

Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure

Amanda L. Lewis^{a,b,1,2}, Nolan Desa^a, Elizabeth E. Hansen^c, Yuriy A. Knirel^d, Jeffrey I. Gordon^c, Pascal Gagneux^{a,e}, Victor Nizet^{a,b,f}, and Ajit Varki^{a,e,g,1}

^aGlycobiology Research and Training Center, Departments of ^bPediatrics, ^gMedicine, and ^cCellular and Molecular Medicine, School of Medicine, and ^fSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093; ^dN.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt 47, 11991 Moscow, Russia; and ^eCenter for Genome Sciences, Washington University, St. Louis, MO 63108

Edited by Sen-itiroh Hakomori, Pacific Northwest Diabetes Research Institute, Seattle, WA, and approved June 19, 2009 (received for review March 9, 2009)

Sialic acids (Sias) are nonulosonic acid (NuLO) sugars prominently displayed on vertebrate cells and occasionally mimicked by bacterial pathogens using homologous biosynthetic pathways. It has been suggested that Sias were an animal innovation and later emerged in pathogens by convergent evolution or horizontal gene transfer. To better illuminate the evolutionary processes underlying the phenomenon of Sia molecular mimicry, we performed phylogenomic analyses of biosynthetic pathways for Sias and related higher sugars derived from 5,7-diamino-3,5,7,9-tetraoxynon-2-ulosonic acids. Examination of $\approx 1,000$ sequenced microbial genomes indicated that such biosynthetic pathways are far more widely distributed than previously realized. Phylogenetic analysis, validated by targeted biochemistry, was used to predict NuLO types (i.e., neuraminic, legionaminic, or pseudaminic acids) expressed by various organisms. This approach uncovered previously unreported occurrences of Sia pathways in pathogenic and symbiotic bacteria and identified at least one instance in which a human archaeal symbiont tentatively reported to express Sias in fact expressed the related pseudaminic acid structure. Evaluation of targeted phylogenies and protein domain organization revealed that the “unique” Sia biosynthetic pathway of animals was instead a much more ancient innovation. Pathway phylogenies suggest that bacterial pathogens may have acquired Sia expression via adaptation of pathways for legionaminic acid biosynthesis, one of at least 3 evolutionary paths for de novo Sia synthesis. Together, these data indicate that some of the long-standing paradigms in Sia biology should be reconsidered in a wider evolutionary context of the extended family of NuLO sugars.

legionaminic acid | phylogeny | pseudaminic acid | neuraminic acid | biosynthetic pathway

Sialic acids (Sias) are displayed in prominent positions on vertebrate cells and are critical for such physiological processes as cellular repulsion, renal filtration, and neuronal plasticity (1, 2). Many other Sia-dependent functions occur in conjunction with Sia-binding lectins, including down-modulation of complement activity and the regulation of leukocyte activation, migration, and apoptosis (1, 3). The divergence of the superphyla protostomes and deuterostomes[†] created a dichotomy in animal Sia expression, and heralded the emergence of widespread Sia-dependent biological functions in deuterostomes. Nearly 5 decades of research have confirmed that with few exceptions, these unique 9-carbon backbone sugars are conspicuously absent from many eukaryotic lineages, including most protostomes, plants, fungi, and protists (1, 4). Sia decoration by de novo biosynthesis or via metabolic scavenging pathways has been reported in more than a dozen pathogenic bacterial species (5, 6) and also was recently described in a human gut-associated methanogenic archaeon (7). During infection, microbes displaying Sia mimicry can exploit host factor H and/or Siglec-9 to down-regulate alternative complement deposition and neutrophil bactericidal activities (8–11).

Sias are 9-carbon backbone derivatives of neuraminic (Neu) and ketodeoxynonulosonic (Kdn) acids. They are actually part of a larger family of carbohydrate structures collectively called nonulosonic acids (NuLOs)[‡]. A number of NuLO sugars other than Sias have been found in microbes, all of which are derivatives of 4 isomeric 5,7-diamino-3,5,7,9-tetraoxynon-2-ulosonic acids (12). At least 2 of these, the D-glycero-d-galacto isomer [legionaminic acid (Leg)] (13, 14) and L-glycero-l-manno isomer [pseudaminic acid (Pse)] (15, 16), have striking structural and biosynthetic similarities to Sias (Fig. 1). These commonalities among NuLO pathways reflect the structural similarity of all of the NuLO sugars, as well as their uniqueness compared with other monosaccharides. Similar steps in each NuLO biosynthetic (NAB)[§] pathway are catalyzed by homologous enzymes, including the condensation of a 6-carbon sugar intermediate with 3-carbon phosphoenolpyruvate (3C) to generate the 9-carbon backbone NuLO sugar, followed by the activation of free NuLO residues using cytidine triphosphate to form cytidine monophosphate (CMP)-NuLO intermediates (Fig. 1). In *Campylobacter* species and other ϵ -proteobacteria, Pse modifications play critical roles in flagellar assembly and, consequently, motility (16, 17), an important physiological function in aquatic environments and for association with animals. Leg modifications also have been identified on flagellar subunits, but have less well-defined functions (14, 18, 19). Both Leg and Pse also have

Author contributions: A.L.L., V.N., and A.V. designed research; A.L.L., N.D., and E.E.H. performed research; Y.A.K. and J.I.G. contributed new reagents/analytic tools; A.L.L., E.E.H., P.G., V.N., and A.V. analyzed data; and A.L.L., Y.A.K., J.I.G., P.G., V.N., and A.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence may be addressed. E-mail: allewis@wustl.edu or alvarki@ucsd.edu.

