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]

[a]Glycobiology Research and Training Center, Departments of [b]Pediatrics, [g]Medicine, and [e]Cellular and Molecular Medicine, School of Medicine, and [f]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093; [d]N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt 47, 11991 Moscow, Russia; and [c]Center 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-tetradeoxynon-2-ulosonic acids. Examination of ≈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-tetradeoxynon-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.

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

[2]Current 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.

 www.pnas.org/cgi/doi/10.1073/pnas.0902431106

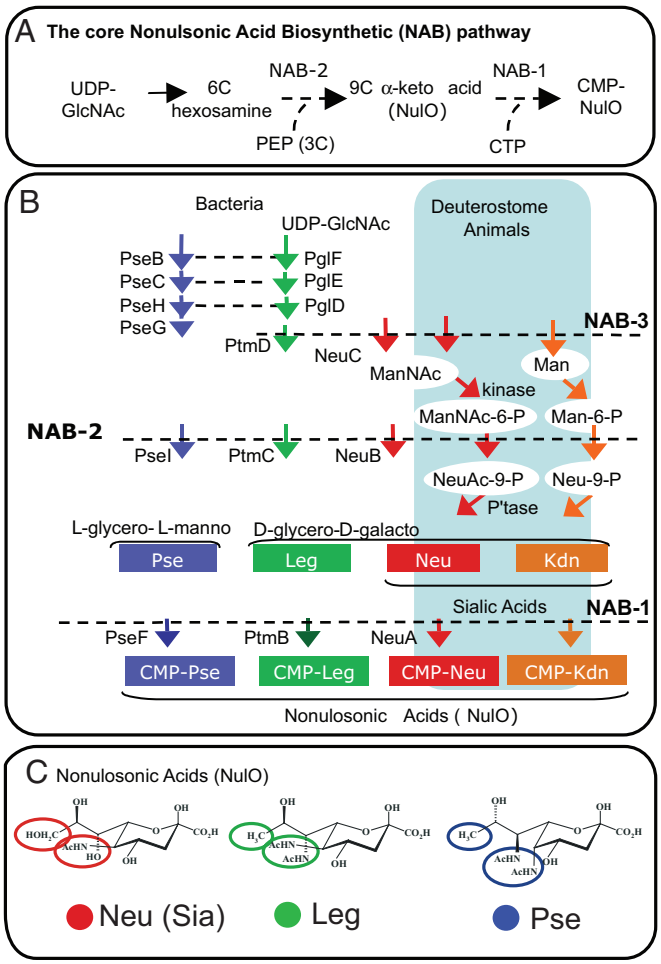

Fig. 1. Related NAB pathways synthesize chemically related sugars. NulOs include all 9-carbon backbone α-keto acid sugars. NulO sugars described to date conform to one of several core backbones that can be further modified by epimerization or modification (1, 12). Major core backbones include the Neu, 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.

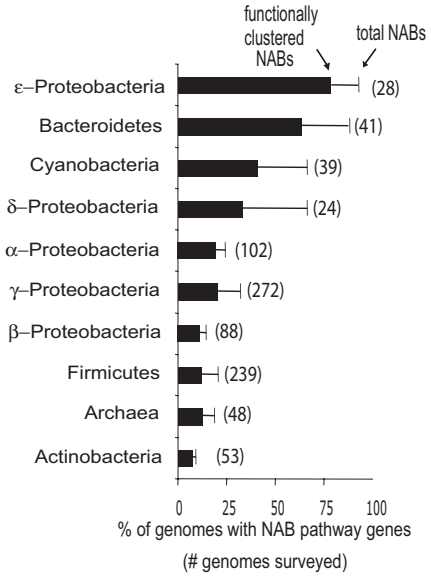

Fig. 2. Distribution of predicted NAB pathways among bacteria and archaea. NAB enzymes NAB-1 and NAB-2 were identified by BLASTp in the genomes of various bacterial phyla and in sequenced archaeons. The number of genomes in each group is given in parentheses. Bars reflect the percentage of genomes in 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. 3*A*; 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.

MICROBIOLOGY

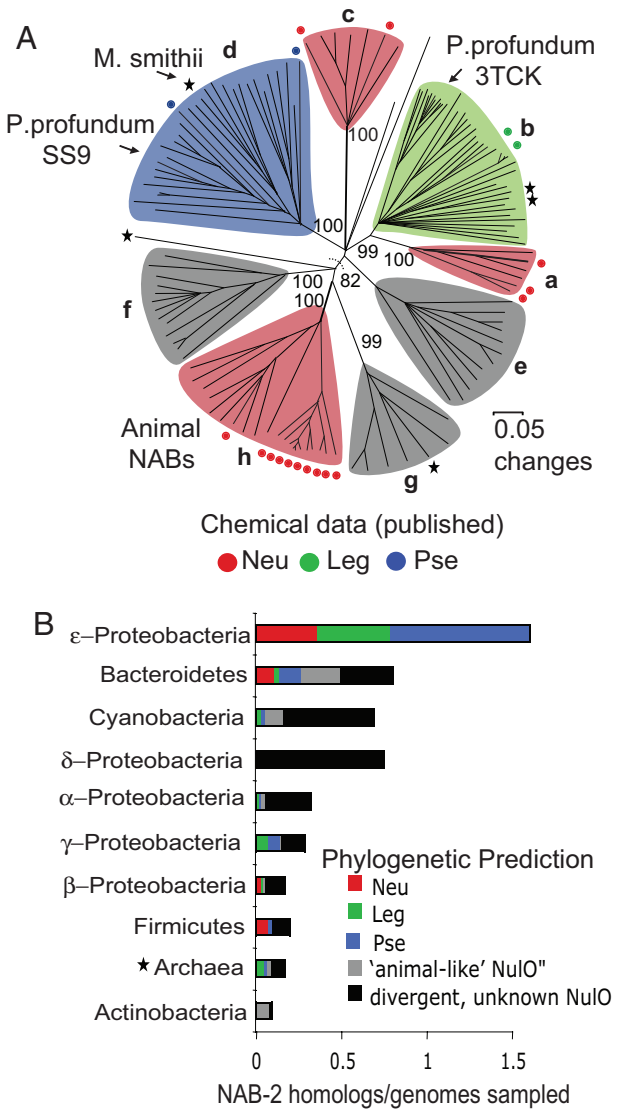

Fig. 3. NAB-2 phylogeny for predicting NulO structure. (*A*), A distance-based neighbor-joining tree was constructed as described in *Materials and Methods*, and published biochemical data were overlaid onto the tree (colored circles). A ''cohesion group'' approach (21) was used to infer enzyme specificities 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. 3*A*). Specifically, the analysis revealed 2 groups of bacterial Neu synthases, represented by clades ''a'' and ''c'' (Fig. 3*A*), 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. 3*A* 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. 3*A* 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 3*B*), 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. 3*A*) 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	*S.agalactiae*	4	Human	MG, SP, PN	Neu
	S. suis	4	Swine, human	MG, SP, endocarditis	Neu
	E. coli K1	3	Human, avian	MG, SP, GE, cystitis	Neu
	R. gnavus	4	Human	NK, gut symbiont	NK
	F. nucleatum	5	Human, animal	Periodontal diseases	NK
c	*C. jejuni*	1	Chicken, human	GE, autoimmunity	Neu
	N. meningitidis	2	Human	MG, SP	Neu
	B. cereus G9241	4	Soil, animal	Anthrax-like PN	NK
	H. acinonychis	1	Feline	Gastritis in cheetah	NK
	F. johnsoniae	3	Soil, aquatic	NK	NK
	F. psychrophilum	3	Aquatic, fish	Fry syndrome, fish	NK
	Algoriphagus sp. *PR1*	3	Aquatic	NK	NK
	Flavobacterium sp. *MED217*	3	Seawater	NK	NK
	I. ioihiensis	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. 3*A*.
*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. 4*A* 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. 4*C* 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 immunohistochemical 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. 3*A*). To test this hypothesis, we first compared the retention times and masses of DMB-derivatized α-keto acid standards Neu and Pse using

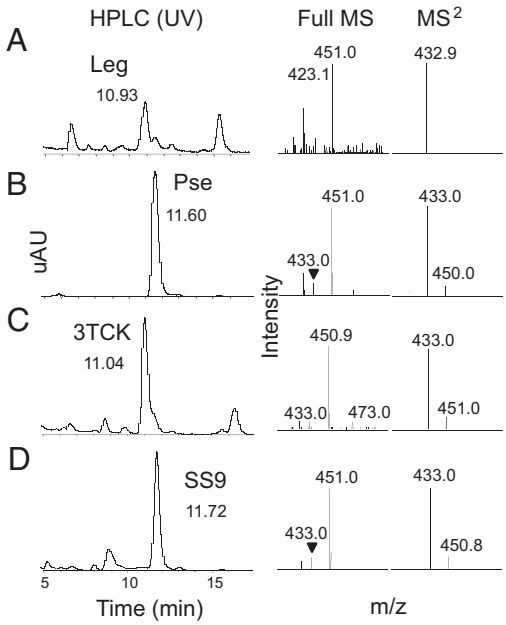

Fig. 4. Targeted chemical validation of *Photobacterium profundum* NAB pathways. LCMS analysis of NulO sugars isolated from purified LPS of Leg (*A*) or Pse (*B*) or from cultures of *P. profundum* genome strain (*C*) 3TCK or SS9 (*D*). Different HPLC retention times enable differentiation between Leg and Pse and reliable identification of NulO in 3TCK and SS9. MS and MS^2 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*.

MICROBIOLOGY

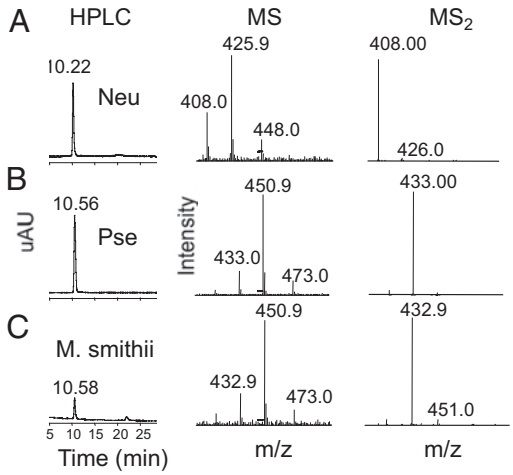

Fig. 5. The human gut methanogen *M. smithii* synthesizes pseudaminic, but not *N*-acetylneuraminic, acid. LCMS analysis of NulO sugars *N*-acetylneuraminic acid (Neu5Ac) (*A*), Pse from *Pseudoalteromonas atlantica* LPS (*B*), or isolated from cell pellets of *M. smithii* (*C*), as described in *Materials* 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 to 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* SS9 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 SS9 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 Eloe for performing high-pressure cultivation of SS9, and Henning Seedorf for assisting with the GeneChip hybridizations.

We appreciate the efforts of Olga Zagnitko in updating annotations of NAB pathways in The SEED, an annotation/analysis tool provided by the Fellowship for Interpretation of Genomes, as well as insightful comments from Aaron Carlin, Yung-Chi Chang, and Shannon Weiman. This work was funded by generous support from the University of California President's Postdoctoral Fellowship Program (to A.L.L.) and the Gianinni Family Foundation (to A.L.L.), along with National Institutes of Health grants HL057345 (to A.V.), DK30292 (to J.I.G.), and R01-HD051796 (to V.N.).

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-tetradeoxynon-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′-diacetylbacillosamine 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 characterisation of N-acetylneuraminate 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-tetradeoxy-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 Mn^{2+}, phosphoenolpyruvate, and N-acetylmannosaminitol. *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-tetradeoxy-l-*glycero*-d-*galacto*-non-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. Eloe 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.

Supporting Information

Lewis et al. 10.1073/pnas.0902431106

SI Materials and Methods

GeneChip-Based Studies of *M. smithii*. This custom GeneChip was also used for whole genome transcriptional profiling of the type strain. Cells were grown at 37 °C, with or without agitation (100 rpm), in 125-mL serum bottles containing 15 mL of supplemented MBC medium (7) under an atmosphere of H_2 and CO_2 (4:1) that was replenished every 6 h, and harvested during the log or stationary phase (log phase: OD_{600} of 1.10 and 0.36 for agitated and static cultures, respectively; stationary phase: OD_{600} of 3.14 and 0.57, respectively). RNA was isolated, and cDNAs were prepared and then hybridized to GeneChips as described previously (7) (n = 9–13 GeneChips/condition). GeneChip-wide normalization (to an intensity of 500) was carried out with an Affymetrix MAS5. The significance of observed differences in gene expression was determined using a 2-tailed Student t test.

