MICROBIAL GENOMICS

RESEARCH ARTICLE

Severi *et al.*, *Microbial Genomics* 2021;7:000614
DOI 10.1099/mgen.0.000614

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

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

Emmanuele Severi[1,2,*,†], Michelle Rudden[1†], Andrew Bell[3], Tracy Palmer[2], Nathalie Juge[3] and Gavin H. Thomas[1,*]

Abstract

Located at the tip of cell surface glycoconjugates, sialic acids are at the forefront of host–microbe interactions and, being easily liberated by sialidase enzymes, are used as metabolites by numerous bacteria, particularly by pathogens and commensals living on or near diverse mucosal surfaces. These bacteria rely on specific transporters for the acquisition of host-derived sialic acids. Here, we present the first comprehensive genomic and phylogenetic analysis of bacterial sialic acid transporters, leading to the identification of multiple new families and subfamilies. Our phylogenetic analysis suggests that sialic acid-specific transport has evolved independently at least eight times during the evolution of bacteria, from within four of the major families/superfamilies of bacterial transporters, and we propose a robust classification scheme to bring together a myriad of different nomenclatures that exist to date. The new transporters discovered occur in diverse bacteria, including *Spirochaetes*, *Bacteroidetes*, *Planctomycetes* and *Verrucomicrobia*, many of which are species that have not been previously recognized to have sialometabolic capacities. Two subfamilies of transporters stand out in being fused to the sialic acid mutarotase enzyme, NanM, and these transporter fusions are enriched in bacteria present in gut microbial communities. Our analysis supports the increasing experimental evidence that competition for host-derived sialic acid is a key phenotype for successful colonization of complex mucosal microbiomes, such that a strong evolutionary selection has occurred for the emergence of sialic acid specificity within existing transporter architectures.

DATA SUMMARY

For all phylogenetic trees included in the study, the local tag or National Center for Biotechnology Information (NCBI) protein identifier is provided to retrieve the sequences from the NCBI database. All gene layout figures also contain locus tags enabling direct retrieval from the NCBI databases.

INTRODUCTION

'Sialic acid' is a generic term covering a family of over 50 related sugar acids that are ubiquitous among vertebrates, where they occur as terminal sugars of cell-surface-exposed glycoconjugates in several tissues [1–4]. The most commonly studied sialic acid, *N*-acetyl-neuraminic acid (Neu5Ac; Fig. 1), is the only form synthesized by humans, whereas other animals can also make C5-group variants such as *N*-glycolyl-neuraminic acid (Neu5Gc) and 3-keto-3-deoxy-D-glycero-D-galactonononic acid (KDN) [2, 3, 5]. Various *O*-substitutions at different hydroxyl moieties expand the diversity of sialic acid in nature [2]. The location of structurally diverse sialic acids at the cellular surface underpins the wide variety of biological roles this molecule plays in the host. These often rely on physical interactions between sialylated glycoproteins

Received 05 March 2021; Accepted 20 May 2021; Published 29 June 2021
Author affiliations: [1]Department of Biology, University of York, York, UK; [2]Microbes in Health and Disease, Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK; [3]Gut Microbes and Health Institute Strategic Programme, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK.
***Correspondence:** Emmanuele Severi, emmanuele.severi@newcastle.ac.uk; Gavin H. Thomas, gavin.thomas@york.ac.uk
Keywords: bacteria; Neu5Ac; sialic acid; SSS; transporter; Mutarotase.
Abbreviations: ABC, ATP-binding cassette; GPH, glycoside–pentoside–hexuronide:cation symporter; KDN, 3-keto-3-deoxy-D-glycero-D-galactonononic acid; MCP, methyl-accepting chemotaxis protein; MFS, major facilitator superfamily; NBD, nucleotide-binding domain; Neu5Ac, *N*-acetyl-neuraminic acid; Neu5Gc, *N*-glycolyl-neuraminic acid; SBP, solute-binding protein; SSS, sodium-solute symporter; TMH, transmembrane helix; TRAP, tripartite ATP-independent periplasmic.
†These authors contributed equally to this work
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Twelve supplementary figures are available with the online version of this article.

and various partners, and they include embryonic development and regulation of the immune system, where sialic acid contributes to the recognition of the 'self' [1, 3].

Being surface located also means that sialic acid comes into direct contact with micro-organisms at the host mucosal surface. While human viruses such as influenza and MERS (Middle East respiratory syndrome) coronavirus are most (in)famous for their ability to target sialylated receptors to cause disease [6, 7], bacteria can also utilize sialic acid as a mediator of interactions with the host [4, 8–10]. Both partnership-establishing bacteria (i.e. symbionts or commensals) and pathogens can release and metabolize host sialic acid, which can then be incorporated into surface structures such as capsule and lipopolysaccharide (LPS) that may help evade the host's immune system through so-called 'molecular mimicry', i.e. by hijacking the role of sialic acid in the host's recognition of the self. Furthermore, sialic acid can be used as a metabolic substrate to sustain growth and enhance the ability of bacterial species to establish themselves in target niches, in health or disease [4, 8–10].

While a small minority of sialic acid-utilizing microbes can synthesize Neu5Ac *de novo* [11], all others rely on host-derived sialic acid, whether for growth or cell surface sialylation [4, 8, 9], acquired through dedicated sialic acid transporters [8, 10, 12]. Although not ubiquitous among prokaryotes, sialic acid transport is a widespread trait across all types of sialic acid-utilizing bacteria that predominantly inhabit mucosal surfaces [4, 5, 8, 13], where it plays a critical role in virulence and host colonization [9, 14–19]. The role of sialic acid transporters in host–microbe interactions has been the target of extensive research, including from ourselves [4, 8, 10, 12, 13, 20], and today six types of bacterial sialic acid transporters have been characterized experimentally (Fig. 1) [12, 14]. These uptake systems differ by a number of features, including mode of energization, subunit composition and substrate specificity (Fig. 1) [12], indicating that sialic acid transport has evolved independently multiple times and that there is selective pressure for the acquisition of this trait in numerous, taxonomically diverse prokaryotes [12].

The first five types of transporters to have been studied all include systems specific for Neu5Ac (Fig. 1), with some able to take up Neu5Gc or KDN [9, 20–24]. The MFS (major facilitator superfamily) transporter NanT was the first sialic acid transporter reported during ground-breaking work by the Vimr group on Neu5Ac catabolism in *Escherichia coli* [25, 26], which also elucidated the canonical prokaryotic metabolism pathway for sialic acid comprising Nan and Nag enzymes [27] (Fig. 1). As an MFS protein, NanT is a secondary transporter that uses the proton gradient to drive concentrative uptake of Neu5Ac, and has been studied both *in vivo* and *in vitro* in some detail [21, 25, 28, 29]. The MFS transporter from the *Bacteroidetes Bacteroides fragilis* and *Tannerella forsythia*, also called NanT, is normally considered as being of the same family as the enterobacterial NanT group, despite notable differences between it and *E. coli* NanT [12, 30, 31]. The TRAP (tripartite ATP-independent periplasmic) transporter

Impact Statement

Sialic acid is an important molecule involved in the interplay of bacteria with their hosts. Many bacteria that have evolved to live on or near mucosal surfaces in, for example, the human gut, nasopharynx or urinary tract, have evolved ways to exploit abundant host-derived sialic acid in their own biology. The capturing of this sugar acid requires the bacteria to synthesize specific active transporters. In this study, our research reveals the strong selection for this process in bacterial evolution, reflected in the multiple times that the same process has evolved independently. Our data suggest that this has happened at least eight times to recognize sialic acid present in different niches. As well as revealing this striking multiplicity of sialic acid transporters, we also propose a robust and extendable classification system for the naming of sialic acid transporters, which is currently confusing. The article will be of interest to researchers studying evolution and microbial transporters, and to all microbiologists who discover a sialic acid transporter involved in the colonization potential of their particular bacterium of study.

SiaPQM, first characterized in *Haemophilus influenzae* [15, 32] and also well-studied in *Vibrio* spp. [28, 33, 34], is among the best-studied sialic acid transporters to date, with a wealth of *in vivo* and *in vitro* data based on the use of native and heterologous hosts, as well as reconstituted purified systems [15, 28, 35, 36]. SiaPQM is a secondary transporter too, being energized by a sodium rather than a proton gradient [29, 36], but it also depends on the solute-binding protein (SBP), SiaP, for function [15, 35]. While use of an SBP is a feature most commonly associated with primary (e.g. ATP-dependent) transporters [35], SiaP here functions together with two membrane components that bear no relationships with those of primary systems (the 4 transmembrane helices [TMH] SiaQ and the 12 TMH SiaM, fused in *H. influenzae*) [37]. Discovered the same year as SiaPQM, SatABCD is another SBP-dependent system, but from the ABC (ATP-binding cassette) transporter superfamily, i.e. a primary transporter that uses ATP binding and hydrolysis to energize uptake. The best characterized SatABCD system to date is that of *Haemophilus ducreyi* (HdSatABCD) [38], but genetic studies have been carried out in *Actinobacteria* too [39, 40]. As also reported for SiaP [35, 41–43], crystallographic and mutational studies identified key residues for Neu5Ac-binding and selectivity by the SBP component SatA [23]. Not to be confused with the *H. ducreyi* transporter is a different ABC system discovered in *Streptococcus pneumoniae* [17], normally referred to as SatABC too, but also as SAT3 [13]. To date, information on this transporter derives solely from genetic and transcriptional studies [17, 44–46] and, despite the distinction from HdSatABCD dating back to 2009 [13], this system remains to be functionally characterized. The

