



# Host and Microbiome Genome-Wide Association Studies: Current State and Challenges

Denis Awany<sup>1†</sup>, Imane Allali<sup>2†</sup>, Shareefa Dalvie<sup>3</sup>, Sian Hemmings<sup>4</sup>, Kilaza S. Mwaikono<sup>2</sup>, Nicholas E. Thomford<sup>1</sup>, Andres Gomez<sup>5</sup>, Nicola Mulder<sup>2</sup> and Emile R. Chimusa<sup>1\*</sup>

<sup>1</sup> Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, <sup>2</sup> Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, <sup>3</sup> Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa, <sup>4</sup> Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, <sup>5</sup> Department of Animal Science, University of Minnesota-Twin Cities, St. Paul, MN, United States

## OPEN ACCESS

### Edited by:

Sara Leslie Pulit,  
University Medical Center Utrecht,  
Netherlands

### Reviewed by:

Omer Weissbrod,  
Harvard T.H. Chan School of Public  
Health, United States  
Alexander Teumer,  
University of Greifswald, Germany

### \*Correspondence:

Emile R. Chimusa  
emile.chimusa@uct.ac.za

<sup>†</sup> These authors have contributed  
equally to this work as first authors

### Specialty section:

This article was submitted to  
Human Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 24 August 2018

**Accepted:** 27 November 2018

**Published:** 22 January 2019

### Citation:

Awany D, Allali I, Dalvie S,  
Hemmings S, Mwaikono KS,  
Thomford NE, Gomez A, Mulder N  
and Chimusa ER (2019) Host  
and Microbiome Genome-Wide  
Association Studies: Current State  
and Challenges. *Front. Genet.* 9:637.  
doi: 10.3389/fgene.2018.00637

The involvement of the microbiome in health and disease is well established. Microbiome genome-wide association studies (mGWAS) are used to elucidate the interaction of host genetic variation with the microbiome. The emergence of this relatively new field has been facilitated by the advent of next generation sequencing technologies that enable the investigation of the complex interaction between host genetics and microbial communities. In this paper, we review recent studies investigating host-microbiome interactions using mGWAS. Additionally, we highlight the marked disparity in the sampling population of mGWAS carried out to date and draw attention to the critical need for inclusion of diverse populations.

**Keywords:** genome-wide association study, microbiome, microbiome-GWAS, host-genetic, host-microbiome interaction

## INTRODUCTION

The past two decades have seen tremendous advancement in our understanding of human genetic variation and its implication in health and disease. This has, in part, been facilitated by extensive scientific collaboration and the exponential increase of technical and methodological advancements (**Figure 1**). Examples of notable large scale scientific collaboration include the Human Genome Project (Sawicki et al., 1993) which published the DNA sequence of the entire human genome; the International Haplotype Map (HapMap) Project (Thorisson et al., 2005) which cataloged the patterns of common polymorphisms (typically minor allele frequency (MAF) larger than 1%) in the human genome and its linkage disequilibrium (LD) structure across multiple ancestral populations. Further, advances in genotyping made feasible, at a relatively low cost, the genotyping of hundreds of thousands (or even millions) of common variants across the human genome. Together these factors catapulted the genome-wide association study (GWAS) in humans (herein referred to as “host”) population.

Genome-wide association study in host populations (hGWAS) has identified hundreds of genetic variants associated with many complex human traits and diseases, novel biological

mechanisms and drug targets for infectious and non-infectious diseases (Relling and Evans, 2015). The microbiome, which is the collection of bacteria, archaea, fungi, protozoa, and viruses that colonize our body surface and their respective genome (Blum, 2017), has shown to play a major role in human health and disease. The success of hGWAS approach provided an optimistic outlook for and eventual implementation to the microbiome. The microbiome genome-wide association study (mGWAS) aims to identify the host's genetic polymorphisms that interact with its microbiome. Recently, mGWASs have identified and validated many heritable bacterial taxa, including the *Christensenellaceae* and *Methanogens* families (Goodrich et al., 2017). Moreover, mGWAS has linked host genotypes and identified pathways with inter-individual variability in microbiome composition in states of health and disease (Hall et al., 2017; Imhann et al., 2018). These findings corroborate the common view that the microbiome plays a significant role in a host's traits, disease susceptibility and resistance, and treatment response.

Even though multiple lines of evidence have indicated significant host–microbiome interactions (Goodrich et al., 2017; Kurilshikov et al., 2017; Weissbrod et al., 2018), the relative strength of these interactions is unclear, with studies yielding somewhat contrasting results (Rothschild et al., 2018). This is perhaps unsurprising given the plasticity of the microbiome to external factors. In light of this, a key, and yet challenging task, is the establishment of truly causative factors in the observed associations between the environment, host genetics and the microbiome when investigating complex traits and diseases. Including the various microbiome data types, that is, proteomic, metabolomic and transcriptomic, to complement the current mostly used genomic data may help illuminate these interactions. However, combining complex and high dimensional data is not straight forward, introducing yet another challenge. In addition, as in the case of hGWAS (Peprah et al., 2015), the existing disparity in microbiome research, in terms of genomic diversity of the sampling population, further thwarts insights into the complex host–microbiome–environment interaction. As depicted in **Figure 2** and elaborated in the section “*disparity in host-microbiome GWAS*,” there is a striking lack of genomic diversity of the study populations in mGWAS published to date.

Here, we review recent studies investigating host–microbiome interactions, through the concept of mGWAS. Then, we highlight the marked disparity in the sampling population of mGWAS carried out to date and draw attention to the critical need for inclusion of non-European populations. Finally, we explore some pertinent challenges in mGWAS.

## HUMAN VARIATION

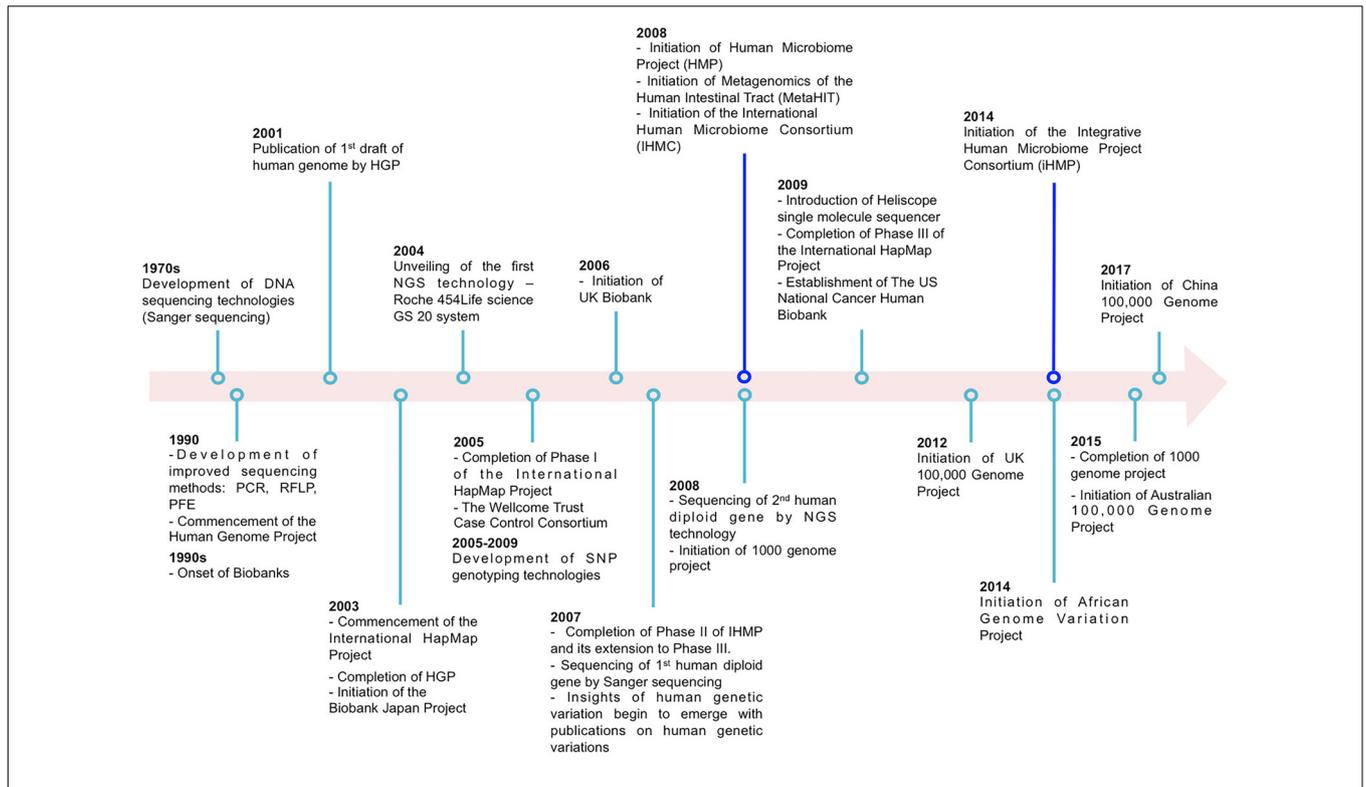
In the realm of genetics, human variation, the variation in allele and/or allele frequency, is inherent in all human populations and underlies population differences in many phenotypic expressions, including resistance and susceptibility to diseases. Genomic variation ranges from large microscopic rearrangements such as insertions and deletions, to smaller submicroscopic variations

