npg

ORIGINAL ARTICLE

Gene-targeted metagenomic analysis of glucan-branching enzyme gene profiles among human and animal fecal microbiota

Sunghee Lee[1], Brandi Cantarel[2], Bernard Henrissat[3], Dirk Gevers[4], Bruce W Birren[4], Curtis Huttenhower[4,5] and GwangPyo Ko[1]

[1]Department of Environmental Health, Graduate School of Public Health, Seoul National University, Gwanak-gu, Seoul, South Korea; [2]Baylor Health, Institute for Immunology Research, Dallas, TX, USA; [3]Aix-Marseille Université, Centre National de la Recherche Scientifique, Marseille, France; [4]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA, USA and [5]Harvard School of Public Health, Boston, MA, USA

Glycoside hydrolases (GHs), the enzymes that breakdown complex carbohydrates, are a highly diversified class of key enzymes associated with the gut microbiota and its metabolic functions. To learn more about the diversity of GHs and their potential role in a variety of gut microbiomes, we used a combination of 16S, metagenomic and targeted amplicon sequencing data to study one of these enzyme families in detail. Specifically, we employed a functional gene-targeted metagenomic approach to the 1-4-α-glucan-branching enzyme (gBE) gene in the gut microbiomes of four host species (human, chicken, cow and pig). The characteristics of operational taxonomic units (OTUs) and operational glucan-branching units (OGBUs) were distinctive in each of hosts. Human and pig were most similar in OTUs profiles while maintaining distinct OGBU profiles. Interestingly, the phylogenetic profiles identified from 16S and gBE gene sequences differed, suggesting the presence of different gBE genes in the same OTU across different vertebrate hosts. Our data suggest that gene-targeted metagenomic analysis is useful for an in-depth understanding of the diversity of a particular gene of interest. Specific carbohydrate metabolic genes appear to be carried by distinct OTUs in different individual hosts and among different vertebrate species' microbiomes, the characteristics of which differ according to host genetic background and/or diet.

Introduction

Glycoside hydrolases (GHs) are one of the dominant enzyme classes in the gut microbiota. They mainly hydrolyze the glycosidic linkage of glycosides and have a crucial role in the digestion of complex carbohydrates such as those found in the plant cell wall. GHs are classified into over 130 sequence-based families, and many of these families group together enzymes of differing substrate specificity (Cantarel *et al.*, 2009). The spectrum of GHs in the gut has been shown to be differential among mammals with distinct feeding habits (Muegge

et al., 2011). Other studies have aimed to associate these enzymes with human diet (Ley *et al.*, 2008) and disease (Turnbaugh *et al.*, 2008). However, these studies essentially bin sequences into GH families and compare the relative abundance of these families across samples, without accounting for the sequence diversity of enzymes within a family. As such binning groups together enzymes of different substrate (or product) specificities, understanding enzyme function in microbes and microbial communities requires targeted analysis within specific GH families.

Of the over 130 GH families, family GH13 possesses one of the broadest distributions among the gut microbiota of highly diverse vertebrate hosts, ranging from birds to mammals (including humans). This family groups together enzymes covering over 20 EC numbers; therefore, inclusion in the family does not directly imply precise functional assignment. However, it has been shown that division of

family GH13 into subfamilies allows a much improved sequence-to-function correspondence (Stam *et al.*, 2006). Family GH13 is present in several starch utilization system (Sus)-like systems, such as SusA and SusG, that carry out the degradation of starch or other carbohydrates (Koropatkin *et al.*, 2012). Bacterial glycogen-branching enzyme (gBE), represented in subfamily 9 of family GH13, is important in glycogen storage and metabolism (van der Maarel and Leemhuis, 2012). Catalytically, the gBE subfamily performs a transglycosylation reaction in order to form α-1,6-glycosidic linkages in glycogen (Kumar *et al.*, 1986) and as such has a crucial role in the glucose cycle, as glycogen energy reserves can be mobilized rapidly to meet a sudden need for glucose (Henning *et al.*, 1996). For example, glycogen storage disease type IV, also known as Andersen's disease, involves a genetic defect in human gBE that causes serious damage to the liver and heart and eventually leads to death because of storage of abnormal glycogen (Fyfe *et al.*, 1992).

Current approaches to studying the gut microbiota have typically provided only indirect evidence regarding the diversity and functional activities of GHs, and gBE in particular, among microbial residents of the vertebrate gut (Muegge *et al.*, 2011). In most studies, either 16S rRNA gene-targeted analysis or metagenomic shotgun sequencing is used to investigate the phylogenetic and metabolic profiles of the gut microbiota (Turnbaugh *et al.*, 2008). Analysis based on 16S rRNA gene sequences is informative regarding the composition and diversity of a microbial community; however, its ability to elucidate the functional gene content and metabolic pathways is limited (Kalyuzhnaya *et al.*, 2008). Metagenomics can provide detailed information on global metabolic functions; however, it typically achieves limited depth of coverage for any one pathway or gene family of interest, and as a result has low ability to detect small polymorphisms or rare genes (Kalyuzhnaya *et al.*, 2008). Neither of these approaches has thus yet provided detailed information on the GH profile of the human gut.

Here, we thus applied a functional gene-targeted metagenomic approach, which overcomes these limitations by amplifying and sequencing a specific gene family of interest in greater depth (Iwai *et al.*, 2009). Specifically, we first assessed both the operational taxonomic units (OTUs) and global metagenomes of the fecal microbiota of human and three animal species (chicken, cow and pig). Based on these microbial and GH profiles, we amplified and sequenced the *gBE* gene families from 69 total samples from these four species using gBE-targeted primers. These data relate the diversity of the gene family to that of the overall microbial community (Iwai *et al.*, 2009), as the function and diversity of *gBE* genes in gut microorganisms was previously uncharacterized (Preiss, 1984). We found that GH and phylogenetic profiles vary independently among individual hosts and across these four species, with alpha- and beta-diversity profiles of OTUs versus operational glucan-branching units (OGBUs), separating host subsets in different ways. This suggests that specific OTUs in the gut may contain different carbohydrate metabolic genes with characteristics selected according to species' genetic complements and individual hosts' backgrounds and diets.

Materials and methods

Study subjects and collection of fecal samples
Human stool samples were collected from patients in two hospitals participating in the Korean healthy twin cohort study in Seoul and Busan, South Korea, as described previously (Sung *et al.*, 2006). Human subjects aged 30–60 years (58.4 ± 14.9) with a BMI in the normal range ($24.8 \pm 4.0 \, \mathrm{kg \, m^{-2}}$) and who had not taken any antibiotics for at least 1 month prior to sample collection were included in the study. Each of the animal fecal samples were collected from 53 separate farms located in the Geongki province, South Korea (Supplementary Table S1).

