CrossMark

Protein & Cell

REVIEW

Of genes and microbes: solving the intricacies in host genomes

Jun Wang[1]✉ ⓘ, Liang Chen[1], Na Zhao[1], Xizhan Xu[1,2], Yakun Xu[1,2], Baoli Zhu[1,2,3]✉

[1] CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Science, Beijing 100101, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Collaborative Innovation Centre for Diagnosis and Treatment of Infectious Diseases First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310058, China
✉ Correspondence: junwang@im.ac.cn (J. Wang), zhubaoli@im.ac.cn (B. Zhu)

ABSTRACT

Microbiome research is a quickly developing field in biomedical research, and we have witnessed its potential in understanding the physiology, metabolism and immunology, its critical role in understanding the health and disease of the host, and its vast capacity in disease prediction, intervention and treatment. However, many of the fundamental questions still need to be addressed, including the shaping forces of microbial diversity between individuals and across time. Microbiome research falls into the classical nature vs. nurture scenario, such that host genetics shape part of the microbiome, while environmental influences change the original course of microbiome development. In this review, we focus on the nature, i.e., the genetic part of the equation, and summarize the recent efforts in understanding which parts of the genome, especially the human and mouse genome, play important roles in determining the composition and functions of microbial communities, primarily in the gut but also on the skin. We aim to present an overview of different approaches in studying the intricate relationships between host genetic variations and microbes, its underlying philosophy and methodology, and we aim to highlight a few key discoveries along this exploration, as well as current pitfalls. More evidence and results will surely appear in upcoming studies, and the accumulating knowledge will lead to a deeper understanding of what we could finally term a "hologenome", that is, the organized, closely interacting genome of the host and the microbiome.

INTRODUCTION

With between three- and ten-fold bacteria colonizing our own body (Sender et al., 2016), most of which are in the gastrointestinal (GI) tract (Qin et al., 2010; Zhu et al., 2010), it is hard to imagine that our genome does not devote a particular set of genes to dealing with all the potential threats, as well as coordinating benefits with our microbiome. Indeed, there are many indications of gene-microbiome cross-talk in humans, other animals (Kurilshikov et al., 2017) and even plants (Lundberg et al., 2012), with a large majority of those identified before the wide application of next-generation sequencing. Those genes function in the immune system (Hooper et al., 2012), with good reason: pathogens, an important part being bacteria, were one of the largest forces shaping the evolution of human genomes and thus the survival of our species and other species that rely on the immune system to defend against those pathogens (Kau et al., 2011).

In natural populations of animals and plants, the occurrence of epidemics constantly wipes out populations at the local (leading to disappearance of a species within an area) or global scale (leading to extinction). However, once there are survivors in those epidemics, there are usually genetic explanations in their genomes, such as natural variations in immune-related genes that lead to the higher resistance and survival of a particular group of individuals (Brinkworth and Pechenkina, 2013). In the next generations, those alleles (one variety of a gene) would usually increase in frequency and lead to changes in population genetics (Prugnolle et al.,

Protein & Cell

2005). There are a lot more pathogens that are not as dramatic as those involved in epidemics but that lead to less lethal infections and only the lower fitness of a few; however, these pathogens can still contribute to the change in allele frequencies (Barreiro and Quintana-Murci, 2010). Of course, pathogens are also involved in the arms race of the host and pathogens; no allele can be the perfect solution, but instead, different alleles are selected and enriched in different periods (Novembre and Han, 2012).

In humans, we know particularly well what the biggest threats were in our past and now, because of historical and medical records, and we see them in our genomes (Barreiro et al., 2008). Bubonic plague used to decimate one-third of European populations at a time, and its effects are visible in the current populations of Europeans, including some unexpected consequences that are summarized in the book the "*Survival of the Sickest*" (Moalem and Prince, 2008). This is still happening, although on a smaller scale nearly every year. At present, tuberculosis (TB) infects millions throughout the world, mainly in undeveloped areas (World Health Organization, 2016); these infections were more widespread in the past, before the invention of antibiotics. A study by Jostins et al. (2012) found surprising results demonstrating that the genes we think are causal in inflammatory bowel disease (IBD) (mainly composed of Crohn's disease and ulcerative colitis, and auto-immunity, which affect a small percentage of the population in Europe) turned out to be consequences of selection by TB. Those genes either promote our immune systems' attempt against TB by lowering the sensitivity to infection or blocking the recognition sites by bacteria, via as-yet-unknown mechanisms; these were consequently under selection by pathogens and are changing in frequency in the population (Jostins *et al.* 2012). Cholera, bacterial and meningitis are among the hundreds of recurring bacterial infections we are aware of, many of which have left a mark in our genome (Gupta, 2016) (Fig. 1).

However, it is not always about bad bacteria. Especially in the last decade, we have started to understand the composition and functions of complex microbial communities in the GI tract of humans and animals (Spor et al., 2011), as well as the skin (Grice and Segre, 2011) and oral microbiome (Dewhirst et al., 2010), reproductive (Ravel et al., 2011) and respiratory systems (Dickson and Huffnagle, 2015). Additionally, we have begun to appreciate the important functions of the normal microbiome in our own health (Fig. 1). We rely on our gut microbiome for digesting food and metabolizing large molecules to smaller ones, so our intestines can take them up more easily (Kau et al., 2011). They produce a large amount of other substances, including vitamins, serotonins, and many other functional molecules that modulate various systems in the host body (Kau et al., 2011; Kostic et al., 2014); thus, the concept of the gut-brain-axis (Foster and Neufeld, 2014), gut-liver-axis (Ray, 2017) and gut-lung-axis (Budden et al., 2017) have been proposed, examined and accepted by wider audiences. The microbiome stimulates the early maturation of the immune systems in infants while

maintaining the normal immune functions of adults; meanwhile, many of the immune-related diseases are primarily caused by dysbiosis in the microbiome (Kamada et al., 2013). The hologenome concept, endorsed by many in this field, can be understood to be the comprehensive inclusion of this whole interaction, cooperation and mutual selection at the genomic and metagenomic level, where the host and its microbiome compose a functional entity and the basis for natural selection and evolution (Zilber-Rosenberg and Rosenberg, 2008).

INDIRECT EVIDENCE

We aim to take the readers along the historical path of discovering the gene-microbiome cross-talk, although the studies we include here are not strictly chronological. For instance, we already knew a number of genes that were critical in maintaining host defence against pathogens (Major Histocompatibility Complex, MHC) (Neefjes et al., 2011), sensing microbes (Toll-Like Receptors, *TLR* for instance, which senses a wide range of molecules produced by microbes) (Kieser and Kagan, 2017), or were involved in other important process that could lead to disease. However, these are largely based on natural knock-out models, i.e., a mutation that leads to loss-of-function of a particular gene. We have studied mice or humans that are usually unhealthy, because critical genes in the host-microbe cross-talk are no longer functional and thus represent the extremities of the gene function spectrum. The more general observations of how variations in the whole genome, especially neutral or near-neutral alleles (those who do not carry as deleterious effect as the loss-of-function) and their association with effect on the microbiome have only come relatively recently (e.g., Hov et al., 2015, Wang et al., 2016, Bonder et al., 2016, Turpin et al., 2016, Goodrich et al., 2016).

The Ochman group published in *Plos Biology* a study on hominoids—primates, including humans, showing that microbiome divergences are well aligned (congruent) with the phylogeny of the mitochondria genome, a relatively simple yet powerful sub-genome for host phylogeny (Ochman et al., 2010). This work was followed by several other in-depth studies (Moeller et al., 2014; Nishida and Ochman, 2017). The microbiome divergence in this study was approximated with the Unifrac distance (Lozupone et al., 2011), which is also a phylogenetic measure of overall bacterial relationships, taking into account both the abundance of bacterial taxa, as well as their positions in a phylogenetic tree. Then, the overall microbiome differences are also clustered to form a "phylogeny" showing their relative similarities, and the congruence with the host phylogeny indicates that the microbiome differences could indeed be shaped by host genetic differences. However, it has to be noted that the evidence here is not without potential confounders, especially considering the natural drift of the microbiome together with its host during evolution and divergence, as well as the dietary differences of different
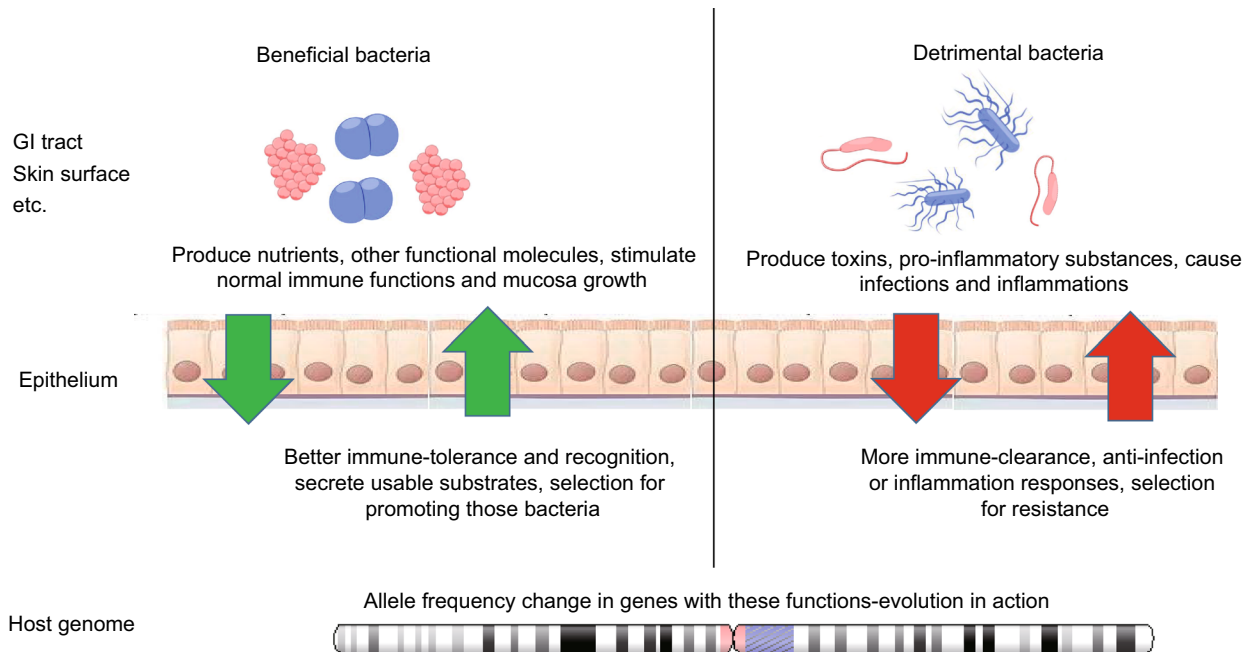
Protein & Cell

