LETTERS
https://doi.org/10.1038/s41564-019-0564-9
nature microbiology

Gut bacteria responding to dietary change encode sialidases that exhibit preference for red meat-associated carbohydrates

Livia S. Zaramela[1,10], Cameron Martino[1,2,10], Frederico Alisson-Silva[3,4,9,10], Steven D. Rees[5], Sandra L. Diaz[3,4], Léa Chuzel[6], Mehul B. Ganatra[6], Christopher H. Taron[6], Patrick Secrest[3,4], Cristal Zuñiga[1], Jianbo Huang[5], Dionicio Siegel[5], Geoffrey Chang[5], Ajit Varki[3,4] and Karsten Zengler[1,7,8*]

Dietary habits have been associated with alterations of the human gut resident microorganisms contributing to obesity, diabetes and cancer[1]. In Western diets, red meat is a frequently eaten food[2], but long-term consumption has been associated with increased risk of disease[3,4]. Red meat is enriched in *N*-glycolylneuraminic acid (Neu5Gc) that cannot be synthesized by humans[5]. However, consumption can cause Neu5Gc incorporation into cell surface glycans[6], especially in carcinomas[4,7]. As a consequence, an inflammatory response is triggered when Neu5Gc-containing glycans encounter circulating anti-Neu5Gc antibodies[8,9]. Although bacteria can use free sialic acids as a nutrient source[10–12], it is currently unknown if gut microorganisms contribute to releasing Neu5Gc from food. We found that a Neu5Gc-rich diet induces changes in the gut microbiota, with Bacteroidales and Clostridiales responding the most. Genome assembling of mouse and human shotgun metagenomic sequencing identified bacterial sialidases with previously unobserved substrate preference for Neu5Gc-containing glycans. X-ray crystallography revealed key amino acids potentially contributing to substrate preference. Additionally, we verified that mouse and human sialidases were able to release Neu5Gc from red meat. The release of Neu5Gc from red meat using bacterial sialidases could reduce the risk of inflammatory diseases associated with red meat consumption, including colorectal cancer[4] and atherosclerosis[13].

Red meat is enriched in Neu5Gc that cannot be synthesized by humans due to an evolutionary loss of a functional CMP-Neu5Ac hydroxylase (CMAH)[5]. Metabolic incorporation of Neu5Gc into tissues requires glycosidically bound-Neu5Gc for reasons currently unknown. In contrast, free-Neu5Gc is utilized by gut microbes or cleared rapidly by the kidneys through the urine[14] (Supplementary Fig. 1). Intestinal bacteria can release host-derived sialic acids from mucosal mucins and glycolipids by expressing sialidases[15–18]. It is also known that gut commensal and pathogenic bacteria can use sialic acids as a carbon source[10–12]. However, to our knowledge, every bacterial sialidase tested prefers *N*-acetylneuraminic acid (Neu5Ac) to Neu5Gc[19–21]. It is currently unknown how bound-Neu5Gc is metabolized in the gut. Once free, sialic acids can be taken up through membrane-associated transporters and used as carbon, nitrogen or energy sources, or used to sialilate bacterial cell surface glycans[16]. In addition, changes in the intestinal concentration of sialic acids, for example induced by inflammation, can alter the expression of bacterial genes involved in sialic acid catabolism promoting intestinal dysbiosis[22]. Due to the importance of sialic acids in microbe–host interaction within the gut, we investigated if a Neu5Gc-rich diet could provoke changes in bacterial metabolism and alter the gut microbiome. We compared the microbiota composition in faecal samples of *Cmah*$^{-/-}$ mice[23] that were fed a sialic acid-free (soy) diet, a Neu5Gc-rich porcine submaxillary mucin (PSM) diet or a Neu5Ac-rich edible bird's nest (EBN) diet[14] (Supplementary Table 1).

Faecal contents were scraped from the mouse colons, and the microbiome of *Cmah*$^{-/-}$ mice and wild-type (WT) mice fed soy, PSM or EBN diets was determined through 16 S rRNA gene amplicon sequencing. Bray–Curtis dissimilarity showed a significant difference in the bacterial genotypes present in *Cmah*$^{-/-}$ and WT mice, indicating that a mouse's inability to synthesize endogenous Neu5Gc significantly impacted microbiome composition (Fig. 1a) and substantiated the need for a human-like *Cmah*$^{-/-}$ mouse model in our experimentation. We found that changes in the microbial composition in *Cmah*$^{-/-}$ mice were diet-dependent, with Clostridiales and Bacteroidales contributing significantly to the variations observed amongst the diets (Fig. 1b and Supplementary Fig. 2). Independent of genotype, the microbiome of PSM-fed mice was significantly less diverse compared to the microbiome of those fed soy and EBN diets ($P < 0.05$) (Supplementary Fig. 3a). Human-like *Cmah*$^{-/-}$ mice revealed similar taxonomic profiles at the family level amongst the three diet groups (Supplementary Fig. 3b). However, at the genus level, *Helicobacter*, *Intestinimonas* and Candidatus *Saccharibacteria genera incertae sedis* were significantly enriched in the EBN group compared to PSM (Fig. 1c). Moreover, *Bacteroides*, *Barnesiella*, *Clostridium* group III, *Parabacteroides*, *Roseburia* and *Turicibacter* were significantly enriched in the PSM group compared to EBN

[1]Department of Pediatrics, University of California, San Diego, CA, USA. [2]Bioinformatics and Systems Biology Program, University of California, San Diego, CA, USA. [3]Department of Medicine and Cellular and Molecular Medicine, University of California, San Diego, CA, USA. [4]Glycobiology Research and Training Center, San Diego, CA, USA. [5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA. [6]New England Biolabs, Ipswich, MA, USA. [7]Departmemt of Bioengineering, University of California, San Diego, CA, USA. [8]Center for Microbiome Innovation, University of California, San Diego, CA, USA. [9]Present address: Paulo de Goes Institute of Microbiology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. [10]These authors contributed equally: Livia S. Zaramela, Cameron Martino, Frederico Alisson-Silva. *e-mail: kzengler@ucsd.edu

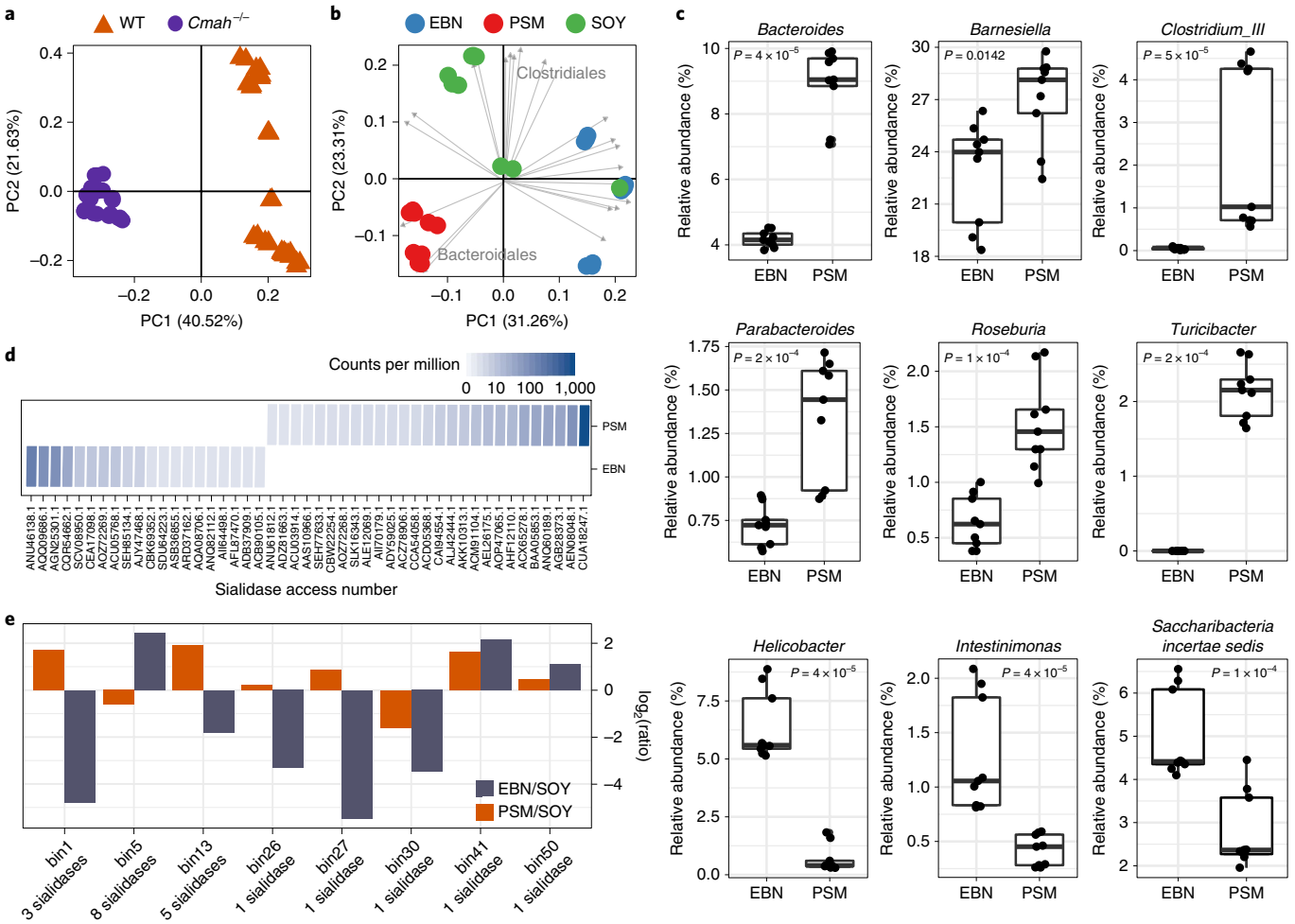


Fig. 1 | Composition of gut microbial community of mice fed on soy, PSM or EBN diet. a, Beta-diversity analysis of WT versus *Cmah*$^{-/-}$ mouse. Pairwise Bray–Curtis dissimilarities were plotted against the first and second principal coordinates (PC1 and PC2) (ANOSIM R = 0.979, *P* = 0.001). WT samples include *n* = 5 biologically independent animals and *n* = 15 independent experiments per diet. *Cmah*$^{-/-}$ samples include *n* = 3 biologically independent animals and *n* = 9 independent experiments. **b**, Beta-diversity analysis of *Cmah*$^{-/-}$ mouse fed on soy, PSM or EBN diet. Pairwise Bray–Curtis dissimilarities were plotted against PC1 and PC2 (ANOSIM R = 0.831, *P* = 0.001). Grey arrows represent the significant vector fitting with each of the PCoA (principal coordinates analysis) ordinations. The most representative taxa are indicated on the plot. **c**, Differentially abundant bacterial genera on PSM and EBN diet. Boxes and whiskers indicate quartiles, and middle makers indicate the median (*n* = 9). *P* values were determined using the two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. **d**, Sialidases diet-dependent using CAZyme database. **e**, Relative abundance ratio (PSM/SOY and EBN/SOY) of the bins with annotated sialidases.

(*P* < 0.05) (Fig. 1c and Supplementary Fig. 4). *Bacteroides* were previously associated to efficiently metabolize carbohydrates from plant as well as animal-based food due to their diverse enzymatic repertoire[24,25]. Additionally, by computationally simulating 773 metabolic models from human gut microbiome members[26] we found that members of the *Bacteroidetes*, including *Bacteroides fragilis*, *Bacteroides cacae* and *Bacteroides thetaiotaomicron*, were some of the most efficient microorganisms using sialic acids as a carbon source (Supplementary Fig. 5, Supplementary Table 2 and Supplementary discussion). Shotgun metagenomic DNA sequencing was performed to evaluate the enzymatic repertoire for using carbohydrates in the microbial community from *Cmah*$^{-/-}$ mice fed soy, PSM or EBN diets, and raw reads were aligned to a carbohydrate active enzymes (CAZymes) dbCAN-seq database[27]. A principal coordinate analysis revealed a near-significant clustering in gene function between diet types (analysis of similarities (ANOSIM) *R* = 0.246, *P* = 0.06), with a higher similarity between the CAZymes on soy and EBN diets (Supplementary Fig. 6). Additionally, the examination of individual sialidase genes revealed several diet-dependent sialidases (Fig. 1d). To evaluate the sialidases genes present in the microbiome, the combined metagenomes were co-assembled and 51 genome bins containing 21 sialidase genes were identified (Supplementary Fig. 7, Supplementary Table 3). Amongst the bins with annotated sialidase genes, bin13, whose closest relative was *B. thetaiotaomicron* (Supplementary Table 3), was the most abundant in PSM compared to EBN diets (Fig. 1e). Bin13 contains five sialidases (sialidase23, sialidase24, sialidase26, sialidase60 and sialidase65). Sialidase26 exhibited high amino acid sequence conservation (81% identity) to sialidase CUA18247.1 from *B. fragilis*, the most abundant sialidase protein in the PSM diet (Fig. 1d, Supplementary Fig. 8).

The increased abundance of bin13 in PSM suggests its prominent role in Neu5Gc metabolism in the gut. As such, bin13 sialidases might possess biochemical properties that make them better suited for Neu5Gc release. To test this notion, each of the five sialidases genes in bin13 were amplified by PCR (Supplementary Fig. 9)

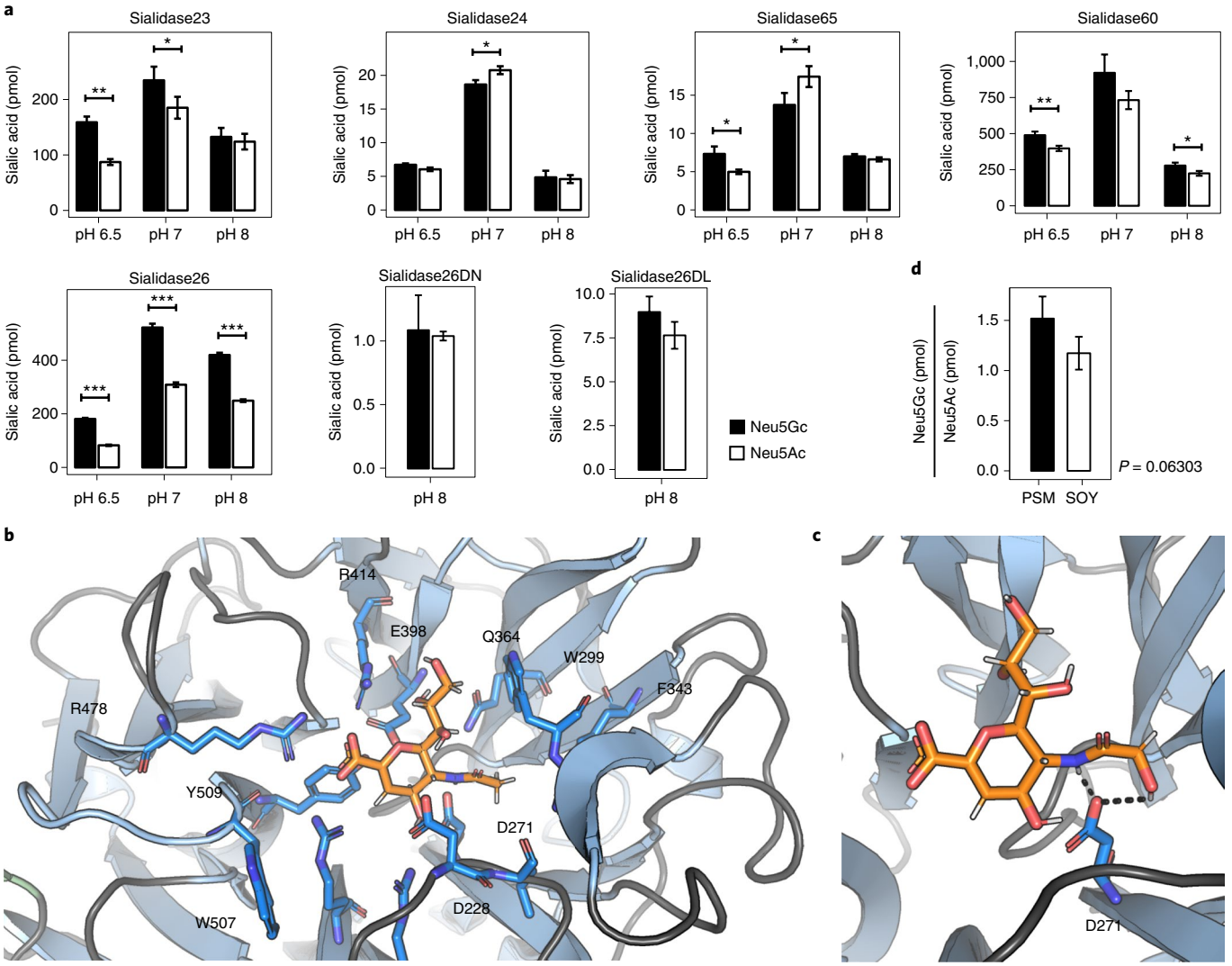


Fig. 2 | Characterization of sialidases preference for Neu5Ac- or Neu5Gc-containing substrates. a, Sialidase activity assay using 0.5 µg of each sialidase in three different pHs (6.5, 7 and 8). A mix of WT and *Cmah*$^{-/-}$ mouse serum containing similar amount of bound-Neu5Ac and bound-Neu5Gc was used as substrate. Data are represented as triplicates ($n = 3$ biologically independent experiments). **b**, Crystal structures of sialidase26 co-crystallized with DANA-Ac. **c**, Crystal structures of sialidase26 co-crystallized with DANA-Gc. Interacting side chains are represented as sticks (blue) and ligands (orange). The protein residues interacting with the sialic acid and the residue position (such as R478 and D271) are indicated in black (**b**,**c**). **d**, Overall ratio sialidase activity in clarified faecal samples from mice fed with PSM ($n = 4$ biologically independent experiments) or Sialic acid-free soy ($n = 3$ biologically independent experiments). Sialidase activity assay was performed using 2 µg of total protein and a mix of WT and *Cmah*$^{-/-}$ mouse serum containing similar amount of bound-Neu5Ac and -Neu5Gc. Sialidase activity is presented as a ratio of Neu5Gc (pmol) and Neu5Ac (pmol). **a**,**b**,**d**, Statistical significance was determined by Student's two-tailed *t*-test. $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$. Bars represent geometric mean ± s.e.m.

and heterologously expressed in vitro, purified and assayed for its substrate preference. Sialidase activity measurements were performed using different enzyme concentrations (0.5–10 µg) at three different pHs (6.5, 7 and 8) (Fig. 2a and Supplementary Fig. 10). Four out of five sialidases showed preferential Neu5Gc activity in at least one of the pHs tested (Fig. 2a). To the best of our knowledge, no previously characterized exo-sialidases have been shown to prefer Neu5Gc over Neu5Ac[21,28–30] (Supplementary discussion). Additionally, we tested in vivo sialidase activity in fresh faecal samples. Clarified faecal pellets from mice fed with PSM tend to preferentially release Neu5Gc compared to clarified faecal pellets from mice fed with soy (Fig. 2d).

