

Diet rapidly and reproducibly alters the human gut microbiome

Lawrence A. David^{1,2†}, Corinne F. Maurice¹, Rachel N. Carmody¹, David B. Gootenberg¹, Julie E. Button¹, Benjamin E. Wolfe¹, Alisha V. Ling³, A. Sloan Devlin⁴, Yug Varma⁴, Michael A. Fischbach⁴, Sudha B. Biddinger³, Rachel J. Dutton¹ & Peter J. Turnbaugh¹

Long-term dietary intake influences the structure and activity of the trillions of microorganisms residing in the human gut^{1–5}, but it remains unclear how rapidly and reproducibly the human gut microbiome responds to short-term macronutrient change. Here we show that the short-term consumption of diets composed entirely of animal or plant products alters microbial community structure and overwhelms inter-individual differences in microbial gene expression. The animal-based diet increased the abundance of bile-tolerant microorganisms (*Alistipes*, *Bilophila* and *Bacteroides*) and decreased the levels of Firmicutes that metabolize dietary plant polysaccharides (*Roseburia*, *Eubacterium rectale* and *Ruminococcus bromii*). Microbial activity mirrored differences between herbivorous and carnivorous mammals², reflecting trade-offs between carbohydrate and protein fermentation. Foodborne microbes from both diets transiently colonized the gut, including bacteria, fungi and even viruses. Finally, increases in the abundance and activity of *Bilophila wadsworthia* on the animal-based diet support a link between dietary fat, bile acids and the outgrowth of microorganisms capable of triggering inflammatory bowel disease⁶. In concert, these results demonstrate that the gut microbiome can rapidly respond to altered diet, potentially facilitating the diversity of human dietary lifestyles.

There is growing concern that recent lifestyle innovations, most notably the high-fat/high-sugar ‘Western’ diet, have altered the genetic composition and metabolic activity of our resident microorganisms (the human gut microbiome)⁷. Such diet-induced changes to gut-associated microbial communities are now suspected of contributing to growing epidemics of chronic illness in the developed world, including obesity^{4,8} and inflammatory bowel disease⁶. Yet, it remains unclear how quickly and reproducibly gut bacteria respond to dietary change. Work in inbred mice shows that shifting dietary macronutrients can broadly and consistently alter the gut microbiome within a single day^{7,9}. By contrast, dietary interventions in human cohorts have only measured community changes on timescales of weeks¹⁰ to months⁴, failed to find significant diet-specific effects¹, or else have demonstrated responses among a limited number of bacterial taxa^{3,5}.

We examined whether dietary interventions in humans can alter gut microbial communities in a rapid, diet-specific manner. We prepared two diets that varied according to their primary food source: a ‘plant-based diet’, which was rich in grains, legumes, fruits and vegetables; and an ‘animal-based diet’, which was composed of meats, eggs and cheeses (Supplementary Table 1). We picked these sources to span the global diversity of modern human diets, which includes exclusively plant-based and nearly exclusively animal-based regimes¹¹ (the latter being the case among some high-latitude and pastoralist cultures). Each diet was consumed *ad libitum* for five consecutive days by six male and four female American volunteers between the ages of 21 and 33, whose body mass indices ranged from 19 to 32 kg m⁻² (Supplementary Table 2). Study volunteers were observed for 4 days before each diet arm to

measure normal eating habits (the baseline period) and for 6 days after each diet arm to assess microbial recovery (the washout period; Extended Data Fig. 1). Subjects’ baseline nutritional intake correlated well with their estimated long-term diet (Supplementary Table 3). Our study cohort included a lifetime vegetarian (see Extended Data Fig. 2, Supplementary Discussion and Supplementary Table 4 for a detailed analysis of his diet and gut microbiota).

Each diet arm significantly shifted subjects’ macronutrient intake (Fig. 1a–c). On the animal-based diet, dietary fat increased from 32.5 ± 2.2% to 69.5 ± 0.4% kcal and dietary protein increased from 16.2 ± 1.3% to 30.1 ± 0.5% kcal ($P < 0.01$ for both comparisons, Wilcoxon signed-rank test; Supplementary Table 5). Fibre intake was nearly zero, in contrast to baseline levels of 9.3 ± 2.1 g per 1,000 kcal. On the plant-based diet, fibre intake rose to 25.6 ± 1.1 g per 1,000 kcal, whereas both fat and protein intake declined to 22.1 ± 1.7% and 10.0 ± 0.3% kcal, respectively ($P < 0.05$ for all comparisons). Subjects’ weights on the plant-based diet remained stable, but decreased significantly by day 3 of the animal-based diet ($q < 0.05$, Bonferroni-corrected Mann–Whitney U test; Extended Data Fig. 3). Differential weight loss between the two diets cannot be explained simply by energy intake, as subjects consumed equal numbers of calories on the plant- and animal-based diets (1,695 ± 172 kcal and 1,777 ± 221 kcal, respectively; $P = 0.44$).

To characterize temporal patterns of microbial community structure, we performed 16S ribosomal RNA gene sequencing on samples collected each day of the study (Supplementary Table 6). We quantified the microbial diversity within each subject at a given time point (α diversity) and the difference between each subject’s baseline and diet-associated gut microbiota (β diversity) (Fig. 1d, e). Although no significant differences in α diversity were detected on either diet, we observed a significant increase in β diversity that was unique to the animal-based diet ($q < 0.05$, Bonferroni-corrected Mann–Whitney U test). This change occurred only 1 day after the diet reached the distal gut microbiota (as indicated by the food tracking dye; Extended Data Fig. 3a). Subjects’ gut microbiota reverted to their original structure 2 days after the animal-based diet ended (Fig. 1e).

Analysis of the relative abundance of bacterial taxonomic groups supported our finding that the animal-based diet had a greater impact on the gut microbiota than the plant-based diet (Fig. 2). We hierarchically clustered species-level bacterial phylotypes by the similarity of their dynamics across diets and subjects (see Methods and Supplementary Tables 7, 8). Statistical testing identified 22 clusters whose abundance significantly changed while on the animal-based diet, whereas only 3 clusters showed significant abundance changes while on the plant-based diet ($q < 0.05$, Wilcoxon signed-rank test; Supplementary Table 9). Notably, the genus *Prevotella*, one of the leading sources of inter-individual gut microbiota variation¹² and hypothesized to be sensitive to long-term fibre intake^{11,13}, was reduced in our vegetarian subject during consumption of the animal-based diet (see Supplementary

¹FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ²Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA. ³Division of Endocrinology, Children’s Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Bioengineering & Therapeutic Sciences and the California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, California 94158, USA. [†]Present address: Molecular Genetics & Microbiology and Institute for Genome Sciences & Policy, Duke University, Durham, North Carolina 27708, USA.

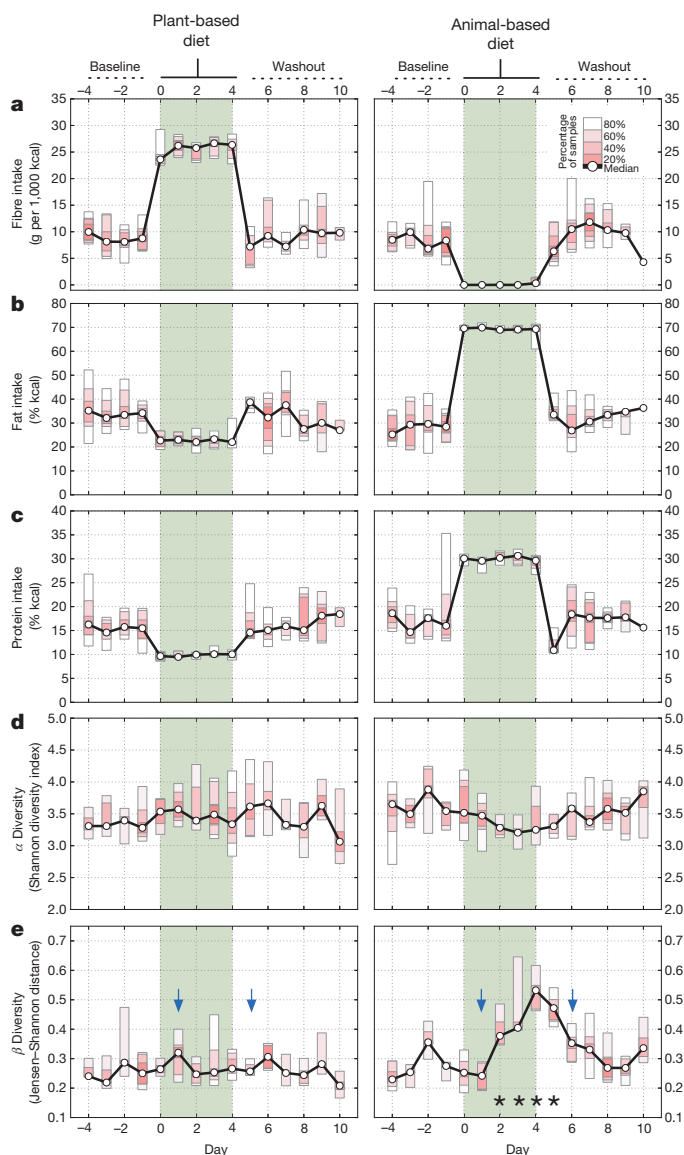


Figure 1 | Short-term diet alters the gut microbiota. **a–e**, Ten subjects were tracked across each diet arm. **a**, Fibre intake on the plant-based diet rose from a median baseline value of 9.3 ± 2.1 to 25.6 ± 1.1 g per 1,000 kcal ($P = 0.007$; two-sided Wilcoxon signed-rank test), but was negligible on the animal-based diet ($P = 0.005$). **b**, Daily fat intake doubled on the animal-based diet from a baseline of $32.5 \pm 2.2\%$ to $69.5 \pm 0.4\%$ kcal ($P = 0.005$), but dropped on the plant-based diet to $22.1 \pm 1.7\%$ kcal ($P = 0.02$). **c**, Protein intake rose on the animal-based diet to $30.1 \pm 0.5\%$ kcal from a baseline level of $16.2 \pm 1.3\%$ kcal ($P = 0.005$), and decreased on the plant-based diet to $10.0 \pm 0.3\%$ kcal ($P = 0.005$). **d**, Within-sample species diversity (α diversity, Shannon diversity index), did not significantly change during either diet. **e**, The similarity of each individual's gut microbiota to their baseline communities (β diversity, Jensen–Shannon distance) decreased on the animal-based diet (dates with $q < 0.05$ identified with asterisks; Bonferroni-corrected, two-sided Mann–Whitney U test). Community differences were apparent 1 day after a tracing dye showed the animal-based diet reached the gut (blue arrows depict appearance of food dyes added to first and last diet day meals; Extended Data Fig. 3a).

Discussion). We also observed a significant positive correlation between subjects' fibre intake over the past year and baseline gut *Prevotella* levels (Extended Data Fig. 4 and Supplementary Table 10).

To identify functional traits linking clusters that thrived on the animal-based diet, we selected the most abundant taxon in the three most-enriched clusters (*Bilophila wadsworthia*, cluster 28; *Alistipes putredinis*, cluster 26; and a *Bacteroides* sp., cluster 29), and performed a literature search for their lifestyle traits. That search quickly yielded a