Figure 1. A simplified illustration of the host gene-microbiome interactions at the interface of various types of epithelia. The mucosal layer of the GI tract, airway, skin surface and reproductive tract surface are the primary interfaces of host-microbe interactions. Those microbes that we consider as beneficial usually produce nutrients, essential functional molecules and maintain the normal functions of the immune systems; thus the primary aim of host genes is to ensure their immune tolerance and facilitate their growth by secreting mucus, etc. as substrates. While harmful bacteria usually produce toxins, pro-inflammatory molecules and lead to infections, the host genes must clear them from the normal community and defend against inflammation and infections.

host species, geographical isolation, etc. (Bodenstein & Theis, 2015). This has also been noted in other comparative studies that try to distinguish signatures of genetic affinity from environmental similarities, especially diet (Ley et al., 2008). Nonetheless, it opens the door to understanding that in the whole genome, variations in the hosts underlie gut microbial variations, at least between different species.

The Ley group carried out another landmark study using the UK Twins biobank (Goodrich et al., 2014). The general idea is quite straightforward: by contrasting twins that are genetically identical (monozygotic twins) or non-identical, but related (dizygotic twins), one can quickly determine if some traits, in this case the microbiome, are genetically related. The assumption is that environmental differences are minimal in twins, or at least not extremely different between monozygotic twins and dizygotic twins. When a trait is more similar in the former than the latter, it must be due to genetic similarities. They indeed found this in the human gut microbiome in UK Twins, through a series of consequent studies using 16S rRNA and metagenomic analysis (Goodrich et al., 2016; Xie et al., 2016). A few particular bacteria also showed considerably high heritability, defined as the similarity of a trait due to the same genetic make-up, including one group, Christenalleaceae, that is inversely correlated to body-mass-index (BMI). Mouse models indeed show that this group of

bacteria has an effect of reducing obesity. However, it is rather disappointing that further analysis locating the genetic loci corresponding to this group of bacteria did not result in a definitive gene, which could be due to the small sample size of twins. This is because genome-wide-association studies (GWAS), as we are going to describe in detail, usually require relatively unrelated individuals, and in twins, the effective sample size is halved and would not reach one thousand. Org et al. (2015) performed similar analysis in 113 strains of different mice, where the microbiome is also more similar within the same strain than between different strains. They estimated the heritability of the microbiome, taking into account the relatedness of those mice strains as well as the pedigree, and concluded that host genetic variation can explain a substantial amount of variation in the gut microbiota.

DIRECT EVIDENCE: DESIGNED GENETIC STUDIES

Resolving confounders

Contrary to the genetic makeup of the hosts, which are (relatively speaking) stable, the microbiome tends to be a dynamic system that has its own natural fluctuations and is highly affected by a variety of environmental factors (Hall et al., 2017); thus, the microbiome observed at different time

points within a particular individual could be vastly different. Additionally, when looking at cross-sectional studies, as most of the large-scale investigations are due to the limitations set by studies of such scale, we are examining a snapshot of the microbiome within different individuals, with a high degree of randomness and noise (Walter and Ley, 2011). However, this is similar to a lot of fields in biology: we depend on biological signals that are substantial enough to be picked up by the appropriate detection methods, which, in this framework, includes statistical methods. Additionally, we depend on a sufficiently large sample size to distinguish statistically significant signals from the rest.

Nonetheless, accounting for the most important confounders is essential for any genetic study, for not doing so would lead to type I errors (false positives, where false genetic loci show up as significant) and type II errors (false negatives, where real genetic loci are covered by noise). In mouse models, we could control those to minimize them, while in humans, this would require a systematic investigation of confounders. A large collection of studies has reported anthropometric measures, including age (Yatsunenko et al., 2012), body mass index (BMI, Dominianni et al., 2015), waist measures and dietary habits (David et al., 2014; Dominianni et al., 2015), other life habits and so on (Yassour et al., 2016). In 2016, two studies appeared simultaneously in a special issue in *Science*, in Belgian Flanders (Flemish Gut Flora Project) (Falony et al., 2016) and Northern Netherlands (LifeLines-DEEP cohort) (Zhernakova et al., 2016), in which scientists carried out population-based analyses of confounding factors in shaping the diversity of the microbiome. In this study, hundreds of different measures were tested, filtered and ranked with their respective contributions to the overall dissimilarity of the gut microbiome (beta-diversity) and taxa abundances—the collective property of which is called alpha-diversity—as well as functional capacities. Many of the factors investigated were partially genetically determined, including gender, BMI, blood chemistry, etc., and thus already indicated a genetic involvement in shaping the microbiome. Other factors, such as age, are certainly not genetically determined, but are some of the top contributors and must be accounted for in studying genetics.

Now it might sound odd, that we would also need to control for genetic confounders while studying genetics. The rationales are as follows: in quantitative genetic studies using either crosses (quantitative trait loci, QTL) or natural populations (genome-wide-associations studies, GWAS), we are aware of the fact that the similarity of a trait could be due to overall relatedness. For instance, mice from the same breeding pair share largely the same growth environment and could have a shared microbiome from maternal transmissions (Benson et al., 2010; Wang et al., 2015). Related human individuals might also share a similar microbiome for the same reasons (Goodrich et al., 2016). Conversely, if the populations we study are not well-mixed, but subpopulations exist, thus providing a distinct population structure, the traits we find to be different between individuals might not be due to the effects of a few genes but rather longer term history of evolution, separation, drifts and so on (Yatsunenko et al., 2012). It is essential to account for kinship in QTL studies and GWAS analysis, and to thoroughly determine if there is distinctive population structure. Usually, all but one related individual are removed in a GWAS, and many try to keep the studied population as homogeneous as possible; however, there are also mathematical solutions that take kinship into account, or population structure via the genetic principle components (Kang et al., 2008; Price et al., 2010).

We quickly discuss the methods to account for confounders but will not go into much technical detail. When we investigate univariate traits, such as richness or taxa abundances, for most of the significant confounders, we use linear model/generalized linear models to remove their "effect" and keep the residues for the genetic analysis. This is relatively straightforward but sometimes cannot be well thought through, as many microbiome responses to a factor are non-linear (Lahti et al., 2014); however, other non-parametric factor do not necessarily perform better and can be misleading in its residues. For overall microbiome dissimilarities or beta-diversities, one can also remove the confounding effects of particular factors using constrained principle coordinates analysis (PcoA) and take its residue (also a distance matrix) (Ruhlemann et al., 2017). We rarely see it being performed, mainly because only a few have worked on the beta-diversity association with the host genome to date, and the field is still in its infancy.

Candidate gene approach

For historical, medical and political reasons, IBD continues to be the central focus of many microbiome investigations. It is a prevalent chronic inflammatory disorder of the GI tract in Europe, with occurrence approximately 1% and is particularly high in certain population of Jewish decedents (Hanauer, 2006). A continuous line of genetic studies have revealed a long list of potential genetic risk factors in IBD patients, including *NOD2*, *CARD9*, *ATG16L1*, *IRGM* and *FUT2*, among others (Xavier and Podolsky, 2007). Since there is a high proportion of microbiome factors in IBD disease, many of those risk genes have been tested to determine if they have impact on the microbiome (Kostic et al., 2014). Many IBD genetic risk factors are indeed are significantly associated with the decrease in the genus *Roseburia*, which plays an essential role in the conversion of acetate-to-butyrate compared to healthy controls, and this genus is known to be decreased in IBD patients (Morgan et al., 2012). We have summarized genes that were hypothesized to have impacts on the microbiome and were consequently tested in either humans (natural genetic variations) or mice (knockout models). As we can see, most studies are still focused on IBD. Of course, this list is by no means complete but contains the most prominent examples we are aware of (Table 1).