such as single-nucleotide polymorphisms (SNPs) and copy number variation (CNV). Analysis of human genomic variation in the 1000 genomes project reported that a typical genome contains ~4.1–5.0 million variants, of which >99.9% are SNPs and indels (1000 Genomes Project Consortium et al., 2015). It is important to note, however, that despite being rare (MAF < 0.5%), structural variants affect more bases, have larger effect sizes (Chiang et al., 2017), and are also thought to be potentially involved in disease pathogenesis due to their enrichment for changes that alter protein sequence and function (Casals and Bertranpetit, 2012; Nelson et al., 2012). Thus, in pursuit of unraveling host genetic variants that interact with the microbiome, knowledge of the abundance and distribution of genetic variants along the genome is critical for characterizing the genetic architecture of common as well as rare traits/diseases, and discerning functionally important variations from the myriad of genomic polymorphisms.

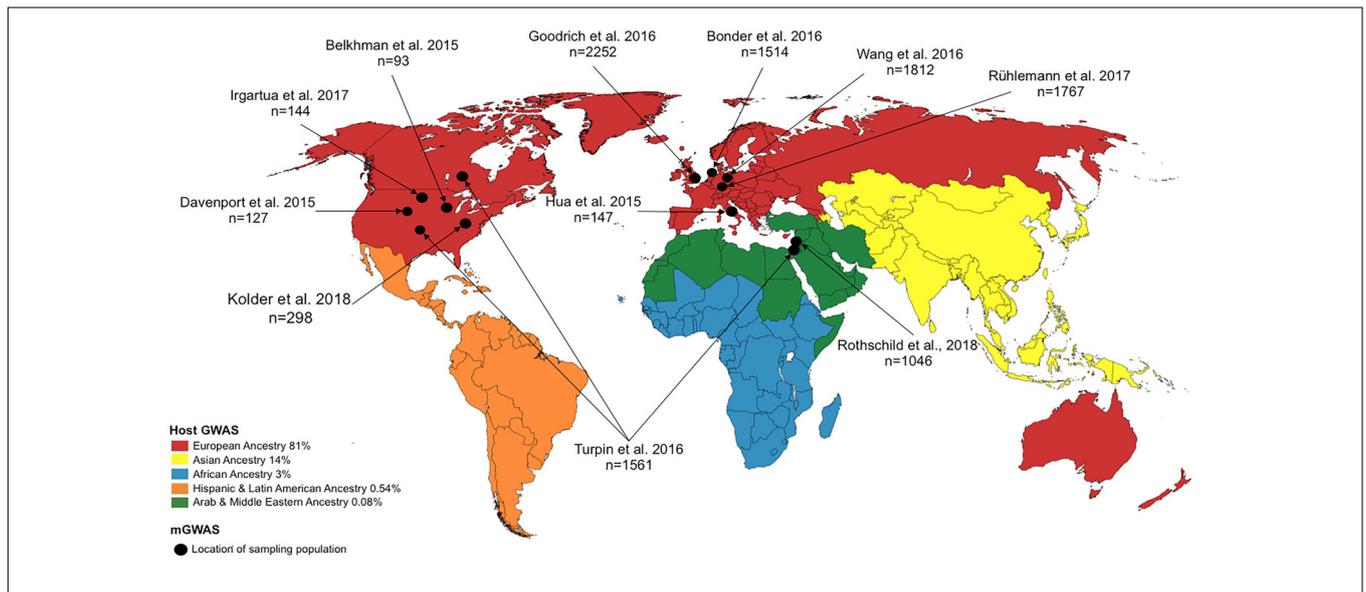
With recent technological advances in genotyping arrays and next-generation sequencing (NGS), it is becoming increasingly feasible to conduct large-scale studies, presenting the opportunity for discovery of both rare and common genetic variation. Deep sequencing offers the opportunity to uncover the complete repertoire of these variations (Marth et al., 2011). However, performing deep sequencing on a genome-wide scale is currently limited due to relatively high cost. To this end, whole-exome sequencing and, particularly, genotype arrays have become methods of choice in the geneticist's arsenal. Intriguingly, these studies, particularly those targeted to protein-coding genes, have revealed the existence of multitude of variants at population and individual level that disrupt protein-coding genes in every human genome (Reijnders et al., 2018), some of which having different phenotypic effects (Blauwendraat et al., 2018; Grarup et al., 2018). These variants generally referred to as loss or gain-of-function variants, occur at low frequency in the genome, and have gene-disrupting ability (MacArthur and Tyler-Smith, 2010); which result in their implications for clinical interpretation of genomic sequences (MacArthur and Tyler-Smith, 2010). It is clear that with NGS technology, many novel genomic variants will be unveiled which, in effect, will facilitate the development of a comprehensive catalog of human genetic diversity.

## HUMAN MICROBIOME DIVERSITY

The human microbiome exhibits both intra- and inter-individual variability (Lax et al., 2014; Gupta et al., 2017). Studies on twins have shown that the microbiota of identical twins are more similar compared to that of their siblings. Also, siblings have a more similar microbiota than that of unrelated individuals (Goodrich et al., 2014). Similar to genetic variation, the human microbiome plays an important role in health and disease (Blekhman et al., 2015; Hall et al., 2017; Gilbert et al., 2018). The microbiome has been associated with variants in host genes involved in immunity and metabolism (Blekhman et al., 2015). The human immune system has a complex



**FIGURE 1 |** Examples illustrating partial major technological and large-scale collaborative projects (excluding data repositories) on host and microbiome genome-wide association studies (GWAS).



**FIGURE 2 |** World map showing study location for host GWAS (represented by continent) and microbiome genome-wide association studies (mGWAS) (represented by country/study site). For host GWAS, the data reflects the state of GWAS in 2016 and the locations refers to continental regions and the proportions of host GWAS using samples recruited from those continental regions are as indicated in the legend [data retrieved from Popejoy and Fullerton (2016)]. For mGWAS, the locations refer to the country/study site where the individual for the study were recruited.

bidirectional relationship with the microbiome. It has been shown that the microbiome is associated with the variability of the immune responses and these responses may also be

involved in modifying the microbiome itself (Belkaid and Hand, 2014; Schirmer et al., 2016; Takiishi et al., 2017). Furthermore, human genetic polymorphisms at various loci are

hypothesized to interact with each other and with an individual's microbiome to impact disease (Blekhman et al., 2015; Hall et al., 2017). In particular, mutations in host genes can influence its interaction with the compositional and functional diversity of the microbiome, potentially modulating an individual's susceptibility to disease (Sandoval-Motta et al., 2017b). In healthy individuals, the microbiome composition is balanced (Rajilic-Stojanovic et al., 2009; Lloyd-Price et al., 2016), and imbalance is now known to be associated with clinical conditions such as diabetes, and inflammatory bowel disease (van Tongeren et al., 2005; Upadhyaya and Banerjee, 2015). Besides host genetics, other factors have been associated with microbial community composition including diet and antibiotic consumption. Many studies have reported that both long- and short-term diet can influence the microbiome composition (Wu et al., 2011; Conlon and Bird, 2014). For example, high-carbohydrate diets have been associated with prevalence of *Prevotella*, while *Bacteroides* are associated with high-fat and high-protein diets (Singh et al., 2017). Additionally, consumption of antibiotics may shift the microbiome composition to a temporally quasi-stable state. This state can be either capable of reverting back to the initial state, or to an alternative irreversible post-antibiotic dysbiosis state (Lozupone et al., 2012). This dysbiosis state is characterized by a loss of taxonomic and functional diversity, which may shift the host's metabolic capacity and reduce the colonization resistance against invading pathogens (Langdon et al., 2016; Lange et al., 2016).

