Resource

Cell Host & Microbe

Genetic Determinants of the Gut Microbiome in UK Twins

Graphical Abstract

Authors

Julia K. Goodrich, Emily R. Davenport, Michelle Beaumont, ..., Jordana T. Bell, Andrew G. Clark, Ruth E. Ley

Correspondence

rel222@cornell.edu

In Brief

Does host genotype shape the microbiome? Goodrich et al. present a gut microbiome analysis of 1,126 twin pairs, which extends the association between host genetics and select bacterial taxa. Lactase nonpersistence was linked to higher levels of Bifidobacteria. Other gene/microbe links relate to diet and barrier defense.

Highlights

- 16S rRNA-based analysis of the gut microbiome in 1,126 twin pairs
- Heritable bacterial taxa are temporally stable
- *Bifidobacterium* associates with lactase gene variants; formate production links to blood pressure
- Gene-microbe links involve genes related to diet, metabolism, olfaction, and defense

Goodrich et al., 2016, Cell Host & Microbe *19*, 731–743
May 11, 2016
http://dx.doi.org/10.1016/j.chom.2016.04.017

CellPress

Cell Host & Microbe
Resource

Genetic Determinants of the Gut Microbiome in UK Twins

Julia K. Goodrich,[1] Emily R. Davenport,[1] Michelle Beaumont,[2] Matthew A. Jackson,[2] Rob Knight,[3] Carole Ober,[4] Tim D. Spector,[2] Jordana T. Bell,[2] Andrew G. Clark,[1] and Ruth E. Ley[1,5,*]
[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14850, USA
[2]Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK
[3]Departments of Pediatrics and Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA
[4]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[5]Department of Microbiome Science, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.chom.2016.04.017

SUMMARY

Studies in mice and humans have revealed intriguing associations between host genetics and the microbiome. Here we report a 16S rRNA-based analysis of the gut microbiome in 1,126 twin pairs, a subset of which was previously reported. Tripling the sample narrowed the confidence intervals around heritability estimates and uncovered additional heritable taxa, some of which are validated in other studies. Repeat sampling of subjects showed heritable taxa to be temporally stable. A candidate gene approach uncovered associations between heritable taxa and genes related to diet, metabolism, and olfaction. We replicate an association between *Bifidobacterium* and the lactase (*LCT*) gene locus and identify an association between the host gene *ALDH1L1* and the bacteria SHA-98, suggesting a link between formate production and blood pressure. Additional genes detected are involved in barrier defense and self/non-self recognition. Our results indicate that diet-sensing, metabolism, and immune defense are important drivers of human-microbiome co-evolution.

INTRODUCTION

The gut microbiome is acquired from the environment starting at birth. Diversity builds up over the first few years of life and, thereafter, is largely shaped by environmental factors such as age, diet, lifestyle, hygiene, and disease state. Patterns of microbiome diversity across human populations can be pronounced, with stark contrasts in membership and structure observed in stool collected from people living on different continents (De Filippo et al., 2010; Martínez et al., 2015; Yatsunenko et al., 2012). Such population differences in microbiomes may be largely driven by diet and lifestyle, but genetic ancestry cannot at present be excluded as an important shaper of microbial diversity. Within a population, genetically encoded differences between hosts, such as those that dictate food preferences, aspects of immunity, or gut physiology, in principle could impact diversity and structure of the microbiome.

For several years, twin studies have provided the basis for asking whether genetic variation in the host associates with genetic variation in the microbiome. The basic premise is the following: if the genetics of the host influence a phenotype, measures of the phenotype will be more similar within monozygotic (MZ) twin pairs compared to within dizygotic (DZ) twin pairs, assuming that since they are reared together, twins within a pair experience similar environments. Given a large enough twin population, these differences can be modeled to yield estimates of heritability, which is the proportion of variance in the phenotype that can be attributed to genetic differences between hosts.

Early twin studies that used culture-based methods (Van de Merwe et al., 1983) and fingerprinting of fecal 16S rRNA gene amplicons (Stewart et al., 2005; Zoetendal et al., 2001) lacked the sample sizes required to estimate heritabilities of specific gut microbiota, but their results did suggest that some part of the gut microbiome was heritable. Two subsequent larger twin studies cast some ambiguity on this conclusion, however (Turnbaugh et al., 2009; Yatsunenko et al., 2012). Using a larger number of twins (~50 pairs of the Missouri twin registry) and 16S rRNA gene sequencing, both studies reported that MZ twin pairs had slightly more similar microbiomes compared to DZ twin pairs, though the differences were not statistically significant. These studies underscored that environmental factors are paramount in shaping the microbiome, yet they also hinted at a small host genetic effect.

In Goodrich et al. (2014), we increased the power to detect heritable microbiota with a 16S rRNA gene-based analysis of 416 twin pairs. As observed for the Missouri twin studies, the UniFrac distances for MZ twins were slightly less than for DZ twins, but due to the greater sample size, the difference reached statistical significance. Importantly, the greater number of subjects allowed us to identify taxa with significant heritabilities. Among common taxa (those found in at least 50% of samples), the most heritable was the family Christensenellaceae (phylum Firmicutes), which forms the hub of a co-occurrence network composed of other heritable taxa and is enriched in lean subjects. Experiments using fecal transplants from twin donors into germ-free mice demonstrated that *Christensenella minuta* addition to a microbiome deficient in Christensenellaceae limited adiposity gain in the recipient mice. Together, these

observations supported the notion that heritable microbes could drive the human phenotypes with which they associate.

A few studies have used quantitative measures derived from the gut microbiome as phenotypes in genetic association studies. Blekhman et al. (2015) performed an analysis of human genetic data generated as a byproduct of the Human Microbiome Project (HMP) for 93 subjects (Consortium, 2012). Without accounting for population structure, ethnicity, or geographic stratification, Blekhman reported correlations between the first principal component of the host genetic variation and the first principal coordinate of the stool UniFrac distances (Blekhman et al., 2015). A second study using the HMP's human DNA data showed that mitochondrial haplotypes correlated with abundances of specific microbiota in stool (Ma et al., 2014). Both results reflect ancestry effects on the composition of the microbiome.

Diet differs among subjects and can rapidly alter microbial community composition (David et al., 2014), casting some doubt over the use of relative abundance data in genetic association studies. Davenport et al. (2015) circumvented this issue with a genome-wide analysis of the fecal microbiome composition in the Hutterites, a founder population that lives and eats communally (Davenport et al., 2015, 2014). These analyses revealed a suite of heritable taxa and highlighted links between microbial taxa and genes involved in barrier defense and immunity.

Studies in mice further reduce the environmental influences on the microbiome. Benson et al. (2010) performed the first quantitative trait loci (QTL) mapping study, using 645 mice from an advanced intercross line (Benson et al., 2010). QTL analysis detected 13 genetic loci that were significantly associated with microbial abundances and five additional suggestive loci. Many of the QTL were in regions with genes that play roles in the immune system. Subsequent studies in mice have confirmed some of these results (Leamy et al., 2014; McKnite et al., 2012). Comparisons across mouse and human studies highlight some common themes: genetic control of certain bacterial taxa (e.g., *Turicibacter*) and associations with immune-related genes.

Here, we report heritability and gene-association analyses for the TwinsUK cohort. We tripled the size of our initial dataset of Goodrich et al. (2014), now including 3,261 fecal samples from 2,731 individuals. This includes 489 dizygotic (DZ) twin pairs, 637 monozygotic (MZ) twin pairs, and 530 samples collected at a second time point. The goals of this study are to (1) calculate the heritabilities of specific components of the gut microbiota and (2) associate abundances of microbes with host gene alleles through candidate gene analysis and genome-wide association. Using this expanded dataset, we improve heritability estimates for taxa previously identified as heritable, expand the list of heritable taxa, and identify associations between a subset of the heritable taxa and candidate genes related to diet and immunity.

RESULTS

Effect of an Expanded Dataset on Heritability Results

In our initial study of 416 twin pairs, we estimated heritability for 909 widely shared taxa (Operational taxonomic units [OTU] and taxonomy bins). Heritability analysis revealed that 5.3% of taxa had a heritability (A, see Experimental Procedures) greater than 0.2 with 95% confidence intervals that did not overlap zero. In the expanded sample set, we examined 945 taxa, and this percentage was increased to 8.8%.

We found a significant correlation between the heritability estimates from the tripled sample set compared to the initial set ($r = 0.57$, p value $< 2.2 \times 10^{-16}$; Figures 1 and S1 and Table S1). Members of the Bacteroidetes phylum were generally not heritable, while taxa with heritability estimates $A > 0.2$ belong to the Firmicutes, Actinobacteria, Tenericutes, and Euryarchaeota (Figure 1). The most heritable taxon in the expanded dataset was again the bacterial family Christensenellaceae ($A = 0.42$, 95% CI = 0.25–0.48). Most of the taxa with a heritability greater that 0.2 in the initial set maintained a high heritability. The main difference between sample sets is a decrease in the size of the confidence intervals around the heritability estimates. The ranking of heritabilities was slightly altered; notably, the genus *Turicibacter* jumped in the ranked heritability list from $A = 0.21$ (with a 95% confidence interval that overlapped zero) to $A = 0.39$ (95% CI = 0.28–0.45; Figure 1).