Protein & Cell

Table 1 Examples of candidate-gene approach studies in host gene-microbe interactions. We performed a literature search centred around the host gene, microbiome and diseases and have listed the most prominent examples where hypothesis-driven studies were carried either in humans (using natural variations) or mice (knock-out models) with respect to changes in the microbiome. We listed the changes observed, as well as the study context (type of disease), which we can see the primary focus on IBD

Gene name	Traits associated with genetic variations	Context of study	References
Human			
IL13/ CD14	Interaction with cesarean delivery and prenatal exposure to antibiotics to affect skin microbiome	Atopic dermatitis	Lee et al. (2014)
FUT2	Airway microbiome (*Pseudomonas aeruginosa*)	Bronchiectasis	Taylor et al. (2017)
IL6	*Helicobacter pylori*	Dyslipidemia	Pohjanen et al. (2016)
ATG16L1	*Fusobacteriaceae*, *Bacteroidaceae*, *Lachnospiraceae*, *Enterobacteriaceae*, *E. coli*	IBD	Sadaghian Sadabad et al. (2015)
CARD9	Gut microbiome composition	IBD	Lamas et al. (2016)
FUT2	Gut microbiome composition, diversity and structure	IBD	Rausch et al. (2011a, b)
NLRP12	Gut microbiome diversity	IBD	Chen et al. (2017a, b)
NOD2	Gut microbiome composition	IBD	de Bruyn et al. (2017)
SLC39A8	Gut microbiome composition	IBD	Li et al. (2016)
TNFSF15	Prevotella	IBD	Nakagome et al. (2017)
SI	Blautia, Oscillibacter, Ruminococcus and unclassified *Enterobacteriaceae*	IBS	Thingholm et al. (2018)
IFN-I	Microbials related to tryptophan metabolism	Multiple sclerosis	Rothhammer et al. (2016)
DEFB-CN	Nasopharyngeal bacterial colonization patterns	Otitis media	Jones et al. (2014)
A2ML1	Middle ear microbiome	Otitis media	Santos-Cortez et al. (2016)
C4B	Gut microbiome composition	Paediatric inflammatory bowel disease	Nissilä et al. (2017)
CARD15	Periodontal microbiota in Crohn's patients	Periodontitis	Stein et al. (2010)
ELANE	Subgingival microbiota	Periodontitis	Ye et al. (2011)
Mouse			
Myd88	Diversity, segmented filamentous bacteria	Anti-microbial signalling	Larsson et al. (2012)
Vdr	*Lactobacillus*, *Clostridium*, *Bacteroides*, *Alistipes*, *Odoribacter*, *Eggerthella*	Bile acid metabolism	Jin et al. (2015)
Tnf	Gut microbiome composition	Colitis	Kozik et al. (2017)
Can	*E. coli*	Colorectal cancer	Peuker et al. (2016)
Lcn2	*Alistipes*	Colorectal cancer	Moschen et al. (2016)
Ifnar1	Gut microbiome composition	IBD	Tschurtschenthaler et al. (2014)
Il10/Tlr4	Gut microbiome composition	IBD	Ward et al. (2016)
Il2	*E. coli* Nissle, *B. vulgatus* and *E. coli mpk/B. vulgatus*	IBD	Bohn et al. (2006)
Nlrp12	Gut microbiome composition	IBD	Chen et al. (2017a, b)
Sirt1	Gut microbiome composition	IBD, colorectal cancer	Lo Sasso et al. (2014)
Muc2	Gut microbiome composition	Ileal homeostasis	Sovran et al. (2015)
Mhc	Gut microbiome composition	Immunology	Kubinak et al. (2015a, b)
B4galnt2	Gut microbiome composition and Salmonella susceptibility	Inflammation	Rausch et al. (2015)

Protein & Cell

Table 1 continued

Gene name	Traits associated with genetic variations	Context of study	References
TREM-1	General dysbiosis in gut microbiome	Inflammation	Kökten et al. ([2018](#))
Nod2	Gut microbiome under high fat diet	Obesity	Rodriguez-Nunez et al. ([2017](#))
Fut2	Multi-generation dynamics of gut microbiome	Susceptibility to enteric infection	Rausch et al. ([2017](#))

Two of the genes are, interestingly, determinants of surface glycans, which serves as the initial contact point/molecule for host-microbe cross-talk. First, the gene *FUT2* encodes an enzyme fucosyltransferase-2 involved in the expression of ABO blood group antigens found on the GI mucosa and secretions. It is found to have two distinct genotypes, one functional secretor and one loss-of-function mutation leading to a non-secretor (McGovern et al., [2010](#)). Recent study revealed that the *FUT2* secretor status (defined by the genotype) has a significant influence on the gut microbiota (Rausch et al., [2011a](#), [b](#)); thus, the genus *Blautia* is lower in group A-secretors compared with non-A-secretors and this reduction is accompanied by higher abundances of members of Rikenellaceae, Peptostreptococcaceae, Clostridiales, and *Turicibacter* (Gampa et al., [2017](#)). Interestingly, the mouse gene B4galnt2 (encoding glycosyl-transferase β-1,4-N-acetylgalactosaminyltransferase 2) has similar function in terms of determining the sugar composition of the intestinal mucosa, and it is tissue-specific when we consider the expression patterns. Its expression in the intestine or not is strongly associated with altered bacterial community composition in the mouse model (Staubach et al., [2012](#)). B4galnt2 intestinal expression changes the gut microbiome and consequently facilitates epithelial invasion of *Salmonella typhimurium*, the underlying mechanism of which could be by increased intestinal inflammatory cytokines and infiltrating immune cells. Additionally, B4galnt2 has an interesting pattern of selection in the mouse population that we will discuss towards the end.

Another set of examples are genes responsible for sensing microbes and triggering down-stream cell signalling pathways. Those are usually components of the innate immune system. For example, exogenous microorganisms can be recognized by pattern recognition receptors (PRR), including but not limited to Toll-like receptors (TLR) and *NOD*-like receptors (Kieser and Kagan, [2017](#)), and the MyD88 protein encoded by the *MYD88* gene as an adaptor can modulate the signal transduction pathway. Those genes all have knockout mouse models, and the impact on the gut microbiome has been observed. In addition, we are aware that MyD88 signalling is critical for the development of type I diabetes (T1D), but the incidence of this disease can be decreased in mice by exposure to microbial stimuli, such as

injection with mycobacteria or various microbial products, suggesting that the cross-talk by specific genes is essential for the healthy development of immune systems in the early stages of life (Wen et al., [2008](#); Kostic et al., [2015](#)).

However, the most intriguing case so far is the MHC loci, wherein humans consistently fail to find significant associations with gut microbial compositions, either in candidate gene approaches or even in the recent GWAS (see later). The largest study so far was carried out in Norway using the bone marrow registry to distinguish candidates of different MHC alleles, and the collected microbiome did not reveal a significant difference (Hov et al., [2015](#)). However, it is another story in mouse models and the signals are much more prominent (Kubinak et al., [2015a](#), [b](#)). This highlights some of the difficulties in studying human genetics in terms of its impact on the microbiome, and the effects of certain genetic variations might be very small (and indeed confirmed in following studies) and may be masked by environmental differences. In mouse models, those factors are better controlled. Additionally, we admit that we do not have the comprehensive list of genes that have been studied using the candidate approach, and we merely touched the classic few (Table [1](#)). However, the principle holds and we do expect to see a larger collection of such studies, each with deeper insights into the mechanisms of gene-microbiome interactions.

Quantitative genetics

The tools of quantitative genetic studies come in handy when we intend to screen the associations between the microbiome and the whole host genome, instead of individual genes. Largely roots from plant and animal breeding science, quantitative genetics aims to find genes or genetic loci that are underlying important biological traits (phenotypes) of studied organisms, providing the basis for causal, mechanistic studies as well as practical applications (improve the production of crops or animals, for example) (McCarthy et al., [2008](#)). Two terms are widely used: QTL and GWAS (Fig. [2](#)). Many argue that, in the strict sense, they are mutually exclusive and that the former applies to quantitative traits such as height in animals or yields in crops, mainly using planned crosses as the study cohort and that
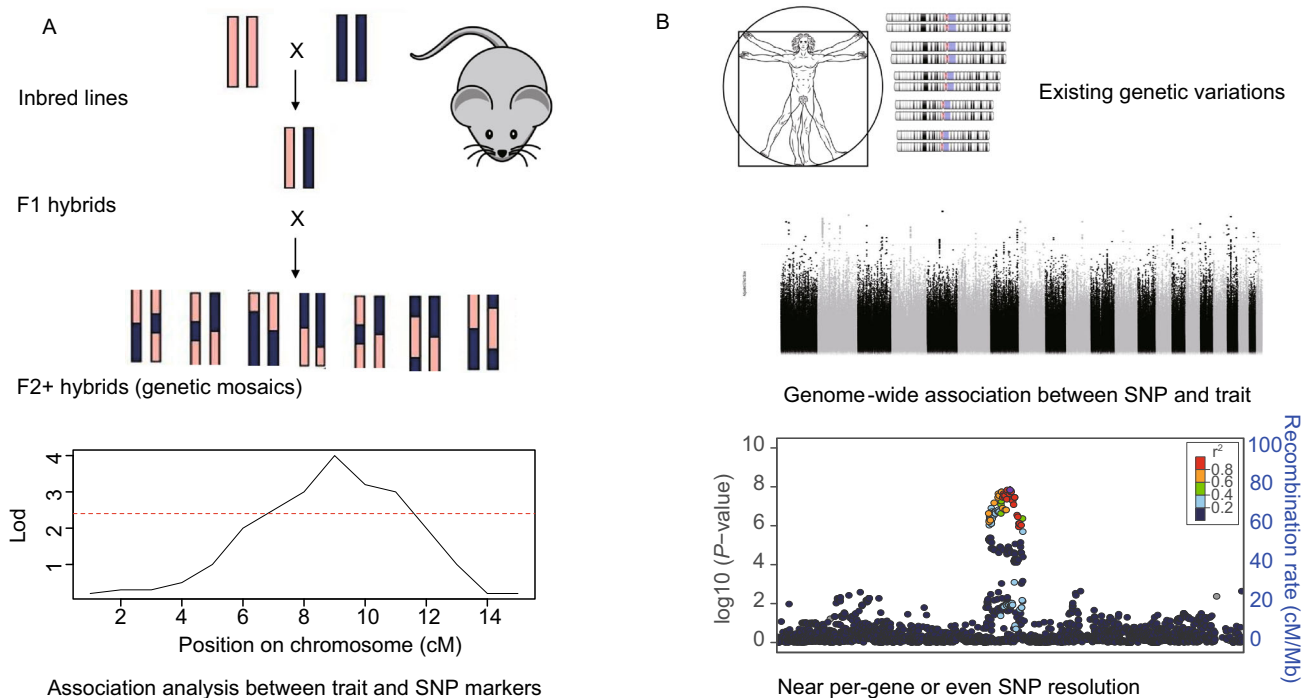
Figure 2. Schematic overview of quantitative trait loci (QTL, panel A) and genome-wide-association studies (GWAS, panel B). Both work on genetic variations, but result from different processes, either by intentional crosses (QTL) or extant (GWAS), and the linkage blocks are of a different size due to the number of recombinations. Association tests were carried out for SNPs and interpolated for a region in QTL analysis, while in GWAS it is done for each SNP and a "peak" in the Manhattan plot indicates a haplotype that might be significantly associated with the trait. In both cases, usually the *P*-values were transformed into −log(*P*) to indicate the significance level, and the genome-wide significance for QTL is usually determined by permutation tests. For GWAS, it is commonly accepted to be set at 1E−08 to 5E−08.

