

# Genomic Characteristics of Invasive *Streptococcus pneumoniae* Serotype 1 in New Caledonia Prior to the Introduction of PCV13

Mariem Hanachi<sup>1,2</sup>, Anmol Kiran<sup>3,4</sup>, Jennifer Cornick<sup>4,5</sup>, Emna Harigua-Souiai<sup>6</sup>, Dean Everett<sup>3,4</sup>, Alia Benkahla<sup>1</sup> and Oussama Souiai<sup>1,7</sup> 

<sup>1</sup>Laboratory of Bioinformatics, Biomathematics and Biostatistics—LR16IPT09, Institut Pasteur de Tunis, University of Tunis El Manar (UTM), Tunis, Tunisia. <sup>2</sup>Faculty of Science of Bizerte, University of Carthage, Jarzouna, Tunisia. <sup>3</sup>Queens Research Institute, University of Edinburgh, Edinburgh, UK. <sup>4</sup>Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi. <sup>5</sup>Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. <sup>6</sup>Laboratory of Molecular Epidemiology and Experimental Pathology—LR16IPT04, Institut Pasteur de Tunis, Université de Tunis El Manar, Tunis, Tunisia. <sup>7</sup>Institut Supérieur des Technologies Médicales de Tunis, Université de Tunis El Manar, Tunis, Tunisia.

Bioinformatics and Biology Insights  
Volume 14: 1–9  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1177932220962106



**ABSTRACT:** *Streptococcus pneumoniae* serotype 1 is a common cause of global invasive pneumococcal disease. In New Caledonia, serotype 1 is the most prevalent serotype and led to two major outbreaks reported in the 2000s. The pneumococcal conjugate vaccine 13 (PCV13) was introduced into the vaccination routine, intending to prevent the expansion of serotype 1 in New Caledonia. Aiming to provide a baseline for monitoring the post-PCV13 changes, we performed a whole-genome sequence analysis on 67 serotype 1 isolates collected prior to the PCV13 introduction. To highlight the *S. pneumoniae* serotype 1 population structure, we performed a multilocus sequence typing (MLST) analysis revealing that NC serotype 1 consisted of 2 sequence types: ST3717 and the highly dominant ST306. Both sequence types harbored the same resistance genes to beta-lactams, macrolide, streptogramin B, fluoroquinolone, and lincosamide antibiotics. We have also identified 36 virulence genes that were ubiquitous to all the isolates. Among these virulence genes, the pneumolysin sequence presented an allelic profile associated with disease outbreaks and reduced hemolytic activity. Moreover, recombination hotspots were identified in 4 virulence genes and more notably in the *cps* locus (*cps2L*), potentially leading to capsular switching, a major mechanism of the emergence of nonvaccine types. In summary, this study represents the first overview of the genomic characteristics of *S. pneumoniae* serotype 1 in New Caledonia prior to the introduction of PCV13. This preliminary description represents a baseline to assess the impact of PCV13 on serotype 1 population structure and genomic diversity.

**KEYWORDS:** *Streptococcus pneumoniae*, serotype 1, whole-genome sequencing, recombination, virulence factors, antibiotics resistance genes, pneumolysin, New Caledonia

**RECEIVED:** July 27, 2020. **ACCEPTED:** August 31, 2020.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was partly funded by H3ABioNet, which is supported by the National Institutes of Health Common Fund (grant no. U41HG006941). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Oussama Souiai, Laboratory of Bioinformatics, Biomathematics and Biostatistics—LR16IPT09, Institut Pasteur de Tunis, University of Tunis El Manar (UTM), 13 place Pasteur, 1004 Tunis, Tunisia. Email: souiai@gmail.com

## Introduction

*Streptococcus pneumoniae* is an invasive gram-positive bacteria, responsible for a high rate of morbidity and mortality, especially for children under 5 years old, in the developing world.<sup>1</sup>

*S. pneumoniae* commonly colonizes the nasopharynx tract.<sup>2</sup> Its migration to other body sites leads to a wide range of non-invasive diseases such as otitis, sinusitis, and numerous invasive diseases including pneumonia, meningitis, and septicemia.<sup>2</sup> Among over the 90 serotypes defined by the structure and antigenicity of the capsular polysaccharide (*cps*), serotype 1 remains the most common cause of invasive pneumococcal disease (IPD) worldwide and was frequently linked to outbreaks.<sup>3–6</sup> Similarly to surrounding areas (Wallis and Futuna and French Polynesia), serotype 1 was reported as predominant in New Caledonia (NC) and led to 2 major outbreaks in the early 2000s.<sup>7,8</sup> The first outbreak occurred from May 1999 to May 2001 and the second

occurred from July to November 2007, respectively affecting children above 5 years old<sup>7</sup> and children under 8 years old.<sup>8</sup>

Prevention strategies aiming to reduce IPD incidence consisted first in the introduction of pneumococcal conjugate vaccine 7 (PCV7) and later the PCV13, targeting a larger range of serotypes, including serotype 1.<sup>9</sup> Even though vaccination campaigns demonstrated a certain efficiency in lowering the IPD, nonvaccine types (NVTs) emerged after PCV7 and PCV13 introduction.<sup>10–14</sup> Regarding the serotype 1, it was reported that some of its clones are still circulating after PCV13 introduction.<sup>15–17</sup> For instance, ST8314 and the newly reported ST9067 respectively expressed less sensitivity to antibiotics and a higher recombination rate.<sup>17</sup> These findings demonstrate the need to monitor the serotype 1 evolution in the NC area to ensure appropriate prevention strategies for IPD, particularly after the introduction of PCV13 in 2010.



Prior to PCV13 introduction, NC *S. pneumoniae* serotype 1 investigations were limited to traditional antibiotic susceptibility testing and serotyping approaches.<sup>7,8</sup> To survey, monitor pathogen outbreaks, and explore genomic features, these conventional comparative methods need to be completed by the next-generation sequencing analyses.<sup>18,19</sup>

Herein, we present an analysis of the genomic characteristics of NC *S. pneumoniae* serotype 1 isolates in circulation prior to the introduction of the PCV13, aiming at a better deciphering of the *S. pneumoniae* population structure. We particularly described the recombination rate, the population structure, the genome content in terms of antibiotic resistance, and virulence genes characterizing NC serotype 1. These genomic features and estimations determined in this retrospective investigation would provide a valuable baseline for further evaluation of PCV13 on *S. pneumoniae* serotype 1 in NC.

## Material and Methods

### *Sample collection, DNA extraction, and whole-genome sequencing*

All 67 Serotype 1 Samples used in this study were collected in NC from 2004 to 2009 by the Pneumococcal African Genomic Consortium (PAGE, <http://www.pagegenomes.org>), and include 11 isolates from July to November 2007 outbreak. The sample collection included 1 carriage and 66 invasive isolates. Invasive isolates were extracted from sterile sites such as blood (46/66), cerebro-spinal fluid (CSF; 4/66), pleural fluid (2/66), and from nonsterile sites including sputum (8/66), ear (5/66), and abscess (1/66) (Supplementary Table S1). Genomic DNA was extracted as previously described.<sup>20</sup> High throughput sequences were generated using Illumina Genome Analyzer-II at Sanger Institute at Wellcome-Trust Campus, United Kingdom. The obtained 100bp paired-end sequences were deposited in ENA, under the PRJEB2102 accession number (Supplementary Table S1).

