Examining the distribution and impact of single nucleotide polymorphisms in the capsular locus of Streptococcus pneumoniae serotype 19A

Arends, D W; Miellet, W R; Langereis, J D; Ederveen, T H A; van der Gaast-de Jongh, C E; van Scherpenzeel, M

Published in:
Infection and Immunity

DOI:
10.1128/IAI.00246-21

Publication date:
2021

Document Version
Peer reviewed version

Examining the distribution and impact of single nucleotide polymorphisms in the capsular locus of *Streptococcus pneumoniae* serotype 19A

D.W. Arends[1], W.R. Miellet[2], J.D. Langereis[1], T.H.A. Ederveen[3], C.E. van der Gaast – de Jongh[1], M. van Scherpenzeel[4, 5], M.J. Knol[2], N.M. van Sorge[6], D.J. Lefeber[5], K. Trzciński[7], E.A.M. Sanders[2, 7], H.C. Dorfmueller[8], H.J. Bootsma[1, 2], M.I. de Jonge[1]

[1]Laboratory of Medical Immunology, Radboud Center for Infectious Diseases, Radboud Institute for Molecular Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

[2]National Institute for Public Health and the Environment, Bilthoven, The Netherlands.

[3]Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands.

[4]GlycoMScan, 5349 AB Oss, The Netherlands

[5]Translational Metabolic Laboratory, Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands.

[6]Department of Medical Microbiology and Infection Prevention, Netherlands Reference Laboratory for Bacterial Meningitis, Amsterdam Institute for Infection and Immunity, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

[7]Department of Paediatric Immunology and Infectious Diseases, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands.

[8]Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee, DD1 5EH, United Kingdom

Corresponding authors: marien.dejonge@radboudumc.nl

Abstract

Streptococcus pneumoniae serotype 19A prevalence has increased after implementation of PCV7 and PCV10 vaccines. In this study, we have provided, with high accuracy, the genetic diversity of the 19A serotype in a cohort of Dutch invasive pneumococcal disease patients and asymptomatic carriers obtained in the period 2004-2016. Whole genomes of the 338 pneumococcal isolates in this cohort were sequenced and their capsule (*cps*) loci compared to examine the diversity and determine the impact on the production of CPS sugar precursors and CPS shedding. We discovered 79 types with a unique CPS locus sequence. Most variation was observed in the *rmlB* and *rmlD* genes of the TDP-Rha synthesis pathway, and in the *wzg* gene, of unknown function. Interestingly, gene variation in the *cps* locus was conserved in multiple alleles. Using RmlB and RmlD protein models, we predict that enzymatic function is not affected by the single nucleotide polymorphisms as identified. To determine if RmlB and RmlD function was affected, we analyzed nucleotide sugar levels using UHPLC-MS. CPS precursors differed between 19A *cps* locus subtypes, including TDP-Rha, but no clear correlation was observed. Also, a significant difference in multiple nucleotide sugar levels was observed between phylogenetically branched groups. Because of indications of a role for Wzg in capsule shedding, we analyzed if this was affected. No clear indication of a direct role in shedding was found. We thus describe genotypic variety in *rmlB*, *rmlD* and *wzg* in serotype 19A the Netherlands, for which we have not discovered an associated phenotype.

52 **Introduction**

53

54 *Streptococcus pneumoniae*, is a common resident of the human upper respiratory tract. It
55 can disseminate into the lungs, causing pneumonia, and invade into the bloodstream
56 leading to sepsis and meningitis. Invasive infections give rise to high morbidity and mortality
57 rates worldwide, especially in young children, the elderly and immuno-compromised
58 individuals. Pneumococcal disease is often occurring as co- or secondary infection especially
59 in influenza, or influenza-like-illness. It is also a common cause of otitis media in children (1).

60 *S. pneumoniae* produces capsular polysaccharides (CPS), which are an important
61 virulence factor. The CPS was shown to protect the bacterium against complement-
62 mediated opsonophagocytosis and multiple other antibacterial pathways, by shielding its
63 immunogenic surface proteins from binding by host factors, such as complement factors or
64 antibodies (2-5). Virulence is affected by capsule thickness, charge and chemical properties
65 (1, 6, 7). The negatively charged CPS is thought to promote colonization of the upper
66 respiratory tract, by repelling host mucopolysaccharides which reduces mucosal clearance
67 (8). *S. pneumoniae* serotype is determined by the polysaccharide antigen, and at present
68 around 100 have been described (9). For almost all serotypes, genes encoding the formation
69 of the capsule are located in the capsular gene locus, an operon regulated through a single
70 promoter region. The *cps* locus is located between *dexB* and *aliA* (10). Because of its
71 important role in virulence and its accessibility to the immune system, the CPS, has been the
72 main target for vaccine development.

73 Multiple *S. pneumoniae* vaccines have been developed, of which pneumococcal
74 conjugate vaccines (PCVs), targeting the CPS of specific serotypes, not only protect against
75 disease, but also interfere with transmission by prevention of colonization of the
76 nasopharynx (11, 12). After implementation of the PCV 7 vaccine, targeting *S. pneumoniae*
77 serotypes 4, 6B, 9V, 14, 18C, 19F and 23F, prevalence of invasive pneumococcal disease
78 (IPD) cases caused by non-vaccine types have increased across multiple sites (12) including
79 the Netherlands (13), as a consequence of serotype replacement and/or capsule switching.
80 In the early 2010s, new conjugate vaccines were developed that covered additional
81 serotypes: PCV10 (PCV7 serotypes plus 1, 5 and 7F) and PCV13 (PCV10 serotypes plus 3, 6A
82 and 19A). Most Western European countries have implemented the PCV13 vaccine, whilst
83 the Netherlands has implemented and is still using PCV10. There is limited evidence for

vaccine cross-protection: protection provided by antibodies, raised against a specific serotype, binding to CPS of another serotype with a similar structure. Despite high structural similarity with 19F, limited protection against 19A infection was found after PCV10 vaccination (14-16). In other studies a complete absence of cross-protection was observed (17, 18) and several studies were inconclusive (19, 20). In most countries that have implemented PCV10, 19A prevalence has been on the rise (12). In countries in Western Europe using PCV10, a marked decrease in PCV10-serotype IPD was reported; however, the proportion of IPD due to PCV13-unique serotypes remained high at 58-64%, predominantly due to serotypes 19A and 3. Although PCV13 implementation has diminished the prevalence of 19A in most of these countries, 19A is one of the most frequently found emerging serotypes, even in some countries that have implemented PCV13 (21). Also, several cases of 19A infections in hospitalized individuals vaccinated with PCV13 have been reported (22). Therefore, examination of the 19A serotype remains relevant, despite an increasing number of countries that are implementing PCV13.

Serotype 19A belongs to the serotype group 19, containing 19A, 19B, 19C and 19F. The capsules of 19F and 19A differ only in the link between glucose (Glc) and rhamnose (Rha): 1->2 or 1->3, respectively (23). However, the genes in the *cps* locus are relatively different (70-99% similarity) (23). Based on different alleles within the *cps* locus (Fig. 1), multiple 19A subtypes have been described (24). Generally, the highest sequence diversity is in the regulatory gene *wzg* and the TDP-Rha biosynthesis genes *rmlC*, *rmlB* and *rmlD* (24). In some subtypes the *rmlD* gene has been flipped and is in reverse position (3'-5') in the locus (24). The *rml* genes are all required for TDP-Rha biosynthesis, which in turn is required for 19A polysaccharide capsule production. RmlA forms a tetramer and converts glucose-1-phosphate into TDP-glucose, which is oxidized and dehydrated by a RmlB dimer to form TDP-4-keto-6-deoxy-D-glucose. Dimeric RmlC in turn catalyzes a double epimerization reaction, after which RmlD monomers forms TDP-L-rhamnose (25). The end product of the biosynthetic pathway of the 19A CPS is a polymer of a repeat unit trisaccharide TDP-Rha-P-ManNAc-Glc (Fig. 1), which is connected to the cell wall peptidoglycan (23).

The function of *wzg* is unknown, making it difficult to speculate about the effect of variation within *wzg* sequence. Wzg is required for normal CPS levels. It was shown to possess enzymatic properties to anchor the CPS to the peptidoglycan cell wall (26, 27), which might be involved in capsule shedding, the release of CPS by the bacteria. This might

function as a decoy for antibody binding, leading to immune evasion. Increased capsule shedding could also increase the likelihood of epithelial cell invasion and host-host transmission (27-29). The consequences of the observed genetic diversity in the *cps* locus of 19A are still unclear. In this study, we describe the genetic diversity of the 19A serotype in the Dutch population and our attempt to reveal its consequences and driving forces.