resolution is proportional to the generations of recombination, while the latter applies mostly to humans where the natural population history of a thousand generations has led to considerable recombination. However, it does not reach the per gene level. Instead, human genomes still have large blocks of genes that are linked, and little recombination has occurred yet. On the contrary, linkage disequilibrium (LD) also occurs as a result, which can lead to gene-level or even SNP-level resolutions for associations; in many cases, the studied trait is binary, especially about disease. However, we would like to note that mathematically, the two approaches are essentially the same. Quantitative genetic studies are about finding significant associations between genetic variations, either a single SNP or a large chromosome region. Both cases consider LD information, and variations in a defined phenotype and different types of traits only affect the model of the association tests. Binary traits require logistic regression, and the result of the associations are usually denoted as an odds ratio (OR): compared to the basal frequency of a trait, a particular SNP could change the frequency of that trait by OR fold. Thus, it is enriched, if OR is higher than 1, and vice versa. While continuously distributed

traits require linear or similar regressions and the "effect" of SNP/haplotype block are beta-values or z-scores, depending on the model used. This means that the mean value associated with a particular SNP/haplotype block deviates from the overall mean, measured by the variance (Hirschhorn and Daly, 2005).

To date, we are aware of six QTLs (Fig. 3), using crosses, that were carried out in mice as the model organism with the microbiome as the targeted trait. Four QTLs were done for the gut microbiome, while the remaining two focused on the skin microbiota. Benson et al. in 2010 published the first proof-of-concept study, showing that in a mouse cross of several generations, we can indeed locate the genetic variations at certain chromosomal regions to the variations of gut microbiome. Even when the resolution is not high, there are some interesting hints about the potential genes involved that correlate to specific microbiome abundances, in particular genes downstream of toll-like-receptor 2 (Tlr2), a gene that is mainly responsible for sensing gram-positive bacteria and downstream genes, including Irak3, Lyz1, Lyz2, IL-22, and IFN-gamma, while the correlated microbiome traits are indeed Coriobacteriaceae and *Lactococcus* (Benson et al.,

Belheouane et al.
Leamy et al.
Wang et al.
McKnite et al.
Srinivas et al.
Benson et al.
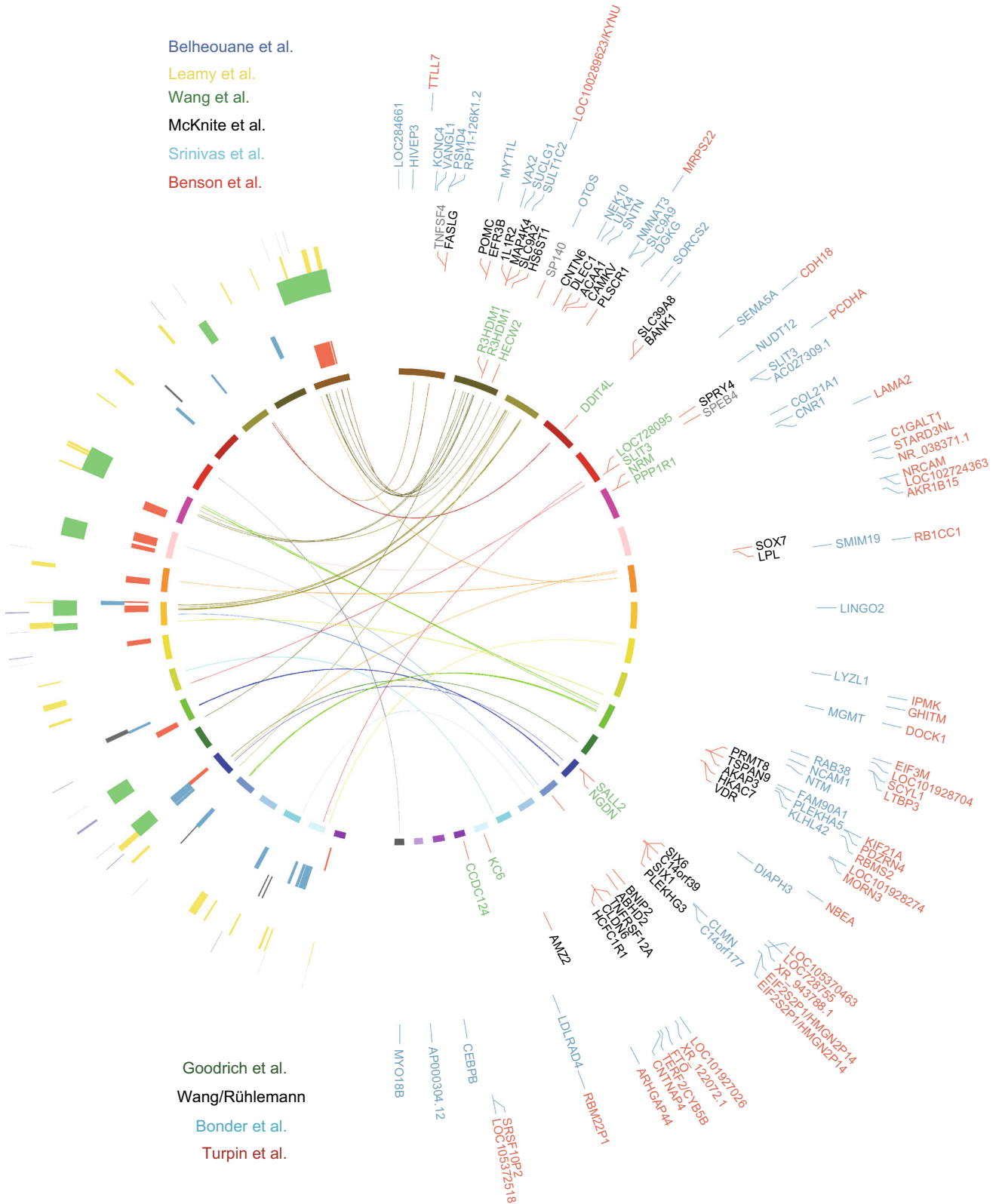
Goodrich et al.
Wang/Rühlemann
Bonder et al.
Turpin et al.

◀ **Figure 3. Overview of microbiome QTL in mouse and GWAS in humans to date.** Left half shows the six QTLs in mice, coloured by different studies and the confidence intervals are marked on the mouse chromosome. Please note that Belheouane and Srinivas studies are skin microbiome QTLs. The right half shows the genes implicated in human GWAS, including UK Twins, FoCus/PopGen (both original publication and later with modified methods), LifeLines-DEEP and GEM studies. Links in the middle show potential overlapping genes that showed up in human GWASs and fall into a confidence interval in mouse QTLs, which might be supportive of each other in terms of the association with microbiome variations.

2010). McKnite et al. (2012) and Leamy et al. (2014) followed up using different cross schemes and identified more regions with limited overlaps between the studies, each with some interesting insights about the genes that might be involved. Wang et al. (2015) published another interesting study using hybrid mice as the QTL cross cohort, where two subspecies of house mice are crossed to the second generation, and many regions are found to be correlated to microbial diversity. Lab mice are essentially *Mus musculus domesticus*, while its eastern European counterpart is *Mus musculus musculus*. They occur naturally in central Europe and have a well-studied evolution and speciation system. Currently, it seems that the microbiome is also affected by this hybridization. Moreover, the type of association is interesting. Half of the associations are transgressive, meaning that heterozygotes for a particular genetic locus have abnormally high or low values, out of the range for both homozygotes (Wang et al., 2015). This and one potential epistatic interaction that follows the Dobzhansky-Muller incompatibility model reveals further insights that the microbiome shape the genome evolution of hosts. Details can be found in the publications and in additional literature by the Bodenstein group (Brucker and Bordenstein, 2013; Bordenstein and Theis, 2015).