Additional taxa reached a heritability threshold of 0.2 when we tripled the size of the dataset (Figure 1). Importantly, the genus *Methanobrevibacter*, the predominant human gut archaeon, reached a significant level of heritability. Other taxa included the genera *SMB53* and *Actinomyces* and the families Clostridiaceae and Peptococcaceae. Measures of alpha-diversity were also heritable in the expanded dataset (Table S2; PD whole tree: $A = 0.37$, 95% CI = 0.17–0.44). Alpha-diversity is positively correlated with the relative abundances of the genus *Methanobrevibacter* ($r = -0.37$, Benjamini-Hochberg (BH) adjusted p value $< 1 \times 10^{-10}$), the family Christensenellaceae ($r = -0.59$, BH adjusted p value $< 1 \times 10^{-10}$), and other members of its co-occurrence consortium (Figure 2). Alpha-diversity is also highly correlated with the first principal coordinate (PC) of principal coordinates analysis (PCoA) of the unweighted UniFrac distances ($r = -0.90$, BH adjusted p value $< 1 \times 10^{-10}$; Figure 2), most likely because its associated taxa drive differentiation of samples along PC1.

Heritable Microbes and Temporal Stability

We collected repeat samples for a subset of the twins (530 samples were spaced a mean 946 ± SEM 15 days; Figure 3). Although non-heritable taxa displayed a wide range of stability, highly heritable taxa were associated with higher levels of stability and absent from the low end of the stability gradient.

Candidate Genes and SNPs

2,139 individuals (including 472 MZ twin pairs and 418 DZ twins pairs) had both genotype and 16S rRNA gene sequence data. We assembled a list of 37 sets of candidate genes or SNPs with suspected roles in shaping the microbiome (Tables 1, S3, and S4). Included are genes implicated in previous mouse or human studies, genetic variants and genes associated with gut microbiome-related diseases, and genes involved in taste reception. To reduce the burden of multiple testing, we performed association analysis on 20 heritable taxa ($A > 0.2$; Figure 1). Table 2 details the SNP-microbe associations for the significant candidate sets. We validated the association between *Bifidobacterium* and *LCT* reported by Blekhman et al. (2015) (*LCT* permutation p value < 0.001). Additionally, the candidate set analysis revealed an association between *Bifidobacterium* and a SNP within the gene *RABGAP1*, which is in linkage disequilibrium (LD) with variants in

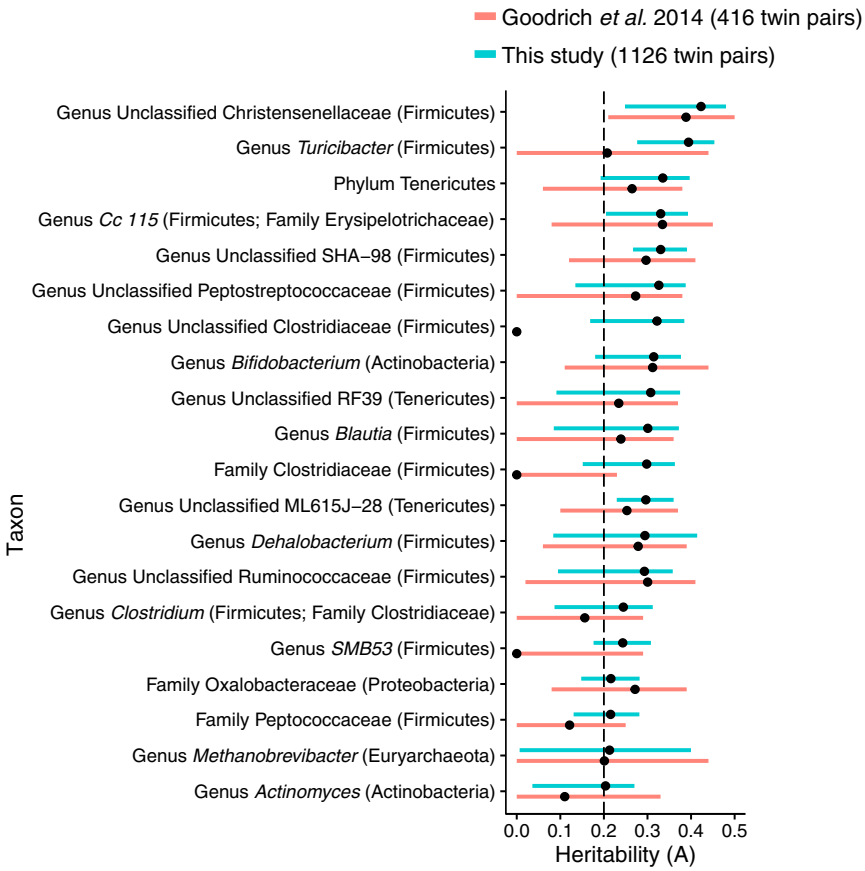


Figure 1. Expansion of the TwinsUK Dataset Results in Similar Microbial Abundance Heritability Estimates with Narrower Confidence Intervals

Each point represents the heritability estimated as the proportion of variance in the microbial abundances (OTU abundances collapsed by taxonomic classification) that can be attributed to genetic effects (A). The bars show the 95% confidence intervals around the heritability estimates. Bars are colored by the dataset: pink indicates heritability estimates reported in Goodrich et al. (2014) (171 MZ, 245 DZ), and blue indicates heritability estimates using an additional 710 twin pairs (blue; 637 MZ, 489 DZ twin pairs). The figure includes only taxa that are present in at least 50% of the TwinsUK participants and have A > 0.2 in the increased dataset. We also excluded any taxon that was highly correlated ($r > 0.9$) with another taxon at a lower taxonomic level. The *Clostridium* genus is known to be polyphyletic; however, further examination revealed that the genus *Clostridium* within the Clostridiaceae family consists of mostly one Greengenes OTU 4465124 that is shared by 82.6% of the samples. See also Figure S1 and Table S1.

LCT (Figure 4). Given that these genes are in LD, it is not possible to know which of the two is truly driving the association.

We observed a significant association between heritable taxa and the olfactory receptor gene OR6A2 (linked to cilantro soapy taste; Eriksson et al., 2012; permutation p value = 0.011), which was driven by the Erysipelotrichaceae *Cc 115* genus. The gene *CD36* had a significant association with the genus *Blautia* (permutation p value = 0.009). *CD36* is involved in a variety of functions, including long-chain fatty acid tasting on the tongue (Silverstein and Febbraio, 2009). We also observed an association between the order SHA-98 (a component of the Christensenellaceae consortium) and the gene *ALDH1L1* in the one-carbon metabolism gene set (permutation p value = 0.006). Within a SNP set associated with inflammatory bowel disease (Jostins et al., 2012), we detected a suggestive association (permutation p value = 0.056) with the genus *SMB53* and variants in the gene *GNA12* involved in barrier defense and associated with ulcerative colitis (Lees et al., 2011).

GWAS with the Expanded TwinsUK Dataset

The above testing with candidate genes was sufficiently successful to suggest that genome-wide association analysis may reveal additional associations. When using the full dataset (945 taxa and 1,300,091 SNPs), no associations reached study-wide significance (correction for all tests; Table S5). The strongest GWA signal among the 20 heritable taxa used in the gene set analysis was the "unclassified Clostridiaceae" with SNP rs10055309 in the gene *SLIT3* (p value 1.20×10^{-8}, BH adjusted p value = 0.016). This taxon bin consists of several OTUs, but the OTU making up the majority of sequence counts is Greengenes OTU 4434334. The "unclassified Clostridiaceae" taxon bin is also highly correlated with the Clostridiaceae family (r = 0.89), which is also associated with the same *SLIT3* SNP (p value = 4.21×10^{-6}, BH adjusted p value = 0.14).

Interestingly, GWA revealed that the SNP with the strongest association to *Bifidobacterium* lies within the gene *R3HDM1* (p value = 4.38×10^{-8}, BH adjusted p value = 0.057). However, several genes in this locus are in strong LD, including *LCT* and the gene *RABGAP1* mentioned above (Figure 4). Although our genotype dataset does not contain the SNP (rs4988235) that correlates directly with lactase persistence, it is in LD with the SNP (rs1446585; 1000 Genomes Phase 3 GBR population r^2 = 0.89) that, in our dataset, has the greatest association with *Bifidobacterium*. In our analysis, individuals who carry the rs1446585(G) allele associated with lactase nonpersistence have higher levels of Bifidobacteria in stool.

