-Zeev N, Swarthout TD, Everett DB, Alaerts M, Msefula J, Brown C, et al. Impact and
509 effectiveness of 13-valent pneumococcal conjugate vaccine on population incidence of
510 vaccine and non-vaccine serotype invasive pneumococcal disease in Blantyre, Malawi,
511 2006–18: prospective observational time-series and case-control studies. Lancet Glob
512 Health. 2021 Jul 1;9(7):e989–98.

513 12. Koenraads M, Swarthout TD, Bar-Zeev N, Brown C, Msefula J, Denis B, et al. Changing
514 Incidence of Invasive Pneumococcal Disease in Infants Less Than 90 Days of Age Before
515 and After Introduction of the 13-Valent Pneumococcal Conjugate Vaccine in Blantyre,
516 Malawi: A 14-Year Hospital Based Surveillance Study. Pediatr Infect Dis J. 2022 Sep
517 1;41(9):764–8.

518 13. Swarthout TD, Fronterre C, Lourenço J, Obolski U, Gori A, Bar-Zeev N, et al. High residual
519 carriage of vaccine-serotype Streptococcus pneumoniae after introduction of
520 pneumococcal conjugate vaccine in Malawi. Nat Commun. 2020 May 6;11(1):2222.

521 14. Swarthout TD, Henrion MYR, Thindwa D, Meiring JE, Mbewe M, Kalizang'Oma A, et al.
522 Waning of antibody levels induced by a 13-valent pneumococcal conjugate vaccine,
523 using a 3 + 0 schedule, within the first year of life among children younger than 5 years
524 in Blantyre, Malawi: an observational, population-level, serosurveillance study. Lancet
525 Infect Dis. 2022 Dec 1;22(12):1737–47.

526 15. Mostowy RJ, Croucher NJ, De Maio N, Chewapreecha C, Salter SJ, Turner P, et al.
527 Pneumococcal Capsule Synthesis Locus cps as Evolutionary Hotspot with Potential to
528 Generate Novel Serotypes by Recombination. Mol Biol Evol. 2017 Oct;34(10):2537–54.

16. Tsang R. Capsule switching and capsule replacement in vaccine-preventable bacterial diseases. Lancet Infect Dis. 2007 Sep 1;7(9):569–70.

17. Roca A, Bojang A, Bottomley C, Gladstone RA, Adetifa JU, Egere U, et al. Effect on nasopharyngeal pneumococcal carriage of replacing PCV7 with PCV13 in the Expanded Programme of Immunization in The Gambia. Vaccine. 2015 Dec 16;33(51):7144–51.

18. Arends DW, Miellet WR, Langereis JD, Ederveen THA, van der Gaast-de Jongh CE, van Scherpenzeel M, et al. Examining the Distribution and Impact of Single-Nucleotide Polymorphisms in the Capsular Locus of Streptococcus pneumoniae Serotype 19A. Infect Immun. 2021 Oct 15;89(11):e0024621.

19. Gladstone RA, Lo SW, Lees JA, Croucher NJ, van Tonder AJ, Corander J, et al. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine. 2019 Apr 16;43:338–46.

20. Obolski U, Swarthout TD, Kalizang'oma A, Mwalukomo TS, Chan JM, Weight CM, et al. The metabolic, virulence and antimicrobial resistance profiles of colonising Streptococcus pneumoniae shift after PCV13 introduction in urban Malawi. Nat Commun. 2023 Nov 17;14(1):7477.

21. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 2019 Feb;29(2):304–16.

22. Enright MC, Spratt BG. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiol Read Engl. 1998 Nov;144 (Pt 11):3049–60.

23. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genomics. 2021;7(11):000685.

24. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020 May 1;37(5):1530–4.

25. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun;14(6):587–9.

26. Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, et al. Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. iMeta. 2022;1(4):e56.

27. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W20–5.

28. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018 Sep 15;34(18):3094–100.

29. Glanville DG, Gazioglu O, Marra M, Tokars VL, Kushnir T, Habtom M, et al. Pneumococcal capsule expression is controlled through a conserved, distal cis-regulatory element during infection. PLOS Pathog. 2023 Jan 31;19(1):e1011035.

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

31. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci Publ Protein Soc. 2018 Jan;27(1):135–45.

32. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 2003 Jul 1;31(13):3381–5.

33. Bernhofer M, Rost B. TMbed: transmembrane proteins predicted through language model embeddings. BMC Bioinformatics. 2022 Aug 8;23(1):326.

34. Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J Cheminformatics. 2018 Aug 14;10(1):39.

35. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J Mol Biol. 2019 May 17;431(11):2197–212.

36. Jakubec D, Skoda P, Krivak R, Novotny M, Hoksza D. PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. Nucleic Acids Res. 2022 Jul 5;50(W1):W593–7.

37. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015 Feb 18;43(3):e15.

38. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genomics. 2016 Nov 30;2(11):e000093.

39. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020 Jul 22;21(1):180.

40. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016 Nov 25;17(1):238.

41. Nobbs AH, Lamont RJ, Jenkinson HF. Streptococcus Adherence and Colonization. Microbiol Mol Biol Rev MMBR. 2009 Sep;73(3):407–50.

42. Chan WT, Domenech M, Moreno-Córdoba I, Navarro-Martínez V, Nieto C, Moscoso M, et al. The Streptococcus pneumoniae yefM-yoeB and relBE Toxin-Antitoxin Operons Participate in Oxidative Stress and Biofilm Formation. Toxins. 2018 Sep 18;10(9):378.

601 43. Chan WT, Moreno-Córdoba I, Yeo CC, Espinosa M. Toxin-Antitoxin Genes of the Gram-
602 Positive Pathogen Streptococcus pneumoniae: So Few and Yet So Many. Microbiol Mol
603 Biol Rev MMBR. 2012 Dec;76(4):773–91.

604 44. Maricic N, Anderson ES, Opipari AE, Yu EA, Dawid S. Characterization of a Multipeptide
605 Lantibiotic Locus in Streptococcus pneumoniae. mBio. 2016 Jan
606 26;7(1):10.1128/mbio.01656-15.

607 45. Bandara M, Skehel JM, Kadioglu A, Collinson I, Nobbs AH, Blocker AJ, et al. The accessory
608 Sec system (SecY2A2) in Streptococcus pneumoniae is involved in export of pneumolysin
609 toxin, adhesion and biofilm formation. Microbes Infect. 2017;19(7–8):402–12.

610 46. Rose L, Shivshankar P, Hinojosa E, Rodriguez A, Sanchez CJ, Orihuela CJ. Antibodies
611 against PsrP, a Novel Streptococcus pneumoniae Adhesin, Block Adhesion and Protect
612 Mice against Pneumococcal Challenge. J Infect Dis. 2008 Aug 1;198(3):375–83.

613 47. Lizcano A, Akula Suresh Babu R, Shenoy AT, Saville AM, Kumar N, D'Mello A, et al.
614 Transcriptional organization of pneumococcal psrP-secY2A2 and impact of GtfA and GtfB
615 deletion on PsrP-associated virulence properties. Microbes Infect. 2017 Jun;19(6):323–
616 33.

617 48. S L, Nj C, F B, C F. Epidemiological dynamics of bacteriocin competition and antibiotic
618 resistance. Proc Biol Sci [Internet]. 2022 Dec 10 [cited 2024 Feb 23];289(1984). Available
619 from: https://pubmed.ncbi.nlm.nih.gov/36196547/

620 49. Glover DT, Hollingshead SK, Briles DE. Streptococcus pneumoniae Surface Protein PcpA
621 Elicits Protection against Lung Infection and Fatal Sepsis. Infect Immun. 2008
622 Jun;76(6):2767.

623 50. Blue CE, Paterson GK, Kerr AR, Bergé M, Claverys JP, Mitchell TJ. ZmpB, a Novel
624 Virulence Factor of Streptococcus pneumoniae That Induces Tumor Necrosis Factor
625 Alpha Production in the Respiratory Tract. Infect Immun. 2003 Sep;71(9):4925–35.

626 51. Cornick JE, Everett DB, Broughton C, Denis BB, Banda DL, Carrol ED, et al. Invasive
627 Streptococcus pneumoniae in Children, Malawi, 2004–2006. Emerg Infect Dis. 2011
628 Jun;17(6):1107–9.

629 52. Donkor ES, Adegbola RA, Wren BW, Antonio M. Population Biology of Streptococcus
630 pneumoniae in West Africa: Multilocus Sequence Typing of Serotypes That Exhibit
631 Different Predisposition to Invasive Disease and Carriage. PLOS ONE. 2013 Jan
632 16;8(1):e53925.

633 53. Ma X, Yao KH, Yu SJ, Zhou L, Li QH, Shi W, et al. Genotype replacement within serotype
634 23F Streptococcus pneumoniae in Beijing, China: characterization of serotype 23F.
635 Epidemiol Infect. 2013 Aug;141(8):1690–6.

636 54. Manenzhe RI, Dube FS, Wright M, Lennard K, Mounaud S, Lo SW, et al. Characterization
637 of Pneumococcal Colonization Dynamics and Antimicrobial Resistance Using Shotgun

638 Metagenomic Sequencing in Intensively Sampled South African Infants. Front Public
639 Health. 2020 Sep 22;8:543898.

640 55. Wen Z, Liu Y, Qu F, Zhang JR. Allelic Variation of the Capsule Promoter Diversifies
641 Encapsulation and Virulence In Streptococcus pneumoniae. Sci Rep. 2016 Jul
642 28;6(1):30176.

643 56. Jiang SM, Wang L, Reeves PR. Molecular Characterization of Streptococcus pneumoniae
644 Type 4, 6B, 8, and 18C Capsular Polysaccharide Gene Clusters. Infect Immun. 2001
645 Mar;69(3):1244–55.

