

ORIGINAL ARTICLE

Expression of the blood-group-related glycosyltransferase *B4galnt2* influences the intestinal microbiota in mice

Fabian Staubach^{1,2}, Sven Künzel¹, Andrea C Baines³, Andrew Yee⁴, Beth M McGee⁵, Fredrik Bäckhed⁶, John F Baines^{1,7,10} and Jill M Johnsen^{8,9,10}

¹Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Plön, Germany;

²Department of Biology, Stanford University, Stanford, CA, USA; ³Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA; ⁴Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA;

⁵Howard Hughes Medical Institute, Ann Arbor, MI, USA; ⁶Department of Molecular and Clinical Medicine, Sahlgrenska Center for Cardiovascular and Metabolic Research/Wallenberg Laboratory, University of Gothenburg, Gothenburg, Sweden; ⁷Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany; ⁸Department of Medicine, Research Institute, Puget Sound Blood Center, Seattle, WA, USA and ⁹Department of Medicine, University of Washington, Seattle, WA, USA

Glycans on mucosal surfaces have an important role in host–microbe interactions. The locus encoding the blood-group-related glycosyltransferase β -1,4-*N*-acetylgalactosaminyltransferase 2 (*B4galnt2*) is subject to strong selective forces in natural house-mouse populations that contain a common allelic variant that confers loss of *B4galnt2* gene expression in the gastrointestinal (GI) tract. We reasoned that altered glycan-dependent intestinal host–microbe interactions may underlie these signatures of selection. To determine whether *B4galnt2* influences the intestinal microbial ecology, we profiled the microbiota of wild-type and *B4galnt2*-deficient siblings throughout the GI tract using 16S rRNA gene pyrosequencing. This revealed both distinct communities at different anatomic sites and significant changes in composition with respect to genotype, indicating a previously unappreciated role of *B4galnt2* in host–microbial homeostasis. Among the numerous *B4galnt2*-dependent differences identified in the abundance of specific bacterial taxa, we unexpectedly detected a difference in the pathogenic genus, *Helicobacter*, suggesting *Helicobacter* spp. also interact with *B4galnt2* glycans. In contrast to other glycosyltransferases, we found that the host intestinal *B4galnt2* expression is not dependent on presence of the microbiota. Given the long-term maintenance of alleles influencing *B4galnt2* expression by natural selection and the GI phenotypes presented here, we suggest that variation in *B4galnt2* GI expression may alter susceptibility to GI diseases such as infectious gastroenteritis.

The ISME Journal (2012) 6, 1345–1355; doi:10.1038/ismej.2011.204; published online 26 January 2012

Subject Category: microbe–microbe and microbe–host interactions

Keywords: *B4galnt2*; blood group; glycosylation; intestine; microbiota; mouse

Introduction

Host glycans have been shown to participate directly in specific host–microbe interactions (Bishop and Gagneux, 2007), and signatures of selection observed at the loci of carbohydrate blood group-related genes in humans (Calafell *et al.*, 2008; Ferrer-Admetlla *et al.*, 2009) are most likely the result of host–pathogen interactions (Anstee, 2010). The Sd(a)/Cad carbohydrate determinant is a

polymorphic blood group antigen of unknown function expressed on the red cells of 90% of human populations (Conte and Serafini-Cessi, 1991) and detectable in several other human tissues, including the intestinal mucosa and kidney in 98% of humans (Morton *et al.*, 1988). The glycosyltransferase β -1,4-*N*-acetylgalactosaminyltransferase 2 (*B4galnt2*) is responsible for catalyzing the last step in the biosynthesis of the Sd(a)/Cad antigen by the addition of an *N*-acetylgalactosamine (GalNAc) residue via a β -1,4 linkage to a subterminal galactose residue substituted with an α -2,3-linked sialic acid (Lo *et al.*, 2003; Montiel *et al.*, 2003). In the intestine, there is evidence of a gradient in GalNAc residues conferred by *B4galnt2* from the ileum to the colon (Robbe *et al.*, 2003). The character of intestinal mucus itself also changes with location,

Correspondence: JM Johnsen, Department of Medicine, Research Institute, Puget Sound Blood Center, Seattle, WA 98104, USA.
E-mail: jillj@psbcresearch.org

¹⁰These authors contributed equally to this work.

Received 16 May 2011; revised 22 September 2011; accepted 30 November 2011; published online 26 January 2012

both in variation in the types of mucins present (Corfield *et al.*, 2001) and in the balance of glycolipids and glycoproteins that may be capable of acting as B4galnt2 substrates (Dohi *et al.*, 1996; Kawamura *et al.*, 2005). Thus, distinct carbohydrate mucosal phenotypes are likely to result from B4galnt2 expression at different intestinal anatomic sites.

Intestinal expression of B4galnt2 is observed throughout vertebrates from fish (Stuckenholtz *et al.*, 2009) to humans, including mice. Although conservation across species is evidence for a functional constraint preventing the loss of intestinal expression, mice genetically deficient in B4galnt2 (B4galnt2^{-/-}) display no obvious phenotype, gastrointestinal (GI) or otherwise, under specified pathogen-free (SPF) conditions in the laboratory (Mohlke *et al.*, 1999; Johnsen *et al.*, 2008).

The absence of intestinal B4galnt2 expression is also observed in mice homozygous for a spontaneous *cis*-regulatory mutation termed *Modifier of von Willebrand Factor-1* (Mvwf1). Mvwf1 specifically turns off B4galnt2 expression in the intestinal epithelium and instead directs B4galnt2 expression in vascular endothelium. Endothelial cell expression of B4galnt2 results in aberrant posttranslational modification of the blood-clotting glycoprotein, von Willebrand factor (VWF), resulting in accelerated VWF clearance and low circulating VWF levels (Mohlke *et al.*, 1999), similar to the common human bleeding disorder, von Willebrand disease (Sweeney *et al.*, 1990). As in the B4galnt2 knockout mouse, loss of intestinal B4galnt2 expression due to Mvwf1 does not result in an obvious GI phenotype in the laboratory. Mvwf1 is common among laboratory mouse strains (Johnsen *et al.*, 2008), and is present at intermediate frequencies in wild *Mus musculus domesticus* (house mouse) populations, where there is strong evidence of both recent and long-term selection at the B4galnt2 locus (Johnsen *et al.*, 2009).

These observations seem paradoxical: Mvwf1 confers what would be expected to be a detrimental phenotype in *M. musculus* (a mild bleeding diathesis), yet the Mvwf1 allele is common in wild mice. The prevalence of natural murine B4galnt2 variants is further illustrated in our recent survey of multiple wild-mouse populations, in which we found loss of B4galnt2 intestinal glycans conferred by the Mvwf1 allele class to be frequent, but not always accompanied by the gain of B4galnt2 vascular expression (Linnenbrink *et al.*, 2011).

