MAJOR ARTICLE

Evolution of the Capsular Regulatory Genes in *Streptococcus pneumoniae*

Sirkka-Liisa Varvio,[1,2] Kari Auranen,[1,2] Elja Arjas,[1,3] and P. Helena Mäkelä[2]

[1]Department of Mathematics and Statistics, University of Helsinki, and [2]Department of Vaccination and Immune Protection, [3]National Institute for Health and Welfare, Helsinki, Finland

The major pneumococcal virulence determinant is its capsule, and pneumococcal epidemiology is based on 91 capsular serotypes, each corresponding to the structure of the capsular polysaccharide determined by the type-specific capsular genome. Here, we provide the beginnings of an approach to intertwine serotype epidemiology, capsular regulatory gene characteristics on the basis of existing sequence information, and the reanalysis of published epidemiological data. We present an approach to explain epidemiological characteristics of serotypes on the basis of genetic differences in their capsular regulatory genes. The part of the capsular genome that regulates capsular expression falls into 2 highly divergent sequence clans: the ancestral pneumococcal capsular regulatory gene sequences (present in 49 serotypes) and laterally transferred sequences (present in 32 serotypes). Our survey of epidemiological data showed a tendency of the ancestral type of the capsular regulatory genome to be associated with carriage and the laterally transferred sequences to be associated with invasive disease isolates. The regulatory gene region shows mosaic structures that have signatures of recent recombination events, reminiscent of structures known from antibiotic resistance genes.

The pneumococcus (*Streptococcus pneumoniae*) is notorious for its success as a human parasite. Since its discovery in Pasteur's time 130 years ago, it has kept its position as a major pathogen and a commensal in the human nasopharyngeal microflora. At any one time, 30%–90% of healthy children <5 years old are colonized by pneumococci, and the annual child death toll due to pneumococcal invasive diseases is estimated to be at least 1 million [1]. The major virulence factor of the pneumococcus is its capsule, of which 91 antigenically distinct serotypes are known at present. Remarkably, many serotypes coexist in the pneumococcal population at any time and geographical location. Serotypes isolated from invasive diseases have a frequency distribution different from that for carriage isolates [2–4].

The capsule consists of a layer of capsular polysaccharide (CPS) that helps the bacteria escape from phagocytic killing and allows their survival in invasive pneumococcal disease (IPD) [5, 6]. This traditional view has recently been complemented by the realization that the capsule also plays important—although complex—roles during the early phase of colonization of the upper airways. There it is needed to facilitate the transit of the bacteria through the viscous nasal mucus to reach their site of sojourn in association with the endothelial cells of the nasopharyngeal mucosa [7]. At the same time, the shielding function of the capsule becomes an obstacle to the adherence of the bacteria to their cellular receptors, and "transparent" colony morphology variants with reduced amount of CPS adhere better than do "opaque" variants [6, 8, 9]. Obviously, careful regulation of CPS expression in response to changing environmental clues is a prerequisite for the pneumococcus in association with the human host.

The pneumococcus has highly developed transfor-

Received 4 February 2009; accepted 12 May 2009; electronically published 25 August 2009.

Potential conflicts of interest: none reported.

Financial support: Finnish Academy (grant 1211489 to E.A. [a part of the Center of Excellence of Population Genetic Analyses] and grant 210550 to K.A.). This work is also part of the research of the PneumoCarr consortium, which is funded by a grant from the Bill and Melinda Gates Foundation through the Grand Challenges in Global Health initiative.

Reprints or correspondence: Sirkka-Liisa Varvio, Dept of Mathematics and Statistics, PO 68, Gustav Hällströminkatu 2b, 00014-University of Helsinki, Finland (sirkka-liisa.varvio@helsinki.fi).

The Journal of Infectious Diseases 2009; 200:1144–51

0022-1899/2009/20007-0018$15.00
DOI: 10.1086/605651

mation mechanisms, and a typical feature of the pneumococcal genome is its plasticity, which facilitates its adaptation to changes in the environment [10, 11]. Uptake and integration of both homologous and foreign DNA has resulted in a large pangenome [12]. The development of resistance to multiple antimicrobial agents within a few decades of their extensive medical use is a widely cited example of pneumococcal evolution through lateral gene transfer (LGT) and multiple recombination events resulting in mosaic genes [13]. A new evolutionary challenge facing the pneumococcus is presented by the polysaccharide conjugate vaccines, which are highly immunogenic and effective against IPDs. These vaccines are, however, protective for only those serotypes that are included in the vaccine (at present 7 serotypes, with 10 and 13 serotypes in products in advanced stages of clinical development). Because of this, they act as a selective force that encourages the replacement of vaccine serotypes with others not in the vaccine [14].

The genes required for the pneumococcal capsular synthesis form a tightly linked cluster of 10–20 genes, of which the majority encode the serotype-specific capsular genome [15]. Capsular synthesis is based on the *wzy* polymerase-dependent pathway (with the exception of 2 serotypes, 3 and 37). The 4 genes located at the 5′ end of the capsular genome (*wzg*, *wzh*, *wzd*, *wze*, or *cpsABCD*), which are common to all serotypes, have been implicated in capsular regulation and posttranslational synthesis modulation [16, 17].

