

## *Supplementary Material*

### **Extracellular pneumococcal serine proteases affect nasopharyngeal colonization**

**Murtadha Q. Ali<sup>1</sup>, Thomas P. Kohler<sup>1</sup>, Gerhard Burchhardt<sup>1</sup>, Andreas Wüst<sup>1</sup>, Nadin Henck<sup>1</sup>, Robert Bolzmann<sup>1</sup>, Franziska Voß<sup>1</sup>, and Sven Hammerschmidt<sup>1\*</sup>**

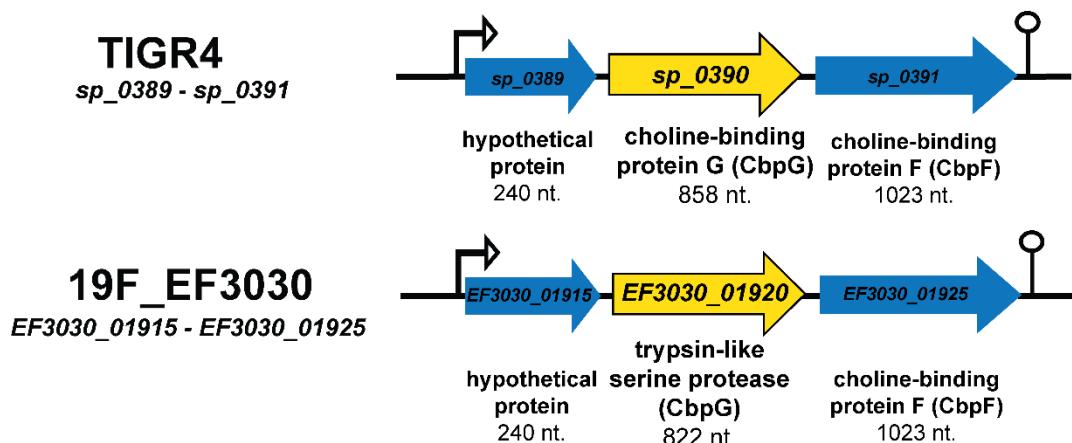
<sup>1</sup>Department of Molecular Genetics and Infection Biology, Interfaculty Institute of Genetics and Functional Genomics, Center for Functional Genomics of Microbes, University of Greifswald, Greifswald, Germany

**\* Correspondence:**

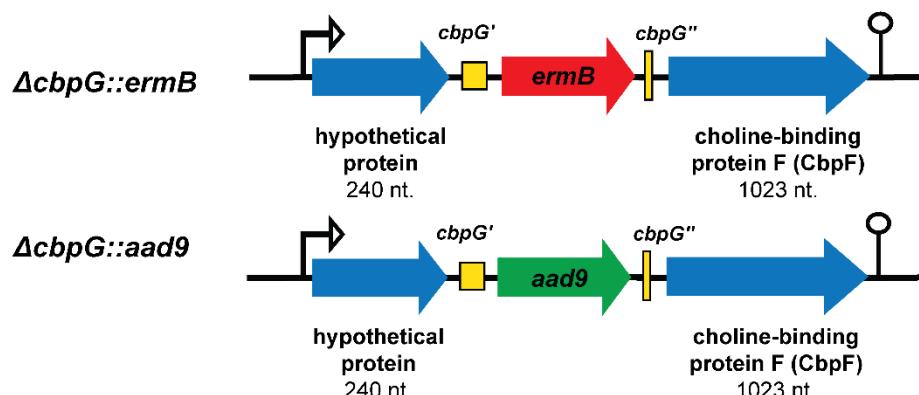
Prof. Dr. Sven Hammerschmidt

[sven.hammerschmidt@uni-greifswald.de](mailto:sven.hammerschmidt@uni-greifswald.de)

## A Gene organization of *cbpG* in *S. pneumoniae* TIGR4 and 19F\_EF3030



## B *S. pneumoniae* $\Delta$ cbpG mutation in TIGR4 and 19F\_EF3030



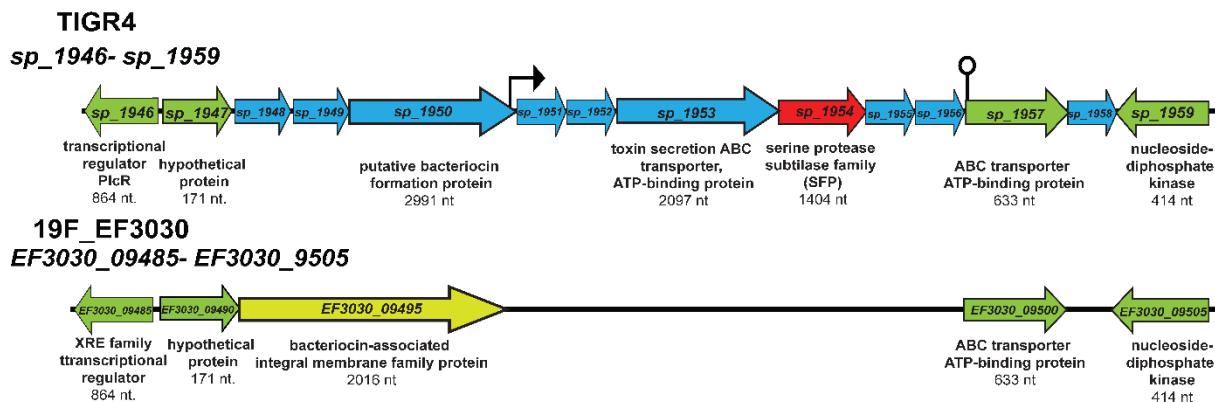
**Figure S1. Genetic organization of *cbpG* in *S. pneumoniae* TIGR4 and 19F\_EF3030.**

(A) Chromosomal localization of the *cbpG* gene (*sp\_0390*, *EF3030\_01920*) shows a 99.6% homology. For the homology search, BlastN and BlastP were used (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Arrowheads indicate the orientation of the genes. The Neural network predicted putative promoters (black arrowheads) by promotor prediction program. The transcription termination signals are shown as loops.

