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.	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 <i>E. coli</i> -K12 (MG1655) and the newly isolated <i>E. coli</i> strain.		
Fig. 5.	Sequence alignment of ZH-C2_04397 to other <i>E. coli</i> proteins.		
Fig. 6.	Sequence alignment of ZH-C2_04599 to other <i>E. coli</i> proteins.		
Supp. Table 1 7 tabs			

Supplementary Table 1 legend: Additional bioinformatic data analyses of *E. coli* genome composition. Related to Figure 6. <u>Tab 1</u>: *E. coli* genome assembly statistics and comparisons to reference strains. <u>Tab 2</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to K-12 MG1655. <u>Tab 3</u>: Biological process, molecular function, and cellular component GO Terms associated with unique *E. coli* genomes. <u>Tab 4</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to UMEA 3174-1. <u>Tab 5</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* ZH-C2 genes relative to VR50. <u>Tab 6</u>: List of low-homology and unique *E. coli* 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 μM 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 WP_087634241	1	NARLFTLSESKYYLNALDAGTGSIRAVIPDLEGNQIAVGQAEWRHLAVPDVPGSMEFDLN MARLFTLSESKYYLNALDAGTGSIRAVIPDLEGNQIAVGQAEWRHLAVPDVPGSMEPDLN
04397 WP_087634241	61 61	KNWQLACECHKQALHNAGIAPEYIAAVSACSKREGIVLYNNEGAPIWACANVDARAAREV KNWQLACECHKQALHNAGIAPEYIAAVSACSKREGIVLYNNEGAPIWACANVDARAAREV
04397 WP_087634241	121 121	SELKELHNNTFENEVYRATGOTLALSA IPRLLWLAHHRSDIYRQASTITNISDWLAYNLS SELKELHNNTFENEVYRATGOTLALSA IPRLLWLAHHRSDIYRQASTITNISDWLAYNLS
04397 WP_087634241	181 181	GELAVDPSNAGTTGLLDLTTRDWKPALLDMAGLRADILSPVKETGTLLGVVSSQAAELCG GELAVDPSNAGTTGLLDLTTRDWKPALLDMAGLRADILSPVKETGTLLGVVSSQAAELCG
04397 WP_087634241	241 241	LKAGTPVVVGGGDVQLGCLGLGVVRPAQTAVLGGTPKQQVVNLAAPVTDPENNVRVNPEV LKAGTPVVVGGGDVQLGCLGLGVVRPAQTAVLGGTPKQQVVNLAAPVTDPENNVRVNPEV
04397 WP_087634241	301 301	I PGNVQAESISPFTGLTMRNPRDAPCAEEXLIAERLGIDTYTLLEENASRVPPGSMGVNP IPGNVQAESISPFTGLTMRNPRDAFCAEEXLIAERLGIDTYTLLEENASRVPPGSMGVNP
04397 WP_087634241	361 361	IFSDRMRFRTWYHAAPSFINLSIDPDKCNKATLFRALEENAAIVSACNLOOIADFSNIHF IFSDRMRFRTWYHAAPSFINLSIDPDKCNKATLFRALEENAAIVSACNLOOIADFSNIHF
04397 WP_087634241	421 421	SSLVPAGGGSKGKLWSGILADVSGLPVNIPVVKEATALGCAIAAGVGAGIPSSHAETGER SSLVPAGGGSKGKLWSGILADVSGLPVNIPVVKEATALGCAIAAGVGAGIPSSHAETGER
04397 WP_087634241	481 481	LVRNERTHTPDPEKHELILAGTNSAGDYTLTNALKGDGIN <mark>RVOM</mark> SSSDKNP <mark>O</mark> FTHATGTE LVRNERTHTPDPEKHELVODGRDKNQAIRRAVMARLSGKRV.ASG
04397 WP_087634241	541	FAGVAQLKDSTFTLERDNTAALTHANLQSDSENTTSVKVGEQSIGGLANNGGTIIFDTDI
04397 WP_087634241	601	PAATLAEGYISVDTLVVGAGDYTWKGRNYQVNGTGDVLIDVPKPWNDPMANNPLTTLNLL
04397 WP_087634241	661	EHDDSHVGVQLVKAQTVIGSGGSLTLRDLQGDEVEADKTLHIAQNGTVVAEGDIGPRLTT
04397 WP_087634241	721	APGNGLYVNYGLKALNIHGGQKLTLAEHGGAYGATADHSAKIGGEGDLAINTVRQVSLSN
04397 WP_087634241	781	GQNDYQGATYVQNGTLRTDADGALGNTRELNISNAAIVDLNGSTQTVETPTGQNGSTVLP
04397 WP_087634241	841	KEGALTVNKGGI SOGELTGGGNLNVTGGTLA I EGLNARVNALTS I SPNAEVSLDNTOGLG
04397 WP_087634241	901	RGNIANDGLLTLKFDLLSEIGRDSVGAVTLIPEDETVTHPIMAWEKLTEARLEEVLTAYK
04397 WP_087634241	961	ADIPLGMIR BENDFRISVA Ö A QEKTALLR I GNDWCIPKG I TPTTHIIKL PIGEIR QPNAT
04397 WP_087634241	1021	LDLSQSVDNEYYCLLLAKELGLNVPDAETIKAGNVRALAVERFDRRWNAERTVLLRLPQE
04397 WP_087634241	1081	DMCQTFGLPSSVKYESDGGPGIARIMAPLMGSSEALKDRYDPMKFQVFQMLIGATDGHAK
04397 WP_087634241	1141	NFSVFIQAGGSYRLTPFYDIISAFPVLGGTGINI