## HOST AND MICROBIOME GWAS

### Host Genome-Wide Association Studies

Genome-wide association study aims to determine the link between genotypic and phenotypic variabilities. This is achieved by obtaining genome-wide genotypic data and phenotypic measurements from a number of subjects, and comparing the frequency of these variants across phenotypic values. GWAS has undoubtedly had successes, identifying thousands of genetic variants associated with hundreds of traits (Visscher et al., 2017), providing valuable insights into the genetic basis of many common traits. Due to LD, any identified associated variant is not necessarily causal as it may simply be “tagging” the causal variant. In addition, most genomic variants are located outside protein-coding regions and are of unknown biological functions (Yang et al., 2017). Consequently, for most traits, little is known about the biological mechanism underlying the associations detected by GWAS.

A critical step toward the elucidation of the underlying biological mechanism is to discern the causal variants. Pinpointing the putative causal variant is, however, challenging for several reasons, including the fact that: (i) most risk regions encompass and implicate multiple variants in the case of complex traits, which without the functional information of the variants, makes it extremely difficult to pinpoint the true causative variant, and (ii) risk variants may reside outside risk regions, and their effects are propagated through regulatory elements (Lin et al., 2016). To this end, several post-GWAS approaches have been

introduced (Wang et al., 2010), driven by the need to leverage GWAS summary statistics to account for polygenicity at the SNP, gene or pathway levels to determine the functional role of the identified variants, uncover their biological mode of action and illuminate their regulatory mechanism (Chen et al., 2015; Chimusa et al., 2015).

Also pertinent to host GWAS is the “missing heritability” problem, which describes the observation that the proportion of heritability explained by the GWAS-associated variants is much less than calculated from familial studies. The reasons for this are still unknown and remains controversial (Marian, 2012; Zaitlen and Kraft, 2012; Sandoval-Motta et al., 2017b), with possible reasons being cited to include epistasis, epigenetics, small effect sizes of the variants, poor coverage of genetic variations on genotyping platforms (Hou and Zhao, 2013). Meanwhile, some researchers have attributed this discrepancy to the fact that GWAS only accounts for genetic variation in human cells and does not consider the effects of the microbiome on phenotype (Marian, 2012; Sandoval-Motta et al., 2017a,b). In light of this, incorporating the microbiome into host GWAS has been hypothesized to significantly reduce the missing heritability gap for microbiome-associated traits (Sandoval-Motta et al., 2017b). However, given that it is not yet known why microbiome is more similar in monozygotic than in dizygotic twins, incorporating other potential sources of variability such as diet, behavior – which are usually assumed to be homogenous across the subjects – may help to explain the missing heritability.

### Microbiome Genome-Wide Association Studies: Approaches and Applications

The microbes that inhabit the human body exist in a synergistic relationship with the human host, performing several important roles in metabolism, detoxification, homeostasis, immunity and epithelial development (Tremaroli and Bäckhed, 2012; Belkaid and Hand, 2014). The human microbiome composition varies widely across different body sites and shows some stability at adulthood for the predominant bacterial communities (Blekhman et al., 2015; Lloyd-Price et al., 2016). Host-microbiome interactions soon establish an equilibrium, which determines the state of health of an individual (Tremaroli and Bäckhed, 2012; Blekhman et al., 2015). Thus, understanding the interactions of both the microbiome and host genetics may provide more insights on disease diagnosis, treatment, and prevention. A number of studies have linked the human microbiome at the various body sites to the development of a wide range of complex traits and diseases, including weight gain, obesity, inflammatory bowel disease, diabetes, cardiovascular disease, cancer, major depression, autism spectrum disorder, and asthma (Kostic et al., 2014; Jiang et al., 2015; Hall et al., 2017; Allali et al., 2018; Qiao et al., 2018). These phenotypic expressions are related to the changes in the overall taxonomic composition, as well as presence or absence of specific bacterial.

Owing to the role played by the microbiome in the pathogenesis of many diseases, there has been a surge of interest in understanding host DNA sequence variations that modulate

the human microbiome (Marchesi et al., 2016). Early insights into interaction between host genome and microbiome were obtained from animal based studies (Blekhman et al., 2015; Wang et al., 2016). For example, Rawls et al. (2006) showed that the observed difference between the microbiotas of zebrafish and mice is due to the underlying host genetics. Host genetic loci that shape diversity in skin microbiota and confer susceptibility to disease in mice were also identified (Srinivas et al., 2013). In humans, studies involving monozygotic and dizygotic twins have shown that the abundance of certain microbial taxa are more correlated amongst monozygotic than with dizygotic twin pairs indicating that host genetic factors are involved in modulating gut microbiome composition across human populations (Goodrich et al., 2014; Xie et al., 2016).

The observations above motivated the advent of microbiome genome-wide association study (mGWAS). Using microbiome attributes (such as alpha diversity, beta diversity or relative abundance of bacterial taxa) as the response variable and host's genotype data as the explanatory variable. mGWAS measures and analyses DNA sequence variations across the host's genome in order to identify genetic factors that modulate the composition and functional diversity of the microbiome. To date, studies published on mGWAS have raised interest and provided new insights (Blekhman et al., 2015; Davenport et al., 2015; Hua et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Wang et al., 2016; Igartua et al., 2017; Rühlemann et al., 2018; Rothschild et al., 2018). The three first studies using mGWAS were conducted on a relatively small sample size. Blekhman et al. conducted the first mGWAS in 93 individuals using human microbiome data and host genetic information gleaned from the Human Microbiome Project (Gilbert et al., 2010); microbiome data and host DNA were from 15 body sites (Blekhman et al., 2015). The authors identified significant associations between several host genes and pathways with microbiome composition. Following this work, Davenport et al. (2015) reported the second mGWAS which investigated host genetic effects on the gut microbiome of 127 Hutteries (North America) and found host SNPs are associated with the abundance of several bacterial taxa. In the third study, Hua et al. have developed the microbiome-GWAS tool that has been tested on 16S rRNA microbiome data from 147 non-malignant lung tissue samples (Yu et al., 2016) to establish the microbiome composition in terms of cancer risk SNPs. The authors found significant associations between six previously established lung cancer risk SNPs and microbiome composition. Subsequent studies have used larger samples of ~300–2000 individuals and have reported significant (Bonder et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Wang et al., 2016). However, Kolde et al. (2018) did not identify significant associations in their untargeted genome-wide analysis in contrast with the findings of Blekhman et al. who used the same cohort and have reported 83 significant associations. Kolde et al. (2018) have reported that the main reason for this difference is the choice of significance thresholds; they used a more stringent Bonferroni correction while Blekhman et al. used false discovery rate (FDR) multiple hypothesis test correction. Additionally, a recent study of 1046 healthy Israeli individuals, with several

different ancestral origins and who share a relatively common environment, did not find any significant associations between (Rothschild et al., 2018). The results of the above studies suggest that some bacterial taxa are heritable but the results of one study cannot be replicated except for the bacterial taxa *Bifidobacterium* which were found to be significantly associated with the lactase *LCT* gene locus (Blekhman et al., 2015; Bonder et al., 2016; Goodrich et al., 2016; Rothschild et al., 2018). **Table 1** summarizes the mGWAS carried out to date.