Isolation and Characterization of NulOs. After resuspension in 2N acetic acid, cell suspensions were incubated at 80 °C for 3 h to release cell surface NulO residues. Insoluble cell debris was pelleted at maximum speed on a tabletop centrifuge, and material released into the soluble fraction was passed over a 10-K molecular weight cutoff filtration unit (Centricon). Purified LPS samples containing Leg (25) or Pse (26) acids were processed similarly by mild acid hydrolysis and filtration. Low molecular weight fractions or commercially available Neu were derivatized with DMB, and NulO–DMB adducts were resolved by HPLC using a reverse-phase C18 column (Varian) eluted isocratically at a rate of 0.9 mL/min over 50 min using 85% MQ water, 7% methanol, and 8% acetonitrile. DMB-derivatized extracts or individually isolated HPLC peaks were analyzed by LCMS using a Finnigan-MAT HPLC system with a tandem LCQ mass spectrometer (46). Detection of fluorescently labeled NulO sugars was achieved at excitation and emission wavelengths of 373 nm and 448 nm, respectively.

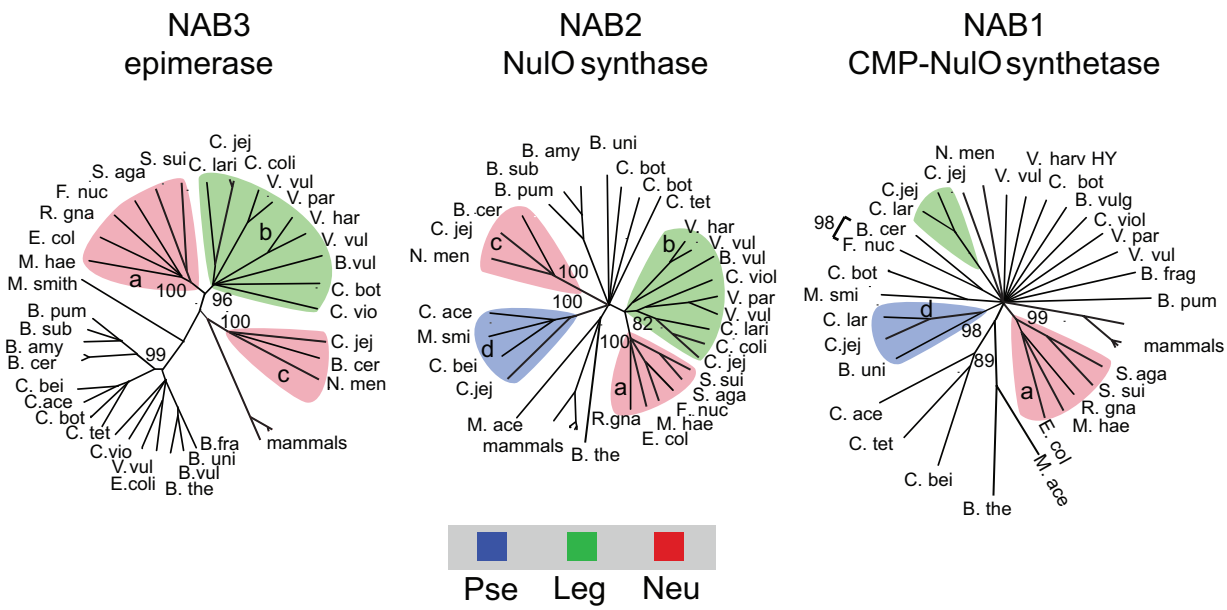

Fig. S1. Phylogenies of NAB pathways reveal distinct bacterial innovations of Sia mimicry. Phylogenetic trees were constructed based on NAB-1, NAB-2, and NAB-3 amino acid sequences collected from pathogenic bacteria represented in clades ''a'' and ''c,'' as well as a defined subset of organisms representative of other phylogenetic clades shown in Fig. 3. Colored branches indicate published biochemical data for specific NulO residues, as shown in the color key. Shading reflects monophyletic clades with high bootstrap support.

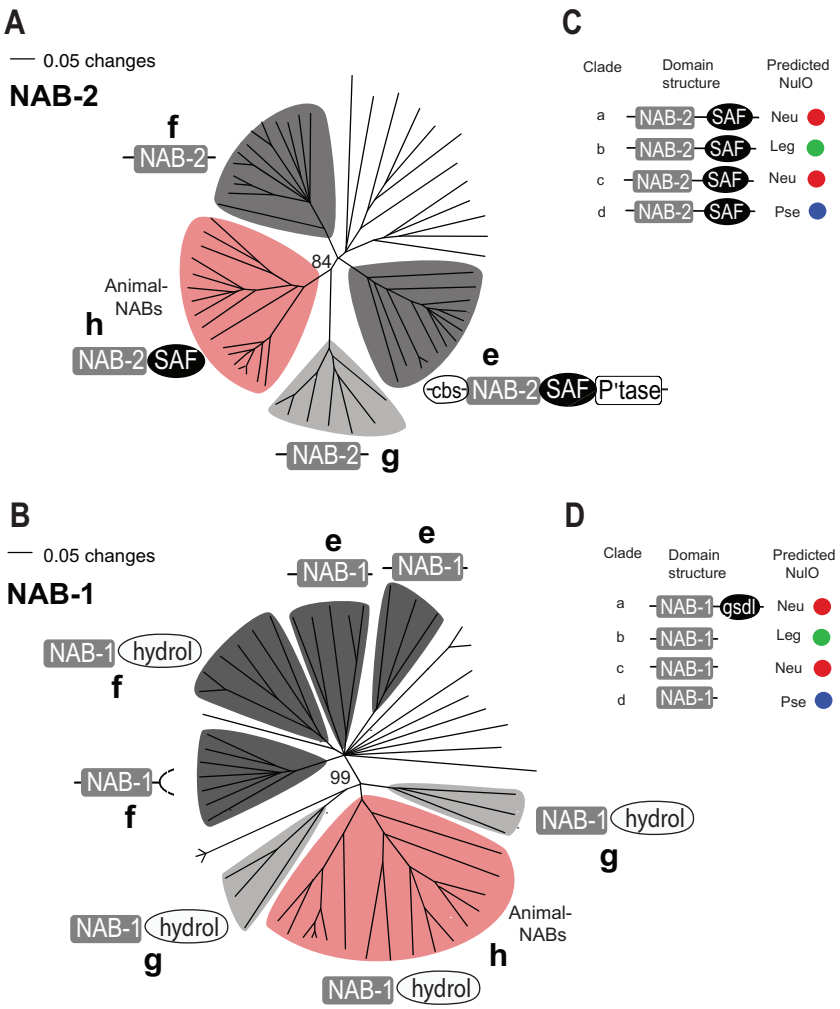

Fig. S2. Phylogenetic analysis of animal and animal-like NAB-1 and NAB-2 amino acid sequences are consistent with domain structure. (*A* and *B*), NAB-1 (*A*) and NAB-2 (*B*) amino acid sequences from organisms represented in Fig. 3, phylogenetic clades ''e''–''h,'' were collected by BLASTp and subjected to a phylogenetic analysis that included only the NAB-1 or NAB-2 domains common to all sequences in the alignment. Protein domain organizations for clades based on the Pfam database (47), and our amino acid alignments were overlaid onto the tree. Red shading highlights the animal taxa (clade ''h'') that express Sias on their cell surfaces. All other clades (shown in gray) reflect novel phylogenetic classes of animal-like NAB enzymes for which no biochemical data currently exist. Clade ''g'' is shown in a lighter gray than other animal-like NABs to emphasize its closer phylogenetic relationship with NABs from animals as indicated by bootstrap values (shown). As in other analyses, NAB-1 appears less well conserved but shares similar phylogenetic features with functionally clustered NAB-2 sequences from the same organisms. Note that the ancestral architecture of NAB-2 likely included the C-terminal SAF domain, and that most of the animal-like NAB-1 sequences (similar to animal NAB-1 and in contrast to the sequences in clades ''a''–''d'') contain a C-terminal domain with predicted hydrolase activity.(*C*) and *D*, Protein domain organization for clades ''a''–''d'' shown for comparison. Pfam numbers for domains are as follows: NAB-1, PF02348; NAB-2, PF03102; SAF, PF08666; CBS, PF00571; P'tase, PF01261; GSDL, PF00657; Hydrol, PF00702.

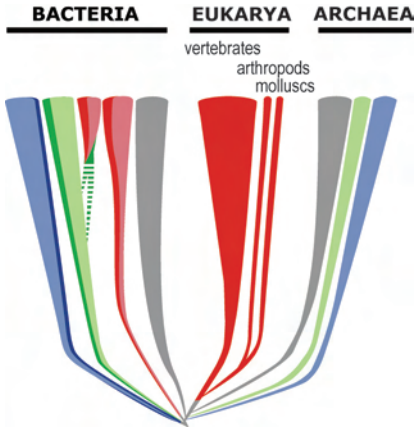

Fig. S3. A model of NulO evolution based on phylogenomic evidence. Based on phylogenetic and genomic evidence, we suggest that an early cellular diversification of NulO sugar structures resulted in the wide variety and distribution of NulO sugars that we find today (darker colors reflect published data; lighter colors indicate phylogenetic predictions). At least 3 distinct semiconvergent evolutionary paths for de novo biosynthesis of Sias are supported by the phylogenetic and biochemical data (i.e., in animals and 2 different groups of microbes often found in close association with Sia-expressing animals).

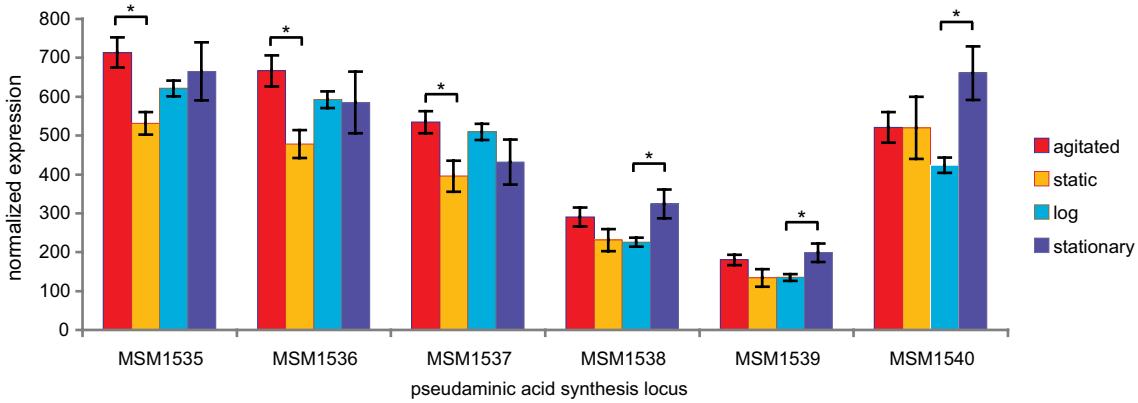

Fig. S4. Expression of predicted *M. smithii* Pse synthesis genes during growth from log phase to stationary phase in standard medium, with or without agitation. Mean values for GeneChip probeset intensities ± SEM are plotted ($n = 9–13$ GeneChips/condition). Asterisks indicate statistically significant differences ($P < .05$ by the Student *t* test). Original annotations of these genes were dTDP-D-glucose 4,6 dehydratase (MSM1535), acylneuraminate cytidylyltransferase (MSM1537), CMP-sialic acid synthetase (MSM1538), and a sialic acid synthase (MSM1539) (7). MSM1536 and MSM1540 encode a pleiotropic regulatory protein DegT and glycerol-3-phosphate dehydrogenase, respectively.

