The ISME Journal (2012) 6, 2091–2106

www.nature.com/ismej

ORIGINAL ARTICLE

Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis

David Berry[1], Clarissa Schwab[2], Gabriel Milinovich[2], Jochen Reichert[1], Karim Ben Mahfoudh[1], Thomas Decker[3], Marion Engel[4], Brigitte Hai[4], Eva Hainzl[5], Susanne Heider[6], Lukas Kenner[6], Mathias Müller[5], Isabella Rauch[3], Birgit Strobl[5], Michael Wagner[1], Christa Schleper[2], Tim Urich[2] and Alexander Loy[1]

[1]*Department of Microbial Ecology, Vienna Ecology Center, Faculty of Life Sciences, University of Vienna, Wien, Austria;* [2]*Department of Genetics in Ecology, Vienna Ecology Center, Faculty of Life Sciences, University of Vienna, Wien, Austria;* [3]*Department of Microbiology, Immunobiology and Genetics, Max F. Perutz Laboratories, University of Vienna, Wien, Austria;* [4]*Research Unit of Environmental Genomics, Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg, Germany;* [5]*Institute of Animal Breeding and Genetics, University of Veterinary Medicine Vienna, Wien, Austria and* [6]*Ludwig Boltzmann Institute for Cancer Research, Medical University of Vienna, Wien, Austria*

Human inflammatory bowel disease and experimental colitis models in mice are associated with shifts in intestinal microbiota composition, but it is unclear at what taxonomic/phylogenetic level such microbiota dynamics can be indicative for health or disease. Here, we report that dextran sodium sulfate (DSS)-induced colitis is accompanied by major shifts in the composition and function of the intestinal microbiota of STAT1$^{-/-}$ and wild-type mice, as determined by 454 pyrosequencing of bacterial 16S rRNA (gene) amplicons, metatranscriptomics and quantitative fluorescence *in situ* hybridization of selected phylotypes. The bacterial families Ruminococcaceae, Bacteroidaceae, Enterobacteriaceae, Deferribacteraceae and Verrucomicrobiaceae increased in relative abundance in DSS-treated mice. Comparative 16S rRNA sequence analysis at maximum possible phylogenetic resolution identified several indicator phylotypes for DSS treatment, including the putative mucin degraders *Akkermansia* and *Mucispirillum*. The analysis additionally revealed strongly contrasting abundance changes among phylotypes of the same family, particularly within the Lachnospiraceae. These extensive phylotype-level dynamics were hidden when reads were grouped at higher taxonomic levels. Metatranscriptomic analysis provided insights into functional shifts in the murine intestinal microbiota, with increased transcription of genes associated with regulation and cell signaling, carbohydrate metabolism and respiration and decreased transcription of flagellin genes during inflammation. These findings (i) establish the first in-depth inventory of the mouse gut microbiota and its metatranscriptome in the DSS colitis model, (ii) reveal that family-level microbial community analyses are insufficient to reveal important colitis-associated microbiota shifts and (iii) support a scenario of shifting intra-family structure and function in the phylotype-rich and phylogenetically diverse Lachnospiraceae in DSS-treated mice.
The ISME Journal (2012) **6,** 2091–2106; doi:10.1038/ismej.2012.39; published online 10 May 2012
Subject Category: microbial ecology and functional diversity of natural habitats
Keywords: gut microbiota; Lachnospiraceae; *Akkermansia*; *Mucispirillum*; dextran sodium sulfate; colitis

Correspondence: T Urich, Department of Genetics in Ecology, Vienna Ecology Center, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, A-1090 Wien, Austria.
E-mail: tim.urich@univie.ac.at
or A Loy, Department of Microbial Ecology, Vienna Ecology Center, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, A-1090 Wien, Austria.
E-mail: loy@microbial-ecology.net
Received 3 November 2011; revised 1 March 2012; accepted 16 March 2012; published online 10 May 2012

Introduction

The human and animal intestinal tract provides a nutrient- and niche-rich ecosystem for trillions of symbiotic microorganisms. Cooperation between the host and its gut microbiota is largely beneficial for all partners, but shifts in microbiota composition can also violate this mutualism (dysbiosis) and result in severe intestinal disease. The lives of about

2 million and 1.4 million persons in Europe and the USA, respectively, are impaired by chronic intestinal disorders such as Crohn's disease or ulcerative colitis (Loftus, 2004), two prominent examples of a range of inflammatory disorders of the intestinal mucosa collectively called inflammatory bowel disease (IBD). IBD is the result of an imbalance in the interaction of symbiotic microorganisms, epithelium and immune system (Braun and Wei, 2007). The composition of the gut microbiota is altered in patients with IBD (reviewed in Packey *et al.* (2009) and Reiff and Kelly (2010)) and animal studies have demonstrated that the presence of microorganisms is important to elicit colitis (Kühn *et al.*, 1993; Dianda *et al.*, 1997).

16S rRNA gene sequencing approaches have provided a more comprehensive view of the individual gut microbiota of IBD patients, though a characteristic IBD microbiota has not been identified, which may be owing to inter-individual variability, different sampling strategies and differences in sequence data acquisition and analysis (Manichanh *et al.*, 2006; Frank *et al.*, 2007; Willing *et al.*, 2010). Many 16S rRNA-based surveys of intestinal microbiota use taxonomic classification as the primary tool for comparative analysis to identify taxa that are indicative of a certain host phenotype. For example, the Firmicutes/Bacteroidetes ratio has been suggested as an indicator of obesity (Ley *et al.*, 2005; Armougom *et al.*, 2009) (though exceptions have been reported (Murphy *et al.*, 2010)) and abundance of Enterobacteriaceae often correlate with inflammation (Lupp *et al.*, 2007). However, the hierarchical level at which taxonomic units become biologically meaningful are not yet defined (Achtman and Wagner, 2008). It thus remains unclear how important within-taxa variation is for different higher-level microbial taxa (that is, families, orders, phyla) in the intestines.

Animal models of disease are an important complement to human studies because they allow for targeted analyses of microbial, pathological and immunological aspects of IBD under well-defined and reproducible conditions (Nell *et al.*, 2010). A widely used approach to trigger intestinal inflammation in animal models is oral administration of the chemical irritant dextran sodium sulfate (DSS) (reviewed in Nell *et al.* (2010) and Okayasu *et al.* (1990)). The DSS colitis mouse model shares some similarity to human ulcerative colitis (Okayasu *et al.*, 1990), though it is rather a model for acute inflammation-induced injury. Antibiotics can prevent colitis in the DSS model, indicating that microbiota have a role in inflammation (Rath *et al.*, 2001). DSS exerts a cytotoxic effect on the gut epithelium, resulting in higher mucus and intestinal permeability and acute barrier damage, which may increase transfer of pro-inflammatory microbial compounds across the gut epithelium (Kitakima *et al.*, 1999; Mueller and Macpherson, 2006; Johansson *et al.*, 2010). Given its widespread use, it is surprising how few studies have investigated the gut microbiota after oral DSS administration. Previous studies using 16S rRNA gene fingerprinting (Heimesaat *et al.*, 2007; Nagalingam *et al.*, 2011) and sequencing of few 16S rRNA gene clones (<100) per animal (Lupp *et al.*, 2007; Nagalingam *et al.*, 2011) have given rise to different conclusions and there has been no investigation of shifts in microbiota functional potential or activity using metagenomic or metranscriptomic approaches. Therefore, a detailed and clear understanding of the structural and functional alterations of the intestinal microbiota in the DSS mouse model is absent.

This study aimed to identify correlations between inflammation, microbiota community composition and the microbial metatranscriptome in the DSS-induced colitis mouse model. The impact of type I and II interferon signaling, which is, among other stimuli, activated by viral and bacterial infection (Decker *et al.*, 2005) and believed to have a role in both human IBD (Schreiber *et al.*, 2002) and acute murine colitis (Bandyopadhyay *et al.*, 2008), was investigated with STAT1$^{-/-}$ and wild-type (wt) C57BL/6N mice (Meraz *et al.*, 1996). 454 pyrosequencing of bacterial 16S rRNA gene and 16S rRNA complementary DNA amplicons was used to characterize the microbial community structure and to identify individual phylotypes significantly associated with health status, genotype or both (indicator phylotypes). Shifts in abundance of selected indicator phylotypes were confirmed by quantitative fluorescence *in situ* hybridization (FISH). Differences between the microbial metatranscriptome of untreated and DSS-treated mice were identified by 454 pyrosequencing of cDNA prepared from the intestinal lumen.

