AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Human Microbiome | Research Article

SARS-CoV-2 infectivity can be modulated through bacterial grooming of the glycocalyx

Cameron Martino,[1,2,3] Benjamin P. Kellman,[1,2] Daniel R. Sandoval,[4] Thomas Mandel Clausen,[4,5] Robert Cooper,[6] Alhosna Benjdia,[7] Feryel Soualmia,[7,8] Alex E. Clark,[9] Aaron F. Garretson,[9] Clarisse A. Marotz,[1] Se Jin Song,[3] Stephen Wandro,[3] Livia S. Zaramela,[1,10] Rodolfo A. Salido,[1,3,6] Qiyun Zhu,[1,11] Erick Armingol,[1,2] Yoshiki Vázquez-Baeza,[3,12] Daniel McDonald,[1] James T. Sorrentino,[1,2] Bryn Taylor,[13] Pedro Belda-Ferre,[1] Promi Das,[1,14] Farhana Ali,[1] Chenguang Liang,[1,6,15] Yujie Zhang,[6,16] Luca Schifanella,[17,18] Alice Covizzi,[19] Alessia Lai,[19] Agostino Riva,[19] Christopher Basting,[17] Courtney Ann Broedlow,[17] Aki S. Havulinna,[20,21] Pekka Jousilahti,[20] Mehrbod Estaki,[1] Tomasz Kosciolek,[1,22] Rayus Kuplicki,[23] Teresa A. Victor,[23] Martin P. Paulus,[23] Kristen E. Savage,[24] Jennifer L. Benbow,[24,25] Emma S. Spielfogel,[24] Cheryl A. M. Anderson,[26] Maria Elena Martinez,[26] James V. Lacey Jr,[24] Shi Huang,[1,3,27] Niina Haiminen,[28] Laxmi Parida,[28] Ho-Cheol Kim,[29] Jack A. Gilbert,[1,3,14] Daniel A. Sweeney,[30] Sarah M. Allard,[1,14] Austin D. Swafford,[3,31] Susan Cheng,[32,33] Michael Inouye,[34,35,36] Teemu Niiranen,[20,37] Mohit Jain,[38] Veikko Salomaa,[20] Karsten Zengler,[1,3,6] Nichole R. Klatt,[17] Jeff Hasty,[6,39] Olivier Berteau,[7] Aaron F. Carlin,[9] Jeffrey D. Esko,[4,40] Nathan E. Lewis,[1,3,6,12,41,42] Rob Knight[1,3,6,43]

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT The gastrointestinal (GI) tract is a site of replication of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and GI symptoms are often reported by patients. SARS-CoV-2 cell entry depends upon heparan sulfate (HS) proteoglycans, which commensal bacteria that bathe the human mucosa are known to modify. To explore human gut HS-modifying bacterial abundances and how their presence may impact SARS-CoV-2 infection, we developed a task-based analysis of proteoglycan degradation on large-scale shotgun metagenomic data. We observed that gut bacteria with high predicted catabolic capacity for HS differ by age and sex, factors associated with coronavirus disease 2019 (COVID-19) severity, and directly by disease severity during/after infection, but do not vary between subjects with COVID-19 comorbidities or by diet. Gut commensal bacterial HS-modifying enzymes reduce spike protein binding and infection of authentic SARS-CoV-2, suggesting that bacterial grooming of the GI mucosa may impact viral susceptibility.

IMPORTANCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for coronavirus disease 2019, can infect the gastrointestinal (GI) tract, and individuals who exhibit GI symptoms often have more severe disease. The GI tract's glycocalyx, a component of the mucosa covering the large intestine, plays a key role in viral entry by binding SARS-CoV-2's spike protein via heparan sulfate (HS). Here, using metabolic task analysis of multiple large microbiome sequencing data sets of the human gut microbiome, we identify a key commensal human intestinal bacteria capable of grooming glycocalyx HS and modulating SARS-CoV-2 infectivity *in vitro*. Moreover, we engineered the common probiotic *Escherichia coli* Nissle 1917 (EcN) to effectively block SARS-CoV-2 binding and infection of human cell cultures. Understanding these microbial interactions could lead to better risk assessments and novel therapies targeting viral entry mechanisms.

KEYWORDS SARS-CoV-2, Covid, human microbiome, aging, Heparan Sulfate

Although the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for respiratory symptoms associated with coronavirus disease 2019 (COVID-19), the gastrointestinal (GI) epithelium has also been shown to be directly

Editor Samuel I. Miller, University of Washington, Seattle, Washington, USA

Address correspondence to Rob Knight, robknight@ucsd.edu.

Cameron Martino, Benjamin P. Kellman, Daniel R. Sandoval, and Thomas Mandel Clausen contributed equally to this article. Author order was determined in order of increasing seniority.

Jeffrey D. Esko, Nathan E. Lewis, and Rob Knight contributed equally to this article.

The authors declare a conflict of interest (see Acknowledgments).

See the funding table on p. 16.

Received 23 December 2024
Accepted 30 January 2025
Published 25 February 2025

[This article was published on 25 February 2025 with an error in Michael Inouye's surname. The byline was corrected in the current version, posted on 20 March 2025.]

infected (1). GI symptoms are reported in 20.3% of COVID-19 cases (2, 3), GI symptoms frequently appear early in infection before other symptoms develop (4, 5), and individuals with COVID-19 who exhibit GI symptoms often have more severe disease (6). Furthermore, there is evidence that enteric tissue acts as a replicative reservoir for SARS-CoV-2 that can prolong and increase the infectious burden on the host (7–14). Therefore, uncovering the mechanisms and predisposing factors governing potential GI infection by SARS-CoV-2, may lead to better risk stratification and possibly lead to improved therapeutic interventions for COVID-19.

Cell entry and infection by SARS-CoV-2 is dependent upon spike (S) protein binding to both heparan sulfate (HS) and angiotensin-converting enzyme 2 (ACE2) (15, 16), which are present in the GI tract (17). HS is a highly negatively charged component of the glycocalyx, a dense forest of glycans and glycoconjugates that coats all living cells (18). This dense surface of exposed glycocalyx is often the first point of contact for many viruses, including SARS-CoV-2 (19). Nearly all human epithelial surfaces are also home to microbes. The abundance and composition of these microbes, termed the human microbiome, have a close relationship with human health and disease (20–22).

The glycocalyx component of the mucosa covering the large intestine consists of two layers. The inner layer consists of tightly packed glycosaminoglycans (GAGs), many of which are O-glycosylated, directly anchored to epithelial cells and the outer layer which is a loose assembly of movable GAGs (23). Several host bacteria can produce enzymes that modify specific classes of GAGs, including HS (24) (Fig. 1A). Such glycan-metabolizing genes are often co-localized and co-regulated in polysaccharide utilization loci (PUL) (25). PULs have been annotated in human gut (e.g., *Bacteroides thetaiotaomicron*) (26–28), as well as in fresh water and soil isolates (e.g., *Pedobacter heparinus*) (29). It has been observed that removal of cell-surface HS via heparin lyase (HSase) purified from *P. heparinus* effectively eliminates SARS-CoV-2 virus infection and S protein binding (15). Consequently, we hypothesized that those bacteria that encode the ability to break down mucosa, such as *B. thetaiotaomicron* (26, 30, 31), may modulate entry of SARS-CoV-2 into cells of the GI tract.

RESULTS

To investigate this hypothesis, we sought to identify host bacteria able to modify host HS, and thereby potentially limiting SARS-CoV-2 viral entry, using a metabolic task-based analysis approach (32) to estimate the catabolic capacity of specific glycan types in each strain across microbial communities (see Materials and Methods). First, we curated a metabolic task describing HS modification and catabolism in the microbiome (Fig.

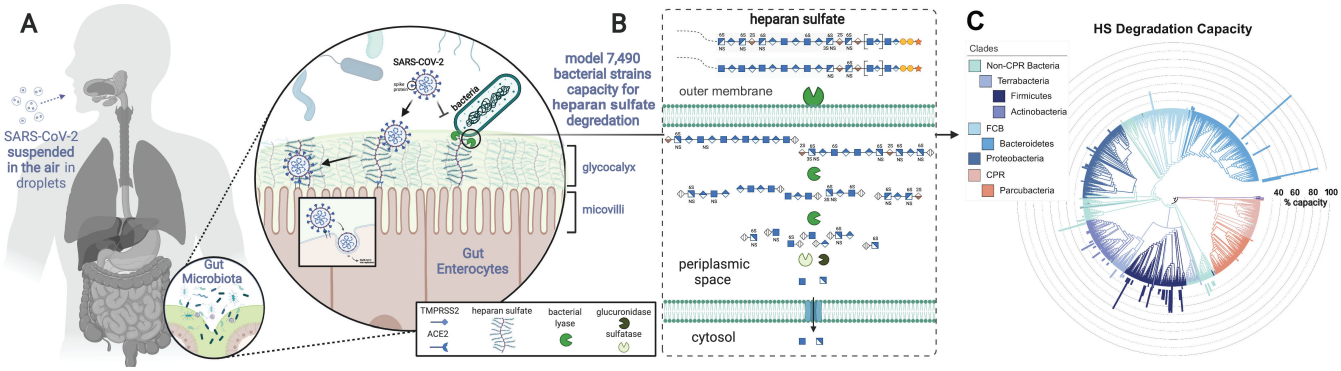

FIG 1 Metabolic task analysis of bacterial species capacity for heparan sulfate (HS) modification across the phylogenetic tree of life. SARS-CoV-2 relies on HS, a major component of the human gut mucosa, for host cell adhesion and entry. Human gut bacteria can alter mucosal surfaces potentially disrupting SARS-CoV-2 cell entry (A). Here, we examine the HS degrading genes in 7,490 bacterial strains measured in the FINRISK 2002 data set and quantify the HS catabolic capacity of each strain based on the set of tasks defined in degradation (B). Bacterial tree of life colored by superphylum groups and phyla containing predicted HS-modifying species. The bar chart represents the predicted capacity for HS modification of each species colored by phyla (C). Created in BioRender. Martino, C. (2025) https://BioRender.com/x09k421.

1B). We then quantified the total counts aligned to each glycosylation-associated gene in each microbe in the FINRISK 2002 data set, a fecal shotgun metagenomic data set spanning over 6,000 samples from participants ranging from 25 to 74 years old with a balanced sex ratio (55% female) (33, 34). Specific inclusion criteria for HS modification genes and HS catabolic measures are described in Materials and Methods. Of the 7,490 species identified in the FINRISK 2002 data set, 463 species of bacteria, which we refer to as HS-modifying bacteria, contained the complete HS catabolic task (completeness = 1) task with high HS catabolic capacity (capacity >90th percentile) (Fig. 1C). For example, three *Bacteroides* species, *B. xylanisolvens*, *B. thetaiotaomicron*, and *B. vulgatus*, were predicted to have extremely high capacity (99th percentile) for HS catabolism consistent with prior reports (26–28).

To determine if natural abundances of HS-modifying bacteria are associated with known risk factors and comorbidities of SARS-CoV-2 infection, we compared the relationship of age, sex, diabetes, smoking, body mass index (BMI), cancer, asthma, liver fibrosis or cirrhosis, cardiovascular disease, and autoimmune diseases (35, 36) to the log-ratio of bacteria predicted to modify HS versus all other bacteria (N-strains numerator = 687, denominator = 5,987) in the FINRISK 2002 study. Through an ordinary least squares (OLS) model of HS versus all other bacteria, significant changes by age (*t*-statistic = −2.84, *P* value = 0.005) and sex (*t*-statistic = −5.28, *P* value = 1.3×10^{-7}) but no other COVID-19 comorbidities were observed (Fig. 2A; Table S1). This finding was validated in an independent data set by matching the Web of Life (WoL) microbial genomes (37) to 16S rRNA gene amplicon sequence variants (ASVs) (see Materials and Methods) from over 20,000 samples in the American Gut Project (AGP), a citizen-science data set with participants ranging in age from 1 to 80 years (20). A similar age- and sex-dependent decrease in HS-modifying bacteria were observed (N-strains numerator = 244, denominator = 1,605) (Fig. 2B). Red meat and diet-derived accessible carbohydrates from dietary fiber are known to generally alter the abundance of mucin glycan degrading bacteria (38, 39). As quantified in the National FINRISK Study 2002 (FINRISK 2002) survey at the time of sample collection, after accounting for age and sex, there was not an overall significant difference in bacteria specifically encoded to degrade HS vs. all other bacteria except in men aged 45–55 by dietary fiber (Fig. 2C and D; Table S1). The alpha diversity across age and sex exhibited previously observed trends (40) (Fig. S1A and B), but HS-modifying bacteria had only a weak negative correlation with alpha diversity (Table S2).