The full sequences of the capsular genomes of the 91 serotypes have been determined and published [18, 19]. However, no attempts to explain the epidemiological characteristics of serotypes on the basis of their genetic differences have been made as of yet. Here, we provide the beginnings of such an approach. We intertwine pneumococcal serotype epidemiology and capsular regulatory gene evolution on the basis of existing sequence information with the reanalysis of published epidemiological data.

METHODS

We exploited European Molecular Biology Laboratory/GenBank data accumulated from several studies of pneumococcal capsular genes and genomes and compared different sequences from the same serotypes to confirm that they were virtually identical. Streptococcal capsular regulatory sequences were retrieved by means of BLASTn and BLASTx, using published pneumococcal sequences for queries. The accession numbers for pneumococcal sequences are CR931632–CR931722, EF538714–EF538718, AF532632–AF532715, AY5085586–AY508641, AY621659, AY621660, AY661448–AY661457, Z83335, AF026471, AF316639, AE00562, AY336008, AF316641, AJ239004, AF402095, X85787, AF094575, U09239, AF030373, AF057294, AJ006986, and AF106132–AF106138. The complete capsular sequences have been determined for at least 2 independent isolates of serotypes 1, 2, 4, 5, 6A, 6B, 8, 9V, 14, 18C, 19F, 19A, 19B, 23F, and 33F, and partial regulatory gene sequences (the genes *wzg* and *wzh*) have been determined for 90 serotypes (6–10 isolates per serotype). The accession numbers for other streptococcal sequences that were used as references in phylogenetic analyses were as follows: for *Streptococcus oralis*, AB289547; for *Streptococcus mitis*, AB181235; for *Streptococcus gordonii*, AY147914; for *Streptococcus thermophilus*, AF448502, CP000023, CP000419, AF454499, AF454495, AF454500, AF454498, AF454496, AY057915, AF454501, and AF448502; for *Streptococcus suis*, AF118389; for *Streptococcus iniae*, AY904444; and for *Streptococcus agalactiae*, AB028896, AB050723, AY375362, AF163833, AF355776, AF349539, AF337958, AY376403, and AY375363.

Sequence alignments were performed using ClustalX [20] and were curated manually. Phylogenetic analyses were performed using MEGA (version 4) [21] and MrBayes (version 3.1) [22] software. Nucleotide diversities, percentage of guanine-cytosine (GC) content, and codon bias indices were calculated using DnaSP4 software [23].

RESULTS

Formation of 2 major clans by the capsular regulatory genes. The first capsular regulatory gene, *wzg*, of *S. pneumoniae* comprised 1 cluster of related sequences. In contrast, the 3 subsequent genes, *wzh*, *wzd*, and *wze*, exhibited deeply dichotomous phylogenetic patterns, with 2 distinct and statistically highly supported sequence clusters (figure 1*A*). For convenience, we call these sequence clusters the red and the blue clans. The same red-blue split applied to each of the 3 genes and extended further to *wchA* (the first type-specific gene), which codes for the initial glycosyl transferase in 66 serotypes but not beyond it (figure 1*A*). Many serogroups (clusters of antigenically related serotypes) were red-blue dichotomized (figure 1*B*). Three serotypes (25F, 25A, and 38) were exceptional with respect to both the location [18] and sequence of the regulatory genes, indicating their derivation from a nonstreptococcal gene pool (figure 1*C*).

The ancestral blues and laterally transferred reds. The analysis of the DNA sequence motifs revealed a structure consistent with recombination events that have replaced blocks of blue-type sequences with the corresponding sequences of the red type (figure 2). We surmise that in this recombination process the blue sequence represents the ancestral form of the capsular regulatory genome, because its percentage of GC content corresponds to the average 39.7% for *S. pneumoniae* [24] and to the 38% for the *wzg* gene. The red-sequence blocks have a percentage of GC content ~5 percentage points lower, and we thus conclude that they were derived by LGT from a different gene pool.

