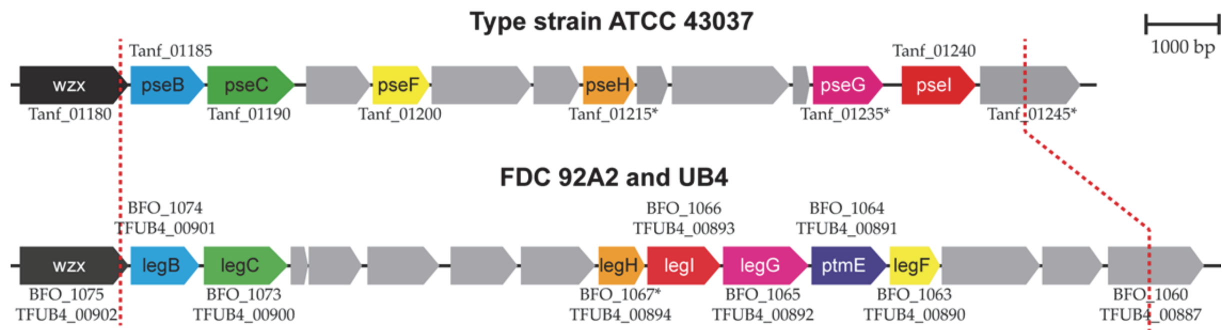
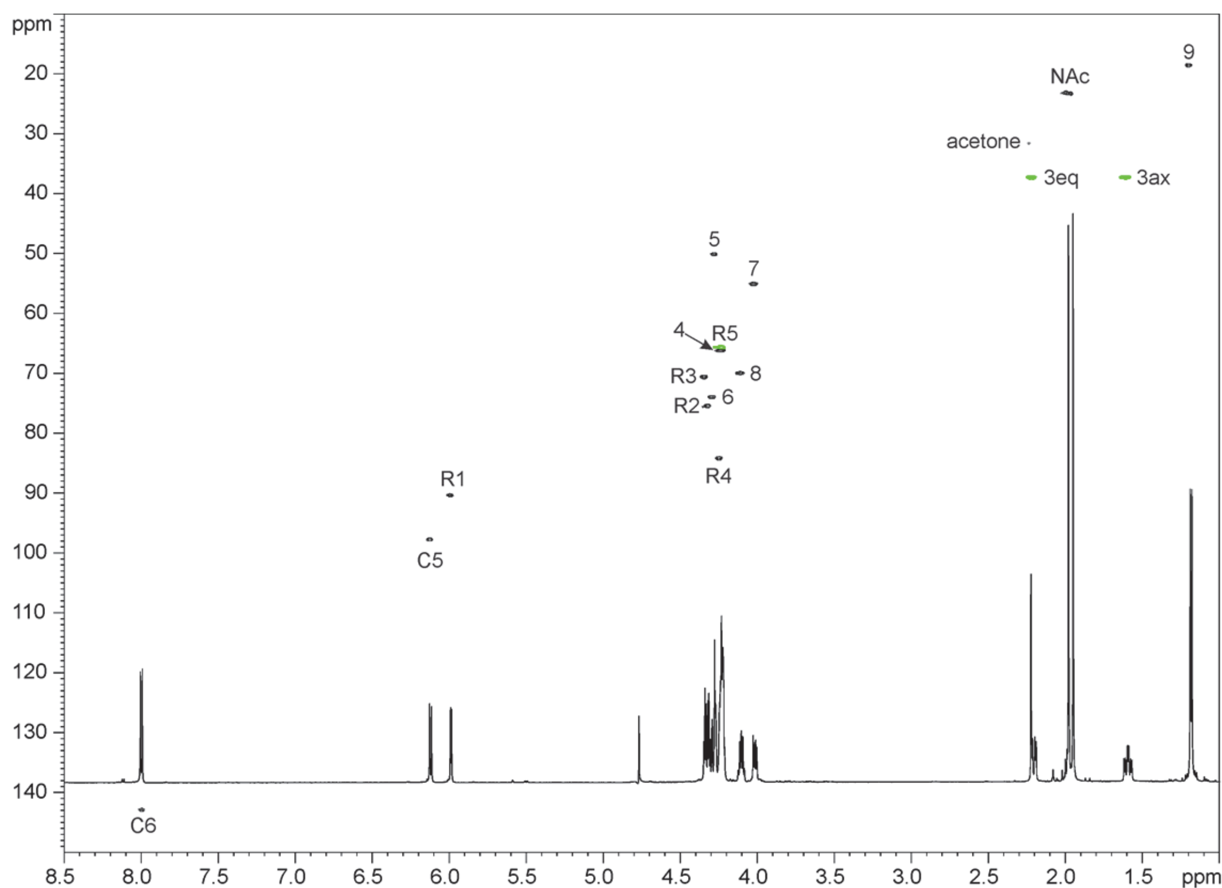
Supplementary Data

Tannerella forsythia strains display different cell-surface nonulosonic acids: biosynthetic pathway characterization and first insight into biological implications