646 57. Islam ST, Huszczynski SM, Nugent T, Gold AC, Lam JS. Conserved-residue mutations in
647 Wzy affect O-antigen polymerization and Wzz-mediated chain-length regulation in
648 Pseudomonas aeruginosa PAO1. Sci Rep. 2013 Dec 1;3:3441.

649 58. Daniels C, Vindurampulle C, Morona R. Overexpression and topology of the Shigella
650 flexneri O-antigen polymerase (Rfc/Wzy). Mol Microbiol. 1998;28(6):1211–22.

651 59. Zhu J, Abruzzo AR, Wu C, Bee GCW, Pironti A, Putzel G, et al. Effects of Capsular
652 Polysaccharide amount on Pneumococcal-Host interactions. PLOS Pathog. 2023 Aug
653 4;19(8):e1011509.

654 60. Hawkins PA, Chochua S, Lo SW, Belman S, Antonio M, Kwambana-Adams B, et al. A
655 global genomic perspective on the multidrug-resistant Streptococcus pneumoniae 15A-
656 CC63 sub-lineage following pneumococcal conjugate vaccine introduction. Microb
657 Genomics. 2023;9(4):000998.

658 61. Moore CE, Giess A, Soeng S, Sar P, Kumar V, Nhoung P, et al. Characterisation of Invasive
659 Streptococcus pneumoniae Isolated from Cambodian Children between 2007 – 2012.
660 PLoS ONE. 2016 Jul 22;11(7):e0159358.

661 62. Yamba Yamba L, Uddén F, Fuursted K, Ahl J, Slotved HC, Riesbeck K.
662 Extensive/Multidrug-Resistant Pneumococci Detected in Clinical Respiratory Tract
663 Samples in Southern Sweden Are Closely Related to International Multidrug-Resistant
664 Lineages. Front Cell Infect Microbiol. 2022 Mar 22;12:824449.

665 63. Manna S, Spry L, Wee-Hee A, Ortika BD, Boelsen LK, Batinovic S, et al. Variants of
666 Streptococcus pneumoniae Serotype 14 from Papua New Guinea with the Potential to
667 Be Mistyped and Escape Vaccine-Induced Protection. Microbiol Spectr. 2022 Jul
668 12;10(4):e01524-22.

669 64. van Tonder AJ, Bray JE, Quirk SJ, Haraldsson G, Jolley KA, Maiden MCJ, et al. Putatively
670 novel serotypes and the potential for reduced vaccine effectiveness: capsular locus
671 diversity revealed among 5405 pneumococcal genomes. Microb Genomics.
672 2016;2(10):e000090.

673 65. Muñoz R, Coffey TJ, Daniels M, Dowson CG, Laible G, Casal J, et al. Intercontinental
674 spread of a multiresistant clone of serotype 23F Streptococcus pneumoniae. J Infect Dis.
675 1991 Aug;164(2):302–6.

676 66. De Lencastre H, Tomasz A. From ecological reservoir to disease: the nasopharynx, day-
677 care centres and drug-resistant clones of Streptococcus pneumoniae. J Antimicrob
678 Chemother. 2002 Dec;50 Suppl S2:75–81.

679 67. Coffey TJ, Dowson CG, Daniels M, Zhou J, Martin C, Spratt BG, et al. Horizontal transfer
680 of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural
681 populations of Streptococcus pneumoniae. Mol Microbiol. 1991 Sep;5(9):2255–60.

682 68. Wyres KL, Lambertsen LM, Croucher NJ, McGee L, von Gottberg A, Liñares J, et al. The
683 multidrug-resistant PMEN1 pneumococcus is a paradigm for genetic success. Genome
684 Biol. 2012 Nov 16;13(11):R103.

685 69. Gladstone RA, Lo SW, Lees JA, Croucher NJ, Tonder AJ van, Corander J, et al.
686 International genomic definition of pneumococcal lineages, to contextualise disease,
687 antibiotic resistance and vaccine impact. eBioMedicine. 2019 May 1;43:338–46.

688 70. Weiser JN, Ferreira DM, Paton JC. Streptococcus pneumoniae: transmission,
689 colonization and invasion. Nat Rev Microbiol. 2018 Jun;16(6):355–67.

690 71. Kirolos A, Swarthout TD, Mataya AA, Bonomali F, Brown C, Msefula J, et al. Invasiveness
691 potential of pneumococcal serotypes in children after introduction of PCV13 in Blantyre,
692 Malawi. BMC Infect Dis. 2023 Jan 26;23(1):56.

693 72. Lourenço J, Obolski U, Swarthout TD, Gori A, Bar-Zeev N, Everett D, et al. Determinants
694 of high residual post-PCV13 pneumococcal vaccine-type carriage in Blantyre, Malawi: a
695 modelling study. BMC Med. 2019 Dec 5;17(1):219.

696

697
698