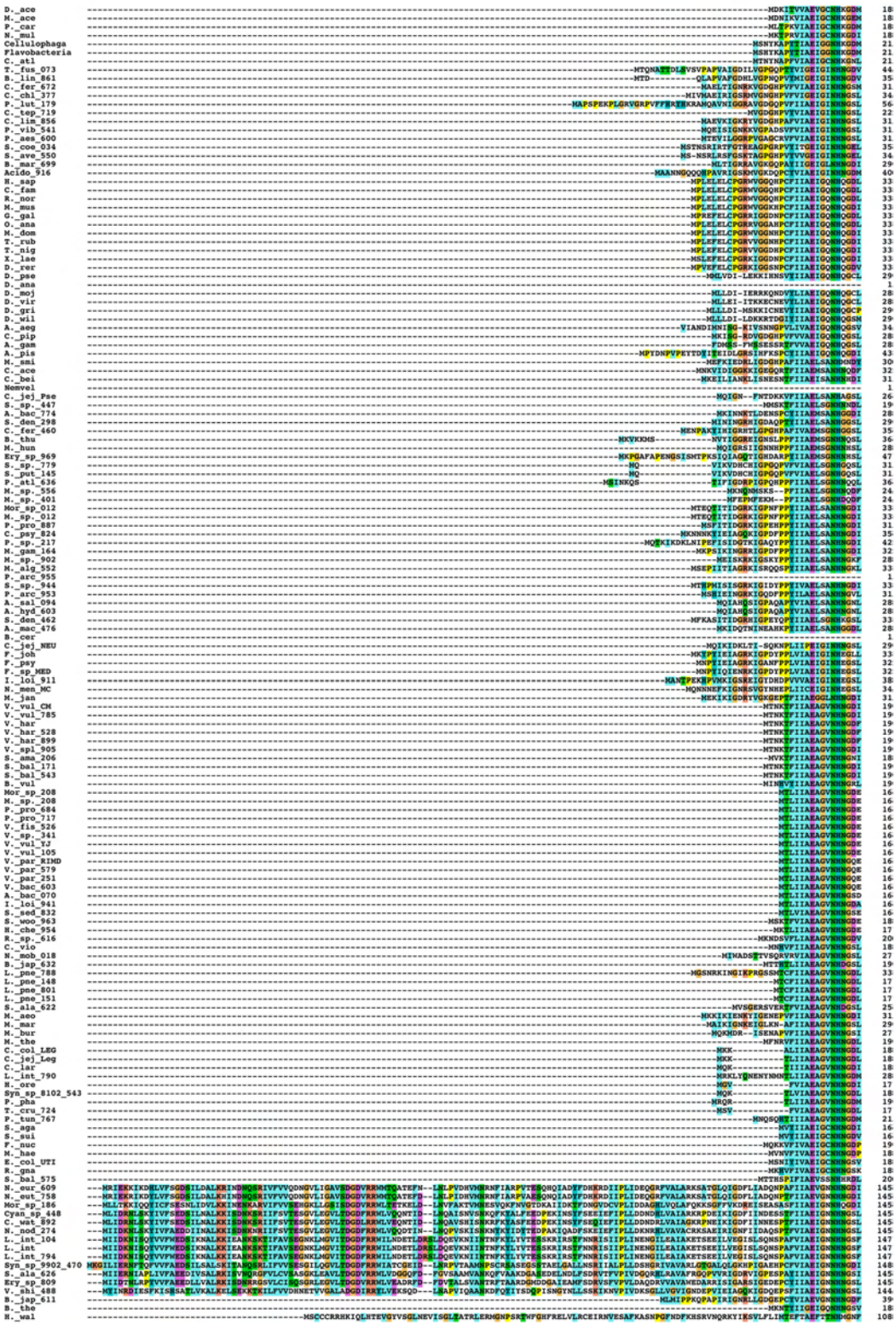

Fig. S5. Multiple sequence alignment of NAB-2 amino acid sequences used for construction of the phylogenetic tree in Fig. 3. Clustal Qt alignment of NAB-2 sequences. With the exception of the more limited sequences from sea anemone (*Nematostella*) and *P. arctica*, all gaps in the alignment were excluded from the phylogenetic analysis.

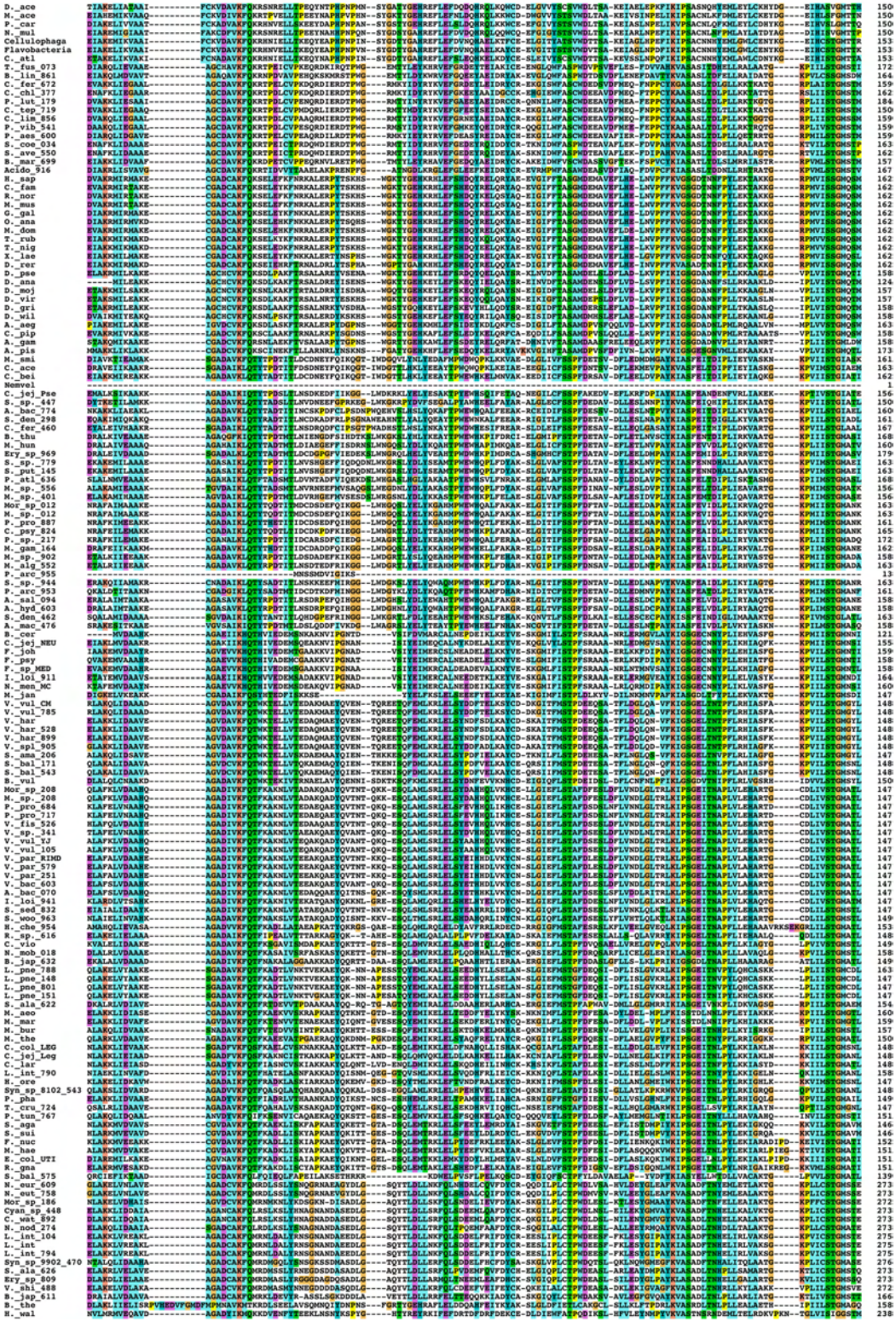

Fig. S5. Continued.

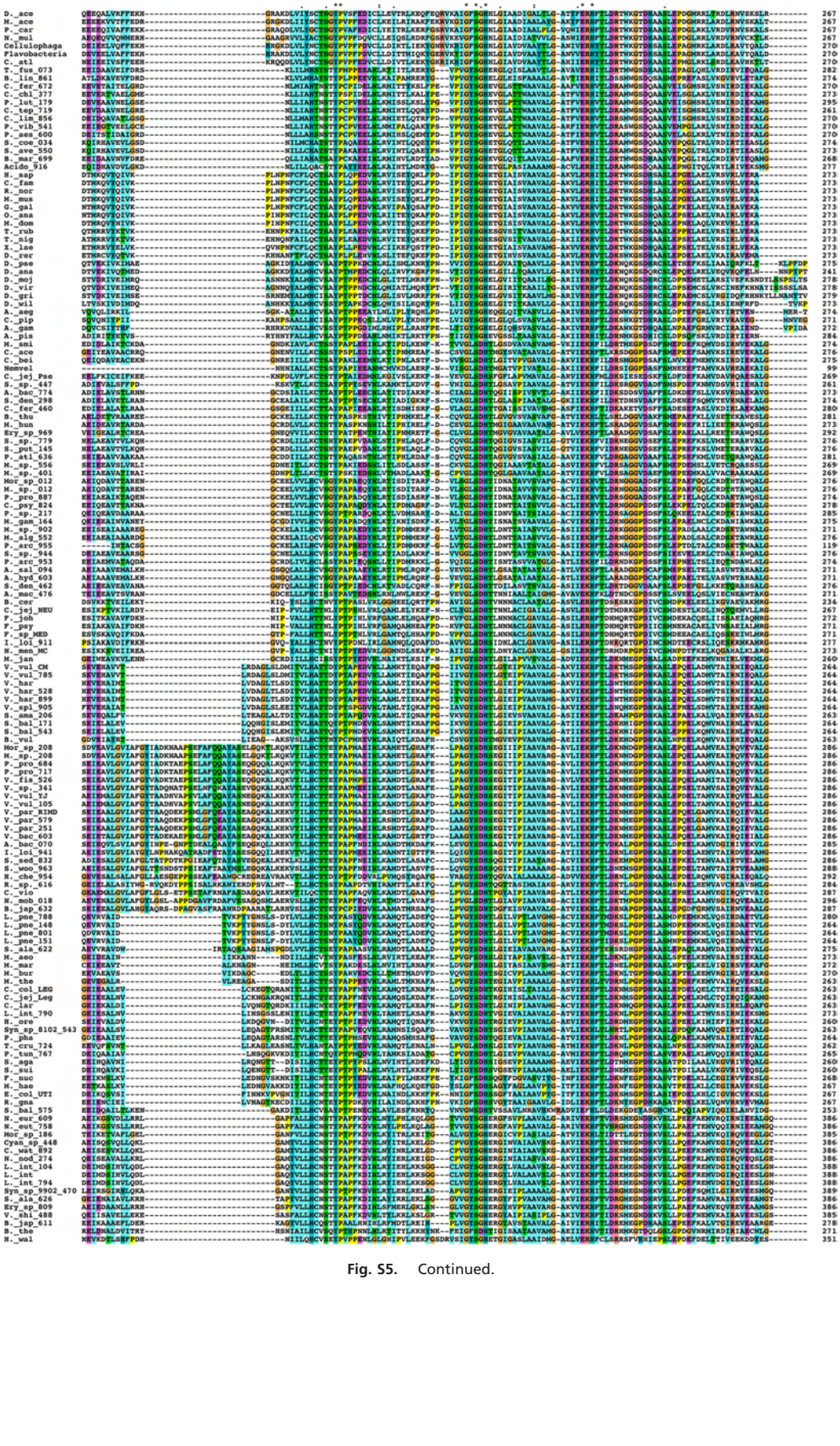

Fig. S5. Continued.

