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

Zhen Han1, Peter S. Thuy-Boun1, Wayne Pfeiffer2, Vincent F. Vartabedian1, Ali Torkamani1,4,5, John R. Teijaro3 & Dennis W. Wolan1,5*

N-Acetylneuraminic acid is the most abundant sialic acid (SA) in humans and is expressed as the terminal sugar on intestinal mucus glycans. Several pathogenic bacteria harvest and display host SA on their own surfaces to evade Siglec-mediated host immunity. While previous studies have identified bacterial enzymes associated with SA catabolism, no reported methods permit the selective labeling, tracking, and quantitation of SA-presenting microbes within complex multi-microbial systems. We combined metabolic labeling, click chemistry, 16S rRNA gene, and whole-genome sequencing to track and identify SA-presenting microbes from a cultured human fecal microbiome. We isolated a new strain of Escherichia coli that incorporates SA onto its own surface and encodes for the nanT, neuA, and neuS genes necessary for harvesting and presenting SA. Our method is applicable to the identification of SA-presenting bacteria from human, animal, and environmental microbiomes, as well as providing an entry point for the investigation of surface-expressed SA-associated structures.

Several methodologies and model systems have been critical to understanding gut microbiome taxonomic composition and the influence these compositions exert on host physiology and intestinal disease pathology1–3. Metagenomics and metabolomics as well as germ-free and gnotobiotic mouse models have helped establish how specific microbes and small-molecule metabolites modulate host responses to a variety of diseases4–7. These methods, in combination with the genetic manipulation of specific bacterial strains, have revealed key catabolizing enzymes and pathogen-associated molecular patterns (PAMPs) associated with microbiome-related metabolism, host immune activation, and bacterial infection8–10. Despite these methods and models, the ability to selectively label and track bacteria with specific catabolic functionalities in a highly complex and metabolically active microbiome remains limited.

N-Acetylneuraminic acid (Neu5Ac, termed sialic acid/SA here) is one of over thirty known sialic acid analogs. It is the predominant form of sialic acid in humans and is presented as the terminal residue on surface-exposed glycans, glycoproteins, and glycolipids11. SA is recognized by immunoinhibitory sialic acid-binding immunoglobulin-like lectins (Siglecs) to prevent autoimmunity12. Interestingly, select human pathogenic microbes have evolved to express this human SA epitope on their own cellular surface to evade host immune surveillance and clearance13–15. For example, group B streptococcus (GBS), a common cause of sepsis in human newborns, presents terminal α2,3-linked SA on its capsular polysaccharide (CPS) to bind Siglecs expressed by neutrophils, macrophages, and platelets and block immune activation16,17. Sequencing analysis revealed that GBS uses a tripartite transporter to translocate environmental SA onto its CPS to augment this immune evasion18. We posited that certain gut commensal organisms likely apply similar SA-mediated protective mechanisms to gain survival advantages. As such, the selective labeling and identification of these microbes from complex microbiome samples may reveal new SA-regulated host-microbe crosstalk mechanisms that could potentially be exploited for therapeutic development aimed at gut microbiome-related diseases.

Here, we report the application of an SA-based azide-containing probe N-acetyl-9-azido-9-deoxy-neuraminic acid (Neu5Ac9N3/Sia9N3), combined with flow cytometry and metagenomic and whole-genome sequencing.
to selectively label, isolate, and identify SA-presenting bacteria from a complex cultured human fecal microbiome sample. Broadly, applications of metabolic probe labeling followed by click chemistry-tagging for visualization have been used to image and track pre-labeled bacterial components, including peptidoglycans and lipopolysaccharides19–21. However, there is limited use of SA-based probes to label bacterial glycans, especially in a highly complex microbial sample. With respect to eukaryotic labeling, “clickable” SA-based probes and its metabolic precursor (i.e., sialic acid-alkyne, sialic acid-azide, and mannosamine-alkyne) have been widely used for the study of dialglycans in both in vitro and in vivo systems22–26. We sought to expand the application of SA-based metabolic probes to selectively label SA-presenting bacteria from a distal gut microbiome. Fecal microbiome samples collected from a healthy human volunteer were cultivated in the presence of the Sia9N3 metabolic probe27 and bacterial incorporation of Sia9N3 was examined by flow cytometry analysis following copper-free azide coupling with azadibenzocyclooctyne-conjugated biotin (ADIBO-BTN/DBCO-BTN) and streptavidin-Alexa 647 staining28. We identified a fecal microbiome sample from a healthy human donor containing bacteria that can readily incorporate Sia9N3. With fluorescence activated cell sorting (FACS) and 16S rRNA gene sequencing analyses, we found that the Sia9N3-incorporating bacteria belonged to the Escherichia genus. Isolation of the SA-incorporating bacteria via colony screening with whole-genome sequencing analysis suggested that the isolated bacteria is a new strain of E. coli that likely employs the NanT-NeuA-NeuS system to integrate environmental sialic acid to form polysialic acid on its capsular polysaccharide structure29,30.