### *Draft genome assembly*

Multilocus sequence typing (MLST) was performed by mapping short reads against housekeeping gene sequences available on the MLST database (<https://pubmlst.org/>) using short read sequence typing (SRST) tool.<sup>21</sup>

To assemble a circular draft genome, we opted for a mixed assembly strategy (*de novo* assembly and reference-based assembly). The *de novo* genome assembly was implemented using an optimized pipeline for prokaryotes.<sup>22</sup> This pipeline uses Velvet, version 1.2.09,<sup>23</sup> and velvetOptimiser, version 2.2.5<sup>24</sup> (with k-mers ranging from 66% to 90% of the read length). The mean number of contigs was equal to 71.94 (SD: 10.55) and the average contig length equal to 29.479bp (SD: 3951) (Supplementary Table S2). The average sequencing coverage was equal to 338bp (SD: 37.21). We used BLAST to determine the closest reference genome to NC isolates, thus determining *S. pneumoniae*

INV104 (NCBI RefSeq: NC\_017591.1), as the closest to our isolates.<sup>25</sup> Finally, CONTIGuator (default parameters) was used to generate circular genomes with a mean size of 2082132bp (sd: 7223).<sup>26</sup> On average, there were 6.5 unmapped scaffolds having a mean size of 831bp (SD: 730.890) (Supplementary Table S3). The mapped scaffolds were interspersed with short gaps ranging from 115 to 236 and replaced by "N" using CONTIGuator.

For assessment purposes, newly assembled circular genomes were aligned against *S. pneumoniae* INV104 genome by MUMmer v4.0 beta.<sup>27</sup> The generated alignment validated CONTIGuator outputs (See Supplementary Figure S1). The downstream analyses were performed using the newly built circular genomes.

### *Genome annotation and pan-genome analysis*

Functional annotation of the newly assembled genomes was performed using PROKKA, version 1.2 (default parameters).<sup>28</sup> The annotation GFF files served as input to the Roary, version 3.11.2 pipeline (default parameters), allowing the pan-genome size estimation.<sup>29,30</sup>

### *Analysis of the accessory genome*

Virulence genes and antibiotic resistance genes (ARGs) were identified by BLASTing (BLASTX with e-value, identity, and coverage cutoffs set, respectively, to 0.1, 90% and 75%), the assembled genomes against the protein sequences from the virulence factors database (VFDB) and the comprehensive antibiotic resistance database (CARD).<sup>31-33</sup> Within this framework, particular interest was devoted to the allelic variation of the pneumolysin gene in NC isolates. To achieve this, we compared the pneumolysin sequences from NC samples to 22 pre-existing and fully sequenced serotype 1 pneumolysin.<sup>34</sup> The multiple alignments of nucleotide and protein sequences were generated using the AliView software, version 1.26.<sup>35</sup> Sequence variations detection was performed using the Clustal Omega web server (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).<sup>36</sup> Prophages in the analyzed genomes were first screened using PHASTER<sup>37</sup> and VirSorter<sup>38</sup> web servers. Simultaneously, integrative and conjugative elements (ICEs) were investigated with a BLASTN (e-value, identity, and coverage cutoffs set, respectively, to 0.1, 90%, and 75%) search against ICEberg database and using the ICEfinder web tool (<https://db-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html>).<sup>39</sup>

### *Recombination rate estimation and hotspots identification*

We used progressiveMauve, version 2.4.0, to generate a whole-genome alignment of the circularized genomes.<sup>40</sup> The resulting alignment was then used as input for Gubbins, version 2.3.5, to detect recombination events. RAXML was used for phylogeny

analysis (100 bootstrap iterations).<sup>41</sup> Detected recombination events were visualized using the Phandango web tool (<https://jameshadfield.github.io/phandango/>).<sup>42</sup> Identified recombination hotspots were annotated using the consensus sequences generated from 67 genomes alignment.

#### Lineage dating by Bayesian evolutionary analysis

As an initial step of the lineage dating process, a core-genome single nucleotide polymorphism (SNP) alignment was achieved by Snippy, an open-source software available on GitHub: <https://github.com/tseemann/snippy>. Gubbins (default parameters) was used to generate a recombinant-free Snippy alignment. The resulting nonrecombinant core SNP alignment was introduced in BEAUti, version 2.4.8, after temporal signal assessment by Tempest software, version 1.5.1.<sup>43,44</sup> Tip dates were specified as years of sample collection. Through the use of the smart model selection web server,<sup>45</sup> we determined that the general time reversible (GTR) with 4 discrete gamma-distributed rate categories was the most appropriate model. The substitution rates were fixed to  $1.57e^{-6}$  as previously determined.<sup>12</sup> We used the strict clock model in combination with the coalescent Bayesian skyline demographic model. BEAST, version 2.4.8, was then run with 800 000 000 Markov Chain Monte Carlo (MCMC) generations with a 10% burn-in. The log files generated by BEAST were summarized by Tracer, version 1.7.1.<sup>46</sup> A 200 cutoff of the estimated sample size (ESS) was used to retain concluding simulations. Summary and visual trees were respectively generated using TreeAnnotator, version 2.4.8<sup>43</sup> (available from <http://beast.bio.ed.ac.uk>) and FigTree, version 1.4.4<sup>47</sup> (available from <http://tree.bio.ed.ac.uk/software/figtree/>).

## Results and Discussion

### Population structure of NC *S. pneumoniae* serotype 1

New Caledonia serotype 1 is subdivided into two sequence types (ST): ST306 and ST3717, also known as SLV-306 and corresponding to a single locus variant of ST306 (G301A) in the *aroE* housekeeping gene. ST306 is the predominant ST with 62 isolates, while ST3717 is represented by only 5 isolates. The ST306 was reported as the most prevalent serotype 1 ST in NC and was associated with the 2000s outbreaks.<sup>7,8</sup> The ST306 is a worldwide distributed serotype, suspected to be an important determinant behind the increase of the serotype 1 IPD expansion.<sup>5,48</sup> The ST3717 was not ubiquitously distributed and exclusively reported in NC according to the MLST database <https://pubmlst.org/> and Hello and collaborators.<sup>8</sup> As reported here, the particular population structure of serotype 1 corroborates with the results of Brueggemann and Spratt.<sup>5</sup> The authors suggested that a low carriage rate of this serotype might explain this observation, by decreasing the likelihood of clone transfer between countries and continents.<sup>5</sup>