²Current address: Washington University School of Medicine, St. Louis, MO 63110.

[†]Protostomes and deuterostomes are bilaterian animals with distinct developmental programs for gut tube formation in which the first opening of the embryo, the blastopore, becomes either the mouth (protostomes, “mouth first”) or the anus (deuterostomes, “mouth second”).

[‡]There currently is no established nomenclature defining the 9 carbon α -keto acids as a group. Following discussions with Hans Kamerling, Roland Schauer, and Nathan Sharon, here we use the abbreviation “NuLO” for non-2-ulosonic acids, which assumes NuL for non-2-uloses and maintains the discrimination between aldonic acids and uronic acids, such as glucuronic acid. We suggest that the use of the term “sialic acid” (Sia) continue to be limited to its original use in describing Neu, Kdn, and their derivatives in deuterostomes and their pathogens, and that NuLO be used to encompass the entire group of 9 carbon α -keto acids.

[§]NAB pathways have been defined in various bacterial pathogens. Enzymes in each of these pathways have been given different designations, beginning with “Neu” for neuraminic, “Pse” for pseudaminic, and “Ptm” for posttranslational modification (see Fig. 1 for published designations). To avoid confusion, here we use “NAB” to refer to homologous steps in these related biosynthetic pathways, numbering the enzymatic steps as shown in Fig. 1.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902431106/DCSupplemental.

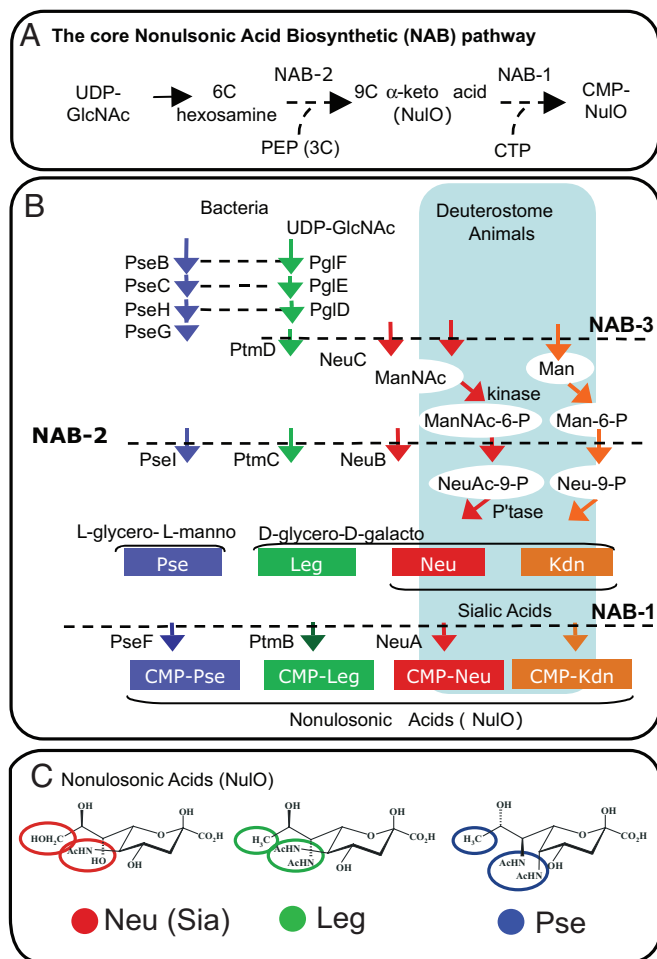


Fig. 1. Related NAB pathways synthesize chemically related sugars. NuLOs include all 9-carbon backbone α -keto acid sugars. NuLO sugars described to date conform to one of several core backbones that can be further modified by epimerization or modification (1, 12). Major core backbones include the Neu, Kdn, Leg, and Pse acids. (A), The unique and shared “core” features of all NAB pathways include UDP-*N*-acetylglucosamine as a starting point, condensation of a 6-carbon intermediate with phosphoenolpyruvate to yield a 9-carbon α -keto acid (NAB-2), and formation of a CMP-activated NuLO intermediate (NAB-1). (B), Architecture and nomenclature of NuLO pathways. Horizontal dashed lines denote NAB enzymes in different pathways that share a common ancestor as deduced by amino acid sequence similarity. (C), Chemical structures of *N*-acetyl derivatives of Neu, Leg, and Pse.

been identified as part of lipopolysaccharide (LPS) O antigens in some Gram-negative bacteria (12), where they conceivably could contribute to biofilm formation, resistance to phage predation, or animal associations. Despite the similarities of Leg and Pse to Sias, the potential roles of these sugars in host-pathogen interactions remain poorly defined, and their distribution among microbes has not yet been systematically investigated.

In the present work, we probed the existing paradigms of Sia evolution using genomic, phylogenetic, and biochemical approaches to ask whether Sias were a unique innovation of the deuterostome lineage, whether bacterial mimicry of host Sias was the result of lateral gene transfer from an animal host or convergent evolution from microbial Sia-like biosynthetic pathways, and whether the chemical structure of Sias and related sugars can be predicted from genomic sequence information.