**Results**

**FACS-based screening helps identify microbiome constituents capable of Sia9N3 incorporation.** We surveyed fecal samples collected from one healthy human volunteer over the course of three months. Fecal microbes were cultured in a Gifu media31 supplemented with Sia9N3 at a physiologically relevant concentration (200 μM)32 under anaerobic atmosphere for 20 h. Microbes were centrifuged, extensively washed, subjected to DBCO-BTN conjugation with streptavidin-Alexa 647, and examined for fluorophore incorporation with flow cytometry analysis (Fig. 1). Among three microbiota samples tested (designated H1, H2, and H3), more than 60% of the bacterial cells from one of the cultured human fecal samples (e.g., H2) readily displayed fluorescence in response to Sia9N3 treatment (Fig. 1). Furthermore, a dose-dependent assay revealed that bacterial fluorescence labeling is dependent on concentrations of both Sia9N3 and DBCO-BTN. Use of 200 μM Sia9N3 in culture media with 50 μM DBCO-BTN for azide conjugation resulted in the maximum percentage of labeled bacterial cells (Fig. 2a,b). Further increases in the concentration of either labeling reagent only amplified the fluorescence intensity on the labeled bacterial cells without increasing background fluorescence or numbers of labeled cells, which suggested that Sia9N3 selectively labeled a subset of SA-catabolizing bacteria (Supplementary Fig. 1).

We next assessed if Neu5Ac would compete with Sia9N3 for surface expression by bacteria when supplemented into the 20 h cultures. Addition of excess Neu5Ac resulted in reduced labeling of bacteria by Sia9N3 (Fig. 2c; Supplementary Fig. 2a), which suggested that bacteria may use the same or partially overlapped transporters and enzymes for SA and Sia9N3 transportation and catabolism. Labeled bacteria were also subjected
Bacterial surface Sia9N₃ is removed by microbial sialidases. Many gut bacteria can express sialidases to release SA from sialoglycans, providing themselves or other bacteria with a source of carbon, nitrogen, and building blocks for cell wall biosynthesis. To determine if the Sia9N₃ probe is presented on bacteria to form sialoglycans, we treated the metabolically labeled H₂ cultured sample with a purified recombinant sialidase (BT0455, NCBI gene ID: 1071627) from Bacteroides thetaiotaomicron, followed by sequential DBCO-BTN conjugation, streptavidin-Alexa 647 staining, and flow cytometry. Sialidase treatment dramatically reduced the number of fluorescently labeled bacterial cells and fluorescence intensity (Fig. 3a; Supplementary Fig. 2b). This result suggested that the majority of bacterial fluorescent labeling is due to Sia9N₃ transported into certain bacte-
Sialic acid can be released from sialoglycans by microbial sialidases as demonstrated (Fig. 3a), which can promote the growth of certain SA-utilizing pathogens in the gut and lead to inflammatory diseases. To explore if Sia9N3 can also be released from the bacterially presented sialoglycans by other bacteria within the cultured sample, we tracked Sia9N3 incorporation over time during the plateau phase of bacterial growth. Briefly, the H2 fecal microbiome sample was grown with Sia9N3 for 14–66 h and the bacteria were quantitated for Sia9N3 incorporation by flow cytometry as described. Interestingly, detection of Sia9N3 labeling was maximally present at 14 h and gradually decreased until almost all labeling was eliminated by 40 h of incubation (Fig. 3b). Cultured media from each sample for every timepoint was assessed for sialidase activity, as measured by catalytic turnover of 4-methylumbelliferyl-N-acetylneuraminic acid (4-MUANA). Measurable sialidase activity was observed across all samples and suggests that bacterially secreted sialidases could remove Sia9N3 from bacterial sialoglycans over time (Fig. 3c). In parallel to the sialidase analysis, 16S rRNA gene sequencing showed that the compositional diversity of the cultured microbiota increased over time (Fig. 3d). At 14 h, the microbiome collection was dominated by Escherichia/Shigella (58% relative abundance), Clostridium (21%), and Enterococcus (15%). Common commensal organisms, including Citrobacter, Peptostreptococcus, Bacteroides, and Parabacteroides, began to appear after 24 h of culture and their relative abundances peaked at 16%, 13%, 10%, and 5%, respectively, at 66 h of culture. The corresponding relative abundance of Escherichia/Shigella, Clostridium, and Enterococcus dropped to 33%, 2%, and 0%, respectively. The bloom of other commensal bacteria at later time points may suppress the detection of Sia9N3-presenting microbes, causing the elimination of fluorescently labeled populations. Of note, Escherichia and Enterococcus may be outcompeted at later time points during the anaerobic culturing as these bacteria are facultative anaerobes. It is also likely that complete elimination of Sia9N3 detection is due to sialidase activity. The sialidases responsible for the removal of Sia9N3 may not be present at early time points and are introduced by late-blooming commensal organisms, such as Bacteroides, that appear at later time points, as determined by 16S rRNA gene sequencing analysis. Importantly, this shift in taxonomic composition may reflect protective mechanisms by which commensal bacteria target virulence factors expressed on pathogenic/pathobiont bacteria by exposing surface antigens for host recognition. Alternatively, commensal bacteria could