We subjected strains collected from the nasopharyngeal or oropharyngeal cavity in healthy individuals (carriage strains) and strains collected from clinical cases of *S. pneumoniae* infection (IPD strains) to whole genome sequencing. Samples ranged from the period before and after implementation of the PCV7 and PCV10 vaccine. The *cps* locus sequences were compared to examine whether certain single nucleotide polymorphisms (SNPs) and/or 19A subtypes were associated with carriage or disease, and to characterize the diversity over time upon potential selective pressure by both PCV vaccines. Furthermore, nucleotide sugars, including TDP-Rha, were measured and structural modelling on RmlB and D was performed in order to understand the impact of SNP accumulation in genes encoding these proteins. Additionally, we examined capsule shedding for the different Wzg proteins we observed.

Results

Sequence variation in the *cps* locus and the identification of 19A subtypes

Considerable DNA sequence variation within the 19A *cps* locus was reported previously (24, 30), referred to as capsular subtypes. Elberse *et al*. (24) described that after implementation of the PCV7 vaccine (2006), an apparent shift in the prevalence of 19A capsular subtypes in invasive pneumococcal disease (IPD) was observed in the Netherlands, from a majority of 19A-I to a majority of 19A-II (2004-2005 vs. 2008-2009). A similar shift after PCV7 implementation was also observed by Brugger *et al.* (30) in Switzerland. We investigated whether the post-PCV7 emergence 19A-II continued using the same allele-specific screening PCR methodology. However, the distribution of capsular subtypes in recent years (2013-2016) seems to be returning to the distribution as observed before implementation of PCV vaccination (Suppl. Fig. 1).

For a more detailed view on the differences in 19A *cps* locus sequence, we subjected pneumococcal strains isolated from carriage (n = 148) and IPD isolates (n = 188, PBCN cohort, (31)) to whole-genome sequencing and extracted the complete *cps* locus sequence from the assemblies (Table 1). Initial comparison of the *cps* gene cluster sequences was performed by cpsMLST, a serotype-specific MLST-like scheme based on *cps* genes. Among the 338 isolates from which complete *cps* sequences were available, 100 unique cpsMLST types were detected, providing higher resolution than the previously defined capsular subtypes identified by PCR (Suppl. Fig. 2). The cpsMLST types correlated with PCR subtypes, but additional genetic variation within the subtype could be observed. No particular cpsMLST type could be linked to either isolation period (pre- and post PCV vaccination; Fig.2A) or isolation site (carrier *vs*. IPD; Fig. 2B). Phylogenetic analysis using the whole genome sequence (Suppl. Fig. 3) revealed a diverse genetic background for each 19A *cps* subtype (Suppl. Fig. 3B-C), suggesting that the current distribution of 19A *cps* loci is partially a result of recombination events.

The distribution of non-synonymous SNPs across the capsular locus genes

To identify potential functional consequences of the sequence variations, we identified all non-synonymous single-nucleotide polymorphisms (nsSNPs) within the *cps* operon using the

sequence from an IPD isolate from the Netherlands as reference. This analysis initially revealed **49** unique *cps*-nsSNP types for which most experiments were performed. Later, additional isolates were examined resulting in **79** nsSNP types within 338 isolates. Table 2 shows to which previously described subtypes each SNP type belongs. Several mutation hotspots were observed, most notably within *rmlB* and *rmlD* (respectively 87.7-94.2% and 77.2-99.5% sequence similarity compared to Hungary-19A-6), which are responsible for the synthesis of TDP-L-rhamnose, which is a component of the 19A CPS; and within the *wzg* gene (92.3-94.6%), a predicted regulator of CPS synthesis, with the possible enzymatic function of linking the CPS to the cell wall peptidoglycan (27). Our analysis reveals that although these genes are mutation hotspots, the 19A *cps* subtypes (19A-I, II, II-Ins, III, IV, V, VI and VIII) generally show a high similarity for the *rmlA-D* gene cluster within each subtype, suggesting sequence conservation within the *cps* loci of the different 19A subtypes. Interestingly, the *rmlD* in subtypes 19A-I, -II, -II-Ins, and IV is situated in reverse direction and on the complement strand. Previously, Morona *et al.* 1999 have reported a potential promoter upstream of *rmlD* on the opposite strand as that for *aliA*. They also described a stem-loop structure between *rmlB* and *rmlD* which could potentially function as a transcription terminator (23). Perhaps this reverse complement *rmlD* is regulated by another promoter. These subtypes also share the same *rmlB* and *rmlD* genes (100% similarity; except one point mutation in *rmlB* of SNP1). Subtypes 19A-III and -V share a similar *rmlB* (99.90-100% similarity) and the same *rmlD* (100% similarity; except one point mutation for SNP30) genes. 19A-VI and -VIII, and Hungary 19A-6 and Bentley 19A reference strain sequences comprise the other group containing similar *rmlB* and *rmlD* (99.4-100%) genes. The *wzg* sequences also show high similarity within the subtype, but fewer similarities are found between subtypes. Only subtypes 19A-III, -V and -VI share the same *wzg* sequence (except one point mutation in SNP48). We observed that subtype V had an additional [YGX]-repeat (4 *vs*. 3 repeats) in the $[YGX]_3$-repeat domain of the *wze* gene compared to the other subtypes. There is evidence that suggests that phosphorylation of tyrosine residues in that domain affects capsule levels (32).

Variations in the promoter region

We also compared the promoter region of the *cps* operon based on the four reference promoter elements described by Wen *et al.* (33): insertional element (IE), repeat unit of

pneumococcus (RUP), spacing sequence (SS) and a core promoter sequence. All isolates had a promoter containing a largely conserved 31 nucleotide region prior to the start codon, followed by a completely conserved core promoter. Differences in the remainder of the promoter between the isolates can be largely grouped into two distinct promoter types. Type I starts with the IS630 IE, followed by the RUP and the SS. Some strains in this group only contain a small partial IS630 IE, or do not have it at all. One isolate did not have a RUP either and only a partial spacing sequence. In isolates of the subtype 19A-IV, repeats of parts of the SS were found. Type II had its IE (IS1201) after its largely intact RUP, and only a partial spacing sequence, showing the same differences as observed by Wen *et al.* The RUP in this group had an altered binding site, compared to the one described by Wu *et al.* (34), for transcription repressor CpsR, possibly affecting its binding. The distribution of the majority of the promoters, but not all, was linked to CPS locus sequence and 19A *cps* subtype grouping (Supplementary figure 3D). Two thirds of subtype I had a type II promoter, the remainder type I. Strikingly, SNP types SNP9 and SNP10 had identical CPS sequences, but either a type I or type II promoter. All the other 19A *cps* subtypes had a type I promoter.

The effect of nsSNPs on protein structure

As the *rmlB* and *rmlD* genes contained the highest rates of non-synonymous SNPs (nsSNPs) within the gene cluster, we decided to further investigate the protein sequence of RmlB and RmlD. We examined the amino acid sequence and structure of these enzymes of the TDP-rhamnose biosynthesis pathway (Fig. 3). Subtypes 19A-I, -II and -IV all share identical RmlB and RmlD proteins, as do subtypes 19A-III and -V (19A-VIII also shares the 19A-III and -V RmlD). To determine the number of proteins that would be translated from the different alleles, the nucleotide sequence was converted to an amino acid sequence. For Wzg, RmlB and RmlD, 9, 7 and 6 different proteins were found, respectively.

Next, we examined whether the non-synonymous SNPs would affect protein structure and function. Using protein structure analysis, we investigated the location of the mutations and their potential role in catalytic activity. The streptococcal RmlB proteins form a functional homo dimer (35), whilst RmlD enzymes are monomeric (36). We systematically investigated the location of the 12 amino acid substitutions compared to the Hungary reference sequence that are the result of the SNPs in a structural model of RmlB, using the *Streptococcus suis* RmlB (90-92% sequence identity) structure as a reference (37). All

residues are located on the surface of the RmlB homodimer and none, except K38, are in proximity to the substrate or co-factor binding site (Fig. 3B). A K38R mutations is found in the Hungary strain, with its sidechain sitting on top of the aromatic ring of the nucleotide (Fig. 3B). However, the K to R mutation is a conserved mutation and should not impact the enzymes ability to bind the substrate. Importantly, non-synonymous SNPs in RmlB were also not found at the dimer interface, which would potentially disrupt functionality. Therefore, the mutations appear to be subtle and are likely not to affect the RmlB protein structure, protein activity and dimerization.