We hypothesized that loss of intestinal B4galnt2 expression is at least in part responsible for the striking signatures of selection at the B4galnt2 locus in wild house mice. In this study, we sought to determine the influence of B4galnt2 expression on the resident microbiota throughout the GI tract as evidence of an intestinal B4galnt2 phenotype, as shifts in the intestinal microbiota are known to affect host susceptibility to pathogens and disease (Bishop and Gagneux, 2007; Sekirov *et al.*, 2010;

Stecher and Hardt, 2010). We performed high-throughput 16S rRNA gene profiling at multiple intestinal sites in mouse sibling pairs that differed only in the presence or absence of B4galnt2 expression. In addition, we analyzed B4galnt2 expression patterns in germ-free versus conventional mice. Here we describe this detailed characterization of mouse GI bacterial communities in seven distinct locations from the duodenum to the colon and provide evidence for a significant effect of B4galnt2 expression on intestinal bacterial populations.

Materials and methods

Animal material and tissue sampling

All animal protocols (except for the germ-free protocols, see below) were approved by the University of Michigan University Committee on the Use and Care of Animals (UCUCA). C57BL6/J animals were purchased from the Jackson Laboratory (Bar Harbor, ME, USA). B4galnt2 knockout animals first engineered by Dr John Lowe were provided with permission and courtesy of Dr David Ginsburg. Genetic background, maternal effect, housing conditions, gender and diet were accounted for in the following mating scheme: B4galnt2^{-/-} animals bred for more than 20 generations to a C57BL6/J background were mated to C57BL6/J animals to generate heterozygous B4galnt2^{+/-} parents. B4galnt2^{+/-} parents were then intercrossed to generate B4galnt2^{+/+} and B4galnt2^{-/-} male sibling offspring. Four 'sibpair' cages containing at least one B4galnt2^{+/+} and one B4galnt2^{-/-} sibling were raised and housed together with standard mouse chow in the same room in a specified pathogen-free (SPF) animal facility at the same time.

Studies were performed at 10 weeks of age to ensure adequate time for the colonization and development of mature, stable microbiota (Rehman *et al.*, 2011). Tissues were harvested by fresh dissection of 1 cm of bowel at each of the following defined anatomic sites: mid-duodenum (D2), mid-jejunum, terminal ileum (ending at the cecal valve), cecum and mid-descending colon. With the exception of the cecum, luminal contents were separated from mucosal specimens by flaying open the bowel, washing twice with 0.7 ml ice-cold RNALater (Ambion, Carlsbad, CA, USA) and the 1.4 ml effluent containing stool pooled and marked as 'luminal contents'. The bowel was then placed in 1.4 ml ice-cold RNALater and marked as 'mucosa'. For the cecal samples, the cecal pouch was dissected and flayed open, the entire specimen placed in a 15-ml tube with 4 ml ice-cold RNALater, and the specimen gently shaken to remove the cecal contents. The mucosa was retrieved with forceps and rinsed with an additional 0.7 ml of ice-cold RNALater. This rinse was added to the 4-ml stool solution and marked 'luminal contents'. The rinsed cecal tissue was then placed in 1.4 ml ice-cold

RNA Later and marked 'mucosa'. Owing to variability in the presence of luminal contents in the sections of the small intestine, only the contents of the cecum and descending colon were analyzed. Instruments were cleaned between each anatomic site to avoid cross-contamination.

Germ-free animals

C57BL/6J animals were raised under conventional or germ-free conditions at the University of Gothenburg. All animal protocols were approved by the Research Animal Ethics Committee in Gothenburg. At 6 weeks of age, small bowel was harvested for *Dolichos biflorus* agglutinin (DBA) lectin histochemistry as described below.

DBA lectin histochemistry

DBA lectin is reactive with terminal *N*-acetyl-D-galactosamine (GalNAc) residues and specifically detects the *Mvwf1* switch in *B4galnt2* expression from the intestine to the blood vessel (Mohlke *et al.*, 1999). Fresh small bowel harvested from C57BL/6J and *B4galnt2*^{-/-} animals was fixed in Z-fix (Anatech Ltd, Battle Creek, MI, USA) at room temperature overnight, then paraffin embedded. DBA lectin histochemical staining was performed on embedded tissues using horseradish peroxidase-conjugated DBA (EY Laboratories, Inc., San Mateo, CA, USA) as previously described (Mohlke *et al.*, 1999).

DNA extraction

Approximately 100 mg of tissue or luminal contents was transferred to a 2-ml screwcap tube containing glass beads 0.1, 0.5 and 1 mm in size, each 50 mg (BioSpec Products, Bartlesville, OK, USA), and 1.4 ml lysis buffer argininosuccinate lyase from the QIAmp DNA stool mini kit (Qiagen, Hilden, Germany). After bead beating in a Precellys (Peqlab, Erlangen, Germany) bead beater (3 × 15 s at 6500 r.p.m.), samples were heated to 95 °C for 10 min and constantly shaken in a thermomixer (Eppendorf). Bacterial DNA was then extracted using the QIAmp DNA stool mini kit (Qiagen) following the manufacturer's instructions.

PCR and pyrosequencing

Universal bacterial primers for the V3 16S rRNA variable region were used to amplify bacterial 16S RNA as described by Dethlefsen *et al.* (2008). PCR amplicon primers were fused to molecular identifier (MID) tags and 454 sequencing adapters according to the manufacturer's instructions (Roche, Basel, Switzerland). All PCR reactions were performed in 50-μl duplicates and combined after PCR. The products were extracted with the Qiagen MinElute Gel Extraction Kit and quantified with the Quant-iT (Invitrogen, Karlsruhe, Germany) dsDNA BR Assay Kit on a Nanodrop 3300 fluorometer. Equimolar

amounts of purified PCR product from seven knock-out and six wild-type individuals were pooled to generate libraries for each anatomic location. Pooled amplicon libraries were further purified using Ampure Beads (Agencourt, Bernried, Germany). A sample of each library was run on an Agilent (Waldbronn, Germany) bioanalyzer as described in the Roche Titanium Amplicon Sequencing protocol before entering emulsion PCR and sequencing. Each library was sequenced on one-eighth of a picotiter plate on a 454 GS-FLX (Roche).

Sequence analysis

Raw reads obtained from the sequencer were filtered using a perl script according to the following criteria: average quality ≥ 25, no ambiguous bases, length between 110 and 200 bp, excluding MID tags and bacterial primers, and a perfect match to the MID and bacterial primer. Primer and tag sequences were removed using alignment with the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970). Sequences were sorted by individuals to produce group files for analysis in MOTHUR v.1.12.3 (Schloss, 2009). Classification into bacterial phyla was performed with the Ribosomal Database Project classifier tool (Wang *et al.*, 2007) and the RDP taxonomy database as implemented in MOTHUR. Reads were aligned with the kmer algorithm available under the align.seqs command (Schloss, 2009) in MOTHUR to the silva reference database (Pruesse *et al.*, 2007). Sequences that did not match the reference alignment in the expected positions were removed.

