Supplemental Figure S1. Organization of pneumococcal CPS gene clusters and their surrounding regions. Type-specific genes are shadowed in gray whereas genes unrelated to CPS biosynthesis are showed as open arrows. Thick and thin arrows represent genes or pseudogenes, respectively. With the exception of type-specific genes, regions showing more than 90% identical nucleotides when comparing sequences of different serotypes, are represented by identical color and shading. Curved arrows indicate promoter regions. The inverted matchsticks represent putative transcriptional terminators.

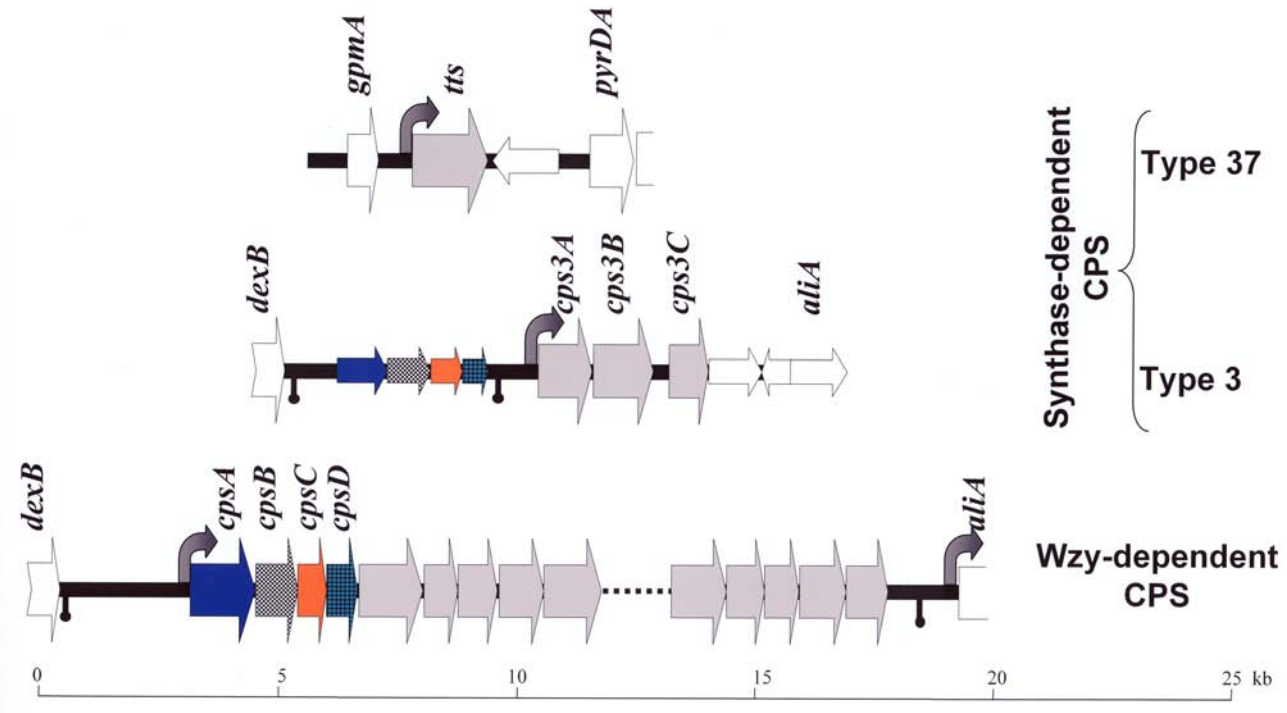
Supplemental Figure S2. Schematic representation of the different sequence organizations (SO) of the DNA region spanning from the termination codon of *dexB* to the initiation codon of *cpsA* in Wzy-dependent CPS of *S. pneumoniae*. DNA regions present in some (but not all) sequence organizations are depicted as blue boxes, solid arrows, etc. Identical shading represents DNA regions showing more than 95% identical nucleotides. Small, black and red vertical bars inside a green rectangle indicate, respectively, the position of the −35 and −10 promoter boxes. The solid straight arrows correspond to a 163-bp fragment of an *aliB*-like gene.