Materials and methods

Animal experiments

Wt ($n=5$) and STAT1$^{-/-}$ ($n=5$) mice (Durbin *et al.*, 1996) 6–8 weeks of age were provided with 2% DSS (molecular weight: 36–50 kDa, MP Biomedicals, Santa Ana, CA, USA) in autoclaved drinking water *ad libitum* for 7 days, after which they received DSS-free autoclaved drinking water for an additional 3 days before sampling. Three of the five DSS-treated wt mice succumbed prematurely on day 9 and were thus not included in the sampling and analysis. Wt ($n=5$) and STAT1$^{-/-}$ ($n=5$) control mice of the same age received DSS-free autoclaved drinking water for the entire 10 days. Upon euthanization, the cecum and colon were immediately removed and flushed with 7 ml of sterile, anoxic phosphate-buffered saline (136 mM NaCl, 2.6 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, pH 7.2) pre-warmed to 37 °C. The excised cecum and colon were flushed together by injecting phosphate-buffered saline into the cecum via a syringe inserted into the hole derived from the separation of the cecum from the ileum. The luminal contents that then exited the

rectum, including fecal pellets, were collected. Biopsy samples (~10 mg) prepared from colon tissue were snap frozen in liquid nitrogen. The intestine was fixed in 4% paraformaldehyde and prepared for both immunohistochemistry and pathology evaluation using an established method (Stevceva *et al.*, 1999; Williams *et al.*, 2001; Pflegerl *et al.*, 2009). Flushed lumen contents were homogenized, collected by centrifugation and either snap frozen for DNA and RNA purification or fixed in 2% paraformaldehyde overnight at 4 °C and then stored in 70% ethanol/30% phosphate-buffered saline at −20 °C until analysis (see Supplemental Methods for more details). All animal experiments were discussed and approved by the institutional ethics committee and conducted in accordance with protocols approved by the Austrian laws (BMWF-66.006/0002-II/10b/2010).

DNA and RNA purification
Nucleic acids were extracted using a phenol-chloroform bead-beating procedure (Griffiths *et al.*, 2000). DNA and RNA were purified with a kit (Qiagen AllPrep DNA/RNA Mini kit, Qiagen, Hilden, Germany), and for RNA isolation with an on-column DNase digestion performed according to the manufacturer's instructions (Qiagen). DNA and RNA quality and quantity were assessed with agarose gel electrophoresis, spectrophotometry (NanoDrop 1000, Thermo Scientific, Waltham, MA, USA) and microfluidic electrophoresis (Experion, Bio-Rad, Hercules, CA, USA). DNA and RNA purified from the same extractions were used for primer-targeted 16S rRNA (gene) as well as metatranscriptomic analyses.

Preparation of 16S rRNA (gene) amplicons
PCR primers targeting a fragment of the 16S rRNA gene (V6–V9 region) of most bacteria were employed (909F, 5′-ACTCAAAKGAATWGACGG-3′ and 1492R, 5′-NTACCTTGTTACGACT-3′) (Berry *et al.*, 2011). In addition to the specific primers, pyrosequencing primers included the sequencing primer and an 8-nt barcode (Hamady *et al.*, 2008). Amplicon libraries were produced from DNA and RNA of both lumen contents (pooled cecum and colon) and biopsies. RNA was reverse transcribed and amplified using the Access RT-PCR System (Promega, Madison, WI, USA) and specific primers (909F/1492R). A two-step, low cycle number PCR procedure was used to amplify template DNA and cDNA, and to minimize bias associated with barcoded pyrosequencing primers as described previously (Berry *et al.*, 2011). PCR amplicons from triplicate amplifications were pooled and purified using Agencourt AMPure beads (Beckman Coulter Genomics, Brea, CA, USA) and quantified with a fluorescent-stain-based kit (Quant-iT PicoGreen, Invitrogen, Carlsbad, CA, USA).

Preparation of metatranscriptomic libraries
Total RNA from lumen (cecum and colon) biomass of replicate mice was pooled. Microbial rRNA was depleted and/or mRNA was enriched from pooled total RNA of control and DSS wt mice using a combination of the RiboMinus (Invitrogen), MicrobeExpress (Ambion, Austin, TX, USA) and/or MicrobeEnrich (Ambion) kits as indicated in Supplementary Table S1. RNA of processed and non-processed samples was reverse transcribed using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) with modifications. Briefly, first-strand cDNA was synthesized via incubation at 37 °C for 4 h and immediately used for second-strand synthesis (16 °C, 4 h). Double-stranded cDNA was purified by phenol extraction, residual RNA was digested (RNase A, Fermentas, Glen Burnie, MD, USA), and the remaining cDNA was again purified by phenol extraction.

Pyrosequencing
Pyrosequencing was performed with Titanium reagents on a 454 genome sequencer FLX (Roche, Basel, Switzerland) as recommended by the manufacturer. Pyrosequencing reads were quality filtered using the automatic amplicon pipeline of the GS Run Processor (Roche) to remove adapter sequences and low-quality reads. Reads were also quality filtered using LUCY (Chou and Holmes, 2001).

16S rRNA (gene) amplicon data analysis
Sequencing reads were de-multiplexed using QIIME (Caporaso *et al.*, 2010) and clustered with UCLUST (Edgar, 2010) into phylotypes (operational taxonomic units) at 97% identity (with a minimum length of 250 nt). Further details about data processing and quality control steps can be found in Supplemental Methods.

Taxonomic classification was made using the Ribosomal Database Project naïve Bayesian classifier (Wang *et al.*, 2007). Rarefaction curves, alpha diversity metrics and UniFrac distances (Lozupone and Knight, 2005) calculated using QIIME (Caporaso *et al.*, 2010) employed re-sampling (bootstrapping and jackknifing: 1000 re-samples) at below the size of the smallest library to avoid sample size-based artifacts (Lozupone *et al.*, 2011). Relative abundance correlation between pairs of phylotypes detected in at least two samples was performed in MATLAB (MathWorks, Natick, MA, USA) by calculating the Pearson correlation coefficient using amplicon libraries produced from lumen DNA. This analysis was also conducted with phylotypes divided into taxonomic families based on the results of the Ribosomal Database Project naïve Bayesian classifier. Non-parametric permutational multivariate analysis of variance (perMANOVA) was conducted using the 'vegan' package (Oksanen *et al.*, 2010) was performed using the 'indicspecies' package in R (De Cáceres and Legendre, 2009). The indicator species analysis determines the strength of the association between a phylotype and a condition

and considers the relative frequency and abundance of phylotypes in target versus non-target conditions (De Cáceres and Legendre, 2009). To focus on dominant indicators, indicator phylotypes were selected that (1) were significantly associated with DSS-treated or healthy mice using indicator species analysis or correlation analysis ($P<0.05$), and (2) had an arithmetic average difference of $>0.5\%$ relative abundance between healthy and DSS-treated mice. Genotype-insensitive as well as genotype-sensitive (that is, specific to either wt or STAT1$^{-/-}$) indicator phylotypes were identified in separate analyses using data from DNA and cDNA templates. Representative indicator phylotype sequences were added to a bootstrapped RAxML phylogenetic tree as described in Supplemental Methods.

Metatranscriptomic data analysis

Metatranscriptomic sequencing data were analyzed following the double RNA analysis pipeline described by Urich *et al*. (2008). Briefly, rRNA tags present in non-rRNA depleted samples (IDs 1, 6, 8, 9 Supplementalry Table S1) were taxonomically binned using MEGAN (Huson *et al*., 2007) and a custom reference database of small subunit rRNA sequences (Urich *et al*., 2008). mRNA tags were compared against the NCBI non-redundant database using BlastX and taxonomically classified using MEGAN. Multiple mRNA libraries from the same condition were combined (wt control: IDs 1–5, wt DSS: IDs 6 and 7) (Supplementary Table S1 and S5). SEED categories were assigned using the MG-RAST Server (Version 2, significance threshold: E-value 10^{-5}) (Meyer *et al*., 2008). Statistically significant differences between metatranscriptomes of the pooled samples were identified using STAMP (STatistical Analysis between Metagenomic Profiles) (Parks and Beiko, 2010). The abundance of selected transcripts was measured with quantitative PCR, as described in Supplemental Methods.

All pyrosequencing data in the study, including amplicon and metatranscriptomic data, are archived at NCBI Sequence Read Archive under Accession SRP008057.

Fluorescence in situ hybridization

Oligonucleotide probes specific for the 16S rRNA of target organisms were hybridized as described previously (Daims *et al*., 2005) (Supplementary Table S2). Details about newly designed probes and re-evaluated published probes are available in Supplemental Methods. Hybridized samples were imaged on a confocal scanning laser microscope (Zeiss 510 Meta, Oberkochen, Germany) and duplicate hybridizations were performed on each sample for quantitative FISH. For each quantification, at least 20 fields of view ($\times 63$) from each sample were analyzed with *daime* image analysis software (Daims *et al*., 2006).

Results

Colitis development in the murine model

Mice began to lose weight 6–7 days after the start of DSS treatment and weight loss continued until the end of the experiment on day 10. DSS treatment affected wt mice more than STAT1$^{-/-}$ mice as demonstrated by the following observations: wt animals lost more weight on days 8, 9 and 10 (paired *t*-test, $P<0.05$), three succumbed prematurely on day 9 (Figure 1a); and the intestinal tissue of the two remaining mice had more inflammatory infiltrate and more crypt damage on day 10, as determined by pathology scoring (Figure 1b, Supplementary Figure S1). Immunohistochemistry staining targeting phosphorylated-STAT1 confirmed that STAT1 was activated in the epithelial tissue and inflammatory infiltrates of DSS-treated wt mice. Phosphorylated-STAT1 was not detectable in STAT1$^{-/-}$ mice (Figures 1c and d).

