Unravelling the effects of the environment and host genotype on the gut microbiome

Aymé Spor, Omry Koren* and Ruth Ley*

Abstract | To what extent do host genetics control the composition of the gut microbiome? Studies comparing the gut microbiota in human twins and across inbred mouse lines have yielded inconsistent answers to this question. However, candidate gene approaches, in which one gene is deleted or added to a model host organism, show that a single host gene can have a tremendous effect on the diversity and population structure of the gut microbiota. Now, quantitative genetics is emerging as a highly promising approach that can be used to better understand the overall architecture of host genetic influence on the microbiota, and to discover additional host genes controlling microbial diversity in the gut. In this Review, we describe how host genetics and the environment shape the microbiota, and how these three factors may interact in the context of chronic disease.

Microbiota
A microbial community or assemblage.

Dysbiosis
A shift in the relative abundances of the microbial taxa compared with the adundances that are observed in healthy animals.

Microbiome
The complete set of genes within a microbiota.

Germ-free
Pertaining to an animal: lacking a microbiome; born and raised under sterile conditions for research purposes.

Adiposity
The property of containing fat.

Department of Microbiology, Cornell University, Ithaca, New York 14853, USA.
**These authors contributed equally to this article.*
Correspondence to R.L.
e-mail: rel222@cornell.edu
doi:10.1038/nrmicro2540

The colonization of the human gut begins at birth and is characterized by a succession of microbial consortia, the composition of which is influenced by changes in diet and by life events[1,2]. The diversity and richness of the microbiota reach adult levels in early childhood, and the composition is thought to then remain relatively stable and resilient to stresses, such as antibiotic treatments[3]. The gut microbiota is highly variable from person to person[4–7] and in different body sites in a single host (FIG. 1), but family members tend to harbour more similar microbiota than unrelated individuals and, indeed, the same bacterial strains can be shared among family members[7–10]. Similarities between the microbiota of related individuals could be due to a shared environment, but may also reflect host genetic relatedness. Environmental and stochastic factors can strongly affect the composition of the microbiota, but the effects of host genetics on shaping this vital 'microbial organ' are less clear[11–13].

A better understanding of how the gut microbiota are assembled and maintained is increasingly relevant to the treatment of complex chronic diseases. A growing number of studies highlight the fact that certain microbiota can be harmful to host health. Dysbioses of the microbiome are associated with an expanding list of chronic diseases that includes obesity[7], inflammatory bowel disease (IBD)[14,15] and diabetes[16]. Examples of dysbioses are shown in FIG. 2: studies have shown that patients with IBD (FIG. 2a) harbour fewer bacteria from the phylum Bacteroidetes, fewer from the phylum Firmicutes and more from the phyla Actinobacteria and Proteobacteria than healthy subjects[17]. People with type 2 diabetes (FIG. 2b) have reduced numbers of bacteria from the Firmicutes and instead harbour a greater proportion of bacteria belonging to the Bacteroidetes[16]. The microbiota of infants with necrotizing enterocolitis (FIG. 2c) is composed of members of the Proteobacteria and Firmicutes only, whereas healthy infants also harbour species from the Bacteroidetes and Fusobacteria[18]. These types of correlative observations raise the question of whether the microbiota has a causative role in disease, or whether dysbiosis is a by-product of the disease. For several diseases, recent work shows the answer to be that the microbiota does contribute to disease. Transplantation experiments in which the microbiota of a diseased animal is grafted into a germ-free healthy recipient have demonstrated that several disease phenotypes could be transferred by the microbiota. These include excess adiposity[19], metabolic syndrome[20] and colitis[21], all of which are traits of complex diseases that are also affected by host genetic and environmental factors.

Recent work has also highlighted a direct beneficial effect of gut microbial communities on the host. For instance, the microbiota can be protective against the onset of type 1 diabetes: germ-free non-obese diabetic (NOD) mice are genetically susceptible to developing type 1 diabetes, but the diabetes prevalence in these

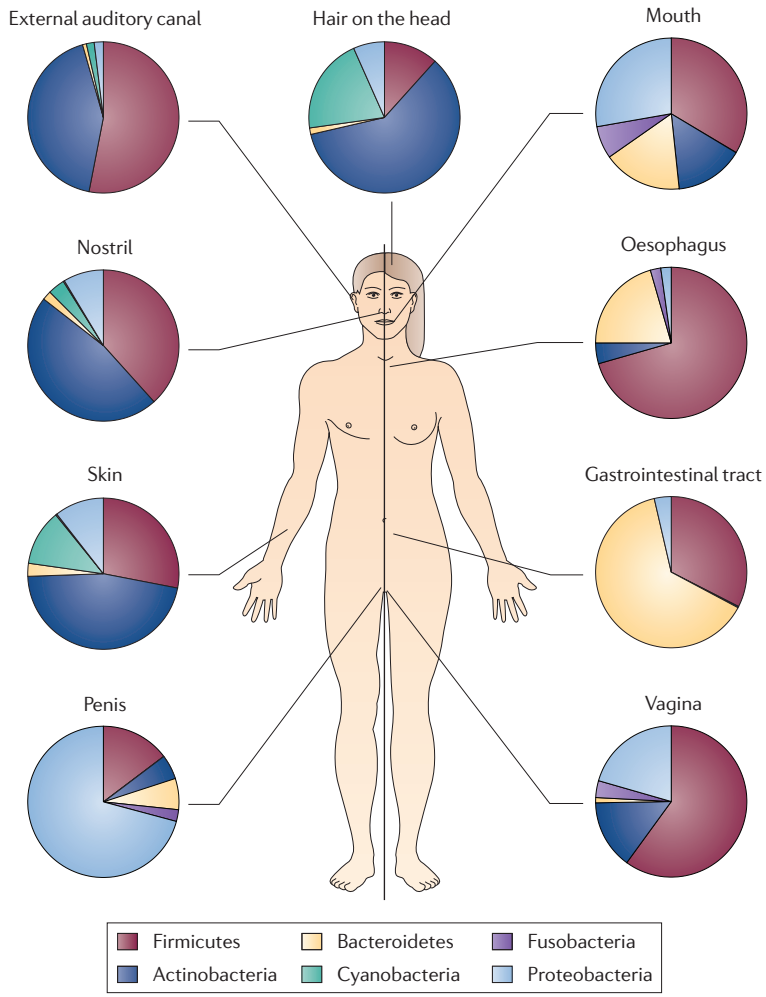

Figure 1 | **Microbial community composition at different body locations in a healthy human.** The relative abundances of the six dominant bacterial phyla in each of the different body sites: the external auditory canal (nine subjects), the hair on the head (nine subjects), the mouth (ten subjects), the oesophagus (four subjects), the gastrointestinal tract (nine subjects), the vagina (eight subjects), the penis (12 subjects), the skin (nine subjects) and the nostril (nine subjects). Data taken from REFS 93–97.

β-diversity
A measure of diversity that describes the differences between any two ecosystems (for example, the UniFrac distance metric). Related to α-diversity and γ-diversity, which are measures of the diversity in a single ecosystem and across a group of ecosystems, respectively.

UniFrac
A β-diversity measure that is phylogeny based. Microbial communities are more similar if they are composed of members that are more closely related, phylogenetically, as this implies a shared evolutionary past. UniFrac units range from 0 (identical communities) to 1 (totally different communities).

mice can been reduced by exposure to the microbiota from NOD mice lacking myeloid differentiation primary response protein 88 (MYD88)[22]. In humans, anecdotal evidence suggests that specific microbiota can effectively treat IBD, and microbial transplantation of whole microbial communities is sometimes used as therapy in pseudomembranous and ulcerative colitis, chronic constipation, Crohn's disease and *Clostridium difficile*-associated diarrhoea[23,24]. In a recent case study of *Clostridium difficile*-associated colitis, the amelioration of symptoms was associated with the persistence of the transplanted healthy community in the patient[25].

Together, these recent studies suggest that a specific combination of microorganisms in the gut can affect host health; therefore, host control over the microbiota could help maximize fitness. Biogeographical patterns of diversity in a single host show that the physiochemical properties of the human gut habitat are very important selection pressures for its microbial constituents (FIG. 1). Variations in those host genes that contribute to properties of the gut habitat therefore have strong potential to affect the variation in the microbiome. Evidence to support a contribution of host genetics to the diversity of the microbial community has been scarce, so the strength of the effect is controversial. However, an increasing number of studies are now evaluating this effect, and the analysis of host genetics is just beginning to be incorporated into studies of how the diversity of the gut bacteria relates to host susceptibility to disease.

In this Review, we describe how environmental factors can contribute to variation in the diversity and composition of the microbiota, and we explore the role of host genes in this process. We also highlight an emerging view of the microbiota: one in which the microbiota itself may be considered as a complex trait that is under host genetic control and that interacts with environmental and host factors in a number of chronic inflammatory diseases.

Environmental impact on the microbiota

To measure the impact of host genetics on microbial diversity, it is useful to have an understanding of the factors that can influence variation in the microbiota in the absence of host genetic variation, as these environmental factors constitute the 'noise' that can mask host genetic effects. Model organisms provide a system for controlling variation between identical hosts: genetically inbred animals act as replicate hosts, allowing the impact of environmental factors on the variation in the microbiota to be assessed. Mice are useful models for studies of human microbial ecology because the intestines of mice harbour communities that are grossly similar in composition (that is, have similar phylum and family level abundances) to those of human intestines, diverging mainly at the genus level (BOX 1). Husbandry conditions can be standardized across mice, and experiments can incorporate full factorial designs for testing the effects of various parameters on microbial diversity.