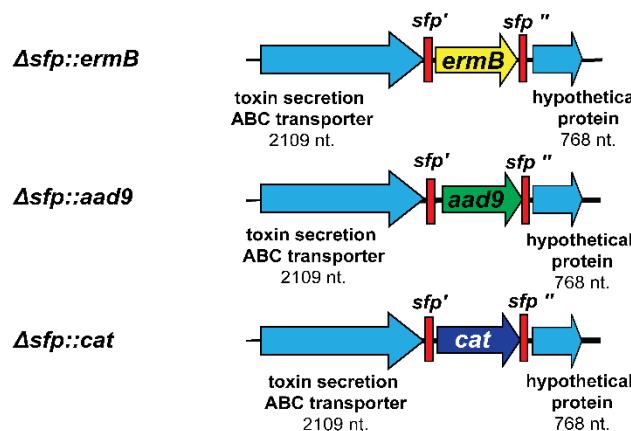
(B) Schematic model of *cbpG* mutants constructed by insertion-deletion mutagenesis shows the *erm* or *aad9* gene cassette insertion sites in the *cbpG* gene.

**A**

### Gene organization and comparison of *sfp* in *S. pneumoniae* TIGR4 and 19F\_EF3030

**B**

### *S. pneumoniae* TIGR4 $\Delta sfp$ mutation with different resistance genes

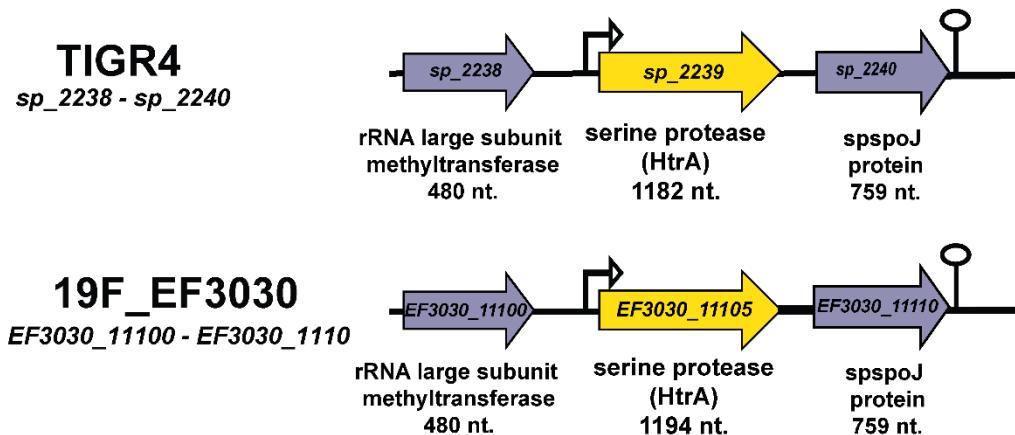


**Figure S2. Genetic organization of *sfp* in *S. pneumoniae* TIGR4 and EF3030.**

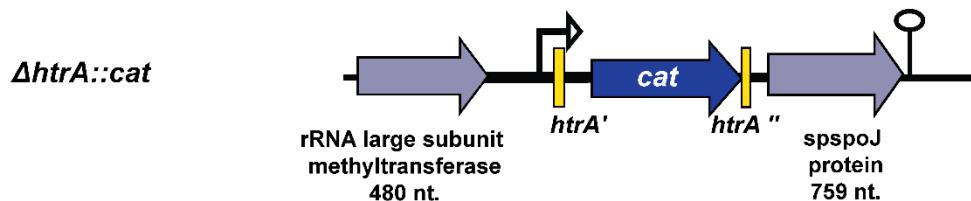
(A) Comparison of genome regions in TIGR4 *sp\_1946-sp\_1959* with 19F *EF3030\_09485-EF3030\_09505*. The *sfp* gene in TIGR4 *sp\_1954* is shown as a red arrow with 1404 nt, while in 19F\_EF3030, the *sfp* gene and 6 genes upstream are absent. The homologous analysis was done (Clustal Omega tool) to check whether the *sfp* gene is located in a different gene locus in 19F. We found that the upstream genes *sp\_1947* (171 nt) and *sp\_1946* (864 nt) encoding the hypothetical protein and transcriptional regulator, respectively, are homolog (100.0% identical) to the gene *EF3030\_09490*, *EF3030\_09485*. The genes *sp\_1957* and *sp\_1959* (light green) encoding for an ABC transport system (ATP-binding protein) and nucleoside-diphosphate kinase were identical to *EF3030\_09500* and *EF3030\_09505*, respectively. Therefore, three genes downstream of *sfp* are missing in EF3030. Arrowheads indicate the orientation of the gene, the predicted putative promoters (black arrowheads), and the transcription termination signals are shown as loops.

(B) Schematic model of *sfp* mutants constructed by insertion-deletion mutagenesis shows the *erm*, *aad9*, or *cat* genes cassette inserted in the *sfp* gene.

