**BRIEFINGS IN BIOINFORMATICS.** VOL 13. NO 6. 769–780
Advance Access published on 15 May 2012
doi:10.1093/bib/bbs022

# Computational systems biology and *in silico* modeling of the human microbiome

*Elhanan Borenstein*

Submitted: 22nd November 2011; Received (in revised form): 10th April 2012

## Abstract

The human microbiome is a complex biological system with numerous interacting components across multiple organizational levels. The assembly, ecology and dynamics of the microbiome and its contribution to the development, physiology and nutrition of the host are clearly affected not only by the set of genes or species in the microbiome but also by the way these genes are linked across numerous pathways and by the interactions between the various species. To date, however, most studies of the human microbiome have focused on characterizing the composition of the microbiome and on comparative analyses, whereas significantly less effort has been directed at elucidating, characterizing and modeling these interactions and on studying the microbiome as a complex, interconnected and cohesive system. Here, specifically, I highlight the pressing need for the development of predictive system-level models and for a system-level understanding of the microbiome, and discuss potential computational frameworks for metagenomic-based modeling of the microbiome at the cellular, ecological and supra-organismal level. I review some preliminary attempts at constructing such models and examine the challenges and hurdles that such modeling efforts face. I also discuss possible future applications and research avenues that such metagenomic systems biology and predictive system-level models may facilitate.

***Keywords:*** *systems biology; metabolic models; human microbiome; metagenomics; reverse ecology; ecosystomics*

## INTRODUCTION

Biological systems are inherently complex, whether they are molecular mechanisms in a living cell, neuronal circuits in the brain or populations of organisms in an ecosystem. This complexity is mostly encapsulated not in the components that constitute the system but rather in the intricate and often highly nonuniform way these numerous components are linked to one another and interact to produce an emergent phenotype. The human microbiome is no different and in many ways embodies a canonical example of a multilevel and hierarchical complex system.

Research of the microbiome in recent years has focused mainly on determining its composition and on examining the variation in composition across a wide range of states. To this end, next-generation sequencing and metagenomics have been used to map both the set of species and the set of genes in numerous microbiome samples. These studies can be generally classified into three main categories. First, much effort has been invested in characterizing normal compositional variation and in better defining the scope of the normal microbiome. Tremendous variation has been observed across time [1–3], across body sites [2–4] and across hosts [5]. Second, multiple studies have examined the response of the microbiome to extreme disruption and its resilience to abrupt population changes by surveying compositional shifts that follow various types of

**Corresponding author.** Elhanan Borenstein, Department of Genome Sciences, University of Washington, 3720 15th Avenue NE, Box 355065, Seattle, WA 98195-5065, USA. Tel: +1 206 685 8165; Fax: +1 206 685 7301; E-mail: elbo@uw.edu
**Elhanan Borenstein** is an Assistant Professor of Genome Sciences and an Adjunct Assistant Professor of Computer Science and Engineering at the University of Washington. He is also an External Professor at the Santa Fe Institute. His research interests include computational and evolutionary systems biology, complex networks and *in silico* modeling, and the application of these techniques to study complex biological systems with an emphasis on the human microbiome.

© The Author 2012. Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

perturbations [6]. Third, and most importantly, extensive efforts have been invested in identifying significant associations between the composition of the microbiome and a variety of diseases or host phenotypes [7–9].

These studies and the characterization of the microbiome across multiple states are clearly a crucial first step in studying the human microbiome and its role in human health, ultimately leading to a better and more profound understanding of the microbiome. To date, however, such studies have emphasized mostly the 'parts list' of the microbiome and have often overlooked the web of interactions between these parts and the complex system-level organization of the microbiome. Clearly, molecular and metabolic interactions between multiple genes within the same microbial cell or across different cells, population-level interactions between microbes of a certain species and ecological interactions between the numerous species comprising the microbiome, all play a key role in the assembly, activity and dynamics of the microbiome and in its impact on the host [10]. Moving beyond the 'bag-of-genes' or 'bag-of-species' viewpoint of comparative studies and accounting also for these interactions is therefore essential for a complete system-level understanding of the microbiome.