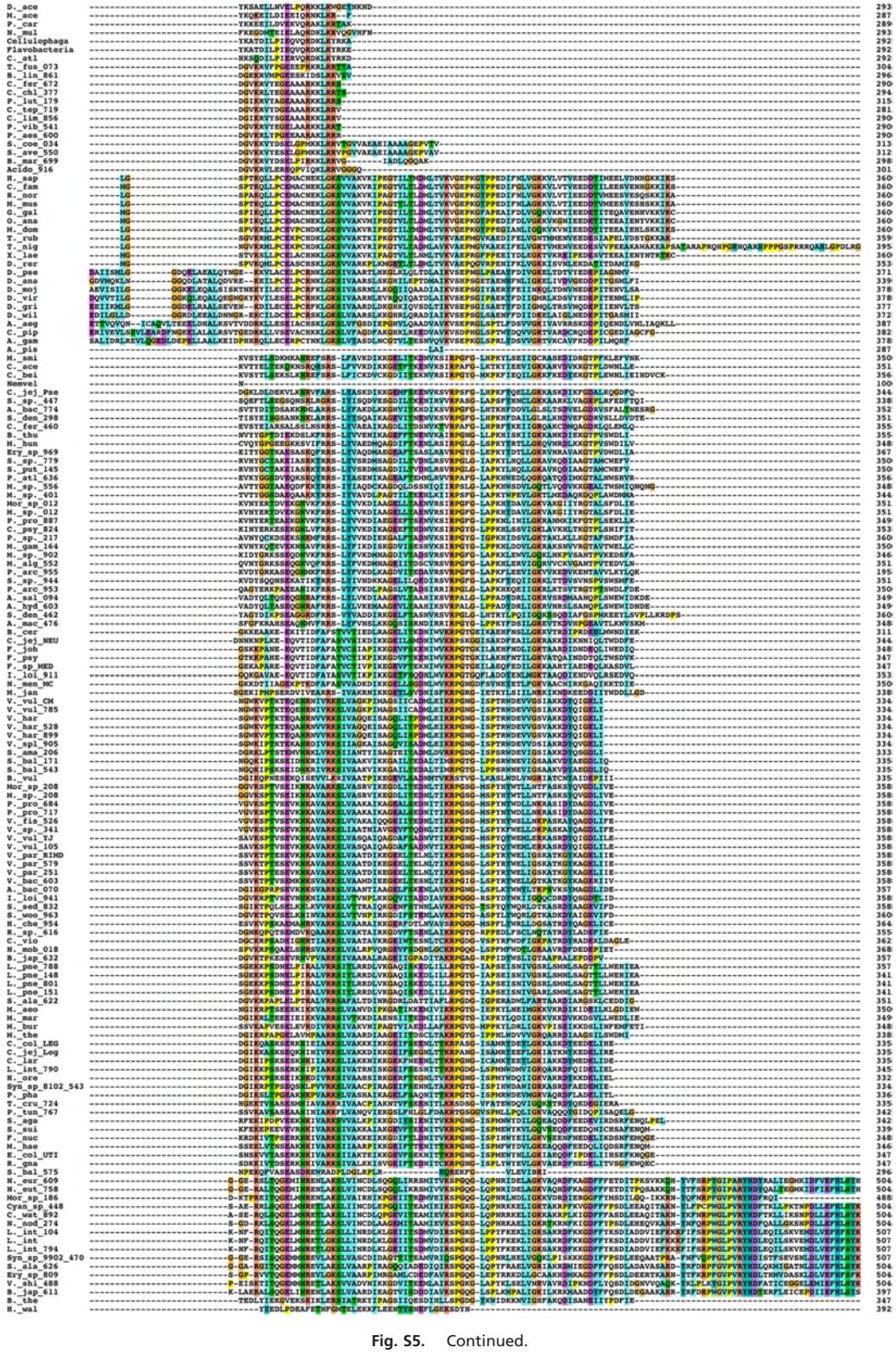

Fig. S5. Continued.

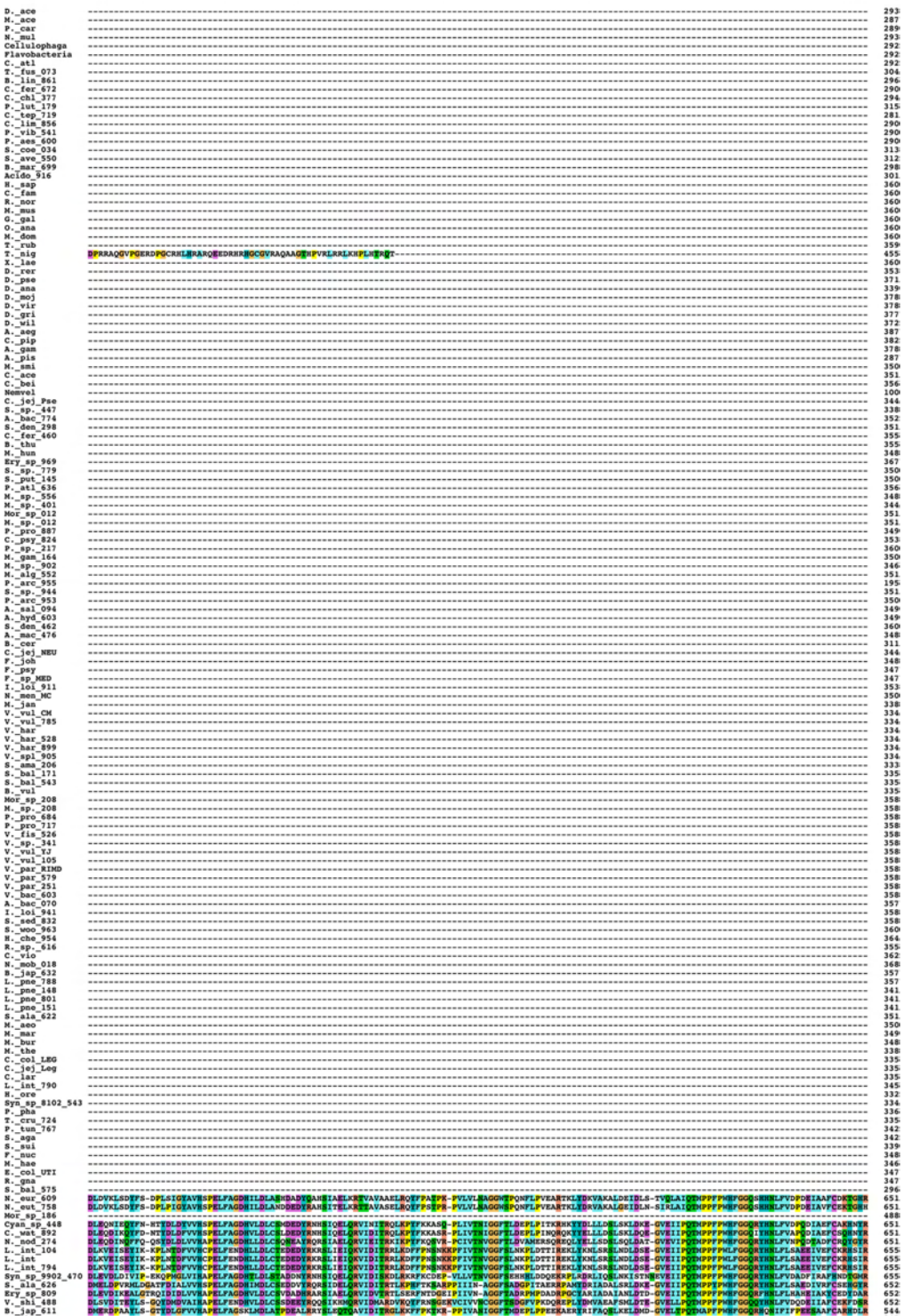

Fig. S5. Continued.

Other Supporting Information

Table S1

Table S1. Distribution of NAB homologs among sequenced bacteria and archaea

NCBI accession numbers for NAB-1 and NAB-2 homologs are color coded according to Fig. 3A predictions, **Neu**; **Leg**; **Pse**; **'animal-like'**, Strains are in alphabetical order with NAB-negative genomes enclosed by double lines below each NCBI class of NAB-positive genomes.

ε-Proteobacteria	NAB-1			NAB-2		
Arcobacter butzleri RM4018	YP_001491128			YP_001491129		
Caminibacter mediatlanticus TB-2	ZP_01871752	ZP_01872215		ZP_01871757	ZP_01872218	
Campylobacter coli RM2228	ZP_00367365	ZP_00367381	ZP_00368095	ZP_00367371	ZP_00367377	
Campylobacter concisus 13826	YP_001467628				YP_001467632	
Campylobacter curvus 525.92	YP_001408965			YP_001408968		
Campylobacter fetus subsp. fetus 82-40	YP_892682			YP_892689		
Campylobacter jejuni RM1221	YP_179491	YP_179505		YP_179497	YP_179501	
Campylobacter jejuni subsp. doylei 269.97	YP_001397604	YP_001397584	YP_001397865	YP_001397596	YP_001397590	
Campylobacter jejuni subsp. jejuni 260.94	ZP_01069032	ZP_01069114		ZP_01068970		ZP_01069267
Campylobacter jejuni subsp. jejuni 81116	YP_001482908			YP_001482827		YP_001482914
Campylobacter jejuni subsp. jejuni 81-176	YP_001000986	ZP_02271650	YP_001000821	YP_001000992	ZP_02271488	YP_001000818
Campylobacter jejuni subsp. jejuni 84-25	ZP_01099325	ZP_01099767	ZP_01099742	ZP_01099311	ZP_01099299	ZP_01099657
Campylobacter jejuni subsp. jejuni CF93-6	ZP_01068249	ZP_01068280	ZP_01068549	ZP_01068260	ZP_01068224	ZP_01068514
Campylobacter jejuni subsp. jejuni CG8486	ZP_01810694	ZP_01810708		ZP_01810700	ZP_01810704	ZP_01809838
Campylobacter jejuni subsp. jejuni HB93-13	ZP_01070901		ZP_01070836	ZP_01070873		ZP_01071011
Campylobacter jejuni subsp. jejuni NCTC 11168	NP_282457	NP_282477	NP_282291	NP_282463	NP_282473	NP_282289
Campylobacter lari RM2100	ZP_00369430	ZP_00369519		ZP_00368650	ZP_00369513	
Campylobacter upsaliensis RM3195	ZP_00371764	ZP_00371495	ZP_00371144	ZP_00370335	ZP_00371499	
Helicobacter acinonychis str. Sheeba	YP_665018	YP_664764		YP_664200		YP_665020
Helicobacter hepaticus ATCC 51449	NP_860431			NP_860439		
Helicobacter pylori 26695	NP_207124			NP_206977		
Helicobacter pylori HPAG1	YP_627070			YP_626916		
Helicobacter pylori J99				NP_222887		
Sulfurimonas denitrificans DSM 1251	YP_393113				YP_393109	
Wolinella succinogenes DSM 1740	NP_908204			NP_908207		
Campylobacter hominis ATCC BAA-381	*Sulfurovum sp. NBC37-1*					
Nitratiruptor sp. SB155-2	*Thiomicrospira denitrificans ATCC 33889*					

Bacteroidetes	NAB-1			NAB-2		
Algoriphagus sp. PR1	ZP_01717386			ZP_01717383		
Bacteroides capillosus ATCC 29799	ZP_02036000					
Bacteroides fragilis NCTC 9343	YP_211444					
Bacteroides fragilis YCH46	YP_101321	YP_099504	YP_099012			
Bacteroides ovatus ATCC 8483	ZP_02066715					
Bacteroides thetaiotaomicron VPI-5482	NP_810626.1			NP_810627		
Bacteroides uniformis ATCC 8492	ZP_02070812	ZP_02071546	ZP_02071964	ZP_02071542	ZP_02070811	
Bacteroides vulgatus ATCC 8482	YP_001301263	YP_001301261	YP_001299934	YP_001301264		
Cellulophaga sp. MED134	ZP_01051503	ZP_01049060		ZP_01051504		
Chlorobium chlorochromatii CaD3	YP_379437	YP_378636				YP_379377

chlorobium ferrooxidans DSM 13031	ZP_01385741	ZP_01385461	ZP_01385231	ZP_01385460		ZP_01385672
Chlorobium limicola DSM 245	ZP_00511412	ZP_00513057				ZP_00512609
Chlorobium phaeobacteroides DSM 266	YP_911960	YP_912515				YP_911647
Chlorobium tepidum TLS	NP_662044	NP_662709				NP_661719
Croceibacter atlanticus HTCC2559	ZP_00950410	ZP_00951263	ZP_00950767		ZP_00950411	
Cytophaga hutchinsonii ATCC 33406	YP_679353				YP_679350	
Flavobacteria bacterium BAL38	ZP_01733186	ZP_01735244		ZP_01733184		
Flavobacteria bacterium BBFL7	ZP_01202896	ZP_01201921	ZP_01203049	ZP_01202899	ZP_01201920	
Flavobacteriales bacterium HTCC2170	ZP_01105687	ZP_01106328		ZP_01105691		
Flavobacterium johnsoniae UW101	YP_001192662	YP_001196831		YP_001192660		
Flavobacterium psychrophilum JIP02/86	YP_001296151	YP_001295888		YP_001296150		
Flavobacterium sp. MED217	ZP_01058929	ZP_01061526		ZP_01058928		
Gramella forsetii KT0803	YP_860622	YP_862055	YP_861860		YP_860619	
Pedobacter sp. BAL39	ZP_01883717				ZP_01883719	
Pelodictyon luteolum DSM 273	YP_374880					YP_375179
Pelodictyon phaeoclathratiforme BU-1	ZP_00590637	ZP_00588167	ZP_00588267	ZP_00588274	ZP_01717383	ZP_00588729
Polaribacter irgensii 23-P	ZP_01118271					
Prosthecochloris aestuarii DSM 271	ZP_00592355	ZP_00590933				ZP_00591854
Prosthecochloris vibrioformis DSM 265	YP_001130443	YP_001129928				YP_001130541
Psychroflexus torquis ATCC 700755	ZP_01254358	ZP_01253608				
Robiginitalea biformata HTCC2501	ZP_01120753					
Salinibacter ruber DSM 13855	YP_444751			YP_444748		
Tenacibaculum sp. MED152	ZP_01052754	ZP_01052336		ZP_01052757		
unidentified eubacterium SCB49	ZP_01889752	ZP_01890888				

