

METHOD



A behavioral model for mapping the genetic architecture of gut-microbiota networks

Libo Jiang^{a,b,*}, Xinjuan Liu^{c*}, Xiaoqing He^{a,b}, Yi Jin^{a,b}, Yige Cao^{a,b}, Xiang Zhan^d, Christopher H. Griffin^e, Claudia Gragnoli^{d,f,g}, and Rongling Wu^{a,b,d}

^aBeijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China; ^bCenter for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China; ^cDepartment of Gastroenterology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China; ^dDepartment of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA; ^eApplied Research Laboratory, The Pennsylvania State University, University Park, PA, USA; ^fDivision of Endocrinology, Diabetes, and Metabolic Disease, Translational Medicine, Department of Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA; ^gMolecular Biology Laboratory, Bios Biotech Multi Diagnostic Health Center, Rome, Italy

ABSTRACT

The gut microbiota may play an important role in affecting human health. To explore the genetic mechanisms underlying microbiota-host relationships, many genome-wide association studies have begun to identify host genes that shape the microbial composition of the gut. It is becoming increasingly clear that the gut microbiota impacts host processes not only through the action of individual microbes but also their interaction networks. However, a systematic characterization of microbial interactions that occur in densely packed aggregates of the gut bacteria has proven to be extremely difficult. We develop a computational rule of thumb for addressing this issue by integrating ecological behavioral theory and genetic mapping theory. We introduce behavioral ecology theory to derive mathematical descriptors of how each microbe interacts with every other microbe through a web of cooperation and competition. We estimate the emergent properties of gut-microbiota networks reconstructed from these descriptors and map host-driven mutualism, antagonism, aggression, and altruism QTLs. We further integrate path analysis and mapping theory to detect and visualize how host genetic variants affect human diseases by perturbing the internal workings of the gut microbiota. As the proof of concept, we apply our model to analyze a published dataset of the gut microbiota, showing its usefulness and potential to gain new insight into how microbes are organized in human guts. The new model provides an analytical tool for revealing the “endophenotype” role of microbial networks in linking genotype to end-point phenotypes.

ARTICLE HISTORY

Received 15 May 2020
Revised 20 August 2020
Accepted 25 August 2020

KEYWORDS

The gut microbiota; competition; cooperation; rule of thumb; network science; qtl

Introduction

In modern translational medicine, increasing recognition has been gained from the potential associations of the gut microbiota with multiple human diseases.^{1–4} To study the genetic mechanisms of such microbiota-host associations, many large-scale genome-wide association studies (GWAS) have been initiated in recent years, leading to the successful identification of a number of host genetic variants that determine the composition of the gut microbiota.^{5–11} The premise of these studies is that the gut microbiome affects human health or diseases through its taxonomic composition and diversity.¹² For example, in a well-designed GWAS, Turpin et al.¹³ detected 58 host SNPs

associated with the relative abundance of 33 microbial taxa. Growing evidence suggests that the extent to which the microbiota impacts human health risks largely relies on how different microbes communicate, cross-feed, and interact with each other in the gut microorganism community.¹⁴ As such, a systematic illustration of host genetic control over microbial interactions is a meaningful way to shed light on the genetic mechanisms for casual relationships between microbiota and host health.

Several authors have emphasized the use of network approaches to model the interactions among microbes.^{15–17} Stein et al.¹⁸ and Coyte et al.¹⁹ developed various mathematical models for estimating the dynamics of microbial interactions in a time course. These approaches are instrumental

CONTACT Rongling Wu ✉ rwu@bjfu.edu.cn; rwu@phs.psu.edu Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China; Department of Public Health Sciences, Penn State College of Medicine, Hershey PA 17033, USA

*These two authors contribute to this work equally.

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

for dissecting the ecological mechanisms underlying how microbes co-exist and co-evolve in the gut, but their utility is limited to those studies in which microbial abundance is measured repeatedly for the same subject at different time points. The longitudinal collection of the gut microbiota is costly but not very informative in some situations since the microbiome is intrinsically stable and resilient in the gut.^{20,21} Furthermore, these approaches were developed for a small group of highly-resolved microbes, yet we encounter the big challenge of charting a high-dimensional landscape of microbial interactions. The microbiome inhabits and aggregates the host gut at an extremely high density, whose diversity represents one of the most complex ecosystems on earth.²¹ Given such a highly packed aggregation of cells, it is impractical to precisely discern and distinguish interactions between each and every pair of microbes. As a result, the development of a rule of thumb for capturing the fundamental principle governing the gut microbiota has become essential.

The gut microbiota exists as communities with complex interacting and communicating networks through the secretion of chemicals or quorum sensing systems.^{22,23} All types of cooperative and competitive interactions known to occur between microbes play a role in consortia of microbes found in the human gut.²⁴ In animals, how each individual responds to the presence of others in a shared territory obeys a rule of action by which it strives to maximize its success of survival and growth,^{25,26} a process that can be explained by behavioral eco-evolutionary theory.^{27–37} We integrated this theory and network science to develop mathematical descriptors for measuring all possible types of microbe–microbe interactions, including mutualism, antagonism, aggression, and altruism, from the abundance data of the gut microbiota.¹¹ In a previous study, we designed culture experiments using fish and bacteria to validate the biological relevance of these descriptors.³⁸ These descriptors present a computational rule of thumbs that can characterize general principles behind microbial community assembly. Its significant merit lies in its simplicity and flexibility to excavate microbial interaction networks of any dimension at any taxonomical level for individual guts. By regressing the emergent properties of microbial networks on host genotypes, here we develop a statistical procedure to

test and estimate how individual SNPs are associated with network properties and further visualize the genetic architecture of internal workings that take place in the human gut.

Results

Biological validation of mathematical descriptors

In our previous study, we derived four mathematical descriptors for microbial interactions, i.e., Z_{mu} for mutualism, Z_{an} for antagonism, Z_{ag} for aggression, and Z_{al} for altruism,¹¹ which are given in equation (1). In this context, aggression may be thought of as a focal species parasitizing another, and altruism may be thought of as a focal species being parasitized by another. Although the biological relevance of these descriptors has been justified from mono- and co-cultural experiments of 45 *E. coli* strains and 45 *S. aureus* strains,³⁸ we performed an additional larger experiment to further validate these descriptors. We collected 100 diverse bacterial strains from each of these two species, grew each strain in monoculture and its interspecific pair with a randomly selected strain from the other species in co-culture, and measured the abundance of each strain at a series of time points in each culture. The comparison of bacterial growth between these two types of environments allows us to quantify the actual strength of ecological interactions, M_u for mutualism, A_n for antagonism, A_g for aggression, and A_l for altruism, at each time point of growth (see the Methods). The plot of Z_{mu} against M_g shows that they are positively associated ($r = 0.32–0.40$; $P < .001$; **Figure 1(a)**) at three phases of microbial growth, lag, log, and stationary, which are determined from fitted growth equations.³⁹ We also found that significant associations occur between Z_{an} and A_n ($r = 0.12–0.34$; $P < .001$; **Figure 1(b)**), between Z_{ag} and A_g ($r = 0.32–0.73$; $P < .001$; **Figure 1(c)**), and between Z_{al} and A_l ($r = 0.30–0.51$; $P < .001$; **Figure 1(d)**) at each growth phase. In most cases, the correlations are stronger at the stationary growth phase than the log growth phase and lag phase. Taken together, the mathematical descriptors, Z_{mu} , Z_{an} , Z_{ag} , and Z_{al} , can well be used as a statistical measure of M_u , A_n , A_g , and A_l , respectively, which facilitates the use of a general mapping or GWAS population to reconstruct the networks of ecological

