

Multiple evolutionary origins reflect the importance of sialic acid transporters in the colonisation potential of bacterial pathogens and commensals.

Emmanuele Severi^{1,+*}, Michelle Rudden^{2,+}, Andrew Bell³, Tracy Palmer¹, Nathalie Juge³, and Gavin H Thomas^{2,*}

¹Microbes in Health & Disease, Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, UK

²Department of Biology, University of York, York, UK

³The Gut Microbes and Health Institute Strategic Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

*gavin.thomas@york.ac.uk, emmanuele.severi@newcastle.ac.uk

+these authors contributed equally to this work

Supplementary Information

Supplementary Figures ([S1](#), [S2](#), [S3](#),[S4](#),[S5](#),[S6](#),[S7](#),[S8](#),[S9](#),[S10](#),[S11](#),[S12](#))

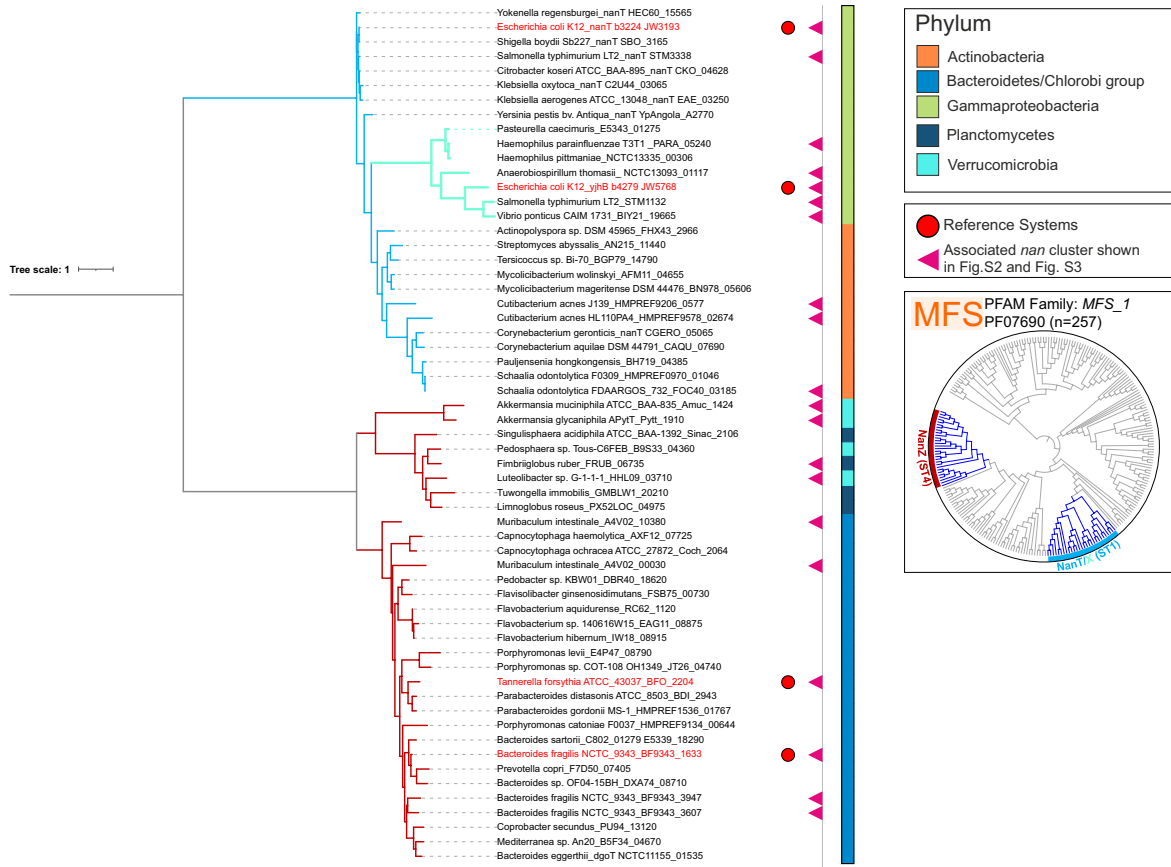


Figure S1 Phylogeny of MFS sialic acid transporters of the ST1 and ST4 families. Maximum likelihood phylogeny of MFS₁ sialic acid transporters that are linked to *nan* genes. ST1 sialic acid transporters clearly separate into two distinct clades, NanT and NanX. This broadly reflects two different substrate affinities where NanX has been shown to transport 2,7-anhydro-Neu5Ac/Neu5Ac2en and NanT selectively transports Neu5Ac (see Text). NanZ is found widely among *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia*, and has been shown to transport Neu5Ac in *Bacteroidetes* (see main text). Reference systems are highlighted in red.