**Figure 3.** Bacterially incorporated surface Sia9N3 can be removed by microbial sialidases over time. (a) Treatment of labeled microbes before DBCO-BTN conjugation with a recombinant sialidase significantly reduces fluorescence labeling. The incorporated Sia9N3 is removed by microbial sialidases over time as fluorescence labeling was gradually removed during microbiota cultivation (b) and microbial sialidase activity was constantly detected in the culture media, 500 pM recombinant BT0455 was used as a positive control (c)). (d) 16S rRNA gene sequencing revealed the change of taxonomic composition during H2 microbiome culture. “Bkg” stands for background, “rSialidase” stands for recombinant sialidase.
be modifying environmental conditions to reduce the growth of pathogenic/pathobiont bacteria. All together, these data demonstrated that Sia9N₉ incorporation by bacteria in culture is dynamically regulated by other bacteria, nutrients, and bacterial sialidases.

The isolated Sia9N₉-presenting bacteria is a new strain of Escherichia coli. We next separated the Sia9N₉ incorporating bacteria from a 20-h culture of the H2 microbiome sample with FACS. The isolated fraction that yielded a high fluorescence signal due to Sia9N₉ incorporation was subjected to 16S rRNA gene sequencing and revealed that the Escherichia genus was significantly enriched post sorting (Fig. 4). We next subjected the cultured and primary H1–H3 microbiome samples to 16S rRNA gene sequencing to determine if differences in the bacterial taxonomic composition accounted for the selective labeling of the H2 microbiome sample only. Of note, 20 h of culture significantly altered the composition of all samples, and therefore, made the Sia9N₉-presenting bacteria detectable in H2 as the Escherichia genus was significantly enriched compared to the starting fecal sample (Supplementary Fig. 3). In addition, although only H2 contains Sia9N₉-presenting bacteria, Escherichia dominated all three cultured microbiome samples as the most abundant genus. These results suggest that Escherichia presented in the H1–3 cultured microbiome samples may differ at the strain level as only the Escherichia in H2 has the catabolic genes to incorporate Sia9N₉. An additional possibility is that the isolated Escherichia organism from H2 is present across all samples; however, the SA-presenting genes are not activated in the cultured H1 and H3 samples.

To identify the bacterial genome of the Sia9N₉-presenting H2 Escherichia, we cultured raw H2 microbiota on a Gifu agar plate overnight in anaerobic environment. Ten colonies were randomly picked and cultured with Gifu media supplemented with 200 μM Sia9N₉ anaerobically for 20 h. The cultured colonies were subjected to click labeling and flow cytometry analysis. Out of 10 colonies screened, liquid culture of colony 2 yielded nearly 100% incorporation of Sia9N₉ ( Colony 2 Passage 1, C2-P2-1, -2, and -3) (Fig. 5a). Furthermore, 3 randomly picked colonies from a C2-P1 culture plate (C2 passage-2, C2-P2-1, -2, and -3) also showed 100% Sia9N₉ incorporation, suggesting that the Sia9N₉ catabolic genes were readily passaged to, and active in, the daughter generations (Fig. 5b). The ZH-C2 colony was subjected to shotgun whole-genome sequencing in an attempt to identify the Escherichia bacterial strain.

The genome of our newly isolated bacteria (ZH-C2) cannot be fully mapped to any previously sequenced Escherichia bacteria in the NCBI RefSeq database, but is most similar to pathogenic E. coli strains UME A3174-1 (accession PRJNA186306) and VR50 (accession PRJEA61445) with average nucleotide identities of 99.85% and 99.76%, respectively (Fig. 6a; Supplementary Table 1). Our E. coli ZH-C2 strain has a 5.23 Mb genome that is similar in size to UMEA 3174-1 and VR50 and is much larger than the 4.64 Mb genome of the prototypi- cal nonpathogenic strain K-12 MG1655 (MG1655, accession PRJNA647163) (Supplementary Table 1). ZH-C2 shares over 90% identity in sequence to most MG1655 genes; however, ZH-C2 consists of 1111 missing or low-homology genes compared to MG1655 (Fig. 6b; Supplementary Table 1). Conversely, the MG1655 genome consists of 347 missing or low-homology genes relative to ZH-C2 (Fig. 6a; Supplementary Table 1). The majority of gene products have no identified function; however, GO term analyses demonstrate ZH-C2 encodes for an increased diversity and number of additional proteins relative to MG1655 involved in DNA repair and replication, polysaccharide transport, virion assembly, as well as type II secretion (Fig. 6b; Supplementary Table 1). With respect to molecular function, a diverse set of additional genes are encoded in ZH-C2, notably predicted cobalamin binding, ATPase, transferase, and porin activities (Fig. 6c; Supplementary Table 1). With a significantly larger collection of ZH-C2 proteins likely localized to the outer membrane compared to MG1655 (Fig. 6d; Supplementary Table 1). Conversely, the MG1655 genome consists

