Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide

Matthew T. Henke[a], Douglas J. Kenny[b], Chelsi D. Cassilly[a], Hera Vlamakis[b], Ramnik J. Xavier[b,c,d], and Jon Clardy[a,1]

[a]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston MA 02115; [b]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [c]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; and [d]Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, MA 02114

Edited by Ralph R. Isberg, Tufts University School of Medicine, Boston, MA, and approved May 16, 2019 (received for review March 9, 2019)

A substantial and increasing number of human diseases are associated with changes in the gut microbiota, and discovering the molecules and mechanisms underlying these associations represents a major research goal. Multiple studies associate *Ruminococcus gnavus*, a prevalent gut microbe, with Crohn's disease, a major type of inflammatory bowel disease. We have found that *R. gnavus* synthesizes and secretes a complex glucorhamnan polysaccharide with a rhamnose backbone and glucose sidechains. Chemical and spectroscopic studies indicated that the glucorhamnan was largely a repeating unit of five sugars with a linear backbone formed from three rhamnose units and a short sidechain composed of two glucose units. The rhamnose backbone is made from 1,2- and 1,3-linked rhamnose units, and the sidechain has a terminal glucose linked to a 1,6-glucose. This glucorhamnan potently induces inflammatory cytokine (TNFα) secretion by dendritic cells, and TNFα secretion is dependent on toll-like receptor 4 (TLR4). We also identify a putative biosynthetic gene cluster for this molecule, which has the four biosynthetic genes needed to convert glucose to rhamnose and the five glycosyl transferases needed to build the repeating pentasaccharide unit of the inflammatory glucorhamnan.

inflammatory bowel disease | microbiome | polysaccharide

Several lines of evidence show an association between changes in the human gut microbiome and disease states in the host, but few studies have probed the molecules and mechanisms underlying these associations. A particularly robust association exists between Crohn's disease (CD) and *Ruminococcus gnavus*, an anaerobic, Gram-positive member of the human gut microbiome. Several studies link an increased relative abundance of *R. gnavus* to increased symptoms of CD, a major form of inflammatory bowel disease (1–5). Patients with CD experience abdominal pain, diarrhea, and bloody stool. Severe CD cases may even involve extraintestinal manifestations of inflammation of the eyes, skin, and spine (6). CD is currently incurable and symptoms can only be managed, often through treatment with antibodies against the major inflammatory cytokine, tumor necrosis factor alpha (TNFα).

The cause of CD is likely a combination of genetic predisposition and environmental triggers that shape the microbiome, such as diet, antibiotic exposure, and infections. Mutations found in genes involved in the innate immune system (*nod2*) and autophagy (*atg16l1*) are among those most associated with susceptibility to Crohn's disease (7). These mutations are thought to impair proper immune recruitment and response to commensal gut microbes by epithelial and dendritic cells, an innate immune cell type found in circulation and in the gut.

Incidence of CD is increasing worldwide (8), and has been linked to adoption of a Western diet (9). This strongly suggests that environmental factors are responsible for development of disease in genetically susceptible individuals. Chief among these environmental factors is the microbial dysbiosis often seen in CD patients. Extensive microbiome sequencing studies have shown that the relative abundance of *R. gnavus* is consistently increased in CD patients compared with healthy individuals (1–5). *R. gnavus* colonizes the intestinal mucosal surface, where it can use sialic acid from mucin glycans as a carbon source (5, 10). *R. gnavus* is a prevalent gut microbe found in the intestines of nearly 90% of people. In a healthy gut *R. gnavus* typically represents <0.1% of the gut microbiota. However, some CD patients experience transient blooms of *R. gnavus* levels that coincide with disease flares. In some severe cases, *R. gnavus* levels peak at 69% of the gut microbiota (3). Additionally, *R. gnavus* blooms in CD are transient and return to healthy levels during remission of disease symptoms. Earlier misclassification as a Ruminococceae (now a Lachnospiraceae) of *R. gnavus* made identification as biomarker for CD difficult. This was further complicated by the overall decreases of both of these clades in CD, such that blooms of *R. gnavus* were not observed because of this general trend at a genus and family level. As the microbiome field has shifted to analyzing more metagenomic data, which allows more robust species-level taxonomic profiling than 16S rRNA amplicon sequencing, the association of *R. gnavus* with CD has become more clear.

The effects of increases in *R. gnavus* level have also been linked to other inflammatory diseases including spondyloarthritis

Significance

The bacteria that live within the human gut play crucial roles in regulating our primary metabolism, protecting us from pathogens, and developing our immune system. Imbalances in bacterial community structure have been implicated in many diseases, such as Crohn's disease, an inflammatory bowel disease. We found and characterized an inflammatory polysaccharide produced by the gut bacterium *Ruminococcus gnavus*, populations of which bloom during flares of symptoms in patients with Crohn's disease. This molecule induces the production of inflammatory cytokines like TNFα by dendritic cells and may contribute to the association between *R. gnavus* and Crohn's disease. This work establishes a plausible molecular mechanism that may explain the association between a member of the gut microbiome and an inflammatory disease.

Author contributions: M.T.H., D.J.K., H.V., R.J.X., and J.C. designed research; M.T.H., D.J.K., and C.D.C. performed research; M.T.H., D.J.K., and C.D.C. analyzed data; and M.T.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jon_clardy@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1904099116/-/DCSupplemental.

Published online June 10, 2019.

(11), itself a frequent symptom of CD (6); eczema in infants (12); and pouchitis in ulcerative colitis patients who have undergone a total colectomy (13). However, despite numerous studies associating *R. gnavus* with various inflammatory diseases, no molecular mechanisms to explain these correlations have been developed, and we began a study to identify a possible mechanism.