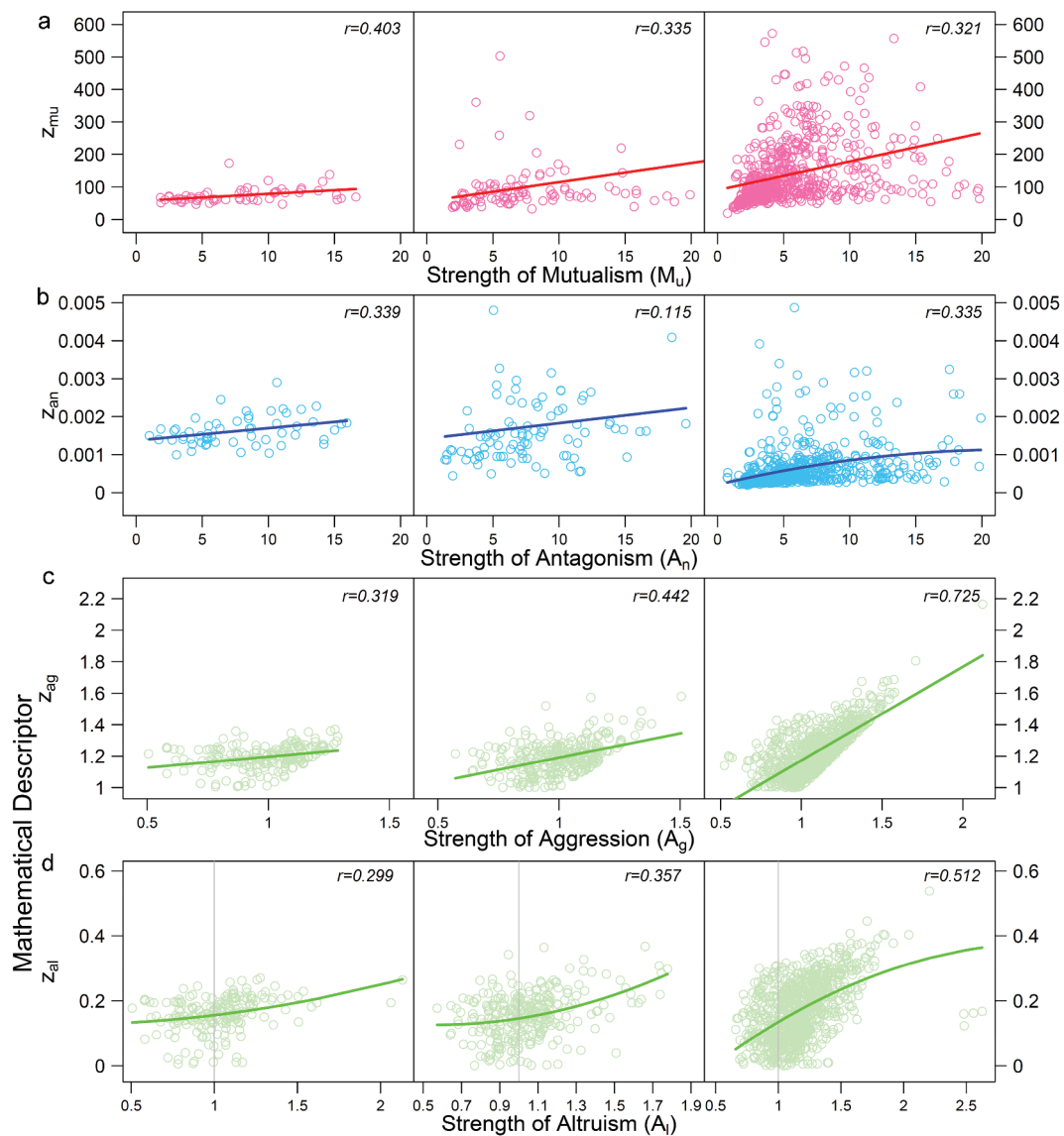


Figure 1. Scatter plots of mathematical descriptors of mutualism (Z_{mu}), antagonism (Z_{an}), aggression (Z_{ag}), and altruism (Z_{al}) against the actual strength of mutualism (M_u) (a), antagonism (A_n) (b), aggression (A_g) (c), and altruism (A_l) (d) across 100 interspecific pairs of strains from *E. coli* and *S. aureus* at three distinct phases of microbial growth (lag, log, and stationary). Dots represent observations of different interspecific strain pairs at each time point. Note that the mutualism dots are those strains whose abundance is larger in co-culture than monoculture for both species, whereas the antagonism dots are those strains whose abundance is larger in monoculture than co-culture for both species. The relationship between two variables is roughly fitted by a curve, with correlation coefficient (r) given within each plot.

interactions and their genetic architecture in the gut microbiota.

Landscaping microbial interaction networks

In a microbial GWAS, Davenport et al.⁶ genotyped 127 hosts of a founder population, the Hutterites, and obtained fecal microbial abundance data (including 101 genera, 50 families, 28 orders, 23

classes, and eight phyla) during two seasons, with 93 hosts in winter, 91 hosts in the following summer, and 57 hosts in both. For each individual, we calculated Z_{mu} , Z_{an} , Z_{ag} , and Z_{al} parameters between each pair of genera and used them to reconstruct four corresponding 101-node networks, each describing gut-microbial interactions based on a different ecological interaction metric. After the significance test of each edge, we obtained a sparse network of each

interaction type averaged over all hosts for each season (Figure 2). One important feature of each network is characterized by those so-called hub or keystone nodes that link with more nodes than a majority of other nodes, thus playing an especially

critical role in maintaining network function. The four networks differ structurally in the pattern of social links and the number of hub microbes. In the mutualism network, 16 and 24 hub microbes, found in winter and summer, respectively, are

