

Supporting Information

Lewis et al. 10.1073/pnas.0902431106

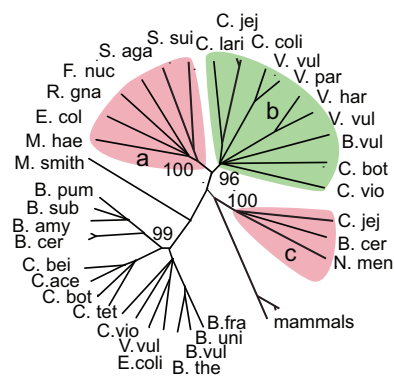
SI Materials and Methods

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 H₂ and CO₂ (4:1) that was replenished every 6 h, and harvested during the log or stationary phase (log phase: OD₆₀₀ of 1.10 and 0.36 for agitated and static cultures, respectively; stationary phase: OD₆₀₀ of 3.14 and 0.57, respectively). RNA was isolated, and cDNAs were prepared and then hybridized to GeneChips as described previously (7) ($n = 9\text{--}13$ 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.

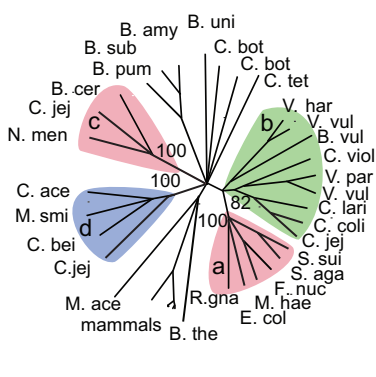
Isolation and Characterization of NulOs. After resuspension in 2N acetic acid, cell suspensions were incubated at 80 °C for 3 h to

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.

NAB3 epimerase



NAB2 NulO synthase



NAB1 CMP-NulO synthetase

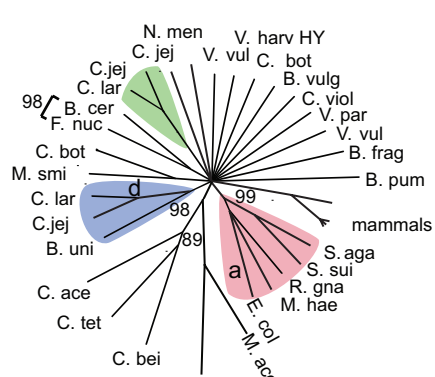


Fig. S1. Phylogenies of NAB pathways reveal distinct bacterial innovations of Sia mimicry. Phylogenetic trees were constructed based on NAB-1, NAB-2, and NAB-3 amino acid sequences collected from pathogenic bacteria represented in clades “a” and “c,” as well as a defined subset of organisms representative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