The other two skin microbiome QTLs, both published by the Baines group, provide interesting observations of host-microbiome homeostasis on the surface of the body (Srinivas et al., 2013; Belheouane et al., 2017). Working on an auto-immune skin disease model called epidermolysis bullosa acquisita (EBA), the group has extended the previous disease-oriented QTLs to include microbiome composition and found that the microbiome could indeed play an important role in determining disease manifestation. Even with roughly the same genetic makeup, developing the disease or not is correlated to the abundance of *Staphylococcus*, a potentially protective species. In contrast, when the bacterial abundance is taken into account in the statistical model, the power of QTL significantly increases, as the "noise", or environmental confounder of bacteria, is controlled (Srinivas et al., 2013). The second skin QTL has innovated the study approach and used 16S rRNA gene transcripts, which

examine the "active" part of the microbiome instead of the standing communities. Together with further cross generations (15th instead of 4th), this resulted in an almost per gene resolution and more significant associations when the transcripts are used. Additionally, some of the loci are involved in carcinogenesis of the skin, which are correlated to similar bacteria that could also lead to cancers in the colon (Belheouane et al., 2017).

We need to mention that the study by Org et al. (2015) discussed above actually carried out a GWAS in a similar fashion to that performed in humans, and several important genes were identified in this process that are associated with the microbial taxa. Contrary to QTLs above, they used standing variety of mouse strains (110 of them) instead of crosses that are specially set up, and the methods carefully considered the population structure. The only limitations were the relative small sample size and the low number of SNPs, for which we cannot really reach a similar genome-wide significance threshold in humans (will discuss below). This limits the resolution in the results.

The microbiome-oriented GWAS in humans, coincidentally also have six cases so far. We would consider at least two not to be sufficiently large to be considered equally as the remaining few. The first approximation of a microbiome GWAS was not really by design. Rather, Blekhman extracted human genome reads from HMP metagenomic raw data, called SNPs from those human reads for each subject, and correlated the genetic variations of the hosts to the microbiome variations. One particular association is between the lactase (*LCT*) gene and *Bifidobacterium*, and both are related to milk consumption and thus could understandably be correlated (Blekhman et al., 2015). However, whether the "fished out" human reads were sufficient to carry out proper SNPs was never clear, and neither was the reliability of the consequent analysis. Davenport et al. (2015) reported a small, but more conventionally designed GWAS study and managed to find some associations, none of which reached the commonly accepted genome-wide significance threshold (which is 5E−08 or 1E−08, the rationale is that when you screen millions of SNPs, the real significance should stand Bonferroni or Benjamini-Hochberg corrections for multiple testing, and thus it is commonly set at this scale). The Ley group also continued with their endeavours in the UK Twins cohort with multiple models for microbiome-SNP associations, and they did manage to actually find some hits that were later rediscovered, including *LCT* and *SLIT3*. However, because of the lack of a central focus on the models or functional studies, the study did not go into sufficient detail in exploring gene-microbiome ties at the genome scale (Goodrich et al., 2016) (Fig. 3).

The real breakthrough in human GWAS on the gut microbiome came as a trio in the November issue of *Nature Genetics* in 2016, where a German cohort (PopGen/FoCus) (Wang et al., 2016), a Dutch cohort (LifeLines-DEEP) (Bonder et al., 2016) and a Canadian cohort (GEM) (Turpin et al., 2016) simultaneously published large-scale GWASs

on the human gut microbiome (Fig. 3). All three cohorts include more than 1,000 unrelated individuals, all have replication cohorts, and all have considered the distribution properties of bacterial taxa (only a small fraction fits normal distributions, the rest are mainly zero-inflated). Among these three studies, the German and Canadian studies used a two-part hurdle model to address zero-inflations, while the Dutch study worked on the none-zero part when this was the case. The other difference is that the German/Canadian cohorts worked on a 16S rRNA based bacterial composition, while the Dutch study also has shot-gun metagenomic data and thus could map certain functional pathways. Beyond using bacterial abundance as the main studied trait, the German study, in particular, proposed a method to associate the overall microbial diversity (beta-diversity) to human genetic variations, and discovered 42 loci that passed the significance threshold, including one Vitamin D receptor (*VDR*) that was known to be involved in bile-acid sensing and homeostasis. Additionally, in this study, a number of functional studies, including bile acid analysis, metagenomic sequencing, cross-checking with different databases and comparing the human transcriptomes vs bacterial abundances ("coupling"), established the validity of *VDR* as a central part of the human-microbiome cross-talk, mostly via bile-acid metabolism and downstream pathways (Wang et al., 2016). The beta-diversity association method was consequently further developed to be less computationally intensive and more adapted to a higher dimensionality, with some further interesting loci discovered in this process (Ruhlemann et al., 2017). The Dutch study mainly confirmed the previous findings of *LCT-Bifidobacterium* associations and showed that environmental influence (in this case, milk intake) also interacts with the genotype of the individuals and shapes the microbial abundance (Bonder et al., 2016). Benson wrote a nice summary on all three of these studies, which was published in the same issue of Nature Genetics (Benson, 2016). Additionally, Kurilshikov and Zhernokova pieced together a wonderful review on this extended topic as well (Kurilshikov et al., 2017).

CONCLUSIONS

We have described the chronicles of genetic investigations in understanding host-microbe interactions, and the main results of the different approaches. We have seen indirect evidence in comparative studies, but those studies have limitations. We could investigate individual genes of interest and gain insights into their importance but could not generate a complete picture. Additionally, there is a quantitative genetic approach, and there are many things that we need to be cautious of. However, this endeavour is, by no means complete. We are just in the preliminary stages of investigation. Here, we would also note the current limitations of the mentioned studies, as well as our own perspectives into future efforts and directions.

Limitations

Our review is very focused human and mouse studies while ignoring the larger context of other model or non-model organisms. The reason is because of the great deal of effort put into the former two models and that the studies in humans and mice are considerably more relevant to our own health. We do know that a vast collection of literature exists for plant gene-microbe interactions, and many are textbook models, such as those genes involved in the invasion and colonization of *Agrobacterium*, which involves a complex interplay that would dwarf some of the host-microbe cross-talk in animals (Nester, 2015; Ellis, 2017). Similarly, a plant GWAS on the microbiome has been published for *Arabidopsis thaliana* and revealed a list of genes that may participate in a wider scale of interactions as well (Horton et al., 2014). However, since many genes lack counterparts in animals or at least do not carry out the same function, the value as a reference to other organisms is limited.

We do not have a shortage of host-microbe cross-talk examples in *C. elegans*, *Drosophila* and Zebra fish and many other common model organisms used in the lab. Most of these are single pathogens, and the difference observed in consequences are due to the genetic variation of the host. This again falls into the category of candidate gene-based approaches, in which one gene was the primary focus of study, and a glimpse into the greater picture of host-microbe cross-talks in those organisms has been observed. In regard to genome-wide, quantitative genetic studies in the microbiome, there have been two carried out in *Drosophila* (Chaston et al., 2015; Dobson et al., 2015), where the authors have pinpointed the interactions of nutrition and the host, and the microbiota serves as an important intermediator for the effects of nutrition to actually occur. Translated into terms that are widely used in human or mouse studies, the microbiome largely determines the metabolomes of the host and consequently the health status. Moreover, there are also a handful of studies, including one on chicken (Zhao et al., 2013), and we apologize if we missed other studies using different studying organisms. All of these studies make important contributions to the field, and by combining those studies, we generate a grander picture and get closer to solving the full puzzle. To achieve this, both the accumulation of data as well as innovation in methods are required.

Still, association does not mean causation, which is a limitation of association-based studies. Functional validation and establishment of real causation is still the bottleneck of many gene-microbe interactions. Moreover, compared to the limited knowledge we have on the host side, we know little to none about which bacterial genes are carrying out the cross-talk with the host. In pathogens, we study the key virulence factors that are part of the invasion process, or pathogenesis, including various toxins, different types of secretion systems, or genes responsible for producing the key metabolites influencing the hosts. We also know a few molecules that play a central role in being recognized by the

Protein & Cell

hosts, such as cell wall components, lipopolysaccharides (LPS). However, we lack a general picture of which part of the bacterial genomes are responsible for establishing and maintaining the connection with the host and which parts underlie the breakdown of such homeostasis. The authors assume that this varies from bacteria to bacteria of course and that environmental bacteria would need fewer genes for this task, while symbiotic bacteria should devote an essential part of the genome; otherwise, they would not be able to maintain a symbiotic relationship with the hosts. The gut microbiome, skin microbiome and bacteria in other body sites are intermediate in the sense that they are not strictly symbiotic but would still need to invest part of the genome. Some studies have shown that long-term intracellular symbionts in insects have lost a large part of their genome in the evolutionary process and only keep a small fraction of the essential genes (Wernegreen, 2002; Bennett et al., 2014). Whether this occurs in the gut or skin is not known, and the authors would argue that this genome reduction, if it exists, would apply to genes that are more maternally transmitted than those that are usually acquired from the environment.