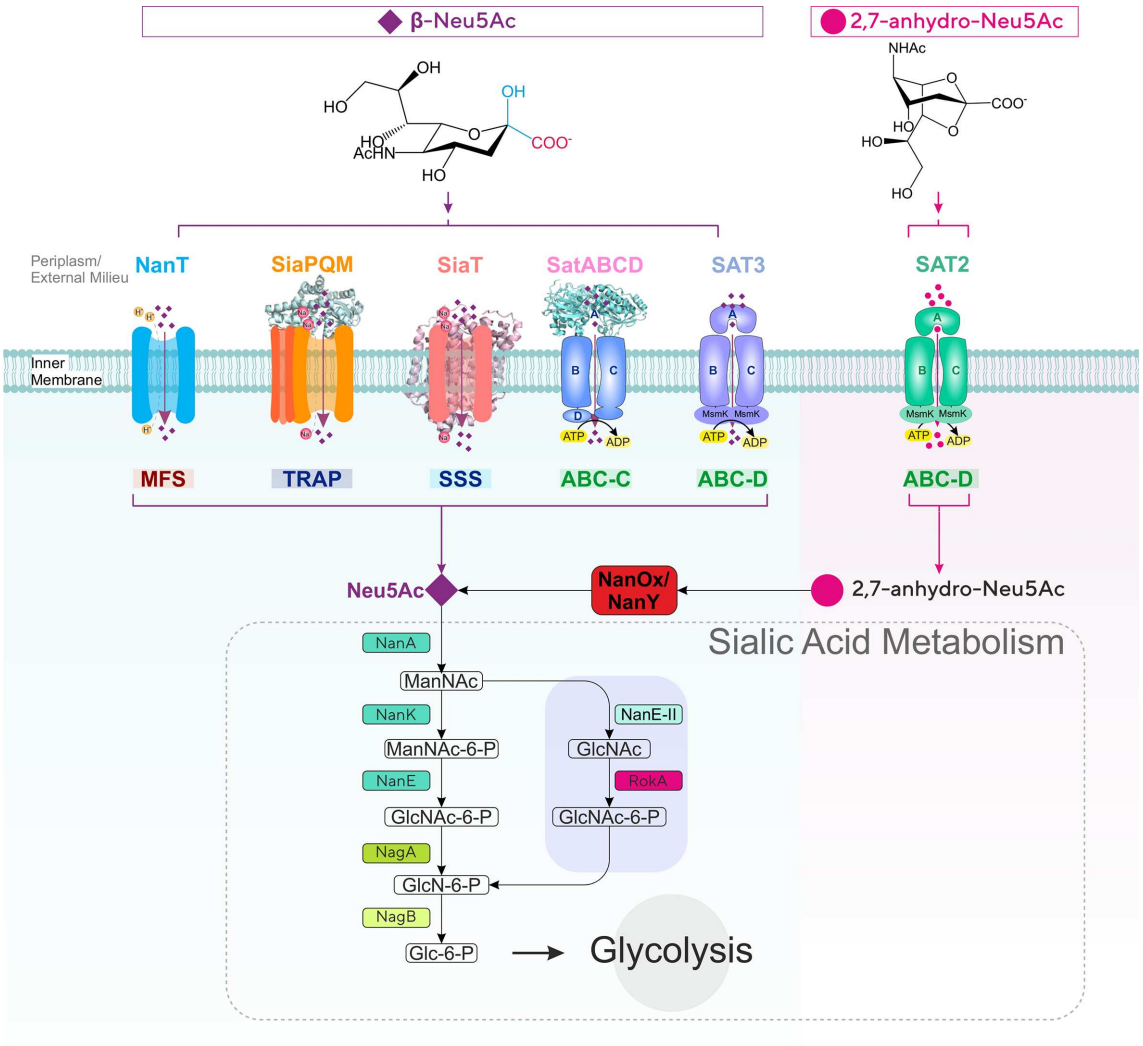

Fig. 1. Diversity of sialic acid transport and catabolism in bacteria. Sialic acid-utilizing bacteria have evolved multiple types of transporters functioning at the inner (cytoplasmic) membrane from four major (super)families (MFS, TRAP, ABC and SSS) differing by mode of energization and subunit composition. Once inside the cell, sialic acid is metabolized to GlcNAc-6-P via one of two characterized pathways to then enter central metabolism by the action of the GlcNAc-catabolic enzymes NagA (GlcNAc-6-P deacetylase) and NagB (GlcN-6-P deaminase). In the *E. coli* paradigm [27], Neu5Ac is converted by the sequential action of three dedicated enzymes, namely NanA (Neu5Ac lyase), NanK (ManNAc kinase) and NanE (ManNAc-6-P epimerase); in the *Bacteroidetes* paradigm [30, 31], NanA and the alternative enzyme NanE-II (GlcNAc epimerase) feed substrate to the glucokinase RokA. Bacteria such as *Ruminococcus gnavus* that take up 2,7-anhydro-Neu5Ac use the cytoplasmic oxidoreductase NanOx (NanY in *E. coli*) to convert the substrate to Neu5Ac, which then enters a canonical Nan pathway. ManNAc, *N*-acetyl-mannosamine; GlcNAc, *N*-acetyl-glucosamine; GlcN, glucosamine; Glc, glucose; MsmK, multitasking ATPase protein serving multiple sugar ABC transporters including sialic acid transporters [44].

fifth group of Neu5Ac-specific transporters include proteins of the SSS (sodium-solute symporter) family of secondary transporters, which are generally referred to as 'SiaT' [12, 22, 47]. First discovered about 10 years ago as a diverse and widespread group of transporters [13, 29], several SiaT transporters from Gram-negative and Gram-positive bacteria have now been functionally characterized *in vivo* and *in vitro* [4, 22, 47, 48], and a high-resolution crystal structure is available for the *Proteus mirabilis* orthologue with bound Neu5Ac [47]. Complementation, mutagenesis and biochemical studies using reconstituted systems confirmed the dependence of these proteins on sodium for function, identified residues involved in substrate-binding and transport of coupled Na^+ ions, and provided insight into their substrate specificity [21, 22, 29, 47].

The sixth and final group of characterized sialic acid transporters has introduced an entirely novel substrate specificity to the field. Early gene cluster analyses had predicted that ABC uptake systems of the 'SAT2' type were sialic acid transporters distinct from the above HdSatABCD ('SAT') and SAT3 ABC systems [13], but SAT2 transporters remained

uncharacterized until recent work [14] studying strains of the gut symbiont *Ruminococcus gnavus*, which have the capacity to produce a unique anhydro form of sialic acid (2,7-anhydro-Neu5Ac; Fig. 1) from Neu5Ac-terminated glycoconjugates via the action of an intramolecular *trans*-sialidase (IT-sialidase) [49, 50]. Working with mutant strains as well as with the purified SBP component of SAT2, Bell *et al.* [14] demonstrated that the *R. gnavus* SBP was specific for 2,7-anhydro-Neu5Ac (Fig. 1). The study established that the IT-sialidase, the SAT2 transporter and the newly discovered oxidoreductase, NanOx, which converts 2,7-anhydro-Neu5Ac back to Neu5Ac before this can be metabolized further (Fig. 1) [14, 51], cooperate to channel an 'exclusive' form of sialic acid into an otherwise canonical Nan catabolic pathway [14] (Fig. 1). This is a compelling instance where prokaryotes have innovated on sialic acid transport and used it to their advantage in their target niche [14, 49]. To date, this is the sole example for this group of ABC sialic acid transporters [14]. However, the characterization of the orthologous oxidoreductases NanOx from *R. gnavus* and NanY (formerly YjhC) from *E. coli* [51, 52] has provided evidence for at least three other potential anhydro-sialic acid transporters. Functional complementation of *E. coli* mutants has demonstrated a role in anhydro-sialic acid acquisition for one of these, namely the NanT-like *E. coli* MFS transporter, NanX (formerly YjhB) [51, 52].

The diversity of bacterial sialic acid transporters has raised various questions regarding structural–functional features, mechanism of transport and exact physiological roles [10, 12]. However, sialic acid transport has seldom been approached from a phylogenetic/evolutionary perspective, in stark contrast with the rich ensemble of phylogenetic studies on the Nan catabolic enzymes, namely NanA, NanE and NanK (Fig. 1) [9, 13, 30, 53]. As these studies have revealed the existence of diverse clades of NanA, NanE and NanK orthologues often combining into mosaic clusters at gene level, equivalent analyses are largely missing for the accompanying transporters, with the sole exception of studies on SiaP [33]. There is a need to update the broad classification of sialic acid transporters put forward by the Boyd group in 2009 [13] taking into account the distribution of individual uptake systems across bacteria [15, 20, 29, 47].

Here, we carried out gene cluster and phylogenetic analyses of all known sialic acid transporters across the bacteria. Using phylogeny as the basis for classification, we first demonstrated that all described sialic acid transporters can be classified into eight distinct families, with validation of the six historically established types (Fig. 1), and with the identification of two new families of phylogenetically distinct MFS transporters. Within the SSS and MFS families, we discovered a novel form of sialic acid transporter, which consists of a fusion with a sialometabolic enzyme, and is widespread among the *Bacteroidetes* and *Planctomycetes*/*Verrucomicrobia*. Overall, the study provides significant new insights into the evolution and function of sialic acid transporters, as well as revealing potential novel aspects for sialic acid transporter components.

METHODS

Sialic acid transporter sequences

The criteria used to collate sialic acid transporter sequences were based on experimental evidence and *in silico* analysis of the sialic acid *nan* operon. Using functionally characterized sialic acid transporters as initial queries, we searched for homologous proteins based on sequence similarity using BLAST [12, 14], and all hits were validated by reciprocal BLAST searches. For SBP-dependent transporters (Fig. 1), we used the SBP component as a search query, as done in our previous work [33]. As we added new clades to the phylogenetic trees, we took a heuristic approach and included the new members in both direct and reciprocal searches, and re-validated the hits found in previous rounds. At the onset of these searches, homologous yet non-orthologous hits were included in order to resolve the phylogeny. Assignment of orthologous transporters was aided with the mapping of Neu5Ac-binding residues when these were known from crystallographic and/or mutational studies (Table 1). All organisms where we found candidate orthologues of sialic acid transporters were confirmed for the presence of complete sialocatabolic pathways (encoded within clusters or separate loci), recapitulating the methodology and the results described in references from the Boyd group [13, 34]. As queries, we used sequences of different organismal origin for the following enzymes: NanA (Neu5Ac lyase), NanK (ManNAc kinase), NanE (ManNAc-6P epimerase) and NanE-II (ManNAc epimerase) (Figs S1–S12, available with the online version of this article). Clusters were then annotated outwards to identify further distinctive sialometabolic genes in the clusters such as, for example, sialidase genes, NanM (Neu5Ac mutarotase), NagA (GlcNAc-6P deacetylase) and NagB (GlcN deaminase). All sialic acid transporter-bearing organisms were searched for further orthologues of the same or different transporter families. Number of TMHs and presence of signal peptides were predicted using TMHMM [54] and SignalP5 [55], respectively. 3D structure prediction for NanM domains was performed with I-TASSER [56] and Swiss-model [57]. Wherever possible, we opted for fully over partially sequenced genomes in order to minimize incomplete genetic information. All sequences were retrieved from the National Center for Biotechnology Information database.

Phylogenetic analysis

Each archetypal sialic acid transporter was used as a search query in the Pfam database. To reduce the sequence dataset within a Pfam family, we downloaded the representative proteome RP15 for each Pfam identifier. Unique accession numbers were retrieved from this list and used to download full length proteins from the UniProt database. Bacterial sequences were filtered from each Pfam family and compiled with our heuristically predicted sialic acid transporters to generate multiple sequence alignments. For large alignments (>500 sequences), sequences were aligned with MUSCLE using default settings. Alignments were manually inspected using Jalview [58]. For <200 sequences, alignments were generated using MAFFT with L-INS-I for an accurate amino acid