Bacteroides caccae ATCC 43185	*Chlorobium phaeobacteroides BS1*	*Parabacteroides merdae ATCC 43184*
Candidatus Sulcia muelleri str. Hc (Homalodisca coagulata)	*Parabacteroides distasonis ATCC 8503*	*Porphyromonas gingivalis W83*

Cyanobacteria	**NAB-1**			**NAB-2**		
Acaryochloris marina MBIC11017	YP_001516521					
Anabaena variabilis ATCC 29413	YP_323517	YP_323463				
Crocosphaera watsonii WH 8501	ZP_00517669			ZP_00514892		
Cyanothece sp. CCY0110-	ZP_01729244			ZP_01731448		
Gloeobacter violaceus PCC 7421	NP_926721					
Lyngbya sp. PCC 8106	ZP_01622329	ZP_01622482				
Nodularia spumigena CCY9414	ZP_01631002	ZP_01631569		ZP_01632274		
Nostoc punctiforme PCC 73102	ZP_00108272	ZP_00110422		ZP_00109033		
Nostoc sp. PCC 7120	NP_485331	NP_485019				
Prochlorococcus marinus str. AS9601	YP_001009733			YP_001009815		
Prochlorococcus marinus str. MIT 9211	YP_001551127	YP_001551172		YP_001551130		
Prochlorococcus marinus str. MIT 9215	YP_001484662	YP_001484645		YP_001484680	YP_001484665	
Prochlorococcus marinus str. MIT 9301	YP_001091637	YP_001091645		YP_001091632		
Prochlorococcus marinus str. MIT 9303	YP_001016135			YP_001016130		

Prochlorococcus marinus str. MIT 9312	YP_397845	YP_397837	YP_397842	YP_397836
Prochlorococcus marinus str. MIT 9313	NP_893935		NP_893939	
Prochlorococcus marinus str. NATL1A	YP_001014697		YP_001014695	
Synechococcus elongatus PCC 7942	YP_401307			
Synechococcus sp. BL107	ZP_01468600	ZP_01468925	ZP_01469027	ZP_01468604
Synechococcus sp. CC9311	YP_729407	YP_729419	YP_729411	YP_729418
Synechococcus sp. CC9605	YP_382511		YP_382515	
Synechococcus sp. CC9902-	YP_376474	YP_376118	YP_376470	YP_376102
Synechococcus sp. RS9916	ZP_01471749		ZP_01471755	
Synechococcus sp. RS9917	ZP_01078981		ZP_01078980	
Synechococcus sp. WH 8102-	NP_896493	NP_896539	NP_896543	NP_896489
Synechocystis sp. PCC 6803	NP_441379		NP_441367	
Trichodesmium erythraeum IMS101	YP_723938		YP_723990	
Leptolyngbya valderiana BDU 20041	*Synechococcus elongatus PCC 6301*		*Synechococcus sp. WH 5701*	
Prochlorococcus marinus str. NATL2A	*Synechococcus sp. JA-2-3B'a(2-13)*		*Synechococcus sp. WH 7803*	
Prochlorococcus marinus ss. marinus CCMP1375	*Synechococcus sp. JA-3-3Ab*		*Synechococcus sp. WH 7805*	
Prochlorococcus marinus ss. pastoris CCMP1986	*Synechococcus sp. RCC307*		*Thermosynechococcus elongatus BP-1*	
δ-Proteobacteria	**NAB-1**		**NAB-2**	
Bdellovibrio bacteriovorus HD100	NP_968563		NP_968560	
delta proteobacterium MLMS-1	ZP_01287684		ZP_01288484	
Desulfovibrio desulfuricans G20	YP_389746		YP_389745	YP_390169
Desulfovibrio vulgaris subsp. vulgaris DP4	YP_968030		YP_968032	YP_968074
Desulfovibrio vulgaris ss. vulgaris Hildenborough			YP_009573	YP_012219
Desulfuromonas acetoxidans DSM 684	ZP_01312106	YP_386919	ZP_01312107	ZP_01311516
Geobacter bemidjiensis Bem			ZP_01773432	
Geobacter lovleyi SZ	YP_001951964			
Geobacter metallireducens GS-15	YP_384236		YP_383423	
Geobacter sulfurreducens PCA	NP_953021		NP_953019	
Geobacter uraniireducens Rf4	YP_001232805		YP_001232802	
Lawsonia intracellularis PHE/MN1-00	YP_595364			YP_965814
Myxococcus xanthus DK 1622	YP_629360			
Pelobacter carbinolicus DSM 2380	YP_356697		YP_356696	YP_356559
Syntrophus aciditrophicus SB	YP_463211			
Anaeromyxobacter dehalogenans 2CP-C	*Desulfotalea psychrophila LSv54*		*Plesiocystis pacifica SIR-1*	
Anaeromyxobacter sp. Fw109-5	*Desulfovibrio vulgaris subsp. vulgaris DP4*		*Stigmatella aurantiaca DW4/3-1*	
Candidatus Desulfococcus oleovorans Hxd3	*Pelobacter propionicus DSM 2379*		*Syntrophobacter fumaroxidans MPOB*	
α-Proteobacteria	**NAB-1**		**NAB-2**	
Bradyrhizobium japonicum USDA 110	NP_772636	NP_772612	NP_772632	NP_772611
Erythrobacter sp. NAP1	ZP_01040808	ZP_01038973	ZP_01040809	ZP_01038969
Fulvimarina pelagi HTCC2506	ZP_01440484		ZP_01440497	
Loktanella vestfoldensis SKA53	ZP_01002614		ZP_01002612	
Magnetospirillum magneticum AMB-1	YP_419451	YP_419454	YP_419443	YP_419455
Magnetospirillum magnetotacticum MS-1	ZP_00056632.2|		ZP_00056630	ZP_00054178
Methylobacterium chloromethanicum CM4	ZP_02059631		ZP_02059639	YP_420076
Nitrobacter sp. Nb-311A	ZP_01045625		ZP_01045622	ZP_01045506
Nitrobacter winogradskyi Nb-255	YP_319003		YP_319002	YP_317265

Oceanicaulis alexandrii HTCC2633	ZP_00952361		ZP_00952353	
Paracoccus denitrificans PD1222	YP_913900	YP_917316	YP_917322	
Rhodopseudomonas palustris BisA53	YP_783144		YP_783141	
Rhodopseudomonas palustris HaA2	YP_485157		YP_485155	
Roseobacter denitrificans OCh 114	YP_684331		YP_684329	
Roseobacter sp. AzwK-3b	ZP_01901532	ZP_01901542	ZP_01901536	ZP_01901525
Roseobacter sp. CCS2	ZP_01750795		ZP_01749990	ZP_01750793
Roseobacter sp. SK209-2-6	ZP_01756746	ZP_01756754	ZP_01756749	ZP_01756734
Sphingopyxis alaskensis RB2256	YP_616624	YP_616627	YP_616622	YP_616626
Stappia aggregata IAM 12614	ZP_01545477	ZP_01550520	ZP_01545475	

Sinorhizobium meliloti 1021	*Methylobacterium extorquens PA1*	*Rickettsia felis URRWXCal2*
Agrobacterium tumefaciens str. C58	*Methylobacterium sp. 4-46*	*Rickettsia massiliae MTU5*
alpha proteobacterium HTCC2255	*Neorickettsia sennetsu str. Miyayama*	*Rickettsia prowazekii str. Madrid E*
Anaplasma marginale str. St. Maries	*Nitrobacter hamburgensis X14*	*Rickettsia rickettsii str. 'Sheila Smith'*
Anaplasma phagocytophilum HZ	*Novosphingobium aromaticivorans DSM 12444*	*Rickettsia sibirica 246*
Bartonella bacilliformis KC583	*Oceanicola batsensis HTCC2597*	*Rickettsia typhi str. Wilmington*
Bartonella henselae str. Houston-1	*Oceanicola granulosus HTCC2516*	*Roseobacter sp. MED193*
Bartonella quintana str. Toulouse	*Ochrobactrum anthropi ATCC 49188*	*Roseovarius nubinhibens ISM*
Brucella abortus biovar 1 str. 9-941	*Orientia tsutsugamushi Boryong*	*Roseovarius sp. 217*
Brucella melitensis 16M	*Parvibaculum lavamentivorans DS-1*	*Roseovarius sp. HTCC2601*
Brucella melitensis biovar Abortus 2308	*Parvularcula bermudensis HTCC2503*	*Roseovarius sp. TM1035*
Brucella ovis ATCC 25840	*Rhizobium etli CFN 42*	*Sagittula stellata E-37*
Brucella suis 1330	*Rhizobium leguminosarum bv. viciae 3841*	*Silicibacter pomeroyi DSS-3*
Candidatus Pelagibacter ubique HTCC1002	*Rhodobacter sphaeroides 2.4.1*	*Silicibacter sp. TM1040*
Candidatus Pelagibacter ubique HTCC1062	*Rhodobacter sphaeroides ATCC 17025*	*Sinorhizobium medicae WSM419*
Caulobacter crescentus CB15	*Rhodobacter sphaeroides ATCC 17029*	*Sphingomonas sp. SKA58*
Caulobacter sp. K31	*Rhodobacterales bacterium HTCC2150*	*Sphingomonas wittichii RW1*
Ehrlichia canis str. Jake	*Rhodobacterales bacterium HTCC2654*	*Sulfitobacter sp. EE-36*
Ehrlichia chaffeensis str. Arkansas	*Rhodopseudomonas palustris BisB18*	*Sulfitobacter sp. NAS-14.1*
Ehrlichia chaffeensis str. Sapulpa	*Rhodopseudomonas palustris BisB5*	*Wolbachia endosymbiont of Drosophila ananassae*
Ehrlichia ruminantium str. Gardel	*Rhodopseudomonas palustris CGA009*	*Wolbachia endosymbiont of Drosophila melanogaster*
Ehrlichia ruminantium str. Welgevonden	*Rhodospirillum rubrum ATCC 11170*	*Wolbachia endosymbiont of Drosophila simulans*
Granulibacter bethesdensis CGDNIH1	*Rickettsia africae ESF-5*	*Wolbachia endosymbiont of Drosophila willistoni TSC#14030-0811.24*
Hyphomonas neptunium ATCC 15444	*Rickettsia akari str. Hartford*	*Wolbachia endosymbiont strain TRS of Brugia malayi*
Jannaschia sp. CCS1	*Rickettsia bellii OSU 85-389*	*Xanthobacter autotrophicus Py2*
Maricaulis maris MCS10	*Rickettsia bellii RML369-C*	*Zymomonas mobilis subsp. mobilis ZM4*
Mesorhizobium loti MAFF303099]	*Rickettsia canadensis str. McKiel*	
Mesorhizobium sp. BNC1	*Rickettsia conorii str. Malish 7*	