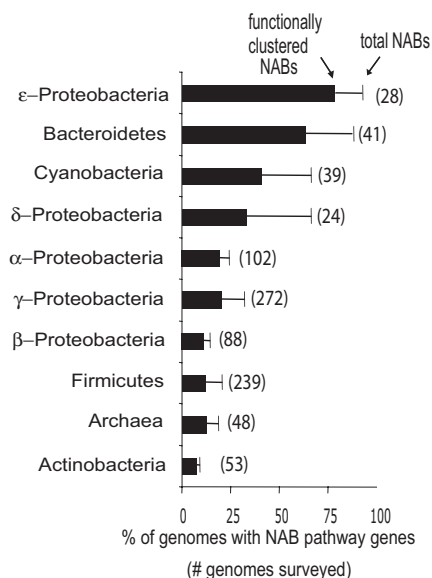


Fig. 2. Distribution of predicted NAB pathways among bacteria and archaea. NAB enzymes NAB-1 and NAB-2 were identified by BLASTp in the genomes of various bacterial phyla and in sequenced archaeons. The number of genomes in each group is given in parentheses. Bars reflect the percentage of genomes in each group with one or more physically clustered NAB gene pairs, and thin line extensions reflect total NAB homolog pairs irrespective of functional clustering.

Results and Discussion

“Functional Clustering” Predicts a Remarkably Wide Distribution of NuLO Sugar Expression Among Bacteria and Archaea. To define the distribution of biosynthetic pathways for NuLO sugars in members of bacteria and archaea, nearly 1,000 sequenced microbial genomes[†] were examined by BLAST for evidence of “functional clusters” (20) of NAB pathway genes. Unexpectedly, about 20% of all microbial phylogenetic types (phylotypes) sequenced to date were found to encode NAB pathway cassettes (Fig. 2).

Many species/subsets, as well as entire phyla in which NuLO sugars have never been documented, were found to have NAB enzymes in their genomes, including remarkably large proportions of available Bacteroidetes (36/41), Cyanobacteria (26/39), and δ -Proteobacteria (16/24), certain pathogenic members of the order *Spirochaetales*, and nearly 19% of sequenced Archaea (9/48) (Fig. 2). NAB pathways were identified in a larger fraction of the available genomes in bacterial phyla (divisions) previously known to include Sia-decorated pathogens (i.e., γ -Proteobacteria, β -Proteobacteria, and Firmicutes). Our analysis reveals a far wider distribution of NuLOs and a deeper evolutionary history of this class of sugars than originally assumed.

Phylogenetic Prediction of NuLO Types Reveals an Evolutionary Context for Sias and Sia-Like Sugars. To better illuminate the evolutionary history of these 9-carbon backbone NuLO sugars and predict their distribution and structure, we performed phylogenetic analysis of the most highly conserved enzyme in the pathway (NAB-2) and overlaid this tree with published biochemical data [Fig. 3A; summarized with strain and sequence identifiers in supporting information (SI) Table S1]. NAB-2 condenses a 6-carbon intermediate with the 3-carbon molecular phosphoenolpyruvate to generate NuLOs of different types (Fig.

[†]These sequences had been previously deposited in the Genbank database. For a list of accession numbers, see Table S1. Annotations have been updated in The SEED, an annotation/analysis tool provided by the Fellowship for Interpretation of Genomes.