Nucleic acid extraction and pyrosequencing of partial 16S rRNA genes
Total DNA was extracted from fecal samples using the bead-beating method, as described previously (Turnbaugh *et al.*, 2008). The V1–V3 regions of the 16S rRNA gene was amplified from extracted total DNA with the help of PCR using primers described previously (Turnbaugh *et al.*, 2008). The forward primer (5′-GCCTTGCCAGCCCGCTCAGTCAGAGT TTGATCCTGGCTCAG-3′) comprised Primer B and the broadly conserved bacterial Primer 8F from 454 Life Sciences (Branford, CT, USA), joined by a four-base linker (TCAG). The reverse primer (Huse *et al.*, 2008) (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAGNN NNNNNNNNNATTACCGCGGCTGCTGGAGT-3′) contained Primer A and the bacterial Primer 518R from 454 Life Sciences, a TCAG linker between the barcode and rRNA primer, and a unique 12-bp error-correcting barcode used to tag each PCR product (designated by NNNNNNNNNNNN) (Hamady *et al.*, 2008). PCR amplification was performed in a total volume of 25 μl containing 2.5 μl of $10 \times$ G-spin PCR buffer (Cosmogene Tech., Seoul, Korea), 0.5 mM modified forward and reverse primers and 100 ng of gel-purified (Qiagen, Valencia, CA, USA) template DNA. The prepared PCR suspension was denatured at 94 °C for 5 min, then amplified using 35 cycles of 94 °C for 45 s, 60 °C for 30 s and 72 °C for 90 s.

Primer design and PCR amplification of the gBE gene followed by 454 pyrosequencing
To design conserved primers for amplification of the partial *gBE* gene, we analyzed 152 protein sequences (Bacteria/GH family 13/GO; EC 2.4.1.18; 1-4-alpha-glucan-branching enzyme) in the UniProt database (Wu *et al.*, 2006). From sequences that were

multiple-aligned using ClustalW (Thompson *et al.*, 2002), the gBE-F and gBE-R primer sets, corresponding to 297S–301Q and 481W–486M of gBE from *Escherichia coli* K-12, were designed to amplify a 569-bp PCR amplicon. The presence of this amplicon was confirmed in amplified total DNA extracted from human, chicken, cow and pig fecal samples. To confirm that this primer set amplified a diverse spectrum of *gBE* genes, an experiment using denaturing gradient gel electrophoresis was carried out as a pre-test (Supplementary Figure S1). The *gBE* genes resolved by denaturing gradient gel electrophoresis were sequenced and identified by comparison with the NCBI and UniProt databases. All denaturing gradient gel electrophoresis bands were confirmed to be gBE. These findings confirm amplification of a broad spectrum of diverse *gBE* genes from all fecal samples of the four host species using the newly developed primer set.

The *gBE* gene was amplified using the same nucleic acid suspension used for the 16S rRNA gene. The forward primer (5′-CCATCTCATCCCTG CGTGTCTCCGACTCAGNNNNNNNNNNTCNTGGG GNTACCAG-3′) contained Primer A and the targeted gene primer gBE-F from Life Sciences, a TCAG linker between the barcode and gBE-F primer, and a unique 10-bp error-correcting barcode used to tag each PCR product (designated by NNNNNNNNNN). The reverse primer (5′-CCTATCCCCTGTGTGC CTTGGCAGTCTCAGCATCCANCCCATRTTCCA-3′) comprised Primer B and the conserved primer gBE-R from 454 Life Sciences, joined by a four-base linker (TCAG). The PCR conditions for the *gBE* gene were as follows: initial denaturation at 94 °C for 5 min, then 35 cycles of amplification at 94 °C for 30 s, 53 °C for 60 s and 72 °C for 7 min. The products were pooled and purified using AMPure magnetic purification beads (Agencourt Bioscience, Beverly, MA, USA), and quantified using a bisBenzimide H assay. Aliquots of each product were incubated for 5 min at room temperature in TNE reagent (10 mM Trizma-HCl, 100 mM NaCl, 1 mM EDTA and 50 ng ml^{-1} freshly prepared bisBenzimide H; pH 8.1; Sigma, St Louis, MO, USA). Sample fluorescence was determined using a fluorometer or plate reader (excitation 365 nm, emission 460 nm). DNA concentration was determined relative to a standard curve constructed using *E. coli* DNA (Sigma). Multiple pools, each containing equimolar amounts of PCR product, were assembled for pyrosequencing analysis using 454 FLX Titanium platform (Roche, Branford, CT, USA).

Metagenomic sequencing
In addition to 16S rRNA and *gBE* gene sequencing, metagenomic sequencing of each host fecal sample was performed using Illumina Hi-Seq 2000 to generate paired end reads of 101 bp, producing on average 6.8 GB (× 2) reads for each of the pooled species fecal samples. Five stool samples from the same host were pooled for each metagenomic sequencing process.

Bioinformatic analysis of 16S rRNA, gBE gene and metagenomic sequences
Computational analyses of 16S rRNA and *gBE* genes were initially implemented with the help of Mothur (Schloss *et al.*, 2009) using a protocol derived from the Human Microbiome Project. The average sequence length was estimated to be ~500 bp using 454 FLX Ti pyrosequencing, as described previously (Balzer *et al.*, 2010). Reads with at least 400 nucleotides (nt) were trimmed and checked for chimerism (Edgar *et al.*, 2011). We obtained consensus OTU clusters and representative sequences using abundant OTU (Ye, 2010). Representative sequences and the OTU table were used for further analysis with the QIIME pipeline as detailed above (Caporaso *et al.*, 2010). Our gBE-targeted PCR, which produced 569-bp amplicons, did not cover the entire PCR product in the raw sequencing file generated by 454 FLX Ti pyrosequencing. Thus, we trimmed the raw data of the *gBE* gene with the forward primer using only Mothur (Schloss *et al.*, 2009). The final 16S rRNA gene set included 322 309 reads (66.0%) of 488 309 initial reads, and the *gBE* sequences secured 84 128 reads (51.9%) of 162 133 initial reads.