ST1 - NanT

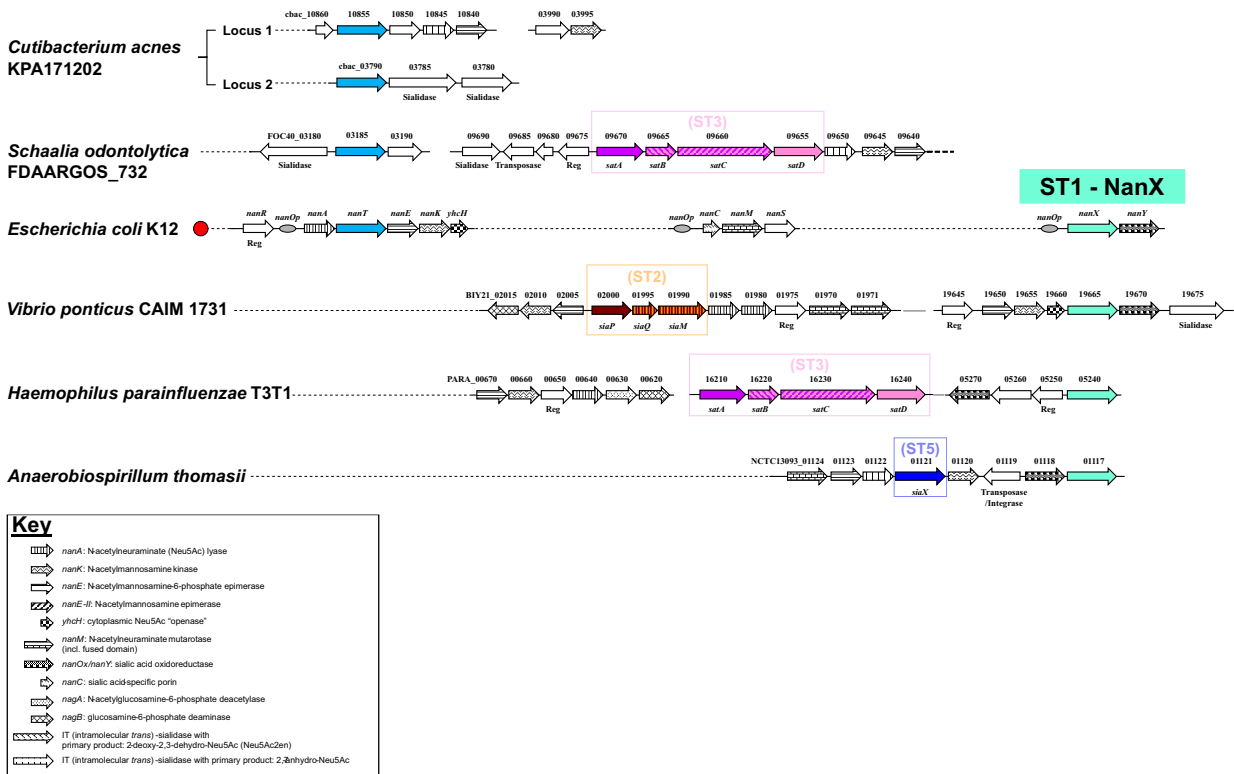


Figure S2 Organisation of ST1 (*nanT/X*) gene clusters in selected bacteria. ST1 transporter genes *nanT* and *nanX* are colour-coded. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted with a red circle. *nanOp* operators highlighted in the *E. coli* ST1 loci emphasise the occurrence of a single NanR regulon in this organism. Reg: regulator gene; HGT: horizontal gene transfer.

ST4 - NanZ

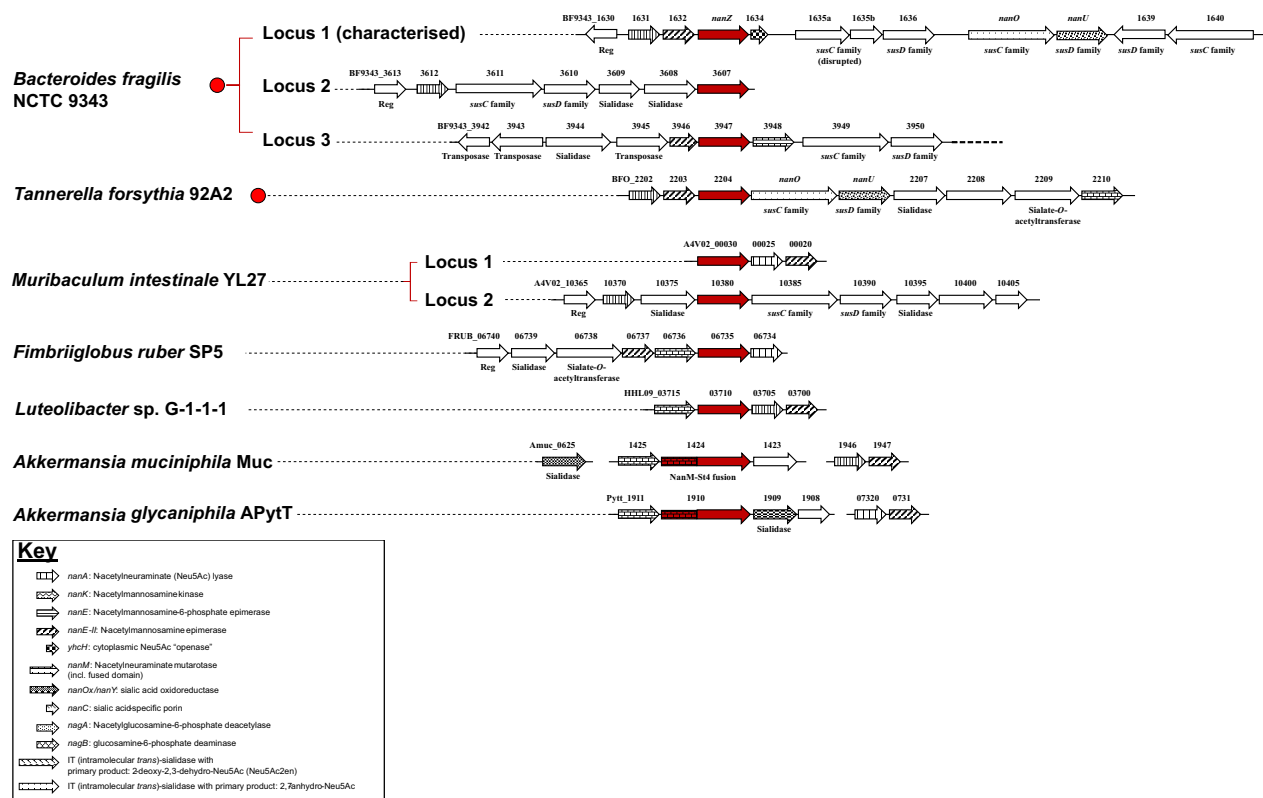
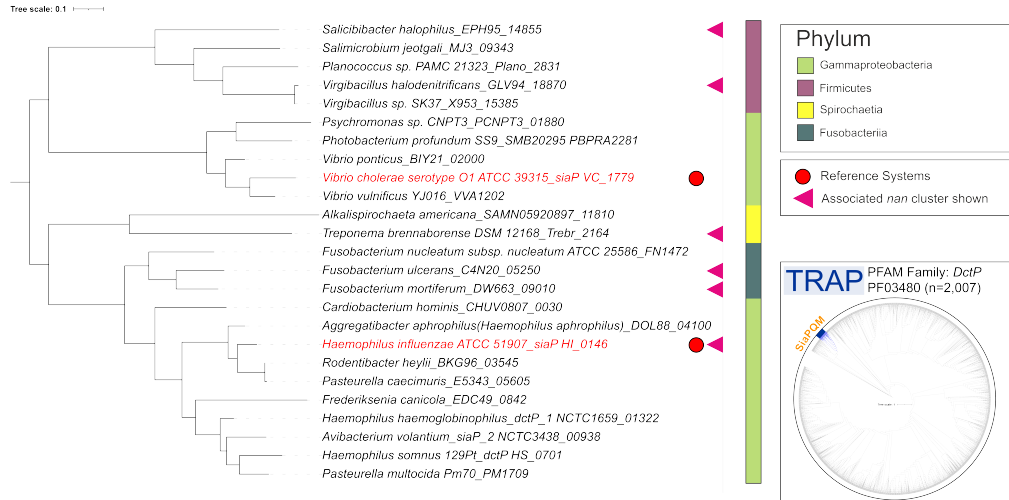


