

Supplemental Materials and Methods, supplementary Tables S1–S5, and
supplementary Figures S1–S9

**Insights into the evolutionary relationships of LytA autolysin and Ply pneumolysin-
like genes in *Streptococcus pneumoniae* and related streptococci**

María Morales, Antonio J. Martín-Galiano, Mirian Domenech and Ernesto García*

Materials and Methods

Bacterial strains, growth conditions and transformation

Pneumococcal strains were grown in C medium (Lacks and Hotchkiss 1960) either supplemented (C + Y) or not supplemented with 0.08% yeast extract. *Escherichia coli* strains were grown in Luria-Bertani medium (Sambrook and Russell 2001). For *in vitro* biofilms assays, we used the non-encapsulated pneumococcal strain R6 and its derivatives: R924 (*lytA::kan*) (Moscoso and Claverys 2004), P206 (*plyA*), and P259 (*plyA lytA::kan*). The following *E. coli* strains were used: DH5 α (*supE44 Δ lacU169 (ϕ 80 lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*) (Hanahan 1983), and BL21 (DE3) (*hsdS gal (λ clts857 ind1 Sam7 nin5 lacUV5-T7 gene 1*) (Novagen). Transformation of *E. coli* strains was performed by the rubidium chloride method (Sambrook and Russell 2001).

Biofilm formation assay and quantification

The optimal conditions for biofilm formation by pneumococcal cells have been previously described (Moscoso, et al. 2006). Cells were grown in C medium to an optical density at 550 nm (OD₅₅₀) of 0.5–0.6, sedimented by centrifugation, resuspended in an equal volume of C medium, and diluted 1/10. Aliquots of 200 μ l were dispensed into each well of polystyrene microtiter plates (Corning, New York, USA). After 6 h of incubation at 34°C, the biofilm formed was stained with 0.2% crystal violet and rinsed three times with distilled water to remove non-adherent bacteria. After solubilizing the biofilm in 95% ethanol, the OD₅₉₅ was determined using an Anthos 2020 microplate absorbance reader (Anthos Labtec Instruments, Wals, Austria).

PCR amplification and cloning of *lytA* alleles

Routine DNA manipulations were performed essentially as described elsewhere (Sambrook and Russell 2001). Restriction endonucleases and T4 DNA ligase were purchased from Takara Bio and used according to the recommendations of the suppliers. DNA fragments were purified using the GeneClean II kit (Bio 101). For PCR reactions, recombinant DNA polymerase from *Thermus thermophilus* HB27 was used (Biotools B&M Laboratories, Spain). For cloning, *lytA_{Spn}* alleles coding for the NAM-amidase alleles 1, 2, 4, 5, or 7 (Supplemental Table S4) were PCR amplified using a mixture of Taq and Pfu DNA polymerases (Biotools B&M Laboratories, Spain) and oligonucleotides LytA-ATG and LytA-STOP (Moscoso, et al. 2010). The products amplified by PCR were purified using a High Pure PCR product purification kit (Roche). The resultant DNA fragments (1,012 bp each) were digested with NdeI and BamHI and cloned into plasmid pT7-7 (Tabor 1990). The ligation mixtures were used to transform *E. coli* DH5 α . Next, the recombinant plasmids were introduced into *E. coli* BL21 (DE3). The accuracy of the different constructs was confirmed by completely sequencing the insert of the corresponding recombinant plasmid. All primers for PCR amplification and nucleotide sequencing were purchased from Sigma-Aldrich.

Overproduction and purification of LytA alleles, and assay of cell wall lytic activity

For overproduction of LytA NAM-amidases, *E. coli* BL21 (DE3) cells harboring the appropriate plasmid were incubated in Luria-Bertani medium with ampicillin (100 $\mu\text{g ml}^{-1}$) up to an optical density at 600 nm of 0.4. At this time point, isopropyl- β -D-thiogalactopyranoside (0.1 mM) was added, and incubation continued for 4 h at 37°C with shaking. The cultures were centrifuged, and the bacteria were resuspended in 20 mM sodium phosphate buffer (pH 6.9) and disrupted by two 10 s periods of sonication. After centrifugation (10,000 $\times g$, 10 min, 4°C), the supernatant was loaded onto a DEAE-cellulose column and NAM-amidases were purified as previously described (Sánchez-Puelles, et al. 1992). LytA concentration was determined

spectrophotometrically using a molar absorption coefficient at 280 nm of 113,750 M⁻¹ cm⁻¹ (Usobiaga, et al. 1996).

Pneumococcal cell walls were radioactively labeled with [*methyl*-³H]choline and assays for cell wall lytic (NAM-amidase) activity were conducted according to standard procedures described elsewhere (Mosser and Tomasz 1970; Romero, et al. 2004). One unit of amidase activity was defined as the amount of enzyme needed to catalyze the hydrolysis (solubilization) of 1 µg of cell wall material in 10 min at 37°C.

Statistical analysis

The data for NAM-amidase activity include the mean ± standard error of at least three independent experiments, each performed in triplicate. Statistical significance was examined using the Student *t* test. For multiples comparisons, one-way analyses of variance (ANOVA) were performed, followed by Dunnet's *post hoc* test when the ANOVA rejected the null hypothesis. Differences were considered statistically significant when P <0.05. The SAS 9.3 statistical package (SAS Institute, Cary, NC) was used for all analyses.

Table S1. Complete and draft genomic sequences of SMG and other relevant streptococci^a.

Species	Complete genomes	Draft genomes	Reference ^b
SMG ^c			
<i>S. australis</i>	0	2	(Willcox, et al. 2001)
<i>S. cristatus</i>	0	2	(Handley, et al. 1991)
<i>S. dentisani</i>	0	2	(Camelo-Castillo, et al. 2014)
<i>S. gordonii</i>	1	0	(Kilian, et al. 1989)
<i>S. infantis</i>	0	6	(Kawamura, et al. 1998)
<i>S. lactarius</i>	0	0	(Martín, et al. 2011)
<i>S. mitis</i>	1	30	(Andrewes and Horder 1906)
<i>S. oligofermentans</i>	1	0	(Tong, et al. 2003)
<i>S. oralis</i>	1	12	(Bridge and Sneath 1982)
<i>S. parasanguinis</i>	2	7	(Whiley, et al. 1990)
<i>S. peroris</i>	0	1	(Kawamura, et al. 1998)
<i>S. pneumoniae</i>	27	275	(Chester 1901)
<i>S. pseudopneumoniae</i>	1	6	(Arbique, et al. 2004)
<i>S. rubneri</i>	0	0	(Huch, et al. 2013)
<i>S. sanguinis</i>	1	21	(White and Niven 1946)
<i>S. sinensis</i>	0	1	(Woo, et al. 2002)
<i>S. tigurinus</i>	0	4	(Zbinden, et al. 2012)
<i>S. troglodytidis</i>	0	0	(Zhang, et al. 2013)
Other streptococci			
<i>S. agalactiae</i>	18	295	
<i>S. anginosus</i>	4	10	
<i>S. mutans</i>	4	161	
<i>S. pyogenes</i>	30	209	
<i>S. suis</i>	20	87	
<i>S. thermophilus</i>	7	13	
Other species ^d	37	167	

a. Genomic sequences available at the National Center for Biotechnology Information (NCBI) (last date accessed, 25 October 2014).

b. Publication corresponds only to SMG.

c. *Streptococcus massiliensis* has also been proposed to belong to SMG but its phylogenetic position is still unclear (Glazunova, et al. 2006).

d. *S. anginosus*, *S. caballi*, *S. canis*, *S. castoreus*, *S. constellatus*, *S. cricetus*, *D. devriesei*, *S. didelphis*, *S. downei*, *S. dysgalactiae*, *S. equi*, *S. equinus*, *S. ferus*, *S. gallolyticus*, *S. henryi*, *S. hyovaginalis*, *S. ictaluri*, *S. infantarius*, *S. iniae*, *S. intermedius*, *S. lutetiensis*, *S. macacae*, *S. macedonicus*, *S. marimammalium*, *S. massiliensis*, *S. merionis*, *S. minor*, *S. orisratti*, *S. ovis*, *S. parauberis*, *S. pasteurianus*, *S. plurextorum*, *S. porci*, *S. porcinus*, *S. pseudoporcinus*, *S. ratti*, *S. salivarius*, *S. thoralensis*, *S. uberis*, *S. urinalis* *S. vestibularis*, and *Streptococcus* sp.

Table S2.

See separate file

Table S3.