The disease associations described above suggest that *R. gnavus* produces molecules that directly induce an inflammatory response by innate immune cells. Through bioactivity-guided fractionation, we found that *R. gnavus* produces a potent, inflammatory polysaccharide with a rhamnose backbone and short glucose sidechains—a glucorhamnan. Culturing in defined medium was necessary to purify the bacterially produced polysaccharide away from the overwhelming amounts of polysaccharides present in many commercial media. We further show that the *R. gnavus* glucorhamnan is recognized by innate immune cells through the toll-like receptor 4 (TLR4). Finally, we identified a biosynthetic gene cluster that is likely responsible for glucorhamnan production in the genomes of *R. gnavus*.

Results

***R. gnavus* Produces an Inflammatory Polysaccharide.** Initially *R. gnavus* was grown in rich media (Brain-Heart Infusion or Todd-Hewitt from Bacto) that are typically used to culture gut anaerobes. However, this led to confounding media-derived materials, and prompted us to culture *R. gnavus* in defined, albeit complex, media. The defined media used were based on a recipe that was recently described (14). Dendritic cells differentiated from mouse bone marrow (mBMDCs) were used in the primary screen to identify inflammatory substances from *R. gnavus*; lipopolysaccharide (LPS), a Gram-negative inducer of the inflammatory response, was used as a positive control. Following treatment of mBMDCs with fractions generated from *R. gnavus*, levels of the inflammatory cytokine TNFα were measured by enzyme-linked immunosorbent assay (ELISA) (Fig. 1*A*).

This assay indicated a dose-dependent potent TNFα-inducing activity in the spent medium, but this activity could not be extracted with organic solvents from either the spent medium or the cell pellet. Size-exclusion filters showed that activity was retained above a 10-kDa molecular-weight cutoff. Collectively these results indicated that the TNFα-inducing factor was a hydrophilic, large-molecular weight, bacterially produced molecule. Further treatment of the spent medium with proteinases, nucleases, lysozyme, or boiling did not abolish activity, indicating that the molecule was a polysaccharide (Fig. 1*B*). The polysaccharide was purified via enzymatic digestion of contaminating biopolymers, followed by size-exchange and anion-exchange chromatographies, while following the TNFα-inducing activity at each step.

***R. gnavus* Polysaccharide Is a Glucorhamnan.** To determine which monosaccharides are present in the purified polysaccharide, gas chromatography–mass spectrometry (GC-MS) monitoring of chemically modified acid-hydrolyzed monosaccharides was performed, which indicated roughly 1:4 glucose to rhamnose (the 6-deoxy form of mannose) (*SI Appendix*, Table S1). Since glucose is particularly susceptible to chemical degradation during chemical derivation steps, this ratio likely represents a lower limit to the amount of glucose present.

Determining how the monosaccharides are linked to each other was performed by generating partially methylated acetylated alditols (15). Linkage analysis showed five major types of monosaccharides to be present: 1-linked (terminally-linked) glucose, 1,6-linked glucose, 1,3-linked rhamnose, 1,2-linked rhamnose, and 1,2,3-linked rhamnose. These were present in roughly 1:1:1:1:1 stoichiometry (*SI Appendix*, Table S2). NMR studies were consistent with the linkage data as the ^{1}H-NMR spectrum (*SI Appendix*, Fig. S1) showed five distinct anomeric signals (H1