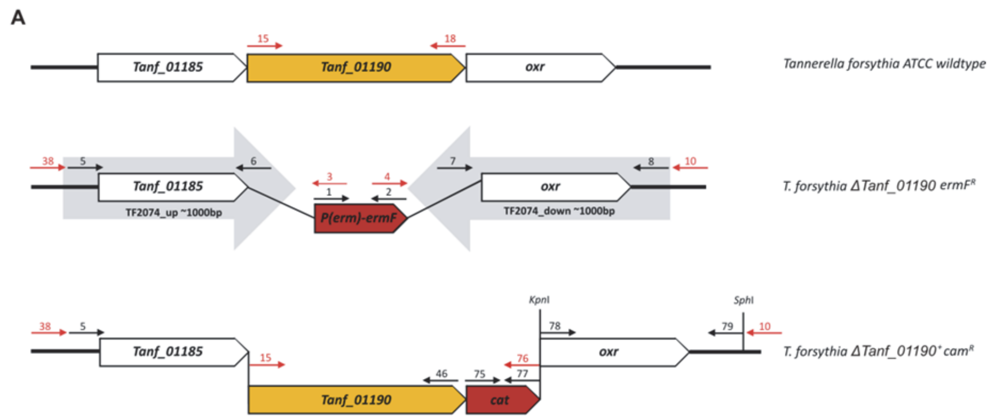
Valentin Friedrich, Bettina Janesch, Markus Windwarder, Daniel Maresch, Matthias L. Braun, Zoë A. Megson, Evgeny Vinogradov, Marie-France Goneau, Ashu Sharma, Friedrich Altmann, Paul Messner, Ian C. Schoenhofen* and Christina Schäffer*

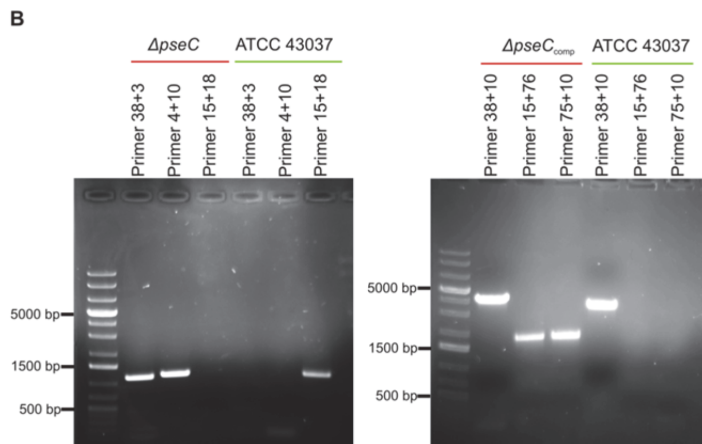


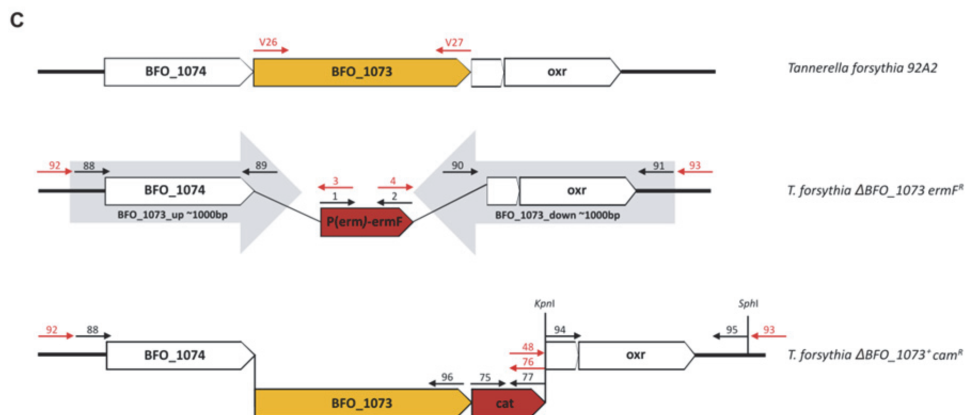
Supp. Fig. 1. NulO biosynthesis loci of the ATCC 43037 type strain and strains FDC 92A2/UB4 (to scale). The type strain was found to possess homologs to the genes of six Pse biosynthesis enzymes (PseB, PseC, PseH, PseG, PseI and PseF) (Schoenhofen, I.C., McNally, D.J., et al. 2006a). The genome of strains FDC 92A2 and UB4 are identical (99.83%) in this region and encode gene homologs for the biosynthesis of Leg (LegB, LegC, PglD/LegH, LegG, LegI and LegF), as well as a predicted nucleotidyl transferase (PtmE) for the formation of GDP-GlcNAc, the precursor substrate for this pathway (Schoenhofen, I.C., Vinogradov, E., et al. 2009). Both loci are located downstream of a putative flippase gene (*wzx*) hypothesized to be involved in the general *O*-glycosylation system. The regions within the red dashed lines show no areas of significant similarity. NCBI locus tags for ATCC 43037, FDC 92A2 and UB4 are displayed under each gene, with asterisks denoting ORFs deviating from their NCBI annotation. Grey arrows represent genes predicted to be unrelated to NulO biosynthesis or of unknown function.