To determine if there was a relationship between the frequency of HS-modifying bacteria and COVID-19 we used a previously published 16S fecal microbiome hospital data set of patients with COVID-19 (41). A subject's disease severity was determined based on their degree of respiratory support at the time of study participation; moderate disease was deemed for those requiring oxygen without invasive ventilation, and severe disease was for those requiring mechanical ventilation. HS-modifying bacteria were significantly decreased in patients with severe COVID-19 compared to those with moderate COVID-19 (*t*-statistic = −2.3, *P* value = 0.03) or uninfected healthcare workers (*t*-statistic = −3.9, *P* value = 0.001) (Fig. S2A). Analysis of a data set of moderate and severe patients with COVID-19 showed that HS-modifying genes relative to housekeeping genes again was significantly decreased in patients with severe COVID-19 (*t*-statistic = 2.9, *P* value = 0.014) (Fig. S2B). To show that abundance of HS-modifying bacteria is an independent risk factor for COVID-19 infection, we divided the FINRISK participants into those with and without comorbidities, and within each group compared the prior samples of subjects who had and had not later been diagnosed with COVID-19. Among the participants who were infected with SARS-CoV2, only those individuals without comorbidities had significantly diminished ratios of HS-modifying bacteria (Welch's *t*-statistic = 5.85, *P* value = 0.027) (Fig. S3A). No significant difference was observed in either HS lyase alone or alpha diversity in those with or without comorbidities (Fig. S3B and C). We then performed shotgun metagenomic sequencing on 416 additional subjects collected after infection through the AGP Microsetta Initiative (20) in collaboration with the California Teachers Study (CTS) and Tulsa 1000 study. This reiterated a

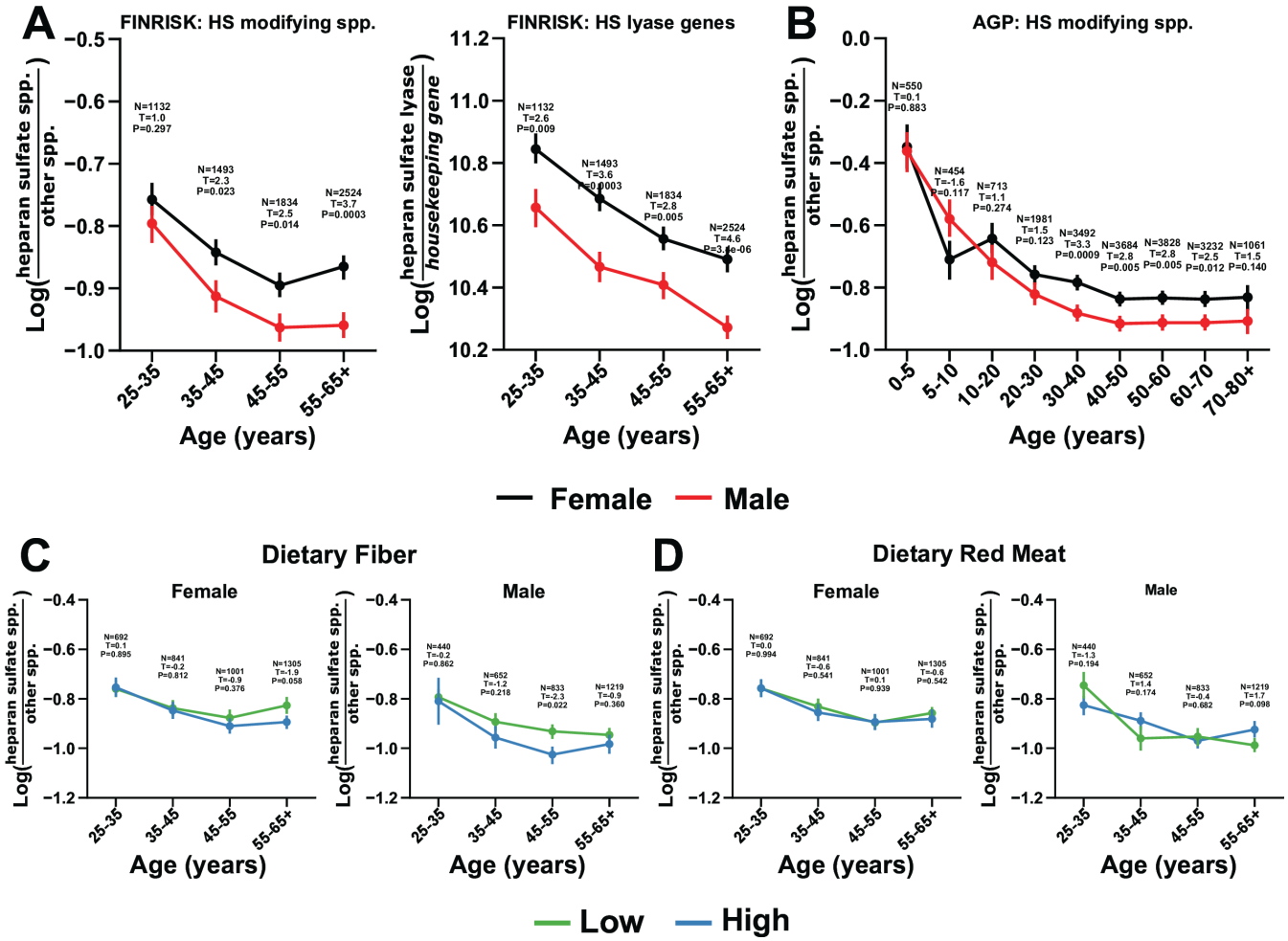

FIG 2 HS-modifying bacteria are inversely enriched according to host age, sex, but not diet. The log-ratio of predicted HS-modifying species relative to those with no predicted capacity for HS degradation (*y*-axis) in the FINRISK 2002 data set compared by host age (*x*-axis). The log-ratio of HS lyase genes relative to a set of housekeeping genes (*y*-axis) in the FINRISK 2002 fecal data compared by host age (*x*-axis) (A). The log-ratio of predicted HS-modifying species relative to those with no predicted capacity for HS degradation (*y*-axis) in the AGP fecal data set, compared over host age (*x*-axis) (B). Log-ratios are colored by participant sex (women, black; men, red). Log-ratio of predicted HS-modifying species relative to those with no predicted capacity (*y*-axis) in the FINRISK 2002 data set compared by dietary fiber and red meat intake above (high; blue) or below (low; green) the median consumption across ages (*x*-axis) and between sex (panel columns) (C and D). All log-ratio plots across age were annotated by the number of subjects at that time point. Error bars represent the standard error of the mean. Presented *P* values and test statistics are from unpaired two-tailed *t* tests evaluated on each host age group between host sex.

significant difference in HS-modifying genes relative to housekeeping genes in those who self-reported having COVID-19 and did not self-report any comorbidities (*t*-statistic = 2.3, *P* value = 0.0224) (Fig. S2D). Taken together, these results show that those with severe COVID-19 have decreased HS-modifying bacteria compared to those with moderate or no disease before, during, and after infection in an age- and sex-dependent manner.

To explore the hypothesis that human gut bacteria modulate HS presentation and therefore SARS-CoV-2 cell entry and infection, *in vitro* experiments were conducted to test the predicted catabolic capacity of these human microbiome species to degrade HS. Axenic cultures of *B. ovatus* and *B. thetaiotaomicron*, highly prevalent human gut bacterial isolates (42) found in 80% of AGP and 99% of FINRISK 2002 participants, were grown on a minimal medium in the presence of heparin (1.4 mg/mL) with and without glucose (22 mM). Both species were able to grow with heparin as the sole carbon source (Fig. 3A). Furthermore, all cultures were verified to catabolize heparin by comparing the concentration of heparin in the medium before and after growth to stationary phase (Fig.

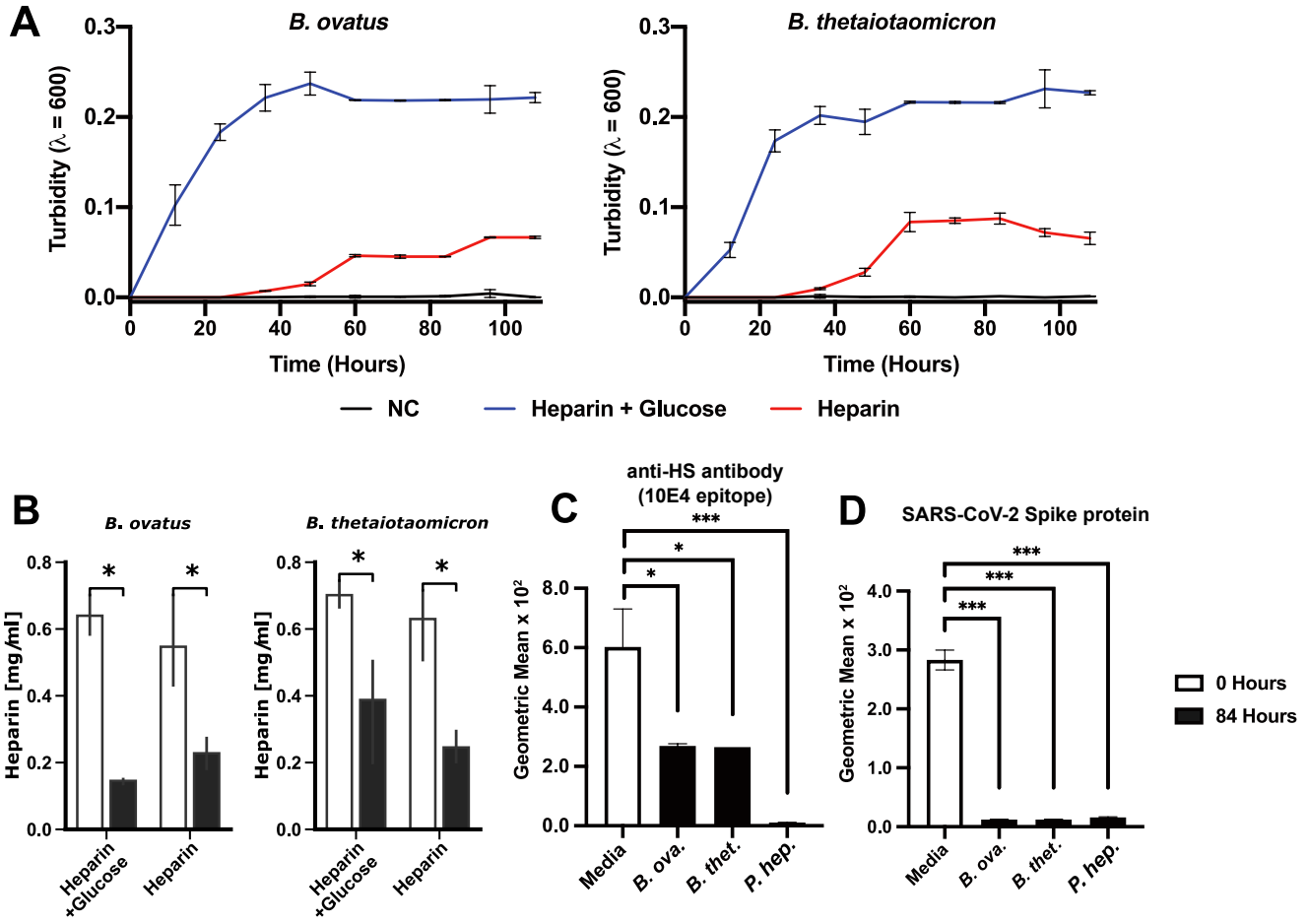

FIG 3 Commensal human gut bacteria block SARS-CoV-2 spike protein binding, through heparan sulfate degradation, in axenic culture supernatant. Growth of *Bacteroides ovatus* and *Bacteroides thetaiotaomicron* (A) measured by optical density (*y*-axis) across time from inoculation (*x*-axis) in minimal media (black; negative control NC), minimal media with 22 mM glucose and 1.4 mg/mL heparin (blue), or minimal media with 1.4 mg/mL heparin alone (red). Comparison of heparin concentration (*y*-axis; mg/mL) before inoculation (white; 0 h) and at stationary phase (gray; 84 h) for *B. ovatus* and *B. thetaiotaomicron* (B). Geometric mean of flow cytometry data (*y*-axis) of cultured human A549 bronchial epithelial cells stained with the anti-HS antibody 10E4 or incubated with biotinylated SARS-CoV-2 spike protein (D). Cells were incubated with culture media (Media), cell-free supernatant of *B. ovatus* (*B. ova*), or *B. thetaiotaomicron* (*B. thet.*) or purified heparin lyase from *Pedobacter heparinus* (*F. hep.*). Presented *P* values are from unpaired *t*-test statistics compared to the untreated control (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$).