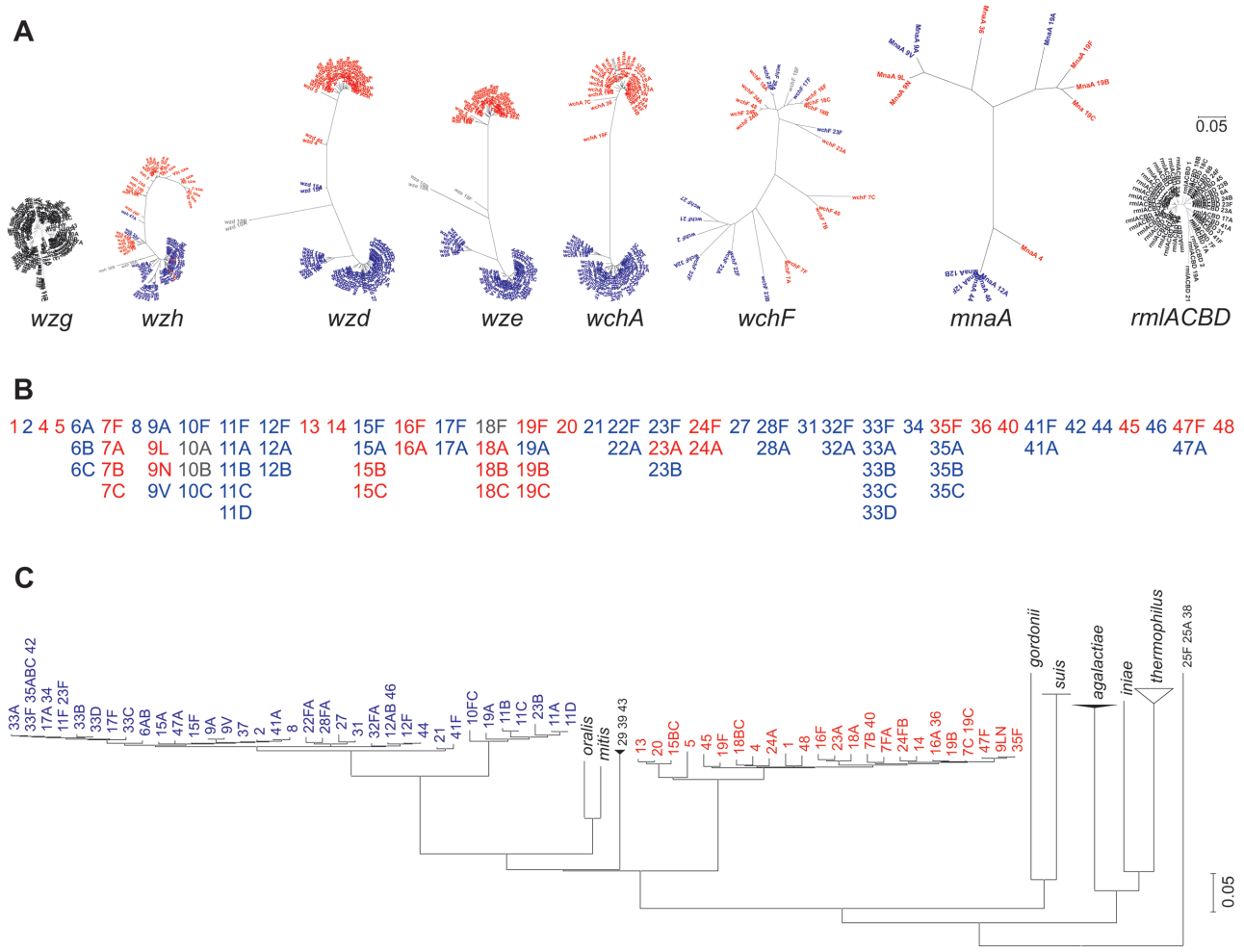

Figure 1. Phylogenetic trees (minimum-evolution, neighbor-joining phylograms) of pneumococcal capsular genes and red and blue serotype characteristics. *A*, Four core capsular genes (*wzg, wzh, wzd,* and *wze*), including 2 examples of serotype-specific glycosyltransferase genes (*wchA* and *wchF*) and 2 other type-specific genes (*mnaA* and *rmlACBD*) (this analysis includes 4 concatenated *rml* genes). The red and blue coloring is also used in the *mnaA* and *wchF* genes to show that, in these cases, the clustering pattern differs from the red-blue classification, defined through capsular regulatory genes. *B*, Pneumococcal serotypes categorized as blue and red on the basis of their capsular regulatory genes. Serotypes shown in gray have an ambiguous classification (figure 2). *C*, Phylogram of the *wze* sequences (excluding the ambiguous serotypes 10A, 10B, and 18F) as a representation of patterns on the species level. All 8 *Streptococcus* species that (on the basis of sequence databases) have capsular synthesis through *wzg, wzh, wzd,* and *wze* are included here, and they represent 3 of the 6 official taxonomic groups. *S. mitis, S. oralis, S. gordonii*, and *S. pneumoniae* are from the mitis group, *S. thermophilus* is from the salivarius group, *S. suis* has equivocal taxonomic grouping, and *S. agalactiae* (group B streptococci) and *S. iniae* are from the pyogenic group. Bayesian and minimum-evolution phylogenies have identical topologies, the former with 1.0 credibility for all species-level branches and the latter with 0.8–1.0 bootsrapping confidence (10,000 iterations).

Several red integration points were identified in the *wzh* gene sequence (figure 2), with the predominant point near the 5′ end of the gene. Some serotypes showed red-blue mosaic structures, 2 serotypes (4 and 45) had a red integration point at *wzd*, and 1 serotype (18F) had a red integration point at the 3′ end of *wze*. In 3 serotypes (29, 39, and 43), a short red integration motif at *wzh* was accompanied by a nonred (and nonblue) streptococcal sequence type. In 1 serotype (20), the red sequence spanning from *wzh* to *wze* was continued by a nonred (and nonblue) sequence at *wchA*. In serotypes 10A and 10B, the *wzd* and *wze* genes were mosaics between blue and *S. mitis*– and/or *S. oralis*–like sequences.

