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.

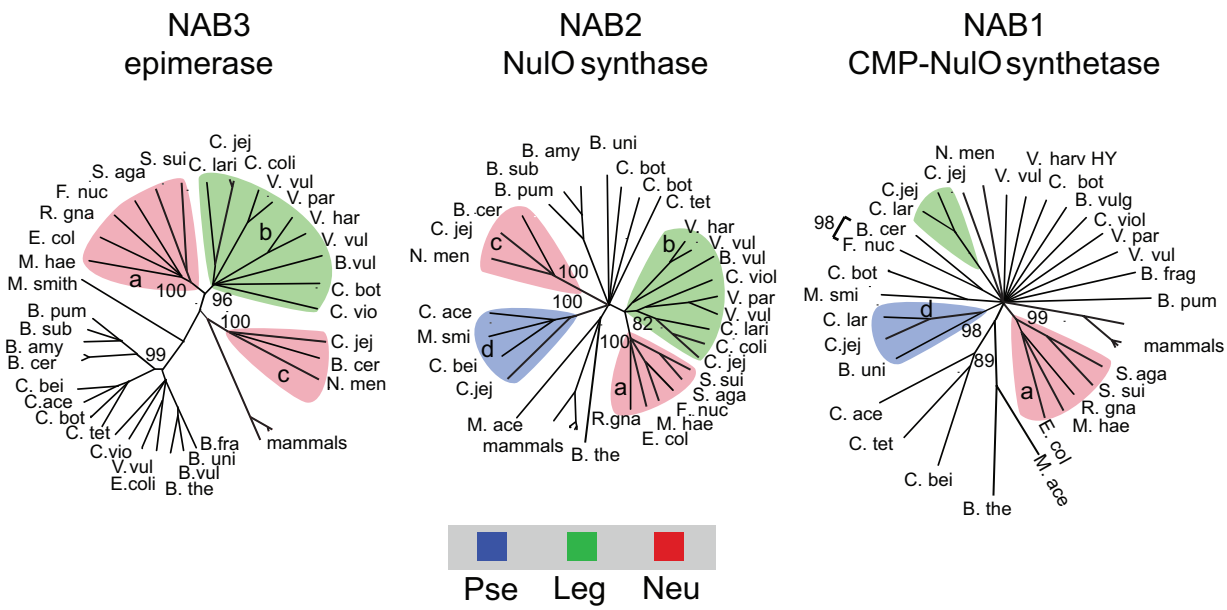

Fig. S1. Phylogenies of NAB pathways reveal distinct bacterial innovations of Sia mimicry. Phylogenetic trees were constructed based on NAB-1, NAB-2, and NAB-3 amino acid sequences collected from pathogenic bacteria represented in clades ''a'' and ''c,'' as well as a defined subset of organisms representative 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.