Statistical analysis

Aligned sequences were used to build a distance matrix and group sequences into operational taxonomic units (OTUs) in MOTHUR. Species richness estimates and collector's curves were generated based on these OTUs and drawn using the R statistics package v.2.11.1 (R Development Core Team, 2010). Phylogenetic trees were built based on the MOTHUR-derived distance matrices with FastTree v. 2.1.3 (Price *et al.*, 2009, 2010) and submitted to the Fast Unifrac online tool for principal coordinate analysis (PCoA) of Unifrac distances (Hamady *et al.*, 2009). Linear models were fitted to the data using the 'lm' function implemented in R. Two mixed effect models were made with the 'lme' function contained in the nlme package v. 3.1-96 (Pinheiro *et al.*, 2009) to account for potential maternal (cage of origin) effects (models 6 and 7 in Table 1). To compare species richness between different sections of the gut, we subsampled the same number of reads per individual and section (*in silico* capping), calculated Chao's richness estimator (Chao, 1984) and compared the number of OTUs observed in the subsamples.

Table 1 Linear model comparisons for the effect of genotype on Unifrac PCOs

Model	Dependent variable	Explanatory variables	AIC	P-value	RSS	Rsq
1	Unifrac PCo1	Null model	16.66		5.66	
2	Unifrac PCo1	Gut section	-145.86	1.20E-32	0.69	0.88
3	Unifrac PCo2	Null model	-115.81		1.15	
4	Unifrac PCo2	Gut section	-213.42	6.87E-20	0.31	0.73
5	Unifrac PCo2	Gut section+genotype	-217.864	0.016 ^a	0.28	0.07 ^a
6 ^b	Unifrac PCo2	Fixed = genotype Random = section/cage	NA	0.017	NA	NA
7 ^b	Unifrac PCo2	Fixed = genotype Random = section/cage/genotype	NA	0.019	NA	NA

Abbreviations: Cage, cage of origin (maternal effect); NA, not applicable in that context; section, location along the intestine.

^aANOVA in comparison with model 4.

^bLinear mixed effects model.

P-values by ANOVA.

Candidate bacterial species distinguishing the genotypes were determined by calculating the point biserial correlation between species-level (97% sequence identity) OTUs and genotype with the 'multipatt' function as part of the 'indicpecies' R package v. 1.5.1 (De Caceres and Legendre, 2009). This approach considers not only the difference in read number assigned to an OTU between genotypes, but also the number of individuals of each genotype in which the OTU is observed, thus drawing statistical power from biological replicates.

The number of sequence reads per sample was capped *in silico* by random subsampling to obtain equal sample sizes for species richness, Unifrac and for detection of candidate species. To provide an initial overview of the sequence reads associated with each region of the GI tract, we classified sequences by aligning them to the SILVA reference database (Pruesse et al., 2007) and obtained taxonomic information from RDP Classifier (Wang et al., 2007).

Results

DBA lectin staining

DBA lectin staining of the small bowel from C57BL6/J animals demonstrates strong DBA reactivity in intestinal epithelial cells, the mucus material contained in goblet cells, and the lining of the intestinal luminal surface. No DBA reactivity is seen in the small bowel of *B4galnt2*^{-/-} animals (Figure 1). These data show that DBA lectin histochemistry is a specific method for detection of *B4galnt2*-carbohydrates in the bowel, which are evident on the intestinal epithelium and mucus of *B4galnt2*-sufficient animals but absent in *B4galnt2*^{-/-} animals.

High-throughput sequencing of the *B4galnt2*^{+/+} and *B4galnt2*^{-/-} intestinal microbiota

To determine the influence of *B4galnt2* expression on the resident microbiota throughout the GI tract, we performed high-throughput pyrosequencing (454) of the bacterial 16S rRNA gene. Our sibpair

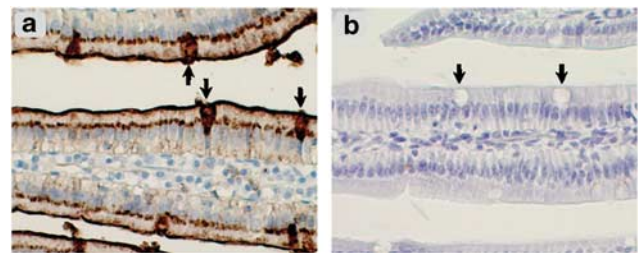


Figure 1 Small-bowel DBA lectin staining. (a) Wild-type (C57BL6/J) intestine exhibits robust intestinal epithelial DBA lectin staining (brown). Mucus (arrows indicate mucus-containing goblet cells) is strongly DBA lectin positive. (b) *B4galnt2*^{-/-} bowel demonstrates complete loss of DBA lectin staining of the epithelium and mucus.

breeding scheme (see Materials and methods) yielded a total of six *B4galnt2*^{+/+} and seven *B4galnt2*^{-/-} individuals. We surveyed the bacterial communities associated with the mucus layer of five major functional compartments of the murine intestine (duodenum, jejunum, ileum, cecum and descending colon) and the luminal contents of the cecum and colon. The 16S rRNA V3 region was successfully amplified from all individuals at all sites, with the exception of the duodenum of one *B4galnt2*^{-/-} individual (see Supplementary Table 1 for distribution of reads per GI tract location). A total of 337 449 reads of the hypervariable region V3 were analyzed after quality filtering. A single *B4galnt2*^{-/-} individual displayed a strongly deviant pattern with respect to the relative abundance of the major bacterial phyla compared with all 12 other mice in the small intestine (Supplementary Figure 1), despite no notable phenotype at the time of dissection. This individual was housed with two other siblings, one *B4galnt2*^{+/+} and one *B4galnt2*^{-/-}, which did not exhibit this pattern. Although we cannot exclude that this aberration is a consequence of loss of *B4galnt2* expression, we speculate that this animal was subclinically ill or otherwise compromised, resulting in a shift distinct from other genotypes. Thus, we removed this individual from further analyses.

Bacterial phyla of the GI tract by anatomic site

The communities within the distinct sections of the GI tract differ largely in their composition and proportion of the major bacterial phyla (Figure 2). The mucosa of the small intestine (duodenum, jejunum and ileum) is dominated by Firmicutes (77%, 78% and 91%, respectively). As previously reported by Ley *et al.* (2005), we also detected sequences of Cyanobacteria origin in the small intestine (1.5% in duodenum and 2.1% in the jejunum). The mucosa of the cecum and colon contained a comparatively lower proportion of Firmicutes (40% and 25%, respectively) and higher proportion of Bacteroidetes (19% and 15%, respectively) and Proteobacteria (37% and 59%, respectively). The cecum samples also contained *Deferribacteres* (2.5%).

The differences between the mucosa and the adjacent luminal contents of the cecum and colon are particularly striking. Cecum and colon mucosal tissues contain 19% and 15% Bacteroidetes, compared with 32% and 64%, respectively, in the lumen. The opposite pattern is found for Proteobacteria, with 37% and 59%, respectively, in the cecum and colon mucosal tissues, compared with 8% and 3% in the lumen. Furthermore, the *Tenericutes* are more abundant in the lumen. This remarkable distinction between the mucosal and luminal communities might be expected in light of previous findings that bacterial populations occupying human colonic mucosa are distinct from those present in fecal samples (Eckburg *et al.*, 2005; Willing *et al.*, 2010). Our results provide direct support for the presence of distinct communities inhabiting the niches found along the GI tract, including those associated with the tissue versus lumen at a given location.