## mGWAS Approaches and Tools

There are many microbiome attributes that may be leveraged as phenotypes for a mGWAS. First, alpha diversity, that is, the diversity of species within community samples (Wilson and Shmida, 1984) may be used as phenotype and an association performed against host genotypes. Second, beta diversity, that is, the diversity between community samples (Wilson and Shmida, 1984), defined using phylogeny-informed or taxa abundance-informed pairwise distance measures, may be used as phenotype. It is important to note, however, that because the microbiota functions as a community, cross-sample analysis using beta diversity measure is more robust compared to alpha diversity (Hua et al., 2016). Third, the relative abundance of each taxon at a given taxonomic level (species, genus, family, order, class, and phylum) may be used as phenotype, and analysis performed to assess the association of each SNP with the taxon. Alternatively, with shotgun metagenomic sequencing that, unlike the 16S rRNA approach, provides functional information, bacterial pathways may be used as trait for mGWAS. The varied and peculiar features of microbiome phenotypes, particularly, the high dimension (Section "Challenges Underpinning mGWAS") limit the application of some host GWAS tools. Nonetheless, host GWAS (Zhou and Stephens, 2012) and (Lippert et al., 2011) have been applied in mGWAS using taxonomic abundance as bacterial phenotype; albeit, it cannot be used for association testing with microbiome distance metrics.

To this end, microbiome-specific tool and method have recently been developed: microbiomeGWAS (Hua et al., 2015), and microbiome-association index (Rothschild et al., 2018). microbiomeGWAS uses standard linear regression with beta diversity metrics and corrects for skewness and kurtosis. It identifies host genetic variants associated with microbiome beta diversity by testing both SNP-microbiome and SNP-environment interactions. Because the statistical power of these distance-based measures depends on the choice of the distance metric, this tool was subsequently improved to accommodate multiple distance matrices. Meanwhile, microbiome-association index ( $b^2$ ) has been specifically developed to quantify the overall association of microbiome to host's phenotype, incorporating the contribution of host genetics. Using this association index, a measure similar to narrow sense heritability in hGWAS, the authors showed that several host phenotypes, including body mass index, fasting glucose levels, glycaemic status, and lactose consumption, exhibited substantial  $b^2$  values in the range of 22–36%. In addition, different statistical methods have been

**TABLE 1** | Summary list of microbiome genome-wide association studies (mGWAS) carried out to date.

Study	Year	Sequencing method	Analysis software	Sample size and location	Microbiome sampling site	Microbiome phenotype studied	N° of associations identified	Comment
Blekhman et al., 2015	2015	Shotgun metagenomic	PLINK	n = 93 United States of America (HIMP)	Multiple sites (15)	Alpha diversity, beta diversity, and bacterial taxa	83 associations identified	Host genetic variants correlated with microbiome composition. Variants in the <i>LCT</i> gene correlated with abundance of <i>Bifidobacterium</i> ( $P = 1.16 \times 10^{-5}$ ). Genes involved in Leptin signalling pathway significantly associated with microbiome composition; Leptin previously implicated in Obesity.
Davenport et al., 2015	2015	16S rRNA	GEMMA	n = 127 Hutterites (North America)	Gut	Bacterial taxa	≥8 bacterial taxa associated with SNPs in host genome in each season	SNPs in regions of the <i>PLD1</i> gene associated with abundance of genus <i>Akkermansia</i> ; the <i>PLD1</i> gene was previously implicated in GWAS of body mass index. one bacterial taxa (genus <i>Bifidobacterium</i> ) correlated with age. ≥4 bacterial taxa differentially abundant by sex
Hua et al., 2015	2015	16S rRNA	microbiome GWAS	N = 147 Italy	Lung	alpha diversity, beta diversity	Six SNPs had suggestive association with beta-diversity	Analysis performed using both weighted and unweighted UniFrac distance matrices. No SNPs were significantly associated after correcting for skewness and kurtosis of beta-diversity distributions.
Goodrich et al., 2016	2016	16S rRNA	microbiome GWAS (for GWAS on the beta diversity measures) GEMMA (for GWAS on taxon)	n = 1,126 twin pairs United Kingdom	Gut	Bacterial taxa, beta diversity	31 associated host loci	<i>LCT</i> gene associated with <i>Bifidobacterium</i> . Also, SNPs in the region of the <i>R3HDM1</i> gene strongly associated with <i>Bifidobacterium</i> ( $P = 4.38 \times 10^{-8}$ ), <i>ALDH1L1</i> gene associated with SHA-98 bacteria. SNPs associated with beta diversity metrics. alpha diversity found to be heritable. Heritable taxa found, including <i>Bifidobacterium</i> , <i>Turricbacter</i> , and <i>Blautia</i> .
Bonder et al., 2016	2016	Shotgun metagenomic	"base" in R	n = 1,514 Dutch	Gut	Bacterial taxa, bacterial pathways	42 associated host loci	Nine host loci associated with bacterial taxa, and 33 loci with bacterial pathways ( $P < 5 \times 10^{-8}$ ), <i>LCT</i> SNP associated with <i>Bifidobacterium</i> ( $P = 3.45 \times 10^{-8}$ ).
Turpin et al., 2016	2016	16S rRNA	Not sated	n = 1,098 (discovery cohort)- Canada and United States n = 463 (replication cohort)- Canada, United States and Israel	Gut	Bacterial taxa and alpha diversity	58 suggestive associations only six of which were significantly associated	Of these six SNPs, one was replicated in the replication cohort. Associated taxa included <i>Rikenellaceae</i> , <i>Faecalibacterium</i> , <i>Lachnospira</i> , and <i>Eubacterium</i> . Many taxa found to be heritable. No significant association with alpha diversity.
Wang et al., 2016	2016	16S rRNA	"envfit" in R	n = 1,812 Germany	Gut	Bacterial taxa, beta diversity	54 significant associations	42 loci (which included variants in <i>VDR</i> gene -encoding vitamin D receptor) associated with beta diversity ( $P < 5 \times 10^{-5}$ ). Host loci associated with the <i>Firmicutes</i> , <i>Proteobacteria</i> , and <i>Bacteroidetes</i> .
Igartua et al., 2017	2017	16S rRNA	GEMMA	n = 144 Hutterites (North America)	Vestibule and Nasopharynx sites	Relative abundance (RA) of bacterial taxa, alpha, and beta-diversity	37 significant associations	Most significant association was between variant (rs117042385) upstream of the <i>TINCR</i> gene ( $p = 1.61 \times 10^{-8}$ ; $q = 0.002$ ). Differences in RA, alpha and beta diversity were observed across sites and seasons. Significant association between host genetic (measured by kinship coefficient) and microbiome Euclidean distance.

(Continued)

TABLE 1 | Continued

Study	Year	Sequencing method	Analysis software	Sample size and location	Microbiome sampling site	Microbiome phenotype studied	N° of associations identified	Comment
Rühlmann et al., 2018	2017	16S rRNA	"snpStats" in R	n (=1,767 Germany)	Gut	beta diversity	4four significant associations	The four loci were significantly associated with variation in beta diversity. Reanalysis using permutation-based analysis were still identified all these loci as genome-wide significant.
Kolde et al., 2018	2018	Shotgun metagenomic	Matrix eQTL	n (=298 North America (HMP))	Multiple sites (Six)	Relative abundance of bacterial taxa, and bacterial pathway	five significant associations with bacterial taxa, and 82 with bacterial pathways	In stool, five species ( <i>Lachnospiraceae bacterium</i> , <i>Roseburia intestinalis</i> , <i>Subdoligranulum</i> (unclassified), <i>Sutterella wadsworthensis</i> ) out of 118 were significantly associated (false discovery rate (FDR)? < ?0.05 by permutation test). In other body sites, <i>Proplombacterium proplonicum</i> , <i>Porphyromonas catoniae</i> , and unclassified <i>Gemella</i> were significantly associated with host genomic variation in buccal mucosa.
Rothschild et al., 2018	2018	16S rRNA and Shotgun metagenomic	"envfit" and "ordiR2step" in R and FaSTLMM	n = 1,046 Israel	Gut	Relative abundance of bacterial taxa, alpha, and beta-diversity	seven suggestive associations	No significant association between host genetic variation and bacterial taxa or beta diversity, after correcting for multiple testing.

used in mGWAS such as the ordination and permutation-based envfit, ordi2step, snpStats, and Spearman's correlation statistical methods (Table 1).