The most compelling enzyme from the substrate specificity study in all tested conditions was sialidase26. Sialidase26 had protein sequence motifs characteristic of the GH33 family of sialidases (Supplementary Figs. 11 and 12). Despite this substrate preference, sequence residues that are predicted to interact with terminal sialic acids in the catalytic site are highly conserved with structurally studied sialidases exhibiting unknown Neu5Gc preference over Neu5Ac (Supplementary Fig. 13 and Supplementary discussion). To explain the structural underpinnings of Neu5Gc preference, we used X-ray crystallography to determine the structure of sialidase26 both alone (PDB 6MRX, 2.0 Å resolution) and in complex with the inhibitors DANA-Ac (*N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid) and DANA-Gc (*N*-glycolyl-2,3-dehydro-2-deoxyneuraminic acid) (Protein Data Bank (PDB) 6MRV, 1.8 Å resolution and PDB 6MYV, 2.2 Å resolution, respectively) (Supplementary Table 4). Sialidase26 structure is common to GH33 sialidases (Fig. 2b), including Y509 nucleophilic engagement of C5 following E398 charge activation, D228 acid–base catalysis of the glycosidic bond at C5, and C2 stabilization by an Arg triad (R203, R414, R478). However, typical sialic acid stabilizing interactions are lost, including Glu engagement of

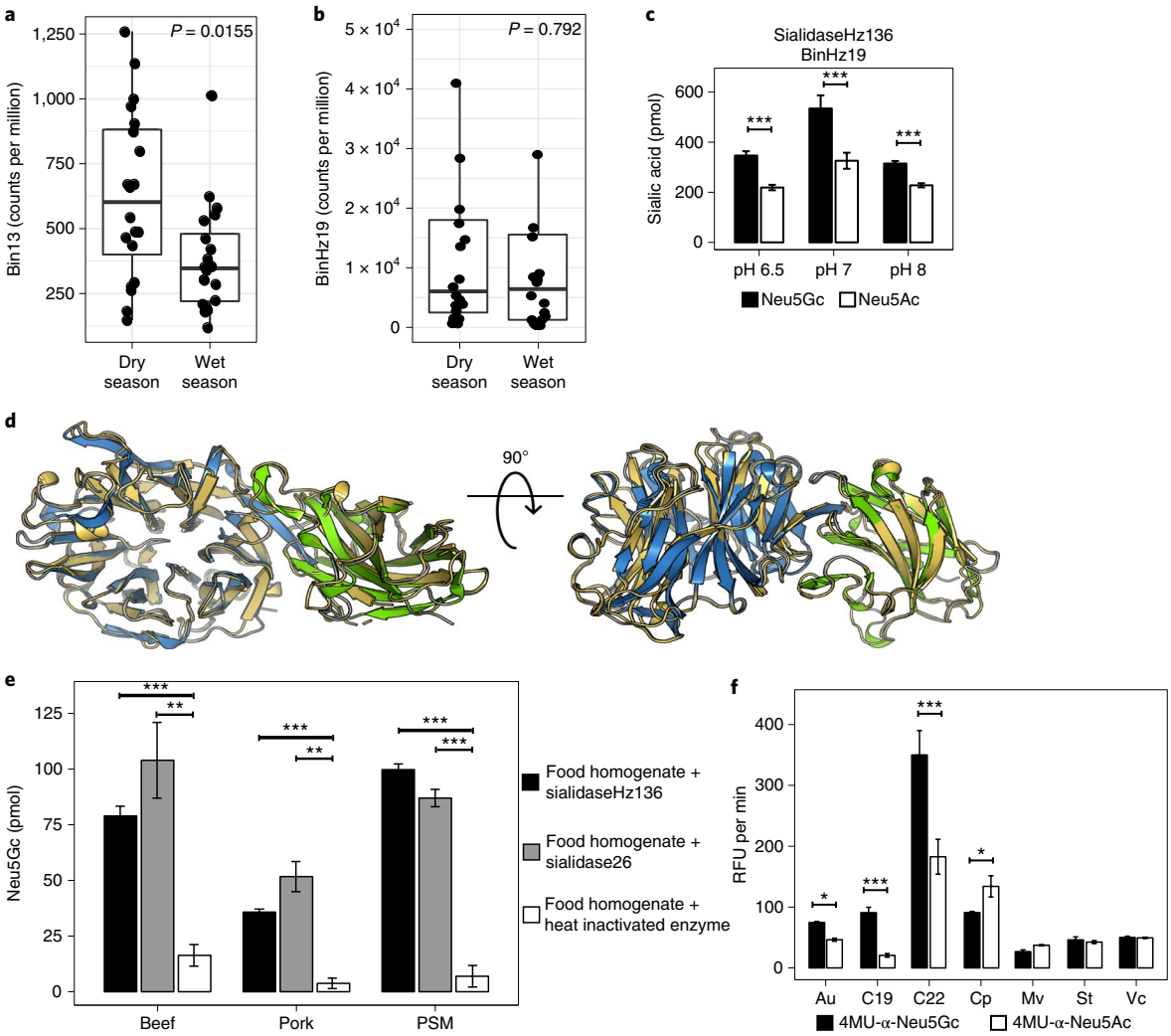


Fig. 3 | Screening for Neu5Gc-preferring sialidases in human and environmental samples. a, Relative abundance of bin13 in samples from Hadza hunter–gatherers population ($n = 20$ biologically independent samples per season). **b**, Relative abundance of binHz19 in samples from Hadza hunter–gatherers population ($n = 20$ biologically independent samples per season). **c**, Activity of sialidaseHz136 retrieved from Hadza shotgun metagenome at three different pHs (6.5, 7 and 8) using 0.5 μg of each enzyme. Data are represented as triplicates ($n = 3$ biologically independent experiments) and the statistical analysis was performed by Student's two-tailed *t*-test. **d**, Crystal structure of sialidaseHz136 (yellow) aligned to the catalytic site (blue) and carbohydrate-binding motif (green) of sialidase26. **e**, Sialidase26 and sialidaseHz136 activity on beef (New York steak), pork (pork sausage) and PSM chow. Data are represented as triplicates ($n = 3$ biologically independent experiments). **f**, Comparison of composting sialidases activity using fluorogenic 4MU-α-Neu5Gc and 4MU-α-Neu5Ac substrates. RFU, relative fluorescence units. The experiments were performed in duplicates ($n = 2$ biologically independent experiments). Au, *Arthrobacter ureafaciens*; C19, composting sialidase C19; C22, composting sialidase C22; Cp, *Clostridium perfringens*; Mv, *Micromonospora viridifaciens*; St, *Salmonella typhimurium*; Vc, *Vibrio cholerae*. **a**,**b**, Statistical significances were determined using the non-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. Boxes and whiskers indicate quartiles and middle maker indicate the median ($n = 20$). **c**,**d**,**f**, Statistical significance was determined by Student's two-tailed *t*-test. $^*P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$. Bars represent geometric mean ± s.e.m.

the glycerol moiety C7-C9 (T397 in sialidase26), C10 stability by an Arg residue (now a Leu), and inward movement of W507 into the binding pocket (although this is restored in the DANA-Gc co-crystal structure). Co-crystallization of sialidase26 with DANA-Gc indicated an overall fold and ligand placement similar to that of DANA-Ac (Fig. 2c). The hydroxyl group at the end of the C5 acetamido group of DANA-Gc points towards the binding pocket residues, forming H contacts (hydrogen bonds) with D271 and probably increasing its stability. Amino acid substitutions at this position to Leu (D271L) or Asn (D271N) significantly lowered sialidase protein activity and eliminated the Neu5Gc preferential cleavage (Fig. 2a). This, in combination with the changes described above, probably explains sialidase26 preference for Neu5Gc.

To extend our findings to humans, we reanalysed faecal shotgun metagenomes from the Hadza (Supplementary Table 5), a genetically distinct indigenous ethnic group that resides in remote Tanzania[31]. The Hadza are hunter–gatherers and change their diet periodically throughout the year according to food availability. In dry seasons, their diet is enriched with meat and tubers, whereas in wet seasons it consists largely of honey and berries[31,32]. Raw reads from Hadza metagenomes were mapped to bin13. We observed that bin13 was significantly more abundant in microbiome samples taken during the dry season compared to the wet season (Fig. 3a). To evaluate the sialidases genes present in the Hadza microbiome, the combined metagenomes were co-assembled and 24 genome bins containing 51 sialidase genes were identified (Supplementary

Fig. 14, Supplementary Table 6). The binHz19, whose closest relative was *Alistipes* sp., was equally abundant in microbiomes from both seasons (Fig. 3b and Supplementary Table 6). BinHz19 also contained a sialidase (sialidaseHz136) with the greatest sequence similarity to sialidase26 identified in our mouse study, suggesting widespread distribution of this sialidase homologue amongst mammals (Supplementary Fig. 15). SialidaseHz136 showed preferential activity for Neu5Gc over Neu5Ac in all tested conditions (Fig. 3c). X-ray crystallography of sialidaseHz136 (PDB 6MNJ, 2.2 Å resolution) exhibited similar predicted engagement of conserved residues with sialic acid substrate when compared to sialidase26 (Fig. 3d), despite a shift in the acetamido-interacting Asp residue. This difference supports the notion that sialidaseHz136 is capable of metabolizing structurally diverse glycans, a concept that would be of benefit to an individual with seasonal variations in diet. To determine if sialidase26 and sialidaseHz136 can release Neu5Gc from food sources directly we tested their activity on beef, pork and PSM chow as substrate. Both enzymes showed pronounced sialidase activity by releasing Neu5Gc from all food tested (Fig. 3e).

Our findings indicate that sialidases exhibiting Neu5Gc preference can be widespread in the mammalian intestine. To determine if Neu5Gc-specific sialidases are restricted to enteric bacteria, we used functional metagenomic screening to identify exo-sialidases with a Neu5Gc specificity in a terrestrial environment. A fosmid library containing ~40 kilobase (kb) inserts of total environmental DNA isolated from soil from an organic composting facility was constructed in *Escherichia coli*. Lysates from 5,376 clones were screened for hydrolysis of 4MU-α-Neu5Gc (Supplementary Fig. 16). Clones showing activity were tested subsequently and re-tested for their ability to cleave 4MU-α-Neu5Ac. Two clones (C19 and C22) showed significant hydrolysis of 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac (Fig. 3f). The DNA sequence of cloned inserts from C19 and C22 encoded a single bacterial sialidase gene. Both enzymes showed exo-sialidase activity and had a marked preference for Neu5Gc hydrolysis in vitro compared to several other known bacterial exo-sialidases (Fig. 3f). Analysis of the deduced amino acid sequences of C19 and C22 showed that they are highly similar to each other (56% identity), both belong to the GH33 family of bacterial exo-sialidases (Supplementary Fig. 17) and are most similar to proteins from terrestrial bacteria of the genera *Rhodopirellula* and *Verrucomicrobia*. Considered together, these data support the conclusion that exo-sialidases with Neu5Gc preference also exist in terrestrial environments where decomposing biological material from animal sources might require such preference.

In summary, we identified several sialidases with previously undescribed preference for Neu5Gc, which are enriched in the gut microbiota of mice and humans on consumption of a Neu5Gc-rich diet. Previous studies from our group showed that dietary free-Neu5Gc is not incorporated in *Cmah*$^{-/-}$ (ref. [14]) mice. Thus, the cleavage of Neu5Gc from foods entering the gut can potentially prevent incorporation of this non-human sugar into the colon tissue. We also hypothesize that the gut microbiome with an under-representation of bacteria with Neu5Gc-preferring sialidases could result in increased xenosialitis and be a potential contributing factor to inflammation-mediated promotion of diseases. Although further in vivo characterization needs to be done, our results lay the foundation to define a strategy for translation of pre- or probiotics to prevent incorporation or to eliminate Neu5Gc from tissues of red meat eaters, thereby reducing the risk of xenosialitis and other diseases associated with red meat consumption.

Methods

In vivo sampling, feeding and animal diet. WT lineage C57BL/6 was purchased from Harlan Laboratories and human-like *Cmah*$^{-/-}$ mice generated as previously described[23]. In brief, the generation of the human-like *Cmah*$^{-/-}$ transgenic mice was performed by target deletion of exon 6 of the *cmah* gene (similar deletion that evolved in humans) using loxP sites and Cre recombinase expression in embryonic stem cells. Transgenic mice were generated by the University of California, San Diego (UCSD), Transgenic Mouse Core. All animal experiments were approved by the UCSD Institutional Animal Care and Use Committee under the protocol number S01227. All animals were maintained in the UCSD vivarium according to guidelines, with 12-h diurnal lighting and access to food and water ad libitum. Sample processing and analysis were not blinded at any step. Samples size was chosen based on the cost, mice, diet and vivarium availability, and the minimum number of samples to perform statistical analysis. WT and *Cmah*$^{-/-}$ mice were raised in the same vivarium room and fed with the same water source. The cages were kept side by side in the same cage rack to minimize external influence on the gut microbiome. Age and sex-matched female mice used in the study were maintained on sialic acid free soy-based diet (Dyets; 110951) from weaning until 8–10 weeks of age to prevent previous exposure to sialic acid. To evaluate the effect of the dietary Neu5Gc in the gut microbiome, sex-matched *Cmah*$^{-/-}$ mice at 10 weeks of age were caged in three groups of five mice each and fed during 4 weeks with the same soy-based diet that was either enriched with Neu5Gc (PSM)[14], Neu5Ac (EBN)[5] or kept in sialic acid-free soy as control. The animals were euthanized in a CO_2 chamber and the colon tissues were cut open with blunted scissors for fresh collection of faecal samples by scraping it straight from the tissue. Colonic faecal scrapings from five mice of each diet type were used for 16S rRNA gene amplicon, and three mice of each diet type were used for metagenomic shotgun sequencing analysis. The feeding protocol was chosen based on previous evidence from our group showing that feeding *Cmah*$^{-/-}$ mice with PSM over a period of weeks can cause mouse tissue incorporation of Neu5Gc at levels histologically similar to the levels seen in adult humans who have eaten red meat for many years[14]. The PSM diet contained 250 µg of Neu5Gc per gram of chow to mimic the amount of Neu5Gc present in beef, the most consumed form of red meat in the Western diet.

Table 1 | Solvents and gradient conditions used for high-performance anion exchange chromatography

Solvents	Gradient setting			
A: Water	Time	%A	%B	%C
B: 100 mM NaOH with 5 mM NaOAc	0	84	16	0
C: 100 mM NaOH with 250 mM NaOAc	20	84	16	0
	21	0	100	0
Pulsed amperometric detector	31	0	100	0
Waveform: standard quad	32	84	16	0
	50	84	16	0

Monosaccharide and amino acid composition of the diets. Monosaccharide composition of the PSM and EBN diets was analysed by High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection (HPAEC–PAD) using Dionex ICS-3000 system (ThermoFisher Scientific) equipped with CarboPac PA1 column 4×250 mm^2, 4 µm, with a 4×50 mm^2 Guard. Briefly, 500 µg of each diet was dissolved in 200 µl of Milli-Q water. The samples were hydrolysed by adding an equal volume of 4 N Trifluoroacetic acid (final concentration of 2 N TFA) at 100 °C for 4 h. The hydrolysed samples were centrifuged at 400*g* for 2 min and evaporated under a flow of dry nitrogen. Once dried, samples were resuspended in 200 µl of Milli-Q water and 50% of each sample was injected. The separation of monosaccharide peaks was achieved using the solvents water (A), 100 mM NaOH with 5 mM NaOAc (B) and 100 mM NaOH with 250 mM NaOAc (C) with pulsed amperometric detector and standard quad waveform, and the gradient conditions in Table 1.

The amino acid composition of both diets was performed by gas chromatography–mass spectrometry (GC–MS) tBDMS (dimethyl-tert-butylsilyl) derivatives quantitation as previously described[33]. HPAEC-PAD and GC–MS analysis were performed by the Glycotechnology Core at the Glycobiology Research and Training Center, UCSD (https://medschool.ucsd.edu/research/GRTC/services/glycoanalytics/Pages/default.aspx).