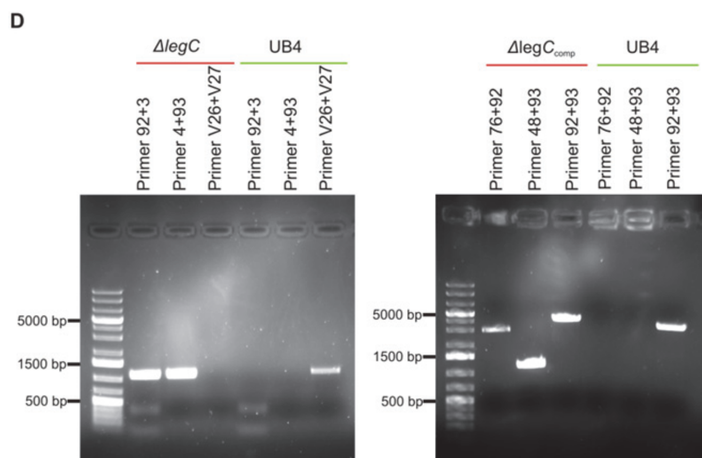


Supp. Fig. 2. The ^{1}H spectrum and ^{1}H-^{13}C HSQC correlation spectrum of CMP-Pse (VII_P). Spectra were recorded on a Varian 600 MHz spectrometer with a Varian 5 mm Z-gradient probe in D_2O at 25°C. C, cytosine; R, ribose; NAc, 5-NHAc and 7-NHAc CH_3 regions of pseudaminic acid; and acetone was included as an internal reference.

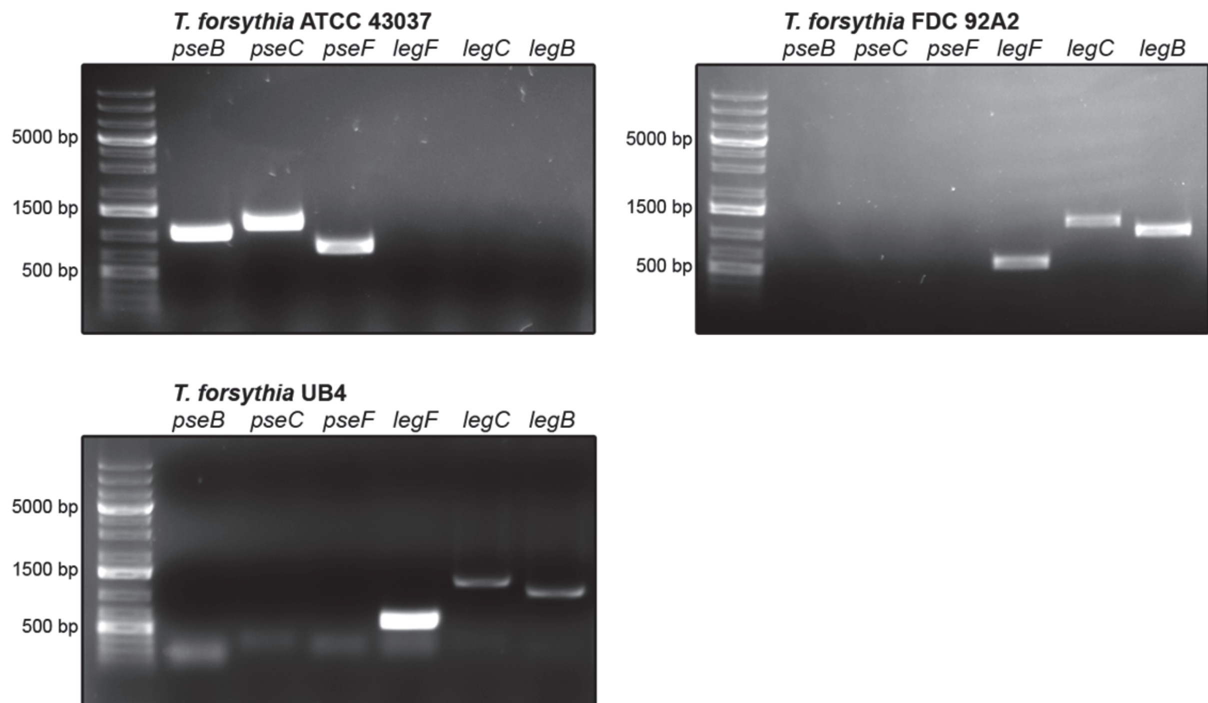

A
15
18
Tanf_01185
Tanf_01190
oxr
Tannerella forsythia ATCC wildtype
38
5
6
3
4
7
8
10
1
2
Tanf_01185
oxr
TF2074_up ~1000bp
P(erm)-ermF
TF2074_down ~1000bp
T. forsythia ΔTanf_01190 ermF^R
KpnI
SphI
38
5
78
79
10
Tanf_01185
15
76
oxr
46
75
77
Tanf_01190
cat
T. forsythia ΔTanf_01190^+ cam^R