3B). Cell-surface HS on H1299, a lung derived human epithelial cell line, was reduced by 60% upon exposure to cell-free supernatants from mid-log phase cultures of *B. ovatus* or *B. thetaiotaomicron* compared to 100% reduction by purified heparin lyase from *P. heparinus* (*F. hep*. HSase; IBEX pharmaceuticals), as measured by binding of the anti-HS monoclonal antibody 10E4 epitope (Fig. 3C). To determine the effect of bacterial HS modification on SARS-CoV-2 binding, H1299 cells were treated with the supernatant of *Bacteroides* cultures or purified *F. hep*. HSase. These were subsequently incubated with biotinylated trimeric SARS-CoV-2 S protein and assessed for cell-surface binding by flow cytometry. Cells treated with *Bacteroides* culture supernatant exhibited a 20- to 30-fold reduction in SARS-CoV-2 S protein binding compared to untreated H1299 cells (*B. ovatus* *t*-statistic = 31.25, *P* value = 7.019×10^{-5}; *B. thetaiotaomicron* *t*-statistic = 30.99, *P* value = 7.023×10^{-5}), similar to the reduction observed by pre-treatment with purified *F. hep*. HSase (15) (*t*-statistic = 23.89, *P* value = 7.59×10^{-5}) (Fig. 3D).

To identify the human commensal microbial HS-modifying enzymes capable of reducing SARS-CoV-2 S binding, purified GAG-specific sulfatases and HS lyases were

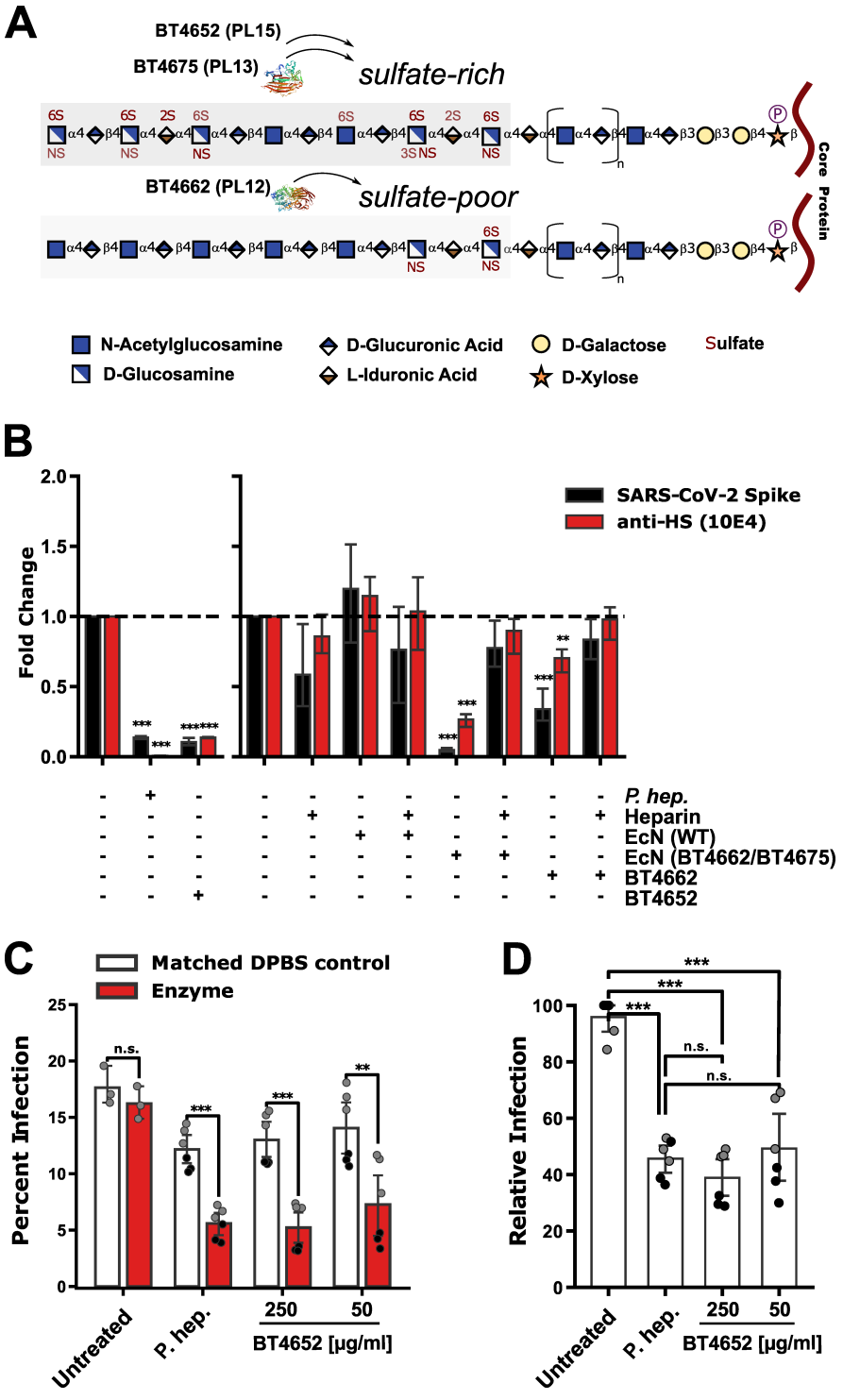

FIG 4 Purified enzymes derived from human commensal gut bacteria block SARS-CoV-2 spike protein binding and viral infection through heparan sulfate modification. Schematic diagram of activity and putative specificity on typical highly and lowly sulfated HS. The bacterial species sourced for enzymes with X-ray crystallographic structures (A). Fold change of the geometric mean of flow cytometry data across two experiments relative to untreated control (*y*-axis) of cultured human A549 cells stained with the HS antibody 10E4 (red) or incubated with biotinylated SARS-CoV-2 spike protein (black) compared with and without 100 µg/mL heparin as a competitive substrate between purified *Pedobacter heparinus* (*F. hep.*) heparin lyase (Hsase), purified *B. thetaiotaomicron* heparin Lyase BT4652 (polysaccharide Lyase family 15; PL15), *Escherichia coli* strain Nissle 1917 (EcN) wild-type (WT) culture supernatant, EcN encoding *B. thetaiotaomicron* HS targeting genes BT4662 (polysaccharide

(Continued on next page)

Fig 4 (Continued)

Lyase family 12; PL12) and BT4675 (polysaccharide Lyase family 13; PL13) (B). SARS-CoV-2 infection of Vero cells performed in the absence and presence of *P. heparinus* Hsase, or with different concentrations of *B. thetaiotaomicron* purified BT4662 heparin lyse. The graph shows the percent infection from a composite of two separate experiments, each performed in triplicate (C). The same data but with the experimental data normalized to the mock infection for each respective experiment (D). Presented *P* values are from unpaired *t*-test statistics w.r.t. no treatment (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$).

tested for their capacity to modulate HS presentation and S protein binding. A total of six human host-resident bacterial exo-active sulfatases from *B. thetaiotaomicron* with different substrate specificities or predicted sulfatases: Δ4,5 hexuronate-2-*O*-sulfatase (BT1596), N-acetylglucosamine-6-sulfatase (BT4656), putative mucin-desulfating sulfatase (BT3177), putative glucosamine N-sulfatase (BT4655), putative iduronate 2-sulfatase (BT0756), and a putative undefined GAG sulfatase (BT1624), were purified after expression in an *Escherichia coli* strain containing a sulfatase maturating system (28) (Fig. S4A). A549 human bronchial epithelial cells were treated with each sulfatase, incubated with S protein, and then assessed for cell-surface binding of S protein by flow cytometry. None of the individual sulfatase treatments significantly decreased S protein binding (Online supplementary file 1) despite the importance of sulfatases for GAG's degradation by *B. thetaiotaomicron* (28).

Next, human commensal bacterial HS lyases were tested for the capacity to reduce SARS-CoV-2 S protein binding. Specifically, three lyases sourced from *B. thetaiotaomicron*: BT4662 (polysaccharide Lyase family 12; PL12) a depolymerizing cell surface lyase that targets sulfate-poor HS, BT4652 (polysaccharide Lyase family 15; PL15), and BT4675 (polysaccharide Lyase family 13; PL13) capable of degrading sulfate-rich HS (Fig. 4A). Enzymes BT4662 and BT4675 were co-expressed in the generally recognised as safe (GRAS) and widely used probiotic *E. coli* strain Nissle 1917 (EcN) with unique purification tags (Fig. S4C). The human cell line A549 was treated with culture supernatant from both the EcN wild type (WT), engineered EcN (BT4662-BT4675), purified lyase BT4652, and a positive control of purified *P. heparinus* HSase (IBEX pharmaceuticals). As demonstrated previously, the positive control, *P. heparinus* HSase eliminated both host-cell HS as detected by 10E4 staining (*t*-statistic = −1,394.6, *P* value = 1.58×10^{-12}) and binding of recombinant SARS-CoV-2 S protein (*t*-statistic = −95.26, *P* value = 7.28×10^{-8}). EcN (WT) culture supernatant produced no significant difference in host cell-surface expression of HS or SARS-CoV-2 S protein binding. On the other hand, EcN (BT4662-BT4675) culture supernatant significantly decreased host-cell surface HS (*t*-statistic = −22.71, *P* value = 2.23×10^{-5}) and significantly reduced SARS-CoV-2 S binding (*t*-statistic = −59.83, *P* value = 4.67×10^{-7}). Similarly, purified BT4662 or BT4652 ablated SARS-CoV-2 S binding (BT4662, *t*-statistic = −80.71, *P* value = 1.41×10^{-7}; BT4652, *t*-statistic = −123.28, *P* value = 6.09×10^{-13}). Finally, both HS degradation and SARS-CoV-2 S protein binding were prevented by including 100 µg/mL heparin as a competitive substrate in the pre-incubation medium for all *B. thetaiotaomicron* lyases (Fig. 4B; Fig. S5).