## A Gene organization of *htrA* in *S. pneumoniae* TIGR4 and 19F\_EF3030



## B *S. pneumoniae* $\Delta$ *htrA* mutation in TIGR4 and 19F\_EF3030

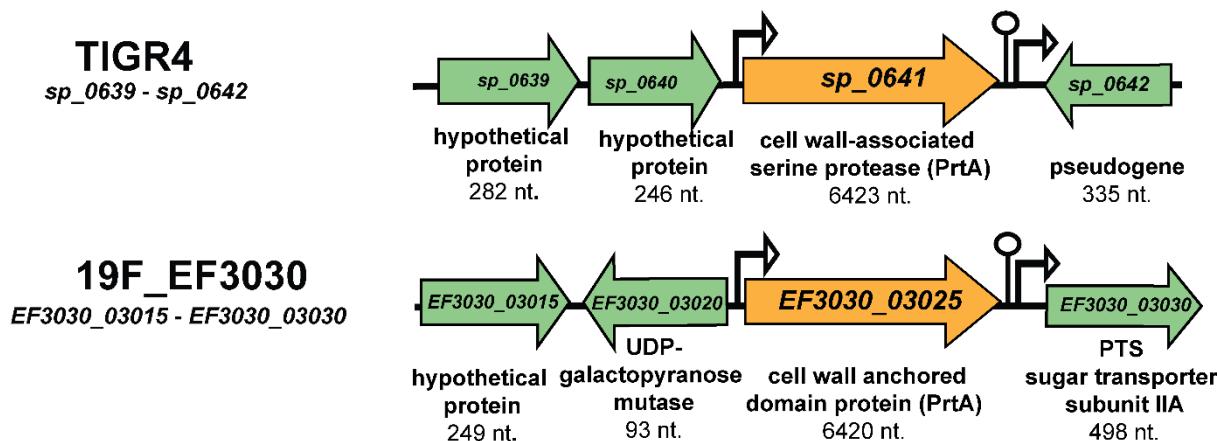


**Figure S3. Genetic organization of *htrA* in *S. pneumoniae* TIGR4 and 19F\_EF3030.**

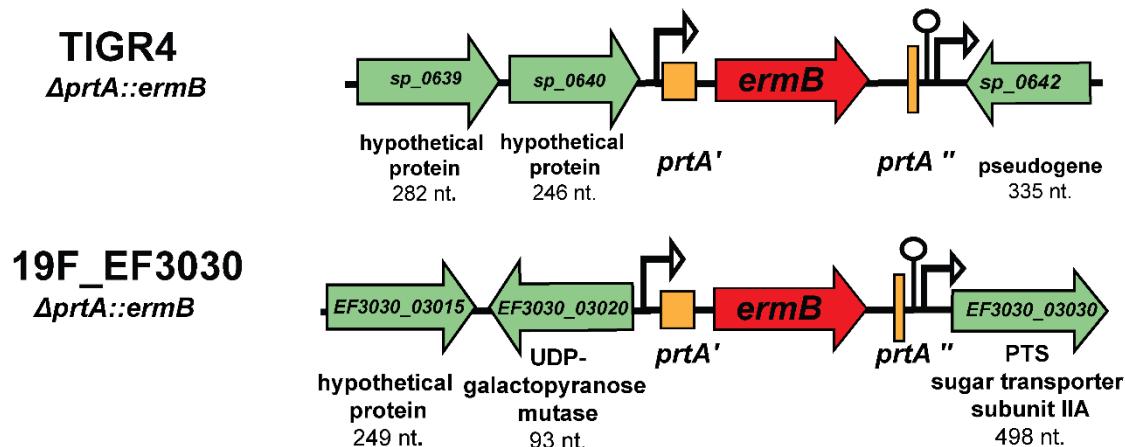
(A) The chromosomal localization of the *htrA* (high-temperature requirement gene: *sp\_2239* in TIGR4 and *EF3030\_11105* in 19F). Gene and protein sequences show a 100.0% identity. The gene upstream of *htrA* encodes for an rRNA subunit methyl-transferase. The downstream gene is annotated as ParB/RepB/Spo0J family partition protein. Arrowheads indicate the gene's orientation, the predicted putative promoters (black arrowheads), and the transcription termination signals are shown as loops.

(B) Schematic model of *htrA* mutants constructed by insertion-deletion mutagenesis in TIGR4 and 19F shows the *cat* gene cassette inserted in the *htrA* gene sequence.

## A Gene organization of *prtA* in *S. pneumoniae* TIGR4 and 19F\_EF3030



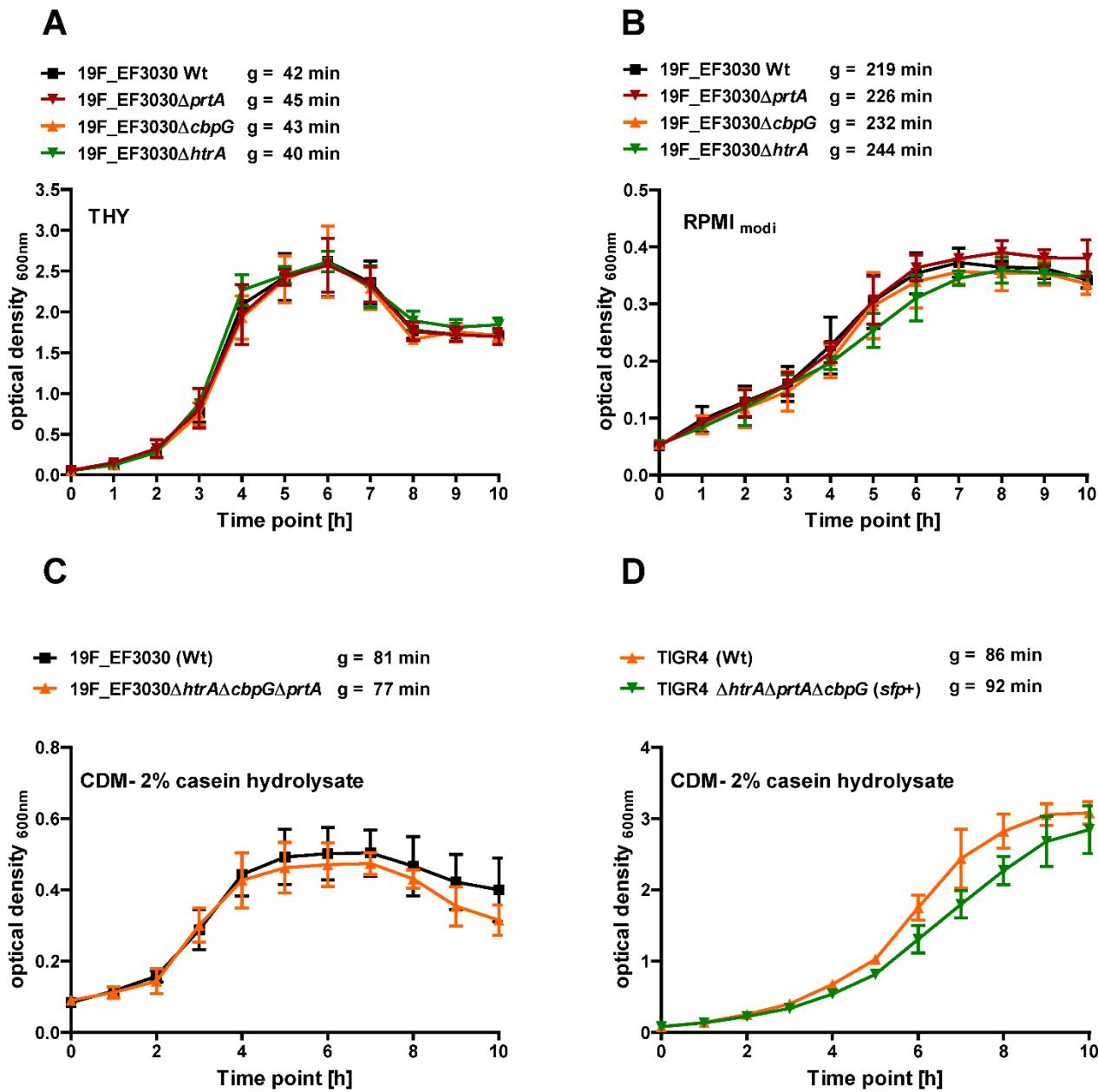
## B *S. pneumoniae* $\Delta$ *prtA* mutation in TIGR4 and 19F\_EF3030