---

**Figure 4.** 16S rRNA gene sequencing analyses. (a) H2 microbiome sample was cultured with or without 200 μM Sia9N₉, which does not significantly alter taxonomic composition of 20-h H2 microbiome culture. Sia9N₉ presenting bacteria from H2 microbiome culture were labeled with fluorescence and isolated with FACS. (a) and (b) 16S rRNA gene sequencing showed that the Escherichia/Shigella genus was significantly enriched after cell sorting whereas other species do not. This suggest that the fluorescently labeled cells are mostly Escherichia/Shigella bacteria.
of a number of additional genes associated with DNA-mediated recombination and integration, as well as cell adhesion, DNA binding and DNA recombinase activities (Fig. 6b–d; Supplementary Table 1).

MG1655 encodes for the sialic acid catabolizing gene cluster that includes nanA, nanE, nanK, the regulatory gene nanR, and sialic acid transporter gene nanT. However, in addition to these sialic acid-related genes, ZH-C2 also encodes for the polysialic acid biosynthesis (kps) cluster consisting of several genes, including: (1) neuA, the cytidine monophosphate (CMP)-sialic acid synthetase which produces the sialyltransferase donor CMP-sialic acid; and (2) neuS, the polysialyltransferase which synthesizes capsular polysialics from CMP-sialic acid. MG1655 lacks neuA, neuS, as well as the other kps cluster genes, which highlights the inability of MG1655 to activate and present Sia9N3 on its cell surface (Supplementary Fig. 4). Moreover, the sugar-phosphate isomerase KpsF/GutQ (accession WP_001296394, ZH-C2_01729), known to be associated with polysialic acid deposition onto the capsule, is also identified in the ZH-C2 genome (with 100% identity to UMEA 3174-1 and VR50) but not in MG1655. Altogether, the genomic analyses suggest that the cell-surface fluorescence labeling we observed is caused by deposition of polySia9N3 on the bacterial capsule.

Despite being isolated from a healthy human fecal sample, the genome of ZH-C2 strain is most similar to pathogenic strains UMEA 3174-1 and VR50, both of which were isolated from urinary tract infections (Supplementary Table 1). Established pathogenic genes, such as the type III secretion system are conserved across pathogenic strains UMEA 3174-1 and VR50, both of which were isolated from urinary tract infections (Supplementary Table 1). Notwithstanding the high conservation, notable differences were also observed. ZH-C2 has 628 and 680 low homology/unique genes relative to UMEA 3174-1 and VR50, respectively. Conversely, UMEA and VR50 have 507 and 393 low homology/unique genes, respectively, to ZH-C2 as determined by sequences with < 90% identity and/or covering < 70% of the sequence length (Supplementary Table 1). Interestingly, there is an enrichment of predicted GO terms associated with DNA-mediated transposition, oxidation–reduction processes, pilus assembly, and carbohydrate metabolic activities in ZH-C2 relative to Table 1).

Figure 5. Isolation and passaging of the Sia9N3 incorporating bacterial strain from H2 microbiome. (a) Colony screening of H2 microbiome isolated one Sia9N3 incorporating bacterial strain. The raw H2 microbiota was diluted and plated on a Gifu-agar plate for overnight anaerobic culture, 10 colonies with different sizes and shapes were picked and cultured with the Gifu media supplemented with 200 μM Sia9N3, and subjected to fluorescence labeling study aforementioned. (b) The Sia9N3 incorporating feature is readily passaged to daughter generations. The cultured C2 media (C2-P1) was plated again and three colonies (C2-Passage 2, C2-P2-1,2,3) were picked for metabolic labeling study. Bacteria cultured from three colonies were almost 100% Sia9N3 incorporating cells, indicating that a single Sia9N3 incorporating bacterial strain is purifited out from the microbiome complex.
Discussion

Azide-modified glucosamine, mannosamine, fucose, and 3-deoxy-d-manno-oct-2-ulosonic acid (Kdo) have been used to metabolically label bacterial cells in vitro and in vivo19,20,42. Very recently, Weiss et al. reported the application of functionalized monosaccharide derivatives including N-azidoacetylglucosamine, N-azidoacetylmannosamine, N-azidoacetylsialic acid in metabolically labeling of bacterial cells from murine intestinal microbiome43. Here, we focus on the application of azide-modified N-acetylmuramic acid to selectively label Neu5Ac-presenting bacteria from a complex human fecal microbiome sample for the first time. Using flow cytometry and metagenomic sequencing, we identified a new strain of *E. coli* that can present environmental Neu5Ac on its surface via the NanT-NeuA-NeuS pathway. Our results also suggest a dynamic process by which...
select distal gut bacteria can exploit environmental sialic acid for self-decoration that, in turn, is likely subject to removal by sialidase activity provided by other microbiome constituents (e.g., *B. thetaiotaomicron*).