For validation, we examined the association of the lactase persistence-associated SNP (rs4988235) with the relative abundance of *Bifidobacterium* using data from a GWAS on fecal microbiome composition in the Hutterites in two seasons (Davenport et al., 2015). As observed in the TwinsUK dataset, nonpersisters have significantly higher levels of Bifidobacteria (winter p value = 0.02, summer p value = 0.001, seasons combined p value = 4×10^{-5}; Figures 4E–4G).

GWAS for Beta-Diversity Measures

To examine the association between genetic variation and beta-diversity metrics, including Bray Curtis dissimilarity, weighted UniFrac distance, and unweighted UniFrac distance, we applied

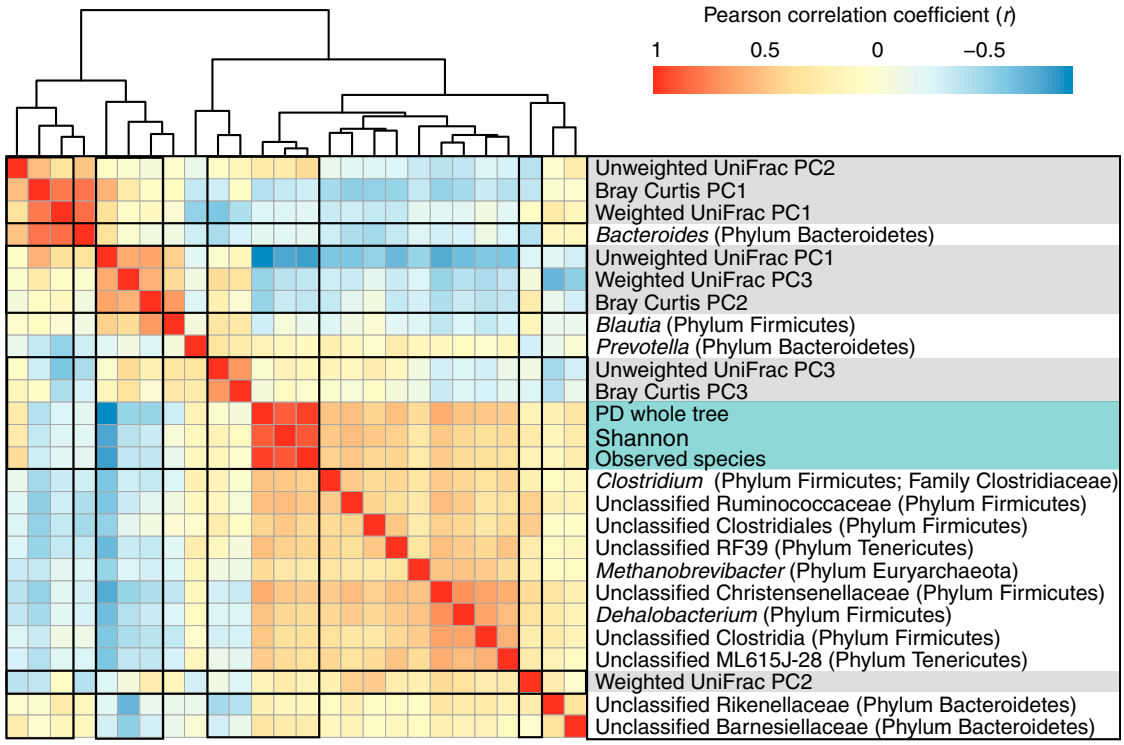


Figure 2. Alpha- and Beta-Diversity Are Correlated with Several Microbiota that Have a Heritability Estimate > 0.2

Heatmap showing the correlation structure between the alpha-diversity metrics, the first three principal coordinates of the beta-diversity distance matrices, and the taxa that are correlated at $|r| > 0.5$ with one of the alpha- or beta-diversity metrics in the heatmap. See also Table S2.

microbiomeGWAS (Hua et al., 2015) to a subset of our dataset that included only unrelated individuals (n = 1,248; see Experimental Procedures). A SNP (rs563779) within the gene *UHRF2* was associated with weighted UniFrac distance at a study-wide significance threshold (p value = 9.77×10^{-9}), while two SNPs in LD on chromosome 4 are associated with Bray Curtis dissimilarity at a relaxed threshold (rs9997915, p value = 1.44×10^{-8}; rs1593554, p value = 1.16×10^{-8}).

Associations with Imputed Gene Expression

We utilized a gene expression-based approach using the PrediXcan framework in which SNPs are used to infer gene expression across a range of tissues (Gamazon et al., 2015). Association testing is then performed between taxon abundance and the imputed gene expression values for each tissue. In addition to reducing the multiple-testing burden, a strength of this approach is the interpretability of the results, as imputed gene expression is a biologically plausible intermediate phenotype through which inter-individual genetic variation may act to influence microbiome composition in the gut. We used PrediXcan to obtain imputed gene expression values for 40 tissues and performed association testing for each of those tissues with each of the taxa. While no genetic associations met study-wide significance, inferred expression of *SIGLEC15* was associated with the abundance of *Akkermansia* at a tissue-wide significance level in transverse colon ($p = 6.21 \times 10^{-9}$). Additionally, several other taxa were associated to genes at a significance threshold of $p = 5 \times 10^{-8}$ (Table S6).

DISCUSSION

We report heritability and genome-wide association analyses for fecal microbiome data obtained from 1,126 twin pairs from the TwinsUK registry. Heritability estimates were broadly similar when tripling the number of subjects from our previous report (Goodrich et al., 2014), and the confidence intervals were narrower with the expanded dataset, making the estimates more robust. The number of heritable taxa was increased, and alpha-diversity was also found to be heritable.

To put microbiome heritabilities in context, they are a little lower than those of other complex traits measured in the same population: systolic blood pressure (0.51; Menni et al., 2013), anxiety (0.44; Davies et al., 2015), and serum Vitamin D (0.43; Hunter et al., 2001). In a meta-analysis paper of twin studies, the average heritability for diseases of the digestive system is 0.31 (Polderman et al., 2015). Although on the low side, these results for 16S rRNA gene data from stool suggest that heritabilities for the microbiome may be refined and increased with sample types other than stool and/or other data types.

We showed that heritable taxa are temporally stable over long periods, suggesting a lesser effect of environmental factors on their relative abundances. This could be explained by a strong dependence of these heritable taxa on host physiology or metabolism. Deeper analysis of strains within these taxa may reveal adaptations to specific host genotypes, as has been shown for *Helicobacter pylori* strains in the stomach and host genotypes (Suerbaum and Josenhans, 2007). Several of the heritable and

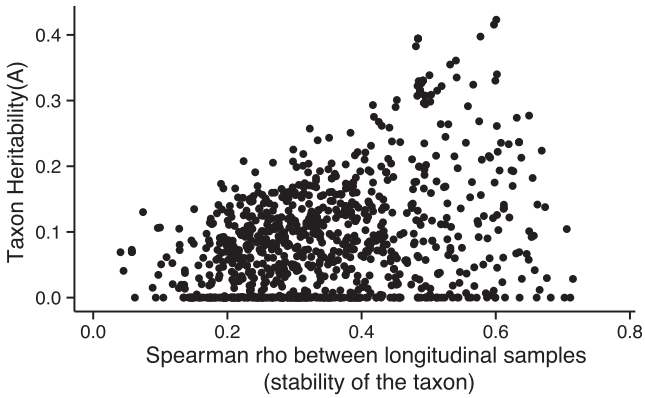


Figure 3. Heritable Taxa Are among the Most Stable Taxa in the TwinsUK Dataset

Heritability (A) of the microbial abundances (OTU abundances and abundances collapsed by taxonomic classification) is plotted against the Spearman correlation between longitudinal samples (530 individuals with samples collected at a second time point spaced 946 ± 15 days [mean ± SEM]).

stable taxa have also been associated with genetic loci in mouse QTL studies. This suggests that the host-microbial connections are widespread in mammals and may be ancient associations.

The twins in this study were genotyped, allowing for tests of association between SNPs and microbiome traits. Because there is a set of candidate genes whose function gives them a strong prior expectation to be relevant to gut function, we focused on these tests first. Candidate gene association tests uncovered associations between heritable taxa and genes related to diet, carbohydrate metabolism and olfaction, barrier defense, and self/non-self recognition. Associations of genes involved in barrier defense and immunity have also been observed in mouse QTL studies (Benson et al., 2010, McKnite et al., 2012, Org et al., 2015) and plant microbiome GWAS (Horton et al., 2014).

Bifidobacterium and the *LCT* Gene Locus

The association for which we have the highest confidence is between *Bifidobacterium* and the *LCT* gene locus (Figure 4). *Bifidobacterium* is heritable in the TwinsUK population, the HMP, the Hutterites, and mouse studies (Figure 5). *LCT* encodes lactase, the enzyme that hydrolyzes lactose in the upper gastrointestinal tract. In mammals, lactase production ceases at weaning, but lactase persistence has evolved independently several times in a subset of humans. Lactase nonpersisters may experience lactose intolerance when lactose reaches the large intestine and is fermented. The direction of the genetic association shows lactase nonpersisters harbor higher levels of *Bifidobacterium*. Since Bifidobacteria are members of the large intestine and metabolize lactose, one can speculate that lactase persisters harbor lower levels of Bifidobacteria because of low lactose availability, whereas nonpersisters who ingest milk provide relatively more lactose to Bifidobacteria.