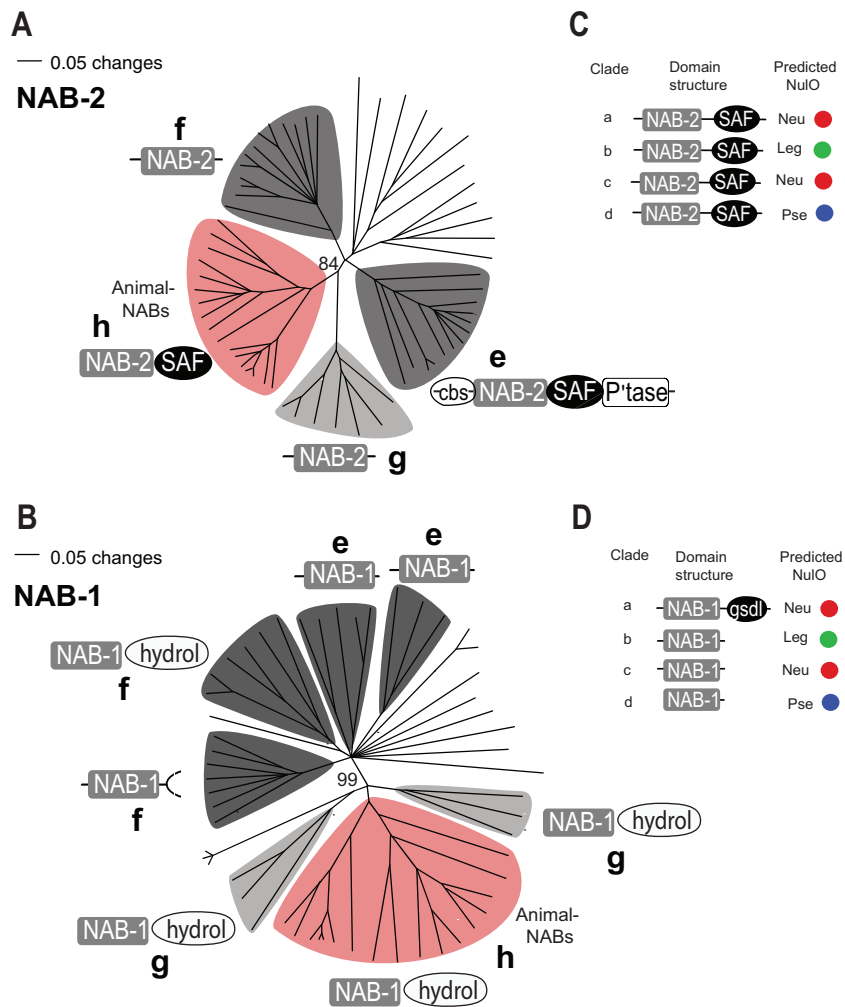


Fig. S2. Phylogenetic analysis of animal and animal-like NAB-1 and NAB-2 amino acid sequences are consistent with domain structure. (A and B), NAB-1 (A) and NAB-2 (B) amino acid sequences from organisms represented in Fig. 3, phylogenetic clades "e"–"h," were collected by BLASTp and subjected to a phylogenetic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

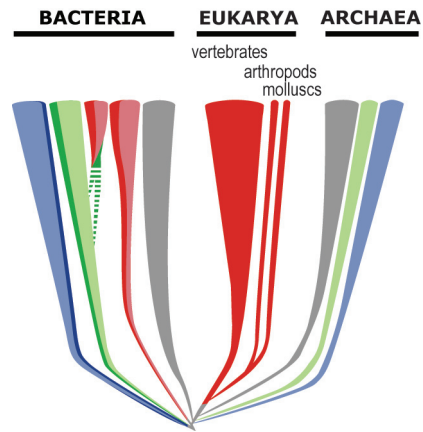


Fig. S3. A model of NuLO evolution based on phylogenomic evidence. Based on phylogenetic and genomic evidence, we suggest that an early cellular diversification of NuLO sugar structures resulted in the wide variety and distribution of NuLO sugars that we find today (darker colors reflect published data;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