Systems biology research has been instrumental in the modern postgenomic era [11]. It has revolutionized the study of complex biological systems, focusing on integrative analysis rather than on a reductionist paradigm and on discovering system-level emergent properties [12, 13]. Specifically, in studying microbial species, genome-scale system-level models have successfully predicted cellular behavior and have suggested novel design principles in various bacterial systems [14, 15]. Extrapolating this approach from single-species to multi-species microbial systems, therefore, represents a promising opportunity and may provide new insights into the workings of various microbial communities and specifically into the function and dynamics of the human microbiome. Surprisingly, however, systems biology research has not yet been extensively applied to the study of the human microbiome and research combining metagenomic data with systems biology methods or with system-level models are lacking. This gap may reflect our limited knowledge of the human microbiome prior to the metagenomic revolution. However, at present, with the massive expansion of genomic and metagenomic data, the explosion of studies examining various aspects of the human microbiome and the advent of computational systems-based methods, the time is ripe for introducing systems biology inspired methods into the study of the microbiome and for developing a system-level predictive understanding of its function [16, 17].

In this article, I underline the need to apply systems biology research to study the microbiome and focus on one promising direction, namely, the construction and analysis of *in silico* system-level metabolic models. I will first briefly review various methods for modeling microbial metabolism that have been extremely fruitful in elucidating the metabolic capacity of microbial species. I will specifically highlight the 'reverse-ecology' concept—an emerging paradigm for inferring species' habitats and ecology from genome-scale metabolic models—which is particularly fitting for studying the microbiome. I will then present two approaches for expanding such modeling frameworks to model microbial communities (Figure 1). First, I will discuss multi-species models that can be used to predict and study metabolic interactions between the various species in a microbial community, and then I will present an alternative approach, modeling the entire community as a single supra-organism. I will finally discuss future directions and potential applications of such models for biomedical research.

## *IN SILICO* MODELS OF MICROBIAL METABOLISM

The construction, study and analysis of *in silico* models of microbial metabolism have proved extremely fruitful over the past few years. As genomic data accumulate, modeling efforts go beyond specific metabolic pathways and focus on whole-cell metabolism and genome-scale analysis [18, 19]. Models can be reconstructed and analyzed using a wide range of modeling frameworks, including topological models, kinetic models, stochastic models and constraint-based methods [15]. These frameworks vary markedly in the amount and type of data required to reconstruct the model, the set of assumptions on which the model is based, the analysis method and the capacity to make accurate predictions. For additional details on modeling microbial metabolism, see Refs [18–21]. In this review, I will focus primarily on two modeling frameworks, 'topology-based' models and 'constraint-based' models. These two