Bacterial species richness differs between sections of the GI tract

To characterize the diversity of these communities at lower taxonomic levels, we tested for differences in bacterial species richness in the *B4galnt2*^{+/+} and *B4galnt2*^{-/-} mice at the five mucosal anatomic

locations of the GI tract and in the luminal contents of the cecum and colon (Supplementary Figures 2 and 3). For this analysis, sequences were grouped into species-level (97% similarity) OTUs and Chao's species richness estimator was calculated.

Although we detected no significant difference in species richness with respect to *B4galnt2* genotype in any of the sections, several interesting patterns are apparent with regard to the individual GI tract locations. Species richness is highest in the luminal samples and on average lower in the mucus layer (335 versus 187 species, Chao's species richness estimator), consistent with the protective function of the mucus layer and a more controlled interaction between host and microbes closer to the intestinal epithelium. The high richness observed in the cecum may relate to its role as a biofermenter offering stable conditions for the growth of diverse species (Savage, 1977). Interestingly, we also observe a progressive decline of bacterial species richness along the small intestine (201, 168 and 82 for duodenum, jejunum and ileum, respectively, Chao's species richness estimator), mainly due to a much lower diversity in the ileum.

Species composition differs between B4galnt2^{+/+} and *B4galnt2*^{-/-} mice

To test whether the composition of bacterial communities is influenced by *B4galnt2* expression, we compared bacterial communities between *B4galnt2*^{+/+} and *B4galnt2*^{-/-} mice using the phylogeny, Germany-based beta-diversity measure UniFrac. This metric represents the distance between bacterial communities by comparing the shared branch length between phylogenetic trees underlying two communities. A matrix of UniFrac distances between the individual mice was analyzed for principal coordinates (PCo) to partition variation among samples into the most important independent components (Figure 3a). The resulting variance along the principal coordinates was analyzed in a linear model framework with respect to the experimental setup (Figure 3b, Table 1).

Our model identifies *B4galnt2* genotype as a significant determinant of PCo2, and hence bacterial

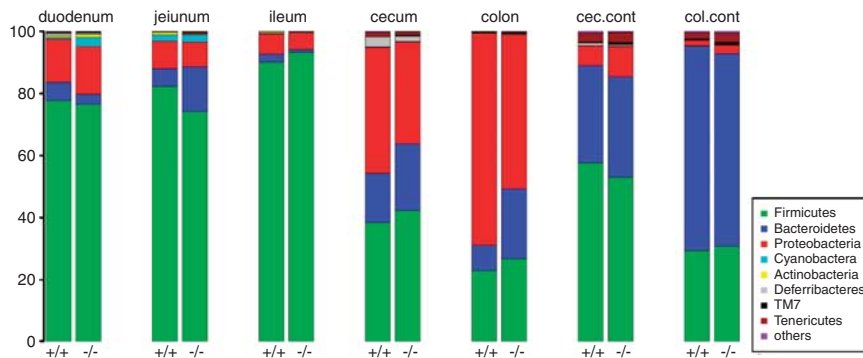


Figure 2 Abundance of major bacterial phyla by anatomic site. Average abundance of sequencing reads from major bacteria phyla in +/+ (*B4galnt2*^{+/+}) and -/- (*B4galnt2*^{-/-}) mice along the intestine in percent of total bacterial sequences obtained. cec.cont, cecum content; col.cont, colon content.

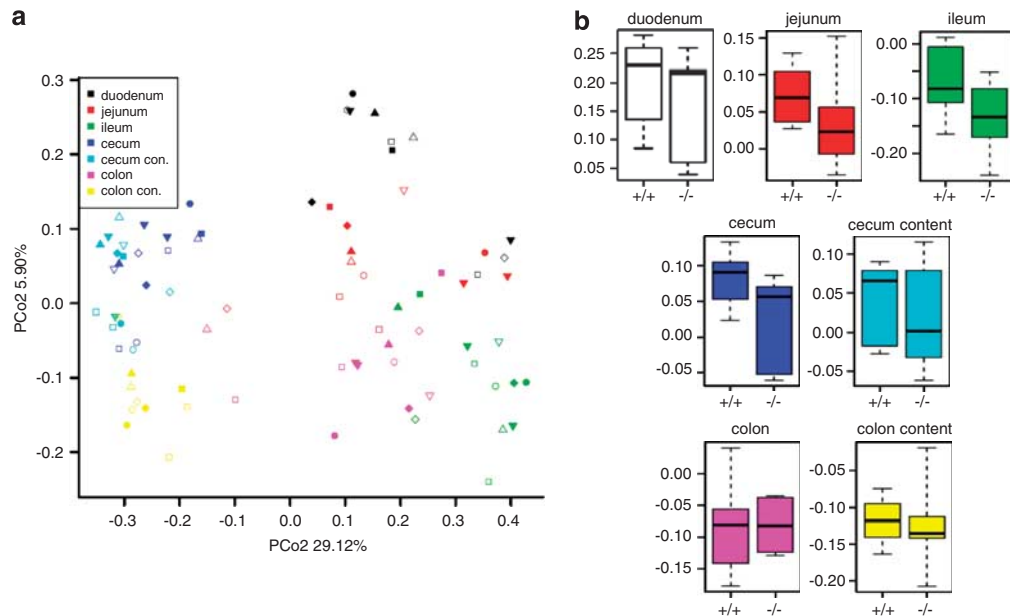


Figure 3 Principal coordinate analysis (PCoA) of Unifrac distances across anatomic sites and genotypes. (a) PCoA of unweighted Unifrac distances for the whole dataset (all anatomic sites sampled). Colors by section, filled symbols = $+/+$; open symbols = $-/-$; symbol shape by cage of origin. (b) Box plot of PCo2 from a, bold horizontal lines represent the median, the box edges mark the upper and lower quartiles, whiskers extend towards the maximum and minimum; color code as in a.

community composition. Both a linear model including section and genotype as additive effects (model 5, Table 1) and the application of mixed-effects models to account for maternal (cage of origin) effects (models 6 and 7) assign genotype as a significant explanatory effect of community composition. The effect is visualized in Figure 3b: The value along PCo2 assigned to $B4galnt2^{+/+}$ mice is higher for all portions of the GI tract except the colon, indicating a systematic effect of $B4galnt2$ expression on bacterial communities along the GI tract.

This analysis also reveals the striking effect of anatomic location on bacterial composition along the GI tract. The individual locations of the gut analyzed in our study differ strongly in the composition of their associated bacteria, accounting for 88% and 73%, respectively, of the variance observed along the first two PCos.