Interestingly, a higher rate of SNPs was found in RmlD (3 per 100 residues). RmlD is a monomeric protein in Gram-positive bacteria, and the structure of the closely related RmlD from *Streptococcus pyogenes* (78-81% sequence identity) was reported recently (36). A structural model of *S. pneumoniae* RmlD was built and analyzed to determine the potential impact of the nsSNPs (Fig 3C). Like the nsSNPs found in RmlB, all non-silent mutations were located on the surface of RmlD, facing away from the catalytic machinery.

The consequence of nsSNPs in *rmlB* and *rmlD* on nucleotide sugar levels

To determine the consequence of nsSNPs in the *rml* genes we conducted sensitive ion-pair UHPLC-mass spectrometry to measure whole cell levels of nucleotide sugars (38). Using the then available 49 isolates, each with a unique cpsSNP type (SNP1-49), we focused on the building blocks of the 19A polysaccharide: TDP-Rha, UDP-Glc, UDP-GlcNAc, UDP-ManNAc and TDP-Glc (see figure 1 for the CPS synthesis by the CPS gene products). Levels of nucleotide sugars not related to the CPS were also analyzed to decrease the effect of changes in levels on the relative amounts measured. As a control, we used serotype 19F isolates EF3030 and BHN100, and their isogenic capsular knockout derivatives (39). Strong variation in amounts of nucleotide sugars produced by the 49 isolates was observed (Figure 4). Strikingly, no clear correlation between 19A *cps* subtype or RmlB/RmlD/Wzg protein type (data not shown) and TDP-Rha production was observed. The EF3030 and BHN100 capsular knockouts showed no production of TDP-Rha, as expected. Interestingly, the unencapsulated seemed to have more of the other nucleotide sugars. When comparing nucleotide sugars as fractions of the total amount of measured sugar, we did not see distinct differences in the relative ratios of the nucleotide sugars (Suppl. Fig. 5).

A significant difference in the median of the following nucleotide sugars was found: UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc and GDP-Man ($p < 0.01$), UDP-ManNAc, TDP-Rha, TDP-Glc and UDP-GlcA ($p < 0.05$), as well as the total amount of nucleotide sugars ($p < 0.005$) (Kruskal-Wallis). Significant differences between specific subtypes was only found recurring in subtype I vs. III (UDP-GlcNAc and UDP-Gal) and II vs III (UDP-GlcNAc, UDP-Gal, UDP-GalNAc, TDP-Rha and total sugar). No significant difference between subtypes was observed for the remainder of the nucleotide sugars.

The phylogenetic tree based on the cpsMLST shows that the cpsSNP types are divided in to two groups: SNPs1-29, 49 and SNPs30-48 (Fig. 4). Multiple Mann-Whitney U tests show that the former group produces significantly more monosaccharides in total ($p < 0.01$) and for the following sugars specifically: GDP-Glc ($p < 0.05$), CMP-Neu5Ac, UDP-Gal, UDP-Glc, UDP-GlcNAc, UDP-GalNAc, ManNAc, GDP-Fuc and TDP-Rha ($p < 0.01$).

Genetic variation of *wzg* and effect on capsule shedding

The role of Wzg has yet to be elucidated but it is thought to regulate expression of the *cps* operon, however the mechanism remains unknown. Previously, others have described a potential enzymatic function for Wzg and other related proteins in *B. subtilis* which suggest that Wzg is involved in coupling teichoic acids and capsular polysaccharides to the peptidoglycan layer (26, 27, 40). We hypothesized that the Wzg proteins (Table 3) translated from the distinct alleles, as observed within this cohort, affected the enzymatic function, which also might affect polysaccharide capsule shedding. To determine whether there was a difference in the amount of CPS shedded between strains with the different *wzg* alleles, supernatants of isolate cultures were collected and transferred in triplicates on a membrane for Western dot blotting (Supplementary Fig. 4), following a method that was similar to one previously described by Kietzmann *et al* (28). Due to substantial proteinaceous background binding observed for anti-serotype 19 serum, we used serum from PCV13-vaccinated mice. Where possible, three to four isolates per Wzg protein type were examined. Although differences in shedding were observed (Fig. 5A), they could not be linked to amino acid sequence variation of Wzg proteins (Fig. 5B). Grouping of Wzg types based on >5 amino acid substitutions, could also not be associated with specific shedding levels (data not shown), which suggests that there is an insignificant role for the observed allelic variation in Wzg in capsule shedding.

Materials and Methods

Genetic analysis of *S. pneumoniae* isolates

The cohort consisted of patients from all age groups diagnosed with invasive pneumococcal disease admitted to 22 Dutch hospitals (having blood cultures assessed in 9 sentinel laboratories) between 2004 and 2016 (31). Blood isolates were obtained from the Netherlands Reference Laboratory of Bacterial Meningitis (NRLBM) and the Pneumococcal Bacteraemia Collection Nijmegen (PBCN). Also isolates of a collection of *S. pneumoniae* carriers were obtained from the National Institute for Public Health and the Environment (RIVM), which was collected for screening serotype prevalence spanning the period of 2006-2016. The isolates were subjected to whole genome sequencing, which was performed by Baseclear using Illumina NGS. CPS loci from Elberse *et al.* 2011 (24), were obtained by Sanger sequencing. Assemblies were constructed in CLC Genomics Workbench/Genome assembly of short reads into contigs was performed using SPAdes, which were subsequently assigned a taxonomic origin with Kraken2. All DNA sequences identified as pneumococcus were used for further analysis. The NGS data were used for cpsMLST and whole genome (wg) MLST analyses using SeqSphere software version 6.0.2 (Ridom GmbH, Münster, Germany). The cpsMLST scheme was based on all genes in the 19A cps operon, the in-house wgMLST scheme was comprised of 1942 genes (1210 core-genome and 732 accessory-genome targets) using S. pneumoniae TIGR4 (NC_003028.3) as a reference genome. Bionumerics was used for MLST minimum spanning tree analysis, for sequence alignment and for calling of non-synonymous single nucleotide polymorphisms. 19A reference strains were obtained from Bentley *et al.* (10) and McGee *et al.* (Hungary 19A-6) (41).

Mass spectrometry analysis of nucleotide sugars

The *Streptococcus pneumoniae* 19A strains were grown overnight on blood agar plates (BD™ Columbia III Agar with 5% Sheep Blood 254098, containing 12 g/L Pancreatic Digest of Casein, 5 g/L Peptic Digest of Animal Tissue, 3 g/L Yeast Extract, 3 g/L Beef Extract, 1 g/L Corn Starch, 5 g/L Sodium Chloride, 13.5 g/L Agar 4g/L Growth factors, 5% Sheep Blood, Defibrinated, pH 7,3 ± 0.2), at 37 °C and 5% CO_2. The next day, single colonies were inoculated into 30 mL THY broth (Difco 249240 and Labconsult SA-NV CON.1702; 3.1 g/L Heart, Infusion from 500 g, 20 g/L Neopeptone, 2 g/L Dextrose, 2 g/L Sodium Chloride, 0.4

g/L Disodium Phosphate, 2.5 g/L Sodium Carbonate, 5 g/L yeast) in 50 mL tubes at 37 °C, 5% CO_2 and grown until an optical density measured at a wavelength of 620 nm of 0.3. Subsequently, samples were placed on ice with NaCl (Merck, 1064041000) immediately and spun down by centrifugation (1 min, pre-cooled centrifuge at 4 °C, 3,220 x *g*). Supernatant was discarded and the pellet was washed with 1.8 mL wash buffer (75 mM Ammonium Carbonate, Sigma 207861) in MilliQ, buffered with acetic acid (Sigma A6283) pH 7.4 and cooled to 4 °C prior to use. The final suspension was transferred to a 2 mL tube and spun down by centrifugation (1 minute at 4 °C, 25,000 x *g*). Supernatant was discarded and the pellet was stored at -80 °C until sample preparation. Frozen cells were extracted at -20˚C with 1 mL pre-cooled 2:2:1 (v/v/v) methanol/acetonitrile/water (methanol: VWR 20846.361, acetonitrile Merck 1000291000, water: B. Braun Medical 3624331) for 5 min. Subsequently, this was centrifuged at 25,000 x *g* for 3 minutes at 4^{o}C. The resulting supernatants were dried overnight using a vacuum centrifuge at room temperature, the pellets were stored at -80 °C.