B
ΔpseC
ATCC 43037
Primer 38+3
Primer 4+10
Primer 15+18
Primer 38+3
Primer 4+10
Primer 15+18
5000 bp
1500 bp
500 bp
ΔpseC_comp
ATCC 43037
Primer 38+10
Primer 15+76
Primer 75+10
Primer 38+10
Primer 15+76
Primer 75+10
5000 bp
1500 bp
500 bp

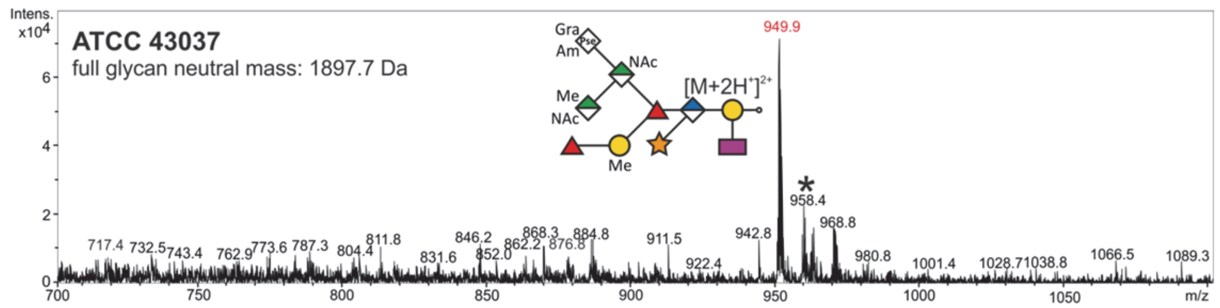

C
V26
V27
BFO_1074
BFO_1073
oxr
Tannerella forsythia 92A2
92
88
89
3
4
90
91
93
1
2
BFO_1074
oxr
BFO_1073_up ~1000bp
P(erm)-ermF
BFO_1073_down ~1000bp
T. forsythia ΔBFO_1073 ermF^R
KpnI
SphI
92
88
94
95
93
BFO_1074
48
76
oxr
96
75
77
BFO_1073
cat
T. forsythia ΔBFO_1073^+ cam^R

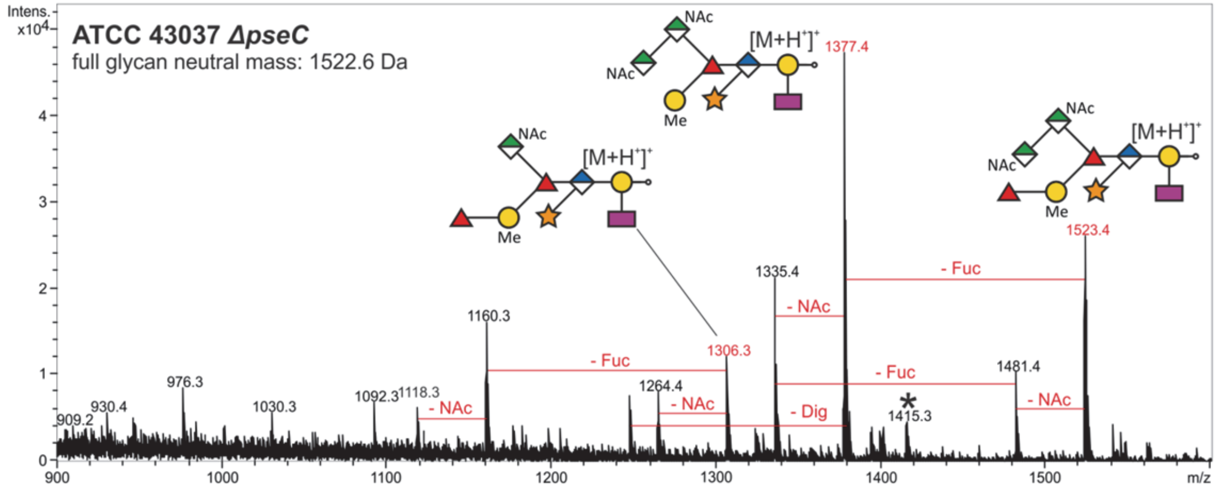

D
ΔlegC
UB4
Primer 92+3
Primer 4+93
Primer V26+V27
Primer 92+3
Primer 4+93
Primer V26+V27
5000 bp
1500 bp
500 bp
ΔlegC_comp
UB4
Primer 76+92
Primer 48+93
Primer 92+93
Primer 76+92
Primer 48+93
Primer 92+93
5000 bp
1500 bp
500 bp