Figure S3 Organisation of ST4 (*nanZ*) gene clusters in selected bacteria. NanZ is encoded in gene clusters with other catabolic *nan* genes required for sialic acid metabolism. *nanZ* is colour-coded, catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Interestingly, there are two novel examples of a NanM-NanZ mutarotase-transporter fusion found in *Akkermansia* spp. (NanM-like domains are indicated by the appropriate pattern). Reg: regulator gene; HGT: horizontal gene transfer. SusCD: outer membrane protein complex for glycan acquisition made of a TonB-dependent transporter (SusC) and an extracytoplasmic lipoprotein (SusD). NanOU is an experimentally confirmed sialic acid-specific SusCD-family complex¹. We did not establish orthologous groups for all other, more divergent SusCD-family complexes in this figure.

A



B

ST2 - SiaPQM

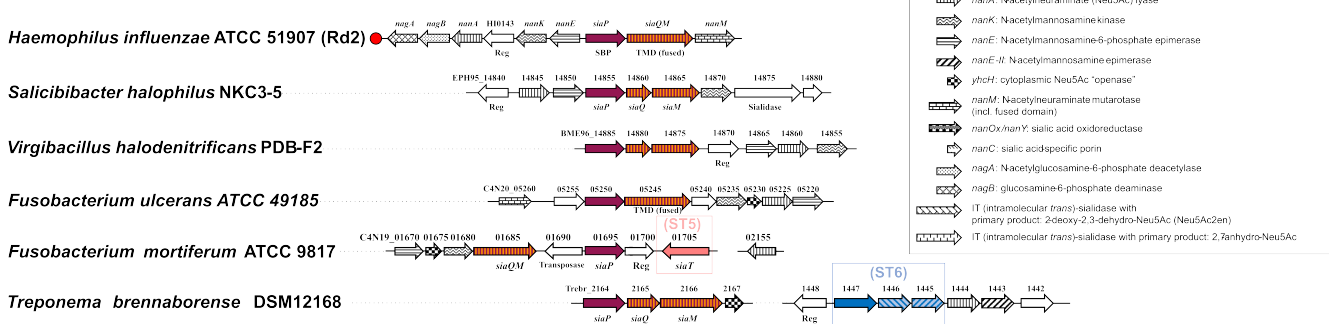


Figure S4 Phylogeny of the ST2-SBP, SiaP, and organisation of ST2 (*siaPQM*) gene clusters in selected bacteria. (A) Maximum likelihood phylogeny of sialic acid-binding extracytoplasmic protein, SiaP. Reference organisms are highlighted in red. (B) Genetic organisation of *siaPQM* transporter genes within *nan* clusters. The figure includes the finding of ST2 genes in *Firmicutes*, reported here for the first time², but also includes *Fusobacterium* spp. and *T. brennaborensis* discussed in the Text. ST2 transporter genes are colour-coded, patterns are used to distinguish among different subunits. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted in red. Reg: regulator gene; SBP: solute-binding protein; TMD: transmembrane domain; HGT: horizontal gene transfer.

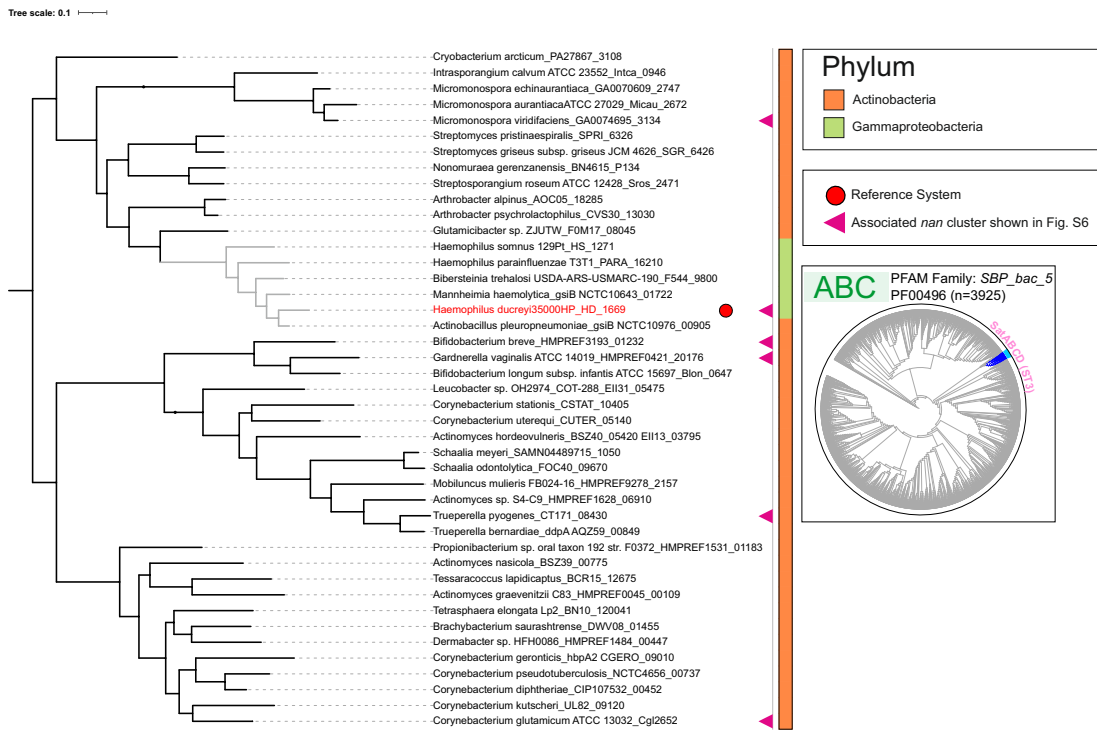


Figure S5 Phylogeny of the ST3-SBP, SatA. Maximum likelihood phylogeny of sialic acid-binding protein component (SatA) of the ST3 ABC transport system, SatABCD. Reference organisms are highlighted in red. Note how sequences from *Pasteurellaceae* form a small clade (in grey) within a diverse family of ST3 transporters that is otherwise exclusive to *Actinobacteria*.