Liquid Chromatography, UHPLC-MRM acquisition and data analysis were performed as previously described (38).

Protein structure analysis

The structural model for RmlB dimer and RmlD monomer based on the *S. pneumoniae* Hungary19A-6 subtype were build using the Swiss-Model server (42). Structural templates were 1ker.pdb for RmlB (37), and 4wpg.pdb and 1kc3.pdb for RmlD (36, 43). Molecular graphics and analyses were performed with the UCSF Chimera package, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311) (44).

Capsule shedding analysis

S. pneumoniae 19A isolates were grown overnight as previously described. The next day, bacteria were harvested from the plates and inoculated into 45% M17 – 45% CAT – 10% FCS broth with catalase (16.3 g/L M17, 4.5 g/L Casamino acids, 2.25 g/L Sodium Chloride, 2.25 g/L Bacto-tryptone, 4.5 g/L yeast extract, 2.25 g/L Glucose, Catalase 22 units/mL (Sigma Aldrich, C3155)) in 10 mL tubes at 37°C, 5% CO_2 and grown until an optical density measured at a wavelength of 620 nm of 0.3-0.5 (starting OD: 0.05). Samples were then spun down by

centrifugation (10 min, room temperature, 3220 x *g*) and supernatant was collected. Then, remaining bacteria were removed using 0.2 µm PES membrane syringe filters. Following, supernatant was diluted in PBS (1:5) before application to a nitrocellulose membrane using a GE Whatman – 10447900 Acrylic Minifold I Dot-Blot 96 Well Plate System. Afterwards the membrane was blocked using 5% BSA in PBS for 1 h at room temperature while shaking. Then the membrane was incubated in 0.5% BSA, 0.1% tween, 0.1-0.2% serum from PCV13 vaccinated mice (containing 22.29 µg/mL anti-19A IgG) in PBS for 1 h at room temperature while shaking. Membranes were then washed 5 x 5 min in PBST and incubated in 0.5% BSA, 0.1% tween, 0.01% rabbit-HRP anti mouse serum in PBS for 1 h at room temperature while shaking. After another round of washing (5 x 5 min in PBST) ECL plus reagent (GE) was applied and luminescence detected using a ChemiDoc XRS+ System (Bio-rad). Intensity was corrected to growth culture OD, to relegate the influence of the difference in number of bacteria per sample.

Discussion

After implementation of PCV7, changes in 19A subtype prevalence were observed (12), suggesting that the vaccination resulted in selective pressure despite the fact that 19F and not 19A polysaccharides were included in this vaccine. We wished to further examine the distribution of the 19A subtypes and to find out what caused the change in prevalence. We discovered that the current prevalence resembled the prevalence of early post-PCV7 vaccination. Possibly the initial change was due to an unfortunate coincidence of a low circulation of 19A, due to natural fluctuation, immediately after PCV7 introduction (45).

Using cpsMLST analysis we found similar clustering of subtypes as observed using PCR, but with greater resolution, showing variation within the subtype. As subtype distribution could not be related to either vaccination or disease, we could not identify the factors that drive the genetic diversity. The protein models of RmlB and RmlD, two of the proteins that vary the most, showed no effect on either dimerization (RmlB only), substrate binding or catalytic activity. Supporting this, TDP-Rha differences could not be linked to specific subtypes or Rml sequences, although a difference in the median of all subtypes was found. Also, it is possible that CPS thickness could be affected, even with similar TDP-Rha levels. However, we could not find, or were able to design, a reliable method for measuring its thickness for such a large number of isolates. Resolution of documented methods, such as fluorescein isothiocyanate (FITC)-dextran exclusion (29) was too low to measure subtle differences in capsule thickness. It could be expected that Wzy, which is responsible for CPS polymerization, would be affected, if immunological pressure is selecting for mutations resulting in reduced or increased thickness of the capsule. However, this gene appeared to be highly conserved, hardly any SNPs were found in *wzy*. TDP-Rha was shown to inhibit the RmlA enzyme in a negative feedback loop, therefore regulating the amount of TDP-Rha being produced in bacteria (46). Therefore, it is also possible that the increased activity of RmlB and RmlD, due to gain of function mutations (SNPs), is downregulated by the product inhibition of RmlA, leading to minimal differences in TDP-Rha synthesis as net result. Thus, differences in TDP-Rha observed could be caused by a multitude of processes, such as an affected synthesis of TDP-Rha or its building blocks, other unknown factors regulating RmlA-D activity, and the thickness or shedding of the CPS. Capsular production could be enhanced, requiring more nucleotide sugars to be synthesized, resulting in higher levels

observed. Problems in the process of capsule formation could also possibly lead to a buildup of nucleotide sugars where the lack of/or inhibition of certain enzymes could form a bottleneck, although buildup of specific monosaccharides would be expected there. Perhaps the observed difference in UDP-Glc and UDP-GlcNAc levels are a result of an impaired capsule formation. In this case dTDP-Rha levels might not be different because of its negative-feedback loop, which possibly might also be occurring for UDP-ManNAc, although no evidence of this exists. It also cannot be excluded that differences in specific and/or total monosaccharide/nucleotide sugar levels could also be attributed to genomic differences in genes outside of the capsular locus.

We found large differences in whole cell total sugar levels between the different serotype 19A strains grown in the same conditions. These surprising differences could point to major differences in sugar metabolism within this serotype, although differences in relative fractions of the individual sugars might be expected then. Although it is possible that the bacteria grown to the same optical density differed in actual number of bacteria, it is unlikely that subtle variations in the actual bacterial load would lead to such significant differences. Possibly, the differences in total amount reflect differences in capsule thickness, where bacteria with thicker capsules have a higher total sugar count. A higher fraction of capsular sugars would be expected here, which was not wat we observed. Possibly, upregulation of total sugar metabolism could be required for thicker capsules. Another possibility for the observed difference is different levels of shedding. Cells that shed more capsule, possibly require a higher sugar metabolism. Or, lower amounts of total sugars, could be the result of higher levels of shedding, where shedded capsule is lost in the sample preparation. However, we did not find a correlation between shedding levels and total sugar levels in the examined SNP types. The significant difference in total sugars between the two phylogenetically branched groups, suggests that the *cps* locus is involved in this observed difference, although it remains unclear in what way. It will be interesting to find out if these differences in total amount of sugars translate to differences in other bacterial processes.

There could be multiple explanations for the existence of the variation in the *rmlB* and *rmlD* genes. Apparently, these genes allow accumulation of nsSNPs without affecting protein function, while this is not the case for *rmlA* and *rmlC*. Those genes might be more vulnerable for loss of function. In *Pseudomonas aeruginosa,* RmlA is a functional tetramer, with multiple multimerization domains and an allosteric pocket, in which mutations also

cause inactivation of the enzyme (47). RmlB and RmlC both form a homodimer, but in RmlC active site residues from both monomers are required (25, 35, 48). RmlD is monomeric and therefore potentially could have the most SNPs not affecting its function (36). However, following this explanation one could expect a more random variation within these genes. However, variation seems to be conserved resulting in a limited number of alleles and protein variants. These alleles are largely distributed over two distinct groups, SNPs1-29 and 49, and SNPs30-48. A difference in nucleotide sugars levels was also observed for these groups. Potentially, this is a result of a selective pressure, in the absence of PCV-13 as a selective pressure. Variation could have been caused by selection for metabolic advantages in certain environments. Another possibility is that we are simply observing the natural genetic drift of the 19A serotype. The observed difference could also be the result of multiple imperfect recombination events: the 19A *cps* locus is found in different genomic backgrounds and the most genetically diverse genes are on the edges of the *cps* locus. More experiments are required to further understand this distribution and the possible selective pressure behind it.

Wzg also showed a high level of SNPs. As it is shown to be involved in the CPS synthesis and required for normal CPS levels (27), it is possible that difference in sequence could affect capsule thickness. Bacteria lacking *wzg* have a decreased amount of capsule (27). A decrease in CPS expression, might also lead to a decrease in shedding (29). The different levels of sugars observed could therefore also be an indicator for differences in shedding. In our experiments, we observed large differences in capsule shedding, however, this was not associated with *wzg* sequence variation.