### Gene contents investigation of NC *S. pneumoniae*

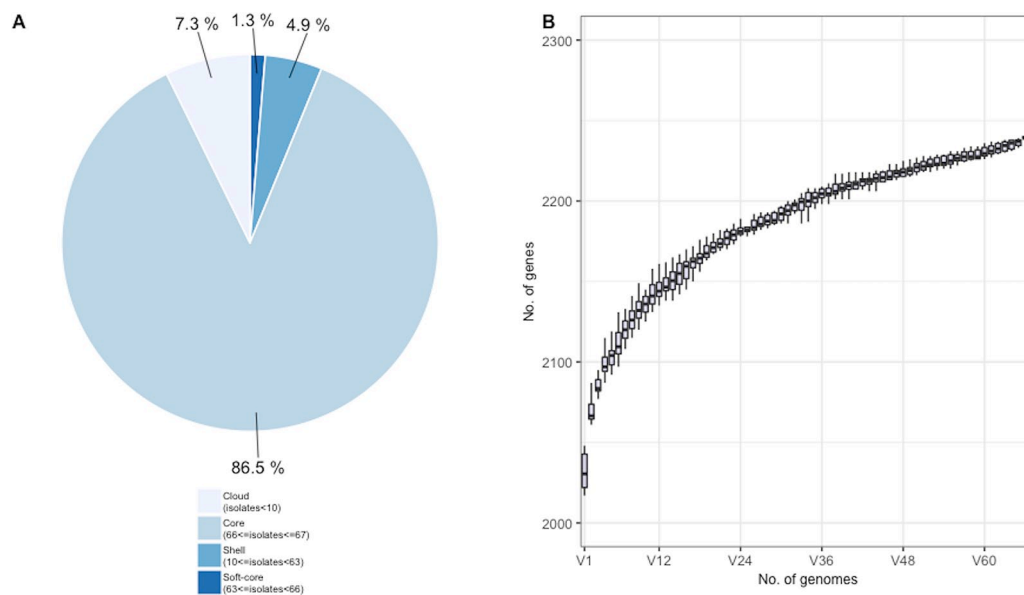
The functional annotation of the newly reconstructed genomes revealed that NC *S. pneumoniae* serotype 1 genomes harbored a mean number of 2052 of protein-coding genes (SD: 14.88). The pan-genome was composed of 2240 genes. 1938 genes (86.5%) are shared by at least 99% of isolates (hard core-genome) (Figure 1A). A summary of shared and unique genes are available in Supplementary Table S4. These results differ from Chaguza and collaborator's findings which reported a smaller core genome composed of 1520 genes across 226 *S. pneumoniae* serotype 1.<sup>50</sup> This difference can be explained by the smaller number of samples in our study (67 vs 226), to the limited geographic location of samples (NC vs African and Asian countries) and to the shorter time of collection (2004–2009 vs 1994–2011).<sup>50</sup> Hence, a larger number of samples would likely give a more accurate estimation of the pan-genome of NC *S. pneumoniae* serotype 1.

Even with latter observations, we can hypothesize that the observed pan-genome is moderately open, reflecting the simultaneous increase of gene pool with the strains number (Figure 1B).<sup>51,52</sup> The pan-genome of *S. pneumoniae* is usually considered as open due to the ability to capture exogenous DNA through horizontal gene transfer.<sup>53</sup> This genome would be less flexible for serotype 1 due to the reduced carriage rate (~9 days), which limits the opportunity of DNA exchange with the nasopharynx microbial community.<sup>4</sup>

### ARGs in the accessory genome

For decades, penicillin was the standard treatment for pneumococcal infections, leading to a substantial increase of resistant clones to beta-lactams.<sup>54–57</sup> Depending on disease severity, clinical manifestations, and local community antibiotic resistance patterns, a wide range of alternative treatments can be administered.<sup>58,59</sup> These treatments contributed to the emergence of resistant clones, and notably to fluoroquinolones and macrolide antibiotics.<sup>56</sup> The serotype 1 was described as rarely resistant to macrolides, penicillin, and quinolones.<sup>4</sup> Conversely, serotype 1 exhibits the highest rate of multidrug resistance (MDR) compared to the other serotypes mainly by expressing resistance to cotrimoxazole, tetracycline, and chloramphenicol.<sup>60</sup>

Our ARG investigation showed that all pre-PCV13 serotype 1 genomes are characterized by the ubiquitous presence of the same ARGs. Among these, we noticed the presence of the efflux pumps encoding genes namely *pmrA* and *patB*.<sup>61</sup> The over-expression of these efflux pumps genes were reported as conferring a low-level fluoroquinolone resistance.<sup>61</sup> We also reported the presence of *RlmA(II)*, a methyltransferase encoding gene, involved in 23S methylation rRNA, linked to high resistance to lincosamides and low resistance to macrolides and streptogramin B antibiotics.<sup>62</sup> While no phenotypic data for lincosamides, fluoroquinolone, and streptogramin B were



**Figure 1.** Pan-genome size of NC *S. pneumoniae* isolates: (A) Proportions of gene coding sequences in the pan-genome divided into 4 categories: the hard-core and the soft-core genome, representing respectively 99% and 95% to 99% of genes shared between isolates. The shell genes are present in 15% to 95% of isolates, whereas the cloud genes are distributed in less than 15% of isolates.<sup>29,30,49</sup> (B) Cumulative number of genes in the pan-genome. Box plots indicate first and third quartiles with medians shown as horizontal lines.

reported, the macrolide resistance was previously reported in the NC surrounding area. In addition, *PBP1a*, *PBP2x*, and *PBP2b* were also detected. Alterations in these genes were associated with amoxicillin and penicillin resistance, in conformance with resistance traits previously observed in NC.<sup>7,8,63</sup>

#### Virulence factors in the accessory genome

Through virulence genes screening, we identified a set of 36 virulence factors among which, some are crucial for the pathogenesis of *S. pneumoniae* (Supplementary Table S5 for virulence genes annotation and VFDB classification). All virulence genes were equally distributed across all isolates, and this independently of the isolation site. This finding supports the idea that isolates from the same geographical location tend to have similar virulence and invasiveness patterns.<sup>64</sup> These “core” virulence genes are known to be implicated in pneumococcal pathogenesis during the colonization and invasion.

Certainly, the most important virulence genes detected are those from the *cps* locus encoding for the polysaccharide capsule (*cpsD*, *cpsG*, *cps2L*, *cps4A*, *cps4D*, *wz*, *wz*, *wze*, and *wzd*).<sup>65</sup> These genes impact on *S. pneumoniae* pathogenesis through various pathways, such as assisting the pathogen to evade the immune system and to colonize the nasopharyngeal tract.<sup>66</sup> In addition, several surface proteins playing a role in host cell adhesion, and therefore, the nasopharyngeal colonization, were retrieved. The latter, are coding for either Choline-Binding Proteins (*cpbD*, *cpbG*, *pce*, *lytA*, *lytC*, and *lytB*),<sup>67</sup> the pilus-encoding pathogenicity islets (*pitA*, *sipA*, *pitB*, and *srtG1*),<sup>68</sup> two neuraminidases (*NanA* and *NanB*),<sup>69</sup> or other surface protein like *eno*, *parvA*, *srtA*, *plr*, *srtH*, *shrA*, and *pfbA*.<sup>2,70</sup> In addition, genes that ensure the host tissues degradation (*hysA*, *eno*, *pce*,

and *gapA*)<sup>71</sup> or provide to *S. pneumoniae* essential nutrients (*piaA*, *piuA*, *PsaA*, *htrA*, *lmb*, and *tig*) were found.<sup>71-74</sup> Moreover, we detected genes involved in the immune invasion. These mechanisms consist in immunoglobulin lysis (*iga*), C3 complement system interference (*ply*, *nanA*, *srtH*, *lytA*, *lytB*, *gapA*, and *cppA*),<sup>75</sup> inflammation induction (*ply*, *lytA*, and *htrA*) and opsonophagocytosis inhibition (CPS genes, *PiaA*, *PiuA*, *nanA*, and *srtH*).<sup>66,69,76,77</sup>

Among the detected virulence genes, we particularly focused on the pneumolysin, one of the most promising protein vaccine antigen.<sup>78</sup> The discovery of clones harboring mutated pneumolysin has questioned the efficiency of pneumolysin-based vaccines, expected to be highly genetically conserved.<sup>78,79</sup> With this in mind, a deepening of the virulence analysis was performed by investigating the pneumolysin allelic profile in NC isolates. Substitutions (K224R, A265S, T172I, and Y150H) and deletions (V270, K271) were observed in both ST306 and ST3717 pneumolysin sequences (Figure 2).