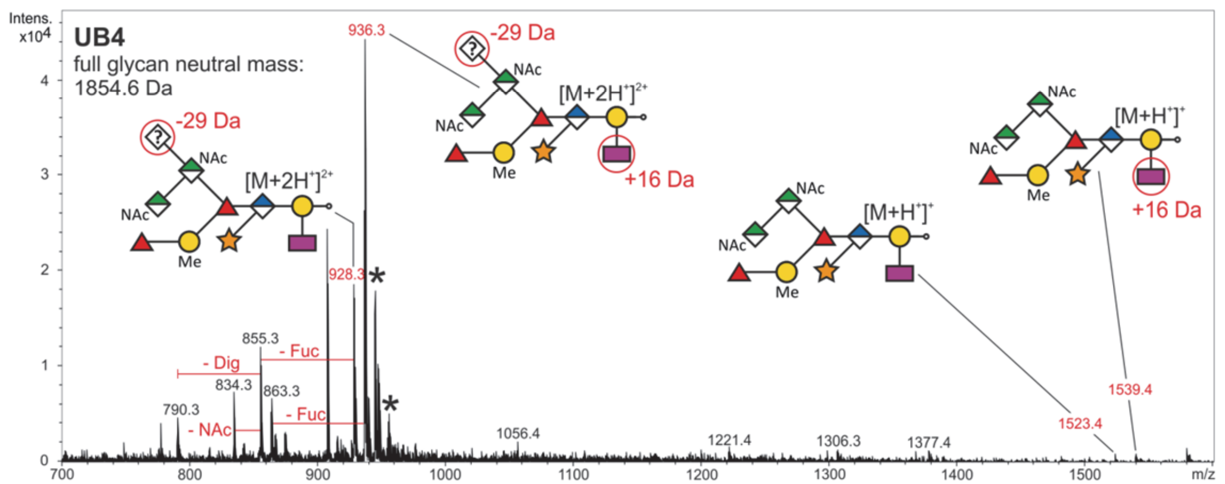
Supp. Fig. 3. Strategy for Tanf_01190 and BFO_1073 gene deletion in *T. forsythia* ATCC 43037 and *T. forsythia* UB4 and confirmation by PCR. (A) Genomic organization of the *pseC* (Tanf_01190) gene showing the upstream and downstream homology regions used for homologous recombination and insertion of the selectable *ermF* cassette disrupting Tanf_01190 (upper lane). Tanf_01190 gene deletion and insertion of the selectable *ermF* cassette is shown in the context of its genomic region (middle lane). Complementation of the Tanf_01190 gene deletion by insertion of the Tanf_01190 gene and the selectable *cat* cassette is shown in the context of its genomic region (lower lane). Primers (compare with Supplemental Table 2) are indicated by arrows. **(B)** Agarose gel electrophoresis (left) to check the deletion of the Tanf_01190 gene upstream (primers 38/3, 1137 bp), downstream (primers 4/10, 1200 bp) and the loss of the gene (primers 15/18, 1164 bp) genomic DNA of *T. forsythia* ATCC 43037 wild-type and mutants with integrated *ermF* cassette. Agarose gel electrophoresis (right) to check complementation of the deleted Tanf_01190 gene using the primers 38/10 (4112 bp for *T. forsythia* ATCC 43037 Δ*pseC* and 3449 bp for *T. forsythia* ATCC 43037 wild-type), primers 15/76 (1815 bp) and primers 75/10 (1838 bp). **(C)** Genomic organization of the *legC* (BFO_1073) gene showing the upstream and downstream homology regions used for homologous recombination and insertion of the selectable *ermF* cassette disrupting BFO_1073 (upper lane). BFO_1073 gene deletion and insertion of the selectable *ermF* cassette is shown in the context of its genomic region (middle lane). Complementation of the BFO_1073 gene deletion by insertion of the BFO_1073 gene and the selectable *cat* cassette is shown in the context of its genomic region (lower lane). Primers (compare with Supplemental Table 2) are indicated with arrows. **(D)** Agarose gel electrophoresis (left) to check the deletion of the BFO_1073 gene upstream (primers 92/3, 1122 bp), downstream (primers 4/93, 1114 bp) and the loss of the gene (primers V26/V27, 1155 bp) genomic DNA of *T. forsythia* UB4 wild-type and mutants with integrated *ermF* cassette. Agarose gel electrophoresis (right) to check complementation of the deleted BFO_1073 gene using the primers 76/92 (2901 bp), primers 48/93 (1132 bp) and primers 92/93 (4002 bp for *T. forsythia* UB4 Δ*legC* and 3340 bp for *T. forsythia* UB4 wild-type).