ZH-C2_04397 N-terminus:

autoinducer-2-kinase LsrK *E. coli* KTE98

ZH-C2_04397 1	NARLFT <mark>U</mark> SESKYYLNALDAGTGSTRAVIFDLEGNQIAVGOAEWRHLAVDDVPGSMEFDL
E0V99460 1	NARLFT <mark>U</mark> SESKYYLNALDAGTGSIRAVIFDLEGNQIAVGOAEWRHLAVDDVPGSMEFDLE
04397 61	KNWQ LAC BOM <mark>R</mark> QALIN NAGTAPETTAR VSACSMREGI VLYN NEG <mark>A</mark> PIWACAN V DARAR EV
E0V99460 61	KNWQ LAC BOM <mark>RQALIN NAGTAPETTAR VSACSMREGI VLYN NEG</mark> PIWACAN V DARAR EV
04397 121	SELKELHNNTFENEVYRATGOTLALSAIPRLLWLAHHRSDIYRQASTITHISDWLAYMLS
E0V99460 121	SELKELHNNTFENEVYRATGOTLALSAIPRLLWLAHHRSDIYRQASTITHISDWLAYMLS
04397 181	GELAVDPSNAGTTGLLDLTTRDWKPALLDNAGLRADILSPVKETGTLLGVVSSQAAELCG
E0V99460 181	Gelavdpsnagtglldlttrdwkpalldnaglradilspvketgtllgvvssqaaelcg
04397 241	LKAGTPVVVGGGDVQLGCLGLGVVRÞAQTAVLGGTPHQQVVNLAAPVTDPENNVRVNPEV
E0V99460 241	LKAGTPVVVGGGDVQLGCLGLGVVRÞAQTAVLGGTPHQQVVNLAAPVTDPENNVRVNPEV
04397 301	I PGHVQAESISFFTGLTHRWFRDAFCAEEKLIAERLGIDTYTLLEENASRVFPGSWGVMF
E0V99460 301	I PGHVQAESISFFTGLTHRWFRDAFCAEEKLIAERLGIDTITLLEENASRVFPGSWGVNF
04397 361	IF SORMRFXTWY NAAP SFILLS ID POKCHKATLFRALEE NAAI VSACHLOOIAD FSWINF
E0V99460 361	IF SORMRFXTWY NAAP SFINLS ID POKCHKATLFRALEE NAAI VSACHLOOIAD FSWINF
04397 421	SELVFAGGGSKCKLMSQILADVSGLPVNIPVVKEATALGCAIAAGVGAGIFSS <mark>M</mark> AETGER
E0V99460 421	SELVFAGGGSKCKLMSQILADVSGLPVNIPVVKEATALGCAIAAGVGAGIFSS <mark>M</mark> AETGER
04397 481 E0V99460 481	LURWERTHPPDPEXHELT. LURWERTHPPDPEXHELT GAAHGEVAADAGFITNATKVEFGLGEGVPVPNHTNNSDAGYQ
04397 E0V99460 541	VDMLITGDDKDGKVMHDAGHTVPNAGNTYSGKTLVNDGLLTIASHTADGVTGMGSSEVTI
04397 499 E0V99460 601	
04397 551	TFTLERDNTAALTHANLOSDSENTTSVKVGEOSIGGLANNGGTIFDTDIFAATLAEGTI
E0V99460 661	TPTLERDNTAALTHANLOSDSENTTSVKVGEOSIGGLANNGGTIFDTDIFAATLAEGTI
04397 611	SVDILVVGAGDITHKGRNYQVAGIGVVLIDVFKPHHDPHANNPLITLNLLEHDDSHVGVQ
E0V99460 721	SVDILVVGAGDITHKGRNYQVNGIGDVLIDVFKPHHDPHANNPLITLNLLEHDDSHVGVQ
04397 671	LVKAQTVIGSGGSLTIRDLQGDEVEADKTLHIAQNOTVVASGDYGFRLTTAPO <mark>H</mark> GLYVNY
E0V99460 781	LVKAQTVIGSGGSLTLRDLQGDEVEADKTLHIAQNGTVVASGDYGFRLTTAPO <mark>H</mark> GLYVNY
04397 731	GLKALNINGQÇKLTLARNGGAYGAYADNSAKIGGRGDLAINYYRQYSLSNGQNDYQGAYY
E0V99460 841	GLKALNINGQKLTLARNGGAYGAYADNSAKIGGRGDLAINYYRQYSLSNGQNDYQGAYY
04397 791	VONGLERDADCALCHTRELNESHAAIVOLNOGENOTVETTIGANGSTVEFKEGALTVNKG
E0V99460 901	VONGLERDADCASKLGLISTRWLKGRÖM <mark>E</mark> RK
04397 851 E0V99460 934	GISQCELTGGGHLNVTGGTLAIEGLMARYNALTSISPNAEVSLDNTQGLCRGNIANDGLL
04397 911 E0V99460	TLKPOLLSEIGRDSVGAVTLIPEDETVTHPINAWEKLTEARLEEVLTAYKADIPLGMIRE
04397 971 E0V99460	ENDFRISVAGAQEKTALLRIGNDWCIPKGITPTTHIIKLPIGEIRQPNATLDLSQSVDNE
04397 1031 EOV99460	YYCLLLAKELGLNYPDAEIIKAGNYKALAVERPDRRWNAERTVLLRLPOEDHCOTFGLPS
04397 1091 E0V99460	SVKYESDGGPGIARIMAPLNGSSEALKDRYDPNKFQVFQWLIGATDGRAKNFSVFIQAGG
04397 1151 E0V99460	SYRLTPFYDIISAFPVLGGTGIHI