To validate the impact of binding on viral infection, we further tested authentic SARS-CoV-2 virus infection using strain USA-WA1/2020. Vero E6 cells were monitored by staining of the cells with antibodies against the SARS-CoV-2 nucleocapsid (N) protein. Cells pretreated for 1 h were infected with virus for 30 min ,followed by the application of methylcellulose overlay to restrict the spread of infection. The impact of infection (MOI of 0.5) was assayed 20 h post-infection (p.i.). Treatment with bacterial heparin lyase in this system has no effect on ACE2 expression (15). Relative infection was calculated by normalizing percent infection values obtained in the absence of any treatment. The positive control treatment of *P. heparinus* HSase significantly reduced percent infection (*t*-statistic = −7.2, *P* value = 2.9×10^{-5}) and relative infection (*t*-statistic = −12.8, *P* value = 7.69×10^{-7}) by ~2-fold on average. The treatment of purified BT4652 alone, on average, significantly reduced percent infection (50 µg/mL, *t*-statistic = −3.4, *P* value = 6.66×10^{-3}; 250 µg/mL, *t*-statistic = −6.2, *P* value = 1.01×10^{-4}) and relative infection (50 µg/mL,

t-statistic = −6.61, *P* value = 2.96×10^{-4}; 250 µg/mL, *t*-statistic = −12.01, *P* value = 1.44×10^{-6}) by ~2-fold and ~2.5-fold at a concentration of 50 and 250 µg/ml, respectively (Fig. 4C and D). Both low and high concentrations of purified BT4652 reduced infection to the same extent as *P. heparinus* HSase (Fig. 4C and D). Taken together, these findings demonstrate the potential of commensal human gut bacteria to reduce HS presentation, SARS-CoV-2 S protein binding, and infection in the absence of high quantities of competitive substrate.

DISCUSSION

In summary, our metabolic task analysis of the human gut microbiome identified a key commensal human intestinal bacteria capable of grooming glycocalyx HS and modulating SARS-CoV-2 infectivity *in vitro*. We observe decreases in predicted HS-modifying bacteria in human gut microbiomes across age and sex in two large survey data sets, in those populations who later contracted SARS-CoV-2, and in severe versus moderate COVID-19 patients in the absence of comorbidities. These data may help to explain age-dependent health differences in the presentation of GI symptoms, infection, and severity of COVID-19. Moreover, the abundance of HS-encoded bacteria is not observed to be consistently enriched in higher fiber diets when accounting for age and sex. Typically, PUL-encoding intestinal bacteria catabolize the outer layer of the mucosa in combination with dietary fiber, leaving the inner layer free of bacteria. In case of decreased dietary fiber, the GI bacteria shift to primarily catabolize host mucosa, including those GAGs found on the inner layer, leaving the mucosa significantly reduced in thickness (31, 39, 43, 44). Our experimental results also demonstrate that the addition of a preferential substrate could reduce the impact of heparin lyase. Therefore, increased fiber diets should be encouraged due to a larger systemic positive impact on COVID-19 severity (45) and overall health (39, 46, 47). However, more work is needed to determine how diet-induced mucosal thickness alteration and HS modification, together impact viral entry and infection across age and gender. Similarly, the microbiome of patients with COVID-19 is also disrupted, with receiving broad-spectrum antibiotics for extended periods of time (41). This further highlights the need for more research to better understand the microbiomes impact, and the impact of the microbiome's interaction with the glycocalyx, as a risk factor during infection. Although not the focus of this work, similar mechanisms may be active in the respiratory system, which has a strikingly different, but equally important, microbial community and functional repertoire from the gut microbiome (22, 29, 48–50). For example, the human oral microbiome from the AGP data set demonstrated an age-dependent change in HS-modifying bacteria, similar to the gut microbiome samples (Fig. S6). Unfortunately, there are no comparably large oral, nasal, or respiratory microbiome data sets that could be used to explore the contribution of other risk factors. Additionally, there is growing evidence to support the idea of a gut-respiratory axis, where gut microbiota influences the respiratory microbiota/pathology and vice versa (49–51). Altering HS in the gut or in the respiratory tract may thus directly lead to respiratory cell entry and infection blocking of SARS-CoV-2. Further, altering HS in the gut or respiratory tract could play a role in other diseases, including infectious diseases. Therefore, this study opens the possibility of studying similar impacts across various sites of the body to predict, diagnose, and even improve treatment strategies.

These results provide a proof-of-concept for applying metabolic task analysis to quantify the catabolic capacities of microbial communities, particularly in glycomics. Through the use of a synthetic biology approach, we validated our metabolic task analysis and observations in large-scale omics survey data *in vitro*. We observed that the common probiotic EcN can be engineered to effectively block SARS-CoV-2 binding and infection of human cell cultures. Future work will be needed to validate these findings *in vivo*. The systems-level approaches and synthetic biology-driven validations used here can be used to explore the trans-kingdom interplay between resident microbes and host-viral infection.

MATERIALS AND METHODS

HS modification capability: completeness and capacity

Many approaches exist to infer or quantify the activities of metabolic gene sets, including topological pathway analyses and mechanistic models (52). However, the use of these methods in the analysis of microbiome omics data can be difficult given the large numbers of microbial species, each with their own set of genes, which require expert curation to define their functions and their collective impact on a species' metabolic capabilities (53, 54). Accounting for glycan metabolism is particularly important since microbiomes are shaped by their ability to metabolize the host and environmental glycome (55, 56). Metabolic task analysis can rapidly estimate the metabolic capabilities of a cell (32). In this, a metabolic task is defined as the genes associated with the reactions necessary to convert a metabolite into other products. Leveraging this concept in glycan metabolism, the HS modification gene set was treated as a metabolic task, that is, a group of reactions necessary to transition between metabolites (32) used to describe a complete pathway of HS modification. Pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG, ec00531) (57), in combination with literature annotation of these pathways for specific bacteria (27, 58, 59), were used to identify reactions (enzyme commission [EC] numbers) associated with the modification task (Table S3). Finally, we used CAZy, dbCAN, and CUPP (51, 60–62) to map EC numbers to microbial genes associated with glycan modification (Table S3).

We created two metrics to describe the HS-catabolic capabilities of microbes. For each microbe in the FINRISK2002 data set, we estimated pathway "Completeness," a binary indication of the presence or absence of all genes in the task, and pathway "Capacity," a continuous indication of the magnitude of flux that could travel through the task. The abundance of genes associated with glycan catabolism was quantified by the total number of reads (total count) mapped between the HS-modification gene set and FINRISK 2002 sample shotgun data through Bowtie2. A pseudocount was added to each count and log-transformed to stabilize the variance.

HS-catabolic completeness for each organism, *i*, with gene, *g*, is an indication that all reactions in a task are represented in an organism. Specifically, for each reaction (EC) in the HS task, we sum the log-counts aligned to glycosylation-related genes within an organism associated with that EC. If the sum of log counts aligned to an EC exceeds a threshold, *t*, the EC was marked as active in that organism. The threshold, *t*, was set at the fifth percentile of log-counts aligned to glycosylation-associated genes in the FINRISK 2002 database, 4.522 log-counts. If all ECs in the HS task were active in an organism, the HS task is considered complete.

$$O_i = \forall_{EC \in \text{Task}} t < \sum_{g}^{EC} \sum_{g}^{EC} \log(\text{read}_g)$$

To quantify the HS-catabolic capacity of each organism, *i*, with genes, *g*, we analyzed the expression of all glycosylation-related genes within the catabolic task. Adapted from metabolic task analysis (32, 63, 64), the catabolic capacity of a microbe is a context-sensitive measure that accommodates rate-limiting or low-abundance enzymes and can therefore accommodate microbial gene transience. Activation of the glycan degrading metabolic task (32), was calculated as the minimum EC activation; the maximum activation of genes performing the EC reaction within an organism. Log of total counts was used to stabilize the variance. The capacity is the minimum EC activation in a pathway where EC activation is the maximum gene activity score for each EC within a microbe. The log of total counts aligned to each gene in the FINRISK 2002 data set was used for the gene activity score:

$$A_i = \min_{EC \in \text{Task}} \max_{g \in EC} \log(\text{reads}_g)$$

Sequencing data

Patient recruitment, sample collection, storage, and molecular processing are described in detail previously (33, 34). Demultiplexed shallow shotgun metagenomic sequences were quality filtered and adapter trimmed using Atropos (65), and human filtered using Bowtie2 (66). For taxonomic assignment, reads were aligned to the WoL database (37) of 10,575 bacterial and archaeal genomes using Shogun (67) in the Bowtie2 alignment mode. For functional assignment of glycosylation-associated genes, the FINRISK 2002 data were aligned to the HS degradation gene set of glycosylation-associated genes using Bowtie2. Subjects who contracted COVID-19 subsequent to sampling were retrospectively collected in September 2020 ($N = 21$).

The prospective observational CTS includes N = 133,477 female participants who have been followed continuously since 1995–1996 (www.calteachersstudy.org). In early 2020, a subset of CTS participants who (i) lived in Los Angeles, Orange, or San Diego counties and (ii) had previously donated biospecimens were invited to participate in the Microsetta Initiative and AGP. All participants provided written informed consent for the data and biospecimens they contributed.

Abundance and expression of HS-modifying bacteria by log-ratios

Differential abundance or expression of HS-modifying bacteria was determined by the log-ratio of mapped reads for each sample of those bacteria with predicted capacity for HS modification relative to those without predicted capacity. Functional gene abundance or expression was determined with a log-ratio of the total mapped reads to that gene relative to the sum of counts of total mapped reads from a set of housekeeping genes in each sample. The housekeeping gene set is comprised of all bacterial nucleotide sequences for the genes atpD, dnaJ, gyrA, gyrB, infB, pheS, proC, rpoA ,rpoB, and rpoD obtained from RefSeq (68). Significance between groups of log-ratios was determined with an unpaired *t* test through SciPy (69).

Mapping of WoL to AGP using 16S ASVs

To reconcile the evolutionary relationships among 16S rRNA gene sequencing ASVs and shotgun metagenomic data, we mapped the ASVs to the WoL (37) reference phylogeny of bacterial and archaeal genomes. First, 16S rRNA genes were annotated from each of the 10,575 genomes included in the phylogeny using RNAmmer 1.2 (70), using domain-specific models (bacteria and archaea, respectively). Second, filtered 150 bp length 16S V4 AGP (20), ASVs (n = 15,486) were aligned to the WoL 16S rRNA genes using BLASTn 2.7.1+ (71), with an *e*-value threshold of 1e−5 and up to 100 target sequences per query. Top hits with identical bit scores of each query were retained and subjected to the taxonomic classification. At each designated rank, taxonomic assignments of all top hits were recorded. For feature table generation, the hits were counted and normalized by the total number of hits. As an example, assuming one ASV aligned equally well to five reference full-length 16S sequences, and they belong to genus A (two sequences) and genus B (three sequences), then the two genera were counted as 2/5 (A) and 3/5 (B), respectively. Per-query counts were summed across each AGP sample and rounded to integers.

Cultivation of *Bacteroides* strains

B. thetaiotaomicron and *Bacteroides ovatus* were cultured in 50 mL anaerobic brain heart infusion (BHI) medium with an overnight incubation at 37°C in an anaerobic serum bottle. The next day, 1 mL of cells from the BHI culture was washed in anaerobic phosphate-buffered saline (PBS) and were passed into 15 mL of minimal medium containing no carbon sources or electron donors other than glucose (22 mM final) and/or heparin (1 mg/mL final). Growth on the minimal medium was measured by optical density at 600 nm every 12 h. Aliquots of 3 mL of culture were taken before

inoculation, immediately after inoculation, at mid-log phase (24 h), and in stationary phase (45 h). Heparin degradation in culture was measured through a Blyscan glycosaminoglycan assay (Biocolor Ltd., Carrickfergus, Northern Ireland) using 100 µL of time 0- and 45-h culture media. All cultivation and HS-modification experiments were conducted in triplicate.

