

Supplementary Information

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

Zhen Han¹, Peter S. Thuy-Boun¹, Wayne Pfeiffer², Vincent F. Vartabedian³, Ali Torkamani^{4,5}, John R. Teijaro³, Dennis W. Wolan^{1,5}

¹Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037

²San Diego Supercomputer Center, University of California, San Diego, La Jolla CA 92093

³Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037

⁴Scripps Research Translational Institute, The Scripps Research Institute, La Jolla, CA 92037

⁵Department 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.** Sia9N₃ 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.** Sia9N₃ 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.

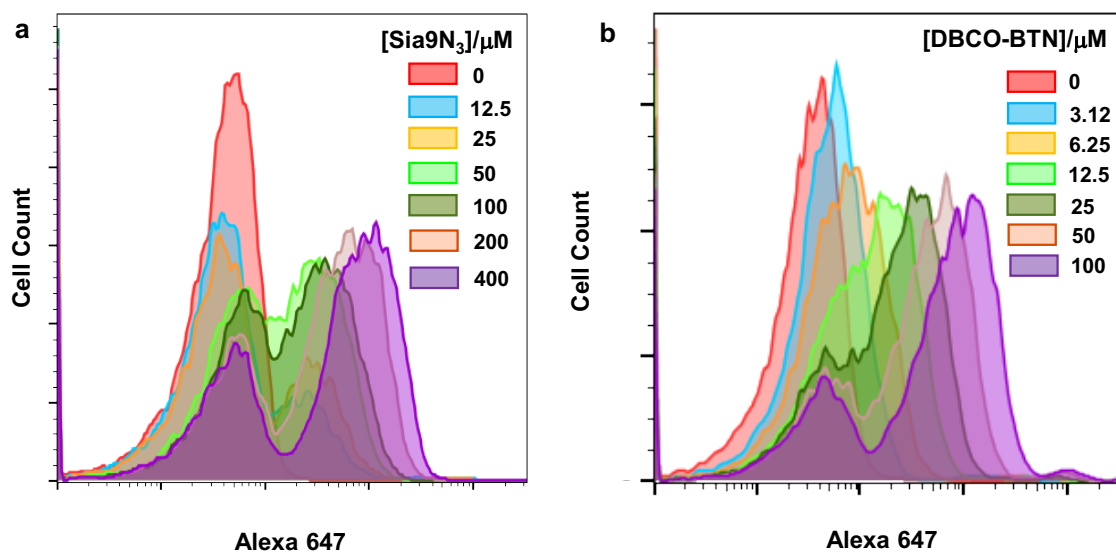


Fig. 1. Sia9N₃ (a) and DBCO-BTN (b) dose-dependent labeling of H2 cultured microbiome sample.