Microbial classification based on 16S rRNA gene sequences was performed using the ribosomal database project classifier naive Bayesian algorithm (Wang *et al.*, 2007). Taxonomic identities of the phylotypes were assigned using the ribosomal database project taxonomic annotation. Complete sequences were aligned by nearest-alignment space termination, with >75% identity based on a non-chimeric core set ⩾1250 nt in length (DeSantis *et al.*, 2006) and filtered by Lanemask to remove columns comprised of only gaps (Lane, 1991) before building the tree. The *gBE* gene was multiple-aligned using ClustalW (Thompson *et al.*, 2002) and filtered by the removal of columns comprised of only gaps with a self parameter. A heatmap was produced using MultiExperiment Viewer (Howe *et al.*, 2011) for comparisons between hosts by searching for significant genes according to representative sequences (OGBU) from *gBE* genes. Both OTU and OGBU groupings were determined with a precision of 0.03 using the furthest-neighbor algorithm. We used network-based analyses to analyze the relationships between OGBUs and OTUs using Cytoscape ver. 2.6.1 (Shannon *et al.*, 2003). LEfSe was used to identify the microbiological markers associated with the host species by LDA effect size (4.5 for 16S rRNA and 4.0 for gBE) via one-against-all option as the strategy for multiclass analysis (Goecks *et al.*, 2010). Phylogenetic trees were produced using the Fasttree method, and alpha- and beta diversity were measured as described previously (Price *et al.*, 2010).

We assessed microbial alpha diversity using the Chao1 measure (Chao, 1984) and the differences among 16S and *gBE* genes by determining beta diversity using unweighted UniFrac (Lozupone and Knight, 2005). Communities of both 16S rRNA and *gBE* genes among hosts were ordinated using

principal coordinates analysis and the UniFrac metric. Differences in diversity among species was determined by analysis of variance and *post-hoc* tests using the statistical analysis program SPSS, version 19.0 (SPSS, Inc., an IBM Company, Chicago, IL, USA). To compare the taxonomic distributions identified using either 16S rRNA or *gBE* gene sequences, we mapped the latter against HMP reference genome sequences. We obtained a set of 1275 reference microbial genomes, which excluded eukaryote, virus, archaea and plasmid genomes, from the human microbiome project (http://hmpdacc.org/HMREFG/) (Nelson *et al.*, 2010). Sequences of both the *gBE* genes were mapped to these reference microbial genomes with similarity levels of at least 0.80, using the CLC long-read alignment program, version 3.2.2 (CLC Bio., Katrinebjerg, Denmark). The resulting CAS format files were converted to the BAM format, which were interpreted using the CLC genomic workbench, version 5.1 (CLC Bio.).

Metagenomic sequences were trimmed by filtering for quality and length using the HMP protocol (Methé *et al.*, 2012) and screened for residual host DNA using the Burrows–Wheeler Alignment tool (Li and Durbin, 2009). Trimmed sequences were searched against the Carbohydrate-Active EnZymes (CAZy) characterized sequence database. This generated relative abundance scores for each CAZy GH and polysaccharide lyase (PL) families, including the relative abundances of housekeeping genes *GT51* and *GT28*, in the host metagenome.

Accession number
The sequence information in this paper has been deposited in the EMBL Sequence Read Archive with study accession number ERP002194.

Results

In this study, we first performed 16S rRNA gene profiling of 69 fecal samples from different host species (human ($N = 16$), chicken ($N = 18$), cow ($N = 15$) and pig ($N = 20$), hereafter abbreviated as 16S) and metagenomic analysis of five pooled samples per species. We subsequently performed *gBE* profiling of all fecal samples using custom primers for the *gBE* family. This allowed us to relate microbial community ecology both to global profiles of species-specific GHs and to individual hosts' *gBE* profiles.

16S rRNA gene and shotgun metagenomic sequencing of vertebrate fecal microbiota
Unsurprisingly, bacterial community composition by 16S rRNA gene profiling differed significantly among human and animal hosts from the family to the whole-phylum level (Figure 1a and Supplementary Figure S2). As expected, human guts were dominated by Bacteroidetes and Firmicutes members, pigs enriched for Lactobacillaceae and Clostridiaceae, chickens also enriched by *Lactobacillus* spp., and the cow gut comprised abundant Clostridiaceae and Ruminococcaceae (Callaway *et al.*, 2010). Our phylogenetic profiles for all four of these communities were thus generally concordant with existing microbial surveys.

We assessed the overall GH compositions of the four host species' communities using the shotgun metagenomic sequencing of pooled samples, and compared our Korean human with 742 published data from HMP (Figure 1b). Each CAZy GH and PL was normalized by the relative abundance of two housekeeping genes (GT51; penicillin-binding protein and GT28; MurG transferase) that consistently occur at constant copy number in bacterial genomes (Cantarel *et al.*, 2012). Strikingly, profiles of GH families based on the CAZy database (Cantarel *et al.*, 2009) revealed a total of 75 GH families and nine PLs (Figure 1b). Similar to microbial community composition, these were also of course highly variable among species, with abundant members including GH2, GH3, GH13 and GH43. Family GH13 (which contains the *gBE* genes) particularly exhibited the broadest representation among species, at its lowest abundance in the chicken fecal microbiota, followed by those of human and pig, and the highest abundance in cow microbiota. GH families from the Korean subject were more abundant than the ones from the HMP database; however, overall profiles of GH families were similar to each other.

Amplicon sequencing of gBE gene family members in the vertebrate gut
Based on this prevalence, we deeply characterized the diversity of *gBE* subfamily sequences within GH13 among all 69 samples using amplicon sequencing of this gene family with custom primers (Figure 1c). Using sequence-clustering methods comparable to those by which 16S sequences are grouped into OTUs, we grouped *gBE* sequences into 172 OGBUs at 97% identity. Although only the most abundant and prevalent OGBUs appear in Figure 1b, the large differences among both species and individual hosts common at the 16S and metagenomic levels also occurred within *gBE* diversity. *gBE* profiles from the same species were generally comparable, with a few exceptions and additional differences remaining among individual hosts. However, only a small number of OGBUs directly mirrored the whole-community differences observed by 16S profiling. OGBU3, OGBU1, OGBU2 and OGBU5, for example, differentiated the microbiota of humans, pig, cow and chicken, respectively.

