MAJOR ARTICLE

Streptococcus pneumoniae Contains 3 *rlrA* Pilus Variants That Are Clonally Related

Monica Moschioni,[1] Claudio Donati,[1] Alessandro Muzzi,[1] Vega Masignani,[1] Stefano Censini,[1] William P. Hanage,[2] Cynthia J. Bishop,[2] Joice Neves Reis,[3] Staffan Normark,[4] Birgitta Henriques-Normark,[4] Antonello Covacci,[1] Rino Rappuoli,[1] and Michèle A. Barocchi[1]

[1]Novartis Vaccines, Siena, Italy; [2]Imperial College, Department of Infectious Disease Epidemiology, London, England, United Kingdom; [3]School of Pharmacy, Federal University of Bahia and Goncalo Moniz Research Center, Oswaldo Cruz Foundation, Salvador, Brazil; [4]Karolinska Institute, Stockholm, Sweden

Background. Pilus components of *Streptococcus pneumoniae* encoded by *rlrA* were recently shown to elicit protection in an animal model of infection. Limited data are available on the prevalence of the *rlrA* operon in pneumococci; therefore, we investigated its distribution and its antigenic variation among disease-causing strains.

Methods. The prevalence of *rlrA* and its association with serotype and genotype were evaluated in a global panel of 424 pneumococci isolates (including the 26 drug-resistant clones described by the Pneumococcal Molecular Epidemiology Network).

Results. The *rlrA* islet was found in 130 isolates (30.6%) of the defined collection. Sequence alignment of 15 *rlrA* islets defined the presence of 3 clade types, with an overall homology of 88%–92%. The presence or absence of a pilus-encoding operon correlated with *S. pneumoniae* genotype ($P < .001$), as determined by multilocus sequence typing, and not with serotype. Further investigation identified a positive trend of *rlrA* occurrence among antimicrobial-resistant pneumococci.

Conclusions. On the basis of *S. pneumoniae* genotype, it is possible to predict the incidence of the *rlrA* pilus operon in a collection of pneumococcal isolates. This will facilitate the development of a protein vaccine.

Streptococcus pneumoniae is a commensal pathogen and often part of the normal flora of the human respiratory tract. The most common pneumococcal infections are otitis media and pneumonia, although *S. pneumoniae* may less frequently cause invasive diseases such as meningitis and sepsis [1]. Prophylactic vaccination with capsular polysaccharide-based vaccines and antibiotic therapy are 2 effective approaches that decrease the risk of disease. However, selective pressures are affecting *S. pneumoniae* population dynamics. In this respect, the treatment and prevention of pneumococcal infections has been complicated by the spread of clones that have increase in the incidence of isolates resistant to antibiotics [2–5], and, following the introduction of the PCV7 glycoconjugate vaccine (Prevnar), the spread of nonvaccine serotypes (NVT) [6, 7]. This evidence provides the impetus to develop vaccines based on protein antigens, which are capable of serotype-independent protection [8, 9].

Recently, in a mouse model of infection, *S. pneumoniae* was shown to express a pilus that has subunits able to elicit protection both by active and passive immunization [10–12]. These structures are encoded by a chromosomal element defined as the *rlrA* pathogenicity islet [13]. The function of pneumococcal pili is currently an area of investigation. To date, pili are known to be involved in adhesion to lung epithelial cells in vitro, as well as in colonization in a murine model of infection [10, 14]. Sequence analysis of available *S. pneumoniae* genomes revealed that the serotype 6B strain 670 contains a region of DNA that shares 89% sequence identity with the *rlrA* islet from TIGR4. However, the locus is absent in strains R6 (a nonencapsulated strain derived from serotype 2 D39), G54 (serotype 19F), and Spain-23F-1 (serotype 23F), suggesting that the *rlrA* islet is not widespread in pneumococcal isolates.

Received 20 June 2007; accepted 7 September 2007; electronically published 12 February 2008.

Presented in part: "Novel Opportunities to Develop Vaccines to Control Antibiotic Resistant Bacteria: From the Trials Back to the Laboratory" (REBAVAC) Workshop, March 2007, Siena, Italy (abstract A7).

Potential conflicts of interest: M.M., C.D., A.M., V.M., S.C., A.C., R.R., and M.A.B. are employed by Novartis Vaccines. All other authors report no potential conflicts.

Reprints or correspondence: Michèle Anne Barocchi, Novartis Vaccines, Via Fiorentina 1 53100 Siena, Italy (Michele.Barocchi@Novartis.com).

The Journal of Infectious Diseases 2008; 197:888–96

0022-1899/2008/19706-0016$15.00
DOI: 10.1086/528375

Table 1. Sources and composition of the strain collection.

Source	Country	Isolates, no.	Invasive disease or carriage	Reference
PMEN	Various	26	ID	[15]
CDC	USA	53	ID	[16]
Oswaldo Cruz Foundation	Brazil	74	ID (32 isolates)/C (42 isolates)	. . .
Imperial College	USA	15	C	. . .
Imperial College	Finland	11	ID (9 isolates)/OM (2 isolates)	. . .
ISS	Italy	40	ID (37 isolates)/C (3 isolates)	. . .
Swiss Tropical Institute	Ghana	15	ID	[17]
Dhaka Shishu Hospital	Bangladesh	14	ID	. . .
Kenyan Medical Research Center	Kenya	12	ID	. . .
Karolinska Institute	Sweden	76	ID	. . .
Novartis	Various	11	ID	. . .
University of Glasgow[a]	Scotland	74	ID	[18]
Sanger Institute (http://www.sanger.ac.uk)	Various	3	ID	. . .

NOTE. C, carriage; CDC, Centers for Disease Control and Prevention; ID, invasive disease; ISS, Istituto Superiore di Sanità; OM, otitis media; PMEN, Pneumococcal Molecular Epidemiology Network.

[a] Results for these strains are derived from previously published work.