Turicibacter and the Peptostreptococcaceae

The genus *Turicibacter* is the second most heritable taxon in our analysis. *Turicibacter* is the sole genus of the family Turicibacteriaceae (Firmicutes), and its relative abundance is correlated with Peptostreptococcaceae ($r = 0.66$). The host genetic associations with these taxa are validated in humans and in mice (Figure 5). Benson et al. (2010) associated *Turicibacter* with a QTL on MMU7 and Peptostreptococcaceae with a QTL on MMU1, but so far we have not detected any allelic associations with *Turicibacter* by GWA either in the candidate gene sets tested or by pathway enrichment.

Both taxa are active members of the small intestinal microbiome (Bouhnik et al., 1999; Kamada et al., 2013; Leimena et al., 2013; Looft et al., 2014; Oh et al., 2012). *Turicibacter* has been linked to host immunity (Kellermayer et al., 2011; Dimitriu et al., 2013; Rausch et al., 2015) and is positively associated with inflammation (Rausch et al., 2015). *Turicibacter* has been isolated from blood of human patients, and analysis of its genome reveals genes consistent with a pathobiont lifestyle (Cuív et al., 2011). Oddly, *Turicibacter* has also been noted as highly abundant at later stages of dead pig decomposition (Yang et al., 2012). *Turicibacter* thus appears to be a pathobiont with the genes necessary to exploit host inflammation and that can also subsist on the host after death.

Akkermansia

We noted an association between *Akkermansia* and predicted tissue-specific expression levels of the gene *SIGLEC15*. This gene encodes a sialic acid-binding immunoglobulin-like lectin that participates in the discrimination of self and non-self and is highly conserved across mammals (Angata et al., 2007; Macauley et al., 2014). The physical niche of the genus *Akkermansia* is the intestinal mucus where it uses mucins as a carbon source (Derrien et al., 2004). Sialic acid is one of the outermost carbohydrate decorations of mucin. *Akkermansia* can cleave sialic acids; its genome encodes four neuraminidase enzymes responsible for the removal of sialic acids from glycoconjugates (Huang et al., 2015). *SIGLEC15* is expressed on macrophages and DCs in humans and is reported to be expressed in ileum tips in mice (Sommer et al., 2015). Another gene reported by Sommer et al. (2015) to be expressed at the villus tips of the small intestine, *PLD1*, was linked genetically with *Akkermansia* in the Hutterites (Davenport et al., 2015). Given that both studies associate *Akkermansia* with genes expressed in its physical niche space, an investigation of host genetic interaction with this important bacterium is warranted.

Other Gene-Taxon Associations

We provide here for future reference other gene-microbe associations from this study requiring validation. The genus *Blautia* is heritable in this dataset and in the mouse study of O'Connor et al. (2014) (Figure 5). *Blautia* associated here with *CD36*, a gene involved in fat sensing on the tongue and the promotion of absorption of long-chain fatty acids in the gut (Silverstein and Febbraio, 2009). Interestingly, Omega-3 fatty acid feeding has been shown to increase levels of *Blautia* in mice (Myles et al., 2014), and mice associated with a simplified community including *Blautia* show upregulation of *CD36* (Woting et al., 2014).

Our strongest signal in the GWAS was between unclassified Clostridiaceae and the SNP rs10055309 in the gene *SLIT3*. SLIT3 is a secreted protein with high expression in the glandular cells of the stomach, duodenum, and small intestine

Table 1. SNP and Gene Candidate Sets with Suspected Roles in Shaping the Gut Microbial Community

Candidate Set Name	Gene or SNP[a]	Number of Genes[b]	Number of SNPs[c]	Min p Value of Tested SNPs[d]	Min FDR Adjusted p Value of Tested SNPs[e]	Candidate Set Permutation p Value[f]	References
Gastric Cancer Risk	SNP		25	8.9×10^{-4}	0.446	0.286	Mocellin et al., 2015
Gastric Cancer Risk	Gene	29	611	8.0×10^{-5}	0.555	0.307	
Inflammatory Bowel Disease	SNP		140	2.4×10^{-5}	0.045	0.056	Jostins et al., 2012
Inflammatory Bowel Disease	Gene	109	4,337	6.8×10^{-6}	0.126	0.203	
Rheumatoid Arthritis	SNP		57	6.9×10^{-4}	0.443	0.511	Okada et al., 2014
Rheumatoid Arthritis	Gene	129	5,555	2.4×10^{-5}	0.599	0.585	
Type 2 Diabetes	SNP		65	0.001	0.527	0.703	Mahajan et al., 2014
Type 2 Diabetes	Gene	78	7,731	4.0×10^{-6}	0.602	0.214	
Non-HLA Associations with Cholestatic Disorders	Gene	87	5,350	2.5×10^{-5}	0.795	0.537	Hirschfield et al., 2013
Cilantro Soapy Taste	SNP		1	0.084	0.446	0.667	Eriksson et al., 2012
Cilantro Soapy Taste	Gene	1	13	9.1×10^{-5}	0.024	0.011	
Bitter Taste	Gene	21	136	0.001	0.717	0.367	Bachmanov and Beauchamp, 2007
Lipid Taste	Gene	1	51	6.1×10^{-5}	0.022	0.009	
Salty Taste	Gene	5	45	0.01	0.965	0.988	
Sour Taste	Gene	7	1,048	1.6×10^{-4}	0.492	0.659	
Sweet Taste	Gene	2	15	0.014	0.508	0.816	
Signal of Selection in Ancient Eurasians	SNP		9	0.003	0.603	0.37	Mathieson et al., 2015
Signal of Selection in Ancient Eurasians	Gene	43	1,122	3.5×10^{-6}	0.063	0.022	
Innate Immunity Genes under Selection	Gene	42	2,217	5.1×10^{-5}	0.425	0.565	Deschamps et al., 2016
LCT	Gene	1	21	3.5×10^{-6}	0.001	<0.001	
Davenport SNPs	SNP		187	1.7×10^{-4}	0.369	0.311	Davenport et al., 2015
NOD2	Gene	1	17	0.007	0.609	0.493	
Org Mouse eQTL	Gene	12	728	0.001	0.987	0.949	Org et al., 2015
Spor et al. (2011) Previous Microbiome Associations	Gene	27	461	6.2×10^{-4}	0.859	0.827	Spor et al., 2011
SNPs Associated with AMY1 CNV	SNP		7	0.007	0.615	0.467	Carpenter et al., 2015
AMY Locus	Gene	1	8	0.007	0.884	0.353	
Caffeine Metabolism	SNP		3	0.035	0.998	0.765	Yang et al., 2010; Josse et al., 2012
Caffeine Metabolism	Gene	3	22	0.015	0.978	0.918	
Choline Metabolism	Gene	10	254	3.8×10^{-5}	0.067	0.065	Smallwood et al., 2016
One Carbon Metabolism	Gene	28	886	1.0×10^{-6}	0.017	0.006	Lim et al., 2007
Alcohol Dehydrogenase	Gene	7	177	0.003	0.999	0.861	

(*Continued on next page*)

Table 1. *Continued*

Candidate Set Name	Gene or SNP[a]	Number of Genes[b]	Number of SNPs[c]	Min p Value of Tested SNPs[d]	Min FDR Adjusted p Value of Tested SNPs[e]	Candidate Set Permutation p Value[f]	References
Blood Lipid Traits	SNP		86	2.1×10^{-7}	3.7×10^{-04}	<0.001	Teslovich et al., 2010
Sugar Transporters	Gene	2	66	0.006	0.789	0.68	
Tight Junctions	Gene	2	207	3.7×10^{-4}	0.262	0.32	
TLRs NLRPs ILs	Gene	15	430	6.0×10^{-4}	0.845	0.773	
SCFA receptors	Gene	2	6	0.005	0.409	0.378	
Sphingolipids	Gene	3	284	2.7×10^{-4}	0.445	0.335	

See also Tables S3 and S4.

[a]Indicates if the candidate set contains the SNPs associated with the disease/trait or all SNPs within 5 kb of any gene found to be associated with the disease/trait.

[b]Number of genes in the candidate gene set that were tested for association in the TwinsUK dataset.

[c]Number of SNPs in the candidate set that were tested for association in the TwinsUK dataset.

[d]Minimum p value (GEMMA likelihood ratio test) for all associations between the 20 heritable taxa and the SNPs in the candidate set.

[e]Minimum FDR adjusted p value within the candidate set (correction for all SNPs in the candidate set and the 20 heritable taxa).