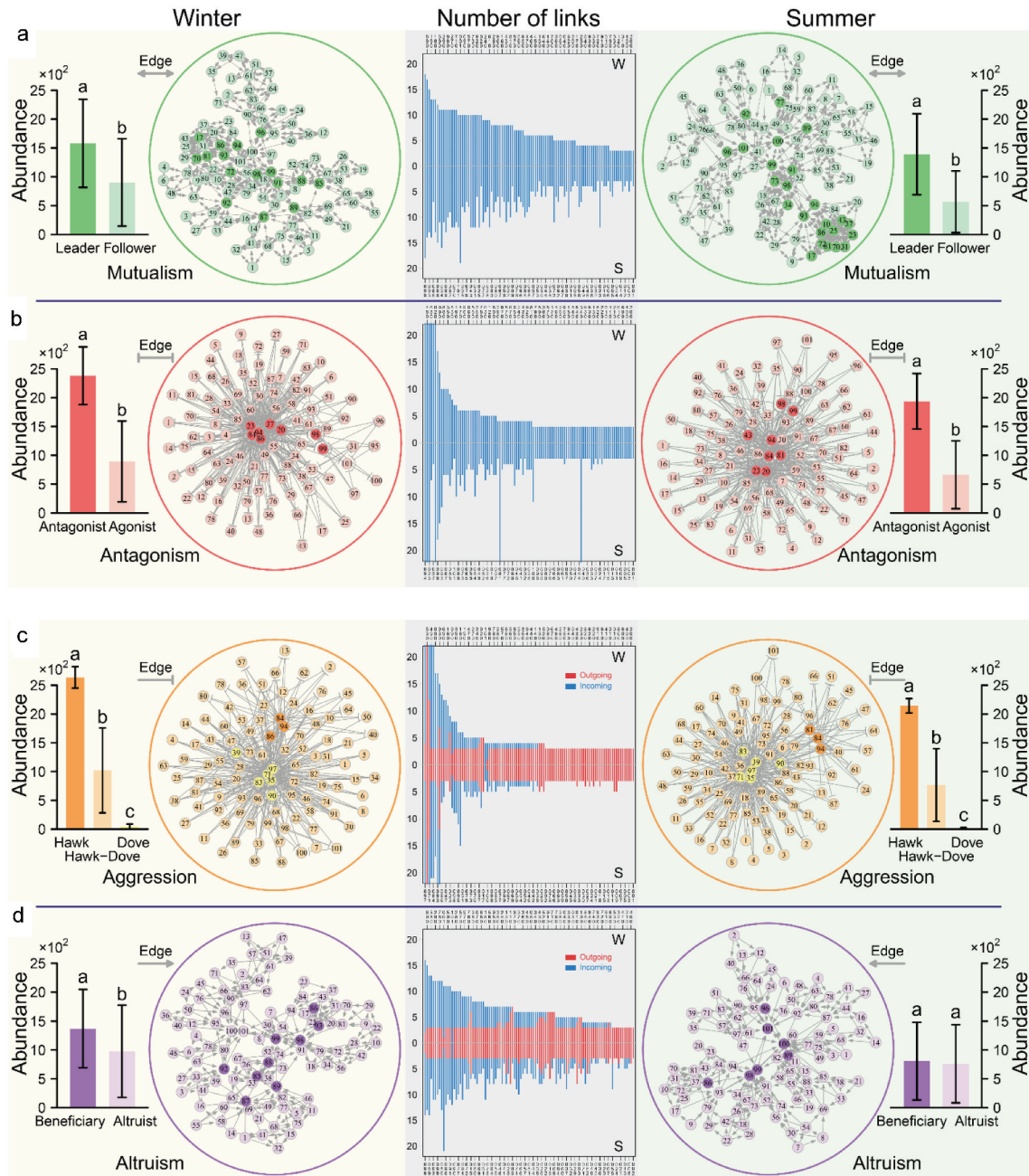


Figure 2. Microbial Z_{mu} -based mutualism (two-way arrowed line) networks (a), Z_{an} -based antagonism (two-way T-shaped line) networks (b), Z_{ag} -based aggression (one-way T-shaped line) networks (c), and Z_{ai} -based altruism (one-way arrowed line) networks (d) at the genus level within the gut microbiota of the Hutterites in winter and summer. In each network, hub microbes are highlighted in dark circles. These hub microbes, expressed as leaders, antagonists, hawks, and beneficiaries in mutualism, antagonism, aggression, and altruism networks, respectively, are compared with other microbes from each network type, called followers, agonists, doves, and altruists, respectively, in bar graphs. The significance of the difference between each pair of these types was tested by a t-test statistic. The identity of each genus is labeled by a number (Table S1). The distribution of links owned by each genus within each network is given in the middle, separately for winter (w) and summer (s).

regarded as being dominant because they are more abundant than the rest (called subordinates) ($P < .05$; [Figure 2\(a\)](#)). The antagonism network includes a few “public enemies” that are antagonistic to many more microbes than others. These antagonistic microbes were observed to be more abundant than the “agonyists” that are less combative ($P < .05$; [Figure 2\(b\)](#)). The aggression network is composed of three hierarchical organizations, the hawk group (those aggressively repressing others), the dove group (those inhibited by others), and the hawk-dove group (those in which a microbe represses its counterpart, but it is also restrained by others). The hawks are much abundant than the hawks-doves ($P < .001$), both being more abundant than the doves ($P < .001$) ([Figure 2\(c\)](#)). The hub microbes of the altruism network act as altruists, which sacrifice their own growth by providing resources exploited by beneficiaries ($P < .01$ for winter) ([Figure 2\(d\)](#)). The test of the distribution of link number by a power law suggests that all networks are scale-free, but show noticeable seasonal difference in each microbe’s link number. Taken together, we have identified remarkable architectural features for each type of interaction network, and microbial networks do not display dramatic season-dependent changes in the main network features. Networks change from season to season mainly in the type of hub microbes.

We calculated six centrality indices, connectivity (Con), closeness (C(u)), betweenness (B(u)), eccentricity (E(u)), eigencentrality (G(u)), and PageRank (P(u)), which describe emergent properties of a network from a different topological perspective for each host in both winter and summer. As can be seen, these indices exhibit pronounced differences among hosts for the same network type and, also, the same index varies dramatically among network types ([Figure 3](#)). All indices are highly season-specific, with the extent depending on network type. All these differences provide a basis for mapping microbial-network QTLs.

Mapping the genetic architecture of microbial interaction networks

By treating each network index as a phenotype, we performed association mapping for the interaction networks. We identified 61 significant host genetic

variants throughout the human genome which are responsible for the centrality indices of each network ([Fig. S1](#)). Because of their role in mediating microbial networks, we call these variants microbial network quantitative trait loci (mnQTLs). First, emergent properties that characterize the same network are under the control of different sets of mnQTLs, suggesting that each property reflects a different topological aspect of a network. Second, in general, no overlapping loci were detected for the four types of networks, from which we conjecture that each type of microbial interactions possesses its own genetic basis. It appears that more mnQTLs control mutualism, antagonism, and altruism than aggression. Third, no common loci were detected between two seasons for each network, indicating that the genetic control of microbial networks is season-dependent. We found that more loci are activated to affect mutualism and altruism in summer than winter, but no such difference was observed for antagonism and aggression mnQTLs.

Based on a gene enrichment analysis, we found that 43 mnQTLs (73%) are located in the regions of candidate genes and that each have specific molecular and developmental functions ([Table 1](#)). For example, mutualism mnQTL for network connection acts like gene FBLN1 encoding fibulin-1, an extracellular matrix and plasma protein,⁴⁰ in winter, but it is more like GABRG2⁴¹ and KDM4C⁴² that encode gamma-aminobutyric acid receptor subunit gamma-2 and Lysine-specific demethylase 4 C in humans, respectively, in summer. The altruism mnQTL for network connection in winter represents gene DIP2A encoding disco-interacting protein 2 homolog A in humans,⁴³ which functions differently from the same season’s mutualism mnQTLs for network connection. mnQTL, rs6110241, is only pleiotropic QTL detected to influence multiple types of microbial networks. This QTL, residing in gene MACROD2 encoding the macrodomain containing 2/mono-ADP ribosylhydrolase 2,⁴⁴ which jointly controls the closeness of mutualism, antagonism, and aggression networks in winter ([Table 1](#)). Those mnQTLs that cannot be annotated may have some unknown functions, deserving of further investigations.