Differential association of the reds and blues with carriage and IPD. The reds, even if newcomers in the pneumococcal genome, have an established standing within the species. Of the 91 serotypes, 32 had a red and 49 a blue capsular regulatory genome type (figures 1*B* and 2). We reviewed a large amount of epidemiological data on the distribution of serotypes (see Hausdorff et al [2] for a recent review of serotype distributions) among healthy carriers and among individuals with IPD. We

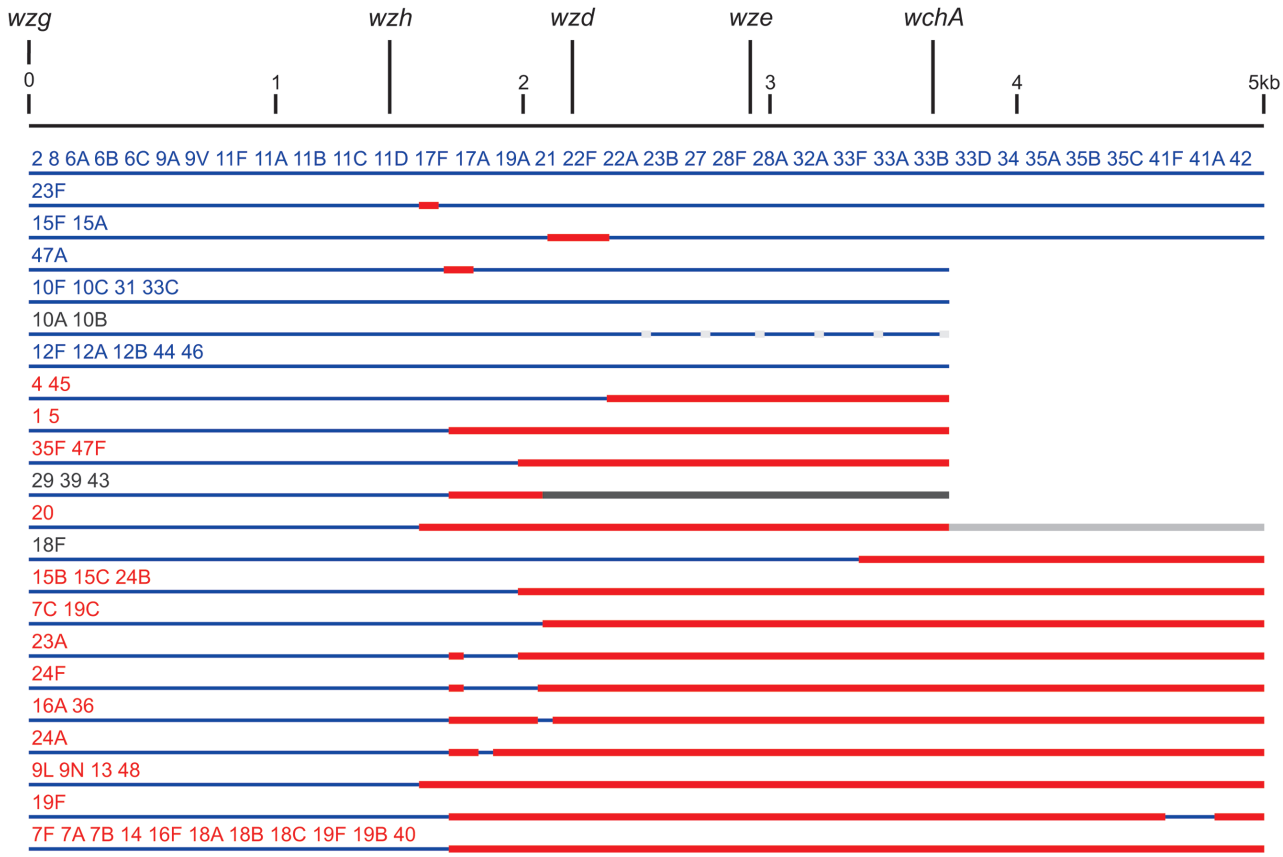

Figure 2. Recombination diagram of the pneumococcal blue and red sequence types at capsular regulatory genes and at *wchA* (the streptococcal initial transferase). Between the most common blue (*top line*) and red (*bottom line*) configurations, several variations exist: different red recombination points, short recombinant stretches (note that serotype 19F is present in 2 versions), and the presence or absence of *wchA*. Sequence motifs—and consequently also serotypes—that do not obey the red-blue classification are depicted in gray or black.

assigned each serotype to the blue or red category according to figure 1*B*. The data from large studies (at least 150 isolates) are shown in table 1. Despite the heterogeneity of these data sets, there was a consistent predominance of the blues among isolates from healthy carriers (two-thirds blue) and, correspondingly, a predominance of the reds among isolates from those with IPD (two-thirds red). Remarkably, these proportions held for different populations and continents and were thus not confined to the prevalence of some particular serotypes. The same applied to other studies not fulfilling the criteria for inclusion in table 1.