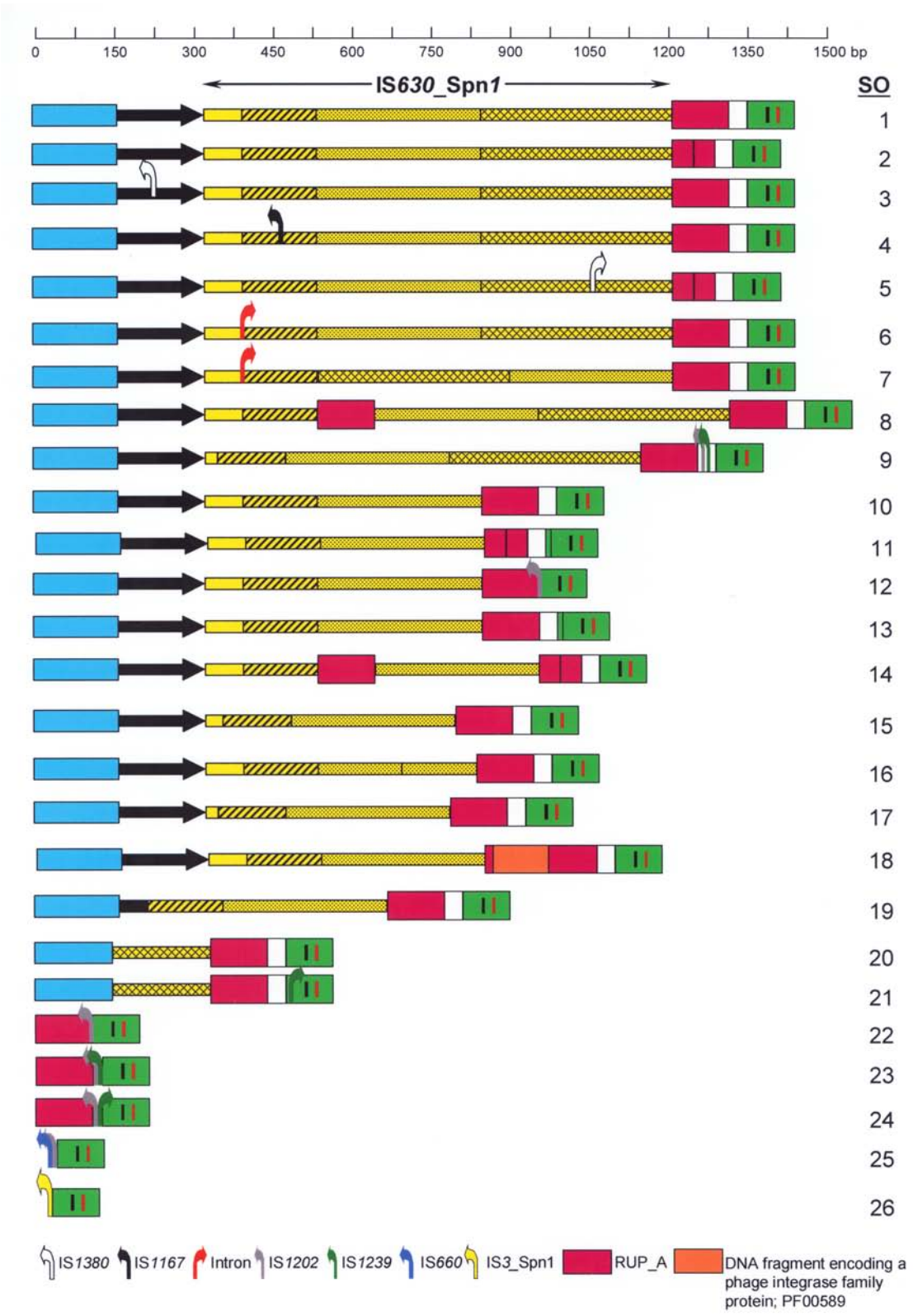
Supplemental Figure S3. Multiple alignment of the conserved sequences located upstream of *cpsA* (A and B) or *cps3A* (C). Multiple gene sequence alignments were conducted using the CLUSTAL W 2.0 program (http://www.ebi.ac.uk/Tools/clustalw2/). Each sequence is identified by its accession number. Serotypes are indicated in parentheses. Asterisks indicate nucleotides identical in all the sequences. Polymorphic positions are shadowed with different colors. Hyphens correspond to deletions introduced to maximize similarity. Eighty six (A), 115 (B), or 4 sequences (C) were aligned. In B and C, the −35 ad −10 boxes of the *cps* promoter are boxed and the transcription initiation site is indicated with white lettering in a black background. The T to C transition at position −8 in the unencapsulated R6 strain (AE008412) has been recently confirmed.[1] However, the existence of minor errors in some sequences cannot be completely discarded. The initiation codon of *cpsA* and *cps3A* is shadowed in gray.