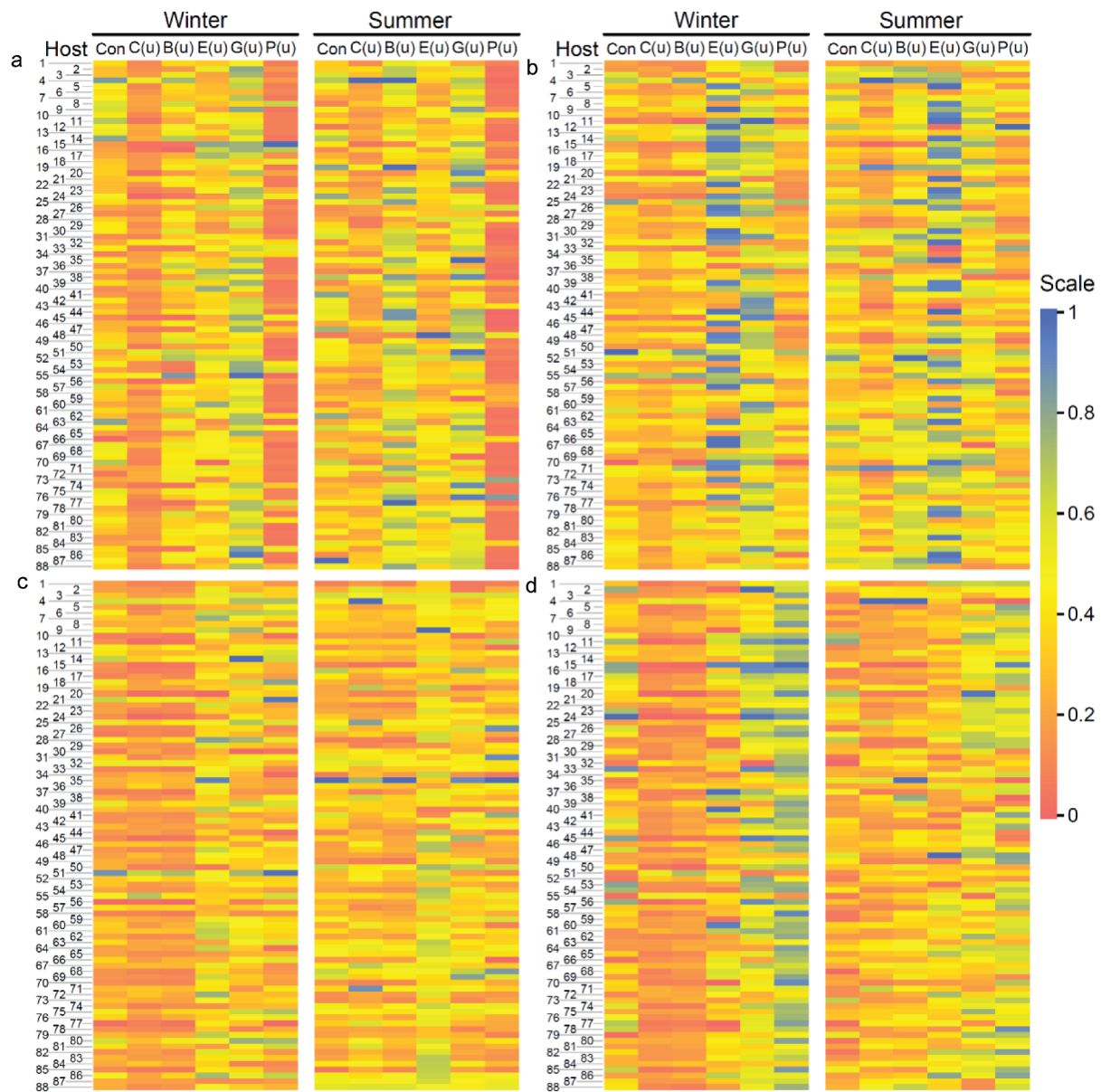


Figure 3. Heatmaps of six indices (showing emergent network properties) constituting mutualism (a), antagonism (b), aggression (c), and altruism networks (d) among 101 genera for network properties for winter and summer.

Impact of network perturbation on genotype-phenotype relationship

Causal links from genotype to phenotype (e.g., SNPs to BMI) may involve the impact of the gut microbiota. We implemented path analysis to dissect how microbial networks influence genotype-phenotype link through their perturbations. Davenport et al.'s⁶ data include 57 hosts whose microbial abundance was measured in both winter and summer. The differences of property parameters for the same host between the two seasons can be used as a measure of the

season-driven perturbation of microbial networks. Although the sample size is quite small, this data can well be used to demonstrate the usefulness of our method for investigating how the significant association between SNPs and BMI is affected by the perturbations of microbial interaction networks. Our analysis is based on five significant SNPs or QTLs that were detected to be associated with BMI.⁶ By viewing a QTL, network perturbation, and BMI as a system, we conducted path analysis to dissect the roadmap from each of these QTLs to BMI

Table 1. Numbers of mutualism, antagonism, aggression, and altruism mnQTLs that affect the emergent properties of ecological networks. The names of genes to which QTLs are annotated are given below.

		Emergent Property					
		Con	C(u)	B(u)	E(u)	G(u)	P(u)
Mutualism mnQTL	Winter	1 <i>FBLN1</i>	2 <i>MACROD2</i>	1 <i>PDXK</i>	0 -	5 <i>RTN4R</i> <i>TNRC6B</i> <i>BIK</i>	0 -
	Summer	8 <i>GABRG2</i> <i>KDM4C</i>	2 -	0 <i>ATP10B</i>	1 -	0 <i>DOCK10</i> <i>LOC105375132</i>	4
Antagonism mnQTL	Winter	0 <i>MACROD2</i>	2 -	0 <i>OAS3</i> <i>DNAH10</i>	4 -	0 <i>SOCS3</i>	1
	Summer	0 -	0 -	0 -	2 <i>IQSEC3</i>	2 <i>LOC339166</i>	2 <i>SCN1A-AS1</i>
Aggression mnQTL	Winter	0 -	1 <i>MACROD2</i>	0 -	0 -	1 -	0 -
	Summer	1 <i>LOC101929231</i>	0 -	1 <i>IQCE</i>	0 -	0 -	1 <i>FCRL3</i>
Altruism mnQTL	Winter	1 <i>DIP2A</i>	0 -	0 -	1 <i>PTPRN2</i>	1 <i>BTBD9</i>	3 <i>SH3GL2</i>
	Summer	0 -	0 -	4 <i>ADAMTS12</i> <i>GABRG2</i> <i>TMPRSS6</i>	0 -	0 -	9 <i>DROSHA</i>

into two path: the direct path and the indirect path through microbial perturbations (Figure 4). A QTL may link with six network properties, but we only chose and incorporated those links that are significant by correlation analysis at or beyond the 10% significance level into the path system. Although all QTLs display a sizable direct effect on BMI, they also affect BMI through the indirect effects of microbial perturbation as enophenotypes. For example, QTL3, residing in the genomic region of gene *LINC01818*, positively affects BMI in a direct way, but it also affects BMI through negative indirect effects of betweenness and eccentricity in the mutualism network (Figure 4(a)). The correlation observed between this QTL and BMI is the overall sum of these direct and indirect effects. In the antagonism network, the pattern of how this QTL affects BMI is determined indirectly by the negative effect of closeness (Figure 4(b)). A similar difference can be found for the aggression (Figure 4(c)) and altruism networks (Figure 4(d)). Taken together, how the perturbations of microbial networks determine the SNP-BMI relationship heavily relies on the network type, network property and QTL.

Monte Carlo simulation

We performed a computer simulation study to examine the statistical power of our model. We mimicked the sampling design of Davenport et al.'s⁶ study by simulating host QTLs that affect the abundance of microbes in the gut. We assume eight microbes that are interact with each other through cooperation or competition. As an example, we focus on two types of interactions, mutualisms and aggression. Three schemes were used to define QTLs in terms of the proportion of its genetic variance to the total phenotypic variance: big QTL (proportion = 0.10–0.20), moderate QTL (proportion = 0.05–0.10), and small QTL (proportion = 0.01–0.05). Traditional GWAS models can only detect the QTLs responsible for the abundance of each microbe, whereas our new model can identify QTLs that regulate how different microbes interact with one another to determine microbial communities using the same data. Based on 1,000 simulation replicates, we calculated and compared the detection power of significant QTLs from both the traditional and the new models (Table 2). We found that the power of detecting mutualism or aggression QTLs by the new model is strikingly larger than the power of detecting abundance QTLs by the traditional model, especially when the QTLs have small effects. For example, the traditional