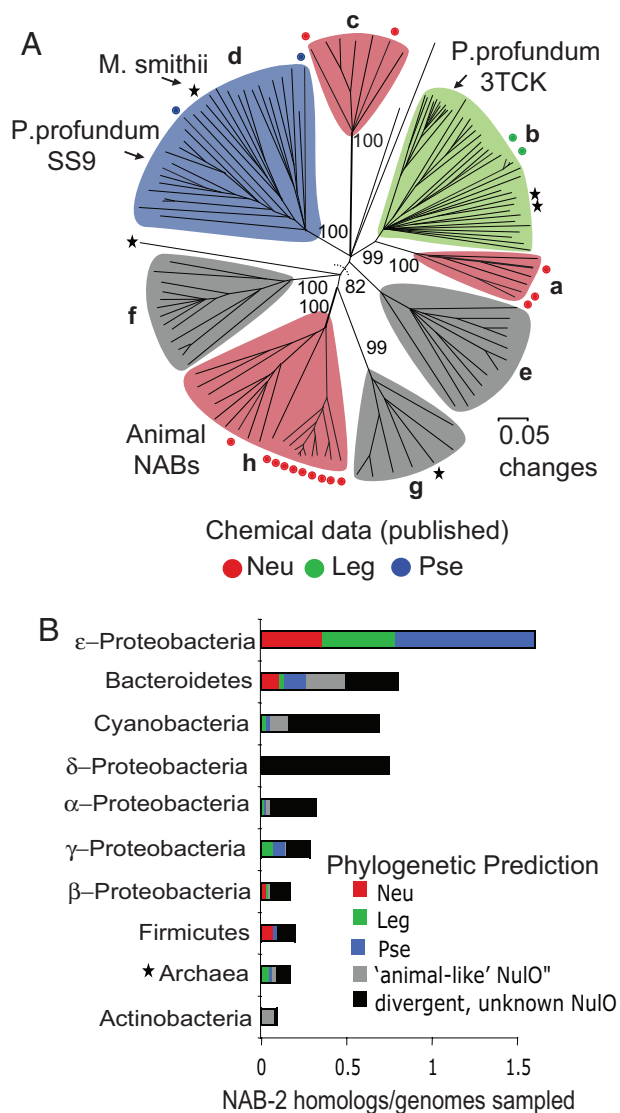


Fig. 3. NAB-2 phylogeny for predicting NuLO structure. (A), A distance-based neighbor-joining tree was constructed as described in *Materials and Methods*, and published biochemical data were overlaid onto the tree (colored circles). A “cohesion group” approach (21) was used to infer enzyme specificities by extrapolation of the published biochemical data (13–15, 22–24) to phylogenetic clades supported by high bootstrap values (shown at relevant nodes). Color shading reflects the phylogenetic predictions of chemical structure for each clade. Clades are designated “a”–“h” for reference. Note that clade “h” is composed entirely of NAB-2 enzymes from animals. Organisms for which biochemical data are presented in later figures are indicated by name. (B), Percentages of NAB-2 clade affiliations were calculated and expressed as a function of the number of genomes surveyed in each NCBI-classified taxonomic group (see *Materials and Methods*). Note that “pruned” tree branches representing diverged NAB-2 enzymes were included in the tabulation of NAB-positive genomes as “unknown” NuLO type (shown in black) if a nearby NAB-1 homolog was identified. Some individual genomes encode multiple NAB pathways leading to NAB-2 homolog/genome sampled values of >1 . These data, with strain and sequence identifiers, are summarized in [Table S1](#).

1). Microbial NAB-2 homologs with published roles in Neu synthesis were identified in multiple clades that are phylogenetically distinct from the clade with animal NAB-2 homologs (Fig. 3A). Specifically, the analysis revealed 2 groups of bacterial Neu synthases, represented by clades “a” and “c” (Fig. 3A), which include sequences from well-known Sia-decorated pathogens (Table 1). These data point to the existence of at least 3

evolutionarily discrete types of Neu synthases that ultimately share a common ancestor (i.e., semiconvergence). Examination of other NAB pathway enzymes revealed similar phylogenetic signatures, indicating that most microbial NAB pathway cassettes persist as evolutionarily conserved units (Fig. S1). These data predict Neu expression in several pathogenic and commensal bacterial species in which NAB biosynthesis has not been characterized previously (Table 1). Moreover, the data strongly suggest that Neu expression is *not* limited to microbes that associate with animals, as has been commonly assumed (see the composition of clade “c” in Table 1).

One evolutionary explanation for the emergence of Sia mimicry is represented by highly supported nodes in the NAB-2 and NAB-3 trees showing that clade “a” (Neu-specific enzymes encoded exclusively in animal-associated bacteria) shares a common ancestor with clade “b” (Leg-specific enzymes encoded mostly by aquatic organisms) (Fig. 3A and Fig. S1). Interestingly, the phylogenetic relationship between NAB pathways represented by clades “a” and “b” do not reflect known evolutionary relationships between organisms represented in these clades. Consistent with the shared ancestry of clades “a” and “b,” organisms represented in these clades (but not those in clade “c”) encode acetyltransferases as part of their NAB gene cassettes. Previous studies have indicated that such acetyltransferases are required for overall NuLO expression (16, 25, 26). For example, *Campylobacter jejuni* (clade “b”) requires PtmH for *N*-acetylation at the 7-carbon position of Pse residues (16), whereas *Streptococcus agalactiae* and *Escherichia coli* (clade “a”) use the homologous NeuD enzyme for *O*-acetylation at the same carbon position of Neu (26). These data indicate that Neu biosynthetic pathways in clade “a” were not acquired by lateral gene transfer from an animal host, but rather the Neu mimicry by organisms represented in clade “a” arose by recruitment and modification of an ancestral NuLO pathway that requires an acetylation reaction at C7 (Fig. 1).

NuLO Biosynthetic Pathways Originated and Diversified Early in the History of Cellular Life. Based on the distribution of Sias among animals, it has been suggested that the biosynthetic pathway for this NuLO sugar may have been an innovation of the deuterostome lineage (1). But the NAB-2-based phylogeny (Fig. 3) revealed novel phylogenetic clusters of microbial NAB homologs highly similar to components of animal Sia pathways (clades “f,” “g,” and possibly “e”), suggesting that Sia synthesis pathways of animals may have much deeper evolutionary roots. Notably, the branching pattern of the limited number of taxa within these clades is consistent with known evolutionary relationships among these organisms. Targeted phylogenetic trees and protein domain analyses of NAB-1 and NAB-2 pairs encoded in the animal and “animal-like” clades further support the deep but firm evolutionary relationship between these biosynthetic pathways (Fig. S2). A few of the organisms represented in these clades associate in various ways with animals, including the spirochetal zoonotic pathogen *Leptospira interrogans* and the Actinobacteria *Brevibacterium linens* and *Thermobifida fusca*, which that cause body odor and farmers’ lung, respectively. All other microbes represented in these clades are environmentally associated, however.