Outlook

As we proclaimed in the beginning of this review, pathogens are driving forces in allele frequency changes in host populations, and we usually observe the results of this selection. However, this rarely occurs in real time, and we have not conducted an in-depth examination of the exact parameters of fitness and the costs. However, Vallier et al. (2017) carried out an astonishing study showing that, in natural populations of western house mice (*Mus musculus domesticus*), two alleles of the B4galnt2 gene co-exist as a result of long-term balancing selection, where one allele confers protection against various pathogens and thus could be favoured by pathogen-driven selection. However, it also leads to bleeding in the GI tract and could potentially reduce host fitness (this is similar to the human bleeding disorder called type 1 von Willebrand disease and could also be selected because it has beneficial effects during pathogen infections). Because this balancing selection is rather recent (from geographical distribution pattern combined with population history), the authors built up evolutionary models and estimated the fitness costs of the bleeding allele. It turns out that the currently observed allele frequency, as well as distribution, could only be explained when there is a heterozygote advantage and an advantage for homozygotes with bleeding alleles, and the costs in fitness of bleeding counts half of pathogenic infections. This is not biologically relevant proof, of course, as both fitness costs and infection costs are extremely difficult to quantify. However, it shows how important selection from microbes can be and how tiny microbes shift our genome, even leading to alleles that are otherwise detrimental to humans to maintain in the population. This is not the only case, as many of the underlying genes/alleles of autoimmune disorders and metabolic

syndromes are believed to be the result of selection by pathogens in the past and will continue to interact and change our genomes in the future (Nielsen et al., 2007; Novembre and Han, 2012; Milot and Pelletier, 2013).

Our review has so far been focused on individual genes, and we could only limit it to the main proof-of-concept studies. An important concept in understanding the host-microbe cross-talk, similar to in any genetic study, is the gene-environment interactions (G by E), where the genetic background manifests different effects when the environmental context changes. This has been shown to be the case in the *LCT* gene and *Bifidobacterium* (Blekhman et al., 2015; Goodrich et al., 2016), where dairy intake serves the environmental background (Bonder et al., 2016). However, we do not have many other examples, since the content of environmental influences is so vast, and many studies have not managed to include a sufficient amount. At the same time, the sample sizes usually do not permit this kind of analysis either. In addition, there is an urgent need to move beyond single gene associations, since for most of the complex traits, the power of the single gene in explaining microbiome diversities as well as functionalities is limited, and conclusions can only be partial and misleading. Integrating multiple genetic variations with respective weight, which results in polygenic scores as used in many diseases (Dudbridge, 2013), could be applied in microbiome research to explain the underlying genetic architecture for a single taxon or the overall diversity and would yield a more complete overview of host-microbe cross-talk. Another important direction is to move beyond single genes to biological pathways, which participate, and examine the association between microbes and certain cellular processes/signalling pathways. This requires enrichment analysis from a collection of single genes (Ramanan et al., 2012). Overall, this fascinating area of research has just revealed its potential in terms of understanding both fundamental biology, as well as application in medicine and human health, with many aspects that have yet to be examined.

ABBREVIATIONS

BMI, body-mass-index; EBA, epidermolysis bullosa acquisita; G by E, gene-environment interactions; GI, gastro-intestinal; GWAS, genome-wide-association studies; IBD, inflammatory bowel disease; LCT, lactase; LD, linkage disequilibrium; LPS, lipopolysaccharides; MHC, major histocompatibility complex; OR, odds ratio; PCoA, principle coordinates analysis; PRR, pattern recognition receptor; QTL, quantitative trait loci; SNP, single nucleotide polymorphisms;

Protein & Cell

T1D, type I diabetes; TB, tuberculosis; TLR, Toll-like receptor; VDR, Vitamin D receptor.

COMPLIANCE WITH ETHICS GUIDELINES

Jun Wang, Liang Chen, Na Zhao, Xizhan Xu, Yakun Xu and Baoli Zhu declare that they have no conflict of interest.

OPEN ACCESS

REFERENCES

Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17–30

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. Nat Genet 40:340–345

Belheouane M, Gupta Y, Kunzel S, Ibrahim S, Baines JF (2017) Improved detection of gene-microbe interactions in the mouse skin microbiota using high-resolution QTL mapping of 16S rRNA transcripts. Microbiome 5:59

Bennett GM, McCutcheon JP, MacDonald BR, Romanovicz D, Moran NA (2014) Differential genome evolution between companion symbionts in an insect-bacterial symbiosis. MBio 5: e01697-14

Benson AK (2016) The gut microbiome-an emerging complex trait. Nat Genet 48:1301–1302

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

Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, Spector TD, Keinan A, Ley RE, Gevers D et al (2015) Host genetic variation impacts microbiome composition across human body sites. Genome Biol 16:191

Bohn E, Bechtold O, Zahir N, Frick JS, Reimann J, Jilge B, Autenrieth IB (2006) Host gene expression in the colon of gnotobiotic interleukin-2-deficient mice colonized with commensal colitogenic or noncolitogenic bacterial strains: common patterns and bacteria strain specific signatures. Inflamm Bowel Dis 12(9):853–862

Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, Deelen P, Vatanen T, Schirmer M, Smeekens SP et al (2016) The effect of host genetics on the gut microbiome. Nat Genet 48:1407–1412

Bordenstein SR, Theis KR (2015) Host biology in light of the microbiome: ten principles of holobionts and hologenomes. Plos Biol 13:e1002226

Brinkworth JF, Pechenkina K (2013) Primates, pathogens, and evolution. Springer, New York

Brucker RM, Bordenstein SR (2013) The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia. Science 341:667–669

Budden KF, Gellatly SL, Wood DL, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM (2017) Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol 15:55–63

Chaston JM, Dobson AJ, Newell PD, Douglas AE (2015) Host genetic control of the microbiota mediates the drosophila nutritional phenotype. Appl Environ Microbiol 82:671–679

Chen L, Wilson JE, Koenigsknecht MJ, Chou WC, Montgomery SA, Truax AD, Brickey WJ, Packey CD, Maharshak N, Matsushima GK, Plevy SE, Young VB, Sartor RB, Ting JP (2017a) NLRP12 attenuates colon inflammation by maintaining colonic microbial diversity and promoting protective commensal bacterial growth. Nat Immunol 18(5):541–551. https://doi.org/10.1038/ni.3690

Chen L, Wilson JE, Koenigsknecht MJ, Chou WC, Montgomery SA, Truax AD, Brickey WJ, Packey CD, Maharshak N, Matsushima GK, Plevy SE, Young VB, Sartor RB, Ting JP (2017b) NLRP12 attenuates colon inflammation by maintaining colonic microbial diversity and promoting protective commensal bacterial growth. Nat Immunol 18(5):541–551. https://doi.org/10.1038/ni.3690

Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y (2015) Genome-wide association studies of the human gut microbiota. PLoS One 10:e0140301

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA et al (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563

de Bruyn M, Vermeire S (2017) NOD2 and bacterial recognition as therapeutic targets for Crohn's disease. Expert Opin Ther Targets 21(12):1123–1139. https://doi.org/10.1080/14728222.2017.1397627

Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG (2010) The human oral microbiome. J Bacteriol 192:5002–5017

Dickson RP, Huffnagle GB (2015) The lung microbiome: new principles for respiratory bacteriology in health and disease. PLoS Pathog 11:e1004923

Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, Westmiller S, Wong AC, Clark AG, Lazzaro BP et al (2015) Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of Drosophila melanogaster. Nat Commun 6:6312

Dominianni C, Sinha R, Goedert JJ, Pei Z, Yang L, Hayes RB, Ahn J (2015) Sex, body mass index, and dietary fiber intake influence the human gut microbiome. PLoS One 10:e0124599

Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. Plos Genet 9:e1003348

Ellis JG (2017) Can plant microbiome studies lead to effective biocontrol of plant diseases? Mol Plant-Microbe Interact 30:190–193

Protein & Cell

Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D et al (2016) Population-level analysis of gut microbiome variation. Science 352:560–564

Foster J, Neufeld KA (2014) Gut-brain axis: How the microbiome influences anxiety and depression. Int J Neuropsychopharmacol 17:27

Gampa A, Engen PA, Shobar R, Mutlu EA (2017) Relationships between gastrointestinal microbiota and blood group antigens. Physiol Genom 49:473–483

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van Treuren W, Knight R, Bell JT et al (2014) Human genetics shape the gut microbiome. Cell 159:789–799

Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE (2016) Genetic determinants of the gut microbiome in UK twins. Cell Host Microbe 19:731–743

Grice EA, Segre JA (2011) The skin microbiome. Nat Rev Microbiol 9:244–253

Gupta S (2016) Infectious disease: something in the water. Nature 533:S114–S115

Hall AB, Tolonen AC, Xavier RJ (2017) Human genetic variation and the gut microbiome in disease. Nat Rev Genet 18:690

Hanauer SB (2006) Inflammatory bowel disease: epidemiology, pathogenesis, and therapeutic opportunities. Inflamm Bowel Dis 12(Suppl 1):S3–S9

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

Hooper LV, Littman DR, Macpherson AJ (2012) Interactions between the microbiota and the immune system. Science 336:1268–1273

Horton MW, Bodenhausen N, Beilsmith K, Meng DZ, Muegge BD, Subramanian S, Vetter MM, Vilhjalmsson BJ, Nordborg M, Gordon JI et al (2014) Genome-wide association study of Arabidopsis thaliana leaf microbial community. Nat Commun 5:5320

Hov JR, Zhong HZ, Qin BC, Anmarkrud JA, Holm K, Franke A, Lie BA, Karlsen TH (2015) The influence of the autoimmunity-associated ancestral HLA haplotype AH8.1 on the human gut microbiota: a cross-sectional study. Plos One 10:e0133804

Jin D, Wu S, Zhang YG, Lu R, Xia Y, Dong H, Sun J (2015) Lack of vitamin D receptor causes dysbiosis and changes the functions of the murine intestinal microbiome. Clin Ther 37(5):996.e7–1009. e7. https://doi.org/10.1016/j.clinthera.2015.04.004