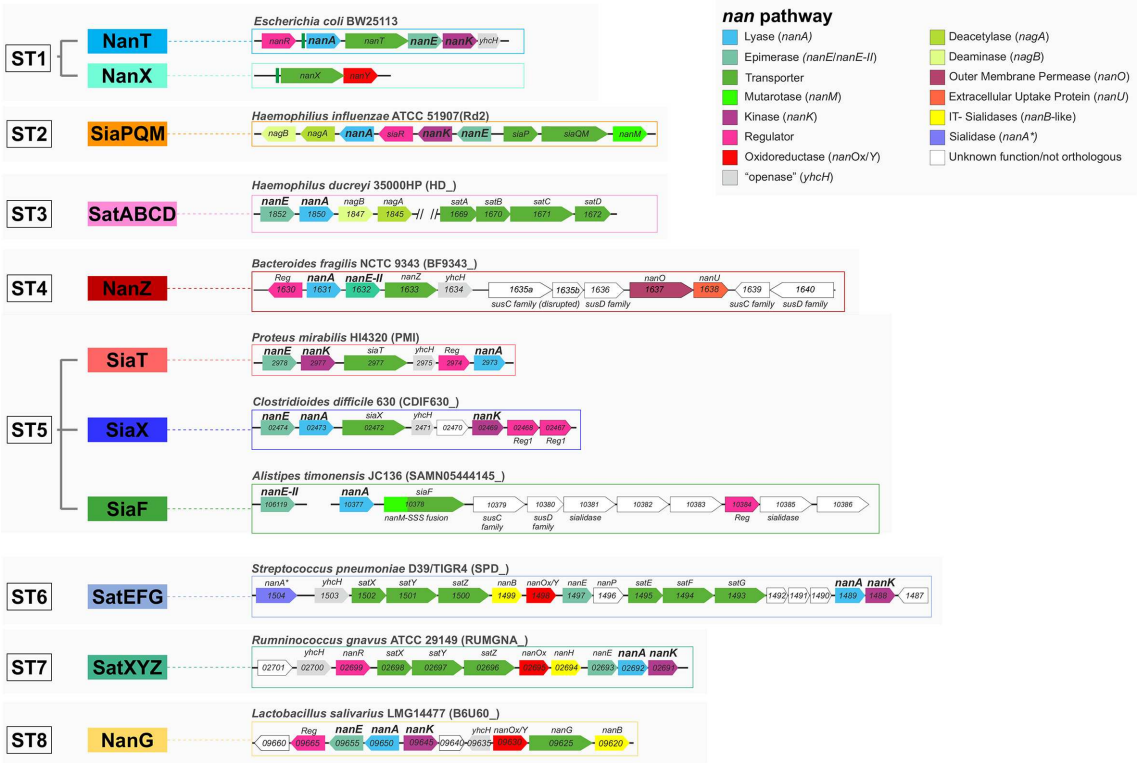

Fig. 2. Structure of *nan* clusters for archetypal transporter families. *nan* clusters are shown for each of the eight newly classified sialic acid transporter families using archetypical organisms as reference. Locus prefixes are denoted in brackets next to the reference organism, gene names are displayed within gene tags. *nanOp* operators highlighted upstream of *nanA* and *nanX* in the *E. coli* ST1 loci emphasize the occurrence of a single NanR regulon in *E. coli* [51]. In TIGR4, the ST6 locus bears minor differences [51]. Note that, as for Post *et al.* [38] before us, we could not find orthologues of *nanK* in *H. ducreyi* 35000HP; thus, ManNAc kinase functions in this organism remain unidentified. YhcH, accessory cytoplasmic Neu5Ac anomerase/'openase' [91]. SusCD, outer membrane protein complex for glycan acquisition made of a TonB-dependent transporter (SusC) and an extracytoplasmic lipoprotein (SusD). NanOU, an experimentally confirmed sialic acid-specific SusCD-family complex [88].

alignment. Phylogenetic reconstruction was performed using IQ-TREE [59]. Phylogeny was inferred by maximum likelihood with automatic model selection to find the best-fit model. Ultrafast bootstrap approximation was used give branch support. Phylogeny was visualized with iTOL [60]. Alignment of selected proteins with crystal structures was made with ESPript [61]. All figures were prepared using CorelDRAW 2020 and PowerPoint.

RESULTS

Phylogenetic analyses identify eight different evolutionary origins for bacterial sialic acid transporters

In light of our recent discovery of two new 2,7-anhydro-sialic acid transporters [14, 51], we undertook the first systematic bioinformatic analysis of sialic acid transporters across bacteria. The approach used, described elsewhere [13], is based on the genetic association of a transporter gene with a complete *nan* catabolic pathway (see Methods) as the primary factor for inclusion. Examples of characterized bacterial *nan* clusters are shown in Fig. 2, with more detailed results of the analysis presented in Figs S2–S12. This bioinformatics approach revealed expanded groups for all known types of sialic acid transporters, as well as additional novel groups linked to known *nan* genes. Within both the MFS superfamily and the large SSS family of transporters, there were potentially multiple sub-families of sialic acid transporters, and their mono- or polyphyletic origin was investigated using a phylogenetic approach (Fig. 3). For the MFS proteins in Pfam family MFS_1, two different evolutionary origins of sialic acid transporters were clearly identified (Fig. 3a), and a third origin was discovered uniquely in the MFS_2 family, which encompassed the GPH (glycoside–pentoside–hexuronide:cation symporter) family [62, 63] (Fig. 3b). This contrasts with the SSS transporters, where three potentially different families appear to form a clear monophyletic group, suggesting that sialic acid specificity emerged once and then these proteins have diversified into related clades (Fig. 3c). For the SBP-dependent transporters, the phylogenetic analysis targeted the SBP component, which defines the substrate specificity, and based on this principle we found a single origin for the TRAP transporters, as previously suggested [33] (Fig. 3d). As for the ABC transporters, which fall into two

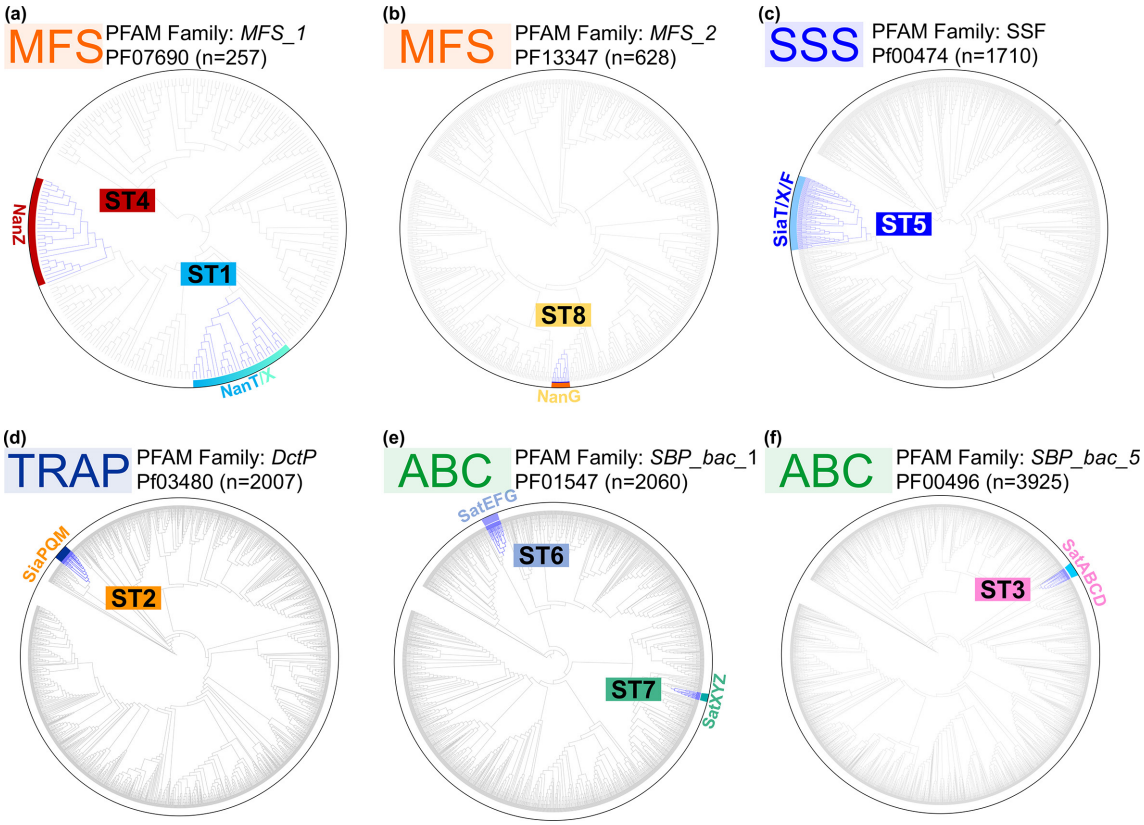

Fig. 3. Phylogenetic classification of sialic acid transporters in bacteria. Global phylogenetic analysis of sialic acid transporters within expanded Pfam families. Clades highlighted in blue are sialic acid transporter sequences that reside in verified *nan* clusters. Among the diverse MFS_1 and ABC (*SBP_bac_1* – cluster D SBPs) families, we observe independent evolutionary origins. In contrast, in the SSS, TRAP, MFS_2 and ABC (*SBP_bac_5* – cluster C SBPs) families appear each to have only single ST origins.

Pfam families, there is evidence of two independent origins for Pfam SBP_bac_1 (cluster D SBPs) (Fig. 3e), while a single origin was identified for the SBP_bac_5 (cluster C SBPs) (Fig. 3f). From this analysis, we can conclude that sialic acid transport specificity has evolved at least eight times during the evolution of bacteria, and to aid in the classification of these diverse transporters we have named these using a sialic acid transporter (ST) family nomenclature, from ST1–ST8 (Table 1).

ST5 (SSS) sialic acid transporters cluster into distinct clades and include candidate anhydro-sialic acid transporters

Although the SSS transporters identified in our *in silico* analyses appear to have a single evolutionary origin (ST5) (Fig. 3c), we noticed that they are the most widespread in bacteria, and we could see distinct clades within them (Fig. 4). Previously characterized ST5 sialic acid transporters [16, 22, 29, 47] map to two of three major clades. The SiaT clade contains the SiaT proteins from enterobacteria and *Staphylococcus aureus*, while the second clade contains the SSS transporter from *Clostridioides difficile*, which having previously been called NanT (Table 1) we propose instead to name SiaX to both stress its ST5 nature and emphasize the distinction among ST5 clades (Fig. 4). The SiaT group has the broadest phylogenetic distribution, while the SiaX group is limited to *Firmicutes* (Fig. 4). A third major clade, which we refer to as 'SiaF' (Fig. 4), contains exclusively newly predicted sialic acid transporters (see later in the text). Of note, within the SiaT clade is a novel sub-clade occurring in some *Brachyspira* spp. (see next section) and a smaller branch featuring transporters encoded within *nan* clusters of *Mycoplasma* spp. [13, 64].