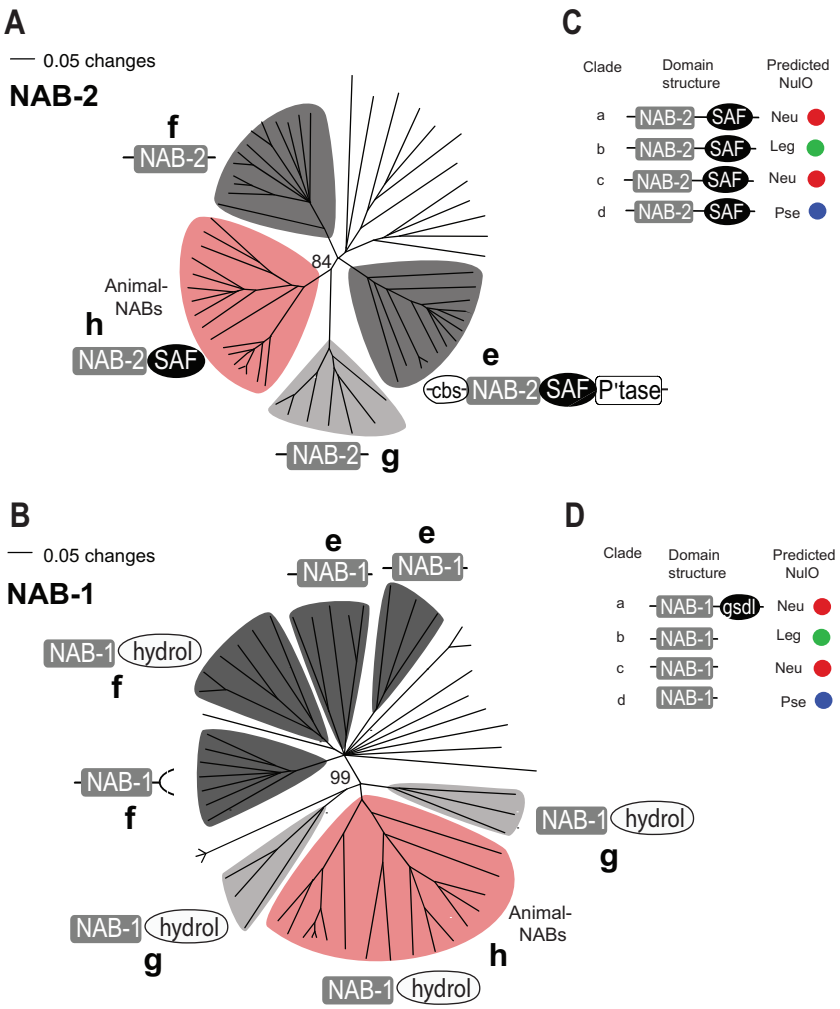

Fig. S2. Phylogenetic analysis of animal and animal-like NAB-1 and NAB-2 amino acid sequences are consistent with domain structure. (*A* and *B*), NAB-1 (*A*) and NAB-2 (*B*) amino acid sequences from organisms represented in Fig. 3, phylogenetic clades ''e''–''h,'' were collected by BLASTp and subjected 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.

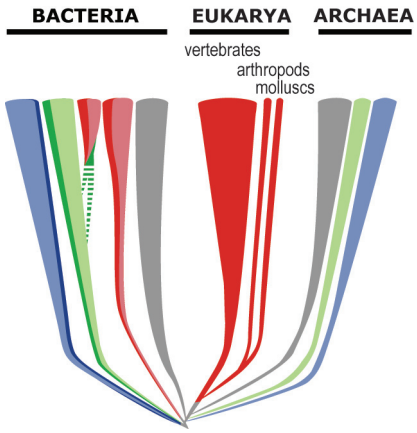

Fig. S3. A model of NulO evolution based on phylogenomic evidence. Based on phylogenetic and genomic evidence, we suggest that an early cellular diversification of NulO sugar structures resulted in the wide variety and distribution of NulO sugars that we find today (darker colors reflect published data; 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).