'Stochastic variability' of the microbiome. One of the earliest factors that can have a profound influence on the microbiota composition is the maternal environment (BOX 2). Several studies have shown that genetically identical mice from the same litters have a more similar microbiota than mice from different litters, even though they may be reared in adjacent cages[22,26,27]. This 'maternal effect' occurs when mouse pups are born vaginally and the birth mother's microbiota is their primary inoculum. Maternal effects can influence bacterial β-diversity[28] (measured by UniFrac) regardless of host genotype[22,27], as well as affecting the relative abundances of phylotypes[26]. The maternal effect has been documented across two[27] and as far out as four generations[26]. As a consequence, the maternal effect can be a major confounding factor when comparing the microbiota of mice with different genotypes or under different treatments.

But, despite shared environments and parental inocula, substantial differences in community composition and structure can exist between littermates reared in the same cage. Although many bacterial phylotypes can be

shared across littermates, often the majority of dominant phylotypes in an animal's gut bacteria are unique or shared with just a subset of other animals[22,27]. Separating littermates into different cages can drive the differences in their microbiota further. For instance, when the abundances of the eight members of the altered Schaedler flora[29] were analysed in isogenic mice, it was found that those mice that were cohabiting at weaning, whether from the same or different litters, had little variation in their microbiota profiles. By contrast, the microbiota of litters split among different cages at weaning diverged in composition. Interestingly, the degree of divergence depended on the genotype of the mouse[30]. Thus, although the initial inoculum may be largely obtained from the mother, stochastic differences in the colonization process between mice, and subtle differences in their environments, interact with the mouse genotype to determine inter-mouse variation in the microbiome[31].

Effect of diet on microbiome variation. Diet is one of the most important factors shaping microbial diversity in the gut, and its effect on the composition of the human microbial community is reviewed elsewhere[32,33]. Here, we highlight how changes in the diet can alter the relative abundances of the taxa that are already present in the community. One family of the Firmicutes in particular, the Erysipelotrichaceae, has been shown by several independent studies to alter in abundance in response to changes in the amount of dietary fat. After inducing obesity in mice by feeding them a 'Western' diet (high in saturated and unsaturated fats), a bloom occurred for an uncultured member of the family Erysipelotrichaceae that is related to the human-associated *Eubacterium dolichum*[19]. The relative abundance of this uncultured phylotype diminished when the mouse diet was changed to the usual mouse chow[19]. In a subsequent set of experiments using 'humanized' mice (formerly germ-free mice harbouring a human faecal microbiota), human-derived erysipelotrichi were found to bloom under a high-fat diet[34]. Other groups have also noted that erysipelotrichi respond to dietary fat: for instance, it has been reported that four clades of this family reacted differently (either increased or decreased in abundance) to high-fat and low-fat diets in mice[35]. In humans, changes in diet composition can also lead to shifts in the abundances of specific gut taxa. For example, changes in the dietry amounts of particular carbohydrates result in changes in population levels of the butyrate-producing *Roseburia* spp.[36]. *Bacteroides* spp. differ in their ability to use specific substrates such as inulin, and these differences can predict the outcomes of competitive interactions between the species[37]. Microbial specialization to diet substrates probably underlies the high species diversity of the gut microbiota, as bacterial species partition the niche space according to their substrate preference and use and, as a result, modulation of the diet composition alters the relative abundances of the taxa that are present.

Host genetics and the heritability of the microbiota

A significant association between variation in the composition of the gut microbiota and variation in the genotype of the host would be a hallmark of genetic control. This type of influence is distinct from inheritance of the microorganisms themselves via 'non-genetic' transmission between generations (for example, the maternal effect). In the simplest scenario, specific host alleles would result in a different microbiota that may be detrimental or beneficial to host health. Studies using human twins, comparisons between mouse lines, and a more recent 'quantitative trait loci' (QTL) detection approach have measured the heritability of the gut microbiota; these studies have yielded contrasting but informative results, as discussed below.

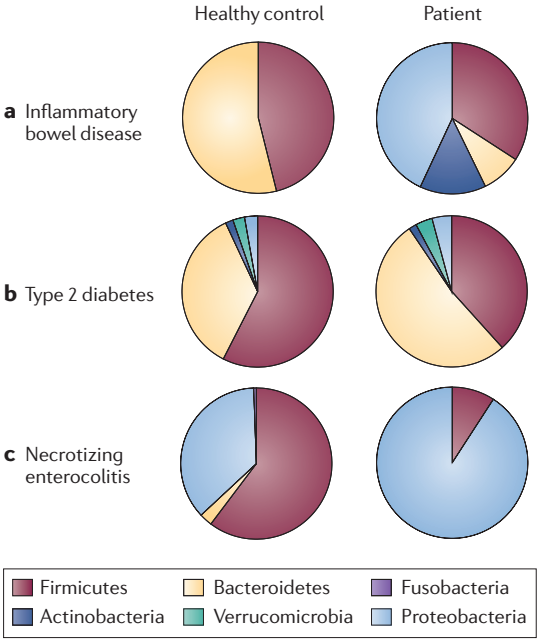

Figure 2 | **Gut microbial dysbiosis associated with disease.** The relative abundances of the predominant bacterial phyla: in caecal samples from patients with inflammatory bowel disease (using clone libraries for bacterial identification)[17] (part **a**); in faecal samples from ten healthy controls and ten patients with type 2 diabetes (using pyrosequencing)[16] (part **b**); and in faecal samples from ten healthy infants and ten infants with necrotizing enterocolitis (using clone libraries)[18] (part **c**).

Human twin studies. Several studies have used comparisons between monozygotic (MZ; identical) and dizygotic (DZ; fraternal) twins to ascertain the heritability of the microbiota[13]. Heritability can be assessed using a classic technique in which a measure of the phenotypic trait of interest is correlated for twin pairs, and the strength of the correlation is compared for MZ versus DZ twin pairs (that is, $h^2 = 2 \times (r_{MZ} - r_{DZ})$, in which h^2 is heritability and r is the correlation between twins). In traditional twin studies, it is assumed that the resemblance between twins that is due to common environmental effects is the same for MZ and DZ twins. For any given component of the microbiota, a greater within-pair similarity for MZ twins than for DZ twins would be an indication of heritability. Heritable aspects of the microbiota that are under host genetic control could include, for

Altered Schaedler flora
A standard enteric flora containing eight species that are known to exhibit tissue tropism, occupying different niches in the mouse gastrointestinal tract.

Quantitative trait locus
A genomic region for which variation is associated with the quantitative variation in a phenotypic trait.

Heritability
The proportion of phenotypic variation in a population that is attributable to genetic variation among individuals.

Box 1 | **Animal models and the microbiome**

Mice are the most widely used mammalian model for microbiome studies[75]. Mice and humans share 99% of their genes and differ by 14% in genome size (2.5 gigabases and 2.9 gigabases, respectively)[76]. Despite their vastly different overall body size, intestinal physiology (mice have a relatively larger caecum) and diet (for example, mice are coprophagous), the same phyla dominate the distal guts of mice and humans: Firmicutes (usually 60–80% of 16S ribosomal RNA gene sequences)[27], Bacteroidetes (usually 20–40% of 16S rRNA gene sequences)[27] and Actinobacteria. Indeed, these same bacterial phyla are found to inhabit the gastrointestinal tracts of many other mammals as well[77]. However, when comparing the bacterial genera in human microbiomes (11,831 colon-associated 16S rRNA sequences) and mice microbiomes (5,088 caecum-associated 16S rRNA sequences), only 15% of the mouse microbiome genera have been found to be represented in humans (see the figure, which shows a comparison of the bacterial diversity from 16S rRNA analysis of mouse caeca and human colons; the bar represents 15% sequence divergence)[27]. Other animal models used for microbiome studies are rats, zebrafish and pigs[75]. These animals can also be reared germ free and then administered with a human microbiota.

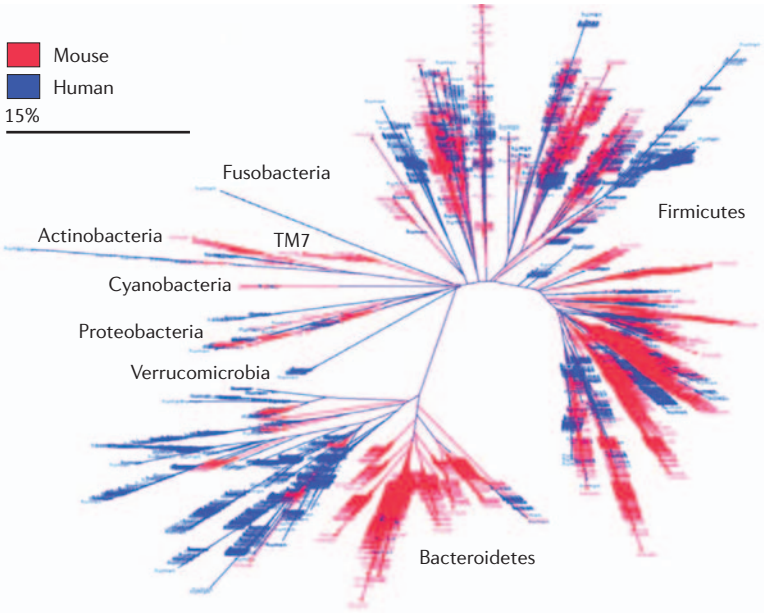

Figure is reproduced, with permission, from REF. 27 © (2005) US National Academy of Sciences.

example, functions, phylotypes, or levels of functions or phylotypes that are more similar within MZ twin pairs than within DZ twin pairs.