To verify strain taxonomy a 1 mL aliquot of each culture was extracted using PowerFecal DNA Isolation Kit (MoBio cat. 12830). Extracted DNA was quantified via Qubit dsDNA HS Assay (Thermo Fisher Scientific), and 5 ng of input DNA was used in a 1:10 miniaturized Kapa HyperPlus protocol (72). The pooled library was sequenced as a paired-end 150-cycle run on an Illumina NovaSeq at the UCSD IGM Genomics Center. The resulting sequences were adapter trimmed using Trimmomatic v0.39 (73) and human read filtered with Bowtie2 (74). Paired-end reads were merged using Flash v1.2.11 (75). Each axenic sample was assembled through SPAdes (76) and verified through average nucleotide identity (77) of greater than 99% between the assembled genome and the putative type-strain genome obtained from NCBI (68).

SARS-CoV-2 spike protein production

Recombinant SARS-CoV-2 spike protein-encoding residues 1–1,138 (Wuhan-Hu-1; GenBank: MN908947.3) with proline substitutions at amino acids positions 986 and 987 and a "GSAS" substitution at the furin cleavage site (amino acid positions 682–682), was produced in ExpiCHO cells by transfection of 6×10^6 cells/mL at 37°C with 0.8 g/mL of plasmid DNA using the ExpiCHO expression system transfection kit in ExpiCHO Expression Medium (ThermoFisher). After 1 day, the cells were refed, then incubated at 32°C for 11 days. The conditioned medium was mixed with cOmplete EDTA-free Protease Inhibitor (Roche). The recombinant protein was purified by chromatography on a Ni2+Sepharose 6 Fast Flow column (1 mL, GE LifeSciences). Samples were loaded in ExpiCHO Expression Medium supplemented with 30 mM imidazole, washed in a 20 mM Tris-HCl buffer (pH 7.4) containing 30 mM imidazole and 0.5 M NaCl. The recombinant protein was eluted with buffer containing 0.5 M NaCl and 0.3 M imidazole. The protein was further purified by size exclusion chromatography (HiLoad 16/60 Superdex 200, prep grade, GE LifeSciences) in 25 mM HEPES buffer (pH 7.5) containing 0.3 M NaCl.

Biotinylation

For binding studies, recombinant SARS-CoV-2 spike protein was conjugated with EZ-Link Sulfo-NHS-Biotin (1:3 molar ratio; Thermo Fisher) in Dulbecco's PBS at room temperature for 30 min. Glycine (0.1 M) was added to quench the reaction, and the buffer was exchanged for PBS using a Zeba spin column (Thermo Fisher).

SARS-CoV-2 spike binding experiment

NCI-H1299 or A549 cells from the American Type Culture Collection were grown in RPMI medium containing 10% FBS and 100 U/mL penicillin and 100 µg/mL streptomycin sulfate under an atmosphere of 5% CO_2 and 95% air. The cells at 50–80% confluence were lifted in 10 mM EDTA in PBS (Gibco) and washed in PBS containing 0.5% BSA. The cells were seeded into a 96-well plate at 10^5 cells per well. The cells were then treated with a mix of 2.5 mU/mL *P. heparinus* Hsase II and 5 mU/mL *P. heparinus* Hsase III (IBEX Pharmaceuticals) in PBS containing 0.5% BSA (100 µL), 100 µL *B. ovatus*, and *thetaiotaomicron* or *E. coli* strain Nissle 1917 (ECN) minimal media culture cell-free supernatant supplemented with 10% BSA, or purified sulfatase or lyase enzymes at the given concentration for 30 min at 37°C. The cells were washed two times in PBS containing 0.5% BSA. The cells were then stained with 25 µg/mL biotinylated S protein (S1/S2) or 1:1,000 Anti-HS (Clone F58-10E4) (Fisher Scientific, NC1183789) in PBS containing 0.5% BSA, for 30 min at 4°C. The cells were washed two times and then stained with Streptavidin-Cy5 (Thermo Fisher) at 1:1,000 (S protein) or Anti-IgM-Alexa488 at 1:1,000 (Anti-HS) in PBS containing 0.5% BSA, for 30 min at 4°C. The cells were washed two

times and then analyzed using a FACSCanto instrument (BD bioscience). Data analysis was performed using FlowJo, and statistical analyses were conducted in Prism 8.

Recombinant sulfatase production

Genes coding for carbohydrate sulfatases *BT0756*, *BT1596*, *BT1624*, *BT3177*, *BT4655*, and *BT4656* from *B. thetaiotaomicron* were cloned into a pET28b or a pRSF plasmid with an *N*-terminal His_6 TAG. The sulfatase-maturating enzyme (anSME) was systematically co-expressed with the different genes to ensure optimal maturation of the sulfatases (42, 78) except for BT4655, which is not predicted as a formylglycine-dependent sulfatase (42). For *BT0756*, *BT1596*, and *BT4656*, the serine residue, target of the post-translational modification catalyzed by anSME, was mutated into a cysteine residue in order to improve conversion into $C_a$formylglycine, as previously reported (28). Plasmids were transformed into *E. coli* BL21(DE3) and the recombinant strains were grown in Luria-Bertani (LB) medium containing the suitable antibiotic (i.e., 100 µg/mL ampicillin or 50 µg/mL kanamycin, for pET28b or pRSF plasmids, respectively) at 37°C with an agitation speed of 170 rpm until OD_{600} reached 0.7. Protein expression was induced by adding 200 µM isopropyl β-D-thiogalactopyranoside and temperature decreased to 20°C for 18 h. Bacterial pellets were harvested by centrifugation (5,500 rpm, 10 min at 4°C) and suspended in buffer A: Tris 50 mM, KCl 100 mM, $MgCl_2$ 10 mM, pH 8, supplemented with 1% vol/vol Triton X-100 and 2-mercaptoethanol. The cells were disrupted by sonication and supernatant clarified by ultracentrifugation (45,000 rpm, 1 h at 4°C). The supernatant was loaded onto a Ni-NTA column previously equilibrated with buffer A. The column was washed with 10 column volumes of buffer A and sulfatases were eluted with buffer A containing 300 mM imidazole. Imidazole was removed using a PD-10 desalting column equilibrated in buffer A. The sulfatase-containing fractions were concentrated with an Amicon concentrator (cutoff of 10 kDa, Millipore), and purity was assayed by SDS–PAGE and mass spectrometry analysis.

Recombinant lyase BT4652 production

The *BT4652* lyase gene (synthesized by *GeneCust*) was ligated into a pET- His_6-MBP-$^{(TEV)}$ vector to produce BT4652 as a fusion protein with an His_6-tagged MBP and a TEV cleavage sequence. *E. coli* BL21 (DE3) was transformed with the plasmid construct (pET-His_6-MBP-$^{(TEV)}$-BT4652) and overexpression was performed in LB medium containing 50 µg/mL kanamycin. Bacterial pellets were harvested by centrifugation and suspended in buffer B: Tris 50 mM, NaCl 300 mM, pH 8, supplemented with 1% vol/vol Triton X-100 and 2-mercaptoethanol. The cells were disrupted by sonication and supernatant clarified by ultracentrifugation (45,000 rpm, 1 h at 4°C). The supernatant was loaded onto an amylose resin column previously equilibrated with buffer B. The column was washed with 10 column volumes of buffer B and MBP-BT4652 protein eluted with the same buffer containing 10 mM maltose. Cleavage of the His_6-MBP tag was performed using the His_6-TEV protease overnight at 4°C. TEV protease and MBP protein were removed by purification on Ni-NTA column and BT4652 was further purified by size-exclusion chromatography using an ÅkTA system and a Superdex S200 column equilibrated with buffer B. Fractions of BT4652 lyase were analyzed by SDS–PAGE and those containing pure BT4652 were pooled and concentrated (Amicon concentrator, cutoff of 10 kDa, Millipore).

Lyase expression in *E. coli* Nissle 1917 (EcN)

Lyases BT4675 and BT4662 were amplified from the genome of *B. thetaiotaomicron*, and inserted into pRSF-Duet using Golden Gate Assembly. The resulting plasmid was transformed into *E. coli* Nissle 1917 via electroporation. For lyase expression, a starter culture of *E. coli* Nissle 1917 containing the lyase plasmid was grown overnight in LB media containing 50 µg/mL of kanamycin, and then used to inoculate 500 mL of the same media. The cells were grown until they reached optical density 0.1–0.2, induced

with 1 mM IPTG, and incubated overnight at room temperature. The cells were pelleted by centrifugation at 5,000 × *g* for 5 min, and the supernatant was removed for further experiments. The recombinant protein was purified by flowing over a Ni2+Sepharose 6 Fast Flow column (1 mL, GE Life Sciences), and eluted with 0.3 M imidazole, 0.5 M NaCl, and 25 mM HEPES pH 7.5. Protein was concentrated and buffer exchanged using a PES Pierce Protein concentrator (Thermo).

Virus

All work with SARS-CoV-2 was conducted in Biosafety Level-3 conditions at the University of California San Diego following the guidelines approved by the Institutional Biosafety Committee. SARS-CoV-2 isolate WA1 (USA-WA1/2020, BEI NR-52281) was passaged through Caco2 cells and then expanded on TMPRSS2-VeroE6 cells. Supernatants were purified and stored at −80°C and titers were determined by fluorescent focus assay on TMPRSS2-VeroE6 cells. Virus stock was verified using deep sequencing analysis.

Lyase assay

VeroE6 cells in 96-well plates were washed with DPBS and pretreated 1 h with *B. thet.* Purified BT4662 heparin lyase, *F. hep.* HSase, or matched volumes of vehicle (DPBS) in 50 µL of DMEM. The cells were then infected with authentic SARS-CoV-2 at a MOI of 0.5 for 30 min in the presence of enzyme or DPBS control. Virus was removed, and the cells were washed two times with 100 µL PBS and overlaid with methylcellulose (1% methylcellulose in MEM + 2% FBS, 1× Pen/Strep, 2 mM L-glutamine, and 1× non-essential amino acids). After 20 h of incubation, the cells were fixed for 30 min in 4% formaldehyde and stained with anti-nucleocapsid primary antibody (GeneTex, gtx135357) and anti-rabbit AlexaFluor 594 secondary with SytoxGreen nuclear counterstain. Plates were imaged on an Incucyte S3 imager. The total cell number and percent of cells infected were measured using the Incucyte onboard software tools.

ACKNOWLEDGMENTS

The authors would like to thank the California Teachers Study Steering Committee that is responsible for the formation and maintenance of the Study within which this research was conducted. A full list of California Teachers Study team members is available at https://www.calteachersstudy.org/team.

This work was supported in part by IBM Research AI through the AI Horizons Network, the UC San Diego Center for Microbiome Innovation, NIH Pioneer award 1DP1AT010885, RAPID grant 2038509 from the National Science Foundation, University of California on COVID-19 grant R00RG2503, and Emerald Foundation 3022 (to R.K.), NIH grants from the NIDDK grant 1P30DK120515, NIGMS R35GM119850, NIAID UH2 AI153029 and Novo Nordisk Foundation grant NNF20SA0066621 (to N.E.L.); RAPID grant 2031989 from the National Science Foundation, Project 3 of NIH P01 HL131474 and a subward from NIH 5U19AI116497 (to J.D.E.); the Alfred Benzon foundation (to T.M.C.); the Academy of Finland grant 321351 and the Emil Aaltonen Foundation (to T.N.); the Finnish Foundation for Cardiovascular Research (to V.S.); the NIH grant R01ES027595 (to M.J.); the ANID Becas Chile Doctorado 2018-72190270 (E.A.); the Academy of Finland grants 321356 and 335525 (A.S.H); the French National Research Agency (ANR) grants ANR-17-CE11-0014 and ANR-20-CE44-0005 (to O.B.); and NIH NIGMS R01GM069811 (to J.H.). M.I. is supported by the Munz Chair of Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) [*]. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was supported by core funding from the British Heart Foundation (RG/13/13/30194, RG/18/13/33946) and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was supported by Health Data Research UK, which is funded by the UK Medical

Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome. The California Teachers Study and the research reported in this publication were supported by the National Cancer Institute of the National Institutes of Health under award number U01-CA199277; P30-CA033572; P30-CA023100; UM1-CA164917; and R01-CA077398 (to J.L. and E.M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

C.M., B.P.K, T.M.C., D.R.S., A.D.S., J.D.E., N.E.L., and R.K. conceived, initiated, and coordinated the project. B.P.K, E.A., J.S., C.L., Y.Z., and N.E.L. conceived and performed pathway analyses. C.M., T.M.C., D.R.S., A.E.C., A.F.G, L.Z., A.B., F.S., O.B., R.C., and R.A.S.B. designed and performed the experimental work. C.M., B.P.K, A.D.S., S.J.S., S.W., E.A., Y.V.B., D.M., L.Z., S.C., Q.Z., P.D., F.A., P.D.F, J.G., D.A.S, S.M.A., B.T., M.I., V.S., A.R., A.L., A.C., C.B., C.A.B., R.K., T.V., M.P., K.E.S., J.L.B., E.S.S., C.A.M.A., M.E.M., J.V.L, M.J., A.S.H., P.J., T.N., and R.K. coordinated, compiled and analyzed sequencing data. B.P.K, T.M.C., K.Z., P.B.F, L.S., N.R.K., J.D.E., F.S., O.B., A.F.C., J.H., and R.K. supplied reagents. C.M., B.P.K., T.M.C., D.R.S., C.A.M., A.D.S., J.D.E., N.E.L., and R.K. wrote the manuscript with input from all authors. All authors discussed the experimental results and read and approved the manuscript.