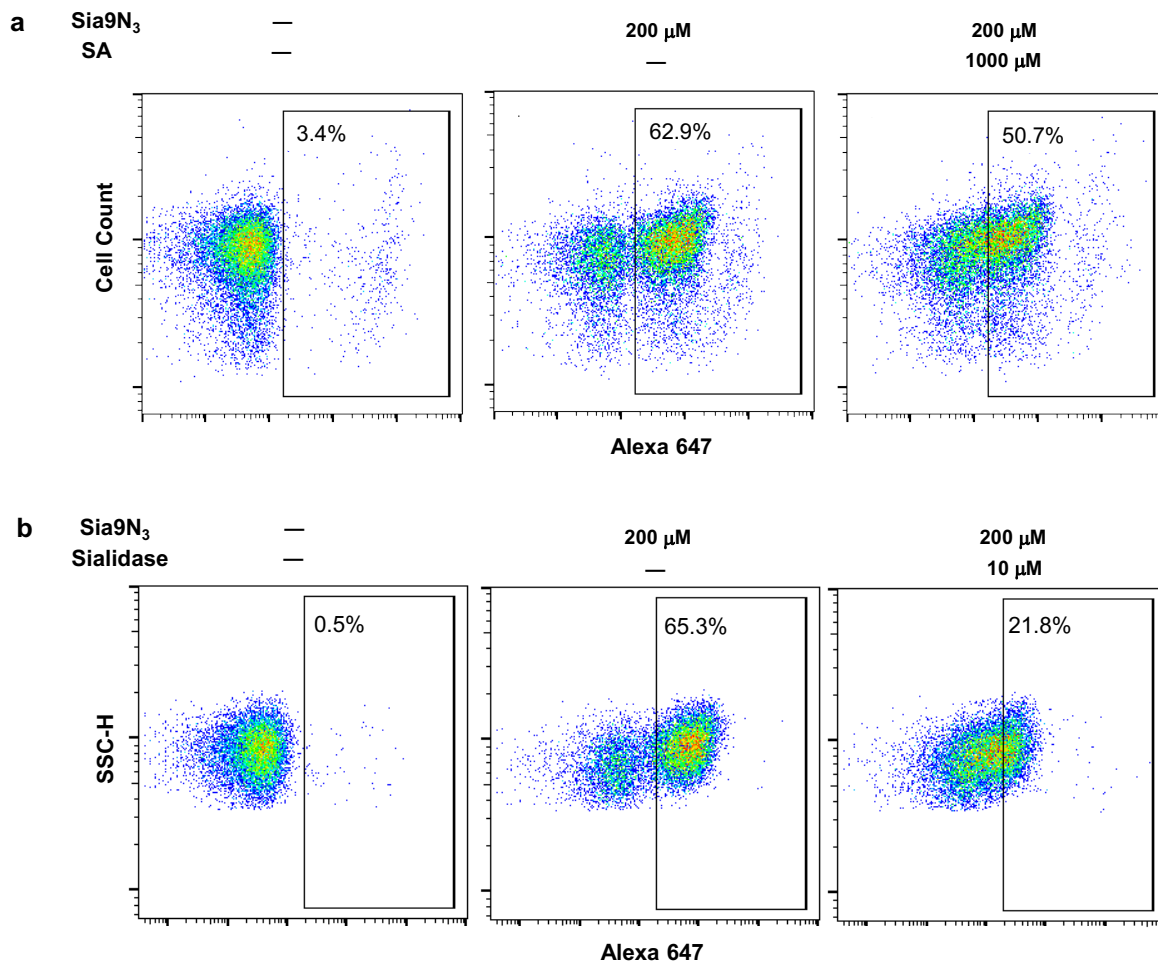


Fig. 2. a Sia9N₃ metabolic labeling of microbiome H2 competed with Neu5Ac. **b** Sia9N₃ metabolic labeling is partially removed by a microbial sialidase BT-0455.

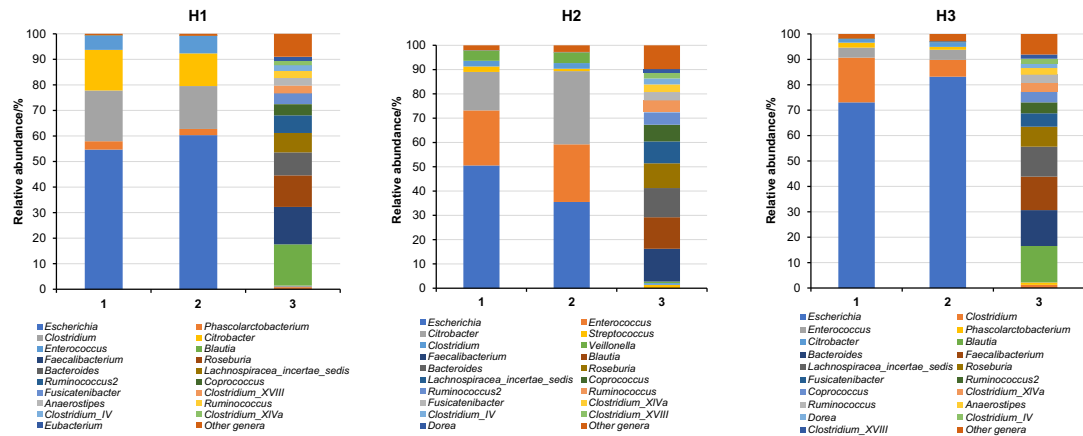


Fig. 3. 16S rRNA gene sequencing revealed taxonomic composition of cultured and primary microbiomes in the presence and absence of 200 μ M Sia9N₃. Cultured human microbiome samples H1-H3 were all tested, column 1: microbiome cultured without Sia9N₃, column 2: microbiome cultured with Sia9N₃, 3: uncultured fecal microbiome samples.