The results of twin studies to date are surprisingly equivocal[13]. Early reports supported the hypothesis that host genotype influences the composition of the human gut microbiota, although the size of the effect is quite small. Fingerprinting-based comparisons of the faecal microbiota of 20 or fewer MZ and DZ twin pairs revealed a slightly greater similarity between the microbiota of MZ twins than between those of DZ twins[38,39], supporting a host genetic influence on the microbiota. However, these studies did not identify which specific members of the microbiota were heritable. By contrast, a more recent study of 31 MZ and 23 DZ twin pairs used metagenomics and 16S ribosomal RNA gene sequence data from clone libraries and from 454-pyrosequencing to characterize the faecal microbiomes and the taxa present, but did not detect any heritable components[7]. The 16S rRNA-based analysis indicated that the bacterial diversity within twin pairs was less for MZ twins than for DZ twins (by unweighted UniFrac), but this difference between twin types was not significant[7]. The different methodologies used by these studies to address the question of microbiome heritability may account for the different outcomes; nevertheless, to date no studies have been adequately powered to specifically identify the heritable components of the microbiome.

Subsequent deep pyrosequencing (1.2–1.5 million reads per sample) of 16S rRNA genes in a pair of MZ twins from this more recent study revealed that ~100,000 16S rRNA sequences were required to observe 60% of the total species-level operational taxonomic units (OTUs) (with 97% identity). With ~1 million sequences, a majority (68% and 79%) of species-level OTUs were shared with those from the co-twin's microbiota[40]. Metagenomic sequencing (3.8–6.3 gigabase pairs per sample) revealed that a majority of reads could not be mapped to reference genomes and only 17% of the genes that could be mapped were shared between the co-twins. Genes with cellulytic activity, mapping to the genus *Faecalibacterium*, were highly enriched in just one twin. These findings further underscore the need for deeper sequencing of 16S rRNA data and metagenomic data in order for the heritability of the microbiome's phylogenetic content and functionality to be estimated.

The lack of a strong effect for host genotype across twin studies is surprising given that, in many ways, MZ twins are more similar to one another than DZ twins are to each other. Importantly for the microbiota, diet and lifestyle preferences have been shown to be heritable in twin studies[41,42]. In addition to diet, which is mentioned above, lifestyle may also impact the microbiome. For instance, levels of exercise have been shown to affect β-diversity in the rat caecal microbiota (as estimated by TGGE (temperature gradient gel electrophoresis)) and to increase caecal butyrate levels[43]. In humans, the effect of exercise on weight loss has been shown to be dependent on the composition of the initial microbiota[44]. If host genotype does exert an effect on the composition of the microbiota — either directly, through secretions (for example, of bile and defensins) into the gut, control of gut motility or modification of epithelial cell surfaces, or indirectly, through food and lifestyle preferences — the effects are likely to be small, and detecting them in a healthy population will require a large number of subjects.

Comparisons between mouse lines. Early studies comparing the microbiota of different genetically inbred lines of mice took steps to reduce the maternal effect that confounds genotype effects. Several approaches have been used that either reduce the maternal effect experimentally or account for it analytically; cross-fostering (swapping offspring between two mothers after birth), uterine transplants of embryos of one genotype into a dam of another genotype, and inoculation of one microbiota into a set of germ-free mice have all been used as methods of standardizing the microbiota, or an adequate statistical modelling framework can be used to analyse the data obtained from non-standardized mice.

Fingerprinting-based comparison
A molecular technique for the study of nucleic acids using either denaturing-gradient gel electrophoresis (DGGE) or temporal temperature gradient gel electrophoresis (TTGE). Another method is terminal restriction fragment length polymorphism (T-RFLP), in which the DNA is amplified using fluorescence-labelled primers, digested using restriction enzymes and then detected. In microbial ecology, these techniques are used to compare microbial communities.

Dam
A female animal parent.

Box 2 | **Assembly and stability of the gut microbiota, and environmental factors affecting the gut microbiome during life**

The microbial colonization process of human body habitats begins at birth, when the baby leaves the uterus. The initial microbial communities are assembled from organisms in the immediate surroundings: for vaginally delivered infants, the mother's vagina is a major source of the initial colonizing bacteria, and for those born by caesarian section, the hands that touch the baby are a dominant source[78]. Specific strains of gut bacteria that are acquired from the mother can be detected initially in the infant but are probably often outcompeted by other strains of less certain origin[10,79,80]. Gut communities start with low phylogenetic and species richness, which increases as a function of time[1,2]. Time in this context may reflect the rate of encounters with new bacteria, the increasing size of the gut or the proliferation of ecological niches that can promote diversity. The introduction of solid foods alters the relative proportions of bacterial phyla in the gut and is followed by the establishment of an adult-like microbiota characterized by a full suite of functions[1] and with greater stability. Thus, infancy is a period of rapid colonization by microbial consortia that can shift in response to events such as illness or changes in diet (see the figure, part **a**, which summarizes the succession of bacterial consortia (in terms of the four dominant phyla) over time (days after birth) in a developing infant gut microbiome from birth to around 2.5 years of age, according to 16S ribosomal RNA sequencing).

How changes in lifestyle, illness, puberty, and so on affect the microbiota and its stability is still a matter of speculation. The idealized Westernized human undergoes certain life stages, during which the diversity (richness) and the temporal stability of the microbiota alters. Throughout the lifetime of this typical Westernized human, various factors are thought to influence the microbiome (see the figure, part **b**).

The most numerically abundant types of bacteria in the mammalian gut are hardly, if ever, detected outside of it[81], suggesting that they are true gut specialists (for example, members of the orders Clostridiales and Bacteroidales in humans). These specialized bacteria have followed a unique evolutionary trajectory: gut bacterial communities are largely phylogenetically distinct from their non-gut (environmental) counterparts[81]. Most importantly, microorganisms have allowed the mammalian dietary repertoire to expand to include plants and other foods that would otherwise be too toxic, too low in energy value and too nutrient poor to support a healthy host[77]. Owing to their dependence on gut microorganisms for nutrition, many mammals have evolved mechanisms to ensure a maximally beneficial consortium.

Comparisons of the genomes of gut microorganisms with the genomes of microorganisms living elsewhere have shown that distantly related members of the Bacteria and even the Archaea use a remarkably similar gene portfolio for life in the gut. These genes include a diverse range of carbohydrate-active genes, adhesins and bile salt hydrolases to facilitate processing of the diet and bacterial retention in the gut[82–84]. Gut microorganisms are also known to interact with the host immune system, to prime it and maintain homeostasis[85,86]. A key immunoregulatory property of gut bacteria is their ability to vary their surface properties; one example is the synthesis of certain polysaccharides that induce an immune system response[87,88]. Beyond general adaptations to the mammalian gut, a few bacterial lineages show host species-specific adaptations. A good example is *Lactobacillus reuteri*, a species composed of genetically distinct strains adapted to different mammalian hosts[89]. Thus, the co-evolution of host and microorganism is ongoing and sometimes quite rapid in evolutionary time.

Many vertebrates have specific behaviours that direct the assembly of the gut microbiota. For instance, the juveniles of some animals (such as horses and iguanas) consume the faeces of the adults (a process known as coprophagy), ensuring that the guts of the young are colonized by appropriate microorganisms[90]. In other animals, mothers deliberately feed faeces to their offspring: koalas inoculate their young directly with a special faecal pellet (a pap) to colonize the infant gut with bacteria that can detoxify secondary compounds of eucalyptus[91]. In many non-coprophagic mammalian species, parental care alone suffices to ensure that offspring are colonized by the 'right' microorganisms[92].