16S rRNA sequencing and analysis. Total genomic DNA was extracted using MoBio PowerFecal DNA isolation kit (MoBio) following the manufacturer's instructions. Purified DNA was amplified and processed according to the Illumina 16S protocol (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). 16S rRNA libraries were generated from three or five biologic replicates and three independent experiments per diet group. Libraries were quality assessed using quantitative PCR (qPCR) and Bioanalyzer (Agilent Technologies), and subsequently sequenced using two MiSeq 600 cycle kits (Illumina). Adaptors were trimmed from the Illumina data using Trimmomatic v.0.36 (ref. [34]). 16S analysis

was performed with Usearch denoising and the RDP (Ribosomal Database Project) 16S rRNA database v.16. Sequences were analysed using the Usearch v.10 (ref. [35]) following the MiSeq 2×250 16S V4 pipeline (https://www.drive5.com/usearch/manual/pipe_examples.html). In brief, paired-end reads were merged using fastq_mergepairs (-fastq_maxdiffs 10; -fastq_pctid 10). Sequences with a distance-based similarity of 97% or greater were grouped into operational taxonomic units (OTU) using cluster_otus (-minsize 2). OTU tables were rarefied to 10,000 observations per sample. OTU-based microbial diversity and dissimilarity metrics were estimated using R package vegan v.2.5-2. Bray–Curtis distance was used for beta-diversity analysis. Vector fitting with each of the PCoA ordinations was performed using the function envfit from R package vegan v.2.5-2. Significant vectors were selected following the criteria ($R \geq 0.7$ and $P \leq 0.01$). Bacteria genera with relative abundance below 1% were not considered for differential abundance analysis. Statistical differences in abundance between diets were calculated using the non-parametric Wilcoxon rank sum test with Holm correction for multiple hypotheses when appropriated.

Shotgun metagenome sequencing and analysis. As described above, total genomic DNA was extracted using MoBio PowerFecal DNA isolation kit (MoBio) following the manufacturer's instructions. Purified DNA from three biological replicates per diet group was prepared for shotgun metagenomic sequencing using the Nextera XT library preparation method with the average fragment size of 450 base pairs (bp) (Illumina). Libraries were quality assessed using qPCR and a Bioanalyzer (Agilent Technologies) and subsequently sequenced using MiSeq 2×250 bp cycle kits (Illumina). An average of 2.5 million non-mouse reads were generated per library. Adaptors were trimmed from the Illumina data using Trimmomatic v.0.36 (ref. [34]). Samples were filtered of possible human and mouse contamination by aligning the trimmed reads against reference databases using Bowtie2 v.2-2.2.3 (ref. [36]) with the following parameters (-D 20 -R 3 -N 1 -L 20 --very-sensitive-local). Overlapped reads were merged using FLASH v.1.2.11 (ref. [37]). Merged and unmerged reads were assembled using Spades v.3.12.0 with the following parameters (-k 21,33,55,77,99,127 --meta --merge)[38]. Differential binning was performed using MetaBat2 v.2.12.1 with minimum contig length of 1,500 bp (ref. [39]). Bin quality (completeness and contamination) was evaluated using CheckM v.1.0.7 (ref. [40]). Taxonomic classification (closest phylogenetic neighbour) was assessed using the RAST (rapid annotation using subsystem technology) online tool[41]. In brief, RAST uses a set of unique proteins to assign the closest related neighbour. Genome annotations were performed using Prokka v.1.11 with default parameters. Amino acid sequences of all genes identified using Prokka were aligned to sialidase sequences obtained from Uniprot up to October 2018. The alignment was performed using Diamond v.0.8.24 with the parameters blastx -k 5 -f 6 --evalue 0.001. Relative bin abundance was obtained by dividing the total number of reads aligned per bin by the total number of reads aligned in all bins. For this analysis, non-mouse trimmed reads were aligned to the binned genomes using Bowtie2 v.2-2.2.3 (ref. [36]) with the parameters set by the flag --very-sensitive. Bins with relevant genes were compared for relative abundance between sample groups using the Wilcoxon rank sum test. The CAZymes database[27] was used to identify diet-dependent sialidases. Non-mouse trimmed reads were aligned against sialidase amino acid sequences using Diamond v.0.8.24 using the parameters -k 5 -f 6 --evalue 0.001. When indicated, counts per million were used to normalize the number reads aligned by protein sequence. The total number of reads by protein sequence were rarefied to 5,000 observations per sample using the R function rrarefy from R package vegan v.2.5-2. Bray–Curtis distance was used for CAZyme diversity analysis. Amino acid sequence similarities were evaluated using BLAST (basic local alignment search tool) online tool and pairwise sequence alignments were generated using Clustal Omega online tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Sialidases per bacterial genome were obtained from PATRIC (Pathosystems Resource Integration Center) database using command-line interface P3-scripts.

Analysis of previously published shotgun metagenomic data. Previously published foecal shotgun metagenomic data from Hadza hunter–gatherer individuals[31] were obtained from Sequence Read Archive repository under the project IDs PRJNA392012 and PRJNA392180. Forty shotgun metagenomic data from individuals were analysed, 20 samples of which were collected during the wet season and 20 during the dry season (Supplementary Table 5). All published shotgun metagenomic data were processed as described in the previous section.

Protein expression and assay. Target sialidase sequences from shotgun metagenomic data were PCR amplified from genomic DNA isolated as described above or ordered from Integrated DNA Technologies, subcloned into a pET19b expression vector with a C-terminal 10xHis tag and N-terminal truncation to remove any signal peptide sequence (predicted by SignalP 4.1, CBS) and transformed into BL21(DE3) *E. coli* (MilliporeSigma) using established heat-shock methods. Cells were grown to optical density 0.6–0.8 (OD 600) in multiple 1-l cultures at 37 °C and induced overnight at 25 °C with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Harvested cells were resuspended in lysis buffer (50 mM HEPES pH 8.0 buffer, 50 mM NaCl and 1 mM TCEP (tris(2-carboxyethyl) phosphine hydrochloride)) with DNaseI and hen egg white lysozyme, lysed with a TS-Series cell disruptor (Constant Systems) at 15 KPSI (kilo-pound per square inch), and spun for 45 min at 186,000*g* with a Ti45 ultracentrifugation rotor (Beckman Coulter) to remove cell debris. Purification was performed as below and based on purification of a putative *Bacteroides* neuraminidase as provided by the Protein Structure Initiative (BACCAC_01090, Joint Center for Structural Genomics, to be published), with modifications to imidazole stringency based on the purified sialidase. Supernatant was loaded on a 5-ml HisTrap Ni affinity column (nickel-charged columns for high resolution histidine-tagged protein purification) on an Akta Explorer purification system (GE Healthcare Life Sciences) with 20–40 mM imidazole added, washed with Running Buffer (50 mM HEPES pH 8.0 buffer, 300 mM NaCl, 40–60 mM imidazole, 10% glycerol and 1 mM TCEP) and eluted with Elution Buffer (20 mM HEPES pH 8.0 buffer, 300 mM imidazole, 10% glycerol and 1 mM TCEP). Samples were concentrated using 10–30 kDa Amicon centrifugal filters (MilliporeSigma) at 1,500*g* to 1 ml and desalted over a 5-ml desalting column using the Akta system into Desalting Buffer (20 mM HEPES pH 8.0 buffer, 200 mM NaCl). The resulting protein sample was diluted as required for functional studies.

Assay for sialidase activity. *In vitro activity*. Sialidases purified as described above were quantified using SDS–PAGE image analysis with BSA (bovine serum albumin) references (Bio-Rad) and absorbance using the Nanophotometer P330 (Implen), with extinction coefficients calculated using ExPASy Translate (https://web.expasy.org/translate/). Sialidase activity assays were performed in a dilution series: 0.5 µg, 2.5 µg, 5 µg and 10 µg of each enzyme were incubated with equal amounts of human-like *Cmah*$^{-/-}$ mouse serum (Neu5Ac: 1,428.64 pmol µl^{-1}, Neu5Gc: 0 pmol µl^{-1}) and WT mouse serum (Neu5Ac: 196.29 pmol µl^{-1}, Neu5Gc: 1,337.45 pmol µl^{-1}) for 1 h at 37 °C. An additional 10 µg of each enzyme was inactivated by heat for 5 min at 95 °C. The samples were kept at −20 °C until derivatization and analysed by HPLC as described below. The DMB reagent was made with the following recipe: 14 mM DMB (1,2-diamino-4,5-methylenedioxybenzene, Sigma D4787), 18 mM sodium hydrosulfite (Sigma 157953), 1.0 M 2-mercaptoethanol (Sigma M3148) and 40 mM trifluoroacetic acid (Sigma T6508), and incubated at 50 °C for 2.5 h. The DMB-derivatized samples were analysed on a Dionex Ultra3000 HPLC System using a Phenomenex Gemini 5µ C18 250×4.6-mm HPLC column at room temperature, eluted in isocratic mode with 85% water, 7% methanol and 8% acetonitrile. The same protocol was used to evaluate the sialidases activity in three different pH ranges.

In vitro activity. In vivo activity using clarified faecal samples: Faecal pellets were collected from 10-week-old *Cmah*$^{-/-}$ mice caged in two groups of four mice and three mice, which were fed over 2 weeks with PSM[14] or sialic acid-free soy, respectively. All the animals' information and maintenance are the same as described in In vivo sampling, feeding and animal diet. Fresh faecal pellets were homogenized (10% w/v) in Tris–HCl pH 7 buffer with protease inhibitor (Roche 11836170001) using Pellet Pestle Cordless Motor (Kimble 749540). Samples were spined at 4,000*g* for 3 min and the supernatant was transferred to SpinX filters (Costar 8160). Samples were centrifuged at 14,000*g* for 10 min and the flow through was transferred to a clean tube. Total protein concentration was determined using Pierce BCA (bicinchoninic acid) Protein Assay (Thermo Scientific 23225). To measure sialidase activity in clarified faecal samples, 2 µg of total protein per sample was incubated with equal amount of human-like *Cmah*$^{-/-}$ mouse serum and WT mouse serum for 2 h at 37 °C. An additional 2 µg of total protein per sample was inactivated by heat for 5 min at 95 °C. Samples were cleaned using 10 kDa Amicon centrifugal filters (MilliporeSigma), 14,000*g* for 25 min at 4 °C. Samples were lyophilized and resuspended in 50 µl of sterile water. Derivatization and HPLC analysis was performed as described in In vitro activity.

Food sources. Beef (New York steak) and pork (breakfast sausage) were purchased at Whole Foods Market and cut into small pieces using sterile blades. PSM chow was powdered using a crucible and pestle. Beef, pork and PSM were homogenized (10% w/v) in Tris–HCl pH 7 buffer using Pellet Pestle Cordless Motor (Kimble 749540), and 50 µl of each suspension was incubated with 2 µg of either sialidase26 or sialidaseHz136 for 2 h at 37 °C. An additional 2 µg of each sialidase was inactivated by heat for 5 min at 95 °C. Reactions were performed in triplicates. Samples were cleaned using 10 kDa Amicon centrifugal filters (MilliporeSigma), 14,000*g* for 25 min at 4 °C. Derivatization and HPLC analysis was performed as described in In vitro activity.

Neu5Gc2en synthesis. Neu5Gc2en (DANA-Gc) was synthesized as previously published[42–45] with minor modifications. Briefly, Neu5Gc (Sigma-Aldrich) was treated with Dowex 50W-X8 (H^+) resin in MeOH for 20 h at 20 °C to form the methyl ester. This ester was then treated with acetic anhydride and pyridine for 42 h at 20 °C to generate the peracetylated methyl ester, which was purified by column chromatography (50:1 $CHCl_3$–MeOH). This sample was treated with TMSOTf under dry nitrogen at 0 °C in MeCN for 6 h to induce elimination. The unsaturated compound was purified by chromatography (toluene/acetone, 3:1 → 2:1). The acetyl groups were cleaved by treatment with NaOH in MeOH over 12 h, followed by neutralization with H^+ resin.

Crystallization and structure determination of sialidase26 and sialidaseHz136. Purified sialidase26 was concentrated to 8 g l^{-1} and set on sitting drop trays in 1:1 volume ratios with the mother liquor (20% PEG (polyethylene glycol) 6000, 0.1 M Tris–HCl pH 8.0) at 16 °C. For ligand co-crystals, concentrated sialidase26 was incubated for 1 h at room temperature with 5 mM *N*-acetyl-2,3-dehydro-2-deoxyneuraminic acid (Neu5Ac2en or DANA, Sigma-Aldrich) or DANA-Gc (synthesized). Crystals appeared after 3–4 d and grew to full size in 8–10 d. Crystals were soaked briefly in the mother liquor, supplemented with 10% glycerol and flash-frozen with liquid nitrogen. Purified sialidaseHz136 was crystallized similarly (20% PEG 3350, 0.1 M Bis-Tris-Propane pH 7.5, 0.2 M sodium citrate). X-ray diffraction data were collected at 100 K at the Lawrence Berkeley National Laboratory Advanced Light Source (8.2.1 and 8.2.2) at a single wavelength. Preliminary diffraction data were collected at the Stanford Synchrotron Radiation Lightsource and Advanced Photon Source. All diffraction data were indexed and integrated with XDS (X-ray detector software) (2018-Jun-08) or MOSFLM (measurement of oscillation FiLMs), processed with AIMLESS v.0.7.2, and truncated with CTRUNCATE within the CCP4 v.7.0.073 suite of programs[46–49]. Phases were estimated via molecular replacement in PHENIX.PHASER v.1.13 using a previously published model of an uncharacterized *Bacteroides*-derived sialidase with high sequence homology to sialidase26 (PDB 4q6k) as a search model. Models underwent rigid-body and restrained positional refinement using PHENIX.REFINE in the PHENIX software suite v.1.13 (ref. [50]) against a maximum likelihood target function and alternated with manual inspection against electron density maps in Coot v.0.8.9.1 (ref. [51]). Geometry restraints for DANA were generated using PHENIX.eLBOW[52], with manual inspections in Coot and refined in the final rounds of refinement, which also included the application of hydrogens to their riding positions and simulated annealing. For the co-crystal structure of sialidase26 and DANA-Gc, diffraction data from two crystals with similar unit cell parameters were processed together in AIMLESS and multi-crystal averaging performed with PHENIX.MULTI_CRYSTAL_AVERAGE for five cycles following Phaser of each before refinement proceeded with PHENIX.REFINE. The resulting refinement statistics for each model are included in Supplementary Table 4. Figures displaying crystal packing were prepared using PyMOL v.1.8.3.2 (http://www.pymol.org), and atomic coordinates and structure factors were deposited with the Protein Data Bank (accession codes 6MNJ, 6MRV, 6MRX and 6MYV). Ramachandran statistics for sialidase26 alone (6MRX), with DANA (6MRV) or with DANA-Gc (6MYV) and SiaHz136 (6MNJ) are, respectively, outliers (0.39, 0.19, 0.24 and 0.39%) and favoured (95.67%, 96.06%, 95.87%, 95.27%). Advanced Light Source beamline v.8.2.2 was used for all datasets, apart from 6MYV that was collected on v.8.2.1. Wavelengths used for the same datasets are 0.99999 Å, 0.99995 Å, 1.00003 Å and 0.99994 Å, respectively. All were collected at 100 K. No Cβ deviations were observed for any dataset.

Enzyme discovery Metagenomic DNA library construction. Envinmental DNA was isolated from soil obtained from a commercial organic composting facility in Hamilton, MA (Brick End Farms) by phenol/chloroform extraction and isopropanol precipitation. A fosmid library was produced using the CopyControl Fosmid Library Production Kit (Lucigen Corporation) as recommended. Briefly, DNA was end-repaired and size-selected using a 1% low melting point agarose gel run overnight at 35 V. DNA fragments from 30–70 kb were isolated from the gel using 1 U of β-agarose I (New England Biolabs) for each 100 µl of melted agarose. The end-repaired and size-selected DNA was ligated to the pCC1FOS cloning vector. Resulting clones were packaged in phage particles. *E. coli* EPI300 T1R cells were transfected with the packaging reaction and plated on lysogeny broth agar medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g MgCl2•6H2O, chloramphenicol 12.5 µg ml^{-1}, 2 ml 2 M NaOH and 20 g l^{-1} agar) and incubated overnight at 37 °C. A total of 5,376 colonies were archived in fourteen 384-well plates in sterile 20% (v/v) glycerol.

Screening for sialidase activity. The compost metagenomic library (theoretically encoding about 215,000 environmental genes) was differentially screened with fluorogenic 2′-(4-methylumbelliferyl)-α-D-*N*-glycolylneuraminic acid (4MU-α-Neu5Gc) (Sussex Research) and 2′-(4-methylumbelliferyl)-α-D-*N*-acetylneuraminic acid 4MU-α-Neu5Ac (Toronto Research Chemicals) substrates. In a primary screen, library clones were grown in 384-well plates containing 50 µl lysogeny broth liquid cultures (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g $MgCl_2$–$6H_2O$, 2 ml of 2 M NaOH l^{-1}, containing chloramphenicol 12.5 µg ml^{-1} and 1× inducing solution (Lucigen Corporation)) overnight at 37 °C. Fifty microlitres of Y-per lysis buffer (Thermo Fischer Scientific) containing 40 µg ml^{-1} of 4MU-α-Neu5Gc was added to each well. The mixtures were incubated overnight at 37 °C in a static incubator. Fluorescence at wavelengths of excitation λ_{ex} = 365 nm and emission λ_{em} = 445 nm was read with a SpectraMax Plus 384 Microplate Reader (Molecular Devices) at 6 h, 24 h and 48 h. Positive clones were defined as those showing fluorescence greater than 3 s.d. above the mean background, and each was re-archived in a fresh 384-well plate in sterile 20% (v/v) glycerol. Each of the positive clones was grown and comparatively re-screened in separate assays containing 4MU-α-Neu5Ac and 4MU-α-Neu5Gc substrates (reactions run in duplicate). Two clones designated C19 and C22 showed significant activity on 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac and were subjected to further study.