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

	W	ith C-terminal HIPA Sen i nr kinase
ZH-C2_0439 WP_001405865	1	MNRIYRVIWNCTLQVFQACSELTRRAGKTSTVNLRKSSGLTTKFSRLTLGVLLALSGSAS
04397 WP_001405865	61	GASLEVDNDQITNIDTDVAYDAYLVGWYGTGVLNILAGGNASLTTITTSVIGANEDSEGT
04397 WP_001405865	121	VNVLGGTWRLYDSGNNARPLNVGQSGTGTLNIKQKGNVDGGYLRLGSSTGGVGTVNVEGE
04397	1	MARTTESESKYYLMADAGTOSIRAVIPDEGWQTAVGQATTRHLAVP
WP_001405865	181	DSVLTIS <mark>H</mark> SIGSYGTGS <u>INI</u> .TDKGYYTSS <mark>I</mark> VAILGYQAGSN <u>Q</u> VV <u>V</u> FRGG <mark>YU</mark> LIK
04397 WP_001405865	50 237	DYPOTHTTDENKNWQLACECHRQALHNATIAPEYIAAVSAC <mark>SHITEO</mark> IVLYNNE
04397 WP_001405865	107 274	NACANVDARAAREVSELEENNINTEEE
04397 WP_001405865	150 334	ŘLLULAHHRSDIYRQASTITNISDNLAVIKLS
04397	205	PALLDWAGLRADILEPWKETGILGUVSSQAAGLCGLKAGTPUVVGGDWQLGCLGLGUV
WP_001405865	381	TGNITYKOKNEVUTNLGINNGYDGRGMNISNQGLVUSNGSSLGVGETGVG
04397	265	RPAQTAVLCGTFNQQVVNLAAPVTDPEWIVRVNPHVIPQNVQAEBFSPT
WP_001405865	433	WVSITTGGMEVNKUVITIGVAGVGNLUISDGGKPVSQNTTLGDKASGIG
04397 WP_001405865	315 485	ULT
04397 WP_001405865	321 545	FRDAFCAEEKTI ÄRR <mark>IG</mark>
04397	349	SRUFPGSWOVRPEPSDRWRFRYWWRANDSPENDSCONKATURALER
WP_001405865	605	MYNGTSGTGTUTUTUNNGTLNVEGER.UCVPEPNVGTUTGAAHGEAAADAG
04397	401	NAIMSACNLOOTADISNIHPSSIVFAGGSKUTLUSOTLADISGLPVNTFVVKE
WP_001405865	661	NTKNEFGLGEGVFVNNHINNSDAGYQVDNLITODKDGTVIHDAGHTVFNAG
04397	455	AMALGCAIAAGUGAGIPSSMAENGCRUVRWERTHIPDZEKHEL <mark>TLASINGAGDYTLIN</mark>
WP_001405865	713	NMISGKIL <mark>U</mark> NDGLLTIASHMADOVICNGSSEVIIAN <mark>D</mark> .GILD <mark>ILASINGAGDYTLIN</mark>
04397	513	ALKGDGLMRVQLSSSDKMFGPTHATGTEFAGVAQLKDSTFTLERDNTAALTHAMLQSDSE