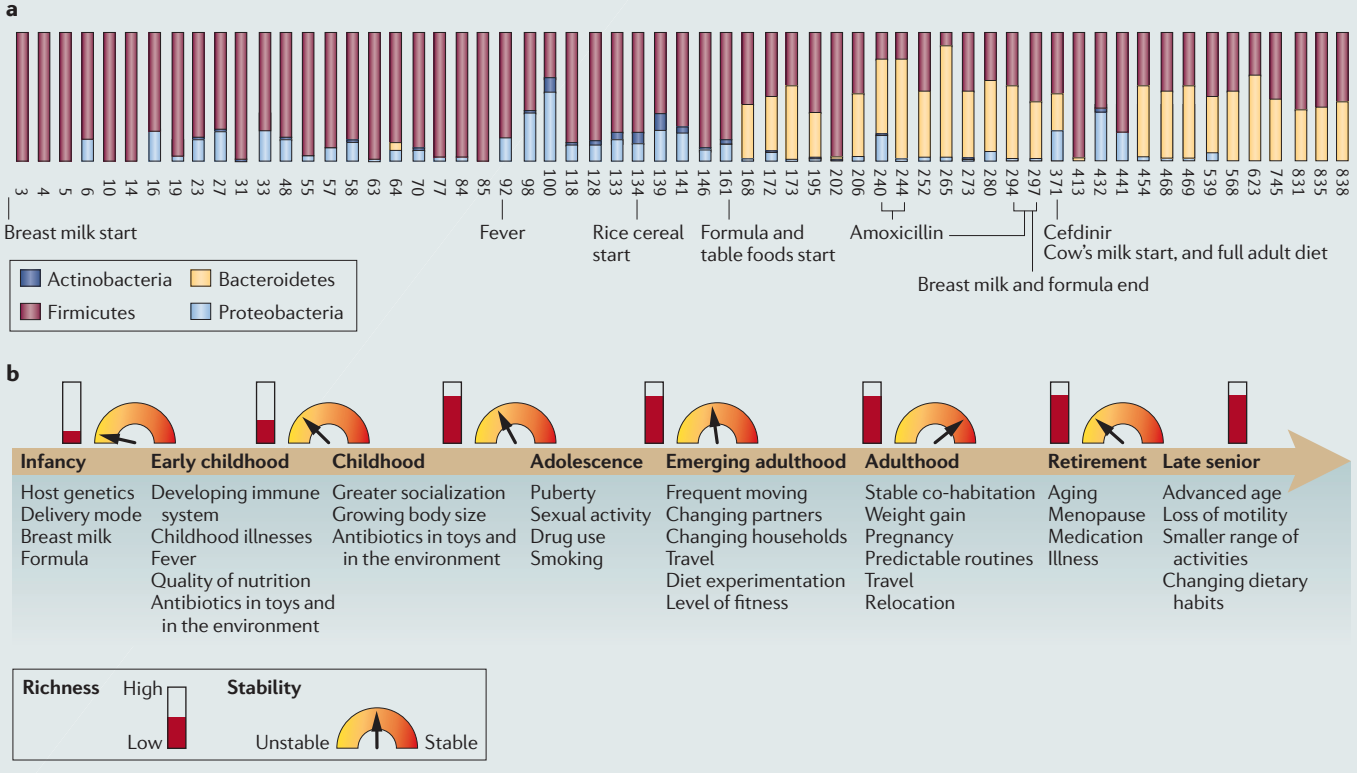

Figure part **a** is modified, with permission, from REF. 1 © (2010) US National Academy of Sciences.

Studies that have reduced the maternal effect experimentally in mice have reported some differences in the gut microbiota of mice of different lines. One study that eliminated the maternal effect by performing uterine transplantations to disassociate mouse pups from the microbiota of their family indicated that mice of different lines that were born together had similar microbiotas[45]. This result suggested that a new maternal effect was introduced from the birth mother. In another approach, three mouse lines (C57BL/10, C3H and BALB/c) were compared by associating germ-free recipients with a faecal slurry from a single mouse donor[46]. Faecal communities were profiled by DGGE (denaturing gradient gel electrophoresis) at several time points over the course of the 13-week experiment. Between 4 and 8 weeks post-inoculation, the C57BL/10 mouse microbiota were distinct from those of CH3 and BALB/c mice; after 8 weeks, the housing of the animals had a greater effect than their genotypes, particularly for BALB/c mice. It should be noted that a maternal effect is not always evident; for example, one recent study of eight mouse lines, using DNA fingerprinting methods, concluded that the microbiota was substantially different in different genetic backgrounds and saw little, if any, maternal effect[47]. In general, however, genetic polymorphisms, rearing and housing conditions — and their interactions — need to be incorporated into models for predicting the impact of host genetics on the microbiota.

QTL detection approach. The influences of the environment (including maternal effects) and of host genetics on compositional features of the gut microbiota have recently been analysed with quantitative genetics[26]. By using a large number of animals from intercross populations, host genetic background and environmental factors can be evaluated systematically; this study used a new, large mouse advanced intercross line[48,49] (AIL) (C57BL/6J crossed with an ICR-derived outbred line[50]) and generated 16S rRNA gene sequences from 645 mice. One result of this work was the detection of a core measurable microbiota (CMM) consisting of 64 conserved taxonomic groups that were present in most or all of the mice[26]. The analysis also revealed that litter and maternal effects explained 26% of the variation in abundances of the CMM taxa.

QTL analysis was used to test whether specific taxa co-segregated as quantitative traits linked with genomic markers (530 fully informative single-nucleotide polymorphism (SNP) markers). The QTL detection approach revealed 13 significant mouse genomic regions and five suggestive QTL for which host genetic variation is significantly associated with relative abundances of 26 of the 64 CMM taxa, including at least one taxon from each of the four major phyla (Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria). Host genetic control was found to affect the tips of the bacterial tree (that is, genus and species levels rather than higher-order taxa), particularly for the Bacteroidetes and Firmicutes.

In addition to showing that host genetic control can be measured using intercrosses, this study also began to unravel the genetic architecture that shapes the composition of the microbiota. In several instances, one QTL was associated with more than one taxon, indicating that host genetic composition can influence population structure. For example, the *Lactobacillus johnsonii–Lactobacillus gasseri* group segregated with two significant QTLs, one of which was adjacent to a QTL for *Turicibacter* spp. on mouse chromosome 7. This chromosome region either encodes genes that affect both groups or contains linked genes that, individually or in combination, affect the composition of the gut microbiota. Interestingly, both of these bacterial groups include species that have been shown to interact with $CD8^+$ T cells[51,52]. With all of the diversity found in the gut microbiota, it is interesting that the QTL screen found significant associations with bacteria that have been shown in other studies to have a direct mechanistic interaction with host immunity. It is tempting to speculate that the QTL highlighted here contain genes related to T cell function.

In another example from this study, a QTL associated with specific bacterial abundances was found to contain genes with important roles in mucosal immunity. Two populations with abundances that were correlated in the data (the genus *Lactococcus* and the family Coriobacteriaceae) segregated with a QTL that included several genes with immune functions: *Irak3* (encoding IL-1 receptor-associated kinase 3, which modulates the MYD88-dependent Toll-like receptor 2 (TLR2) pathway), *Lyz1* and *Lyz2* (two primary mouse lysozyme genes), *Ifng* (the interferon-γ gene) and *Il22* (the interleukin-22 gene). Several of these genes have been shown to affect the community composition or structure of the gut microbiota (see below). The demonstration that a host genotype can have an effect on the composition of its associated gut microbiota is a substantial step forward in our understanding of the how the gut ecosystem functions. The QTL study described above is the first, to our knowledge, to describe significant associations between variation at a given host locus and variation in the abundance of a given microbial taxon.

Genome-wide association studies (GWA studies) are needed to relate variation in the human genome to variation in the human microbiome. These types of studies can also use genotyped twins, as information about the host genotype adds powerful detail to twin studies. For example, the assumption that environmental effects are the same for MZ and DZ twins is not always valid: environmental effects such as the conditions in the maternal womb can differ for MZ and DZ twins, resulting in strong effects on some phenotypes[53] and potentially being important for the microbiota. Instead of estimating how much co-variance is due to environmental or genetic sources, the genetic relatedness of DZ twins can be measured directly from SNP or other host genetic data across the genome[54]. Such studies in large populations will enable the detection of genes for which variation is related to complex diseases that can be triggered by differences in the microbiota.

Single host genes that affect the microbiota

One human gene for which variation has been shown to influence the β-diversity of the gut microbiota is *MEFV*, encoding pyrin[55] (FIG. 3a). In humans, changes in

Advanced intercross line
An experimental mouse population used for the study of quantitative trait loci (QTL). This population is created by randomly intercrossing strains and then intercrossing the mice in subsequent generations until the desired number of generations is achieved. This results in mice with many recombinations in the genome, which is required for QTL studies.

Genome-wide association study
A genetic approach aiming to relate genome-wide variation in genetic markers to variation in any given phenotype. The microbiota, or its components, can be considered as a host phenotype.