[f]p value for the association of the candidate set with the 20 heritable taxa calculated from 1,000 permutations (see Experimental Procedures for details).

(Sanz-Pamplona et al., 2014). *SLIT3* was reported as frequently methylated in colorectal cancers (Dickinson et al., 2004) and has been shown to exhibit dysregulated exons in colorectal adenomas (Pesson et al., 2014). Notably, *SLIT3* expression was shown to be upregulated in colon crypts when germ-free mice are conventionalized (Sommer et al., 2015). Until these gene-microbe links are validated, they are essentially hypothetical; however, it is interesting that the gene products are expressed in the gut epithelium.

The Christensenellaceae Consortium

The family Christensenellaceae remained the most highly heritable taxon. Taxa we reported as heritable and which had relative abundances correlated to those of the Christensenellaceae family (SHA-98, *Dehalobacterium*, RF39, ML615J−28) were again heritable in the larger dataset.

The genus *Methanobrevibacter*, a member of the Christensenellaceae consortium, also reached a significant level of heritability in this analysis. The abundances of taxa within the heritable consortium correlate positively with alpha-diversity, also heritable. Whether the methanogenesis and associated fermentative dynamics drive a more diverse microbiota or higher levels of methanogens and species richness are both results of other factors, such as a high-fiber diet, remains to be ascertained.

Methanobrevibacter smithii is the dominant methanogen in the human gut, and MZ twins have previously been shown to have greater concordance for the carriage of this archaeon than DZ twins (Hansen et al., 2011). Studies across mammal species (Hackstein and van Alen, 1996) and within bovine lines (Roehe et al., 2016) have also suggested that host genetics influence levels of methanogens. Methanogen carriage has been associated with leanness and with a better metabolic profile in obese humans (Le Chatelier et al., 2013). Given a fermentable diet, a genetic predisposition for high methanogen levels may lead to methane production and an overall metabolism that partly explain the leaner phenotype associated with the Christensenellaceae consortium.

We observed an association between the order SHA-98 (a member of the Christensenellaceae consortium) and *ALDH1L1*, which codes for an aldehyde dehydrogenase involved in formate oxidation. Formate is produced endogenously by the host and is also a fermentation product that acts as a major interspecies electron carrier between syntrophs (Boone et al., 1989; Pavlostathis et al., 1990). Production of formate by a *Clostridium* sp. has been shown to promote growth of a syntroph in a gnotobiotic mouse model (Rey et al., 2013). *Methanobrevibacter smithii* can use formate in methanogenesis (Miller et al., 1982). Formate concentration in urine has been shown to be significantly correlated with systolic and diastolic blood pressure in 4,630 subjects in Asia, USA, and Europe (Holmes et al., 2008). Holmes et al. (2008) also reported that urinary formate and Na^+ excretion were positively correlated and, given the importance of Na^+ in blood pressure, suggested an unrecognized role for formate in its regulation. Hypertension is a major risk factor for stroke, and a SNP in *ALDH1L1* has also been associated with ischemic stroke (Xie et al., 2013). Interestingly, the microbiomeGWAS tool, which uses beta-diversity distances in GWAS, also revealed an association between weighted UniFrac

Table 2. Most Significant SNP × Heritable Taxon Associations within Each Significant Candidate Set

Candidate Set Name	Gene or SNP[a]	Gene Set p Value[b]	SNP Driving the Candidate Set Association[c]	Gene[d]	Heritable Taxon Driving the Candidate Set Association[e]
Inflammatory Bowel Disease	SNP	0.056	rs1182182	*GNA12*	*SMB53* (family Clostridiaceae)
Cilantro Soapy Taste	Gene	0.011	rs1506977	*OR6A2*	*Cc 115* (family Erysipelotrichaceae)
Lipid Taste	Gene	0.009	rs1360741	*CD36*	*Blautia*
Signal of Selection in Ancient Eurasians	Gene	0.022	rs2164210	*LCT*	*Bifidobacterium*
LCT	Gene	<0.001	rs2164210	*LCT*	*Bifidobacterium*
One-Carbon Metabolism	Gene	0.006	rs2276731	*ALDH1L1*	Unclassified SHA-98
Blood Lipid Traits	SNP	<0.001	rs6730157	*RAB3GAP1*	*Bifidobacterium*

[a]Indicates if the candidate set contains the SNPs associated with the disease/trait or all SNPs within 5 kb of a gene found to be associated with the disease/trait.
[b]p value for the association of the candidate set with the 20 heritable taxa calculated from 1,000 permutations (see Experimental Procedures for details).
[c]SNP within the candidate set that has the most significant association with one of the 20 heritable taxa.
[d]Gene containing the SNP that is driving the candidate set association.
[e]Heritable taxon that has the most significant association with one of the SNPs within the candidate set.

and a SNP in the gene *UHRF2*, which has been identified as linked to ischemic stroke as well. Our results suggest the hypothesis that the Christensenellaceae-methanogen consortium regulates the thermodynamics of fermentation in the gut, including formate production and consumption. This activity interacts with the host's own enzymatic activities to impact formate levels, with repercussions for blood pressure.

Conclusion

Our results highlight gene-microbe interactions from recent evolutionary adaptation to diet, its sensing, and metabolism in the gut. In the case of the *Bifidobacterium-LCT* link, host genetics most likely shape the microbiome through diet preference, which itself is heritable. These signals contrast with the immune-related genes uncovered in studies where diet is controlled. For links to immune genes to be detected in human populations where diet is unrestricted, very large numbers of subjects may be necessary.

Most microbiome-genetic studies use 16S rRNA gene diversity as the phenotype, and other data types, such as metagenomic data, will focus on functions that can be shared across taxa. However, any sequence-based results are limited by their nature as relative abundances. Ultimately, results of sequence-based studies require validation with quantitative measures such as qPCR. The gene-microbe associations uncovered here will require validation across multiple studies, but they support incorporating measures of diet and microbiome in studies seeking a genetic basis for disease risk susceptibility, particularly for diseases involving chronic conditions of over- and under-nutrition.

EXPERIMENTAL PROCEDURES

Sample Collection

All work involving human subjects was approved by the Cornell University IRB (Protocol ID 1108002388). Fecal samples were collected as described previously (Goodrich et al., 2014; Jackson et al., 2016). The sample set consists of 3,261 fecal samples from 2,731 individuals (530 individuals sampled at a second time point).

Sample Processing and 16S rRNA Gene Sequencing

DNA extraction, amplification of the V4 hypervariable region of the 16S rRNA gene (primers 515F and 806R), purification, and pooling were performed on all fecal samples as previously described (Goodrich et al., 2014). The pooled amplicons were sequenced using the Illumina Miseq platform with 2 × 250 bp paired-end sequencing.

16S rRNA Gene Data Analysis

Mate-pair merging, de-multiplexing, quality control, and OTU picking were performed using QIIME version 1.8 (Quantitative Insights Into Microbial Ecology; Caporaso et al., 2010). Taxa filtering, covariate regression, and details on the correlations between diversity metrics and taxa are described in the Supplemental Information.

Heritability Calculations

The ACE model (Eaves et al., 1978) was used to estimate heritability for 945 taxa, three alpha-diversity metrics, and the first three PCs from PCoA of each of the beta-diversity metrics (details are provided in the Supplemental Information). p value adjustment to correct for all 945 taxa was done using the Benjamini-Hochberg algorithm.

Stability of Microbiota

530 individuals (including 125 DZ twin pairs and 69 MZ twin pairs) supplied two or more serial fecal samples spaced 3–1,632 days apart (946 ± 15 days [mean ± SEM]). For the stability analysis, two samples were randomly chosen from each individual. Spearman correlation coefficients between longitudinal samples were calculated for each taxon.

Host Genetic Association Analyses

The participants in this study were previously genotyped and the genotype data imputed using IMPUTE version 2 (Howie et al., 2009) and quality checked as previously described (Moayyeri et al., 2013). GEMMA (v.0.94) was used to perform SNP-microbe association tests (Zhou and Stephens, 2012). Association analysis details for the candidate gene and SNP sets, taxon genome-wide analysis, and imputed gene expression are provided in the Supplemental Information. The tool microbiomeGWAS (Hua et al., 2015) was used to perform a GWAS on the beta-diversity measures. Only one twin per family was included in this analysis, for a total sample size of 1,248 individuals.