One of the data sets [3] included both invasive and carriage isolates from the same population and was thus suitable for calculation of an empirical odds ratio to compare the probability of invasive disease due to the red propensity. An odds ratio of 3.7 was found, with a 95% confidence interval of 2.4–5.5 ([*ad*]/[*bc*], where *a* is the number of invasive red, *b* is the number of carried red, *c* is the number of invasive blue, and *d* is the number of carried blue).

In all, the traditional naming of certain serotypes as "primary pathogens" or "primarily invasive" [3, 4] gets one putative explanation through the capsular regulatory genes. Serotypes 1, 5, and 7F—all red—are very seldom found carried. Of the common serotypes (~10) named "carried," only 19F is red; all others are blue. Note that of the 91 serotypes that have been characterized as different types (on the basis of antigenic reactions), only ~30 actually circulate. Characterization of the occasionally found (ie, rare) serotypes is not possible.

DISCUSSION

Our work is intended as an initiative to develop an understanding of the pneumococcal capsular evolution, previously regarded as elusive [18, 37]. We show that the 3 shared regulatory genes split most serotypes into 2 highly divergent clans, the red clan and the blue clan. There is no obvious pattern of preferential association between either regulatory type and any identifiable structural feature of the type-specific part of the CPS. Specifically, an examination of serogroups (groups of antigenically and structurally related serotypes) demonstrates

Table 1. Classification of Serotypes as Red or Blue on the Basis of Their Capsular Regulatory Genes

Isolate type, country, age group	Serotype regulatory genes, %					Serotyped isolates, no.	Reference
	Blue	Red	3	25F, 25A, 38	Other		
Carriage isolates							
England, 0–5 years	58.0	35.8	5.1	1.1	…	350	[3]
Sweden, 1–6 years	58.4	37.3	1.6	1.6	1.2	246	[4]
Netherlands, 1–19 years	58.1	35.9	6.0	…	…	472	[25]
Canada, preschool age	68.6	25.4	1.0	1.5	3.5	518	[26]
Kenya, <5 years	58.9	37.1	1.0	…	2.4	207	[27]
Gambia, <18 months	52.2	31.1	2.8	…	14.0	237	[28]
Invasive isolates							
England, 0–5 years	31.6	67.0	1.3	0.1	…	150	[3]
Sweden, 1–6 years	40.3	52.7	5.1	1.1	0.7	246	[4]
Norway, all ages	34.8	58.2	2.2	…	4.9	325	[29]
Germany, adults	36.5	52.8	9.7	1.0	…	647	[30]
Denmark							[31]
<2 years	41.8	55.8	0.7	1.7	0.7	297	
2–59 years	32.9	63.3	3.8	…	…	1820	
>60 years	37.7	54.1	6.4	1.8	…	3048	
Canada, <5 years	38.0	55.0	…	…	7.0	303	[32]
Latin America, <6 years							[33]
Argentina	18.4	65.9	0.7	…	15.0	1006	
Brazil	28.4	57.2	2.0	…	12.4	1203	
Colombia	29.0	58.8	1.0	…	11.2	623	
Uruguay	15.1	73.1	4.5	…	7.3	352	
Taiwan, all ages	33.0	57.0	…	…	10.0	483	[34]
Australia, <15 years	28.9	64.5	1.5	…	5.0	698	[35]
United States, <5 years	39.8	58.9	0.7	0.6	…	855	[36]

NOTE. Serotype data were extracted from tables and/or figures in published epidemiological studies. Data on carriage isolates are from nasopharyngeal swabs obtained from healthy individuals. Data on invasive isolates are from normally sterile body sites, blood, cerebrospinal fluid, and joint, pleural, and peritoneal fluid. Studies including at least 150 serotyped isolates and in which at least 85% of the serotypes were reported in detail are included here. The category "other" includes pooled rare serotypes and isolates that were not typed or were not typeable. Serotype 3 (synthesis through a different pathway) and serotypes 25F, 25A, and 38 (not blue or red) are listed separately.

that both regulatory types can be found in association with closely related capsules (eg, serotypes 9A and 9V in serogroup 9 have blue regulatory genes, whereas types 9L and 9N have red regulatory genes) (figure 1*B*).

Throughout the regulatory gene region and at *wchA*, the blue clan corresponds to the typical pneumococcal genomic content in its percentage of GC. Its sequence diversity is higher than that in the red clan, suggesting that the blues form an older clan. The reds have a percentage of GC content ~5 percentage points lower than that of the blues, consistent with an LGT origin. On the basis of existing sequence information from other species, which have homologues of the pneumococcal capsular genes, the source species of the reds remains unidentified. Phylogenetic inference suggests that the source is within the mitis group of streptococci (figure 1*C*); this group comprises 15 species [38], of which capsular gene sequence information exists only for *S. pneumoniae*, *S. mitis*, *S. oralis*, and *S. gordonii*. It is well known that *S. mitis* and *S. oralis* are LGT partners with *S. pneumoniae* at antibiotic resistance genes (especially *pbp2x*) [13, 39]. They are probably not the source species of the reds (figure 1*C*), because their capsular regulatory gene percentage of GC content is similar to that of the blues and of *S. pneumoniae*. However, *S. mitis* is known to be extremely variable genetically [13, 40, 41]. Capsular sequence information is available only from a few *S. mitis* strains, and thus the possibility that *S. mitis* is the transmitter of the reds cannot be ruled out.