Notable with the SiaX clade are a number of transporters that are genetically linked to the 2,7-anhydro-Neu5Ac oxidoreductase NanOx/NanY (Fig. 4) and in some cases an IT-sialidase, one such transporter being the SP1328 protein from *S. pneumoniae* TIGR4 [51] (Fig. S7). While this suggests that they may function in the uptake of 2,7-anhydro-sialic acid rather than Neu5Ac [51, 52], they occur within a wider range of closely related Neu5Ac transporters, including the characterized *C. difficile* protein (Fig. 4). We noted that the IT-sialidases associated with these NanOx-linked SiaX proteins are orthologues of *S. pneumoniae* NanC (Fig. S7), which has as its primary product Neu5Ac2en (*N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid) [65–67], this being another oxidized form of Neu5Ac found in nature. As the

Table 1. Classification and characteristics of bacterial sialic acid transporters

Family ID*	Pfam†	Updated name	Components‡	Alternative name§	Experimental confirmation‖	Energization¶	Specificity#	Reference system**	UniProt†	PDB†	TCDB†
ST1	MFS_1 PF07690	NanT	Single TMD (14 TMH)	–	1985 [25]	H+ [28]	Neu5Ac	*Escherichia coli*	P41036	–	2.A.1.12.1
	MFS_1 PF07690	NanX	Single TMD (12 TMH)	YjhB, ORF425 [10, 26, 51, 52, 92]	2020 [51, 52]	H+	2,7-Anhydro-Neu5Ac Neu5Ac2en	*Escherichia coli*	P39352	–	–
ST2	DctP PF03480	SiaPQM	SiaP (SBP), SiaQM (TMD1+TMD2) or SiaP (SBP), SiaQ (TMD1), SiaM (TMD2)	SiaPT, NanPU, NeuT [19, 42, 86, 93–97]	2005 [15, 93]	Na+ [28, 36]	Neu5Ac	*Haemophilus influenzae, Vibrio cholerae*	P44542, Q9KR64	2CEX††, 4MAG††	2.A.56.1.3, 2.A.56.1.6
ST3	SBP_bac_5 PF00496	SatABCD	SatA (SBP), SatB (TMD1), SatC (TMD2+NBD2), SatD (NBD1)	SAT, SiaEFGI, NanBCDF, NanABC2 [13, 76, 80, 86]	2005 [38]	ATP	Neu5Ac	*Haemophilus ducreyi*	Q7VL18	5ZA4, 5Z99, 5YYB	–
ST4	MFS_1 PF07690	NanZ	Single TMD (12 TMH)	NanT [30, 31]	2009 [30]	H+	Neu5Ac	*Bacteroides fragilis, Tannerella forsythia*	Q5LEN6, A0A1D3USD0	–	–
ST5	SSF PF00474	SiaT	Single TMD (13 TMH)	STM1128, NanT, NanV, NanX, NanP [4, 29, 48, 73, 86, 98, 99]	2010 [29]	Na+ [22, 29, 47]	Neu5Ac	*Proteus mirabilis, Salmonella typhimurium, Staphylococcus aureus*	B4EZY7, Q8ZQ35, Q2G161§§	5NV9, 5NVA	2.A.21.3.10‡‡
	SSF PF00474	SiaX	Single TMD (13 TMH)	NanT [16, 73]	2013 [16]	Na+	Neu5Ac 2,7-anhydro-Neu5Ac(?) Neu5Ac2en(?)	*Clostridiodes difficile, Streptococcus pneumoniae*	Q185B4, A0A0H2UQE5	–	–
	SSF PF00474	SiaF	CPM+TMD (13 TMH)	–	NC	Na+	Neu5Ac(?)	*Alistipes timonensis*	A0A1H4AP10	–	–
ST6	SBP_bac_1 PF01547	SatEFG	SatE (SBP), SatF (TMD1), SatG (TMD2), SatH (NBD)‖‖	SatABC, NanUVW, SAT3, NanT, NanABC [13, 17, 44–46, 81, 86]	2011 [17]	ATP	Neu5Ac	*Streptococcus pneumoniae*	A0A0H2ZL68	–	–
ST7	SBP_bac_1 PF01547	SatXYZ	SatX (SBP), SatY (TMD1), SatZ (TMD2), SatW (NBD)‖‖	SAT2 [14]	2019 [14]	ATP	2,7-Anhydro-Neu5Ac	*Ruminococcus gnavus*	A7B561	–	–
ST8	MFS_2 PF13347	NanG	Single TMD (11 TMH)	GPH [51]	NC	H+ or Na+	2,7-Anhydro-Neu5Ac(?)	*Lactobacillus salivarius*	A0A1V9QLX9	–	–

*ST families are as defined in Figs 2 and 3. ST families are ordered by date of experimental confirmation (see footnote ‖). ST1 and ST5 have been further subdivided to reflect the functional and phylogenetic differences among clades.
†In the case of SBP-dependent transporters Pfam, UniProt, PDB and TCDB identifiers all refer to the those of the SBP component, consistent with the methodology used for the phylogenetic analyses (see Methods).
‡SBP, solute-binding protein; TMH, transmembrane helix; TMD, transmembrane domain; NBD, nucleotide binding domain; CPM, cyclically permutated mutarotase (NanM-domain). For single component transporters, we indicate the number of TMHs to emphasize structural differences.
§We list here all the names used for ST uptake systems in the literature. These include locus tags used before functional confirmation, identical names used for different transporters, different names used for the same transporter, and group identifiers.
‖For each ST family, we include here only the date when function in sialic acid uptake was first demonstrated. Additional dates are used in the case of transporters of distinct clades (see footnote *). NC, Not confirmed.
¶Mode of energization is predicted from the Pfam of each ST family. References are included for those cases where the identity of the coupling ion used by secondary transporters has been demonstrated.
#Some Neu5Ac-specific transporters can also transport Neu5Gc and/or KDN – these details are not included here for the sake of simplicity. '(?)' means that the substrate is not confirmed and is primarily predicted based on gene clusters analysis.
**Organism(s) of origin for the first-discovered and/or best-characterized transporters of the family. Please note that, while not indicated here, some transporters may be limited to specific strains of a species. In the case of the uncharacterized transporters in this table, we use *Alistipes timonensis* for SiaF (ST5), as this organism's growth is stimulated by Neu5Ac [73], while we use *Lactobacillus salivarius* for NanG (ST8), as this is the only example to date where the *nanG* gene maps to a complete *nan* cluster (Fig. S12).
††Several structures of SiaP have been solved including orthologues from different organisms, complexes with different substrates, and also mutant proteins. The complete list includes: 2CEY, 2CEX, 3B50, 2V4C, 2WX9, 2WYP, 2WYK, 2XA5, 2XWV, 2XWO, 2XXK, 2XWI, 2XWK, 4MAG, 4MMP, 4MNP, 5LTC, 6H76, 6H75.
‡‡TCDB [100] lists, under other names, two further SiaT proteins: 2.A.21.3.7 (reported [29] but uncharacterized, from *Aliivibrio fischeri*) and 2.A.21.3.20 (genetically characterized [48], from *Lactobacillus sakei*). Both feature in our phylogenetic analysis (Fig. 4).
§§The best-characterized SiaT orthologue from *S. aureus* comes from strain RF122 [22] (locus tag SAB0251c), which does not possess a UniProt entry. We here replace it with the entry for strain NCTC 8325, which differs by a single residue.
‖‖ST6 and ST7 transporters generally rely on conserved, multitasking NBD proteins such as MsmK for function [44]. A minority of clusters do include a gene coding for NBD linked to ST6 and ST7 genes (see Figs S10 and S11), and we here propose names for these additional components.

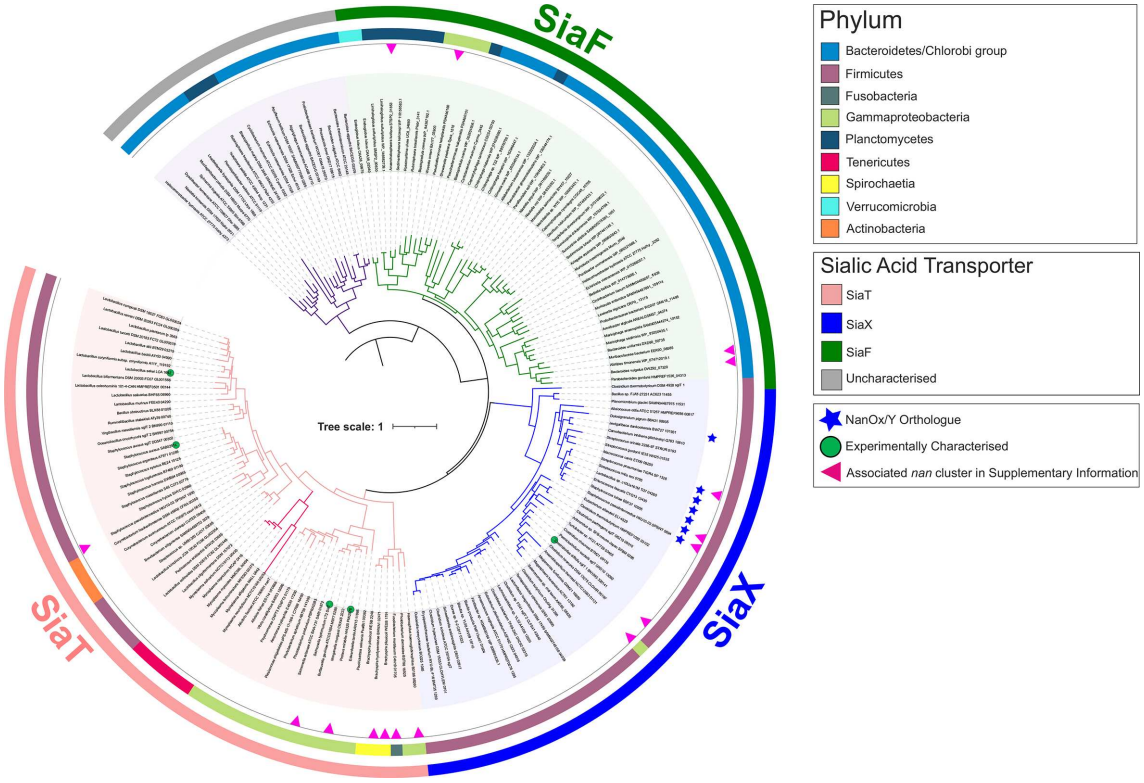

Fig. 4. Phylogenetic distribution of ST5 (SSS) sialic acid transporters in bacteria. Phylogeny of ST5 (SSS) sialic acid transporters at the phylum level. Coloured branches represent three major clades, the historical SiaT clade (pink), the SiaX clade (blue) including both Neu5Ac and putative anhydro-Neu5Ac transporters (asterisk), and the clade of SiaF proteins (green) representing a novel fusion between sialic acid transporter and mutarotase. A fourth group of uncharacterized transporters (grey), including one from *Bacteroides thetaiotaomicron*, is addressed in Discussion. Experimentally characterized transporters are highlighted on the tree with a green circle; the red arrowhead indicates the examples shown in Fig. S7 of associated *nan* sialocatabolic genes (occurring in clusters or at separate loci). SiaT is distributed widely across several bacterial phyla, whereas SiaX is restricted to the *Firmicutes* with isolated exceptions, and SiaF occurs near-exclusively across *Bacteroides* and *Planctomycetes/Verrucomicrobia*. The maximum-likelihood tree was inferred from SSS transporter proteins (n=354) residing within a *nan* cluster containing at least one sialocatabolic *nan* gene. The scale bar represents the number of substitutions per amino acid position.

IT-sialidase acts upstream of the transporter, and the NanOx/NanY proteins can also efficiently convert Neu5Ac2en to Neu5Ac [52], we hypothesize that NanOx-linked SiaX transporters might be able to take up this substrate too. Unlike the characterized anhydro-Neu5Ac transporters of the ST7 and ST1 types, which belong to easily distinguishable phyletic groupings (Figs 3 and S1), and take up only oxidized sialic acid and not Neu5Ac [14, 51, 52], SiaX transporters are all closely related with each other within a single clade regardless of their predicted specificity, and this raises questions as to what might determine the substrate specificity of individual SiaX proteins [22, 47]. Strikingly, all the residues forming the known sialic acid-binding site in the structure of *P. mirabilis* SiaT (Fig. S8) are conserved across the SiaT/X/F proteins, which does corroborate their single origin, but does not help gain insights into their exact substrate specificities. In the absence of detailed structural–functional information about SiaX transporters, their specificity towards Neu5Ac/anhydro-Neu5Ac remains unknown, but it is possible that some of them can transport multiple forms of sialic acid, as reported for other types of sialic transporters. For instance, the ST2-SBP SiaP can bind both Neu5Ac and Neu5Ac2en (though the latter with considerably lower affinity) [15, 35], and the ST1 anhydro-sialic acid transporter NanX has been reported to take up Neu5Ac2en besides 2,7-anhydro-Neu5Ac [52]. These cases provide precedents of sorts for versatile binding sites that recognize different forms of sialic acid.