Bifidobacterium Validation in the Hutterite Dataset

16S rRNA gene sequencing analysis and genotyping is described in the original study (Davenport et al., 2015). Briefly, the dataset consists of data from the Hutterites sampled during two seasons (summer, n = 91; winter,

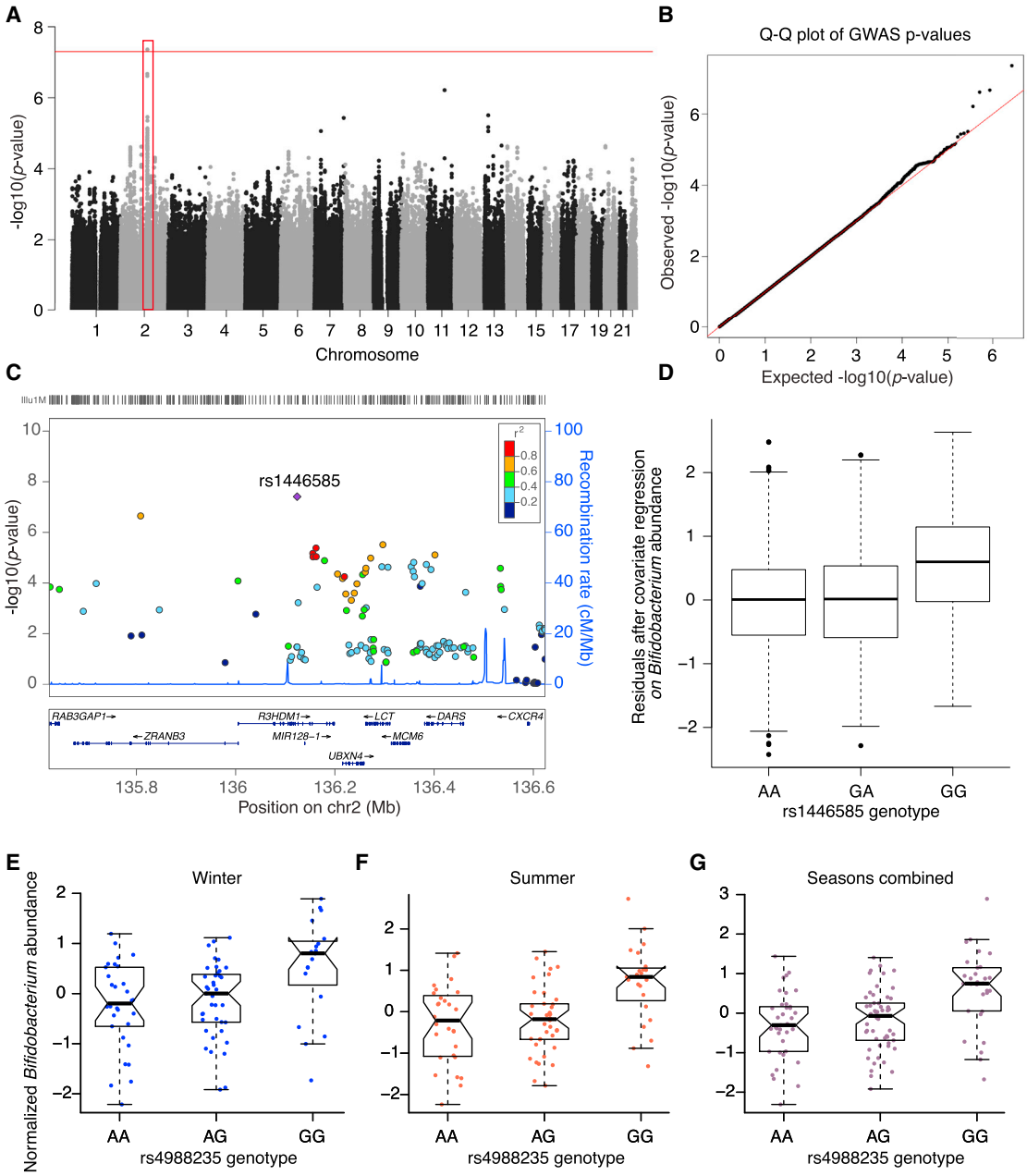


Figure 4. Relative Abundances of the Heritable Genus *Bifidobacterium* Are Associated with Genetic Variants in the Genomic Locus Containing the Gene *LCT*

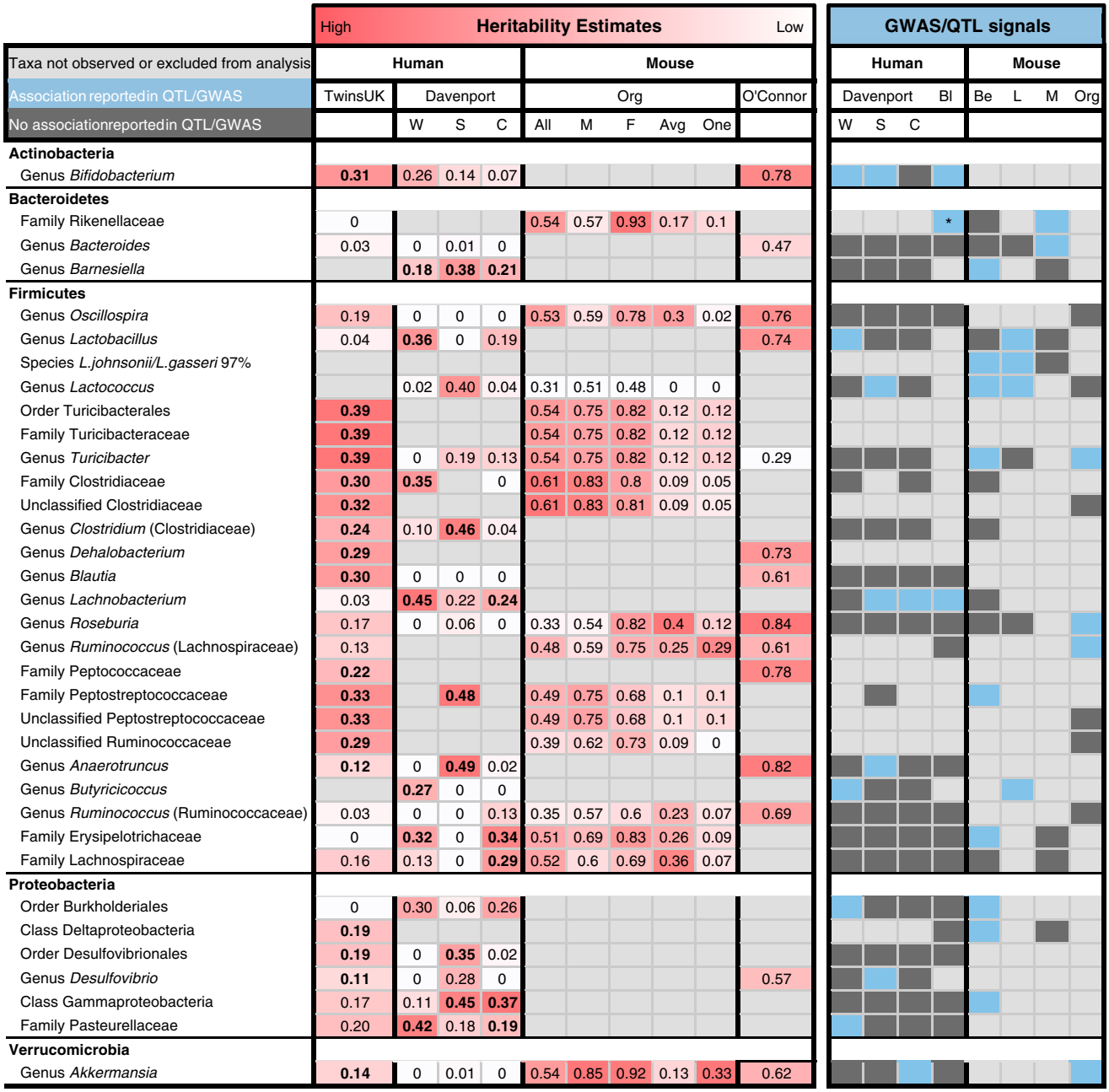
(A) Genome-wide Manhattan plot: the x axis represents the chromosome and position along the chromosome, and the y axis is –log of the p value for the association of the SNP (each dot) with the genus *Bifidobacterium*. The red box highlights the associated locus on chromosome 2 that contains the gene *LCT*.

(B) Quantile-quantile (Q-Q) plots of the p values. The Q-Q plot measures deviation from the expected distribution of p values. The diagonal (red) line represents the expected (null) distribution.

(C) Close-up plots of a 1 Mb window around the SNP with the highest association. The coloring of the points represents r^2 between a SNP and the SNP with the highest association in the locus (rs1446585, denoted by the purple diamond); r^2 is calculated from the 1000 Genomes data on the CEU population.

(D) Box-plot of *Bifidobacterium* normalized abundances within each genotype at the most strongly associated SNP (rs1446585; p value = 4.38×10^{-8}). The y axis depicts the residuals from linear regression of the Box-Cox transformed abundances with the covariates (the number of 16S rRNA gene sequences per sample, age, gender, shipment date, collection method (postal or visit), and ID of technician performing DNA extraction).

(E–G) Normalized *Bifidobacterium* abundance within each genotype at the lactase persistence-associated SNP (rs4988235) in the Hutterite dataset. (E) Winter samples (p value = 0.02). (F) Summer samples (p value = 0.001). (G) Seasons combined (p value = 4×10^{-5}). See also Tables S5 and S6.