See separate file

Table S4. Correspondence between *lytA* and LytA alleles of *S. pneumoniae*

<i>lytA_{Spn}</i> allele (Fam_) ^a (Acc. No.)	LytA _{Spn} Allele ^b	Specific activity	
		(Units mg ⁻¹) ^c	Significant (<i>P</i> < 0.05) ^d
1 (B) (AE007317)	1	1.3 × 10 ⁶ ± 0.97 × 10 ⁵	allele 1 vs. 7: Yes
2 (B) (AE005672)	2	1.3 × 10 ⁶ ± 2.2 × 10 ⁵	allele 2 vs. 7: Yes
3 (A) (CP002176)	3		
4 (B) (FQ312030)	4	1.2 × 10 ⁶ ± 1.5 × 10 ⁵	allele 4 vs. 7: Yes
5 (A) (AGPA01000015)	5	1.6 × 10 ⁶ ± 1.9 × 10 ⁵	allele 5 vs. 7: No
6 (A) (AWTX01000004)	5		
7 (A) (FM211187)	5		
8 (A) (FQ312027)	5		
9 (A) (ALAZ01000017)	6		
10 (A) (FQ312045)	2		
11 (A) (CP002121)	7	1.9 × 10 ⁶ ± 2.1 × 10 ⁵	
12 (B) (AGID01000009)	2		
13 (A) (CP001845)	2		
14 (B) (AIKV01000007)	2		
15 (A) (CP000920)	5		
16 (B) (CP000918)	2		
17 (B) (ALCRO1000016)	2		
18 (B) (AFAX01000007)	8		
19 (B) (ALCV01000008)	5		
20 (B) (ALBF01000021)	2		
21 (B) (AQTO01000004)	9		
22 (A) (AKQY01000003)	10		
23 (A) (ALBM01000008)	11		
24 (A) (AILM01000015)	2		
25 (A) (AILB01000008)	12		
26 (A) (AIKT01000015)	5		
27 (A) (ALBP01000011)	13		
28 (A) (ALBK01000014)	6		
29 (A) (ABAC01000008)	7		
30 (A) (ABAB01000008)	—		

^a Fam_A and Fam_B of *lytA_{Spn}* alleles have been described elsewhere (Morales, et al. 2010).

^b The *lytA_{30-Spn}* allele was not translated into protein since it may contain a sequencing error.

^c Expressed as the mean ± the standard error of the mean.

^d For multiples comparisons, one-way analyses of variance (ANOVA) were performed, followed by Dunnet's *post hoc* test when the ANOVA rejected the null hypothesis. Differences were considered statistically significant when *P* < 0.05.

References

Morales M, et al. 2010. Evidence of localized prophage-host recombination in the *lytA* gene encoding the major pneumococcal autolysin. *J Bacteriol* 192:2624–2632.

TGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTAC
AAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAACGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACT
ACCTCAAACCAGACGGAACACTGGCAGACAAGCCAGAATTCACAGTAGAGCCAGATGGCTTGATTACAGTAAAATAA

>27_SPN

ATGGAAATTAATGTGAGTAAATTAAGAACAGATTTGCCTCAAGTCGGCGTGAACCATATAGGCAAGTACACGCACACTCAACTGGGA
ATCCGCATTAACCGTACAGAATGAAGCGGATTATCACTGGCGGAAAGACCCAGAATTAGGTTTTTCTCGCACATTGTTGGGAACGG
TTGCATCATGCAGGTAGGACTGTTGATAATGGTGCCTGGGACGTTGGGGCGGTTGGAATGCTGAGACCTATGCAGCGGTTGAACGT
ATTGAAAGCCATTCAACTAAAGAAGAGTTCATGACGGACTACCGCTTTATATCGAACTCTTACGCAATCTAGCAGATGAAGCAGGTT
TGCCGAAAACGCTTGATACAGGGAGTTTAGCTGGAATTAACACGCACGAGTATTGCACGAATAACCAACCAACAACCCTCAGACCA
TGTGGATCCATACCCCTTACTTGGCAAAAATGGGGCATTAGCCGTGAGCAGTTAAGTATGATATTGAGAACGGCTTGACGATTGAAACA
GGCTGGCAGAAGAATGACACTGGCTACTGGTACGTACATTACAGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTT
GGTACTACTTTGACAGTTTCAAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTTCGACAACCTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTTGTATGTAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTAC
AAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACT
ACCTCAAACCAGACGGAACACTGGCAGATAAACAGAGTTTACAGTAGAGCCAGATGGCTTGATTACAGTAAAATAA

>28_SPN

ATGGAAATTAATGTGAGTAAATTAAGAACAGATTTGCCTCAAGTCGGCGTGAACCATATAGGCAAGTACACGCACACTCAACTGGGA
ATCCGCATTAACCGTACAGAATGAAGCGGATTATCACTGGCGGAAAGACCCAGAATTAGGTTTTTCTCGCACATTGTTGGGAACGG
TTGCATCATGCAGGTAGGACTGTTGATAATGGTGCCTGGGACGTTGGGGCGGTTGGAATGCTGAGACCTATGCAGCGGTTGAACGT
ATTGAAAGCCATTCAACTAAAGAAGAGTTCATGACGGACTACCGCTTTATATCGAACTCTTACGCAATCTAGCAGATGAAGCAGGTT
TGCCGAAAACGCTTGATACAGGGAGCTTAGCTGGAATTAACACGCACGAGTATTGCACGAATAACCAACCAACAACCCTCAGACCA
TGTGGATCCATACCCCTTACTTGGCAAAAATGGGGCATTAGCCGTGAGCAGTTAAGTATGATATTGAGAACGGCTTGACGATTGAAACA
GGCTGGCAGAAGAATGACACTGGCTACTGGTACGTACATTACAGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTT
GGTACTACTTTGACAGTTTCAAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTTCGACAACCTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGAGAAGTGGTACTATTTTGTATGTAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTAC
AAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACT
ACCTCAAACCAGACGGAACACTGGCAGATAAACAGAGTTTACAGTAGAGCCAGATGGCTTGATTACAGTAAAATAA

>29_SPN

ATGGAAATTAATGTGAGTAAATTAAGAACAGATTTGCCTCAAGTCGGCGTGAACCATATAGGCAAGTACACGCACACTCAACTGGGA
ATCCGCATTAACCGTACAGAATGAAGCGGATTATCACTGGCGGAAAGACCCAGAATTAGGTTTTTCTCGCACATTGTTGGGAACGG
TTGCATCATGCAGGTAGGACTGTTGATAATGGTGCCTGGGACGTTGGGGCGGTTGGAATGCTGAGACCTATGCAGCGGTTGAACGT
ATTGAAAGCCATTCAACTAAAGAAGAGTTCATGACGGACTACCGCTTTATATCGAACTCTTACGCAATCTAGCAGATGAAGCAGGTT
TGCCGAAAACGCTTGATACAGGGAGTTTAGCTGGAATTAACACGCACGAGTATTGCACGAATAACCAACCAACAACCCTCAGACCA
TGTGGATCCATACCCCTTACTTGGCAAAAATGGGGCATTAGCCGTGAGCAGTTAAGTATGATATTGAGAACGGCTTGACGATTGAAACA
GGCTGGCAGAAGAATGACACTGGCTACTGGTACGTACATTACAGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTT
GGTACTACTTTGACAGTTTCAAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTTCGACAACCTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGAGAAGTGGTACTATTTTGTATGTAGAAGGTGCCATGAAGACAGGCTGGGTCAAGTAC
AAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACT
ACCTCAAACCAGACGGAACACTGGCAGATAAACAGAGTTTACAGTAGAGCCAGATGGCTTGATTACAGTAAAATAA

>30_SPN

ATGGAAATTAATGTGAGTAAATTAAGAACAGATTTGCCTCAAGTCGGCGTGAACCATATAGGCAAGTACACGCACACTCAACTGGGA
ATCCGCATTAACCGTACAGAATGAAGCGGATTATCACTGGCGGAAAGACCCAGAATTAGGTTTTTCTCGCACATTGTTGGGAACGG
TTGCATCATGCAGGTAGGACTGTTGATAATGGTGCCTGGGACGTTGGGGCGGTTGGAATGCTGAGACCTATGCAGCGGTTGAACGT
ATTGAAAGCCATTCAACTAAAGAAGAGTTCATGACGGACTACCGCTTTATATCGAACTCTTACGCAATCTAGCAGATGAAGCAGGTT
TGCCGAAAACGCTTGATACAGGGAGTTTAGCTGGAATTAACACGCACGAGTATTGCACGAATAACCAACCAACAACCCTCAGACCA
TGTGGATCCATACCCCTTACTTGGCAAAAATGGGGCATTAGCCGTGAGCAGTTAAGTATGATATTGAGAACGGCTTGACGATTGAAACA
GGCTGGCAGAAGAATGACACTGGCTACTGGTACGTACACTCAGACGGCTCTTATCCAAAAGACAAGTTTGAGAAAATCAATGGCACTT
GGTACTACTTTGACAGTTTCAAGGCTATATGCTTGCAGACCGCTGGAGGAAGCACACAGACGGCAACTGGTACTGGTTTCGACAACCTCAGG
CGAAATGGCTACAGGCTGGAAGAAAATCGCTGATAAGTGGTACTATTTCAACGAAGAAGGTGCCATGAAGACTTGAAGGCTCAAGTAC
AAGGACACTTGGTACTACTTAGACGCTAAAGAAGGCGCCATGGTATCAAATGCCTTTATCCAGTCAGCGGACGGAACAGGCTGGTACT
ACCTCAAACCAGACGGAACACTGGCAGACAAGCCAGAATTCACAGTAGAGCCAGATGGCTTGATTACAGTAAAATAA