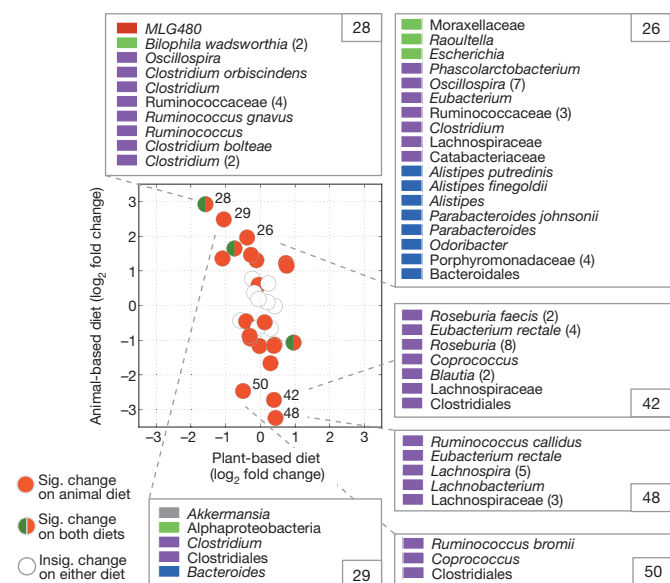


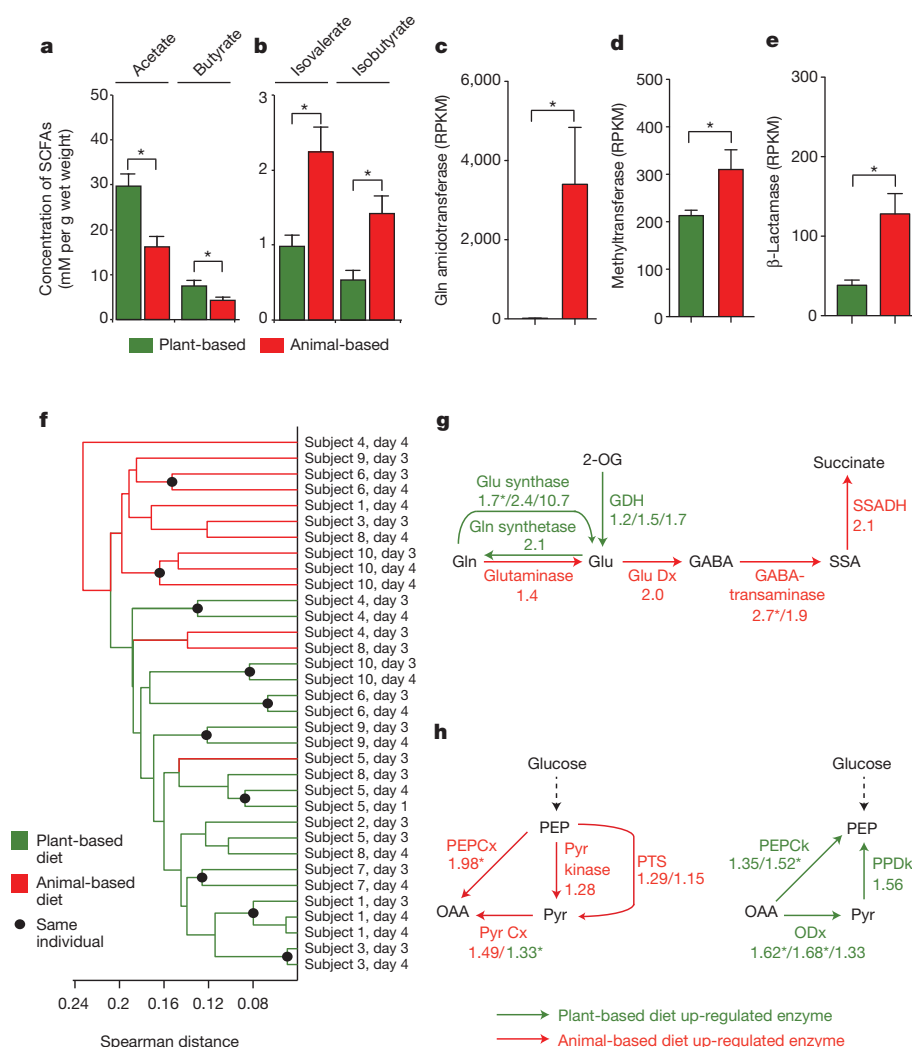
Figure 2 | Bacterial cluster responses to diet arms. Cluster $\log_2$ fold changes on each diet arm were computed relative to baseline samples across all subjects and are drawn as circles. Clusters with significant (Sig.) fold changes on the animal-based diet are coloured in red, and clusters with significant fold changes on both the plant- and animal-based diets are coloured in both red and green. Uncoloured clusters exhibited no significant (Insig.) fold change on either the animal- or plant-based diet ($q < 0.05$, two-sided Wilcoxon signed-rank test). Bacterial membership in the clusters with the three largest positive and negative fold changes on the animal-based diet are also displayed and coloured by phylum: Firmicutes (purple), Bacteroidetes (blue), Proteobacteria (green), Tenericutes (red) and Verrucomicrobia (grey). Multiple operational taxonomic units (OTUs) with the same name are counted in parentheses.

common theme of bile resistance for these taxa, which is consistent with observations that high fat intake causes more bile acids to be secreted¹⁴.

Analysis of faecal short-chain fatty acids (SCFAs) and bacterial clusters suggests that macronutrient shifts on both diets also altered microbial metabolic activity. Relative to the plant-based diet and baseline samples, the animal-based diet resulted in significantly lower levels of the products of carbohydrate fermentation and a higher concentration of the products of amino acid fermentation (Fig. 3a, b and Supplementary Table 11). When we correlated subjects' SCFA concentrations with the same-day abundance of bacterial clusters from Fig. 2, we found significant positive relationships between clusters composed of putrefactive microbes^{15,16} (that is, *Alistipes putredinis* and *Bacteroides* spp.) and SCFAs that are the end products of amino acid fermentation (Extended Data Fig. 5). We also observed significant positive correlations between clusters comprised of saccharolytic microbes³ (for example, *Roseburia*, *E. rectale* and *F. prausnitzii*) and the products of carbohydrate fermentation.

To test whether the observed changes in microbial community structure and metabolic end products were accompanied by more widespread shifts in the gut microbiome, we measured microbial gene expression using RNA sequencing (RNA-seq). A subset of samples was analysed, targeting the baseline periods and the final 2 days of each diet (Extended Data Fig. 1 and Supplementary Table 12). We identified several differentially expressed metabolic modules and pathways during the plant- and animal-based diets (Supplementary Tables 13, 14). The animal-based diet was associated with increased expression of genes for vitamin biosynthesis (Fig. 3c); the degradation of polycyclic aromatic hydrocarbons (Fig. 3d), which are carcinogenic compounds produced during the charring of meat¹⁷; and the increased expression of β -lactamase genes (Fig. 3e). Metagenomic models constructed from our 16S rRNA data¹⁸ suggest that the observed expression differences are due to a combination of regulatory and taxonomic shifts within the microbiome (Supplementary Tables 15, 16).

Figure 3 | Diet alters microbial activity and gene expression. **a, b**, Faecal concentrations of SCFAs from carbohydrate (**a**) and amino acid (**b**) fermentation (* $P < 0.05$, two-sided Mann–Whitney U test; $n = 9$ –11 faecal samples per diet arm; Supplementary Table 11). **c–e**, The animal-based diet was associated with significant increases in gene expression (normalized to reads per kilobase per million mapped (RPKM)) among glutamine amidotransferases (KEGG orthologous group K08681, vitamin B₆ metabolism) (**c**), methyltransferases (K00599, polycyclic aromatic hydrocarbon degradation) (**d**) and β -lactamases (K01467) (**e**). **f**, Hierarchical clustering of gut microbial gene expression profiles collected on the animal-based (red) and plant-based (green) diets. Expression profile similarity was significantly associated with diet ($P < 0.003$; two-sided Fisher's exact test excluding replicate samples), despite inter-individual variation that preceded the diet (Extended Data Fig. 6a, b). **g, h**, Enrichment on animal-based diet (red) and plant-based diet (green) for expression of genes involved in amino acid metabolism (**g**) and central metabolism (**h**). Numbers indicate the mean fold change between the two diets for each KEGG orthologous group assigned to a given enzymatic reaction (Supplementary Table 17). Enrichment patterns on the animal- and plant-based diets agree perfectly with patterns observed in carnivorous and herbivorous mammals, respectively² ($P < 0.001$, Binomial test). GDH, glutamate dehydrogenase; Glu Dc, glutamate decarboxylase; ODC, oxaloacetate decarboxylase; PEPCK, phosphoenolpyruvate carboxylase; PEPCK, PEP carboxykinase; PPDk, pyruvate, orthophosphate dikinase; PTS, phosphotransferase system; Pyr Cx, pyruvate carboxylase; SSADH, succinate-semialdehyde dehydrogenase. Note that Pyr Cx is represented by two groups, which showed divergent fold changes. **c–h**, * $P < 0.05$, Student's t -test. Values in panels **a–e** are mean $\pm$ standard error of the mean (s.e.m.).



Next, we hierarchically clustered microbiome samples based on the transcription of Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologous groups¹⁹, which suggested that overall microbial gene expression was strongly linked to host diet. Nearly all of the diet samples could be clustered by diet arm ($P < 0.003$, Fisher's exact test; Fig. 3f), despite the pre-existing inter-individual variation we observed during the baseline diets (Extended Data Fig. 6a, b). Still, subjects maintained their inter-individual differences on a taxonomic level on the diet arms (Extended Data Fig. 6c). Of the three RNA-seq samples on the animal-based diet that clustered with samples from the plant-based diet, all were taken on day 3 of the diet arm. In contrast, all RNA-seq samples from the final day of the diet arms (day 4) clustered by diet (Fig. 3f).

Remarkably, the plant- and animal-based diets also elicited transcriptional responses that were consistent with known differences in gene abundance between the gut microbiomes of herbivorous and carnivorous mammals, such as the trade-offs between amino acid catabolism versus biosynthesis, and in the interconversions of phosphoenolpyruvate (PEP) and oxaloacetate² (Fig. 3g, h). The former pathway favours amino acid catabolism when protein is abundant², and we speculate that the latter pathway produces PEP for aromatic amino acid synthesis when protein is scarce²⁰. In all 14 steps of these pathways, we observed fold changes in gene expression on the plant- and animal-based diets, the directions of which agreed with the previously reported differences between herbivores and carnivores ($P < 0.001$, Binomial test). Notably, this perfect agreement is not observed when the plant- and animal-based diets are only compared with their respective baseline periods, indicating that the expression patterns in Fig. 3g, h reflect functional changes from both diet arms (Supplementary Table 17).

Our findings that the human gut microbiome can rapidly switch between herbivorous and carnivorous functional profiles may reflect past selective pressures during human evolution. Consumption of animal foods by our ancestors was probably volatile, depending on season and stochastic foraging success, with readily available plant foods offering a fall-back source of calories and nutrients²¹. Microbial communities that could quickly, and appropriately, shift their functional repertoire in response to diet change would have subsequently enhanced human dietary flexibility. Examples of this flexibility may persist today in the form of the wide diversity of modern human diets¹¹.

We next examined whether, in addition to affecting the resident gut microbiota, either diet arm introduced foreign microorganisms into the distal gut. We identified foodborne bacteria on both diets using 16S rRNA gene sequencing. The cheese and cured meats included in the animal-based diet were dominated by lactic acid bacteria commonly used as starter cultures for fermented foods^{22,23}: *Lactococcus lactis*, *Pediococcus acidilactici* and *Streptococcus thermophilus* (Fig. 4a). Common non-lactic-acid bacteria included several *Staphylococcus* taxa; strains from this genus are often used when making fermented sausages²³. During the animal-based diet, three of the bacteria associated with cheese and cured meats (*L. lactis*, *P. acidilactici* and *Staphylococcus*) became significantly more prevalent in faecal samples ($P < 0.05$, Wilcoxon signed-rank test; Extended Data Fig. 7c), indicating that bacteria found in common fermented foods can reach the gut at abundances above the detection limit of our sequencing experiments (on average 1 in 4×10^4 gut bacteria; Supplementary Table 6).

We also sequenced the internal transcribed spacer (ITS) region of the rRNA operon from community DNA extracted from food and

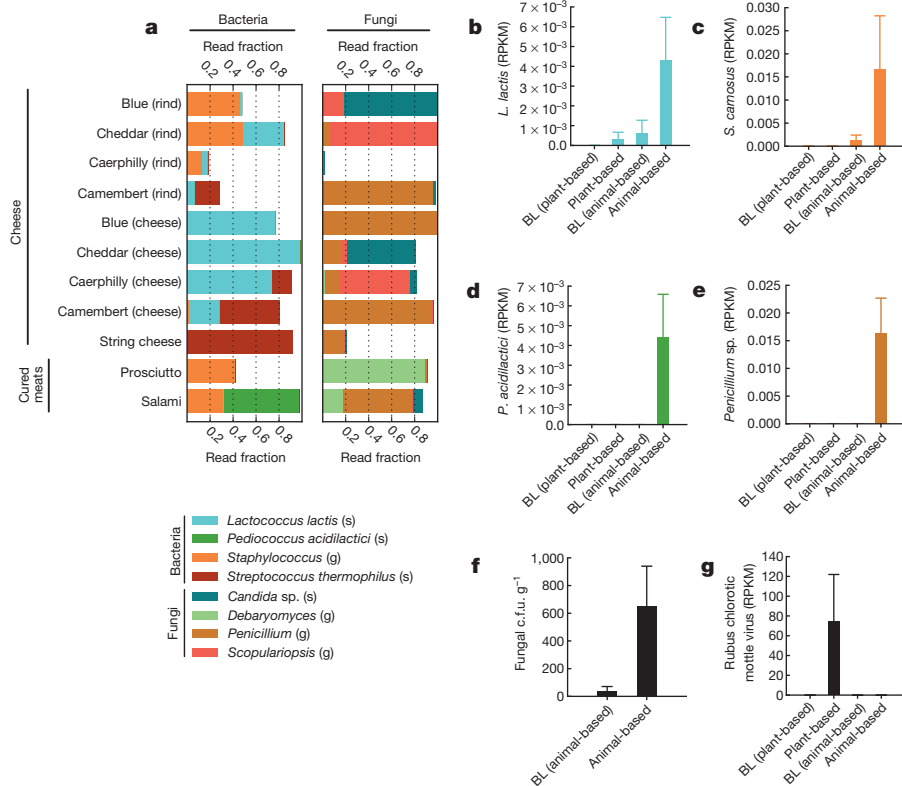


Figure 4 | Foodborne microbes are detectable in the distal gut. **a**, Common bacteria and fungi associated with the animal-based diet menu items, as measured by 16S rRNA and ITS gene sequencing, respectively. Taxa are identified on the genus (g) and species (s) level. A full list of foodborne fungi and bacteria on the animal-based diet can be found in Supplementary Table 21. Foods on the plant-based diet were dominated by matches to the Streptophyta, which derive from chloroplasts within plant matter (Extended Data Fig. 7a). **b–e**, Faecal RNA transcripts were significantly enriched ($q < 0.1$, Kruskal–Wallis test; $n = 6–10$ samples per diet arm) for several food-associated microbes on the animal-based diet relative to baseline (BL) periods, including *Lactococcus lactis* (**b**), *Staphylococcus carnosus* (**c**), *Pediococcus acidilactici* (**d**) and a *Penicillium* sp. (**e**). A complete table of taxa with significant expression differences can be found in Supplementary Table 22. **f**, Fungal concentrations in faeces before and 1–2 days after the animal-based diet were also measured using culture media selective for fungal growth (plate count agar with milk, salt and chloramphenicol). Post-diet faecal samples exhibit significantly higher fungal concentrations than baseline samples ($P < 0.02$; two-sided Mann–Whitney U test; $n = 7–10$ samples per diet arm). c.f.u., colony-forming units. **g**, *Rubus chlorotic mottle virus* transcripts increase on the plant-based diet ($q < 0.1$, Kruskal–Wallis test; $n = 6–10$ samples per diet arm). **b–g**, Bar charts all display mean $\pm$ s.e.m.

faecal samples to study the relationship between diet and enteric fungi, which so far remains poorly characterized (Supplementary Table 18). Menu items on both diets were colonized by the genera *Candida*, *Debaryomyces*, *Penicillium* and *Scopulariopsis* (Fig. 4a and Extended Data Fig. 7a), which are often found in fermented foods²². A *Penicillium* sp. and *Candida* sp. were consumed in sufficient quantities on the animal- and plant-based diets to show significant ITS sequence increases on those respective diet arms (Extended Data Fig. 7b, c).