Identification of ST5 sialic acid transporters in pathogenic *Spirochaetes*

Using our transporter-led approach has also uncovered more fundamental insight into the sialic acid biology of the pathogenic *Spirochaetes*, organisms where sialic acid transport genes have been previously identified only once (namely, ST2/*siaPQM* genes in *Treponema brennaborense* [33]; Fig. S4). The SiaT clade of ST5 transporters also includes a small subclade of highly similar transporters that occur in some species of the *Spirochaete* genus *Brachyspira* (Fig. 4). *Brachyspira hyodysenteriae*, *Brachyspira pilosicoli* B2904 and *B. pilosicoli* WesB

are pathogens of pigs, birds and humans, respectively, and have recently been reported to catabolize Neu5Ac/Neu5Gc and adhere to sialic acid-rich mucin glycoproteins [68, 69]. By looking for homologues of SiaT transporters in these bacteria, we identified a full complement of sialocatabolic genes in two *B. pilosicoli* strains (Fig. S7), and complete sets of *nan* genes scattered over different loci in the other species including *B. hyodysenteriae* (Figs S7 and S11), which is consistent with these species' reported use of sialic acid as a carbon source . The *Brachyspira* spp. *B. pilosicoli* WesB and *Brachyspira hampsonii* NSH-16 [70] had satellite loci carrying extra sialic acid transporter genes, specifically a second similar *siaT* gene in WesB (Figs 4 and S7) and genes for the ST7 2,7-anhydro-sialic acid transporter, SatXYZ (formerly SAT2; Table 1), linked to *nanOx/nanY*, in *B. hampsonii* (Fig. S11).

SiaF clade of ST5 includes SSS transporters fused to sialic acid mutarotases

The third clade of ST5 transporters identified in this analysis, SiaF, is present in a diverse group of bacteria belonging to the phyla *Bacteroidetes*, *Planctomycetes* and *Verrucomicrobia* (Fig. 4). This phylogenetically heterogenous clade is distinguished by a common, hitherto unprecedented feature in that they all contain N-terminal fusions to the NanM Neu5Ac mutarotase protein domain [71] (hence, F for 'fusion'; Fig. S7). The NanM domain carries a predicted leader peptide for its translocation across the membrane, consistent with the periplasmic localization of unfused NanM in *E. coli* [71] and the N_{out}-C_{in} topology of the 13 TMH SSS domain [47] (Table 1), but its six-bladed β-propeller core appears to lack the helical dimerization hairpin present in the homodimeric *E. coli* enzyme (see Methods), suggesting that within the fusion the NanM domain is monomeric, just like the attached SSS moiety [22, 47]. The architecture of these fusion proteins is consistent with the proposed role of NanM to act upstream of sialic acid uptake to provide a faster supply of anomerically correct substrate (β-Neu5Ac) to the transporter [71]. Many SiaF transporters are encoded in *nan* clusters (Figs 2 and S7) with some being the only identified sialic acid transporters, while others co-occur with orthologues of the ST4 MFS transporter (e.g. in *Phocaeicola vulgatus* – ex *Bacteroides vulgatus*; Fig. S7). Very few exceptions are found outside the above phyla, and all are limited to *Gammaproteobacteria* species such as *Pseudoalteromonas haloplanktis* (Fig. S7), where these transporters had been identified previously, but not as fusion proteins [13].

The experimental evidence for the function of these SiaF transporters is very scarce. To our knowledge, there are only two cases where sialic acid metabolism has been investigated (to a degree) in bacteria that bear *siaF* genes, namely the *Bacteroidetes Alistipes timonensis* JC36 and *Capnocytophaga canimorsus* [72, 73]. While the best-understood aspect of sialometabolism in *C. canimorsus* is the involvement of the sialidase SiaC in growth [72], in the case of *A. timonensis* we have evidence of sialic acid acquisition in the form of the stimulating effect that exogenous Neu5Ac has on growth in liquid culture [73]. As SiaF is the only predicted sialic acid transporter in this latter bacterium (Fig. S7), this provides good preliminary evidence for SiaF proteins to function in sialic acid uptake, but more detailed investigation is required to clarify the function of these proteins and to elucidate the role of the NanM-ST5 fusion.

ST4 (NanZ) MFS transporters are distinct from ST1 (NanT/X) proteins and occur in *Bacteroidetes* and other diverse gut commensals

A well-studied gut bacterium that is known to rapidly consume host-derived sialic acid is *Bacteroides fragilis* [29, 30, 64]. As mentioned in Introduction, the transporter from this bacterium is a known MFS transporter and was perhaps understandably called NanT, as the canonical *E. coli* NanT (ST1) is also an MFS transporter. Our analysis, however, places the *B. fragilis* MFS transporter in an evolutionary distinct family, which we have named ST4 with the transporter renamed NanZ (Fig. 3a, Table 1). ST4 proteins are also found in the phyla *Planctomycetes* and *Verrucomicrobia*, forming a distinct clade (Figs S1–S3), and are not seen in any other bacteria. In *B. fragilis*, we identified three *nanZ* genes (*BF1633*, *BF3607* and *BF3947*), encoding proteins of approximately 80% identity, yet only the *BF1633* gene product was previously described as a sialic acid transporter (Table 1), seemingly accounting for all sialic acid acquisition in this bacterium under the conditions tested [30]. Unlike the close orthologue from the oral pathogen *Tannerella forsythia* [31], neither *B. fragilis* NanZ (BF1633) nor its two BF3607 and BF3947 paralogues have been studied in a heterologous host, so it is not known whether these transporters are functionally different or whether the redundancy reflects use in different environmental conditions. It is notable that other bacteria, such as the sialic acid-utilizing commensal *Muribaculum intestinale* [73], also have multiple *nanZ* genes (Fig. S3).

We also found two instances of NanZ sequences in the important gut bacterium *Akkermansia muciniphila* and the closely related species, *A. glycaniphila*, where the ST4 transporter is again fused to a NanM-like domain (Fig. S3), as we have just described for the SiaF proteins of ST5. Similarly, the N-terminal mutarotase domain carries a leader peptide for translocation to the extracytoplasmic space, while the transporter moiety, which in NanZ proteins is normally made of a predicted 12 TMH N_{in}-C_{in} core (Table 1), here possesses an extra N-terminal TMH to adjust the topology accordingly. In both *Akkermansia* species, the corresponding genes are part of a separate locus outside the *nan* cluster (Fig. S3). Whereas some studies reported that *A. muciniphila* can release but not consume sialic acid [74], others reported growth stimulation by Neu5Ac for the same strain in a complex medium [73]. Therefore, it is possible that this fusion protein may act as a Neu5Ac transporter in *Akkermansia* species, under specific growth conditions, although experimental confirmation of its function is warranted. To our knowledge, this is the first identification of a candidate sialic acid transporter for these species, and this also completes the mapping of sialocatabolic genes among *Planctomycetes* and *Verrucomicrobia* [5, 9, 13].

With regard to the ST1 family, which contains NanT and NanX proteins, we observed overlapping yet distinctive distributions for the two clades, with NanT orthologues primarily found in enterobacteria and *Actinobacteria*, and NanX orthologues occurring in different orders of *Gammaproteobacteria* (Figs S1 and S2). To our knowledge, this is the first report of ST1 sialic acid transporters occurring outside the *Gammaproteobacteria*. Notably, all NanX transporters are genetically linked to the NanOx/NanY oxidoreductase in these novel genotypes (Fig. S2), indicating that they might function as 2,7-anhydro-sialic acid transporters in these organisms too. NanT and NanX can also be distinguished as they contain a different number of TMHs, 14 in NanT [26] and 12 in NanX (Table 1), which as proposed elsewhere [4] might account for the different substrate specificity of the transporters.

Expanded distribution of ST2 (SiaPQM) TRAP transporters in bacterial pathogens

So far, we have discussed sialic acid transporters from the ST1, ST4 and ST5 families, accounting for MFS and SSS transporters; from these, the four remaining ST types that feature characterized transporters (ST2, ST3, ST6 and ST7) are distinct in that they use an SBP in their mechanism, which usually correlates with higher affinity transport than that conferred by a classical symporter. For the ST2 proteins, which are TRAP transporters, we expanded their distribution from exclusively Gram-negative organisms (primarily *Gammaproteobacteria* and *Fusobacteria*, but also the spirochaete *T. brennaborense* [33]) to include *Firmicutes*, which are reported here for the first time (Fig. S4). Recent evidence supports an important function of ST2 proteins in bacterial vaginosis, as the pathogen *Fusobacterium nucleatum* is able to use sialic acid liberated by other members of the vaginal microbiota to improve its colonization of this niche [19]. The transporter used by *F. nucleatum* was originally identified as a SiaPQM system in early studies on ST2 sialic acid transporters [15], and confirmation of its role in sialic acid acquisition and growth was obtained by deleting the gene for the fused membrane component, *siaQM* (called *siaT* by Agarwal *et al.* [19]) (Table 1). Interestingly, we found ST2 transporters in *nan* clusters in two other species of *Fusobacterium*, namely, *Fusobacterium ulcerans* and *Fusobacterium mortiferum* (Fig. S4), and recent data suggest that *F. mortiferum* can also consume free sialic acid similar to its cousin *F. nucleatum* [19, 75].

ST3 (SatABCD) ABC transporters are widespread in the *Actinobacteria*

Our analysis of the ST3 proteins (Figs S5 and S6) revealed a significant change in our understanding of the origins of this ABC transporter. While the original SatABCD system was discovered and characterized in the Gram-negative bacterium *H. ducreyi*, with other examples pointed out in related *Pasteurellaceae* [38], it is now clear that their origin lies in the *Actinobacteria* (Figs S5 and S6). Experimental support for this assertion comes from characterization of ST3 transporters in both *Corynebacterium glutamicum* and *Bifidobacterium breve*, where the genes encoding the ST3 system have been disrupted with resulting loss of growth on Neu5Ac [39, 76]. The same genes are in the characterized *nan* cluster of *Bifidobacterium longum* subsp. *infantis* [40]. Also, the vaginal pathogen *Gardnerella vaginalis* ATCC 14019, which is a known sialidase-positive Neu5Ac consumer, contains an ST3 family transporter very similar to the characterized bifidobacterial system [9], which is thus likely the route of sialic acid uptake by this important pathogen [77]. A ST3 system is also seen in the actinobacterium *Micromonospora viridifaciens*, a non-pathogenic bacterium identified as a producer of high levels of an inducible sialidase activity [78]. We describe the full *nan* cluster from this bacterium for the first time (Fig. S6), which contains the ST3 transporter and catabolic genes as well as a gene for a likely sialic acid-responsive transcription factor. Also, this cluster contains the structural gene for the well-studied GH33 family sialidase [79]. Hence, it is highly likely that sialic acid utilization plays a role in the biology of this bacterium in the soil, as has been suggested for the related actinobacterium *C. glutamicum* [39, 80]. Our analysis now suggests that the small *Pasteurellaceae* clade of ST3 sialic acid transporters originated by horizontal gene transfer from actinobacterial clusters.