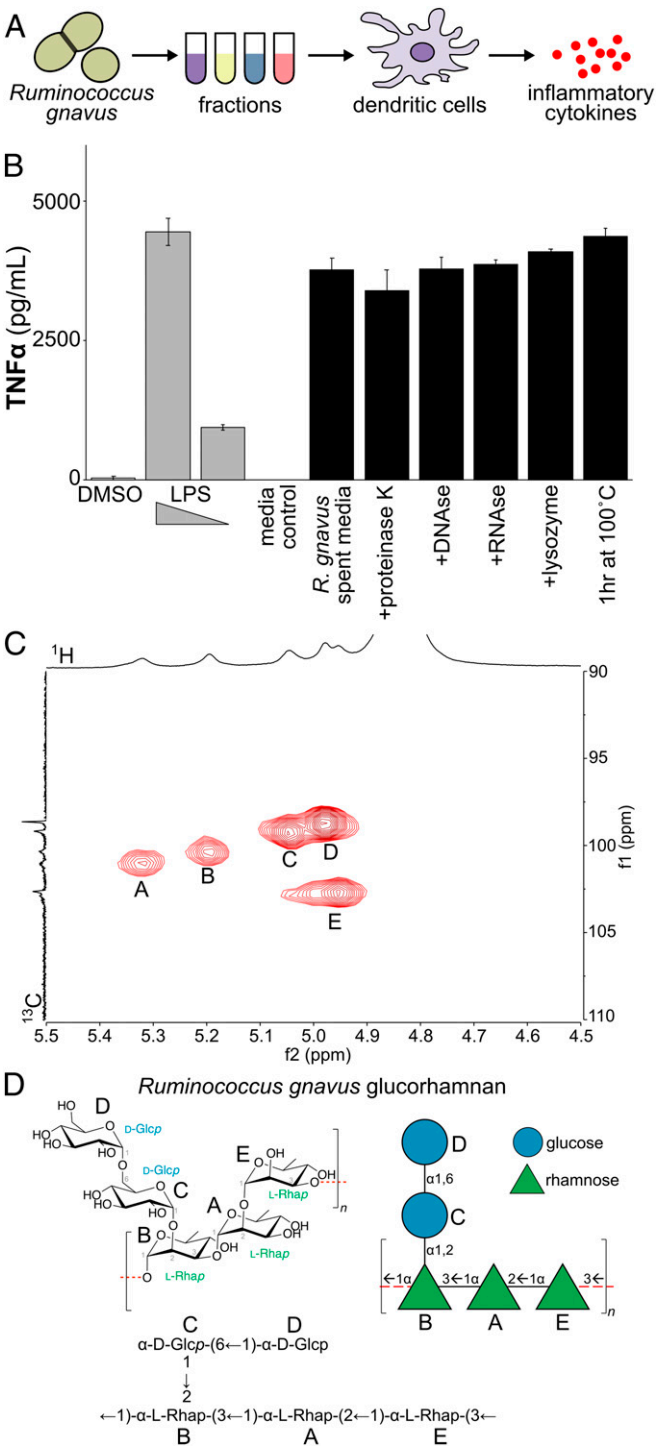

Fig. 1. *R. gnavus* produces an inflammatory glucorhamnan polysaccharide. (*A*) Schematic for generating fractions from *R. gnavus* in culture were screened for the ability to induce the inflammatory cytokine TNFα in mBMDCs. (*B*) TNFα was secreted by mBMDCs that were treated with *R. gnavus* spent medium. This activity was resistant to proteinase-, nuclease-, lysozyme-, and heat treatment, indicating it was a polysaccharide. LPS was a positive control, and DMSO and media alone were negative controls. (*C*) HSQC spectrum zoom-in of anomeric region of purified polysaccharide shows the repeating unit is composed of five distinct monosaccharide types (labeled A through E). (*D*) The purified polysaccharide is a glucorhamnan, with a main backbone of rhamnose residues and short sidechains of glucose residues. The structure of the repeating unit is shown following three separate conventions, overlaid with the corresponding monosaccharide designation A through E.

of a monosaccharide) between 4.9 and 5.4 ppm (Fig. 1*C*). We next turned to ordering the monosaccharides using the anomeric proton chemical shifts.

We assigned 5.32 ppm to H1 of the **A** monosaccharide, 5.20 ppm to H1 of **B**, 5.05 ppm as H1 to **C**, 4.98 ppm to H1 of **D**, and 4.95 ppm as H1 of **E**. The identity of a monosaccharide can often be deduced from the total correlation spectroscopy (TOCSY) cross-peaks from the anomeric proton (16). Since there is only one TOCSY cross-peak from H1 to H2 for **A**, **B**, and **E**, these must be rhamnose. The bond angle between H2 and H3 of rhamnose limits coupling between these protons, thus cross-peaks between H1 and H3 to H6 of rhamnose residues are very faint if present at all. Since all protons of glucose residues are axial, coupling along the entire glucose backbone is often observed in TOCSY. Since H1 of monosaccharide systems **C** and **D** have at least four TOCSY cross-peaks, **C** and **D** are glucose residues.

The H2 protons for monosaccharides **A** through **E** could be assigned by COSY. Carbon chemical shifts for C2 of both the **A** and **B** rhamnose indicated these positions were substituted, while carbon chemical shift at C2 of **E** indicated no substitution, and a faint H2BC signal from **E** H2 to C3 at 79 ppm, **E** was substituted at the three position. Thus, **E** was assigned as 1,3-rhamnose. NOESY showed cross-peaks for H1 of **E** to H2 of **A** and for H1 of **B** to H3 of **E**. Additionally, since H2 of **B** showed a NOESY cross-peak with H1 of **C**, **B** could be assigned to 1,2,3-rhamnose, and **A** could be assigned to 1,3-rhamnose. NOESY showed a cross-peak between H1 of **A** and a methine at 4.03 ppm. Since linkage data showed glucose residues are only attached at the methylene at C6, this 4.03-ppm peak must be H3 of **B**, although no direct correlations could connect H2 and H3 of **B**.

Finally, the assignment of the two glucose residues was largely determined by process of elimination. HSQC-TOCSY showed H1 of **D** in the same spin system as a methylene at 61 ppm, indicating that **D** is unlinked at the C6, and therefore, the terminally linked glucose. Thus, **C** must be the 1,6-linked glucose residue. NOESY from H1 of **D** shows it to be linked to a proton at 3.71 ppm. The only proton at 3.71 that has a carbon shift consistent with being linked is a methylene at 65.64 ppm, whose other proton has a shift at 4.12 ppm. Unfortunately, there is not a clear NOESY peak between H1 of **D** at the other methylene proton at 4.12 ppm. Although there are no data directly tying this carbon signal to the **C** glucose system, by process of elimination, it is the only position shown in linkage analysis not otherwise explained. Key NMR spectral observations are summarized in *SI Appendix*, Table S3. Full NMR spectra and peaks are provided in *SI Appendix*, Figs. S1–S9 and Dataset S1. In summary, the polysaccharide from *R. gnavus* is a glucorhamnan such that the backbone contains 1,3- and 1,2-linked rhamnose residues, with short sidechains of two glucose residues (Fig. 1*D*).