## DISPARITY IN HOST AND MICROBIOME GWAS

Despite the meteoric rise in GWASs over the last few years, the number of studies inclusive of genetically diverse populations is disproportionately low. A 2015 assessment of the number of NIH-funded GWAS focused on or utilizing non-European populations revealed great disparity; for example, of the 4,942 publications, African American, Hispanic, and Jewish ancestry constituted only ~3%, <1%, and <1%, respectively, of the sampling study population (Peprah et al., 2015). A recent analysis of a curated database of genomic variants associated with various traits/diseases, provided by the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI), revealed a bias in genomic diversity and disproportionate representation (in terms of ancestry, and physical and social environments of the study subjects) in published GWAS (Hindorff et al., 2018). As of August 2016, non-European ancestry represented only 19% of all individuals in GWAS. This becomes a pertinent issue given the observation that non-European individuals contribute a larger number of genotype-phenotype associations (Morales et al., 2018; Hindorff et al., 2018), and studies in other (non-European) population groups continue to identify novel genetic variants. This disparity in population representation is not confined to host GWAS. Although a relatively young field, mGWAS carried out since its inception in 2015, show marked disparity in terms of genetic diversity and population representation. As depicted in Figure 2, most studies involved individuals from North America and Europe. With mGWAS findings differing markedly across all these studies, better inclusivity of diverse populations will illuminate any underlying interactions since it is possible that the interaction of host genetic variant(s) with the microbiome may be population-, environment-, or even individual-specific owing to other yet-unknown clinical and environmental factors.

From a statistical genetics perspective, while the inclusion of non-European populations, in particular the African population, is a critical step toward discovery of important host-genetic interactions in traits/diseases, the implementation requires methodological and technological refinement. It is well known that variants associated with diseases found in populations of European descent do not always replicate in non-European, particularly African populations (Peprah et al., 2015). This discrepancy across populations are due to several possible reasons including differences in allelic architecture, LD, and environmental factors across populations (Popejoy and Fullerton, 2016; Bentley et al., 2017). Thus, there is a need to design appropriate novel statistical models that are tailored to leverage the characteristics of non-European subjects. Moreover, most of the current technologies for mining genetic data, for example genotyping arrays, have been designed for populations

of European descent with long-range patterns of LD (Peprah et al., 2015; Popejoy and Fullerton, 2016) or nearly homogeneous environments.

## CHALLENGES UNDERPINNING mGWAS

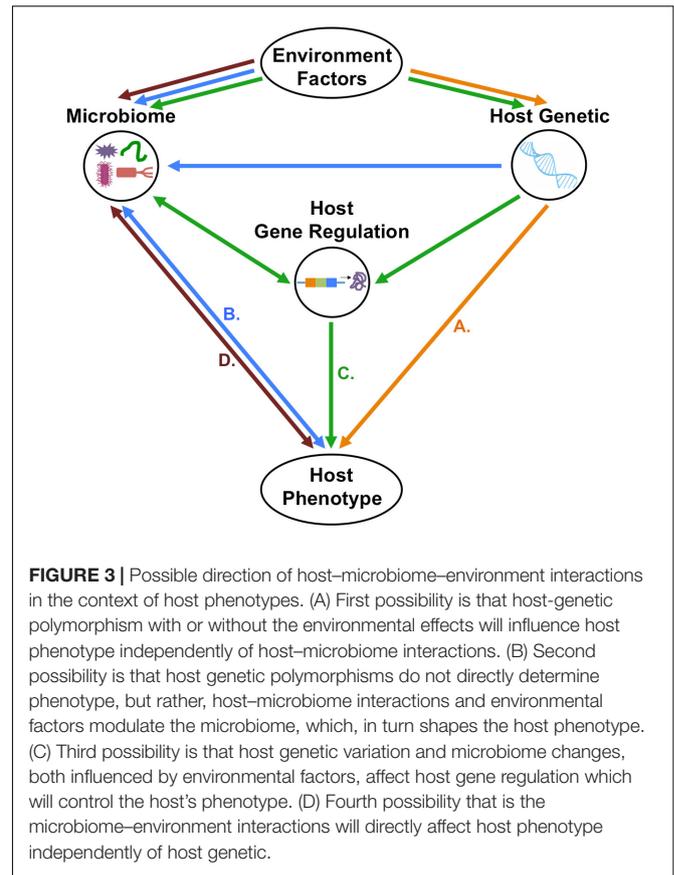
Microbiome genome-wide association studies findings have provided unprecedented views into the association of human host genes with microbes or microbial genes. However, there are several key challenges which also present new opportunities that need to be tackled if we are to assemble a global understanding of host-genetic association with the microbiome.

### Demographic and Environmental Factors

The human microbiome is sensitive to a wide range of demographic and environmental factors. Factors as diverse as gender (Foster et al., 2017), age (Mäkivuokko et al., 2010), and geography (Yatsunenکو et al., 2012) have all been shown to influence the composition and functional diversity of the microbiome. These factors can introduce sampling artifacts or biases in mGWAS which can reduce statistical power. Therefore, the statistical models need to adjust for the effect of these factors. This is particularly important if such factors do not have interaction effect with the genetic variants. Accounting for these factors remains a fundamental challenge for mGWAS. Although it is nearly impractical to adjust for all these factors in a typical setting, it is imperative that enumerable factors be considered as covariates in downstream analyses. Adjustment for potential endogenous and exogenous sources of variability will be key aspects for providing reliable and replicable results.

### The Complexity of Microbiome Data

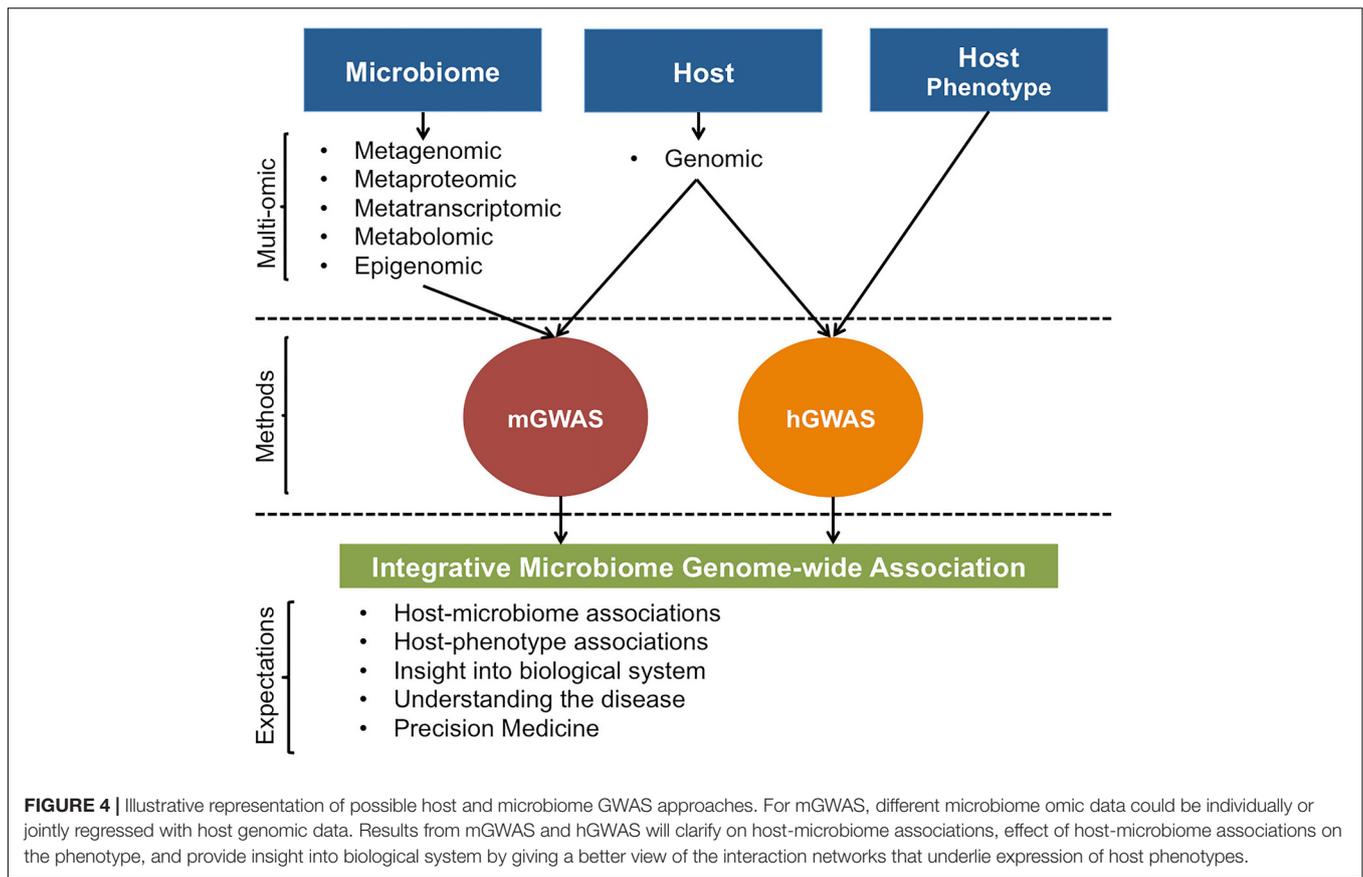
Pertinent to mGWAS, the complexity of microbiome data in terms of dimension, phenotype and correlation structure presents a challenge in the development of robust association frameworks. Microbiome data is highly dimensional, often consisting of hundreds of bacterial taxa. When searching for genetic variants associated with a bacterial trait, multiple tests are carried out, each time testing the null hypothesis of no difference in genotype distribution. With many taxa, this leads to not only high computational cost, but increases in the number of statistical tests. This requires correction for multiple testing to control for the occurrence of false positives. The correction (commonly genome-wide significance, permutation tests, FDR or Bonferroni correction), however, introduces yet another challenge – a potential reduction in statistical power – especially when an underpowered adjustment procedure or an inappropriate error rate is used. The inherent strengths and limitations of each of these correction methods influence association results. For example, (Blekhman et al., 2015) and (Kolde et al., 2018), using FDR and Bonferroni corrections, respectively, obtained contrasting results using the same research cohort. The current solution to circumvent the issue of multiple testing is to focus on only a subset of taxa or variants. For example, (Davenport et al., 2015) reduced the number of bacterial taxa by removing all taxa highly correlated with taxa at the same or lower taxonomic