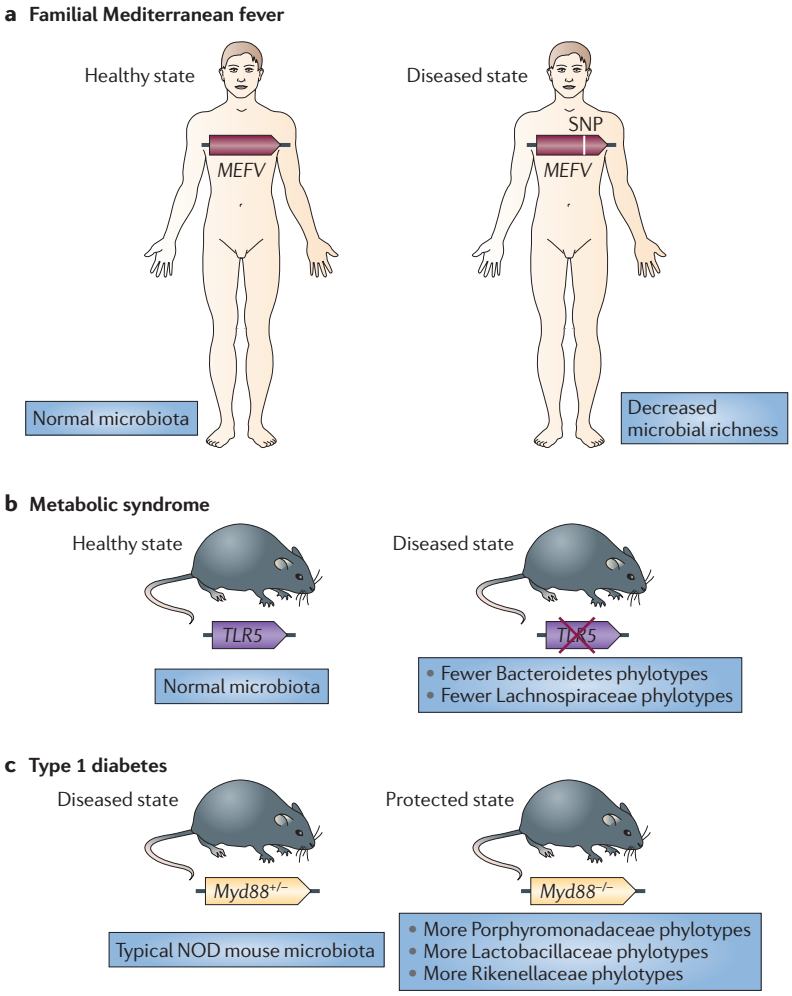

Figure 3 | **Host genetic variants, their effects on the composition of the gut microbiota, and the corresponding host disease state. a** | In humans, genetic variation in the form of a single-nucleotide polymorphism (SNP) at the *MEFV* locus, which encodes pyrin, has been associated with variation of the microbiota and, especially, lower microbial richness in patients with familial Mediterranean fever[55]. **b** | Toll-like receptor 5 (TLR5)-deficient mice display an altered microbiota (a decrease in the abundances of certain Bacteroidetes and Lachnospiraceae phylotypes) and the development of metabolic syndrome[20]. **c** | Myeloid differentiation primary response 88 (*Myd88*)-knockout non-obese diabetic (NOD) mice are protected against the development of type 1 diabetes by changes in the composition of their gut microbiota (increased abundances of Lactobacillaceae, Rikenellaceae and Porphyromonadaceae phylotypes) compared with the microbiota of NOD mice that are heterozygous for *Myd88* (REF. 22).

the gut microbiota are associated with a single mutation in *MEFV* that leads to familial Mediterranean fever (a hereditary autoinflammatory disorder affecting people with Mediterranean ancestors[56]; TABLE 1). For many other host genes, their effect in a mouse model has preceded the search for human gene variants that might co-vary with the microbiome. Most of the genes shown to have an impact on the composition of the gut microbiome thus far are components of the immune system, and a few others have roles in metabolism (TABLE 1). As this field is still in its infancy, many of the studies discussed below describe the effect of either the presence or absence of a gene rather than of variation in that gene. For the full list of genes and their effects on the microbiota, see Supplementary information S1 (table).

Genes with roles in metabolism. A few host genes with roles in metabolism have been studied for their impact on the gut microbiota. One is the gene coding apolipoprotein AI; see TABLE 1 for the effects on the microbiota of variation in this gene. Another example is the leptin-encoding gene, *OB* (also known as *LEP*), which has been studied to a greater degree. Leptin is an adipose-derived hormone that has a key role in energy intake and expenditure, and its secretion is directly proportional to the amount of body fat. Some of the functions of leptin are to regulate appetite, energy expenditure and metabolism[57]. Leptin also acts as a cytokine with effects on immune cells, and reduced leptin levels have been shown to be associated with increased susceptibility to infection[58]. Polymorphisms in the leptin receptor gene (*OBR*; also known as *LEPR*) have been associated with obesity and type 2 diabetes[59,60]. In a recent GWA study of 1,504 women of European ancestry, SNPs at the *OBR* locus were observed to be significantly associated with the plasma levels of soluble leptin receptor, which is inversely associated with diabetes risk factors[61].

The effect of leptin on the gut microbiota has been studied through disruption of the corresponding gene in *ob*/*ob* mice[27] and of the receptor in Zucker rats[62]. In comparisons between obese, leptin-deficient (*ob*/*ob*) mice and their heterozygous ($Ob^{+/-}$) or wild-type ($Ob^{+/+}$) lean littermates, the maternal effect proved to be stronger at shaping β-diversity than either the effect of the genotype at the *Ob* locus or its corresponding associated phenotype. On the basis of the unweighted UniFrac distance metric, mouse caecal microbiota clustered according to their host litter, regardless of host *Ob* genotype. However, the microbiota of the obese hosts showed a dysbiosis: on average, the microbiota of obese mice had lower abundances of the Bacteroidetes than those of lean mice[27]. The dysbiosis had a functional consequence: the obesity-associated microbiome had an increased capacity for energy harvest from the host diet, resulting in greater adiposity in the host[63]. In the Zucker rat model, the loss of leptin receptor (also known as FA in rats) resulted in lower levels of total faecal bacteria, and a different species composition (for example, decreased levels of *Bifidobacterium* spp. and increased levels of *Halomonas* spp.) coupled to a different metabolite profile in urine and serum[62]. This study was further able to discriminate the microbiota of heterozygote rats ($Fa^{+/-}$) from those of wild-type rats ($Fa^{+/+}$), even though rats of both genotypes are lean and have similar metabolic profiles that are unaffected by their different microbiota.

What accounts for the dysbioses observed in these obese animals? Direct effects might include alterations in immune function. Another possibility that remains to be tested empirically is that host fat mass has an impact on the microbiota through the effects of leptin on mucus production. Administration of leptin to the gut has been shown to affect mucus levels[64,65]. Therefore, circulating leptin may indirectly modulate this key biophysical

ob/*ob* mouse
A mouse carrying two copies of a non-functional leptin gene (*Ob*; also known as *Lep*). These mice are leptin deficient and obese.

Zucker rat
A rat that is deficient in the leptin receptor (FA; also known as LEPR and OBR) and is obese.

Table 1 | **Host genes with effects on the composition of the microbiota**

Gene	Role of encoded protein	Disease*	Sampling site	Impact on the microbiota
MEFV‡	Proposed to help control inflammation by interacting with the cytoskeleton in certain white blood cells[56]	Familial Mediterranean fever	Faeces[55]	A study of mutations in *MEFV* revealed that distinct grouping of microbiomes is dependent on the allele carrier status of the host, and that patients with familial Mediterranean fever have a lower microbial richness[55]
APOA1	The main protein component of plasma high-density lipoprotein (HDL), which promotes cholesterol efflux from the liver for excretion; cholesterol is a component of bile acids, which restrict bacterial growth in the small intestine and, in turn, are processed by bacterial enzymatic activity into secondary bile acids that are re-absorbed[82]	Polymorphisms in the human *APOA1* gene have been associated with the risk of obesity and cardiovascular disease[98], and with hyperlipidaemia[99]	Faeces[35]	DGGE analysis indicates that the microbiota of APOAI-deficient mice has a different community structure from that of wild-type mice[35]
MYD88	An important signalling molecule that acts as one of the nodes in the information-gathering system of the innate immune system involved in sensing bacterial products; information from sensors of microbial products (for example, TLRs) is routed through MYD88 in inflammation response pathways[100]	Loss of MYD88 compromises the innate immune response to pathogens[100]	Caecum[22]	Analysis of the microbiota using clone libraries showed that three bacterial families (Lactobacillaceae, Rikenellaceae and Porphyromonadaceae) differ in abundance between MYD88-deficient mice and their wild-type counterparts[22]
			Caecum[19]	Clone libraries were used to compare the microbiota of wild-type and MYD88-deficient mice, and this analysis found no UniFrac-based clustering by genotype; furthermore, MYD88 is not required for a caecal bloom of the family Erysipelotrichaceae when mice are fed a high-fat diet[19]
NOD2	An intracellular pattern recognition receptor that recognizes muramyl dipeptide, a peptidoglycan constituent	Mutations in *NOD2* are among the strongest genetic risk factors for ileal Crohn's disease in humans[68,101]	Intestinal tissue[102]	In humans, a *NOD2* composite genotype is significantly associated with shifts in the composition of the intestinal microbial community (as assessed by UniFrac) for a subset of patients with Crohn's disease and ulcerative colitis[102]
			Faeces and the small intestine[103]	*Dorea* spp. and *Subdoligranulum* spp. abundances are significantly correlated with *Nod2* genotype, and NOD2-deficient mice have been shown to harbour a greater load of commensal bacteria§ belonging to the Firmicutes and the Bacteroidetes than wild-type mice (as quantified by RT-PCR)[103]
Defensin genes	Antimicrobial peptides that are produced by the host and are important for mucosal defence; they are secreted into the crypts of the intestine to eliminate bacteria that may otherwise induce inflammation and compromise the epithelial barrier	An association has been reported between the copy number of the gene encoding β-defensin and Crohn's disease[104]; both β-defensin and α-defensin expression levels are reduced in some patients with Crohn's disease[105,106], and Crohn's disease is associated with an altered microbiota[107]	Small intestine[108]	Mice deficient for matrilysin cannot cleave and activate α-defensin, and their microbiota have a lower percentage of bacteria from the Bacteroidetes than the microbiota of wild-type mice; transgenic mice expressing α-defensin 5 (HD-5) have a higher proportion of bacteria from the Firmicutes[108]
			Caecum and colon[109,110]	Defensins seem to reduce the levels of SFB, which recruit IL-17-producing T cells[109] and stimulate IgA production[110]
RELMB	A cytokine that is expressed in the gastrointestinal tract and has been implicated in innate immunity[111,112]; it has also been shown to regulate the expression of REGIIIγ, an antimicrobial peptide[113]; RELMβ-deficient mice remain relatively lean when fed freely on a high-fat diet that renders wild-type mice obese[114]	Unknown	Faeces[114]	The *Relmb*$^{-/-}$ genotype modestly but significantly affected the abundance of 15 lineages from the Bacteroidetes, one from the Proteobacteria and 15 from the Firmicutes, compared with their abundances in wild-type mice; much of the difference between the two genotypes was due to changes in the abundances of rare lineages[114]