Amino acid sequences of the *LytA*_{Spn} alleles analyzed in this study

>1_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>2_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>3_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIAEKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>4_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>5_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>6_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIAEKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>7_SPN
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIAEKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>8_SPN GA13499
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>9_SPN 801
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>10_SPN PCS81218
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAESYAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>11_SPN 2081074
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>12_SPN GA47597
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

>13_SPN 2082239
MEINVSCLRDLDPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPPELGGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNNAETAAVEL
IESHSTKEEFMTDYRLYIELLRNLDAEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPPYLAKEGWSREQFKHDIENGLTIET
GWQKNDTGYWVHSDGSYPKDKFEKINGTWYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWYFNEEGAMKTGWVKY
KDTWYLLDAKEGAMVSNAFIQSADGTGWYLLKPDGTLADLADPEFTVEPDGLITVK

Table S5. Correspondence between *plyA* and PlyA alleles of *S. pneumoniae*^a

<i>plyA</i> allele (Fam_) ^b	PlyA allele ^c	Acc. No. of <i>plyA</i> alleles in databases	References	Specific activity (HU mg ⁻¹) ^d
1 (1)	2	AE005672	(Tettelin, et al. 2001)	4.4 × 10 ⁵
2 (1)	2	GU968237	(Jefferies, et al. 2010)	
3 (1)	2	GU968219	(Jefferies, et al. 2010)	
4 (1)	2	FQ312045	(Donati, et al. 2010)	
5 (1)	2	GU968217	(Jefferies, et al. 2010)	
6 (1)	2	EF413941	(Jefferies, et al. 2007)	
7 (1)	2	CP000918	Unpublished	
8 (1)	2	DQ251179	(Kirkham, et al. 2006)	
9 (1)	2	GU968385	(Jefferies, et al. 2010)	
10 (1)	1	GU968223	(Jefferies, et al. 2010)	4.1 × 10 ⁵
11 (1)	1	EF413947	(Jefferies, et al. 2007)	
12 (1)	1	GU968317	(Jefferies, et al. 2010)	
13 (1)	1	CP002176	Unpublished	
14 (2)	5	EF413960	(Jefferies, et al. 2007)	Non-hemolytic
15 (1)	1	GU968247	(Jefferies, et al. 2010)	
16 (1)	1	–	–	
17 (1)	1	–	–	
18 (1)	1	CP000936	Unpublished	
19 (1)	2	–	–	
20 (2)	5	EF368014	(Marks, et al. 2007)	
21 (2)	10	EF413934	(Jefferies, et al. 2007)	1.1 × 10 ⁵
22 (1)	11	EF413933	(Jefferies, et al. 2007)	5.9 × 10 ⁵
23 (1)	2	–	–	
24 (1)	9	GU968397	(Jefferies, et al. 2010)	3.1 × 10 ⁵
25 (1)	16	GU968252	(Jefferies, et al. 2010)	Unknown
26 (1)	2	GU968225	(Jefferies, et al. 2010)	
27 (1)	2	GU968238	(Jefferies, et al. 2010)	
28 (1)	2	GU968396	(Jefferies, et al. 2010)	
29 (2)	3	EF413957	(Jefferies, et al. 2007)	6.9 × 10 ³
30 (1)	20 ^e	–	–	Unknown
31 (1)	1	–	–	
32 (1)	2	–	–	
33 (1)	8	GU968401	(Jefferies, et al. 2010)	9.1 × 10 ⁴
34 (2)	3	–	–	
35 (1)	1	GU968221	(Jefferies, et al. 2010)	

^a Novel *plyA* or PlyA alleles are shaded in a light or dark gray background respectively.

^b Fam_1 and Fam_2 designate 1416- and 1410-bp-long *plyA*_{Spn} alleles respectively.

^c Numbering of PlyA alleles correspond to those previously proposed (Jefferies, et al. 2010).

^d The specific activities of the PlyA alleles [in hemolytic units (HU) per milligram of protein] were adapted from Jefferies et al. (2007).

^e Since up to 19 different PlyA alleles has been reported to date (Jefferies, et al. 2007; Jefferies, et al. 2010; Harvey, et al. 2011), this new allele has preliminarily named as allele 20.

References

Donati C, et al. 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol.* 11:R107.

Harvey RM, Ogunniyi AD, Chen AY, Paton JC. 2011. Pneumolysin with low hemolytic activity confers an early growth advantage to *Streptococcus pneumoniae* in the blood. *Infect Immun.* 79:4122–4130.

Jefferies JMC, et al. 2007. Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. *J Infect Dis.* 196:936–944.

Jefferies JMC, et al. 2010. Identification of novel pneumolysin alleles from paediatric carriage isolates of *Streptococcus pneumoniae*. *J Med Microbiol.* 59:808–814.

Kirkham L-AS, et al. 2006. Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *J Clin Microbiol.* 44:151–159.

Marks M, et al. 2007. Influence of neutropenia on the course of serotype 8 pneumococcal pneumonia in mice. *Infect Immun.* 75:1586–1597.

Tettelin H, et al. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 293:498–506.

Nucleotide sequences of novel *plyA* alleles

>*plyA16*

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGGAGAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGGTTGTTATCGAAAGAAAGAAGCGGAGCTTGTGACAAATACAAGTGA
TATTTCTGTAACAGCTACCAACGACAGTCGCCCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTAGAGAATAATCCCACT
CTTCTTGGCGTGCATCGTGTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAGGTGTTCCGGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGAAATCTCTTT
GATATTGATTTTAACTCTGTCCATTACAGGCGAAAAGCAGATTCAGATTGTTAATTTTAAAGCAGATTATTATACAGTCAGCGTAGACG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAGGAGTCAAGGTAGCTCCTCAGACAGAGTGAAGCAGATTTTGGACAATACAGAAGTGAAGGCGGTATTTTAGGGG
GCGACCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGTTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGATGAAT
TATCCTATGATCATCAAGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTGACGGCTCACTTTACCCTAG
TATTCCTTTAAAAGGGAATGTTCTGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAAACCGATTGGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGATTAG

>*plyA17*

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGGAGAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGGTTGTTATCGAAAGAAAGAAGCGGAGCTTGTGACAAATACAAGTGA
TATTTCTGTAACAGCTACCAACGACAGTCGCCCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTAGAGAATAATCCCACT
CTTCTTGGCGTGCATCGTGTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAGGTGTTCCGGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGAAATCTCTTT
GATATTGATTTTAACTCTGTCCATTACAGGCGAAAAGCAGATTCAGATTGTTAATTTTAAAGCAGATTATTATACAGTCAGCGTAGATG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAGGAGTCAAGGTAGCTCCTCAGACAGAGTGAAGCAGATTTTGGACAATACAGAAGTGAAGGCGGTATTTTAGGGG
GCGATCCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGTTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGATGAAT
TATCCTATGATCATCAAGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTGACGGCTCACTTTACCCTAG
TATTCCTTTAAAAGGGAATGTTCTGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAAACCGATTGGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGACTAG

>*plyA19*

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGGAGAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGGTTGTTATCGAAAGAAAGAAGCGGAGCTTGTGACAAATACAAGTGA
TATTTCTGTAACAGCTACCAACGACAGTCGCCCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTAGAGAATAATCCCACT
CTTCTTGGCGTGCATCGTGTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAGGTGTTCCGGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGAAATCTCTTT
GATATTGATTTTAACTCTGTCCATTACAGGCGAAAAGCAGATTCAGATTGTTAATTTTAAAGCAGATTATTATACAGTCAGCGTAGACG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAGGAGTCAAGGTAGCTCCTCAGACAGAGTGAAGCAGATTTTGGACAATACAGAAGTGAAGGCGGTATTTTAGGGG
GCGACCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGTTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGATGAAT
TATCCTATGATCATCAAGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTAAGCGCTCACTTTACCCTAG
TATTCCTTTAAAAGGGAATGTTCTGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAAACCGATTGGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGACTAG

>*plyA23*

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGGAGAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGGTTGTTATCGAAAGAAAGAAGCGGAGCTTGTGACAAATACAAGTGA
TATTTCTGTAACAGCTACCAACGACAGTCGCCCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTAGAGAATAATCCCACT
CTTCTTGGCGTGCATCGTGTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAGGTGTTCCGGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGAAATCTCTTT
GATATTGATTTTAACTCTGTCCATTACAGGCGAAAAGCAGATTCAGATTGTTAATTTTAAAGCAGATTATTATACAGTCAGCGTAGACG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGTCAAGTCTATCTCAAGTTGGAAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAGGAGTCAAGGTAGCTCCTCAGACAGAGTGAAGCAGATTTTGGACAATACAGAAGTGAAGGCGGTATTTTAGGGG
GCGACCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGTTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGAAATGAAT
TATCCTATGATCATCAAGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTAAGCGCTCACTTTACCCTAG
TATTCCTTTAAAAGGGAATGTTCTGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAAACCGATTGGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGACTAG

>*plyA30*

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGGAGAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGGTTGTTATCGAAAGAAAGAAGCGGAGCTTGTGACAAATACAAGTGA
TATTTCTGTAACAGCTACCAACGACAGTCGCCCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTAGAGAATAATCCCACT

CTTCTTGGCGTTCGATCGTGCTCCGATGACTTATAGTATTGATTTGCCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGGGAAGACC
CCAGCAATTCAAGTGTTCGCGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCCAGCTAG
AATCGAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGGAATTTCTCTT
GATATTGATTTTAACTCTGTCCATTACAGCGAAAAGCAGATTAGATTGTTAATTTTAAAGCAGATTATTATACAGTCAGCGTAGACG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCCTTTGGT
CTATATTGGCAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAAGAGTCAAGGTAGCTCCTCACAGCAGAGTGAAGCAGATTGGACAATACAGAAGTGAAGGCGGTATTATTAGGGG
GCGACCCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGATTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGATGAAT
TATCCTATGATCATCAAGGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTGACGGCTCACTTTACCAGTAG
TATTCCTTTAAAAGGGAATGTTTCGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAACCGATTTGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACAACCTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGATTAG

>plyA31

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGAGAAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGTGTTATCGAAAGAAAAGCGGAGCTTGTCGACAAAATACAAGTGA
TATTTCTGTAAACAGTACCAACGACAGTCGCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTTAGAGAATAATCCCACT
CTTCTTGGCGTTCGATCGTGCTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAAGTGTTCGCGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGGAATTTCTCTT
GATATTGATTTTAACTCTGTCCATTACAGCGAAAAGCAGATTAGATTGTTAATTTTAAACAGATTATTATTACAGTCAGCGTAGACG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAAGAGTCAAGGTAGCTCCTCACAGCAGAGTGAAGCAGATTGGACAATACAGAAGTGAAGGCGGTATTATTAGGGG
GCGACCCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAG
ACTAAGGTTACAGCTTACAGAAACGGAGATTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGATGAAT
TATCCTATGATCATCAAGGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTGACGGCTCACTTTACCAGTAG
TATTCCTTTAAAAGGGAATGTTTCGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAACCGATTTGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACAACCTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGACTAG

>plyA32

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGAGAAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGTGTTATCGAAAGAAAAGCGGAGCTTGTCGACAAAATACAAGTGA
TATTTCTGTAAACAGTACCAACGACAGTCGCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTTAGAGAATAATCCCACT
CTTCTTGGCGTTCGATCGTGCTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAAGTGTTCGCGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGGTCAAGTTTGGTCTGACTTTGAAAAGACAGGGGAATTTCTCTT
GATATTGATTTTAACTCTGTCCATTACAGGCAAAAAGCAGATTAGATTGTTAATTTTAAAGCAGATTATTATTACAGTCAGTGTAGATG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAAACAGAGAGGAATTTCTGCAGAGCGCTCCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
GCTTTGATAAAAAGAGTCAAGGTAGCTCCTCACAGCAGAGTGAAGCAGATTGGACAATACAGAAGTGAAGGCGGTATTATTAGGGG
GCGACCCAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCAGA
TCATCCAGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAA
ACTAAGGTTACAGCTTACAGAAACGGAGATTTACTGCTGGATCATAGTGGTGCCTATGTTGCCCAATATTATATTACTTGGGAATGAAT
TATCCTATGATCATCAAGGTAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTAACGGCTCACTTTACCAGTAG
TATTCCTTTAAAAGGGAATGTTTCGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGCTTGCCCTGGGAATGGTGGCGTACGGTTTAT
GAAAAACCGATTTGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACAACCTCTCTATCCTCAGGTAGAGGATAAGGTAGAAA
ATGACTAG

>plyA34 (1410 nt)

ATGGCAAATAAAGCAGTAAATGACTTTATACTAGCTATGAATTACGATAAAAAGAACTCTTGACCCATCAGGGAGAAAAGTATTGAAA
ATCGTTTCATCAAAGAGGGTAATCAGCTACCCGATGAGTTTGTGTTATCGAAAGAAAAGCGGAGCTTGTCGACAAAATACAAGTGA
TATTTCTGTAAACAGTACCAACGACAGTCGCTCTATCCTGGAGCACTTCTCGTAGTGGATGAGACCTTGTTAGAGAATAATCCCACT
CTTCTTGGCGTTCGATCGTGCTCCGATGACTTATAGTATTGATTTGGCTGGTTTGGCAAGTAGCGATAGCTTTCTCCAAGTGAAGACC
CCAGCAATTCAAGTGTTCGCGGAGCGGTAACGATTTGTTGGCTAAGTGGCATCAAGATTATGGTCAGGTCAATAATGTCCCAGCTAG
AATGCAGTATGAAAAATCACGGCTCACAGCATGGAACAACCTCAAGGTCAAGTTTGGTCTGACTTTGAAAAGATAGGGGAATTTCTCTT
GATATTGATTTTAACTCTGTCCATTACAGGAAAAGCAGATTAGATTGTTAATTTTAAAGCAGATTATTATTACAGTCAGCGTAGATG
CTGTTAAAAATCCAGGAGATGTGTTTCAAGATACTGTAACGGTAGAGGATTTAAGGCAGAGAGGAATTTCTGCAGAGCGCTCCTTTGGT
CTATATTTCCGAGTGTGCTTATGGGCGCCAAGTCTATCTCAAGTTGGAACCACGAGTAAGAGTGATGAAGTAGAGGCTGCTTTTGAA
TCTTTGATAAAAAGGAGTAGCTCCTCACAGCAGAGTGAAGCAGATTGGACAATACAGAAGTGAAGGCGGTATTATTAGGGGGCGACC
CAAGTTCGGGTGCCGAGTTGTAACAGGCAAGGTGGATATGGTAGAGGACTTGATTCAAGAAGGCAGTCGCTTTACAGCCGATCATCC
AGGCTTGCCGATTTCTTATACAACCTCTTTTTTACGTGACAATGTAGTTGCGACCTTTCAAACAGTACAGACTATGTTGAGACTAAG
GTTACAGCTTACAGAAACGGAGATTTACTGCTGGATCATAGTGGTGCCTATGTTGCTCAATATTATATTACTTGGGATGAATTAATCCT
ATGATCATCAAGGCAAGGAAGTCTTGACTCCTAAGGCTTGGGACAGAAATGGGCAGGATTTAACGGCTCACTTTACCAGTAGTATTC
TTTTAAAAGGGAATGTTTCGTAATCTCTCTGTCAAATTAGAGAGTGTACCGGCTTGCCCTGGGAATGGTGGCGTACGGTTTATGAAAA
ACCGATTTGCCACTAGTGCCTAAGCGGACGATTTCTATTTGGGGAACAACCTCTCTATCCTCAGGTAGAGGATAAGGTAGAAAATGACT
AG

Amino acid sequence of the novel PlyA_{Spn}-20 allele

>PlyA20

MANKAVNDFILAMNYDKKKLLTHQGESIENRFIKEGNQLPDEFVVIERKKRSLSTNTSDISVTATNDSRLYPGALLVVDETLLENNPT
LLAVDRAPMTYSIDLPLGLASSDSFLQVEDPSNSSVRGAVNDLLAKWHQDYGQVNNVPHARMQYEKITAHSMEQLKVKFGSDFEKTGNSL
DIDFNSVHSGEKQIQIVNFKQIYYTVSVDVKNPGDVFQDTVTVEDLKQRGISAERPLVYIASVAYGRQVYLKLETTSKSDEVEAAFE
ALIKGVKQVAPQTEWKQILDNTEVKAVILGGDPSSGARVVTGKVDMMVEDLIQEGSRFTADHPGLPISYTTSFRLRDNVVATFQNSTDYVE
TKVTAYRNGDLLLDHSGAYVAQYYITWDELSYDHQGKEVLTPKAWDRNGQDLTAHFTTISIPKGNVRNLSVKIRECTGLAWEWRTVY
EKTDLPLVRKRTISIWGTTLYPQVEDKVEN

This novel allele differs from PlyA1 by a S238A conserved substitution.