16S rRNA gene-based surveys of the bacterial community in the murine intestine

In-depth analysis of bacterial communities was performed with amplicon sequence libraries from luminal DNA (comparisons made at a library size of 2500 sequences). Comparative analyses of template (DNA vs RNA) and sample location (biopsy vs lumen) are reported in the supplement (see Supplemental Results, Supplementary Figures S2, S3, Supplementary Table S3). For DNA-based lumen samples, DSS treatment was the largest factor driving community composition (perMANOVA, $P<0.001$), but genotype was also a significant factor in determining intestinal microbiota composition (perMANOVA, $P=0.019$). PCoA also revealed some clustering by genotype for unweighted UniFrac, though not for weighted UniFrac (Supplementary Figure S4), indicating that genotype differences are due to the presence and/or absence of rarer phylotypes. There were no significant differences in alpha diversity metrics between DSS treatment and control groups or between genotypes (Supplementary Table S4) and no consistent differences in rarefaction curves of samples from different genotypes and treatments (Supplementary Figure S5).

16S rRNA gene libraries were used to compare shifts in bacterial taxon abundance between samples from untreated and DSS-treated mice. All relative abundance shifts stated below are arithmetic averages of replicates and statistically significant ($P<0.05$). The overall abundances of the dominant phyla Firmicutes (58–65%) and Bacteroidetes (27–31%) were not affected by DSS treatment, but taxa within these two phyla showed clear changes in abundance. DSS treatment increased abundance of unclassified Clostridiales (1.8–4.1%) in mice of both genotypes and of Ruminococcaceae (6–16%) in STAT1$^{-/-}$ mice (Supplementary Figure S6).

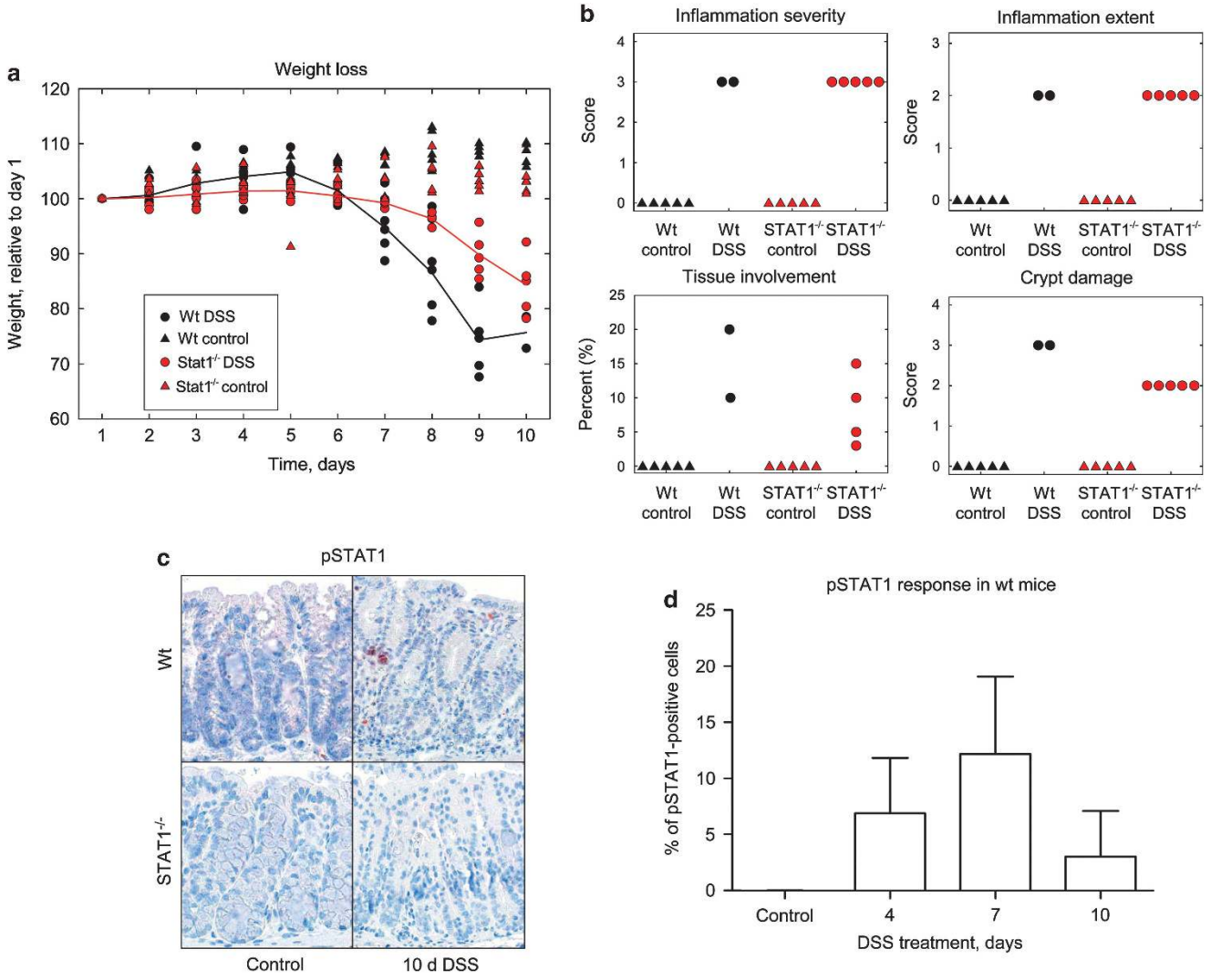

Figure 1 Effect of DSS treatment on weight loss, crypt damage and STAT1 phosphorylation in wt and STAT1$^{-/-}$ mice. (**a**) The body weight of each mouse is expressed as the percent of its weight at the start of the experiment. Weights for each mouse are presented for wt (black) and STAT1$^{-/-}$ (red) control (triangle) and DSS-treated (circle) mice. The arithmetic averages for replicates are plotted as solid lines for DSS-treated mice. (**b**) Pathology scoring of hematoxylin and eosin stained intestinal tissue from DSS-treated wt and STAT1$^{-/-}$ mice in a blinded evaluation according to the literature (Stevceva *et al.*, 1999). (**c**) Representative staining for phosphorylated-STAT1 (pSTAT1) in intestinal tissue of a DSS-treated wt and STAT1$^{-/-}$ mouse. (**d**) Quantification of pSTAT1 in tissues of wt mice over the course of DSS treatment.

The abundance of Bacteroidaceae (1.6–4.4%) increased but abundances of unclassified Bacteroidales (16.5–9.6%) and Rikenellaceae (2.7–1.5%) decreased upon DSS treatment in both genotypes. The less abundant but consistently detected phyla, Proteobacteria, Verrucomicrobia and Deferribacteres, were each dominated by a single family or genus, Enterobacteriaceae, *Akkermansia* and *Mucispirillum*, respectively. *Akkermansia* increased in DSS-treated mice of both genotypes (ND (not detectable) to 2.2%), whereas *Mucispirillum* (0.2–1.6%) and Enterobacteriaceae increased in wt mice only (ND to 4%) (Supplementary Figure S6).

To determine if the degree of phylotype kinship influences phylotype dynamics, we compared the Pearson correlation coefficient of phylotype abundance and the genetic distance (that is, 16S rRNA sequence dissimilarity) between all phylotype pairs (using the 16S rRNA gene sequence reads) in control and DSS-treated wt and STAT1$^{-/-}$ mice. Most phylotype pairs (98.3%) showed no significant correlation. Strong positive and negative correlations were present at varying levels of genetic distance, and there were far fewer strong negative correlations (Figure 2a). The average correlation of phylotypes assigned to the same taxonomic family, which is a measure of the amount of ecological cohesion (Webb *et al.*, 2002), was in all cases positive, but showed substantial diversity (Figure 2b). This diversity was attributable to both the number of phylotypes present in the family (that is, phylotype richness) and average within-family

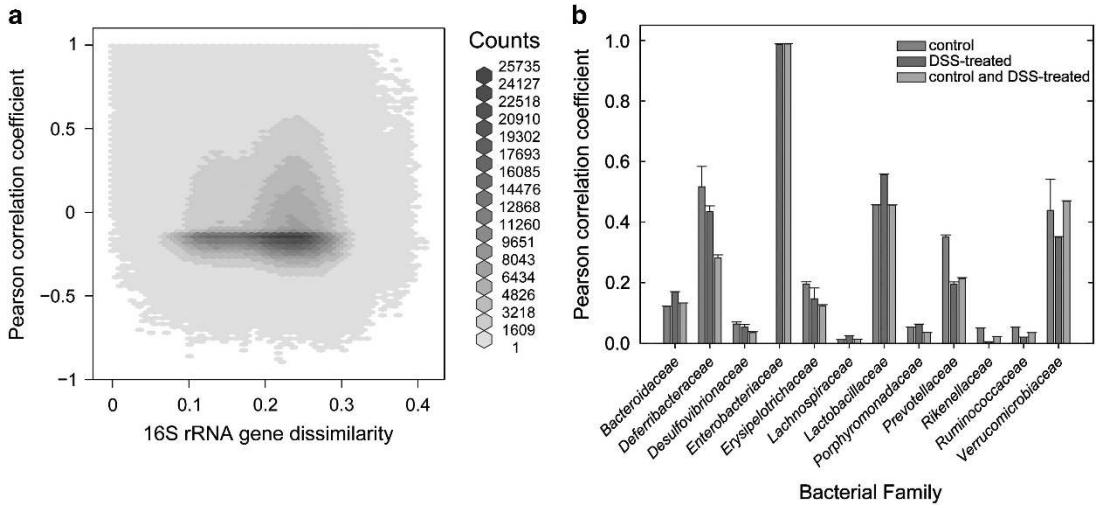

Figure 2 Pearson correlation coefficient of phylotype relative abundance between all phylotype pairs. The Pearson correlation coefficient was calculated for each pair of phylotypes by calculating the correlation of their relative abundances in all DNA-based intestinal lumen libraries. (**a**) The correlation coefficients are plotted against the sequence dissimilarity of the 16S rRNA genes being compared. (**b**) The arithmetic average correlation coefficient (including both positive and negative correlations) of phylotypes assigned to the same taxonomic family, which is a proxy for the amount of ecological cohesion within a family. Whiskers indicate s.e.m.

genetic distance. This variation in within-family characteristics was also observed when samples from control mice and DSS-treated mice were considered separately (Figure 2b). Family-level ecological cohesion decreased with both increasing phylotype richness ($R^2 = 0.76$) and, to a lesser extent, increasing average within-family genetic distance ($R^2 = 0.54$) (Supplementary Figure S7).