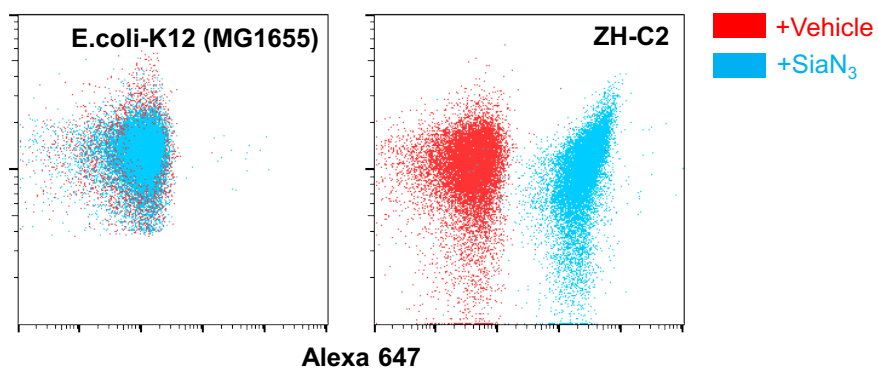


Fig. 4. Test of Sia9N₃ incorporation by *E. coli*-K12 (MG1655) and the newly isolated *E. coli* strain ZH-C2.

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

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

ZH-C2_04397
WP_087634241 1 MARLFTLSSEKTYLMLADAGTGRVAVFDLQSGQIAVGOAEMRRLAVDPVGSMEFPLN
1
1
04397 61 KHMQLACRCHQALNMAQIAPEYIAVAGSASHREGIVLTHNREGAPIMACANVDARAARAV
WP_087634241 61
61
04397 121 SELKELNHTFENEVYRATGQTLALSAIPRLMLAHRSDIYRQASTIMISDMLATHLS
WP_087634241 121
121
04397 181 GELAVDPHAGTGLDLDITRDWPKFALDMAGLRADLSPVKESTGLLVSSQAARLCC
WP_087634241 181
181
04397 241 LKAGTFFVVGQDVLGCLGVVFAQTAVLGGTFQGVVLAAPVDPFPHNVRVPHV
WP_087634241 241
241
04397 301 IPGNVQARSIPFFTLNTHMFRDAPCAEKLALRELGIDITLLEMAHVRVPPGSGVGNH
WP_087634241 301
301
04397 361 IPFDRNRKFTWYHAAPSINLSIDDPKCNKATLFRALENAAIVSACNLQIADFSNIFP
WP_087634241 361
361
04397 421 SELVFAGGSKQLMSQILADVSGLPVIVVKEATALGCATAAGVCAQIPSSHAGTGR
WP_087634241 421
421
04397 481 DVNERTSPPDPKRELGLAARGVEAADAGPITATKVEFLGEGVFPNHTNSDAGTQ
WP_087634241 481
481
04397 541 FAGVAQLKSTFLEDRDAAALTRAMLQSDSEHTSVKVGSSIGLANNQGITPDTDI
WP_087634241 541
541
04397 601 PAATLAEQIISVDLTVVAGDITMKGRNIVQNGQDVLDPKPHNPNMNPITLNL
WP_087634241 601
601
04397 661 EDDSSVGVVLKQAVTISGGELTLRLDGDVEADKTLIAQNGTVVAEGDYQRLIT
WP_087634241 661
661
04397 721 APGCLVYVTKKALNIGGKLTAEHGGATGATADMSAKTGGGLAINTVRSLSN
WP_087634241 721
721
04397 781 GQNDYQAVIYVQHTLRDADGALNTRELINSHAAIVLNGSTVETFTQHGSTVLF
WP_087634241 781
781
04397 841 KEGALTVNKGISQGLTGGQNLVNGVGTALTEGLNARYNALTSPNARELNDQGL
WP_087634241 841
841
04397 901 RQNIANDGLLTKPDLLEISDRSVAQVTLIPDETVTFINAMEKLEAELEVLTA
WP_087634241 901
901
04397 961 ADIFLGHIRENDRFVSAQAKTALLIGMDCIPKGIPTETIILKIPLOEGRPA
WP_087634241 961
961
04397 1021 LDLGQSDMEVYCLLAKELGLMVPDAEIKAGVRAVAVERPDRRNARTVLLLPD
WP_087634241 1021
1021
04397 1081 DHCQTFGLPSVKYTESDGGPIARINAFPLMGSSEALKDRYDFMKFQVFWLIGATD
WP_087634241 1081
1081
04397 1141 NFSVFIQAGSRYLTFPFDIISAPVVGQGTGHI
WP_087634241 1141
1141