ST3 SAT1 - SatABCD

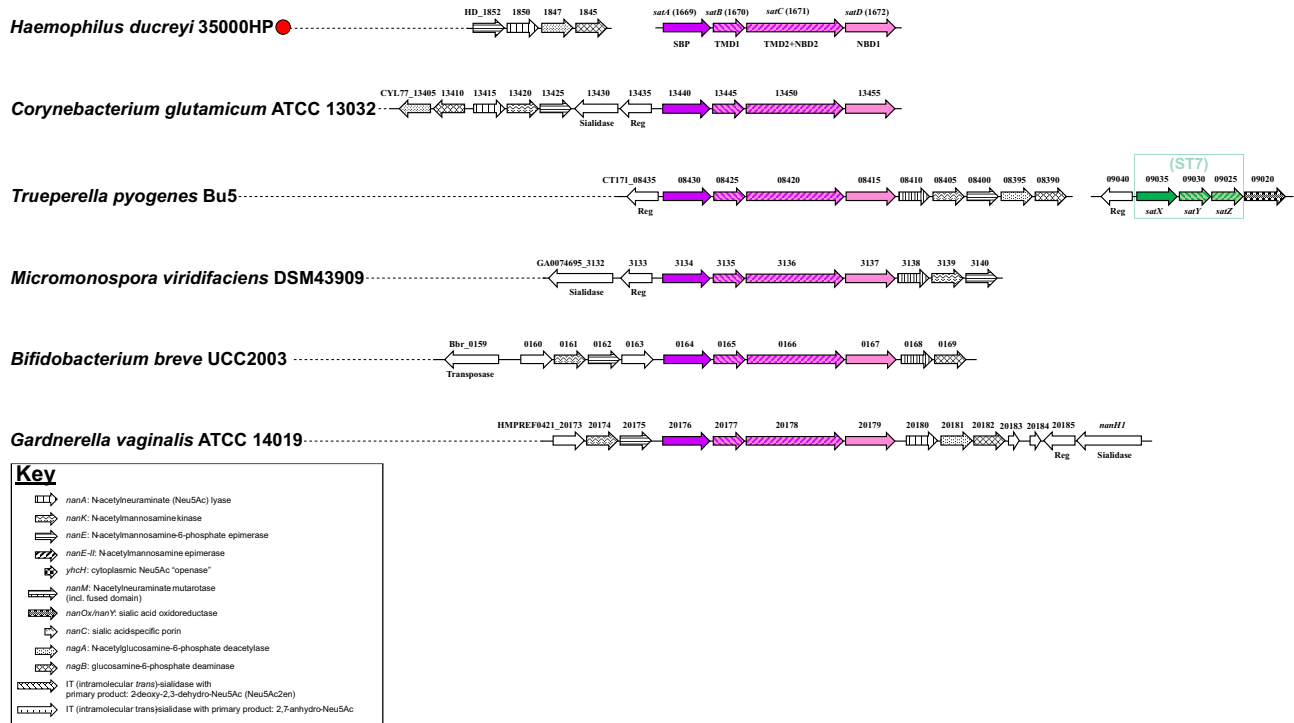
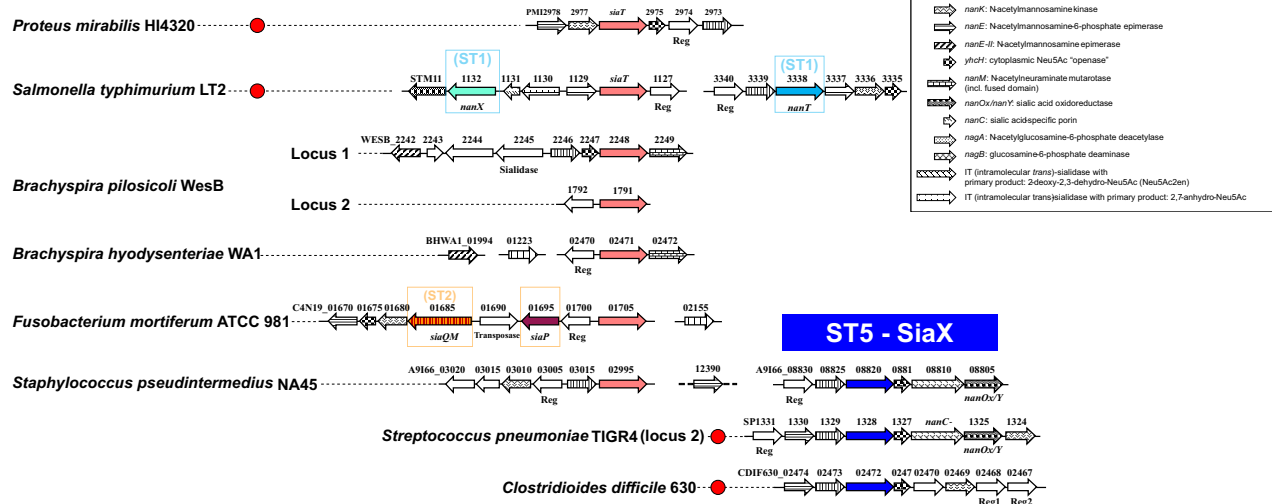


Figure S6 Organisation of ST3 (*satABCD*) gene clusters in selected bacteria. ST3 transporter genes are colour-coded, patterns are used to distinguish among different subunits. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted with a red circle. Note that no *nanK* orthologues occur in *H. ducreyi*³ and this work, and genes encoding for ManNAc kinase function are yet to be identified (see also Fig. 2). Reg: regulator gene; SBP: solute-binding protein; TMD: transmembrane domain; NBD: nucleotide-binding domain (ATP-ase protein); HGT: horizontal gene transfer.