An inspection of the current phylogenetic pattern suggests that the diversification of the reds, as well as that of the blues, are (or have been) concurrent processes. Both clans include major clusters of approximately equidistant sequences (figure 1*C*). The patchiness of red-blue recombination events (figure 2) suggests that new combinations can be expected to emerge. The limited current sequence data already shows the presence

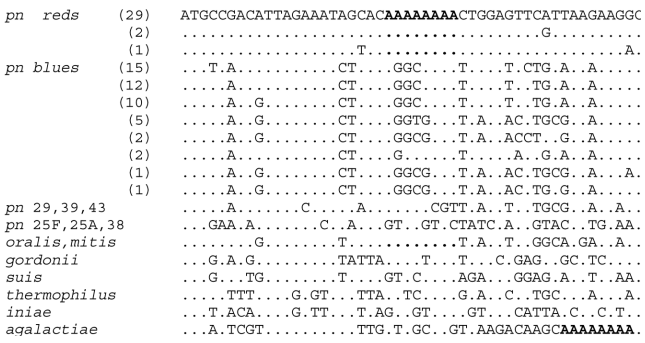

Figure 3. The 5′ sequence motif of the *wze* gene in the pneumococcus (*pn*) and 7 other *Streptococcus* species. The number of pneumococcal serotypes having a given sequence is shown in brackets.

of 2 forms in 1 serotype (19F) (figure 2). That many serotypes do have completely or nearly identical blue or red regulatory gene sequences (figure 1*C*) speaks for very recent recombination events in their history.

Our epidemiologic survey showed a tendency of the blue type of the capsular regulatory genome to be associated with carriage and the red with invasive disease isolates. This conclusion rests on the assumption of a tight association of the assigned (blue or red) regulatory and the type-specific capsular genomes. This assumption is warranted on the basis of the current sequence information for the capsular genomes (even if limited), because the sequences from different isolates of a given serotype have proved to be almost identical (with the exception of serotype 19F; see above). The existence of highly divergent sequences among capsular regulatory genes has been reported from a small set of serotypes [42, 43] but were not considered further (and were not even mentioned in publications citing Morona et al [42] or Jiang et al [43]). A clue about the red-blue grouping has thus existed but has been ignored.

Pneumococci employing the red type of capsular regulation seem better fit to cause IPD. In traditional parlance, they are more virulent for this outcome than the blues. A thick layer of capsule is associated with protection of the bacteria from phagocytosis, the critical host defense in IPD [6]. Accordingly, if the red type of capsular regulation would lead to an increase in CPS production compared with the blue type, this could provide the reds with a pivotal advantage in causing IPD and account for their larger share among IPD isolates. The known and described regulation of the amount of CPS synthesis is a posttranslational tyrosine phosphoregulatory system, the most essential component of which is the regulatory gene *wze* [18, 19]—that is, the gene that unequivocally defines the blues and the reds. Whether the red type of regulation would mean increased expression of the red genes remains to be shown. The reds might be more efficiently translated because their codon bias is higher (on average, 0.5 in the reds and 0.4 in the blues at *wze*). Interestingly, in one study [44] expression of *wzg* in serotypes causing invasive infection (most of them red according to our classification) was reported to be 2-fold higher than serotypes causing nasopharyngeal colonization (most of them blue). In studies comparing gene expression profiles of nasopharyngeal isolates with those of invasive isolates, significant differences in capsular gene expression have not, however, been found (eg, in Oggioni et al [45]).

At the 5′ end of *wze*, an A_8 motif is present in all the reds but is not present in any of the blues (figure 3). This motif has been identified as one of a set of putative virulence factors in the full genome of serotype 4 [24]. Simple sequence repeats, which are prone to strand slippages, are known to lead to phase variation in bacteria (gene transcription on or off), including the pneumococcus [46, 47]. Although the described regulatory mechanism of capsular genes is posttranslational, the presence of the A_8 motif in reds suggests that it could also induce phase variation here and thus provide the reds with an additional capability by acting as a transcription regulatory mechanism.

Only a small minority of the total population of pneumococci causes IPD. The majority is represented by the bacteria colonizing the upper respiratory mucosa of healthy carriers. These isolates are crucial for the survival and spread of the pneumococcus. Because of the complex functions of the capsule in this process, our present knowledge is not sufficient to suggest which type of capsular regulation provides an advantage in colonization, although the epidemiological data suggest that it is the blue type of regulation.