ST6 (SatEFG) ABC transporters in *Spirochaetes* encompass orphan SBP components

The distribution of ST6 ABC transporters identified orthologues primarily in *Firmicutes* (Fig. S9), including the known pathogen *S. pneumoniae* and other important species of *Streptococcus* [17, 81]. A ST6 transporter is also present in the sialic acid-utilizing commensal *Ruthenibacterium lactatiformans* [73] (Fig. S10). We renamed this transporter SatEFG to avoid confusion with ST3 transporters (Table 1); the ATPase component, which is normally provided by the multitasking *msmK* gene [44], should be called SatH when encoded by a dedicated gene in the same cluster (Fig. S10). These transporters are also seen in a small number of *Spirochaetes* including *T. brennaborense* (Fig. S10); however, we also discovered a few instances of a novel genetic linkage between the sole SBP component (SatE) of a ST6 system and a methyl-accepting chemotaxis protein (MCP) [82] (Fig. S10). These instances all occur in some *Treponema* species including the 'Red Complex' [83] pathogen, *Treponema denticola* (Fig. S10). As reported by others [83, 84], we could not find orthologues of any *nan* catabolic and/or other sialic acid transporter genes in *T. denticola*, including those for the cognate membrane components SatF and SatG (unlike in *T. brennaborense*, which bears ST2, ST6 and *nan* genes; Figs S4–S10), though there is some evidence for Neu5Ac consumption by *T. denticola* [84]. Because of the absence of SatF and SatG orthologues, and the conserved link with this uncharacterized MCP protein, we think it is unlikely that the orphan ST6-SBP/SatE components in these *Treponema* species have a direct role in transport, and we instead suggest that the physical interacting partner of SatE might be the MCP itself. Precedents for this are other well-studied SBPs that take part in sensory apparatuses responding to extracellular small molecules, where small

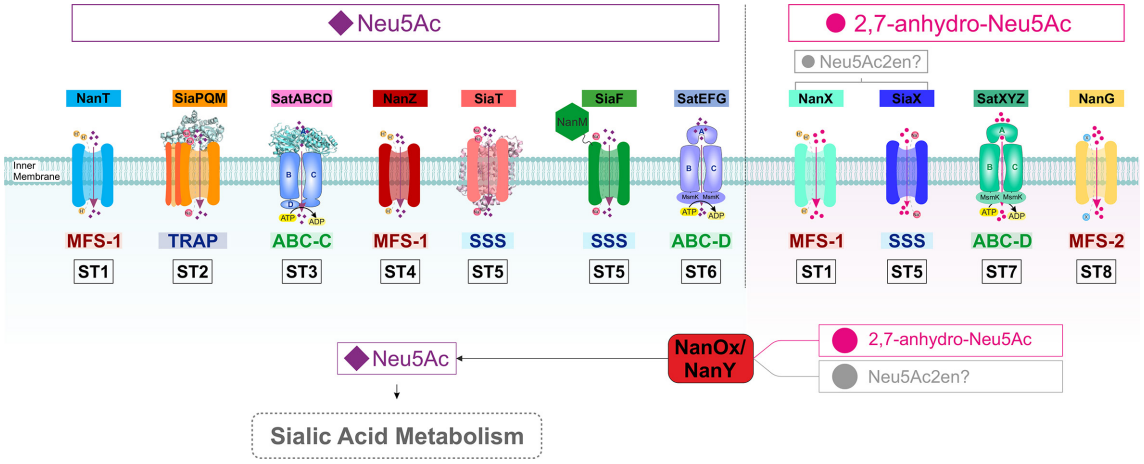

Fig. 5. Sialic acid transporters ST1–ST8. Newly classified sialic acid transporter families (Table 1) are shown. ST groups are separated by substrate specificity for Neu5Ac or 2,7-anhydro-Neu5Ac/Neu5Ac2en. Substrate specificities are discussed in the text and summarized in Table 1.

molecule sensing is always mediated by physical interaction between the SBP capturing the substrate and the MCP [85]. Based on these observations, we speculate that this particular group of SatE proteins, rather than effecting uptake, might help orchestrate some cellular response to exogenous sialic acid in their organisms of origin, but research is warranted to confirm this notion of 'sialic acid sensing' and to indeed discover what the ultimate physiological outcomes might be (see also Discussion).

ST7 (SatXYZ) and ST8 (NanG) families contain additional transporters for 2,7-anhydro-Neu5Ac

Of the final two STs, ST7 includes the recently discovered class of ABC transporters with specificity for 2,7-anhydro-Neu5Ac, exemplified by the characterized system from *Ruminococcus gnavus* [14, 51]. We propose that these are called SatXYZ (Table 1), while the ATPase component should be called SatW when encoded by a dedicated gene in the same cluster (as in, for example, *Haemophilus haemoglobinophilus*; Fig. S11). Novel examples of ST7 transporters are those found in the spirochaete *B. hampsonii* NSH-16, a porcine pathogen [70], and the human oral pathogen *Streptococcus sanguinis* SK36, but intriguingly ST7 genes also occur in the environmental actinobacterium *Beutenbergia cavernae* (Fig. S11), suggesting a yet unidentified role in the environment (as argued above for ST3 systems).

The final family, ST8, encompasses MFS transporters of the GPH type from within the MFS_2 Pfam family, which is distinct from the MFS_1 family featuring ST1 (NanT/X) and ST4 (NanZ) transporters (Fig. 3). We propose to call ST8 transporters NanG (for GPH). The only ST8 transporter identified prior to this work was found in some strains of *Lactobacillus salivarius* [51, 52], and here we discovered more orthologues through the conserved linkage to the NanOx/NanY oxidoreductase, and in some cases also to an IT-sialidase of the type of *S. pneumoniae* NanB, which has 2,7-anhydro-Neu5Ac as its the primary product [66] (Fig. S12). These genetic links suggest that ST8 transporters might specialize in the uptake of 2,7-anhydro-Neu5Ac, but this will need to be confirmed experimentally. The distribution of the newly identified ST8/NanG transporters strongly suggests that this family might have evolved recently, as orthologues occur in a very small number of *Firmicutes* and are highly similar (>70% aa identity) proteins (Fig. S12).

DISCUSSION

This *in silico* analysis provided novel insights into the evolution of bacterial sialic acid transporters, complementing the comprehensive work on the downstream catabolic enzymes [9, 13, 33, 34]. Using phylogeny as the basis for classification, we found that sialic acid transport has evolved no fewer than eight times from four major superfamilies of transporters (Table 1, Fig. 5). The most diversity was found within the ST5/SSS transporters, despite them likely having a single origin, and this family provided the first examples of fusions of transporters to other sialic acid metabolism-related proteins (SiaF). While two of the major superfamilies each contain two different ST groups, namely MFS_1 with ST1 and ST4, and ABC SBP_bac_1 with ST6 and ST7, the individual STs within each pair have completely different placements on the respective trees of these two large Pfam families, which in either case is consistent with the STs having evolved specificity for sialic acid independently, as opposed to them arising from, for example, the horizontal gene transfer into new phyla of a common sialic acid-transporting ancestor. Structure–function considerations (Table 1) match the phylogenetic distance between members of each ST pair, as for ST6 and ST7, they use different forms of sialic acid in an exclusive way (Neu5Ac and anhydro-Neu5Ac, respectively), consistent with independent origins, while in the case of the ST1 and ST4 proteins, all of the Neu5Ac-transporting ST1 proteins (i.e. NanT) contain 14

TMHs rather than the normal 12 TMHs seen in the ST4 transporters (also Neu5Ac-specific) and other MFS_1 proteins, which again is most parsimoniously explained by the independent evolution of substrate specificity from within already diversified groups. An illuminating comparison can be made with the ST3 group, where the global phylogeny within the large ABC SBP_bac_5 superfamily retains the placement of all ST3 proteins within a single group, yet shows a clear example of horizontal gene transfer into an ancestral *Pasteurellaceae* from an actinobacterial ancestor (Figs 3 and S5).

The only time the phylogenetic distribution of sialic acid transporter genes has been considered previously was in the comprehensive analysis of the distribution of the NanA enzyme, where transporter families were mapped onto the NanA phylogeny [9], using the broader classification to four families, namely SSS, MFS (NanT), ABC or TRAP [9]. The detailed and comprehensive study of human gut microbiome organisms by Ravcheev and Thiele [86] used the same four families for general classification, giving all transporters identified, however, arbitrary new names (Table 1). This unique transporter-focussed study now defines eight sialic acid transporter families (ST1–ST8) and proposes a unified naming scheme for these systems to reduce the significant confusion current in the community, and to provide a clear naming framework as new bacteria are identified that use sialic acid uptake in their biology. It is also the first study to include a classification of the recently discovered sialic acid transporters with specificity for the anhydro-form released by IT-sialidases and requiring the NanOx/NanY enzyme for utilization (Fig. 5). Our study now reveals that this latter specificity is present in at least four of the eight STs, suggesting that this is an ancient adaptation that is used by diverse bacteria to consume anhydro-sialic acid released in their environment (Fig. 5).

Our analysis identified for the first time candidate sialic acid transporters in *Planctomycetes* and *Verrucomicrobia* [5, 13], and expands the repertoire of sialic acid transporters in *Spirochaetes*, from one type of transporter [33] to four. Recent research has provideded evidence for Neu5Ac utilization by organisms expressing these novel transporters [68, 69, 73], lending support to our assignment. The recent observation by Pereira and colleagues that *A. muciniphila*'s growth was stimulated by the addition of Neu5Ac *in vitro* [73] supports a potential biological role for the sialic acid transporters identified in this work. Heterologous expression of these novel candidates will help determine their role in sialic acid transport, as done for the original SiaT [29].

One behaviour being better understood in the function of bacteria in the human gut microbiome is the cross-feeding between commensals and its exploitation by pathogens [16, 86]. One established example concerns the well-characterized common gut commensal *Bacteroides thetaiotaomicron* VPI-5482, which is able to liberate sialic acid by using its sialidase but not to consume the released Neu5Ac [87] and is, therefore, thought to act 'altruistically'. In keeping with these results, it has been stated that this bacterium lacks the sialic acid catabolic genes [16]. With regard to this latter statement, however, we note from the analysis presented in this paper that *B. thetaiotaomicron* does contain an ST5 family transporter (Fig. 4), which in fact retains all the residues of the Neu5Ac-binding site identified in *P. mirabilis* SiaT (Fig. S8). This gene, *BT_2813*, sits next to *BT_2814*, which encodes a NanA-like protein, and in some other related *Bacteroidetes*, such as *Cyclobacterium marinum* (Fig. 4), the same two genes are linked to a sialidase gene. Also at a separate locus there is a *nanE-II* gene, *BT_3605*, linked to orthologues of the *nanOU* outer membrane sialic acid transporter genes from *B. fragilis* [88], suggesting that *B. thetaiotaomicron* actually has the required complement of gene functions to take up and catabolize some form of sialic acid or even Neu5Ac. While the evidence for sialic acid cross-feeding in the gut is clear, our findings at least suggest that *B. thetaiotaomicron*'s capacity for sialic acid utilization should be checked in more detail (e.g. by heterologous expression of the transporter gene [29]), before the idea that this bacterium cannot use sialic acid at all is set in concrete [16].