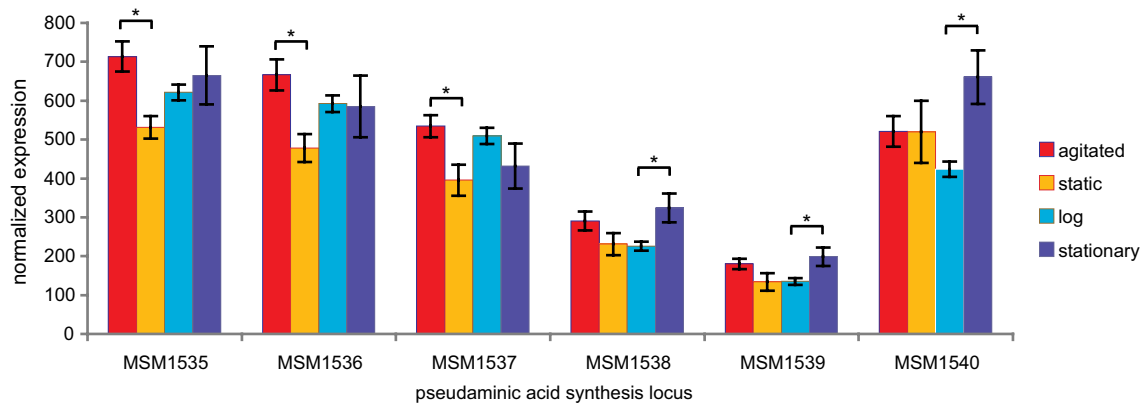


Fig. S4. Expression of predicted *M. smithii* Pse synthesis genes during growth from log phase to stationary phase in standard medium, with or without agitation. Mean values for GeneChip probeset intensities \pm SEM are plotted ($n = 9-13$ GeneChips/condition). Asterisks indicate statistically significant differences ($P < .05$ by the Student t test). Original annotations of these genes were dTDP-D-glucose 4,6 dehydratase (MSM1535), acylneuraminate cytidyltransferase (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.