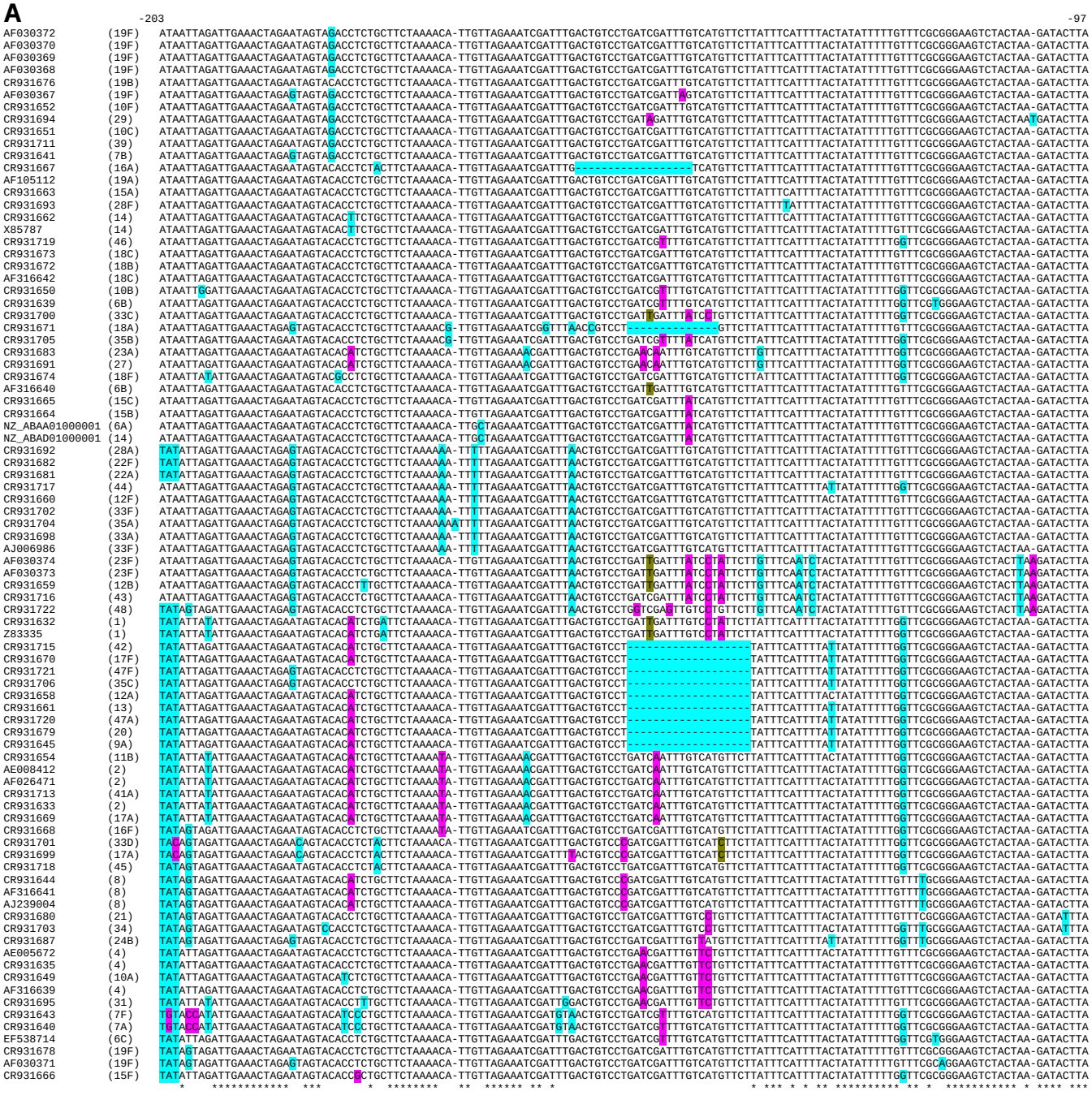
Supplemental Figure S4. Organization of the CPS gene cluster of type 3 pneumococci of the DNA region spanning from the termination codon of *dexB* to the initiation codon of *cps3A*. Regions showing more than 90% identical nucleotides are represented by identical color and shading. Thick and thin arrows represent genes or pseudogenes, respectively. The sequence organization of SO_1 serotypes is shown for comparison. Small, black and red vertical bars indicate, respectively, the position of the −35 and −10 promoter boxes.