In order to evaluate the overall concordance of phylogenetic diversity with *gBE* diversity, we performed principal coordinate analysis using unweighted UniFrac and examined the major
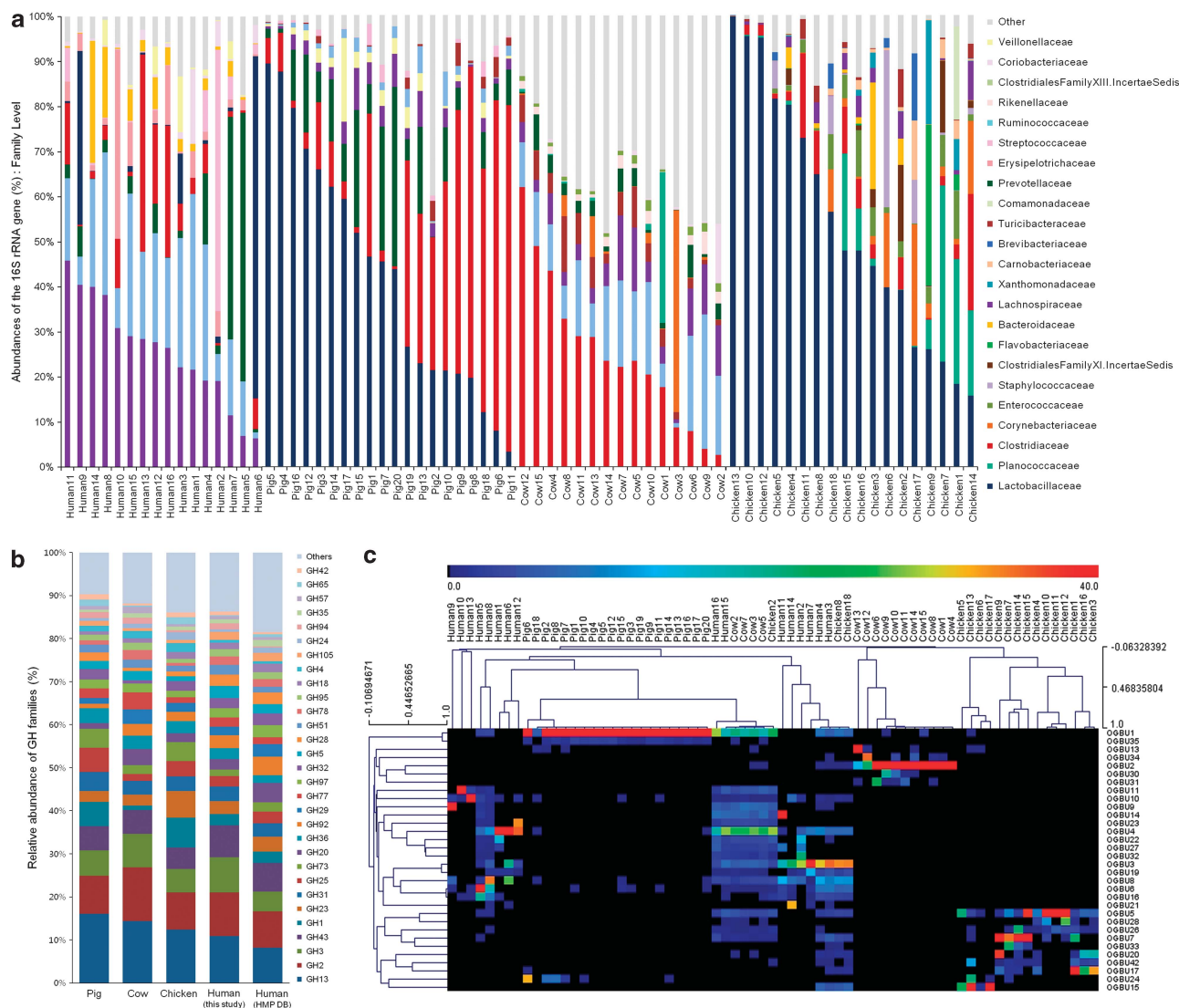
Figure 1 The relative abundance and relationship of bacteria and glycoside hydrolases genes in the gut microbiota from four different hosts. (**a**) The stacked bar graph shows the relative abundance of bacteria at the family level. (**b**) Relative abundances of glycoside hydrolases genes in the gut microbiota from four different hosts and 742 HMP database using whole genome shotgun sequencing of fecal samples. Each CAZy GH and PL is normalized by the relative abundance of two housekeeping genes (*GT51* and *GT28*). (**c**) The heatmap shows relationship between hosts and OGBU genes, and color indicates abundance of the OGBU gene within hosts.

patterns of variability of *gBE* genes as related to the 16S rRNA gene (Figure 2a). The configurations of gBE and phylogenetic diversity were in many cases not concordant, both for individual hosts and when comparing whole species. For example, human and pig 16S profiles were the most similar ones but differed significantly in gBE profiles. Likewise, no individual bovine hosts were unusual outliers with respect to 16S-based community composition; however, four hosts were readily detected to carry distinct gBE repertoires despite phylogenetic similarity. These data suggest that, whereas gene family function certainly correlates with microbial phylogenetic composition, distinct gBE profiles can commonly arise from similar phylogenetic profiles and *vice versa*.

Microbes and gBE sequence families enriched in the human, pig, cow and chicken gut microbiota

To identify which specific microbes and gBEs underly these between-species differences, we tested for significant enrichment of microbial clades in the four vertebrate gut microbial communities (Figure 2b). This resulted first, as expected, in a diverse selection of over- and under-abundant organisms. Enriched microbial clades of the human gut microbiota, as compared with other species, included *Clostridia* sp., *Faecalibacterium* sp. and Lachnospiraceae. Gut microbes enriched in cow, chicken and pig included Bacteroidia, Lactobacillaceae and unclassified *Lactobacillus* spp, respectively. Approximately 25.3% of all identified OTUs ($N = 312$) were common to all four hosts, including
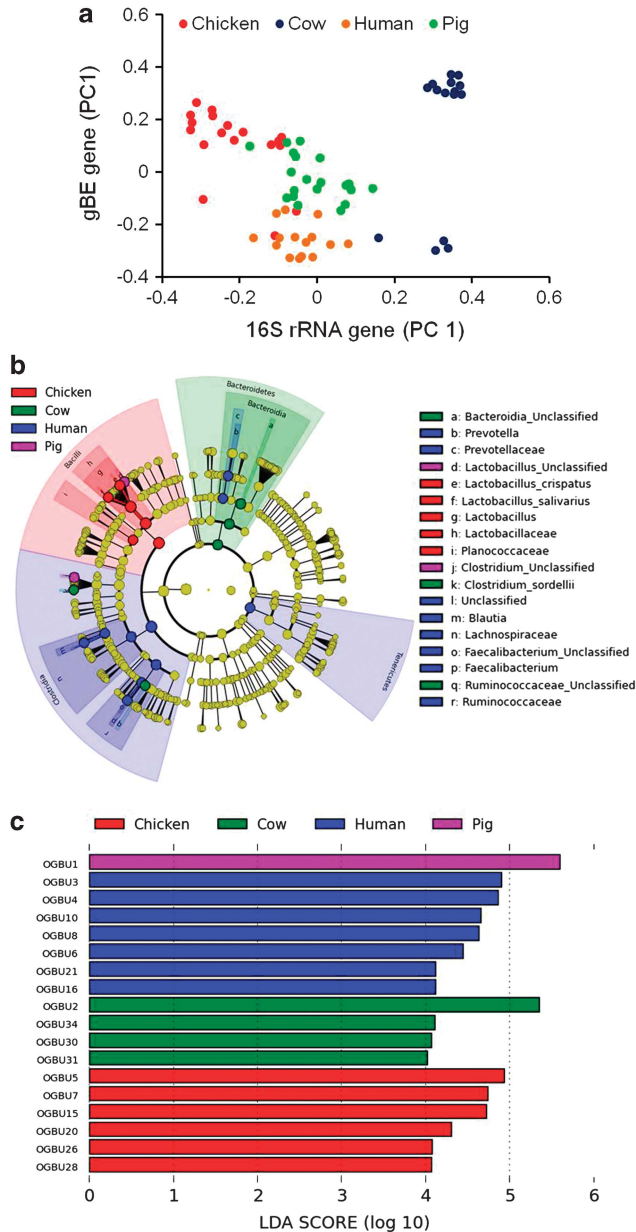
Figure 2 Concordances of 16S rRNA and *gBE* genes and abundance characteristics of OTUs and OGBUs in four different hosts. (**a**) Concordances of 16S rRNA and *gBE* genes using values of principal component 1 (PC1), and PC1 of 16S rRNA and *gBE* gene indicate 11.7% and 16.3%, respectively. (**b**) Cladogram of relevant features of 16S rRNA genes in four different hosts. (**c**) LDA (linear discriminant analysis) plot means biomarkers found by ranking accordingly to their effect size (4.0) of the OGBUs.