Jones EA, Kananurak A, Bevins CL, Hollox EJ, Bakaletz LO (2014) Copy number variation of the beta defensin gene cluster on chromosome 8p influences the bacterial microbiota within the nasopharynx of otitis-prone children. PLoS One. 9(5):e98269. https://doi.org/10.1371/journal.pone.0098269

Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA et al (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491:119–124

Kamada N, Seo SU, Chen GY, Nunez G (2013) Role of the gut microbiota in immunity and inflammatory disease. Nat Rev Immunol 13:321–335

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI (2011) Human nutrition, the gut microbiome and the immune system. Nature 474:327–336

Kieser KJ, Kagan JC (2017) Multi-receptor detection of individual bacterial products by the innate immune system. Nat Rev Immunol 17:376–390

Kökten T, Gibot S, Lepage P, D'Alessio S, Hablot J, Ndiaye NC, Busby-Venner H, Monot C, Garnier B, Moulin D, Jouzeau JY, Hansmannel F, Danese S, Guéant JL, Muller S, Peyrin-Biroulet L (2018) TREM-1 inhibition restores impaired autophagy activity and reduces colitis in mice. J Crohns Colitis 12(2):230–244. https://doi.org/10.1093/ecco-jcc/jjx129

Kostic AD, Xavier RJ, Gevers D (2014) The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146:1489–1499

Kostic AD, Gevers D, Siljander H, Vatanen T, Hyotylainen T, Hamalainen AM, Peet A, Tillmann V, Poho P, Mattila I et al (2015) The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe 17:260–273

Kozik AJ, Nakatsu CH, Chun H, Jones-Hall YL (2017) Age, sex, and TNF associated differences in the gut microbiota of mice and their impact on acute TNBS colitis. Exp Mol Pathol 103(3):311–319. https://doi.org/10.1016/j.yexmp.2017.11.014

Kubinak JL, Stephens WZ, Soto R, Petersen C, Chiaro T, Gogokhia L, Bell R, Ajami NJ, Petrosino JF, Morrison L, Potts WK, Jensen PE, O'Connell RM, Round JL (2015a) MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. Nat Commun. 23(6):8642. https://doi.org/10.1038/ncomms9642

Kubinak JL, Stephens WZ, Soto R, Petersen C, Chiaro T, Gogokhia L, Bell R, Ajami NJ, Petrosino JF, Morrison L et al (2015b) MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. Nat Commun 6:8642

Kurilshikov A, Wijmenga C, Fu J, Zhernakova A (2017) Host genetics and gut microbiome: challenges and perspectives. Trends Immunol 38:633–647

Lahti L, Salojarvi J, Salonen A, Scheffer M, de Vos WM (2014) Tipping elements in the human intestinal ecosystem. Nat Commun. https://doi.org/10.1038/ncomms5344

Lamas B, Richard ML, Leducq V, Pham HP, Michel ML, Da Costa G, Bridonneau C, Jegou S, Hoffmann TW, Natividad JM, Brot L, Taleb S, Couturier-Maillard A, Nion-Larmurier I, Merabtene F, Seksik P, Bourrier A, Cosnes J, Ryffel B, Beaugerie L, Launay JM, Langella P, Xavier RJ, Sokol H (2016) CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat Med. 22(6):598–605. https://doi.org/10.1038/nm.4102

Larsson E, Tremaroli V, Lee YS, Koren O, Nookaew I, Fricker A, Nielsen J, Ley RE, Bäckhed F (2012) Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. Gut 61 (8):1124–1131. https://doi.org/10.1136/gutjnl-2011-301104

Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K, Sinha R, Peterson DA, Walter J, Benson AK et al (2014) Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol 15:552

Lee SY, Yu J, Ahn KM, Kim KW, Shin YH, Lee KS, Hong SA, Jung YH, Lee E, Yang SI, Seo JH, Kwon JW, Kim BJ, Kim HB, Kim WK, Song DJ, Jang GC, Shim JY, Lee SY, Kwon JY, Choi SJ, Lee KJ, Park HJ, Won HS, Yoo HS, Kang MJ, Kim HY, Hong SJ (2014) Additive effect between IL-13 polymorphism and cesarean section delivery/prenatal antibiotics use on atopic dermatitis: a birth cohort study (COCOA). PLoS One. 9(5):e96603. https://doi.org/10.1371/journal.pone.0096603

Ley R, Lozupone CA, Hamady M, Knight R, Gordon JI (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6(10):776–788. https://doi.org/10.1038/nrmicro1978

Li D, Achkar JP, Haritunians T, Jacobs JP, Hui KY, D'Amato M, Brand S, Radford-Smith G, Halfvarson J, Niess JH, Kugathasan S, Büning C, Schumm LP, Klei L, Ananthakrishnan A, Aumais G, Baidoo L, Dubinsky M, Fiocchi C, Glas J, Milgrom R, Proctor DD, Regueiro M, Simms LA, Stempak JM, Targan SR, Törkvist L, Sharma Y, Devlin B, Borneman J, Hakonarson H, Xavier RJ, Daly M, Brant SR, Rioux JD, Silverberg MS, Cho JH, Braun J, McGovern DP, Duerr RH (2016) A pleiotropic missense variant in SLC39A8 is associated with Crohn's Disease and human gut microbiome composition. Gastroenterology 151(4):724–732. https://doi.org/10.1053/j.gastro.2016.06.051

Lo Sasso G, Ryu D, Mouchiroud L, Fernando SC, Anderson CL, Katsyuba E, Piersigilli A, Hottiger MO, Schoonjans K, Auwerx J (2014) Loss of Sirt1 function improves intestinal anti-bacterial defense and protects from colitis-induced colorectal cancer. PLoS One. 9(7):e102495. https://doi.org/10.1371/journal.pone.0102495

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

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Del Rio TG et al (2012) Defining the core Arabidopsis thaliana root microbiome. Nature 488:86–90

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and chal- lenges. Nat Rev Genet 9:356–369

McGovern DPB, Jones MR, Taylor KD, Marciante K, Yan XF, Dubinsky M, Ippoliti A, Vasiliauskas E, Berel D, Derkowski C et al (2010) Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum Mol Genet 19:3468–3476

McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, Bastiaansen JWM, Wang XS, Kachman SD, Auwerx J et al (2012) Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS ONE 7:e39191

Milot E, Pelletier F (2013) Human evolution: new playgrounds for natural selection. Curr Biol 23:R446–R448

Moalem S, Prince J (2008) Survival of the sickest: the surprising connections between disease and longevity. Harper, London

Moeller AH, Li Y, Mpoudi Ngole E, Ahuka-Mundeke S, Lonsdorf EV, Pusey AE, Peeters M, Hahn BH, Ochman H (2014) Rapid changes in the gut microbiome during human evolution. Proc Natl Acad Sci USA 111:16431–16435

Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB et al (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79

Moschen AR, Gerner RR, Wang J, Klepsch V, Adolph TE, Reider SJ, Hackl H, Pfister A, Schilling J, Moser PL, Kempster SL, Swidsinski A, Orth Höller D, Weiss G, Baines JF, Kaser A, Tilg H (2016) Lipocalin 2 protects from inflammation and tumorige- nesis associated with gut microbiota alterations. Cell Host Microbe. 19(4):455–469. https://doi.org/10.1016/j.chom.2016.03.007

Nakagome S, Chinen H, Iraha A, Hokama A, Takeyama Y, Sakisaka S, Matsui T, Kidd JR, Kidd KK, Said HS, Suda W, Morita H, Hattori M, Hanihara T, Kimura R, Ishida H, Fujita J, Kinjo F, Mano S, Oota H (2017) Confounding effects of microbiome on the susceptibility of TNFSF15 to Crohn's disease in the Ryukyu Islands. Hum Genet. 136(4):387–397. https://doi.org/10.1007/s00439-017-1764-0

Neefjes J, Jongsma ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11:823–836

Nester EW (2015) Agrobacterium: nature's genetic engineer. Front Plant Sci 5:730

Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG (2007) Recent and ongoing selection in the human genome. Nat Rev Genet 8:857–868

Nishida AH, Ochman H (2017) Rates of gut microbiome divergence in mammals. Mol Ecol

Nissilä E, Korpela K, Lokki AI, Paakkanen R, Jokiranta S, de Vos WM, Lokki ML, Kolho KL, Meri S (2017) C4B gene influences intestinal microbiota through complement activation in patients with paediatric-onset inflammatory bowel disease. Clin Exp Immunol. 190(3):394–405. https://doi.org/10.1111/cei.13040

Novembre J, Han EJ (2012) Human population structure and the adaptive response to pathogen-induced selection pressures. Philos Trans R Soc B 367:878–886

Ochman H, Worobey M, Kuo CH, Ndjango JB, Peeters M, Hahn BH, Hugenholtz P (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8: e1000546

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

Peuker K, Muff S, Wang J, Künzel S, Bosse E, Zeissig Y, Luzzi G, Basic M, Strigli A, Ulbricht A, Kaser A, Arlt A, Chavakis T, van den Brink GR, Schafmayer C, Egberts JH, Becker T, Bianchi ME, Bleich A, Röcken C, Hampe J, Schreiber S, Baines JF, Blumberg RS, Zeissig S (2016) Epithelial calcineurin controls microbiota- dependent intestinal tumor development. Nat Med. 22(5):506–515. https://doi.org/10.1038/nm.4072