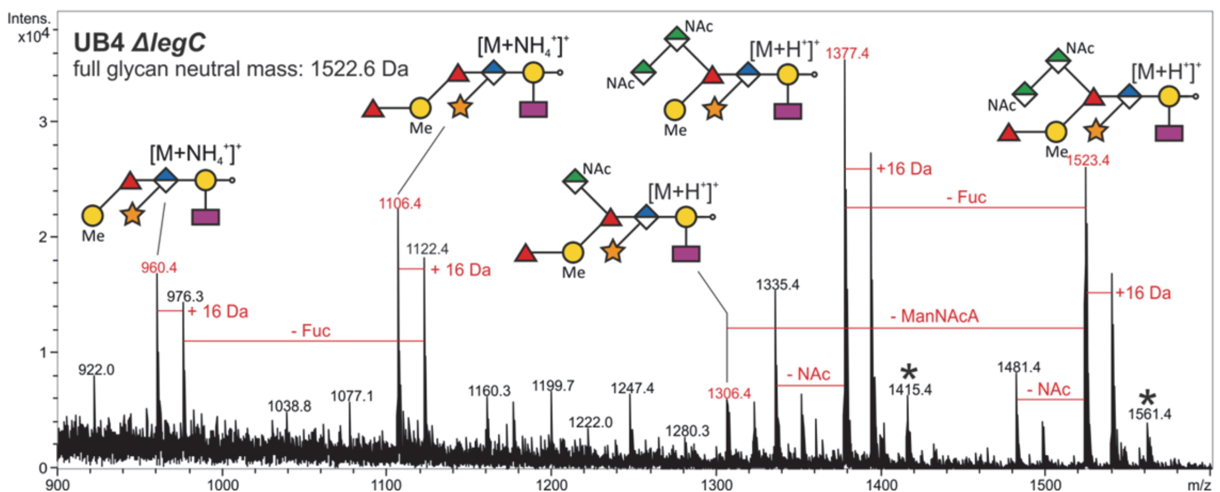


Supp. Fig. 4: Amplification of selected genes from the Pse or Leg biosynthesis pathways in strains *T. forsythia* ATCC 43037, *T. forsythia* FDC 92A2 and *T. forsythia* UB4. In *T. forsythia* ATCC 43037, only genes of the Pse biosynthesis pathway could be amplified. From genomic DNA of the FDC 92A2 and UB4 strains, only genes from the Leg biosynthesis pathway could be amplified.


ATCC 43037
full glycan neutral mass: 1897.7 Da
[M+2H+]2+
949.9
958.4
968.8


ATCC 43037 ΔpseC
full glycan neutral mass: 1522.6 Da
1377.4
1523.4
1306.3
1160.3
1335.4
1481.4
- Fuc
- NAc
- Dig


UB4
full glycan neutral mass:
1854.6 Da
-29 Da
+16 Da
936.3
928.3
1523.4
1539.4
- Fuc
- Dig
- NAc


UB4 ΔlegC
full glycan neutral mass: 1522.6 Da
1377.4
1523.4
1106.4
960.4
1306.4
+16 Da
- Fuc
- ManNAcA
- NAc

Supp. Fig. 5. Deconvoluted ESI-TOF positive-mode MS spectrum of TfsA S-layer protein *O*-glycans in *T. forsythia* ATCC 43037 and UB4 in comparison to the corresponding NulO-negative mutant strains. Compared to the wild-type glycan of ATCC 43037 (m/z 949.9, neutral mass of 1897.7 Da), the glycan of the Δ*pseC* mutant (m/z 1523.4, neutral mass of 1522.6 Da) was found to lack Pse5Am7Gra (m/z 361.14, neutral mass 379.15 Da) and the methyl group on the terminal ManNAcA. This modification is also missing in the *T. forsythia* UB4 wild-type glycan (m/z 928.3, neutral mass 1854.6 Da), which shows an overall similar composition to the ATCC 43037 wild-type, but with a -29 Da mass difference in the terminal NulO (m/z 332.2, neutral mass 350.3 Da). In addition, full-length and truncated glycans exhibiting +16 Da at the digitoxose position were detected (m/z 936.3, neutral mass 1870.6 Da). As expected, glycans of the UB4 Δ*legC* deletion mutant (m/z 1523.4, neutral mass 1522.6 Da) were found to lack the terminal NulO sugar residue and are therefore identical to those of the Δ*pseC* mutant. Asterisks denote adduct peaks. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (PMID 26543186, Glycobiology 25: 1323–1324, 2015).

Supp. Table I. CE-MS data of *T. forsythia* pseudaminic acid biosynthetic intermediates produced in this study. Note, we typically do not observe compound V_P (2,4-diacetamido-2,4,6-trideoxy-L-altropyranose) by this CE-MS analysis, but the expected *m/z* ion for UDP (403.2) was observed for PseG reactions

Compound	**Observed *m/z***	**Calculated mass (Da)**	**Formula (M)**	**Comments**
UDP-GlcNAc (I_P)	606.0	607.4	$C_{17}H_{27}N_3O_{17}P_2$	$[M-H]^-$
UDP-2-acetamido-2,6-dideoxy-β-L-*arabino*-hexos-4-ulose (II_P)	588.0	589.3	$C_{17}H_{25}N_3O_{16}P_2$	$[M-H]^-$
UDP-4-amino-4,6-dideoxy-β-L-AltNAc (III_P)	589.2	590.4	$C_{17}H_{28}N_4O_{15}P_2$	$[M-H]^-$
UDP-2,4-diacetamido-2,4,6-trideoxy-β-L-altropyranose (IV_P)	631.4	632.4	$C_{19}H_{30}N_4O_{16}P_2$	$[M-H]^-$
Pse5,7Ac_2 (VI_P)	333.1	334.3	$C_{13}H_{22}N_2O_8$	$[M-H]^-$