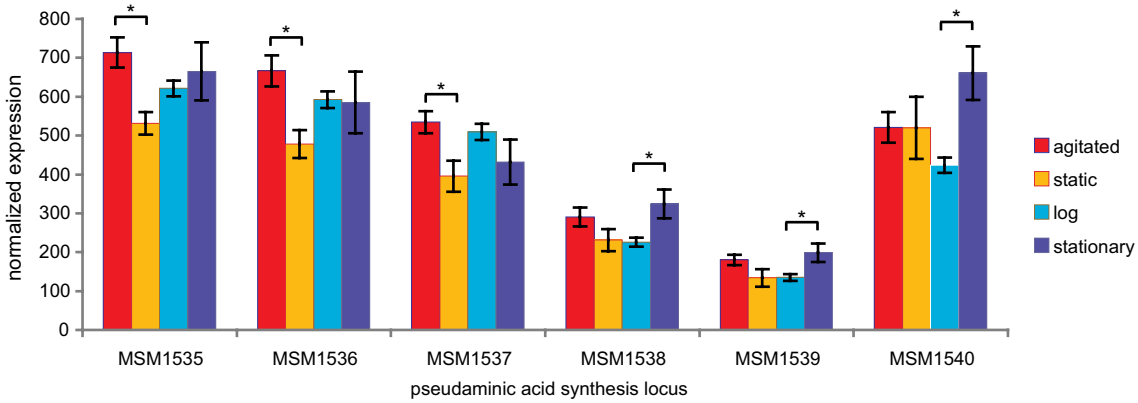

Fig. S4. Expression of predicted *M. smithii* Pse synthesis genes during growth from log phase to stationary phase in standard medium, with or without agitation. Mean values for GeneChip probeset intensities ± SEM are plotted ($n = 9–13$ GeneChips/condition). Asterisks indicate statistically significant 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.