Differences in the bacterial community of $B4galnt2^{+/+}$ and $B4galnt2^{-/-}$ mice by location

Because distinct locations along the GI tract differ largely in their bacterial composition, genotype-dependent differences between the bacterial communities of $B4galnt2^{+/+}$ and $B4galnt2^{-/-}$ mice are likely to be site-specific. Thus, to analyze the potential influence of $B4galnt2$ on bacterial communities independent of variation between sections, we applied UniFrac to each anatomic location separately. We find the largest genotype effect on the bacterial communities associated with the ileum (Figure 4, $P = 7.49 \times 10^{-5}$, $R^2 = 0.81$, $P = 0.001$ after

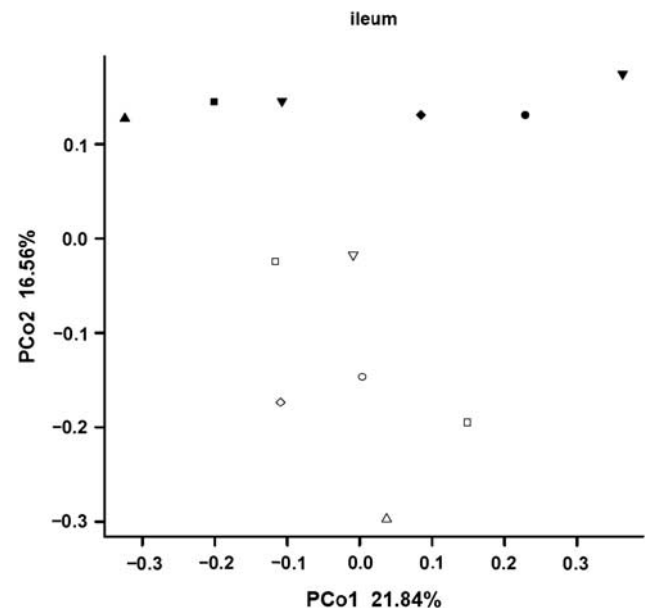


Figure 4 Unweighted Unifrac PCoA of the ileal mucosal bacterial community. Solid symbols = $B4galnt2^{+/+}$, empty symbols = $B4galnt2^{-/-}$. Symbol shape by cage of origin.

Bonferroni correction for testing all sections and the first two PCos, analysis of variance).

Individual bacterial species associated with $B4galnt2^{+/+}$ and $B4galnt2^{-/-}$ genotypes

To shed light on the identity of bacterial species that make up some of the differences between the

intestinal bacterial communities associated with *B4galnt2*^{+/+} and *B4galnt2*^{-/-} mice, we applied a common ecological measure of species habitat association (Dufréne and Legendre, 1997) to OTU clusters (97% sequence identity). OTUs identified as indicators of *B4galnt2* expression ‘habitat’ (that is, those consistently present or more abundant in one genotype compared with the other) were classified using the Ribosomal Database Project tool (Wang *et al.*, 2007). Numerous interesting candidate taxa belonging to the three major phyla Bacteroidetes, Firmicutes and Proteobacteria were identified by this method (Table 2). Members of the Firmicutes appear to be the most widely influenced. A total of 11 indicator OTUs belonging to the classes Clostridiales or Lactobacillales were identified from this group, with at least one member differentiating *B4galnt2* expression habitats in all seven sampled locations of the GI tract. The Bacteroidetes contained eight OTUs distributed across all jejunum, cecum and the luminal contents of the cecum and colon, five of which belonged to the genus *Barnesiella*. Four OTUs belonging to the Proteobacteria were identified in the duodenum, colon and luminal contents of the colon, three of which belonged to the genus *Helicobacter*.

Intestinal expression of B4galnt2 does not require the presence of bacteria

Expression of intestinal glycosyltransferases, such as the *Fut2* glycosyltransferase, has been shown to be influenced by the microbiota present (Bry *et al.*, 1996; Meng *et al.*, 2007). This regulatory mechanism likely has an important role in the host’s ability to alter the mucosal surface in response to the environment. To determine whether *B4galnt2* expression similarly requires the presence of intestinal bacteria,

DBA lectin staining was performed on the small bowels of mice housed under conventional and germ-free conditions. Intestinal DBA lectin staining was present and appeared similar under both conventional and germ-free conditions (data not shown), indicating that *B4galnt2* expression occurs in the absence of intestinal bacteria.

Discussion

At birth the intestinal tract is sterile, but is rapidly colonized by a diverse spectrum of bacteria, in addition to archaea and eukaryotes. These organisms are provided a nutrient-rich and largely stable environment by the host. In turn, the host relies on the microbiota for a variety of metabolic processes and their presence is required for normal host intestinal development, mucosal integrity and maintenance of immunologic balance (Backhed *et al.*, 2005; O’Hara and Shanahan, 2006; Artis, 2008; Fraser *et al.*, 2009). Furthermore, these complex communities protect the host from pathogenic organisms in several ways: by occupying microbial niches resulting in displacement, by production of antimicrobial factors and by competing for nutrients and receptors (O’Hara and Shanahan, 2006). By systematically profiling multiple locations throughout the murine GI tract, we have characterized distinct microbial communities at discrete anatomic sites throughout the intestine and found striking differences between sites and between the mucosa and adjacent luminal contents. Although this is seemingly in contrast to a previous report that found no difference between the populations of the mucosa and lumen in a humanized mouse model (Turnbaugh *et al.*, 2009), this apparent discrepancy may be attributable to the current study being based

Table 2 Candidate bacterial species as determined by OTU genotype correlation

Section	Genotype	OTU	r	P	Seq	Taxonomy
Duodenum	+/+	154	0.76	0.02	6	Campylobacteriales (100); Helicobacteraceae (100); Helicobacter (100)
	+/+	134	0.712	0.03	14	Lactobacillales (100); Streptococcaceae (100); Lactococcus (100)
	-/-	37	0.844	0.003	15	Lactobacillales (100); Streptococcaceae (100); Streptococcus (80)
	-/-	63	0.392	0.037	76	Pseudomonadales (100); Moraxellaceae (100); Acinetobacter (100)
Jejunum	+/+	135	0.579	0.039	171	Clostridiales (100); Clostridiaceae (100)
	-/-	48	0.496	0.048	22	Bacteroidales (100); Porphyromonadaceae (100); Barnesiella (100)
Ileum	-/-	62	0.707	0.015	8	Clostridiales (100); Clostridiaceae (100)
Cecum	+/+	595	0.768	0.018	7	Clostridiales (100); Lachnospiraceae (100); Moryella (86)
	-/-	37	0.622	0.029	59	Clostridiales (100); Lachnospiraceae (100); Moryella (100)
	-/-	103	0.61	0.024	41	Bacteroidales (100); Porphyromonadaceae (100)
	-/-	110	0.481	0.048	70	Bacteroidales (100); Porphyromonadaceae (100); Barnesiella (100)
Colon	+/+	123	0.704	0.046	9	Lactobacillales (100)
	+/+	30	0.597	0.044	15	Campylobacteriales (100); Helicobacteraceae (100); Helicobacter (100)
Cecum cont.	+/+	331	0.763	0.012	44	Clostridiales (100)
	+/+	499	0.742	0.017	9	Clostridiales (100); Lachnospiraceae (100)
	-/-	192	0.671	0.048	8	Clostridiales (100); Lachnospiraceae (100); Syntrophococcus (88)
	-/-	105	0.632	0.047	13	Bacteroidales (100); Porphyromonadaceae (100); Barnesiella (100)
	-/-	103	0.61	0.024	41	Bacteroidales (100); Porphyromonadaceae (100)
Colon cont.	+/+	23	0.632	0.027	17	Bacteroidales (100); Porphyromonadaceae (100); Barnesiella (100)
	-/-	192	0.667	0.046	12	Bacteroidales (100); Porphyromonadaceae (100); Barnesiella (100)

Abbreviation: OTU, unique identifier for the respective OTU (artificial species name); sequence number calculations are based on taxonomy; class, family, genus are listed according to RDP taxonomy; bootstrap values are shown in brackets.

on sampling of unperturbed native bacterial communities.