Therefore, we investigated the distribution and sequence variability of the *rlrA* islet among 424 *S. pneumoniae* clinical isolates (26 of these were derived from the Pneumococcal Molecular Epidemiology Network [PMEN] multidrug-resistant collection) [15]. We evaluated isolates from distinct geographic regions, which were selected for epidemiological and genetic diversity. The isolates were analyzed for the presence and sequence variability of the *rlrA* islet, as well as serotype and genetic background as defined by multilocus sequence typing (MLST). Here we show that *rlrA*-positive pneumococci are found worldwide, and that the operon, defined by 3 variants, is strictly correlated with the genotype of *S. pneumoniae,* rather than the serotype. These findings, along with previous data showing that pili are protective against pneumococcal infection [11], suggest that they may be useful components for a vaccine.

MATERIALS AND METHODS

Strain collection. A total of 424 isolates collected worldwide were analyzed for the presence of *rlrA* (table 1). This collection consisted of 26 clones from the PMEN collection [15]; 53 invasive pneumococcal isolates from the Centers for Disease Control and Prevention in the United States [16]; 74 clinical isolates from Salvador, Brazil (42 isolates recovered from carriers and 32 from patients with meningitis, respectively); 26 clinical isolates from the Unites States and Finland (15 recovered from carriers, 2 from patients with otitis media, and 9 from patients with invasive diseases); 15 invasive clinical isolates from Ghana [17]; 14 invasive clinical isolates from Bangladesh; 40 clinical isolates from Italy; 12 invasive clinical isolates from Kenya; 76 clinical isolates from Sweden; and 11 laboratory strains from the Novartis collection. Additionally, data about *rlrA* presence have been extrapolated for 74 isolates from a previously published report [18] and for 3 isolates from sequence data available at the Sanger Institute Web site (http://www.sanger.ac.uk).

rlrA islet detection and sequencing. TIGR4 [19] and 670 (genomic sequence available at http://www.TIGR.org) *rlrA* nucleotide sequences were analyzed, and a set of 30 oligonucleotide primers was designed (table 2): 22 matching inside the islet, 2 annealing in 2 conserved genes (SP459-SP470 of TIGR4) flanking the operon, and 6 clade-specific primers. The set of primers was defined to detect the presence and location of the operon within all the isolates in the collection, to amplify and sequence the entire locus, and, for *rlrA*-positive strains, to determine the clade type (I, II, or III). In brief, the genomic location of the operon was determined by simultaneously assessing 4 polymerase chain reaction (PCR) amplifications: the first using primers (459 for, 470 rev) matching regions flanking the operon (when *rlrA* was absent, a lower fragment size was detectable); the second with 459 for and 1 rev, matching inside the islet (the fragment was detectable only when the islet was present and inserted in the genomic region between SP459-SP470); the third and fourth amplifying in conserved regions of the operon (amplification detected whenever the islet was present in the genome) (table 2).

Multilocus sequence typing (MLST). MLST was performed as previously described [20]. In brief, internal fragments of the *aroE, gdh, gki, recP, spi, xpt,* and *ddl* genes were amplified by PCR directly from the bacteria by use of the primer pairs indicated at http://spneumoniae.mlst.net/misc/info.asp#experimental. Sequences were obtained on both DNA strands by use of an ABI 3730xl DNA Analyzer. Alleles from the MLST Web site (http://spneumoniae.mlst.net) were downloaded for alignment analyses and sequence type (ST) determination. In MLST, an ST is

Table 2. Nucleotide sequences of the primers used in this study.

Forward primer	Sequence (5'→3')	Reverse primer	Sequence (5'→3')	Purpose
459 for	AACTGAATTGACACAACGTGTCTT	470 rev	GCCACACAAGATGTTGATGCTTTT	*rlrA* detection
459 for	AACTGAATTGACACAACGTGTCTT	P01 rev	AGCGACAAGCCACTGTATCATATT	*rlrA* detection
P01 for	ACTTTCTAATGAGTTGTTTAGGCG	P01 rev	AGCGACAAGCCACTGTATCATATT	*rlrA* sequencing
P02 for	CTGGTCGATAACTCCTTCAATCTT	P02 rev	GTACGACAAAAGTGTGGCTTGTT	*rlrA* sequencing and *rlrA* detection
P03 for	GAATGCGATATTCAGGACCAACTA	P03 rev	ATCTCACTGAGTTAATCCGTTCAC	*rlrA* sequencing
P04 for	TGTATACAAGTGTGTCATTGCCAG	P04 rev	CATCTTCACCTGTTCTCACATTTT	*rlrA* sequencing
P05 for	GCGGTCTTTAGTCTTCAAAAACA	P05 rev	CAAGAGAAAAACACAGAGCCATAA	*rlrA* sequencing
P06 for	TTGCTTAAGTAAGAGAGAAAGGAGC	P06 rev	CAGGAGTATAGTGTCCGCTTTCTT	*rlrA* sequencing
P07 for	GGCAATGTTGACTTTATGAAGGTG	P07 rev	TATCAGCATCCCTTTATCTTCAAAC	*rlrA* sequencing
P08 for	TGAGATTTTCTCGTTTCTCTTAGC	P08 rev	AATAGACGATGGGTATTGATCATGT	*rlrA* sequencing and *rlrA* detection
P09 for	CCGACGAACTTTGATGATTTATTG	P09 rev	ACCAACAGACGATGACTGTTAATC	*rlrA* sequencing
P10 for	AATGACTTTGAGCCTGTCTTGAT	P10 rev	TTCTACAATTTCCTGGCCATTATC	*rlrA* sequencing
P11 for	GCCATTTGGATCAGCTAAAAGTT	P11 rev	TTTTTCAACCCACTACAGTTGACA	*rlrA* sequencing
Clade I for	AACAGATGGGGATATGGATAAAATTG	Clade I rev	AATGGTAATTCAATTTCAATTGGA	clade type determination
Clade II for	AATCCATAAGTTACTGCTCTCAGA	Clade II rev	ATCCATAGCTACATTATTCAAAGT	clade type determination
Clade III for	GACAGATCAAGAGCTTGACGCTTG	Clade III rev	CTGGATCTACGAAACCTGCTGCAG	clade type determination

uniquely determined by the allelic profile. New allelic profiles have been submitted to the MLST database for ST assignment.