The *Sia9N*$_9$-presenting activity of our new *E. coli* ZH-C2 strain is only detected in sample H2. Additional studies will investigate if this new *E. coli* is missing from the H1 and H3 cultures or present with a silenced NanT-NeuA-NeuS pathway during the ex vivo culture. Previous studies suggest both possibilities, as the gut microbiome is regulated by diet and crossstalk amongst bacteria as well as the host and suggests a dynamic composition of different bacterial strains over time$^{44,45}$. Conversely, certain bacteria, including *Bacteroides fragilis*, are able to switch capsular polysaccharide structures enabling the colonization specific intestinal niches$^{46}$. The gradual decrease in *E. coli* Sia9N$_{9}$ surface expression over time will also be further interrogated. We anticipate that the loss of Sia9N$_{9}$ is caused by other commensal organisms that gain prominence over time in the ex vivo culture and remove the *E. coli* poly-SA as a biologically relevant mechanism for host recognition and clearance. Additionally, surface-expressed poly-SA may be harvested by the parental *E. coli* for growth, as environmental conditions and nutrient availability shift over time.

Distal gut bacteria can degrade and ferment complex carbohydrates to provide an energy source and/or immunomodulatory molecules for themselves, the host, and/or other gut bacteria$^{1,32,47,48}$. In return, host cells can produce glycans which can provide nutritional advantages for certain microbes$^{49}$. Such mutualistic relationships make the tracking and study of carbohydrate catabolism by individual bacteria in complex microbiomes complicated and challenging. Consistent with previous studies, we found that cultured microbiome samples do not fully represent the taxonomic diversity of the original source due to growth advantages of certain bacteria under anaerobic systems with limited nutritional sources from the selected media$^{31,50}$. Notwithstanding, our approach took advantage of the bias induced by bacterial culture and enriched for a new SA-presenting strain of *Escherichia coli* isolated from a human gut microbiome. We posit that SA-presenting bacteria can be identified from different microbiomal sources including respiratory, urinary, and vaginal microbiomes in both health and disease (i.e., detection of pathogens in blood and stool clinical samples). Additionally, the scavenging of environmental sialic acid is associated with pathogenic bacterial virulence$^{50}$, and proteins involved in the process may represent new targets for the discovery of antibacterial small molecules. Further, application of different probes, media, and culture conditions (i.e., pH, anaerobic/aerobic growth, time) may assist us in the identification of other bacterial species with catabolic functions of interest from clinically relevant microbiota samples.

Our integrated assay system detects sialic acid-presenting bacteria in real-time from complex microbiome samples. We envision that this labeling and tracking strategy can assist in identifying new SA-presenting bacteria and developing diagnostic tools and drug discovery models. Furthermore, this workflow may be used to investigate the installation of other environmental glycans by microbiota.

**Methods**

**Synthesis, purification, and characterization of Neu5Ac9N$_9$**. ACS reagent grade organic solvents were purchased from Sigma Aldrich and were used directly if not specified. $N$-Acetyl-$\alpha$-azido-$\beta$-deoxy-
neuraminic acid (Neu5Ac9N$_9$/Sia9N$_{9}$) was synthesized, purified, and characterized as previously reported$^{32}$. Briefly, Neu5Ac (3 g, 9.6 mmol) was mixed with Dowex 50W X8-200-400 (H) (2 g, Alfa Aesar, L13922) in 50 mL methanol, followed by overnight stirring at room temperature. The resin was removed by filtration over Celite and the filtrate was concentrated under reduced pressure to 15 mL. 15 mL of ice-cold diethyl ether was layered above the concentrated methanolic solution, and the mixture was chilled at $4\, ^{\circ}$C overnight to yield a pale yellow crystalline solid. Supernatant was carefully decanted, and the pellet was carefully rinsed with ice cold diethyl ether three times, then air dried to give 2.2 g Neu5AcOMe as a white crystalline powder. Neu5AcOMe (2.2 g, 6.8 mmol) was dissolved in 20 mL of anhydrous pyridine, cooled in an ice bath, then treated with $p$-toluenesulfonyl chloride (2.3 g, 12.3 mmol; freshly purified by recrystallization from chloroform and petroleum ether). The mixture was stirred over an ice bath for 30 min, followed by overnight stirring at room temperature. Pyridine was removed by rotary evaporation to form syrup-like crude, which was purified with flash chromatography (EtoAc:MeOH = 20:1 - 20:5), dried, to yield 0.6 g Tosyl-Neu5AcOMe as white solid. Tosyl-Neu5AcOMe (0.6 g, 1.26 mmol) and NaN$_3$ (0.32 g) were suspended in 10 mL of acetonewater (3:1) solution followed by overnight stirring at 70 °C. The crude was dried, acidified with Dowex 50Wx8-200-400 (H), and purified with flash chromatography (60:40 - 25:75 DCM:MeOH), dried to yield 0.31 g pure final product Neu5Ac9N$_9$/Sia9N$_{9}$.