γ-Proteobacteria	**NAB-1**	**NAB-2**
Aeromonas hydrophila subsp. hydrophila ATCC 7966	YP_858601	YP_858603
Aeromonas salmonicida subsp. salmonicida A449	YP_001140096	YP_001140094
Alkalilimnicola ehrlichei MLHE-1	YP_743154	YP_743157
Alteromonadales bacterium TW-7	ZP_01612074	ZP_01612070
Alteromonas macleodii 'Deep ecotype'	ZP_01110485	ZP_01110476
Azotobacter vinelandii AvOP	ZP_00416273	
Colwellia psychrerythraea 34H	YP_268821	YP_268824

Escherichia coli APEC O1	YP_854393		YP_854394	
Escherichia coli UTI89	YP_542349		YP_542350	
Francisella tularensis subsp. holarctica			YP_513685	
Haemophilus ducreyi 35000HP	NP_873215			
Haemophilus influenzae 22.1-21	ZP_01785141			
Haemophilus influenzae 22.4-21	ZP_01786684			
Haemophilus influenzae 3655	ZP_01788996			
Haemophilus influenzae 86-028NP	YP_249309			
Haemophilus influenzae PittAA	ZP_01790101			
Haemophilus influenzae PittEE	YP_001290615			
Haemophilus influenzae PittGG	YP_001291778			
Haemophilus influenzae PittHH	ZP_01792523			
Haemophilus influenzae PittII	ZP_01794452			
Haemophilus influenzae R2846	ZP_00154934.2|			
Haemophilus influenzae R2866	ZP_00157363.2|			
Haemophilus influenzae R3021	ZP_01797675	ZP_01797708		
Haemophilus influenzae Rd KW20	NP_439432			
Haemophilus somnus 129PT	YP_718918			
Haemophilus somnus 2336	YP_001784447			
Hahella chejuensis KCTC 2396	YP_435951	YP_436364	YP_435954	YP_436365
Idiomarina loihiensis L2TR	YP_154910	YP_154944	YP_154941	YP_154911
Legionella pneumophila str. Corby	YP_001251802		YP_001251801	YP_001251785
Legionella pneumophila str. Lens	YP_126150		YP_126151	YP_126168
Legionella pneumophila str. Paris	YP_123147		YP_123148	YP_123162
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	YP_094787		YP_094788	YP_094804
marine gamma proteobacterium HTCC2207			ZP_01225164	
Marinobacter algicola DG893			ZP_01894552	
Marinobacter sp. ELB17	ZP_01738899		ZP_01738902	
Marinomonas sp. MED121	ZP_01075554		ZP_01075556	
Marinomonas sp. MWYL1	YP_001342399		YP_001342401	
Moritella sp. PE36	ZP_01898014	ZP_01896204&187	ZP_01896208	ZP_01896186
Nitrococcus mobilis Nb-231	ZP_01127539	ZP_01126021	ZP_01126018	ZP_01126018
Oceanobacter sp. RED65	ZP_01306785		ZP_01306787	
Oceanospirillum sp. MED92	ZP_01167341		ZP_01167812	ZP_01167342
Pasteurella multocida subsp. multocida str. Pm70	NP_245124			
Photobacterium profundum 3TCK	ZP_01218688	ZP_01218721	ZP_01218684	ZP_01218717
Photobacterium profundum SS9	YP_130889		YP_130887	
Pseudoalteromonas atlantica T6c	YP_662638		YP_662636	
Pseudoalteromonas tunicata D2	ZP_01132761		ZP_01132767	
Pseudomonas entomophila L48	YP_607270			
Pseudomonas fluorescens Pf-5	YP_258749		YP_258751	
Pseudomonas fluorescens PfO-1	YP_347253		YP_347255	
Pseudomonas putida F1	YP_001269217		YP_001266824	
Pseudomonas putida GB-1	YP_001667719			
Pseudomonas putida KT2440	NP_743946			
Pseudomonas putida W619	YP_001750560		YP_001750558	
Pseudomonas stutzeri A1501	YP_001174303		YP_001174300	

Psychrobacter arcticus 273-4	YP_263949		YP_263953	YP_263955
Psychromonas sp. CNPT3	ZP_01215215		ZP_01215217	
Reinekea sp. MED297	ZP_01114619		ZP_01114616	
Shewanella amazonensis SB2B	YP_928211		YP_928206	
Shewanella baltica OS155			YP_001051309	
Shewanella baltica OS185	YP_001367175		YP_001367171	
Shewanella baltica OS195	YP_001555547		YP_001555543	
Shewanella baltica OS223			ZP_01840575	
Shewanella denitrificans OS217	YP_562297		YP_562298	YP_564113
Shewanella frigidimarina NCIMB 400	YP_751516			
Shewanella loihica PV-4	YP_001093456		YP_001093457	
Shewanella oneidensis MR-1			NP_718815	
Shewanella pealeana ATCC 700345	YP_001499916		YP_001499913	
Shewanella putrefaciens CN-32	YP_001184147		YP_001184145	
Shewanella sediminis HAW-EB3	YP_001474836		YP_001474832	
Shewanella sp. ANA-3	YP_868949		YP_868944	
Shewanella sp. MR-7	YP_737443		YP_737447	
Shewanella sp. W3-18-1	YP_962777		YP_962779	
Shewanella woodyi ATCC 51908	YP_001759959		YP_001759963	
Thiomicrospira crunogena XCL-2	YP_391722		YP_391724	
Vibrio fischeri ES114	YP_203530		YP_203526	
Vibrio harveyi ATCC BAA-1116	YP_001443906		YP_001443899	
Vibrio harveyi HY01	ZP_01987522		ZP_01987528	
Vibrio parahaemolyticus AQ3810	ZP_01993064	ZP_01992254	ZP_01992251	
Vibrio parahaemolyticus RIMD 2210633	NP_796582		NP_796579	
Vibrio shilonii AK1	ZP_01867489		ZP_01867488	
Vibrio sp. Ex25	ZP_01475338	ZP_01475320	ZP_01475341	
Vibrio splendidus 12B01	ZP_00989910		ZP_00989905	
Vibrio vulnificus CMCP6	NP_759780		NP_759785	
Vibrio vulnificus YJ016	NP_933109		NP_933105	
Vibrionales bacterium SWAT-3	ZP_01813606		ZP_01813603	

Acinetobacter baumannii ATCC 17978	*Francisella tularensis ss. novicida GA99-3549*	*Shewanella putrefaciens 200*
Acinetobacter sp. ADP1	*Francisella tularensis ss. novicida U112*	*Shewanella sp. MR-4*
Actinobacillus pleuropneumoniae serovar 1 str. 4074	*Francisella tularensis ss. tularensis FSC033*	*Shigella boydii CDC 3083-94*
Actinobacillus succinogenes 130Z	*Francisella tularensis subsp. tularensis FSC198*	*Shigella boydii Sb227*
Aeromonas salmonicida subsp. salmonicida A449	*Francisella tularensis ss. tularensis SCHU S4*	*Shigella dysenteriae 1012*
Alcanivorax borkumensis SK2	*Francisella tularensis ss. tularensis WY96-3418*	*Shigella dysenteriae Sd197*
Baumannia cicadellinicola str. Hc (Homalodisca coagulata)	*gamma proteobacterium KT 71*	*Shigella flexneri 2a str. 2457T*
Beggiatoa sp. PS	*Halorhodospira halophila SL1*	*Shigella flexneri 2a str. 301*
Buchnera aphidicola str. APS (Acyrthosiphon pisum)	*Idiomarina baltica OS145*	*Shigella flexneri 5 str. 8401*
Buchnera aphidicola str. Bp (Baizongia pistaciae)	*Klebsiella pneumoniae subsp. pneumoniae MGH 78578*	*Shigella sonnei Ss046*
Buchnera aphidicola str. Cc (Cinara cedri)	*Mannheimia haemolytica PHL213*	*Sodalis glossinidius str. 'morsitans'*
Buchnera aphidicola str. Sg (Schizaphis graminum)	*Mannheimia succiniciproducens MBEL55E*	*Vibrio alginolyticus 12G01*
Candidatus Blochmannia floridanus	*marine gamma proteobacterium HTCC2080*	*Vibrio angustum S14*
Candidatus Blochmannia pennsylvanicus str. BPEN	*marine gamma proteobacterium HTCC2143*	*Vibrio cholerae 1587*
Candidatus Carsonella ruddii PV	*Marinobacter aquaeolei VT8*	*Vibrio cholerae 2740-80*

Candidatus Ruthia magnifica str. Cm (Calyptogena magnifica)	*Methylococcus capsulatus str. Bath*	*Vibrio cholerae 623-39*
Candidatus Vesicomyosocius okutanii HA	*Neptuniibacter caesariensis*	*Vibrio cholerae AM-19226*
Chromohalobacter salexigens DSM 3043	*Nitrosococcus oceani ATCC 19707*	*Vibrio cholerae B33*
Citrobacter koseri ATCC BAA-895	*Photobacterium sp. SKA34*	*Vibrio cholerae MAK 757*
Coxiella burnetii Dugway 5J108-111	*Photorhabdus luminescens subsp. laumondii TTO1*	*Vibrio cholerae MO10*
Coxiella burnetii Dugway 7E9-12	*Pseudoalteromonas haloplanktis TAC125*	*Vibrio cholerae MZO-2*
Coxiella burnetii 'MSU Goat Q177'	*Pseudomonas aeruginosa 2192*	*Vibrio cholerae MZO-3*
Coxiella burnetii RSA 331	*Pseudomonas aeruginosa C3719*	*Vibrio cholerae NCTC 8457*
Coxiella burnetii RSA 334	*Pseudomonas aeruginosa PA7*	*Vibrio cholerae O1 biovar eltor str. N16961*
Coxiella burnetii RSA 493	*Pseudomonas aeruginosa PACS2*	*Vibrio cholerae O395*
ctinobacillus pleuropneumoniae L20	*Pseudomonas aeruginosa PAO1*	*Vibrio cholerae RC385*
Dichelobacter nodosus VCS1703A	*Pseudomonas aeruginosa UCBPP-PA14*	*Vibrio cholerae V51*
Endoriftia persephone 'Hot96_1+Hot96_2'	*Pseudomonas mendocina ymp*	*Vibrio cholerae V52*
Enterobacter sakazakii ATCC BAA-894	*Pseudomonas syringae pv. phaseolicola 1448A*	*Vibrio sp. MED222*
Enterobacter sp. 638	*Pseudomonas syringae pv. syringae B728a*	*Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis*
Erwinia carotovora subsp. atroseptica SCRI1043	*Pseudomonas syringae pv. tomato str. DC3000*	*Yersinia bercovieri ATCC 43970*
Escherichia coli 101-1	*Psychrobacter cryohalolentis K5*	*Yersinia enterocolitica subsp. enterocolitica 8081*
Escherichia coli 536	*Psychrobacter sp. PRwf-1*	*Yersinia frederiksenii ATCC 33641*
Escherichia coli 53638	*Psychromonas ingrahamii 37*	*Yersinia intermedia ATCC 29909*
Escherichia coli B	*Rickettsiella grylli*	*Yersinia mollaretii ATCC 43969*
Escherichia coli B171	*Saccharophagus degradans 2-40*	*Yersinia pestis Angola*
Escherichia coli B7A	*Salmonella enterica subsp. enterica serovar 4,[5],12:i:- str. CVM23701*	*Yersinia pestis Antiqua*
Escherichia coli CFT073	*S. enterica subsp. enterica serovar Agona str. SL483*	*Yersinia pestis biovar Antiqua str. B42003004*
Escherichia coli E110019	*S. enterica subsp. enterica serovar Choleraesuis str. SC-B67*	*Yersinia pestis biovar Antiqua str. E1979001*
Escherichia coli E22	*S. enterica subsp. enterica serovar Dublin str. CT_02021853*	*Yersinia pestis biovar Antiqua str. UG05-0454*
Escherichia coli E24377A	*S. enterica subsp. enterica serovar Heidelberg str. SL476*	*Yersinia pestis biovar Mediaevalis str. K1973002*
Escherichia coli F11	*S. enterica subsp. enterica serovar Javiana str. GA_MM04042433*	*Yersinia pestis biovar Microtus str. 91001*
Escherichia coli HS	*S. enterica subsp. enterica serovar Kentucky str. CDC 191*	*Yersinia pestis biovar Microtus str. 91001*
Escherichia coli O157:H7 EDL933	*S. enterica subsp. enterica serovar Kentucky str. CVM29188*	*Yersinia pestis biovar Orientalis str. F1991016*
Escherichia coli O157:H7 str. Sakai	*S. enterica subsp. enterica serovar Newport str. SL254*	*Yersinia pestis biovar Orientalis str. IP275*
Escherichia coli SECEC SMS-3-5	*S. enterica subsp. enterica serovar Newport str. SL317*	*Yersinia pestis biovar Orientalis str. MG05-1020*
Escherichia coli str. K-12 substr. MG1655	*S. enterica subsp. enterica serovar Paratyphi A str. ATCC 9150*	*Yersinia pestis CA88-4125*
Escherichia coli W3110	*S. enterica subsp. enterica serovar Saintpaul str. SARA23*	*Yersinia pestis CO92*
Francisella tularensis subsp. holarctica 257	*S. enterica subsp. enterica serovar Saintpaul str. SARA29*	*Yersinia pestis FV-1*
Francisella tularensis subsp. holarctica FSC022	*S. enterica subsp. enterica serovar Schwarzengrund str. CVM19633*	*Yersinia pestis KIM*
Francisella tularensis subsp. holarctica FSC200	*S. enterica subsp. enterica serovar Schwarzengrund str. SL480*	*Yersinia pestis Nepal516*
Francisella tularensis subsp. holarctica FTA	*S. enterica subsp. enterica serovar Typhi str. CT18*	*Yersinia pestis Pestoides F*
Francisella tularensis subsp. holarctica FTNF002-00	*S. enterica subsp. enterica serovar Typhi Ty2*	*Yersinia pseudotuberculosis IP 31758*
Francisella tularensis subsp. holarctica OSU18	*Salmonella typhimurium LT2*	*Yersinia pseudotuberculosis IP 3295*
Francisella tularensis subsp. novicida GA99-3548	*Serratia proteamaculans 568*	