References

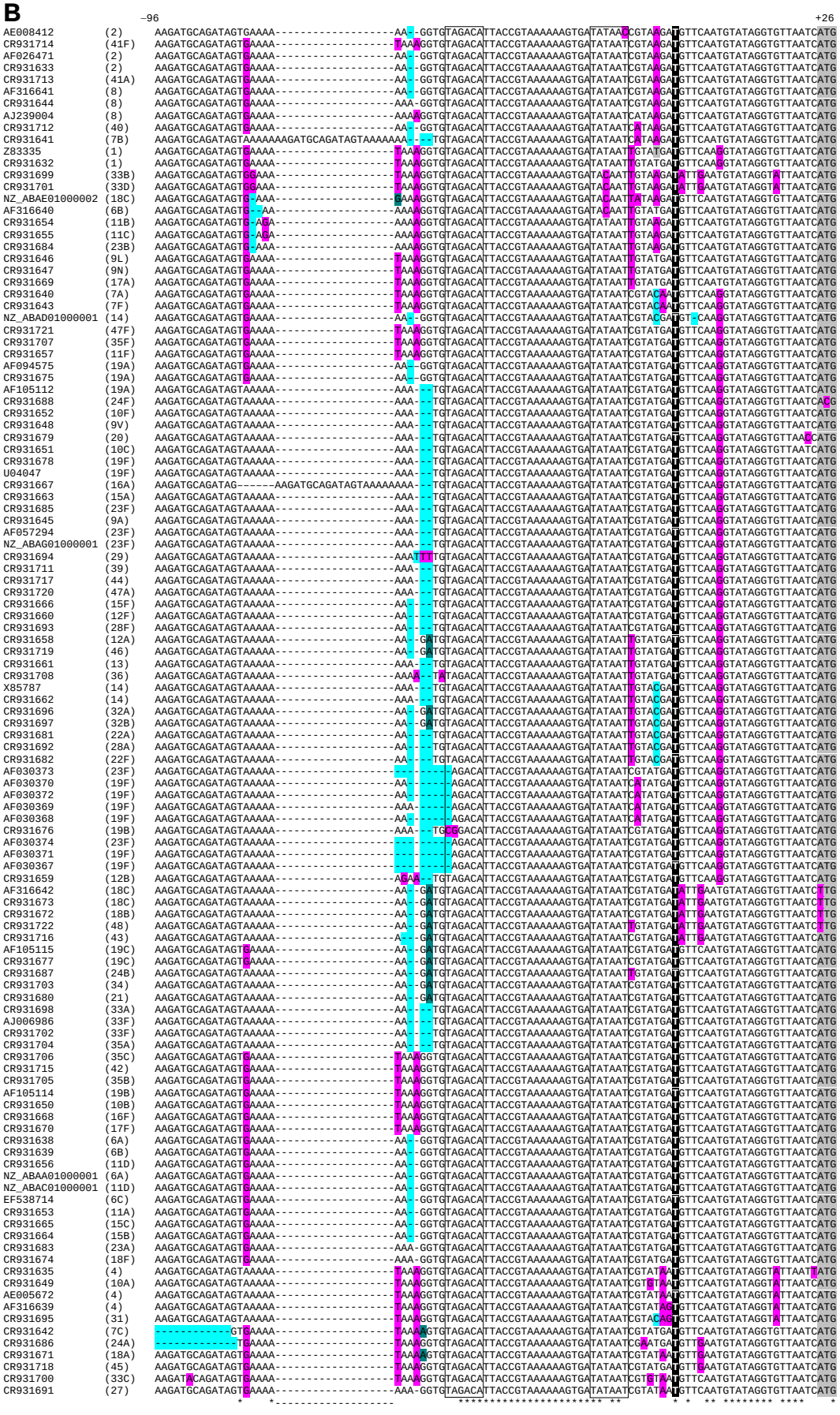
1. Lanie, J. A., Ng, W.-L., Kazmierczak, K. M., *et al*. 2007, Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6, *J. Bacteriol.*, **189,** 38−51.


gpmA
tts
pyrDA
dexB
cps3A
cps3B
cps3C
aliA
Synthase-dependent CPS
Type 37
Type 3
dexB
cpsA
cpsB
cpsC
cpsD
aliA
Wzy-dependent CPS
0
5
10
15
20
25 kb


0
150
300
450
600
750
900
1050
1200
1350
1500 bp
IS630_Spn1
SO
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
IS1380
IS1167
Intron
IS1202
IS1239
IS660
IS3_Spn1
RUP_A
DNA fragment encoding a phage integrase family protein; PF00589