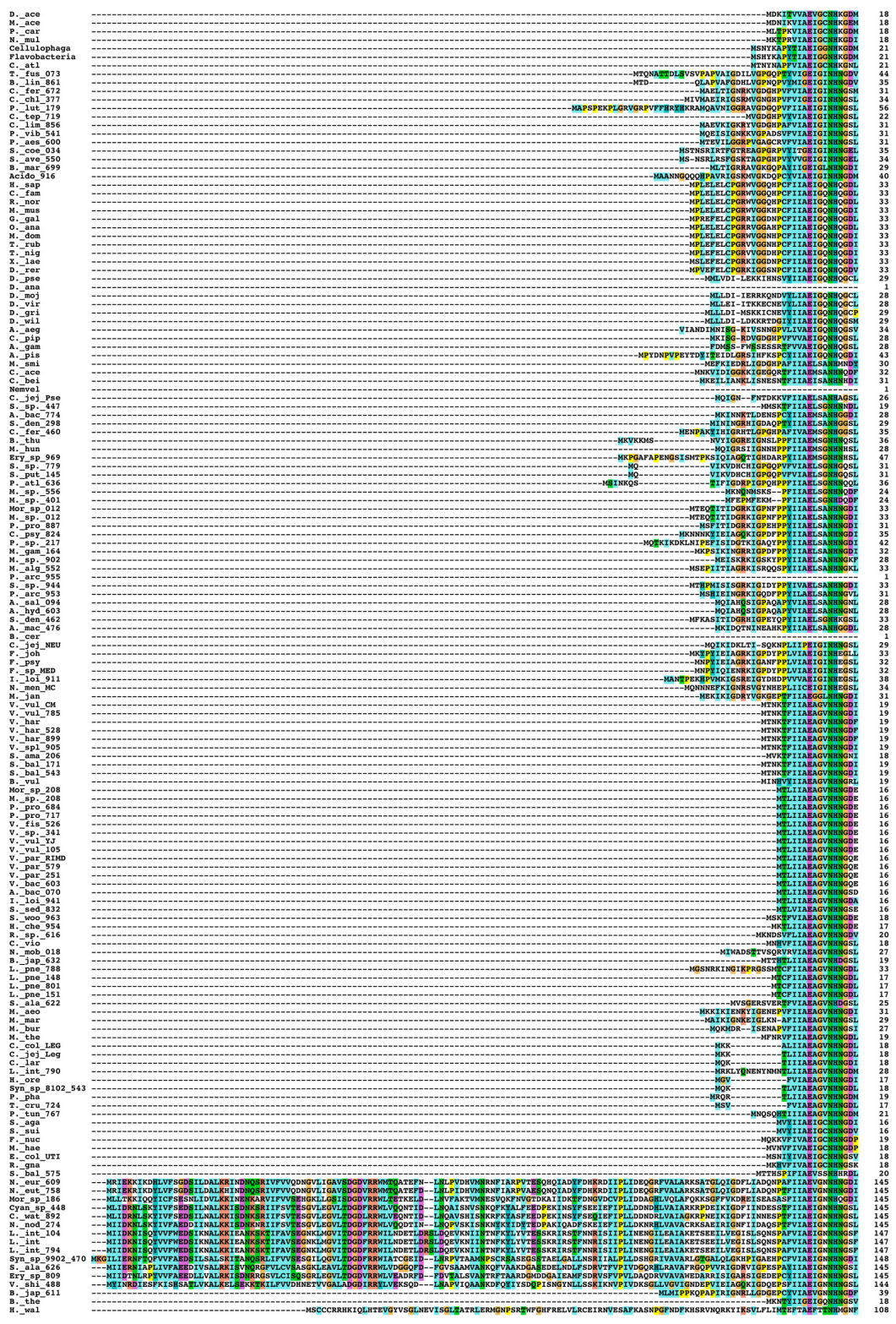


Fig. S5. Multiple sequence alignment of NAB-2 amino acid sequences used for construction of the phylogenetic tree in Fig. 3. Clustal Qt alignment of NAB-2 sequences. With the exception of the more limited sequences from sea anemone (*Nematostella*) and *P. arctica*, all gaps in the alignment were excluded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