Supp. Table II. Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype and/or relevant characteristics	Source
Tannerella forsythia ATCC 43037	wild-type isolate	ATCC
Tannerella forsythia FDC 92A2	clinical isolate	Graham Stafford
Tannerella forsythia UB4	clinical isolate	Ashu Sharma
Escherichia coli DH5α	F^- φ80d*lacZ* M15 (*lacZYA-argF*)*U169 deoR recA1 endA1 hsdR17* (rK^-mK^-) *phoA supE44 thi-1 gyrA96 relA1*$^-$	Invitrogen
Escherichia coli BL21 (DE)	F^-*, ompT, hsdS* (rB^-mB^-), *gal, dcat* (DE3)	Invitrogen
Escherichia coli BL21-CodonPlus(DE3)-RIL	F^-*, ompT, hsdS* (rB^-mB^-), *dcm*$^+$, Tet^r, *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^r]	Novagen
T. forsythia ATCC 43037 Δ*pseC*	*T. forsythia* knockout of the *pseC* gene; Erm^r	This study
T. forsythia ATCC 43037 Δ*pseC*$_{comp}$	Complemented *T. forsythia* ATCC 43037 Δ*pseC*; Cat^r	This study
T. forsythia UB4 Δ*legC*	*T. forsythia* knockout of the *legC* gene; Erm^r	This study
T. forsythia UB4 Δ*legC*$_{comp}$	Complemented *T. forsythia* UB4 Δ*legC*; Cat^r	This study
pJET1.2_pseC_ermF	pJET1.2 carrying the *pseC*_P(ermF)_ermF knockout cassette	This study
pJET1.2_legC_ermF	pJET1.2 carrying the *legC*_P(ermF)_ermF knockout cassette	This study
pJET1.2_pseC_up+gene_cat	pJET1.2 carrying the *pseC*_up region, *pseC* gene and the *cat* gene	This study
pJET1.2_pseC_cat	pJET1.2 carrying the *pseC_cat* complementation cassette	This study

pJET1.2_legC_up+gene_cat	pJET1.2 carrying the *legC*_up region, *legC* gene and the *cat* gene	This study
pJET1.2_legC_cat	pJET1.2 carrying the *legC_cat* complementation cassette	This study
pET28a	Expression vector with a His_6-tag, Kan^r	Novagen
pET22b	Expression vector with a His_6-tag, Amp^r	Novagen
pEXALV	*Paenibacillus alvei* expression vector used for the amplificationof the *cat* gene	(Zarschler, K., Janesch, B., et al. 2009)
pMAL_2cE	Expression vector with maltose-binding protein (N-term), Amp^r	NEB
pCR2.1	Cloning vector, Amp^r, Kan^r, oriColE1, *lac* promoter	Invitrogen
pET28a_pseB_His	pET28a carrying the His_6-tagged *pseB* gene, Kan^r	This study
pMAL_2cE_pseC	pMAL_2cE carrying fusion construct of maltose binding protein and *pseC* gene, Amp^r	This study
pET28a_pseH_His	pET28a carrying the His_6-tagged *pseH* gene, Kan^r	This study
pMAL_2cE_pseG	pMAL_2cE carrying fusion construct of MBP and *pseG* gene, Amp^r	This study
pET28a_pseI_His	pET28a carrying the His_6-tagged *pseI* gene, Kan^r	This study
pMAL_2cE_pseF	pMAL_2cE carrying fusion construct of MBP and *pseF* gene, Amp^r	This study
TF2075C-pet22b	pET22b carrying the His_6-tagged *legB* gene, Amp^r	This study
TF2074N-pet22b	pET22b carrying the His_6-tagged *legC* gene, Amp^r	This study
BFO_1065N-pet22b	pET22b carrying the His_6-tagged *legG* gene, Amp^r	This study

Supp. Table III. Oligonucleotide primers used for PCR amplification reactions

#	Primers	Sequence (5'->3')
11	pseB_C_NcoI_for	aatcaCCATGGTAAATAATAAATCCATATTGATTACCGGG
12	pseB_C_XhoI_rev	aatcaCTCGAGTTAATGGTGATGGTGATGGTGTGCGTTAAAATGAGGATCAACATATTTTC
80	pseC_BamHI_for	aatcaGGATCCATGATAACAAGACCGATACC
86	pseC_HindIII_rev	aatcaAAGCTTCTATTTGATAAAAAAATTAATCACTTTCTC
28	pseH_C_NcoI_for new	aatcaCCATGGCAAAAAAAAAGATATTCGAAATTGGAG
22	pseH_N_XhoI_rev	aatcaCTCGAGTCAGATTTTTGATTCCCTACAAAACCG
81	pseG_BamHI_for	aatcaGGATCCATGGGCAAACGTCAGATTATATTTCG
105	pseG_HindIII_rev full	aatcaAAGCTTTTATATTCCGTGAAAAGCATGGA
33	pseI_C_BsaINcoI_for	aatcaGGTCTCCCATGGAGACAACAATTGTAGCAGAACTATCAGC
34	pseI_C_XhoI_rev	aatcaCTCGAGTCAATGGTGATGGTGATGGTGTTTCTTAATGTAATCTAAGTCCAACC
23	pseF_C_NcoI_for	aatcaCCATGGAAATATTGGCTATTACTCAGGC
24	pseF_C_XhoI_rev	aatcaCTCGAGTTAATGGTGATGGTGATGGTGTCTTATTTTTAAATCATTATTTATTGAC
101	pseF_BamHI_for	aatcaGGATCCATGAAAATATTGGCTATTACTCAGGC
102	pseF_XbaI_rev	aatcaTCTAGATTATCTTATTTTTAAATCATTATTTATTGAC
V83	BFO1074_synth_for	tgcaCATATGCAAACAATACTAGTTACC
V84	BFO1074_synth_rev	ggcCTCGAGGGCCTCGAGCACGTTGTATATCCCGGATTTA
V133	BFO1073_synth_for	catgCATATG**GGCCATCATCATCATCATCACGGCAGC**ATGGAGAGCAGCTATAAGAAA
V134	BFO1073_synth_rev	catgCATGCTCGAGTCACTTTTTCAGATTTACACTAC
V114	BFO1065_synth_for	gtactaCATATG**GGCCATCATCATCATCATCACGGCAGC**CGTAAAGTTTGTGTTGTTACC