Reconstructed metabolic models analysis. Genome-scale network reconstructions combine detailed biochemical and physiological information, providing insights into the metabolism for subsequent manipulation strategies[53] or to control metabolism[54,55]. The scope of these models encompasses the characterization of the metabolic behaviour of target microorganisms. We evaluated growth phenotypes of microorganisms of the gut microbiota associated with sialidase metabolism (Enzyme Commission number 3.2.1.18). Seventeen microorganisms of the gut microbiota containing sialidases were identified by scanning the repository of the gut microbiome metabolic models[26]. Growth rates were simulated using flux balance analysis. All metabolic models were constrained using a Western diet (45% fat, 35% carbohydrate, 20% protein), containing experimental constraints for 20 sugars, 24 fibre-related metabolites, 12 fatty acids, 20 amino acids and 88 minerals, vitamins and other metabolites. Experimental constraints are reported in detail in the gut microbiota repository[26]. All metabolic models were simulated using the Gurobi Optimizer v.5.6.3 (Gurobi Optimization) solver in MATLAB (The MathWorks) with the COBRA (constraint-based reconstruction and analysis) Toolbox[56]. Additionally, the contribution to growth of the metabolites associated with sialidase activity (for example, *N*-acetylneuraminate) was determined using shadow prices simulations[56].

Statistics and reproducibility. Statistical analysis was performed using R programming language. The statistical significance of differential relative abundance (16S rRNA amplicon and shotgun metagenomics) was computed using no-parametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. The statistical significance of sialidase activity was computed using Student's two-tailed *t*-test. The significance levels are $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$. All analysis was performed in biologically independent animals or independent experiments as indicated in the text. Samples were not randomized. All data were used in the analysis.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data supporting the findings of this study are available under accession number PRJNA505660. X-ray crystallographic data that support the findings of this study have been deposited in the RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank (accession codes: 6MRX, 6MRV, 6MYV and 6MNJ).

Code availability

The code used to generate the figures and for statistical analysis can be accessed from the corresponding author upon request.

Received: 16 January 2019; Accepted: 16 August 2019;
Published online: 23 September 2019

References

1. Hall, A. B., Tolonen, A. C. & Xavier, R. J. Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* **18**, 690–699 (2017).
2. Mann, N. Dietary lean red meat and human evolution. *Eur. J. Nutr.* **39**, 71–79 (2000).
3. Etemadi, A. et al. Mortality from different causes associated with meat, heme iron, nitrates, and nitrites in the NIH-AARP Diet and Health Study: population based cohort study. *BMJ* **357**, 1–11 (2017).
4. Alisson-Silva, F., Kawanishi, K. & Varki, A. Human risk of diseases associated with red meat intake: analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid. *Mol. Aspects Med.* **51**, 16–30 (2016).
5. Samraj, A. N. et al. A red meat-derived glycan promotes inflammation and cancer progression. *Proc. Natl Acad. Sci. USA* **112**, 542–547 (2015).
6. Tangvoranuntakul, P. et al. Human uptake and incorporation of an immunogenic non-human dietary sialic acid. *Proc. Natl Acad. Sci. USA* **100**, 12045–12050 (2003).
7. Samraj, A. N., Laubli, H., Varki, N. & Varki, A. Involvement of a non-human sialic acid in human cancer. *Front. Oncol.* **4**, 1–13 (2014).
8. Varki, A. Uniquely human evolution of sialic acid genetics and biology. *Proc. Natl Acad. Sci. USA* **107**, 8939–8946 (2010).
9. Dhar, C., Sasmal, A. & Varki, A. From 'serum sickness' to 'xenosialitis': past, present, and future significance of the non-human sialic acid Neu5Gc. *Front. Immunol.* **10**, 807 (2019).
10. Almagro-Moreno, S. & Boyd, E. F. Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. *Infect. Immun.* **77**, 3807–3816 (2009).
11. McDonald, N. D., Lubin, J.-B., Chowdhury, N. & Boyd, E. F. Host-derived sialic acids are an important nutrient source required for optimal bacterial fitness in vivo. *MBio* **7**, e02237–15 (2016).

12. Lewis, A. L. & Lewis, W. G. Host sialoglycans and bacterial sialidases: a mucosal perspective. *Cell. Microbiol.* **14**, 1174–1182 (2012).
13. Kawanishi, K. et al. Human species-specific loss of CMP-*N*-acetylneuraminic acid hydroxylase enhances atherosclerosis via intrinsic and extrinsic mechanisms. *Proc. Natl Acad. Sci. USA* **116**, 16036–16045 (2019).
14. Banda, K., Gregg, C. J., Chow, R., Varki, N. M. & Varki, A. Metabolism of vertebrate amino sugars with *N*-glycolyl groups: mechanisms underlying gastrointestinal incorporation of the non-human sialic acid xeno-autoantigen *N*-glycolylneuraminic acid. *J. Biol. Chem.* **287**, 28852–28864 (2012).
15. Almagro-Moreno, S. & Boyd, E. F. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol. Biol.* **9**, 118 (2009).
16. Almagro-Moreno, S. & Boyd, E. F. Bacterial catabolism of non-ulosonic (sialic) acid and fitness in the gut. *Gut Microbes* **1**, 45–50 (2010).
17. Li, J. & McClane, B. A. NanI sialidase can support the growth and survival of *Clostridium perfringens* strain F4969 in the presence of sialyated host macromolecules (mucin) or Caco-2 cells. *Infect. Immun.* **86**, e00547–17 (2018).
18. Tailford, L. E. et al. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat. Commun.* **6**, 7624 (2015).
19. Kim, S., Oh, D. B., Kang, H. A. & Kwon, O. Features and applications of bacterial sialidases. *Appl. Microbiol. Biotechnol.* **91**, 1–15 (2011).
20. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: a mini-review. *Biochem. Soc. Trans.* **44**, 166–175 (2016).
21. Chokhawala, H. A., Yu, H. & Chen, X. High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries. *Chembiochem* **8**, 194–201 (2007).
22. Huang, Y. L., Chassard, C., Hausmann, M., Von Itzstein, M. & Hennet, T. Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nat. Commun.* **6**, 8141 (2015).
23. Hedlund, M. et al. *N*-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution. *Mol. Cell. Biol.* **27**, 4340–4346 (2007).
24. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
25. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289–306 (2012).
26. Magnúsdóttir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
27. Huang, L. et al. DbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* **46**, D516–D521 (2018).
28. Owen, C. D. et al. Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont *Ruminococcus gnavus*. *Nat. Commun.* **8**, 2196 (2017).
29. Inoue, S. et al. A unique sialidase that cleaves the Neu5Gcα2→5-OglycolylNeu5Gc linkage: comparison of its specificity with that of three microbial sialidases toward four sialic acid dimers. *Biochem. Biophys. Res. Commun.* **280**, 104–109 (2001).
30. Davies, L. R. L. et al. Metabolism of vertebrate amino sugars with *N*-glycolyl groups: resistance of α2-8-linked *N*-glycolylneuraminic acid to enzymatic cleavage. *J. Biol. Chem.* **287**, 28917–28931 (2012).
31. Smits, S. A. et al. Seasonal cycling in the gut microbiome of the Hadza hunter–gatherers of Tanzania. *Science* **357**, 802–805 (2017).
32. Schnorr, S. L. et al. Gut microbiome of the Hadza hunter–gatherers. *Nat. Commun.* **5**, 3654 (2014).
33. Wood, P. L., Khan, M. A. & Moskal, J. R. Neurochemical analysis of amino acids, polyamines and carboxylic acids: GC–MS quantitation of tBDMS derivatives using ammonia positive chemical ionization. *J. Chromatogr. B* **831**, 313–319 (2006).
34. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
35. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
37. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
38. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
39. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
40. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
41. Aziz, R. K. et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genom.* **9**, 75 (2008).
42. Li, Y. et al. Identifying selective inhibitors against the human cytosolic sialidase NEU2 by substrate specificity studies. *Mol. Biosyst.* **7**, 1060–1072 (2011).
43. Bill Cai, T., Lu, D., Landerholm, M. & Wang, P. G. Sialated diazeniumdiolate: a new sialidase-activated nitric oxide donor. *Org. Lett.* **6**, 4203–4205 (2004).
44. Ercégovic, T. & Magnusson, G. Highly stereoselective α-sialylation. Synthesis of GM3-saccharide and a bis-sialic acid unit. *J. Org. Chem.* **60**, 3378–3384 (1995).
45. Numata, M., Sugimoto, M., Shibayama, S. & Ogawa, T. A total synthesis of hematoside, α-NeuGc-(2→3)-β-Gal-(1→4)-β-Glc-(1→1)-Cer. *Carbohydr. Res.* **174**, 73–85 (1988).
46. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281 (2011).
47. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D.* **62**, 72–82 (2006).
48. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D.* **67**, 235–242 (2011).
49. Kabsch, W. XDS. *Acta Crystallogr. D.* **66**, 125–132 (2010).
50. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
52. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
53. Tan, J., Zuniga, C. & Zengler, K. Unraveling interactions in microbial communities—from co-cultures to microbiomes. *J. Microbiol.* **53**, 295–305 (2015).
54. Zengler, K. & Zaramela, L. S. The social network of microorganisms—how auxotrophies shape complex communities. *Nat. Rev. Microbiol.* **16**, 383–390 (2018).
55. Zuñiga, C., Zaramela, L. & Zengler, K. Elucidation of complexity and prediction of interactions in microbial communities. *Microb. Biotechnol.* **10**, 1500–1522 (2017).
56. Schellenberger, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).

Acknowledgements

We thank all Zengler-, Varki- and Chang-lab members for helpful discussions. Research was supported in part by the National Institutes of Health under award no.R01GM32373 (to A.V.) and by the National Science Foundation under award no. IOS-1444435 (to G.C.). C.M. was supported by grants from the National Institutes of Health, USA (NIH grant no. T32GM8806) and by a Chancellor's Research Excellence Scholarship (UCSD). F.A.-S. was partly supported by the Program Science Without Borders Bex 9254-13-7-Capes Brazil.

Author contributions

L.S.Z., C.M., F.A.-S., A.V. and K.Z. conceptualized the study. L.S.Z., F.A.-S. and K.Z. wrote the manuscript with input from all authors. L.S.Z. and C.M. performed and analysed the microbiome experiments. L.S.Z. and S.L.D. performed the enzymatic characterization. L.S.Z., F.A.-S. and P.S. performed the animal work. S.D.R. performed the protein expression and crystallization experiments, assisted by J.H. and D.S. C.Z. performed the metabolic model analysis. L.C., M.B.G. and C.H.T. constructed the fosmid library and performed the compost sialidase activity assays. C.H.T., G.C., A.V. and K.Z. provided resources and supervised the study.

Competing interests

L.S.Z., C.M., F.A.-S., S.R., S.L.D., G.C., A.V. and K.Z. have filed a patent application (number pending) that claims the use of sialidases to reduce Neu5Gc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0564-9.

Correspondence and requests for materials should be addressed to K.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Supplementary Information for:

Gut bacteria responding to dietary change encode sialidases that exhibit preference for red meat-associated carbohydrates

Livia S. Zaramela[1#], Cameron Martino[1,2#], Frederico Alisson-Silva[3,4#§], Steven D. Rees[5], Sandra L. Diaz[3,4], Léa Chuzel[6], Mehul B. Ganatra[6], Christopher H. Taron[6], Patrick Secrest[3,4], Cristal Zuñiga[1], Jianbo Huang[5], Dionicio Siegel[5], Geoffrey Chang[5], Ajit Varki[3,4] and Karsten Zengler[1,7,8*]

[1]Department of Pediatrics, University of California, San Diego; [2]Bioinformatics and Systems Biology Program, University of California, San Diego; [3]Department of Medicine and Cellular & Molecular Medicine, University of California, San Diego; [4]Glycobiology Research and Training Center (GRTC); [5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego; [6]New England Biolabs; [7]Departmemt of Bioengineering, University of California, San Diego; [8]Center for Microbiome Innovation, University of California, San Diego.

[#]These authors contributed equally to this work.

*Corresponding author.

[§]Current affiliation: Paulo de Goes Institute of Microbiology, Federal University of Rio de Janeiro, Brazil.

Contents

Supplementary Discussions

Evaluating growth phenotypes associated with the exo-alpha-sialidase activity in the gut microbiome

Genome-scale network reconstructions combine detailed biochemical and physiological information, providing new insights into the metabolism for subsequent manipulation strategies to enhance productivity or to control the metabolism[1]. The scope of these models encompasses the characterization of the metabolic behavior of target microorganisms. Here, we evaluated the growth phenotypes associated with the sialidase activity (EC: 3.2.1.18) in microorganisms of the human gut microbiome. We identified microorganisms containing sialidase activity and simulated changes in the growth rate among different taxonomic groups.

Out of 773 metabolic models available for microorganism in the gut microbiome[2] (Supplementary Figure 5a) we identified 17 microorganism containing putative sialidase activity (Supplementary Figure 5b). Most of the identified organisms fall into the taxonomic group of Bacteroidetes. Phenotypes regarding growth rate showed that microorganism carrying sialidase activity can growth around 0.1-0.51 1/h, being Bacteroidetes the most efficient growing microorganisms with an average growth rate of 4.8±0.35 1/h. Modeling predictions show that six out of seven Bacteroidetes can reshape their growth rate depending on the sialidase activity. The maximum change in the growth rate is 11±2% of the wild type growth rate. The growth change due to sialidase activity is shown in shaded red in Supplementary Figure 5c.

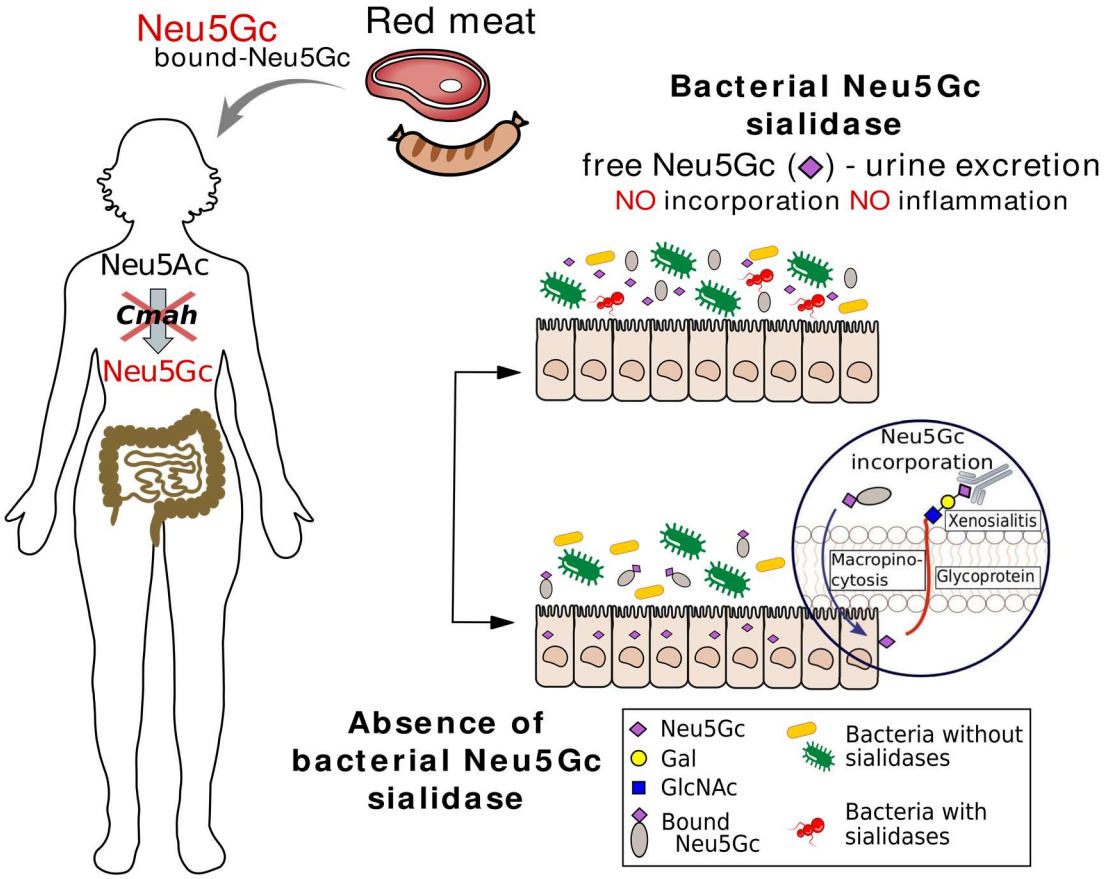
Sialidases substrate preference

To the best of our knowledge, no previously characterized exo-sialidases have been shown to prefer Neu5Gc over Neu5Ac. A single report has shown that a sialidase from the oyster species *Crassostrea virginica* cleaves the very rare glycan structure Neu5Gcα2,5-O(glycolyl)Neu5Gc more efficiently than Neu5Acα2-8Neu5Ac. Because the authors did not compare enzyme substrate preference using a glycan structure with same terminal linkage but differing in Neu5Gc vs Neu5Ac, it cannot be concluded that the oyster sialidases actually prefers Neu5Gc-containing substrates[3]. Similarly, the sialyltransferase from the pathogenic bacteria *Pasteurella multocida*[4] has been shown to cleaves Neu5Gc more efficiently than Neu5Ac, however, no assay was performed in order to evaluate preferential cleavage. In fact, previous studies revealed that many bacterial species such as *Arthrobacter ureafaciens*, *Clostridium perfringens*, *Streptococcus sp*., *Salmonella typhimurium, Streptococcus pneumonia, Ruminococcus gnavus* and *Vibrio cholera* express sialidases that hydrolyse Neu5Ac-terminated