A model of early cellular diversification of NAB pathways (Fig. S3), including those for Sias, is supported by multiple lines of evidence, including phylogenetic and protein domain comparisons (Fig. 3A and Fig. S2), the wide distribution of NAB pathways within members of Bacteria and the considerable diversity of their predicted sugar structures (Figs. 2 and 3B), and the diverse composition of taxa in the animal-like clades (Table S1). The presence of archaeal NAB sequences in several distinct clades of the phylogenetic tree (Fig. 3A) also supports the conclusion that paralogous gene duplications with divergence of

Table 1. Phylogenetic prediction of sialic acid synthesis in bacteria

	Organism	NCBI class*	Host/environment	Disease	NuLO known
a	<i>S. agalactiae</i>	4	Human	MG, SP, PN	Neu
	<i>S. suis</i>	4	Swine, human	MG, SP, endocarditis	Neu
	<i>E. coli K1</i>	3	Human, avian	MG, SP, GE, cystitis	Neu
	<i>R. gnavus</i>	4	Human	NK, gut symbiont	NK
	<i>F. nucleatum</i>	5	Human, animal	Periodontal diseases	NK
c	<i>C. jejuni</i>	1	Chicken, human	GE, autoimmunity	Neu
	<i>N. meningitidis</i>	2	Human	MG, SP	Neu
	<i>B. cereus G9241</i>	4	Soil, animal	Anthrax-like PN	NK
	<i>H. acinonychis</i>	1	Feline	Gastritis in cheetah	NK
	<i>F. johnsoniae</i>	3	Soil, aquatic	NK	NK
	<i>F. psychrophilum</i>	3	Aquatic, fish	Fry syndrome, fish	NK
	<i>Algoriphagus</i> sp. PR1	3	Aquatic	NK	NK
	<i>Flavobacterium</i> sp. MED217	3	Seawater	NK	NK
	<i>I. ioihiensis</i>	4	NK, hydrothermal vent	NK	NK

MG; meningitis. SP, septicemia. PN, pneumonia. GE, gastroenteritis. NK; none known. Here "a" and "c" refer to phylogenetic clades shown in Fig. 3A.

*NCBI classifications are as follows: 1, ϵ -Proteobacteria; 2, β -Proteobacteria; 3, Bacteroidetes; 4, Firmicutes; 5, Fusobacteria.

enzymatic function occurred very early in cellular evolution. We conclude that NAB enzymes, which at first glance appear to be animal-like, should be more accurately considered present-day components of an ancient pathway that was universally adopted by deuterostome animals. The distant sequence relationship between animal and animal-like clades suggests that Sia biosynthesis did not arise by lateral transfer into the animal lineage, but rather that Sia synthesis was likely inherited in the traditional sense, accompanied by multiple gene losses along other eukaryotic lineages (27, 28).

Chemical Validation of the Phylogenomic/Phyloglycomic Approach.

To provide functional validation of our phylogenomic findings, we present 2 striking examples that illustrate the utility of a phylogenetic approach for predicting chemical structure and inferring the evolutionary history of this class of monosaccharides.

Photobacterium profundum Strains Encode Phylogenetically Distinct NAB Enzymes and Biochemically Distinct NuLO Sugars.

The phylogenetic data predict that 2 otherwise closely related strains of *Photobacterium profundum* (3TCK and SS9) encode biosynthetic pathways for distinct NuLO sugar structures (Fig. 3). We tested this prediction using a well-validated approach for Sia structure identification. Leg and Pse acid standards were isolated from purified LPS preparations (29, 30) and, after fluorescent derivatization with 1,2-diamino-4,5-methylene dioxybenzene (DMB) (26), they eluted at distinct HPLC retention times (Fig. 4A and B). Tandem mass spectrometry confirmed the expected masses of these eluted Leg and Pse standards, showing that the DMB-HPLC approach can be effectively applied to the broader class of NuLO sugars. In another experiment, *P. profundum* SS9 and 3TCK strains were grown under optimized conditions, and mild acid hydrolysis was used to release NuLO sugars. NuLO sugars isolated from *P. profundum* genome strains 3TCK and SS9 exhibited the expected retention times and masses of Leg and Pse standards based on the phylogenetic prediction (Fig. 4C and D).

***Methanobrevibacter smithii*: A Case of Mistaken Identity.** The published genome sequence of the principal human gut-associated methanogen, *Methanobrevibacter smithii*, contains a gene cluster originally annotated to encode a Sia (Neu) biosynthesis pathway. This archaeon has a prominent capsule when grown in vitro, and previous biochemical assays plus lectin-binding immunohisto-

chemical studies suggested that NuLO sugars, presumably Neu5Ac, are expressed by the cultured-type strain (7). Further analyses using custom *M. smithii* GeneChips (see *Materials and Methods*) revealed that this gene cluster was present in 7/7 *M. smithii* isolates examined, and that expression of these genes is differentially regulated in vitro (Fig. S4).