ZH-C2_04397
WP_001405865 1 NBRITVYHWCTLQVFPACBELTRAQRQSEVTVLRSSGLTTFEELFLVGLLALSGSA
1
1
04397 61 GAELEVNDQITRIQTVATDQAVLQKQVTLVLAQGRSGLTITVTVVQARDEQET
WP_001405865 61
61
04397 121 VNVLGGTNRLYDGGNARPLVVGSGTGLTIKQKHVDGGLTLGSGTGGVTVNVEGE
WP_001405865 121
121
04397 181 MARLFTLSSEKTYLMLADAGTGRVAVFDLQSGQIAVGOAEMRRLAVDPVGSMEFPLN
WP_001405865 181
181
04397 237 DSVLTFEASSTGLALDAGTGRVAVFDLQSGQIAVGOAEMRRLAVDPVGSMEFPLN
WP_001405865 237
237
04397 274 TQIGLQVGGDQDGLTTRDWPKALDMAGLRADLSPVKESTGLLVSSQAARLCC
WP_001405865 274
274
04397 334 RLLLAHRSDIYRQASTITMISDMLATHLS...GTAADPSNAGTGLDITTRDKK
WP_001405865 334
334
04397 381 PALDQAGLRAQILDPVKESTGLLVSSQAARLCCGLKATPFFVGGDGLDITTRDKK
WP_001405865 381
381
04397 433 RPAQTLAQLTQVGVVLAAPVDPFPHNVRVPHV...GKVFVQDGLDAGRAGIG
WP_001405865 433
433
04397 485 TLNLMDATSPDVT...VGVNFGSIVVNSGATLSTGQVPIGQNSAKGKIVISTDLSL
WP_001405865 485
485
04397 545 FDRAPCAEKLALRELGIDITLLEMAHVRVPPGSGVGNH...LLEMA
WP_001405865 545
545
04397 605 SRKIPFGS...FSDRNRKFTWYHAAPSINLSIDDPKCNKATLFRALENAAIVSACNLQIADFSNIFP
WP_001405865 605
605
04397 661 RALACALQAD...PSSVPAAGGK...LMSGLDGLP...IPVVK
WP_001405865 661
661
04397 713 RALACALQAD...PSSVPAAGGK...LMSGLDGLP...IPVVK
WP_001405865 713
713
04397 769 ALKGGDLNRVQLSSSDKMPFTRATGTFAGVQLKDSFTPLERDNTAALTRAMLQSDS
WP_001405865 769
769
04397 829 NTSVVGQDSIGGLANNQGITPDTIPATLAEQIISVDLTVVAGDITMKGRNIVQ
WP_001405865 829
829
04397 889 GQDVLDPFPHNVRVPHV...LLEMAHVRVPPGSGVGNH...LLEMA
WP_001405865 889
889
04397 949 EYADKTLIAQNGTVVAEGDYQRLITAPGCLVYVTKKALNIGGKLTAEHGGAT
WP_001405865 949
949
04397 1009 GATADMSAKTGGGLAINTVRSLSNQGNDTQAGTIVQKGLTADAGLGNTRLENI
WP_001405865 1009
1009
04397 1049 RAAIVDNGSSQVETFTQNGSTVLFKKAALTVNKGQISQELTGGDLNVTGGLAI
WP_001405865 1049
1049
04397 1129 ROLNARTNALTSPHAEVSLDNTQGLGRGIANDDLLTKPDLLEISDRSVAQVTLIP
WP_001405865 1129
1129
04397 1189 RDETVTFINAMEKLEAELEVLTAADIPLQGHIRENDRFVSAQAKTALLIGMDCIP
WP_001405865 1189
1189
04397 1249 DNCIPGIIPTETIILKIPLOEGRPA...DGGVDRVITCLLAKRELWVPAEITKA
WP_001405865 1249
1249
04397 1309 QHVRALAVRFRDRNARTVLLRQSDHCQTFGLPSVKYTESDGGPIARINAFPLMGS
WP_001405865 1309
1309
04397 1369 SEALKDRYDFMKFQVFWLIGATDGHAKRFVFIQAGSRYLTFPFDIISAPVVGQGTGHI
WP_001405865 1369
1369
04397 1429 H
WP_001405865 1429
1429