A
-203
-97

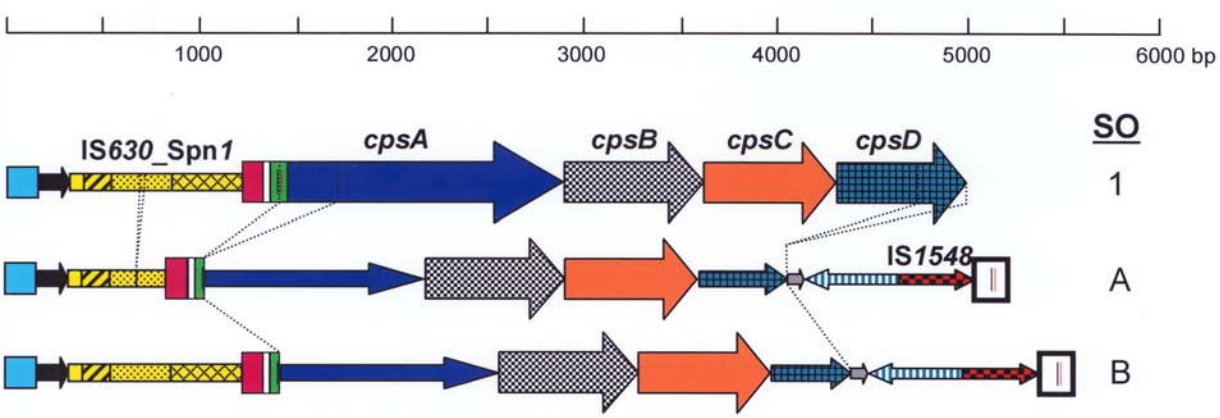
B



C

```
Z47210    −121   GGAAACAGAGGTTAGGAAAGTAATCAGTTTAACGGGATATCTTTTCAAAGCTGATACTAAGGCACAAAAAAAAGTTTGATATTCCCC  −35
CR931634         GGAAACAGAGGTTAGGAAAGTAATTAGTTTAACGGGATATCTTTTCAAAGCTGATACTAAGGCAC-AAAAAAAGTTTGATATTCCCC
U15171           GGAAACAGAGGTTAGGAAAGTAATCAG-TTAACGGGATATCTTTTCAAAGCTGATACTAAGGCACAAAAAAAAGTTTGATATTCCCC
NZ_AAZZ01000001  GGAAACAGAGGTTAGGAAAGTAATTAGTTTAACGGGATATCTTTTCAAAGCTGATACTAAGGCAC-AAAAAAAGTTTGATATTCCCC
                 ********************** ** ************************************ ********************

Z47210     −34   TTGACAATAGATAAAATTATTATATAATTAAACTATTGCTTTTTAAATAAAGTGAGAATATTAATAATGCAGAGAAAGAGGACTGTAGTAAAATG   +61
CR931634         TTGACAATAGATAAAATTATTATATAATTAAACTATTGCTTTTTAAATAAAGTGAGAATATTAATAATTCAGAGAAAGAGGACTGTAGTAAAATG
U15171           TTGACAATAGATAAAATTATTATATAATTAAACTATTGCTTTTTAAATAAAGTGAGAATATTAATAATGCAGAGAAAGAGGACTGTAGTAAAATG
NZ_AAZZ01000001  TTGACAATAGATAAAATTATTATATAATTAAACTATTGCTTTTTAAATAAAGTGAGAATATTAATAATTCAGAGAAAGAGGACTGTAGTAAAATG
                 ************************************************************************ ************************
```


1000
2000
3000
4000
5000
6000 bp
IS630_Spn1
cpsA
cpsB
cpsC
cpsD
IS1548
SO
1
A
B

Supplemental Table S1. *cps* gene clusters of *S. pneumoniae* analyzed in this study