ST5 - SiaT



ST5 - SiaF

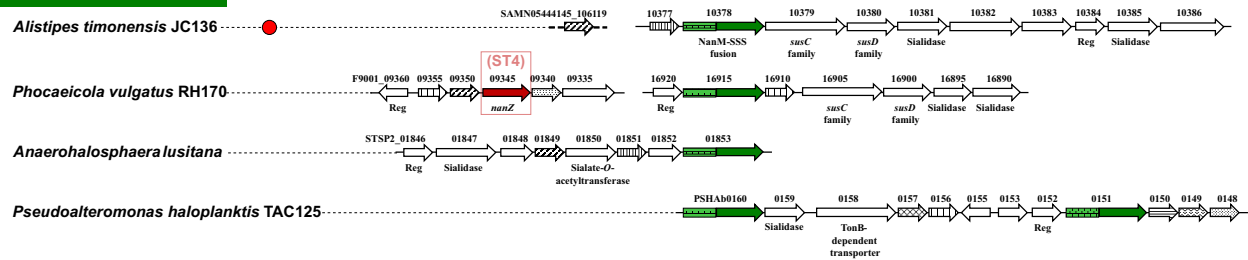


Figure S7 Organisation of ST5 (*siaT/siaX/siaF*) gene clusters in selected bacteria. Based on phylogeny, the ST5/SSS transporters can be resolved into three distinct clades (Fig. 4). The different transporter genes are colour-coded based on their clade within the ST5 phylogeny. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted by a red circle. SiaT clusters are widespread in bacteria and all characterised transporters from these clusters have been shown to transport Neu5Ac. The *siaT* sialic acid utilisation cluster of *B. pilosicoli* B2904 (not shown) is identical to locus 1 of strain WesB, but locus 2 is present only in WesB. The SiaX clade is common among *Firmicutes* and includes both Neu5Ac and potential anhydro-Neu5Ac transporters. As for the Neu5Ac transporters of this clade, there is an example of demonstrated Neu5Ac specificity for the *C. difficile* transporter. Candidate anhydro-Neu5Ac transporters are genetically linked to anhydro-Neu5Ac metabolic enzymes, namely the oxidoreductase NanOx/NanY and an IT-sialidase, as seen here in *S. pneumoniae* TIGR4 and *S. pseudintermedius* NA45 (the latter organism also possesses a *siaT* orthologous gene at a different locus). The SiaF clade, occurring near-exclusively in *Bacteroidetes* and *Planctomycetes/Verrucomicrobia*, feature proteins that are all fusions between the sialic acid mutarotase NanM (pattern-coded accordingly) and a ST5 transporter. Reg: regulator gene; HGT: horizontal gene transfer. SusCD: outer membrane protein complex for glycan acquisition made of a TonB-dependent transporter (SusC) and an extracytoplasmic lipoprotein (SusD).