Here we show that the amino acid sequence of *M. smithii* NAB-2 clusters is within clade "d," and thus is predicted to synthesize Pse rather than Neu acids (Fig. 3A). To test this hypothesis, we first compared the retention times and masses of DMB-derivatized α -keto acid standards Neu and Pse using

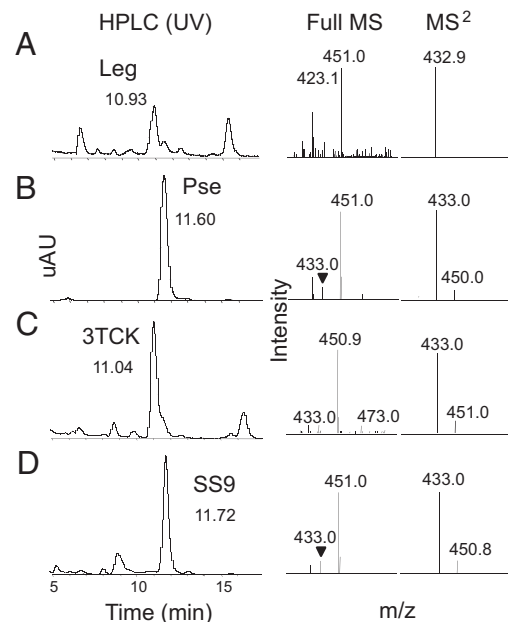


Fig. 4. Targeted chemical validation of *Photobacterium profundum* NAB pathways. LCMS analysis of NuLO sugars isolated from purified LPS of Leg (A) or Pse (B) or from cultures of *P. profundum* genome strain (C) 3TCK or SS9 (D). Different HPLC retention times enable differentiation between Leg and Pse and reliable identification of NuLO in 3TCK and SS9. MS and MS² analyses provide additional confirmation of the expected mass for Leg and Pse for both standards and unknowns. In all cases, MS data are shown from 400–500 *m/z*.

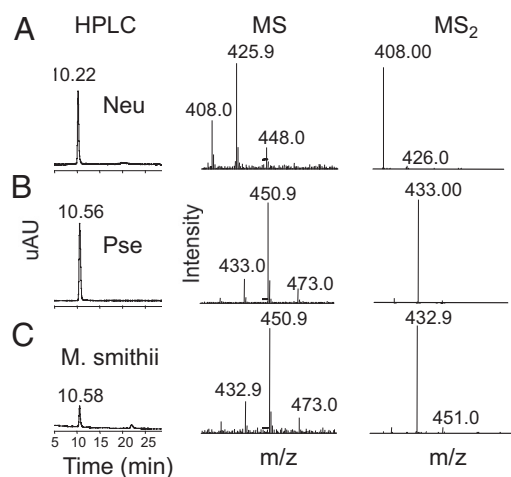


Fig. 5. The human gut methanogen *M. smithii* synthesizes pseudaminic, but not *N*-acetylneuraminic acid. LCMS analysis of NuLO sugars *N*-acetylneuraminic acid (Neu5Ac) (A), Pse from *Pseudoalteromonas atlantica* LPS (B), or isolated from cell pellets of *M. smithii* (C), as described in *Materials and Methods*. In all cases, MS data are shown from 400–500 *m/z*.

reverse-phase HPLC with tandem mass spectrometry, and found that Neu and Pse have similar but distinct retention times and can be discriminated on the basis of mass-to-charge (*m/z*) ratio (Fig. 5*A* and *B*). Parallel preparation and analysis of *M. smithii* NuLO sugars in parallel with Neu and Pse standards clearly demonstrated a retention time and mass consistent with Pse expression (Fig. 5*C*). This finding again validates the phylogenetic approach to predicting NuLO structure, and emphasizes that available methods for Sia detection can be successfully expanded to consider the larger family of NuLO sugars.

Prospectus. In summary, our results provide insight into the evolutionary history of Sias by considering them in the larger phylogenetic context of related NuLO sugars. We emphasize that the surprisingly wide distribution of NAB pathways among the 3 domains of life (Bacteria, Archaea, and Eukarya) is a reflection of many interwoven evolutionary processes, including gene duplications with functional divergence, gene loss, lateral gene transfer, and more specific adaptations of biosynthetic pathways. Clearly, much remains to be done to understand the biology and evolution of these remarkably common and diverse carbohydrate molecules found at the surfaces of contact between many bacteria and their external environments. These findings serve as a proof of principle for the utility of a phylogenomic/phyloglycomic approach to predicting NuLO sugar types and strongly suggest that the expression of Sias and Sia-like sugars by bacteria may be advantageous in a wide range of animal body habitats. Determining the advantages of Sia mimicry in these different host contexts, as well as identifying those factors contributing to lateral dissemination of Sia gene cassettes (i.e., functional clusters) and their incorporation into various LPS or capsular polysaccharide biosynthetic pathways, are important areas for further investigation. These studies should provide a path for future investigation concerning the contributions of Sias and related sugars to the survival and persistence of microbes in both host and environmental reservoirs, as well as in disease pathogenesis.

Materials and Methods

Identification of Physically Clustered NAB Genes. BLASTp, on the “BLAST with microbial genomes” webpage at the National Center for Biotechnology Institute (NCBI) website (www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), was used to query 960 complete microbial genomes. Multiple NAB-1 and NAB-2

amino acid sequences were used for genomic query, with an emphasis on NAB enzymes of defined function or those encoded in organisms known to express specific NuLO structures. Incomplete genomes also were queried using the nonredundant protein database and were included in the data set if deemed NAB-positive. Accession numbers for homologous NAB sequences were cataloged according to the NCBI taxonomic classification and examined for “functional clustering” (20) of NAB-1 and NAB-2 enzymes, as judged by proximal accession numbers. NAB-1 homologs were validated by phylogenetic analysis, and “contaminating” CMP-Kdo synthetases were removed; these distant NAB-1 homologs were not found in functional clusters with NAB-2 homologs. Note that Kdo is an 8-carbon α -keto acid that follows a similar biosynthetic pathway involving condensation of a 5-carbon sugar with phosphoenolpyruvate, followed by activation of CMP-Kdo (1). Results of the genomic profiling were compiled, and the proportion of genomes with one or more physical clusters of NAB-1 and NAB-2 genes was expressed as a function of total genomes surveyed in the different microbial taxa (Fig. 2).