Serotype	Strain[a]	Accession No.	SO[b]
1	13868	Z83335	4
	519/43	CR931632	4
2	R6	AE008412	1
	D39	AF026471	1
	pn2L	CR931633	8
3[c]	406	Z47210	A
	524/62	CR931634	B
	SP3-BS71	NZ_AAZZ01000001	A
4	TIGR4	AE005672	3
	WCH35	AF316639	1
	600/62	CR931635	1
6A	34351=Rodriques	CR931638	18
	SP6-BS73	NZ_ABAA01000001	10
6B	WCH18	AF316640	10
	2626/39	CR931639	10
6C	CHPA388	EF538714	1
7A	2040/37	CR931640	1
7B	Johnson	CR931641	13
7C	Sutcliff	CR931642	22
7F	554/62	CR931643	1
8	WCH56	AF316641	1
	6028/1995	AJ239004	1
	573/62	CR931644	1
9A	Wilder	CR931645	2
9L	T9233/128/68	CR931646	22
9N	533/62	CR931647	22
9V	980/68	CR931648	22
10A	10061/38	CR931649	17
10B	423/82	CR931650	10
10C	Gro Norge	CR931651	15
10F	34355	CR931652	10
11A	1813/39	CR931653	25
11B	8087/40	CR931654	1
11C	Eddy nr. 53	CR931655	24
11D	79/86	CR931656	22
	SP11-BS70	NZ_ABAC01000001	22
11F	34356	CR931657	22
12A	559/66	CR931658	5
12B	Gambia 1/81	CR931659	10
12F	6312	CR931660	10
13	34357	CR931661	2
14	NCTC 11902	X85787	10
	34359	CR931662	10
	SP14-BS69	NZ_ABAD01000001	10
15A	389/39	CR931663	10
15B	7904/39	CR931664	10

15C	553/62	CR931665	10
15F	688/63	CR931666	1
16A	R105	CR931667	11
16F	34361	CR931668	1
17A	4704	CR931669	20
17F	Rose	CR931670	2
18A	8609/43	CR931671	14
18B	1033/41	CR931672	10
18C	WCH94	AF316642	10
	4593/40	CR931673	10
	SP18-BS74	NZ_ABAE01000002	26
18F	Gethens	CR931674	1
19A	1777/39	AF094575	22
	19A2	AF105112	10
	141/68	CR931675	22
19B	–	AF105114	22
	4594	CR931676	10
19C	–	AF105115	23
	7588/39	CR931677	23
19F	SSZ	U04047	12
	NCTC 11906	AF030367	7
	SP-496	AF030368	10
	SP-VA92	AF030369	10
	SP-GA71	AF030370	10
	PO-329	AF030371	8
	SP-VA96	AF030372	10
	485/61	CR931678	6
20	34365	CR931679	2
21	546/62	CR931680	1
22A	3405/39	CR931681	16
22F	1772/40	CR931682	16
23A	1196/45	CR931683	10
23B	1039/41	CR931684	25
23F	SP-264	AF030373	10
	UK-577	AF030374	10
	Him18	AF057294	22
	Dr. Melchior	CR931685	22
	SP23-BS72	NZ_ABAG01000001	22
24A	2748/40	CR931686	21
24B	2236/42	CR931687	1
24F	24F L	CR931688	22
27	34371	CR931691	10
28A	1982/45	CR931692	16
28F	34372	CR931693	10
29	34373	CR931694	10
31	34374	CR931695	1
32A	2813/41	CR931696	9
32F	34375	CR931697	9
33A	Biehl	CR931698	10
33B	E294	CR931699	1

33C	7098/41	CR931700	10
33D	CSF/79	CR931701	1
33F	SSISP33F	AJ006986	10
	3084/37	CR931702	10
34	676/74	CR931703	1
35A	1936/39	CR931704	10
35B	4356/39	CR931705	10
35C	7765/43	CR931706	2
35F	361/39	CR931707	22
36	1095/39	CR931708	22
39	203/40	CR931711	10
40	Colemore	CR931712	22
41A	6803	CR931713	1
41F	8211/40	CR931714	22
42	198/71	CR931715	2
43	2427/48	CR931716	19
44	Hammer	CR931717	10
45	Eddy nr. 72	CR931718	1
46	Eddy nr. 73	CR931719	10
47A	L351	CR931720	2
47F	Eddy nr. 52	CR931721	2
48	656/63	CR931722	19

[a] The sequences (either complete or partial) of the *cps* loci compared were obtained from the EMBL/GenBank/DDBJ databases (5 January 2009; last date accessed). Among them, only those including in a continuous sequence both the termination codon of *dexB* and the initiation codon of *cpsA* were analyzed further. For the reasons mentioned above, two isolates of serotype 37 (AJ131984 and CR931709) were not considered here. Neither did we include sequences corresponding to serotypes 25A (accession number CR931689), 25F (CR931690), and 38 (CR931710) since the first four genes of their *cps* loci, which are very similar, are interrupted by a transposase gene and rearranged in order. We also excluded from the analysis the sequences of two serotype 5 isolates (AY336008 and CR931636/CR931637) due to the probable existence of gene reorganizations at the *cps* locus. For reasons already described, serotype 3 sequences were analyzed separately from Wzy-polymerase-dependent capsular serotypes.

[b] SO, sequence organization.

[c] For clarity, SO of serotype 3 strains is designated as A or B.