Our discovery of protein fusions between the sialic acid mutarotase, NanM, and two unrelated types of sialic acid transporters (ST4 and ST5) is the first ever report of a covalent coupling between a metabolic enzyme and a transporter in sialic acid biology. It is intriguing that, in both of these independent instances of fusion, the fusion takes place at the N-terminus of the transporter, rather than at the C-terminus where is most frequent in bacteria [89, 90]; however, the extracytoplasmic positioning of the NanM moiety makes mechanistic sense and can be understood thinking of the location of NanM as a periplasmic enzyme. The biological function of NanM in the periplasm has been proposed to be to help bacteria acquire more rapidly the transported anomer of Neu5Ac, which is the β-anomer [71] (Fig. 5). In this line of reasoning, it makes teleological sense for a transporter to recruit NanM as an integral domain located in the periplasm, as this domain would feed β-Neu5Ac directly into the transporter as opposed to releasing it in solution where diffusion would limit its acquisition by the transporter.

Our work uncovered the unprecedented occurrence of orphan (i.e. devoid of cognate membrane components) sialic acid-binding proteins, all of which are orthologues of the ST6-SBP, SatE, found in *T. denticola* and related species, and which are instead genetically linked to a MCP protein predicted to function in small molecule sensing. Together with *Porphyromonas gingivalis* and *T. forsythia*, *T. denticola* forms the so-called Red Complex, a consortium of micro-organisms that underlie the inflammatory condition known as periodontitis leading to tooth loss [83]. While the role of sialic acid metabolism in periodontal pathogens is still under investigation, all three organisms use extracellular sialidases for the release of free Neu5Ac, which *T. forsythia* can also consume [31, 83, 84]. In the light of our finding, it is unlikely that the SatE orthologue of *T. denticola* functions as part of a transporter, and we suggest instead that it mediates some yet uncharacterized cellular response to sialic acid via physical interaction with the associated MCP. The occurrence in the *T. denticola* ST6 locus of a supernumerary copy of the gene for the flagellar motor switch protein FliG

(Fig. S10) might be taken as a hint for a role in chemotaxis, but research is warranted to explore all aspects of this newly introduced (and admittedly speculative) notion of sialic acid sensing.

We have previously considered the consequences of using different types of transporter mechanisms for sialic acid acquisition and attempted to rationalize this to the biological niche that the bacterium inhabits [12]. Given this larger dataset, it is worth considering this again in, for example, the anaerobic environment of the human colon. Bacteria here are growing by fermentation by scavenging dietary and host-derived carbon and energy sources. While one might expect transport to use the most energetically efficient systems in this niche, the reality is that the bacteria that live there use all varieties of transporter mechanism for sialic acid uptake. This analysis is complicated now by the occurrence of often more than one sialic acid transporter in a single bacterium, which potentially allows the organism to use different transporters under different environmental conditions. The reason to have a Neu5Ac transporter and a separate 2,7-anhydro-Neu5Ac transporter for a gut bacterium appears clear, as both substrates will be liberated from host glycans depending on which sialidases are made by the community, and it does seem advantageous for a species to be able to take up whichever form is available. However, it is not clear why bacteria have multiple Neu5Ac transporters, although again this might relate to individual preferences for other forms of sialic acid, such as Neu5Gc available through the diet. Also, having multiple transporters for Neu5Ac with varying affinity could be a good reason to have apparently redundant systems, as again it would allow the bacterium to scavenge sialic acid most efficiently, if its concentration were changing in the environment.

As microbiome-sequencing projects resolve an ever-growing number of microbial communities in health and disease, the diversity among sialic acid transporters is likely to increase and research is warranted to decipher the functional role and substrate specificity of these important transporters. Our naming system, which matches the eight phylogenetically different families of transporters, now provides a clear framework to accommodate newly identified members within these families and can be expanded as new sialic acid transporter families are discovered in the future.

Funding information
The authors gratefully acknowledge the support of the BBSRC Institute Strategic Program Gut Microbes and Health BB/R012490/1 and BBSRC Responsive Mode grant BB/P008895/1.

Author contributions
The study was conceived and undertaken by E. S., M. R. and G. H. T., with input on data analysis and interpretation from A. B., N. J. and T. P. The manuscript was written by E. S., M. R. and G. H. T., with additional input from all authors.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Varki A. Uniquely human evolution of sialic acid genetics and biology. *Proc Natl Acad Sci USA* 2010;107:8939–8946.
2. Angata T, Varki A. Chemical diversity in the sialic acids and related α-keto acids: an evolutionary perspective. *Chem Rev* 2002;102:439–469.
3. Ghosh S. Sialic acid and biology of life: an introduction. In: *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease*. Cambridge, MA: Academic Press, 2020. pp. 1–61.
4. Vimr ER. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol* 2013;2013:816713.
5. Haines-Menges BL, Whitaker WB, Lubin JB, Boyd EF. Host sialic acids: a delicacy for the pathogen with discerning taste. *Microbiol Spectr* 2015;3:MBP-0005-2014.
6. Matrosovich M, Herrler G, Klenk HD. Sialic acid receptors of viruses. *Top Curr Chem* 2015;367:1–28.
7. Li W, Hulswit RJG, Widjaja I, Raj VS, McBride R, *et al*. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci USA* 2017;114:E8508–E8517.
8. Severi E, Hood D, Thomas G. Sialic acid utilization by bacterial pathogens. *Microbiology* 2007;153:2817–2822.
9. McDonald ND, Lubin JB, Chowdhury N, Boyd EF. Host-derived sialic acids are an important nutrient source required for optimal bacterial fitness in vivo. *mBio* 2016;7:e02237-15.
10. North RA, Horne CR, Davies JS, Remus DM, Muscroft-Taylor AC, *et al*. "Just a spoonful of sugar...": import of sialic acid across bacterial cell membranes. *Biophys Rev* 2018;10:219–227.
11. McDonald ND, Boyd EF. Structural and biosynthetic diversity of nonulosonic acids (Nulos) that decorate surface structures in bacteria. *Trends Microbiol* 2021;29:142–157.
12. Thomas GH. Sialic acid acquisition in bacteria-one substrate, many transporters. *Biochem Soc Trans* 2016;44:760–765.
13. Almagro-Moreno S, Boyd EF. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol Biol* 2009;9:118.
14. Bell A, Brunt J, Crost E, Vaux L, Nepravishta R, *et al*. Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nat Microbiol* 2019;4:2393–2404.
15. Severi E, Randle G, Kivlin P, Whitfield K, Young R, *et al*. Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol Microbiol* 2005;58:1173–1185.
16. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, *et al*. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 2013;502:96–99.
17. Marion C, Burnaugh AM, Woodiga SA, King SJ. Sialic acid transport contributes to pneumococcal colonization. *Infect Immun* 2011;79:1262–1269.
18. Huang Y-L, Chassard C, Hausmann M, Von Itzstein M, Hennet T. Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nat Commun* 2015;6:8141.
19. Agarwal K, Robinson LS, Aggarwal S, Foster LR, Hernandez-Leyva A, *et al*. Glycan cross-feeding supports mutualism between *Fusobacterium* and the vaginal microbiota. *PLoS Biol* 2020;18:e3000788.
20. Stafford G, Roy S, Honma K, Sharma A. Sialic acid, periodontal pathogens and *Tannerella forsythia*: stick around and enjoy the feast! *Mol Oral Microbiol* 2012;27:11–22.
21. Hopkins AP, Hawkhead JA, Thomas GH. Transport and catabolism of the sialic acids N-glycolylneuraminic acid and 3-keto-3-deoxy-D-glycero-D-galactonononic acid by *Escherichia coli* K-12. *FEMS Microbiol Lett* 2013;347:14–22.
22. North RA, Wahlgren WY, Remus DM, Scalise M, Kessans SA, *et al*. The sodium sialic acid symporter from *Staphylococcus aureus* has altered substrate specificity. *Front Chem* 2018;6:233.
23. Setty TG, Mowers JC, Hobbs AG, Maiya SP, Syed S, *et al*. Molecular characterization of the interaction of sialic acid with the

periplasmic binding protein from *Haemophilus ducreyi*. *J Biol Chem* 2018;293:20073–20084.