**Figure S4. Genetic organization of *prtA* in *S. pneumoniae* in TIGR4 and 19F\_EF3030.**

(A) The chromosomal localization of the *prtA* gene encoding for the cell wall-associated serine protease (*sp\_0641*, *EF3030\_03025*). Gene and protein sequences are 100.0% identical. In TIGR4, the gene downstream of *prtA* is a pseudogene, while in 19F\_EF3030, the gene encodes for a PTS sugar transporter subunit. The genes upstream of *prtA* (*sp\_0640*, 246 nt., and *sp\_0639*, 282 nt.) in TIGR4 encode for hypothetical proteins. 19F\_EF3030 shows different upstream genes compared to TIGR4; these genes are *EF3030\_03015* and *EF3030\_03020*, which encode for a hypothetical protein or UDP-galactopyranose mutase. Arrowheads indicate the gene's orientation, the predicted putative promoters (black arrowheads), and the transcription termination signals are shown as loops.

(B) Schematic model of *prtA* mutants constructed by insertion-deletion mutagenesis in TIGR4 and 19F shows the *ermB* gene cassette inserted in the *prtA* gene sequence.



**Figure S5. Growth of pneumococcal serine protease mutants in different culture media**

(A, B) *S. pneumoniae* 19F\_EF3030 wild-type and single isogenic mutants were cultured at 37°C in THY and chemically defined medium (RPMI<sub>modi</sub>).

(C, D) 19F\_EF3030 or its isogenic mutant lacking all serine proteases as well as TIGR4 wild-type and its SFP positive triple mutant were grown in CDM without the essential amino acid solution I and II, supplemented instead with 2% casein hydrolysate. Error bars represent SD (n = 4). The symbol "g" indicates the generation time, calculated from four biological replicates. The data were statistically analyzed using a two-way ANOVA analysis. **L-amino acid I:** alanine, asparagine acid, asparagine, glutamine acid, glycine, isoleucine, leucine, lysine, proline, hydroxyproline, serine, valine, glutamine, and threonine. **L-amino acid II:** arginine, histidine, methionine, phenylalanine, tryptophan, and tyrosine.

**Table S1. Growth rates of *Streptococcus pneumoniae* wild-type and isogenic mutants in THY and RPMI<sub>modi</sub> medium.**

Pneumococcal strain	Growth rate ( $\mu$ ) in THY [min <sup>-1</sup> ]	Growth rate ( $\mu$ ) in RPMI <sub>modi</sub> [min <sup>-1</sup> ]
<b>19F_EF3030 (wild type)</b>	0.016	0.0031
19F_EF3030ΔprtA	0.015	0.0030
19F_EF3030ΔhtrA	0.017	0.0028
19F_EF3030ΔcbpG	0.016	0.0028
<b>19F_EF3030 (wild type)</b>	0.0146	0.0045
19FΔhtrAΔcbpG (prtA+)	0.0149	0.0045
19FΔprtAΔcbpG (htrA+)	0.0176	0.0048
19FΔprtAΔhtrA (cbpG+)	0.0135	0.0041
19FΔhtrAΔcbpGΔprtA (All proteases)	0.0142	0.0052
<b>TIGR4Δcps (wild type)</b>	0.0099	0.018
TIGR4ΔcpsΔhtrAΔcbpGΔsfp (prtA+)	0.0096	0.013
TIGR4ΔcpsΔprtAΔcbpGΔsfp (htrA+)	0.0091	0.011
TIGR4ΔcpsΔprtAΔhtrAΔsfp (cbpG+)	0.0090	0.009
TIGR4ΔcpsΔhtrAΔprtAΔcbpG (sfp+)	0.0085	0.010
<b>TIGR4lux (wild type)</b>	0.0119	0.0041
TIGR4luxΔhtrAΔcbpGΔsfp (prtA+)	0.0123	0.0041
TIGR4luxΔprtAΔcbpGΔsfp (htrA+)	0.0113	0.0041
TIGR4luxΔhtrAΔprtAΔsfp (cbpG+)	0.0081	0.0045
TIGR4luxΔhtrAΔprtAΔcbpG (sfp+)	0.0083	0.0045

**Table S2. Colony-forming units (CFU) of adherent pneumococci counted on blood agar plates after 4 h of infection of Detroit-562 cells**