We find that variation in the expression of a single glycosyltransferase, *B4galnt2*, is associated with significant shifts in the composition of the intestinal microbiota. These differences are consistent with host carbohydrate-specific selection on colonizing microbial populations, as *B4galnt2*^{-/-} animals and their control *B4galnt2*^{+/+} littermates shared the same microbe exposures over their lifetime. The fact that the ileum displayed the most clear separation in overall composition as measured by the unweighted UniFrac metric also suggests *B4galnt2*-dependent immune activity, as strong antibacterial activity is present in this tissue (Petnicki-Ocwieja et al., 2009).

In addition to overall changes in composition, we identify a number of specific bacterial lineages influenced by host *B4galnt2* expression. Although it could be argued that many bacterial orders found in the bowel have at least one member that has been shown to be important for gut health, nearly all the candidate lineages we identified as being influenced by *B4galnt2* expression have been previously identified as significant to intestinal communities. For example, we discovered differences in several OTUs belonging to the order Clostridiales in the jejunum, ileum, cecum and cecum content, similar to two recent reports that found distinct Clostridiales OTUs in inflammatory bowel disease (IBD) samples compared with the controls (Willing et al., 2010; Frank et al., 2011). Furthermore, we observe differences in the family Lachnospiraceae (Order: Clostridiales), which was identified to be decreased in patients with Crohn's disease localized to the ileum compared with controls (Willing et al., 2010) and also reported to be disparate between healthy and

diseased mice in an IL10 knockout model of IBD (Ye et al., 2008). Interestingly, Lachnospiraceae are major determinants of the recently reported 'enterotypes' described in a large study of human fecal metagenomes (Arumugam et al., 2011). Ye et al. (2008) also identified differences in their murine IBD model in the genus *Barnesiella*, which we too found to be influenced by *B4galnt2* genotype. Others have associated *Barnesiella* with CD8+ T-cell function in mice (Presley et al., 2010), suggesting more than a bystander role for this genus in intestinal inflammation.

Intriguingly, we found that the indicator OTUs for *B4galnt2*^{+/+} and *B4galnt2*^{-/-} mice at the same location of the gut often were members of the same bacterial family or genus (Figure 5). *Streptococcus* and *Lactococcus* were indicative of *B4galnt2*^{+/+} and *B4galnt2*^{-/-} mucosal communities in the duodenum, as were two different *Moryella* OTUs in the cecum. Likewise, closely related indicator OTUs distinguished by *B4galnt2* genotype were found in the luminal content of cecum and colon (Lachnospiraceae and *Barnesiella*, respectively). Thus, similar species appear to substitute each other depending on *B4galnt2* genotype, indicating that many closely related species have the potential to occupy distinct *B4galnt2* glycan-defined niches in the mucosa.

Unexpectedly, among the bacteria identified were several OTUs belonging to the genus *Helicobacter*. *Helicobacter* spp. are known to naturally infect and cause disease in laboratory mice, most commonly *H. hepaticus*, *H. bilis* and *H. typhlonius* (Feng et al., 2005). Although a difference in *Helicobacter* abundance would be expected to result in an altered risk for enteritis or other hepatobiliary inflammatory processes, in a protected laboratory environment

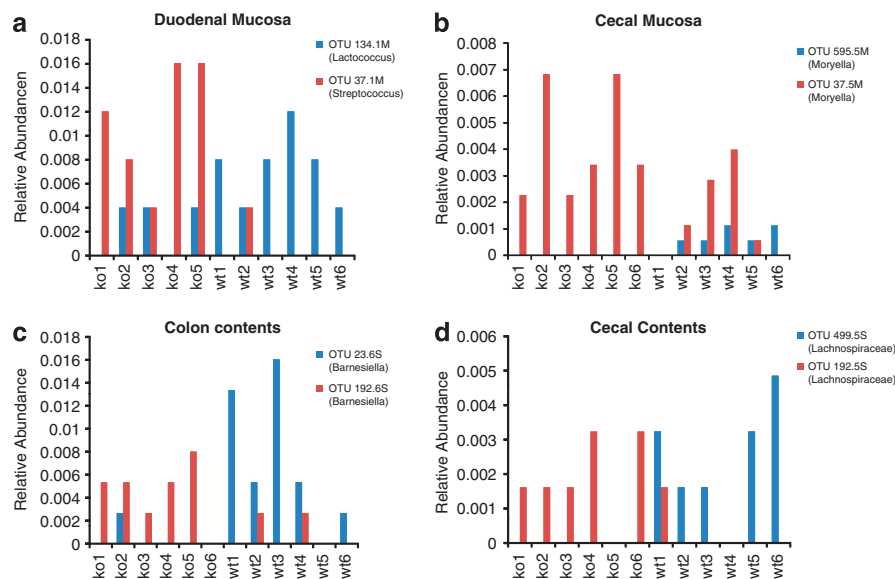


Figure 5 Individual bacterial species associated with *B4galnt2* genotype. OTUs of closely related species appear to substitute by *B4galnt2* genotype (*B4galnt2*^{+/+} = wt, *B4galnt2*^{-/-} = ko). (a) Duodenal mucosa: OTU 134.1M (*Lactococcus* sp.), OTU 37.1M (*Streptococcus* sp.), (b) cecal mucosa: OTU 595.5M (*Moryella* sp.), OTU 37.5M (*Moryella* sp.), (c) colon contents: OTU 23.6S (*Barnesiella* sp.), OTU 192.6S (*Barnesiella* sp.), (d) cecal contents: OTU 499.5S (Lachnospiraceae), OTU 192.5S (Lachnospiraceae).

animals with *Helicobacter* may exhibit a more subtle or subclinical phenotype. In humans, although *H. hepaticus* and *H. bilis* are both associated with hepatocentric disease (Goldman and Mitchell, 2010), the most prevalent and best-studied pathogenic *Helicobacter* is *H. pylori*.