Based on ^{31}P and ^{1}H-^{31}P HMBC NMR experiments, some level of poly-1,3-phosphoglycerol (the canonical teichoic acid) is attached to the glucorhamnan. The exact connectivity between the phosphoglycerol and the glucorhamnan has not been elucidated, as no NOESY or ^{1}H-^{31}P HMBC correlations between the glucorhamnan and the phosphorglycerol portions were observed. Upon mild hydrolysis with hydrofluoric acid (HF), which hydrolyzes phosphate bonds, but not glycosidic, ester, or peptide bonds, the corresponding phosphoglycerol signals are lost, but the repeating unit remains intact (*SI Appendix*, Fig. S10), with exactly the correct number of methyl, methane, and methylene NMR signals expected from the proposed repeating unit. Analytical size-exclusion chromatography indicates the mass of the intact glucorhamnan is roughly 9 kDa, while the HF-treated glucorhamnan is roughly 8 kDa. Thus, the phosphoglycerol is unlikely to be attached along the main backbone of the glucorhamnan, but rather as extensions or sidechains thereto. The phosphoglycerol is also unlikely to be contaminating teichoic acid, as this latter molecule can be separated from the glucorhamnan by anion exchange as a distinct fraction.

Dendritic Cells Recognize Inflammatory Glucorhamnan Through Toll-Like Receptor 4. Having purified the active polysaccharide, we showed that the *R. gnavus* glucorhamnan induced secretion of TNFα in a dose-dependent manner, and that the *R. gnavus* glucorhamnan was as potent as the well-characterized yeast mannan, a potent inflammatory polysaccharide also recognized by the innate immune system (Fig. 2*A*). We also show that this response was dependent on the toll-like receptor 4 (TLR4), as mBMDCs generated from *tlr4*-knockout mice failed to respond to the glucorhamnan. As a control, mBMDCs were generated from *tlr2*-knockout mice [toll-like receptor 2 (TLR2) recognizes pathogen-associated molecular patterns with lipid tails] still secreted TNFα in response to treatment with *R. gnavus* glucorhamnan.

Identification of the Putative Glucorhamnan Biosynthetic Gene Cluster. Since linking the glucorhamnan to its biosynthetic gene cluster would enable both the mining of microbiota sequence data to identify other bacterial producers and of patient transcriptomic data to assess disease association we examined *R. gnavus* genomes. In the genome of the *R. gnavus* type strain ATCC 29149 (NZ_AAYG02000000) there are three gene clusters annotated to make polysaccharides (17). Only one of them (contig NZ_AAYG02000032) from RUMGNA_03512 to RUMGNA_03534 is likely to encode the biosynthetic pathway for the glucorhamnan (17). This cluster contains all four genes necessary for rhamnose biosynthesis as well as five glycosyltransferases

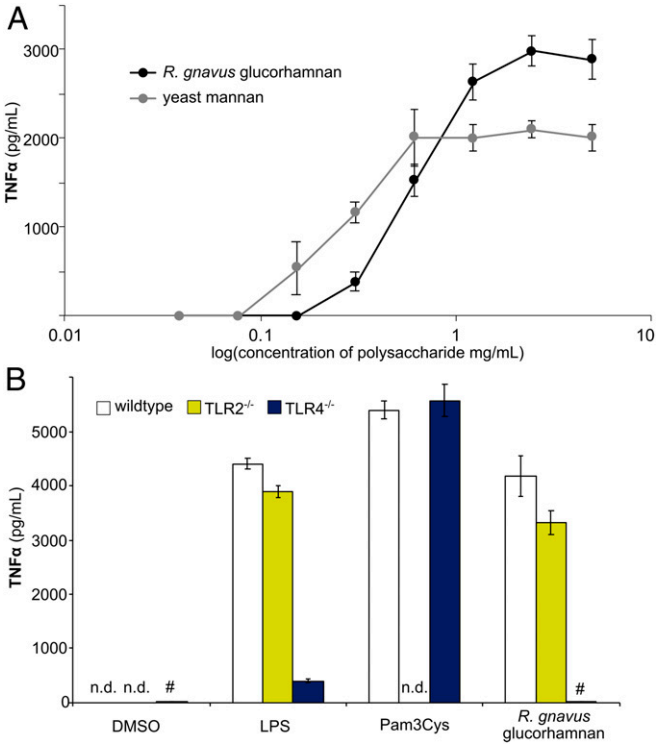

Fig. 2. *R. gnavus* glucorhamnan signals through TLR4 in a dose-dependent fashion in innate immune cells. (*A*) *R. gnavus* glucorhamnan stimulates mBMDCs to produce TNFα in dose-dependent manner roughly as potently as yeast mannan, a well-characterized inflammatory polysaccharide [error bars = SD of technical replicates (n = 4)]. (*B*) Secretion of TNFα by mBMDCs is lost in mice lacking TLR4 but not TLR2 (LPS) is a TLR4 ligand control; Pam3Cys is a synthetic TLR2 ligand control), n.d. = not detected; # indicates levels below 25 pg/mL detected.