Pneumococcal strain	Adherent bacteria CFU/ml per 2.5 x 10 <sup>5</sup> cells	
	Mean ± SD	P-value
<b>19F_EF3030 (wild type)</b>	1162 ± 928.4	-
19FΔhtrAΔcbpG (prtA+)	144.2 ± 140.1	0,0162
19FΔprtAΔcbpG (htrA+)	154.8 ± 151.4	0,0162
19FΔprtAΔhtrA (cbpG+)	147.1 ± 175.3	0,0162
19FΔhtrAΔcbpGΔprtA (non-function serine proteases )	98.71 ± 109.6	0,0040
<b>TIGR4Δcps (wild type)</b>	4870 ± 1294	-
TIGR4ΔcpsΔhtrAΔcbpGΔsfp (prtA+)	2365 ± 1428	0.0591
TIGR4ΔcpsΔprtAΔcbpGΔsfp (htrA+)	1491 ± 1253	0.0294
TIGR4ΔcpsΔprtAΔhtrAΔsfp (cbpG+)	1424 ± 8630	0.0294
TIGR4ΔcpsΔhtrAΔprtAΔcbpG (sfp+)	8307 ± 2650	0.0294

\*P value less than 0.05 was taken as statistically significant.

**Table S3. Colony-forming units (CFU) from nasopharyngeal lavage counted on blood agar plates**

Pneumococcal strains	CFU/ml in the nasopharyngeal lavage							
	Median Day 2	P- value	Median Day3	P- value	Median Day 7	P- value	Median Day 14	P- value
							23300	
<b>19F_EF3030 (wild type)</b>	291000		210000		147000		23300	
19F $\Delta$ htrA $\Delta$ cbpG (prtA+)	31700	0,0004	27600	0,0004	25800	0,0005	7070	0,0163
19F $\Delta$ prtA $\Delta$ cbpG (htrA+)	53700	0,0025	69200	0,0153	25200	0,0009	9950	0,0630
19F $\Delta$ prtA $\Delta$ htrA (cbpG+)	21600	0,0005	16700	0,0003	20000	0,0009	6060	0,0066
19F $\Delta$ htrA $\Delta$ cbpG $\Delta$ prtA	33800	0,0005	35400	0,0012	36000	0,0081	1860	0,0015

\*P value less than 0.05 was taken as statistically significant.

**Table S4. Colony-forming units from bronchoalveolar lavage counted on blood agar plates**

Pneumococcal strain	CFU/ml in the nasopharyngeal wash							
	Median Day 2	P- value	Median Day3	P- value	Median Day 7	P- value	Median Day 14	P- value
							33,00	
<b>19F_EF3030 (wild type)</b>	564,0		1795		6408		33,00	
19F $\Delta$ htrA $\Delta$ cbpG (prtA+)	15,00	0,1443	215,8	0,5753	30,00	0,0040	1,000	0,7819
19F $\Delta$ prtA $\Delta$ cbpG (htrA+)	175,0	0,8519	50,00	0,0224	16,65	0,0852	233,0	0,5841
19F $\Delta$ prtA $\Delta$ htrA (cbpG+)	207,0	0,4327	108,0	0,1916	133,0	0,0733	1,000	0,0370
19F $\Delta$ htrA $\Delta$ cbpG $\Delta$ prtA	1240	0,8813	425,0	0,6815	1,000	0,0911	1,000	0,0156

\*P value less than 0.05 was taken as statistically significant.

## 1 Pneumococcal serine protease amino acid sequences

- 1.1 HtrA amino acid sequences of *S. pneumoniae* sp\_2239 in TIGR4 and EF3030\_11105 in 19F proteins are identical (protein accession no. AAK76286.1 and QBF69928.1). The signal peptide sequence (31 aa) is marked in blue, the serine protease catalytic domain (182 aa) position 96-277 is marked in red, PDZ domain (87 aa) position 289-375 is marked in green.

MKHLKTFYKKWFQLLVVIVISFFSGALGSFS ITQLTQKSSVNNNSNNNSTITQTAYKNENSTTQAVNKVKDAVVSVITYSAN  
RQNSVFGNDTDQDTSQRISSEGSGVIYKKNDKEAYIYTNNHVGASKVDIRLSGDTKVPGIEVGADTFSDIAVVKISSEK  
VTTVAEFGDSSKLTVGETAIAIGSPLGSEYANTVTQGIVSSLNRNVLKSEDGQAISTKAIQTDATINPGNGGGPLINIQG  
QVIGITSSKIATNGGTSVEGLFAIPANDAINIEQLEKNGKVTTPALGIQMVNLSNVSTDIRRNLIPSNVTSGVIVRSV  
QSNMPANGHLEKYDVITKDDKEIASSTDLQSALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES

- 1.2 PrtA amino acid sequences of *S. pneumoniae* sp\_0641 in TIG4 and EF3030\_03025 in 19F proteins are identical (protein accession no. AAK74791.1 and QBF68585.1). The signal peptide sequence (27 aa) is marked in blue, the serine protease catalytic domain (542 aa) position 223-764 is marked in red, DUF 1034 domain (140 aa) position 795-934 is marked in green. The C-terminal LPKTG anchoring motif (42) is marked in yellow.