The results of the correlation analysis suggested that family-level characterization could hide important health-state indicator taxa, and therefore a phylotype-level indicator analysis was subsequently performed on both 16S rRNA gene and cDNA data sets to identify phylotypes indicative of specific conditions (healthy vs DSS-treated and/or wt vs STAT1$^{-/-}$ mice). Although 468 indicators were identified as statistically significant ($P < 0.05$), only numerically dominant indicators, phylotypes with a mean relative abundance that was at least 0.5% greater in the condition for which they were indicators (for example, a shift from not detected to 0.5%, or of 1–1.5%), were examined to restrict the analysis to the more abundant members of the microbiota. In all, 61 numerically dominant indicator phylotypes were identified: 26 for healthy mice and 35 for DSS-treated mice (Figure 3). A higher number of genotype-specific indicators were identified for DSS-treated mice (63%) than for untreated mice (27%) because a relatively large number of specific indicator phylotypes were identified in DSS-treated wt mice (17/35). Roughly half of the phylotypes were indicators for both RNA and DNA template (52%), but 39% appeared as indicators only for DNA and not RNA, which is most likely attributable to deeper sequencing of 16S rRNA amplicon libraries generated from DNA than from RNA. Most indicators of healthy gut microbiota were shared between both genotypes (19/26), though genotype-specific indicators were also observed (Figure 3).

All indicator phylotypes in the Ruminococcaceae and unclassified Clostridiales (that is, annotated as Clostridiales in Figure 3) were indicators for DSS treatment and all non-Bacteroidaceae Bacteroidales were indicators for untreated mice (Figure 3), which was consistent with the shifts observed at the family level (above). Most of the indicator phylotypes, however, belonged to the Lachnospiraceae and did not cluster in phylogenetic analysis according to the condition for which they were an indicator (Figure 4). Quantitative FISH using specific probes confirmed the sequencing data for four indicator phylotypes: *Akkermansia muciniphila*, *Mucispirillum schaedleri* and two Lachnospiraceae phylotypes (Figure 5). A strong correlation was observed between relative abundances as measured by FISH and by sequencing libraries ($R^2 = 0.80$, Supplementary Figure S8), confirming that relative shifts in abundant 16S rRNA (gene) phylotypes can be reliably monitored using the two-step PCR pyrosequencing approach (Berry *et al.*, 2011). Abundant phylotype targets, however, were consistently under-represented in sequencing libraries relative to FISH quantification, which may reflect PCR suppression of abundant sequences.

Metatranscriptomic analysis

Pyrosequencing was performed on cDNA obtained from total RNA as well as mRNA-enriched RNA. In the latter case, different enrichment strategies were tested, which resulted in up to 24% mRNA

Figure 3 Dominant indicator phylotypes of health state and genotype. Dominant indicator phylotypes are presented with their relative abundance (as percent) in each sample of DNA- and RNA-based sequencing libraries. Each phylotype is annotated with the closest reference sequence from the Living Tree Project (Yarza *et al.*, 2008), and information is in bold font when they share at least 97% sequence similarity. Dominant indicators are significantly associated with a specific condition and show >0.5% difference in relative abundance between two conditions (for details see Materials and Methods). The phylotypes are grouped and color coded according to the condition for which they are an indicator for each combination of treatment and genotype. Each column in the 'DNA template' and 'RNA template' blocks represents a sample from a single mouse. A representative sequence for each indicator phylotype is given in the supplement. Detailed information on sample numbers and sequencing depth is available in Supplementary Figure S3.

(Supplementary Table S1). The relative abundances of higher-level taxonomic groups determined with total RNA-derived 16S rRNA (data set sizes between 18 889 and 49 388 ribo-tags) were similar to 16S rRNA gene and cDNA amplicon libraries and also showed an increase in the orders Enterobacteriales, Verrucomicrobiales and Deferribacterales in DSS-treated mice (Figure 6). No ribo-tags of archaea, protists or fungi were detected in any samples. Confirming this result, no yeasts were detected by FISH analysis (probe PF2, Supplementary Table S2. FISH data not shown), which is consistent with the reports of low and non-detection of fungi in human (Gosalbes *et al.*, 2011) and piglet metatranscriptomes (Poroyko *et al.*, 2010), respectively. Taxonomic binning of mRNA with MEGAN produced relative abundance profiles similar to those observed using 16S rRNA, with substantial contributions from the phyla Firmicutes, Bacteroidetes and Proteobacteria (Figure 6). This allowed us to assign these mRNAs with high confidence to the respective taxa. A global classification of mRNA using MG-RAST revealed several functional categories (level 1 subsystems) differentially abundant between untreated and DSS-treated mice of both genotypes (STAMP analysis: $P<0.01$, 10 for wt, 2 for STAT1$^{-/-}$) (Supplementary Figure S9, A and C). Notably, DSS treatment was associated with a marked decrease in flagellum-associated transcripts (level 3 subsystem), a shift that was attributable to

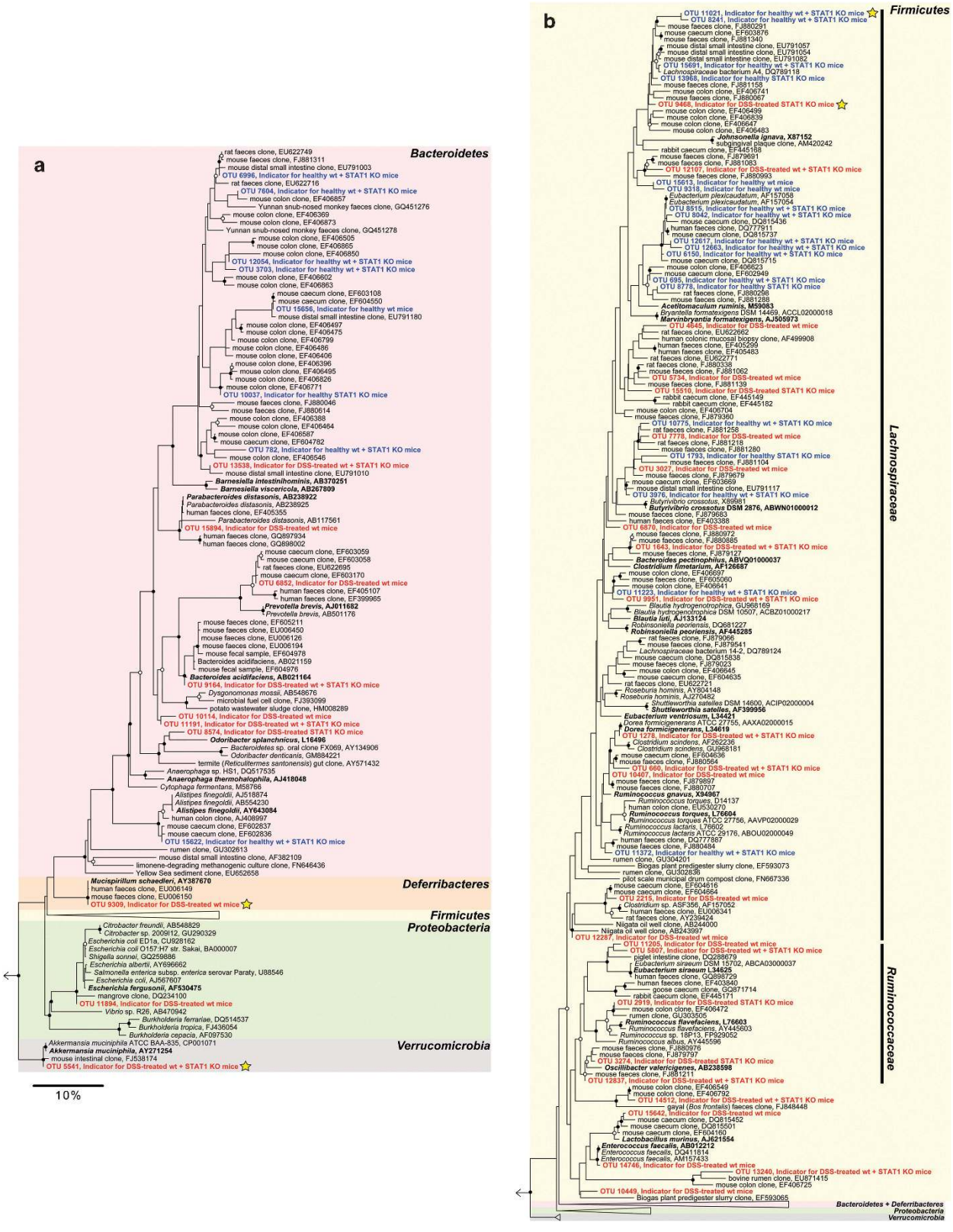

Figure 4 Phylogenetic tree of dominant indicator OTUs. Indicator phylotypes for DSS-treated (red) and control (blue) mice were added to a bootstrapped RAxML tree of full-length sequences via quick-add parsimony without modifying tree topology to demonstrate the approximate phylogenetic placement of indicators. The tree is expanded to show most phyla in (**a**) and the Firmicutes in (**b**). Living Tree Project reference sequences are indicated in bold font. Nodes with bootstrap support of at least 75% (○) and 90% (●) are marked and phylotypes that were quantified with FISH are indicated with a star.

