Supplementary Information

Identification of an N-acetylneuraminic acid presenting-bacteria isolated from a human microbiome

Zhen Han[1], Peter S. Thuy-Boun[1], Wayne Pfeiffer[2], Vincent F. Vartabedian[3], Ali Torkamani[4,5], John R. Teijaro[3], Dennis W. Wolan[1,5]

[1]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037
[2]San Diego Supercomputer Center, University of California, San Diego, La Jolla CA 92093
[3]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037
[4]Scripps Research Translational Institute, The Scripps Research Institute, La Jolla, CA 92037
[5]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037

Supplementary Figures and Tables

Fig. 1.	Dose-dependent labeling of H2 cultured microbiome sample.
Fig. 2.	Sia9N_3 metabolic labeling of microbiota H2 competed with Neu5Ac and removed by a microbial sialidase BT-0455.
Fig. 3.	16S rDNA sequencing of primary and cultured microbiome samples.
Fig. 4.	Sia9N_3 incorporation by *E. coli*-K12 (MG1655) and the newly isolated *E. coli* strain.
Fig. 5.	Sequence alignment of ZH-C2_04397 to other *E. coli* proteins.
Fig. 6.	Sequence alignment of ZH-C2_04599 to other *E. coli* proteins.
Supp. Table 1	7 tabs

Supplementary Table 1 legend: Additional bioinformatic data analyses of *E. coli* genome composition. Related to Figure 6. Tab 1: *E. coli* genome assembly statistics and comparisons to reference strains. Tab 2: List of low-homology and unique *E. coli* ZH-C2 genes relative to K-12 MG1655. Tab 3: Biological process, molecular function, and cellular component GO Terms associated with unique *E. coli* genomes. Tab 4: List of low-homology and unique *E. coli* ZH-C2 genes relative to UMEA 3174-1. Tab 5: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. Tab 6: List of low-homology and unique *E. coli* ZH-C2 genes relative RefSeq collection. Low homology hit proteins are best match spans with <70% of query protein length and/or <90% positive identity rate. Tab 7: List of genes and protein products of ZH-C2.

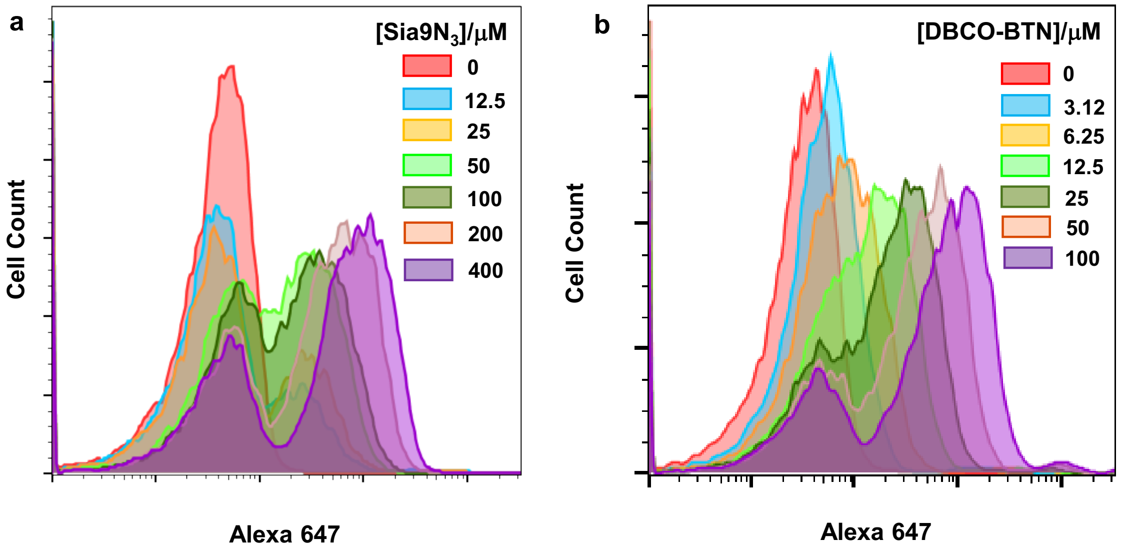


Fig. 1. Sia9N$_3$ (**a**) and DBCO-BTN (**b**) dose-dependent labeling of H2 cultured microbiome sample.