H. pylori predominantly inhabits the gastric mucosa and is associated with a spectrum of human intestinal diseases ranging from gastritis to malignancy (Timothy and Martin, 2009). Most *H. pylori* is located within the overlying gastric mucus layer and does not interact with the underlying epithelium. However, under some conditions *H. pylori* can adhere to the gastric mucosa, triggering virulence factors in the bacterium and an inflammatory response from the host (Timothy and Martin, 2009). Thus, the mere presence of *Helicobacter* may result in a bystander effect by which the nearby-resident microbiota may be influenced not only by niche competition with *Helicobacter* but also by more general environmental changes due to mucosal inflammation and upregulation of host defenses.

Adhesion to host glycans presented on the gastric mucosa, notably carbohydrate antigens of the ABH and Lewis blood group systems, is a critical step in the pathogenesis of *H. pylori* (Kobayashi et al., 2009). Furthermore, *H. pylori* is known to be capable of specifically binding additional carbohydrate moieties, including terminal β -1,3-GalNAc residues (Miller-Podraza et al., 2005), while other murine enterohepatic *Helicobacter* spp. isolates also demonstrate evidence of carbohydrate-specific adhesion (Hynes et al., 2003). In the duodenal and colonic mucosa, two indicator *Helicobacter* spp. were detected in the *B4galnt2*^{+/+} animals, yet these *Helicobacter* were largely undetectable in *B4galnt2*^{-/-} individuals (Figure 6). Thus, we postulate a novel direct interaction between *Helicobacter* and host mucosal *B4galnt2*-derived β -1,4-GalNAc residues to be a likely mechanism responsible for the significant increase in abundance of these *Helicobacter* species observed in the *B4galnt2*^{+/+} mice.

The conservation of intestinal *B4galnt2* expression across species suggests an important functional role for *B4galnt2* in the GI tract. This hypothesis is supported by our finding that the loss of *B4galnt2* expression significantly impacts the composition of the resident microbiota. We propose that the complex signatures of natural selection observed in house mice (Johnsen et al., 2009) are at least in part due to the variation of *B4galnt2*-GalNAc residues on intestinal mucosal surfaces (the loss of glycans during evolution is discussed by Bishop and Gagneux (2007)). Taken together, these data support a scenario in which the loss of intestinal *B4galnt2* expression offers a significant fitness advantage in the face of pathogens reliant on the presence of the otherwise ubiquitous intestinal mucosal *B4galnt2*-derived GalNAc. Host *B4galnt2* glycans may serve as a carbon source for both symbiotic and

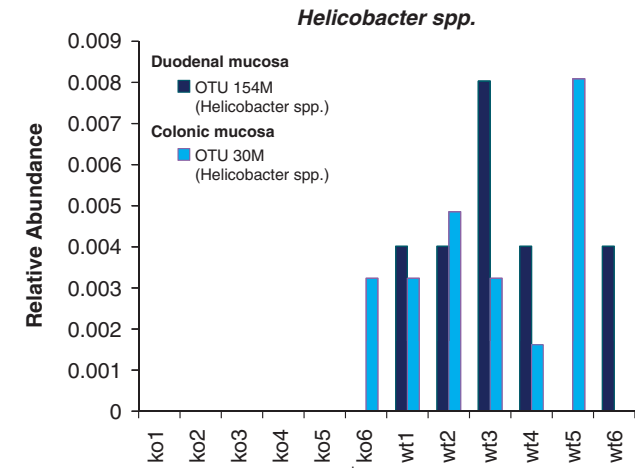


Figure 6 Abundance of *Helicobacter* spp. by *B4galnt2* genotype. Relative abundance of OTUs of two *Helicobacter* spp. in the duodenum and colon segregated by *B4galnt2* genotype (individuals are the same as in Figure 5, *B4galnt2*^{+/+} = wt and *B4galnt2*^{-/-} = ko). *ko6 duodenal mucosal sample was not available for analysis (see Results).

disease-causing organisms and/or change specific binding targets for GI pathogens. For example, even in this study intended to characterize the symbiotic microbiota, we detected a difference in the pathogenic genus *Helicobacter*, which is known to adhere to other host blood group carbohydrate structures during pathogenesis.

In summary, we have found that *B4galnt2* expression influences the intestinal microbiota. *Helicobacter* and other enteric pathogens known to interact with GalNAc residues, including protozoans (*Entamoeba histolytica* (Frederick and Petri, 2005)), viruses (Norovirus (Shirato et al., 2008)) and a spectrum of bacteria (Krivan et al., 1988; Karlsson, 1995), are excellent candidates to underlie the striking signatures of selection at the *B4galnt2* locus in wild-mouse populations.

Acknowledgements

We wish to thank David Ginsburg for his assistance with the *B4galnt2*^{-/-} mice, Philipp Rausch, Ateequr Rehman and Sri Girish Srivinas Murthy for helpful discussions and assistance in data analysis, and Silke Carstensen for excellent technical assistance. This work was supported by the Puget Sound Blood Center (JM), the National Institutes of Health RO1-HL039693 (JM), the Deutsche Forschungsgemeinschaft (DFG) Excellence Cluster 'Inflammation at Interfaces' (JFB) and DFG grants BA 2863/2-1 (JFB) and STA 1154/1-1 (FS) and an intramural grant from the CAU Kiel Medical Faculty (JFB).

References

Anstee DJ. (2010). The relationship between blood groups and disease. *Blood* **115**: 4635–4643.