Microbial culturing and re-analysis of our RNA-seq data suggested that foodborne microbes survived transit through the digestive system and may have been metabolically active in the gut. Mapping RNA-seq reads to an expanded reference set of 4,688 genomes (see Methods) revealed a significant increase on the animal-based diet of transcripts expressed by food-associated bacteria (Fig. 4b–d) and fungi (Fig. 4e; $q < 0.1$, Kruskal–Wallis test). Many dairy-associated microbes remained viable after passing through the digestive tract, as we isolated 19 bacterial and fungal strains with high genetic similarity ($>97\%$ ITS or 16S rRNA) to microbes cultured from cheeses fed to the subjects (Supplementary Table 19). Moreover, *L. lactis* was more abundant in faecal cultures sampled after the animal-based diet, relative to samples from the preceding baseline period ($P < 0.1$; Wilcoxon signed-rank test). We also detected an overall increase in the faecal concentration of viable fungi on the animal-based diet (Fig. 4f; $P < 0.02$; Mann–Whitney U test). Interestingly, we detected RNA transcripts from multiple plant viruses (Extended Data Fig. 8). One plant pathogen, *Rubus chlorotic*

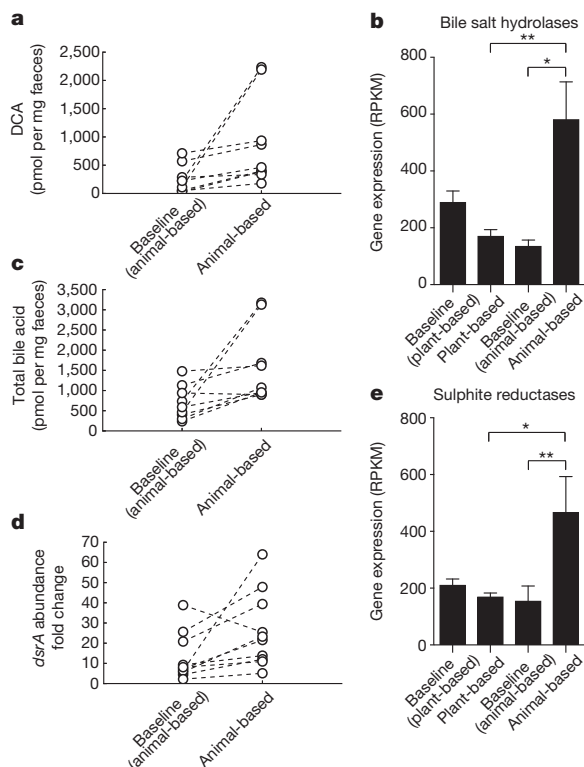


Figure 5 | Changes in the faecal concentration of bile acids and biomarkers for *Bilophila* on the animal-based diet. **a**, DCA, a secondary bile acid known to promote DNA damage and hepatic carcinomas²⁶, accumulates significantly on the animal-based diet ($P < 0.01$, two-sided Wilcoxon signed-rank test; see Supplementary Table 23 for the diet response of other secondary bile acids). **b**, RNA-seq data also supports increased microbial metabolism of bile acids on the animal-based diet, as we observed significantly increased expression of microbial bile salt hydrolases (K01442) during that diet arm ($*q < 0.05$, $**q < 0.01$, Kruskal–Wallis test; normalized to RPKM; $n = 8–21$ samples per diet arm). **c**, Total faecal bile acid concentrations also increase significantly on the animal-based diet, relative to the preceding baseline period ($P < 0.05$, two-sided Wilcoxon signed-rank test), but do not change on the plant-based diet (Extended Data Fig. 9). Bile acids have been shown to cause inflammatory bowel disease in mice by stimulating the growth of the bacterium *Bilophila*⁶, which is known to reduce sulphite to hydrogen sulphide via the sulphite reductase enzyme DsrA (Extended Data Fig. 10). **d**, **e**, Quantitative polymerase chain reaction (PCR) showed a significant increase in microbial DNA coding for *dsrA* on the animal-based diet ($P < 0.05$; two-sided Wilcoxon signed-rank test) (**d**), and RNA-seq identified a significant increase in sulphite reductase expression ($*q < 0.05$, $**q < 0.01$, Kruskal–Wallis test; $n = 8–21$ samples/diet arm) (**e**). **b**, **e**, Bar graphs display mean $\pm$ s.e.m.

mottle virus, was only detectable on the plant-based diet (Fig. 4g). This virus infects spinach²⁴, which was a key ingredient in the prepared meals on the plant-based diet. These data support the hypothesis that plant pathogens can reach the human gut via consumed plant matter²⁵.

Finally, we found that microbiota changes on the animal-based diet could be linked to altered faecal bile acid profiles and the potential for human enteric disease. Recent mouse experiments have shown that high-fat diets lead to increased enteric deoxycholic concentrations (DCA); this secondary bile acid is the product of microbial metabolism and promotes liver cancer²⁶. In our study, the animal-based diet significantly increased the levels of faecal DCA (Fig. 5a). Expression of bacterial genes encoding bile salt hydrolases, which are prerequisites for gut microbial production of DCA²⁷, was also significantly higher on the animal-based diet (Fig. 5b). Elevated DCA levels, in turn, may have contributed to the microbial disturbances on the animal-based diet, as this bile acid can inhibit the growth of members of the Bacteroidetes and Firmicutes phyla²⁸.

Mouse models have also provided evidence that inflammatory bowel disease can be caused by *B. wadsworthia*, a sulphite-reducing bacterium whose production of H₂S is thought to inflame intestinal tissue⁶. Growth of *B. wadsworthia* is stimulated in mice by select bile acids secreted while consuming saturated fats from milk. Our study provides several lines of evidence confirming that *B. wadsworthia* growth in humans can also be promoted by a high-fat diet. First, we observed *B. wadsworthia* to be a major component of the bacterial cluster that increased most while on the animal-based diet (cluster 28; Fig. 2 and Supplementary Table 8). This *Bilophila*-containing cluster also showed significant positive correlations with both long-term dairy ($P < 0.05$; Spearman correlation) and baseline saturated fat intake (Supplementary Table 20), supporting the proposed link to milk-associated saturated fats⁶. Second, the animal-based diet led to significantly increased faecal bile acid concentrations (Fig. 5c and Extended Data Fig. 9). Third, we observed significant increases in the abundance of microbial DNA and RNA encoding sulphite reductases on the animal-based diet (Fig. 5d, e). Together, these findings are consistent with the hypothesis that diet-induced changes to the gut microbiota may contribute to the development of inflammatory bowel disease. More broadly, our results emphasize that a more comprehensive understanding of diet-related diseases will benefit from elucidating links between nutritional, biliary and microbial dynamics.

METHODS SUMMARY

All experiments were performed under the guidance of the Harvard Committee on the Use of Human Subjects in Research; informed consent was obtained from all 11 subjects. Nine of those subjects participated in both diet arms, which were separated by 1 month. Each day, subjects logged their food intake and non-invasively sampled their gut microbiota. Each sample was either frozen immediately at -80°C or briefly stored in personal -20°C freezers before transport to the laboratory. Menu items for the animal- and plant-based diets were purchased at grocery stores and a restaurant, prepared by the experimenters, and distributed to the subjects daily. DNA was extracted from all faecal samples as previously described²⁹, sequenced using 16S rRNA- and ITS-specific primers, and analysed with the Quantitative Insights Into Microbial Ecology (QIIME) software package³⁰ and custom Python scripts. Three baseline days and 2 days on each experimental diet were selected for RNA-seq analysis²⁹. Each RNA-seq data set was mapped to a functional database comprising 539 human-associated microbial genomes and a taxonomic identification database comprising 4,688 eukaryotic, prokaryotic and viral genomes. SCFA analysis was performed by gas chromatography, and bile-acid analysis used enzymatic assays and mass spectrometry.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 April; accepted 29 October 2013.

Published online 11 December 2013.

- Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
- Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).

- Duncan, S. H. *et al.* Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl. Environ. Microbiol.* **73**, 1073–1078 (2007).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5**, 220–230 (2011).
- Devkota, S. *et al.* Dietary-fat-induced tauracholic acid promotes pathobiont expansion and colitis in *IL10*^{−/−} mice. *Nature* **487**, 104–108 (2012).
- Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333**, 101–104 (2011).
- Russell, W. R. *et al.* High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am. J. Clin. Nutr.* **93**, 1062–1072 (2011).
- Cordain, L. *et al.* Plant-animal subsistence ratios and macronutrient energy estimations in worldwide hunter-gatherer diets. *Am. J. Clin. Nutr.* **71**, 682–692 (2000).
- Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107**, 14691–14696 (2010).
- Reddy, B. S. Diet and excretion of bile acids. *Cancer Res.* **41**, 3766–3768 (1981).
- Smith, E. A. & Macfarlane, G. T. Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol. Ecol.* **25**, 355–368 (1998).
- Smith, E. A. & Macfarlane, G. T. Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. *J. Appl. Bacteriol.* **81**, 288–302 (1996).
- Sinha, R. *et al.* High concentrations of the carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) occur in chicken but are dependent on the cooking method. *Cancer Res.* **55**, 4516–4519 (1995).
- Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.* **31**, 814–821 (2013).
- Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Pittard, J. & Wallace, B. J. Distribution and function of genes concerned with aromatic biosynthesis in *Escherichia coli*. *J. Bacteriol.* **91**, 1494–1508 (1966).
- Hawkes, K., O'Connell, J. F. & Jones, N. G. Hunting income patterns among the Hadza: big game, common goods, foraging goals and the evolution of the human diet. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **334**, 243–250 (1991).
- Bourdichon, F., Berger, B. & Casaregola, S. Safety demonstration of microbial food cultures (MFC) in fermented food products. *Bull. Int. Dairy Fed.* **455**, 1–66 (2012).
- Nychas, G. J. & Arkoudelos, J. S. Staphylococci: their role in fermented sausages. *Soc. Appl. Bacteriol. Symp. Ser.* **19**, 167S–188S (1990).
- McGavin, W. J. & Macfarlane, S. A. Rubus chlorotic mottle virus, a new sobemovirus infecting raspberry and bramble. *Virus Res.* **139**, 10–13 (2009).
- Zhang, T. *et al.* RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* **4**, e3 (2006).
- Yoshimoto, S. *et al.* Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* **499**, 97–101 (2013).
- Ridlon, J. M., Kang, D. J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**, 241–259 (2006).
- Islam, K. B. *et al.* Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* **141**, 1773–1781 (2011).
- Maurice, C. F., Haider, H. J. & Turnbaugh, P. J. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152**, 39–50 (2013).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank A. Murray, G. Guidotti, E. O'Shea, J. Moffitt and B. Stern for insightful comments; M. Delaney for biochemical analyses; C. Daly, M. Clamp and C. Reardon for sequencing support; N. Fierer for providing ITS primers; A. Luong and K. Bauer for technical assistance; J. Brulic and R. Menon for nutritional guidelines; A. Rahman for menu suggestions; A. Must and J. Queenan for nutritional analysis; and our diet study volunteers for their participation. This work was supported by the National Institutes of Health (P50 GM068763), the Boston Nutrition Obesity Research Center (DK0046200), and the General Mills Bell Institute of Health and Nutrition.

Author Contributions L.A.D., R.J.D. and P.J.T. designed the study, and developed and prepared the diets. L.A.D., C.F.M., R.N.C., D.B.G., J.E.B., B.E.W. and P.J.T. performed the experimental work. A.V.L., A.S.D., Y.V., M.A.F. and S.B.B. conducted bile acid analyses. L.A.D. and P.J.T. performed computational analyses. L.A.D. and P.J.T. prepared the manuscript.

Author Information RNA-seq data have been deposited in the Gene Expression Omnibus under accession GSE46761; 16S and ITS rRNA gene sequencing reads have been deposited in MG-RAST under accession 6248. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.J.T. (pturnbaugh@fas.harvard.edu).

METHODS

Sample collection. All experiments were performed under the guidance of the Harvard Committee on the Use of Human Subjects in Research. We recruited 11 unrelated subjects ($n = 10$ per diet; 9 individuals completed both arms of the study). One participant suffered from a chronic gastrointestinal disease, but all other volunteers were otherwise healthy. The volunteers' normal bowel frequencies ranged from three times a day to once every other day. Three participants had taken antibiotics in the past year. Additional subject information is provided in Supplementary Table 2. Gut microbial communities were sampled and analysed from faeces^{29,30}. Subjects were instructed to collect no more than one sample per day, but to log all bowel movements. No microbiota patterns were observed as a function of sampling time of day (data not shown). Subjects collected samples by placing disposable commode specimen containers (Claflin Medical Equipment) under their toilet seats before bowel movements. CultureSwabs (BD) were then used to collect faecal specimens for sequencing analyses, and larger collection tubes were provided for harvesting larger, intact stool samples (~10 g) for metabolic analyses. Each sample was either frozen immediately at -80°C or briefly stored in personal -20°C freezers before transport to the laboratory.