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

ZH-C2_04397
E0V99460 1 MARLFTLSSEKTYLMLADAGTGRVAVFDLQSGQIAVGOAEMRRLAVDPVGSMEFPLN
1
1
04397 61 KHMQLACRCHQALNMAQIAPEYIAVAGSASHREGIVLTHNREGAPIMACANVDARAARAV
E0V99460 61
61
04397 121 SELKELNHTFENEVYRATGQTLALSAIPRLMLAHRSDIYRQASTIMISDMLATHLS
E0V99460 121
121
04397 181 GELAVDPHAGTGLDLDITRDWPKFALDMAGLRADLSPVKESTGLLVSSQAARLCC
E0V99460 181
181
04397 241 LKAGTFFVVGQDVLGCLGVVFAQTAVLGGTFQGVVLAAPVDPFPHNVRVPHV
E0V99460 241
241
04397 301 IPGNVQARSIPFFTLNTHMFRDAPCAEKLALRELGIDITLLEMAHVRVPPGSGVGNH
E0V99460 301
301
04397 361 IPFDRNRKFTWYHAAPSINLSIDDPKCNKATLFRALENAAIVSACNLQIADFSNIFP
E0V99460 361
361
04397 421 SELVFAGGSKQLMSQILADVSGLPVIVVKEATALGCATAAGVCAQIPSSHAGTGR
E0V99460 421
421
04397 481 DVNERTSPPDPKRELGLAARGVEAADAGPITATKVEFLGEGVFPNHTNSDAGTQ
E0V99460 481
481
04397 541 VDMILITDQDKGVHMDAGTVPNAGTISGKTLVNDGLLITASHTADQVQNGSSEVIT
E0V99460 541
541
04397 599LASTNSAGDITLNLKGGDLNRVQLSSSDKMPFTRATGTFAGVQLKDS
E0V99460 599
599
04397 661 RPTLEDRDNTAALTRAMLQSDSSEHTSVKVGSSIGLANNQGITPDTDIIPATLAEQI
E0V99460 661
661
04397 721 SVDLTVVAGDITMKGRNIVQNGQDVLDPKPHNPNMNPITLNLLEDDSSVGVQ
E0V99460 721
721
04397 781 LVRAQTVISGGSGLTLRLDGDVEADKTLIAQNGTVVAEGDYQRLITAPGCLVYV
E0V99460 781
781
04397 841 GLKALNIGGKLTAEHGGATGATADMSAKTGGGLAINTVRSLSNQGNDTQAGT
E0V99460 841
841
04397 901 VQHTLRDADGALGNTRLEINSHAAIVLNGSTVETFTQHGSTVLFKEGALTVNKG
E0V99460 901
901
04397 934 G...GQELTGGQNLVNGVGTALTEGLNARYNALTSPNARELNDQGLRGNANDGL
E0V99460 934
934
04397 971 TLKFDLLEISDRSVAQVTLIPDETVTFINAMEKLEAELEVLTAADIPLQGHIRE
E0V99460 971
971
04397 991 RDNFRVSAQAKTALLIGMDCIPKGIPTETIILKIPLOEGRPA...DGGVDRVITCLLAKRELWVPAEITKA
E0V99460 991
991
04397 1031 TYCLLAKELGLMVPDAEIKAGVRAVAVERPDRRNARTVLLRQSDHCQTFGLPS
E0V99460 1031
1031
04397 1091 SVKYESDGGPIARINAFPLMGSSEALKDRYDFMKFQVFWLIGATDGHAKRFVFIQAGG
E0V99460 1091
1091
04397 1151 SYRLTFPFDIISAPVVGQGTGHI
E0V99460 1151
1151