Supplemental Table S2. Proven and putative streptococcal operators analyzed in this study

Organism	Transcriptional regulator	Operator	Pneumococcal ortholog	Reference
S. pneumoniae				
	AdcR	TAACYRGTTAA	–	[1]
	CiaR	TTTAAGN$_5$TTTAAG	–	[2]
	CodY	AATTTTCWGAAAATT	–	[3, 4]
	ComE	ACACTTTGGN$_{12}$ACAGTTGAG	–	[5]
	ComX1	TACGAATA	–	[6, 7]
	CopY	KACAN$_2$TGTA	–	[8, 9]
	FabT	RYTTTGAYTGTMAAAKT	–	[10]
	GlnR	TGTNAN$_7$TNACA	–	[11]
	MalR	AAACGTTT	–	[12]
	NmlR	CTTGGAGN$_4$CTCAAAG	–	[13]
	PsaR	AAAATTAAN$_6$TTAATTTT	–	[14-16]
	RitR	WNATTANW$_3$RWYRR	–	[17]
	SczA	TGTTCAGWAWTGAAYA	–	[18]
	SpxR	ATAGAGAATAGAGA	–	[19]
S. agalactiae				
	RovS	AWAAWVHTDAWN$_{6/7}$WTKWWAMDWAK	SPD_0939	[20]
S. mutans				
	HrcA	TTAGCACTCN$_9$GAGTGCTAA	SPD_0458	[21-23]
S. pyogenes				
	CovR/CsrR	DDHHATTARAR	SPD_0344	[24, 25]
S. suis/*S. gordonii*				
	ArcR	WNTGAATW$_4$ATTCA	SPD_0786	[26-28]

References

1. Panina, E. M., Mironov, A. A. and Gelfand, M. S. 2003, Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins, *Proc. Natl. Acad. Sci. USA*, **100,** 9912–9917.
2. Halfmann, A., Kovács, M., Hakenbeck, R. and Brückner, R. 2007, Identification of the genes directly controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15 promoters drive expression of small non-coding RNAs, *Mol. Microbiol.*, **66,** 110–126.
3. den Hengst, C. D., van Hijum, S. A. F. T., Geurts, J. M. W., Nauta, A., Kok, J. and Kuipers, O. P. 2005, The *Lactococcus lactis* CodY regulon: Identification of a conserved *cis*-regulatory element, *J. Biol. Chem.*, **280,** 34332–34342.
4. Hendriksen, W. T., Bootsma, H. J., Estevão, S.*, et al.* 2008, CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization, *J. Bacteriol.*, **190,** 590-601.
5. Ween, O., Gaustad, P. and Håvarstein, L. S. 1999, Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*, *Mol. Microbiol.*, **33,** 817–827.
6. Lee, M. S. and Morrison, D. A. 1999, Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation, *J. Bacteriol.*, **181,** 5004–5016.
7. Luo, P., Li, H. and Morrison, D. A. 2003, ComX is a unique link between multiple quorum sensing outputs and competence in *Streptococcus pneumoniae*, *Mol. Microbiol.*, **50,** 623–633.
8. Portmann, R., Poulsen, K. R., Wimmer, R. and Solioz, M. 2006, CopY-like copper inducible repressors are putative 'winged helix' proteins, *BioMetals*, **19,** 61–70.
9. Reyes, A., Leiva, A., Cambiazo, V., Méndez, M. A. and González, M. 2006, *Cop-like* operon: Structure and organization in species of the *lactobacillale* order, *Biol. Res.*, **39,** 87–93.
10. Lu, Y.-J. and Rock, C. O. 2006, Transcriptional regulation of fatty acid biosynthesis in *Streptococcus pneumoniae*, *Mol. Microbiol.*, **59,** 551–566.