Diet design. We constructed two diet arms, each of which consisted mostly of plant- or animal-based foods (Extended Data Fig. 1). Subjects on the plant-based diet ate cereal for breakfast and precooked meals made of vegetables, rice and lentils for lunch and dinner (see Supplementary Table 1 for a full list of diet ingredients). Fresh and dried fruits were provided as snacks on this diet. Subjects on the animal-based diet ate eggs and bacon for breakfast, and cooked pork and beef for lunch. Dinner consisted of cured meats and a selection of four cheeses. Snacks on this diet included pork rinds, cheese and salami. Ingredients for the plant-based diet, dinner meats and cheeses for the animal-based diet, and snacks for both diets were purchased from grocery stores. Lunchmeats for the animal-based diet were prepared by a restaurant that was instructed to not add sauce to the food. On each diet arm, subjects were instructed to eat only provided foods or allowable beverages (water or unsweetened tea for both diets; coffee was allowed on the animal-based diet). They were also allowed to add one salt packet per meal, if desired for taste. Subjects could eat unlimited amounts of the provided foods. Outside of the 5-day diet arms, subjects were instructed to eat normally.

Food logs, subject metadata and dietary questionnaires. Subjects were given notepads to log their diet, health and bowel movements during the study. Subjects transcribed their notepads into digital spreadsheets when the study ended. Each ingested food (including foods on the diet arm) was recorded, as well as data on time, location, portion size, and food brand. Subjects were provided with pocket digital scales (American Weigh) and a visual serving size guide to aid with quantifying the amount of food consumed. Each day, subjects tracked their weight using either a scale provided in the lab, or their own personal scales at home. While on the animal-based diet, subjects were requested to measure their urinary ketone levels using provided Ketostix strips (Bayer; Extended Data Fig. 1). If subjects recorded a range of ketone levels (the Ketostix colour key uses a range-based reporting system) the middle value of that range was used for further analysis. Subjects were encouraged to record any discomfort they experienced while on either diet (for example, bloating, constipation). Subjects tracked all bowel movements, regardless of whether or not they collected samples, recording movement time, date and location, and qualitatively documented stool colour, odour and type³¹. Subjects were also asked to report when they observed stool staining from food dyes consumed at the beginning and end of each diet arm (Extended Data Fig. 3a).

Diet quantification. We quantified subjects' daily nutritional intake during the study using CalorieKing and Nutrition Data System for Research (NDSR). The CalorieKing food database was accessed via the CalorieKing Nutrition and Exercise Manager software (version 4.1.0). Subjects' food items were manually transferred from digital spreadsheets into the CalorieKing software, which then tabulated each food's nutritional content. Macronutrient content per serving was calculated for each of the prepared meals on the animal- and plant-based diet using lists of those meals' ingredients. Nutritional data was outputted from CalorieKing in CSV format and parsed for further analysis using a custom Python script. NDSR intake data were collected and analysed using Nutrition Data System for Research software version 2012, developed by the Nutrition Coordinating Center (NCC), University of Minnesota. We estimated subjects' long-term diet using the National Cancer Institute's Diet History Questionnaire II (DHQ)³². We used the DHQ to quantify subjects' annual diet intake, decomposed into 176 nutritional categories. Subjects completed the yearly, serving-size-included version of the DHQ online using their personal computers. We parsed the survey's results using Diet*Calc software (version 1.5; Risk Factor Monitoring and Methods Branch, National Cancer Institute) and its supplied 'Food and Nutrient Database' and 'dhqweb.yearly.withserv.2010.qdd' QDD file. There was good agreement between subjects' diets as measured by CalorieKing, the NDSR, and the DHQ: 18 of 20 nutritional comparisons between pairs of databases showed significant correlations (Supplementary

Table 3). Unless specified, nutritional data presented in this manuscript reflect CalorieKing measurements.

16S rRNA gene sequencing and processing. Temporal patterns of microbial community structure were analysed from daily faecal samples collected across each diet (Extended Data Fig. 1). Samples were kept at -80°C until DNA extraction with the PowerSoil bacterial DNA extraction kit (MoBio). The V4 region of the 16S rRNA gene was PCR amplified in triplicate, and the resulting amplicons were cleaned, quantified and sequenced on the Illumina HiSeq platform according to published protocols^{33,34} and using custom barcoded primers (Supplementary Table 6). Raw sequences were processed using the QIIME software package³⁰. Only full-length, high-quality reads ($-r = 0$) were used for analysis. OTUs were picked at 97% similarity against the Greengenes database³⁵ (constructed by the nested_gg_workflow.py QiimeUtils script on 4 February, 2011), which we trimmed to span only the 16S rRNA region flanked by our sequencing primers (positions 521–773). In total, we characterized an average of $43,589 \pm 1,826$ 16S rRNA sequences for 235 samples (an average of 0.78 samples per person per study day; Supplementary Table 6). Most of the subsequent analysis of 16S rRNA data, including calculations of α and β diversity, were performed using custom Python scripts, the SciPy Python library³⁶, and the Pandas Data Analysis Library³⁷. Correction for multiple hypothesis testing used the fdrtool³⁸ R library, except in the case of small test numbers, in which case the Bonferroni correction was used.

OTU clustering. We used clustering to simplify the dynamics of thousands of OTUs into a limited number of variables that could be more easily visualized and manually inspected. Clustering was performed on normalized OTU abundances. Such abundances are traditionally computed by scaling each sample's reads to sum to a fixed value (for example, unity); this technique is intended to account for varying sequencing depth between samples. However, this standard technique may cause false relationships to be inferred between microbial taxa, as increases in the abundance of one microbial group will cause decreases in the fractional abundance of other microbes (this artefact is known as a 'compositional' effect³⁹). To avoid compositional biases, we used an alternative normalization approach, which instead assumes that no more than half of the OTUs held in common between two microbial samples change in abundance. This method uses a robust (outlier-resistant) regression to estimate the median OTU fold change between communities, by which it subsequently rescales all OTUs. To simplify community dynamics further, we only included in our clustering model OTUs that comprised 95% of total reads (after ranking by normalized abundance).

Abundances for each included OTU were then converted to log space and median centred. We computed OTU pairwise distances using the Pearson correlation (OTU abundances across all subjects and time points were used). The resulting distance matrix was subsequently inputted into Scipy's hierarchical clustering function ('fcluster').

Default parameters were used for fcluster, with the exception of the clustering criterion, which was set to 'distance', and the clustering threshold, which was set to '0.7'. These parameters were selected manually so that cluster boundaries visually agreed with the correlation patterns plotted in a matrix of pairwise OTU distances. Statistics on cluster abundance during baseline and diet periods were computed by taking median values across date ranges. Baseline date ranges were the 4 days preceding each diet arm (that is, days -4 to -1). Date ranges for the diet arms were chosen so as to capture the full effects of each diet. These ranges were not expected to perfectly overlap with the diet arms themselves, due to the effects of diet transit time. We therefore chose diet arm date ranges that accounted for transit time (as measured by food dye; Extended Data Fig. 3a), picking ranges that began 1 day after foods reached the gut, and ended 1 day before the last diet arm meal reached the gut. These criteria led microbial abundance measurements on the plant-based diet to span days 2–4 of that study arm, and animal-based diet measurements to span days 2–5 of that diet arm.

RNA-seq sample preparation and sequencing. We measured community-wide gene expression using meta-transcriptomics^{7,29,40,41} (Supplementary Table 12). Samples were selected on the basis of our prior 16S rRNA gene-sequencing-based analysis, representing 3 baseline days and 2 time points on each diet ($n = 5$ –10 samples per time point; Extended Data Fig. 1). Microbial cells were lysed by a bead beater (BioSpec Products), total RNA was extracted with phenol:chloroform:isoamyl alcohol (pH 4.5, 125:24:1, Ambion 9720) and purified using Ambion MEGAClear columns (Life Technologies), and rRNA was depleted via Ambion MICROBExpress subtractive hybridization (Life Technologies) and custom depletion oligonucleotides. The presence of genomic DNA contamination was assessed by PCR with universal 16S rRNA gene primers. cDNA was synthesized using SuperScript II and random hexamers ordered from Invitrogen (Life Technologies), followed by second-strand synthesis with RNaseH and *E. coli* DNA polymerase (New England Biolabs). Samples were prepared for sequencing with an Illumina HiSeq instrument after enzymatic fragmentation (NEBE6040L/M0348S). Libraries were quantified by quantitative reverse transcriptase PCR (qRT-PCR) according to the Illumina

protocol. qRT-PCR assays were run using ABsolute™ QPCR SYBR Green ROX Mix (Thermo Scientific) on a Mx3000P QPCR System instrument (Stratagene). The size distribution of each library was quantified on an Agilent HS-DNA chip. Libraries were sequenced using the Illumina HiSeq platform.

Functional analysis of RNA-seq data. We used a custom reference database of bacterial genomes to perform functional analysis of the RNA-seq data²⁹. This reference included 538 draft and finished bacterial genomes obtained from human-associated microbial isolates⁴², and the *Escherichia coli* DSM22 (ref. 7) reference genome. All predicted proteins from the reference genome database were annotated with KEGG¹⁹ orthologous groups (KOs) using the KEGG database (version 52; BLASTX e value $< 10^{-5}$, bit score > 50 , and $> 50\%$ identity). For query genes with multiple matches, the annotated reference gene with the lowest e value was used. When multiple annotated genes with an identical e value were encountered after a BLAST query, we included all KOs assigned to those genes. Genes from the database with significant homology (BLASTN e value $< 10^{-20}$) to non-coding transcripts from the 539 microbial genomes were excluded from subsequent analysis. High-quality reads (see Supplementary Table 12 for sequencing statistics) were mapped using SSAHA2 (ref. 43) to our reference bacterial database and the Illumina adaptor sequences (SSAHA2 parameters: '-best 1 -score 20 -solexa'). The number of transcripts assigned to each gene was then tallied and normalized to RPKM. To account for genes that were not detected owing to limited sequencing depth, a pseudocount of 0.01 was added to all samples. Samples were clustered in Matlab (version 7.10.0) using a Spearman distance matrix (commands: `pdist`, `linkage`, and `dendrogram`). Genes were grouped by taxa, genomes and KOs by calculating the cumulative RPKM for each sample. HUMAnN⁴⁴ was used for metabolic reconstruction from metagenomic data followed by LefSe⁴⁵ analysis to identify significant biomarkers. A modified version of the 'SConstruct' file was used to input KO counts into the HUMAnN pipeline for each RNA-seq data set. We then ran LefSe on the resulting KEGG module abundance file using the '-o 1000000' flag.

Taxonomic analysis of RNA-seq data. We used Bowtie 2 read alignment program⁴⁶ and the Integrated Microbial Genomes (IMG; version 3.5) database⁴⁷ to map RNA-seq reads to a comprehensive reference survey of prokaryotic, eukaryotic and viral genomes. Our reference survey included all 2,809 viral genomes in IMG (as of version 3.5), a set of 1,813 bacterial and archaeal genomes selected to minimize strain redundancy⁴⁸, and 66 genomes spanning the Eukarya except for the plants and non-nematode Bilateria. Reads were mapped to reference genomes using Bowtie 2, which was configured to analyse mated paired-end reads, and return fragments with a minimum length of 150 bp and a maximum length of 600 bp. All other parameters were left to their default values. The number of base pairs in the reference genome data set exceeded Bowtie's reference size limit, so we split the reference genomes into four subsets. Each read was mapped to each of these four sub-reference data sets, and the results were merged by picking the highest-scoring match across the sub-references. We settled tied scores by randomly choosing one of the best-scoring matches. To measure more precisely the presence or absence of specific taxa, we filtered out reads that mapped to more than one reference sequence. Raw read counts were computed for each reference genome by counting the number of reads that mapped to coding sequences according to the IMG annotations; these counts were subsequently normalized using RPKM scaling. Our analysis pipeline associated several sequences with marine algae, which are unlikely to colonize the human gut. We also detected a fungal pathogen exclusively in samples from subjects consuming the animal-based diet (*Neosartorya fischeri*); this taxon was suspected of being a misidentified cheese fungus, owing to its relatedness to *Penicillium*. We thus reanalysed protist and *N. fischeri* reads associated with potentially mis-annotated taxa using BLAST searches against the NCBI non-redundant database, and we assigned taxonomy manually based on the most common resulting hits (Extended Data Fig. 8).

qPCR. Community DNA was isolated with the PowerSoil bacterial DNA extraction kit (MoBio). To determine the presence of hydrogen consumers, PCR was performed on faecal DNA using the following primer sets: sulphite reductase⁶ (*dsrA*), forward, 5'-CCAACATGCACGGYTCCA-3', reverse, 5'-CGTCAACTTGAACCTGAACCTGTAGG-3'; and sulphate reduction^{49,50} (*aps* reductase), forward, 5'-TGGCAGATMATGATYMACGG-3', reverse, 5'-GGGCCGTAAACCGTCCTTGAA-3'. qPCR assays were run using ABsolute™ QPCR SYBR Green ROX Mix (Thermo Scientific) on a Mx3000P QPCR System instrument (Stratagene). Fold changes were calculated relative to the 16S rRNA gene using the $2^{-\Delta\Delta C_t}$ method and the same primers used for 16S rRNA gene sequencing.

SCFA measurements. Faecal SCFA content was determined by gas chromatography. Chromatographic analysis was carried out using a Shimadzu GC14-A system with a flame ionization detector (FID) (Shimadzu Corp). Fused silica capillary columns 30 m $\times$ 0.25 mm coated with 0.25 μ m film thickness were used (Nukol for the volatile acids and SPB-1000 for the nonvolatile acids (Supelco Analytical). Nitrogen was used as the carrier gas. The oven temperature was 170 °C and the FID and injection port was set to 225 °C. The injected sample volume was 2 μ l and the

run time for each analysis was 10 min. The chromatograms and data integration was carried out using a Shimadzu C-R5A Chromatopac. A volatile acid mix containing 10 mM of acetic, propionic, isobutyric, butyric, isovaleric, valeric, isocaproic, caproic and heptanoic acids was used (Matreya). A non-volatile acid mix containing 10 mM of pyruvic and lactic and 5 mM of oxalacetic, oxalic, methyl malonic, malonic, fumaric and succinic acids was used (Matreya). A standard stock solution containing 1% 2-methyl pentanoic acid (Sigma-Aldrich) was prepared as an internal standard control for the volatile acid extractions. A standard stock solution containing 50 mM benzoic acid (Sigma-Aldrich) was prepared as an internal standard control for the non-volatile acid extractions.