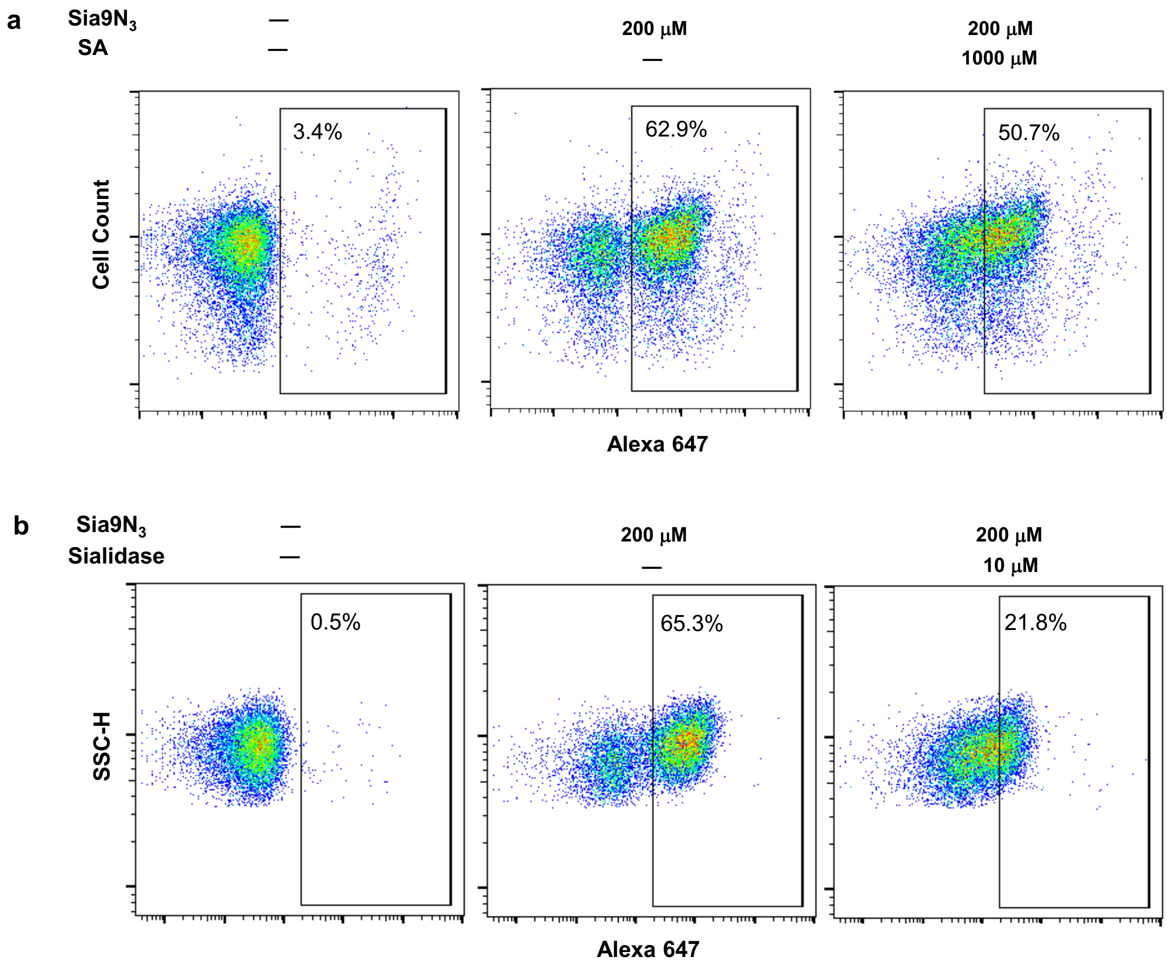


Fig. 2. a $Sia9N_3$ metabolic labeling of microbiome H2 competed with Neu5Ac. **b** $Sia9N_3$ metabolic labeling is partially removed by a microbial sialidase BT-0455.