These mutations (K224R, A265S, V270, and K271) are known to play a crucial role in diminishing the hemolytic activity.<sup>34,48,79</sup> Added to T172I and Y150H mutations, the hemolytic activity became depleted.<sup>79</sup> The hemolytic function depletion, even primordial, confirms its uncorrelation to invasiveness loss and lower epidemic-prone capacity.<sup>34</sup>

#### Contribution of mobile genetic elements in *S. pneumoniae* serotype 1 diversification

Mobile genetic elements (MGEs) can deeply influence bacterial pathogenic potential.<sup>80,81</sup> When querying the ICEberg or using the ICEfinder web tools, we did not detect any ICE.<sup>56,82,83</sup>

CLUSTAL O(1.2.4) multiple sequence alignment

D39	MANKAVNDFILAMNYDKKKLLTHQGESIENRFIKEGNQLPDEFVVIERKKRSLSTNTSDI	60
Allele5	MANKAVNDFILAMNYDKKKLLTHQGESIENRFIKEGNQLPDEFVVIERKKRSLSTNTSDI	60
ply2	MANKAVNDFILAMNYDKKKLLTHQGESIENRFIKEGNQLPDEFVVIERKKRSLSTNTSDI	60
*****		
D39	SVTATNDSRLYPGALLVVDETLLENNPTLLAVDRAPMTYSIDLPGASSDSFQLQVEDPSN	120
Allele5	SVTATNDSRLYPGALLVVDETLLENNPTLLAVDRAPMTYSIDLPGASSDSFQLQVEDPSN	120
ply2	SVTATNDSRLYPGALLVVDETLLENNPTLLAVDRAPMTYSIDLPGASSDSFQLQVEDPSN	120
*****		
D39	SSVRGAVNDLLAKWHQDYGVVNNVPRM <sup>Q</sup> Y <sup>E</sup> KI <sup>T</sup> AHSM <sup>E</sup> QLK <sup>V</sup> KFGSD <sup>F</sup> EK <sup>I</sup> GNSLDIDF	180
Allele5	SSVRGAVNDLLAKWHQDYGVVNNVPRM <sup>Q</sup> HEK <sup>I</sup> AHSM <sup>E</sup> QLK <sup>V</sup> KFGSD <sup>F</sup> EK <sup>I</sup> GNSLDIDF	180
ply2	SSVRGAVNDLLAKWHQDYGVVNNVPRM <sup>Q</sup> HEK <sup>I</sup> AHSM <sup>E</sup> QLK <sup>V</sup> KFGSD <sup>F</sup> EK <sup>I</sup> GNSLDIDF	180
*****		
D39	NSVHSGEKQIQIVNFKQIYYTVSVDVAVK <sup>N</sup> PGD <sup>V</sup> FQD <sup>T</sup> V <sup>T</sup> VEDL <sup>K</sup> QRGISAERPLVYISSV	240
Allele5	NSVHSGEKQIQIVNFKQIYYTVSVDVAVK <sup>N</sup> PGD <sup>V</sup> FQD <sup>T</sup> V <sup>T</sup> VEDL <sup>R</sup> QRGISAERPLVYISSV	240
ply2	NSVHSGEKQIQIVNFKQIYYTVSVDVAVK <sup>N</sup> PGD <sup>V</sup> FQD <sup>T</sup> V <sup>T</sup> VEDL <sup>R</sup> QRGISAERPLVYISSV	240
*****		
D39	AYGRQVYLKLETTSKSDEVEAAFE <sup>A</sup> L <sup>I</sup> K <sup>G</sup> V <sup>K</sup> APQTEWKQILDNTEVKAVILGGDPSSGA	300
Allele5	AYGRQVYLKLETTSKSDEVEAAFE <sup>S</sup> L <sup>I</sup> K <sup>G</sup> V--APQTEWKQILDNTEVKAVILGGDPSSGA	298
ply2	AYGRQVYLKLETTSKSDEVEAAFE <sup>S</sup> L <sup>I</sup> K <sup>G</sup> V--APQTEWKQILDNTEVKAVILGGDPSSGA	298
*****		
D39	RVVTGKVD <sup>M</sup> VEDL <sup>I</sup> QEGSR <sup>F</sup> TADHPGL <sup>P</sup> ISY <sup>T</sup> TSFL <sup>R</sup> DNV <sup>V</sup> ATFQ <sup>N</sup> ST <sup>D</sup> YVET <sup>K</sup> VTAY <sup>R</sup> N	360
Allele5	RVVTGKVD <sup>M</sup> VEDL <sup>I</sup> QEGSR <sup>F</sup> TADHPGL <sup>P</sup> ISY <sup>T</sup> TSFL <sup>R</sup> DNV <sup>V</sup> ATFQ <sup>N</sup> ST <sup>D</sup> YVET <sup>K</sup> VTAY <sup>R</sup> N	358
ply2	RVVTGKVD <sup>M</sup> VEDL <sup>I</sup> QEGSR <sup>F</sup> TADHPGL <sup>P</sup> ISY <sup>T</sup> TSFL <sup>R</sup> DNV <sup>V</sup> ATFQ <sup>N</sup> ST <sup>D</sup> YVET <sup>K</sup> VTAY <sup>R</sup> N	358
*****		
D39	GDLLLDHSGAYVAQYYITWDELSYDHQ <sup>G</sup> KEV <sup>L</sup> TPKAWDRNGQDLTAHFTT <sup>S</sup> IPLKGN <sup>V</sup> RN	420
Allele5	GDLLLDHSGAYVAQYYITWDELSYDHQ <sup>G</sup> KEV <sup>L</sup> TPKAWDRNGQDLTAHFTT <sup>S</sup> IPLKGN <sup>V</sup> RN	418
ply2	GDLLLDHSGAYVAQYYITWDELSYDHQ <sup>G</sup> KEV <sup>L</sup> TPKAWDRNGQDLTAHFTT <sup>S</sup> IPLKGN <sup>V</sup> RN	418
*****		
D39	LSVKIRECTGLAW <sup>E</sup> W <sup>R</sup> T <sup>V</sup> Y <sup>E</sup> K <sup>T</sup> DL <sup>P</sup> L <sup>V</sup> R <sup>K</sup> R <sup>T</sup> IS <sup>I</sup> W <sup>G</sup> T <sup>T</sup> L <sup>P</sup> Q <sup>V</sup> ED <sup>K</sup> V <sup>E</sup> N <sup>D</sup> X	472
Allele5	LSVKIRECTGLAW <sup>E</sup> W <sup>R</sup> T <sup>V</sup> Y <sup>E</sup> K <sup>T</sup> DL <sup>P</sup> L <sup>V</sup> R <sup>K</sup> R <sup>T</sup> IS <sup>I</sup> W <sup>G</sup> T <sup>T</sup> L <sup>P</sup> Q <sup>V</sup> ED <sup>K</sup> V <sup>E</sup> N <sup>D</sup> X	470
ply2	LSVKIRECTGLAW <sup>E</sup> W <sup>R</sup> T <sup>V</sup> Y <sup>E</sup> K <sup>T</sup> DL <sup>P</sup> L <sup>V</sup> R <sup>K</sup> R <sup>T</sup> IS <sup>I</sup> W <sup>G</sup> T <sup>T</sup> L <sup>P</sup> Q <sup>V</sup> ED <sup>K</sup> V <sup>E</sup> N <sup>D</sup> X	470
*****		