Table 1 (cont.) | **Host genes with effects on the composition of the microbiota**

Gene	Role of encoded protein	Disease*	Sampling site	Impact on the microbiota
IgA locus	An antibody that has an important role in mucosal immunity	Humans lacking IgA have a higher incidence of inflammatory bowel diseases[115]	Small intestine[116]	Mice deficient in IgA harbour an increased abundance of SFB[116]; mice that are unable to export IgA shed more *S.* Typhimurium when infected than wild-type mice[117]
			Faeces[118]	Weight loss has been associated with a decrease in the proportion of IgA-coated bacteria in humans[118]
HLA genes	HLA genes of the major histocompatibility complex genomic region encode cell-surface antigen-presenting proteins; the class II HLAs present antigens to helper ($CD4^+$) T cells that direct the differentiation of antibody-producing B cells	In humans, the risk of Coeliac's disease is strongly linked with variation in HLA genes: class II HLA serotypes HLA-DQ2 and HLA-DQ8 have a major role in predisposing individuals to the development of Coeliac's disease, and inheritance of specific HLA-DQ genotypes explains 40% of the genetic predisposition in this disease[67]	Faeces[119,120]	Mouse studies have noted differences in the faecal microbiota using fingerprinting techniques, but the strength of the cohort effect was generally underestimated at that time and may have confounded the results[119,120]
			Faeces[121]	FISH-based comparison of the faecal microbiota of infants with low, medium and high risks of Coeliac's disease (based on HLA types) showed small but significant differences between the risk groups in the abundances of common gut bacteria such as the *Bacteroides* spp.–*Prevotella* spp. group[121]

APOAI, apolipoprotein AI; DGGE, denaturing-gradient gel electrophoresis; FISH, fluorescence *in situ* hybridization; HLA, human leukocyte antigen; IgA, immunoglobulin A; IL-17, interleukin-17; MYD88, myeloid differentiation primary response protein 88; RELMβ, resistin-like molecule-β; RT-PCR, real-time PCR ; SFB, segmented filamentous bacteria; *S.* Typhimurium, *Salmonella enterica* subsp. *enterica* serovar Typhimurium; TLR, Toll-like receptor; *Diseases or adverse phenotypes associated with loss or variants of the gene. ‡Encodes pyrin. §Bacteria that are associated with the host such that one party benefits and the other is not affected.

element of the gut habitat. Indirect effects might include host behavioural changes, such as alterations to food intake and eating patterns. Leptin-deficient or leptin-insensitive animals overeat, and hyperphagia alone may alter the composition of the gut microbial community. Microbiota of fasted mice have shown a significant alteration in the caecal Bacteroidetes/Firmicutes ratio compared with the ratio in mice that were able to eat freely: the levels of bacteria from the Bacteroidetes increased, whereas the levels of bacteria from the Firmicutes decreased in the fasting mice[66]. Thus, excess adiposity may reshape the microbiota in several non-exclusive direct and indirect ways — including the induction of low-grade inflammation, changes in food intake patterns or an altered physiochemical environment — helping to maintain a positive feedback loop between microbiota composition and host phenotype.

Genes with a role in innate and adaptive immunity. Genes that are implicated in innate or adaptive immunity have been shown in GWA studies to be associated with inflammatory diseases[67–69]. In a recent study of rheumatoid arthritis, genetic variation in genes that are integral to TLRs and nuclear factor-κB signalling pathways (notably, genes encoding TLR2, TLR4 and MYD88) contributed to the observed variation in the response to tumour necrosis factor (TNF)-blocking agents[70]. Interestingly, several of these genes have been shown (mostly by gene deletions in mice) to have a role in shaping the gut microbial community (FIG. 3b,c). Here, we focus on some of the more drastic changes in the microbiome that are associated with variation in these genes. We elaborate below on the effects of genes encoding TLRs and IFNs, and summarize in TABLE 1 and Supplementary information S1 (table) the effects of variation in genes encoding MYD88, NOD2, resistin-like molecule-β (RELMβ), immunoglobin A (IgA), and human leukocyte antigen (HLA) complexes.

TLR5 is expressed basolaterally in gut epithelial cells and recognizes flagellin, a highly conserved protein unit in the bacterial flagellum. In one study, a high proportion of one group of related mice lacking TLR5 developed metabolic syndrome, whereas another group, derived previously, tended to be colitic[20]. This discrepancy in symptoms between the two groups of otherwise genetically identical mice implicated the microbiota in disease. The loss of TLR5 resulted in an altered caecal microbiota in the mice with metabolic syndrome compared with that in wild-type mice from the same derivation group: the mice with metabolic syndrome displayed an altered community diversity (by unweighted UniFrac analysis) and an altered abundance of the dominant phylotypes (mostly uncultured OTUs from the phylum Firmicutes that are typical of the mouse microbiota), as well as a greater caecal bacterial load. This remodelled microbiota was sufficient to cause symptoms of metabolic syndrome when administered to germ-free wild-type recipients by gavage[20].

Studies of the effect of TLR2 and TLR4 on microbial diversity have yielded inconsistent results, probably owing to the different methodologies used to describe the microbiota. As mentioned above, when germ-free *Tlr2*$^{-/-}$*Tlr4*$^{-/-}$ mice and germ-free wild-type mice were fed with a faecal suspension from a single donor mouse, there was no detectable effect of the TLR deficiency on the faecal microbiota, as determined using DGGE analysis[46]. Similarly, terminal restriction fragment length polymorphism (T-RFLP) analysis has shown no effect of *Tlr2* and *Tlr4* on the diversity of the caecal microbiota in rats[71]. By contrast, when the caecal microbiota

Hyperphagia
Over-consumption of food.

of TLR2-deficient mice were characterized using clone libraries, it was found that the mutant mice harboured altered microbiota and increased levels of *Helicobacter* spp. compared with those of wild-type mice[72].

IFNs have roles in a range of immune functions, and mice that are deficient in IFN signalling pathways are highly susceptible to microbial infections. The potential for IFN signalling pathways to modulate the composition of the microbiota has been demonstrated recently in mice. In one study, two types of mice with different deficiencies in IFN signalling pathways were analysed: mice that lack signal transducer and activator of transcription 1 (STAT1), which is essential for the signalling of both type I IFNs (IFNα and IFNβ) and type II IFNs (IFNγ), and mice that lack IFN regulatory factor 9 (IRF9), which is primarily involved in type I IFN signalling[73]. DGGE analysis of faecal samples collected daily over several weeks revealed that mice deficient in IRF9 had greater intra-strain variability over time (that is, greater instability) in the composition of the gut microbiota, both daily and over 5-day intervals, than controls and *Stat1*-knockout mice[73]. A recent GWA study linked ulcerative colitis, but not Crohn's disease, to the risk locus *IFNG* in humans[74]; furthermore, the level of methylation at the *IFNG* locus was correlated with the immune response to microbial components and with the expression of IFNγ[74]. Instability may be a characteristic of the microbiota that compromises its beneficial nature; the concept that stability is a feature of the microbiome that may be partly under host genetic control is a fascinating idea that warrants further exploration.

Concluding remarks

Environmental factors, in a broad sense, and host genetics clearly interact to control the acquisition and to maintain the stability of a healthy gut microbiota. In turn, these three components — environment, host genetics and microbiome — interact to maintain homeostasis in the gut. The disruption of this stability by modifying one or more of the three interacting components may be a trigger for the development of diseases.

Although most of the studies presented above show an effect of the host genotype on the microbiome using faecal and caecal samples, bacteria inhabiting specific locations of the gut, such as mucosal surface-associated communities, might be even more influenced by the host genetics, as they are more 'tightly' associated with the host. Furthermore, different body habitats can harbour profoundly different microbiota. Shifting from faecal surveys to other location-specific surveys will help delineate the influence of host genetics on the microbiota, but this remains a huge challenge. The rapidly decreasing cost of sequencing will allow future studies to include whole-genome sequencing of the host, so that the effects of rare human genes are also taken into account, as well as information about the methylation status of the genome, so that epigenetic effects can be incorporated into disease models. A deeper characterization of the microbiome through metagenomic, metatranscriptomic and metabolomic approaches, as well as strain resequencing to characterize populations of specific taxa within individuals, will also add powerfully to the models.

Considering the gut microbiota as a host phenotypic trait in GWA studies is the next step towards a better understanding of complex diseases, such as Crohn's disease, ulcerative colitis, type 1 diabetes, rheumatoid arthritis and even obesity. The effects that have been documented so far in terms of host genotype influencing microbial diversity are due to very powerful gene effects that also have other large phenotypic effects. Smaller effects, and effects due to interactions between host genes (epistasis), remain to be discovered using the principles of quantitative genetics. To find the host genomic loci that are implicated in the abundances of individual microbial species or groups of microorganisms, GWA studies must include very large cohorts combined with either experimental conditions designed to overcome the variation in microbial community composition that is caused by factors other than the genetic variation of the host (such as diet, maternal effects, technical replicates and other factors), or at least ways to accurately account for this variation. When the influence of the host genome on its associated microbiome is clearly documented, multi-level models taking into account the environment, the host genome and the microbiome can be developed to better predict the outcome of perturbations to the gut ecosystem, such as diet change, the onset of a disease or the administration of antibiotics.