β-Proteobacteria	**NAB-1**	**NAB-2**

Burkholderia cenocepacia MC0-3	YP_001765854				
Burkholderia phymatum STM815	YP_001859225			YP_001859224	
Burkholderia pseudomallei S13	ZP_01329806				
Chromobacterium violaceum ATCC 12472	NP_903698			NP_903701	
Dechloromonas aromatica RCB	YP_284470			YP_284473	
Herminiimonas arsenicoxydans				YP_001099432	
Methylophilales bacterium HTCC2181	ZP_01552361	NP_841608		NP_841609	
Neisseria meningitidis FAM18	YP_974195			YP_974194	
Neisseria meningitidis MC58	NP_273133			NP_273132	
Nitrosomonas europaea ATCC 19718	NP_841608			NP_841609	
Nitrosomonas eutropha C91	YP_747757			YP_747758	
Nitrosospira multiformis ATCC 25196	YP_411110			YP_411108	
Ralstonia pickettii 12D				ZP_02007588	
Verminephrobacter eiseniae EF01-2	YP_999560			YP_999561	

Acidovorax avenae subsp. citrulli AAC00-1	*Burkholderia oklahomensis EO147*	*Burkholderia thailandensis E264*
Acidovorax sp. JS42	*Burkholderia phytofirmans PsJN*	*Burkholderia thailandensis TXDOH*
Azoarcus sp. BH72	*Burkholderia pseudomallei 1106a*	*Burkholderia ubonensis Bu*
Azoarcus sp. EbN1	*Burkholderia pseudomallei 1106b*	*Burkholderia vietnamiensis G4*
Bordetella bronchiseptica RB50	*Burkholderia pseudomallei 112*	*Burkholderia xenovorans LB400*
Bordetella parapertussis 12822	*Burkholderia pseudomallei 14*	*Comamonas testosteroni KF-1*
Bordetella pertussis Tohama I	*Burkholderia pseudomallei 1655*	*Delftia acidovorans SPH-1*
Burkholderia ambifaria AMMD	*Burkholderia pseudomallei 1710a*	*ethylibium petroleiphilum PM1*
Burkholderia ambifaria MC40-6	*Burkholderia pseudomallei 1710b*	*Janthinobacterium sp. Marseille*
Burkholderia cenocepacia AU 1054	*Burkholderia pseudomallei 305*	*Limnobacter sp. MED105*
Burkholderia cenocepacia HI2424	*Burkholderia pseudomallei 381*	*Methylobacillus flagellatus KT*
Burkholderia dolosa AUO158	*Burkholderia pseudomallei 406e*	*Neisseria gonorrhoeae FA 1090*
Burkholderia mallei 2002721280	*Burkholderia pseudomallei 668*	*Neisseria meningitidis Z2491*
Burkholderia mallei ATCC 10399	*Burkholderia pseudomallei 7894*	*Polaromonas naphthalenivorans CJ2*
Burkholderia mallei ATCC 23344	*Burkholderia pseudomallei 9*	*Polaromonas sp. JS666*
Burkholderia mallei FMH	*Burkholderia pseudomallei 91*	*Polynucleobacter sp. QLW-P1DMWA-1*
Burkholderia mallei GB8 horse 4	*Burkholderia pseudomallei B7210*	*Ralstonia eutropha H16*
Burkholderia mallei JHU	*Burkholderia pseudomallei BCC215*	*Ralstonia eutropha JMP134*
Burkholderia mallei NCTC 10229	*Burkholderia pseudomallei DM98*	*Ralstonia metallidurans CH34*
Burkholderia mallei NCTC 10247	*Burkholderia pseudomallei K96243*	*Ralstonia pickettii 12J*
Burkholderia mallei PRL-20	*Burkholderia pseudomallei NCTC 13177*	*Ralstonia solanacearum GMI1000*
Burkholderia mallei SAVP1	*Burkholderia pseudomallei Pasteur 52237*	*Ralstonia solanacearum UW551*
Burkholderia multivorans ATCC 17616	*Burkholderia sp. 383*	*Rhodoferax ferrireducens T118*
Burkholderia oklahomensis C6786	*Burkholderia thailandensis Bt4*	*Thiobacillus denitrificans ATCC 25259*

Firmicutes	**NAB-1**			**NAB-2**	
Alkaliphilus oremlandii OhILAs	YP_001514040			YP_001514037	
Bacillus amyloliquefaciens FZB42				YP_001423072	
Bacillus cereus G9241	ZP_00236340	ZP_00239762		ZP_00236342	
Bacillus pumilus SAFR-032	YP_001487128			ZP_01724639	YP_001488643
Bacillus sp. B14905	ZP_01724636			ZP_01724639	
Bacillus subtilis subsp. subtilis str. 168				NP_391666	
Bacillus thuringiensis serovar israelensis ATCC 35646				ZP_00741540	
Clostridium acetobutylicum ATCC 824					NP_348805
Clostridium beijerinckii NCIMB 8052	YP_001311347				YP_001311344
Clostridium botulinum A str. ATCC 19397	YP_0013849	YP_001384963			YP_001384922

	23			YP_001384962		
Clostridium botulinum A str. ATCC 3502	YP_001255180	YP_001255220		YP_001255219	YP_001255179	
Clostridium botulinum A str. Hall	YP_001388392	YP_001388433		YP_001388391	YP_001388432	
Clostridium botulinum F str. Langeland	YP_001392001	YP_001391979		YP_001392003	YP_001391978	
Clostridium difficile QCD-32g58	ZP_01801682					
Clostridium kluyveri DSM 555	YP_001395849	YP_001395014		YP_001395855	YP_001395492	
Clostridium novyi NT	YP_877856			YP_877857	YP_877964	
Clostridium tetani E88	NP_782308		NP_781273	NP_782309		
Clostridium thermocellum ATCC 27405	YP_001039031	YP_001038625		YP_001039033	YP_001038624	
Desulfitobacterium hafniense DCB-2				ZP_01372474		
Desulfitobacterium hafniense Y51	YP_519540			YP_520419		
Enterococcus faecium DO	ZP_00604822			ZP_00604816		
Geobacillus kaustophilus HTA426	YP_148973			YP_148976		
Halothermothrix orenii H 168				ZP_01188511	ZP_01188516	
Lactobacillus johnsonii NCC 533		NP_965637				
Lactococcus lactis ssp. cremoris SK1		YP_808727				
Lactococcus lactis subsp. lactis Il140		NP_266866				
Moorella thermoacetica ATCC 39073	YP_429619			YP_429617		
Ruminococcus gnavus ATCC 29149	ZP_02041060			ZP_02041061		
Streptococcus agalactiae 18RS21	ZP_00781880	ZP_00780163	ZP_00781966	ZP_00781900		
Streptococcus agalactiae 2603V/R	NP_688167	NP_688113		NP_688170		
Streptococcus agalactiae 515	ZP_00788998	ZP_00789247		ZP_00789031		
Streptococcus agalactiae A909	YP_329861	YP_329810		YP_329864		
Streptococcus agalactiae CJB111	ZP_00788538	ZP_00786919		ZP_00788547		
Streptococcus agalactiae COH1	ZP_00785707	ZP_00785426		ZP_00785697		
Streptococcus agalactiae H36B	ZP_00782978	ZP_00782941		ZP_00782986		
Streptococcus agalactiae NEM316	NP_735677	NP_735617		NP_735680		
Streptococcus mutans UA159		NP_721495				
Streptococcus suis 05ZYH33	YP_001197947			YP_001197944		
Streptococcus suis 89/1591	ZP_00875903	ZP_00875443		ZP_00875906		
Streptococcus suis 98HAH33	YP_001200143			YP_001200140		
Streptococcus thermophilus CNRZ1		YP_141453				
Streptococcus thermophilus LMD-9		YP_820445				
Streptococcus thermophilus LMG 18		YP_139528				
Syntrophomonas wolfei subsp. wolfei str. Goettingen				YP_752933		
Thermoanaerobacter pseudethanolicus ATCC 33223				YP_001664975		
Thermoanaerobacter sp. X514				YP_001663050		
Thermoanaerobacter tengcongensis MB4				NP_622664		
Thermosinus carboxydivorans Nor1	ZP_01667693					

Alkaliphilus metalliredigens QYMF	*Eubacterium dolichum DSM 3991*	*Pelotomaculum thermopropionicum SI*
Anaerostipes caccae DSM 14662	*Eubacterium siraeum DSM 15702*	*Peptostreptococcus micros ATCC 33270*
Bacillus anthracis str. A1055	*Eubacterium ventriosum ATCC 27560*	*Ruminococcus obeum ATCC 29174*
Bacillus anthracis str. A2012	*Exiguobacterium sibiricum 255-15*	*Ruminococcus torques ATCC 27756*
Bacillus anthracis str. Ames	*Faecalibacterium prausnitzii M21/2*	*Staphylococcus aureus RF122*
Bacillus anthracis str. 'Ames Ancestor'	*Geobacillus thermodenitrificans NG80-2*	*Staphylococcus aureus subsp. aureus COL*