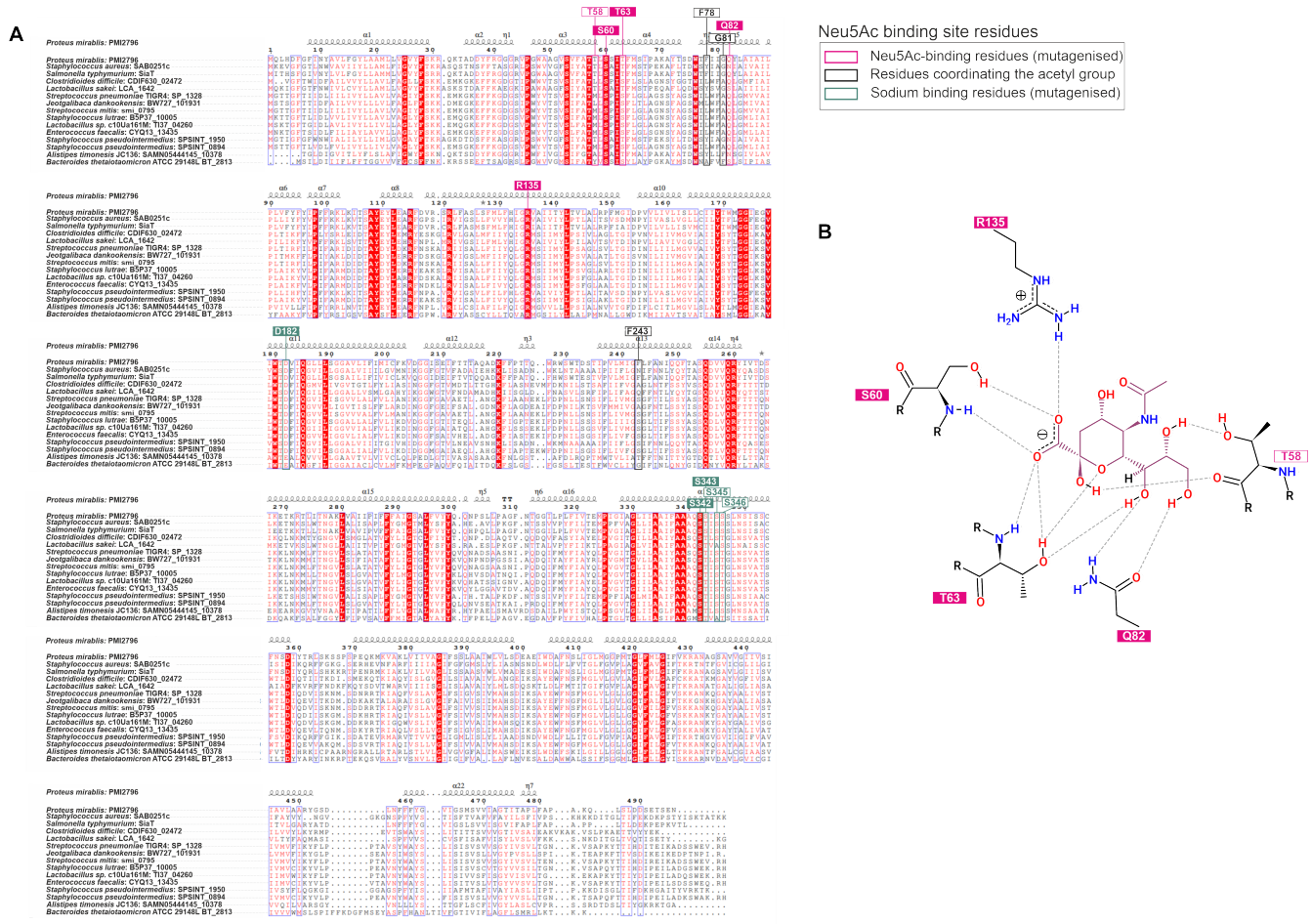


Figure S8 The sialic acid binding-site is conserved across ST5 sialic acid transporters. (A) Clustal alignment of ST5 sequences highlighting the Neu5Ac-binding residues (red and black boxes) and selected Na⁺ binding residues (green boxes) in PmSiaT⁴. Residues in red and green boxes have been subjected to mutagenesis, with deleterious effects only seen for residues in block colour⁴. The alignment, visualised with ESPrnt (see Methods) for mapping to the PmSiaT structure (Table 1), includes all candidate anhydro-Neu5Ac transporters of the SiaX clade listed in Fig. 4, all characterised SiaT and SiaX Neu5Ac transporters (*P. mirabilis*, *S. typhimurium*, *S. aureus*, *L. sakei*, and *C. difficile*), the SiaF protein from *A. timonensis*, and finally the uncharacterised ST5 transporter from *B. thetaiotaomicron* (see Discussion). **(B)** Co-ordination of Neu5Ac by the binding site residues in PmSiaT (adapted from⁴).

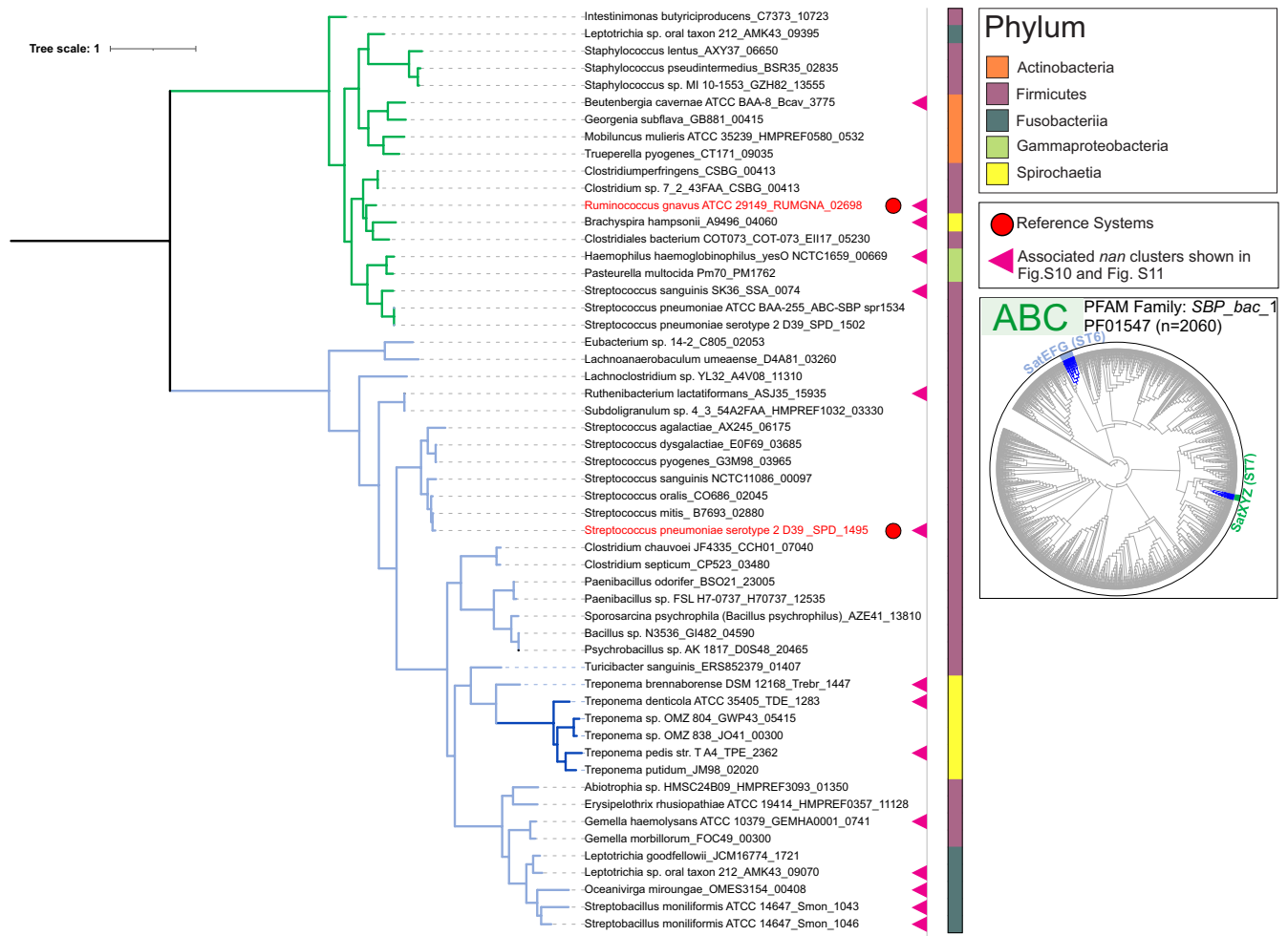


Figure S9 Phylogeny of the ST6-SBP (SatE) and the ST7-SBP (SatX) components of ABC sialic acid transporters. Maximum likelihood phylogeny of sialic acid-binding protein components (SatE and SatX, respectively) of the ST6 and ST7 ABC transport systems. Reference organisms are highlighted in red. *Treponema* spp. possessing “orphan” *satE* genes linked to genes for methyl-accepting proteins (see main text) are marked out by dark blue branches.

ST6 SAT3 – SatEFG(H)