D_ace	---	YKSAELNVELPQRKLMGKLNKD	293
M_ace	---	YKREIDIDIBETORNKLR - F	287
P_cnr	---	YKREIDIDIBETORNKLR - F	287
N_mul	---	FRESDMELIADLDKLRVGVFVN	293
Cellulophaga	---	YKATDILPISQVDRKLRKRA	292
Flavobacteriia	---	YKATDILPISQVDRKLRKRA	292
C_atl	---	HKSDFILPFRVDRKLRKRD	292
T_fus_073	---	DQVRRVFTGAAARAKLR	292
B_lin_861	---	DQVRRVFTGAAARAKLR	296
C_fer_672	---	DQVRRVFTGAAARAKLR	294
C_chl_177	---	DQVRRVFTGAAARAKLR	294
P_lut_179	---	DQVRRVFTGAAARAKLR	315
C_tep_719	---	DQVRRVFTGAAARAKLR	281
C_lin_856	---	DQVRRVFTGAAARAKLR	290
P_vib_541	---	DQVRRVFTGAAARAKLR	290
P_ace_600	---	DQVRRVFTGAAARAKLR	290
S_cen_034	---	DQVRRVFTGAAARAKLR	290
S_ave_550	---	DQVRRVFTGAAARAKLR	290
B_mai_599	---	DQVRRVFTGAAARAKLR	290
Acidib_916	---	DQVRRVFTGAAARAKLR	290
H_sap	---	DQVRRVFTGAAARAKLR	290
C_fam	---	DQVRRVFTGAAARAKLR	290
R_nor	---	DQVRRVFTGAAARAKLR	290
M_ana	---	DQVRRVFTGAAARAKLR	290
G_gal	---	DQVRRVFTGAAARAKLR	290
M_mus	---	DQVRRVFTGAAARAKLR	290
M_dom	---	DQVRRVFTGAAARAKLR	290
T_rub	---	DQVRRVFTGAAARAKLR	290
N_nig	---	DQVRRVFTGAAARAKLR	290
L_lae	---	DQVRRVFTGAAARAKLR	290
D_ror	---	DQVRRVFTGAAARAKLR	290
X_pse	---	DQVRRVFTGAAARAKLR	290
D_mir	---	DQVRRVFTGAAARAKLR	290
D_vir	---	DQVRRVFTGAAARAKLR	290
D_wil	---	DQVRRVFTGAAARAKLR	290
A_ace	---	DQVRRVFTGAAARAKLR	290
C_pip	---	DQVRRVFTGAAARAKLR	290
A_gam	---	DQVRRVFTGAAARAKLR	290
M_mil	---	DQVRRVFTGAAARAKLR	290
C_ace	---	DQVRRVFTGAAARAKLR	290
C_ace	---	DQVRRVFTGAAARAKLR	290
Nemvel	---	DQVRRVFTGAAARAKLR	290
C_jaj_760	---	DQVRRVFTGAAARAKLR	290
F_sp_447	---	DQVRRVFTGAAARAKLR	290
A_bac_774	---	DQVRRVFTGAAARAKLR	290
S_den_298	---	DQVRRVFTGAAARAKLR	290
C_fer_460	---	DQVRRVFTGAAARAKLR	290
B_thu	---	DQVRRVFTGAAARAKLR	290
M_hun	---	DQVRRVFTGAAARAKLR	290
Ery_sp_969	---	DQVRRVFTGAAARAKLR	290
S_sp_779	---	DQVRRVFTGAAARAKLR	290
P_pu_145	---	DQVRRVFTGAAARAKLR	290
P_ncl_636	---	DQVRRVFTGAAARAKLR	290
M_sp_556	---	DQVRRVFTGAAARAKLR	290
M_sp_401	---	DQVRRVFTGAAARAKLR	290
Mor_sp_012	---	DQVRRVFTGAAARAKLR	290
M_sp_012	---	DQVRRVFTGAAARAKLR	290
P_pro_887	---	DQVRRVFTGAAARAKLR	290
C_psy_824	---	DQVRRVFTGAAARAKLR	290
P_sp_217	---	DQVRRVFTGAAARAKLR	290
M_gam_164	---	DQVRRVFTGAAARAKLR	290
M_sp_902	---	DQVRRVFTGAAARAKLR	290
M_ala_552	---	DQVRRVFTGAAARAKLR	290
P_ace_955	---	DQVRRVFTGAAARAKLR	290
S_sp_944	---	DQVRRVFTGAAARAKLR	290
S_den_953	---	DQVRRVFTGAAARAKLR	290
A_sal_094	---	DQVRRVFTGAAARAKLR	290
A_hyd_603	---	DQVRRVFTGAAARAKLR	290
S_den_462	---	DQVRRVFTGAAARAKLR	290
A_mac_476	---	DQVRRVFTGAAARAKLR	290
B_cer	---	DQVRRVFTGAAARAKLR	290
F_joi_NEU	---	DQVRRVFTGAAARAKLR	290
F_joh	---	DQVRRVFTGAAARAKLR	290
P_mf	---	DQVRRVFTGAAARAKLR	290