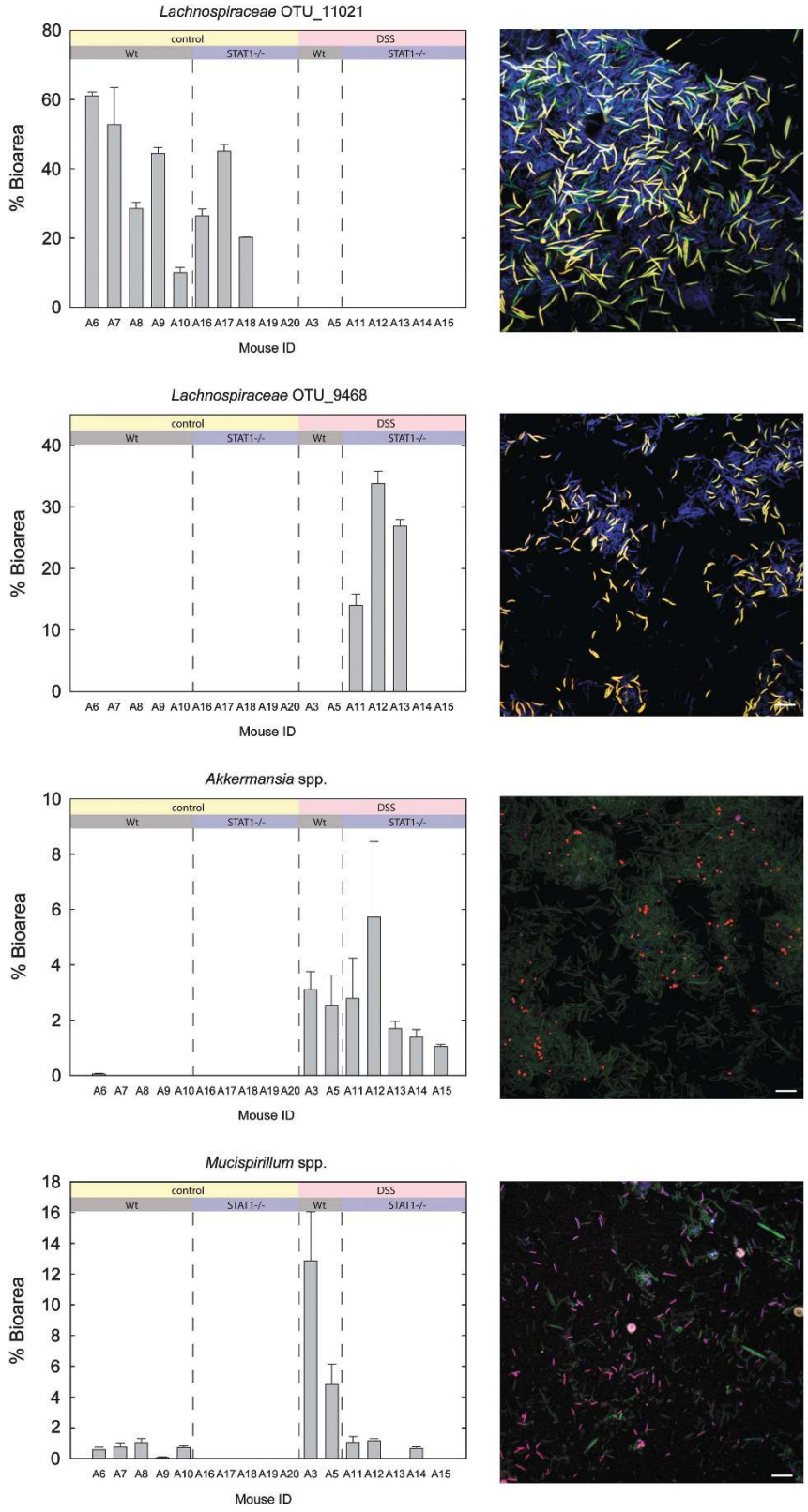

Figure 5 Quantitative FISH for selected indicator phylotypes. The relative abundance, represented as percent bacterial bio-area (the area of all microbes detected with the EUB338 probe set), of selected indicators is shown for each sample. Error bars indicate the s.d. from technical replication of target quantification. Samples are lumen biomass pooled from the cecum and colon. Each sample is shown to illustrate the significant inter-individual variability. Example images are shown for each probe. The specific probe is shown in yellow for Lachnospiraceae OTU_11021 and Lachnospiraceae OTU_9468, red for *Akkermansia* and purple for *Mucispirillum*. Other bacteria are shown in either blue or green. The white bar on each image indicates a length of 10 μm. No fluorescence was observed when samples were hybridized with the NONEUB probe (negative control) under these conditions.

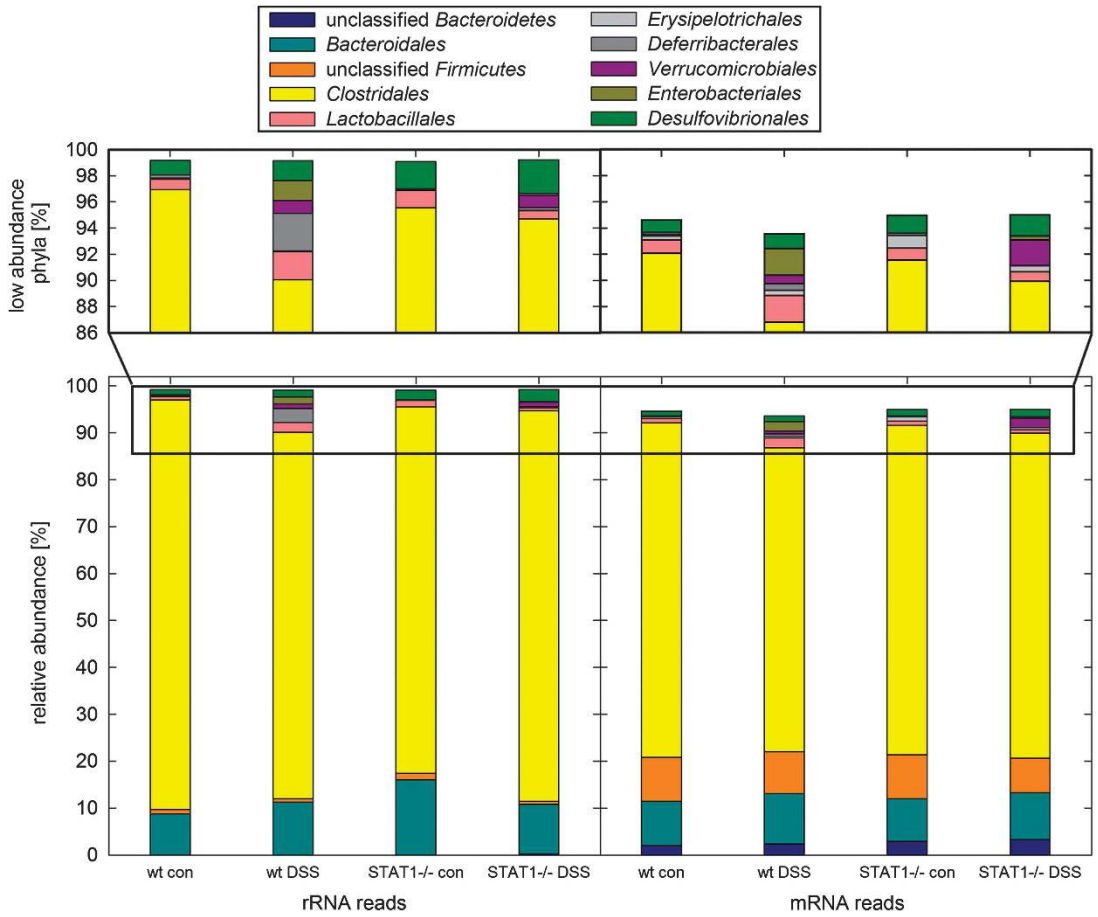

Figure 6 Taxonomic composition of mRNA and rRNA reads from metatranscriptome libraries. Assignment of rRNA (left) and mRNA (right) on the 'order' level was determined with MEGAN and the relative abundance of the major groups is presented as a color-coded stacked column graph for each of the pooled samples. Five replicate mice were pooled for all conditions except DSS-treated wt mice, which had two replicates. The abundance of reads that could not be classified is the difference between the total of each stacked column and 100%. The inset at the top of the figure is a magnification to more clearly show the change in relative abundance of less abundant groups. Con, control.

transcripts derived from the Lachnospiraceae (Supplementary Figure S9, B and E).