We also reported differences in the CPS locus promoter, even within 19A-subtypes. We have yet to pursue further examination of its relevance. It will be interesting to determine which consequences these differences have on gene expression and whether it has an evolutionary advantage. Possibly invasiveness, disease progression or colonization are affected.

To conclude, we have shown, with increased resolution, the variation within the 19A serotype in a cohort from the Netherlands, observing 79 nonsynonymous cpsSNP types. We were unable to find a correlation between cpsSNP type and isolate origin. However, we did observe conserved *wzg*, *rmlB* and *rmlD* alleles within the previously described 19A subtypes. Further studies are required to investigate the drivers of genetic variation in the *cps* locus as

described in this study. This will lead to the improved understanding of the effects of host-pathogen interactions that result in pneumococcal capsule variation and its dynamics.

469 **Funding**

470 This work was funded by the RIVM, Bilthoven, The Netherlands, project S/115002/01/SP,
471 (Sp-cas). Helge Dorfmueller is funded by Royal Society and Wellcome Trust Sir Henry Dale
472 Fellowship [109357/Z/15/Z].

475

476 **References**

477 1. Engholm DH, Kilian M, Goodsell DS, Andersen ES, Kjærgaard RS. 2017. A visual review of the
478 human pathogen Streptococcus pneumoniae. FEMS Microbiol Rev 41:854-879.
479 2. Abeyta M, Hardy GG, Yother J. 2003. Genetic alteration of capsule type but not PspA type
480 affects accessibility of surface-bound complement and surface antigens of Streptococcus
481 pneumoniae. Infect Immun 71:218-25.
482 3. de Vos AF, Dessing MC, Lammers AJ, de Porto AP, Florquin S, de Boer OJ, de Beer R, Terpstra
483 S, Bootsma HJ, Hermans PW, van 't Veer C, van der Poll T. 2015. The polysaccharide capsule
484 of Streptococcus pneumonia partially impedes MyD88-mediated immunity during
485 pneumonia in mice. PLoS One 10:e0118181.
486 4. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. 2010. The Streptococcus pneumoniae
487 capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms.
488 Infect Immun 78:704-15.
489 5. Wartha F, Beiter K, Albiger B, Fernebro J, Zychlinsky A, Normark S, Henriques-Normark B.
490 2007. Capsule and D-alanylated lipoteichoic acids protect Streptococcus pneumoniae against
491 neutrophil extracellular traps. Cell Microbiol 9:1162-71.
492 6. Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of Streptococcus pneumoniae
493 virulence factors in host respiratory colonization and disease. Nat Rev Microbiol 6:288-301.
494 7. Austrian R. 1981. Some observations on the pneumococcus and on the current status of
495 pneumococcal disease and its prevention. Rev Infect Dis 3 Suppl:S1-17.
496 8. Nelson AL, Roche AM, Gould JM, Chim K, Ratner AJ, Weiser JN. 2007. Capsule enhances
497 pneumococcal colonization by limiting mucus-mediated clearance. Infect Immun 75:83-90.
498 9. Project GPS. 2021. https://www.pneumogen.net/gps/serotypes.html. Accessed
499 10. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K,
500 Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kaltoft MS, Barrell B, Reeves PR,
501 Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular biosynthetic locus from all 90
502 pneumococcal serotypes. PLoS Genet 2:e31.
503 11. Bosch A, van Houten MA, Bruin JP, Wijmenga-Monsuur AJ, Trzciński K, Bogaert D, Rots NY,
504 Sanders EAM. 2016. Nasopharyngeal carriage of Streptococcus pneumoniae and other
505 bacteria in the 7th year after implementation of the pneumococcal conjugate vaccine in the
506 Netherlands. Vaccine 34:531-539.
507 12. Isturiz R, Sings HL, Hilton B, Arguedas A, Reinert RR, Jodar L. 2017. Streptococcus
508 pneumoniae serotype 19A: worldwide epidemiology. Expert Rev Vaccines 16:1007-1027.

13. Vissers M, Wijmenga-Monsuur AJ, Knol MJ, Badoux P, van Houten MA, van der Ende A, Sanders EAM, Rots NY. 2018. Increased carriage of non-vaccine serotypes with low invasive disease potential four years after switching to the 10-valent pneumococcal conjugate vaccine in The Netherlands. PLoS One 13:e0194823.
14. Jokinen J, Rinta-Kokko H, Siira L, Palmu AA, Virtanen MJ, Nohynek H, Virolainen-Julkunen A, Toropainen M, Nuorti JP. 2015. Impact of ten-valent pneumococcal conjugate vaccination on invasive pneumococcal disease in Finnish children--a population-based study. PLoS One 10:e0120290.
15. Domingues CM, Verani JR, Montenegro Renoiner EI, de Cunto Brandileone MC, Flannery B, de Oliveira LH, Santos JB, de Moraes JC. 2014. Effectiveness of ten-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in Brazil: a matched case-control study. Lancet Respir Med 2:464-71.
16. Peckeu L, van der Ende A, de Melker HE, Sanders EAM, Knol MJ. 2021. Impact and effectiveness of the 10-valent pneumococcal conjugate vaccine on invasive pneumococcal disease among children under 5 years of age in the Netherlands. Vaccine 39:431-437.
17. Hammitt LL, Akech DO, Morpeth SC, Karani A, Kihuha N, Nyongesa S, Bwanaali T, Mumbo E, Kamau T, Sharif SK, Scott JA. 2014. Population effect of 10-valent pneumococcal conjugate vaccine on nasopharyngeal carriage of Streptococcus pneumoniae and non-typeable Haemophilus influenzae in Kilifi, Kenya: findings from cross-sectional carriage studies. Lancet Glob Health 2:e397-405.
18. Naucler P, Galanis I, Morfeldt E, Darenberg J, Örtqvist Å, Henriques-Normark B. 2017. Comparison of the impact of pneumococcal conjugate vaccine 10 or pneumococcal conjugate vaccine 13 on invasive pneumococcal disease in equivalent populations. Clin Infect Dis 65:1780-1789.
19. Hausdorff WP, Hoet B, Schuerman L. 2010. Do pneumococcal conjugate vaccines provide any cross-protection against serotype 19A? BMC Pediatr 10:4.
20. Knol MJ, Wagenvoort GH, Sanders EA, Elberse K, Vlaminckx BJ, de Melker HE, van der Ende A. 2015. Invasive pneumococcal disease 3 years after introduction of 10-valent pneumococcal conjugate vaccine, the Netherlands. Emerg Infect Dis 21:2040-4.
21. Agudelo CI, Castañeda-Orjuela C, Brandileone MCC, Echániz-Aviles G, Almeida SCG, Carnalla-Barajas MN, Regueira M, Fossati S, Alarcón P, Araya P, Duarte C, Sánchez J, Novas M, Toraño-Peraza G, Rodríguez-Ortega M, Chamorro-Cortesi G, Kawabata A, García-Gabarrot G, Camou T, Spadola E, Payares D, Andrade AL, Di Fabio JL, Castañeda E. 2020. The direct effect of pneumococcal conjugate vaccines on invasive pneumococcal disease in children in the

Latin American and Caribbean region (SIREVA 2006-17): a multicentre, retrospective observational study. Lancet Infect Dis doi:10.1016/s1473-3099(20)30489-8.

22. Ozkaya-Parlakay A, Polat M, Bedir Demirdag T, Gulhan B, Kanik-Yuksek S, Nar-Otgun S. 2020. Vaccine failures in pediatric cases caused by streptococcus pneumoniae serotype 19A. Hum Vaccin Immunother doi:10.1080/21645515.2020.1767450:1-2.

23. Morona JK, Morona R, Paton JC. 1999. Comparative genetics of capsular polysaccharide biosynthesis in Streptococcus pneumoniae types belonging to serogroup 19. J Bacteriol 181:5355-64.

24. Elberse K, Witteveen S, van der Heide H, van de Pol I, Schot C, van der Ende A, Berbers G, Schouls L. 2011. Sequence diversity within the capsular genes of Streptococcus pneumoniae serogroup 6 and 19. PLoS One 6:e25018.

25. van der Beek SL, Zorzoli A, Çanak E, Chapman RN, Lucas K, Meyer BH, Evangelopoulos D, de Carvalho LPS, Boons GJ, Dorfmueller HC, van Sorge NM. 2019. Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. Mol Microbiol 111:951-964.