Alistipes was in high correlation with Rikenellaceae so only the genus was tested

Figure 5. Comparison of Taxa Estimated as Heritable or Linked to Genes in at Least Two Human GWA or Mouse QTL Studies

The color gradient over the heritability estimates ranges from the lowest heritability estimate (white) to the highest heritability estimate (red) in the given study. For the TwinsUK heritability estimates, the bold values indicate heritability estimates with a 95% confidence interval not overlapping 0. The estimates for Davenport et al. (2015) are the proportion of variance explained (PVE) estimates ("chip heritability"). We report the winter (W), summer (S), and seasons combined (C) datasets. For the Davenport study, bold values indicate heritability estimates with a standard error not overlapping 0. For Org et al. (2015) we report results using all mice (All), just males (M), just females (F), an average per strain (Avg), and a single mouse per strain (One). No significance value was reported for the Org et al. (2015) and O'Connor et al. (2014) heritability estimates. The coloring over the QTL/GWAS studies indicates if each taxon had a significant association (blue) or not (dark gray) in the given study. Light gray indicates that the taxon was not observed in the given study or was excluded from the study analysis for other reasons. The figure shows taxa with an association from the following gut microbiota GWA and QTL studies: Davenport et al. (2015) (Davenport), Blekhman et al. (2015) (Bl), Benson et al. (2010) (B), Leamy et al. (2014) (L), McKnite et al. (2012) (M), and Org et al. (2015) (Org). See Supplemental Information for more details.

n = 93; seasons combined, n = 127). GEMMA was used to perform the association analysis between *Bifidobacterium* and the SNP rs4988235.

Cross-Study Comparisons of Taxa Influenced by Host Genetics

We compiled heritability and QTL/GWAS results from published human and mouse studies and summarize these results in Figure 5. For the human studies, this includes the current study, the Hutterites (Davenport et al., 2015), and the HMP (Blekhman et al., 2015). The mouse studies examined are the advanced intercross lines (Benson et al., 2010; Leamy et al., 2014), the Hybrid Mouse Diversity Panel (Org et al., 2015), collaborative cross/diversity outbred mapping panels (O'Connor et al., 2014), and recombinant inbred strains (McKnite et al., 2012). Details are provided in the Supplemental Information.

ACCESSION NUMBERS

The accession number for the 16S rRNA gene sequences reported in this paper is European Nucleotide Archive (ENA): ERP015317.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2016.04.017.

AUTHOR CONTRIBUTIONS

R.E.L. and A.G.C. supervised the study, and J.T.B. and T.D.S. helped design the study and provided comments and discussion. J.T.B. and T.D.S. oversaw collection of samples; J.K.G., R.E.L., and E.R.D. oversaw microbial data generation; J.K.G. and E.R.D performed the analysis with contributions from C.O., R.K., M.B., and M.A.J.; and J.K.G., E.R.D., and R.E.L. prepared the manuscript, with comments from A.G.C., T.D.S., J.T.B., and M.A.J.

ACKNOWLEDGMENTS

We thank the participants of the TwinsUK registry. We thank Jessica Sutter, Qiaojuan Shi, Jillian Waters, Gabriela Surdulescu, Martin Voetsch, and Ayrun Nessa. This work was funded by NIH RO1 DK093595, DP2 OD007444, a David and Lucile Packard Foundation Fellowship (R.E.L.), the Arnold and Mabel Beckman Foundation (R.E.L.), The Cornell Center for Comparative Population Genomics (R.E.L. and J.K.G.), a National Science Foundation Graduate Fellowship (J.K.G.), The Wellcome Trust, and the European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

Received: March 7, 2016
Revised: April 20, 2016
Accepted: April 22, 2016
Published: May 11, 2016

REFERENCES

Angata, T., Tabuchi, Y., Nakamura, K., and Nakamura, M. (2007). Siglec-15: an immune system Siglec conserved throughout vertebrate evolution. Glycobiology *17*, 838–846.

Bachmanov, A.A., and Beauchamp, G.K. (2007). Taste receptor genes. Annu. Rev. Nutr. *27*, 389–414.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. USA *107*, 18933–18938.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., and Clark, A.G. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Boone, D.R., Johnson, R.L., and Liu, Y. (1989). Diffusion of the interspecies electron carriers H(2) and formate in methanogenic ecosystems and its implications in the measurement of K(m) for H(2) or formate uptake. Appl. Environ. Microbiol. *55*, 1735–1741.

Bouhnik, Y., Alain, S., Attar, A., Flourié, B., Raskine, L., Sanson-Le Pors, M.J., and Rambaud, J.C. (1999). Bacterial populations contaminating the upper gut in patients with small intestinal bacterial overgrowth syndrome. Am. J. Gastroenterol. *94*, 1327–1331.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Carpenter, D., Dhar, S., Mitchell, L.M., Fu, B., Tyson, J., Shwan, N.A.A., Yang, F., Thomas, M.G., and Armour, J.A.L. (2015). Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. Hum. Mol. Genet. *24*, 3472–3480.

Consortium, H.; Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Cuív, P.Ó., Klaassens, E.S., Durkin, A.S., Harkins, D.M., Foster, L., McCorrison, J., Torralba, M., Nelson, K.E., and Morrison, M. (2011). Draft genome sequence of Turicibacter sanguinis PC909, isolated from human feces. J. Bacteriol. *193*, 1288–1289.

Davenport, E.R., Mizrahi-Man, O., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2014). Seasonal variation in human gut microbiome composition. PLoS ONE *9*, e90731.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. PLoS ONE *10*, e0140301.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Davies, M.N., Verdi, S., Burri, A., Trzaskowski, M., Lee, M., Hettema, J.M., Jansen, R., Boomsma, D.I., and Spector, T.D. (2015). Generalised anxiety disorder–A twin study of genetic architecture, genome-wide association and differential gene expression. PLoS ONE *10*, e0134865.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. USA *107*, 14691–14696.

Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. *54*, 1469–1476.

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E., and Quintana-Murci, L. (2016). Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. Am. J. Hum. Genet. *98*, 5–21.

Dickinson, R.E., Dallol, A., Bieche, I., Krex, D., Morton, D., Maher, E.R., and Latif, F. (2004). Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers. Br. J. Cancer *91*, 2071–2078.

Dimitriu, P.A., Boyce, G., Samarakoon, A., Hartmann, M., Johnson, P., and Mohn, W.W. (2013). Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. Environ. Microbiol. Rep. *5*, 200–210.

Eaves, L.J., Last, K.A., Young, P.A., and Martin, N.G. (1978). Model-fitting approaches to the analysis of human behaviour. Heredity (Edinb) *41*, 249–320.

Eriksson, N., Wu, S., Do, C.B., Kiefer, A.K., Tung, J.Y., Mountain, J.L., Hinds, D.A., and Francke, U. (2012). A genetic variant near olfactory receptor genes influences cilantro preference. Flavour *1*, 1–7.

Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., and Im, H.K.; GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Hackstein, J.H.P., and van Alen, T.A. (1996). Fecal methanogens and vertebrate evolution. Evolution *50*, 559–572.

Hansen, E.E., Lozupone, C.A., Rey, F.E., Wu, M., Guruge, J.L., Narra, A., Goodfellow, J., Zaneveld, J.R., McDonald, D.T., Goodrich, J.A., et al. (2011). Pan-genome of the dominant human gut-associated archaeon, Methanobrevibacter smithii, studied in twins. Proc. Natl. Acad. Sci. USA *108* (*Suppl 1*), 4599–4606.

Hirschfield, G.M., Chapman, R.W., Karlsen, T.H., Lammert, F., Lazaridis, K.N., and Mason, A.L. (2013). The genetics of complex cholestatic disorders. Gastroenterology *144*, 1357–1374.

Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., et al. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. Nature *453*, 396–400.

Horton, M.W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B.D., Subramanian, S., Vetter, M.M., Vilhjálmsson, B.J., Nordborg, M., Gordon, J.I., and Bergelson, J. (2014). Genome-wide association study of Arabidopsis thaliana leaf microbial community. Nat. Commun. *5*, 5320.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.

Hua, X., Song, L., Yu, G., Goedert, J.J., Abnet, C.C., Landi, M.T., and Shi, J. (2015). MicrobiomeGWAS: a tool for identifying host genetic variants associated with microbiome composition. bioRxiv.

Huang, K., Wang, M.M., Kulinich, A., Yao, H.L., Ma, H.Y., Martínez, J.E.R., Duan, X.C., Chen, H., Cai, Z.P., Flitsch, S.L., et al. (2015). Biochemical characterisation of the neuraminidase pool of the human gut symbiont Akkermansia muciniphila. Carbohydr. Res. *415*, 60–65.

Hunter, D., De Lange, M., Snieder, H., MacGregor, A.J., Swaminathan, R., Thakker, R.V., and Spector, T.D. (2001). Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J. Bone Miner. Res. *16*, 371–378.