Samples were kept frozen at -80 °C until analysis. The samples were removed from the freezer and 1,200 μ l of water was added to each thawed sample. The samples were vortexed for 1 min until the material was homogenized. The pH of the suspension was adjusted to 2–3 by adding 50 μ l of 50% sulphuric acid. The acidified samples were kept at room temperature (~21 °C) for 5 min and vortexed briefly every minute. The samples were centrifuged for 10 min at 5,000g. Five-hundred microlitres of the clear supernatant was transferred into two tubes for further processing. For the volatile extraction, 50 μ l of the internal standard (1% 2-methyl pentanoic acid solution) and 500 μ l of ethyl ether anhydrous were added. The tubes were vortexed for 30 s and then centrifuged at 5,000g for 10 min. One microlitre of the upper ether layer was injected into the chromatogram for analysis. For the nonvolatile extraction, 50 μ l of the internal standard (50 mM benzoic acid solution) and 500 μ l of boron trifluoride-methanol solution (Sigma-Aldrich) were added to each tube. These tubes were incubated overnight at room temperature. One millilitre of water and 500 μ l of chloroform were added to each tube. The tubes were vortexed for 30 s and then centrifuged at 5,000g for 10 min. One microlitre of the lower chloroform layer was injected into the chromatogram for analysis. Five-hundred microlitres of each standard mix was used and the extracts prepared as described for the samples. The retention times and peak heights of the acids in the standard mix were used as references for the sample unknowns. These acids were identified by their specific retention times and the concentrations determined and expressed as mM concentrations per gram of sample.

Bulk bile acid quantification. Faecal bile acid concentration was measured as described previously⁵¹. One-hundred milligrams of lyophilized stool was heated to 195 °C in 1 ml of ethylene glycol KOH for 2 h, neutralized with 1 ml of saline and 0.2 ml of concentrated HCl, and extracted into 6 ml of diethyl ether three times. After evaporation of the ether, the sample residues were dissolved in 6 ml of methanol and subjected to enzymatic analysis. Enzymatic reaction mixtures consisted of 66.5 mmol l⁻¹ Tris, 0.33 mmol l⁻¹ EDTA, 0.33 mol l⁻¹ hydrazine hydrate, 0.77 mmol l⁻¹ NAD (N 7004, Sigma-Aldrich), 0.033 U ml⁻¹ α -hydroxysteroid dehydrogenase (Sigma-Aldrich) and either sample or standard (taurocholic acid; Sigma-Aldrich) dissolved in methanol. After 90 min of incubation at 37 °C, absorbance was measured at 340 nm.

Measurement of primary and secondary bile acids. Profiling of faecal primary and secondary bile acids was performed using a modified version of a method described previously⁵². To a suspension of ~100 mg of stool and 0.25 ml of water in a 4 ml Teflon-capped glass vial was added 200 mg of glass beads. The suspension was homogenized by vortexing for 60–90 s. Ethanol (1.8 ml) was added, and the suspension was heated with stirring in a heating block at 80 °C for 1.5 h. The sample was cooled, transferred to a 2 ml Eppendorf tube, and centrifuged at 12,225g for 1–2 min. The supernatant was removed and retained. The pellet was resuspended in 1.8 ml of 80% aqueous ethanol, transferred to the original vial, and heated to 80 °C for 1.5 h. The sample was centrifuged again, and the supernatant was removed and added to the first extraction supernatant. The pellet was resuspended in 1.8 ml of chloroform:methanol (1:1 v/v) and refluxed for 30–60 min. The sample was centrifuged, and the supernatant removed and concentrated to dryness on a rotary evaporator. The ethanolic supernatants were added to the same flask, the pH was adjusted to neutrality by adding aqueous 0.01N HCl, and the combined extracts were evaporated to dryness. The dried extract was resuspended in 1 ml of 0.01N aqueous HCl by sonication for 30 min. A BIO-RAD Poly-Prep chromatography column (0.8 $\times$ 4 cm) was loaded with Lipidex 1000 as a slurry in MeOH, allowed to pack under gravity to a final volume of 1.1 ml, and washed with 10 ml of distilled water. The suspension was filtered through the bed of Lipidex 1000 and the effluent was discarded. The flask was washed with 3 $\times$ 1 ml of 0.01N HCl, the washings were passed through the gel, and the bed was washed with 4 ml of distilled water. Bile acids and sterols were recovered by elution of the Lipidex gel bed with 8 ml of methanol. A BIO-RAD Poly-Prep chromatography column (0.8 $\times$ 4 cm) was loaded with washed SP-Sephadex as a slurry in 72% aqueous MeOH to a final volume of 1.1 ml. The methanolic extract was passed through the SP-Sephadex column, and the column was washed with 4 ml of 72% aqueous methanol. The extract and wash were combined, and the pH was brought to neutral with 0.04N aqueous NaOH. A BIO-RAD Poly-Prep chromatography column (0.8 $\times$ 4 cm) was loaded with Lipidex-DEAP, prepared in the acetate form,

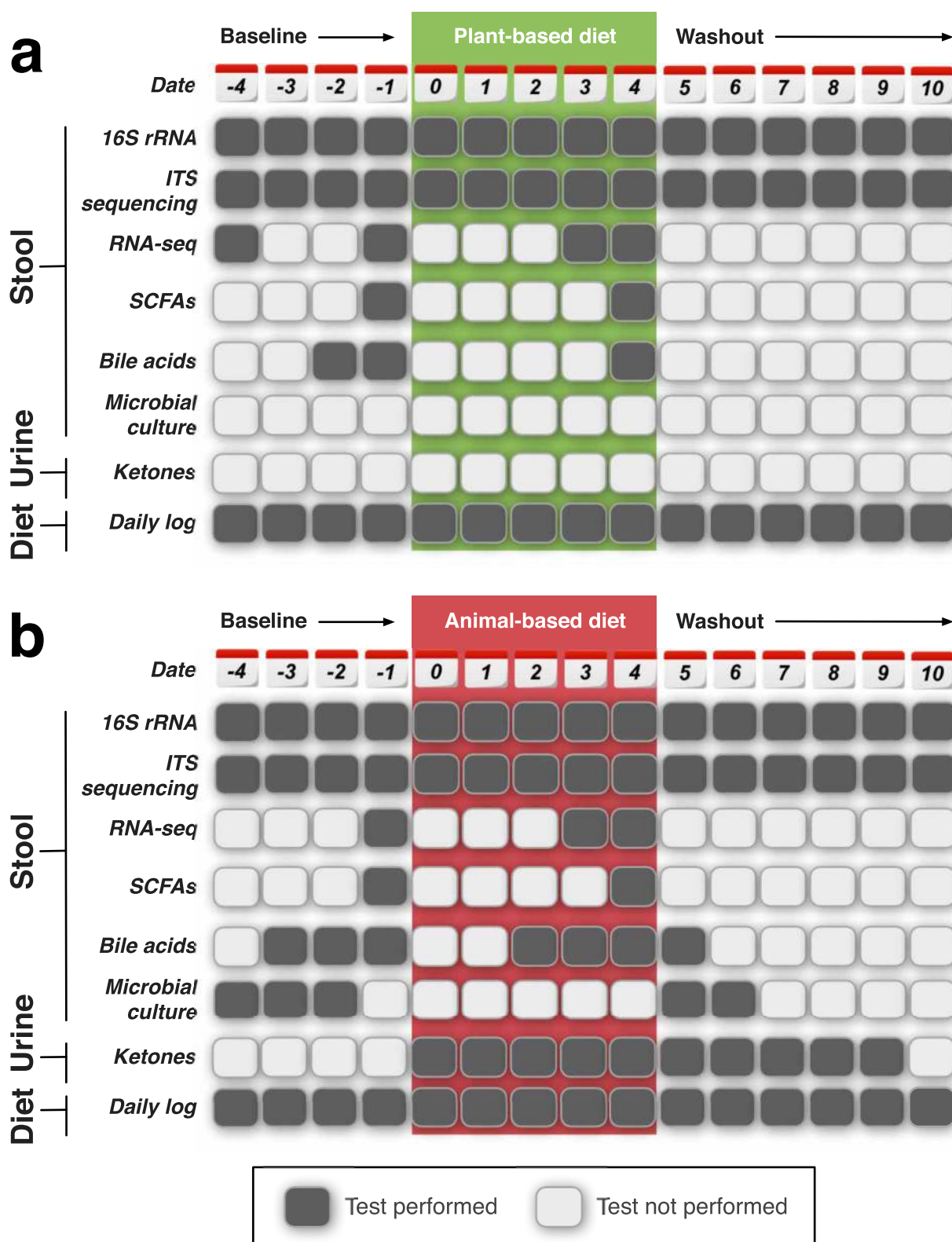
as a slurry in 72% aqueous MeOH to a final volume of 1.1 ml. The combined neutralized effluent was applied to the column, and the solution was eluted using air gas pressure (flow rate $\sim 25 \text{ ml h}^{-1}$). The flask and column were washed with $2 \times 2 \text{ ml}$ of 72% aqueous ethanol, and the sample and washings were combined to give a fraction of neutral compounds including sterols. Unconjugated bile acids were eluted using 4 ml of 0.1 M acetic acid in 72% (v/v) aqueous ethanol that had been adjusted to pH 4.0 by addition of concentrated ammonium hydroxide. The fraction containing bile acids was concentrated to dryness on a rotary evaporator. The bile acids were converted to their corresponding methyl ester derivatives by the addition of 0.6 ml of MeOH followed by 40 μl of a 2.0 M solution of (trimethylsilyl)-diazomethane in diethyl ether. The solution was divided in half, and each half of the sample was concentrated to dryness on a rotary evaporator. The bile acids in the first half of the sample were converted to their corresponding trimethylsilyl ether derivatives by the addition of 35 μl of a 2:1 solution of *N,O*-bis(trimethylsilyl)trifluoroacetamide and chlorotrimethylsilane and analysed by gas chromatography-mass spectrometry (GC-MS). The identities of individual bile acids were determined by comparison of retention time and fragmentation pattern to known standards. Both the ratio of cholest-3-ene to deoxycholic acid in the sample and the amount of internal standard to be added were determined by integrating peak areas. A known amount of the internal standard, 5-cholestane-3-ol (5-coprostanol), was added to the second half of the sample (0.003–0.07 mmol). The bile acids in the second half of the sample were converted to their corresponding trimethylsilyl ether derivatives by the addition of 35 μl of a 2:1 solution of *N,O*-bis(trimethylsilyl)trifluoroacetamide and chlorotrimethylsilane and analysed by GC-MS. Amounts of individual bile acids were determined by dividing integrated bile acid peak area by the internal standard peak area, multiplying by the amount of internal standard added, and then dividing by half of the mass of faecal matter extracted. In the event that the first half of the sample contained cholest-3-ene, the coprostanol peak area in the second half of the sample was corrected by subtracting the area of the cholest-3-ene peak, determined by applying the cholest-3-ene:deoxycholic acid ratio calculated from the first half of the sample.

ITS sequencing. Fungal amplicon libraries were constructed with primers that target the ITS, a region of the nuclear ribosomal RNA cistron shown to promote successful identification across a broad range of fungal taxa⁵³. We selected primers—ITS1f (ref. 54) and ITS2 (ref. 55)—focused on the ITS1 region because it provided the best discrimination between common cheese-associated fungi in preliminary *in silico* tests. Multiplex capability was achieved by adding Golay barcodes to the ITS2 primer. Owing to relatively low concentrations, fungal DNA was amplified in three serial PCR reactions, with the first reaction using 1 μl of the PowerSoil DNA extract, and the subsequent two reactions using 1 μl of the preceding PCR product as the template. In each round of PCR, sample reactions were performed in triplicate and then combined. Barcoded amplicons were cleaned, quantified and pooled to achieve approximately equal amounts of DNA from each sample using methods identical to those used for 16S. We gel purified the pool, targeting amplicons between 150 bp and 500 bp in size, and submitted it for Illumina sequencing. Preliminary taxonomic assignments of ITS reads using the 12_11 UNITE OTUs ITS database (see <http://qiime.org>) resulted in many unassigned reads. To improve the percentage of reads assigned, we created our own custom database of ITS1 sequences. We extracted ITS sequences from GenBank by targeting specific collections of reliable ITS sequences (for example, AFTOL, Fungal Barcoding Consortium) and by searching for sequences of yeasts and filamentous fungi that have been previously isolated from dairy and other food ecosystems. We also retrieved a wider range of fungi for our database by searching GenBank with the query “internal transcribed spacer[All Fields] AND fungi NOT ‘uncultured’”. Sequences that did not contain the full ITS1 were removed. We also included reference OTUs that were identified as widespread cheese fungi in a survey of cheese rinds (B.E.W., J.E.B. and R.J.D., unpublished observations), but were not in public databases.