NAB-2 Phylogeny for Prediction of NuLO Sugar Types. Amino acid sequence comparisons indicated that NAB-2 sequences are better conserved than other NAB enzymes and form the most conclusive basis for prediction of specific NuLO sugars in different organisms. NAB-2 amino acid sequences were collected using BLASTp at the NCBI nonredundant protein database and aligned using ClustalQt (Fig. S5) (31). Nexus files from the alignment were uploaded into PAUP* 4.0b10 (32) for exclusion of gaps and domains not found in all sequences, followed by construction of a neighbor-joining tree using the bootstrap/jackknife option with 1,000 replicates. Less-conserved sequences were apparent from visual inspection of the alignment and clustering of branches in a star shape at the base of the phylogenetic tree. Such branches were “pruned” from the tree in successive analyses to reveal significant monophyletic clades and improve aspects of the alignment that could better resolve sequence relationships between different clades. Analysis of the same alignment based on parsimony produced a nearly identical phylogenetic pattern as that from the distance-based approach. NuLO types were predicted from extrapolations of published biochemical data (1, 12–16, 22–24, 33–35) to other members of clades supported by high bootstrap values (shown in Fig. 3*A*). To determine the distribution of NuLO types among different microbial phylotypes, the percentages of NAB-2 clade affiliations (Fig. 3*A–G*) were calculated and expressed as a function of the number of genomes surveyed in each NCBI-classified taxonomic group (Fig. 3*B*). “Pruned” tree branches representing diverged NAB-2 enzymes were included in the tabulation of NAB-positive genomes as “unknown” NuLO type (shown in black in Fig. 3*B*) if a nearby NAB-1 homolog was identified. Note that many individual genomes encode multiple NAB pathways (e.g., ϵ -Proteobacteria, Bacteroidetes; see Table S1), sometimes leading to a ratio of NAB-2 homologs:genomes sampled of >1 in Fig. 3*B*.

Strains and Culture Conditions. The 2 strains of *Photobacterium profundum* with complete genome sequences are close phylogenetic relatives but are adapted to different aquatic ecosystems (36). *P. profundum* 559 is a piezophilic (pressure-loving) strain that grows optimally under low-temperature, high-pressure conditions. In contrast, *P. profundum* 3TCK is adapted to an aquatic niche much closer to the surface. High-pressure growth of 559 was performed anaerobically at 16 °C in 2216 medium (Difco) supplemented with 20 mM glucose and 100 mM Hepes buffer (pH 7.5) (Sigma). Late-exponential phase cultures were diluted 500-fold into fresh medium and used to fill 4.5-mL polyethylene transfer pipettes (Samco). Transfer pipettes were heat-sealed with a hand-held heat-sealing clamp (Nalge) and incubated at 30 MPa in a stainless steel pressure vessel (37). 3TCK was cultivated in the same media but incubated at room temperature without shaking.

Three strains of *M. smithii* were obtained from the DSMZ culture collection (2374, 2375, and 11975), 4 strains were isolated from a single human fecal sample by selective culturing, and the sequenced-type strain (PS) was obtained from ATCC. *M. smithii* was grown in supplemented MBC medium under anaerobic conditions for 6 days at 37 °C as described previously (7).

GeneChip-Based Studies of *M. smithii*. Genomic DNA was prepared and hybridized to a custom Affymetrix GeneChip containing probesets that recognize 99% of its 1,795 predicted protein coding genes. Similarly, RNA isolated under different in vitro growth conditions was hybridized to GeneChips as described previously (See *SI Text*) (7).

Chemical Analysis of NuLO Acids. *M. smithii* and *P. profundum* were harvested from cultures by centrifugation and washed twice with PBS. NuLO residues were released from cells or purified LPS samples containing Leg (29) or Pse (30) acids by mild acid hydrolysis, and low molecular weight fractions were sub-

jected to derivatization with DMB, followed by HPLC analysis and liquid chromatography–mass spectrometry (LCMS) as described previously (see *SI Text*) (38).

ACKNOWLEDGMENTS. We thank Russell F. Doolittle for many helpful discussions, Sandra Diaz for assisting with mass spectrometry, Doug Bartlett for providing *P. profundum*, Emiley Eloë for performing high-pressure cultivation of SS9, and Henning Seedorf for assisting with the GeneChip hybridizations.