Discussion

Global shifts in microbial community structure in DSS-induced colitis

To develop a deeper understanding of the structure and function of intestinal microbiota in the DSS model, this study investigated differences in lower intestinal microbiota in wt and STAT1$^{-/-}$ mice with and without severe DSS-induced colitis. Wt mice experienced increased weight loss and colitis symptoms compared with STAT1$^{-/-}$ mice (Figures 1a and b), as expected from a previous report (Bandyopadhyay *et al.*, 2008). Non-treated STAT1$^{-/-}$ mice had a significantly different microbiota composition compared with wt mice (perMANOVA, $P=0.019$). This may indicate that STAT1 influences the interaction between commensals and the immune system, which is similar to the effect of a deficiency in the T-box transcription factor T-bet (Garrett *et al.*, 2007), and which may in part be due to role of STAT1 in regulating T-bet expression (Afkarian *et al.*, 2002). A confirmation of this hypothesis would require additional testing with littermate controls to rule out any possible maternal effects (Spor *et al.*, 2011). DSS-treated mice had a distinctive microbial community composition in the amplicon pyrosequencing libraries (Supplementary Figures S2-S4), which was also observed in a study that used a shallower sequencing depth (based on 16S rRNA gene clone libraries) (Nagalingam *et al.*, 2011). Although it is most likely that shifts in the microbial community are caused by the severe inflammation, it cannot be ruled out that shifts are not also due to either a direct or indirect effect of the DSS itself. It has been shown, for example, that DSS increases intestinal mucus permeability and bacterial penetration to the epithelial tissue (Johansson *et al.*, 2010), which could be expected to open up new niches for the bacterial community before the onset of inflammation. Importantly, however, gut

microbiota is unable to degrade DSS under anoxic conditions (Kitajima *et al.*, 2002), a finding that we confirmed with *in vitro* incubations (data not shown). Loss in bacterial species richness has been reported in both DSS-induced colitis (Nagalingam *et al.*, 2011) and human IBD (Willing *et al.*, 2010; Walker *et al.*, 2011). In the present study, the mean bacterial phylotype richness was reduced in DSS-treated mice, though the reduction was not statistically significant owing to large inter-mouse variability (Supplementary Table S4, Supplementary Figure S5). This may be due to the relatively quick induction of acute colitis in the present experiment; loss of phylotype richness may occur most significantly over longer time scales in chronic, relapsing inflammatory disease.

Abundance of the two major phyla, Firmicutes and Bacteroidetes, and the dominant orders within these phyla, Clostridiales and Bacteroidales, respectively, remained stable irrespective of treatment or genotype. However, abundance of some higher-level bacterial groups (approximately family level) shifted in treated mice. DSS-treated mice of both genotypes had increased Bacteroidaceae, unclassified Clostridiales, Verrucomicrobaceae, and Deferribacteraceae, and decreased unclassified Bacteroidales and Rikenellaceae (Supplementary Figure S6). Ruminococcaceae increased in DSS-treated STAT1$^{-/-}$ mice and Enterobacteriaceae increased in DSS-treated wt mice. As facultative anaerobes, members of the Enterobacteriaceae may profit from increased oxygen availability or resistance to reactive oxygen species produced during inflammation. Increases in Enterobacteriaceae have been observed in human IBD and murine colitis as well as after antibiotic treatment and infection by enteric pathogens (Lupp *et al.*, 2007; Hill *et al.*, 2010; Stecher *et al.*, 2010). Because the relative abundance of Enterobacteriaceae increases under a variety of treatments, this group seems to be a general indicator of a disrupted intestinal microbiota, though not necessarily a trigger of colitis (Bloom *et al.*, 2011). Differences between DSS-treated wt and STAT1$^{-/-}$ mice most likely reflect the differences observed in the degree of inflammation.

Dramatic dynamics within the abundant Lachnospiraceae

Molecular surveys of intestinal microbiota commonly classify sequences to higher-level taxonomic groups (for example, phylum, order and family) to characterize community composition and dynamics (Ley *et al.*, 2005; Frank *et al.*, 2007; Lupp *et al.*, 2007). This approach assumes that all organisms belonging to the same taxon fill the same, or similar, ecological niche and display weak intra-taxon competition. A recent study of the mouse gut microbiota provided partial support for this assumption by demonstrating a relationship between phylogenetic similarity and co-occurrence, which has been succinctly captured by the maxim 'like will to like' (Stecher *et al.*, 2010). There is, however, no clear threshold for the level of genetic relatedness necessary for cohesive or repulsive ecological forces to act upon organisms (Achtman and Wagner, 2008), and it is therefore impossible to predict *a priori* if all members of a phylogenetic group have similar functional properties.

To examine this at the level of phylogenetic resolution currently attainable with 454 sequencing reads of 16S rRNA genes, we used a similar approach as Stecher *et al.* (2010) to compare the correlation of phylotype relative abundances across mouse gut microbiota samples with their 16S rRNA sequence dissimilarity for all phylotype pairs. Relationships between most phylotype pairs at all levels of genetic distance seemed to be neutral in these mouse gut communities (Figure 2a). Positive and negative correlations were nevertheless evident at all levels of genetic distance, but strong negative correlations, indicating antagonistic interactions, were hardly apparent for very similar phylotypes (>95%). One example of antagonism between distantly related organisms in the human intestines is how *Bacillus thuringiensis* produces a narrow-spectrum bacteriocin, thuricin CD, which specifically targets *Clostridium difficile* (Rea *et al.*, 2010). Strong positive correlations of closely related phylotypes (>95%) hint to the presence of groups of organisms that occupy a similar realized niche but do not compete with each other, whereas strong positive correlation of distantly related phylotypes are better explained either through convergence (independently evolved to occupy a similar niche) or through synergistic interactions (Webb *et al.*, 2002). An example of the latter would be functional cooperation of fermenting and hydrogen-consuming microorganisms from different bacterial or archaeal lineages in syntrophic substrate degradation, hypothetically exemplified in this study by the correlation ($R = 0.77$) of a putatively fermenting phylotype (Bacteroidales OTU_12071) and a putatively hydrogen-scavenging, sulfate-reducing phylotype (*Desulfovibrio* OTU_9264).

Similar results were attained when repeating the correlation analysis using the Spearman coefficient (data not shown). It is unlikely that intragenomic variation in 16S rRNA gene copies bias our analysis because the most typical level of variation (<1% sequence divergence) (Acinas *et al.*, 2004) is smaller than the phylotype clustering threshold (3% divergence).

We found that within-family correlations of phylotype pairs were strikingly divergent between families (Figure 2). The strength of within-family cohesion decreased with increasing within-family phylotype richness ($R^2 = 0.76$), and, to a lesser extent, with increasing average genetic distance ($R^2 = 0.54$) (Supplementary Figure S7), such that low richness families (for example,

Enterobacteriacae, Verrucomicrobaceae, Deferribacteraceae and Lactobacillaceae) were more cohesive than high richness families (for example, Lachnospiraceae and Ruminococcaceae) (Figure 2). The lack of cohesiveness in high richness families warns against inferring ecological properties based on high-level taxonomic groupings, such as is sometimes recommended (Philippot *et al.*, 2010), because it has the effect of reducing detection sensitivity by 'hiding' organisms that behave differently but are grouped in the same family.

We addressed this issue by applying an indicator phylotype approach, which uses the 16S rRNA (gene) 454 pyrosequencing data to its maximum resolution to identify significant changes in abundance for each phylotype. The phylotype indicator analysis produced 61 numerically important indicators of different 'ecological states' of the mouse-intestinal microbiota symbiosis: treatment (untreated vs exposure to DSS) and/or genotype (wt vs STAT1$^{-/-}$) (Figure 3). There is evidence of phylogenetic clustering of some of these indicators. For example, all the indicator phylotypes in the Ruminococacceae and Bacteroidaceae were indicators for DSS-treated mice and non-Bacteroidaceae Bacteroidales were indicators for untreated mice (Figure 3). In agreement with the variability in the phylotype correlation analysis, however, the abundant clostridial family Lachnospiraceae included indicators for both DSS-treated and -untreated mice (Figure 4).