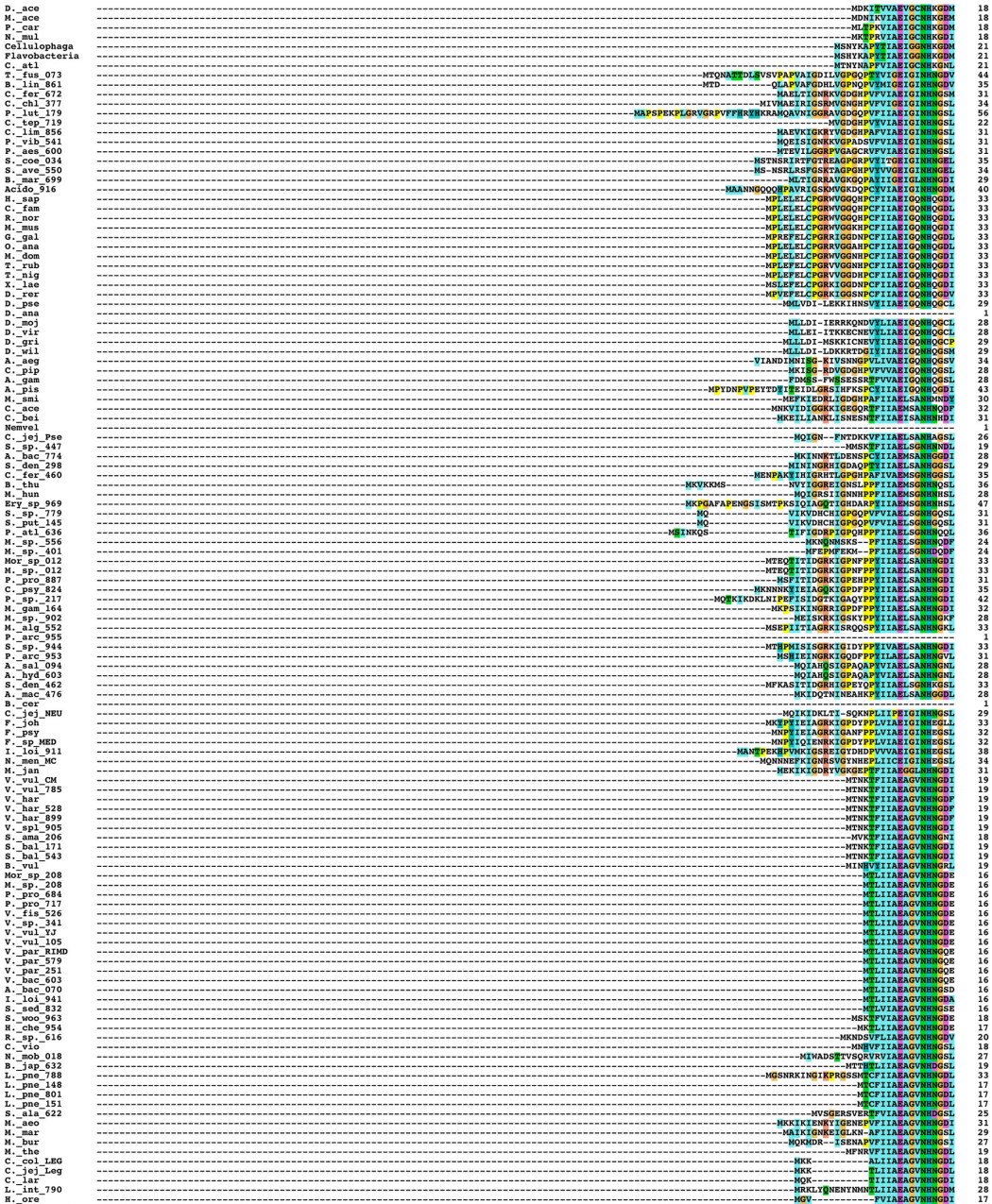

Fig. S5. Multiple sequence alignment of NAB-2 amino acid sequences used for construction of the phylogenetic tree in Fig. 3. Clustal Qt alignment of NAB-2 sequences. With the exception of the more limited sequences from sea anemone (*Nematostella*) and *P. arctica*, all gaps in the alignment were excluded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