- Artis D. (2008). Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat Rev Immunol* **8**: 411–420.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* **473**: 174–180. (Erratum in: *Nature* 2011; **474**: 666).
- Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.
- Bishop JR, Gagneux P. (2007). Evolution of carbohydrate antigens—microbial forces shaping host glycomes? *Glycobiology* **17**: 23R–34R.
- Bry L, Falk PG, Midtvedt T, Gordon JI. (1996). A model of host-microbial interactions in an open mammalian ecosystem. *Science* **273**: 1380–1383.
- Calafell F, Roubinet F, Ramírez-Soriano A, Saitou N, Bertranpetit J, Blancher A. (2008). Evolutionary dynamics of the human ABO gene. *Hum Genet* **124**: 123–135.
- Chao A. (1984). Nonparametric estimation of the number of classes in a population. *Scand J Stat* **11**: 265–270.
- Conte R, Serafini-Cessi F. (1991). Comparison between the erythrocyte and urinary Sda antigen distribution in a large number of individuals from Emilia-Romagna, a region of northern Italy. *Transfus Med* **1**: 47–49.
- Corfield AP, Carroll D, Myerscough N, Probert CS. (2001). Mucins in the gastrointestinal tract in health and disease. *Front Biosci* **6**: D1321–D1357.
- De Caceres M, Legendre P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566–3574.
- Dethlefsen L, Huse S, Sogin ML, Relman DA. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**: e280.
- Dohi T, Yuyama Y, Natori Y, Smith PL, Lowe JB, Oshima M. (1996). Detection of N-acetylgalactosaminyltransferase mRNA which determines expression of Sda blood group carbohydrate structure in human gastrointestinal mucosa and cancer. *Int J Cancer* **67**: 626–631.
- Dufréne M, Legendre P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345–366.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.
- Feng S, Ku K, Hodzic E, Lorenzana E, Freet K, Barthold SW. (2005). Differential detection of five mouse-infecting helicobacter species by multiplex PCR. *Clin Diagn Lab Immunol* **12**: 531–536.
- Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A *et al.* (2009). A natural history of FUT2 polymorphism in humans. *Mol Biol Evol* **26**: 1993–2003.
- Frank DN, Robertson CE, Hamm CM, Kpadeh Z, Zhang T, Chen H *et al.* (2011). Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* **17**: 179–184.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741–746.
- Frederick JR, Petri Jr WA. (2005). Roles for the galactose-/N-acetylgalactosamine-binding lectin of *Entamoeba* in parasite virulence and differentiation. *Glycobiology* **15**: 53R–59R.
- Goldman CG, Mitchell HM. (2010). Helicobacter spp. other than *Helicobacter pylori*. *Helicobacter* **15**: 69–75.
- Hamady M, Lozupone C, Knight R. (2009). Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* **4**: 17–27.
- Hynes SO, Teneberg S, Roche N, Wadstrom T. (2003). Glycoconjugate binding of gastric and enterohepatic *Helicobacter* spp. *Infect Immun* **71**: 2976–2980.
- Johnsen JM, Levy GG, Westrick RJ, Tucker PK, Ginsburg D. (2008). The endothelial-specific regulatory mutation, Mvfw1, is a common mouse founder allele. *Mamm Genome* **19**: 32–40.
- Johnsen JM, Teschke M, Pavlidis P, McGee BM, Tautz D, Ginsburg D *et al.* (2009). Selection on cis-regulatory variation at B4galnt2 and its influence on von Willebrand factor in house mice. *Mol Biol Evol* **26**: 567–578.
- Karlsson KA. (1995). Microbial recognition of target-cell glycoconjugates. *Curr Opin Struct Biol* **5**: 622–635.
- Kawamura YI, Kawashima R, Fukunaga R, Hirai K, Toyama-Sorimachi N, Tokuhara M *et al.* (2005). Introduction of Sd(a) carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res* **65**: 6220–6227.
- Kobayashi M, Lee H, Nakayama J, Fukuda M. (2009). Roles of gastric mucin-type O-glycans in the pathogenesis of *Helicobacter pylori* infection. *Glycobiology* **19**: 453–461.
- Krivan HC, Roberts DD, Ginsburg V. (1988). Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc beta 1-4Gal found in some glycolipids. *Proc Natl Acad Sci USA* **85**: 6157–6161.
- Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA* **102**: 11070–11075.
- Linnenbrink M, Johnsen JM, Montero I, Brzezinski CR, Harr B, Baines JF. (2011). Long-term balancing selection at the blood group-related gene B4galnt2 in the genus *Mus* (Rodentia; Muridae). *Mol Biol Evol* **28**: 2999–3003.
- Lo PL, Cabuy E, Chiricolo M, Dall’Olio F. (2003). Molecular cloning of the human beta1,4 N-acetylgalactosaminyltransferase responsible for the biosynthesis of the Sd(a) histo-blood group antigen: the sequence predicts a very long cytoplasmic domain. *J Biochem* **134**: 675–682.
- Meng D, Newburg DS, Young C, Baker A, Tonkonogy SL, Sartor RB *et al.* (2007). Bacterial symbionts induce a FUT2-dependent fucosylated niche on colonic epithelium via ERK and JNK signaling. *Am J Physiol Gastrointest Liver Physiol* **293**: G780–G787.
- Miller-Podraza H, Lanne B, Angström J, Teneberg S, Milh MA, Jovall PA *et al.* (2005). Novel binding epitope for *Helicobacter pylori* found in Neolacto carbohydrate chains. *J Biol Chem* **280**: 19695–19703.
- Mohlke KL, Purkayastha AA, Westrick RJ, Smith PL, Petryniak B, Lowe JB *et al.* (1999). Mvfw, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell* **96**: 111–120.
- Montiel MD, Krzewinski-Recchi MA, Delannoy P, Harduin-Lepers A. (2003). Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: evidence for an unusual extended cytoplasmic domain. *Biochem J* **373**: 369–379.
- Morton JA, Pickles MM, Vanhegan RI. (1988). The Sda antigen in the human kidney and colon. *Immunol Invest* **17**: 217–224.

- Needleman SB, Wunsch CD. (1970.). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443–453.
- O'Hara AM, Shanahan F. (2006). The gut flora as a forgotten organ. *EMBO Rep* **7**: 688–693.
- Petnicki-Ocwieja T, Hrnčir T, Liu YJ, Biswas A, Hudcovic T, Tlaskalova-Hogenova H et al. (2009). Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc Natl Acad Sci USA* **106**: 15813–15818.
- Pinheiro J, Bates D, Debroy S, Sarkar D, The R Core Team. (2009). *Linear and Nonlinear Mixed Effects Models*.
- Presley LL, Wei B, Braun J, Borneman J. (2010). Bacteria associated with immunoregulatory cells in mice. *Appl Environ Microbiol* **76**: 936–941.
- Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.
- Price MN, Dehal PS, Arkin AP. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*.
- Rehman A, Sina C, Gavrilova O, Häsler R, Ott S, Baines JF et al. (2011). Nod2 is essential for temporal development of intestinal microbial communities. *Gut* **60**: 1354–1362.
- Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP et al. (2003). Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *J Biol Chem* **278**: 46337–46348.
- Savage DC. (1977). Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* **31**: 107–133.
- Schloss PD. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One* **4**: e8230.
- Sekirov I, Russell SL, Antunes LC, Finlay BB. (2010). Gut microbiota in health and disease. *Physiol Rev* **90**: 859–904.
- Shirato H, Ogawa S, Ito H, Sato T, Kameyama A, Narimatsu H et al. (2008). Noroviruses distinguish between type 1 and type 2 histo-blood group antigens for binding. *J Virol* **82**: 10756–10767.
- Stecher B, Hardt WD. (2010). Mechanisms controlling pathogen colonization of the gut. *Curr Opin Microbiol* **14**: 82–91.
- Stuckenholtz C, Lu L, Thakur P, Kaminski N, Bahary N. (2009). FACS-assisted microarray profiling implicates novel genes and pathways in zebrafish gastrointestinal tract development. *Gastroenterology* **137**: 1321–1332.
- Sweeney JD, Novak EK, Reddington M, Takeuchi KH, Swank RT. (1990). The RIIS/J inbred mouse strain as a model for von Willebrand disease. *Blood* **76**: 2258–2265.
- Timothy LC, Martin JB. (2009). Helicobacter pylori in health and disease. *Gastroenterology* **136**: 1863–1873.
- Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**: 6ra14.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z et al. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**: 1844–1854.
- Ye J, Lee JW, Presley LL, Bent E, Wei B, Braun J et al. (2008). Bacteria and bacterial rRNA genes associated with the development of colitis in IL-10(–/–) mice. *Inflamm Bowel Dis* **14**: 1041–1050.



This work is licensed under the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)