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: transposase (plasmid)

```

ZH-C2_04599 1 MVVEHAGLRQRGPNKSSLHKHLAPDAGLQAPGASLSFAQGSHPFESLDRGIGRGR
ACQ42061 1 MVVEHAGLRQRGPNKSSLHKHLAPDAGLQAPGASLSFAQGSHPFESLDRGIGRGR
04599 61 FETAHRLDQYLELSVIGLDHVEIILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
ACQ42061 61 FETAHRLDQYLELSVIGLDHVEIILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
04599 121 FVVLQASQGLAQEPLRCLGAAGRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
ACQ42061 121 FVVLQASQGLAQEPLRCLGAAGRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
04599 181 ATPPEPAQPLLHLRGVALDPAIDRRMVDNRNAAFRQHFLKVAIADRIATIPARRPQDHITL
ACQ42061 181 ATPPEPAQPLLHLRGVALDPAIDRRMVDNRNAAFRQHFLKVAIADRIATIPARRPQDHITL
04599 241 EMAPLEIRHRSVRPISAKHAQASRFLOQSPPEVHVQLHQKGMISLSPPTICNSARLMA
ACQ42061 241 EMAPLEIRHRSVRPISAKHAQASRFLOQSPPEVHVQLHQKGMISLSPPTICNSARLMA
04599 301 VLGAICGTAGVALLVLTNPAALDPVGAAGLAGAVSNAPGTVLTRKMQPPVPLLTFTAW
ACQ42061 301 VLGAICGTAGVALLVLTNPAALDPVGAAGLAGAVSNAPGTVLTRKMQPPVPLLTFTAW
04599 361 LAAGGLLVPVALVDFDPPIMPPTGTVLGLANLGLIGAGLTYFLNFRGISRLEPTVVSL
ACQ42061 361 LAAGGLLVPVALVDFDPPIMPPTGTVLGLANLGLIGAGLTYFLNFRGISRLEPTVVSL
04599 421 GFLSPGTAVLLGWFLDQTLQIIGVLLVIGSINLWQSRNRTPRARIACRSK
ACQ42061 421 GFLSPGTAVLLGWFLDQTLQIIGVLLVIGSINLWQSRNRTPRARIACRSK
    
```

ZH-C2_04599 C-terminus: EamA transporter

```

ZH-C2_04599 1 MVVEHAGLRQRGPNKSSLHKHLAPDAGLQAPGASLSFAQGSHPFESLDRGIGRGR
WP_175065871 1 MVVEHAGLRQRGPNKSSLHKHLAPDAGLQAPGASLSFAQGSHPFESLDRGIGRGR
04599 61 FETAHRLDQYLELSVIGLDHVEIILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
WP_175065871 61 FETAHRLDQYLELSVIGLDHVEIILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRL
04599 121 FVVLQASQGLAQEPLRCLGAAGRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
WP_175065871 121 FVVLQASQGLAQEPLRCLGAAGRQVEIDRVAPLVDCPVQVGPLAPHLDVGFIQAPARIK
04599 181 ATPPEPAQPLLHLRGVALDPAIDRRMVDNRNAAFRQHFLKVAIADRIATIPARRPQDHITL
WP_175065871 181 ATPPEPAQPLLHLRGVALDPAIDRRMVDNRNAAFRQHFLKVAIADRIATIPARRPQDHITL
04599 241 EMAPLEIRHRSVRPISAKHAQASRFLOQSPPEVHVQLHQKGMISLSPPTICNSARLMA
WP_175065871 241 EMAPLEIRHRSVRPISAKHAQASRFLOQSPPEVHVQLHQKGMISLSPPTICNSARLMA
04599 300 AVLGAICGTAGVALLVLTNPAALDPVGAAGLAGAVSNAPGTVLTRKMQPPVPLLTFTAW
WP_175065871 52 AVLGAICGTAGVALLVLTNPAALDPVGAAGLAGAVSNAPGTVLTRKMQPPVPLLTFTAW
04599 360 QLAAGGLLVPVALVDFDPPIMPPTGTVLGLANLGLIGAGLTYFLNFRGISRLEPTVVSL
WP_175065871 112 QLAAGGLLVPVALVDFDPPIMPPTGTVLGLANLGLIGAGLTYFLNFRGISRLEPTVVSL
04599 420 LGFLSPGTAVLLGWFLDQTLQIIGVLLVIGSINLWQSRNRTPRARIACRSK
WP_175065871 172 LGFLSPGTAVLLGWFLDQTLQIIGVLLVIGSINLWQSRNRTPRARIACRSK
    
```

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.