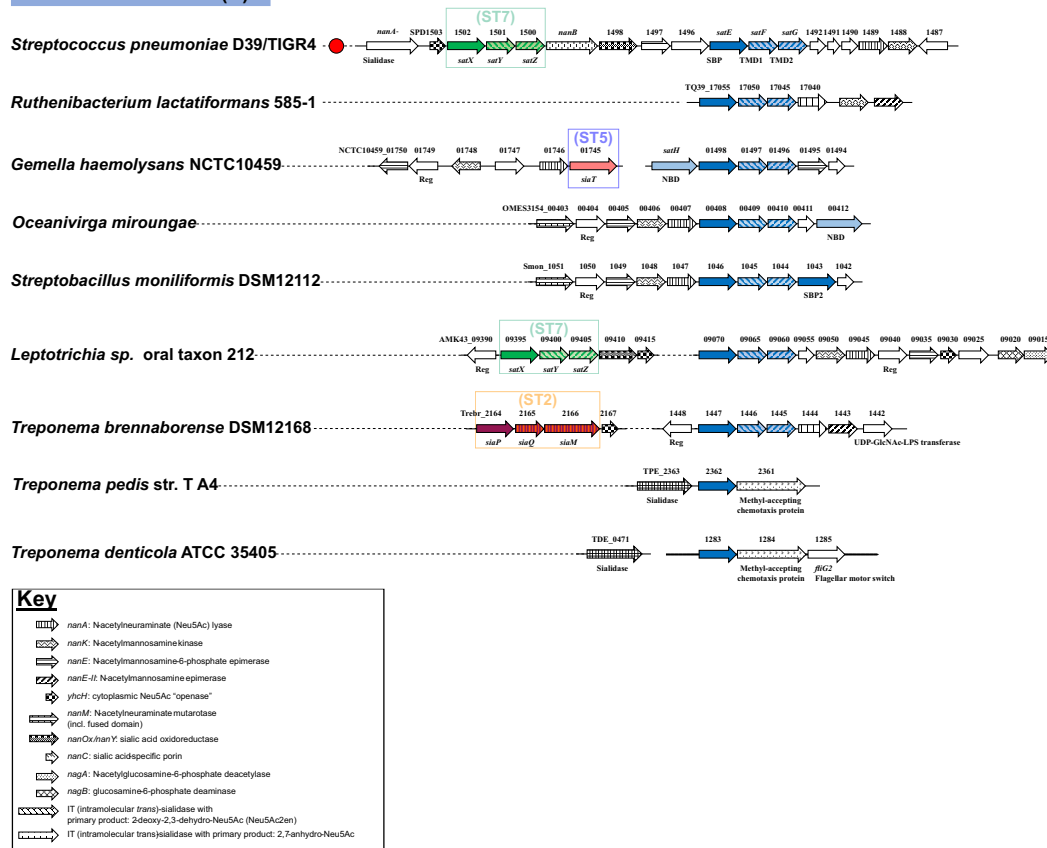
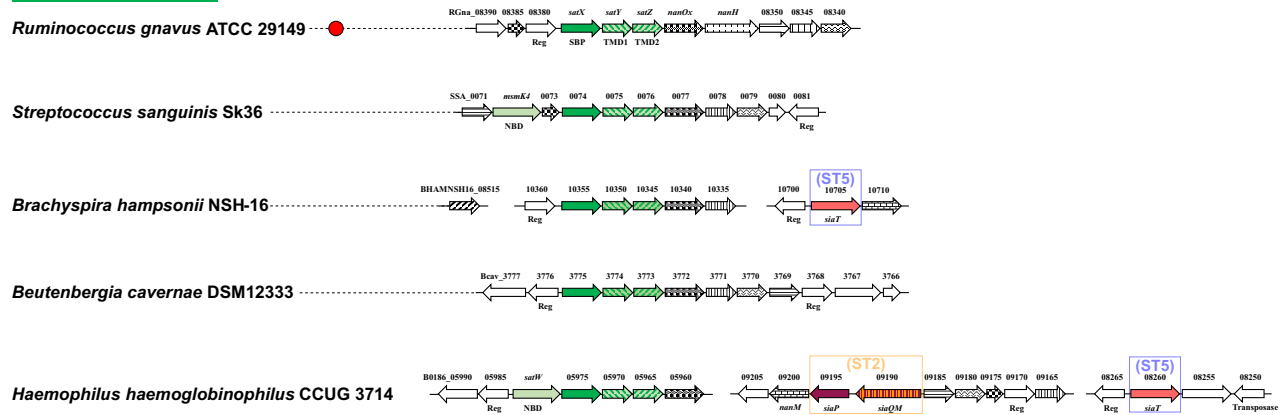


Figure S10 Organisation of ST6 (*satEFG*) gene clusters in selected bacteria. ST6 transporter genes are colour-coded, patterns are used to distinguish among different subunits. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted with a red circle. While ST6 transporter loci normally do not contain an NBD gene, relying instead on multitask *msmK*-like genes for function⁵, note here two cases where specific NBD genes have been recruited to the cluster. The incomplete SAT3 cluster of the sialic acid-utilising organism, *Ruthenibacterium lactatiformans* strain 585-1⁶ was completed by matching it to a near-identical locus in the sequenced genome of strain 82B1, which was then used to annotate the remaining genes. Some *Treponema* spp. only contain the *satE* gene coding for the SBP component (Fig. S9), which however possesses a conserved genetic link with a methyl-accepting chemotaxis protein (MCP), suggesting a role in sialic acid sensing independent of a role in uptake. In *T. denticola*, the ST6 locus also contains an allelic variant of the flagellar motor switch protein gene, *fliG*. Reg: regulator gene; SBP: solute-binding protein; TMD: transmembrane domain; NBD: nucleotide-binding domain (ATP-ase protein); HGT: horizontal gene transfer.

ST7 – SatXYZ(W)



Key

	<i>nanA</i> : N-acetylneuraminase (Neu5Ac) lyase
	<i>nanK</i> : N-acetylmannosamine kinase
	<i>nanE</i> : N-acetylmannosamine-6-phosphate epimerase
	<i>nanE-II</i> : N-acetylmannosamine epimerase
	<i>yhcH</i> : cytoplasmic Neu5Ac "openase"
	<i>nanM</i> : N-acetylneuraminase mutarotase (incl. fused domain)
	<i>nanOx/nanY</i> : sialic acid oxidoreductase
	<i>nanC</i> : sialic acid-specific porin
	<i>nagA</i> : N-acetylglucosamine-6-phosphate deacetylase
	<i>nagB</i> : glucosamine-6-phosphate deaminase
	IT (intramolecular trans)-sialidase with primary product: 2-deoxy-2,3-dehydro-Neu5Ac (Neu5Ac2en)
	IT (intramolecular trans)sialidase with primary product: 2,7-anhydro-Neu5Ac

Figure S11 Organisation of ST7 (*satXYZ*) gene clusters in selected bacteria. ST7 transporter genes are colour-coded, patterns are used to distinguish among different subunits. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted with a red circle. While ST7 loci are normally devoid of NBD genes, presumably relying on multitask *msmK*-like for function as shown for *S. pneumoniae* SatEFG⁵, note here two cases where different NBD genes have been recruited to the cluster, namely in *H. haemoglobinophilus* (*satW*) and *S. sanguinis* SK36 (*msmK4* coding for supernumerary allelic variant of *msmK*). Reg: regulator gene; SBP: solute-binding protein; TMD: transmembrane domain; NBD: nucleotide-binding domain (ATP-ase protein); HGT: horizontal gene transfer.