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

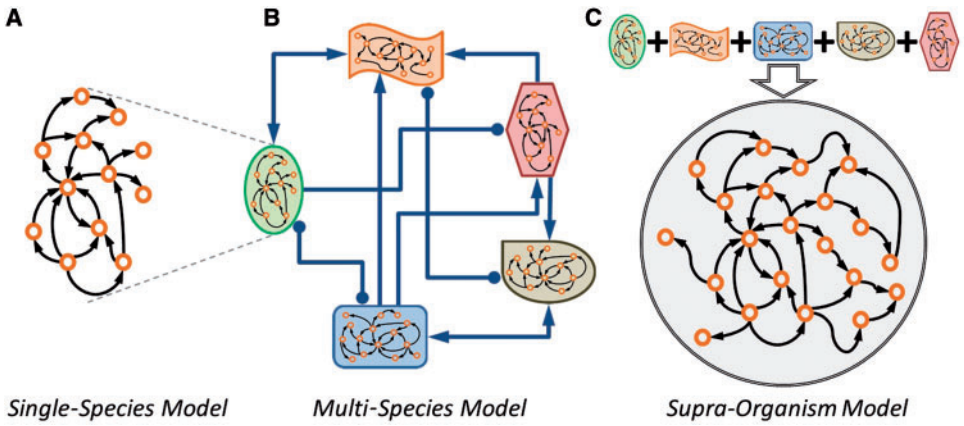


**Figure 1:** *In silico* models of the human microbiome. (**A**) Single-species metabolic models represent the set of chemical reactions that take place within the boundaries of a single cell. Here, a simple connectivity-based model is illustrated where nodes represent metabolites and edges connect substrates to products. (**B**) Multiple single-species models can be integrated into an ecosystem community model. Using various computational frameworks (such as reverse ecology), the type and magnitude of the interactions within the community can be predicted, inferring, for example, whether pairs of species directly compete for nutrients and hence hinder the growth of one another (illustrated as circle-headed arrows) or cooperate via cross-feeding (illustrated as pointed arrows). (**C**) Alternatively, supra-organism models can be reconstructed, ignoring boundaries between species altogether and modeling community-level metabolism. In many cases, such models are a necessity since genomic information is not available for all the species in the community. Supra-organism models can then be reconstructed directly from shotgun metagenomic data.

frameworks are the ones most commonly used to model microbial metabolism and are most likely to make a significant contribution to the study of the human microbiome. Specifically, both these frameworks proved extremely powerful in offering novel insights into microbial behavior and in successfully predicting various aspects of microbial function. Moreover, these frameworks do not require hard-to-obtain kinetic parameters, which are clearly not available for recently characterized microbiome species.

## Topology-based models

The underlying premise of any topology-based analysis of complex biological networks is that the structure and topology of a system are major determinants of the system's capacity and function [22]. Analyzing the topology of a biological system may therefore provide valuable insights into its behavior and help to explain observed phenotypes. In the context of metabolism, a large body of work has focused on analyzing the topology of simple network-based models that describe the metabolic process in a given species. The construction of such models often relies on an automated homology-based computational inference of the set of enzymatic genes in a given genome and the derived set of biochemical reactions that an organism can potentially catalyze. This process relies heavily on cross-species metabolic databases such as KEGG [23] or MetaCyc [24]. The set of reactions can then be represented, for example, as a simple directed graph where nodes denote metabolites and edges connect substrates to products (Figure 1A). Alternative representations, including an enzyme-based graph, a bipartite graph or a hypergraph, are similarly useful and are commonly used [25]. The topology of the reconstructed network is then examined, and topological features that correlate with metabolic phenotypes are identified.

These models are clearly extreme simplifications of an organism's metabolism, taking into account only the presence or absence of various metabolic reactions and ignoring many properties of the metabolic process such as reaction stoichiometry or rate. Yet, probing the topology of such models has proved successful in gaining insights into the metabolic capacity of an organism and into a plethora of other phenotypic attributes [26–30]. The most significant advantage of these models, however, is the relative ease with which they can be reconstructed and the scale on which they can be obtained [31], facilitating large-scale and cross-species analysis.

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

## Constraint-based models

Constraint-based models aim to define the total set of constraints that govern metabolic fluxes in a given species [15]. These include mass balance (stoichiometric constraints), reaction reversibility (thermodynamic constraints) or limitations on the capacity of each reaction. The reconstruction of genome-scale constraint-based models usually involves four major steps [32]: (i) metabolic gene annotation (as in the construction of topology-based models); (ii) manual curation according to the literature and conversion into a mathematical model; (iii) validation of the model's predictions by comparison to phenotypic data; and (iv) improvement of the model by cycling through computational and experimental work. Such manually curated, high-quality models accordingly entail an extremely involved process and follow a meticulous reconstruction protocol [33], and are therefore bound in scale. An automated framework for the generation and optimization of such models has recently been put forward [34], suggesting a possible large-scale alternative.