The pneumococcus is currently exposed to a profound change in environment, which will provide pressures for new responses. The powerful new conjugate vaccines selectively disadvantage the serotypes represented in the vaccine. Their effect is seen in the pneumococcal population in the nasopharynx of carriers as a decrease in the frequency of vaccine serotypes and a concomitant increase in nonvaccine serotypes [48]. Among the serotypes replacing the vaccine types, the so-called vaccine escape recombinants are a concrete example of evolution through recombination having replaced the original capsular genome coding for the vaccine serotype 4 with the capsular genome coding for the unrelated nonvaccine serotype 19A [49]. Sequence mapping of the capsular regulatory genome in further vaccine escape recombinants is an urgent need, to give us a better understanding of pneumococcal evolution under the current vaccination pressure. To date, the recombination break points have been identified upstream and downstream of the capsular genome; however, as yet no sequence information on the capsular genomes exists.

Acknowledgments

We thank Bob O'Hara for checking the language and Joose Väinölä for help with artwork.

References

1. World Health Organization. Pneumococcal vaccine for childhood immunization—WHO position paper. Wkly Epidemiol Rec **2007**; 82: 93–104.
2. Hausdorff WP, Brueggemann AB, Hackell JG, Scott AG. Pneumococcal serotype epidemiology. In: Silver GR, Klugman KP, Mäkelä PH, eds. Pneumococcal vaccines: the impact of cojugate vaccine. Washington, DC: ASM Press, **2008**:139–58.
3. Brueggemann AB, Griffiths DT, Meats E, et al. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J Infect Dis **2003**; 187:1424–32.
4. Sandgren A, Sjöström K, Olsson-Liljequist B, et al. Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae.* J Infect Dis **2004**; 189:785–96.
5. Weiser JN, Austrian R, Sreenivasan PK, Masure HR. Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. Infect Immun **1994**; 62:2582–9.
6. Nelson AL, Roche AM, Gould JM, et al. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. Infect Immun **2007**; 75:83–90.
7. Weiser JN, Bae D, Epino H, et al. Changes in the availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of *Streptococcus pneumoniae.* Infect Immun **2001**; 69:5430–9.
8. Magee AD, Yother J. Requirement for capsule in colonization by *Streptococcus pneumoniae.* Infect Immun **2001**; 69:3755–61.
9. Cundell DR, Weiser JN, Shen J, Young A, Tuomanen EI. Relationship between colonial morphology and adherence of *Streptococcus pneumoniae.* Infect Immun **1995**; 63:757–61.
10. Claverys JP, Prudhomme M, Mortier-Barriere I, Martin B. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity. Mol Microbiol **2000**; 35:251–9.
11. Mortier-Barriere I, Velten M, Dupaigne P, et al. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell **2007**; 130:824–36.
12. Hiller NL, Janto B, Hogg JS, et al. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. J Bacteriol **2007**; 189:8186–95.
13. Bruckner R, Nuhn M, Reichmann P, Weber B, Hakenbeck R. Mosaic genes and mosaic chromosomes—genomic variation in *Streptococcus pneumoniae.* Int J Med Microbiol **2004**; 294:157–68.
14. Whitney GC, Moore MR. Direct and indirect effectiveness and safety of pneumococcal conjugate vaccine in practice. In: Silver GR, Klugman KP, Mäkelä PH, eds. Pneumococcal vaccines: the impact of conjugate vaccine. Washington, DC: ASM Press, **2008**:353–68.
15. Yother J, Bentley SD, Hennessey JP Jr. Genetics, biosynthesis, and chemistry of pneumococcal capsular polysaccharides In: Silver GR, Klugman KP, Mäkelä PH, eds. Pneumococcal vaccines: the impact of cojugate vaccine. Washington, DC: ASM Press, **2008**:33–46.
16. Morona JK, Morona R, Miller DC, Paton JC. Mutational analysis of the carboxy-terminal $(YGX)_4$ repeat domain of CpsD, an autophosphorylating tyrosine kinase required for capsule biosynthesis in *Streptococcus pneumoniae.* J Bacteriol **2003**; 185:3009–19.
17. Morona JK, Morona R, Paton JC. Attachement of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. Proc Natl Acad Sci U S A **2006**; 103:8505–10.
18. Bentley SD, Aanensen DM, Mavroidi A, et al.Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet **2006**; 2:e31.
19. Park IH, Park S, Hollingshead SK, Nahm MH. Genetic basis for the new pneumococcal serotype, 6C. Infect Immun **2007**; 75:4482–9.
20. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **1997**; 25: 4876–82.
21. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **2007**; 24:1596–9.
22. Rohnquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **2003**; 19:1572–4.
23. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **2003**; 19:2496–7.
24. Tettelin H, Nelson KE, Paulsen IT, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae.* Science **2001**; 293: 498–506.
25. Bogaert D, Sluijter M, Toom NL, et al. Dynamics of pneumococcal colonization in healthy Dutch children. Microbiology **2006**; 152:377–85.
26. Kellner JD, Ford-Jones EL. *Streptococcus pneumoniae* carriage in children attending 59 Canadian child care centers. Arch Pediatr Adolesc Med **1999**; 153:495–502.
27. Abdullahi O, Nyiro J, Lewa P, Slack M, Scott AG. The descriptive epidemiology of *Streptococcus pneumoniae* and *Haemophilus influenzae* nasopharyngeal carriage in children and adults in Kilifi district, Kenya. Pediatr Infect Dis J **2008**; 27:59–64.
28. Hill PC, Cheung YB, Akisanya A, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a longitudinal study. Clin Infect Dis **2008**; 46:807–14.
29. Pedersen MK, Høiby EA, Frøholm LO, et al. Systemic pneumococcal disease in Norway 1995–2001: capsular serotypes and antimicrobial resistance. Epidemiol Infect **2004**; 132:167–75.
30. Reinert RR, Haupts S, van der Linden M, et al. Invasive pneumococcal disease in adults in North-Rhine Westphalia, Germany, 2001–2003. Clin Microbiol Infect **2005**; 11:985–91.
31. Konradsen HB, Kaltoft MS. Invasive pneumococcal infections in Denmark from 1995–1999: epidemiology, serotypes and resistence. Clin Diagn Lab Immunol **2002**; 9:358–65.
32. Lovgren M, Spika JS, Talbot JA. Invasive *Streptococcus pneumoniae* infections: serotype distribution and antimicrobial resistence in Canada, 1992–1995. CMAJ **1998**; 158:327–31.
33. di Fabio JL, Castañeda E, Agudelo CI, et al. Evolution of *Streptococcus pneumoniae* serotypes and penicillin susceptibility in Latin America, Sireva-Vigia Group, 1993 to 1999. Pediatr Infect Dis J **2001**; 20:959–67.
34. Chen YY, Yao SM, Chou CY, et al. Surveillance of invasive *Streptococcus pneumoniae* in Taiwan, 2002–2003. J Med Microbiol **2006**; 55:1109–14.
35. Watson M, Brett M, Brown M, Stewart MG, Warren S. Pneumococci responsible for invasive disease and discharging ears in children in Sydney, Australia. J Med Microbiol **2007**; 56:819–23.
36. Gertz RE, McEllistrem MC, Boxrud DJ, et al. Clonal distribution of invasive pneumococcal isolates from children and selected adults in the United States prior to 7-valent conjugate vaccine introduction. J Clin Microbiol **2003**; 41:4194–216.
37. Mavroidi A, Aanensen DM, Godoy D, et al. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol **2007**; 189:7841–55.
38. Köhler W. The present state of species within the genera *Streptococcus* and *Enterococcus.* Int J Med Microbiol **2007**; 297:133–50.
39. Chi F, Nolte O, Bergmann C, Ip M, Hakenbeck R. Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in *Streptococcus pneumoniae, S. mitis* and *S. oralis.* Int J Med Microbiol **2007**; 297: 503–12.
40. Whatmore AM, Efstratiou A, Pickerill AP, Broughton K, Woodard G. Genetic relationships between clinical isolates of *Streptococcus pneumoniae, Streptococcus oralis,* and *Streptococcus mitis* harboring *S. pneumoniae* virulence factor-encoding genes. Infect Immun **2000**; 68:1374–82.
41. Kilian M, Poulsen K, Blomqvist T, et al. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. PLoS One **2008**; 3:e2683.
42. Morona JK, Morona R, Paton JC. Analysis of the 5′ portion of the type 19A capsule locus identifies two classes of cpsC, cpsD and cpsE genes in *Streptococcus pneumoniae.* J Bacteriol **1999**; 181:3599–605.
43. Jiang SM, Wang L, Reeves PM. Molecular characterization of *Strep-*