**Sample collection.** Human fecal sample collection methods were carried out in accordance with relevant guidelines and regulations. Experimental protocols were approved by the Institutional Review Board (IRB) at Scripps Research and Scripps Health; protocol #: IRB-14-635 and informed consent was obtained. Three human fecal samples (H1, H2, and H3) were collected from one healthy donor longitudinally at days 1, 18, and 80. Immediately after collection, fecal samples were transported in an anaerobic bag on dry ice into an anaerobic chamber (Coy Laboratory Products) filled with 3% hydrogen and 97% N$_2$. Fecal samples were then re-suspended in filtered phosphate buffer saline (PBS) buffer, pH 7.4. Large non-bacterial particles were removed, and the cloudy supernatant was carefully aspirated out. Bacterial cells were then centrifuged down at 16,000×g for 5 min and washed with PBS 3 times. Fecal bacteria were cultured in the Gifu media in anaerobic atmosphere at 37 °C with an initial optical density at 600 nm (OD$_{600}$) of 0.005. The rest of bacterial cells were re-suspended in 20% glycerol/PBS solution at an OD$_{600}$ of 10, flash frozen, and stored at −80 °C for future use.

**Sequential fluorescence labeling of bacterial cells and flow cytometry.** Fecal microbces cultured with or without Sia9N$_{9}$ were collected by centrifugation at 16,000×g and washed with filtered PBS for 3 times. Bacterial cells were then re-suspended to an OD$_{600}$ of 1.0, treated with DBCO-Biotin for 2 h at room temperature in anaerobic atmosphere. The bacterial cells were washed with PBS for 3 times with centrifugation at 16,000×g
for 5 min, followed by streptavidin Alexa 647 staining at 1:200 dilution for 1 h with OD_{600} at 1.0. Stained bacterial cells were washed with PBS for 3 times, followed by analyzing on a NovoCyte 3000 with NovoSampler Pro model. Singlet cells were selected by plotting SSC-H and FSC-H versus SSC-A and FSC-A, respectively, followed by gated on Alexa 647 channel. 15,000 total events were collected for data analysis with the percentage of singlets at ~ 90%.

**Fluorescence microscopy.** Cultured H2 bacterial cells were thoroughly washed and subjected to click labeling with the biotin tag. Streptavidin Alexa 647 was then used to stain the cells. Fluorescently labeled bacterial cells were spotted on a glass slide, air-dried for 15 min, and covered with a cover slip with ~ 1 mm thickness. Confocal microscopy was performed on a Zeiss LSM 880 Airyscan confocal laser scanning microscope. Imaging analyses were processed with FlowJo_v10.

**Expression and purification of sialidase BT0455.** The gene fragment encodes for the *Bacteroides thetaiotaomicron* sialidase BT0455 (Uniprot ID: Q8AAK9) was ordered from Integrated DNA Technologies (IDT), with the N-terminal signal peptide removed. The BT0455 encoding sequence was cloned into a customized pT7HMT vector25, followed by expression in BL21(DE3) cells as an N-terminal His, tag fusion. BT0455 was purified from bacterial cell lysate with a nickel–nitrilotriacetic acid (Ni–NTA) affinity column, followed by size-exclusion chromatography. Finally, BT0455 was aliquoted and stored at − 80 °C in a storage buffer containing 20 mM Tris–HCl and 100 mM NaCl with pH at 8.0.

**Sialidase activity assay.** 14–66 h after fecal microbiota culture, the bacterial cells were pelleted, and the supernatants (cultured media) were assayed for sialidase activity. Specifically, a 200 μL PBS mixture containing 500 pM purified BT0455 sialidase or cultured media and 100 μM turn-on fluorescent substrate 4-methylumbelliferyl-α-d-N-acetylneuraminic acid (4-MUNANA) was pre-incubate for 60 s, followed by fluorescence measurement on a microplate reader (excitation 365 nm, emission 450 nm). The fluorescence intensities were measured every 60 s over the linear range of enzymatic reaction. The presented data is representative of three individual activity assays. In each assay, fluorescence intensity at different time points were measured in duplicate.

**FACS.** Fluorescence activated cell sorting (FACS) of fluorescently labeled bacteria from H2 microbiota culture was performed on a Sony MA900 multi-application cell sorter. Singlets were selected by plotting SSC-H and FSC-H versus SSC-A and FSC-A, respectively. Approximately 1 × 10^7 fluorescently labeled bacterial cells were sorted and collected for 16S rDNA sequencing analyses.