Bacteroides sp. and *Clostridium* sp. from Clostridiaceae family. The cow microbiota had the highest proportion of unique OTUs (22.4%), followed by chicken (5.8%), pig (3.5%), and human (1.3%) (Supplementary Figure S3a and Supplementary Table S2).

A comparable enrichment analysis of gBE families among these four species revealed that some OGBUs were unique to a single species, with a majority consistently shared among two or more (Figure 2c).

The cow gBE profiles were most distinct from those of other hosts, whereas the gBE profiles of human fecal microbiota were less globally distinct and frequently overlapped with those of the other fecal microbiota. This was true in particular for the human and pig gBE profile, which shared 39 OGBUs. Approximately 14.0% of all identified *gBE* genes were found in all four host species. Again, the cow fecal microbiota had the highest proportion of unique *gBE* genes (16.3%), followed as above in OTUs by chicken (8.7%), pig (2.3%) and human (1.2%) (Supplementary Figure S3b and Supplementary Table S3).

Association of microbes with gBE genes identified in the fecal microbiota

We applied two approaches to assess which microbes were associated with *gBE* sequence families beginning by examining OGBU/OTU co-occurrence patterns among microbial communities across all hosts (Figure 3 and Supplementary Figure S4a). Among the major gut microbiota, relatively few OGBUs co-varied in abundance in association with *Bacteroides* spp., *Enterococcus* spp. and *Ruminococcus* spp. On the other hand, *Clostridium* spp. from the Clostridiaceae family, *Lactobacillus* spp. and *Prevotella* spp. were associated with high abundance of many distinct OGBUs, particularly the joint co-occurrence of OGBU1 and OGBU35. We performed more detailed network analyses in order to assess the full complement of associations between *gBE* genes and each of the several key gut microorganisms including *Bacteroides* spp., *Clostridium* spp. from Clostridiaceae family, *Lactobacillus* spp., *Prevotella* spp., *Enterococcus* spp. and *Ruminococcus* spp. (Supplementary Figures S4b–g). Although some OTU/OGBU associations were retained among multiple species, this was the exception rather than the rule. Instead, OGBUs were most often either species-specific or, when present in multiple species, associated with different OTUs in different species' guts.

In addition to observing these gBE/OTU co-occurrence patterns *in vivo*, we also used 1275 reference genomes as nucleotide mapping targets for individual *gBE* amplicons. We aligned *gBE* sequences against these reference genomes using the CLC mapper, requiring 80% nucleotide similarity. This found high-confidence targets for 91.0% of human *gBE* sequences and 87.2% of pig. However, the *gBE* sequences from the fecal microbiota of chicken (57.9%) and cow (29.4%) were much less commonly mapped to reference genomes, suggesting less comprehensive phylogenetic coverage of these organisms in current genome databases. This also allowed us to compare the abundance of microbes as estimated from 16S sequences with the abundance of their genomes' *gBE* sequences (Supplementary Figure S5). These abundances were remarkably different in some species and microbes; for example,

Lactobacillus spp., which were abundant in both chicken and pig based on 16S rRNA genes, carried very few gBEs in chicken but retained many in pig. If the differences above suggest a need for more reference genomes covering these organisms, this finding indicates that the gBEs within such genomes may be quite distinct from those in current gut genera genomes.

Alpha- and beta-diversity profiles of gBE amplicons among hosts

We next considered the ecological profiles of these four hosts' gut microbiomes at a broader level
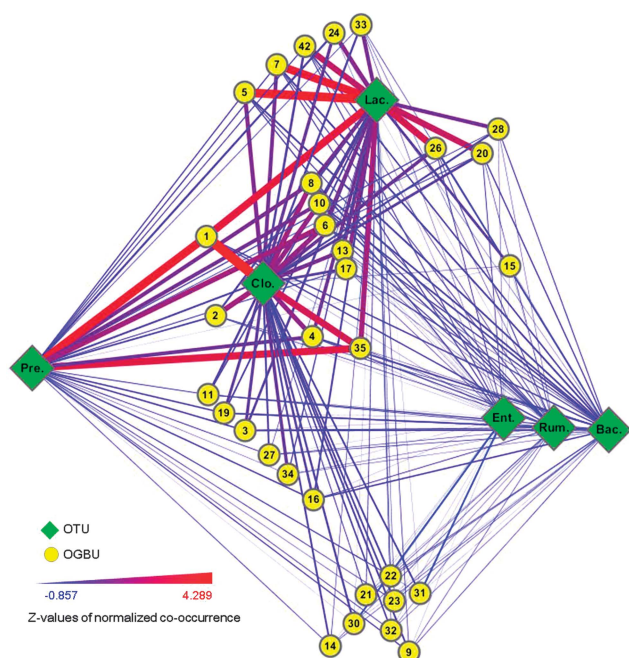

Figure 3 Associations between OGBUs and six major gut microbiota. Nodes, green diamond and yellow circles, indicate OTUs and OGBUs, respectively. Edge colors indicate increasing co-occurrence from thin blue to thick red. Each network of *Bacteroides* spp., *Clostridium* spp. from Clostridiaceae family, *Lactobacillus* spp., *Prevotella* spp., *Enterococcus* spp. and *Ruminococcus* spp. is presented in Supplementary Figure S4.