Clonal complex (CC). CCs are groups of STs that share a recent common ancestor. The eBURST algorithm defines clonal complexes by partitioning the MLST data set into groups of single-locus variants (SLVs), i.e., profiles that differ at 1 of the 7 MLST loci [21]. This partitioning associates each ST with a clonal complex and identifies the most likely founder ST, which is defined as being the ST with the greatest number of SLVs within the CC. To explore the relationship between *rlrA* presence in our data set and CC, we ran eBURST with default settings on the entire MLST database and subsequently assigned each ST in our data set to a CC. In this work, we have named CCs in accordance with the ST number of the founder predicted by eBURST.

Statistical analysis. All statistical analyses of association were performed using R (version 2.4.0; CRAN) and the vcd package (version 1.0–3; CRAN) for the analysis of categorical data. Pearson χ^2 *P* values and Cramer V coefficients were reported to test the null hypothesis of independence and to measure the strength of the observed correlations [22].

Gene prediction and multiple alignment. The 7 TIGR4 *rlrA* islet gene sequences were used to perform Smith-Waterman searches against the 13 *rlrA* sequences of *rlrA*-positive PMEN clones to find the genes in each locus. Multiple alignments of the homologous genes were obtained by translating the predicted nucleotide sequences into peptides, aligning, and then back-translating to the original nucleotide sequences with T-Coffee (version 4.70) [23]. The *rlrA* multiple alignment was reconstructed by concatenating the 7 aligned genes with the corresponding intergenic regions. Phylogenetic trees were reconstructed from nucleotidic multiple alignment by use of MEGA (version 3.1) [24], by use of unweighted pair group method with arithmetic mean and K2P distance correction. Multiple alignments were used with MEGA (version 3.1) [24] to measure the nucleotide and amino acid identity within and between clades I, II, and III.

RESULTS

***Prevalence of the rlrA operon in* S. pneumoniae.** To define the *rlrA* islet distribution, a panel of 424 isolates was analyzed. The collection was obtained by sampling clinical isolates worldwide; it includes 70 serotypes and 200 different STs, which can be grouped into 76 CCs [20]. *rlrA* islet presence was determined by PCR and confirmed (for the 26 PMEN clones) by Southern blot analysis with a conserved fragment of the *srtB* gene (data not shown). Out of the entire collection, 130 (30.6%) of the *S. pneumoniae* isolates were *rlrA* positive and showed the pilus operon consistently inserted in the same genomic region.

Association of the rlrA islet with serotype. To identify common features among *rlrA*-positive isolates, we investigated its association with capsular serotype. Figure 1*A* represents a histogram of isolates stratified by serotype [25]; the dark area corresponds to the fraction of *rlrA*-positive isolates. Of the most common serotypes (1, 3, 4, 5, 6B, 7F, 9V, 14, 18C, 19A, 19F, and 23F, which accounted for 257 (60.6%) of all pneumococcal isolates), serotypes 1, 3, 5, 7F, and 18C were not associated with the presence of the *rlrA* operon. However, with the exception of Serotype 9V, most *rlrA*-positive serotypes included variable fractions of *rlrA*-negative strains. Therefore, although there was an association between the capsular serotype and the presence of the islet, the strength of this association was minimal (Cramer $V = 0.68$).