**DNA extraction and sequencing analysis.** Pre-sorted or post-sorted bacteria were subjected to DNA extraction with the ENZA bacterial DNA extraction kit (Omega biotek, D3350) based on manufacturer’s protocol. 5–22 ng DNA was used with the NEXTFLEX 16S V4 Amplicon-Seq Kit 2.0 library prep kit (PerkinElmer) following manufacturer’s recommended protocol with between 11 and 20 PCR cycles depending on sample input. PCR amplicons were all pooled and gel purified to select the target amplicon region. Libraries were loaded onto a flowcell and sequenced on an Illumina MiSeq sequencer to generate 2 × 300-bp paired-end reads. We collected 0.2 and 3 million reads per sample for 16S rRNA and whole genome sequencing analyses, respectively. 16S rRNA gene sequencing data was processed with Illumina Basespace 16S metagenomic sequencing application for microbiome taxonomic annotation. The taxonomic classification step used ClassifyReads, a high-performance implementation of the Ribosomal Database Project (RDP) Classifier developed by Wang et al.53.

**Colony screening.** Human microbiota H2 was diluted with PBS and plated on a Gifu-agar plate for overnight culture in an anaerobic atmosphere. 10 colonies were randomly picked and inoculated in 1 mL Gifu-Sia9N1 media for 20 h anaerobic culture at 37 °C. Then, bacterial cells were collected, washed, and tested for fluorescence labeling with flow cytometry as described. Bacteria cultured from C2 colony (C2 passage-1, C2-P1) showed nearly 100% fluorescence labeling. To test the homogeneity of C2 and C2-P1 bacteria, C2-P1 bacteria were cultured on a Gifu-agar plate and 3 colonies were randomly picked and inoculated in Gifu-Sia9N1 media followed by cultivation and fluorescence labeling experiments. As a result, 100% bacteria cultured from all three colonies (C2-P2-1 ~ 3) were fluorescently labeled.

**Shotgun whole-genome sequencing.** Colony C2-P2-2 and C2-P2-3 were cultured in 6 mL Gifu media, respectively. Bacterial cells were then pelleted and washed, and bacterial DNA was extracted with ENZA bacterial DNA extraction kit (Omega biotek, D3350) based on manufacturer’s protocol. For genomic preps, DNA was sheared to ~ 500 bp size using a Covaris S2 Focused-ultrasonicator, after which 10 ng of sheared DNA was used to make sequencing libraries with the NEBNext Ultra II DNA Library Prep Kit following manufacturer’s recommended protocol with 9 cycles of PCR. Libraries were size selected using AmpureXP beads to 500–700 bp, loaded onto a flowcell, and sequenced on an Illumina MiSeq sequencer to generate 2 × 300-bp paired-end reads.

**Whole-genome assembly of ZH-C2.** The reads for ZH-C2 were processed by Trimmomatic 0.36 to remove the Illumina adapters and parts of the reads with low quality using the following filters: ILLUMINA-CLIP: <path-to-adapters>/TruSeq3-PE-2.fa:2:30:10 LEADING:30 SLIDINGWINDOW:10:20 MINLEN:100. The trimmed reads were then assembled using SPAdes 3.13.0 with the following k values: 21, 33, 55, 77, 99, 127. Next, the blastn command of BLAST 2.7.1 with -max_target_seqs 1 was used to align the assembled
ZH-C2 contigs against the NCBI RefSeq genome database (https://www.ncbi.nlm.nih.gov/genome). Two custom Perl scripts binned the voluminous BLAST output by species and strain. The average nucleotide identity (ANI) between ZH-C2 and the top-scoring strains was computed using blastn with -best_hit_overhang 0.1 -best_hit_score_edge 0.1 and another custom Perl script to determine the closest *E. coli* genomic strains: UMEA 3174-1 (closest overall), VR50 (second closest), and K-12 MG1655 (closest of three reference genomes). Some 11% of the ZH-C2 contigs did not align with *E. coli*. These were presumed to be from contaminants and were removed from the assembled genome along with another 5% of the contigs that were shorter than 200 bp to comply with guidelines for submission to GenBank.

**Identification of genes and their protein products.** The proteins produced by UMEA 3174-1, VR50, and K-12 MG1655 are available from the RefSeq database. For ZH-C2, Prokka 1.14.6 was used to identify the genes and associated proteins. Many were identified only as “hypothetical proteins”. To get further information on these as well as more consistent naming, a database of RefSeq bacterial proteins was generated using the commands given at https://dmnfarrell.github.io/bioinformatics/local-refseq-db. The blastp command of BLAST with -max_target_seqs 1 was then used to align the proteins from Prokka against the new database. This provided additional information on more than half of the hypothetical proteins examined; 19 proteins from Prokka were not found in RefSeq.

**Identification and removal of duplicate proteins.** To check for the presence of duplicate proteins in a single genome, the proteins in each genome were aligned against themselves using blastp with -max_target_seqs 15. Most matches had less than 100% identity and 100% coverage, and these were ignored. The remaining perfect matches were mostly unique but also revealed 6 duplicate proteins in ZH-C2 from Prokka and 41 duplicate proteins in K-12 MG1655 from RefSeq. These duplicates were removed for subsequent analysis. No duplicates were found in the proteins for UMEA 3174-1 or VR50 from RefSeq.