Endpoint PCR-based and pyrosequencing-based approaches have inherent biases that can reduce the ability to consider the results quantitative (Polz and Cavanaugh, 1998; Gomez-Alvarez *et al.*, 2009). To confirm that members of the highly divergent and phylotype-rich Lachnospiraceae family have contrasting dynamics, we quantified the abundances of the two most abundant phylotypes (which share 93% sequence similarity over the V8–V9 region of the 16S rRNA) using FISH. These closely related phylotypes represent two as-yet-uncultivated taxa that were identified as indicators for different conditions: one phylotype (OTU_11021) drastically decreased during inflammation in both genotypes, whereas the other one (OTU_9468) drastically increased during inflammation in STAT1$^{-/-}$ mice (Figure 5). Quantitative FISH analysis with specific probes confirmed that the phylotypes dominated bacterial biomass in some mouse lumen samples (up to $>$60% of the bacterial bio-area) and were undetectable in other samples, and that the dramatic shifts in abundance occurred under different conditions for the two phylotypes (Figure 5). Consistent with this observation, the metatranscriptome data also suggested shifting activity within the Lachnospiraceae, with increased abundance of Lachnospiraceae-affiliated transcripts related to regulation and cell signaling, carbohydrate metabolism and respiration, as well as decreased abundance for flagellum-associated transcripts in DSS-treated mice (Supplementary Figure S9). The most parsimonious explanation for these changes in the metatranscriptome is the major shift in the abundance of the two Lachnospiraceae phylotypes, although it may theoretically also be caused by differential transcription of the same cell population and not shifting composition. The loss of potential Lachnospiraceae flagellin transcripts in colitis may be due to a response of the immune system. Flagellins bind to and activate the membrane bound toll-like receptor 5 and the cytosolic receptor IPAF, thus triggering a transcriptional cascade (Hayashi *et al.*, 2001). Lachnospiraceae flagellins have been identified as dominant antigens in Crohn's disease (Duck *et al.*, 2007) and in spontaneous colitis in IL-10$^{-/-}$ mice, during which there is a loss of a flagellated Lachnospiraceae with high similarity (99%) to Lachnospiraceae bacterium A4 (DQ789118) (Ye *et al.*, 2008). Although Lachnospiraceae phylotype OTU_11021, which in our study was an abundant colonizer of the healthy mouse gut but absent in DSS-treated mice, is only moderately related to Lachnospiraceae bacterium A4 (94% 16S rRNA sequence similarity), it is tempting to speculate that OTU_11021 cells are also flagellated.

Putative mucus-associated bacteria in DSS-induced colitis

A subset of the intestinal microbiota benefits either obligately or facultatively from the degradation of host-derived mucus secretions (Hoskins and Boulding, 1981; Miller and Hoskins, 1981). Mucus composition and secretion are altered in inflammation (Ehsanullah *et al.*, 1982; Shirazi *et al.*, 2000; Einerhand *et al.*, 2002) and loss of mucus secretion and assembly of the major structural component, the mucin glycoprotein, can increase sensitivity to DSS and in some cases induce spontaneous colitis (Van der Sluis *et al.*, 2006; Heazlewood *et al.*, 2008; Park *et al.*, 2009). Additionally, there is some evidence that mucolytic bacteria and mucolytic activity are increased in inflammation (Png *et al.*, 2010). Consistent with this, we identified via pyrosequencing two mucus-degrading and/or -inhabiting phylotypes belonging to the genera *Mucispirillum* and *Akkermansia,* respectively, as indicators of DSS treatment and confirmed this by quantitative FISH (Figure 5). The closest cultured representative of the *Mucispirillum* phylotype is *M. schaedleri*, a distinctive spiral-shaped bacterium, which has been observed to be physically associated with the secreted mucus layer (Robertson *et al.*, 2005). Another member of this genus is a part of the gnotobiotic 'Altered Schaedler Flora' model (ASF 457) (Dewhirst *et al.*, 1999). Using specific FISH probes, we confirmed that the *Mucispirillum* phylotype in our mice also possesses a spiral morphology (Figure 5), which has been suggested to be an asset for movement through mucus (Berg and Turner, 1979). *Akkermansia* is a frequently observed member of the human and

animal gut microbiota and is believed to be a specialist in degradation of host-derived mucin (Derrien *et al.*, 2004). There are contradicting reports about shifts in the abundance of *Akkermansia* in human IBD (Derrien *et al.*, 2008; Png *et al.*, 2010). *Akkermansia* benefits from both the disrupted environment caused by antibiotic treatment (Hill *et al.*, 2010) and decreased supply of dietary derived nutrients, as has been observed in the fasting state of Syrian hamsters (Sonoyama *et al.*, 2009) and pythons (Costello *et al.*, 2010). In our study, the increased abundance of *Akkermansia* in DSS-treated mice was accompanied by detection of *Akkermansia* transcripts associated with mucin degradation, including glycosyl hydrolases (families 2 and 31) and beta-*N*-acetylhexosaminidase (Supplementary Figure S10). Furthermore, *Akkermansia* ruberythrin, an oxidative stress response protein that has been detected in the piglet gut (Poroyko *et al.*, 2010), increased in DSS-treated wt mice (Supplementary Figure S10), which likely reflects that the mucin-degrading *Akkermansia* are exposed to reactive oxygen or nitrogen species during inflammation. Although abundance shifts of known mucin degraders and transcript profiling of mucin-foraging genes strongly implicates a shift in mucus degradation, a direct demonstration of mucus foraging *in vivo* is urgently needed to confirm these hypotheses and to determine whether other phylotypes also contribute to mucus degradation.

Conclusions

Acute inflammation induced by DSS treatment had a major impact on the intestinal microbiota of wt and STAT1$^{-/-}$ mice marked by abundance shifts of higher-level taxonomic groups as well as phylotypes within the same family. Indicator phylotypes were identified for DSS-treated and -untreated mice, and among them putative mucus-degrading bacteria were increased in DSS-treated mice. Metatranscriptomic analysis demonstrated that the mucus degrader *Akkermansia muciniphila* expresses genes needed for mucin foraging and responds to inflammation by expressing the oxidative stress response protein ruberythrin.

Importantly, within-family variation in phylotype dynamics was observed and was particularly evident for the Lachnospiraceae, a family in which contrasting abundance shifts of dominant phylotypes were associated with a shift in transcriptional activity profiles. These data support our assertion that gut microbiota surveys strongly benefit from characterization at the level of maximum possible phylogenetic resolution and also highlight that lower resolution analyses (that is, at higher taxonomic levels) can fall short in detecting even large changes in phylotype abundances. In this context, it is worth bearing in mind that in some cases the 16S rRNA (gene)-based phylogenetic approach, even at the maximum possible level of resolution, still has insufficient resolution for detecting genetically similar organisms with divergent ecophysiology and health significance, such as between different virulent and commensal strains of *Escherichia coli* (Wirth *et al.*, 2006) and even between *Escherichia coli* and *Shigella* species (Fukushima *et al.*, 2002).

In summary, this study provides the first comprehensive view into the structure and function of intestinal microbiota in the DSS-induced acute colitis mouse model and demonstrates the power of integrating maximum phylogenetic resolution (phylotype level) 16S rRNA sequence analysis with quantitative FISH and metatranscriptomics—a combination that promises to greatly increase our understanding of the roles of the intestinal microbiota in health and disease.

Acknowledgements

This work was financially supported by the Austrian Federal Ministry of Science and Research (GEN-AU III InflammoBiota). We greatly thank Michael Schloter for access to pyrosequencing facilities. We also thank Julia Ramesmayer for technical assistance and Michael Pester for helpful discussions on bioinformatics and phylogenetic analysis.

References

Achtman M, Wagner M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* **6**: 431–440.

Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J Bacteriol* **186**: 2629–2635.

Afkarian M, Sedy JR, Yang J, Jacobson NG, Cereb N, Yang SY *et al.* (2002). T-bet is a STAT 1-induced regulator of IL-12 R expression in naive CD 4+ T cells. *Nat Immunol* **3**: 549–557.

Armougom F, Henry M, Vialettes B, Raccah D, Raoult D. (2009). Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and *Methanogens* in anorexic patients. *PLoS One* **4**: e7125.

Bandyopadhyay SK, CAdl Motte, Kessler SP, Hascall VC, Hill DR, Strong SA. (2008). Hyaluronan-mediated leukocyte adhesion and dextran sulfate sodium-induced colitis are attenuated in the absence of signal transducer and activator of transcription 1. *Am J Pathol* **173**: 1361–1368.

Berg HC, Turner L. (1979). Movement of microorganisms in viscous environments. *Nature* **278**: 349–351.

Berry D, Ben Mahfoudh K, Wagner M, Loy A. (2011). Barcoded pyrosequencing primers used in multiplex amplicon sequencing bias amplification. *Appl Environ Microbiol* **77**: 7846–7849.

Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL *et al.* (2011). Commensal *Bacteroides* species induce colitis in host-genotype-specific

fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**: 390–403.
Braun J, Wei B. (2007). Body traffic: Ecology, genetics, and immunity in inflammatory bowel disease. *Annu Rev Pathol: Mech Dis* **2**: 401–429.
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.
Chou H-H, Holmes MH. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics* **17**: 1093–1104.
Costello EK, Gordon JI, Secor SM, Knight R. (2010). Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME J* **4**: 1375–1385.
Daims H, Lücker S, Wagner M. (2006). daime, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* **8**: 200–213.
Daims H, Stoecker K, Wagner M. (2005). Fluorescence *in situ* hybridisation for the detection of prokaryotes. In: Osborn A, Smith C (eds) *Advanced Methods in Molecular Microbial Ecology*. Bios-Garland: Abingdon, UK, pp 213–239.
De Cáceres M, Legendre P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566–3574.
Decker T, Müller M, Stockinger S. (2005). The yin and yang of type I interferon activity in bacterial infection. *Nat Rev Immunol* **5**: 675–687.
Derrien M, Collado MC, Ben-Amor K, Salminen S, de Vos WM. (2008). The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Appl Environ Microbiol* **74**: 1646–1648.
Derrien M, Vaughan EE, Plugge CM, de Vos WM. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**: 1469–1476.
Dewhirst FE, Chien C-C, Paster BJ, Ericson RL, Orcutt RP, Schauer DB *et al.* (1999). Phylogeny of the defined murine microbiota: Altered Schaedler flora. *Appl Environ Microbiol* **65**: 3287–3292.
Dianda L, Hanby AM, Wright NA, Sebesteny A, Hayday AC, Owen MJ. (1997). T cell receptor-alpha beta-deficient mice fail to develop colitis in the absence of a microbial environment. *Am J Pathol* **150**: 91–97.
Duck LW, Walter MR, Novak J, Kelly D, Tomasi M, Cong Y *et al.* (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* **13**: 1191–1201.
Durbin JE, Hackenmiller R, Simon MC, Levy DE. (1996). Targeted disruption of the mouse *stat1* gene results in compromised innate immunity to viral disease. *Cell* **84**: 443–450.
Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
Ehsanullah M, Filipe MI, Gazzard B. (1982). Mucin secretion in inflammatory bowel disease: Correlation with disease activity and dysplasia. *Gut* **23**: 485–489.
Einerhand AWC, Renes IB, Makkink MK, van der Sluis M, Büller HA, Dekker J. (2002). Role of mucins in inflammatory bowel disease: Important lessons from experimental models. *Eur J Gastroenterol Hepatol* **14**: 757–765.
Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* **104**: 13780–13785.
Fukushima M, Kakinuma K, Kawaguchi R. (2002). Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J Clin Microbiol* **40**: 2779–2785.
Garrett WS, Lord GM, Punit S, Lugo-Villarino G, Mazmanian SK, Ito S *et al.* (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* **131**: 33–45.
Gomez-Alvarez V, Teal TK, Schmidt TM. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J* **3**: 1314–1317.
Gosalbes MJ, Durbán A, Pignatelli M, Abellan JJ, Jiménez-Hernández N, Pérez-Cobas AE *et al.* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* **6**: e17447.
Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* **66**: 5488–5491.
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.
Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR *et al.* (2001). The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**: 1099–1103.
Heazlewood CK, Cook MC, Eri R, Price GR, Tauro SB, Taupin D *et al.* (2008). Aberrant mucin assembly in mice causes endoplasmic reticulum stress and spontaneous inflammation resembling ulcerative colitis. *PLoS Medicine* **5**: e54.
Heimesaat MM, Fischer A, Siegmund B, Kupz A, Niebergall J, Fuchs D *et al.* (2007). Shift towards pro-inflammatory intestinal bacteria aggravates acute murine colitis via Toll-like Receptors 2 and 4. *PLoS One* **2**: e662.
Hill DA, Hoffmann C, Abt MC, Du Y, Kobuley D Kirn TJ *et al.* (2010). Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. *Mucosal Immunol* **3**: 148.
Hoskins LC, Boulding ET. (1981). Mucin degradation in human colon ecosystems: evidence for the existence and role of bacterial subpopulations producing glycosidases as extracellular enzymes. *J Clin Invest* **67**: 163.
Huson DH, Auch AF, Qi J, Schuster SC. (2007). MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.
Johansson MEV, Gustafsson JK, Sjöberg KE, Petersson J, Holm L, Sjövall H *et al.* (2010). Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS One* **5**: e12238.
Kitajima S, Morimoto M, Sagara E. (2002). A model for dextran sodium sulfate (DSS)-induced mouse colitis: bacterial degradation of DSS does not occur after incubation with mouse cecal contents. *Exp Anim* **51**: 203–206.
Kitakima S, Takuma S, Morimoto M. (1999). Changes in colonic mucosal permeability in mouse colitis induced with dextran sulfate sodium. *Exp Anim* **48**: 137–143.

Kühn R, Löhler J, Rennick D, Rajewsky K, Müller W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**: 263–274.

Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA* **102**: 11070.

Loftus EVJ. (2004). Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* **126**: 1504–1517.

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

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J* **5**: 169–172.

Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC *et al.* (2007). Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of *Enterobacteriaceae*. *Cell Host Microbe* **2**: 119–129.

Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L *et al.* (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205–211.

Meraz MA, White JM, Sheehan KCF, Bach EA, Rodig SJ, Dighe AS *et al.* (1996). Targeted disruption of the *stat1* gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway. *Cell* **84**: 431–442.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass E Kubal M *et al.* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinf* **9**: 386.

Miller R, Hoskins L. (1981). Mucin degradation in human colon ecosystems. Fecal population densities of mucin-degrading bacteria estimated by a 'most probable number' method. *Gastroenterology* **81**: 759–765.

Mueller C, Macpherson AJ. (2006). Layers of mutualism with commensal bacteria protect us from intestinal inflammation. *Gut* **55**: 276–284.

Murphy EF, Cotter PD, Healy S, Marques TM, O'Sullivan O, Fouhy F *et al.* (2010). Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut* **59**: 1635–1642.

Nagalingam NA, Kao JY, Young VB. (2011). Microbial ecology of the murine gut associated with the development of dextran sodium sulfate-induced colitis. *Inflamm Bowel Dis* **17**: 917–926.

Nell S, Suerbaum S, Josenhans C. (2010). The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nat Rev Micro* **8**: 564–577.

Okayasu I, Hatakeyama S, Yamada M, Ohkusa T, Inagaki Y, Nakaya R. (1990). A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* **98**: 694–702.

Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL *et al.* (2010). Vegan: community ecology package, R package version 1.17-4 (http://cran.r-project.org/).

Packey C, Sartor D, Balfour R. (2009). Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Curr Opin Infect Dis* **22**: 292–301.

Park SW, Zhen G, Verhaeghe C, Nakagami Y, Nguyenvu LT, Barczak AJ *et al.* (2009). The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proc Natl Acad Sci USA* **106**: 6950–6955.

Parks DH, Beiko RG. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715–721.

Pflegerl P, Vesely P, Hantusch B, Schlederer M, Zenz R, Janig E *et al.* (2009). Epidermal loss of JunB leads to a SLE phenotype due to hyper IL-6 signaling. *Proc Natl Acad Sci USA* **106**: 20423–20428.

Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB *et al.* (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Micro* **8**: 523–529.

Png CW, Linden SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI *et al.* (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. *Am J Gastroenterol* **105**: 2420–2428.

Polz MF, Cavanaugh CM. (1998). Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724–3730.

Poroyko V, White JR, Wang M, Donovan S, Alverdy J, Liu DC *et al.* (2010). Gut microbial gene expression in mother-fed and formula-fed piglets. *PLoS One* **5**: e12459.

Rath HC, Schultz M, Freitag R, Dieleman LA, Li F, Linde H-J *et al.* (2001). Different subsets of enteric bacteria induce and perpetuate experimental colitis in rats and mice. *Infect Immun* **69**: 2277–2285.

Rea MC, Sit CS, Clayton E, O'Connor PM, Whittal RM, Zheng J *et al.* (2010). Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proc Natl Acad Sci USA* **107**: 9352–9357.

Reiff C, Kelly D. (2010). Inflammatory bowel disease, gut bacteria and probiotic therapy. *Int J Med Microbiol* **300**: 25–33.

Robertson BR, O'Rourke JL, Neilan BA, Vandamme P, On SLW, Fox JG *et al.* (2005). *Mucispirillum schaedleri* gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *Int J Syst Evol Microbiol* **55**: 1199–1204.

Schreiber S, Rosenstiel P, Hampe J, Nikolaus S, Groessner B, Schottelius A *et al.* (2002). Activation of signal transducer and activator of transcription (STAT) 1 in human chronic inflammatory bowel disease. *Gut* **51**: 379–385.

Shirazi T, Longman RJ, Corfield AP, Probert CSJ. (2000). Mucins and inflammatory bowel disease. *Postgrad Med J* **76**: 473–478.

Sonoyama K, Fujiwara R, Takemura N, Ogasawara T, Watanabe J, Ito H *et al.* (2009). Response of gut microbiota to fasting and hibernation in Syrian hamsters. *Appl Environ Microbiol* **75**: 6451–6456.

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

Stecher B, Chaffron S, Käppeli R, Hapfelmeier S, Freedrich S, Weber TC *et al.* (2010). Like will to like: Abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathog* **6**: e1000711.

Stevceva L, Pavli P, Buffinton G, Wozniak A, Doe W. (1999). Dextran sodium sulphate induced colitis activity varies with mouse strain but develops in lipopolysaccharide unresponsive mice. *J Gastroen Hepatol* **14**: 54–60.

Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**: e2527.

Van der Sluis M, De Koning BAE, De Bruijn ACJM, Velcich A, Meijerink JPP, Van Goudoever JB *et al.* (2006). *Muc2*-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**: 117–129.

Walker A, Sanderson J, Churcher C, Parkes G, Hudspith B, Rayment N *et al.* (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol* **11**: 7.

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.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. (2002). Phylogenies and community ecology. *Ann Rev Ecol Sys* **33**: 475–505.

Williams KL, Randall Fuller C, Dieleman LA, DaCosta CM, Haldeman KM, Balfour Sartor R *et al.* (2001). Enhanced survival and mucosal repair after dextran sodium sulfate-induced colitis in transgenic mice that overexpress growth hormone. *Gastroenterology* **120**: 925–937.

Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z *et al.* (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**: 1844–1854.

Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH *et al.* (2006). Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* **60**: 1136–1151.

Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH *et al.* (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**: 241–250.

Ye J, Lee JW, Presley LL, Bent E, Wei B, Braun J *et al.* (2008). Bacteria and bacterial rRNA genes associated with the development of colitis in IL-10 – / – mice. *Inflamm Bowel Dis* **14**: 1041–1050.

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