WP_001405865	769	Alkgdglmrvqlsssdkmfgpthatgtefagvaqlkdstftlerdntaalthamlqsdse
04397	573	NTTSVKVGEQSIGGLANNGGTIIPDTDIPAATLAEGYISVDTLVVGAGDYTNKGRNYQVN
WP_001405865	829	NTTSVKVGEQSIGGLANNGGTIIPDTDIPAATLAEGYISVDTLVVGAGDYTNKGRNYQVN
04397	633	GTGDVLIDVPKPNNDPMANNPLTINLENDDSNVGVQLVKAQTVIGSGGSLTLRDLQGD
WP_001405865	889	GTGDVLIDVPKPNNDPMANNPLTINLENDDSNVGVQLVKAQTVIGSGGSLTLRDLQGD
04397	693	EVERDKTLHIRQNGTVVAEGDYGPRLTTAPGNGLYVNYGLKALNTHGGOKLTLAEHGGAY
WP_001405865	949	EVERDKTLHIRQNGTVVAEGDYGPRLTTAPGNGLYVNYGLKALNIHGGOKLTLAEHGGAY
04397	753	GATADMSAKIGGEGDLAINTVRQVSLSNGQNDYQGATYVQNGTLRTDADGALGNTRELNI
WP_001405865	1009	GATADMSAKIGGEGDLAINTVRQVSLSNGQNDYQGATYVQNGTLRTDADGALGNTRELNI
04397	813	SNAAIVDLNGSTOTVETFTGONGSTVLFKEGALTVNKGGISOGELTGGGNLNVTGGTLAI
WP_001405865	1069	SNAAIVDLNGSTOTVETFTGONGSTVLFKEGALTVNKGGISOGELTGGGNLNVTGGTLAI
04397	873	EGLMARYMALTSISPMAEVSLDWTOGLGRONIANDOLLTLKFDLLSEIGRDSVGAVTLIP
WP_001405865	1129	EGLMARYMALTSISPMAEVSLDWTOGLGRONIANDGLLTLKFDLLSEIGRDSVGAVTLIP
04397	933	EDETYTHPINANSKLTEARLEEVLTAYKADIPLGNIREENDPRISVAGAQEKTALLRIGN
WP_001405865	1189	EDETYTHPINANSKLTEARLEEVLTAYKADIPLGNIREENDPRISVAGAQEKTALLRIGN
04397	993	DWCIPKGITPTTHIIKLPIGEIROPNATLDLBOSVDNEYYCLLLAKELGLNVPDAEIIKA
WP_001405865	1249	DWCIPKGITPTTHIIKLPIGEIROPNATLDLBOSVDNEYYCLLLAKELGLNVPDAEIIKA
04397	1053	GN VRALAVERPDRRWNAERTVLLRLPQEDMCQTPGLPSSVKYESDGGGGIARINAPLMGS
WP_001405865	1309	GN VRALAVERPDRRWNAERTVLLRLPQEDMCQTPGLPSSVKYESDGGGGIARINAPLMGB
04397	1113	SEALKDRYDPHKFQVFQHLIGATDGHAKNPSVPIQAGGSYRLTPPYDIISAPPVLGGTGI
WP_001405865	1369	SEALKDRYDPHKFQVFQHLIGATDGHAKNPSVPIQAGGSYRLTPPYDIISAPPVLGGTGI
04397 WP_001405865	1173 1429	H