26. Kawai Y, Marles-Wright J, Cleverley RM, Emmins R, Ishikawa S, Kuwano M, Heinz N, Bui NK, Hoyland CN, Ogasawara N, Lewis RJ, Vollmer W, Daniel RA, Errington J. 2011. A widespread family of bacterial cell wall assembly proteins. Embo j 30:4931-41.

27. Eberhardt A, Hoyland CN, Vollmer D, Bisle S, Cleverley RM, Johnsborg O, Håvarstein LS, Lewis RJ, Vollmer W. 2012. Attachment of capsular polysaccharide to the cell wall in Streptococcus pneumoniae. Microb Drug Resist 18:240-55.

28. Kietzman CC, Gao G, Mann B, Myers L, Tuomanen EI. 2016. Dynamic capsule restructuring by the main pneumococcal autolysin LytA in response to the epithelium. Nat Commun 7:10859.

29. Zafar MA, Hamaguchi S, Zangari T, Cammer M, Weiser JN. 2017. Capsule type and amount affect shedding and transmission of Streptococcus pneumoniae. mBio 8.

30. Brugger SD, Troxler LJ, Rüfenacht S, Frey PM, Morand B, Geyer R, Mühlemann K, Höck S, Thormann W, Furrer J, Christen S, Hilty M. 2016. Polysaccharide capsule composition of pneumococcal serotype 19A subtypes is unaltered among subtypes and independent of the nutritional environment. Infect Immun 84:3152-3160.

31. Cremers AJH, Mobegi FM, van der Gaast-de Jongh C, van Weert M, van Opzeeland FJ, Vehkala M, Knol MJ, Bootsma HJ, Välimäki N, Croucher NJ, Meis JF, Bentley S, van Hijum S, Corander J, Zomer AL, Ferwerda G, de Jonge MI. 2019. The contribution of genetic variation of Streptococcus pneumoniae to the clinical manifestation of invasive pneumococcal disease. Clin Infect Dis 68:61-69.

32. Morona JK, Miller DC, Morona R, Paton JC. 2004. The effect that mutations in the conserved capsular polysaccharide biosynthesis genes cpsA, cpsB, and cpsD have on virulence of Streptococcus pneumoniae. J Infect Dis 189:1905-13.

33. Wen Z, Liu Y, Qu F, Zhang JR. 2016. Allelic variation of the capsule promoter diversifies encapsulation and virulence in Streptococcus pneumoniae. Sci Rep 6:30176.

34. Wu K, Xu H, Zheng Y, Wang L, Zhang X, Yin Y. 2016. CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in Streptococcus pneumoniae. Scientific Reports 6:29255.

35. Beis K, Allard ST, Hegeman AD, Murshudov G, Philp D, Naismith JH. 2003. The structure of NADH in the enzyme dTDP-d-glucose dehydratase (RmlB). J Am Chem Soc 125:11872-8.

36. van der Beek SL, Le Breton Y, Ferenbach AT, Chapman RN, van Aalten DM, Navratilova I, Boons GJ, McIver KS, van Sorge NM, Dorfmueller HC. 2015. GacA is essential for Group A Streptococcus and defines a new class of monomeric dTDP-4-dehydrorhamnose reductases (RmlD). Mol Microbiol 98:946-62.

37. Allard ST, Beis K, Giraud MF, Hegeman AD, Gross JW, Wilmouth RC, Whitfield C, Graninger M, Messner P, Allen AG, Maskell DJ, Naismith JH. 2002. Toward a structural understanding of the dehydratase mechanism. Structure 10:81-92.

38. van Scherpenzeel M, Conte F, Büll C, Ashikov A, Hermans E, Willems A, van Tol W, Kragt E, Moret EE, Heise T, Langereis JD, Rossing E, Zimmermann M, Rubio-Gozalbo ME, de Jonge MI, Adema GJ, Zamboni N, Boltje T, Lefeber DJ. 2020. Dynamic analysis of sugar metabolism reveals the mechanisms of action of synthetic sugar analogs. bioRxiv doi:10.1101/2020.09.15.288712:2020.09.15.288712.

39. Habets MN, van Selm S, van der Gaast-de Jongh CE, Diavatopoulos DA, de Jonge MI. 2017. A novel flow cytometry-based assay for the quantification of antibody-dependent pneumococcal agglutination. PLoS One 12:e0170884.

40. Gale RT, Li FKK, Sun T, Strynadka NCJ, Brown ED. 2017. B. subtilis LytR-CpsA-Psr Enzymes Transfer Wall Teichoic Acids from Authentic Lipid-Linked Substrates to Mature Peptidoglycan In Vitro. Cell Chem Biol 24:1537-1546.e4.

41. McGee L, McDougal L, Zhou J, Spratt BG, Tenover FC, George R, Hakenbeck R, Hryniewicz W, Lefévre JC, Tomasz A, Klugman KP. 2001. Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. J Clin Microbiol 39:2565-71.

42. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296-w303.

43. Blankenfeldt W, Kerr ID, Giraud M-F, McMiken HJ, Leonard G, Whitfield C, Messner P, Graninger M, Naismith JH. 2002. Variation on a theme of SDR: dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode. Structure 10:773-786.

44. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-12.

45. Prins-van Ginkel AC, Berbers GAM, Grundeken LH, Tcherniaeva I, Wittenberns JI, Elberse K, Mollema L, de Melker HE, Knol MJ. 2016. Dynamics and determinants of pneumococcal antibodies specific against 13 vaccine serotypes in the pre-vaccination era. PLOS ONE 11:e0147437.

46. Blankenfeldt W, Asuncion M, Lam JS, Naismith JH. 2000. The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). Embo j 19:6652-63.

47. Alphey MS, Pirrie L, Torrie LS, Boulkeroua WA, Gardiner M, Sarkar A, Maringer M, Oehlmann W, Brenk R, Scherman MS, McNeil M, Rejzek M, Field RA, Singh M, Gray D, Westwood NJ, Naismith JH. 2013. Allosteric competitive inhibitors of the glucose-1-phosphate thymidylyltransferase (RmlA) from Pseudomonas aeruginosa. ACS Chem Biol 8:387-96.

48. Dong C, Major LL, Allen A, Blankenfeldt W, Maskell D, Naismith JH. 2003. High-resolution structures of RmlC from Streptococcus suis in complex with substrate analogs locate the active site of this class of enzyme. Structure 11:715-23.

Figure legends

Figure 1: (top) The capsule locus of *S. pneumoniae* serotype 19A and the function of its genes in the production of the polysaccharide capsule (CPS). (bottom) Biosynthetic pathway of the serotype 19A capsular polysaccharide. Derived from Morona *et al.* 1999 (23).

Figure 2: Minimum spanning tree of an MLST analysis of the 19A serotype strains of our cohort (n = 338). MLST analysis was based on the sequence of the *cps* locus. Strains are grouped based on origin of samples: by time period (A) or type of host (B).

Figure 3: (A) Alignment of amino acid sequence from RmlB, RmlD and Wzg of the different serotype 19A cpsSNP variants. Showing substitutions in colors: (blue) nonpolar amino acids, (red) polar amino acids, (green) negatively charged amino acids and (purple) positively charged amino acids. All variants identified are shown. Structural model of an RmlB dimer (B) and RmlD monomeric protein (C). Differences of protein residues in the context of the RmlB and RmlD structures are shown in blue for the Hungary19A-6 reference protein sequence vs. SNP2. The dimeric RmlB model is shown with one domain in pale and the second domain color coded according to the sequence conservation.

Figure 4: Results of mass spectrometry quantification of whole cell monosaccharide and nucleotide sugar content in samples of 19A serotype isolates with a unique cpsSNP type (Top): quantities of the individual CPS monosaccharides or nucleotide sugars with SE-bars. The median differed significantly for all: UDP-Glc and UDP-GlcNAc ($p < 0.01$); UDP-ManNAc, TDP-Rha, TDP-Glc ($p < 0.05$). Significant differences between subtypes I vs. III (UDP-GlcNAc only), and II vs. III (UDP-GlcNAc and TDP-Rha) are shown in the graph (* = $p < 0.05$) (Kruskal-Wallis). (Bottom): Phylogenetic tree of a CPS locus MLST analysis with quantities of multiple monosaccharides and nucleotide sugars. From left to right: Phylogenetic tree, genomic background (MLST group), prevalence in cohort (n), 19A CPS locus subtype, levels of monosaccharides. Two phylogenetic groups can be distinguished: SNPs1-29, 49 and SNPs30-48. A significant difference in total amount of sugars between the two groups is shown (** = $p < 0.01$). Also, GDP-Glc ($p < 0.05$), CMP-Neu5Ac, UDP-Gal, UDP-Glc, UDP-GlcNAc, UDP-GalNAc, ManNAc, GDP-Fuc and TDP-Rha ($p < 0.01$) differed significantly (multiple Mann Whitney U tests). Data shown represents the results of three independent experiments.