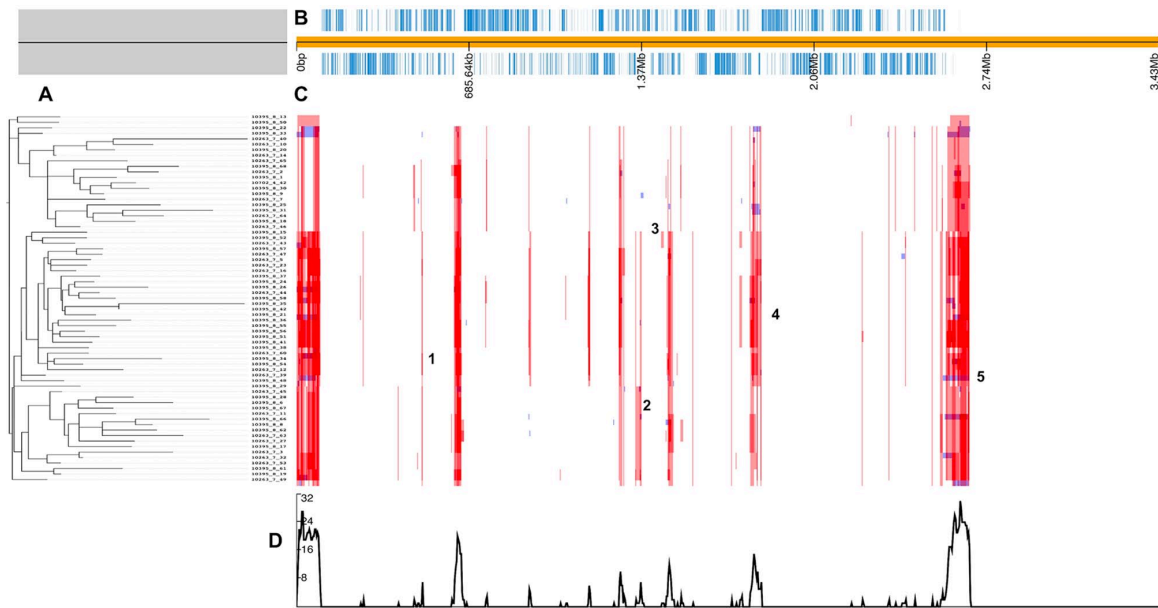
**Figure 2.** Amino acid multiple sequence alignment of pneumolysin sequences. D39, Allele 5, and ply2 referred to pneumolysin sequences of respectively the reference strain (D39), NC serotype 1, and 22 serotype 1 retrieved from Jefferies et al's<sup>34</sup> study. The mutation positions compared to the reference pneumolysin D39 are highlighted in purple.

The MGE investigation was then broadened to prophage sequences. Those elements are also linked to the transmission of virulence traits (ie, *pblA* and *pblB*) and resistance gene (ie, *tetM*) in *S. pneumoniae*.<sup>84,85</sup> Consistent with previous findings, we barely retrieve prophages in serotype 1 *S. pneumoniae*.<sup>86</sup> Indeed, two “questionable” and 37 putative prophage regions were detected using PHASTER and VirSorter, respectively.<sup>37,38</sup> Those detected by PHASTER correspond to the Enterobacteria phage phi92, having a region size of ~11 KB. The 37 viral putative regions detected by VirSorter had a mean region size of 13 KB.<sup>38</sup> Considering that the average of a complete prophage sequence is about 45 Kb,<sup>87</sup> we hypothesized whether the “questionable” and “putative” regions detected by both tools might be potential “remnants” prophages.<sup>84</sup>

#### Homologous recombination rate and hotspots in NC *S. pneumoniae* serotype 1

Homologous recombination is a key factor for the acquisition of ARGs, virulence traits, and new metabolic properties by pathogens.<sup>80,88</sup> Previous studies depicted the crucial role of recombination in the transmission of fluoroquinolones, beta-lactams, trimethoprim, and sulfamethoxazole resistance in *S.*

*pneumoniae*.<sup>56,83,89</sup> Herein, NC *S. pneumoniae* full-genome screening showed that no recombination events were detected in ARGs. In contrast, screening of virulence genes highlighted a recent recombination event occurring in the lamin-binding protein (*lmb*) gene, and more ancestral recombination events in *piaA* (pneumococcal iron acquisition A), *shrA* (Streptococcal Lipoprotein Rotamase A), *pce* (Choline-binding protein E) and *cps2L* (Figure 3). These genes are likely to be beneficial to *S. pneumoniae*, either during colonization or during the invasion process. The *pce* gene has a dual role in the virulence of *S. pneumoniae* by being associated with the process of neutrophil decreasing activity and plasminogen binding.<sup>71</sup> The *ShrA* is yet another virulence gene, promoting upper airways colonization during the initial step of pneumococcal infection.<sup>71,90</sup> The *piaA* and *lmb* genes provide, respectively, the Fe and Zn nutrients necessary for the adaptation of *S. pneumoniae* to different microenvironments while progressing to an invasive status.<sup>71,91</sup> The most notable virulence gene detected in the recombination hotspot is the *cps2L*, part of the *cps* gene cluster coding for *S. pneumoniae* capsular polysaccharide. The pneumococcal capsular polysaccharide is targeted by the PCV vaccine and defines the *S. pneumoniae* serotype, its carriage duration, recombination rate, and therefore, its virulence and invasiveness.<sup>92,93</sup> However, the *cps* locus organization



**Figure 3.** Recombination hotspots in whole-genome alignment: (A) the phylogenetic tree was computed on recombination sites free alignment and was generated by GUBBINS. (B) Functional annotation of the consensus sequence where the light blue spots represent the mapping of the predicted CDS. (C) The detected recombination hotspots: red blocks represent ancestral recombination while the blue blocks represent specific recombination to one isolate. Recombination hotspots overlapping with virulence genes are labeled with a number: 1, 2, 3, 4, and 5, respectively, corresponding to *cps2L*, *slrA*, *piaA*, *cpbE/pce*, and *lmb*, respectively. (D) The bottom part of the figure corresponds to a graph measuring SNP density.

facilitates the recombination events favoring the emergence of new capsules variants clones, more commonly known as NVT through capsule switching mechanism.<sup>82,93</sup> This phenomenon was highly promoted by the introduction of PCV7. For instance, the replacement of the serotype 23F and serotype 4 clones by 19A variants following PCV7 introduction.<sup>12,94</sup> Similar events occurred in NC, where an increase of serotype 15B and 19A prevalence were described after-PCV7 introduction in 2008.<sup>13</sup>