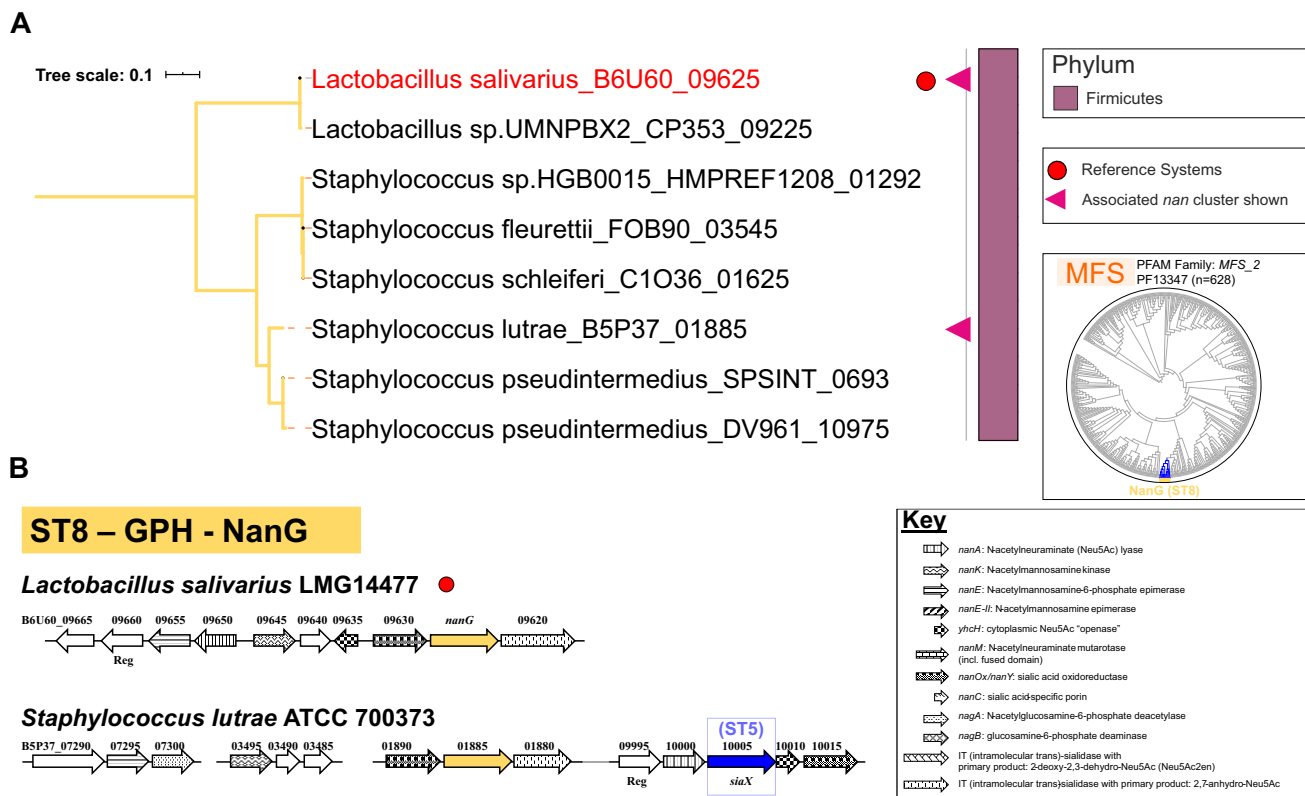


Figure S12 Phylogeny of ST8 (GPH) sialic acid transporters, and organisation of ST8 (*nanG*) clusters in selected bacteria. ST8 transporter genes are colour-coded. Catabolic *nan* genes are pattern-coded and described in the key. For clarity other conserved genes recurring in these clusters are not pattern-coded except when emphasising specific genetic links. The function of other notable genes is reported (without pattern) if relevant for cluster annotation or to exemplify instances of potential HGT. Reference gene clusters are highlighted in red. Note that *S. lutrae* possesses two predicted anhydro-sialic acid transporter genes. Reg: regulator gene; HGT: horizontal gene transfer.

References

1. Phansopa, C. *et al.* Structural and functional characterization of NanU, a novel high-affinity sialic acid-inducible binding protein of oral and gut-dwelling Bacteroidetes species. *Biochem. J.* **458**, 499–511, DOI: [10.1042/BJ20131415](https://doi.org/10.1042/BJ20131415) (2014).
2. Chowdhury, N. *et al.* The VC1777–VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*. *Microbiology* **158**, 2158–2167, DOI: [10.1099/mic.0.059659-0](https://doi.org/10.1099/mic.0.059659-0) (2012).
3. Post, D. M. B., Mungur, R., Gibson, B. W. & Munson, R. S. Identification of a novel sialic acid transporter in *Haemophilus ducreyi*. *Infect. Immun.* **73**, 6727–6735, DOI: [10.1128/IAI.73.10.6727-6735.2005](https://doi.org/10.1128/IAI.73.10.6727-6735.2005) (2005).
4. Wahlgren, W. Y. *et al.* Substrate-bound outward-open structure of a Na⁺-coupled sialic acid symporter reveals a new Na⁺ site. *Nat. Commun.* **9**, 1753, DOI: [10.1038/s41467-018-04045-7](https://doi.org/10.1038/s41467-018-04045-7) (2018).
5. Marion, C., Aten, A. E., Woodiga, S. A. & King, S. J. Identification of an ATPase, MsmK, Which Energizes Multiple Carbohydrate ABC Transporters in *Streptococcus pneumoniae*. *Infect. Immun.* **79**, 4193–4200, DOI: [10.1128/IAI.05290-11](https://doi.org/10.1128/IAI.05290-11) (2011).
6. Pereira, F. C. *et al.* Rational design of a microbial consortium of mucosal sugar utilizers reduces *Clostridiodes difficile* colonization. *Nat. Commun.* **11**, 5104, DOI: [10.1038/s41467-020-18928-1](https://doi.org/10.1038/s41467-020-18928-1) (2020).