C.M. is the founder of Leaven Foods, Inc., receives income and has equity. B.P.K. is a co-founder of Augment Biologics and has equity. D.R.S. is a full-time employee for Pfizer and receives income. C.A.M. is a full-time employee for Native Microbials, Inc., has equity, and receives income. Y.V.B is a full-time employee for BiomeSense, Inc., has equity, and receives income. D.M. is a consultant for BiomeSense, Inc., has equity, and receives income. The terms of these arrangements have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies. P.B.F. is a full-time employee of Element Biosciences, has equity, and receives income. C.L. is a full-time employee for Merck and receives income. M.E. is a full-time paid employee and has equity at Innovate Phytoceuticals Inc. and is a scientific advisor and holds equity at Melius Microbiomics Inc. T.K. is a co-founder of Onebiome Sp. zo.o. and has equity. J.A.G. is on the SAB for BiomeSense Inc., Holobiome Inc., SunGenomics, and Bened Life. M.I. is a trustee of the Public Health Genomics (PHG) Foundation and a member of the Scientific Advisory Board of Open Targets. He has research collaborations with AstraZeneca, Nightingale Health, and Pfizer which are unrelated to this work. V.S. has received honoraria for consultations from Novo Nordisk and Sanofi. He also has ongoing research collaboration with Bayer Ltd (all unrelated to the present study). K.Z. is cofounder of Isolatio Bio, Native Microbials, and Guilden Corporation. J.H. is a cofounder, board member of, and has equity in GenCirq Inc., which focuses on cancer therapeutics. J.D.E. is the founder of TEGA Therapeutics and a consultant for NeuImmune. J.D.E, T.M.C., and D.R.S. are consultants of Covicept Therapeutics, Inc. (unrelated to the present study). N.E.L. is a co-founder of Augment Biologics, Inc., and NeuImmune, Inc. and a consultant for Biogen and Regeneron. R.K. is a scientific advisory board member and consultant for BiomeSense, Inc., has equity, and receives income. He is a scientific advisory board member and has equity in GenCirq. He is a consultant for DayTwo and receives income. He has equity in and acts as a consultant for Cybele. He is a co-founder of Biota, Inc., and has equity. He is a cofounder of Micronoma, has equity, and is a scientific advisory board member. The terms of these arrangements have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies.

AUTHOR AFFILIATIONS

[1]Department of Pediatrics, University of California San Diego School of Medicine, La Jolla, California, USA
[2]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, California, USA

[3]Center for Microbiome Innovation, University of California San Diego, La Jolla, California, USA
[4]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California, USA
[5]Copenhagen Center for Glycomics, Department of Molecular and Cellular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[6]Department of Bioengineering, University of California San Diego, La Jolla, California, USA
[7]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, ChemSyBio, 78350, Jouy-en-Josas, France
[8]Sorbonne Université, Faculty of Sciences and Engineering, IBPS, UMR 8263 CNRS-SU, ERL INSERM U1345, Development, Adaptation and Ageing, F-75252 Paris, France
[9]Department of Medicine, University of California San Diego, La Jolla, California, USA
[10]Department of Biochemistry, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil
[11]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[12]Jacobs School of Engineering, University of California San Diego, La Jolla, California, USA
[13]Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, California, USA
[14]Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA
[15]Merck & Co., Inc., Rahway, NJ, 07065, USA
[16]Department of Biological & Medical Informatics, University of California San Francisco, School of Pharmacy, San Francisco, California, USA
[17]Department of Surgery, Division of Surgical Outcomes and Precision Medicine Research, Medical School, University of Minnesota, Minneapolis, Minnesota, USA
[18]National Institutes of Health, National Cancer Institute, Center for Cancer Research, Animal Models and Retroviral Vaccine Section, Bethesda, Maryland, USA
[19]Department of Infectious diseases, Luigi Sacco Hospital, Milan and Department of Biomedical and Clinical Sciences (DIBIC), University of Milan, Milan, Italy
[20]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki and Turku, Finland
[21]Institute for Molecular Medicine Finland, FIMM - HiLIFE, Helsinki, Finland
[22]Sano Centre for Computational Medicine, Krakow, Poland
[23]Laureate Institute for Brain Research, Tulsa, Oklahoma, USA
[24]Division of Health Analytics, Department of Computational and Quantitative Medicine, City of Hope, Duarte, California, USA
[25]UC Health Data Warehouse, University of California Irvine, Irvine, California, USA
[26]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, California, USA
[27]Faculty of Dentistry, The University of Hong Kong, Hong Kong, China
[28]IBM T. J. Watson Research Center, Yorktown Heights, New York, USA
[29]AI and Cognitive Software, IBM Research-Almaden, San Jose, California, USA
[30]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of California San Diego, La Jolla, California, USA
[31]International Biomedical Research Alliance, Bethesda, Maryland, USA
[32]Division of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts, USA
[33]Cedars-Sinai Medical Center, Los Angeles, California, USA
[34]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, United Kingdom
[35]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Australia

[36]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
[37]Division of Medicine, Turku University Hospital and University of Turku, Turku, Finland
[38]Department of Pharmacology, University of California, San Diego, La Jolla, California, USA
[39]Molecular Biology Section, Division of Biological Science, University of California San Diego, La Jolla, California, USA
[40]Glycobiology Research and Training Center, University of California San Diego, La Jolla, California, USA
[41]Center for Molecular Medicine, Complex Carbohydrate Research Center, and Dept of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA
[42]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark
[43]Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, USA

AUTHOR ORCIDs

Cameron Martino http://orcid.org/0000-0001-9334-1258
Rodolfo A. Salido http://orcid.org/0000-0003-0631-2817
Bryn Taylor http://orcid.org/0000-0002-4957-5417
Promi Das http://orcid.org/0000-0001-6733-2059
Mehrbod Estaki http://orcid.org/0000-0003-4556-8444
Niina Haiminen http://orcid.org/0000-0002-8663-1019
Austin D. Swafford http://orcid.org/0000-0001-5655-8300
Karsten Zengler http://orcid.org/0000-0002-8062-3296
Nichole R. Klatt http://orcid.org/0000-0003-2968-5480
Rob Knight http://orcid.org/0000-0002-0975-9019

FUNDING

Funder	Grant(s)	Author(s)
IBM | IBM Research	AI Horizons Network	Cameron Martino Qiyun Zhu Yoshiki Vázquez-Baeza Shi Huang Niina Haiminen Laxmi Parida Ho-Cheol Kim Rob Knight
HHS | National Institutes of Health (NIH)	P01 HL131474	Jeffrey D. Esko
Novo Nordisk Foundation	NNF20SA0066621	Nathan E. Lewis
National Science Foundation (NSF)	2031989	Jeffrey D. Esko
Alfred Benzon Foundation (The Alfred Benzon Foundation)		Thomas Mandel Clausen
Academy of Finland	321351 and 354447	Teemu Niiranen
Emil Aaltosen Säätiö (Emil Aaltonen Foundation)		Teemu Niiranen
Sydäntutkimussäätiö (Finnish Foundation for Cardiovascular Research)		Veikko Salomaa
HHS | National Institutes of Health (NIH)	R01ES027595	Mohit Jain
ANID Becas Chile Doctorado	2018-72190270	Erick Armingol

Funder	Grant(s)	Author(s)
Academy of Finland	321356	Aki S. Havulinna
UC San Diego Center for Microbiome Innovation	N/A	Se Jin Song Stephen Wandro Daniel McDonald Promi Das
Academy of Finland	335525	Aki S. Havulinna
Union's Horizon Europe Research and Innovation 459 Actions	101046041	Agostino Riva
French National Research Agency	ANR-17-CE11-0014	Olivier Berteau
French National Research Agency	ANR-20-CE44-0005	Olivier Berteau
HHS | National Institutes of Health (NIH)	R01GM069811	Jeff Hasty
Munz Chair of Cardiovascular Prediction and Prevention		Michael Inoyue
NIHR | NIHR Cambridge Biomedical Research Centre (NIHR Cambridge BRC)		Michael Inoyue
British Heart Foundation (BHF)	RG/13/13/30194	Michael Inoyue
British Heart Foundation (BHF)	RG/18/13/33946	Michael Inoyue
NIHR | NIHR Cambridge Biomedical Research Centre (NIHR Cambridge BRC)	BRC-1215-20014	Michael Inoyue
HHS | National Institutes of Health (NIH)	1DP1AT010885	Cameron Martino Clarisse A. Marotz Rodolfo A. Salido Qiyun Zhu Daniel McDonald Bryn Taylor Pedro Belda-Ferre Mehrbod Estaki Tomasz Kosciolek Rob Knight
Health Data Research UK (HDR UK)		Michael Inoyue
HHS | National Institutes of Health (NIH)	U01-CA199277	Kristen E. Savage Jennifer L. Benbow Emma S. Spielfogel Cheryl A. M. Anderson Maria Elena Martinez James V. Lacey Jr
HHS | National Institutes of Health (NIH)	P30-CA033572	Kristen E. Savage Jennifer L. Benbow Emma S. Spielfogel Cheryl A. M. Anderson Maria Elena Martinez James V. Lacey Jr
HHS | National Institutes of Health (NIH)	P30-CA023100	Kristen E. Savage Jennifer L. Benbow Emma S. Spielfogel Cheryl A. M. Anderson

Funder	Grant(s)	Author(s)
		Maria Elena Martinez
		James V. Lacey Jr
HHS | National Institutes of Health (NIH)	UM1-CA164917	Kristen E. Savage
		Jennifer L. Benbow
		Emma S. Spielfogel
		Cheryl A. M. Anderson
		Maria Elena Martinez
		James V. Lacey Jr
HHS | National Institutes of Health (NIH)	R01-CA077398	Kristen E. Savage
		Jennifer L. Benbow
		Emma S. Spielfogel
		Cheryl A. M. Anderson
		Maria Elena Martinez
		James V. Lacey Jr
National Science Foundation (NSF)	2038509	Daniel McDonald
		Rob Knight
UC | University of California, San Diego (UCSD)	R00RG2503	Daniel McDonald
		Rob Knight
Emerald Foundation	3022	Pedro Belda-Ferre
		Rob Knight
HHS | National Institutes of Health (NIH)	1P30DK120515	Nathan E. Lewis
HHS | National Institutes of Health (NIH)	R35GM119850	Nathan E. Lewis
HHS | National Institutes of Health (NIH)	UH2AI153029	Nathan E. Lewis