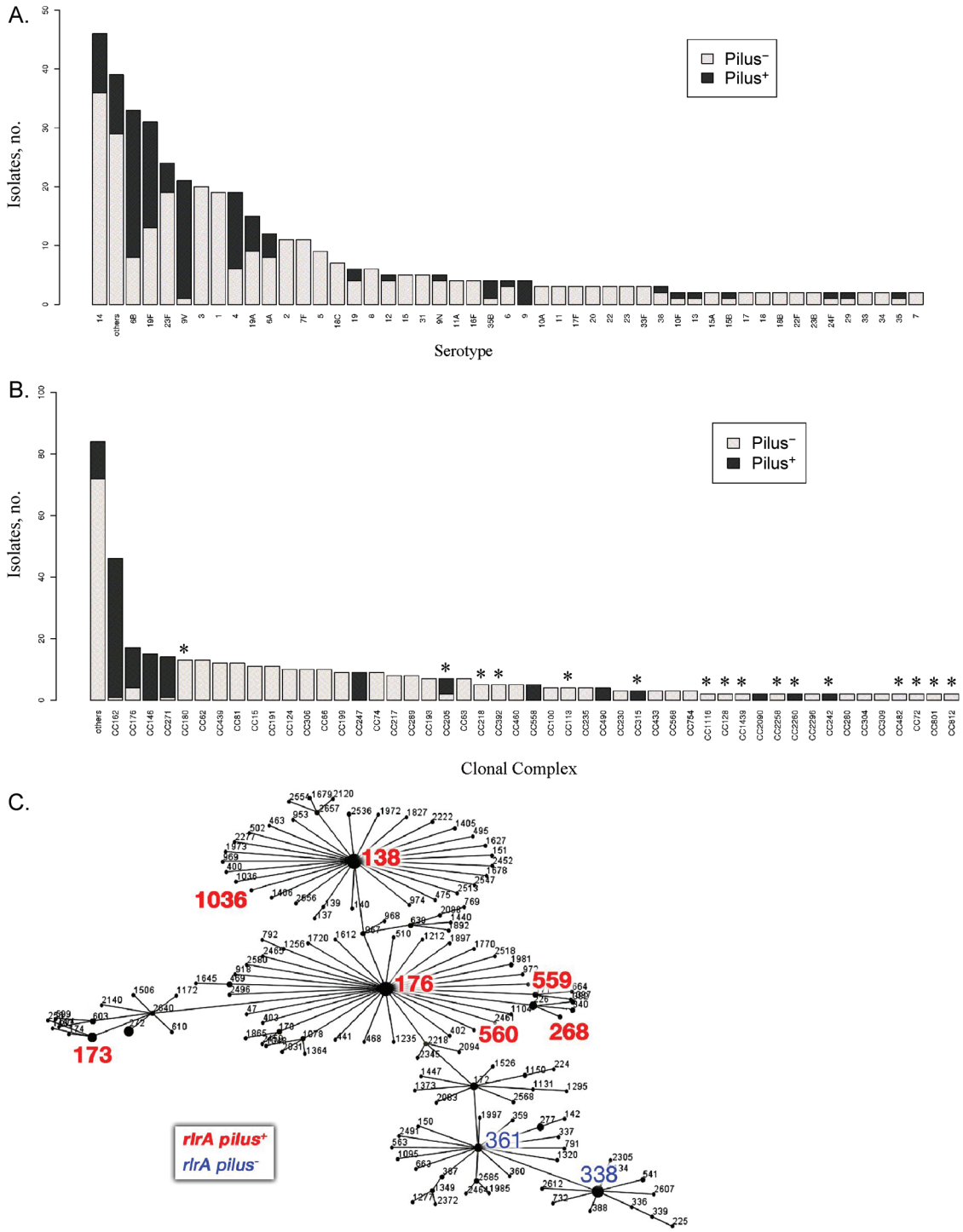


Figure 1. Distribution of *rlrA*-positive *Streptococcus pneumoniae* by serotype *(A)* and by clonal complex (CC) *(B)*. Histograms show the relative frequency of *rlrA*-positive *(black)* and *rlrA*-negative *(gray)* isolates in the collection, stratified by serotype or CC. In panel *A*, the column "others" contains serotypes represented in our collection by a single isolate; in panel *B*, "others" clusters together CCs represented by a single isolate and isolates (singletons) that could not be assigned to a CC. *Asterisks*, CCs for which 1 sequence type was tested. Total number of isolates, $N = 424$. *C*, eBURST graphic representation of CC 176, obtained by eBURST analysis on entire MLST *S. pneumoniae* database. Sequence types analyzed for *rlrA* presence are represented by larger font.

Association of the rlrA islet with genotype. The genetic structure of the *S. pneumoniae* collection was determined by MLST [26]. As described above, each isolate was assigned to a CC after ST determination. Figure 1*B* shows a histogram representing the number of isolates analyzed for each CC and illustrates the fraction of *rlrA*-positive and *rlrA*-negative *S. pneumoniae* isolates. Among the CCs most represented in this study (figure 1*B*), at least 10 different STs were analyzed for the pres-

Table 3. Presence of the *rlrA* islet in the Pneumococcal Molecular Epidemiology Network collection.

Clone number	Strain	Country of origin	Serotype	MLST type	Clonal complex	*rlrA* islet	Clade	Accession Number
1	Spain 23F-1	Spain	23F	81	CC81	−	. . .	. . .
2	Spain 6B-2	Spain	6B	90	CC90	+	II	EF560636
3	Spain 9V-3	France	9V	156	CC156	+	I	EF560637
4	Tennessee 23F-4	USA	23F	37	CC439	−	. . .	. . .
5	Spain 14–5	Spain	14	18	CC15	−	. . .	. . .
6	Hungary 19A-6	Hungary	19A	268	CC176	+	II	EF560625
7	South Africa 19A-7	South Africa	19A	75	S	−	. . .	. . .
8	South Africa 6B-8	South Africa	6B	185	CC185	−	. . .	. . .
9	England 14–9	England	14	9	CC15	−	. . .	. . .
10	CSR 14–10	Slovakia	14	20	CC15	−	-	-
11	CSRA 19A-11	Slovakia	19A	175	S	−	. . .	. . .
12	Finland 6B-12	Finland	6B	270	CC90	+	II	EF560632
13	South Africa 19A-13	South Africa	19A	41	CC41	−	. . .	. . .
14	Taiwan 19F-14	Taiwan	19F	236	CC271	+	I	EF560627
15	Taiwan 23F-15	Taiwan	23F	242	CC242	+	III	EF560629
16	Poland 23F-16	Poland	23F	173	CC176	+	II	EF560628
17	Maryland 6B-17	USA	6B	384	CC90	+	II	EF560634
18	Tennessee 14–18	USA	14	67	CC66	−	. . .	. . .
19	Colombia 5–19	Colombia	5	289	CC289	−	. . .	. . .
20	Poland 6B-20	Poland	6B	315	CC315	+	III	EF560635
21	Portugal 19F-21	Portugal	19F	177	CC177	+	III	EF560626
22	Greece 6B-22	Greece	6B	273	CC146	+	II	EF560633
23	N. Carolina 6A-23	USA	6A	376	CC2090	+	II	EF560631
24	Utah 35B-24	USA	35B	377	CC558	+	III	EF560630
25	Sweden 15A-25	Sweden	15A	63	CC63	−	. . .	. . .
26	Colombia 23F-26	Colombia	23F	338	CC176	−	. . .	. . .

NOTE. MLST, multilocus sequence typing; S, singleton.

ence of *rlrA* . Almost all the STs analyzed were homogeneous for *rlrA* presence, with a few exceptions. In brief, of 18 ST156 (CC162) and 7 ST205 (CC205) isolates analyzed, 1 and 2 isolates, respectively, were negative for the *rlrA* islet. Figure 1*B* clearly shows that CCs are composed of either *rlrA*-positive or *rlrA*-negative isolates, which indicates that its presence is dependent on the genetic background $P < .001$; Cramer $V = 0.96$). CC176 represents an exception to the scenario described above (figure 1*C*). Eight distinct STs were tested (comprising 16 isolates): 5 were *rlrA* positive and 3 were *rlrA* negative. By mapping the STs analyzed onto a predictive eBURST diagram (which displays hypothetical patterns of descent), we observed that *rlrA*-positive and *rlrA*-negative STs were grouped into 2 defined areas of the CC.

rlrA pilus islet sequence variability. In the TIGR4 strain, pili are encoded within the *rlrA* islet, a 12 kb genomic region flanked by a pair of *IS1167* elements, a feature characteristic of mobile genetic elements [13, 27]. The islet consists of 7 genes, namely, *rlrA,* which encodes a RofA-like transcriptional regulator, *rrgA, rrgB* and *rrgC*, which encode LPXTG cell-wall anchored proteins that constitute the structure of the pilus, and *srtB*, *srtC* and *srtD*, which encode for the 3 sortase enzymes that catalyze the pilus polymerization reaction [10, 12, 28].