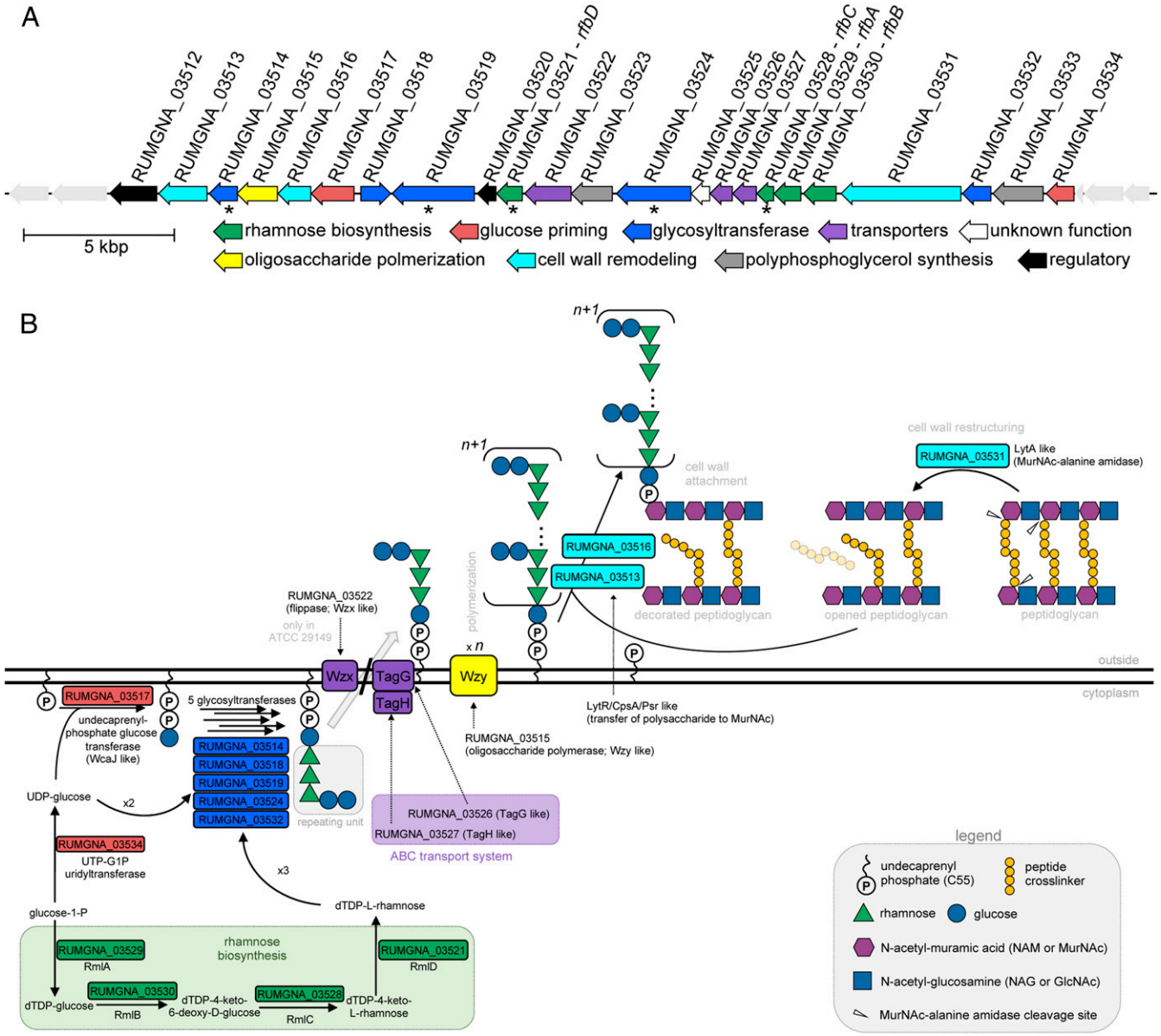

Fig. 3. Proposed biosynthetic pathway for *R. gnavus* glucorhamnan (*A*) A gene cluster (RUMGNA_03512 to RUMGNA_03534) encoding five glycosyltransferases and the biosynthetic enzyme for rhamnose is likely responsible for glucorhamnan synthesis. Five genes (designated by *) were transcribed during culture conditions (*SI Appendix*, Fig. S11). (*B*) A proposal for the biosynthesis of the glucorhamnan. Sequentially, monosaccharide precursors are synthesized, assembled into the repeating unit, transported across the membrane, where they are polymerized to form the full-length glucorhamnan, which is then covalently attached to the peptidoglycan.

(Fig. 3*A* and *SI Appendix*, Dataset S2). This number of glycosyltransferases is consistent with the number of different monosaccharides seen in the glucorhamnan repeating unit, and the majority of genes in this cluster. Of the 23 genes in the potential gene cluster, 18 have been assigned to likely functions in the biosynthetic pathway for the glucorhamnan (Fig. 3*B*). The biosynthesis can largely be summarized as: (*i*) monosaccharide synthesis and priming, (*ii*) assembly by glycosyltransferases of the pentasaccharide repeat unit on an undecaprenylphosphate lipid anchor, (*iii*) flipping of the assembled pentasaccharide repeat unit to the extracellular leaflet of the membrane, (*iv*) polymerization of the pentasaccharide repeat units to form a polysaccharide, (*v*) which is then transferred to the peptidoglycan.

While the glucorhamnan was initially isolated from the cell-free supernatant, two genes in the potential gene cluster encode LytR/CpsA/Psr-like proteins (RUMGNA_03513 and RUMGNA_03516). LytR/CpsA/Psr-like proteins catalyze the transfer of polysaccharides from undecaprenylphosphate lipid anchors to *N*-acetyl–muramic acid of peptidoglycan. While trace levels of the glucorhamnan could be recovered from lipid anchored cell-associated polysaccharides, treatment of the cell pellet of *R. gnavus* with mutanolysin (a lysozyme-like enzyme), liberated significantly greater amounts of the glucorhamnan into the supernatant. This suggests that the majority of the glucorhamnan is actually a cell-wall polysaccharide, covalently linked to the peptidoglycan. The exact nature of this linkage has not yet been determined.

The five remaining genes in the gene cluster have not been assigned to a specific role in the biosynthesis of the glucorhamnan. Of these, two are likely to be involved in regulatory functions, as the

only detectable domains were a germane domain (implicated in sporulation) (RUMGNA_03520), and a carbohydrate binding domain (RUMGNA_03512). Two genes are homologs of enzymes involved in cell-wall teichoic acid assembly [LtaA (RUMGNA_03523) and LtaS (RUMGNA_03533)]. While their role in glucorhamnan synthesis is unknown, their presence, combined with the presence of phosphoglycerol NMR signals in the purified glucorhamnan, suggest that there is some potential to transfer phosphoglycerol units onto the glucorhamnan repeating unit core. And finally, one gene (RUMGNA_03525) has no known homologs, and could not be annotated with a potential function.