In the second step of this analysis, we estimated the recombination rate of NC *S. pneumoniae* isolates. The recombination to mutation ( $r/m$ ) ratio was equal to 2.51 (with 2.45 and 3.24, respectively for ST306 and ST 3717) (Supplementary Table S6). This low recombination rate supports the previous findings.<sup>50</sup> Indeed, the serotype 1 recombination rate was estimated to be 1.51 in Southeast Africa, 0.05 in South Africa, and 4.22 in West Africa and Southeast Asia isolates.<sup>50</sup> In contrast, recombinant multidrug-resistant *S. pneumoniae* clones such as the PMEN1, PMEN2, and PMEN14, exhibited a higher  $r/m$  value of, respectively, 7.2, 14.9, and 34.06.<sup>11,41,95</sup>

#### Dating lineage by Bayesian evolutionary analysis

Prior to BEAST analysis, we tested the clock-likeness of the data set by Tempest software.<sup>44</sup> Using a best-fit root, we obtained a chi-square value of  $6.4508e^{-2}$  and a positive coefficient correlation value ( $R^2$ : 0.254) (Supplementary Figure S2). The tree height was estimated to 1976.7363 using FigTree meaning that serotype 1 emerged prior to the PCV7 introduction in 2008 (Supplementary Figure S3).

The most common ancestor (tMRCA) of ST306 was estimated to 2000.731 (95% highest posterior density [HPD]):

1988.2523–2004.1678). This date corresponds to the first outbreak period caused by ST306 in NC.<sup>8</sup> In addition, mutation rate median value was estimated to 0.5733 (substitutions/site/year [95% HPD:  $8.093e^{-3}$ , 4.8331]), meaning that one mutation occurs every 2 years. One possible consequence of the emergence of ST3717 was caused by the Guanine-Adenine mutation in the ST306 *aroE* housekeeping gene. This finding is corroborated by the tMRCA (2002.5071) estimated for the ST3717.

#### Conclusion

Despite its high prevalence and association with the 2000s outbreaks, the NC serotype 1 genomic specificity was investigated using only conventional low throughput methods. In this study, we provided further insights by describing the population structure and genomic features of NC *S. pneumoniae* serotype 1 pre-PCV13 using the whole-genome sequencing. The results shown here confirm that the serotype 1 NC population was extensively dominated by ST306. The genomes of the whole population harbor multiple genes conferring resistance to antibiotics and contain a large number of virulence genes. NC genomes are also characterized by a relatively low rate of recombination and lack of MGE elements, involved in genomes diversification.

The absence of genomic data from other serotypes and phenotypic (*in vivo/in vitro*) experimental data from our samples, limited the spectrum of the analysis to the study of genomic, virulence, and antibiotic resistance features of NC *S. pneumoniae* serotype 1. Thus, investigations on *S. pneumoniae* serotype 1 post-PCV13, and more broadly the overall pneumococcal populations in NC need to be performed to highlight the PCV13 impact.

## Acknowledgements

The authors thank Mrs. Samar Hannachi for the preliminary analysis undertaken in this project. They thank Nidhal Ghanmi and Aymen Ben Chaalia for technical assistance during the project.

## Author Contributions

M.H: was responsible for the bioinformatics analyses and drafted the manuscript and figures. A.K: aided in the bioinformatics analysis and assisted in writing and reviewing the manuscript. J.C: aided in interpreting the results and provided manuscript revision. E.H.S: assisted in drafting and manuscript reviewing. D.E: aided in manuscript revision. A.B.K: participated in results interpretation, worked on writing and reviewing the manuscript. O.S: supervised the work, participated in results interpretation, drafted, and reviewed the manuscript.

## ORCID iD

Oussama Souiai  <https://orcid.org/0000-0003-2443-114X>

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

- Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. <https://www.ncbi.nlm.nih.gov/pubmed/19748398>. Accessed March 21, 2020.
- Henriques-Normark B, Tuomanen EI. The pneumococcus: epidemiology, microbiology, and pathogenesis. *Cold Spring Harb Perspect Med*. 2013;3:a010215. doi:10.1101/cshperspect.a010215.
- Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. *Lancet Infect Dis*. 2005;5:83-93.
- Ritchie ND, Mitchell TJ, Evans TJ. What is different about serotype 1 pneumococci? *Future Microbiol*. 2012;7:33-46. doi:10.2217/fmb.11.146.
- Brueggemann AB, Spratt BG. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. *J Clin Microbiol*. 2003;41:4966-4970. doi:10.1128/JCM.41.11.4966-4970.2003.
- Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis*. 2000;30:100-121. doi:10.1086/313608.
- Michel N, Watson M, Baumann F, Perolat P, Garin B. Distribution of *Streptococcus pneumoniae* serotypes responsible for penicillin resistance and the potential role of new conjugate vaccines in New Caledonia. *J Clin Microbiol*. 2005;43:6060-6063. doi:10.1128/JCM.43.12.6060-6063.2005.
- Le Hello S, Watson M, Levy M, et al. Invasive serotype 1 *Streptococcus pneumoniae* outbreaks in the South Pacific from 2000 to 2007. *J Clin Microbiol*. 2010;48:2968-2971. doi:10.1128/JCM.01615-09.
- Daniels CC, Rogers PD, Shelton CM. A review of pneumococcal vaccines: current polysaccharide vaccine recommendations and future protein antigens. *J Pediatr Pharmacol Ther*. 2016;21:27-35. doi:10.5863/1551-6776-21.1.27.
- Usuf E, Bottomley C, Bojang E, et al. Persistence of nasopharyngeal pneumococcal vaccine serotypes and increase of nonvaccine serotypes among vaccinated infants and their mothers 5 years after introduction of pneumococcal conjugate vaccine 13 in The Gambia. *Clin Infect Dis*. 2019;68:1512-1521. doi:10.1093/cid/ciy726.
- Croucher NJ, Finkelstein JA, Pelton SI, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet*. 2013;45:656-663. doi:10.1038/ng.2625.
- Croucher NJ, Harris SR, Fraser C, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2013;331:430-434. doi:10.1126/science.1198545.
- Scotet J, Baumann F, Dupont-Rouzeyrol M, et al. Impact of Prevnar vaccination on nasopharyngeal carriage of *Streptococcus pneumoniae* in healthy children in New Caledonia. *BMC Proc*. 2011;5:P13. doi:10.1186/1753-6561-5-s1-p13.
- Pai R, Moore MR, Pilishvili T, et al. Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. *J Infect Dis*. 2005;192:1988-1995. doi:10.1086/498043.
- Ludwig G, Garcia-Garcia S, Lanasa M, et al. Serotype and clonal distribution dynamics of invasive pneumococcal strains after PCV13 introduction (2011-2016): surveillance data from 23 sites in Catalonia, Spain. *PLoS ONE*. 2020;15:e0228612. doi:10.1371/journal.pone.0228612.
- Kambiré D, Soeters HM, Ouédraogo-Traoré R, et al. Early impact of 13-valent pneumococcal conjugate vaccine on pneumococcal meningitis—Burkina Faso, 2014–2015. *J Infect*. 2018;76:270-279. doi:10.1016/j.jinf.2017.12.002.
- du Plessis M, Allam M, Tempia S, et al. Phylogenetic analysis of invasive serotype 1 *Pneumococcus* in South Africa, 1989 to 2013. *J Clin Microbiol*. 2016;54:1326-1334. doi:10.1128/JCM.00055-16.
- Schürch AC, Arredondo-Alonso S, Willems RJJ, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect*. 2018;24:350-354. doi:10.1016/j.cmi.2017.12.016.
- Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. *Pathology (Phila)*. 2015;47:199-210. doi:10.1097/PAT.0000000000000235.
- Everett DB, Mukaka M, Denis B, et al. Ten years of surveillance for invasive *Streptococcus pneumoniae* during the era of antiretroviral scale-up and cotrimoxazole prophylaxis in Malawi. *PLoS ONE*. 2011;6:e17765. doi:10.1371/journal.pone.0017765.
- Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST): multi-locus sequence types from short reads. *BMC Genomics*. 2012;13:338. doi:10.1186/1471-2164-13-338.
- Page AJ, De Silva N, Hunt M, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom*. 2016;2:e000083. doi:10.1099/mgen.0.000083.
- Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18:821-829. doi:10.1101/gr.074492.107.
- Using the Velvet de novo assembler for short-read sequencing technologies. <https://www.ncbi.nlm.nih.gov/pubmed/?term=20836074>. Accessed February 3, 2020.
- Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36:W5-W9. doi:10.1093/nar/gkn201.
- Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med*. 2011;6:11. doi:10.1186/1751-0473-6-11.
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14:e1005944. doi:10.1371/journal.pcbi.1005944.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068-2069. doi:10.1093/bioinformatics/btu153.
- Roary: rapid large-scale prokaryote pan genome analysis. <https://pubmed.ncbi.nlm.nih.gov/26198102/>. Accessed July 4, 2020.
- Sitto F, Battistuzzi FU. Estimating pangenomes with roary. *Mol Biol Evol*. 2020;37:933-939. doi:10.1093/molbev/msz284.
- Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. doi:10.1186/1471-2105-10-421.
- Chen L, Yang J, Yu J, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. 2005;33:325-328. doi:10.1093/nar/gki008.
- McArthur AG, Waglechner N, Nizam F, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother*. 2013;57:3348-3357. doi:10.1128/AAC.00419-13.
- Jefferies JM, Johnston CH, Kirkham LA, et al. Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. *J Infect Dis*. 2007;196:936-944. doi:10.1086/520091.
- Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics (Oxford, England)*. 2014;30:3276-3278. doi:10.1093/bioinformatics/btu531.
- Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol*. 2014;1079:105-116. doi:10.1007/978-1-62703-646-7\_6.
- Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44:W16-W21. doi:10.1093/nar/gkw387.
- Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ*. 2015;3:e985. doi:10.7717/peerj.985.
- Bi D, Xu Z, Harrison EM, et al. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res*. 2012;40:D621-D626. doi:10.1093/nar/gkr846.
- Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. 2010;5:e11147. doi:10.1371/journal.pone.0011147.
- Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res*. 2015;43:e15. doi:10.1093/nar/gku1196.
- Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *BioRxiv*. 2017:1-6. doi:10.1101/119545.

43. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2014;10:e1003537. doi:10.1371/journal.pcbi.1003537.
44. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2:vew007. doi:10.1093/ve/vew007.
45. SMS: smart model selection in PhyML. <https://www.ncbi.nlm.nih.gov/pubmed/?term=28472384>. Accessed February 3, 2020.
46. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018;67:901-904. doi:10.1093/sysbio/syy032.
47. Rambaut A. FigTree v1.4.2. <http://tree.bio.ed.ac.uk/software/figtree/>. Published 2014. Accessed July 19, 2020.
48. Kirkham L-AS, Jefferies JMC, Kerr AR, et al. Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *J Clin Microbiol*. 2006;44:151-159. doi:10.1128/JCM.44.1.151-159.2006.
49. Livingstone PG, Morphey RM, Whitworth DE. Genome sequencing and pan-genome analysis of 23 *Coralloccoccus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front Microbiol*. 2018;9:3187. doi:10.3389/fmicb.2018.03187.
50. Chaguza C, Cornick JE, Harris SR, et al. Understanding pneumococcal serotype 1 biology through population genomic analysis. *BMC Infect Dis*. 2016;16:649-649. doi:10.1186/s12879-016-1987-z.
51. Rouli L, Merhej V, Fournier P-E, Raoult D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect*. 2015;7:72-85. doi:10.1016/j.nmni.2015.06.005.
52. Guimarães LC, Florczak-Wypianska J, de Jesus LB, et al. Inside the Pan—genome—methods and software overview. *Curr Genomics*. 2015;16:245-252. doi:10.2174/1389202916666150423002311.
53. Donati C, Hiller NL, Tettelin H, et al. Structure and dynamics of the pangenome of *Streptococcus pneumoniae* and closely related species. *Genome Biol*. 2010;11:R107. doi:10.1186/gb-2010-11-10-r107.
54. File TM. Appropriate use of antimicrobials for drug—resistant pneumonia: focus on the significance of  $\beta$ -lactam—resistant *Streptococcus pneumoniae*. *Clin Infect Dis*. 2002;34:S17-S26. doi:10.1086/324526.
55. McGee L, McDougal L, Zhou J, et al. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol*. 2001;39:2565-2571. doi:10.1128/JCM.39.7.2565-2571.2001.
56. Cornick JE, Bentley SD. *Streptococcus pneumoniae*: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes Infect*. 2012;14:573-583. doi:10.1016/j.micinf.2012.01.012.
57. Reinert RR. The antimicrobial resistance profile of *Streptococcus pneumoniae*. *Clin Microbiol Infect*. 2009;15:7-11. doi:10.1111/j.1469-0691.2009.02724.x.
58. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44:S27-S72. doi:10.1086/511159.
59. Yildirim I, Shea KM, Pelton SI. Pneumococcal disease in the era of pneumococcal conjugate vaccine. *Infect Dis Clin North Am*. 2015;29:679-697. doi:10.1016/j.idc.2015.07.009.
60. Chaguza C, Cornick JE, Andam CP, et al. Population genetic structure, antibiotic resistance, capsule switching and evolution of invasive pneumococci before conjugate vaccination in Malawi. *Vaccine*. 2017;35:4594-4602. doi:10.1016/j.vaccine.2017.07.009.
61. Garvey MI, Baylay AJ, Wong RL, Piddock LJV. Overexpression of patA and patB, which encode ABC transporters, is associated with fluoroquinolone resistance in clinical isolates of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother*. 2011;55:190-196. doi:10.1128/AAC.00672-10.
62. Liu M, Douthwaite S. Resistance to the macrolide antibiotic tylosin is conferred by single methylations at 23S rRNA nucleotides G748 and A2058 acting in synergy. *Proc Natl Acad Sci USA*. 2002;99:14658-14663. doi:10.1073/pnas.232580599.
63. Kosowska K, Jacobs MR, Bajaksouzian S, Koeth L, Appelbaum PC. Alterations of penicillin-binding proteins 1A, 2X, and 2B in *Streptococcus pneumoniae* isolates for which amoxicillin MICs are higher than penicillin MICs. *Antimicrob Agents Chemother*. 2004;48:4020-4022. doi:10.1128/AAC.48.10.4020-4022.2004.
64. Blumental S, Granger-Farbos A, Moisi JC, et al. Virulence factors of *Streptococcus pneumoniae*. Comparison between African and French invasive isolates and implication for future vaccines. *PLoS ONE*. 2015;10:e0133885. doi:10.1371/journal.pone.0133885.
65. *Streptococcus pneumoniae* capsular polysaccharide. <https://pubmed.ncbi.nlm.nih.gov/30977464/>. Accessed July 20, 2020.
66. Yother J. Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annu Rev Microbiol*. 2011;65:563-581. doi:10.1146/annurev.micro.62.081307.162944.
67. Maestro B, Sanz JM. Choline binding proteins from *Streptococcus pneumoniae*: a dual role as enzybiotics and targets for the design of new antimicrobials. *Antibiotics*. 2016;5:21. doi:10.3390/antibiotics5020021.
68. Bagnoli F, Moschioni M, Donati C, et al. A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. *J Bacteriol*. 2008;190:5480-5492. doi:10.1128/JB.00384-08.
69. Janesch P, Rouha H, Badarau A, et al. Assessing the function of pneumococcal neuraminidases NanA, NanB and NanC in vitro and in vivo lung infection models using monoclonal antibodies. *Virulence*. 2018;9:1521-1538. doi:10.1080/21505594.2018.1520545.
70. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. *Microbiology (Reading)*. 2006;152:295-303. doi:10.1099/mic.0.28610-0.
71. Weiser JN, Ferreira DM, Paton JC. *Streptococcus pneumoniae*: transmission, colonization and invasion. *Nat Rev Microbiol*. 2018;16:355-367. doi:10.1038/s41579-018-0001-8.
72. Ibrahim YM, Kerr AR, McCluskey J, Mitchell TJ. Role of HtrA in the virulence and competence of *Streptococcus pneumoniae*. *Infect Immun*. 2004;72:3584-3591. doi:10.1128/IAI.72.6.3584-3591.2004.
73. Trigger factor in *Streptococcus mutans* is involved in stress tolerance, competence development, and biofilm formation. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC538946/>. Accessed July 6, 2020.
74. Moulin P, Patron K, Cano C, et al. The Adc/Lmb system mediates zinc acquisition in *Streptococcus agalactiae* and contributes to bacterial growth and survival. *J Bacteriol*. 2016;198:3265-3277. doi:10.1128/JB.00614-16.
75. Andre GO, Converso TR, Politano WR, et al. Role of *Streptococcus pneumoniae* proteins in evasion of complement-mediated immunity. *Front Microbiol*. 2017;8:224. doi:10.3389/fmicb.2017.00224.
76. Jomaa M, Yuste J, Paton JC, Jones C, Dougan G, Brown JS. Antibodies to the iron uptake ABC transporter lipoproteins PiaA and PiuA promote opsonophagocytosis of *Streptococcus pneumoniae*. *Infect Immun*. 2005;73:6852-6859. doi:10.1128/IAI.73.10.6852-6859.2005.
77. Dalia AB, Standish AJ, Weiser JN. Three surface exoglycosidases from *Streptococcus pneumoniae*, NanA, BgaA, and StrH, promote resistance to opsonophagocytic killing by human neutrophils. *Infect Immun*. 2010;78:2108-2116. doi:10.1128/IAI.01125-09.
78. Feldman C, Anderson R. Review: current and new generation pneumococcal vaccines. *J Infect*. 2014;69:309-325. doi:10.1016/j.jinf.2014.06.006.
79. Harvey RM, Ogunniyi AD, Chen AY, Paton JC. Pneumolysin with low hemolytic activity confers an early growth advantage to *Streptococcus pneumoniae* in the blood. *Infect Immun*. 2011;79:4122-4130. doi:10.1128/IAI.05418-11.
80. Gyles C, Boerlin P. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. *Vet Pathol*. 2014;51:328-340. doi:10.1177/0300985813511131.
81. Alvarenga DO, Moreira LM, Chandler M, Varani AM. Chapter 7. A practical guide for comparative genomics of mobile genetic elements in prokaryotic genomes. In: Setubal J, Stoye J, Stadler P, eds. *Comparative genomics. Methods in molecular biology*. Vol 1704. New York, NY: Humana Press; 2018:213-242.
82. Andam CP, Hanage WP. Mechanisms of genome evolution of *Streptococcus*. *Infect Genet Evol*. 2015;33:334-342. doi:10.1016/j.meegid.2014.11.007.
83. Chaguza C, Cornick JE, Everett DB. Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*. *Comput Struct Biotechnol J*. 2015;13:241-247. doi:10.1016/j.csbj.2015.03.007.
84. Brueggemann AB, Harrold CL, Rezaei Javan R, van Tonder AJ, McDonnell AJ, Edwards BA. Pneumococcal prophages are diverse, but not without structure or history. *Sci Rep*. 2017;7:42976. doi:10.1038/srep42976.
85. Wyres KL, van Tonder A, Lambertsen LM, et al. Evidence of antimicrobial resistance-conferring genetic elements among pneumococci isolated prior to 1974. *BMC Genomics*. 2013;14:500. doi:10.1186/1471-2164-14-500.
86. Ramirez M, Severina E, Tomasz A. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. *J Bacteriol*. 1999;181:3618-3625.
87. Ramisetty BCM, Sudhakari PA. Bacterial “grounded” prophages: hotspots for genetic renovation and innovation. *Front Genet*. 2019;10:65. doi:10.3389/fgene.2019.00065.
88. Lu B, Leong HW. Computational methods for predicting genomic islands in microbial genomes. *Comput Struct Biotechnol J*. 2016;14:200-206. doi:10.1016/j.csbj.2016.05.001.
89. Heyderman RS, Bentley SD, Pluschke G, et al. Region-specific diversification of the highly virulent serotype 1 *Streptococcus pneumoniae*. *Microb Genom*. 2015;1:e000027. doi:10.1099/mgen.0.000027.
90. Hermans PWM, Adrian PV, Albert C, et al. The streptococcal lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization. *J Biol Chem*. 2006;281:968-976. doi:10.1074/jbc.M510014200.
91. Shafeeq S, Kuipers OP, Kloosterman TG. The role of zinc in the interplay between pathogenic streptococci and their hosts. *Mol Microbiol*. 2013;88:1047-1057. doi:10.1111/mmi.12256.



92. Chaguza C, Andam CP, Harris SR, et al. Recombination in *Streptococcus pneumoniae* lineages increase with carriage duration and size of the polysaccharide capsule. *mBio*. 2016;7:e01053-16. doi:10.1128/mBio.01053-16.
93. Geno KA, Gilbert GL, Song JY, et al. Pneumococcal capsules and their types: past, present, and future. *Clin Microbiol Rev*. 2015;28:871-899. doi:10.1128/CMR.00024-15.
94. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog*. 2007;3:e168. doi:10.1371/journal.ppat.0030168.
95. Croucher NJ, Hanage WP, Harris SR, et al. Variable recombination dynamics during the emergence, transmission and “disarming” of a multidrug-resistant pneumococcal clone. *BMC Biol*. 2014;12:49. doi:10.1186/1741-7007-12-49.