The well-characterized PMEN collection was used as a reference collection for initial assessment of the variability of the *rlrA* islet in a global panel of pneumococcal isolates [15]. By PCR, the *rlrA* islet was found to be present in 13 (50%) of 26 PMEN clones (table 3). Sequencing of the full-length locus was carried out for the 13 PMEN strains (GenBank accession numbers EF560625–EF560637) and an additional 18 isolates from the collection (data not shown). Multiple alignment of entire nucleotide sequences revealed the presence of 3 main families, herein referred to as clades I, II, and III (figure 2).

rlrA sequences within each clade shared at least 98.8% DNA sequence identity, whereas sequences belonging to different clades shared from 88.1% to 92.1% identity (clades I and III being the most similar) (table 4). Data on protein sequence similarity between clades highlights the fact that the transcriptional regulator RlrA was the most conserved (100%), followed by the 3 sortases and the RrgC pilus subunit (98%–99%). Conversely, the RrgA (84%–98%) and RrgB (49%–67%) pilus components are the least conserved, with RrgB existing in 3 different variants, corresponding to

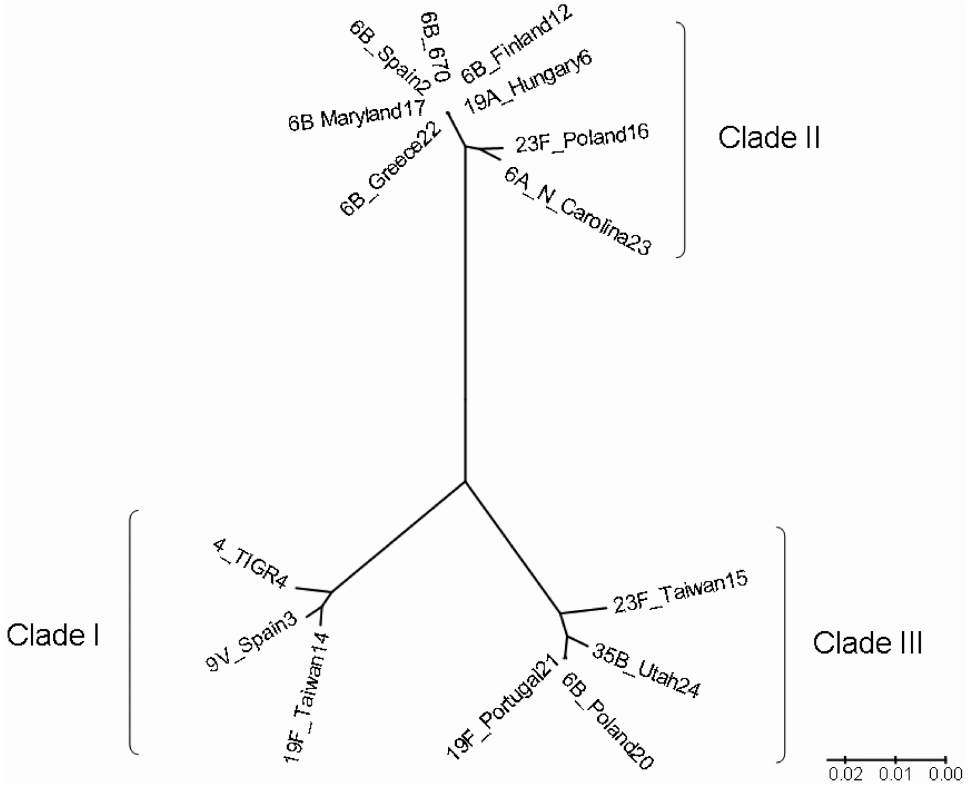


Figure 2. Genetic variability among *rlrA* islets in *Streptococcus pneumoniae* isolates. Distance-based tree obtained from the multiple sequence alignment of 15 *rlrA* islets (TIGR4, 6B 670, 13 *rlrA*-positive Pnemococcal Molecular Epidemiology Network clones).

the 3 clades of the *rlrA* operon (table 4). A set of clade-specific primers designed from the most variable regions of *rrgB* (table 1*B*) allowed the determination of *rlrA* clade type by PCR. Of the 130 *rlrA*-positive isolates, 81 (62.2%) were defined as clade I; 35 (26.8%), clade II; and 14 (10.7%), clade III. Furthermore, all strains belonging to the same CC were homogeneous for *rlrA* clade type.

Table 4. Sequence variability of the *rlrA* operon.

	Similarity between clades, % ± SE		
	I–II	I–III	II–III
rlrA islet	88.1 ± 0.3	92.1 ± 0.3	90.3 ± 0.3
Gene product, annotation			
RlrA, transcriptional regulator	100 ± 0	100 ± 0	100 ± 0
SrtB, sortase B	98.1 ± 0.6	98.0 ± 0.6	99.2 ± 0.3
SrtC, sortase C	95.4 ± 0.9	96.9 ± 0.6	94.9 ± 1
SrtD, sortase D	93.8 ± 1.1	97.7 ± 0.4	94.7 ± 0.9
RrgA, pilus subunit	84.2 ± 1.2	98.9 ± 0.3	84.0 ± 1.3
RrgB, pilus subunit	49.3 ± 1.9	53.9 ± 1.8	67.3 ± 1.8
RrgC, pilus subunit	98.3 ± 0.5	99.1 ± 0.3	97.8 ± 0.6

NOTE. Nucleotide homology of the entire *rlrA* locus and amino acid sequence homology of the *rlrA* islet gene products among 15 sequenced strains (TIGR4, 6B 670, 13 pilus-positive Pneumococcal Molecular Epidemiology Network clones) in the 3 clades identified by multiple sequence alignment and phylogenetic reconstruction. SE computed by 500 bootstraps with MEGA 3.1.

Sequence analysis of IS1167 target site duplications. The genomic region flanking the operon was analyzed both in *rlrA*-positive and *rlrA*-negative isolates. The objective was to investigate whether the presence of the *rlrA* islet was the result of a single insertion event, which may have evolved into 3 separate clades, or whether different clones acquired the islet independently.

In TIGR4, the *rlrA* islet is flanked by 2 *IS1167* mobile genetic elements characteristic of *S. pneumoniae.* These 2 elements, *IS1167-1* (SP0460) and *IS1167-2* (SP0469), share 61% identity at the nucleotide level. *IS1167* has been shown to be stable during laboratory passage and to exhibit a high degree of variability [29]. Moreover, *IS1167* sequences are characterized by the presence of imperfect terminal inverted repeats (which are responsible for homologous recombination) and are bracketed by 8-bp direct repeats, generated by the duplication of the *IS1167* target sequence.