1. Angata T, Varki A (2002) Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chem Rev* 102:439–469.
2. Eylar EH, Doolittle RF, Madoff MA (1962) Sialic acid from blood cells of the lamprey eel. *Nature* 193:1183–1184.
3. Varki A (2007) Glycan-based interactions involving vertebrate sialic acid-recognizing proteins. *Nature* 446:1023–1029.
4. Warren L (1963) The distribution of sialic acids in nature. *Comp Biochem Physiol* 10:153–171.
5. Vimr E, Lichtensteiger C (2002) To sialylate, or not to sialylate: That is the question. *Trends Microbiol* 10:254–257.
6. Kelm S, Schauer R (1997) Sialic acids in molecular and cellular interactions. *Int Rev Cytol* 175:137–240.
7. Samuel BS, et al. (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci USA* 104:10643–10648.
8. Damian RT (1965) Molecular mimicry in biological adaptation. *Science* 147:824.
9. Carlin AF, Lewis AL, Varki A, Nizet V (2007) Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *J Bacteriol* 189:1231–1237.
10. Ram S, et al. (1998) A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*. *J Exp Med* 187:743–752.
11. Carlin AF, et al. (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* 113:3333–3336.
12. Knirel YA, Shashkov AS, Tsvetkov YE, Jansson PE, Zähringer U (2003) 5,7-Diamino-3,5,7,9-tetraoxynon-2-ulosonic acids in bacterial glycopolymers: Chemistry and biochemistry. *Adv Carbohydr Chem Biochem* 58:371–417.
13. Glaze PA, Watson DC, Young NM, Tanner ME (2008) Biosynthesis of CMP-N,N'-diacetyllegionaminic acid from UDP-N,N'-diacetylbaucillosamine in *Legionella pneumophila*. *Biochemistry* 47:3272–3282.
14. McNally DJ, et al. (2007) Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J Biol Chem* 282:14463–14475.
15. Schoenhofen IC, McNally DJ, Brisson JR, Logan SM (2006) Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: Synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. *Glycobiology* 16:8C–14C.
16. McNally DJ, et al. (2006) Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81–176 using a focused metabolomics approach. *J Biol Chem* 281:18489–18498.
17. Schirm M, et al. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol Microbiol* 48:1579–1592.
18. Logan SM, et al. (2009) Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics-based approaches. *FEBS J* 276:1014–1023.
19. Twine SM, et al. (2008) Flagellar glycosylation in *Clostridium botulinum*. *FEBS J* 275:4428–4444.
20. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 96:2896–2901.
21. Bonner CA, et al. (2008) Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities. *Microbiol Mol Biol Rev* 72:13–53.
22. Chou WK, Dick S, Wakarchuk WW, Tanner ME (2005) Identification and characterization of NeuB3 from *Campylobacter jejuni* as a pseudaminic acid synthase. *J Biol Chem* 280:35922–35928.
23. Hao J, Balagurumoorthy P, Sarilla S, Sundaramoorthy M (2005) Cloning, expression, and characterization of sialic acid synthases. *Biochem Biophys Res Commun* 338:1507–1514.
24. Suryanti V, Nelson A, Berry A (2003) Cloning, over-expression, purification, and characterization of N-acetylneuraminic acid synthase from *Streptococcus agalactiae*. *Protein Expr Purif* 27:346–356.
25. Daines DA, Wright LF, Chaffin DO, Rubens CE, Silver RP (2000) NeuD plays a role in the synthesis of sialic acid in *Escherichia coli* K1. *FEMS Microbiol Lett* 189(2):281–284.
26. Lewis AL, Hensler ME, Varki A, Nizet V (2006) The group B streptococcal sialic acid O-acetyltransferase is encoded by neuD, a conserved component of bacterial sialic acid biosynthetic gene clusters. *J Biol Chem* 281:11186–11192.
27. Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292:1903–1906.
28. Doolittle RF (2002) Gene transfers between distantly related organisms. *Horizontal Gene Transfer* eds Syvanen M, Kado C (Academic, New York), 2nd Ed, pp 269–275.
29. Knirel YA, Rietschel ET, Marre R, Zähringer U (1994) The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. *Eur J Biochem* 221:239–245.
30. Perepelov AV, et al. (2005) Structure of an acidic polysaccharide from the agar-decomposing marine bacterium *Pseudoalteromonas atlantica* strain IAM 14165 containing 5,7-diacetamido-3,5,7,9-tetraoxo-L-glycero-L-manno-non-2-ulosonic acid. *Carbohydr Res* 340:69–74.
31. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinform* Unit 2.3.
32. Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods) (Sinauer Associates, Sunderland, MA), Version 4.
33. Sundaram AK, et al. (2004) Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *Biochem J* 383(Pt 1):83–89.
34. Gunawan J, et al. (2005) Structural and mechanistic analysis of sialic acid synthase NeuB from *Neisseria meningitidis* in complex with Mn²⁺, phosphoenolpyruvate, and N-acetylmannosaminol. *J Biol Chem* 280:3555–3563.
35. Shashkov AS, et al. (2007) Structure of the O-antigen of *Providencia stuartii* O20, a new polysaccharide containing 5,7-diacetamido-3,5,7,9-tetraoxo-L-glycero-D-galactono-2-ulosonic acid. *Carbohydr Res* 342:653–658.
36. Campanaro S, et al. (2005) Laterally transferred elements and high pressure adaptation in *Photobacterium profundum* strains. *BMC Genomics* 6:122.
37. Eloë EA, Lauro FM, Vogel RF, Bartlett DH (2008) The deep-sea bacterium *Photobacterium profundum* SS9 utilizes separate flagellar systems for swimming and swarming under high-pressure conditions. *Appl Environ Microbiol* 74:6298–6305.
38. Lewis AL, Nizet V, Varki A (2004) Discovery and characterization of sialic acid O-acetylation in group B *Streptococcus*. *Proc Natl Acad Sci USA* 101:11123–11128.