**Comparison of proteins.** The proteins in ZH-C2 were compared with those in each of the three other genomes of interest, and vice versa, using blastp. Even though -max_target_seqs 1 was specified, some proteins had more than one match with varying lengths. Thus, a second pass was made through the blastp output retaining only the longest match when there was more than one. Finally, the matches for each of the six pairwise comparisons were sorted into four groups corresponding to (1) perfect matches between the two genomes, (2) high-similarity matches with at least 90% amino acid identity covering at least 70% of the protein length, (3) low-similarity matches that did not meet the previous criteria, and (4) non-matches.

Venn diagrams were generated using output from CD-HIT (4.8.1) clustering of 4 combined proteomes (ZH-C2, MG1655, UMEA 3174-1, VR50)8,8. The following command line input was used to generate desired cluster files: "cd-hit -i inputfile.fasta -o outputfile -c %cutoff -g 1 -d 0" where “%cutoff” is the percent similarity cut-off used to define a cluster. In our analysis, we typically set percent similarity cut-offs at 65%, 75%, 85%, 95%, and 99%. CD-HIT output was organized into tables using the Python Pandas package and Venn diagrams were generated using venn/pyvenn packages.

Gene Ontology (GO) terms were generated using InterProScan (5.40-77.0)60. Briefly, protein sequences of interest were concatenated into a protein fasta file then submitted for InterProScan analysis using the following command line input: "/interproscan.sh -i input_file.fasta -f tsv -d -goterms". Output tab separated files were organized using the Python Pandas package and bar plots were generated in Matplotlib. GO terms were mapped to their name spaces using the comprehensive "go.obo" listing available at geneontology.org 61. GO terms were organized such that all protein sequences contributed at least 1 count to each of the 3 "Biological Process," "Molecular Function," and "Cellular Component" GO namespaces. If sequences could not be annotated, 1 count of "None" was contributed to each of the name spaces. For sequences annotated multiple times, each annotation contributed 1 count to the GO term’s respective namespace.

**Data availability**
The raw sequence reads and assembled genome are available from NCBI under BioProject PRJNA693345.

Received: 6 November 2020; Accepted: 1 February 2021
Published online: 26 February 2021

**References**
1. Rooks, M. G. & Garrett, W. S. Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* 16, 341–352 (2016).
2. Valdes, A. M., Walter, J., Segal, E. & Spector, T. D. Role of the gut microbiota in nutrition and health. *BMJ* 361, k2179 (2018).
3. Boulange, C. L., Neves, A. L., Chilloux, J., Nicholson, J. K. & Dumas, M. E. Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med.* 8, 42 (2016).
4. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65 (2010).
5. Korem, T. et al. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 349, 1101–1106 (2015).
6. Shepherd, E. S., DeLoach, W. C., Pruss, K. M., Whitaker, W. R. & Sonnenburg, J. L. An exclusive metabolic niche enables strain engrafment in the gut microbiota. *Nature* 557, 434–438 (2018).
7. Tanoue, T. et al. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature* 565, 600–605 (2019).
8. Uchiyama, S. et al. Dual actions of group B streptococcus capsular sialic acid provide resistance to platelet-mediated antimicrobial killing. *Proc. Natl. Acad. Sci. U. S. A.* 116, 7465–7470 (2019).
9. Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W. & Apicella, M. A. Novel sialic acid transporter of *Haemophilus influenzae*. *Infect. Immun.* 73, 5291–5300 (2005).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
56. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
58. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
59. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
60. Zdobnov, E. M. & Apweiler, R. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
61. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

**Acknowledgements**

We thank Jenny Cornell, Jessica Ledesma, and Steven Robert Head at Next Generation Sequencing Core at The Scripps Research Institute for 16S rRNA gene and whole-genome sequencing experiments. We thank Padmaja Natarajan and Alain Domissy at Center for Computational Biology at The Scripps Research Institute for analyzing the 16S rRNA gene sequencing data. We thank Qiyun Zhu at the University of California, San Diego for assistance with the whole-genome sequencing data analysis. This work was supported by NIH award R21 AI139744 to D.W.W. and Scripps Research Translational Institute, an NIH-NCATS Clinical and Translational Science Award (CTSA; UL1 TR002550) to W.P. and A.T.

**Author contributions**

Z.H. and D.W.W. designed the research; Z.H., P.T.B., V.F.V., and J.R.T. performed the experiments; Z.H., W.P., P.T.B., A.T., and D.W.W. analyzed the data; Z.H. and D.W.W. composed the paper and all authors edited and approved its contents.

**Competing interests**

The authors declare no competing interests.

**Additional information**

The online version contains supplementary material available at [https://doi.org/10.1038/s41598-021-83875-w](https://doi.org/10.1038/s41598-021-83875-w).

**Correspondence** and requests for materials should be addressed to D.W.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

© The Author(s) 2021