V115	BFO1065_synth_rev	agtcatCTCGAGTCATTACAGGTTATAGAAGCTTTTCT
1	P(ermF)_for	GGTACCCCCGATAGCTTCCGCTATTGC
2	ermF_rev	CTACGAAGGATGAAATTTTTCAGGG
3	P(ermF)_rev	GCAATAGCGGAAGCTATCGGGGGTACC
4	ermF_for	CCCTGAAAAATTTCATCCTTCGTAG
5	pseC_up_for	CCATATTGATTACCGGGGGCAC
6	pseC_up_OE_rev	GCAATAGCGGAAGCTATCGGGGGTACCTATGCGTTAAAATGAGGATCAAC
7	pseC_down_OE_for	GGAAGTTGTCCCTGAAAAATTTCATCCTTCGTAGACAGGATATAACAAATCGAAAGGAAACGATTCC
8	pseC_down_rev	CGCCGGTAAACGTGTAGAACCATATCTTGCCTGAG
38	KO_pseC_for3	GTTGCTGCCTTTTATATTGGGATATTATGG
10	KO_pseC_rev	GCTCCCTCTTCGTTAGTCGTTGC
75	cat_for neu	ATGAACTTTAATAAAATTGATTTAGACAATTGG
76	cat_rev_neu	TTATAAAAGCCAGTCATTAGGCCTATCTGAC
46	pseC_down_OE_cat_for	GGCCTAATGACTGGCTTTTATAAACAGGATATAACAAATCGAAAGG
77	cat_KpnI_SphI_rev	aatcaGCATGCGGTACCTTATAAAAGCCAGTCATTAGGCCTATCTGAC
78	pseC_down_KpnI_for	aatcaGGTACCACAGGATATAACAAATCGAAAGGAAACG
79	pseC_down_SphI_rev	aatcaGCATGCCGCCGGTAAACGTGTAGAACCATATCTTGCC
15	pseC_C_NcoI_for	aatcaCCATGGTAACAAGACCGATACCATATGG
18	pseC_N_XhoI_rev	aatcaCTCGAGCTATTTGATAAAAAAATTAATCACTTTCTC
88	legC_up_for	GCCTTGGTTTTGTGGCAAGAAAAATC
89	legC_up_OE_rev	GCAATAGCGGAAGCTATCGGGGGTACCAGACGCTACACGTTGTATATCC

90	legC_down_OE_for	CCCTGAAAAATTTCATCCTTCGTAGATTATGGAATTAAAAGAGTTTATTG
91	legC_down_rev	CGCTTCCATCTGTTGATACTTACG
92	KO_legC_for	CTGCTAAGATCATATTAGTTGCTGC
93	KO_legC_rev	CAGACCATTCCGGAAACTCTTCGGC
94	legC_down_KpnI_for	aatcaGGTACCATTATGGAATTAAAAGAGTTTATTG
95	legC_down_SphI_rev	aatcaGCATGCCGCTTCCATCTGTTGATACTTACG
96	legC_OE_cat_rev	CTAAATCAATTTTATTAAAGTTCATTCACTTTTTCAGATTTACACTAC
48	cat_down_for	GTCAGATAGGCCTAATGACTGGC
V26	BFO1073_NdeI_for	gactCATATGGAGAGCAGCTATAAGAAAATAACAG
V27	BFO1073_XhoI_rev	gcatCTCGAGCTTTTTCAGATTTACACTACTAGGG
V63	BFO1063_NdeI_for	gccgCATATGATAAATACGGATAAAAAGATTTTGG
V64	BFO1063_XhoI_rev	gcatCTCGAGGATATAGATTGCTCCATTATATTCGTAATAAAC

[a] Introduced restriction sites are underlined, primer-encoded His-tags are in **bold**, lower case letters indicate artificially introduced bases to improve restriction enzyme cutting.