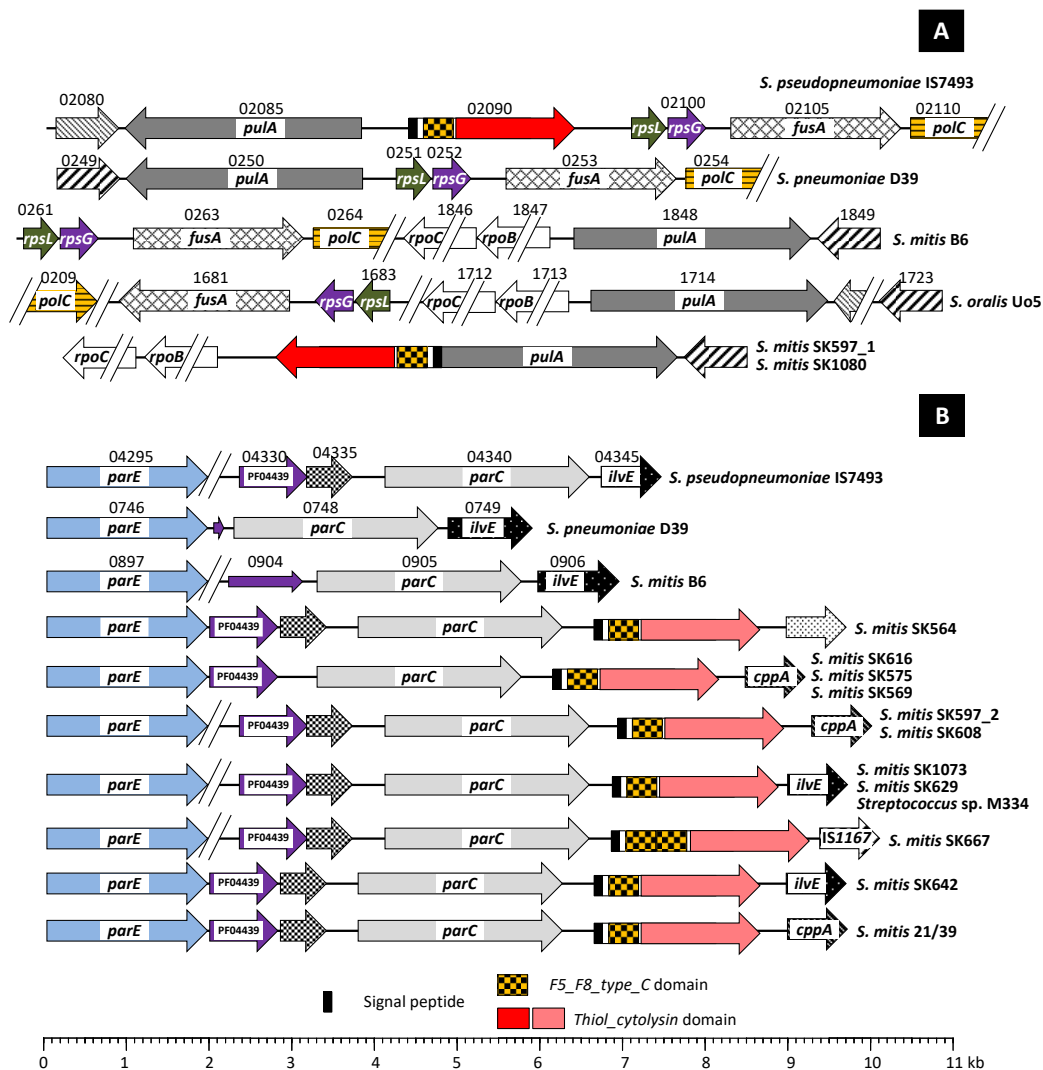


Fig. S1. Diagram of SMG genes homologous to *S. pseudopneumoniae* SPPN_02090 (*IlyA1* and *IlyA2* subfamilies).

A. With the exception of this gene, *S. pneumoniae* and *S. pseudopneumoniae* genomes are syntenic in this region, whereas disparities are characteristic of *S. mitis* and *S. oralis* genomes. A SPPN_02090 ortholog is also present upstream of *pulA* in the *S. mitis* strains SK597 and SK1080.

B. A different subfamily of SPPN_02090 orthologous genes (*IlyA2*) is located downstream of *parC* in many different *S. mitis* strains (also including SK597). Regions showing $\geq 90\%$ sequence identity are represented by identical color and shading.

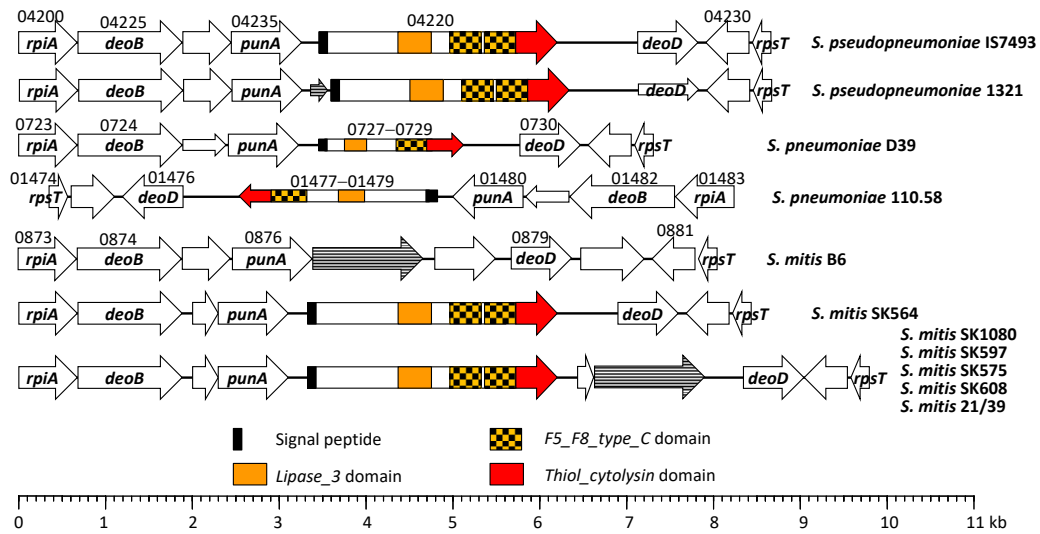


Fig. S2. Diagram of SMG genes (and gene products) homologous to *S. pseudopneumoniae* SPPN_04220 (*IlyB* subfamily).

Thin arrows represent interrupted genes (pseudogenes). Regions showing $\geq 90\%$ sequence identity are represented by identical color and/or shading.

S. mitis

B6
NCTC 12261
11/5
13/39
17/34
18/56
29/42
F0392
SK95
SK321
SK569
SK575
SK579
SK616

S. oralis

Uo5
ATCC 35037
ATCC 49296
SK10
SK100
SK313

S. infantis

ATCC 700779
X

S. tigurinus

AZ 3A
1366
2425
2426

Streptococcus sp.

F0407
GMD1S
GMD2S
GMD4S
GMD6S
SK643
SPAR10

B6
NCTC 12261
11/5
13/39
17/34
18/56
29/42
F0392
SK95
SK321
SK569
SK575
SK579
SK616
Uo5
ATCC 35037
ATCC 49296
SK10
SK100
SK313
ATCC 700779
X
AZ 3A
1366
2425
2426
F0407
GMD1S
GMD2S
GMD4S
GMD6S
SK643
SPAR10

B6
NCTC 12261
11/5
13/39
17/34
18/56
29/42
F0392
SK95
SK321
SK569
SK575
SK579
SK616
Uo5
ATCC 35037
ATCC 49296
SK10
SK100
SK313
ATCC 700779
X
AZ_3A
1366
2425
2426
F0407
GMD1S
GMD2S
GMD4S
GMD6S
SPAR10
SK643

B6
NCTC 12261
11/5
13/39
17/34
18/56
29/42
F0392
SK95
SK321
SK569
SK575
SK579
SK616
Uo5
ATCC 35037
ATCC 49296
SK10
SK100
SK313
ATCC 700779
X
AZ_3A
1366
2425
2426
F0407
GMD1S
GMD2S
GMD4S
GMD6S
SPAR10
SK643

-----TAGCAAATACAGCTAGGATTGCACTTTTGTGGCCAAGATAGCCTTG
-----CAGTAAATACAACTAGGATAGCACTTTAGTTGCTAAATTTCCCTTG
-----CAGTAAATACAACTAGGATAGCACTTTAGTTGCTAAATTTCCCTTG
-----TAGCAAATACAGCTAGGATGAGCACTTTAATTGCCAAGATGCACTTG
-----CAGTAAATCAAGCTAGGATGACACTTTTGTGGCCAAGATGCCTTG
-----TAGCAAATCAAGCTAGGATGGTACTTTTGTACCAAGATTTCCCTTG
-----CAGTAAATCAAGCTAGGATAGCACTTTAGTTGCTAAATTTCCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTTCCAGATTTCCCTTG
-----TCCTAAATCCAGCTAGGATTATCCTTTAGCTGCCAAGATTGCCTTG
-----CAGTAAATCCAACTAGGATAGCACTTTGGTTGCCAAGATGACCTTA
-----TAGTAAATACAGCTAGGATTGCACTTTGGTTACCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATGACAGTTTGTGGCCAAGATGCCTTG
-----TAGCAAATACAGCTAGGATATCACCTTTGGTTGCCAAGATTTCCCTTG
TTTTTCGAAGAGTATTAGCAAATCCAGCTAGGAGTGTCTTTTAGCAGCCAAGACTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
TTTTTCGAAGAGTATTAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTT
-----CAGCAAATCCAGCTAGGATTGTCCTTTGTGGCCAAGATTGCCTTG
-----TTGCAAACCAAGCTAGGCTTTGCTCTCCGACTGCCAGATTTCTTG
-----TTATAAATGAAGCTAGGATCGTCCGCCGACAGCCAAGATTTCTTG
-----TTATAAATGAAGCTAGGATTGTCGCCGACAGCCAAGATTTCTTG
-----TTGCAAAGCAAGCTAGGATTTCCCTCCGACTGCCAGATTTCTTG
-----TTGCAAAGCAAGCTAGGATTTCCCTCCGACTGCCAGATTTCTTG
-----TTGCAAAGCAAGCTAGGATTTCCCTCCGACTGCCAGATTTCTTG
-----TTGCAAAGCAAGCTAGGATTTCCCTCCGACTGCCAGATTTCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----CAGCAAATCCAGCTAGGATTGTCCTTTGGCTGCCAAGATTGCCTTG
-----TTATAAATGAAGCTAGGATTGTCGCCGACAGCCAAGATTTCTTG
-----CAGCAAATCCAGCTAGGATGACACTTTTGTGGCCAAGATGCCTTG
*** * ** * * *

```

B6          CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGGATAGTCATGGCAACC
NCTC 12261  CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
11/5       CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
13/39      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
17/34      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
18/56      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
29/42      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
F0392     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK95       CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK321     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK569     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK575     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK579     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK616     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
Uo5        CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
ATCC 35037 CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
ATCC 49296 CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK10       CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK100     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK313     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
ATCC 700779 CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
X          CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
AZ_3A     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
1366      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
2425      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
2426      CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
F0407     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
GMD1S     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
GMD2S     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
GMD4S     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
GMD6S     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SK643     CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
SPAR10    CCTTCCTTTTATCAAAGGGTGACGAAACAGTGAGAAATACAGTTGAATGGTCATGGCAACC
** ***** ** ** * * * * * * * * * * * * * * * * * * * * * * * *