24. Saha S, Coady A, Sasmal A, Kawanishi K, Choudhury B, *et al*. Exploring the impact of ketodeoxynonulosonic acid in host-pathogen interactions using uptake and surface display by nontypeable *Haemophilus influenzae*. *mBio* 2021;12:e03226-20.
25. Vimr ER, Troy FA. Identification of an inducible catabolic system for sialic acids (nan) in *Escherichia coli*. *J Bacteriol* 1985;164:845–853.
26. Martinez J, Steenbergen S, Vimr E. Derived structure of the putative sialic acid transporter from *Escherichia coli* predicts a novel sugar permease domain. *J Bacteriol* 1995;177:6005–6010.
27. Plumbridge J, Vimr E. Convergent pathways for utilization of the amino sugars. *Microbiology* 1999;181:47–54.
28. Mulligan C, Leech AP, Kelly DJ, Thomas GH. The membrane proteins SiaQ and SiaM form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter SiaPQM (VC1777-1779) from *Vibrio cholerae*. *J Biol Chem* 2012;287:3598–3608.
29. Severi E, Hosie AHF, Hawkhead JA, Thomas GH. Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and in vivo comparison with known bacterial sialic acid transporters. *FEMS Microbiol Lett* 2010;304:47–54.
30. Brigham C, Caughlan R, Gallegos R, Dallas MB, Godoy VG, *et al*. Sialic acid (N-acetyl neuraminic acid) utilization by *Bacteroides fragilis* requires a novel N-acetyl *Mannosamine epimerase*. *J Bacteriol* 2009;191:3629–3638.
31. Roy S, Douglas CWI, Stafford GP. A novel sialic acid utilization and uptake system in the periodontal pathogen *Tannerella forsythia*. *J Bacteriol* 2010;192:2285–2293.
32. Allen S, Zaleski A, Johnston JW, Gibson BW, Apicella MA. Novel sialic acid transporter of *Haemophilus influenzae*. *Infect Immun* 2005;73:5291–5300.
33. Chowdhury N, Norris J, McAlister E, Lau SYK, Thomas GH, *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* 2012;158:2158–2167.
34. Lubin JB, Kingston JJ, Chowdhury N, Boyd EF. Sialic acid catabolism and transport gene clusters are lineage specific in *Vibrio vulnificus*. *Appl Environ Microbiol* 2012;78:3407–3415.
35. Müller A, Severi E, Mulligan C, Watts AG, Kelly DJ, *et al*. Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from *Haemophilus influenzae*. *J Biol Chem* 2006;281:22212–22222.
36. Mulligan C, Geertsma ER, Severi E, Kelly DJ, Poolman B, *et al*. The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc Natl Acad Sci USA* 2009;106:1778–1783.
37. Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. Tripartite ATP-independent periplasmic (TRAP) transporters and tripartite tricarboxylate transporters (TTT): from uptake to pathogenicity. *Front Cell Infect Microbiol* 2018;8:33.
38. Post DMB, Mungur R, Gibson BW, Munson RS. Identification of a novel sialic acid transporter in *Haemophilus ducreyi*. *Infect Immun* 2005;73:6727–6735.
39. Gruteser N, Marin K, Krämer R, Thomas GH. Sialic acid utilization by the soil bacterium *Corynebacterium glutamicum*. *FEMS Microbiol Lett* 2012;336:131–138.
40. Sela DA, Li Y, Lerno L, Wu S, Marcobal AM, *et al*. An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J Biol Chem* 2011;286:11909–11918.
41. Fischer M, Hopkins AP, Severi E, Hawkhead J, Bawdon D, *et al*. Tripartite ATP-independent periplasmic (TRAP) transporters use an arginine-mediated selectivity filter for high affinity substrate binding. *J Biol Chem* 2015;290:27113–27123.
42. Gangi Setty T, Cho C, Govindappa S, Apicella MA, Ramaswamy S. Bacterial periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr D Biol Crystallogr* 2014;70:1801–1811.
43. Johnston JW, Coussens NP, Allen S, Houtman JCD, Turner KH, *et al*. Characterization of the N-acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of nontypeable *Haemophilus influenzae* strain 2019. *J Biol Chem* 2008;283:855–865.
44. Marion C, Aten AE, Woodiga SA, King SJ. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. *Infect Immun* 2011;79:4193–4200.
45. Hentrich K, Löfling J, Pathak A, Nizet V, Varki A, *et al*. *Streptococcus pneumoniae* senses a human-like sialic acid profile via the response regulator CIAR. *Cell Host Microbe* 2016;20:307–317.
46. Afzal M, Shafeeq S, Ahmed H, Kuipers OP. Sialic acid-mediated gene expression in *Streptococcus pneumoniae* and role of NanR as a transcriptional activator of the nan gene cluster. *Appl Environ Microbiol* 2015;81:3121–3131.
47. Wahlgren WY, Dunevall E, North RA, Paz A, Scalise M, *et al*. Substrate-bound outward-open structure of a Na+-coupled sialic acid symporter reveals a new Na+ site. *Nat Commun* 2018;9:1753.
48. Anba-Mondoloni J, Chaillou S, Zagorec M, Champomier-Vergès MC. Catabolism of N-acetylneuraminic acid, a fitness function of the food-borne lactic acid bacterium *Lactobacillus sakei*, involves two newly characterized proteins. *Appl Environ Microbiol* 2013;79:2012–2018.
49. Crost EH, Tailford LE, Monestier M, Swarbreck D, Henrissat B, *et al*. The mucin-degradation strategy of *Ruminococcus gnavus*: the importance of intramolecular trans-sialidases. *Gut Microbes* 2016;7:302–312.
50. Tailford LE, Owen CD, Walshaw J, Crost EH, Hardy-Goddard J, *et al*. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* 2015;6:7624.
51. Bell A, Severi E, Lee M, Monaco S, Latousakis D, *et al*. Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria. *J Biol Chem* 2020;295:13724–13736.
52. Kentache T, Thabault L, Peracchi A, Frédérick R, Bommer GT, *et al*. The putative *Escherichia coli* dehydrogenase YJHC metabolises two dehydrated forms of n-acetylneuraminate produced by some sialidases. *Biosci Rep* 2020;40:BSR20200927.
53. Almagro-Moreno S, Boyd EF. Bacterial catabolism of nonulosonic (sialic) acid and fitness in the gut. *Gut Microbes* 2010;1:45–50.
54. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 2001;305:567–580.
55. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, *et al*. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019;37:420–423.
56. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008;9:40.
57. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, *et al*. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296–W303.
58. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25:1189–1191.
59. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 2016;44:W232–W235.
60. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
61. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014;42:320–324.
62. Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH. The major facilitator superfamily (MFS) revisited. *FEBS J* 2012;279:2022–2035.

63. Wang SC, Davejan P, Hendargo KJ, Javadi-Razaz I, Chou A, *et al*. Expansion of the Major Facilitator Superfamily (MFS) to include novel transporters as well as transmembrane-acting enzymes. *Biochim Biophys Acta Biomembr* 2020;1862:183277.
64. Michaels DL, Moneypenny CG, Shama SM, Leibowitz JA, May MA, *et al*. Sialidase and N-acetylneuraminate catabolism in nutrition of *Mycoplasma alligatoris*. *Microbiol (United Kingdom)* 2019;165:662–667.
65. Owen CD, Lukacik P, Potter JA, Sleator O, Taylor GL, *et al*. *Streptococcus pneumoniae* NanC: Structural insights into the specificity and mechanism of a sialidase that produces a sialidase inhibitor. *J Biol Chem* 2015;290:27736–27748.
66. Xu G, Kiefel MJ, Wilson JC, Andrew PW, Oggioni MR, *et al*. Three *Streptococcus pneumoniae* sialidases: Three different products. *J Am Chem Soc* 2011;133:1718–1721.
67. Xiao K, Wang X, Yu H. Comparative studies of catalytic pathways for *Streptococcus pneumoniae* sialidases NanA, NanB and NanC. *Sci Rep* 2019;9:2157.
68. Quintana-Hayashi MP, Venkatakrishnan V, Haesebrouck F, Lindén S. Role of sialic acid in brachyspira hyodysenteriae adhesion to pig colonic mucins. *Infect Immun* 2019;87:1–11.
69. Mappley LJ, Black ML, AbuOun M, Darby AC, Woodward MJ, *et al*. Comparative genomics of *Brachyspira pilosicoli* strains: genome rearrangements, reductions and correlation of genetic compliment with phenotypic diversity. *BMC Genomics* 2012;13:454.
70. Mirajkar NS, Phillips ND, La T, Hampson DJ, Gebhart CJ. Characterization and recognition of *Brachyspira hampsonii* sp. nov., novel intestinal spirochete that Is pathogenic to pigs. *J Clin Microbiol* 2016;54:2942–2949.
71. Severi E, Müller A, Potts JR, Leech A, Williamson D, *et al*. Sialic acid mutarotation is catalyzed by the *Escherichia coli* β-propeller protein YjhT. *J Biol Chem* 2008;283:4841–4849.
72. Mally M, Shin H, Paroz C, Landmann R, Cornelis GR. *Capnocytophaga canimorsus*: A human pathogen feeding at the surface of epithelial cells and phagocytes. *PLoS Pathog* 2008;4:e1000164.
73. Pereira FC, Wasmund K, Cobankovic I, Jehmlich N, Craig W, *et al*. Rational design of a microbial consortium of mucosal sugar utilizers reduces *Clostridiodes difficile* colonization. *Nat Commun* 2020;11:1–21.
74. Kostopoulos I, Elzinga J, Ottman N, Klievink JT, Blijenberg B, *et al*. *Akkermansia muciniphila* uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Sci Rep* 2020;10:14330.
75. Yoneda S, Loeser B, Feng J, Dmytryk J, Qi F, *et al*. Ubiquitous sialometabolism present among oral fusobacteria. *PLoS One* 2014;9:e99263.
76. Egan M, O'Connell Motherway M, Ventura M, van Sinderen D. Metabolism of sialic acid by *Bifidobacterium breve* ucc2003. *Appl Environ Microbiol* 2014;80:4414–4426.
77. Lewis WG, Robinson LS, Gilbert NM, Perry JC, Lewis AL. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted actinobacterium Gardnerella vaginalis. *J Biol Chem* 2013;288:12067–12079.
78. Aisaka K, Uwajima T. Production of neuraminidase by *Micromonospora viridifaciens*. *FEMS Microbiol Lett* 1987;44:289–291.
79. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a β-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* 1995;3:1197–1205.
80. Uhde A, Brühl N, Goldbeck O, Matano C, Gurow O, *et al*. Transcription of sialic acid catabolism genes in *Corynebacterium glutamicum* is subject to catabolite repression and control by the transcriptional repressor NanR. *J Bacteriol* 2016;198:2204–2218.
81. Pezzicoli A, Ruggiero P, Amerighi F, Telford JL, Soriani M. Exogenous sialic acid transport contributes to group B *Streptococcus* infection of mucosal surfaces. *J Infect Dis* 2012;206:924–931.
82. Salah Ud-Din AIM, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell Mol Life Sci* 2017;74:3293–3303.
83. Stafford GP, Sharma A. Periodontal pathogen sialometabolic activity in periodontitis. In: *Emerging Therapies in Periodontics*. Cham: Springer, 2020. pp. 187–194.
84. Kurniyati K, Zhang W, Zhang K, Li C. A surface-exposed neuraminidase affects complement resistance and virulence of the oral spirochaete *Treponema denticola*. *Mol Microbiol* 2013;89:842–856.
85. Zhang Y, Gardina PJ, Kuebler AS, Kang HS, Christopher JA, *et al*. Model of maltose-binding protein/chemoreceptor complex supports intrasubunit signaling mechanism. *Proc Natl Acad Sci U S A* 1999;96:939–944.
86. Ravcheev DA, Thiele I. Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Front Genet* 2017;8:1.
87. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, *et al*. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* 2011;10:507–514.
88. Phansopa C, Roy S, Rafferty JB, Douglas CWI, Pandhal J, *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* 2014;458:499–511.
89. Willson BJ, Chapman LN, Thomas GH. Evolutionary dynamics of membrane transporters and channels: enhancing function through fusion. *Curr Opin Genet Dev* 2019;58–59:76–86.
90. Willson BJ, Dalzell L, Chapman LNM, Thomas GH. Enhanced functionalisation of major facilitator superfamily transporters via fusion of C-terminal protein domains is both extensive and varied in bacteria. *Microbiol (United Kingdom)* 2019;165:419–424.
91. Kentache T, Thabault L, Deumer G, Haufroid V, Frédérick R, *et al*. The metalloprotein YhcH is an anomerase providing N-acetylneuraminate aldolase with the open form of its substrate. *J Biol Chem* 2021;296:100699.
92. Kalivoda KA, Steenbergen SM, Vimr ER. Control of the *Escherichia coli* sialoregulon by transcriptional repressor NanR. *J Bacteriol* 2013;195:4689–4701.
93. Allen S, Zaleski A, Johnston JW, Gibson BW, Apicella MA. Novel sialic acid transporter of *Haemophilus influenzae*. *Infect Immun* 2005;73:5291–5300.
94. Tatum FM, Tabatabai LB, Briggs RE. Sialic acid uptake is necessary for virulence of *Pasteurella multocida* in turkeys. *Microb Pathog* 2009;46:337–344.
95. Steenbergen SM, Lichtensteiger CA, Caughlan R, Garfinkle J, Fuller TE, *et al*. Sialic acid metabolism and systemic pasteurellosis. *Infect Immun* 2005;73:1284–1294.
96. Robinson LS, Schwebke J, Lewis WG, Lewis AL. Identification and characterization of NanH2 and NanH3, enzymes responsible for sialidase activity in the vaginal bacterium *Gardnerella vaginalis*. *J Biol Chem* 2019;294:5230–5245.
97. Johnston JW, Coussens NP, Allen S, Houtman JCD, Turner KH, *et al*. Characterization of the N -Acetyl-5-neuraminic acid-binding site of the extracytoplasmic solute receptor (SiaP) of Nontypeable *Haemophilus influenzae* strain 2019. *J Biol Chem* 2008;283:855–865.
98. Olson ME, King JM, Yahr TL, Horswill AR. Sialic acid catabolism in *Staphylococcus aureus*. *J Bacteriol* 2013;195:1779–1788.
99. Gelfand MS, Rodionov DA. Comparative genomics and functional annotation of bacterial transporters. *Phys Life Rev* 2008;5:22–49.
100. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, *et al*. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res* 2021;49:D461–D467.