663 **Figure 5:** Western blot analysis of the effect of Wzg sequence on capsule shedding.
664 Supernatants of growth cultures blotted on a membrane were stained using serum from
665 mice vaccinated with PCV13 (which includes serotype 19A). Shown intensities were
666 corrected for with a medium blank, and normalized per blot using purified 19A CPS. (A)
667 Mean intensity and SE of the different examined SNP types. Results of multiple independent
668 experiments (n=3) are shown. (B) Scatter blot of the different SNP types showing intensity
669 variation within the Wzg protein type. No significant differences were found (Kruskal-
670 Wallis).

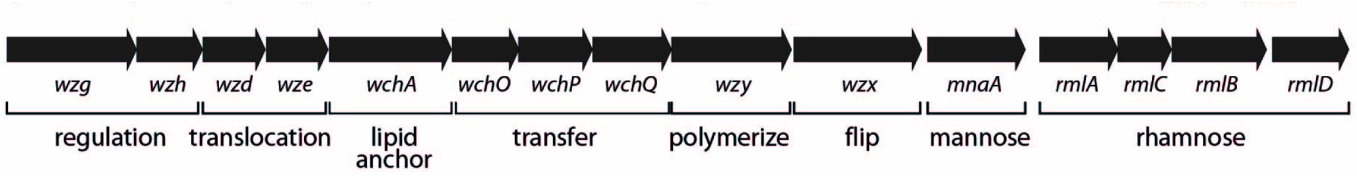
671

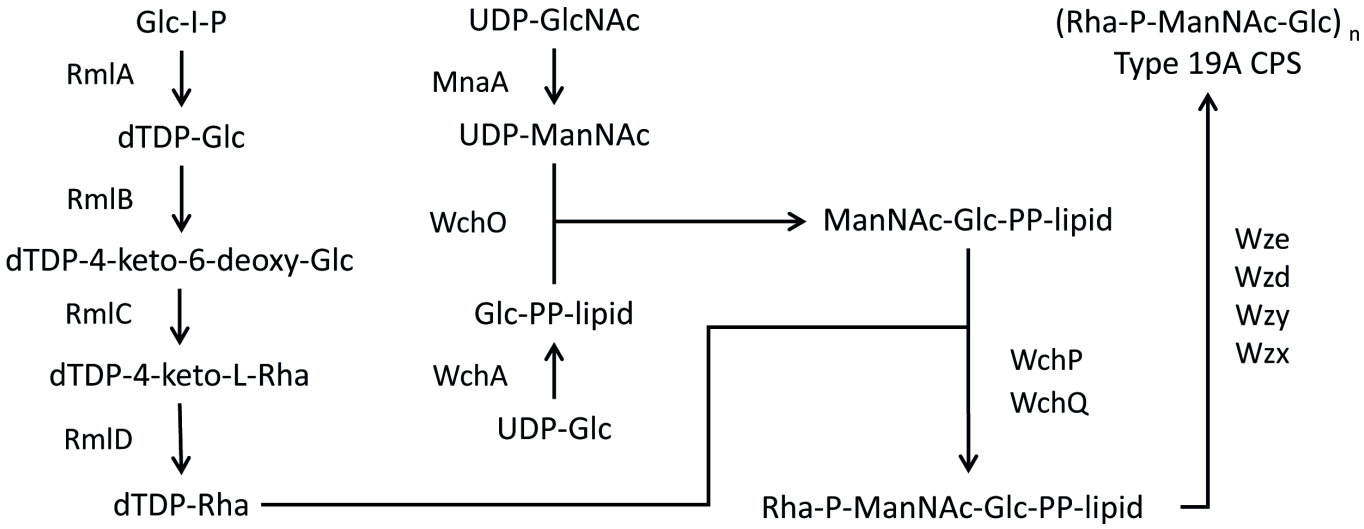
672 **Table footnotes**

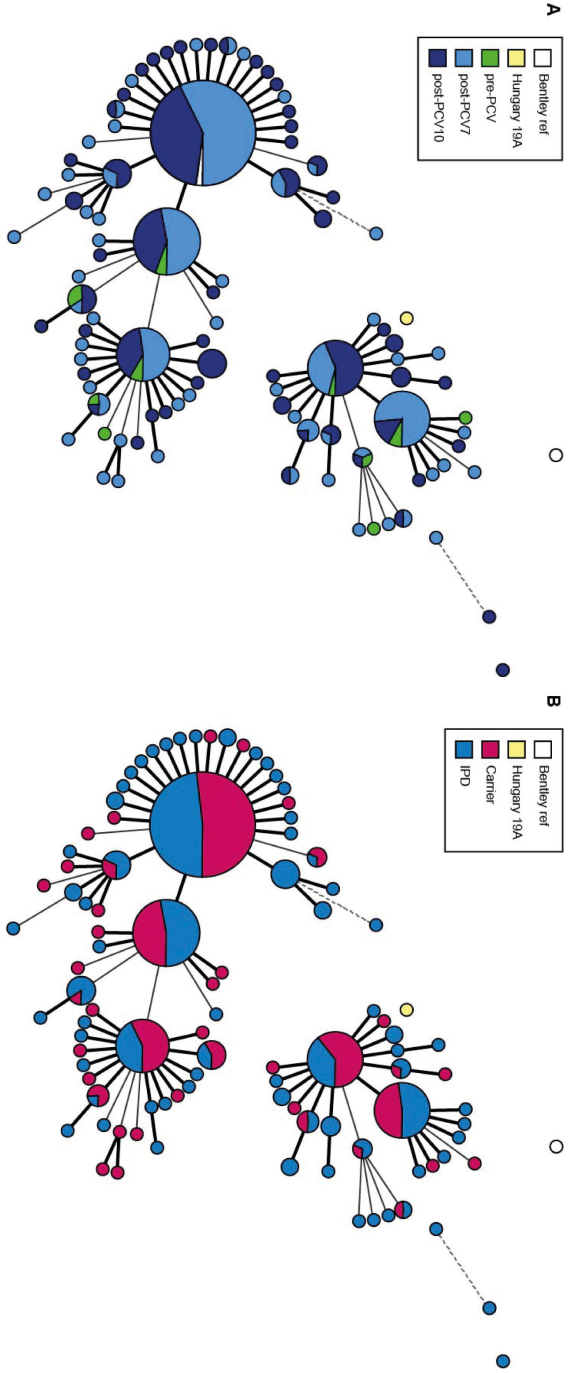
673 **Table 1:** *S. pneumoniae* isolates with a complete 19A CPS locus, included in this study.