```

Fig. S3. Multiple alignment of the DNA region located between the 3' end of *dinF* and the gene located immediately downstream in selected SMG lacking a *lytA*_{SMG} gene. The orientation of the sequences corresponds to that in the genome of the *S. pneumoniae* D39 strain. The nucleotides that match those of the *p*/REP copy overlapping the 3' end of *dinF* are shown in a black background. Asterisks indicate identical nucleotides in all the sequences. Hyphens indicate gaps introduced to maximize similarity. The termination codon of *dinF* is marked in a red background whereas the initiation codon of the downstream gene is highlighted in a green background. In *S. oralis* strains Uo5 and ATCC 49296, an intervening BOX element is present (boxC is labeled in blue; boxA is boldfaced and labeled in gray).

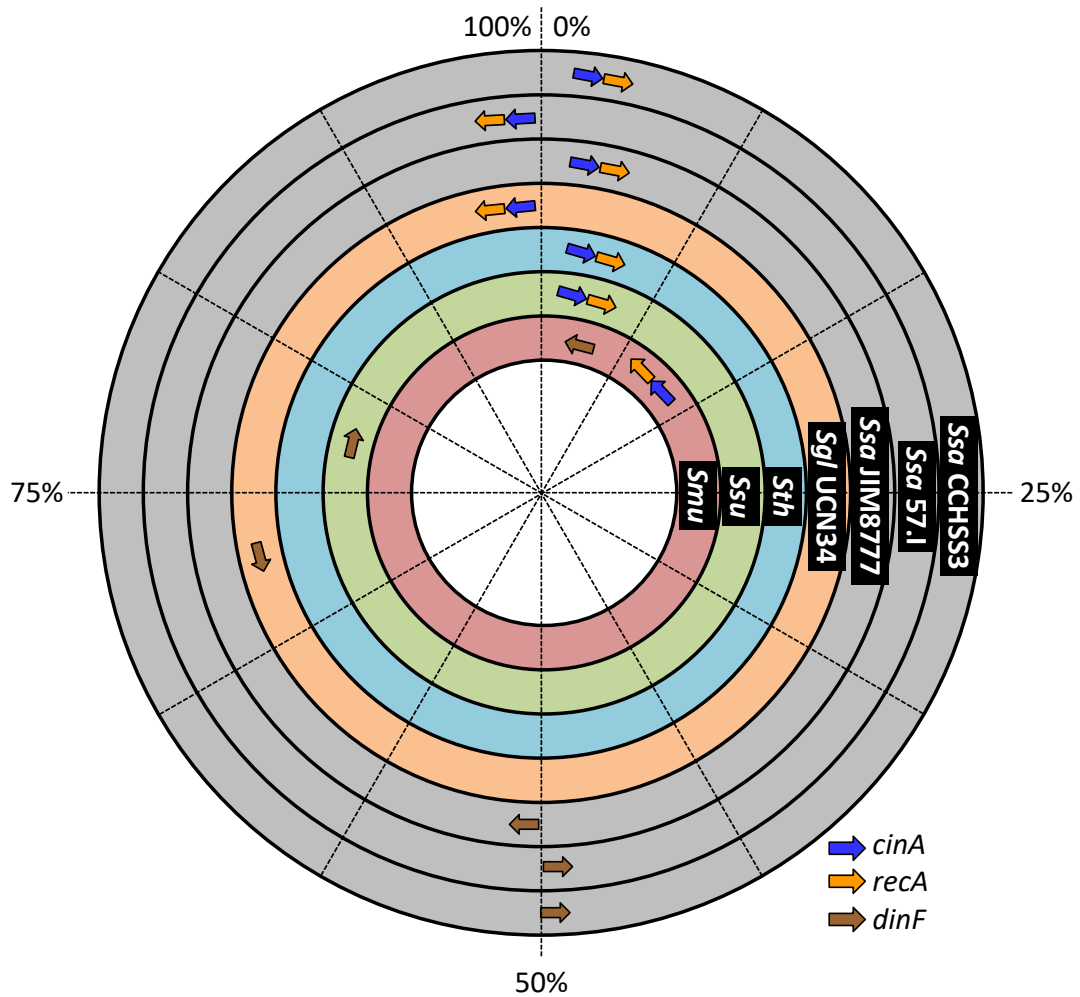


Fig. S4. Diagram of the approximate location and orientation of genes *cinA*, *recA*, and *dinF* in the chromosomes of different streptococci.

Only complete genomes were analyzed. The first gene of each genome is *dnaA*. The positions were calculated as the percentage of the respective genome. Interestingly, *S. thermophilus* strains lack a functional *dinF* gene. Note that in *S. salivarius*, the location and orientation of the three genes appear to be strain-specific. *Ssa*, *S. salivarius*; *Sgl*, *S. gallolyticus*; *Sth*, *S. thermophilus*; *Ssu*, *S. suis*; *Smu*, *S. mutans*.

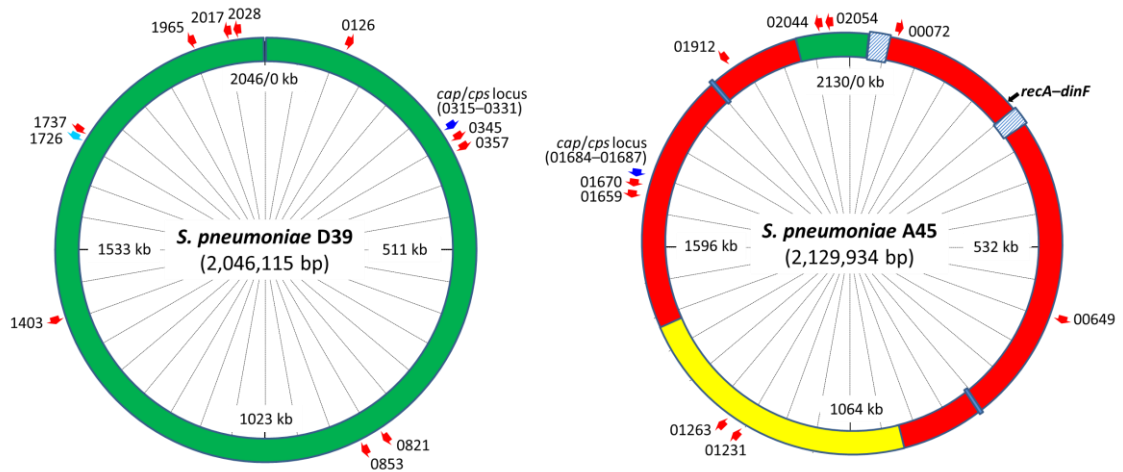


Fig. S5. Chromosomal rearrangement in the serotype 3 equine *S. pneumoniae* A45 strain.

The genome of the D39 strain is shown for comparison. The *plyA* gene and the capsular cluster (*cap/cps*) are shown in light and dark blue respectively. Several genes encoding choline-binding proteins are shown as red arrows that indicate the direction of transcription. Genes are identified as ordered in D39 (prefix SPD_)/A45 (SPNA45_) and correspond to the following: 0126/01912 (*pspA*); 0345/01670 (*cbpC*); 0357/01659 (*cbpF*); 0821/01231 (*cbpE*); 0853/01263 (*lytB*); 1403/00649 (*lytC*); 1726/– (*plyA*); 1737/– (*lytA_{spn}*); 1965/00072 (*pcpA*); 2017/02044 (*cbpA*); 2028/02054 (*cbpD*). The open reading frames SPNA45_01659 and SPNA45_01670 appear to correspond to pseudogenes. The chromosomal fragments of strain A45 that are inverted with respect to D39 are labeled in red. The chromosomal fragment that has been translocated (but not inverted) is marked in yellow. Hatched and blue boxes correspond to complete or defective prophages respectively.