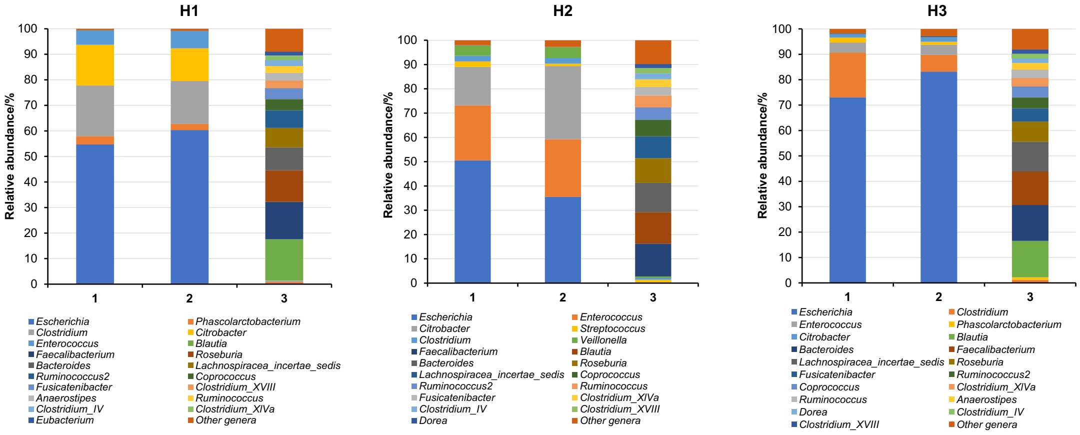


Fig. 3. 16S rRNA gene sequencing revealed taxonomic composition of cultured and primary microbiomes in the presence and absence of 200 μM Sia9N$_3$. Cultured human microbiome samples H1-H3 were all tested, column 1: microbiome cultured without Sia9N$_3$, column 2: microbiome cultured with Sia9N$_3$, 3: uncultured fecal microbiome samples.