1. Koenig, J. E. *et al.* Microbes and Health Sackler Colloquium: succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* 28 Jul 2010 (doi:10.1073/pnas.1000081107).
A pilot study describing the colonization process of the infant gut microbiome from birth to 2.5 years of age; this study reveals that seemingly chaotic shifts in the microbiome are explained by events in the infant's life, such as diet change, antibiotic treatment and others.
2. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, e177 (2007).
3. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**, e280 (2008).
4. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
5. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
6. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
A large metagenomics-based characterization of the microbiome in multiple subjects finds a set of widely shared bacterial species.
7. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
A key paper using human twin data to measure the effect of host genotype on the microbiome and the link between dysbiosis and obesity.
8. Moodley, Y. *et al.* The peopling of the Pacific from a bacterial perspective. *Science* **323**, 527–530 (2009).
9. Schwarz, S. *et al.* Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog.* **4**, e1000180 (2008).
10. Vaishampayan, P. A. *et al.* Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biol. Evol.* **6**, 53–66 (2010).
11. Dethlefsen, L., Eckburg, P. B., Bik, E. M. & Relman, D. A. Assembly of the human intestinal microbiota. *Trends. Ecol. Evol.* **21**, 517–523 (2006).
12. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848 (2006).
13. Tims, S., Zoetendal, E. G., de Vos, W. M. & Kleerebezem, M. in: *Metagenomics of the Human Body* (ed. Nelson, K. E.) 15–41 (Springer, Berlin, 2011).
14. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
15. Sartor, R. B. Microbial influences in inflammatory bowel diseases. *Gastroenterology* **134**, 577–594 (2008).
16. Larsen, N. *et al.* Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **5**, e9085 (2010).
17. Peterson, D. A., Frank, D. N., Pace, N. R. & Gordon, J. I. Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe* **3**, 417–427 (2008).
18. Wang, Y. *et al.* 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J.* **3**, 944–954 (2009).

19. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
20. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
This article shows that TLR5 deficiency in mice results in an altered microbiota that is associated with metabolic syndrome. Transplanting this altered microbiome confers the disease state to germ-free recipient wild-type mice.
21. Garrett, W. S. *et al.* Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* **131**, 33–45 (2007).
22. Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**, 1109–1113 (2008).
23. Borody, T. J. *et al.* Bacteriotherapy using fecal flora: toying with human motions. *J. Clin. Gastroenterol.* **38**, 475–483 (2004).
24. You, D. M., Franzos, M. A. & Holman, R. P. Successful treatment of fulminant *Clostridium difficile* infection with fecal bacteriotherapy. *Ann. Intern. Med.* **148**, 632–633 (2008).
25. Grehan, M. J. *et al.* Durable alteration of the colonic microbiota by the administration of donor fecal flora. *J. Clin. Gastroenterol.* **44**, 551–561 (2010).
26. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex, polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107**, 18933–18938 (2010).
The first study taking a QTL-based approach to relate genomic variation in the host to microbial diversity in the gut. Testing of the CMM abundances for co-segregation with 530 SNPs identifies 18 host-associated QTLs that show a linkage with the abundances of specific microbial taxa.
27. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
28. Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213–251 (1972).
29. Orcutt, R. P., Gianni, F. J. & Judge, R. J. Development of an "Altered Schaedler Flora" for NCI gnotobiotic rodents. *Microecol. Ther.* **17**, 59 (1987).
30. Deloris Alexander, A. *et al.* Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. *Mamm. Genome* **17**, 1093–1104 (2006).
31. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
32. Louis, P., Scott, K. P., Duncan, S. H. & Flint, H. J. Understanding the effects of diet on bacterial metabolism in the large intestine. *J. Appl. Microbiol.* **102**, 1197–1208 (2007).
33. Tilg, H. Obesity, metabolic syndrome, and microbiota: multiple interactions. *J. Clin. Gastroenterol.* **44**, S16–S18 (2010).
34. 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).
35. Zhang, C. *et al.* Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *ISME J.* **4**, 232–241 (2010).
36. 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).
37. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
38. Stewart, J. A., Chadwick, V. S. & Murray, A. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J. Med. Microbiol.* **54**, 1239–1242 (2005).
39. Zoetendal, E. G., Akkermans, A. D., Akkermans-van Vliet, W. M., de Visser, J. A. & de Vos, W. M. The host genotype affects the bacterial community in the human gastronintestinal tract. *Microb. Ecol. Health Dis.* **13**, 129–134 (2001).
40. Turnbaugh, P. *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. Teucher, B. *et al.* Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Res. Hum. Genet.* **10**, 734–748 (2007).
42. Vinkhuyzen, A. A., van der Sluis, S., de Geus, E. J., Boomsma, D. I. & Posthuma, D. Genetic influences on 'environmental' factors. *Genes Brain Behav.* **9**, 276–287 (2010).
43. Matsumoto, M. *et al.* Voluntary running exercise alters microbiota composition and increases n-butyrate concentration in the rat cecum. *Biosci. Biotechnol. Biochem.* **72**, 572–576 (2008).
44. Santacruz, A. *et al.* Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity* **17**, 1906–1915 (2009).
45. Friswell, M. K. *et al.* Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *PLoS ONE* **5**, e8584 (2010).
46. Loh, G., Brodziak, F. & Blaut, M. The Toll-like receptors TLR2 and TLR4 do not affect the intestinal microbiota composition in mice. *Environ. Microbiol.* **10**, 709–715 (2008).
47. Kovacs, A. *et al.* Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microb. Ecol.* **6**, 423–428 (2011).
48. Darvasi, A. & Soller, M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207 (1995).
49. Wang, X. *et al.* Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. *Genome Res.* **13**, 1654–1664 (2003).
50. Kelly, S. A. *et al.* Genetic architecture of voluntary exercise in an advanced intercross line of mice. *Physiol. Genomics* **42**, 190–200 (2010).
51. Mohamadzadeh, M. *et al.* Lactobacilli activate human dendritic cells that skew T cells toward T helper 1 polarization. *Proc. Natl Acad. Sci. USA* **102**, 2880–2885 (2005).
52. Presley, L. L., Wei, B., Braun, J. & Borneman, J. Bacteria associated with immunoregulatory cells in mice. *Appl. Environ. Microbiol.* **76**, 936–941 (2010).
53. Devlin, B., Daniels, M. & Roeder, K. The heritability of IQ. *Genetics* **137**, 597–606 (1994).
54. Visscher, P. M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
55. Khachatryan, Z. A. *et al.* Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS ONE* **3**, e3064 (2008).
Genetic variation of the *MEFV* gene is associated with shifts in the bacterial populations within the phyla Bacteroidetes, Firmicutes and Proteobacteria.
56. Ting, J. P., Kastner, D. L. & Hoffman, H. M. CATERPILLERs, pyrin and hereditary immunological disorders. *Nature Rev. Immunol.* **6**, 183–195 (2006).
57. Kelesidis, T., Kelesidis, I., Chou, S. & Mantzoros, C. S. Narrative review: the role of leptin in human physiology: emerging clinical applications. *Ann. Intern. Med.* **152**, 93–100 (2010).
58. Fernandez-Riejos, P. *et al.* Role of leptin in the activation of immune cells. *Mediators Inflamm.* **2010**, 568343 (2010).
59. Jeon, J. P. *et al.* Copy number variation at leptin receptor gene locus associated with metabolic traits and the risk of type 2 diabetes mellitus. *BMC Genomics* **11**, 426 (2010).
60. Park, K. S. *et al.* Polymorphisms in the leptin receptor (LEPR)-putative association with obesity and T2DM. *J. Hum. Genet.* **51**, 85–91 (2006).
61. Sun, Q. *et al.* Genome-wide association study identifies polymorphisms in LEPR as determinants of plasma soluble leptin receptor levels. *Hum. Mol. Genet.* **19**, 1846–1855 (2010).
62. Waldram, A. *et al.* Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *J. Proteome Res.* **8**, 2361–2375 (2009).
63. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
64. El Homsi, M. *et al.* Leptin modulates the expression of secreted and membrane-associated mucins in colonic epithelial cells by targeting PKC, PI3K, and MAPK pathways. *Am. J. Physiol. Gastrointest. Liver Physiol.* **293**, G365–G373 (2007).
65. Plaisancie, P. *et al.* Luminal leptin activates mucin-secreting goblet cells in the large bowel. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290**, G805–G812 (2006).
66. Crawford, P. A. *et al.* Regulation of myocardial ketone body metabolism by the gut microbiota during nutrient deprivation. *Proc. Natl Acad. Sci. USA* **106**, 11276–11281 (2009).
67. Barbosa, T. & Rescigno, M. Host-bacteria interactions in the intestine: homeostasis to chronic inflammation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 80–97 (2010).
68. Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genet.* **40**, 955–962 (2008).
A GWA study that reveals many gene variants associated with Crohn's disease.
69. Imielinski, M. *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nature Genet.* **41**, 1335–1340 (2009).
70. Potter, C. *et al.* Association between anti-tumour necrosis factor treatment response and genetic variants within the TLR and NFκB signalling pathways. *Ann. Rheum. Dis.* **69**, 1315–1320 (2010).
71. Teran-Ventura, E. *et al.* Characterization of housing-related spontaneous variations of gut microbiota and expression of toll-like receptors 2 and 4 in rats. *Microb. Ecol.* **60**, 691–702 (2010).
72. Albert, E. J., Sommerfeld, K., Gophna, S., Marshall, J. S. & Gophna, U. The gut microbiota of toll-like receptor 2-deficient mice exhibits lineage-specific modifications. *Environ. Microbiol. Rep.* **1**, 65–70 (2009).
73. Thompson, C. L., Hofer, M. J., Campbell, I. L. & Holmes, A. J. Community dynamics in the mouse gut microbiota: a possible role for IRF9-regulated genes in community homeostasis. *PLoS ONE* **5**, e10335 (2010).
74. Gonsky, R. *et al.* Distinct IFNG methylation in a subset of ulcerative colitis patients based on reactivity to microbial antigens. *Inflamm. Bowel Dis.* **17**, 171–178 (2011).
75. Gootenberg, D. B. & Turnbaugh, P. J. Humanized animal models of the microbiome. *J. Anim. Sci.* 10 Sep 2010 (doi:10.2527/jas.2010-3371).
76. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
77. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).
78. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* **107**, 11971–11975 (2010).
79. Matsumiya, Y., Kato, N., Watanabe, K. & Kato, H. Molecular epidemiological study of vertical transmission of vaginal *Lactobacillus* species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction. *J. Infect. Chemother.* **8**, 43–49 (2002).
80. Tannock, G., Fuller, R., Smith, S. & Hall, M. Plasmid profiling of members of the family Enterobacteriaceae, lactobacilli, and bifidobacteria to study the transmission of bacteria from mother to infant. *J. Clin. Microbiol.* **28**, 1225–1228 (1990).
81. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Rev. Microbiol.* **6**, 776–788 (2008).
82. Jones, B. V., Begley, M., Hill, C., Gahan, C. G. & Marchesi, J. R. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc. Natl Acad. Sci. USA* **105**, 13580–13585 (2008).
83. Lozupone, C. A. *et al.* The convergence of carbohydrate active gene repertoires in human gut microbes. *Proc. Natl Acad. Sci. USA* **105**, 15076–15081 (2008).
84. Samuel, B. S. *et al.* Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc. Natl Acad. Sci. USA* **104**, 10643–10648 (2007).
85. Clarke, T. B. *et al.* Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. *Nature Med.* **16**, 228–231 (2010).
86. Hooper, L. V. Do symbiotic bacteria subvert host immunity. *Nature Rev. Microbiol.* **7**, 367–374 (2009).
87. Comstock, L. E. & Kasper, D. L. Bacterial glycans: key mediators of diverse host immune responses. *Cell* **126**, 847–850 (2006).
88. Peterson, D. A., McNulty, N. P., Guruge, J. L. & Gordon, J. I. IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe* **2**, 328–339 (2007).
89. Walter, J., Britton, R. A. & Roos, S. Microbes and Health Sackler Colloquium: Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proc. Natl Acad. Sci. USA* 25 Jun 2010 (doi: 10.1073/pnas.1000099107).
A study describing the evolutionary processes that shape the symbiotic interactions between microorganisms and the host gut, focusing on *Lactobacillus reuteri*.