MKKSTVLSLTAAVILAAYAPNEVVLA DTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVS  
NPVIDNNTSNEEAIKEENSNKSQGDYTDTSFVNKNTEPKKEDKVYYIAEFKDKESEGEKAIKESSLKNTKVLYTYDRIF  
NGSAIETTPDNLDKIKQIEGSISSVERAQKVQPMNHARKEIGVEEADYLKSINAPFGKNFDGRGMVISNI  
DTGTDYRHK AMRIDDDAKASMRFKKEDLKGTDKNWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFN  
GIDGIAPNAQIFS YKMSDAGSGFAGDETMFHAIEDSIKHNVDVSVSSGFTGTGLVGEKYWQAIRALKAGIPMVVATG  
NYATSASSSSWDLVANNHLMKMTDTGNVTRTA  
AHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITT  
NEDGT KAPSLKLFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARA  
IMVNTVNYN RDNWTEL  
PAMYEADE GTKSQVFSISGDDGVKLWNMINPD  
KKTEVKRNNKEDFKDKLEQYYPIDM  
ESFNSNKPNVGDEKEIDFKFAPDT  
DKELYKE DII  
VPAGSTSWGPRIDLLKPDVS  
APGKNIKSTLN  
VINGKSTYGYMSGTS  
MATPIVA  
ASTVLIRPKL  
KEMLERPV  
KLN  
GDDKIDL  
TS  
LT  
KIAL  
QNT  
ARP  
MM  
DAT  
SW  
KE  
KS  
QY  
F  
AS  
PR  
QQ  
GAG  
LIN  
VAN  
AL  
R  
NE  
VV  
V  
AT  
FK  
NT  
DS  
K  
GL  
V  
NS  
Y  
GS  
IS  
L  
KE  
KG  
DK  
KY  
FT  
IK  
L  
H  
N  
TS  
N  
R  
P  
L  
T  
F  
K  
V  
S  
A  
I  
T  
D  
S  
L  
T  
D  
R  
L  
K  
L  
D  
E  
T  
Y  
K  
D  
Q  
Y  
K  
N  
E  
T  
L  
F  
A  
R  
D  
Q  
K  
E  
H  
P  
E  
K  
F  
D  
E  
I  
A  
N  
E  
V  
W  
Y  
A  
G  
A  
L  
V  
N  
E  
D  
G  
E  
V  
K  
N  
L  
E  
V  
T  
Y  
A  
G  
E  
Q  
G  
R  
N  
R  
K  
L  
D  
K  
G  
N  
T  
I  
Y  
E  
I  
K  
G  
A  
G  
D  
L  
R  
G  
K  
I  
I  
E  
V  
A  
L  
D  
G  
S  
S  
N  
F  
T  
K  
I  
H  
R  
I  
K  
F  
A  
N  
Q  
A  
D  
E  
K  
G  
M  
I  
S  
Y  
Y  
L  
V  
D  
P  
D  
Q  
S  
S  
K  
Y  
Q  
K  
L  
G  
E  
I  
A  
E  
S  
K  
F  
K  
N  
L  
G  
N  
G  
K  
E  
G  
S  
L  
K  
K  
D  
T  
T  
G  
L  
I  
N  
P  
R  
G  
R  
E  
E  
N  
A  
P  
E  
S  
K  
D  
N  
Q  
D  
P  
A  
T  
K  
I  
R  
G  
Q  
F  
E  
P  
I  
A  
E  
G  
Q  
Y  
F  
Y  
K  
F  
K  
Y  
R  
L  
T  
K  
D  
Y  
P  
Q  
V  
S  
Y  
I  
P  
V  
K  
I  
D  
N  
T  
A  
P  
K  
I  
V  
S  
V  
D  
F  
S  
N  
P  
E  
K  
I  
R  
G  
K  
I  
S  
D  
K  
G  
G  
F  
N  
W  
E  
L  
R  
V  
N  
E  
S  
V  
V  
D  
N  
Y  
L  
I  
Y  
G  
D  
L  
H  
I  
D  
N  
T  
R  
D  
F  
N  
I  
K  
L  
N  
V  
K  
D  
G  
D  
I  
M  
D  
W  
G  
M  
K  
D  
Y  
K  
A  
N  
G  
F  
P  
D  
K  
V  
T  
D  
M  
G  
N  
V  
Y  
L  
Q  
T  
G  
Y  
S  
D  
L  
N  
A  
K  
A  
V  
G  
V  
H  
Y  
Q  
F  
L  
Y  
D  
N  
V  
K  
P  
E  
V  
N  
I  
D  
P  
K  
G  
N  
T  
S  
I  
E  
Y  
A  
D  
G  
K  
S  
V  
V  
F  
N  
I  
N  
D  
K  
R  
N  
N  
G  
F  
D  
G  
E  
I  
Q  
E  
Q  
H  
I  
Y  
I  
N  
G  
K  
E  
M  
S  
S  
T  
I  
V  
S  
E  
E  
D  
F  
I  
L  
P  
V  
Y  
K  
G  
E  
L  
E  
K  
G  
Y  
Q  
F  
D  
G  
W  
E  
I  
S  
G  
F  
E  
G  
K  
K  
D  
A  
G  
Y  
V  
I  
N  
L  
S  
K  
D  
T  
F  
I  
K  
P  
V  
F  
K  
I  
E  
E  
K  
K  
E  
E  
N  
K  
P  
T  
F  
D  
V  
S  
K  
K  
D  
N  
P  
Q  
V  
N  
H  
S  
Q  
L  
N  
E  
S  
H  
R  
K  
E  
D  
L  
Q  
R  
E  
E  
H  
S  
Q  
K  
S  
D  
S  
T  
K  
D  
V  
T  
A  
T  
V  
L  
D  
K  
N  
N  
I  
S  
S  
K  
S  
T  
T  
N  
N  
P  
N  
K  
L  
P  
K  
T  
G  
T  
A  
S  
G  
A  
Q  
T  
I  
L  
L  
A  
A  
G  
I  
M  
F  
I  
V  
G  
I  
F  
L  
G  
L  
K  
K  
K  
N  
Q  
D

- 1.3 SFP amino acid sequences of *S. pneumoniae sp\_1954* in TIGR4 (protein accession number [ABC75782.1](#)). The signal peptide sequence (22 aa) is marked in blue, the serine protease catalytic domain (295 aa) position 167-461 is marked in red**