Given the set of constraints defined by the model, several methods have been developed to explore the space of allowable solutions and to predict specific metabolic solutions that the cell may exhibit. For example, 'pathway analysis' methods examine the range and variability of flux distribution by considering all possible flux pathways in a given network [35]. Predicting a single solution often relies on the assumption that metabolic regulation evolved to produce metabolic fluxes that maximize the organism's growth [21]. Specifically, flux balance analysis (FBA) predicts fluxes across the model such that a 'biomass' objective function is optimized [36]. Such constraint-based models and pertinent analysis methods have repeatedly generated accurate predictions concerning the growth and activity of an organism under various environmental and genetic conditions [21].

## Reverse ecology

The studies described above generally use metabolic models to infer metabolic function and dynamics. As stated above, much of systems biology relies on the assumption that the organization of systems captures their function. However, as systems adapt to their environments, their structure clearly reflects not only their capacity but also the environment in which they evolved. Recent studies, for example, have revealed a marked association between the modularity of metabolic networks across a large collection of microbial species and the level of variability in their environments [28, 37]. Reverse ecology—an emerging research paradigm—aims to identify such universal structural signatures that can then be used to obtain insights into the ecology of poorly characterized species [38–40]. Specifically, systems-based reverse ecology attempts to develop computational tools for analyzing genome-scale models, characterizing the natural habitat of microbial species on a large scale and predicting the interaction of these species with their environments and with other species. For a full review of this reverse ecology research paradigm, see ref. [40].

Specifically, following this reverse ecology approach, a computational framework for analyzing genome-scale metabolic models and for inferring the set of compounds that organisms extract from their surroundings, has been introduced [38]. This computationally derived set, termed the 'seed set' of the network, was shown to accurately describe the effective biochemical environments of hundreds of microbial species, providing a proxy for their natural habitats. This framework has now been utilized to explore and characterize multiple aspects of microbial ecology, including host–parasite interactions (see also below) [41], universal strategies governing microbial metabolism [42], environmental robustness [30] and metabolic exchanges between co-resident endosymbionts [43]. A web-based tool for calculating the seed set of a network has recently been presented [44].

Reverse ecology seems a research avenue especially suited to studying the microbiome, allowing the translation of high-throughput genomic and metagenomic data into ecological data [17, 40]. The human microbiome, with the massive amounts of data that are now being generated to characterize it [4, 45] on the one hand and our limited understanding of its ecology on the other hand, represents a unique opportunity for reverse ecology research.

# MULTI-SPECIES ECOSYSTEM MODELS

As in any ecosystem, the various species inhabiting the human microbiome form a complex set of interactions [46]. A telling sign of these mutual dependencies is the difficulty (and often failure) in culturing the vast majority of these microbial species in isolation [47, 48]. Such interactions play a key role

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

in community assembly, dynamics and response to perturbation. An extensive body of work, for example, relates the organization of ecological interaction networks to ecosystem robustness [49, 50]. While such studies mostly focus on communities of macro-organisms, a similar approach can be applied to microbial communities.

## Understanding inter-species interactions

To better understand the nature and magnitude of species interactions, considerable effort has been invested in experimentally studying a variety of simple co-culture systems [10, 51]. Growth assays, for example, can provide a detailed account of the way one species affects the growth rate of another [52]. Yet, considering the enormous diversity of the human microbiota and the difficulties involved in isolating and culturing many gut dwelling species, this experimental approach is clearly limited and does not support a large scale, systematic framework for studying species interactions in the microbiome. Alternatively, to obtain a phenomenological understanding of the outcome of such interactions, numerous studies have analyzed the co-occurrence of species in the human microbiome and in other microbiomes of interest [53–56]. These studies regularly find prominent co-occurrence patterns, suggesting a nonrandom assembly of the microbiome [57] and generating hypotheses concerning the driving forces underlying these patterns. Such co-occurrence patterns, however, do not offer a mechanistic understanding of these driving forces or direct evidence for specific species interactions.