90. Crowell-Davis, S. & Caudle, A. Coprophagy by foals: recognition of maternal feces. *Appl. Anim. Behav. Sci.* **24**, 267–272 (1989).
91. Osawa, R., Blanshard, W. H. & Ocallaghan, P. G. Microbiological studies of the intestinal microflora of the koala, *Phascolarctos cinereus*. II. Pap, a special maternal feces consumed by juvenile koalas. *Aust. J. Zool.* **41**, 611–620 (1993).
92. Lombardo, M. Access to mutualistic endosymbiotic microbes: an underappreciated benefit of group living. *Behav. Ecol. Sociobiol.* **62**, 479–497 (2008).
93. Bik, E. M. *et al.* Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J.* **4**, 962–974 (2010).
94. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
95. Kim, T. K. *et al.* Heterogeneity of vaginal microbial communities within individuals. *J. Clin. Microbiol.* **47**, 1181–1189 (2009).
96. Pei, Z. *et al.* Bacterial biota in the human distal esophagus. *Proc. Natl Acad. Sci. USA* **101**, 4250–4255 (2004).
97. Price, L. B. *et al.* The effects of circumcision on the penis microbiome. *PLoS ONE* **5**, e8422 (2010).
98. Chen, E. S. *et al. APOA1/A5* variants and haplotypes as a risk factor for obesity and better lipid profiles in a Brazilian elderly cohort. *Lipids* **45**, 511–517 (2010).
99. Liu, Z. K., Hu, M., Baum, L., Thomas, G. N. & Tomlinson, B. Associations of polymorphisms in the apolipoprotein A1/C3/A4/A5 gene cluster with familial combined hyperlipidaemia in Hong Kong Chinese. *Atherosclerosis* **208**, 427–432 (2010).
100. Asquith, M. J., Boulard, O., Powrie, F. & Maloy, K. J. Pathogenic and protective roles of MyD88 in leukocytes and epithelial cells in mouse models of inflammatory bowel disease. *Gastroenterology* **139**, 519–529.e2 (2010).
101. Hampe, J. *et al.* Association of NOD2 (CARD 15) genotype with clinical course of Crohn's disease: a cohort study. *Lancet* **359**, 1661–1665 (2002).
102. Frank, D. N. *et al.* Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **17**, 179–184 (2011).
103. Petnicki-Ocwieja, T. *et al.* Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc. Natl Acad. Sci. USA* **106**, 15813–15818 (2009).
104. Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.* **79**, 439–448 (2006).
105. Wehkamp, J. *et al.* Reduced Paneth cell α-defensins in ileal Crohn's disease. *Proc. Natl Acad. Sci. USA* **102**, 18129–18134 (2005).
106. Wehkamp, J. *et al.* The Paneth cell α-defensin deficiency of ileal Crohn's disease is linked to Wnt/Tcf-4. *J. Immunol.* **179**, 3109–3118 (2007).
107. Baker, P. I., Love, D. R. & Ferguson, L. R. Role of gut microbiota in Crohn's disease. *Expert Rev. Gastroenterol. Hepatol.* **3**, 535–546 (2009).
108. Salzman, N. H. *et al.* Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunol.* **11**, 76–83 (2010).
This article demonstrates that mice which are genetically engineered to express a human defensin in the intestine have an altered gut community composition, indicating that host-derived antimicrobial peptides can control microbial diversity in the gut.
109. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
110. Ivanov, I. I & Littman, D. R. Segmented filamentous bacteria take the stage. *Mucosal Immunol.* **3**, 209–212 (2010).
111. Artis, D. *et al.* RELMβ/FIZZ2 is a goblet cell-specific immune-effector molecule in the gastrointestinal tract. *Proc. Natl Acad. Sci. USA* **101**, 13596–13600 (2004).
112. Hogan, S. P. *et al.* Resistin-like molecule β regulates innate colonic function: barrier integrity and inflammation susceptibility. *J. Allergy Clin. Immunol.* **118**, 257–268 (2006).
113. Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126–1130 (2006).
114. Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716–1724.e2 (2009).
115. Yel, L. Selective IgA deficiency. *J. Clin. Immunol.* **30**, 10–16 (2010).
116. Suzuki, K. *et al.* Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc. Natl Acad. Sci. USA* **101**, 1981–1986 (2004).
117. Wijburg, O. L. *et al.* Innate secretory antibodies protect against natural *Salmonella typhimurium* infection. *J. Exp. Med.* **203**, 21–26 (2006).
118. Nadal, I. *et al.* Shifts in clostridia, bacteroides and immunoglobulin-coating fecal bacteria associated with weight loss in obese adolescents. *Int. J. Obes.* **33**, 758–767 (2009).
119. Toivanen, P., Vaahtovuo, J. & Eerola, E. Influence of major histocompatibility complex on bacterial composition of fecal flora. *Infect. Immun.* **69**, 2372–2377 (2001).
120. Vaahtovuo, J., Toivanen, P. & Eerola, E. Bacterial composition of murine fecal microflora is indigenous and genetically guided. *FEMS Microbiol. Ecol.* **44**, 131–136 (2003).
121. De Palma, G. *et al.* Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. *Curr. Issues Mol. Biol.* **12**, 1–10 (2010).

Acknowledgements
We thank A. Benson, D. Pomp, L. Angenent and the three reviewers for helpful comments on the manuscript, and J. Koenig for early discussions. We are grateful for a Beckman Young Investigator Award, and to The Hartwell Foundation, the David and Lucile Packard Foundation, the US National Science Foundation (IOS-0958,184), and the US National Institutes of Health Human Microbiome Project Data Analysis and Coordination Center (U01 HG004866) for support.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Ruth Ley's homepage:
http://microbe.calsnet.cornell.edu/leylab/

SUPPLEMENTARY INFORMATION
See online article: S1 **(table)**

ALL LINKS ARE ACTIVE IN THE ONLINE PDF