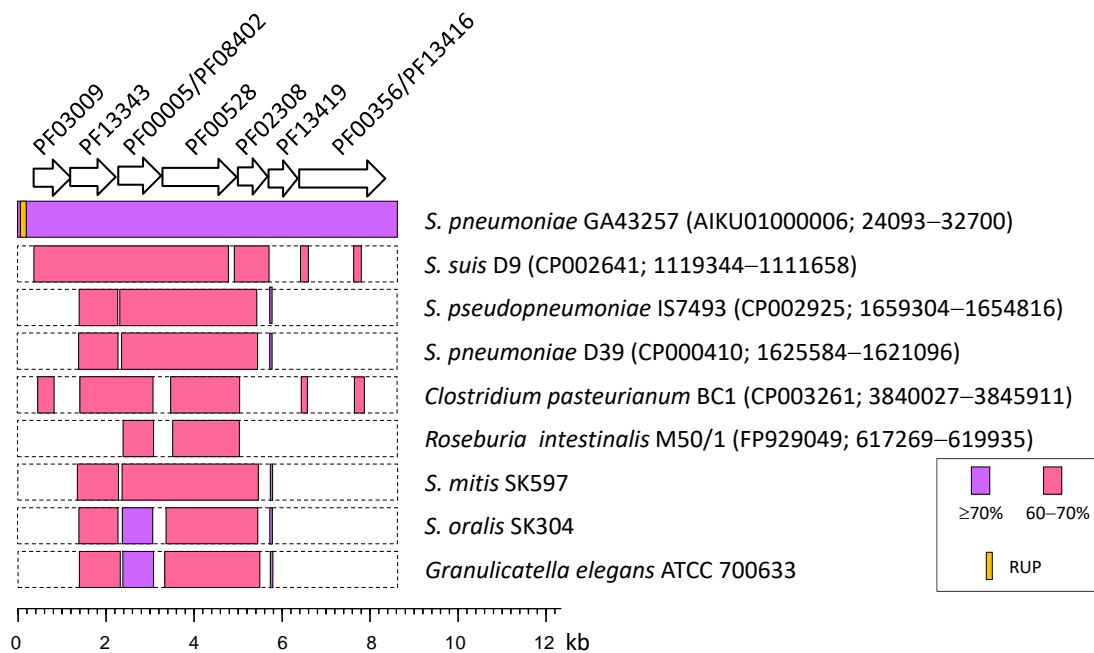


Fig. S6. Alignment of the 8.6 kb DNA fragment found in the *plyA*–*lytA* island of various pneumococcal isolates with the sequences deposited in the databases.

For complete genomes, the accession no. and nucleotide positions are indicated in parentheses. The inset shows the percentages of nucleotide similarity compared with the fragment of strain GA43257. Predicted proteins were compared with the Pfam database: PF03009, *GDPD* (glycerophosphoryl diester phosphodiesterase family); PF13343, *SBP_bac_6* (bacterial extracellular solute-binding protein); PF00005, *ABC_tran* (ATP-binding domain of ABC transporters); PF08402, *TOBE_2* (TOBE domain); PF00528, *BPD_transp_1* (binding-protein-dependent transport system inner membrane component); PF02308, *MgtC* (MgtC family); PF13419, *HAD_2* (haloacid dehalogenase-like hydrolase); PF00356, *LacI* (bacterial regulatory proteins, lacI family); PF13416, *SBP_bac_8* (bacterial extracellular solute-binding protein).

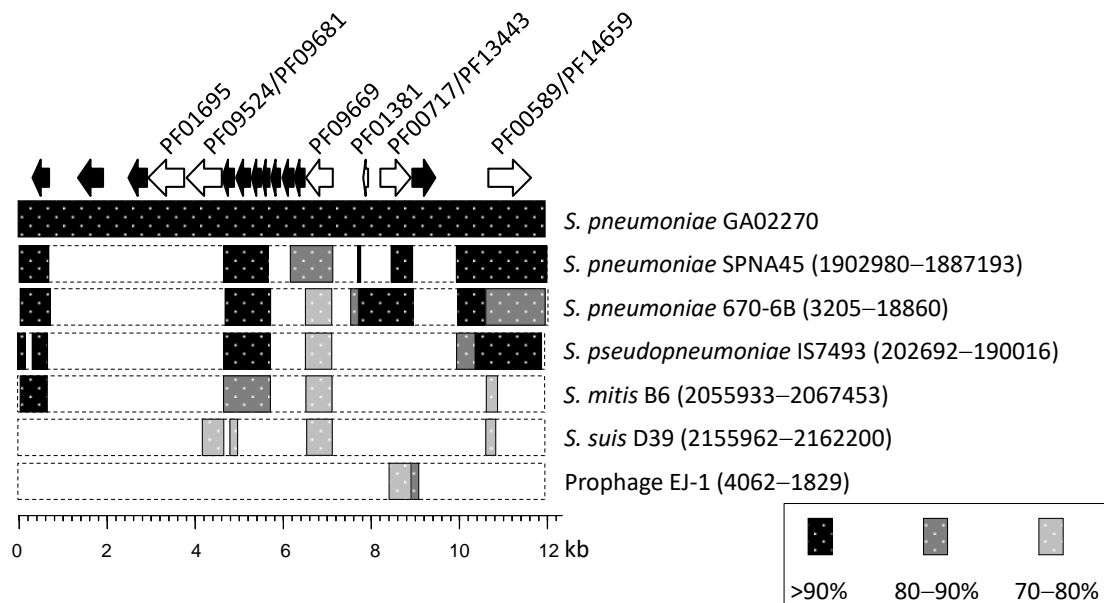


Fig. S7. Alignment of the putative, defective prophage found in the *plyA–lytA* island of various pneumococcal isolates with the sequences deposited in the databases.

Potential phage genes are also shown as arrows indicating the direction of transcription. Blackened arrows account for genes encoding hypothetical proteins. The inset shows the percentages of nucleotide similarity compared with the defective prophage of strain GA02270. Predicted proteins were compared with the Pfam database: PF01695, *IstB_IS21* (IstB-like ATP binding protein); PF09524, *Phg_2220_C* (Conserved phage C-terminus); PF09681, *Phage_rep_org_N* (N-terminal phage replisome organiser); PF09669, *Phage_pRha* (Phage regulatory protein Rha); PF01381, *HTH_3* (Helix-turn-helix); PF00717, *Peptidase_S24* (Peptidase S24-like); PF13443, *HTH_26* (Helix-turn-helix); PF00589, *Phage_integrase* (Phage integrase family); PF14659, *Phage_int_SAM_3* (Phage integrase, N-terminal SAM-like domain).

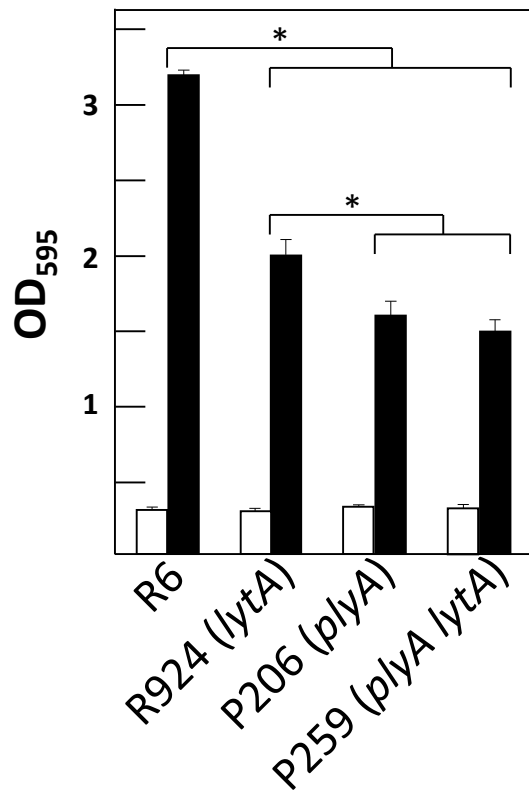


Fig. S8. Biofilm formation capacity of *lytA*, *plyA*, and *plyA lytA* mutants of the *S. pneumoniae* strain R6.

The non-encapsulated R6 strain was used as a control. Cells were grown in C medium for 6 h at 34°C. Open and filled bars indicate growth and biofilm formation, respectively. In all panels, the results represent the means \pm standard errors from at least four independent experiments, each performed in triplicate. *, $P < 0.05$ (compared to the R6 or R924 strains; ANOVA with *post-hoc* tests).