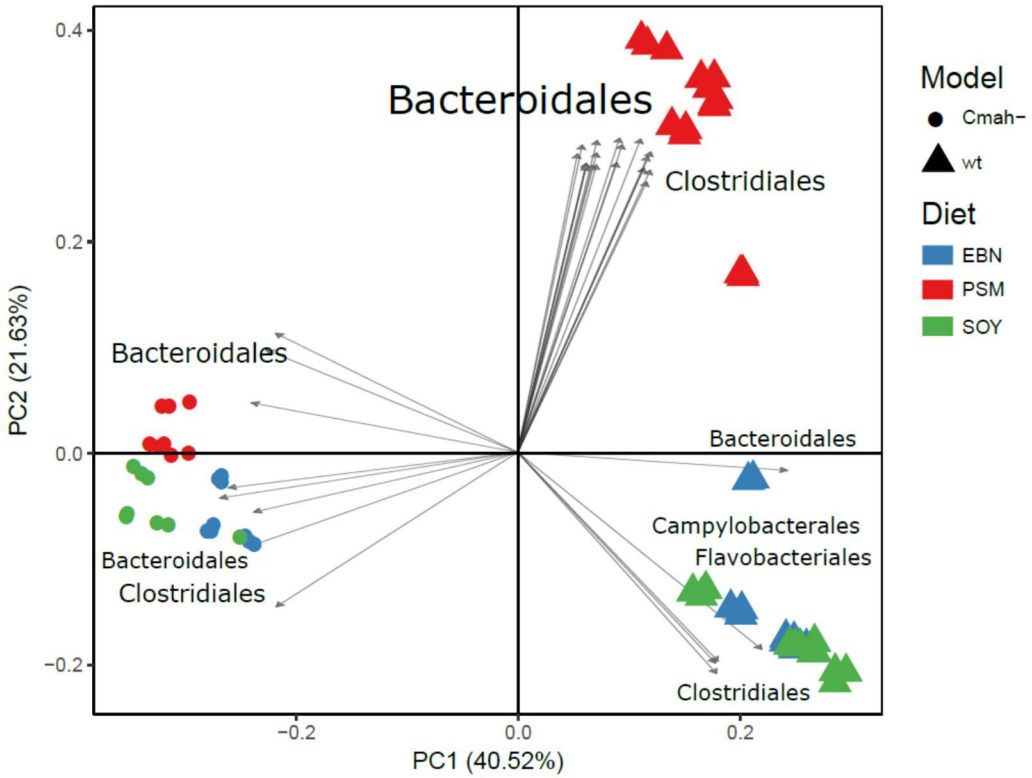
glycans more efficiently than Neu5Gc-terminated glycans[4–6], a substrate preference that is also shared by the murine sialidases Neu1, Neu2 and Neu4[4].

Comparison between sialidase26 and existing sialidases structures

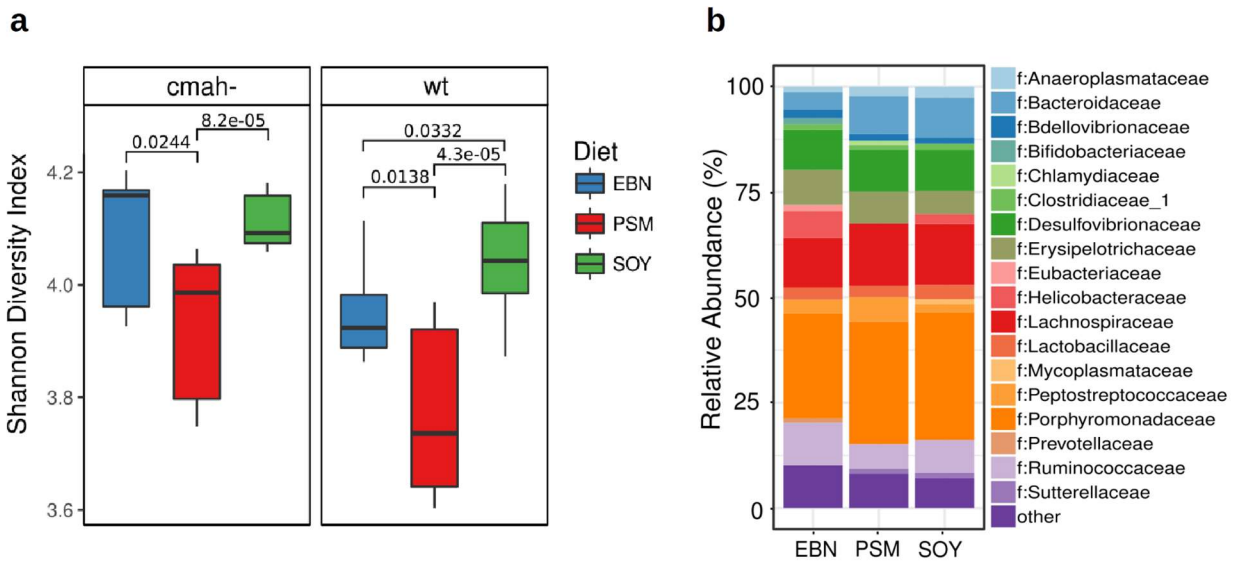
The sialidase26 identified in this study was compared with published sialidase structures. Despite this novel substrate preference, sequence residues that are predicted to interact with terminal Sias in the catalytic site are highly also fully conserved with structurally studied sialidases exhibiting no known Neu5Gc preference over Neu5Ac, including those from *B. caccae* (PDB 4Q6K), *B. thetaiotamicron*[7], and *P. distasonis* (PDB 4FJ6). Other known sialidases with very low sequence identity (<30%) nevertheless shared 50-75% identity with residues predicted to interact with sialic acid, including NanA and NanB from *S. pneumoniae*[8–10] NanI from *C. perfringens*[11], an IT-sialidase from *R. gnavus*[6], and a trans-sialidase from *M. decora*, but suggest no clear sequence motif or mutation causing Neu5Gc preference.



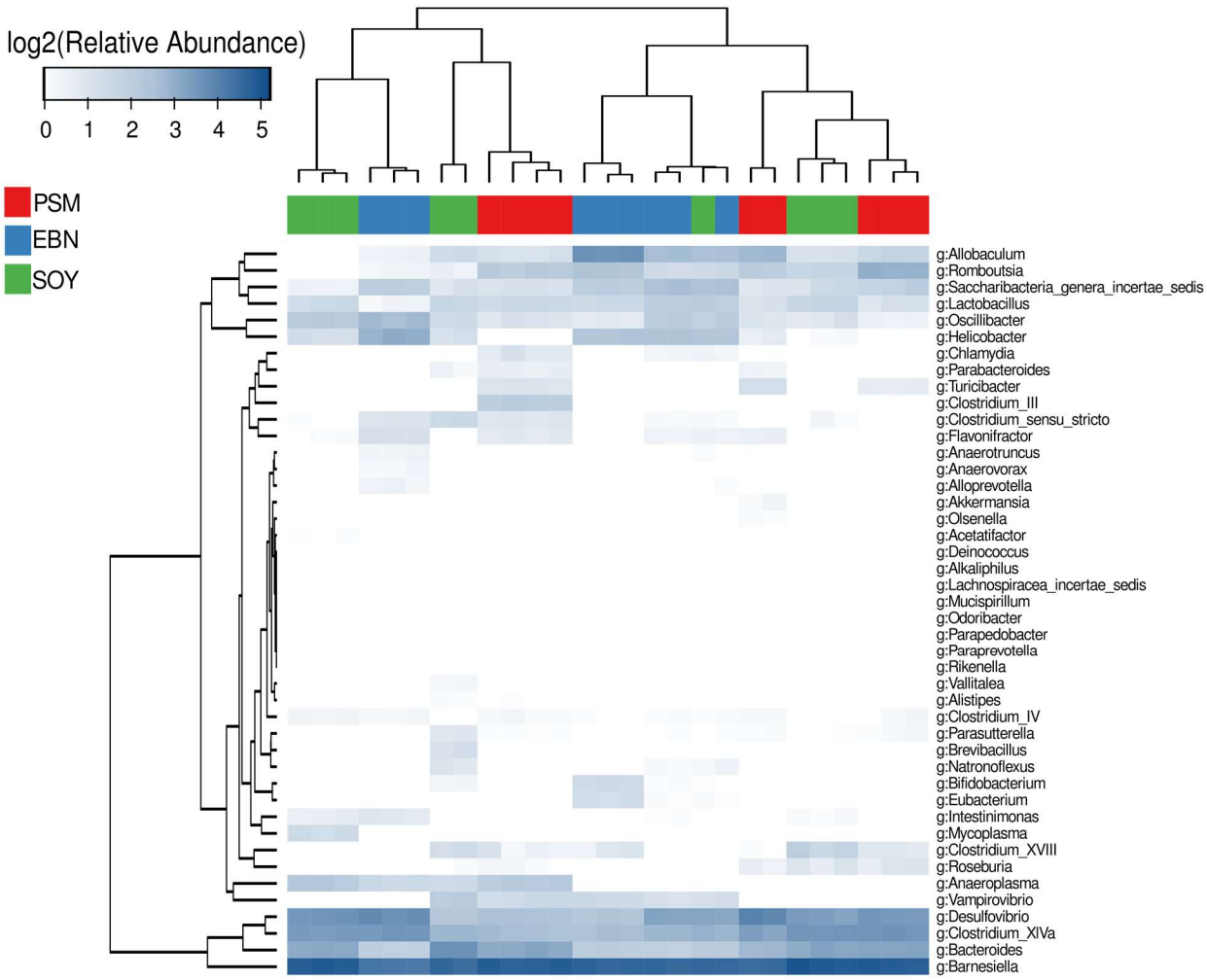
Supplementary Figure 1: Proposed scheme for the incorporation of bound-Neu5Gc. In absence of sialidases, Neu5Gc is incorporated into human tissue triggering xenosialitis. In the presence of bacterial sialidases, Neu5Gc is released from red meat glycoproteins and metabolized by bacteria or excreted in the urine.



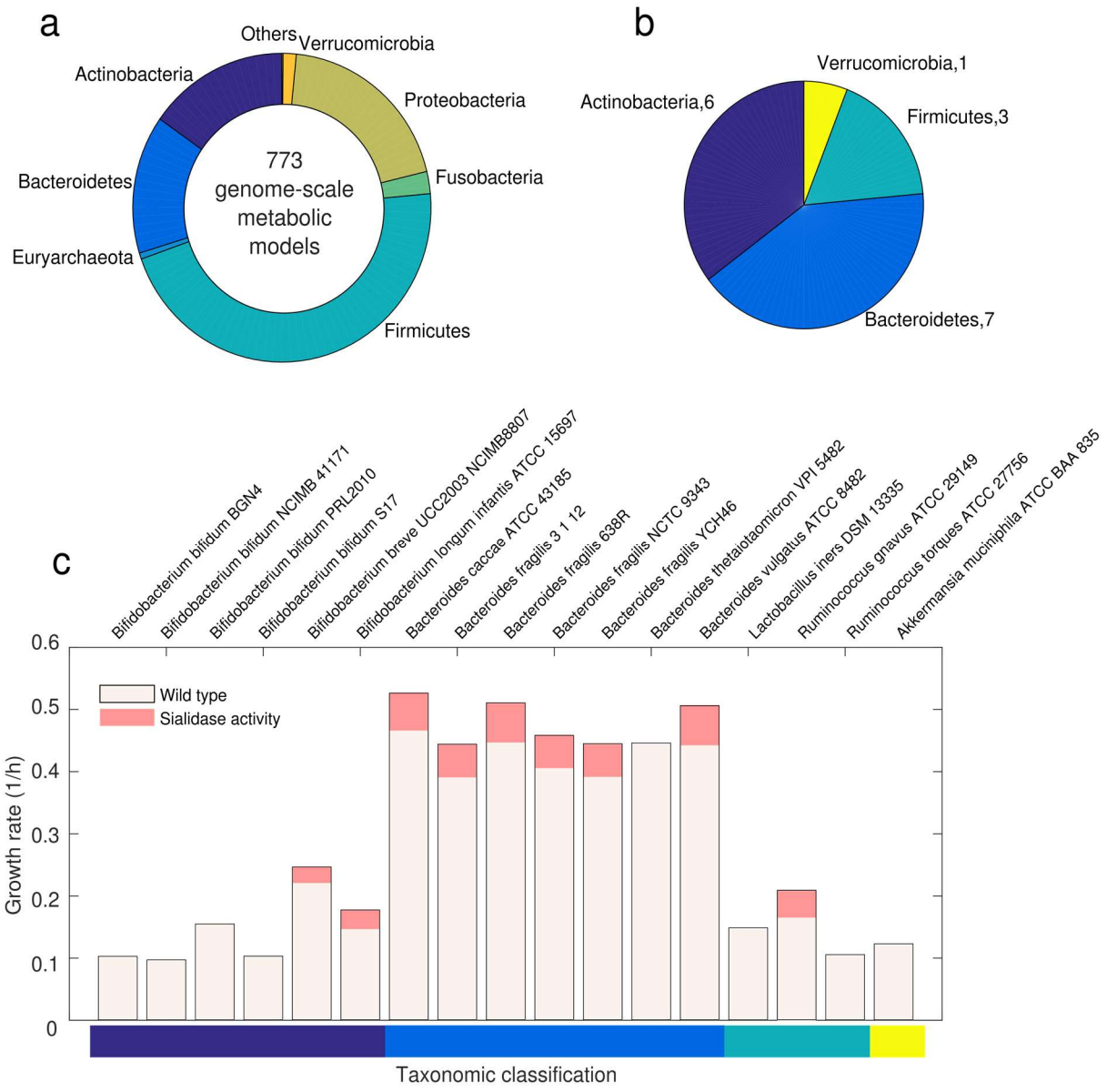
Supplementary Figure 2: The microbiome of *Cmah*$^{-/-}$ mice and WT mice fed soy, PSM, and EBN diets was determined through 16S rRNA gene amplicon sequencing. Beta-diversity analysis of WT vs *Cmah*$^{-/-}$ mouse fed with Sias-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R = 0.979, pvalue = 0.001). Significant vectors in grey (R >= 0.7, pvalue =< 0.01) were obtained using *envfit* function from the R package vegan version 2.5-2 and the most representative taxa are indicated on the plot. WT samples enclose n=5 biologically independent animals and n=15 independent experiments per diet. *Cmah*$^{-/-}$ samples enclose n=3 biologically independent animals and n=9 independent experiments.



Supplementary Figure 3: The microbiome of *Cmah*$^{-/-}$ mice and WT mice fed soy, PSM, and EBN diets was determined through 16S rRNA gene amplicon sequencing. a. Alpha-diversity analysis of WT vs *Cmah*$^{-/-}$ mouse fed with Sias-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet. Statistical differences in diversity were calculated using the nonparametric Wilcoxon rank sum test with Holm correction for multiple hypotheses. The significance levels are indicated as follow: *p-value* ≤ 0.05 (*), *p-value* ≤ 0.01 (**), and *p-value* ≤ 0.001 (***). Boxes and whiskers indicate quartiles and middle maker indicate the median (cmah- n=9; wt n=15). **b.** Relative abundance at family level of *Cmah*$^{-/-}$ mouse fed with Sias-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet.

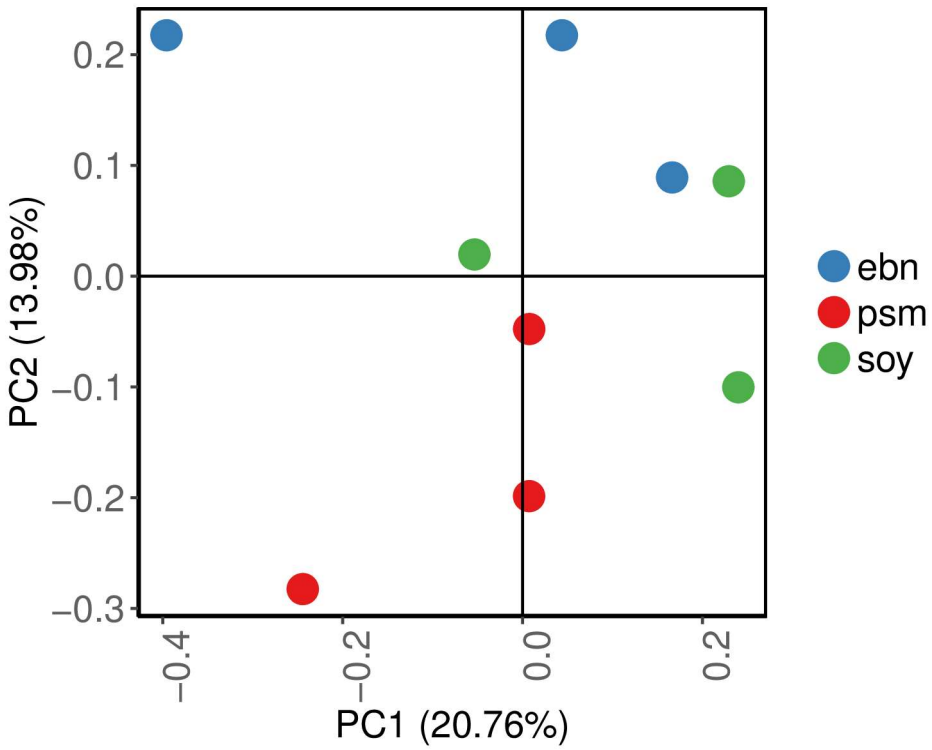


Supplementary Figure 4: OTUs taxonomic relative abundance at genus level. 16S rRNA gene amplicon sequencing was performed for *Cmah*$^{-/-}$ mice fed soy, PSM, and EBN diets. Bacteria genus with relative abundance below than 1% were not considered for differential abundance analysis. Relative abundance is log2 transformed. Columns are coloured label according with diet (red – PSM, blue – EBN, and green – SOY). Hierarchical clustering using Euclidian distance were applied to both row and column as indicated for the phylogenetic trees.