**FIGURE 3 |** Possible direction of host–microbiome–environment interactions in the context of host phenotypes. (A) First possibility is that host-genetic polymorphism with or without the environmental effects will influence host phenotype independently of host–microbiome interactions. (B) Second possibility is that host genetic polymorphisms do not directly determine phenotype, but rather, host–microbiome interactions and environmental factors modulate the microbiome, which, in turn shapes the host phenotype. (C) Third possibility is that host genetic variation and microbiome changes, both influenced by environmental factors, affect host gene regulation which will control the host’s phenotype. (D) Fourth possibility that is the microbiome–environment interactions will directly affect host phenotype independently of host genetic.

level. Meanwhile, in another study by Bonder et al. (2016), the authors performed a targeted association analysis focusing only on SNPs in genes related to immunity and metabolism. While robust in detecting true associations, using only a selection of taxa leads to incomplete representation of microbiome composition and, consequently, limits the opportunity to discover novel associations. This issue is critical as there may be specific rare taxa that can interact with the host genetics. Likewise, although powerful at dissecting association at a biologically plausible region of interest, reducing the number of host genes variants examined can result in exclusion of relevant genes from the analysis.

In addition, the genetic architecture, the landscape of contributions of host genetics to a given microbiome phenotype, is at best poorly understood although studies of host–genetic interactions with microbiome suggests polygenicity. Moreover, the reported percentage of microbiome variation explained by the associated alleles are generally very small (for example 0.65–0.97% in Wang et al., 2016), and thus unable to explain much of the variability in microbiome phenotype. Given this effect size, large sample sizes will be required to detect modestly associated variants. Meanwhile, the high level of trait collinearity coupled with complex correlation structure (Kurilshikov et al., 2017) also make it challenging for statistical methods. Even though parametric linear models remain the cornerstone for genetic association studies and have played pivotal role in mGWAS



carried out to date, they are limited to detecting non-linear interaction patterns. In particular, when modeling complex structures such as varying effect and non-linear interactions, the exponential rise in the number of parameters increases computational cost and reduces statistical power (Moore et al., 2010). Moreover, linear models generally treat interaction effects as factors with independent marginal effects; a strategy that lowers its power in the presence of interaction effects (Millstein et al., 2006). Given these limitations, there is a need to develop mGWAS-adapted statistical methods to complement existing linear models.

## Replicability of Results

Besides the challenges toward achieving reliable results, replication of mGWAS results has been poor. The first mGWAS, (Blekhman et al., 2015), using 93 individuals and bacterial taxa as phenotypes, detected 83 associations between genetic polymorphisms in host coding genes and the abundance of specific bacterial taxa. Of note, was the association of immune-related genes, *HLA-DRA* and *TLRI*, with abundance of *Selenomonas* and *Lautropia*, respectively, and the strong link between SNPs in the lactase persistence gene, *LCT*, and abundance of *Bifidobacterium*. In subsequent studies, for example, (Davenport et al., 2015; Bonder et al., 2016; Goodrich et al., 2016), many immune and metabolism-related host genes were found to be significantly associated with abundance of

bacterial taxa and beta diversity measures. However, there was little congruence between the results across these studies. This can possibly be attributed to factors including differences in statistical methods, multiple-testing corrections, lifestyle, diet, demographic and environmental conditions of the samples. Nonetheless, the enrichment of microbiome-associated variants with immunity and metabolism related genes, and the generally small effect size, the percentage of microbiome variation explained by the genes, (<1%) remain the most consistent. This suggests that a large proportion of heritability for most bacterial traits may be accounted for by many small effect genetic polymorphisms in immunity and metabolism-encoding genes. A corollary is that bacterial traits likely have an infinitesimal genetic architecture, requiring meta-analyses to detect the associated variants. A combination of larger sample sizes, unified robust analysis methods, and inter-cohort analyses facilitated by collaborations such as MiBioGen consortium (Wang et al., 2018) are crucial for both the attainment of power to detect small to moderate host genetic effects on microbiome traits and replicability of findings.

## CONCLUSION AND PERSPECTIVES

Discovery of the role of the microbiome in normalcy and disease status spurred efforts to elucidate the interaction of host genetic variation with the microbiome, leading to the

development of mGWAS. Initial foray into mGWAS has led to the discovery of host genetic variants that contribute to variability in compositional and functionality diversity of the microbiome. Despite the significant strides made in this field, further developments are still needed to elucidate the extent, direction, and mechanism of host-microbiome association and how this association ultimately impact on host's phenotypic expression (Figure 3).

Moving forward, the importance of sufficient sample sizes cannot be overstated. This will require a collaborative approach, pooling samples from across different geographic regions of the world to generate sufficiently powered studies for discovery and replication. In doing so, the trade-off between sample size and between-sample heterogeneity must be carefully assessed, given the myriad of factors that can reduce the association power of mGWAS. Otherwise, a wide between-sample differences resulting from temporal and spatial heterogeneities will reduce the power to detect true association. In addition to sample size, maximizing bacterial trait information will potentially increase detection power. Current mGWAS has focused on independent analysis of various bacterial traits. This is probably due to the current lack of known software applications that can enable a joint analysis of microbial taxa/pathway and microbial diversity. Given the relative etiological similarity of these traits, it is likely that such joint multiple phenotype analysis will maximize discovery power.

In addition, even though mGWAS to date have primarily focused on the genomic level, regressing host's genetic variation with microbiome's transcriptomic, metabolomic, and proteomic data types or integrating them in a joint host genotype-microbiome association analysis (Figure 4) will be an exciting venture. mGWAS using these multi-level data types can potentially yield insights into whether host-microbiome interactions are, if any, universal or more pronounced to a specific microbial attribute; facilitating identification of particular host genetic polymorphisms that interacts with the microbiome on a molecular level. This is crucial if mGWAS results are to have any utility at level of understanding host's clinical and biological states.

## REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Allali, I., Boukhatem, N., Bouguenouch, L., Hardi, H., Boudouaya, H. A., Cadenas, M. B., et al. (2018). Gut microbiome of moroccan colorectal cancer patients. *Med. Microbiol. Immunol.* 207, 211–215. doi: 10.1007/s00430-018-0542-5
- Belkaid, Y., and Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141. doi: 10.1016/j.cell.2014.03.011
- Bentley, A. R., Callier, S., and Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *J. Commun. Genet.* 8, 255–266. doi: 10.1007/s12687-017-0316-6
- Blauwendraat, C., Reed, X., Kia, D. A., Gan-Or, Z., Lesage, S., Pihlström, L., et al. (2018). Frequency of loss of function variants in *Irrk2* in parkinson disease. *JAMA Neurol.* 75, 1416–1422. doi: 10.1001/jamaneurol.2018.1885
- Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* 16, 1–12. doi: 10.1186/s13059-015-0759-1

Furthermore, the integration of environmental factors in mGWAS will lead to an exciting starting point and perspective for a more comprehensive and robust analysis of host-microbiome interaction in more powered studies. However, the current challenges include difficulty of adjusting for potential environmental factors, complexity of microbiome data, and lack of robust and unified analytical frameworks to handle the diverse and peculiar properties of microbial attributes as quantitative traits. Together, addressing these challenges coupled with increased sample size, independent replication, and meta-analysis in multiple populations will provide a more complete understanding of human variation and microbial diversity in connection with health and disease.

## AUTHOR CONTRIBUTIONS

EC, DA, and IA conceived and structured the manuscript. DA, IA, SD, SH, KM, NT, AG, NM, and EC generated the contents and wrote the manuscript.

## FUNDING

The authors would like to thank DAAD, the German Academic Exchange Programme, for the financial support under Reference number 91653117. Some of the authors are funded in part by the National Institutes of Health Common Fund under grant number U41HG006941, the H3ABioNet Project grant number U24HG006941, SADaCC grant number 1U01HG007459-01, and Wellcome Trust/AESA grant H3A/18/001.

## ACKNOWLEDGMENTS

We are grateful to Oyekanm Nashiru and Hassan Ghazal for helpful discussions.

- Blum, H. E. (2017). The human microbiome. *Adv. Med. Sci.* 62, 414–420. doi: 10.1016/j.advms.2017.04.005
- Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., et al. (2016). The effect of host genetics on the gut microbiome. *Nature genetics* 48, 1407. doi: 10.1038/ng.3663
- Casals, F., and Bertranpetit, J. (2012). Human genetic variation, shared and private. *Science* 337, 39–40. doi: 10.1126/science.1224528
- Chen, H.-S., Hutter, C. M., Mechanic, L. E., Amos, C. I., Bafna, V., Hauser, E. R., et al. (2015). Genetic simulation tools for post-genome wide association studies of complex diseases. *Genet. Epidemiol.* 39, 11–19. doi: 10.1002/gepi.21870
- Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., et al. (2017). The impact of structural variation on human gene expression. *Nat. Genet.* 49, 692. doi: 10.1038/ng.3834
- Chimusa, E. R., Mbiyavanga, M., Mazandu, G. K., and Mulder, N. J. (2015). ancgwas: a post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. *Bioinformatics* 32, 549–556. doi: 10.1093/bioinformatics/btv619

- Conlon, M. A., and Bird, A. R. (2014). The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* 7, 17–44. doi: 10.3390/nu7010017
- Davenport, E. R., Cusanovich, D. A., Michelini, K., Barreiro, L. B., Ober, C., and Gilad, Y. (2015). Genome-wide association studies of the human gut microbiota. *PLoS ONE* 10:e0140301. doi: 10.1371/journal.pone.0140301
- Foster, J. A., Rinaman, L., and Cryan, J. F. (2017). S\*stress & the gut-brain axis: Regulation by the microbiome. *Neurobiology of stress*
- Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., and Knight, R. (2018). Current understanding of the human microbiome. *Nat. Med.* 24, 392. doi: 10.1038/nm.4517
- Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., et al. (2010). Meeting report: the terabase metagenomics workshop and the vision of an earth microbiome project. *Standards in genomic sciences* 3, 243. doi: 10.4056/sigs.1433550
- Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., et al. (2016). Genetic determinants of the gut microbiome in uk twins. *Cell host & microbe* 19, 731–743. doi: 10.1016/j.chom.2016.04.017
- Goodrich, J. K., Davenport, E. R., Clark, A. G., and Ley, R. E. (2017). The relationship between the human genome and microbiome comes into view. *Annu. Rev. Genet.* 51, 413–433. doi: 10.1146/annurev-genet-110711-155532
- Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekman, R., et al. (2014). Human genetics shape the gut microbiome. *Cell* 159, 789–799. doi: 10.1016/j.cell.2014.09.053
- Grarup, N., Moltke, I., Andersen, M. K., Dalby, M., Vitting-Seerup, K., Kern, T., et al. (2018). Loss-of-function variants in *adcy3* increase risk of obesity and type 2 diabetes. *Nat. Genet.* 50, 172. doi: 10.1038/s41588-017-0022-7
- Gupta, V. K., Paul, S., and Dutta, C. (2017). Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Frontiers in microbiology* 8:1162. doi: 10.3389/fmicb.2017.01162
- Hall, A. B., Tolonen, A. C., and Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* 18, 690. doi: 10.1038/nrg.2017.63
- Hindorf, L. A., Bonham, V. L., Brody, L. C., Ginoza, M. E. C., Hutter, C. M., Manolio, T. A., et al. (2018). Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* 19, 175–185. doi: 10.1038/nrg.2017.89
- Hou, L., and Zhao, H. (2013). A review of post-gwas prioritization approaches. *Frontiers in genetics* 4:280. doi: 10.3389/fgene.2013.00280
- Hua, X., Goedert, J. J., Landi, M. T., and Shi, J. (2016). Identifying host genetic variants associated with microbiome composition by testing multiple beta diversity matrices. *Hum. Hered.* 81, 117–126. doi: 10.1159/000448733
- Hua, X., Song, L., Yu, G., Goedert, J. J., Abnet, C. C., Landi, M. T., et al. (2015). Micro\*biomegwas: a tool for identifying host genetic variants associated with microbiome composition. *Biorxiv* doi: 10.1159/000448733
- Igartua, C., Davenport, E. R., Gilad, Y., Nicolae, D. L., Pinto, J., and Ober, C. (2017). Host genetic variation in mucosal immunity pathways influences the upper airway microbiome. *Microbiome* 5, 16. doi: 10.1186/s40168-016-0227-5
- Imhann, F., Vila, A. V., Bonder, M. J., Fu, J., Gevers, D., Visschedijk, M. C., et al. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 67, 108–119. doi: 10.1136/gutjnl-2016-312135
- Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., et al. (2015). Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav. Immun.* 48, 186–194. doi: 10.1016/j.bbi.2015.03.016
- Kolde, R., Franzosa, E. A., Rahnvard, G., Hall, A. B., Vlamakis, H., Stevens, C., et al. (2018). Host genetic variation and its microbiome interactions within the human microbiome project. *Genome medicine* 10, 6. doi: 10.1186/s13073-018-0515-8
- Kostic, A. D., Xavier, R. J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146, 1489–1499. doi: 10.1053/j.gastro.2014.02.009
- Kurilshikov, A., Wijmenga, C., Fu, J., and Zhernakova, A. (2017). Host genetics and gut microbiome: challenges and perspectives. *Trends in immunology* 38, 633–647. doi: 10.1016/j.it.2017.06.003
- Langdon, A., Crook, N., and Dantas, G. (2016). The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome medicine* 8, 39. doi: 10.1186/s13073-016-0294-z
- Lange, K., Buerger, M., Stallmach, A., and Bruns, T. (2016). Effects of antibiotics on gut microbiota. *Dig. Dis.* 34, 260–268. doi: 10.1159/000443360
- Lax, S., Smith, D. P., Hampton-Marcell, J., Owens, S. M., Handley, K. M., Scott, N. M., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345, 1048–1052. doi: 10.1126/science.1254529
- Lin, J.-R., Cai, Y., Zhang, Q., Zhang, W., Nogales, R., and Zhang, Z. (2016). In\*tegrated post-gwas analysis shed new light on the disease mechanisms of schizophrenia. *Genetics, genetics*
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835. doi: 10.1038/nmeth.1681
- Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. *Genome medicine* 8, 51. doi: 10.1186/s13073-016-0307-y
- Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220. doi: 10.1038/nature11550
- MacArthur, D. G., and Tyler-Smith, C. (2010). Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* 19, R125–R130. doi: 10.1093/hmg/ddq365
- Mäkivuokko, H., Tiihonen, K., Tynkkynen, S., Paulin, L., and Rautonen, N. (2010). The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *British journal of nutrition* 103, 227–234. doi: 10.1017/S0007114509991553
- Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D., Hirschfield, G. M., Hold, G., et al. (2016). The gut microbiota and host health: a new clinical frontier. *Gut* 65, 330–339. doi: 10.1136/gutjnl-2015-309990
- Marian, A. J. (2012). Elements of “missing heritability”. *Curr. Opin. Cardiol* 27, 197–201. doi: 10.1097/HCO.0b013e328352707d
- Marth, G. T., Yu, F., Indap, A. R., Garimella, K., Gravel, S., Leong, W. F., et al. (2011). The functional spectrum of low-frequency coding variation. *Genome Biol.* 12, R84. doi: 10.1186/gb-2011-12-9-r84
- Millstein, J., Conti, D. V., Gilliland, F. D., and Gauderman, W. J. (2006). A testing framework for identifying susceptibility genes in the presence of epistasis. *The American Journal of Human Genetics* 78, 15–27. doi: 10.1086/498850
- Moore, J. H., Asselbergs, F. W., and Williams, S. M. (2010). Bioinformatics challenges for genome wide association studies. *Bioinformatics* 26, 445–455. doi: 10.1093/bioinformatics/btp713
- Morales, J., Welter, D., Bowler, E. H., Cerezo, M., Harris, L. W., McMahon, A. C., et al. (2018). A standardized framework for representation of ancestry data in genomics studies, with application to the nhgri-ebi gwas catalog. *Genome Biol.* 19, 21. doi: 10.1186/s13059-018-1396-2
- Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., Jean, P. S., Verzilli, C., et al. (2012). A\*n abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* doi: 10.1126/science.1217876
- Peprah, E., Xu, H., Tekola-Ayele, F., and Royal, C. D. (2015). Genome-wide association studies in Africans and African americans: expanding the framework of the genomics of human traits and disease. *Public Health Genomics* 18, 40–51. doi: 10.1159/000367962
- Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538, 161. doi: 10.1038/538161a
- Qiao, Y., Wu, M., Feng, Y., Zhou, Z., Chen, L., and Chen, F. (2018). Alterations of oral microbiota distinguish children with autism spectrum disorders from healthy controls. *Scientific reports* 8, 1597. doi: 10.1038/s41598-018-19982-y
- Rajilic-Stojanovic, M., Heilig, H. G., Molenaar, D., Kajander, K., Surakka, A., Smidt, H., et al. (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* 11, 1736–1751. doi: 10.1111/j.1462-2920.2009.01900.x
- Rawls, J. F., Mahowald, M. A., Ley, R. E., and Gordon, J. I. (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 127, 423–433. doi: 10.1016/j.cell.2006.08.043
- Reijnders, M. R., Miller, K. A., Alvi, M., Goos, J. A., Lees, M. M., De Burca, A., et al. (2018). De novo and inherited loss-of-function variants in *tlk2*: Clinical and genotype-phenotype evaluation of a distinct neurodevelopmental disorder.