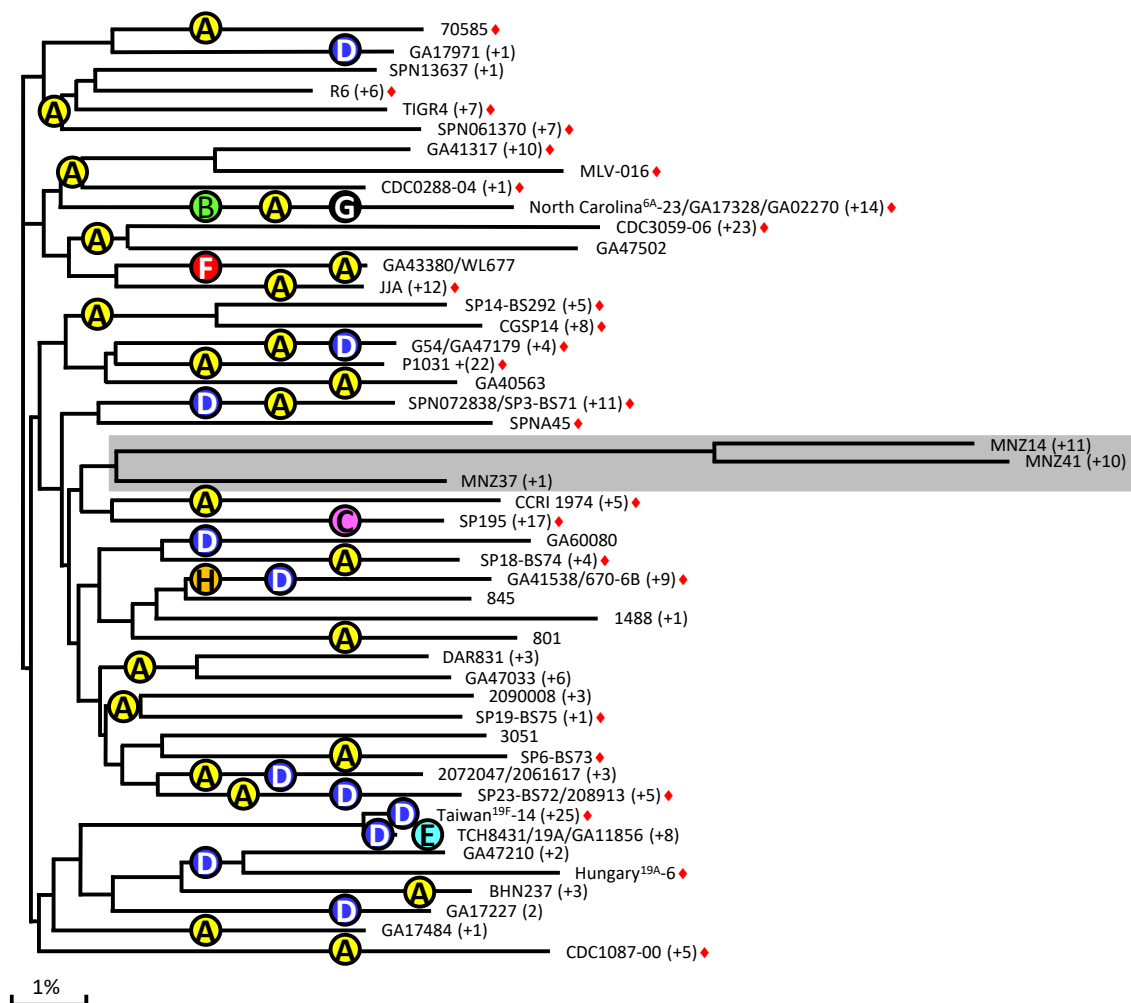


Fig. S9. Dendrogram of *S. pneumoniae* genomes showing the arrangement distribution of the *plyA*—*lytA* island.

The number of additional genomes represented in each branch is shown in parentheses. Red diamonds indicate the location of the 44 *S. pneumoniae* genomes analyzed by Donati et al. (2010). Group II non-Ec-*Spn* are indicated in a gray background. The figure was slightly modified from <http://www.ncbi.nlm.nih.gov/genome/?term=CP000410> (Tatusova, et al. 2014). The names of the different arrangements (A—H) are as in Fig. 6.

References

- Andrewes FW, Horder TJ 1906. A study of the streptococci pathogenic for man. *Lancet*. 168:708–713.
- Arbique JC, et al. 2004. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol*. 42:4686–4696.
- Bridge PD, Sneath PHA 1982. *Streptococcus gallinarum* sp. nov. and *Streptococcus oralis* sp. nov. *Int J Syst Bacteriol*. 32:410–415.
- Camelo-Castillo A, Benítez-Páez A, Belda-Ferre P, Cabrera-Rubio R, Mira A 2014. *Streptococcus dentisani* sp. nov., a new member of the mitis group. *Int J Syst Evol Microbiol*. 64:60–65.
- Chester FD. 1901. A manual of determinative bacteriology. New York: The MacMillan Co.
- Donati C, et al. 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol*. 11:R107.
- Glazunova OO, Raoult D, Roux V 2006. *Streptococcus massiliensis* sp. nov., isolated from a patient blood culture. *Int J Syst Evol Microbiol*. 56:1127–1131.
- Hanahan D 1983. Studies on transformation of *E. coli* with plasmids. *J Mol Biol*. 166:557–580.
- Handley P, Coykendall A, Beighton D, Hardie JM, Whiley RA 1991. *Streptococcus crista* sp. nov., a viridans streptococcus with tufted fibrils, isolated from the human oral cavity and throat. *Int J Syst Bacteriol*. 41:543–547.
- Huch M, et al. 2013. *Streptococcus rubneri* sp. nov., isolated from the human throat. *Int J Syst Evol Microbiol*. 63:4026–4032.
- Kawamura Y, et al. 1998. *Streptococcus peroris* sp. nov. and *Streptococcus infantis* sp. nov., new members of the *Streptococcus mitis* group, isolated from human clinical specimens. *Int J Syst Bacteriol*. 48:921–927.

- Kilian M, Mikkelsen L, Henrichsen J 1989. Taxonomic study of viridans streptococci: description of *Streptococcus gordonii* sp. nov. and emended descriptions of *Streptococcus sanguis* (White and Niven 1946), *Streptococcus oralis* (Bridge and Sneath 1982), and *Streptococcus mitis* (Andrewes and Horder 1906). *Int J Syst Bacteriol.* 39:471–484.
- Lacks S, Hotchkiss RD 1960. A study of the genetic material determining an enzyme activity in *Pneumococcus*. *Biochim Biophys Acta.* 39:508–518.
- Martín V, Mañes-Lázaro R, Rodríguez JM, Maldonado-Barragán A 2011. *Streptococcus lactarius* sp. nov., isolated from breast milk of healthy women. *Int J Syst Evol Microbiol.* 61:1048–1052.
- Moscoso M, Claverys JP 2004. Release of DNA into the medium by competent *Streptococcus pneumoniae*: kinetics, mechanism and stability of the liberated DNA. *Mol Microbiol.* 54:783–794.
- Moscoso M, Domenech M, García E 2010. Vancomycin tolerance in clinical and laboratory *Streptococcus pneumoniae* isolates depends on reduced enzyme activity of the major LytA autolysin or cooperation between CiaH histidine kinase and capsular polysaccharide. *Mol Microbiol.* 77:1052–1064.
- Moscoso M, García E, López R 2006. Biofilm formation by *Streptococcus pneumoniae*: role of choline, extracellular DNA, and capsular polysaccharide in microbial accretion. *J Bacteriol.* 188:7785–7795.
- Mosser JL, Tomasz A 1970. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *J Biol Chem.* 245:287–298.
- Romero P, López R, García E 2004. Characterization of LytA-like *N*-acetylmuramoyl-L-alanine amidases from two new *Streptococcus mitis* bacteriophages provides insights into the properties of the major pneumococcal autolysin. *J Bacteriol.* 186:8229–8239.
- Sambrook J, Russell DW. 2001. *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

- Sánchez-Puelles JM, Sanz JM, García JL, García E 1992. Immobilization and single-step purification of fusion proteins using DEAE-cellulose. *Eur J Biochem.* 203:153–159.
- Tabor S. 1990. Expression using the T7 RNA polymerase/promoter system. In: Ausubel FM, et al., editors. *Current protocols in molecular biology*. New York: Green Publishing Associates Inc. and John Wiley & Sons. p. 16.12.11–16.12.11.
- Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I 2014. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 42:D553–D559.
- Tong H, Gao X, Dong X 2003. *Streptococcus oligofermentans* sp. nov., a novel oral isolate from caries-free humans. *Int J Syst Evol Microbiol.* 53:1101–1104.
- Usobiaga P, et al. 1996. Structural organization of the major autolysin from *Streptococcus pneumoniae*. *J Biol Chem.* 271:6832–6838.
- Whiley RA, et al. 1990. *Streptococcus parasanguis* sp. nov., an atypical viridans *Streptococcus* from human clinical specimens. *FEMS Microbiol Lett.* 68:115–122.
- White JC, Niven CF, Jr. 1946. *Streptococcus* s.b.e.: a streptococcus associated with subacute bacterial endocarditis. *J Bacteriol.* 51:717–722.
- Willcox MD, Zhu H, Knox KW 2001. *Streptococcus australis* sp. nov., a novel oral streptococcus. *Int J Syst Evol Microbiol.* 51:1277–1281.
- Woo PCY, et al. 2002. *Streptococcus sinensis* sp. nov., a novel species isolated from a patient with infective endocarditis. *J Clin Microbiol.* 40:805–810.
- Zbinden A, et al. 2012. *Streptococcus tigurinus* sp. nov., isolated from blood of patients with endocarditis, meningitis and spondylodiscitis. *Int J Syst Evol Microbiol.* 62:2941–2945.
- Zhang M, et al. 2013. Isolation and characterization of *Streptococcus troglodytidis* sp. nov., from a foot abscess of a chimpanzee (*Pan troglodytes*). *Int J Syst Evol Microbiol.* 63:449–453.