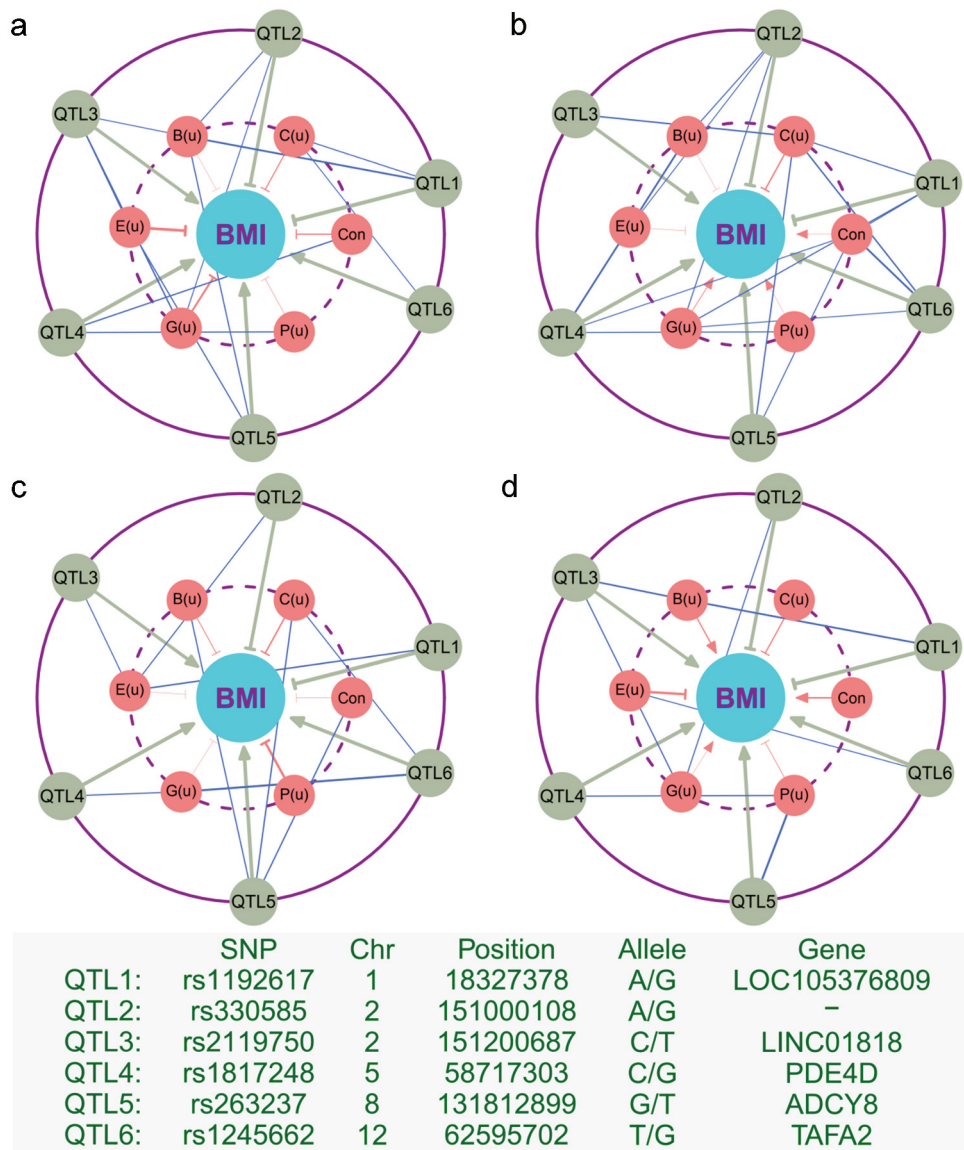


Figure 4. Path analysis revealing how QTLs (outer) affect BMI as a final phenotype (inner) through the season-driven perturbation of microbial networks as an “endophenotype” (middle) (described by differences of six emergent property indices between winter and summer). Path coefficients are denoted by directed lines from QTLs to BMI (gray) and from networks to BMI (blue). Arrowed line and T-shaped line represent a positive and negative impact of path, respectively. Correlation coefficients between QTLs and networks are denoted by non-directed lines. In all cases, the magnitudes of path and correlation coefficients are proportional to the thickness of lines.

Table 2. Power comparison of host QTL detection by a traditional model (aimed to detect abundance QTLs) and our new model (aimed to identify microbial interactions including mutualism or aggression).

	Traditional Model	New Model	
		Mutualism	Aggression
Big QTL	0.89 ± 0.032	0.99 ± 0.003	0.99 ± 0.005
Moderate QTL	0.42 ± 0.051	0.92 ± 0.024	0.89 ± 0.031
Small QTL	0.19 ± 0.046	0.72 ± 0.046	0.67 ± 0.047

model has the power of only about 0.20 to detect a small QTL, whereas the new model increases the power of QTL detection to approximately 0.70. We

also found that the new model shows reasonably good precision in parameter estimate and relatively low false positive rates for QTL detection (< 0.08).

Discussion

Despite a vast amount of data increasingly available to explore the associations of the gut microbiota and human health risk, we are faced with the big challenge of identifying a complete picture of these associations and interpreting their underlying biological

mechanisms. As the gut microbiota is an ecosystem inhabited by highly dense and highly diverse microorganisms, no constituent members function in isolation. Rather, different members affect host health through a complex network of microbial interactions.⁴⁵ As such, community ecology concerned with the composition and structure of biotic assemblages on our planet is believed to be useful for dissecting microbial interactions, and some analytical tools, such as network reconstruction, have been introduced to the microbiome community.^{15,17} However, the traditional theory and corresponding tools may be inadequate to capture the mechanistic details of how ecological interactions are generated and how they are at play within the gut microbiota existing as a heterogeneous biological community.

To quantify internal workings within the gut that contains a typical highly-dense microbial community, we proposed simple mathematical descriptors of pairwise interactions.³⁸ While there is substantial debate as to the role of higher-order interactions in microbial communities,⁴⁶ we believe our pairwise interaction results provide a critical starting point to investigate these higher order interactions more deeply. In this study, we performed a large-scale cultural experiment using diverse strains of two bacterial species to validate the biological relevance of these descriptors. We further integrated these mathematical descriptors into a mapping or GWAS setting to unveil the genetic and molecular mechanisms underlying microbial interactions in the host gut. In a well-designed GWAS study, Davenport et al.⁶ found a few significant QTLs associated with the abundance of eight bacterial taxa. To show how our new model can be used in practice, we reanalyzed this data, identifying many more QTLs that participate in the mediation of microbial interactions. Results from simulation studies show that our model outperforms traditional approaches for GWAS data analysis in several key statistical criteria, including power, estimation precision, and false discovery rate. Most of these so-called microbial interaction QTLs (mnQTLs) (70%) can be annotated to candidate genes with known biological functions (Table 1). One of the mnQTL at candidate gene, *MACROD2*, which exerts a pleiotropic effect on an emergent property of mutualism, antagonism, and aggression networks, deserves further investigation. Recent studies show

that *MACROD2* is related to neurodevelopmental disorder⁴⁷ and cancer.⁴⁸ For example, Mohseni et al.⁴⁴ identified the overexpression of this gene in metastatic tamoxifen-resistant breast cancer. Other studies show that microbial composition has a tight link with autism disorders⁴⁹ and breast cancer.⁵⁰ As such, although our discoveries should be interpreted with caution because of use of imputed SNP data (see the Methods), it may be reasonable to formulate and test a novel hypothesis regarding the paper-rock-scissors game of *MACROD2*, microbes, and diseases.

Compared to the QTLs detected by a traditional model, those by our model affords biologically more meaningful interpretations of microbial networks. Our model can discern the ecological discrepancies of genetic effects exerted by specific mnQTLs; for example, mutualism mnQTLs play a role in controlling how microbes cooperate with each other through secreted chemicals or quorum sensing,^{22,23} whereas antagonism mnQTLs are responsible for how different microbes choose to compete for resources in limited space. Aggression mnQTLs determine whether and how those aggressive microbes maximize their growth by exploiting the resources of other submissive microbes. Altruism mnQTLs guide how microbes sacrifice themselves to benefit other microbes. Further computational and experimental approaches are needed to infer the casual effects of these mnQTLs on microbial community network structure. With such information, we can better understand how mnQTLs determine the stability and dynamics of microbial community assemble through affecting the emergent properties of networks. Also, by altering the function of some mnQTLs, gut-microbial communities may be driven toward a direction desirable to human wellness.

As an increasingly recognized determinant of human phenotype, the gut microbiota can be viewed as the “endophenotypes” that bridge the link of genotype to phenotype. By integrating our model and path analysis, we can characterize how QTLs affect host phenotypes through perturbing gut-microbial networks. Through the structural dissection of genotype-phenotype relationship, path analysis can identify how QTLs determine host phenotype through their direct effects or the indirect effect of endophenotypic microbial

networks. Using Davenport et al.'s⁶ data, we identified these two different paths for the BMI QTLs. We found that indirect genetic effects related to these QTLs are mediated through multiple microbial interaction types. Although our results should be interpreted with caution because of the use of a small sample size, our model does show its usefulness to dissect the microbe-driven roadmap from genotype to phenotype. Taken together, a holistic, systems-oriented approach is needed to comprehend the mechanistic basis for QTL-microbiota-phenotype systems, generating useful information for practical translational medicine.

Obviously, the application of our theory is not limited to the human microbiota. It may find its immediate implications for understanding ecological interactions of other biofilms, such as soil microbiota, ocean microbiota, and atmosphere microbiota.⁵¹ The framework can identify common principles that guide the structure and function of these biofilms and build predictive models for linking microbial compositions to the function of various ecosystems. As pinpointed by a group of microbiologists, despite our remarkable capacity to identify species compositions of microbial communities, we are not yet able to precisely predict and manipulate the function of microbiota because of the lack of fundamental knowledge about their inner workings.⁵¹ Our integrative theory of behavioral ecology, network science, and, a mapping model provides a novel attempt to advance toward the goals of predicting the behavior and properties of microbiota and thus possibly engineering synthetic microbial consortia with novel function.