Sequence analysis of this region in 30 isolates revealed the absence of target site duplication on the outer extremities of the operon (including IS elements), as well as on either side of the single IS elements. The lack of direct repeats suggests that the region was not actively inserted by these IS elements as a composite transposon, but could have been acquired by homologous recombination. To date, 2 different insertion sites have been found for the *IS1167-1* element (these sites are clade-type independent, but CC related). No differences were found for *IS1167-2* insertion sites in *rlrA*-positive isolates.

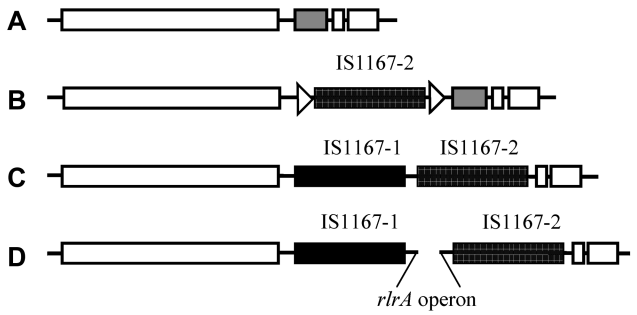


Figure 3. Schematic representation of the genomic region in *rlrA*-negative *Streptococcus pneumoniae* isolates *(A–C)* and in *rlrA*-positive isolates *(D). A,* genomic background (R6-like); *B,* genetic arrangement derived from *IS1167* insertion on an R6 genomic background; *C,* genetic arrangement potentially derived from loss of *rlrA. White triangles,* direct repeats formed by target site duplication. *Gray boxes,* open reading frame coding for a hypothetical protein; *white boxes,* open reading frames flanking the region.

***Sequence analysis of rlrA-negative* S. pneumoniae.** *rlrA*-negative pneumococci revealed 3 genetically distinct arrangements (figure 3): (1) a region of ~800 bp that codes for a hypothetical protein is present (as seen in R6 and in a majority of *rlrA*-negative strains), (2) an *IS1167-2* bracketed by 8-bp direct repeats is inserted in the genomic context described in the first scenario, and (3) the pilus operon is absent and both *IS1167-1* and *IS1167-2* are present. The latter situation, found in an ST156 isolate and in 2 ST205 isolates mentioned previously, could represent an event of *rlrA* islet loss, whereas the second scenario, described in 20 clonally related strains, supports the idea that this locus represents a hot spot region for the insertion of *IS1167* family members.

Antimicrobial-resistant clones are more likely to carry the pilus islet. Given the high prevalence of *rlrA*-positive strains in the antibiotic-resistant PMEN collection (50%), the relationship between antimicrobial resistance and *rlrA* presence was assessed further. In this analysis, we used the complete *S. pneumoniae* database downloaded from the MLST web site (http://www.mlst.net) instead of our collection, because antimicrobial resistance data was incomplete for our collection. All pneumococci for which minimum inhibitory concentration (MIC) values were available were classified as susceptible (S) or nonsusceptible (NS) to 3 major antibiotic compounds: penicillin, erythromycin, and tetracycline [30]. STs were subsequently classified as NS if they were composed of at least 50% NS isolates. STs not fulfilling this requirement were classified as S. In most cases, a single ST was composed of either S or NS strains (Cramer V correlation coefficients between ST and antibiotic resistance were 0.94, 0.78, and 0.77 for penicillin, erythromycin and tetracycline, respectively). In contrast to the distribution of the *rlrA* islet, several CCs were heterogeneous for resistance (the Cramer V correlation coefficients between CC and resistance were 0.81, 0.73, and 0.67 for penicillin, erythromycin and tetracycline, respectively).

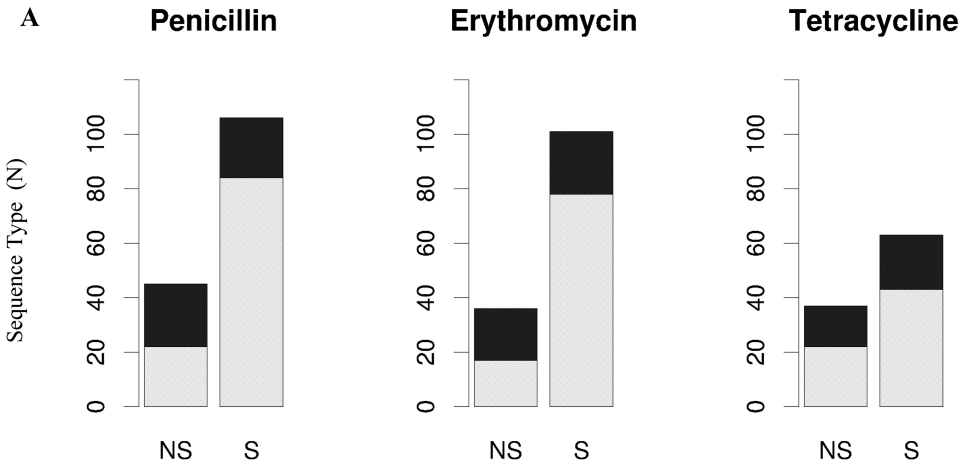
The presence of the *rlrA* islet was then projected onto the MLST data set (figure 4), by classifying as *rlrA* positive those STs in our collection that were composed of at least 50% *rlrA*-positive isolates. The percentage of *rlrA*-positive STs among strains susceptible to penicillin, erythromycin, or tetracycline was 21%, 23%, and 32%, respectively, whereas the percentage of *rlrA*-positive STs among clones not susceptible to penicillin, erythromycin, or tetracycline was 51%, 53%, and 41%, respectively. The Pearson χ^2 statistic showed that for penicillin and erythromycin, the presence of the *rlrA* islet was significantly associated with resistance ($P < .001$ and $P < .001$, respectively). A less significant association was found for tetracycline ($P < .37$), probably due to the lower number of strains for which MIC data were available.