Jackson, M.A., Goodrich, J.K., Maxan, M.-E., Freedberg, D.E., Abrams, J.A., Poole, A.C., Sutter, J.L., Welter, D., Ley, R.E., Bell, J.T., et al. (2016). Proton pump inhibitors alter the composition of the gut microbiota. Gut *65*, 749–756.

Josse, A.R., Da Costa, L.A., Campos, H., and El-Sohemy, A. (2012). Associations between polymorphisms in the AHR and CYP1A1-CYP1A2 gene regions and habitual caffeine consumption. Am. J. Clin. Nutr. *96*, 665–671.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IIBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Kamada, N., Chen, G.Y., Inohara, N., and Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. Nat. Immunol. *14*, 685–690.

Kellermayer, R., Dowd, S.E., Harris, R.A., Balasa, A., Schaible, T.D., Wolcott, R.D., Tatevian, N., Szigeti, R., Li, Z., Versalovic, J., and Smith, C.W. (2011). Colonic mucosal DNA methylation, immune response, and microbiome patterns in Toll-like receptor 2-knockout mice. FASEB J. *25*, 1449–1460.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al.; MetaHIT consortium (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Leamy, L.J., Kelly, S.A., Nietfeldt, J., Legge, R.M., Ma, F., Hua, K., Sinha, R., Peterson, D.A., Walter, J., Benson, A.K., and Pomp, D. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol. *15*, 552.

Lees, C.W., Barrett, J.C., Parkes, M., and Satsangi, J. (2011). New IBD genetics: common pathways with other diseases. Gut *60*, 1739–1753.

Leimena, M.M., Ramiro-Garcia, J., Davids, M., van den Bogert, B., Smidt, H., Smid, E.J., Boekhorst, J., Zoetendal, E.G., Schaap, P.J., and Kleerebezem, M. (2013). A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics *14*, 530.

Lim, U., Wang, S.S., Hartge, P., Cozen, W., Kelemen, L.E., Chanock, S., Davis, S., Blair, A., Schenk, M., Rothman, N., and Lan, Q. (2007). Gene-nutrient interactions among determinants of folate and one-carbon metabolism on the risk of non-Hodgkin lymphoma: NCI-SEER case-control study. Blood *109*, 3050–3059.

Looft, T., Allen, H.K., Cantarel, B.L., Levine, U.Y., Bayles, D.O., Alt, D.P., Henrissat, B., and Stanton, T.B. (2014). Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. ISME J. *8*, 1566–1576.

Ma, J., Coarfa, C., Qin, X., Bonnen, P.E., Milosavljevic, A., Versalovic, J., and Aagaard, K. (2014). mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. BMC Genomics *15*, 257.

Macauley, M.S., Crocker, P.R., and Paulson, J.C. (2014). Siglec-mediated regulation of immune cell function in disease. Nat. Rev. Immunol. *14*, 653–666.

Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C., Prokopenko, I., et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. *46*, 234–244.

Martínez, I., Stegen, J.C., Maldonado-Gómez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015). The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell Rep. *11*, 527–538.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS ONE *7*, e39191.

Menni, C., Mangino, M., Zhang, F., Clement, G., Snieder, H., Padmanabhan, S., and Spector, T.D. (2013). Heritability analyses show visit-to-visit blood pressure variability reflects different pathological phenotypes in younger and older adults: evidence from UK twins. J. Hypertens. *31*, 2356–2361.

Miller, T.L., Wolin, M.J., Conway de Macario, E., and Macario, A.J. (1982). Isolation of Methanobrevibacter smithii from human feces. Appl. Environ. Microbiol. *43*, 227–232.

Moayyeri, A., Hammond, C.J., Valdes, A.M., and Spector, T.D. (2013). Cohort Profile: TwinsUK and healthy ageing twin study. Int. J. Epidemiol. *42*, 76–85.

Mocellin, S., Verdi, D., Pooley, K.A., and Nitti, D. (2015). Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. Gut *64*, 1209–1219.

Myles, I.A., Pincus, N.B., Fontecilla, N.M., and Datta, S.K. (2014). Effects of parental omega-3 fatty acid intake on offspring microbiome and immunity. PLoS ONE *9*, e87181.

O'Connor, A., Quizon, P.M., Albright, J.E., Lin, F.T., and Bennett, B.J. (2014). Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm. Genome *25*, 583–599.

Oh, P.L., Martínez, I., Sun, Y., Walter, J., Peterson, D.A., and Mercer, D.F. (2012). Characterization of the ileal microbiota in rejecting and nonrejecting recipients of small bowel transplants. Am. J. Transplant. *12*, 753–762.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381.

Org, E., Parks, B.W., Joo, J.W.J., Emert, B., Schwartzman, W., Kang, E.Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., et al. (2015). Genetic and environmental control of host-gut microbiota interactions. Genome Res. *25*, 1558–1569.

Pavlostathis, S.G., Miller, T.L., and Wolin, M.J. (1990). Cellulose fermentation by continuous cultures of Ruminococcus albus and Methanobrevibacter smithii. Appl. Microbiol. Biotechnol. *33*, 109–116.

Pesson, M., Volant, A., Uguen, A., Trillet, K., De La Grange, P., Aubry, M., Daoulas, M., Robaszkiewicz, M., Le Gac, G., Morel, A., et al. (2014). A gene expression and pre-mRNA splicing signature that marks the adenoma-adenocarcinoma progression in colorectal cancer. PLoS ONE *9*, e87761.

Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709.

Rausch, P., Steck, N., Suwandi, A., Seidel, J.A., Künzel, S., Bhullar, K., Basic, M., Bleich, A., Johnsen, J.M., Vallance, B.A., et al. (2015). Expression of the blood-group-related gene B4galnt2 alters susceptibility to Salmonella infection. PLoS Pathog. *11*, e1005008.

Rey, F.E., Gonzalez, M.D., Cheng, J., Wu, M., Ahern, P.P., and Gordon, J.I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. Proc. Natl. Acad. Sci. USA *110*, 13582–13587.

Roehe, R., Dewhurst, R.J., Duthie, C.-A., Rooke, J.A., McKain, N., Ross, D.W., Hyslop, J.J., Waterhouse, A., Freeman, T.C., Watson, M., and Wallace, R.J. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. PLoS Genet. *12*, e1005846.

Sanz-Pamplona, R., Berenguer, A., Cordero, D., Molleví, D.G., Crous-Bou, M., Sole, X., Paré-Brunet, L., Guino, E., Salazar, R., Santos, C., et al. (2014). Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer. Mol. Cancer *13*, 46.

Silverstein, R.L., and Febbraio, M. (2009). CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. Sci. Signal. *2*, re3.

Smallwood, T., Allayee, H., and Bennett, B.J. (2016). Choline metabolites: gene by diet interactions. Curr. Opin. Lipidol. *27*, 33–39.

Sommer, F., Nookaew, I., Sommer, N., Fogelstrand, P., and Bäckhed, F. (2015). Site-specific programming of the host epithelial transcriptome by the gut microbiota. Genome Biol. *16*, 62.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. *9*, 279–290.

Stewart, J.A., Chadwick, V.S., and Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. J. Med. Microbiol. *54*, 1239–1242.

Suerbaum, S., and Josenhans, C. (2007). Helicobacter pylori evolution and phenotypic diversification in a changing host. Nat. Rev. Microbiol. *5*, 441–452.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Van de Merwe, J.P., Stegeman, J.H., and Hazenberg, M.P. (1983). The resident faecal flora is determined by genetic characteristics of the host. Implications for Crohn's disease? Antonie van Leeuwenhoek *49*, 119–124.

Woting, A., Pfeiffer, N., Loh, G., Klaus, S., and Blaut, M. (2014). Clostridium ramosum promotes high-fat diet-induced obesity in gnotobiotic mouse models. MBio *5*, e01530–14.

Xie, W., Wood, A.R., Lyssenko, V., Weedon, M.N., Knowles, J.W., Alkayyali, S., Assimes, T.L., Quertermous, T., Abbasi, F., Paananen, J., et al.; MAGIC Investigators; DIAGRAM Consortium; GENESIS Consortium; RISC Consortium (2013). Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. Diabetes *62*, 2141–2150.

Yang, A., Palmer, A.A., and de Wit, H. (2010). Genetics of caffeine consumption and responses to caffeine. Psychopharmacology (Berl.) *211*, 245–257.

Yang, S.-H., Hong, S.H., Cho, S.B., Lim, J.S., Bae, S.E., Ahn, H., and Lee, E.Y. (2012). Characterization of microbial community in the leachate associated with the decomposition of entombed pigs. J. Microbiol. Biotechnol. *22*, 1330–1335.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

Zoetendal, E.G., Akkermans, A.D.L., Akkermans-van Vliet, W.M., de Visser, J.A.G.M., and de Vos, W.M. (2001). The host genotype affects the bacterial community in the human gastrointestinal tract. Microb. Ecol. Health Dis. *13*, 129–134.