Microbial culturing. Faecal samples were cultured under conditions permissive for the growth of food-derived microbes. Faecal samples were suspended in a volume of PBS equivalent to ten times their weight. Serial dilutions were prepared and plated on brain heart infusion agar (BD Biosciences) supplemented with 100 $\mu\text{g ml}^{-1}$ cycloheximide, an antifungal agent, and plate count agar with milk and salt (per litre: 5 g tryptone, 2.5 g yeast extract, 1 g dextrose, 1 g whole milk powder, 30 g NaCl, 15 g agar) supplemented with 50 $\mu\text{g ml}^{-1}$ chloramphenicol, an antibacterial agent. Plates were incubated under aerobic conditions at room temperature

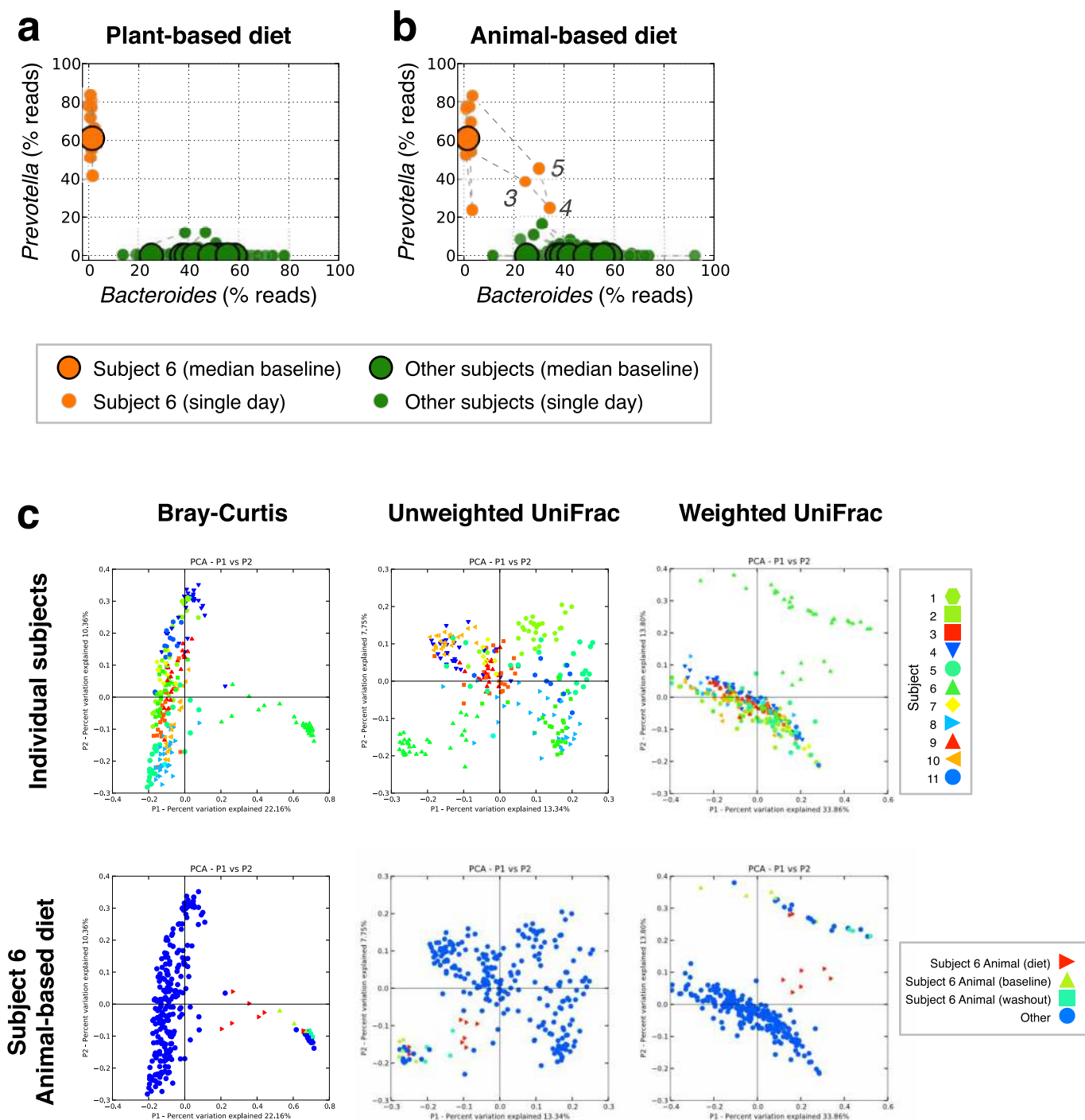
for 7 days. Plates supplemented with chloramphenicol that yielded significant growth of bacteria, as determined by colony morphology, were excluded from further analysis. Plates were examined by eye for bacterial colonies or fungal foci whose morphological characteristics were similar to previously characterized food-derived microbes. Candidate food-derived microbes were isolated and identified by Sanger sequencing of the 16S rRNA gene (for bacteria; primers used were 27f, 5-AGAGTTTGTATCTGGCTCAG and 1492r, 5-GGTTACCTTGTACGACTT) or ITS region (for fungi; primers used were ITS1f, 5-CTTGGTCATTAGAGGAA GTAA and ITS4, 5 TCCTCCGCTTATTGATATGC). After select colonies had been picked for isolation, the surface of each plate was scraped with a razor blade to collect all remaining colonies, and material was suspended in PBS. Dilutions were pooled, and DNA was extracted from the resulting pooled material using a PowerSoil kit (MoBio). The remaining pooled material was stocked in 20% glycerol and stored at -80°C .

31. Lewis, S. J. & Heaton, K. W. Stool form scale as a useful guide to intestinal transit time. *Scand. J. Gastroenterol.* **32**, 920–924 (1997).
32. National Institutes of Health. *Diet History Questionnaire Version 2.0* (National Institutes of Health, Applied Research Program, National Cancer Institute, 2010).
33. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4516–4522 (2011).
34. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
35. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
36. Jones, E. *et al.* SciPy: open source scientific tools for Python (2001).
37. McKinney, W. Data structures for statistical computing in Python. *Proc. 9th Python Sci. Conf.* 51–56 (2010).
38. Strimmer, K. fdrtool: a versatile R package for estimating local and tail areabased false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).
39. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
40. Turnbaugh, P. J. *et al.* Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl Acad. Sci. USA* **107**, 7503–7508 (2010).
41. Rey, F. E. *et al.* Dissecting the *in vivo* metabolic potential of two human gut acetogens. *J. Biol. Chem.* **285**, 22082–22090 (2010).
42. Nelson, K. E. *et al.* A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).
43. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
44. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).
45. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
47. Markowitz, V. M. *et al.* IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
48. Martin, J. *et al.* Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. *PLoS ONE* **7**, e36427 (2012).
49. Deplancke, B. *et al.* Molecular ecological analysis of the succession and diversity of sulfate-reducing bacteria in the mouse gastrointestinal tract. *Appl. Environ. Microbiol.* **66**, 2166–2174 (2000).
50. Stewart, J. A., Chadwick, V. S. & Murray, A. Carriage, quantification, and predominance of methanogens and sulfate-reducing bacteria in faecal samples. *Lett. Appl. Microbiol.* **43**, 58–63 (2006).
51. Porter, J. L. *et al.* Accurate enzymatic measurement of fecal bile acids in patients with malabsorption. *J. Lab. Clin. Med.* **141**, 411–418 (2003).
52. Setchell, K. D., Lawson, A. M., Tanida, N. & Sjovall, J. General methods for the analysis of metabolic profiles of bile acids and related compounds in feces. *J. Lipid Res.* **24**, 1085–1100 (1983).
53. Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl Acad. Sci. USA* **109**, 6241–6246 (2012).
54. Gardes, M. & Bruns, T. D. ITS primers with enhanced specificity for basidiomycetes application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).
55. White, T. J., Bruns, T., Lee, S. & Taylor, J. in *PCR Protocols: A Guide to Methods and Applications* (eds Gelfand, D. H., Innis, M. A., Shinsky, J. J. & White, T. J.) 315–322 (1990).
56. Walker, H. K., Hall, W. D., Hurst, J. W., Comstock, J. P. & Garber, A. J. *Ketoneuria 3rd edn* (Butterworths, 1990).



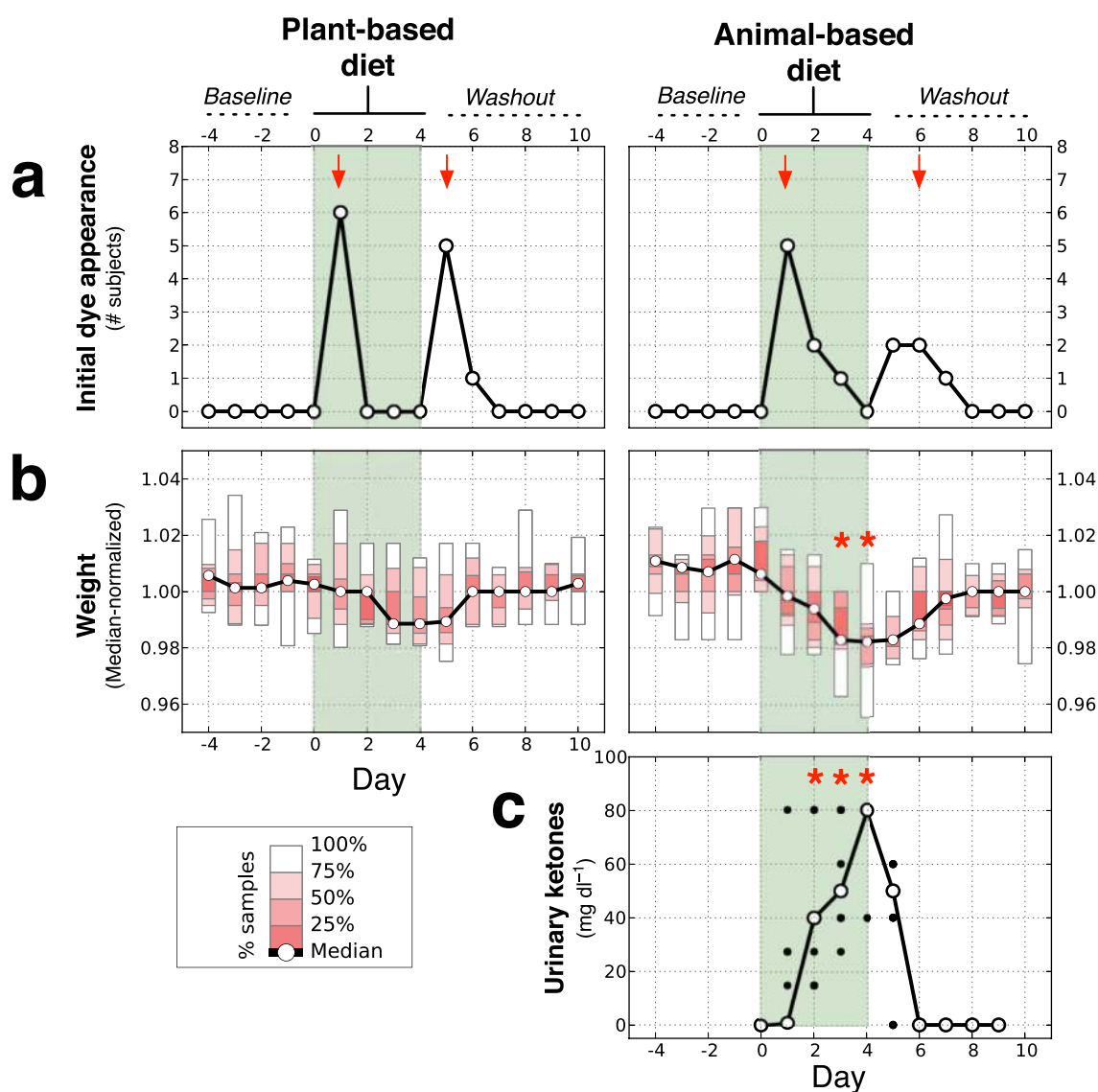
Extended Data Figure 1 | Study design. **a**, **b**, The plant-based (**a**) and animal-based (**b**) diets were fed to subjects for five consecutive days. All dates are defined relative to the start of these diet arms (day 0). Study volunteers were observed for 4 days before each diet (the baseline period, days -4 to -1) and for 6 days after each diet arm (the washout period, days 5 to 10) in order to measure subjects' eating habits and assess their recovery from each diet arm. Subjects were instructed to eat normally during both the baseline and washout periods. Stool samples were collected daily on both diet arms and 16S rRNA and fungal ITS sequencing was performed on all available samples. Subjects

also kept daily diet logs. Several analyses (RNA-seq, SCFAs and bile acids) were performed primarily using only two samples per person per diet (that is, a baseline and diet arm comparison). Comparative sampling did not always occur using exactly the same study days owing to limited sample availability for some subjects. Because we expected the animal-based diet to promote ketogenesis, we only measured urinary ketones on the animal-based diet. To test the hypothesis that microbes from fermented foods on the animal-based diet survived transit through the gastrointestinal tract, we cultured bacteria and fungi before and after the animal-based diet.



Extended Data Figure 2 | A vegetarian's microbiota. a–c, One of the study subjects is a lifelong vegetarian (subject 6). a, Relative abundances of *Prevotella* and *Bacteroides* are shown across the plant-based diet for subject 6 (orange circles), as well as for all other subjects (green circles). Consecutive daily samples from subject 6 are linked by dashed lines. For reference, median baseline abundances are depicted using larger circles. b, Relative abundances are also shown for samples taken on the animal-based diet. Labelled points correspond to diet days where subject 6's gut microbiota exhibited an increase in the relative abundance of *Bacteroides*. c, A principal-coordinates-based

characterization of overall community structure for subject 6, as well as all other subjects. QIIME³⁰ was used to compute microbial β diversity with the Bray–Curtis, unweighted UniFrac and weighted UniFrac statistics. Sample similarities were projected onto two dimensions using principal coordinates analysis. Top, when coloured by subject, samples from subject 6 (green triangles) partition apart from the other subjects' samples. Bottom, of all of subject 6's diet samples, the ones most similar to the other subjects' are the samples taken while consuming the animal-based diet.