## Integrating multiple single-species models into an ecosystem model

Genome-scale *in silico* models provide an alternative approach for studying species interactions. Since such models accurately predict the behavior of various microorganisms and their interactions with the environment, they seem ideal for modeling and studying the interactions between the various species in a community of microbial organisms. Integrating multiple single-species models that represent the community members and developing computational methods for inferring interspecific metabolic dependencies will allow us to reconstruct predicted microbial 'food webs' (Figure 1B), similar to the ones commonly used to describe macro-organisms' ecosystems [50, 58]. Surprisingly, however, to date very few models of multi-species microbial systems have been introduced [21], presenting mostly simple *in silico* models of microbe–microbe or microbe–host interactions.

Specifically, applying topology-based analysis, Christian *et al.* [59] examined the level of cooperation between microbial organisms, using the network expansion algorithm [60] to compare the biosynthetic capabilities of a two-species unified model with the biosynthetic capabilities of each species in isolation. Extending the single-species reverse-ecology framework described above, Borenstein and Feldman [41] introduced a pair-wise, topology-based measure of biosynthetic support, assessing the extent to which the nutritional needs of a putative endosymbiont can be provided by a host and facilitating the prediction of such interactions on a large scale. Freilich *et al.* [61] further extended this framework, introducing a topology-based measure of competition and examining its correlation with the co-occurrence of microbial species in scientific literature.

Additional attempts have been made to apply constraint-based modeling to various two-species systems. Stolyar *et al.* [62] introduced one of the first examples of such a two-species stoichiometric metabolic model, using fully sequenced genomes of *Desulfovibrio vulgaris* and *Methanococcus maripaludis*, to analyze the mutualistic interactions between sulfate-reducing bacteria and methanogens. Bordbar *et al.* [63] constructed and examined a host–pathogen model, integrating a *Mycobacterium tuberculosis* model with that of the human alveolar macrophage, and simulated the metabolic changes during infection. Zhuang *et al.* [64] modeled the competition between Rhodoferax and Geobacter species under diverse conditions. Wintermute and Silver [65] used a stoichiometric model of interacting strains and examined >1000 pairs of auxotrophic *Escherichia coli* mutant strains, focusing on the prevalence of synthetic mutualism. Finally, Klitgord and Segre [66] extended the FBA framework and considered pair-wise combinations of seven microbial species to examine how the environment in which two species are placed affects their metabolic interaction.

These constraint-based modeling studies have applied a variety of approaches to combine single-species stoichiometric models into an ecosystem model and to account for the transfer of metabolites between the two species and between the species and the environment. Yet, a standard framework for constructing and analyzing stoichiometric ecosystem models is still lacking and several conceptual

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

challenges should be addressed before such a modeling framework can be introduced. For example, it is not clear whether some of the assumptions often associated with constraint-based modeling approaches (including the optimality of the biomass objective function) still hold for community metabolism (see, for example, the detailed discussion in ref. [67]).

An additional property that repeatedly appears as a key element in constructing such multi-species models is compartmentalization—the partitioning of reactions and metabolites between different compartments in the model that may represent different organelles or different species. Klitgord and Segre [68], for example, investigated how compartmentalization affects genome-scale flux balance models, comparing a compartmentalized model of yeast organelles to a de-compartmentalized version. Taffs *et al.* [69] examined three different compartmentalization schemes to study mass and energy flows in microbial communities. The first scheme utilized a compartmentalized model wherein each species occupied a distinct compartment as in some of the studies described above [62]. The second scheme, referred to as the pooled reactions model, is similar in spirit to the supra-organismal approach discussed later in this article. Finally, a third scheme, termed a nested consortium analysis, employed two rounds of processing, the first operating on individual models and the second on manually selected ecologically interesting pathways. Taffs *et al.* compared these schemes using models based on three distinct microbial guilds, identifying specific advantages and disadvantages of each scheme and highlighting criteria for selecting a modeling approach appropriate for a given microbial system.