AUTHOR CONTRIBUTIONS

Cameron Martino, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing | Benjamin P. Kellman, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Daniel R. Sandoval, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Thomas Mandel Clausen, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Robert Cooper, Investigation, Writing – review and editing | Alhosna Benjdia, Investigation, Writing – review and editing | Feryel Soualmia, Investigation, Writing – review and editing | Alex E. Clark, Investigation, Writing – review and editing | Aaron F. Garretson, Investigation, Writing – review and editing | Clarisse A. Marotz, Writing – review and editing | Se Jin Song, Data curation, Formal analysis, Writing – review and editing | Stephen Wandro, Data curation, Formal analysis, Writing – review and editing | Livia S. Zaramela, Data curation, Formal analysis, Writing – review and editing | Rodolfo A. Salido, Investigation, Writing – review and editing | Qiyun Zhu, Data curation, Formal analysis, Writing – review and editing | Erick Armingol, Data curation, Formal analysis, Writing – review and editing | Yoshiki Vázquez-Baeza, Data curation, Formal analysis, Writing – review and editing | Daniel McDonald, Data curation, Formal analysis, Writing – review and editing | James T. Sorrentino, Writing – review and editing | Bryn Taylor, Writing – review and editing | Pedro Belda-Ferre, Methodology, Writing – review and editing | Promi Das, Data curation, Writing – review and editing | Farhana Ali, Data curation, Writing – review and editing | Chenguang Liang, Formal analysis, Writing – review and editing | Yujie Zhang, Formal analysis, Writing – review and editing | Luca Schifanella, Methodology, Writing – review and editing | Alice Covizzi, Data curation,

Writing – review and editing | Alessia Lai, Data curation, Writing – review and editing | Agostino Riva, Data curation, Writing – review and editing | Christopher Basting, Data curation, Writing – review and editing | Courtney Ann Broedlow, Data curation, Writing – review and editing | Aki S. Havulinna, Data curation, Writing – review and editing | Pekka Jousilahti, Data curation, Writing – review and editing | Mehrbod Estaki, Data curation, Writing – review and editing | Tomasz Kosciolek, Data curation, Writing – review and editing | Rayus Kuplicki, Data curation, Writing – review and editing | Teresa A. Victor, Data curation, Writing – review and editing | Martin P. Paulus, Data curation, Supervision, Writing – review and editing | Kristen E. Savage, Data curation, Writing – review and editing | Jennifer L. Benbow, Data curation, Writing – review and editing | Emma S. Spielfogel, Data curation, Writing – review and editing | Cheryl A. M. Anderson, Data curation, Supervision, Writing – review and editing | Maria Elena Martinez, Data curation, Supervision, Writing – review and editing | James V. Lacey Jr, Data curation, Supervision, Writing – review and editing | Shi Huang, Data curation, Writing – review and editing | Niina Haiminen, Resources, Writing – review and editing | Laxmi Parida, Resources, Writing – review and editing | Ho-Cheol Kim, Resources, Supervision, Writing – review and editing | Jack A. Gilbert, Supervision, Writing – review and editing | Daniel A. Sweeney, Data curation, Supervision, Writing – review and editing | Sarah M. Allard, Data curation, Writing – review and editing | Austin D. Swafford, Conceptualization, Supervision, Writing – original draft, Writing – review and editing | Susan Cheng, Data curation, Writing – review and editing | Michael Inouye, Data curation, Writing – review and editing | Teemu Niiranen, Data curation, Writing – review and editing | Mohit Jain, Data curation, Writing – review and editing | Veikko Salomaa, Data curation, Supervision, Writing – review and editing | Karsten Zengler, Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review and editing | Nichole R. Klatt, Resources, Writing – review and editing | Jeff Hasty, Resources, Writing – review and editing | Olivier Berteau, Conceptualization, Investigation, Methodology, Supervision, Writing – review and editing | Aaron F. Carlin, Resources, Writing – review and editing | Jeffrey D. Esko, Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing | Nathan E. Lewis, Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing | Rob Knight, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

DIRECT CONTRIBUTION

This article is a direct contribution from Rob Knight, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by Deepak Shukla, University of Illinois Chicago College of Applied Health Sciences, and Vaibhav Tiwari, Midwestern University - Downers Grove Campus.

DATA AVAILABILITY

All American Gut Project sequence data and de-identified participant responses can be found in EBI under project PRJEB11419 and Qiita (https://qiita.ucsd.edu/) study ID 10317. The COVID-19 patient data are available through EBI under accession ERP124721 associated feature tables are publicly available in Qiita under study ID 13092. The FINRISK data that support the findings of this study are available from the THL Biobank based on a written application and following relevant Finnish legislation. Details of the application process are described on the website of the Biobank: https://thl.fi/en/web/thl-biobank/for-researchers. All of the data associated with this publication and in the California Teachers Study are available for research use. The California Teachers Study welcomes all such inquiries and encourages individuals to visit https://www.calteachersstudy.org/for-researchers. The source code for the analyses can be found at https://doi.org/10.5281/zenodo.3973506. The following reagent was deposited by the Centers

for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, NR-52281.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental Material (mBio04015-24-s0001.pdf). Supplemental figures and tables.

REFERENCES

1. Han Y, Duan X, Yang L, Nilsson-Payant BE, Wang P, Duan F, Tang X, Yaron TM, Zhang T, Uhl S, et al. 2021. Identification of SARS-CoV-2 inhibitors using lung and colonic organoids. Nature New Biol 589:270–275. https://doi.org/10.1038/s41586-020-2901-9
2. Elshazli RM, Kline A, Elgaml A, Aboutaleb MH, Salim MM, Omar M, Munshi R, Mankowski N, Hussein MH, Attia AS, Toraih EA, Settin A, Killackey M, Fawzy MS, Kandil E. 2021. Gastroenterology manifestations and COVID-19 outcomes: a meta-analysis of 25,252 cohorts among the first and second waves. J Med Virol 93:2740–2768. https://doi.org/10.1002/jmv.26836
3. Vespa E, Pugliese N, Colapietro F, Aghemo A. 2021. Stay (GI) Healthy: COVID-19 and gastrointestinal manifestations. Tech Innov Gastrointest Endosc 23:179–189. https://doi.org/10.1016/j.tige.2021.01.006
4. Villapol S. 2020. Gastrointestinal symptoms associated with COVID-19: impact on the gut microbiome. Transl Res 226:57–69. https://doi.org/10.1016/j.trsl.2020.08.004
5. Cao W. 2020. What do we know so far about gastrointestinal and liver injuries induced by SARS-CoV-2 virus? Gastroenterol Res 13:225–226. https://doi.org/10.14740/gr1350
6. Hayashi Y, Wagatsuma K, Nojima M, Yamakawa T, Ichimiya T, Yokoyama Y, Kazama T, Hirayama D, Nakase H. 2021. The characteristics of gastrointestinal symptoms in patients with severe COVID-19: a systematic review and meta-analysis. J Gastroenterol 56:409–420. https://doi.org/10.1007/s00535-021-01778-z
7. Guo M, Tao W, Flavell RA, Zhu S. 2021. Potential intestinal infection and faecal-oral transmission of SARS-CoV-2. Nat Rev Gastroenterol Hepatol 18:269–283. https://doi.org/10.1038/s41575-021-00416-6
8. Chu H, Chan J-W, Wang Y, Yuen T-T, Chai Y, Shuai H, Yang D, Hu B, Huang X, Zhang X, Hou Y, Cai J-P, Zhang AJ, Zhou J, Yuan S, To K-W, Hung I-N, Cheung TT, Ng A-L, Hau-Yee Chan I, Wong I-H, Law S-K, Foo D-C, Leung W-K, Yuen K-Y. 2021. SARS-Cov-2 induces a more robust innate immune response and replicates less efficiently than SARS-CoV in the human intestines: an *ex vivo* study with implications on pathogenesis of COVID-19. Cell Mol Gastroenterol Hepatol 11:771–781. https://doi.org/10.1016/j.jcmgh.2020.09.017
9. Neurath MF, Überla K, Ng SC. 2021. Gut as viral reservoir: lessons from gut viromes, HIV and COVID-19. Gut 70:1605–1608. https://doi.org/10.1136/gutjnl-2021-324622
10. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, HLH Across Speciality Collaboration, UK. 2020. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet 395:1033–1034. https://doi.org/10.1016/S0140-6736(20)30628-0
11. Natarajan A, Zlitni S, Brooks EF, Vance SE, Dahlen A, Hedlin H, Park RM, Han A, Schmidtke DT, Verma R, Jacobson KB, Parsonnet J, Bonilla HF, Singh U, Pinsky BA, Andrews JR, Jagannathan P, Bhatt AS. 2022. Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection. Med 3:371–387. https://doi.org/10.1016/j.medj.2022.04.001
12. Woo MS, Shafiq M, Fitzek A, Dottermusch M, Altmeppen H, Mohammadi B, Mayer C, Bal LC, Raich L, Matschke J, Krasemann S, Pfefferle S, Brehm TT, Lütgehetmann M, Schädler J, Addo MM, Schulze Zur Wiesch J, Ondruschka B, Friese MA, Glatzel M. 2023. Vagus nerve inflammation contributes to dysautonomia in COVID-19. Acta Neuropathol 146:387–394. https://doi.org/10.1007/s00401-023-02612-x
13. Jin JC, Ananthanarayanan A, Brown JA, Rager SL, Bram Y, Sanidad KZ, Amir M, Baergen RN, Stuhlmann H, Schwartz RE, Perlman JM, Zeng MY. 2023. SARS CoV-2 detected in neonatal stool remote from maternal COVID-19 during pregnancy. Pediatr Res 93:1375–1382. https://doi.org/10.1038/s41390-022-02266-7
14. Cheung CCL, Goh D, Lim X, Tien TZ, Lim JCT, Lee JN, Tan B, Tay ZEA, Wan WY, Chen EX, Nerurkar SN, Loong S, Cheow PC, Chan CY, Koh YX, Tan TT, Kalimuddin S, Tai WMD, Ng JL, Low JG-H, Yeong J, Lim KH. 2022. Residual SARS-CoV-2 viral antigens detected in GI and hepatic tissues from five recovered patients with COVID-19. Gut 71:226–229. https://doi.org/10.1136/gutjnl-2021-324280
15. Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, Narayanan A, Majowicz SA, Kwong EM, McVicar RN, et al. 2020. SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. Cell 183:1043–1057. https://doi.org/10.1016/j.cell.2020.09.033
16. Kearns FL, Sandoval DR, Casalino L, Clausen TM, Rosenfeld MA, Spliid CB, Amaro RE, Esko JD. 2022. Spike-heparan sulfate interactions in SARS-CoV-2 infection. Curr Opin Struct Biol 76:102439. https://doi.org/10.1016/j.sbi.2022.102439
17. Zhang H, Kang Z, Gong H, Xu D, Wang J, Li Z, Li Z, Cui X, Xiao J, Zhan J, Meng T, Zhou W, Liu J, Xu H. 2020. Digestive system is a potential route of COVID-19: an analysis of single-cell coexpression pattern of key proteins in viral entry process. Gut 69:1010–1018. https://doi.org/10.1136/gutjnl-2020-320953
18. Esko JD, Selleck SB. 2002. Order out of chaos: assembly of ligand binding sites in heparan sulfate. Annu Rev Biochem 71:435–471. https://doi.org/10.1146/annurev.biochem.71.110601.135458
19. Cagno V, Tseligka ED, Jones ST, Tapparel C. 2019. Heparan sulfate proteoglycans and viral attachment: true receptors or adaptation bias? Viruses 11:596. https://doi.org/10.3390/v11070596
20. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, et al. 2018. American gut: an open platform for citizen science microbiome research. mSystems 3:e00031-18. https://doi.org/10.1128/mSystems.00031-18
21. Thaiss CA, Zmora N, Levy M, Elinav E. 2016. The microbiome and innate immunity. Nature New Biol 535:65–74. https://doi.org/10.1038/nature18847
22. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C. 2017. Erratum: strains, functions and dynamics in the expanded human microbiome project. Nature New Biol 551:256. https://doi.org/10.1038/nature24485
23. Johansson MEV, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC. 2008. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci U S A 105:15064–15069. https://doi.org/10.1073/pnas.0803124105
24. Linhardt RJ, Galliher PM, Cooney CL. 1986. Polysaccharide lyases. Appl Biochem Biotechnol 12:135–176. https://doi.org/10.1007/BF02798420
25. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide utilization loci: fueling microbial communities. J Bacteriol 199:e00860-16. https://doi.org/10.1128/JB.00860-16
26. Cartmell A, Lowe EC, Baslé A, Firbank SJ, Ndeh DA, Murray H, Terrapon N, Lombard V, Henrissat B, Turnbull JE, Czjzek M, Gilbert HJ, Bolam DN. 2017. How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. Proc Natl Acad Sci U S A 114:7037–7042. https://doi.org/10.1073/pnas.1704367114
27. Ndeh D, Baslé A, Strahl H, Yates EA, McClurgg UL, Henrissat B, Terrapon N, Cartmell A. 2020. Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. Nat Commun 11:646. https://doi.org/10.1038/s41467-020-14509-4
28. Ulmer JE, Vilén EM, Namburi RB, Benjdia A, Beneteau J, Malleron A, Bonnaffé D, Driguez P-A, Descroix K, Lassalle G, Le Narvor C, Sandström C, Spillmann D, Berteau O. 2014. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont bacteroides

thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. J Biol Chem 289:24289–24303. https://doi.org/10.1074/jbc.M114.573303
29. Galliher PM, Cooney CL, Langer R, Linhardt RJ. 1981. Heparinase production by *Flavobacterium heparinum*. Appl Environ Microbiol 41:360–365. https://doi.org/10.1128/aem.41.2.360-365.1981
30. Weiss RJ, Esko JD, Tor Y. 2017. Targeting heparin and heparan sulfate protein interactions. Org Biomol Chem 15:5656–5668. https://doi.org/10.1039/C7OB01058C
31. Earle KA, Billings G, Sigal M, Lichtman JS, Hansson GC, Elias JE, Amieva MR, Huang KC, Sonnenburg JL. 2015. Quantitative imaging of gut microbiota spatial organization. Cell Host Microbe 18:478–488. https://doi.org/10.1016/j.chom.2015.09.002
32. Richelle A, Kellman BP, Wenzel AT, Chiang AWT, Reagan T, Gutierrez JM, Joshi C, Li S, Liu JK, Masson H, Lee J, Li Z, Heirendt L, Trefois C, Juarez EF, Bath T, Borland D, Mesirov JP, Robasky K, Lewis NE. 2021. Model-based assessment of mammalian cell metabolic functionalities using omics data. Cell Rep Methods 1:100040. https://doi.org/10.1016/j.crmeth.2021.100040
33. Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, Kuulasmaa K, Laatikainen T, Männistö S, Peltonen M, Perola M, Puska P, Salomaa V, Sundvall J, Virtanen SM, Vartiainen E. 2018. Cohort profile: the national FINRISK study. Int J Epidemiol 47:696–696i. https://doi.org/10.1093/ije/dyx239
34. Salosensaari A, Laitinen V, Havulinna AS, Meric G, Cheng S, Perola M, Valsta L, Alfthan G, Inouye M, Watrous JD, Long T, Salido R, Sanders K, Brennan C, Humphrey GC, Sanders JG, Jain M, Jousilahti P, Salomaa V, Knight R, Lahti L, Niiranen T. 2020. Taxonomic signatures of long-term mortality risk in human gut microbiota. Epidemiology (Sunnyvale). https://doi.org/10.1101/2019.12.30.19015842
35. Redd WD, Zhou JC, Hathorn KE, McCarty TR, Bazarbashi AN, Thompson CC, Shen L, Chan WW. 2020. Prevalence and characteristics of gastrointestinal symptoms in patients with severe acute respiratory syndrome coronavirus 2 infection in the united states: a multicenter cohort study. Gastroenterology 159:765–767. https://doi.org/10.1053/j.gastro.2020.04.045
36. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, et al. 2020. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature New Biol
37. Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, Belda-Ferre P, Al-Ghalith GA, Kopylova E, McDonald D, et al. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. Nat Commun 10:5477. https://doi.org/10.1038/s41467-019-13443-4
38. Zaramela LS, Martino C, Alisson-Silva F, Rees SD, Diaz SL, Chuzel L, Ganatra MB, Taron CH, Secrest P, Zuñiga C, Huang J, Siegel D, Chang G, Varki A, Zengler K. 2019. Gut bacteria responding to dietary change encode sialidases that exhibit preference for red meat-associated carbohydrates. Nat Microbiol 4:2082–2089. https://doi.org/10.1038/s41564-019-0564-9
39. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, Pudlo NA, Kitamoto S, Terrapon N, Muller A, Young VB, Henrissat B, Wilmes P, Stappenbeck TS, Núñez G, Martens EC. 2016. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell 167:1339–1353. https://doi.org/10.1016/j.cell.2016.10.043
40. de la Cuesta-Zuluaga J, Kelley ST, Chen Y, Escobar JS, Mueller NT, Ley RE, McDonald D, Huang S, Swafford AD, Knight R, Thackray VG. 2019. Age- and sex-dependent patterns of gut microbial diversity in human adults. mSystems 4:e00261-19. https://doi.org/10.1128/mSystems.00261-19
41. Marotz C, Belda-Ferre P, Ali F, Das P, Huang S, Cantrell K, Jiang L, Martino C, Diner RE, Rahman G, et al. 2021. SARS-CoV-2 detection status associates with bacterial community composition in patients and the hospital environment. Microbiome 9:132. https://doi.org/10.1186/s40168-021-01083-0
42. Benjdia A, Martens EC, Gordon JI, Berteau O. 2011. Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, *Bacteroides thetaiotaomicron*. J Biol Chem 286:25973–25982. https://doi.org/10.1074/jbc.M111.228841
43. Hedemann MS, Theil PK, Bach Knudsen KE. 2009. The thickness of the intestinal mucous layer in the colon of rats fed various sources of non-digestible carbohydrates is positively correlated with the pool of SCFA but negatively correlated with the proportion of butyric acid in digesta. Br J Nutr 102:117–125. https://doi.org/10.1017/S0007114508143549
44. Brownlee IA, Havler ME, Dettmar PW, Allen A, Pearson JP. 2003. Colonic mucus: secretion and turnover in relation to dietary fibre intake. Proc Nutr Soc 62:245–249. https://doi.org/10.1079/pns2003206
45. Tadbir Vajargah K, Zargarzadeh N, Ebrahimzadeh A, Mousavi SM, Mobasheran P, Mokhtari P, Rahban H, Găman M-A, Akhgarjand C, Taghizadeh M, Milajerdi A. 2022. Association of fruits, vegetables, and fiber intake with COVID-19 severity and symptoms in hospitalized patients: a cross-sectional study. Front Nutr 9:934568. https://doi.org/10.3389/fnut.2022.934568
46. Zmora N, Suez J, Elinav E. 2019. You are what you eat: diet, health and the gut microbiota. Nat Rev Gastroenterol Hepatol 16:35–56. https://doi.org/10.1038/s41575-018-0061-2
47. Martino C, Dilmore AH, Burcham ZM, Metcalf JL, Jeste D, Knight R. 2022. Microbiota succession throughout life from the cradle to the grave. Nat Rev Microbiol 20:707–720. https://doi.org/10.1038/s41579-022-00768-z
48. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature New Biol 486:207–214. https://doi.org/10.1038/nature11234
49. Beck JM, Schloss PD, Venkataraman A, Twigg H 3rd, Jablonski KA, Bushman FD, Campbell TB, Charlson ES, Collman RG, Crothers K, et al. 2015. Multicenter comparison of lung and oral microbiomes of HIV-infected and HIV-uninfected Individuals. Am J Respir Crit Care Med 192:1335–1344. https://doi.org/10.1164/rccm.201501-0128OC
50. Ali F, Sweeney DA. 2020. In pursuit of microbiome-based therapies for acute respiratory failure. Am J Respir Crit Care Med 202:1616–1618. https://doi.org/10.1164/rccm.202008-3146ED
51. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM. 2017. Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol 15:55–63. https://doi.org/10.1038/nrmicro.2016.142
52. Reed JL, Famili I, Thiele I, Palsson BO. 2006. Towards multidimensional genome annotation. Nat Rev Genet 7:130–141. https://doi.org/10.1038/nrg1769
53. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 35:81–89. https://doi.org/10.1038/nbt.3703
54. Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res 46:7542–7553. https://doi.org/10.1093/nar/gky537
55. Eilam O, Zarecki R, Oberhardt M, Ursell LK, Kupiec M, Knight R, Gophna U, Ruppin E. 2014. Glycan degradation (GlyDeR) analysis predicts mammalian gut microbiota abundance and host diet-specific adaptations. MBio 5:e01526-14. https://doi.org/10.1128/mBio.01526-14
56. Koropatkin NM, Cameron EA, Martens EC. 2012. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol 10:323–335. https://doi.org/10.1038/nrmicro2746
57. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092
58. Hobbs JK, Pluvinage B, Boraston AB. 2018. Glycan-metabolizing enzymes in microbe-host interactions: the *Streptococcus pneumoniae* paradigm. FEBS Lett 592:3865–3897. https://doi.org/10.1002/1873-3468.13045
59. Robb M, Hobbs JK, Woodiga SA, Shapiro-Ward S, Suits MDL, McGregor N, Brumer H, Yesilkaya H, King SJ, Boraston AB. 2017. Molecular characterization of N-glycan degradation and transport in *Streptococcus pneumoniae* and its contribution to virulence. PLoS Pathog 13:e1006090. https://doi.org/10.1371/journal.ppat.1006090
60. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–5. https://doi.org/10.1093/nar/gkt1178
61. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–51. https://doi.org/10.1093/nar/gks479
62. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418

63. Richelle A, Joshi C, Lewis NE. 2019. Assessing key decisions for transcriptomic data integration in biochemical networks. PLoS Comput Biol 15:e1007185. https://doi.org/10.1371/journal.pcbi.1007185
64. Richelle A, Chiang AWT, Kuo C-C, Lewis NE. 2019. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. PLoS Comput Biol 15:e1006867. https://doi.org/10.1371/journal.pcbi.1006867
65. Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ 5:e3720. https://doi.org/10.7717/peerj.3720
66. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
67. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Gohl DM, Beckman KB, Knight R, Knights D. 2018. Evaluating the information content of shallow shotgun metagenomics. mSystems 3:e00069-18. https://doi.org/10.1128/mSystems.00069-18
68. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65. https://doi.org/10.1093/nar/gkl842
69. Johansson R. 2018. *Numerical python: scientific computing and data science applications with Numpy, SciPy and Matplotlib*. Apress.
70. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160
71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
72. Sanders JG, Nurk S, Salido RA, Minich J, Xu ZZ, Zhu Q, Martino C, Fedarko M, Arthur TD, Chen F, et al. 2019. Optimizing sequencing protocols for leaderboard metagenomics by combining long and short reads. Genome Biol 20:226. https://doi.org/10.1186/s13059-019-1834-9
73. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
74. Langdon WB. 2015. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min 8:1. https://doi.org/10.1186/s13040-014-0034-0
75. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. https://doi.org/10.1093/bioinformatics/btr507
76. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021
77. Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760
78. Benjdia A, Leprince J, Sandström C, Vaudry H, Berteau O. 2009. Mechanistic investigations of anaerobic sulfatase-maturating enzyme: direct Cbeta H-atom abstraction catalyzed by a radical AdoMet enzyme. J Am Chem Soc 131:8348–8349. https://doi.org/10.1021/ja901571p