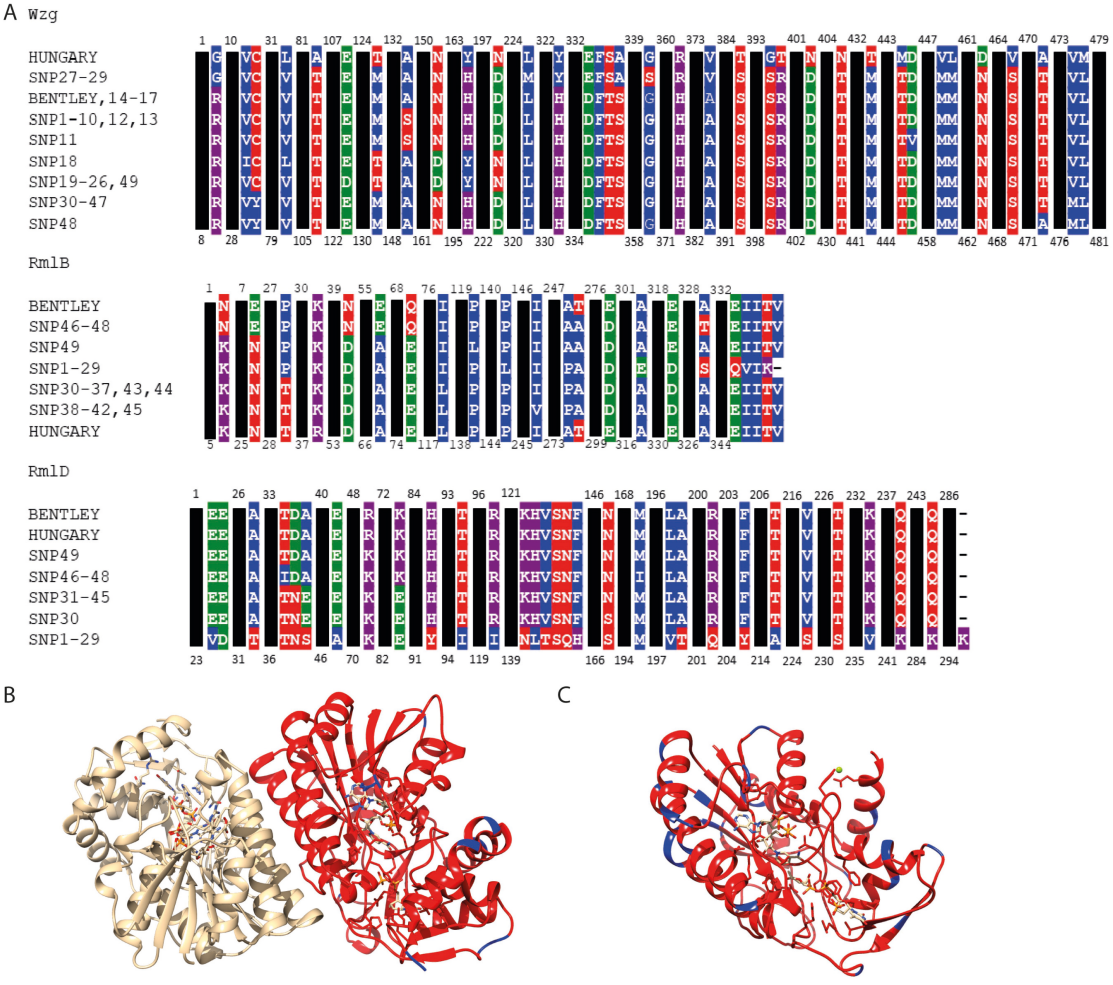
674 **Table 2:** Distribution of the observed SNP types across the 19A subtypes as described by
675 Elberse *et al.* 2011 (24).

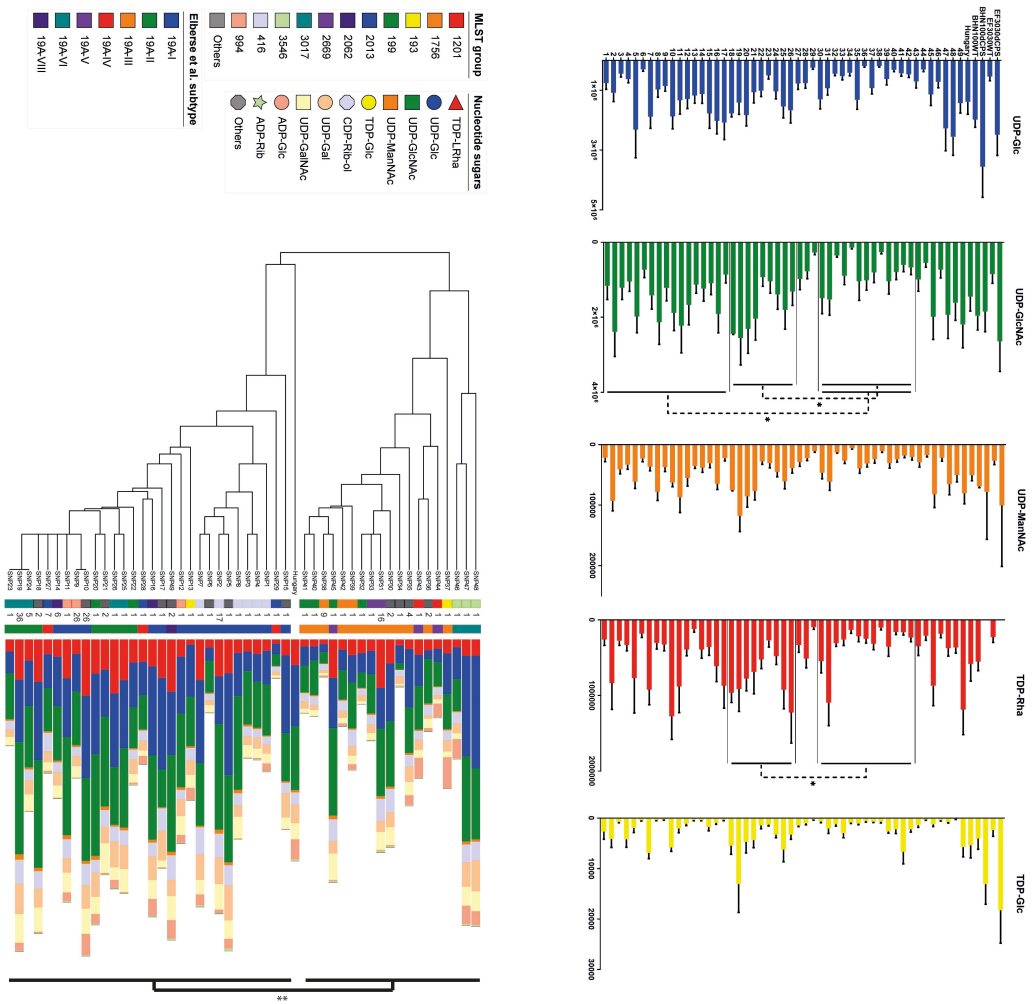
676 **Table 3:** Distribution of the observed SNP types across the Wzg types.











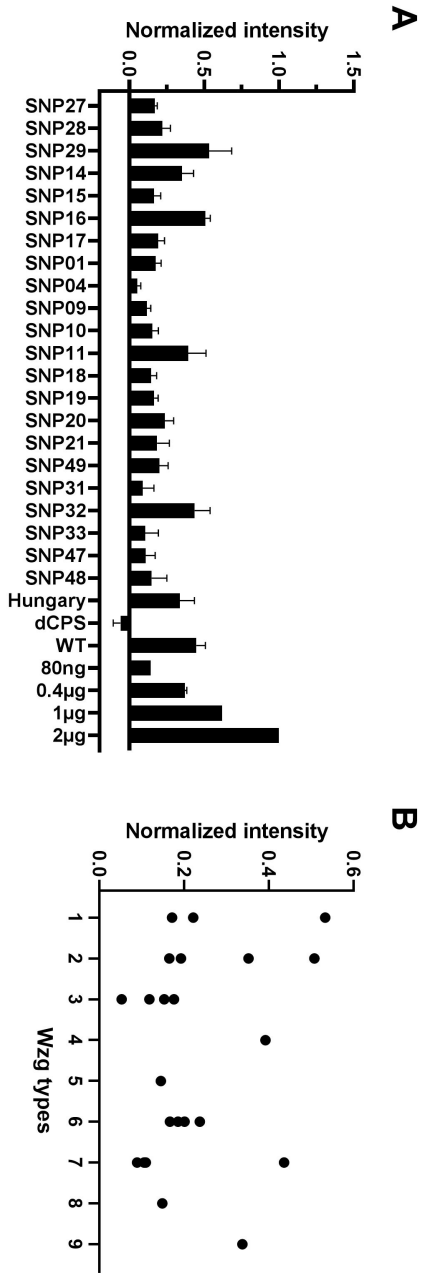


Table 1: *S. pneumoniae* isolates with complete 19A CPS locus, included in this study.

Origin	Total	Pre-PCV (2004-2006)	Post-PCV7 (2006-2011)	Post-PCV10 (2011-2016)
Invasive pneumococcal disease	188	16	55	117
Carrier	148	-	115	33
Reference	2	-	-	-

Table 2: Distribution of the observed SNP types across the 19A subtypes as described by Elberse *et al*. [11].

19A subtype	SNP types
I	SNP1-17
II	SNP18-23
II-ins	SNP24-26
III	SNP30-42
IV	SNP27-29
V	SNP43-45
VI	SNP46-48
VIII	SNP49

Table 3: Distribution of the observed SNP types across the Wzg types.

Wzg types	SNP no.
1	SNP27-29
2	SNP14-17
3	SNP1-10, 12, 13
4	SNP11
5	SNP18
6	SNP19-26, 49
7	SNP30-47
8	SNP48
9	Hungary