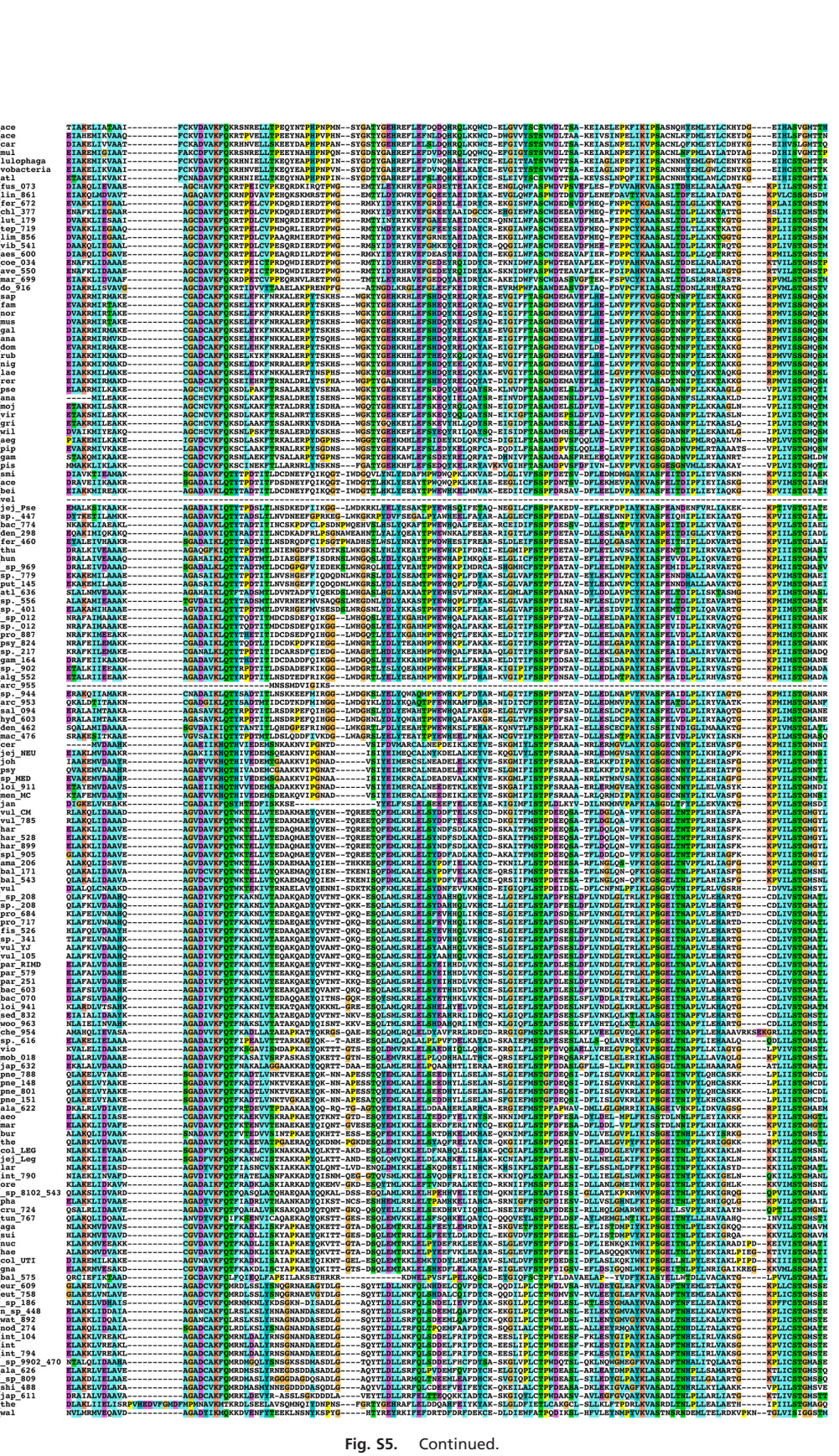

Fig. S5. Continued.

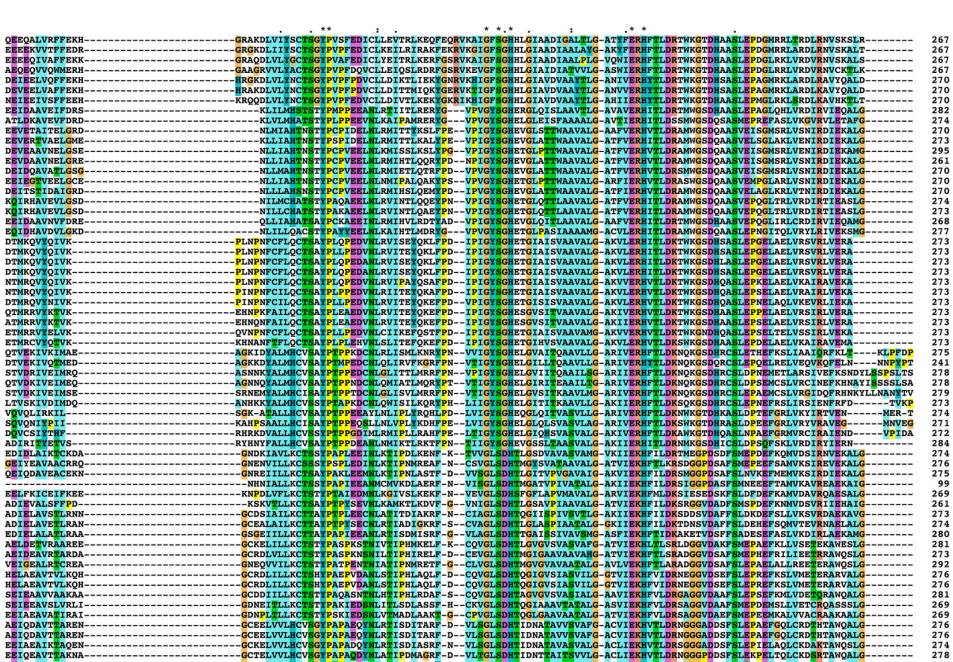

Fig. S5. Continued.

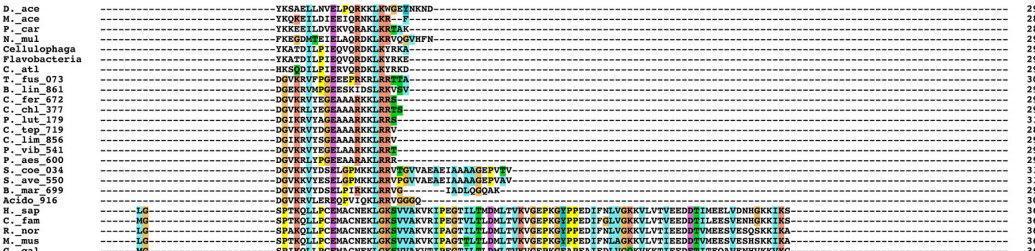

Fig. S5. Continued.