DISCUSSION

The polysaccharide conjugate vaccine for *S. pneumoniae* has decreased the incidence of invasive disease caused by vaccine types, although the increased incidence disease due to nonvaccine serotypes makes difficult the estimate of polysaccharide conjugate vaccine coverage [6, 7]. Moreover, the epidemiology of the pneumococci is complicated by the diversity of circulating genotypes often disguised by multiple capsular types. The introduction of a protein-based vaccine could overcome limitations due to serotype replacement and serotype-dependent coverage. A tool to rapidly estimate the distribution and variability of protein antigens in *S. pneumoniae* could facilitate vaccine development.

In this study, a collection of isolates that represented both the capsular and genetic diversity of *S. pneumoniae* was used to determine the prevalence and antigenic variability of the *rlrA* pilus. In this article, we report 3 major findings. First, we demonstrate that the *rlrA* operon previously identified in laboratory strains is found in multiple isolates from diverse parts of the world. In fact, 30% of isolates in our pneumococcal collection contained this genetic islet, confirming a recent report about strains isolated from the Native American population [31]. This suggests that pili could become one of several components in a protein-based vaccine for *S. pneumoniae.*

Second, we show that presence of the *rlrA* islet segregates according to genotype, rather than phenotype. Moreover, the islet was probably acquired prior to the formation of the clonal complexes and steadily maintained during clonal diversification, even in CCs that show evidence of a complex evolutionary history. Following these observations, the occurrence of *rlrA*-negative STs in CC 176, the founder of which is *rlrA* positive, could be explained as an instance of pilus operon loss. This event can be inferred to have taken place between ST 176 and ST 361 (figure 1*C*), the *rlrA*-negative genotype then having been inherited by the descendant STs. Alternative explanations could be acquisition of the *rlrA* pilus islet, subsequent to the formation of distinct subclones of the CC; or artificial grouping by eBURST of isolates from 2 independent clonal expansions, one of which



B

	Penicillin	Erythromycin	Tetracycline
NS	0.51 (0.36-0.67)	0.53 (0.36-0.70)	0.41 (0.25-0.58)
S	0.21 (0.14-0.30)	0.23 (0.15-0.32)	0.32 (0.21-0.45)

Figure 4. Distribution of the *rlrA* islet in antimicrobial-resistant *S. pneumoniae* sequence types (STs). *A,* Histogram showing the relative frequency of *rlrA*$^{+}$ *(black)* and *rlrA*$^{-}$ *(gray)* STs. Strains within the multilocus sequence typing database collection for which minimum inhibitory concentration (MIC) values were provided were analyzed and stratified by ST and antibiotic susceptibility class (susceptible [S] and nonsusceptible [NS]). NS strains were defined by MIC values of >0.12 μg/mL, >0.5 μg/mL, and >4 μg/mL for penicillin, erythromycin, and tetracycline, respectively. Those STs in which >50% of the strains were NS were defined as NS. The total number of STs analyzed was 151 for penicillin, 137 for erythromycin, and 100 for tetracycline. The fraction of *rlrA*-positive STs within S and NS groups (95% confidence interval [CI], based on the binomial distribution of probability) were as follows: the fraction of *rlrA*-positive STs among strains susceptible to penicillin, erythromycin, or tetracycline was 0.21 (95% CI, 0.14–0.30), 0.23 (95% CI, 0.15–0.32), and 0.32 (0.21–0.45), respectively, whereas the fraction of *rlrA*-positive STs among clones not susceptible to penicillin, erythromycin, or tetracycline was 0.51 (95% CI, 0.36–0.67), 0.53 (0.36–0.70), and 0.41 (0.25–0.58), respectively.

originated from an *rlrA*-positive strain and the other from an *rlrA* negative strain, although testing of the eBURST algorithm suggests its robustness under the evolutionary processes [32].

Third, the genetic variability of the *rlrA* pilus islet can be organized into 3 clades, as demonstrated by sequence analysis of the operons from genetically diverse pneumococci. Single protein alignments highlight RrgA and RrgC as the most promising components for a serotype-independent vaccine, while the variability of RrgB makes this protein less attractive. Additionally, the *rlrA* clade type within a CC was determined to be identical, strongly suggesting clonal inheritance of this islet; this was in contrast with serotype, which frequently varies within a CC. These findings clearly reveal that the association between the *rlrA* islet and serotype depends on the genetic link between serotype and genotype. In fact, there is a significant correlation between serotype and the presence of the operon only for those serotypes that correspond to a restricted number of CCs, such as serotype 9V (CC 162) and serotype 3 (CC 180).

Furthermore, sequence analysis of the genomic region in *rlrA*-positive and *rlrA*-negative isolates suggests that the islet could have been acquired by homologous recombination during multiple events mediated by *IS1167* insertions. This latter observation raises questions regarding the potential genetic source of the *rlrA* genetic element in other *Streptococcus* species. To date, 25 *S. mitis* isolates have been analyzed and tested negative for the *rlrA* islet by Southern blot analysis and PCR (data not shown). Moreover, the probability that alternative elements could code for a pilus in *S. pneumoniae* strains lacking the *rlrA* islet, as is the case for *S. agalactiae* and *S. pyogenes* [33, 34], will be verified by sequencing new genomes.

In our study we also noted a correlation between the presence of the *rlrA* islet and antibiotic resistance, suggesting its clinical relevance with respect to the development and spread of pneumococcal disease. The reasons for the apparent association between the *rlrA* pilus operon and antibiotic resistance are not clear. It could be that lineages that contain both the *rlrA* islet and specific resistance genes are more commonly subject to recombination and more likely to have taken up both elements [35]. Other scenarios may be suggested in which pili aid adhesion during the colonization of a nasopharynx devoid of other bacterial flora as a result of antibiotic treatment. Elucidation of this relationship is beyond the scope of the current work. We note however, that this association is another attractive feature of pilus proteins as a vaccine component, as they could provide protection against drug-resistant clones.

This analysis provides the potential to predict the actual distribution of *rlrA* among a collection of isolates with defined genotypes, even given the current limits dictated by epidemiological surveillance. The efficacy of the pilus as a protective antigen furthermore supports a global effort to understand its epidemiology in the design of a protein-based vaccine.