More generally, additional research is still required in order to make topology- and constraint-based models applicable to the human microbiome. The above studies were mostly done on a very small scale, considering simple two-species models, and focusing on pair-wise metabolic interactions. With hundreds of species comprising the gut microbiota, it is necessary to scale such models up. Many more species should be modeled and analyzed to allow sufficient coverage of the species comprising the microbiome and to provide a solid infrastructure for modeling it. Combining such multi-species genome-based models with data on the co-occurrence of species in various samples can reveal forces driving community assembly. Furthermore, since the type and extent of the interaction between two species may strongly depend on the presence of other species in the environment or on additional contextual factors, modeling frameworks should go beyond two-species interactions. Finally, the majority of these studies take a static view of metabolism, assuming an overall metabolic steady state or addressing questions concerning metabolic potential rather than specific dynamic activation of various metabolic pathways. These studies further assume a static community composition and fixed relative abundances of the interacting species. The introduction of dynamics, both molecular and ecological, is essential for making these modeling frameworks accurate and useful for addressing questions concerning observed temporal patterns in the microbiome, its assembly and its dynamic response to perturbations. Preliminary studies, incorporating molecular and ecological dynamics, have only recently been introduced [64, 70] and much work is still ahead before a full comprehensive framework is available.

## SUPRA-ORGANISM MODELS

The metabolism of the microbiome is a complex composite of the metabolic activity of numerous microbial cells from many different species. Each species (and in fact, each strain) in the microbiome encodes a unique set of metabolic functions with unique metabolic capacities. Accordingly, microbial cells of different species represent, in reality, compartmentalized metabolic units, with specific limitations on the transfer of metabolites across cell membranes [68, 69]. The models described above attempt, to varying degrees, to capture this compartmentalization as well as the autonomous nature of each species.

One can, however, apply an alternative and fundamentally different modeling approach and model the entire microbiome as a single supra-organism [71, 72]. This abstraction is in fact common practice in comparative metagenomic analyses where the entire set of genes found via shotgun metagenomics is studied as a proxy for the microbiome's capacity and metabolic strategy [73]. Such a gene-centric approach completely ignores the gene's species of origin, and is regularly used, for example, to compare the set of genes found in microbiome samples associated with different host states. In the context of the human gut microbiome, this abstraction is further justified by the relative consistency of the functions

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

encoded in the microbiome, compared to a much higher variation at the species level [7], suggesting microbiome-level niche adaptation.

Following this supra-organismal approach, *in silico* models of the microbiome can be reconstructed directly from shotgun metagenomic data, representing community-level metabolism and totally ignoring cell boundaries or the shuttling of metabolites across species (Figure 1C). Such metagenomics-based models were used in a recent study to identify system-level variation associated with obesity and inflammatory bowel disease (IBD) [74]. Reconstructing community-level metabolic networks of the gut microbiome and projecting shotgun metagenomic data onto these networks, this study demonstrated that most of the enzymatic genes that are differentially abundant in obese (or IBD) microbiomes tend to be located at the periphery of the metabolic network. This suggests that obesity and IBD are associated with modifications in the way the microbiome interacts with the gut environment rather than variation in core metabolism. Furthermore, using these community-level metabolic networks, it was shown that obese microbiome models are less modular than models representing lean microbiomes, a characteristic feature of adaption to a low-variability environment.

