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

Mariem Hanachi1,2, Anmol Kiran3,4, Jennifer Cornick4,5, Emna Harigua-Souiai6, Dean Everett3,4, Alia Benkahla1 and Oussama Souiai1,7

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

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

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.1

*S. pneumoniae* commonly colonizes the nasopharynx tract.2 Its migration to other body sites leads to a wide range of noninvasive diseases such as otitis, sinusitis, and numerous invasive diseases including pneumonia, meningitis, and septicemia.2 Among over the 90 serotypes defined by the structure and antigenicity of the capsular polysaccharide (φp), serotype 1 remains the most common cause of invasive pneumococcal disease (IPD) worldwide and was frequently linked to outbreaks.3,4 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.5,6 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 old7 and children under 8 years old.8

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 1.9 Even though vaccination campaigns demonstrated a certain efficiency in lowering the IPD, nonvaccine types (NVTs) emerged after PCV7 and PCV13 introduction.10-14 Regarding the serotype 1, it was reported that some of its clones are still circulating after PCV13 introduction.15-17 For instance, ST8314 and the newly reported ST9067 respectively expressed less sensitivity to antibiotics and a higher recombination rate.18 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.7,8 To survey, monitor pathogen outbreaks, and explore genomic features, these conventional comparative methods need to be completed by the next-generation sequencing analyses.18,19

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.20 High throughput sequences were generated using Illumina Genome Analyzer-II at Sanger Institute at Wellcome-Trust Campus, United Kingdom. The obtained 100 bp 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.21 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.22 This pipeline uses Velvet, version 1.2.09,23 and velvetOptimiser, version 2.2.524 (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 29479 bp (SD: 3951) (Supplementary Table S2). The average sequencing coverage was equal to 338 bp (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.25 Finally, CONTIGuator (default parameters) was used to generate circular genomes with a mean size of 2082132 bp (sd: 7223).26 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.27 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).28 The annotation GFF files served as input to the Roary, version 3.11.2 pipeline (default parameters), allowing the pan-genome size estimation.29,30

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).31-33 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.34 The multiple alignments of nucleotide and protein sequences were generated using the AliView software, version 1.26.35 Sequence variations detection was performed using the Clustal Omega web server (https://www.ebi.ac.uk/Tools/msa/clustalo/).36 Prophages in the analyzed genomes were first screened using PHASTER37 and VirSorter38 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).39

Recombination rate estimation and hotspots identification

We used progressiveMauve, version 2.4.0, to generate a whole-genome alignment of the circularized genomes.40 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). Detected recombination events were visualized using the Phandango web tool (https://jameshadfield.github.io/phandango/). 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 1.5.1.43,44 Tip dates were specified as years of sample collection. Through the use of the smart model selection web server,45 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.12 We used the strict clock model in combination with the coalescent Bayesian skyline demographic model. BEAST, version 2.4.8, was then run with 800000000 Markov Chain Monte Carlo (MCMC) generations with a 10% burn-in. The log files generated by BEAST were summarized by Tracer, version 1.7.1.46 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 2.4.841 (available from http://beast.bio.ed.ac.uk) and FigTree, version 1.4.447 (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.7,8 The ST306 is a worldwide distributed serotype, suspected to be an important determinant behind the increase of the serotype 1 IPD expansion.5,48 The ST3717 was not ubiquitously reported as an important determinant behind the increase of the serotype 1 IPD expansion.5,48 The ST3717 was not ubiquitously reported as an important determinant behind the increase of the serotype 1 IPD expansion.5,48

**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 1.50 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).50 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).51,52 The pan-genome of S. pneumoniae is usually considered as open due to the ability to capture exogenous DNA through horizontal gene transfer.51 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.