Conclusions

The gut microbiota determine human health through their complex inner interactions, a process encoded by the human genome. However, modeling microbial interactions within such highly dense community assemblies represents a major challenge. We integrate behavioral ecology and network science to develop a rule of thumbs for defining and quantifying the networks of various interaction types for microbial communities. We map the host genetic architecture of how microbes interact and work together to determine

microbial community behavior. We implement path analysis to reveal the direct effects of the genotype, as well as its indirect effects through microbial networks as the “endophenotype,” on host phenotype. Our model provides a tool that potentially can be used to understand the mechanistic basis of structural-functional relationships within the gut microbiota.

Methods

Calculating microbial interactions

Suppose there is a genome-wide association study (GWAS) in which a sample of n human subjects are randomly drawn from a natural population. To study how an environmental factor influences the gut microbiota, stool samples are collected under different conditions. For each sample, 16S rRNA gene sequencing is used to monitor the abundance of bacteria at different levels of taxa from genus, families, orders, and classes to phyla. Meanwhile, all sampled subjects are measured for health- or disease-related phenotypes, such as body mass index (BMI), diabetes, gastroenteritis, or Crohn's disease, under the conditions considered. To investigate the genetic control of the gut microbiota, all hosts are genotyped for SNPs by a high-throughput genotyping technique for subsequent genome-wide association studies (GWAS).

A linear mixed model is used to correct for microbial abundance and host phenotypic data due to the relatedness between individual subjects and other demographical effects resulting from age and sex among others. We use the corrected microbial abundance data to identify interaction networks at a particular taxonomical level. Let x_u and x_v ($x_u > x_v$) denote the abundance levels of two microbes u and v ($u, v = 1, \dots, m; u \neq v$) from a total of m microbes, respectively. Based on animal behavioral ecology theory, Wang et al.¹¹ argue that the edges of the networks reconstructed by m microbes can be calculated as

$$\begin{cases} Z_{mu} = \frac{x_u x_v}{x_u - x_v} & \text{Mutualism} \\ Z_{an} = \frac{1}{(x_u x_v)(x_u - x_v)} & \text{Antagonism} \\ Z_{ag} = \frac{x_u}{x_v} & \text{Aggression} \\ Z_{al} = 1 - \frac{x_v}{x_u} & \text{Altruism} \end{cases} \quad (1)$$

with $u, v = 1, \dots, m$ ($u \neq v$). Based on equation (1), we can use Z_{mu} to quantify the extent to which two microbes u and v cooperate with each other (mutualism), Z_{an} to quantify the extent to which these two microbes compete against each other (antagonism), Z_{ag} to quantify the extent to which the more abundant microbe u parasitizes or exploits the less abundant microbe v (aggression from u to v), and Z_{al} to quantify the extent to which the more abundant microbe u is parasitized to benefit the less abundant microbe v (altruism from u to v).

Experimental validation of interaction descriptors

We designed an experiment to validate equation (1)'s descriptors. Consider a pair of microbes that are cultured separately and jointly in different bearers but with the same growth condition. In the shared environment, organisms need to interact with each other by choosing either cooperation or competition.^{31–38} We denote the more abundant microbe in co-culture as u and the less abundant microbe as v . Let x_u and x_v as well as w_u and w_v denote the abundance levels of microbes u and v in co-culture as well as monoculture, respectively. If two microbes cooperate with each other, then the relative abundance of each microbe should be larger in co-culture than in monoculture, i.e., $x_u/w_u > 1$ and $x_v/w_v > 1$. The magnitude of the product of these two ratios $(x_u/w_u)(x_v/w_v)$, adjusted by $|x_u/w_u - x_v/w_v|$, is positively proportional to the degree of mutualism between the two microbes. If two microbes compete against each other, then the relative abundance of each microbe should be larger in monoculture than in co-culture, i.e., $w_u/x_u > 1$ and $w_v/x_v > 1$. Then, the size of $(w_u/x_u)(w_v/x_v)$, adjusted by $|w_u/x_u - w_v/x_v|$, is positively proportional to the degree of antagonism between the two microbes. If one microbe is aggressive on the other, i.e., the former grows at a cost of the latter, then the relative abundance of the former over the latter should be greater when the two microbes grow in the common environment than when they grow alone. Accordingly, if one microbe is functionally altruistic toward the other, i.e. the former grows poorly while the other benefits from this, then the relative abundance of the latter in co-culture over monoculture should be larger than the relative abundance of the former in co-culture over

monoculture. Based on these lines of reasoning, the actual strength of interactions can be quantified using the following equations:

$$\begin{cases} M_u = \frac{(x_u/w_u)(x_v/w_v)}{|x_u/w_u - x_v/w_v|} & \text{Mutualism(symmetrical)} \\ A_n = \frac{(w_u/x_u)(w_v/x_v)}{|w_u/x_u - w_v/x_v|} & \text{Antagonism(symmetrical)} \\ A_g = \frac{x_u/x_v}{w_u/w_v} & \text{Aggression(uaggressivetov)} \\ A_l = \frac{x_v/w_v}{x_u/w_u} & \text{Altruism(ualtruistictov)} \end{cases} \quad (2)$$

where we use the first two letters to represent each type of interactions. By analyzing the abundance data from co-culture and monoculture, we use equation (2) to calculate the strength of each interaction type and compare it with the descriptor estimated by equation (1). If they are positively significantly correlated by statistical testing, we conclude that the descriptors can be used as proxies to measure the strength of interactions.

Emergent properties of microbial networks

By calculating mathematical descriptors for each pair from m microbes in each host gut using equation (1), we reconstructed an m -node network for mutualism, antagonism, aggression, and altruism, respectively. Network science methods use six indices to describe these host-specific networks:

$$C(u) = 1 / \sum_{1 \leq v, v \neq u} D(u, v), \quad (3)$$

- (i) **Connectivity** is the average number of nodes (microbes) with which a microbe links⁵² and describe the density of links within a network;
- (ii) **Closeness** describes the degree of linkage between one microbe and others, calculated as

where m is the number of microbes, $D(u, v)$ is the minimum distance between microbes u and v ;⁵³

$$B(u) = \sum_{1 \leq v < u} \frac{g_{vw}(u)}{g_{vw}(4)} \quad (4)$$

- (i) **Betweenness** reflects the importance of a microbe as a bridge across the network, calculated as

where g_{vw} is the number of the shortest paths between microbes v and w , $g_{vw}(u)$ is the number of the shortest paths of microbe u on g_{vw} ;⁵⁴

$$E(u) = \frac{1}{\max(D(u, v))} \quad (5)$$

- (i) **Eccentricity** is the longest distance of one microbe to others,⁵⁵ expressed as

where $u = 1, \dots, v-1, v+1, \dots, m$;

$$G(u) = \left(\frac{1}{\lambda}\right) \sum_{1 \leq v} a_{uv} G(v) \quad (6)$$

- (i) **Eigenvector** describes the importance of a microbe to its neighboring microbes,⁵⁶ calculated as

where λ is a constant solving the equation $AG = \lambda G$, a_{uv} describes whether microbes are linked with each other in the network;

$$P(u) = 1 - d + d \sum_{1 \leq v, u \neq v} a_{uv} P(v) / K_u \quad (7)$$

- (i) **PageRank** is a node ranking method, recursively defined by the equation

: where d is the damping coefficient, usually set as 0.85, and K_u is the number of outbound links for microbe v .⁵⁷

Of these property parameters, connectivity describes the overall structure of a network, whereas the other five are microbe-specific. To reflect how the latter characterize the global network, we may take their average over all microbes.

Mapping microbial network properties

To study how host genes affect the gut microbiota, we genotype the n samples at p genome-wide distributed polymorphic loci. Consider a SNP whose three genotypes are denoted as AA (coded as 1), Aa (coded as 0), and aa (coded as -1), with observations n_1 , n_2 , and n_3 , respectively. We use a linear mixed model to correct network property values for any possible covariates. Let y_i denote the corrected value of a network property on individual i ($i = 1, \dots, n$). The association between the SNP and the

network variable can be tested by formulating a likelihood, expressed as

$$L(y) = \sum_{i=1}^{n_{AA}} f_{AA}(y_i : \mu_{AA}, \sigma^2) \sum_{i=1}^{n_{Aa}} f_{Aa}(y_i : \mu_{Aa}, \sigma^2) \sum_{i=1}^{n_{aa}} f_{aa}(y_i : \mu_{aa}, \sigma^2) \quad (8)$$

where $f(\cdot)$ is the distribution density function of the network property variable, assumed to follow a normal distribution with genotype-dependent means μ . and variance σ^2 . By estimating μ_{AA} , μ_{Aa} , and μ_{aa} and testing their difference based on a log-likelihood ratio via genome-wide permutation resampling, we can assess whether and how the SNP influences the network property.