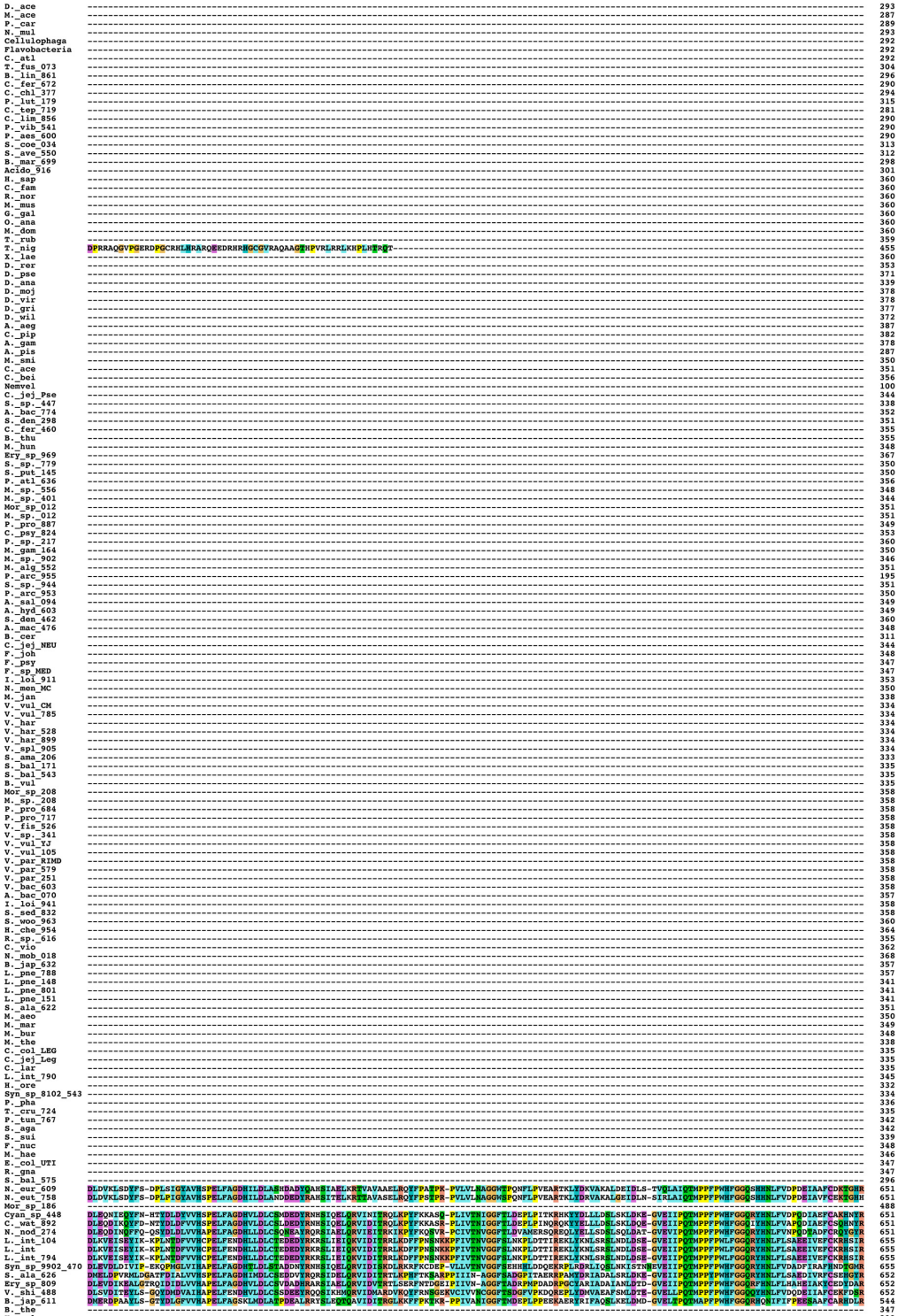

Fig. S5. Continued.

Other Supporting Information

Table S1