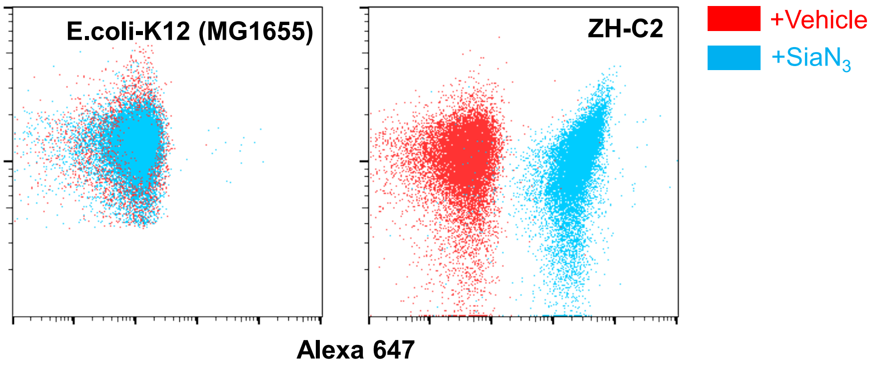
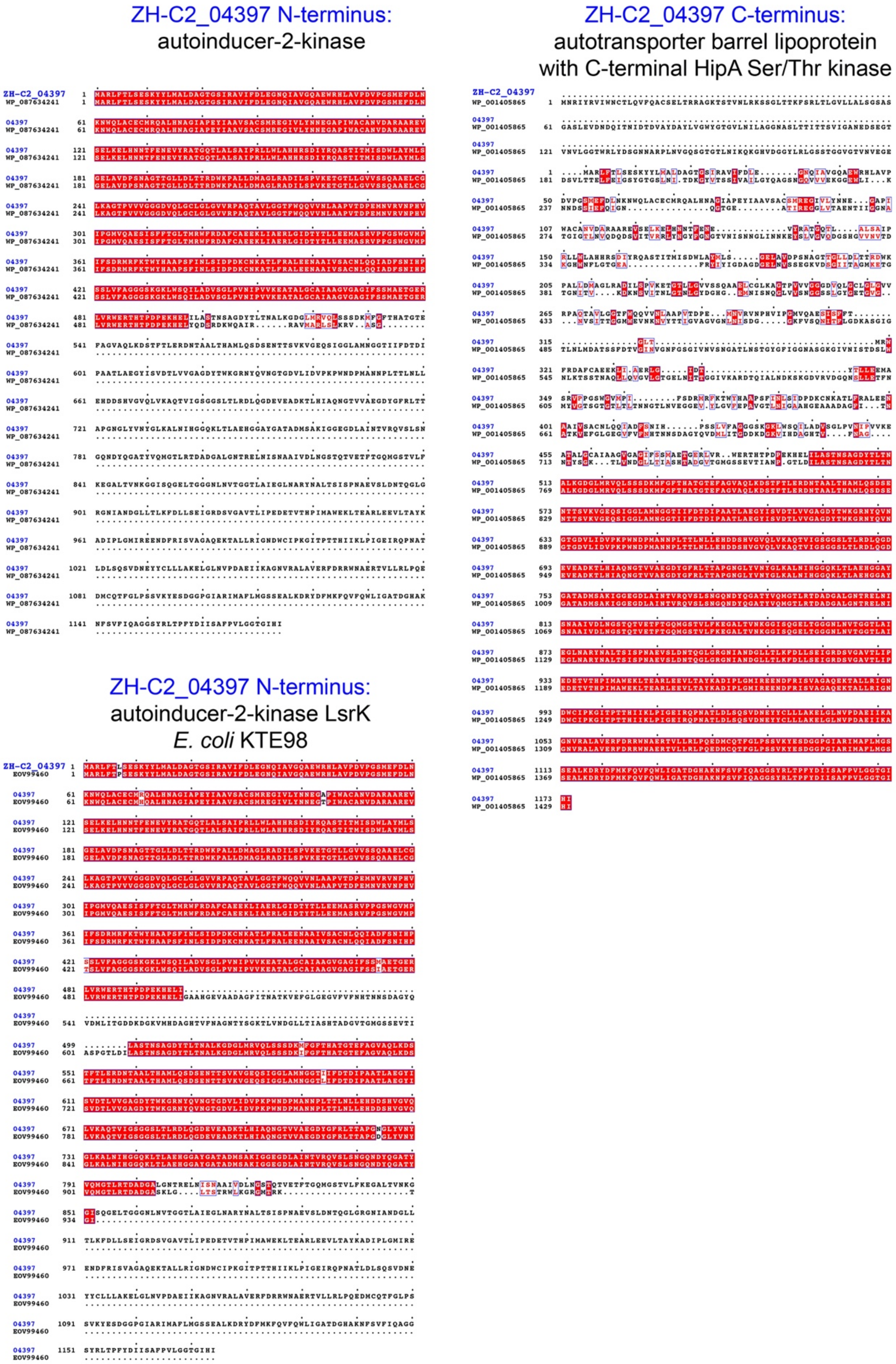


Fig. 4. Test of Sia9N$_3$ incorporation by *E. coli*-K12 (MG1655) and the newly isolated *E. coli* strain ZH-C2.



ZH-C2_04397 N-terminus:
autoinducer-2-kinase

ZH-C2_04397 N-terminus:
autoinducer-2-kinase LsrK
E. coli KTE98

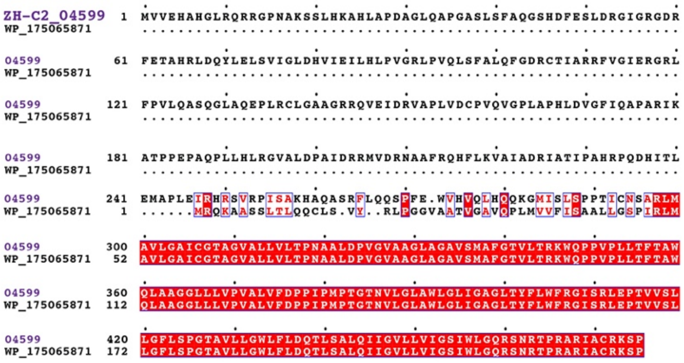
ZH-C2_04397 C-terminus:
autotransporter barrel lipoprotein
with C-terminal HipA Ser/Thr kinase

Fig. 5. Unique multi-domain ZH-C2_04397 with an N-terminal transposase, autotransporter barrel lipoprotein, and C-terminal HipA-like Ser/Thr kinase aligns to single-domain *E. coli* proteins.