**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.54-57 Depending on disease severity, clinical manifestations, and local community antibiotic resistance patterns, a wide range of alternative treatments can be administered.58,59 These treatments contributed to the emergence of resistant clones, and notably to fluoroquinolones and macrolide antibiotics.56 The serotype 1 was described as rarely resistant to macrolides, penicillin, and quinolones.4 Conversely, serotype 1 exhibits the highest rate of multidrug resistance (MDR) compared to the other serotypes mainly by expressing resistance to ceftriaxone, tetracycline, and chloramphenicol.60

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*.61 The over-expression of these efflux pumps genes were reported as conferring a low-level fluoroquinolone resistance.52 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.62 While no phenotypic data for lincosamides, fluoroquinolone, and streptogramin B were...
Bioinformatics and Biology Insights

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.7,8,63

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.64 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, wzh, wzg, wze, and wzd).65 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.66 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 (cppD, cppG, pce, lytA, lytC, and lytB),67 the pilus-encoding pathogenicity islets (pitA, sipA, pitB, and srtGI),68 two neuraminidases (NanA and NanB),69 or other surface protein like eno, parA, srtA pil, srtH, sleA, and pfbA.2,70 In addition, genes that ensure the host tissues degradation (hyxA, eno, pce, and gapA)71 or provide to S. pneumoniae essential nutrients (piaA, piaM, PiaA, btrA, lmb, and tig) were found.71-74 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),75 inflammation induction (ply, lytA, and btrA) and opsonophagocytosis inhibition (CPS genes, PiaA, PiaM, nanA, and srtH).66,69,76,77

Among the detected virulence genes, we particularly focused on the pneumolysin, one of the most promising protein vaccine antigen.78 The discovery of clones harboring mutated pneumolysin has questioned the efficiency of pneumolysin-based vaccines, expected to be highly genetically conserved.78,79 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.34,48,79 Added to T172I and Y150H mutations, the hemolytic activity became depleted.79 The hemolytic function depletion, even primordial, confirms its uncorrelation to invasiveness loss and lower epidemic-prone capacity.34

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

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

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.29,30,49 (B) Cumulative number of genes in the pan-genome. Box plots indicate first and third quartiles with medians shown as horizontal lines.
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.84,85 Consistent with previous findings, we barely retrieve prophages in serotype 1 S. pneumoniae.86 Indeed, two “questionable” and 37 putative prophage regions were detected using PHASTER and VirSorter, respectively.37,38 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.38 Considering that the average of a complete prophage sequence is about 45 Kb,87 we hypothesized whether the “questionable” and “putative” regions detected by both tools might be potential “remnants” prophages.84

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.80,88 Previous studies depicted the crucial role of recombination in the transmission of fluoroquinolones, beta-lactams, trimethoprim, and sulfamethoxazole resistance in S. pneumoniae.56,83,89 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), slrA (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.71 The SlrA is yet another virulence gene, promoting upper airways colonization during the initial step of pneumococcal infection.71,90 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.71,91 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.92,93 However, the cps locus organization
facilitates the recombination events favoring the emergence of new capsules variants clones, more commonly known as NVT through capsule switching mechanism. 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. Similar events occurred in NC, where an increase of serotype 15B and 19A prevalence were described after PCV7 introduction in 2008.

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. 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.

**Dating lineage by Bayesian evolutionary analysis**

Prior to BEAST analysis, we tested the clock-likeness of the data set by Tempest software. 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.

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. 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. 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.

**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
1. 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.
2. Henriques-Normark B, Tuomanen EI. The pneumococcus: epidemiology, microbiology, and pathogenesis. Cold Spring Harb Perspect Med. 2013;3:a010215. doi:10.1101/cshperspect.a010215.
3. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. Lancet Infect Dis. 2005;5:83-93.
4. Ritchie ND, Mitchell TJ, Evans TJ. What is different about serotype 1 pneumococci? Future Microbiol. 2002;7:43-46. doi:10.2174/fmicb.7.1.146.
5. Brueggemann AB, Spera 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.
6. Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serotypes cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. J Clin Infect Dis. 2000;30:100-121. doi:10.1086/313608.
7. Michel N, Watson M, Baumann F, Perilet 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.
8. Le Hello S, Wason M, Levy M, et al. Invasive serotype 1 Streptococcus pneu-

7
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, Pvi 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.