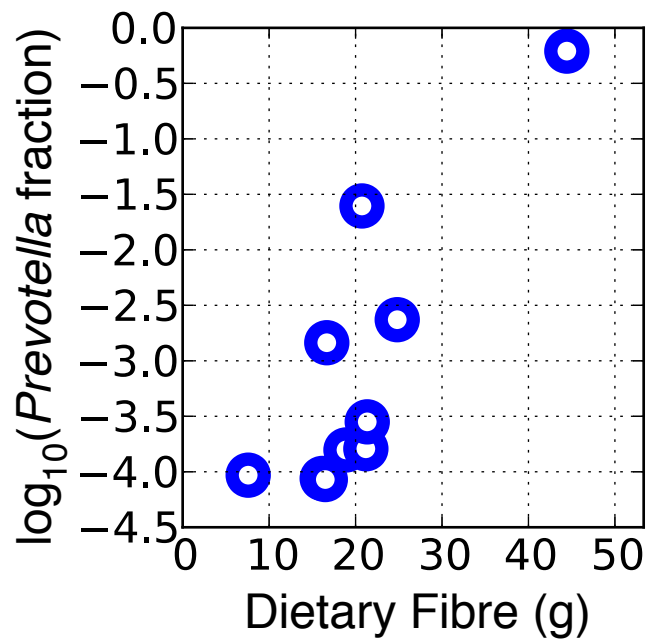
Extended Data Figure 3 | Subject physiology across diet arms.

a, Gastrointestinal motility, as measured by the initial appearance of a non-absorbable dye added to the first and last lunch of each diet. The median time until dye appearance is indicated with red arrows. Subject motility was significantly lower ($P < 0.05$, Mann–Whitney U test) on the animal-based diet (median transit time of 1.5 days) than on the plant-based one (1.0 days).

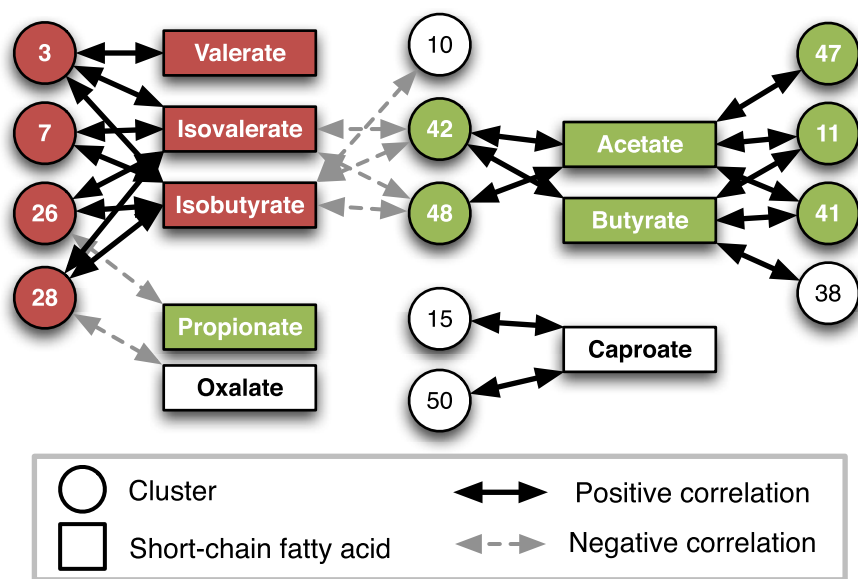
b, Range (shaded boxes) and median (solid line) of subjects' weights over time. Subjects' weight did not change significantly on the plant-based diet relative to baseline periods, but did decrease significantly on the animal-based diet (asterisks denote $q < 0.05$, Bonferroni-corrected Mann–Whitney U test).

Subjects lost a median of 1.6% and 2.5% of body weight by days 3 and 4, respectively, of the animal-based diet arm. **c**, Measurements of subjects' urinary

ketone levels. Individual subjects are shown with black dots, and median values are connected with a black solid line. Urinary ketone readings were taken from day 0 of the animal-based diet onwards. Ketone levels were compared to the readings on day 0, and asterisks denote days with significant ketone increases ($q < 0.05$, Bonferroni-corrected Mann–Whitney U test; significance tests were not carried out for days on which less than four subjects reported their readings.). All subjects on the animal-based diet showed elevated levels of ketones in their urine by day 2 of the diet ($\geq 15 \text{ mg dl}^{-1}$ as compared to 0 mg dl^{-1} during initial readings), indicating that they experienced ketonuria during the diet arm. This metabolic state is characterized by the restricted availability of glucose and the compensatory extraction of energy from fat tissue⁵⁶.



Extended Data Figure 4 | Baseline *Prevotella* abundance is associated with long-term fibre intake. *Prevotella* fractions were computed by summing the fractional 16S rRNA abundance of all OTUs whose genus name was *Prevotella*. Daily intake of dietary fibre over the previous year was estimated using the Diet History Questionnaire³² (variable name "TOTAL_DIETARY_FIBER_G_NDSR"). There is a significant positive correlation between subjects' baseline *Prevotella* abundance and their long-term dietary fibre intake (Spearman's $\rho = 0.78$, $P = 0.008$).



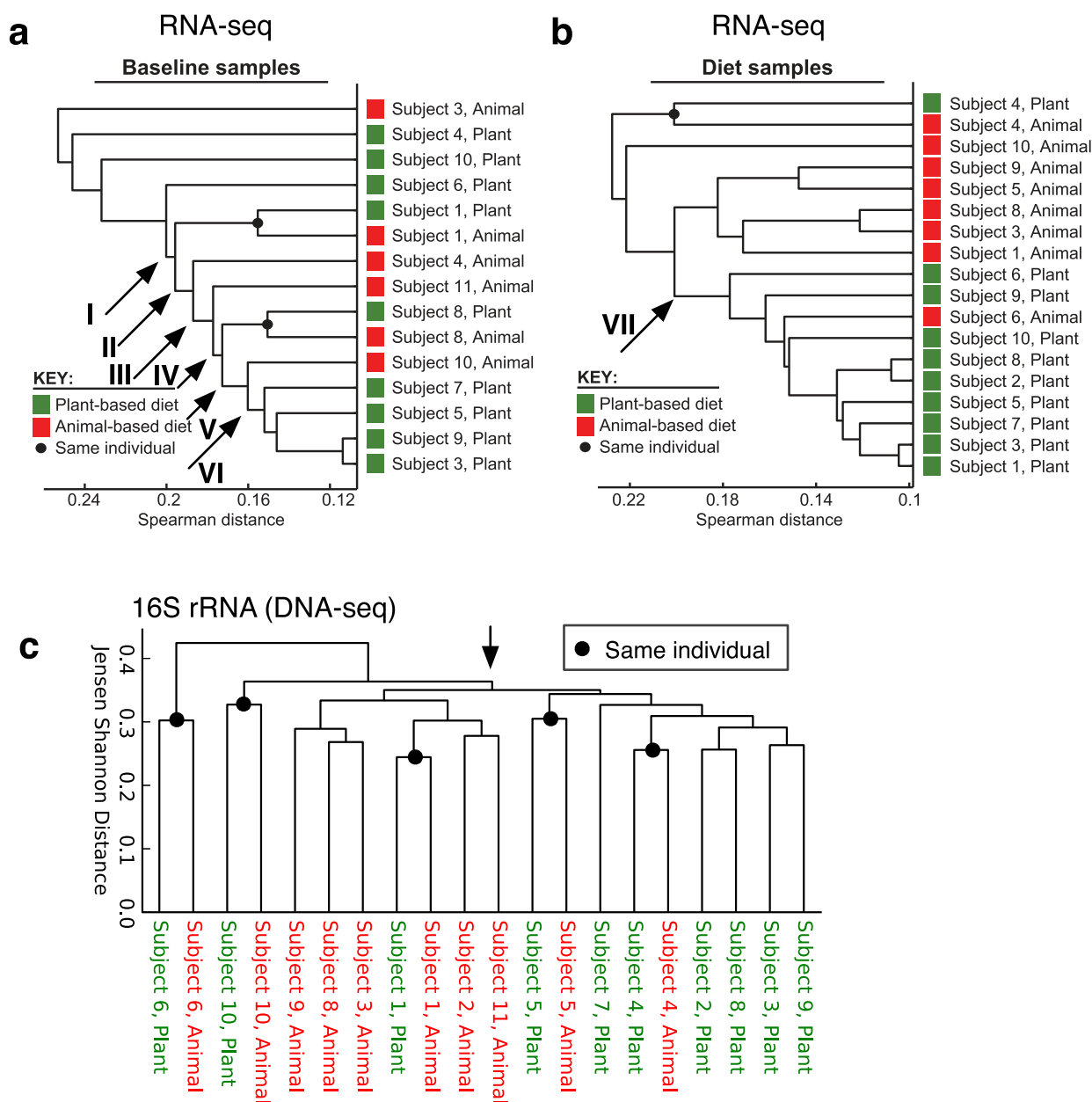
Extended Data Figure 5 | Significant correlations between SCFAs and cluster abundances across subjects. SCFAs are drawn in rectangles and coloured maroon or green if they are produced from amino acid or carbohydrate fermentation, respectively. Clusters whose members include known bile-tolerant or amino-acid-fermenting bacteria^{15,16} are coloured

maroon, whereas clusters including known saccharolytic bacteria³ are coloured green. Uncoloured clusters and SCFAs are not associated with saccharolytic or putrefactive pathways. Significant positive and negative correlations are shown with black arrows and grey arrows, respectively ($q < 0.05$; Spearman correlation).

Cluster membership

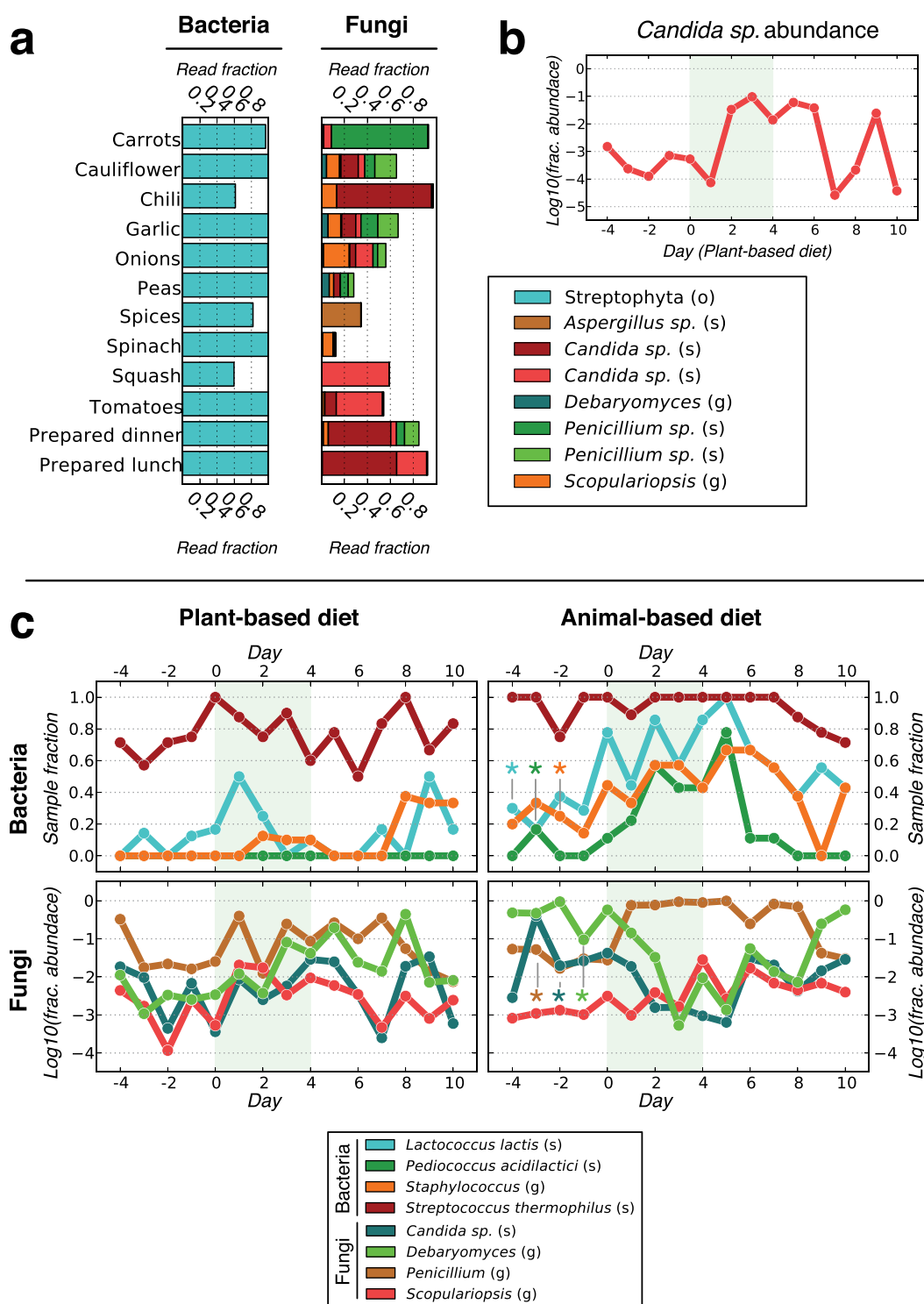
3	7	26	28	Bile / Putrefactive
				<i>Bilophila</i>
				<i>Alistipes</i>
				<i>Bacteroides</i>

11	41	47	42	48	Saccharolytic
					<i>Roseburia</i>
					<i>E. rectale</i>
					<i>F. prausnitzii</i>
					<i>Bifidobacterium</i>