- The American Journal of Human Genetics* 102, 1195–1203. doi: 10.1016/j.ajhg.2018.04.014
- Relling, M. V., and Evans, W. E. (2015). P<sup>\*</sup>armacogenomics in the clinic. *Nature* 526. doi: 10.1038/nature15817
- Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555, 210. doi: 10.1038/nature25973
- Rühlemann, M. C., Degenhardt, F., Thingholm, L. B., Wang, J., Skieceviäinen, J., Rausch, P., et al. (2018). 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* 9, 68–75. doi: 10.1080/19490976.2017.1356979
- Sandoval-Motta, S., Aldana, M., and Frank, A. (2017a). Evolving ecosystems: Inheritance and selection in the light of the microbiome. *Arch. Med. Res.* 48, 780–789. doi: 10.1016/j.arcmed.2018.01.002
- Sandoval-Motta, S., Aldana, M., Martínez-Romero, E., and Frank, A. (2017b). The human microbiome and the missing heritability problem. *Frontiers in genetics* 8:80. doi: 10.3389/fgene.2017.00080
- Sawicki, M. P., Samara, G., Hurwitz, M., and Passaro, E. (1993). Human genome project. *The American journal of surgery* 165, 258–264. doi: 10.1016/S0002-9610(05)80522-7
- Schirmer, M., Smeekens, S. P., Vlamakis, H., Jaeger, M., Oosting, M., Franzosa, E. A., et al. (2016). Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell* 166, 1125–1136. doi: 10.1016/j.cell.2016.10.020
- Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., et al. (2017). Influence of diet on the gut microbiome and implications for human health. *Journal of translational medicine* 552, 73. doi: 10.1186/s12967-017-1175-y
- Srinivas, G., Möller, S., Wang, J., Künzel, S., Zillikens, D., Baines, J. F., et al. (2013). Genome wide mapping of gene–microbiota interactions in susceptibility to autoimmune skin blistering. *Nature communications* 4, 2462. doi: 10.1038/ncomms3462
- Takiishi, T., Fenero, C. I. M., and Câmara, N. O. S. (2017). Intestinal barrier and gut microbiota: shaping our immune responses throughout life. *Tissue Barriers* 5, e1373208. doi: 10.1080/21688370.2017.1373208
- Thorisson, G. A., Smith, A. V., Krishnan, L., and Stein, L. D. (2005). The international hapmap project web site. *Genome Res.* 15, 1592–1593. doi: 10.1101/gr.4413105
- Tremaroli, V., and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature* 489, 242. doi: 10.1038/nature11552
- Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* 48, 1413. doi: 10.1038/ng.3693
- Upadhyaya, S., and Banerjee, G. (2015). Type 2 diabetes and gut microbiome: at the intersection of known and unknown. *Gut Microbes* 6, 85–92. doi: 10.1080/19490976.2015.1024918
- van Tongeren, S. P., Slaets, J. P., Harmsen, H., and Welling, G. W. (2005). Fecal microbiota composition and frailty. *Appl. Environ. Microbiol.* 71, 6438–6442. doi: 10.1128/AEM.71.10.6438-6442.2005
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005
- Wang, J., Kurilshikov, A., Radjabzadeh, D., Turpin, W., Croitoru, K., Bonder, M. J., et al. (2018). Meta-analysis of human genome–microbiome association studies: the microbiome consortium initiative. *Microbiome* 6, 101. doi: 10.1186/s40168-018-0479-3
- Wang, J., Thingholm, L. B., Skieceviäinen, J., Rausch, P., Kummen, M., Hov, J. R., 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. doi: 10.1038/ng.3695
- Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11, 843. doi: 10.1038/nrg2884
- Weissbrod, O., Rothschild, D., Barkan, E., and Segal, E. (2018). Host genetics and microbiome associations through the lens of genome wide association studies. *Curr. Opin. Microbiol.* 44, 9–19. doi: 10.1016/j.mib.2018.05.003
- Wilson, M. V., and Shmida, A. (1984). M<sup>\*</sup>easuring beta diversity with presence-absence data. *The Journal of Ecology* 1055–1064. doi: 10.2307/2259551
- Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344
- Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., et al. (2016). Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell systems* 3, 572–584. doi: 10.1016/j.cels.2016.10.004
- Yang, J., Fritsche, L. G., Zhou, X., Abecasis, G., and International Age-Related Macular Degeneration Genomics Consortium. (2017). A scalable bayesian method for integrating functional information in genome-wide association studies. *The American Journal of Human Genetics* 101, 404–416. doi: 10.1016/j.ajhg.2017.08.002
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222. doi: 10.1038/nature11053
- Yu, G., Gail, M. H., Consonni, D., Carugno, M., Humphrys, M., Pesatori, A. C., et al. (2016). Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genome Biol.* 17, 163. doi: 10.1186/s13059-016-1021-1
- Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. *Hum. Genet.* 131, 1655–1664. doi: 10.1007/s00439-012-1199-6
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821. doi: 10.1038/ng.2310

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Awany, Allali, Dalvie, Hemmings, Mwaikono, Thomford, Gomez, Mulder and Chimusa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.