11. Kloosterman, T. G., Hendriksen, W. T., Bijlsma, J. J. E.*, et al.* 2006, Regulation of glutamine and glutamate metabolism by GlnR and GlnA in *Streptococcus pneumoniae*, *J. Biol. Chem.*, **281,** 25097−25109.
12. Puyet, A., Ibañez, A. M. and Espinosa, M. 1993, Characterization of the *Streptococcus pneumoniae* maltosaccharide regulator MalR, a member of the LacI-GalR family of repressors displaying distinctive genetic features, *J. Biol. Chem.*, **268,** 25402−25408.
13. Stroeher, U. H., Kidd, S. P., Stafford, S. L., Jennings, M. P., Paton, J. C. and McEwan, A. G. 2007, A pneumococcal MerR−like regulator and *S*−nitrosoglutathione reductase are required for systemic virulence, *J. Infect. Dis.*, **196,** 1820−1826.
14. Jakubovics, N. S., Smith, A. W. and Jenkinson, H. F. 2000, Expression of the virulence-related Sca (Mn^{2+}) permease in *Streptococcus gordonii* is regulated by a diphtheria toxin metallorepressor-like protein ScaR, *Mol. Microbiol.*, **38,** 140−153.
15. Johnston, J. W., Briles, D. E., Myers, L. E. and Hollingshead, S. K. 2006, Mn^{2+}-dependent regulation of multiple genes in *Streptococcus pneumoniae* through PsaR and the resultant impact on virulence, *Infect. Immun.*, **74,** 1171−1180.
16. Kloosterman, T. G., Witwicki, R. M., van der Kooi-Pol, M. M., Bijlsma, J. J. E. and Kuipers, O. P. 2008, Opposite effects of Mn^{2+} and Zn^{2+} on PsaR-mediated expression of the virulence genes *pcpA*, *prtA*, and *psaBCA* of *Streptococcus pneumoniae*, *J. Bacteriol.*, **190,** 5382−5393.
17. Ulijasz, A. T., Andes, D. R., Glasner, J. D. and Weisblum, B. 2004, Regulation of iron transport in *Streptococcus pneumoniae* by RitR, an orphan response regulator, *J. Bacteriol.*, **186,** 8123−8136.
18. Kloosterman, T. G., van der Kooi-Pol, M. M., Bijlsma, J. J. E. and Kuipers, O. P. 2007, The novel transcriptional regulator SczA mediates protection against Zn^{2+} stress by activation of the Zn^{2+}-resistance gene *czcD* in *Streptococcus pneumoniae*, *Mol. Microbiol.*, **65,** 1049−1063.
19. Ramos-Montañez, S., Tsui, H.-C. T., Wayne, K. J.*, et al.* 2008, Polymorphism and regulation of the *spxB* (pyruvate oxidase) virulence factor gene by a CBS-HotDog

domain protein (SpxR) in serotype 2 *Streptococcus pneumoniae*, *Mol. Microbiol.*, **67,** 729–746.

20. Samen, U. M., Eikmanns, B. J. and Reinscheid, D. J. 2006, The transcriptional regulator RovS controls the attachment of *Streptococcus agalactiae* to human epithelial cells and the expression of virulence genes, *Infect. Immun.*, **74,** 5625–5635.
21. Kim, S.-N., Bae, Y.-G. and Rhee, D.-K. 2008, Dual regulation of *dnaK* and *groE* operons by HrcA and Ca^{++} in *Streptococcus pneumoniae*, *Arch. Pharm. Res.*, **31,** 462–467.
22. Kwon, H.-Y., Kim, S.-N., Pyo, S.-N. and Rhee, D.-K. 2005, Ca^{2+}-dependent expression of the CIRCE regulon in Streptococcus pneumoniae, *Mol. Microbiol.*, **55,** 456–468.
23. Lemos, J. A. C., Chen, Y.-Y. M. and Burne, R. A. 2001, Genetic and physiologic analysis of the *groE* operon and role of the HrcA repressor in stress gene regulation and acid tolerance in *Streptococcus mutans*, *J. Bacteriol.*, **183,** 6074–6084.
24. Churchward, G., Bates, C., Gusa, A. A., Stringer, V. and Scott, J. R. 2009, Regulation of streptokinase expression by CovR/S in *Streptococcus pyogenes*: CovR acts through a single high-affinity binding site, *Microbiology*, **155,** 566–575.
25. Federle, M. J. and Scott, J. R. 2002, Identification of binding sites for the group A streptococcal global regulator CovR, *Mol. Microbiol.*, **43,** 1161–1172.
26. Dong, Y., Chen, Y.-Y. M., Snyder, J. A. and Burne, R. A. 2002, Isolation and molecular analysis of the gene cluster for the arginine deiminase system from *Streptococcus gordonii* DL1, *Appl. Environ. Microbiol.*, **68,** 5549–5553.
27. Gruening, P., Fulde, M., Valentin-Weigand, P. and Goethe, R. 2006, Structure, regulation, and putative function of the arginine deiminase system of *Streptococcus suis*, *J. Bacteriol.*, **188,** 361–369.
28. Zeng, L., Dong, Y. and Burne, R. A. 2006, Characterization of *cis*-acting sites controlling arginine deiminase gene expression in *Streptococcus gordonii*, *J. Bacteriol.*, **188,** 941–949.