(Figure 4). In an assessment of phylogenetic alpha diversity, or the number and distribution of distinct organisms, the OTUs in cow microbiota were again most diverse as observed using both OTU count and the Chao1 richness measures; the chicken gut was least diverse, with greater variability among both pig and human individuals (Figures 4a and c). An analysis of variance test followed by Games–Howell comparisons in particular indicated that the richness of cow fecal microbiota was significantly higher than those of the other species ($P < 0.001$). Unlike OTUs, the alpha diversity pattern for OGBUs was not significantly different among hosts, even though the richness of cow OGBUs trended slightly higher (Figure 4b).

Beta-diversity profiles were subsequently computed in two ways: between individual hosts within each species (human, pig, cows and chicken, Figure 5a) and in aggregate across species' hosts (Figure 5b). The average unweighted UniFrac distances of 16S profiles among hosts within the same species were 0.718 (chicken), 0.652 (cow), 0.643 (human) and 0.628 (pig) (Figure 5a), significantly differing only in chicken and marginally between cow and pig. The average unweighted UniFrac distances of *gBE* genes were 0.622 (chicken), 0.654 (cow), 0.560 (human) and 0.573 (pig); these showed a greater range of differences among within-species hosts. Specifically, significant differences occurred between human and chicken fecal microbiota and between human and cow and the overall pattern mirrored that observed globally in Figure 2a. However, it remained surprising that differences among hosts did not appear to mirror either the short-term (lifetime) or long-term (evolutionary) dietary patterns of these species with respect to starch utilization and glycogen metabolism.

Between species (Figure 5b), there were several interesting differences in beta diversities as calculated based on 16S rRNA versus *gBE* gene sequences. Overall, both sequence types showed relatively high similarities by this measure between the human and pig microbiota. However, based on OGBU overlap (unweighted UniFrac), the human gBE complement was to some degree a subset of that
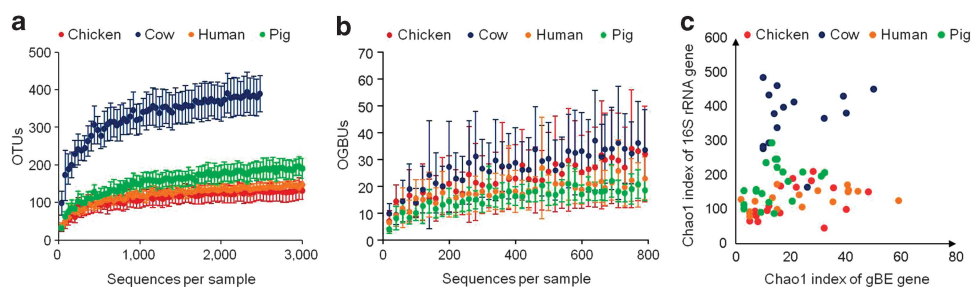
Figure 4 Alpha diversities of OTUs and *gBE* genes in four different hosts. (**a**) Phylogenetic diversity was estimated from the average Chao1 values of gut microbiota samples from chicken (red), cow (blue), human (orange) and pig (green). The results are based on 3000 sequences of 16S rRNA genes per sample. (**b**) The *gBE* genes per sample are based on 800 sequences. Bars indicate means ± 95% confidence intervals. (**c**) The scatter plot means concordance of alpha diversity between 16S rRNA and *gBE* genes.
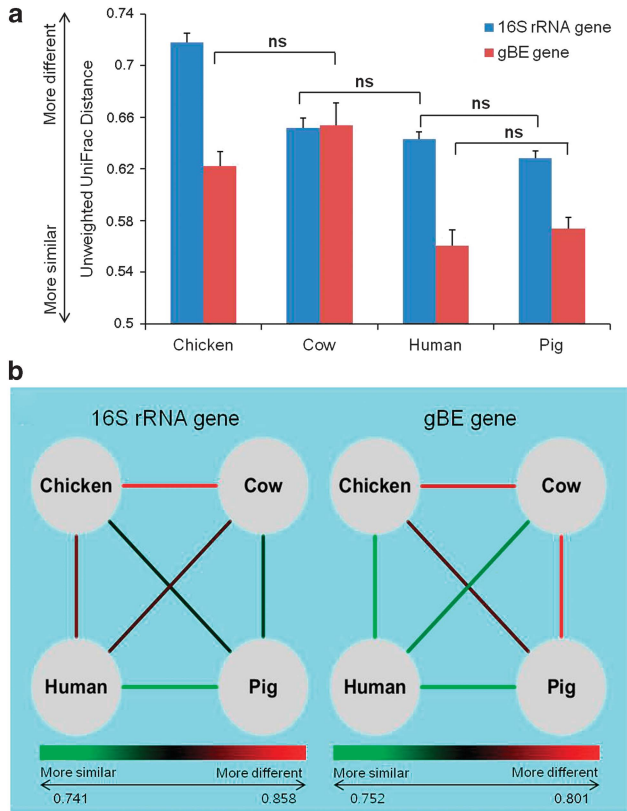
Figure 5 Average unweighted UniFrac distances within (**a**) and between (**b**) chicken, cow, human and pig estimated from 16S rRNA and *gBE* gene PCR amplicon sequences. (**a**) The ns indicates nonsignificant differences between hosts, and the all other comparisons except ns are statistically significant ($P < 0.05$). (**b**) Hosts (nodes) connected by normalized average unweighted UniFrac distance. Edge colors indicate increasing similarity from red to green. Similarity of 16S rRNA are in 0.741 (human–pig) to 0.858 (chicken–cow), and gBE are in 0.752 (chicken–human) to 0.801 (cow–pig) distance range.

of all other species. No other species pair showed such similarity either phylogenetically (16S) or by gBE overlap. The data thus suggest that, whereas the functional diversity of a microbial community can be quite distinct from the diversity of its phylogenetic profile, it remains to be determined what evolutionary or environmental factors drive the convergence or divergence of functional gene repertoires.