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 045	5991	MVVEHAHGLRORRGPNAKSSLHKAHLAPDAGLOAPGASLSFAOGSHDFESLDRGIGRGD	2
AC042061	1	MUVERANGLRORRGPNAKSSI.HKAHLAPDAGLOAPGASI.SFAOGSHDFESLDRGIGRGDI	
negaroor	-		•
		· · · · · · · · · · · · · · · · · · ·	•
04599	61	FETAHRLDQYLELSVIGLDHVIEILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRI	
ACQ42061	61	FETAHRLDQYLELSVIGLDHVIEILHLPVGRLPVQLSFALQFGDRCTIARRFVGIERGRI	
			•
04599	121	FPVLOASOGLAOEPLRCLGAAGRROVEIDRVAPLVDCPVOVGPLAPHLDVGFIOAPARI	4
AC042061	121	FPVLOASOGLAOEPLECLGAAGEROVETDEVAPLVDCPVOVGPLAPHLDVGFTOAPAET	
meganeer		· · · »Eue Seeufer exception (* * * * * * * * * * * * * * * * * * *	
04500	101	S MODED S OD T UT D CUST D D S T D D D VUD D VS S D D O U D T S D T S D T S D T D S UD D O D U T D V	
04399	101	A TPPEPA QPLLALKGVALDPAIDKKMVDKNAAFKQAFLKVA TADKTATTPAAKPQDATT	
ACQ42061	181	ATPPEPAQPLLHLRGVALDPAIDRRMVDRNAAFRQHFLKVAIADRIATIPAHRPQDHITI	
		<u> </u>	•
04599	241	EMAPLEIRHRSVRPISAKHAQASRFLQQSPFEWVHVQLHQQKGMISLSPPTICNSARLM	A
AC042061	241	EMAPLEIRHRSVRPISAKHAOASRFLOOSP <mark>H</mark> EN	
04500	201	UT CAT COMACUALT UT MENAAT DEVCUAACT ACAUCHAECOUT MEVNOEDUBT T MEMAN	
04399	301	VLGAICGIAGVALLVLIPNAALDPVGVAAGLAGAVSMAFGIVLIKK#QPPVFLLIPIA#	2
ACQ42061			•
 • • • • • • • • • • • • • • • • • • •	0.00		•
04599	361	LAAGGLLLVPVALVFDPPIPMPTGTNVLGLAWLGLIGAGLTYFLWFRGISRLEPTVVSL	4
ACQ42061			
-			
04599	421	GFLSPGTAVLLGWLFLDOTLSALOIIGVLLVIGSIWLGORSNRTPRARIACRKSP	
10042061		All and a second s	
ACQ42061			

ZH-C2_04599 C-terminus: EamA transporter

			•	• • •
ZH-C2_04599 WP 175065871	9 1	MVVEHAHGLRQRRGPNAKSSI	HKAHLAPDAGLQAPGA	SLSFAQGSHDFESLDRGIGRGDR
04599	61			FALOFGDRCTTARREVGTERGRI.
WP_175065871				
04599	121	FPVLQASQGLAQEPLRCLGA	GRRQVEIDRVAPLVDC	PVQVGPLAPHLDVGFIQAPARIK
WP_1/50658/1				
04599	181	ATPPEPAQPLLHLRGVALDP	IDRRMVDRNAAFROHF	LKVAIADRIATIPAHRPQDHITL
WP_175065871				
WP_175065871	1	MRQKAASSLTLQQCI	S.VYRLPGGVAAT	GAVQPLMVVFISAALLGSPIRLM
04599	300	AVLGAICGTAGVALLVLTPNA	ALDPVGVAAGLAGAVS	MAFGTVLTRKWOPPVPLLTFTAW
WP_175065871	52	AVLGAICGTAGVALLVLTPN#	ALDPVGVAAGLAGAVS	MAFGTVLTRKWQPPVPLLTFTAW
04599	360	QLAAGGLLLVPVALVFDPPI	MPTGTNVLGLAWLGLI	GAGLTYFLWFRGISRLEPTVVSL
mr_1/30030/1			AFTOIN COLAN LOLI	
04599 WP_175065871	420 172	LGFLSPGTAVLLGWLFLDQTI LGFLSPGTAVLLGWLFLDQTI	SALQIIGVLLVIGSIW SALQIIGVLLVIGSIW	ILGQRSNRTPRARIACRKSP ILGQRSNRTPRARIACRKSP

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.