As genetic manipulation of *R. gnavus* is not yet possible, quantitative PCR was performed on a cDNA library generated from *R. gnavus* to assess if the gene cluster is expressed. Three of the glycosyltransferases (RUMGNA_03514, RUMGNA_03519, and RUMGNA_03524), and two rhamnose biosynthesis genes [RUMGNA_03528 (rfbC analog) and RUMGNA_03521 (rfbD analog)] were shown to be expressed during growth of *R. gnavus* (*SI Appendix*, Fig. S11). These genes were no longer expressed when cells reached stationary phase, consistent with the observation that no additional glucorhamnan appears to be generated in culture after maximum OD_{600} is reached (~1.4 after 24 h).

Further suggesting that this gene cluster is the likely candidate, the other two polysaccharide biosynthetic gene clusters in the *R. gnavus* ATCC 29149 genome contain more glycosyltransferase genes than are required for the biosynthesis of a pentasaccharide repeat unit. The polysaccharide gene cluster from RUMGNA_01817 to RUMGNA_01849 contains seven glycosyltransferase genes as well as several genes involved in the biosynthesis of sialic acid, which is not present in the glucorhamnan. And, the polysaccharide gene cluster from RUMGNA_02393 to RUMGNA_02407 contains nine glycosyltransferase genes. In addition, only the proposed glucorhamnan gene cluster is conserved among many clinical isolates of *R. gnavus* (*SI Appendix*, Fig. S12), many of which produce the glucorhamnan in culture.

Discussion

We have identified and described a potent, TLR4-dependent, inflammatory glucorhamnan polysaccharide made by *R. gnavus*, a gut microbe associated with Crohn's disease and a variety of other inflammatory diseases. Initial culturing of *R. gnavus* was done in rich commercial media. Unfortunately, the TNFα-inducing polysaccharide produced by *R. gnavus* could not be purified away from contaminating yeast polysaccharides from the media. As we wanted to focus on inflammatory signals produced de novo by bacteria, not media components that might be modified by bacteria, a defined medium (based on ref. 14) was developed for the purification of the TNFα-inducing glucorhamnan that would be produced in the gut (*SI Appendix*, Dataset S3).

Bacteria produce a variety of cell-associated polysaccharides that play roles in cell morphology, cell-wall rigidity, evasion of the host immune system, and preventing dessication. Polysaccharides as mediators of host-microbe communication is well documented in polysaccharide A from *Bacteroides fragilis* (18, 19), but polysaccharide-mediated interactions are likely to be widespread among gut microbes. The polysaccharides potentially involved in these interactions are difficult to study, and the ability of many bacteria to exert effects by modifying environmental polysaccharides is also a complicating factor. The molecular level analysis is especially challenging as polysaccharides are enzymatically, not genetically, encoded. Their molecular structures are described as a population average: the composition of the repeating unit and the number of those repeating units. For these reasons the precise molecular structure of these biomolecules is often impossible to determine with atomic-level specificity.

Cell-wall–associated rhamnans that are structurally similar to the *R. gnavus* glucorhamnan have been broadly described in many Lactobacilliales, which are a group of distantly related, aerotolerant, Gram-positive bacteria that produce lactic acid. These include species of *Lactococcus* (20), *Streptococcus* (21), and several *Enterococcus* species (22). These cell-wall rhamnose-glucose polysaccharides (RGPs) all share the same 1,3- and 1,2-linked rhamnose backbone with the *R. gnavus* glucorhamnan described here. However, between species, and even strains of the same species, there is tremendous variability in the number, linkage, and identity of the monosaccharides that make up the sidechains. Among these, the RGP from pathobiont *Streptococcus mutans* serotype c is the most similar to the *R. gnavus* glucorhamnan.

Unlike the distantly related Lactobacilliales, *R. gnavus* is in the order Clostridiales which are Gram-positive, obligate anaerobes. Within this group, only a single example of a high-rhamnose polysaccharide has been reported: the anti-tumorigenic "clostrhamnan" from *Clostridium saccharoperbutylacetonicum*, which is simply the rhamnose backbone with infrequent sidechains of a single glucose monomer (23).

These Gram-positive, cell-wall rhamnans have previously been identified as playing a role in establishing cell shape, cell-wall rigidity, and pathogenesis. For instance, mutants in *Enterococcus faecalis* enterococcal polysaccharide antigen (epa) gene cluster impair colonization of the human GI tract (24), and when Group A *Streptococcus* can no longer add sidechains to the rhamnan backbone, they are not impaired in their ability to grow, but are much less virulent in models of infections (25).

The last common ancestor to *R. gnavus* and the Lactobacilliales likely lived around three billion years ago (26), which indicates that the rhamnose polysaccharide backbone is conserved from an ancient ancestor, evolved independently in these branches, or spread among these organisms through horizontal gene transfer. Further studies on the distribution of these polysaccharides within Clostridiales and gene cluster evolution will be needed to distinguish between these possibilities. While the gene cluster presented above is likely responsible for the biosynthesis of the *R. gnavus* glucorhamnan, this will need to be confirmed by knocking out production in the wild-type strain, for which there is currently no genetic system; discovering a naturally occurring mutant that lacks production; or by recombinantly expressing the gene cluster in a rhamnan negative species.