tococcus pneumoniae type 4, 6B, 8 and 18C capsular polysaccharide gene clusters. Infect Immun **2001**; 69:1244–55.

44. Hathaway LJ, Bättig P, Muhlemann K. In vitro expression of the first capsule gene of *Streptococcus pneumoniae, cpsA* is associated with serotype-specific colonization prevalence and invasiveness. Microbiology **2007**; 153:2465–71.
45. Oggioni MR, Trappetti C, Kadioglu A, et al. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. Mol Microbiol **2006**; 61:1196–210.
46. van Selm S, van Cann LM, Kolkman MA, van der Zeijst BA, van Putten JP. Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. Infect Immun **2003**; 71:6192–8.
47. Pericone CD, Bae D, Shchepetov M, McCool T, Weiser JN. Short-sequence tandem and nontandem DNA repeats and endogeneous hydrogen peroxide production contribute to genetic instability of *Streptococcus pneumoniae.* J Bacteriol **2002**; 184:4392–9.
48. O'Brien KL, Dagan R, Mäkelä PH. Nasopharyngeal carriage. In: Silver GR, Klugman KP, Mäkelä PH, eds. Pneumococcal vaccines: the impact of cojugate vaccine. Washington, DC: ASM Press, **2008**:279–300.
49. Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog **2007**; 3:e168.