Bacillus anthracis str. Australia 94	*Lactobacillus acidophilus NCFM*	*Staphylococcus aureus subsp. aureus JH1*
Bacillus anthracis str. CNEVA-9066	*Lactobacillus brevis ATCC 367*	*Staphylococcus aureus subsp. aureus JH9*
Bacillus anthracis str. Kruger B	*Lactobacillus casei ATCC 334*	*Staphylococcus aureus subsp. aureus MRSA252*
Bacillus anthracis str. Sterne	*Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842*	*Staphylococcus aureus subsp. aureus MSSA476*
Bacillus anthracis str. Vollum	*Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365*	*Staphylococcus aureus subsp. aureus Mu3*
Bacillus anthracis str. Western North America USA6153	*Lactobacillus gasseri ATCC 33323*	*Staphylococcus aureus subsp. aureus Mu50*
Bacillus anthracis Tsiankovskii-I	*Lactobacillus plantarum WCFS1*	*Staphylococcus aureus subsp. aureus MW2*
Bacillus cereus 03BB108	*Lactobacillus reuteri 100-23*	*Staphylococcus aureus subsp. aureus N315*
Bacillus cereus AH1134	*Lactobacillus reuteri F275*	*Staphylococcus aureus subsp. aureus NCTC 8325*
Bacillus cereus AH187	*Lactobacillus sakei subsp. sakei 23K*	*Staphylococcus aureus subsp. aureus str. Newman*
Bacillus cereus AH820	*Lactobacillus salivarius subsp. salivarius UCC118*	*Staphylococcus aureus subsp. aureus USA300*
Bacillus cereus ATCC 10987	*Lactococcus lactis subsp. cremoris MG1363*	*Staphylococcus epidermidis ATCC 12228*
Bacillus cereus ATCC 14579	*Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293*	*Staphylococcus epidermidis RP62A*
Bacillus cereus B4264	*Listeria innocua Clip11262*	*Staphylococcus haemolyticus JCSC1435*
Bacillus cereus E33L	*Listeria monocytogenes 10403S*	*Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305*
Bacillus cereus H3081.97	*Listeria monocytogenes EGD-e*	*Streptococcus gordonii str. Challis substr. CH1*
Bacillus cereus NVH0597-99	*Listeria monocytogenes F6900*	*Streptococcus pneumoniae D39*
Bacillus cereus subsp. cytotoxis NVH 391-98	*Listeria monocytogenes FSL F2-515*	*Streptococcus pneumoniae R6*
Bacillus cereus W	*Listeria monocytogenes FSL J1-175*	*Streptococcus pneumoniae SP11-BS70*
Bacillus clausii KSM-K16	*Listeria monocytogenes FSL J1-194*	*Streptococcus pneumoniae SP14-BS69*
Bacillus coagulans 36D1	*Listeria monocytogenes FSL J1-208*	*Streptococcus pneumoniae SP18-BS74*
Bacillus halodurans C-125	*Listeria monocytogenes FSL J2-003*	*Streptococcus pneumoniae SP19-BS75*
Bacillus licheniformis ATCC 14580	*Listeria monocytogenes FSL J2-064*	*Streptococcus pneumoniae SP23-BS72*
Bacillus sp. NRRL B-14911	*Listeria monocytogenes FSL J2-071*	*Streptococcus pneumoniae SP3-BS71*
Bacillus sp. SG-1	*Listeria monocytogenes FSL N1-017*	*Streptococcus pneumoniae SP6-BS73*
Bacillus thuringiensis serovar konkukian str. 97-27	*Listeria monocytogenes FSL N3-165*	*Streptococcus pneumoniae SP9-BS68*
Bacillus thuringiensis str. Al Hakam	*Listeria monocytogenes FSL R2-503*	*Streptococcus pneumoniae TIGR4*
Bacillus weihenstephanensis KBAB4	*Listeria monocytogenes HPB2262*	*Streptococcus pyogenes M1 GAS*
Caldicellulosiruptor saccharolyticus DSM 8903	*Listeria monocytogenes J0161*	*Streptococcus pyogenes M49 591*
Carboxydothermus hydrogenoformans Z-2901	*Listeria monocytogenes J2818*	*Streptococcus pyogenes MGAS10270*
Clostridium bolteae ATCC BAA-613	*Listeria monocytogenes LO28*	*Streptococcus pyogenes MGAS10394*
Clostridium botulinum A str. Hall	*Listeria monocytogenes str. 1/2a F6854*	*Streptococcus pyogenes MGAS10750*
Clostridium botulinum Bf	*Listeria monocytogenes str. 4b F2365*	*Streptococcus pyogenes MGAS2096*
Clostridium botulinum C str. Eklund	*Listeria monocytogenes str. 4b H7858*	*Streptococcus pyogenes MGAS315*
Clostridium botulinum G	*Listeria welshimeri serovar 6b str. SLCC5334*	*Streptococcus pyogenes MGAS5005*
Clostridium botulinum NCTC 2916	*Mesoplasma florum L1*	*Streptococcus pyogenes MGAS6180*
Clostridium botulinum str. Iwanei E	*Mycoplasma agalactiae PG2*	*Streptococcus pyogenes MGAS8232*
Clostridium butyricum 5521	*Mycoplasma capricolum subsp. capricolum ATCC 27343*	*Streptococcus pyogenes MGAS9429*
Clostridium cellulolyticum H10	*Mycoplasma gallisepticum R*	*Streptococcus pyogenes SSI-1*
Clostridium difficile 630	*Mycoplasma genitalium G37*	*Streptococcus pyogenes str. Manfredo*
Clostridium difficile QCD-66c26	*Mycoplasma hyopneumoniae 232*	*Streptococcus sanguinis SK36*
Clostridium leptum DSM 753	*Mycoplasma hyopneumoniae 7448*	*Streptococcus thermophilus CNRZ1066*
Clostridium perfringens ATCC 13124	*Mycoplasma hyopneumoniae J*	*Symbiobacterium thermophilum IAM 14863*
Clostridium perfringens B str. ATCC 3626	*Mycoplasma mobile 163K*	*Ureaplasma parvum serovar 1*
Clostridium perfringens C str. JGS1495	*Mycoplasma mycoides subsp. mycoides LC str. GM12*	*Ureaplasma parvum serovar 14*
Clostridium perfringens CPE str. F4969	*Mycoplasma mycoides subsp. mycoides SC str. PG1*	*Ureaplasma parvum serovar 3*

Clostridium perfringens E str. JGS1987	*Mycoplasma penetrans HF-2*	*Ureaplasma parvum serovar 3 str. ATCC 700970*
Clostridium perfringens NCTC 8239	*Mycoplasma pneumoniae M129*	*Ureaplasma parvum serovar 6*
Clostridium perfringens SM101	*Mycoplasma pulmonis UAB CTIP*	*Ureaplasma urealyticum serovar 10*
Clostridium perfringens str. 13	*Mycoplasma synoviae 53*	*Ureaplasma urealyticum serovar 11*
Clostridium phytofermentans ISDg	*Oceanobacillus iheyensis HTE831*	*Ureaplasma urealyticum serovar 12*
Clostridium sp. L2-50	*Oenococcus oeni ATCC BAA-1163*	*Ureaplasma urealyticum serovar 13*
Desulfotomaculum reducens MI-1	*Oenococcus oeni PSU-1*	*Ureaplasma urealyticum serovar 4*
Dorea longicatena DSM 13814	*Onion yellows phytoplasma OY-M*	*Ureaplasma urealyticum serovar 5*
Enterococcus faecalis V583	*Paenibacillus larvae subsp. larvae BRL-230010*	*Ureaplasma urealyticum serovar 7*
Epulopiscium sp. 'N.t. morphotype B'	*Pasteuria nishizawae str. North American*	*Ureaplasma urealyticum serovar 8*
	Pediococcus pentosaceus ATCC 25745	*Ureaplasma urealyticum serovar 9*

Archaea	**NAB-1**	**NAB-2**
Haloquadratum walsbyi DSM 16790	YP_659191	YP_659192
Methanobrevibacter smithii ATCC 35061	YP_001273517	YP_001274112
Methanocaldococcus jannaschii DSM 2661	NP_247475	NP_248059
Methanococcoides burtonii DSM 6242	YP_566239	YP_566240
Methanococcus aeolicus Nankai-3	YP_001324615	YP_001324618
Methanococcus maripaludis C5	YP_001097036	YP_001097035
Methanosarcina acetivorans C2A	NP_618639	NP_618640
Methanosarcina barkeri str. Fusaro	YP_306910	
Methanospirillum hungatei JF-1	YP_504500	YP_504504

Methanococcus vannielii SB	*Methanococcus maripaludis C7*	*Pyrobaculum calidifontis JCM 11548*
Methanoculleus marisnigri JR1	*Methanococcus maripaludis S2*	*Pyrobaculum islandicum DSM 4184*
Aeropyrum pernix K1	*Methanocorpusculum labreanum Z*	*Pyrococcus abyssi GE5*
Archaeoglobus fulgidus DSM 4304	*Methanopyrus kandleri AV19*	*Pyrococcus furiosus DSM 3638*
Caldivirga maquilingensis IC-167	*Methanosaeta thermophila PT*	*Pyrococcus horikoshii OT3*
Candidatus Methanoregula boonei 6A8	*Methanosarcina mazei Go1*	*Staphylothermus marinus F1*
Ferroplasma acidarmanus fer1	*Methanosphaera stadtmanae DSM 3091*	*Sulfolobus acidocaldarius DSM 639*
Haloarcula marismortui ATCC 43049	*Methanothermobacter thermautotrophicus str. Delta H*	*Sulfolobus solfataricus P2*
Halobacterium sp. NRC-1	*Nanoarchaeum equitans Kin4-M*	*Sulfolobus tokodaii str. 7*
Halorubrum lacusprofundi ATCC 49239	*Natronomonas pharaonis DSM 2160*	*Thermococcus kodakarensis KOD1*
Hyperthermus butylicus DSM 5456	*Picrophilus torridus DSM 9790*	*Thermofilum pendens Hrk 5*
Ignicoccus hospitalis KIN4/I	*Pyrobaculum aerophilum str. IM2*	*Thermoplasma acidophilum DSM 1728*
Metallosphaera sedula DSM 5348	*Pyrobaculum arsenaticum DSM 13514*	*Thermoplasma volcanium GSS1*

Actinobacteria	**NAB-1**	**NAB-2**
Brevibacterium linens BL2	ZP_00379860	ZP_00379861
Mycobacterium gilvum PYR-GCK		YP_001136025
Streptomyces avermitilis MA-4680	NP_824551	NP_824550
Streptomyces coelicolor A3(2)]	NP_629034	NP_629033
Thermobifida fusca YX	YP_288072	YP_288073

Acidothermus cellulolyticus 11B	Frankia sp. CcI3	Mycobacterium tuberculosis CDC1551
Actinomyces odontolyticus ATCC 17982	Frankia sp. EAN1pec	Mycobacterium tuberculosis F11
Arthrobacter aurescens TC1	Janibacter sp. HTCC2649	Mycobacterium tuberculosis H37Ra
Arthrobacter sp. FB24	Kineococcus radiotolerans SRS30216	Mycobacterium tuberculosis H37Rv
Bifidobacterium adolescentis ATCC 15703	Leifsonia xyli subsp. xyli str. CTCB07	Mycobacterium tuberculosis str. Haarlem
Bifidobacterium adolescentis L2-32	*marine actinobacterium PHSC20C1*	*Mycobacterium ulcerans Agy99*
Bifidobacterium longum DJO10A	*Mycobacterium avium 104*	*Mycobacterium vanbaalenii PYR-1*
Bifidobacterium longum NCC2705	*Mycobacterium avium subsp. paratuberculosis K-10*	*Nocardia farcinica IFM 10152*

Clavibacter michiganensis subsp. michiganensis NCPPB 382	*Mycobacterium bovis AF2122/97*	*Nocardioides sp. JS614*
Collinsella aerofaciens ATCC 25986	*Mycobacterium bovis BCG str. Pasteur 1173P2*	*Propionibacterium acnes KPA171202*
Corynebacterium diphtheriae NCTC 13129	*Mycobacterium leprae TN*	*Rhodococcus sp. RHA1*
Corynebacterium efficiens YS-314	*Mycobacterium smegmatis str. MC2 155*	*Rubrobacter xylanophilus DSM 9941*
Corynebacterium glutamicum ATCC 13032	*Mycobacterium sp. JLS*	*Saccharopolyspora erythraea NRRL 2338*
Corynebacterium glutamicum R	*Mycobacterium sp. KMS*	*Salinispora arenicola CNS-205*
Corynebacterium jeikeium K411	*Mycobacterium sp. MCS*	*Salinispora tropica CNB-440*
Frankia alni ACN14a	*Mycobacterium tuberculosis C*	*Tropheryma whipplei str. Twist*
		Tropheryma whipplei TW08/27

Spirochetales	**NAB-1**		**NAB-2**	
Leptospira borgpetersenii serovar Hardjo-bovis JB197	YP_800486		YP_800485	YP_800448
Leptospira borgpetersenii serovar Hardjo-bovis L550	YP_797605		YP_797604	YP_797566
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	YP_002112	YP_002102	YP_002108 no ptase	YP_002104
Leptospira interrogans serovar Lai str. 56601	NP_711786	NP_711796	NP_711790	NP_711794
Treponema denticola ATCC 35405			NP_971570	
Treponema pallidum subsp. pallidum str. Nichols	NP_218729		NP_219001	

Borrelia afzelii ACA-1	*Borrelia burgdorferi B31*	*Borrelia garinii PBi*	
Borrelia afzelii PKo	*Borrelia burgdorferi Bol26*	*Borrelia valaisiana VS116*	
Borrelia burgdorferi 156a	*Borrelia burgdorferi ZS7*		