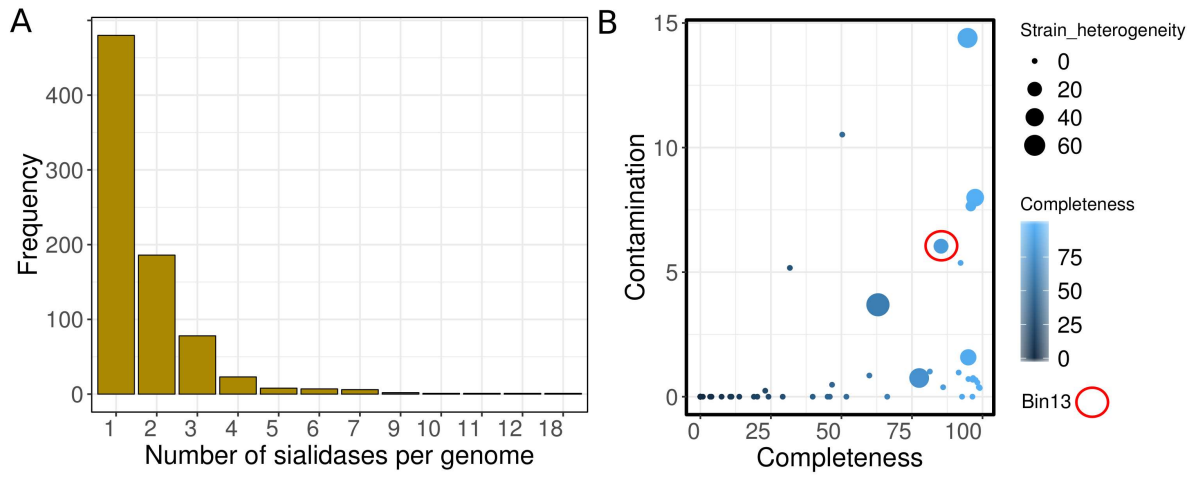


Supplementary Figure 5: Evaluating growth phenotypes associated with the exo-alpha-sialidase activity in the gut microbiome. Genome-scale network reconstructions combine detailed biochemical and physiological information, providing new insights into the metabolism for subsequent manipulation strategies to enhance productivity or to control the metabolism. The scope of these models encompasses the characterization of the metabolic behaviour of target microorganisms. We evaluated growth phenotypes of microorganisms of the gut microbiota associated with sialidase metabolism (EC: 3.2.1.18). **a,** Gut-microbiome microorganisms with genome-scale metabolic models available based on 773 bacterial genomes. **b,** We identified 17

microorganism containing putative sialidase activity. Most of the identified organisms fall into the taxonomic group of Bacteroidetes. **c,** Phenotypes regarding growth rate showed that microorganism carrying sialidase activity can growth around 0.1-0.51 1/h, being Bacteroidetes the most efficient growing microorganisms with an average growth rate of 4.8±0.35 1/h. Modeling predictions show that six out of seven Bacteroidetes can reshape their growth rate depending on the sialidase activity. The maximum change in the growth rate is 11±2% of the wild type growth rate. The growth change due to sialidase activity is shown in shaded red (top of barplot) in Supplementary Figure 5c.



Supplementary Figure 6: Principal coordinate analysis (PCoA) applied to CAZymes of *Cmah*$^{-/-}$ mouse fed with Sias-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R = 0.246, pvalue = 0.06). Samples enclose n=3 biologically independent animals per diet.



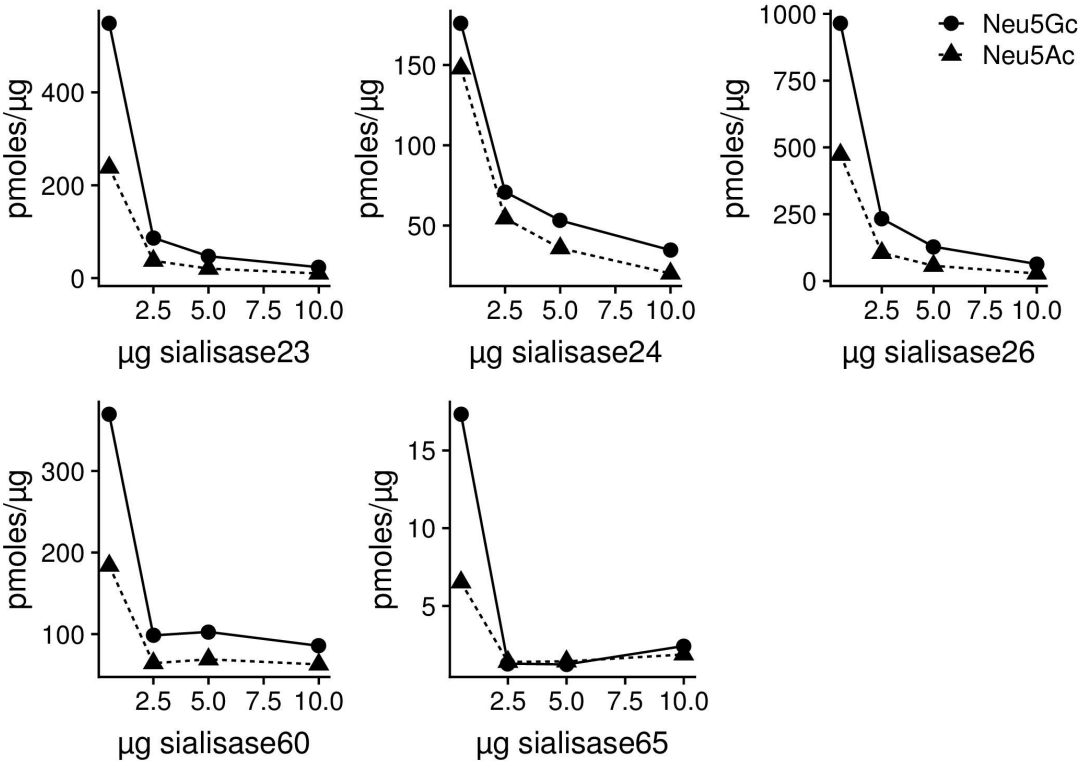
Supplementary Figure 7: Pan-genomes analysis indicated the presence of 1,371 sialidases genes on bacterial genomes available at PATRIC database until the present date (October 12th, 2018). a. Number of sialidases per bacterial genome available at PATRIC database. While most bacterial genomes contain only one sialidase, we identified hundreds of examples of bacterial genomes with multiple sialidases in PATRIC[7]. **b. Evaluation of contamination, strain heterogeneity, and completeness's of all 51 bins retrieve from the shotgun metagenome.** The presence of multiple sialidases per genome is likely to be a result of bacterial strains or close related species co-assembling. Although, we have dramatically improved our ability to correctly assemble genomes from complex microbial communities, the distinction between genomes from different strains still challenge. Evaluating the 51 bins assembled from our shotgun metagenome, the bin13, containing the 5 sialidase genes characterized here, has shown one of the highest strain heterogeneity (36.36%) amongst all the bins.

```
CLUSTAL O(1.2.4) multiple sequence alignment


Sialidase26      MKKNL-FLSIIFSFCVILQAFASDTVFVRETQIPVLIERQDNVLFMLRLNAKESHTLDEV   59
CUA18247.1       MKKAVILFSLFCFLCAIPVVQAADTIFVRETRIPILIERQDNVLFYLRLDAKESQTLNDV   60
                 *** : ::*::  :*.*  . *:**:*****:**:********** ***:****:**::*

Sialidase26      VLNFGKDVNMSDIQSVKLYYSGTEARQNYGKNFFAPVSYISSHTPGKTLAANPSYSINKS   119
CUA18247.1       VLNLGEGVNLSEIQSIKLYYGGTEALQDSGKKRFAPVGYISSNTPGKTLAANPSYSIKKS   120
                 ***:*:.**:*:***:****.**** *: **: ****.****:**************:**

Sialidase26      QVNNPKRKVALKANQKLFPGINYFWISLQMKPDASLLDKVAAKIAAIKVDNKEALMHTVS   179
CUA18247.1       EVTNPGNQVVLKGDQKLFPGINYFWISLQMKPGTSLTSKVTADIASITLDGKKALLDVVS   180
                 :*.** .:*.**.:******************.:** .**:*.**:*.:*.*:**:..**

Sialidase26      PENIVHRVGVGVRHAGDDGSASFRIPGLVTTNKGTLLGVYDVRYNNSADLQEHVDIGLSR   239
CUA18247.1       ENGIEHRMGVGVRHAGDDNSAAFRIPGLVTTNKGTLLGVYDVRYNSSVDLQEHVDVGLSR   240
                  :.* **:**********.**:**********************.*.*******:****

Sialidase26      SVDGGKTWEKMRLPLAFGETGDLPAAQNGVGDPSILVDTKTNTVWVVAAWTHGMGNQRAW   299
CUA18247.1       STDGGKTWEKMRLPLAFGEFGGLPAGQNGVGDPSILVDTKTNNVWVVAAWTHGMGNQRAW   300
                 *.**************** *.***.***************.*****************

Sialidase26      WSSYPGMDMNHTAQLVLSKSTDDGKTWSKPINITEQVKDPSWYFLLQGPGRGITMQDGTL   359
CUA18247.1       WSSHPGMDMNHTAQLVLAKSTDDGKTWSAPINITEQVKDPSWYFLLQGPGRGITMSDGTL   360
                 ***:*************:********** *************************.****

Sialidase26      VFPIQFIDSTRVPNAGIMYSKDRGETWKIHNYARTNTTEAQVAEVEPGVLMLNMRDNRGG   419
CUA18247.1       VFPTQFIDSTRVPNAGIMYSKDGGKNWKMHNYARTNTTEAQVAEVEPGVLMLNMRDNRGG   420
                 *** ***************** *:.**:*******************************

Sialidase26      SRAISTTKDLGKTWTEHSSSRKALQEPVCMASLISVKAKDNVLNKDILLFSNPNTVKGRH   479
CUA18247.1       SRAVAITKDLGKTWTEHESSRKALPESVCMASLISVKAKDNVLGKDLLIFSNPNTTKGRY   480
                 ***:: ***********.****** * ****************.**:*:******.***:

Sialidase26      HITIKASLDGGVTWLPEHQVMLDEGEGWGYSCLTMIDKETIGILYESSVAHMTFQAVQLR   539
CUA18247.1       NTTIKISLDGGVTWSPEHQLLLDEGNNWGYSCLSMIDKETIGILYESSVAHMTFQAVKLK   540
                 : *** ******** ****::****:.******:**********************:*:

Sialidase26      DIIK   543
CUA18247.1       DIIK   544
                 ****
```

Supplementary Figure 8: Amino acid sequence alignment between the sialidase26 and the sialidase CUA18247.1.

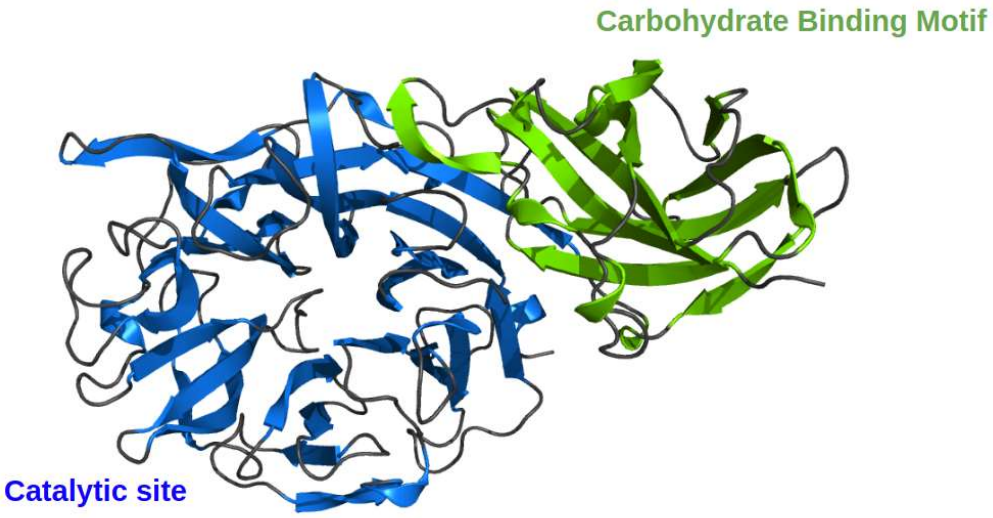
```
sialidase26metagenome      MKKNLFLSIIFSFCVILQAFASDTVFVRETQIPVLIERQDNVLFMLRLNAKESHTLDEVV 60
sialidase26sanger          --------------------MSDTVFVRETQIPVLIERQDNVLFMLRLNAKESHTLDEVV 40
                                                ************************************

sialidase26metagenome      LNFGKDVNMSDIQSVKLYYSGTEARQNYGKNFFAPVSYISSHTPGKTLAANPSYSINKSQ 120
sialidase26sanger          LNFGKDVNMSDIQSVKLYYSGTEARQNYGKNFFAPVSYISSHTPGKTLAANPSYSINKSQ 100
                           ************************************************************

sialidase26metagenome      VNNPKRKVALKANQKLFPGINYFWISLQMKPDASLLDKVAAKIAAIKVDNKEALMHTVSP 180
sialidase26sanger          VNNPKRKVALKANQKLFPGINYFWISLQMKPDASLLDKVAAKIAAIKVDNKEALMHTVSP 160
                           ************************************************************

sialidase26metagenome      ENIVHRVGVGVRHAGDDGSASFRIPGLVTTNKGTLLGVYDVRYNNSADLQEHVDIGLSRS 240
sialidase26sanger          ENIVHRVGVGVRHAGDDGSASFRIPGLVTTNKGTLLGVYDVRYNNSADLQEHVDIGLSRS 220
                           ************************************************************

sialidase26metagenome      VDGGKTWEKMRLPLAFGETGDLPAAQNGVGDPSILVDTKTNTVWVVAAWTHGMGNQRAWW 300
sialidase26sanger          VDGGKTWEKMRLPLAFGETGDLPAAQNGVGDPSILVDTKTNTVWVVAAWTHGMGNQRAWW 280
                           ************************************************************

sialidase26metagenome      SSYPGMDMNHTAQLVLSKSTDDGKTWSKPINITEQVKDPSWYFLLQGPGRGITMQDGTLV 360
sialidase26sanger          SSYPGMDMNHTAQLVLSKSTDDGKTWSKPINITEQVKDPSWYFLLQGPGRGITMQDGTLV 340
                           ************************************************************

sialidase26metagenome      FPIQFIDSTRVPNAGIMYSKDRGETWKIHNYARTNTTEAQVAEVEPGVLMLNMRDNRGGS 420
sialidase26sanger          FPIQFIDSTRVPNAGIMYSKDRGETWKIHNYARTNTTEAQVAEVEPGVLMLNMRDNRGGS 400
                           ************************************************************

sialidase26metagenome      RAISTTKDLGKTWTEHSSSRKALQEPVCMASLISVKAKDNVLNKDILLFSNPNTVKGRHH 480
sialidase26sanger          RAISTTKDLGKTWTEHSSSRKALQEPVCMASLISVKAKDNVLNKDILLFSNPNTVKGRHH 460
                           ************************************************************

sialidase26metagenome      ITIKASLDGGVTWLPEHQVMLDEGEGWGYSCLTMIDKETIGILYESSVAHMTFQAVQLRD 540
sialidase26sanger          ITIKASLDGGVTWLPEHQVMLDEGEGWGYSCLTMIDKETIGILYESSVAHMTFQAVQLRD 520
                           ************************************************************

sialidase26metagenome      IIK----------        543
sialidase26sanger          IIKHHHHHHHHHH        533
                           ***
```

Supplementary Figure 9: Sialidase26 sequence alignment. Alignment between the amino acid sequence obtained from the metagenomics data (*de novo* genome assembling) and the amino acid sequence obtained from the PCR amplified gene (Sanger sequencing). For PCR, DNA extracted from mouse fecal samples were used as template.