Discussion

We present here the first targeted metagenomic characterization of the gBE family (subfamily 9 within family GH13) *in vivo* among 69 gut microbiomes in human, pig, cow and chicken hosts. The study was performed using an innovative tiered approach including an initial 16S rRNA gene sequence survey, targeted metagenomic shotgun sequencing on pooled samples and finally *gBE* amplicon profiling of the original complete sample set. This allowed the derivation of 172 'operational glucan-branching units' or OGBUs based on

OTU-like clustering of the resulting *gBE* amplicons, which we associated with each host and species, specific gut microbes and with overall ecological diversity of the gut microbiota. Similar gene-targeted metagenomics may prove useful for other gene families for which appropriate conserved regions capable of specific, verified amplification are available (Kuczynski *et al.*, 2011).

Our initial phylogenetic profiles based on 16S rRNA gene sequences were consistent with existing microbial surveys of diverse vertebrate gut microbiota. The human gut in this Korean population was rich in *Clostridia*, comparable to other urban populations (Huttenhower *et al.*, 2012). The family Ruminococcaceae and the genus *Prevotella* spp. were abundant in cow rumens, where they have a major role in the complex lignocellulosic degradation system including microbial attachment to and digestion of plant biomass (Brulc *et al.*, 2009). *Lactobacillus* spp. were frequently detected in the fecal microbiota of pigs both by culture-based- and culture-independent methods in a previous study as well (Isaacson and Kim, 2012). In some cases, *Bacillus* spp. and *Lactobacillus* spp. are used as animal probiotics (Barbosa *et al.*, 2005), which might partially determine their abundance in the gut microbiota. It is thus not surprising that the differences found here between both OTUs and OGBUs across species were extreme. More interestingly, the same microbial organisms in different hosts appeared to contain different OGBUs by both co-variation and genomic analyses. As neither differences between host species nor differences among individual hosts were sufficient to explain this, a combination of both short-term (for example, diet and horizontal transfer (Hehemann *et al.*, 2010)) and long-term (for example, genetic (Ley *et al.*, 2008; Muegge *et al.*, 2011)) factors is likely involved.

Diet is likely one of the main factors that determine the *gBE* gene profile. In addition, previous studies have reported that the gut microbiota is a horizontal gene transfer 'hot spot' because of an abundance of conjugal elements such as CTnRINT (Kurokawa *et al.*, 2007; Smillie *et al.*, 2011). Extensive genetic exchange took place between lactic acid bacteria such as *L. acidophilus* and *L. johnsonni*, in which horizontal gene transfer also had a role (Nicolas *et al.*, 2007). A recent study reported horizontal gene transfer of porphyranase from ocean bacteria (*Zobellia* sp.) to members of the gut microbiota, such as *Bacteroides plebeius* (Hehemann *et al.*, 2010). This, in tandem with our results, strongly suggests that bacteria identified as identical by 16S sequencing may possess highly varying copy numbers and/or types of GH. Horizontal gene transfer is thus a central evolutionary strategy for maintenance of microbial community homeostasis and regulation of microbial and host–microbial metabolic balance.

Dynamic interactions occur between the host and the gut microbiota, and diet, host genetics and the

immune system are all crucial factors that determine the fecal microbiota (Guarner and Malagelada, 2003). In addition to the gBE and GH examples above, the gut microbiota produces many other enzymes involved in carbohydrate utilization, reduction in cholesterol levels and vitamin biosynthesis (Hooper et al., 2002). In addition, host genetics and immunological responses, particularly mucosal immunity in the gut, exert constant selective pressure on the gut microbiota (Kovacs et al., 2011). In turn, the gut microbiota interacts with the host and helps to shape the mucosal immune response (Sartor, 2011), and it can finally be influenced by transient organisms such as pathogenic bacteria as well (Reid et al., 1990).

However, the characteristics of gBEs in the gut, and the modality of their resulting influence on microbial dynamics, are likely associated with diet, as carbohydrates are of course the direct target of gBE activity (Supplementary Figure S6). Depending on the levels of polysaccharides and nutritional environments, different microbes would have different roles in both catabolic and anabolic metabolic activities in the gut. When bacteria thrive on carbohydrates, they have to store these carbohydrates, and GBE is part of the bacterial glycogen pathway. Thus, the GBE subfamily of GH13 would be indirectly linked to the diet. Additionally, when gut microbes were dead, their intracellular polysaccharide such as glycogen should be released and made available to other gut microbes. As a result, both catabolic and anabolic metabolic activities of microbiota in the gut should be highly dynamic and interact with each other. Therefore, the activities of gBE and other GH enzymes affect the utilization of carbohydrates by gut microbiota, specifically intracellular glycogen storage (although not direct breakdown of food). Sugar availability affects the dynamics of the microbiota, however, indirectly influencing host metabolism as a result. Several studies have reported how the gut microbiota helps the host breakdown non-digestible diet sources (Xu and Gordon, 2003). For example, most Firmicute species breakdown subsets of difficult-to-digest dietary polysaccharides, allowing their digestion and absorption, one possible route by which gut microbes may be associated with metabolic disorders (Das, 2010). In addition, to facilitate polysaccharide catabolism, glycogen biosynthesis by the gut microbiota may control glucose accessibility and solubility. Glycogen accumulates in a number of bacteria as an energy-reserve compound, and its synthesis usually occurs during the stationary phase of growth or under carbon limitation (Preiss, 1984).

Glycogen synthesis by bacteria is not fully understood. Previous studies have suggested that it has a role in prolonging viability by storage of a source of energy (Slock and Stahly, 1974). One advantage of using glycogen as a reserve compound is that it has little effect on internal osmotic pressure and provides a stored source of energy and carbon (Strange, 1968). It is well known that gBE affects polysaccharide branch levels and the water solubility of glucose polymers (Smith, 2001). Therefore, it is likely that the characteristics and abundance of gBE and other GH enzymes determine the levels and types of monomers and polymers of carbohydrates inside microorganisms and are free in the gut. To better determine the functions of glycogen-degrading enzymes, future work should evaluate the effect of glycogen average chain length, a core factor that influences glycogen metabolic rates and import/export to and from microorganisms (Wang and Wise, 2011).

Although overall patterns of diversity between microbiomes were roughly comparable when considering either 16S-based OTUs or gBE-based OGBUs, the structural versus functional organization of these gut communities differed in several ways. Alpha diversities of gBE genes among the four host species were not significantly different, despite clear differences in 16S-based microbial diversity. Ruminants, for example, have been shown to have particularly high species diversity (Wright and Klieve, 2011), which was reflected here. A diversity of bacterial enzymes in the gastrointestinal tract would be advantageous for foregut fermentation to digest the plant cell wall by microbial processing (Brulc et al., 2009). Although quite distinct ranges of individual gBE genes were observed within each of the four host species, as well as of GH families overall, individual hosts' diversities of OGBUs remained stable. This pattern of gBE genes could be due to selection for particular diet- and host-adapted gBE profiles in the fecal microbiota regardless of the accompanying microbial phylogenetic profile.