Most of the current GWAS characterize the association between genotype and high-order phenotypes, such as complex traits or diseases. However, this association may be determined by a certain “black box” behind the causal link from genotype to phenotype. Such a black box is widely recognized as a series of regulatory processes that drive DNA to genes to proteins to metabolites. We argue that the black box may involve the mediation of the gut microbiota because of increasing evidence that the gut microbiota is associated with human phenotypes. Here, we can test whether microbial networks serves as a black box to modulate genotype-phenotype relationship. Let z_i denote the high-order phenotypic value of a quantitative trait on individual i . We calculate the Pearson correlation between y and z across individuals, denoted as r_{yz} to quantify and test the association between the network property and host phenotype. Meanwhile, we use Huo et al.’s⁵⁸ mutual information approach to calculate the correlations between (discrete) genotype (g) and (continuous) network variable (y) and high-order phenotypic trait (z), denoted as r_{gy} and r_{gz} , respectively.

We use path analysis⁵⁹ to dissect r_{gz} into the direct effect of SNP on z and its indirect effect on z through y . Let $P_{z \leftarrow g}$ and $P_{z \leftarrow y}$ denote the path coefficients of SNP and microbial network toward host phenotype. Thus, we have

$$r_{gz} = P_{z \leftarrow g} + P_{z \leftarrow y} r_{gy} \quad (9)$$

where $P_{z \leftarrow g}$ is the direct path from SNP to phenotype, and $P_{z \leftarrow y}$ r_{gy} is the indirect path through microbial network. In the system constituted by SNP, microbial network, and phenotype, SNP is a fixed variable, which is unchanged with other variables. Thus, the correlation between microbial network and phenotype can only include a single direct path, which implies $r_{yz} = P_{z \leftarrow y}$. To characterize how well phenotypic variation is determined by SNP and the microbial network, we calculate the coefficient of determination,

$$R^2 = P_{z \leftarrow g}^2 + P_{z \leftarrow y}^2 + 2r_{gy}P_{z \leftarrow g}P_{z \leftarrow y} \quad (10)$$

and the path coefficient of all unknown variables (e) that contribute to host phenotype,

$$P_{z \leftarrow e} = \sqrt{1 - R^2} \quad (11)$$

We will use R^2 and $P_{z \leftarrow e}$ to evaluate the effect of the microbial network as a black box to determine host phenotype.

Data collection

To validate the mathematical descriptors of equation (1), we performed a microbial experiment by cultivating 100 interspecific pairs of strains from *Escherichia coli* and *Staphylococcus aureus* in monoculture and co-culture. These strains were sampled from the National Infrastructure of Microbial Resources, China. Both cultural experiments used the same two-times diluted brain heart infusion medium (OXOID, Basingstoke, England). In co-culture, two strains were mixed with a 1:1 ratio. We measured the abundance of each strain in each culture once every 0.5 h during the first 2 h of cultivation, followed by once every 2 h till 12 h and once every 4 h till 36 h. Quantitative PCR (qPCR) measurements of strains were detailed in Jiang et al.³⁸ We used three commonly used growth equations, Gompertz, logistic, and Richards,³⁹ to fit the growth data of each strain and then chose an optimal one that suits this strain by statistical reasoning.³⁸ We divided growth trajectories into lag, log, and stationary phases for each strain based on the optimal growth equation.

To justify our mapping model, we used it to reanalyze a published GWAS data for the human gut microbiota. In this GWAS study, the fecal

microbiome from the Hutterites, a religious isolate living in North America, were examined by 16S rRNA gene sequencing in a winter and the next summer. To compare the seasonal difference of microbial interactions, we used 127 subjects (79 females and 48 males) who had data in both seasons. The abundance of microbes was read at the genus, family, order, class and phylum levels. Prior to any analysis, the abundance data were corrected for age and sex. We retrieved the BMI data of the Hutterite samples based on the correlation coefficient (0.51) of BMI with the relative abundance of the genus *Akkermansia*, and used BMI to test how microbial interactions are associated with human health. We were not able to obtain human SNP data, but imputed SNP genotypes based on significant associations reported in the original study. Because of this limitation, this data analysis was better used as the proof of concept for our new theory, whose results should be interpreted with caution.

Availability of data and material

The data and code can be freely uploaded at <https://github.com/LiboJiang/MicrobialNetworks/StaticModel> and used by researchers worldwide. They can also be requested from the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by grant 31401900 from National Natural Science Foundation of China (L.J.) and NICHD 5R01HD086911-02 from the National Institute of Health (C.G.).

References

1. Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012;148:1258–1270. doi:10.1016/j.cell.2012.01.035.
2. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol*. 2012;9:577–589. doi:10.1038/nrgastro.2012.156.
3. Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, et al. The gut microbiota and

- host health: a new clinical frontier. *Gut*. 2015;65:330–339. doi:10.1136/gutjnl-2015-309990.
4. Shanahan F. The gut microbiota—a clinical perspective on lessons learned. *Nat Rev Gastroenterol Hepatol*. 2012;9:609–614. doi:10.1038/nrgastro.2012.145.
 5. Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vila AV, Deelen P, Vatanen T, Schirmer M, Smeekens SP, et al. The effect of host genetics on the gut microbiome. *Nat Genet*. 2016;48:1407–1412.
 6. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y, White BA. Genome-wide association studies of the human gut microbiota. *PLoS ONE*. 2015;10:e0140301. doi:10.1371/journal.pone.0140301.
 7. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhan R, Beaumont M, Van Treuren W, Knight R, Bell JT, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159:789–799. doi:10.1016/j.cell.2014.09.053.
 8. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG, Ley RE. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe*. 2016;19:731–743. doi:10.1016/j.chom.2016.04.017.
 9. Goodrich JK, Davenport ER, Clark AG, Ley RE. The relationship between the human genome and microbiome comes into view. *Annu Rev Genet*. 2017;51:413–433. doi:10.1146/annurev-genet-110711-155532.
 10. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. *Nat Rev Genet*. 2017;18:690–699.
 11. Wang Q, Jiang LB, Cao YG, Zhan X, Griffin CH, Wu RL. 2019. Play a game through a behavioral network? An interrogation of internal workings in microbial community assemble. *mSystems*.
 12. Awany D, Allali I, Dalvie S, Hemmings S, Mwaikono KS, Thomford NE, Gomez A, Mulder N, Chimusa ER. Host and microbiome genome-wide association studies: current state and challenges. *Front Genet*. 2019;9:637. doi:10.3389/fgene.2018.00637.
 13. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A, et al. GEM project research consortium, Paterson AD, Croitoru K. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet*. 2016;48:1413–1417. doi:10.1038/ng.3693.
 14. Wasielewski H, Alcock J, Aktipis A. Resource conflict and cooperation between human host and gut microbiota: implications for nutrition and health. *Ann NY Acad Sci*. 2016;1372:20–28. doi:10.1111/nyas.13118.
 15. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–550. doi:10.1038/nrmicro2832.
 16. Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomics time series using sparse linear regression. *PLoS ONE*. 2014;9(7):e102451. doi:10.1371/journal.pone.0102451.
 17. Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: A network perspective. *Trends Microbiol*. 2017;25:217–228.
 18. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rättsch G, Pamer EG, Sander C, Xavier JB. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLoS Comput Biol*. 2013;9(12):e1003388. doi:10.1371/journal.pcbi.1003388.
 19. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science*. 2015;350:663–666. doi:10.1126/science.aad2602.
 20. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–214.
 21. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489:220–230. doi:10.1038/nature11550.
 22. Asfahl KL, Schuster M, Gibbs K. Social interactions in bacterial cell-cell signaling. *FEMS Microbiol Rev*. 2017;41:92–107. doi:10.1093/femsre/fuw038.
 23. Scherlach K, Hertweck C. Mediators of mutualistic microbe–microbe interactions. *Nat Prod Rep*. 2018;35:303–308. doi:10.1039/C7NP00035A.
 24. Nadell CD, Drescher K, Foster KR. Spatial structure, cooperation and competition in biofilms. *Nat Rev Microbiol*. 2016;14:589–600.
 25. McFarland DJ. Decision-making in animals. *Nature*. 1977;269:15–21. doi:10.1038/269015a0.
 26. Parker GA, Smith MJ. Optimality theory in evolutionary biology. *Nature*. 1990;348:27–33. doi:10.1038/348027a0.
 27. Axelrod R, Hamilton WD. The evolution of cooperation. *Science*. 1981;211:1390–1396. doi:10.1126/science.7466396.
 28. Kim MS, Zhang H, Shim WB. Application of game theory to explore the dynamics of host–pathogen association in phytobiomes. *Phytobiomes*. 2018;2:111–116. doi:10.1094/PBIOMES-04-18-0019-P.
 29. Nordbotten JM, Levin SA, Szathmáry E, Stenseth NC. Ecological and evolutionary dynamics of interconnectedness and modularity. *PNAS*. 2018;115:750–755. doi:10.1073/pnas.1716078115.
 30. Wu A, Ross D. Evolutionary game between commensal and pathogenic microbes in intestinal microbiota. *Game*. 2016;7(3):26. doi:10.3390/g7030026.
 31. Barraclough TG. How do species interactions affect evolutionary dynamics across whole communities? *Annu Rev Ecol Evol Syst*. 2015;46:25–48. doi:10.1146/annurev-ecolsys-112414-054030.
 32. Stachowicz JJ. Mutualism, facilitation, and the structure of ecological communities: positive interactions play a critical, but underappreciated, role in ecological communities by reducing physical or biotic stresses in existing habitats and by creating new habitats on which many species depend. *BioScience* 2001;51:235–246.