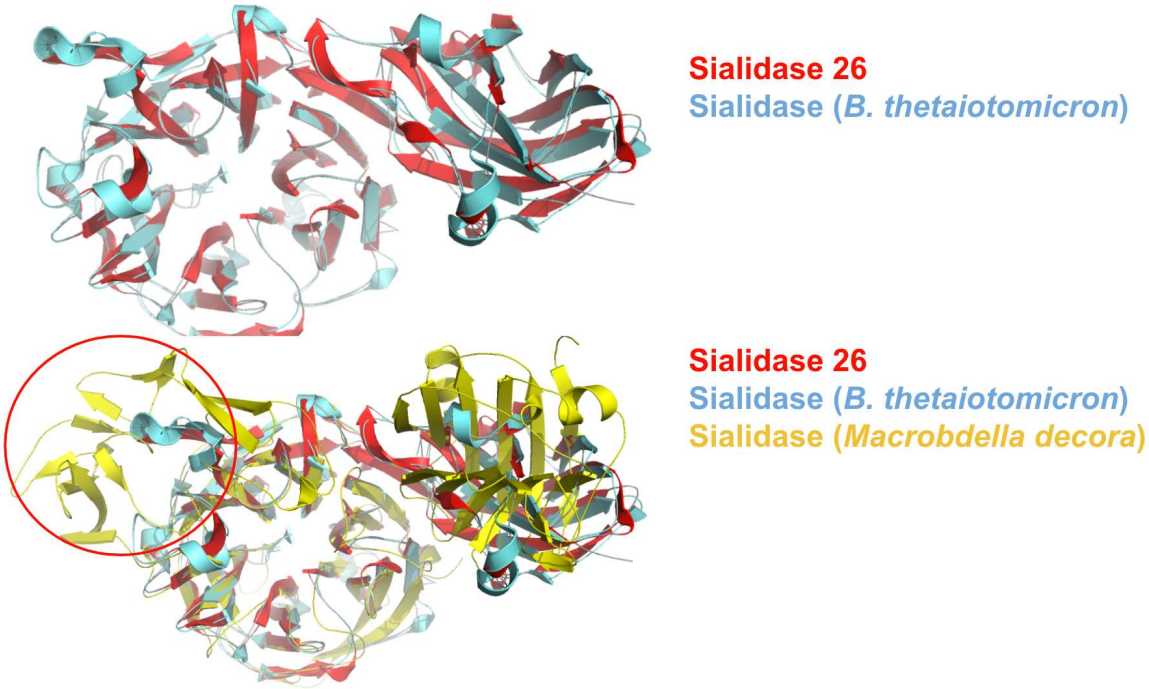


Supplementary Figure 10: Sialidases substrate preference. To the best of our knowledge, no previously characterized exo-sialidases have been shown to prefer Neu5Gc over Neu5Ac. A single report has shown that a sialidase from the oyster species *Crassostrea virginica* cleaves the very rare glycan structure Neu5Gcα2,5-O(glycolyl)Neu5Gc more efficiently than Neu5Acα2-8Neu5Ac. Because the authors did not compare enzyme substrate preference using a glycan structure with same terminal linkage but differing in Neu5Gc vs Neu5Ac, it cannot be concluded that the oyster sialidases actually prefers Neu5Gc-containing substrates[3]. Similarly, the sialyltransferase from the pathogenic bacteria *Pasteurella multocida*[4] has been shown to cleaves Neu5Gc more efficiently than Neu5Ac, however, no assay was performed in order to evaluate preferential cleavage. In fact, previous studies revealed that many bacterial species such as *Arthrobacter ureafaciens*, *Clostridium perfringens*, *Streptococcus sp*., *Salmonella typhimurium, Streptococcus pneumonia, Ruminococcus gnavus* and *Vibrio cholera* express sialidases that hydrolyse Neu5Ac-terminated glycans more efficiently than Neu5Gc-terminated glycans[4–6], a substrate preference that is also shared by the murine sialidases Neu1, Neu2 and Neu4[4]. Sialidase activity using different enzyme concentrations (0.5 µg, 2.5 µg, 5 µg, and 10 µg). For protein expression and purification. The sialidase genes were PCR amplified using DNA extracted from mouse faecal samples as the template. Independent experiments were performed in duplicate with similar results.

Supplementary Figure 11: Sialidase26 shares common sequence motifs with other sialidases. BTSA (sialidase from *Bacteroides thetaiotomicron*) is shown as an example, aligned to Sialidase26. Common sialidase traits include a RIP/RLP motif (AAs 203-205) and four Asp-containing consensus sequences (AAs 240-247, 319-326, 379-387, and 486-493; S/T-x-D-x-G-x-T-W/F), which electrostatically interact with the carboxylate groups of sialic acids. It also contains a conserved nucleophilic tyrosine and acid/based glutamic acid residues at AAs 509 and 398, respectively, and a N-terminal signal peptide, which flags them for secretion across the membrane. Yellow shades indicate the signal sequence, RIP/RLP motif, and Asp boxes as described on the left.

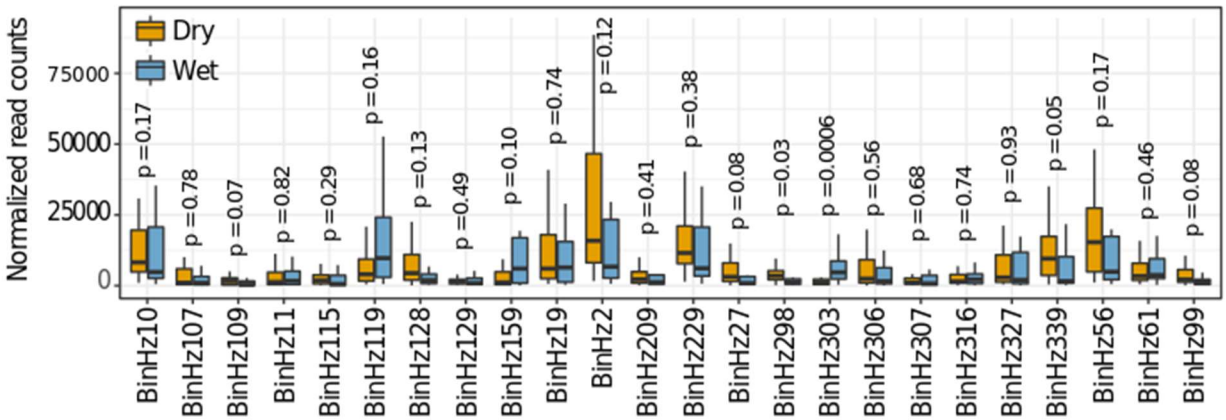


Supplementary Figure 12: Overall structure of GH33 sialidases. Catalytic site is represented in blue and the carbohydrate binding motif is represented in green.

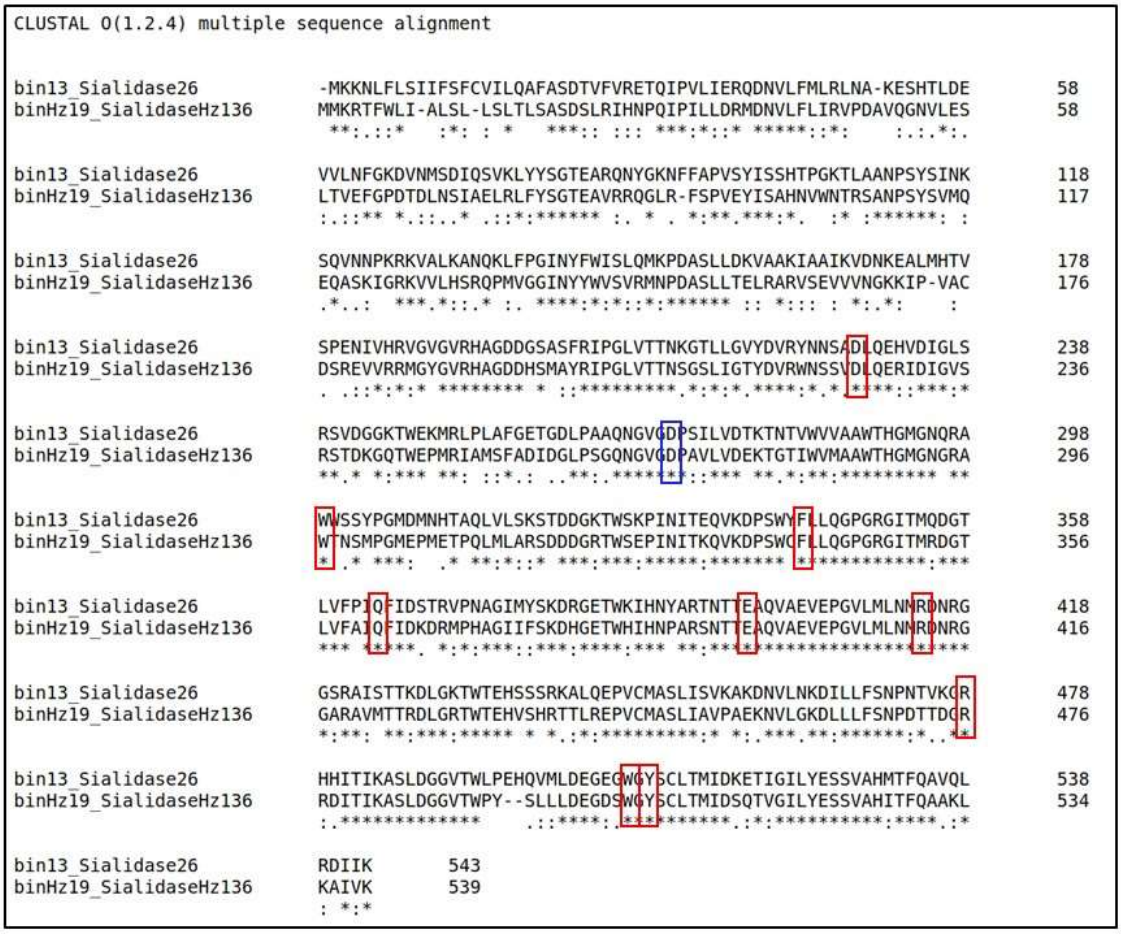


Supplementary Figure 13: Comparison between sialidase26 and existing sialidases structures

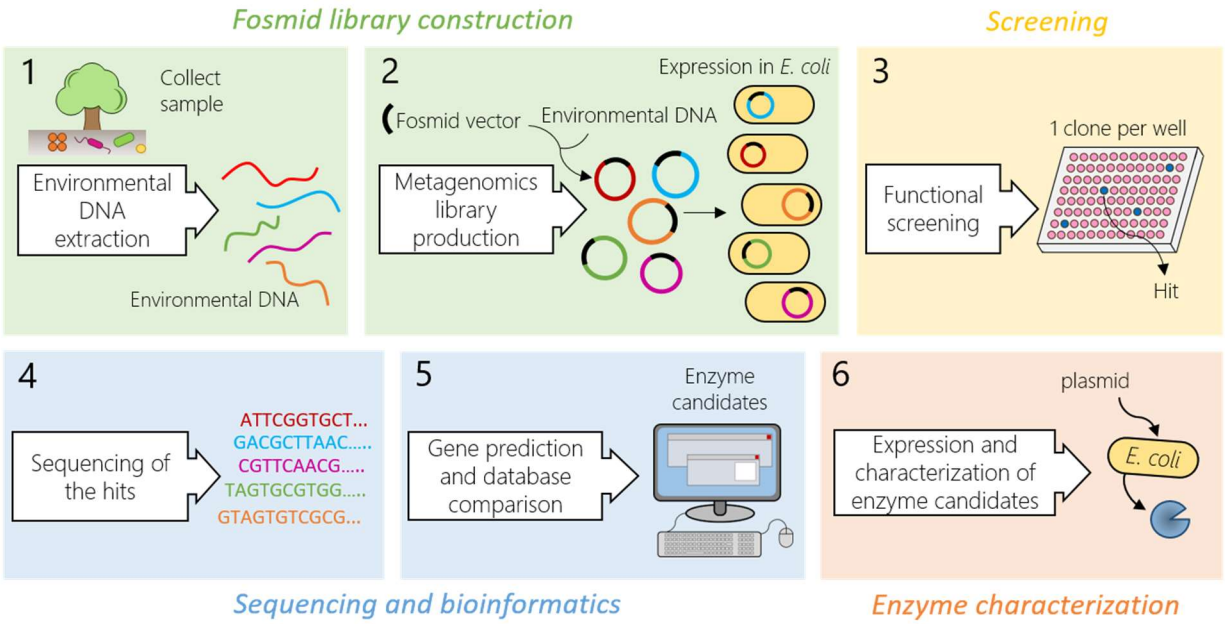
The sialidase26 identified in this study was compared with published sialidase structures. Despite this novel substrate preference, sequence residues that are predicted to interact with terminal Sias in the catalytic site are highly also fully conserved with structurally studied sialidases exhibiting no known Neu5Gc preference over Neu5Ac, including those from *B. caccae* (PDB 4Q6K), *B. thetaiotamicron*[7], and *P. distasonis* (PDB 4FJ6). Other known sialidases with very low sequence identity (<30%) nevertheless shared 50-75% identity with residues predicted to interact with sialic acid, including NanA and NanB from *S. pneumoniae*[8–10] NanI from *C. perfringens*[11], an IT-sialidase from *R. gnavus*[6], and a trans-sialidase from *M. decora*, but suggest no clear sequence motif or mutation causing Neu5Gc preference. Sialidase26 exhibits high similarity to other sialidases, including a sialidase of *B. thetaiotomicron* (BTSA). Both sialidases contain a wide binding site for broad substrate accommodation; BTSA exhibits hydrolysis activity for α2,6-, 2,3-, and 2,8-linked sialic acid moieties. For both proteins, the CBM is found N-terminally. Missing from both sialidases is an insertion motif (red circle), found in the *Macrobedella decora* sialidase.



Supplementary Figure 14: Hadza shotgun metagenome. Normalized read counts (counts per million) of the bins with annotated sialidases. *p-values* were computed using the two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses. n=20 biological independent sample per condition (dry or wet). Boxes and whiskers indicate quartiles and middle maker indicate the median (n=20).



Supplementary Figure 15: Amino acid sequence alignment between the sialidase26 and the sialidaseHz136. The crystal structure suggests that the D271 residue (blue rectangle) plays a key role in substrate preference. However, all amino acids residues in contact with the Sias (red rectangles) might play a contributing role due to contact with the substrate.



Supplementary Figure 16: Fosmid library construction, screening, enzyme selection, and characterization.

```
CLUSTAL O(1.2.4) multiple sequence alignment


sialidase_C19      MKRTGWLLLLVSGPALCQ--PPRGYSIPLIDLAAEAWRQTVVDREPGQYLGHPTTVLLED 58
sialidase_C22      -MVIAWM-----AGVQCTAAALVDLGVTILDISGEPERQVIVDREPGQYLGHVSTVLLED 54
                       .*:     . . *      . .: ::*::.*  **.:********** :******

sialidase_C19      GRTVLAVYPKGHGRGAIVYKRSRDGGRTWSARLPVPENWATSQETPTIHRVVDPRGRKRL 118
sialidase_C22      DKTILAVYPKGHGRGAIVLKRSEDGGRTWSGRLPVPASWATSKETPTIHRVVDARGRRRL 114
                   .:*:************** ***.*******.***** .****:********** ***:**

sialidase_C19      ILFSGLYPIRMSVSEDDGETWTPLAPIGNFGGVVAMASVERL-RDGRYMALFHDDGRFLR 177
sialidase_C22      LLWSGLYPARRALSDDDGRTWTELEPVGEWGGIVVMGFVEATRQPGHYVAMFHDDGRYFA 174
                   :*:***** * ::*:***.*** * *:*::**:*.*. **   : *:*:*:******::

sialidase_C19      GGGKPD-RFVVYKTLSGDGGLTWSEPVPILSHPQAHLCEPGLLRSPDGRRIAILLRENSR 236
sialidase_C22      AQPSTNRSMTLYLTRSSDGGVSWSSPVAVWSNSAVHLCEPGGIWSPDRRRLAVLLRENRR 234
                   .  . :  :.:* * *.***::**.** : *:  .****** : *** **:*:***** *

sialidase_C19      KFNSFVSFSDDEGETWSEPRELPGALTGDRHTAVYARDGRLFISFRDTT----------- 285
sialidase_C22      VNNSHIMFSDDEGHTWTTPVEMPLSRAGDRHTLRYTPDGRIVCVFRAVTPVGMRGRSFGQ 294
                     **.: ******.**: * *:* : :*****  *: ***:.  ** .*

sialidase_C19      --------HESPTRGDWVAWVGRFGDIENGREGQYRVRLMKNHKDLDCCYPGVLRLPDDT 337
sialidase_C22      NEDVDTLGVGSPFEGDCVAWVGTWDDLVHNRPGQYYVRLLNNRKGWDTTYPGVEVLPDGT 354
                             ** .** ***** :.*: :.* *** ***::*:*. *  ****  ***.*

sialidase_C19      ILTTTYGHWTPGEPPYIVSIRLRLSELDRRAQGAGR--       373
sialidase_C22      VVVTTYGHWNAGEPPYIRSVRFRLEELDRMAAKASQKP       392
                   ::.******. ****** *:*:**.**** *  *.:
```

Supplementary Figure 17: Amino acid sequence alignment between the sialidaseC19 and the sialidaseC22.

Supplementary Table 1: Composition of amino acids. neutral monosaccharides, and sialic acid between EBN and PSM.