Acknowledgments

We would like to acknowledge those people who provided invaluable *S. pneumoniae* clinical isolates for this study: Active Bacterial Core surveillance ([ABCs] a collaboration between the Centers for Disease Control and Prevention [CDC], state health departments, and universities) http://www.cdc.gov/ncidod/dbmd/abcs/index.htm) and Bernard Beall from the CDC for supplying genotyped and serotyped ABCs strains; Annalisa Pantosti (ISS Italy); Albert Ko and Mitermayer G. Reis (Oswaldo Cruz Foundation, Brazil); Gerd Plushke (Swiss Tropical Institute); Samir Saha (Dhaka Shishu Hospital, Bangladesh); and Anthony Scott (Kenyan Medical Research Center, Kenya). We would like to thank Mogen Kilian (University of Aarhus, Denmark) for providing us with *S. mitis* genomic DNA. We would also like to thank Julian Parkhill from the Sanger Institute for *rlrA* sequence information and helpful discussions on mobile genetic elements, as well as the Novartis Sequencing Facility, Beatrice Rogolino, Silvia Guidotti, and Monica Giraldi for technical assistance.

References

1. Ortqvist A, Hedlund J, Kalin M. *Streptococcus pneumoniae:* epidemiology, risk factors, and clinical features. Semin Respir Crit Care Med **2005**; 26:563–74.
2. Hanage WP, Huang SS, Lipsitch M, et al. Diversity and antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae* carriage isolates in the post-heptavalent conjugate vaccine era. J Infect Dis **2007**; 195:347–52.
3. Sogstad MK, Hoiby EA, Caugant DA. Molecular characterization of non-penicillin–susceptible *Streptococcus pneumoniae* in Norway. J Clin Microbiol **2006**; 44:3225–30.
4. Farrell DJ, Klugman KP, Pichichero M. Increased antimicrobial resistance among nonvaccine serotypes of *Streptococcus pneumoniae* in the pediatric population after the introduction of 7-valent pneumococcal vaccine in the United States. Pediatr Infect Dis J **2007**; 26:123–8.
5. Sjostrom K, Blomberg C, Fernebro J, et al. Clonal success of piliated penicillin nonsusceptible pneumococci. Proc Natl Acad Sci U S A **2007**; 104:12907–12.
6. Gonzalez BE, Hulten KG, Lamberth L, Kaplan SL, Mason EO Jr. *Streptococcus pneumoniae* serogroups 15 and 33: an increasing cause of pneumococcal infections in children in the United States after the introduction of the pneumococcal 7-valent conjugate vaccine. Pediatr Infect Dis J **2006**; 25:301–5.
7. Pai R, Moore MR, Pilishvili T, Gertz RE, Whitney CG, Beall B. Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. J Infect Dis **2005**; 192:1988–95.
8. Barocchi MA, Censini S, Rappuoli R. Vaccines in the era of genomics: the pneumococcal challenge. Vaccine **2007**; 25:2963–73.
9. Ogunniyi AD, Grabowicz M, Briles DE, Cook J, Paton JC. Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae.* Infect Immun **2007**; 75:350–7.
10. Barocchi MA, Ries J, Zogaj X, et al. A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci U S A **2006**; 103:2857–62.
11. Gianfaldoni C, Censini S, Hilleringmann M, et al. *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. Infect Immun **2007**; 75:1059–62.
12. LeMieux J, Hava DL, Basset A, Camilli A. RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae rlrA* pathogenicity islet. Infect Immun **2006**; 74:2453–6.
13. Hemsley C, Joyce E, Hava DL, Kawale A, Camilli A. MgrA, an orthologue of Mga, acts as a transcriptional repressor of the genes within the *rlrA* pathogenicity islet in *Streptococcus pneumoniae.* J Bacteriol **2003**; 185:6640–7.
14. Nelson AL, Ries J, Bagnoli F, et al. RgrA is a pilus-associated adhesin in *Streptococcus pneumoniae.* Mol Microbiol **2007**; 66:329–40.
15. McGee L, McDougal L, Zhou J, et al. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. J Clin Microbiol **2001**; 39:2565–71.
16. Beall B, McEllistrem MC, Gertz RE Jr, et al. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. J Clin Microbiol **2006**; 44:999–1017.
17. Leimkugel J, Adams Forgor A, et al. An outbreak of serotype 1 *Streptococcus pneumoniae* meningitis in northern Ghana with features that are characteristic of *Neisseria meningitidis* meningitis epidemics. J Infect Dis **2005**; 192:192–9.
18. Paterson GK, Mitchell TJ. The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. Microbes Infect **2006**; 8:145–53.
19. Tettelin H, Nelson KE, Paulsen IT, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae.* Science **2001**; 293:498–506.
20. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae:* identification of clones associated with serious invasive disease. Microbiology **1998**; 144:3049–60.
21. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **2004**; 186:1518–30.
22. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in FORTRAN: the art of scientific computing. 2nd ed. New York: Cambridge University Press, 1992.
23. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol **2000**; 302:205–17.
24. Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **2004**; 5:150–63.
25. Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. Lancet Infect Dis **2005**; 5:83–93.
26. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A **1998**; 95:3140–5.
27. Hava DL, Hemsley CJ, Camilli A. Transcriptional regulation in the *Streptococcus pneumoniae rlrA* pathogenicity islet by RlrA. J Bacteriol **2003**; 185:413–21.
28. Hava DL, Camilli A. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. Mol Microbiol **2002**; 45:1389–406.
29. Zhou L, Hui FM, Morrison DA. Characterization of IS1167, a new insertion sequence in *Streptococcus pneumoniae.* Plasmid **1995**; 33:127–38.
30. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing: 16th informational supplement, document M100-S16. Wayne, PA: Clinical and Laboratory Standards Institute; **2006**.
31. Basset A, Trzcinski K, Hermos C, et al. Association of the pneumococcal pilus with certain capsular serotypes but not with increased virulence. J Clin Microbiol **2007**; 45:1684–9.
32. Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol **2007**; 7:30.
33. Lauer P, Rinaudo CD, Soriani M, et al. Genome analysis reveals pili in Group B *Streptococcus.* Science **2005**; 309:105.
34. Mora M, Bensi G, Capo S, et al. Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci U S A **2005**; 102:15641–6.
35. Hsieh YC, Wang JT, Lee WS, et al. Serotype competence and penicillin resistance in *Streptococcus pneumoniae.* Emerg Infect Dis **2006**; 12:1709–14.