ZH-C2_04599 N-terminus:
tranposase (plasmid)

```
ZH-C2_04599 1   MVVEHAHGLRQRRGPNAKSSLHKAHLAPDAGLQAPGASLSFAQGSHDFESLDRGIGRGDR
ACQ42061    1   MVVEHAHGLRQRRGPNAKSSLHKAHLAPDAGLQAPGASLSFAQGSHDFESLDRGIGRGDR

04599       61  FETAHRLDQYLELSVIGLDHVIEILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
ACQ42061    61  FETAHRLDQYLELSVIGLDHVIEILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL

04599       121 FPVLQASQGLAQEPLRCLGAAGRRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
ACQ42061    121 FPVLQASQGLAQEPLRCLGAAGRRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK

04599       181 ATPPEPAQPLLHLRGVALDPAIDRRMVDRNAAFRQHFLKVAIADRIATIPAHRPQDHITL
ACQ42061    181 ATPPEPAQPLLHLRGVALDPAIDRRMVDRNAAFRQHFLKVAIADRIATIPAHRPQDHITL

04599       241 EMAPLEIRHRSVRPISAKHAQASRFLQQSPFEWVHVQLHQQKGMISLSPPTICNSARLMA
ACQ42061    241 EMAPLEIRHRSVRPISAKHAQASRFLQQSPHEY...........................

04599       301 VLGAICGTAGVALLVLTPNAALDPVGVAAGLAGAVSMAFGTVLTRKWQPPVPLLTFTAWQ
ACQ42061        ............................................................

04599       361 LAAGGLLLVPVALVFDPPIPMPTGTNVLGLAWLGLIGAGLTYFLWFRGISRLEPTVVSLL
ACQ42061        ............................................................

04599       421 GFLSPGTAVLLGWLFLDQTLSALQIIGVLLVIGSIWLGQRSNRTPRARIACRKSP
ACQ42061        .......................................................
```

ZH-C2_04599 C-terminus:
EamA transporter



```
ZH-C2_04599 1   MVVEHAHGLRQRRGPNAKSSLHKAHLAPDAGLQAPGASLSFAQGSHDFESLDRGIGRGDR
WP_175065871    ............................................................

04599       61  FETAHRLDQYLELSVIGLDHVIEILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
WP_175065871    ............................................................

04599       121 FPVLQASQGLAQEPLRCLGAAGRRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
WP_175065871    ............................................................

04599       181 ATPPEPAQPLLHLRGVALDPAIDRRMVDRNAAFRQHFLKVAIADRIATIPAHRPQDHITL
WP_175065871    ............................................................

04599       241 EMAPLEIRHRSVRPISAKHAQASRFLQQSPFE.WVHVQLHQQKGMISLSPPTICNSARLM
WP_175065871 1  ......MRQKAASSLTLQQCLS.VY..RLPGGVAATVGAVQPLMVVFISAALLGSPIRLM

04599       300 AVLGAICGTAGVALLVLTPNAALDPVGVAAGLAGAVSMAFGTVLTRKWQPPVPLLTFTAW
WP_175065871 52 AVLGAICGTAGVALLVLTPNAALDPVGVAAGLAGAVSMAFGTVLTRKWQPPVPLLTFTAW

04599       360 QLAAGGLLLVPVALVFDPPIPMPTGTNVLGLAWLGLIGAGLTYFLWFRGISRLEPTVVSL
WP_175065871 112 QLAAGGLLLVPVALVFDPPIPMPTGTNVLGLAWLGLIGAGLTYFLWFRGISRLEPTVVSL

04599       420 LGFLSPGTAVLLGWLFLDQTLSALQIIGVLLVIGSIWLGQRSNRTPRARIACRKSP
WP_175065871 172 LGFLSPGTAVLLGWLFLDQTLSALQIIGVLLVIGSIWLGQRSNRTPRARIACRKSP
```

Fig. 6. Unique multi-domain ZH-C2_04599 with an N-terminal transposase and C-terminal EamA transporter aligns to single-domain *E. coli* proteins.