In conclusion, our approach in this study, employing targeted metagenomics for gBE genes, facilitated the characterization of a specific metabolic gene family in the gut microbiota of different host species. The characteristics of gBE genes were associated with differences among host species independently of the phylogenetic characteristics of the microbiome, likely as a result of selection induced by carbohydrate metabolism in the gut. Our data suggest that even very similar gut bacteria often possess different gBE genes, indicative of both short-term (for example, diet) and long-term (for example, genetic) pressures. Comparison of mammals with distinct evolutionary backgrounds and diets will broaden our understanding of the functional and ecological dynamics of the gut microbiota.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Balzer S, Malde K, Lanzén A, Sharma A, Jonassen I. (2010). Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim. *Bioinformatics* **26**: i420–i425.

Barbosa TM, Serra CR, La Ragione RM, Woodward MJ, Henriques AO. (2005). Screening for *Bacillus* isolates in the broiler gastrointestinal tract. *Appl Environ Microbiol* **71**: 968–978.

Brulc JM, Antonopoulos DA, Berg Miller ME, Wilson MK, Yannarell AC, Dinsdale EA *et al.* (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci USA* **106**: 1948–1953.

Callaway T, Dowd S, Edrington T, Anderson R, Krueger N, Bauer N *et al.* (2010). Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *J Anim Sci* **88**: 3977–3983.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233–D238.

Cantarel BL, Lombard V, Henrissat B. (2012). Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* **7**: e28742.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Chao A. (1984). Nonparametric estimation of the number of classes in a population. *Scand J Stat* **11**: 265–270.

Das UN. (2010). Obesity: genes, brain, gut, and environment. *Nutrition* **26**: 459–473.

DeSantis TZ, Hugenholtz P, Keller K, Brodie E, Larsen N, Piceno Y *et al.* (2006). NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**: W394–W399.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Fyfe JC, Giger U, Van Winkle TJ, Haskins ME, Steinberg SA, Wang P *et al.* (1992). Glycogen storage disease type IV: inherited deficiency of branching enzyme activity in cats. *Pediatr Res* **32**: 719–725.

Goecks J, Nekrutenko A, Taylor J, Team TG. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**: R86.

Guarner F, Malagelada JR. (2003). Gut flora in health and disease. *Lancet* **361**: 512–519.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**: 235–237.

Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**: 908–912.

Henning SL, Wambolt RB, Schönekess BO, Lopaschuk GD, Allard MF. (1996). Contribution of glycogen to aerobic myocardial glucose utilization. *Circulation* **93**: 1549–1555.

Hooper LV, Midtvedt T, Gordon JI. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* **22**: 283–307.

Howe EA, Sinha R, Schlauch D, Quackenbush J. (2011). RNA-Seq analysis in MeV. *Bioinformatics* **27**: 3209–3210.

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.

Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT *et al.* (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207–214.

Isaacson R, Kim HB. (2012). The intestinal microbiome of the pig. *Anim Health Res Rev* **13**: 100–109.

Iwai S, Chai B, Sul WJ, Cole JR, Hashsham SA, Tiedje JM. (2009). Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *ISME J* **4**: 279–285.

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E *et al.* (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* **26**: 1029–1034.

Koropatkin NM, Cameron EA, Martens EC. (2012). How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**: 323–335.

Kovacs A, Ben-Jacob N, Tayem H, Halperin E, Iraqi FA, Gophna U. (2011). Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microbial Ecol* **61**: 423–428.

Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D *et al.* (2011). Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet* **13**: 47–58.

Kumar A, Larsen CE, Preiss J. (1986). Biosynthesis of bacterial glycogen. Primary structure of Escherichia coli ADP-glucose: alpha-1, 4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the glgA gene. *J Biol Chem* **261**: 16256–16259.

Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A *et al.* (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**: 169–181.

Lane DJ. (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) *Nucleic Acid Techniques in Bacterial Systematics*. Wiley: New York, pp 115–175.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al.* (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.

Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lozupone CA, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C *et al.* (2012). A framework for human microbiome research. *Nature* **486**: 215–221.

Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L *et al.* (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**: 970–974.

Nelson KE, Weinstock GM, Highlander SK, Worley KC, Creasy HH, Wortman JR *et al.* (2010). A catalog of reference genomes from the human microbiome. *Science* **328**: 994–999.

Nicolas P, Bessières P, Ehrlich SD, Maguin E, Van De Guchte M. (2007). Extensive horizontal transfer of core genome genes between two Lactobacillus species found in the gastrointestinal tract. *BMC Evol Biol* **7**: 141.

Preiss J. (1984). Bacterial glycogen synthesis and its regulation. *Ann Rev Microbiol* **38**: 419–458.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Reid G, Bruce AW, McGroarty JA, Cheng K, Costerton JW. (1990). Is there a role for lactobacilli in prevention of urogenital and intestinal infections? *Clin Microbiol Rev* **3**: 335–344.

Sartor R. (2011). Key questions to guide a better understanding of host–commensal microbiota interactions in intestinal inflammation. *Mucosal Immunol* **4**: 127–132.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504.

Slock J, Stahly D. (1974). Polysaccharide that may serve as a carbon and energy storage compound for sporulation in Bacillus cereus. *J Bacteriol* **120**: 399–406.

Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**: 241–244.

Smith AM. (2001). The biosynthesis of starch granules. *Biomacromolecules* **2**: 335–341.

Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Eng Des Sel* **19**: 555–562.

Strange R. (1968). Bacterial ''glycogen'' and survival. *Nature* **220**: 606–607.

Sung J, Cho SI, Lee K, Ha M, Choi EY, Choi JS *et al.* (2006). Healthy twin: A twin-family study of Korea-Protocols and current status. *Twin res Hum Genet* **9**: 844–848.

Thompson JD, Gibson TJ, Higgins DG. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2: Unit 2.3.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE *et al.* (2008). A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

van der Maarel MJEC, Leemhuis H. (2012). Starch modification with microbial alpha-glucanotransferase enzymes. *Carbohydr Polym* **93**: 116–121.

Wang L, Wise MJ. (2011). Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften* **98**: 719–729.

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.

Wright ADG, Klieve AV. (2011). Does the complexity of the rumen microbial ecology preclude methane mitigation? *Anim Feed Sci Technol* **166**: 248–253.

Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B *et al.* (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* **34**: D187–D191.

Xu J, Gordon JI. (2003). Honor thy symbionts. *Proc Natl Acad Sci USA* **100**: 10452–10459.

Ye Y. (2010). Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. *Proceedings of BIBM 153–157.*

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