MKKKYWTLATLFFCLFNNSVTAQEIPKNLDGNIHTHTQSESFSDEKQVDYSNKNQEEVDQNKFRIQIDKTELFVTTDKH  
LEKNCCCKLELEPQINNDIVNSESNNLGEDNLNDKIKENVSHLDNRGGNIEHDKDNLNESSIVRKYEWDIDKVTTGGGESYKL  
YSKSNSKVSIAILDSGVDLQNTGLLKNLSNHSKNYVPNKGYLGKEEGEEGIISDIQDRLGHGTAVVAQIVGDDNINGVNPH  
VNINVYRIFGKSSASPWIVKAIFDAVDDGNDIINLSTGQYLMIDGEYEDGTNDFTFLKYKKAIDYANQKGVIIVAALGN  
DSLNVSNQSDLLKLISSRKVKRPGLVVDVPSYFSSTISVGGIDRLGNLSDFSNKGDSDAIYAPAGSTLSLSELGLNNFIN  
AEKYKEDWIFSATLGGYTYLYGNSFAAPKVSGAIAMIIDKYKLKDQPYNYMFVKKFWKKHYQ

- 1.4 CbpG amino acid sequences of *S. pneumoniae sp\_0390* in TIGR4 and *EF3030\_01920* in 19F protein are highly homologous in TIGR4 and 19F (protein accession no. [AAK74556.1](#) and [QBF69943.1](#)). The trypsin-like serine protease catalytic domain (184 aa) position 14-197 is marked in red, the repeats of the CBM (CW1 position 207-221), (CW2 position 226-245), and (CW3 position 246-265) are marked in orange. At the C-terminal region (in light blue, position 267-285), it is probably involved in binding to choline residues of teichoic acids.**

MVL SKYYGVADGMNVEGRGSANFIKDNVLITAHHNYYRHGYGKEADDIYVLP AVSPS QEPFGKIKVKEVRYLKEFRNLNSK  
DAREYDLALLILEEPIGAKLGLPTSQKNLTGITVTITGYP SYNFKIHQMYTDKKQVLSDDGMFLDYQVDTLEGSSGST  
VYDASHRVVGVHTLGDGANQINSAVKLERNLPFIYSLKGYSLEGWKKINGSWYHYRQHDKQ**TGWQEINDTWYVLDSSCK**  
**MLTDWQKVNGKWYYLN SNGAMV**TGSQTIDGKV**YNFASSGEWI**

## TIGR4 and 19F\_EF3030 HtrA protein sequence alignment

sp_2239	-----MKHLKTFYKKWFQLLVIVISFFSGALGSFSITQLTQKSSVNNNNSTITQTAYK	56
EF3030_11105	MEANMKHLKTFYKKWFQLLVIVISFFSGALGSFSITQLTQKSSVNNNNSTITQTAYK	60
	*****	
sp_2239	NENSTTQAVNKVKDAVVSVITYSANRQNSVFGNDDTDTSQRISSEGSGVIYKKNDKEAY	116
EF3030_11105	NENSTTQAVNKVKDAVVSVITYSANRQNSVFGNDDTDTSQRISSEGSGVIYKKNDKEAY	120
	*****	
sp_2239	IVTNNHVIN GASKVDI RLSDGT KVPGEIVGADTFSDIAVV KISSEKVTTVAEFGDSSKL T	176
EF3030_11105	IVTNNHVIN GASKVDI RLSDGT KVPGEIVGADTFSDIAVV KISSEKVTTVAEFGDSSKL T	180
	*****	
sp_2239	VGETAIAIGSPLGSEYANTVTQGIVSSLNRNVSLKSEDGQAISTKAIQDTAINPGNSGG	236
EF3030_11105	VGETAIAIGSPLGSEYANTVTQGIVSSLNRNVSLKSEDGQAISTKAIQDTAINPGNSGG	240
	*****	
sp_2239	PLINI QGVIGITSSKIATNGG TSVEGLGFAIPANDAINIIEQLEKNGKVTRP ALGIQM V	296
EF3030_11105	PLINI QGVIGITSSKIATNGG TSVEGLGFAIPANDAINIIEQLEKNGKVTRP ALGIQM V	300
	*****	
sp_2239	NLSNVSTSDIRRLNIPSNVTSGVIVRSVQSNMPANGHLEKYDVITKVDDKEIASSTD LQS	356
EF3030_11105	NLSNVSTSDIRRLNIPSNVTSGVVRSVQSNMPANGHLEKYDVITKVDDKEIASSTD LQS	360
	*****:*****	
sp_2239	ALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES	393
EF3030_11105	ALYNHSIGDTIKITYYRNGKEETTSIKLNKSSGDLES	397
	*****	