Currently, the degree of structural conservation of the glucorhamnan(s) among strains of *R. gnavus* is unknown, but if the conservation of the putative biosynthetic gene cluster is an indicator (*SI Appendix*, Fig. S12), the structure will be highly conserved among most clinically sequenced isolates. Future analysis into how clinical isolates vary in this respect, along with transcriptomic analysis in healthy and CD patients, will help establish the clinical relevance of the glucorhamnan in Crohn's disease.

R. gnavus can utilize mucin as a carbon source (5, 10, 27), and may directly contribute to breakdown in gut barrier function. The degradation of this barrier, or the host's ability to maintain a bacteria-free zone within the mucin layer, may increase the immune system's exposure to triggers like the *R. gnavus* glucorhamnan, which would otherwise not be sensed by the immune system. In this way, mucin destruction and glucorhamnan production by *R. gnavus* could have profound effects on human inflammatory responses. Ultimately, we have identified a molecular hypothesis that can be interrogated to explain the link between the human gut commensal, *R. gnavus*, and Crohn's disease pathology.

Methods and Materials

Bacterial Growth Conditions. *R. gnavus* ATCC 29149 was grown in Bacto Todd-Hewitt broth, Bacto Brain-Heart infusion, and defined medium (*SI Appendix*, Dataset S3, based on ref. 14). Starter cultures were grown overnight for 16 h in Todd-Hewitt or Brain-Heart Infusion broths, or for 2 d for defined medium from glycerol stocks or agar plates of Todd Hewitt or defined medium. Large-scale (1 L) cultures were inoculated with 1:1,000 ratio

of starter culture in the same medium, and grown anaerobically (~5% H_2, 85% N_2, 10% CO_2) at 37 °C for 2 d.

Purification of Polysaccharide. After 1–3 d of growth (growth time made no significant differences on yield of polysaccharide), cells were separated from supernatant by centrifugation. Supernatant was sterile-filtered using 0.2-μm aPES filters. The sterile supernatant was processed in a tangential flow-filtration apparatus from Pall using a 10-kDa cutoff membrane, such that molecules smaller than this are discarded, and molecules larger than 10 kDa are retained in the reservoir. By this method, samples can be concentrated, diluted with fresh water, and reconcentrated, and, in effect, be dialyzed quickly. In this manner, 5 L of sterile, cell-free media were concentrated to ~500 mL, diluted to 5 L with fresh distilled, deionized water, and reconcentrated to ~500 mL for a total of three cycles. Crude, washed polysaccharides were resuspended in 1× TBS with 1 mM $MgCl_2$, $MnCl_2$, and $CaCl_2$ and residual nucleic acids were degraded with RNase and DNase for 2 h at 37 °C, followed by proteinase K digestion at 37 °C for 3 h at 37 °C. The crude polysaccharide fraction was concentrated in a 10-kD MWCO filter, and washed excessively with fresh water to remove peptides and nucleotides from enzymatic digestion. The rententate was purified by size-exclusion chromatography (SEC) over a Superdex Increase 200 column with water as the mobile phase. The active fraction from SEC was further purified by anion-exchange chromatography over an HF Q column.

Composition Analysis. Composition analysis was performed by GC-MS of per-O-trimethylsilyl (TMS) derivatives of the monosaccharide methyl glycosides produced from purified polysaccharide by acidic methanolysis as described previously (28). Briefly, the polysaccharide (400 μg) and inositol (20 μg as internal standard) were freeze-dried. Dry residue was heated with methanolic HCl in a sealed screw-top glass test tube for 16 h at 80 °C. After cooling and removal of the solvent under a stream of nitrogen, the sample was treated with a mixture of methanol, pyridine, and acetic anhydride for 30 min. The solvents were evaporated, and the sample was derivatized with Tri-Sil (Pierce) at 80 °C for 30 min. GC-MS analysis of the TMS methyl glycosides was performed on an Agilent 7890A GC interfaced to a 5975C MSD, using a Supelco Equity-1 fused silica capillary column (30 m × 0.25 mm ID).

Linkage Analysis. For glycosyl linkage analysis, dry sample (>1 mg) was permethylated, depolymerized, reduced, and acetylated. The resulting partially methylated alditol acetates (PMAAs) were analyzed by GC-MS as described previously (15). Briefly, dry polysaccharide was suspended in 200 μL DMSO and left to stir for a week. Permethylation was performed by two rounds of treatment with sodium hydroxide (15 min) and methyl iodide (45 min). Following sample workup, the permethylated material was hydrolyzed using 2 M TFA (2 h in sealed tube at 121 °C), reduced with $NaBD_4$, and acetylated using acetic anhydride/TFA. The resulting PMAAs were analyzed on an Agilent 7890A GC interfaced to a 5975C MSD, using a 30-m Supelco SP-2331 bonded phase fused silica capillary column.

NMR Measurements. NMR spectra were recorded at 298 K in D_2O with 0.1% acetone for internal referencing (^1H 2.225 ppm, ^{13}C 30.89 ppm) on a 500-MHz Bruker with cryoprobe. Phosphorous NMR spectra were recorded on a 400-MHz Varian. NOESY were collected with mixing time of 100 ms, and TOCSY were collected with mixing time of 200 ms.

Cell Culture. Bone marrow was isolated from C57BL/6 mice and differentiated in the presence of GM-CSF (Peprotech) for 7 d to generate mBMDCs. mBMDCs were treated with fractions overnight. Concentrations of TNF-α in culture media were detected using the TNF alpha mouse ELISA Kit from ThermoFisher. For receptor identification, $TLR2^{-/-}$ and $TLR4^{-/-}$ mice were used to generate mBMDCs. LPS was used as a positive control for a TLR4 agonist and Pam3Cys was used as a positive control for a TLR2 agonist. The Institutional Review Board and Institutional Animal Care and Use Committee of Massachusetts General Hospital approved the methods used in this study.