33. Herbert-Read JE, Perna A, Mann RP, Schaerf TM, Sumpter DJ, Ward AJ. Inferring the rules of interaction of shoaling fish. *Proc Natl Acad Sci USA*. 2011;108:18726–18731. doi:10.1073/pnas.1109355108.
34. Hoare DJ, Krause J, Peuhkuri N, Godin JGJ. Body size and shoaling in fish. *J Fish Biol*. 2000;57(6):1351–1366. doi:10.1111/j.1095-8649.2000.tb02217.x.
35. Chance MRA, Larsen RR. The social structure of attention. New York: John Wiley & Sons. 1976.
36. Desjardins JK, Hofmann HA, Fernald RD, de Polavieja GG. Social context influences aggressive and courtship behavior in a cichlid fish. *PLoS ONE*. 2012;7(7):e32781. doi:10.1371/journal.pone.0032781.
37. Romensky M, Herbert-Read JE, Ward AJW, Sumpter DJT. Body size affects the strength of social interactions and spatial organization of a schooling fish (*Pseudomugil signifer*). *R Soc Open Sci*. 2017;4:161056. doi:10.1098/rsos.161056.
38. Jiang LB, Xu J, Zhang Y, Sang MM, Ye MX, Wei K, Xu P, Tai R, Zhao Z, Jiang YL, et al. 2019. Drive to driven: how genes play a part in intraspecific interaction networks. *iScience*.
39. Zwietering MH, Jongenburger I, Rombouts FM, van Riet K. Vant Riet K. Modeling of the bacterial growth curve. *Appl Environ Microbiol*. 1990;56:1875–1881. doi:10.1128/AEM.56.6.1875-1881.1990.
40. Balbona K, Tran H, Godyna S, Ingham KC, Strickland DK, Argraves WS. Fibulin binds to itself and to the carboxyl-terminal heparin-binding region of fibronectin. *J Biol Chem*. 1992;267:20120–20125.
41. Nymann-Andersen J, Wang H, Chen L, Kittler JT, Moss SJ, Olsen RW. Subunit specificity and interaction domain between GABA(A) receptor-associated protein (GABARAP) and GABA(A) receptors. *J Neurochem*. 2002 March;80:815–823. doi:10.1046/j.0022-3042.2002.00762.x.
42. Nagase T, Ishikawa K, Suyama M, Kikuno R, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O. Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res*. 1998;5:277–286.
43. Zhang L, Mabwi HA, Palange NJ, Jia R, Ma J, Bah FB, Sah RK, Li D, Wang D, Bah FBM, et al. Expression patterns and potential biological roles of Dip2a. *PLoS ONE*. 2015;10(11):e0143284.
44. Mohseni M, Cidado J, Croessmann S, Cravero K, Cimino-Mathews A, Wong HY, Scharpf R, Zabransky DJ, Abukhdeir AM, Garay JP, et al. MACROD2 overexpression mediates estrogen independent growth and tamoxifen resistance in breast cancers. *Proc Natl Acad Sci U S A*. 2014;111:17606–17611.
45. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol*. 2016;14:20–32. doi:10.1038/nrmicro3552.
46. Guo X, Boedicker JQ. The contribution of high-order metabolic interactions to the global activity of a four-species microbial community. *PLoS Comput Biol*. 2016;12(9):e1005079. doi:10.1371/journal.pcbi.1005079.
47. Ito H, Morishita R, Mizuno M, Kawamura N, Tabata H, Nagata KI. Biochemical and morphological characterization of a neurodevelopmental disorder-related mono-ADP-ribosylhydrolase, MACRO domain containing 2. *Dev Neurosci*. 2018;40:278–287. doi:10.1159/000492271.
48. Sakthianandeswaren A, Parsons MJ, Mouradov D, MacKinnon RN, Catimel B, Liu S, Palmieri M, Love C, Jorissen RN, Li S, et al. MACROD2 haploinsufficiency impairs catalytic activity of PARP1 and promotes chromosome instability and growth of intestinal tumors. *Cancer Discov*. 2018;8:988–1005. doi:10.1158/2159-8290.CD-17-0909.
49. De Angelis M, Francavilla R, Piccolo M, De GA, Gobetti M. Autism spectrum disorders and intestinal microbiota. *Gut Microbes*. 2015;6:207–213. doi:10.1080/19490976.2015.1035855.
50. Whisner CM, Aktipis AC. The role of the microbiome in cancer initiation and progression: how microbes and cancer cells utilize excess energy and promote one another's growth. *Curr Nutr Rep*. 2019;8:42–51. doi:10.1007/s13668-019-0257-2.
51. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou W, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *Isme J*. 2016;10(11):2557–2568. doi:10.1038/ismej.2016.45.
52. Horvath S, Dong J, Miyano S. Geometric interpretation of gene coexpression network analysis. *PLoS Comp Biol*. 2008;4:e1000117. doi:10.1371/journal.pcbi.1000117.
53. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks*. 1978;1:215–239. doi:10.1016/0378-8733(78)90021-7.
54. Brandes U. A faster algorithm for betweenness centrality. *J Math Sociol*. 2001;25:163–177. doi:10.1080/0022250X.2001.9990249.
55. Hage P, Harary F. Eccentricity and centrality in networks. *Soc Networks*. 1995;17:57–63. doi:10.1016/0378-8733(94)00248-9.
56. Bonacich P. Some unique properties of eigenvector centrality. *Soc Networks*. 2007;29:555–564. doi:10.1016/j.socnet.2007.04.002.
57. Page L The PageRank citation ranking: bringing order to the web. Stanford Digital Libraries Working Paper. 1998;9:1–14.
58. Huo L, Lv C, Fei SM, Zhou D. The correlation analysis of discrete variables and continuous variables based on mutual information. *Appl Mech Mater*. 2012;121-126:4203–4207. doi:10.4028/www.scientific.net/AMM.121-126.4203.
59. Wright S. Path coefficients and path regressions: alternative or complementary concepts. *Biometrics*. 1960;16:189–202. doi:10.2307/2527551.