---

**TIGR4 and 19F\_EF3030 PrtA protein sequence alignment**

sp_0641TIG4 EF3030_03025	MKKSTVLSLTAAVILAAYAPNEVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMET MKKSTVLSLTAAVILAAYAPNEVVLADTSSSEDALSISDKEKVAENKEKHDIHNAIET *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	60 60
sp_0641TIG4 EF3030_03025	SQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENSNKSQGDYTDASFVNKTENPK SKDTEEKTTIIEKEVVSKNPVIDTKTSNEEAKIKEENSNKSQGDHTDSFVNKTENPK *:*:*****:*****:*****:*****:*****:*****:*****:*****:*****:	120 120
sp_0641TIG4 EF3030_03025	KEDKVYYIAEFKDKESGEKAICELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEG KEDKVYYIAEFKDKESGSKAICELSSLKNTKVLYTYDRIFNGGAIETTQDNLNKIKQIEG *****:*****:*****:*****:*****:*****:*****:*****:*****:	180 180
sp_0641TIG4 EF3030_03025	ISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNI DTGTDYRHK ITSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNI DTGTDYRHK *:*****:*****:*****:*****:*****:*****:*****:*****:	240 240
sp_0641TIG4 EF3030_03025	AMRIDDDAKASMRFKKEDLKGTDKNYWLSDKI PHAFNYYNGGKITVEKYDDGRDYFDPHG AMRIDDDAKASMRFKKEDLKGTDKNYWLSDKI PHAFNYYNGGKITVEKYDDGRDYFDPHG *****:*****:*****:*****:*****:*****:*****:*****:	300 300
sp_0641TIG4 EF3030_03025	MHIAGILAGNDTEQDIKINFNGIDGIAPNAQIFS YKMYSDAGSGFAGDETMFHAI EDSIKH MHIAGILAGNDTEQDIKINFNGIDGIAPNAQIFS YKMYSDAGSGFAGDETMFHAI EDSIKH *****:*****:*****:*****:*****:*****:*****:*****:	360 360
sp_0641TIG4 EF3030_03025	NVDVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMV VATGNYATSASSSWDLVANNHLK NVDVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMV VATGNYATSASSSWDLVANNHLK *****:*****:*****:*****:*****:*****:*****:	420 420
sp_0641TIG4 EF3030_03025	MTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITT NEDGT MTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKN KITT NEDGT *****:*****:*****:*****:*****:*****:*****:	480 480
sp_0641TIG4 EF3030_03025	KAPSKLKFYIYGKQDQDLIGLDLRKIAVM DRIYTKDLKNAFKKAMDKGAR AIMVNTV KAPSKLKFYIYGKQDQDLIGLDLRKIAVM DRIYTKDLKNAFKKAMDKGAR AIMVNTV *****:*****:*****:*****:*****:*****:*****:	540 540
sp_0641TIG4 EF3030_03025	YYNRDNWTEL PAMGYEADEG TKSQVFSISGDDGVKLWNMINPDK KTEVKRNNKEDFKDK YYNRDNWTEL PAMGYEADEG TKSQVFSISGDDGVKLWNMINPNKKTEVKRNNKEDFKDK *****:*****:*****:*****:*****:*****:*****:	600 600
sp_0641TIG4 EF3030_03025	LEQYYPIDMESFSNPKNVGDEKEIDFKFAPDTDKELYKEDIIVPAGST SWGP RIDLLLK LEQYYPIDMESFSNPKNVGDEKEIDFKFAPDTDKELYKEDIIVPAGST SWGP RIDLLLK *****:*****:*****:*****:*****:*****:	660 660
sp_0641TIG4 EF3030_03025	PDV Sapgknikstlnvingkstygymsgtsmat piva astvlirpk lkemlerpv lkn lk PDV Sapgknikstlnvingkstygymsgtsmat piva astvlirpk lkemlerpv lkn lk *****:*****:*****:*****:*****:*****:	720 720
sp_0641TIG4 EF3030_03025	GDDKIDLTS LTKIALQNTARPMDATSWKEKSQYFASPRQQGAGLINVAN ALRNEVVATF GDDKIDLTS LTKIALQNTARPMDATSWKEKSQYFASPRQQGAGLINVAN ALRNEVVATF *****:*****:*****:*****:*****:*****:	780 780
sp_0641TIG4 EF3030_03025	KNTDSKGLVNSYGSISLKEIKGD KYFTIKLHN TSNRPLTFKVSASA ITTDS LTDRLKLD KNTDSKGLVNSYGSISLKEIKGD KYFTIKLHN TSNRPLTFKVSASA ITTDS LTDRLKLD *****:*****:*****:*****:*****:*****:	840 840
sp_0641TIG4 EF3030_03025	ETYKDEKSPDGKQIVPEIHPKVKGANITFEHD TFTIGANSSFDLNAV INVGEAKKNKF ETYKDEKSPDGKQIVPEIHPKVKGANITFEHD TFTIGANSSFDLNAV INVGEAKKNKF *****:*****:*****:*****:*****:*****:	900 900



## Supplementary Material

EF3030_03025	FLYDNVKPEVNIDPKGNTSIEYADGKSVVFNINDKRNNNGFDGEIQEQHIIYVNGKEYTSFD *****;	1919
sp_0641TIG4 EF3030_03025	DIKQIIDKTLNIKIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIV DIKQITDKTLNIKIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIV *****	1980 1979
sp_0641TIG4 EF3030_03025	SEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKKEEE SEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKKEEE *****	2040 2039
sp_0641TIG4 EF3030_03025	NKPTFDVSKKKDNPQVNHSQLNESHRKEDLQREEHSQKS DSTKDVTATVLDKNNISSKST NKPTFDVSKKKDNPQVNHSQLNESHRKEDLQREDHSQKS DSTKDVTATVLDKNNISSKST *****:	2100 2099
sp_0641TIG4 EF3030_03025	TNNPNKLPKTGTASGAQTLLAAGIMFIVGIFLGLKKKNQD 2140 TNNPNKLPKTGTASGAQTLLAAGIMFIVGIFLGLKKKNQD 2139 *****	

---

## TIGR4 and 19F\_EF3030 CbpG protein sequence alignment

sp_0390 EF3030_01920	MVLSKYVGADGMNVEGRGSANFIKDNVLITAAHNYYRHGYKEADDIYVLPAVSPSQEP -----MNVEGRGSANFIKDNVLITAAHNYYRHGYKEADDIYVLPAVSPSQEL *****	60 48
sp_0390 EF3030_01920	FGKIKVKEVRYLKEFRNLNSKDAREYDLALLILEEPIGAKILGTGLPLTSQKNLTGIFTVI FGKIKVKEVRYLKEFRNLNSKDAREYDLALLILEEPIGAKILGTGLPLTSQKNLTGIFTVI *****	120 108
sp_0390 EF3030_01920	TGYPYSNFKIHQMYTDKKQVLSDDGMFLDYQVDTLEGSSGTVYDASHRVVGVHTLGDGA TGYPYSNFKIHQMYTDKKQVLSDDGMFLDYQVDTLEGSSGTVYDASHRVVGVHTLGDGA *****	180 168
sp_0390 EF3030_01920	NQINSAVKLNERNLPFIYSVLKGYSLEGWKKINGSWYHYRQHDQQTGWQEINDTWYLYDS NQINSAVKLNERNLPFIYSVLKGYSLEGWKKINGSWYHYRQHDQQTGWQEINDTWYLYDS *****	240 228
sp_0390 EF3030_01920	SGKMLTDWQKVNGKWYYLNSNGAMVTGSQTIDGKVYNFASSGEWI SGKMLTDWQKVNGKWYYLNSNGAMVTGSQTIDGKVYNFASSGEWI	285 273