Quantitative Reverse Transcription PCR. Total RNA was purified by chloroform-phenol extraction from cell pellets of replicate cultures of *R. gnavus* grown in defined medium, aliquots of which were taken at the time points indicated. RNA was Dnase treated and cDNA was prepared using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with or without reverse transcriptase. Transcripts of interest were amplified from the cDNA library with gene-specific primers and visualized on an agarose gel, as well as quantified by real-time PCR carried out using iTaq Universal SYBR Green Supermix (Bio-Rad). All qPCRs were normalized to 16S rRNA gene expression at 24 h. Primers used are listed in *SI Appendix*, Tables S6 and S7.

ACKNOWLEDGMENTS. We thank the East Quad NMR Core at Harvard Medical School, and Prof. A. Sloan Devlin for GC-MS access. This work was funded by National Institutes of Health R01-AT009708 (J.C. and R.J.X.) and F32-GM126650 (M.T.H.). Polysaccharide structural analysis was supported by the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, US Department of Energy Grant (DE-SC0015662) to the Complex Carbohydrate Research Center.

1. B. P. Willing *et al.*, A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854.e1 (2010).
2. M. Joossens *et al.*, Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* **60**, 631–637 (2011).
3. A. B. Hall *et al.*, A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med.* **9**, 103 (2017).
4. K. Nishino *et al.*, Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *J. Gastroenterol.* **53**, 95–106 (2018).
5. C. W. Png *et al.*, Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am. J. Gastroenterol.* **105**, 2420–2428 (2010).
6. J. S. Levine, R. Burakoff, Extraintestinal manifestations of inflammatory bowel disease. *Gastroenterol. Hepatol. (N. Y.)* **7**, 235–241 (2011).
7. T.-C. Liu, T. S. Stappenbeck, Genetics and pathogenesis of inflammatory bowel disease. *Annu. Rev. Pathol.* **11**, 127–148 (2016).
8. N. A. Molodecky *et al.*, Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46–54 (2012).
9. D. Lee *et al.*, Diet in the pathogenesis and treatment of inflammatory bowel diseases. *Gastroenterology* **148**, 1087–1106 (2015).
10. E. H. Crost *et al.*, Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *PLoS One* **8**, e76341 (2013).
11. M. Breban *et al.*, Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. *Ann. Rheum. Dis.* **76**, 1614–1622 (2017).
12. H. Zheng *et al.*, Altered gut microbiota composition associated with eczema in infants. *PLoS One* **11**, e0166026 (2016).
13. K. Machiels *et al.*, Specific members of the predominant gut microbiota predict pouchitis following colectomy and IPAA in UC. *Gut* **66**, 79–88 (2017).
14. M. Tramontano *et al.*, Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol.* **3**, 514–522 (2018).
15. C. Heiss, J. S. Klutts, Z. Wang, T. L. Doering, P. Azadi, The structure of *Cryptococcus neoformans* galactoxylomannan contains β-D-glucuronic acid. *Carbohydr. Res.* **344**, 915–920 (2009).
16. K. Gheysen, C. Mihai, K. Conrath, J. C. Martins, Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chemistry* **14**, 8869–8878 (2008).
17. P. Sudarsanam *et al.*, [Ruminococcus] gnavus ATCC 29149, whole genome shotgun sequencing project. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/AAYG00000000.2. Accessed 2 January 2018.
18. S. K. Mazmanian, J. L. Round, D. L. Kasper, A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).
19. L. E. Comstock, D. L. Kasper, Bacterial glycans: Key mediators of diverse host immune responses. *Cell* **126**, 847–850 (2006).
20. I. Sadovskaya *et al.*, Another brick in the wall: A rhamnan polysaccharide trapped inside peptidoglycan of *Lactococcus lactis*. *MBio* **8**, e01303-17 (2017).
21. F. St Michael *et al.*, Investigating the candidacy of the serotype specific rhamnan polysaccharide based glycoconjugates to prevent disease caused by the dental pathogen *Streptococcus mutans*. *Glycoconj. J.* **35**, 53–64 (2018).
22. M. Y. Mistou, I. C. Sutcliffe, N. M. van Sorge, Bacterial glycobiology: Rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol. Rev.* **40**, 464–479 (2016).
23. Y. Watanabe, H. Misaki, Structure of a new rhamnose polysaccharide, clostrhamnan, isolated from antitumor hot water extract of *Clostridium saccharoperbutylacetonicum* cells. *Agric. Biol. Chem.* **51**, 931–932 (1987).
24. L. Rigottier-Gois *et al.*, The surface rhamnopolysaccharide epa of *Enterococcus faecalis* is a key determinant of intestinal colonization. *J. Infect. Dis.* **211**, 62–71 (2015).
25. N. M. van Sorge *et al.*, The classical lancefield antigen of group a Streptococcus is a virulence determinant with implications for vaccine design. *Cell Host Microbe* **15**, 729–740 (2014).
26. F. U. Battistuzzi, A. Feijao, S. B. Hedges, A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* **4**, 44 (2004).
27. L. E. Tailford *et al.*, Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat. Commun.* **6**, 7624 (2015).
28. J. Santander *et al.*, Mechanisms of intrinsic resistance to antimicrobial peptides of *Edwardsiella ictaluri* and its influence on fish gut inflammation and virulence. *Microbiology* **159**, 1471–1486 (2013).