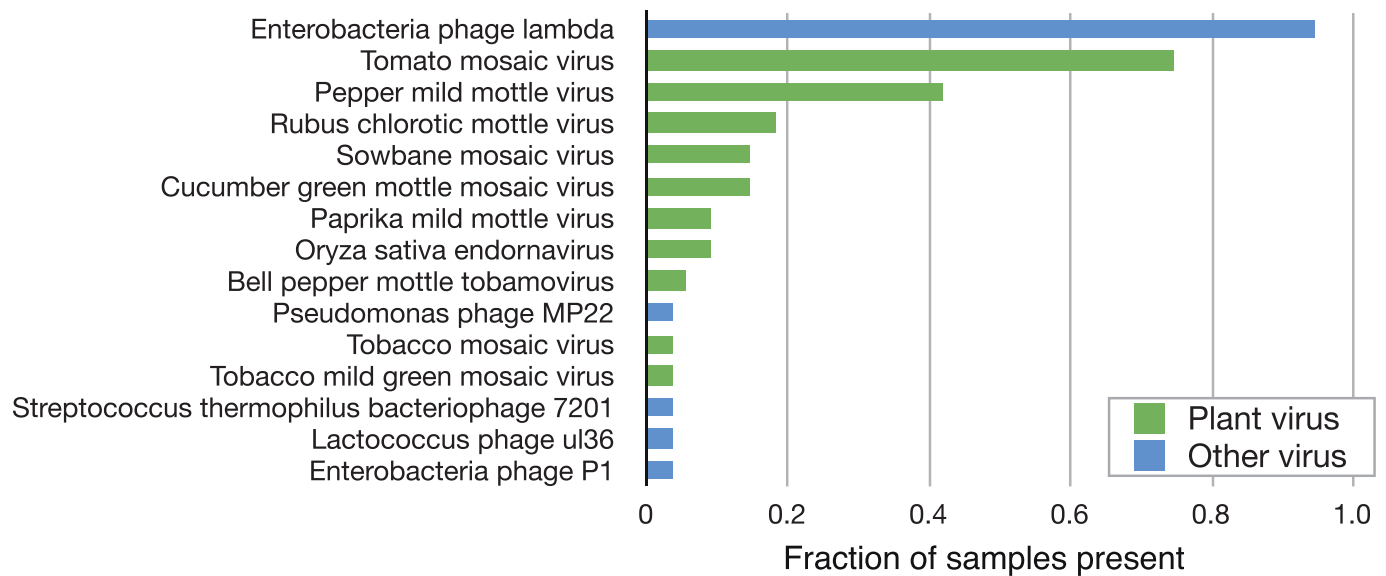
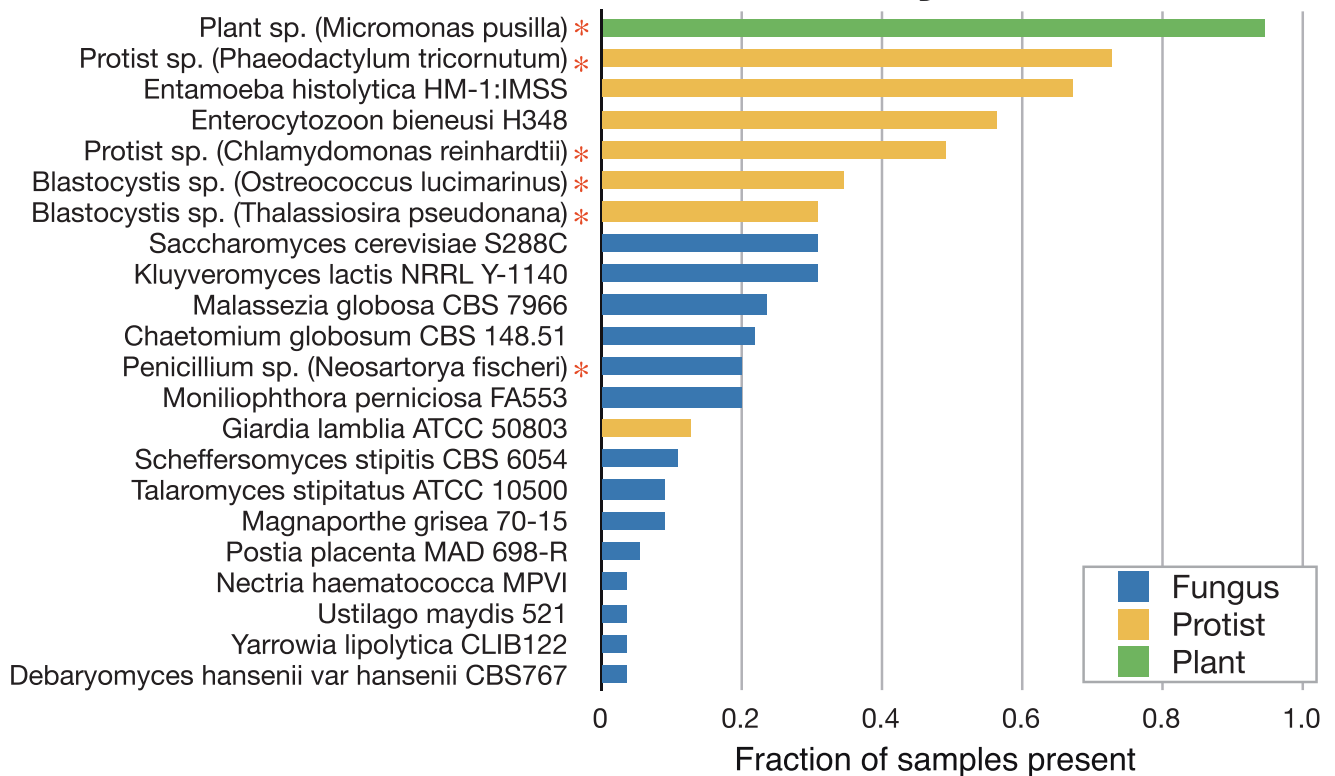
Extended Data Figure 6 | Inter-individual microbial community variation according to diet and sequencing technique. **a, b,** To measure the degree to which diet influences inter-individual differences in gut microbial gene expression, we clustered RNA-seq profiles from baseline (**a**) and diet (**b**) periods. Dots indicate pairs of samples that cluster by subject. The potential for diet to partition samples was measured by splitting trees at the arrowed branches and testing the significance of the resulting 2×2 contingency table (diet versus partition; Fisher's exact test). To avoid skewed significance values caused by non-independent samples, we only clustered a single sample per subject, per diet period. In the case of multiple baseline samples, the sample closest to the diet intervention was used. In the case of multiple diet samples, the last sample during the diet intervention was kept. A single sample was

randomly chosen if there were multiple samples from the same person on the same day. No association between diet and partitioning was found for partitions I–VI ($P > 0.05$). However, a significant association was observed for partition VII ($P = 0.003$). **c,** To determine whether diet affects inter-individual differences in gut microbial community structure, we hierarchically clustered 16S rRNA data from the last day of each diet arm. Samples grouped weakly by diet: sub-trees partitioned at the arrowed node showed a minor enrichment for plant-based diet samples in one sub-tree and animal-based diet samples in the other ($P = 0.07$; Fisher's exact test). Still, samples from five subjects grouped by individual, not diet (indicated by black nodes), indicating that diet does not reproducibly overcome inter-individual differences in gut microbial community structure.



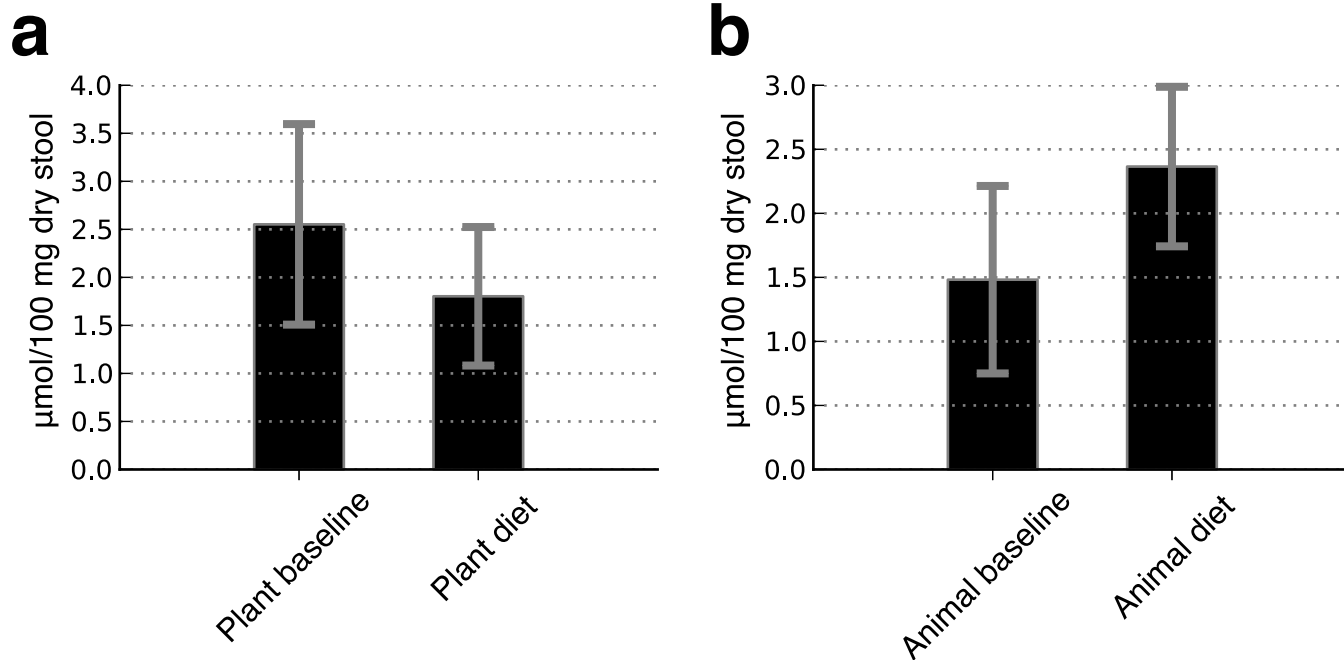
Extended Data Figure 7 | Food-associated microbes and their enteric abundance over time. **a**, Major bacterial and fungal taxa found in plant-based diet menu items were determined using 16S rRNA and ITS sequencing, respectively, at the species (s), genus (g) and order level (o). The majority of 16S rRNA gene sequences are Streptophyta, representing chloroplasts from the ingested plant matter. **b**, One of the fungi from **a**, *Candida sp.*, showed a significance increase in faecal abundance on the plant-based diet ($P < 0.05$, Wilcoxon signed-rank test). **c**, Levels of bacteria and fungi associated with the animal-based diet are plotted over the plant- and animal-based diet arms. Taxa are identified on the genus (g) and species (s) level. The abundance of foodborne bacteria was near our detection limit by 16S rRNA gene sequencing; to minimize resulting measurement errors, we have plotted the fraction of

samples in which bacteria are present or absent. *Lactococcus lactis*, *Pediococcus acidilactici* and *Staphylococcus*-associated reads all show significantly increased abundance on the animal-based diet ($P < 0.05$, Wilcoxon signed-rank test). Fungal concentrations were measured using ITS sequencing and are plotted in terms of log-fractional abundance. Significant increases in *Penicillium*-related fungi were observed, along with significant decreases in the concentration of *Debaryomyces* and a *Candida sp.* ($P < 0.05$, Wilcoxon signed-rank test). One possible explanation for the surprising decrease in the concentration of food-associated fungi is that the more than tenfold increase in *Penicillium* levels lowered the relative abundance of all other fungi, even those that increased in terms of absolute abundance.

a**Viruses****b****Eukaryotes**

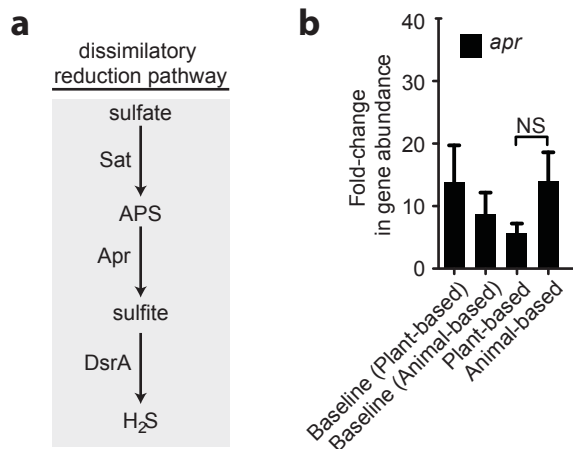
Extended Data Figure 8 | Eukaryotic and viral taxa detected via RNA-seq.
a. Identified plant and other viruses. The most common virus is a DNA virus (lambda phage) and may be an artefact of the sequencing process. **b.** Identified fungi, protists and other eukaryotes. Taxa that were re-annotated using

manually curated BLAST searches are indicated with asterisks and their original taxonomic assignments are shown in parentheses (see Methods for more details).



Extended Data Figure 9 | Faecal bile acid concentrations on baseline, plant- and animal-based diets. a, b, Median bulk bile acid concentrations are shown for all individuals on the plant-based (a) and animal-based (b) diets (error bars denote median absolute deviations). For detailed experimental protocols, see Methods. Bile acid levels did not significantly change on the

plant-based diet relative to baseline levels ($P > 0.1$, Mann–Whitney U test). However, bile acid levels trended upwards on the animal-based diet, rising from $1.48 \mu\text{mol}$ per 100 mg dry stool during the baseline period to $2.37 \mu\text{mol}$ per 100 mg dry stool ($P < 0.10$, Mann–Whitney U test).



Extended Data Figure 10 | The dissimilatory sulphate reduction pathway.

a, Microbes reduce sulphate to hydrogen sulphide by first converting sulphate to adenosine 5'-phosphosulphate (APS) via the enzyme ATP sulphurylase (Sat). Next, APS is reduced to sulphite by the enzyme APS reductase (Apr). Finally, the end product hydrogen sulphide is reached by reducing sulphite through the enzyme sulphite reductase (DsrA). This last step of the pathway can be performed by *Bilophila* and is thought to contribute to intestinal inflammation⁶. **b**, No significant changes in *apr* gene abundance were observed on any diet ($P > 0.05$, Mann-Whitney U test; $n = 10$ samples per diet arm). Values are mean $\pm$ s.e.m. However, *dsrA* abundance increased on the animal-based diet (Fig. 5d). NS, not significant.