% of aa per microgram		
	EBN	**PSM**
Ala	5.35	11.57
Gly	7.82	5.51
Val	13.15	9.63
Leu	11.44	10.3
Ile	4.42	4.72
Pro	11.64	6.96
Ser	13.99	14.35
Thr	5.85	8.46
Phe	6.5	3.82
Asp	9.99	8.82
Glu	5.28	12.74
Lys	0.95	1.83
Tyr	3.63	1.29
monossacharide per microgram of diet (in ng)		
	EBN	**PSM**
Fuc	3,166	16,958
GalNAc	44,498	61,266
GlcNAc	57,294	4,102
Gal	57,838	38.14
Glc	306.4	315
Man	5,956	0
sialic acid per gram of diet (in mg)		
	EBN	**PSM**
Neu5Ac	0.25	0
Neu5Gc	0	0.25

Supplementary Table 2: Metabolic model simulation results

Microorganism	Taxonomy		Growth rate	Adjusted growth rate	Sialidase activity
Bifidobacterium_bifidum_BGN4	Bacteria	Actinobacteria	0.1021	0.1021	0
Bifidobacterium_bifidum_NCIMB_41171	Bacteria	Actinobacteria	0.0964	0.0964	0
Bifidobacterium_bifidum_PRL2010	Bacteria	Actinobacteria	0.1541	0.1541	0
Bifidobacterium_bifidum_S17	Bacteria	Actinobacteria	0.1024	0.1024	0
Bifidobacterium_breve_UCC2003_NCIMB8807	Bacteria	Actinobacteria	0.247	0.2208	0.0262
Bifidobacterium_longum_infantis_ATCC_15697	Bacteria	Actinobacteria	0.178	0.1468	0.0312
Bacteroides_caccae_ATCC_43185	Bacteria	Bacteroidetes	0.5272	0.4662	0.061
Bacteroides_fragilis_3_1_12	Bacteria	Bacteroidetes	0.4443	0.3908	0.0535
Bacteroides_fragilis_638R	Bacteria	Bacteroidetes	0.5099	0.447	0.063
Bacteroides_fragilis_NCTC_9343	Bacteria	Bacteroidetes	0.4591	0.4056	0.0534
Bacteroides_fragilis_YCH46	Bacteria	Bacteroidetes	0.4441	0.3916	0.0526
Bacteroides_thetaiotaomicron_VPI_5482	Bacteria	Bacteroidetes	0.4454	0.4454	0
Bacteroides_vulgatus_ATCC_8482	Bacteria	Bacteroidetes	0.5071	0.4424	0.0647
Lactobacillus_iners_DSM_13335	Bacteria	Firmicutes	0.1479	0.1479	0
Ruminococcus_gnavus_ATCC_29149	Bacteria	Firmicutes	0.2075	0.1643	0.0432
Ruminococcus_torques_ATCC_27756	Bacteria	Firmicutes	0.1049	0.1049	0
Akkermansia_muciniphila_ATCC_BAA_835	Bacteria	Verrucomicrobia	0.1222	0.1222	0

Supplementary Table 3: Genome bins mouse metagenome

Bin_ID	Size	GC%	N50	L50	Completeness	Contamination	Strain heterogeneity	Closest neighbor
bin.21	1850494	33.2	24064	23	99.03	0.35	0	*Helicobacter winghamensis* ATCC BAA-430
bin.27	2269403	54.9	80475	9	98.87	0.38	0	*Porphyromonas gingivalis* ATCC 33277
bin.47	2568469	50.8	93332	10	98.68	0.38	0	*Bacteroides vulgatus* ATCC 8482
bin.25	3208820	47.9	41655	25	98.18	0.57	0	*Porphyromonas gingivalis* ATCC 33277
bin.50	2102753	38.6	16945	37	97.54	0.67	0	*Anaerotruncus colihominis* DSM 17241
bin.13	6594969	42.9	14791	130	97.4	7.99	36.36	*Bacteroides thetaiotaomicron* VPI-5482
bin.48	2449761	51.6	57972	15	96.67	0.75	0	*Porphyromonas gingivalis* ATCC 33277
bin.19	4078864	45.1	47956	27	96.5	0.68	0	*Coprococcus comes* ATCC 27758
bin.40	2326744	52.8	77390	9	96.42	0	0	*Bacteroides fragilis* ATCC 25285
bin.23	1180283	40.3	9017	40	95.85	7.65	5.88	*Chlamydia muridarum* Nigg
bin.41	2668268	48.8	66931	12	95	0.71	0	*Alistipes putredinis* DSM 17216
bin.26	2635792	52.3	18023	40	94.95	1.58	28.57	*Bacteroides thetaiotaomicron* VPI-5482
bin.33	4542794	59.3	17331	81	94.7	14.4	52.17	*Clostridium cellulolyticum* H10
bin.44	1546950	34	20898	25	92.74	0	0	*Clostridium cellulolyticum* H10
bin.18	3999932	59.1	12835	90	92.23	5.37	0	*Eubacterium siraeum* V10Sc8a
bin.5	2531697	46	35719	25	91.55	0.97	0	*Dorea formicigenerans* ATCC 27755
bin.36	2123123	55.1	47701	13	86.04	0.38	0	*Bacteroides fragilis* ATCC 25285
bin.16	2765019	60.2	9836	83	85.3	6.04	21.43	*Clostridium cellulolyticum* H10
bin.1	1743578	53.6	16637	33	81.33	1.01	0	*Bacteroides vulgatus* ATCC 8482

bin.24	1958324	53.8	79876	9	77.55	0.75	50	*Porphyromonas gingivalis* ATCC 33277
bin.3	1438390	50.8	84511	7	66.23	0	0	*Bacteroides fragilis* ATCC 25285
bin.38	2554082	58.8	9223	81	62.94	3.69	77.78	*Anaerotruncus colihominis* DSM 17241
bin.31	839425	46.7	141422	3	59.89	0.85	0	*Geobacillus kaustophilus* HTA426
bin.4	2121811	47.5	17632	34	51.75	0	0	*Bryantella formatexigens* DSM 14469
bin.45	5361881	42.9	4933	270	50.29	10.52	0	*Coprococcus comes* ATCC 27758
bin.49	737196	33.7	17545	14	46.67	0.48	0	*Acholeplasma laidlawii* PG-8A
bin.29	821437	34	6183	41	45.87	0	0	*Thermincola* sp. JR
bin.15	875044	47.6	94357	3	45.09	0	0	*Anaerotruncus colihominis* DSM 17241
bin.14	822161	42.9	58321	5	39.81	0	0	*Clostridium perfringens* ATCC 13124
bin.28	1108277	51.4	3959	89	31.72	5.17	0	*Bacteroides vulgatus* ATCC 8482
bin.12	1293986	47.1	48036	9	29.18	0	0	*Bacteroides eggerthii* DSM 20697
bin.30	627666	44.3	64511	3	24.14	0	0	*Porphyromonas gingivalis* ATCC 33277
bin.37	444059	39	2631	53	22.99	0.24	0	*Erysipelotrichaceae bacterium* 5_2_54FAA
bin.11	443662	46.2	3668	38	20.18	0	0	*Erysipelotrichaceae bacterium* 5_2_54FAA
bin.20	1150156	47.7	12470	20	18.97	0	0	*Bacteroides vulgatus* ATCC 8482
bin.43	216870	53.5	39370	3	13.79	0	0	*Bacteroides vulgatus* ATCC 8482
bin.10	547001	55.5	5527	33	11.03	0	0	*Ethanoligenens harbinense* YUAN-3
bin.8	270586	57.5	12196	8	10.58	0	0	*Alistipes putredinis* DSM 17216
bin.42	243180	61.1	3209	27	7.52	0	0	*Ethanoligenens harbinense* YUAN-3

bin.39	513508	48.1	27413	7	3.93	0	0	*Symbiobacterium thermophilum* IAM 14863
bin.9	514825	44.5	15600	9	3.45	0	0	*Bacteroides intestinalis* DSM 17393
bin.34	502466	45.9	32140	6	0.86	0	0	*Bacteroides thetaiotaomicron* VPI-5482
bin.7	584478	47.5	4526	35	0	0	0	*Dorea formicigenerans* ATCC 27755
bin.6	298825	47.2	6605	12	0	0	0	*Clostridium thermocellum* ATCC 27405
bin.51	340372	48.7	5053	18	0	0	0	*Clostridium saccharolyticum* WM1
bin.46	299296	50.2	20062	5	0	0	0	*Desulfitobacterium* sp. Y51
bin.35	466484	49.3	7616	17	0	0	0	*Eubacterium rectale* ATCC 33656
bin.32	322877	43.7	14422	5	0	0	0	*Bacteroides intestinalis* DSM 17393
bin.22	301236	49.3	26150	4	0	0	0	*Erysipelotrichaceae bacterium* 5_2_54FAA
bin.2	390337	49	7096	17	0	0	0	*Bacteroides pectinophilus* ATCC 43243
bin.17	251249	43.4	13443	5	0	0	0	*Bacteroides intestinalis* DSM 17393

Supplementary Table 4: X-ray crystallography. Data collection and refinement statistics (molecular replacement)

	Sia26 (6MRX)	Sia26:DANA (6MRV)	Sia26:DANA-Gc (6MYV)	SiaHz136 (6MNJ)
Data collection				
Space group	*I*121	*I*121	*P*1	*I*222
Cell dimensions				
a, *b*, *c* (Å)	79.81, 115.48, 260.81	92.04, 116.70, 113.64	85.64, 89.75, 95.81	119.67, 145.39, 150.29
α, β, γ (°)	90.00, 100.92, 90.00	90.00, 106.34, 90.00	64.70, 75.65, 88.58	90.00, 90.00, 90.00
Resolution (Å)	48.42(2.00) *	80.82(1.80) *	48.50(2.20) *	75.14(2.20) *
R_{sym} or R_{merge}	11.0(92.6)	10.3(81.9)	13.5(87.2)	21.1(87.3)
I / σI	6.9(1.2)	6.6(1.1)	6.0(1.3)	10.5(3.7)
Completeness (%)	100.0(100.0)	99.7(99.9)	98.3(97.2)	100.0(99.8)
Redundancy	3.5(3.2)	3.1(3.1)	3.9(3.9)	12.7(11.3)
Refinement				
Resolution (Å)	48.42-2.00	60.78-1.80	48.50-2.20	75.14-2.20
No. reflections	157,315	106,029	124,154	66,624
R_{work} / R_{free}	0.22/0.24	0.20/0.23	0.23/0.26	0.21/0.24
No. atoms				
Protein	16,280	8.309	8,309	8,225
Ligand/ion	-	72	148	-
Water	282	421	789	142
B-factors				
Protein	28.23	21.64	34.33	27.60
Ligand/ion	-	18.88	38.32	-
Water	24.71	20.94	33.75	22.38
R.m.s. deviations				
Bond lengths (Å)	0.002	0.004	0.006	0.002
Bond angles (°)	0.591	0.780	0.823	0.593

*One crystal per dataset was used.

Supplementary Table 5: Relative abundance of Hadza data aligned to bin13.

Accession	SRA	Season	Experiment	Total number trimmed nonhuman reads	Reads mapped bin13	Normalized
SRX2963153	SRR5763445	dry	TZ_47979_R1	6273252	7900	1259.31
SRX2963152	SRR5763446	wet	TZ_93321_R1	11025059	3919	355.46
SRX2963151	SRR5763447	wet	TZ_67706_R1	6178345	3427	554.68
SRX2963150	SRR5763448	dry	TZ_21460_R1	3330172	977	293.38
SRX2963149	SRR5763449	wet	TZ_41014_R1	9107663	3216	353.11
SRX2963148	SRR5763450	wet	TZ_10742_R1	7466471	2867	383.98
SRX2963147	SRR5763451	wet	TZ_73414_R1	12370910	7183	580.64
SRX2963146	SRR5763452	wet	TZ_83893_R1	7635604	2318	303.58
SRX2963145	SRR5763453	wet	TZ_65172_R1	13353637	3822	286.21
SRX2963144	SRR5763454	wet	TZ_65642_R1	18721999	18973	1013.41
SRX2963143	SRR5763455	dry	TZ_90808_R1	10356039	6961	672.17
SRX2963142	SRR5763456	dry	TZ_16581_R1	2529052	701	277.18
SRX2963141	SRR5763457	wet	TZ_24827_R1	4383172	791	180.46
SRX2963140	SRR5763458	wet	TZ_70090_R1	2510510	761	303.13
SRX2963139	SRR5763459	dry	TZ_86768_R1	4155868	1938	466.33
SRX2963138	SRR5763460	dry	TZ_37609_R1	3947819	1929	488.62
SRX2963137	SRR5763461	dry	TZ_39227_R1	3207470	1398	435.86
SRX2963136	SRR5763462	dry	TZ_81781_R1	3040744	1484	488.04
SRX2963135	SRR5763463	dry	TZ_42864_R1	2708678	2452	905.24
SRX2963134	SRR5763464	dry	TZ_29321_R1	2563465	2915	1137.13
SRX2963133	SRR5763465	dry	TZ_87532_R1	3758101	555	147.68
SRX2963132	SRR5763466	dry	TZ_76219_R1	3081596	2994	971.57
SRX2963131	SRR5763467	dry	TZ_83985_R1	2619653	1731	660.77
SRX2963130	SRR5763468	dry	TZ_25806_R1	2097908	1833	873.73
SRX2963129	SRR5763469	dry	TZ_83788_R1	4712957	4714	1000.22
SRX2963128	SRR5763470	dry	TZ_38244_R1	2205033	1480	671.19
SRX2963127	SRR5763471	dry	TZ_43638_R1	3806024	3041	799.00
SRX2963126	SRR5763472	dry	TZ_99300_R1	3325735	607	182.52
SRX2963125	SRR5763473	dry	TZ_42041_R1	4596873	1210	263.22
SRX2963124	SRR5763474	wet	TZ_86771_R1	3953617	886	224.10
SRX2963123	SRR5763475	wet	TZ_60579_R1	4494910	1534	341.27
SRX2963122	SRR5763476	wet	TZ_83627_R1	4292298	2283	531.88
SRX2963121	SRR5763477	wet	TZ_91827_R1	2797766	1294	462.51
SRX2963120	SRR5763478	wet	TZ_40311_R1	2204802	263	119.29
SRX2963119	SRR5763479	wet	TZ_39371_R1	3293821	1388	421.40
SRX2963118	SRR5763480	wet	TZ_21618_R1	3602167	677	187.94
SRX2963117	SRR5763481	wet	TZ_14139_R1	3024516	633	209.29
SRX2963116	SRR5763482	wet	TZ_12739_R1	3207143	2003	624.54
SRX2963115	SRR5763483	wet	TZ_61717_R1	3764610	741	196.83
SRX2963114	SRR5763484	dry	TZ_27689_R1	5648186	3072	543.89

Table 6: Genome bins Hadza metagenome. Only bins with annotated sialidase gene.

Bin_ID	Size	GC%	N50	L50	Completeness	Contamination	Strain heterogeneity	Closest neighbor
binHz107	2820482	47.19	8077	115	70.86	5.43	76.92	*Bacteroides fragilis* ATCC 25285
binHz109	264436	47.46	7540	13	0	0	0	*Bacteroides thetaiotaomicron* VPI-5482
binHz10	2925305	47.19	9453	98	58.62	9.48	83.33	*Prevotella ruminicola* 23
binHz115	2033063	55.06	16916	36	70.78	0.77	75	*Prevotella ruminicola* 23
binHz119	2936672	47.61	9285	99	90.53	5.6	72.22	*Bacteroides thetaiotaomicron* VPI-5482
binHz11	2663757	49.62	10264	81	88.7	6.64	63.33	*Alistipes putredinis* DSM 17216
binHz128	2793467	46.76	15277	57	86.29	4.7	66.67	*Bacteroides intestinalis* DSM 17393
binHz129	2691906	43.31	9809	88	91.36	3.39	41.67	*Bacteroides caccae* ATCC 43185
binHz159	2314390	27.67	8565	88	92.16	2.11	0	*Clostridium perfringens* ATCC 13124
binHz19	2253882	51.55	11736	57	62.71	3.97	70	*Alistipes* sp.
binHz209	2568830	47.68	13754	59	83.81	2.6	45.45	*Prevotella ruminicola* 23
binHz229	3246489	44.17	5551	166	27.03	5.87	0	*Bacteroides thetaiotaomicron* VPI-5482
binHz27	2394901	52.12	16214	49	70.69	8.62	60	*Prevotella ruminicola* 23
binHz298	2959682	48.18	16224	57	94.1	2.76	71.43	*Prevotella ruminicola* 23
binHz2	3207967	46.21	7641	131	52.35	6.03	100	*Prevotella ruminicola* 23
binHz303	2261674	64.46	21132	33	97.58	3.35	18.18	*Atopobium parvulum* DSM 20469
binHz306	1649945	53.8	12205	44	68.29	1.43	60	*Bacteroides vulgatus* ATCC 8482
binHz307	2438559	50.63	9789	82	82.81	2.58	60	*Prevotella ruminicola* 23
binHz316	1888527	28.91	5338	128	79.6	4	16.67	*Brachyspira murdochii* DSM 12563
binHz327	1038399	51.49	13669	28	29.76	0.95	100	*Bacteroides vulgatus* ATCC 8482
binHz339	3266281	45.62	17634	58	94.94	3.33	56.25	*Prevotella ruminicola* 23
binHz56	3289260	43.4	11369	92	91.94	8.59	45.24	*Prevotella ruminicola* 23
binHz61	4226746	40.58	5623	221	17.54	1.12	14.29	*Roseburia intestinalis* XB6B4
binHz99	2863804	45.31	20114	46	91.36	5.19	80	*Prevotella ruminicola* 23

Supplementary References

1. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).

2. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).

3. Inoue, S. *et al.* A unique sialidase that cleaves the Neu5Gcα2 →5-OglycolylNeu5Gc linkage: Comparison of its specificity with that of three microbial sialidases toward four sialic acid dimers. *Biochem. Biophys. Res. Commun.* **280**, 104–109 (2001).

4. Chokhawala, H. A., Yu, H. & Chen, X. High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries. *ChemBioChem* **8**, 194–201 (2007).

5. Davies, L. R. L. *et al.* Metabolism of vertebrate amino sugars with *N*-glycolyl groups: Resistance of α2-8-linked *N*-glycolylneuraminic acid to enzymatic cleavage. *J. Biol. Chem.* **287**, 28917–28931 (2012).

6. Owen, C. D. *et al.* Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont *Ruminococcus gnavus*. *Nat. Commun.* **8**, 2196 (2017).

7. Wattam, A. R. *et al.* Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* **45**, D535–D542 (2017).

8. Xu, G. *et al.* Crystal structure of the NanB Sialidase from *Streptococcus pneumoniae*. *J. Mol. Biol.* **384**, 436–449 (2008).

9. Xu, G., Li, X., Andrew, P. W. & Taylor, G. L. Structure of the catalytic domain of *Streptococcus pneumoniae* sialidase NanA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64**, 772–775 (2008).

10. Yang, L., Connaris, H., Potter, J. A. & Taylor, G. L. Structural characterization of the carbohydrate-binding module of NanA sialidase, a pneumococcal virulence factor. *BMC Struct. Biol.* **15**, 15 (2015).

11. Newstead, S. L. *et al.* The structure of *Clostridium perfringens* NanI sialidase and its catalytic intermediates. *J. Biol. Chem.* **283**, 9080–9088 (2008).