Pohjanen VM, Koivurova OP, Niemelä SE, Karttunen RA, Karttunen TJ (2016) Role of Helicobacter pylori and interleukin 6–174 gene

polymorphism in dyslipidemia: a case-control study. BMJ Open. 6 (1):e009987. https://doi.org/10.1136/bmjopen-2015-009987

Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11:459–463

Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F (2005) Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol 15:1022–1027

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T et al (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59–65

Ramanan VK, Shen L, Moore JH, Saykin AJ (2012) Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet 28:323–332

Rausch P, Rehman A, Künzel S, Häsler R, Ott SJ, Schreiber S, Rosenstiel P, Franke A, Baines JF (2011a) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci USA 108(47):19030–19035. https://doi.org/10.1073/pnas.1106408108

Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, Rosenstiel P, Franke A, Baines JF (2011b) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. Proc Natl Acad Sci USA 108:19030–19035

Rausch P, Steck N, Suwandi A, Seidel JA, Künzel S, Bhullar K, Basic M, Bleich A, Johnsen JM, Vallance BA, Baines JF, Grassl GA (2015) Expression of the blood-group-related gene B4galnt2 alters susceptibility to salmonella infection. PLoS Pathog 11(7): e1005008. https://doi.org/10.1371/journal.ppat.1005008

Rausch P, Künzel S, Suwandi A, Grassl GA, Rosenstiel P, Baines JF (2017) Multigenerational influences of the Fut2 gene on the dynamics of the gut microbiota in mice. Front Microbiol. 8:991. https://doi.org/10.3389/fmicb.2017.00991

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO et al (2011) Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci USA 108:4680–4687

Ray K (2017) Alcoholic liver disease: gut-liver axis: PPIs, entero-coccus and promotion of alcoholic liver disease. Nat Rev Gastroenterol Hepatol 14:689

Rodriguez-Nunez I, Caluag T, Kirby K, Rudick CN, Dziarski R, Gupta D (2017) Nod2 and Nod2-regulated microbiota protect BALB/c mice from diet-induced obesity and metabolic dysfunction. Sci Rep. 7(1):548. https://doi.org/10.1038/s41598-017-00484-2

Rothhammer V, Mascanfroni ID, Bunse L, Takenaka MC, Kenison JE, Mayo L, Chao CC, Patel B, Yan R, Blain M, Alvarez JI, Kébir H, Anandasabapathy N, Izquierdo G, Jung S, Obholzer N, Pochet N, Clish CB, Prinz M, Prat A, Antel J, Quintana FJ (2016) Type I interferons and microbial metabolites of tryptophan modulate astrocyte activity and central nervous system inflammation via the aryl hydrocarbon receptor. Nat Med. 22(6):586–597. https://doi.org/10.1038/nm.4106 **Epub 2016 May 9**

Ruhlemann MC, Degenhardt F, Thingholm LB, Wang J, Skiecevi-ciene J, Rausch P, Hov JR, Lieb W, Karlsen TH, Laudes M et al (2017) Application of the distance-based F test in an mGWAS investigating beta diversity of intestinal microbiota identifies

variants in SLC9A8 (NHE8) and 3 other loci. Gut Microbes 48:1–8

Sadaghian Sadabad M, Regeling A, de Goffau MC, Blokzijl T, Weersma RK, Penders J, Faber KN, Harmsen HJ, Dijkstra G (2015) The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. Gut. 64(10):1546–1552. https://doi.org/10.1136/gutjnl-2014-307289

Santos-Cortez RL, Hutchinson DS, Ajami NJ, Reyes-Quintos MR, Tantoco ML, Labra PJ, Lagrana SM, Pedro M, Llanes EG, Gloria-Cruz TL, Chan AL, Cutiongco-de la Paz EM, Belmont JW, Chonmaitree T, Abes GT, Petrosino JF, Leal SM, Chiong CM (2016) Middle ear microbiome differences in indigenous Filipinos with chronic otitis media due to a duplication in the A2ML1 gene. Infect Dis Poverty 5(1):97

Sender R, Fuchs S, Milo R (2016) Are we really vastly outnum-bered? Revisiting the ratio of bacterial to host cells in humans. Cell 164:337–340

Sovran B, Loonen LM, Lu P, Hugenholtz F, Belzer C, Stolte EH, Boekschoten MV, van Baarlen P, Kleerebezem M, de Vos P, Dekker J, Renes IB, Wells JM (2015) IL-22-STAT3 pathway plays a key role in the maintenance of ileal homeostasis in mice lacking secreted mucus barrier. Inflamm Bowel Dis. 21(3):531–542. https://doi.org/10.1097/MIB.0000000000000319

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

Srinivas G, Moller S, Wang J, Kunzel S, Zillikens D, Baines JF, Ibrahim SM (2013) Genome-wide mapping of gene-microbiota interactions in susceptibility to autoimmune skin blistering. Nat Commun 4:2462

Staubach F, Kunzel S, Baines AC, Yee A, McGee BM, Backhed F, Baines JF, Johnsen JM (2012) Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. ISME J 6:1345–1355

Stein JM, Lammert F, Zimmer V, Granzow M, Reichert S, Schulz S, Ocklenburg C, Conrads G (2010) Clinical periodontal and microbiologic parameters in patients with Crohn's disease with consideration of the CARD15 genotype. J Periodontol. 81 (4):535–545

Taylor SL, Woodman RJ, Chen AC, Burr LD, Gordon DL, McGuckin MA, Wesselingh S, Rogers GB (2017) FUT2 genotype influences lung function, exacerbation frequency and airway microbiota in non-CF bronchiectasis. Thorax 72(4):304–310. https://doi.org/10.1136/thoraxjnl-2016-208775

Thingholm L, Rühlemann M, Wang J, Hübenthal M, Lieb W, Laudes M, Franke A, D'Amato M (2018) Sucrase-isomaltase 15Phe IBS risk variant in relation to dietary carbohydrates and faecal microbiota composition. Gut. https://doi.org/10.1136/gutjnl-2017-315841

Tschurtschenthaler M, Wang J, Fricke C, Fritz TM, Niederreiter L, Adolph TE, Sarcevic E, Künzel S, Offner FA, Kalinke U, Baines JF, Tilg H, Kaser A (2014) Type I interferon signalling in the intestinal epithelium affects Paneth cells, microbial ecology and epithelial regeneration. Gut. 63(12):1921–1931. https://doi.org/10.1136/gutjnl-2013-305863

Protein & Cell

Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A et al (2016) Association of host genome with intestinal microbial composition in a large healthy cohort. Nat Genet 48:1413–1417

Vallier M, Abou Chakra M, Hindersin L, Linnenbrink M, Traulsen A, Baines JF (2017) Evaluating the maintenance of disease-associated variation at the blood group-related gene B4galnt2 in house mice. BMC Evol Biol 17:187

Walter J, Ley R (2011) The human gut microbiome: ecology and recent evolutionary changes. Ann Rev Microbiol 65(65):411–429

Wang J, Kalyan S, Steck N, Turner LM, Harr B, Kunzel S, Vallier M, Hasler R, Franke A, Oberg HH et al (2015) Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. Nat Commun 6:6440

Wang J, Thingholm LB, Skieceviciene J, Rausch P, Kummen M, Hov JR, Degenhardt F, Heinsen FA, Ruhlemann MC, Szymczak S et al (2016) Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat Genet 48:1396–1406

Ward MA, Pierre JF, Leal RF, Huang Y, Shogan B, Dalal SR, Weber CR, Leone VA, Musch MW, An GC, Rao MC, Rubin DT, Raffals LE, Antonopoulos DA, Sogin ML, Hyman NH, Alverdy JC, Chang EB (2016) Insights into the pathogenesis of ulcerative colitis from a murine model of stasis-induced dysbiosis, colonic metaplasia, and genetic susceptibility. Am J Physiol Gastrointest Liver Physiol. 310(11):G973–G988. https://doi.org/10.1152/ajpgi.00017.2016

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA et al (2008) Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature 455:1109–1113

Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3:850–861

World Health Organization (2016) Global tuberculosis report 2016. WHO, Geneva

Xavier RJ, Podolsky DK (2007) Unravelling the pathogenesis of inflammatory bowel disease. Nature 448:427–434

Xie HL, Guo RJ, Zhong HZ, Feng Q, Lan Z, Qin BC, Ward KJ, Jackson MA, Xia Y, Chen X et al (2016) Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. Cell Syst 3:572

Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, Franzosa EA, Vlamakis H, Huttenhower C, Gevers D et al (2016) Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci Transl Med 8:343ra381

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP et al (2012) Human gut microbiome viewed across age and geography. Nature 486:222–227

Ye Y, Carlsson G, Wondimu B, Fahlén A, Karlsson-Sjöberg J, Andersson M, Engstrand L, Yucel-Lindberg T, Modéer T, Pütsep K (2011) Mutations in the ELANE gene are associated with development of periodontitis in patients with severe congenital neutropenia. J Clin Immunol. 31(6):936–945. https://doi.org/10.1007/s10875-011-9572-0 **Epub 2011 Jul 29**

Zhao L, Wang G, Siegel P, He C, Wang H, Zhao W, Zhai Z, Tian F, Zhao J, Zhang H et al (2013) Quantitative genetic background of the host influences gut microbiomes in chickens. Sci Rep 3:1163

Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S et al (2016) Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352:565–569

Zhu B, Wang X, Li L (2010) Human gut microbiome: the second genome of human body. Protein Cell 1:718–725

Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol Rev 32:723–735

Protein & Cell