F_sp_MED	---	DQVRRVFTGAAARAKLR	290
I_loi_111	---	DQVRRVFTGAAARAKLR	290
I_loi_112	---	DQVRRVFTGAAARAKLR	290
M_jan	---	DQVRRVFTGAAARAKLR	290
V_vul_CM	---	DQVRRVFTGAAARAKLR	290
V_vul_785	---	DQVRRVFTGAAARAKLR	290
V_har	---	DQVRRVFTGAAARAKLR	290
V_har_528	---	DQVRRVFTGAAARAKLR	290
V_har_599	---	DQVRRVFTGAAARAKLR	290
V_sp1_905	---	DQVRRVFTGAAARAKLR	290
S_ama_206	---	DQVRRVFTGAAARAKLR	290
S_bal_171	---	DQVRRVFTGAAARAKLR	290
S_bal_543	---	DQVRRVFTGAAARAKLR	290
S_vu	---	DQVRRVFTGAAARAKLR	290
Mor_sp_208	---	DQVRRVFTGAAARAKLR	290
M_sp_208	---	DQVRRVFTGAAARAKLR	290
P_pro_584	---	DQVRRVFTGAAARAKLR	290
P_pro_717	---	DQVRRVFTGAAARAKLR	290
V_fis_526	---	DQVRRVFTGAAARAKLR	290
V_sp_241	---	DQVRRVFTGAAARAKLR	290
V_vul_17	---	DQVRRVFTGAAARAKLR	290
V_vul_105	---	DQVRRVFTGAAARAKLR	290
V_pnt_180	---	DQVRRVFTGAAARAKLR	290
V_sp_579	---	DQVRRVFTGAAARAKLR	290
V_sp_551	---	DQVRRVFTGAAARAKLR	290
A_bac_603	---	DQVRRVFTGAAARAKLR	290
A_bac_070	---	DQVRRVFTGAAARAKLR	290
S_woc_963	---	DQVRRVFTGAAARAKLR	290
H_che_954	---	DQVRRVFTGAAARAKLR	290
S_sed_832	---	DQVRRVFTGAAARAKLR	290
I_loi_941	---	DQVRRVFTGAAARAKLR	290
R_sp_616	---	DQVRRVFTGAAARAKLR	290
C_vio	---	DQVRRVFTGAAARAKLR	290
N_mob_018	---	DQVRRVFTGAAARAKLR	290
B_jap_632	---	DQVRRVFTGAAARAKLR	290
L_pne_788	---	DQVRRVFTGAAARAKLR	290
L_pne_148	---	DQVRRVFTGAAARAKLR	290
L_pne_801	---	DQVRRVFTGAAARAKLR	290
S_ala_622	---	DQVRRVFTGAAARAKLR	290
M_aeo	---	DQVRRVFTGAAARAKLR	290
M_mur	---	DQVRRVFTGAAARAKLR	290
M_bur	---	DQVRRVFTGAAARAKLR	290
M_che	---	DQVRRVFTGAAARAKLR	290
C_col_LEG	---	DQVRRVFTGAAARAKLR	290
C_jej_1e9	---	DQVRRVFTGAAARAKLR	290
C_jar	---	DQVRRVFTGAAARAKLR	290
L_int_790	---	DQVRRVFTGAAARAKLR	290
H_ore	---	DQVRRVFTGAAARAKLR	290
Syr_sp_8102_543	---	DQVRRVFTGAAARAKLR	290
P_phi	---	DQVRRVFTGAAARAKLR	290
T_cru_724	---	DQVRRVFTGAAARAKLR	290
P_tun_767	---	DQVRRVFTGAAARAKLR	290
S_aga	---	DQVRRVFTGAAARAKLR	290
S_aui	---	DQVRRVFTGAAARAKLR	290
F_ruc	---	DQVRRVFTGAAARAKLR	290
M_hae	---	DQVRRVFTGAAARAKLR	290
R_col_VIT	---	DQVRRVFTGAAARAKLR	290
R_gna	---	DQVRRVFTGAAARAKLR	290
S_bal_575	---	DQVRRVFTGAAARAKLR	290
N_eut_509	---	DQVRRVFTGAAARAKLR	290
N_eut_578	---	DQVRRVFTGAAARAKLR	290
Cyan_sp_448	---	DQVRRVFTGAAARAKLR	290
C_wat_892	---	DQVRRVFTGAAARAKLR	290
N_nod_274	---	DQVRRVFTGAAARAKLR	290
L_int_104	---	DQVRRVFTGAAARAKLR	290
L_int	---	DQVRRVFTGAAARAKLR	290
L_int_194	---	DQVRRVFTGAAARAKLR	290
Syr_sp_9902_470	---	DQVRRVFTGAAARAKLR	290
S_ala_626	---	DQVRRVFTGAAARAKLR	290
Ery_sp_809	---	DQVRRVFTGAAARAKLR	290
V_shl_488	---	DQVRRVFTGAAARAKLR	290
B_jap_511	---	DQVRRVFTGAAARAKLR	290
H_wal	---	DQVRRVFTGAAARAKLR	290

Fig. 55. Continued.