The study discussed above highlights the promise of 'metagenomic systems biology' research. More generally, the motivation for studying such supra-organism models is 3-fold: First and foremost, since many gut dwelling species resist isolation and sequencing, such community-level models are often a necessity. These models allow us to consider the entire set of genes found in a microbiome sample even when the species encoding some of these genes remain elusive. Second, while multi-species models are especially fitting for studying the ecology of the microbiome and the interactions between its various members, supra-organism models seem uniquely apt for studying the activity of the microbiome as a whole and specifically the interaction between the microbiome and the host. Such models can be used, for example, to examine possible exchange of metabolites between the community and the gut environment. Furthermore, integrating such models with models of human metabolism [75, 76] will allow us to study the metabolic dependencies between the gut microbiome and the host in an analogous manner to the study of the interaction between a single microbial endosymbiont and its host [41]. Finally, in principle, supra-organism models can be reconstructed and analyzed using the diverse toolset developed to study single species metabolic models, and therefore have tremendous potential for elucidating fundamental questions concerning community metabolism.

To date, however, studies of supra-organism models are still scarce and further development is needed before a fully comprehensive framework for metagenomic systems biology is introduced. For example, it is not clear to what degree the identified principles and observations from genome-scale single species models can be extrapolated to metagenomic-scale supra-organism models. Metagenomic coverage is another hurdle since rare, but potentially important functions may be missed. Yet, even with these limitations, as shotgun metagenomic data continue to accumulate for both the human microbiome and many other microbiomes of interest, supra-organism models are an increasingly attractive alternative to genome-scale models and are bound to lead to exciting discoveries and to a better understanding of the microbiome.

## FUTURE CHALLENGES AND OPPORTUNITIES

The various studies discussed so far represent the first exciting steps toward the development of a complete systems-biology framework for studying the human microbiome. Much work remains to be done to make each of the two modeling approaches outlined above (namely, multi-species models and supra-organism models) a viable and comprehensive modeling framework. Probably, the most daunting task is to combine these two modeling schemes into a single unified framework, bridging the gap between genome-based single-species models and metagenomics-based supra-organism models and introducing a multiscale model of the microbiome from the molecular to the ecological level.

Moreover, these conceptual modeling approaches and the specific computational techniques described above are clearly not the only routes for modeling the microbiome. System-level models can be constructed with varying levels of abstraction and analyzed using a wide range of computational methods. The choice of a specific modeling framework should be informed by both the type of data available and, more importantly, the questions one wishes to address.

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

Successful system-level modeling often relies on abundant data of multiple types. Genomic and metagenomic sequencing are therefore essential for the continued development of better and more accurate models. Specifically, accurate characterization of species composition (e.g. using 16S as the chosen marker) and genomic data on member species are important for correctly modeling the microbiome ecosystem. Currently, more than a thousand human-associated genomes have been sequenced (mostly by the Human Microbiome Project) and hundreds of additional microbiome species will be sequenced soon. These species may still reflect a potentially small set of species compared to the number of species inhabiting the human body and forthcoming advances in next-generation sequencing technologies are predicted to further make thousands of microbial genomes available. As more human-associated reference genomes become available, ecosystem models could be scaled up to eventually account for most of the relevant species in the microbiome. Similarly, high-coverage shotgun metagenomic data will promote improved supra-organism models that accurately capture the metabolic capacity of the community. Modeling efforts should in turn inform data collection, providing testable predictions and pointing to genes and species of interest. Multi-species ecosystem models, for example, can identify putative keystone species and prioritize sequencing pipelines [77].

As next generation sequencing continues to improve and new technologies emerge, the rate at which genomic and metagenomic data accumulate will further increase. These technologies clearly pose huge informatics challenges [78], many of which arise from the short read length generated by platforms such as Illumina and SOLiD. Specifically, in the context of metabolic modeling, accurate annotation is key for the successful reconstruction of both multi-species and supra-organism models. Further progress in these technologies therefore must go hand in hand with the development of novel algorithms for the assembly, mapping and annotation of billions of short reads, facilitating high-resolution functional characterization of microbiomes and ultimately the assembly of full microbial genomes directly from shotgun metagenomic data [79]. A partial list of relevant online resources useful for system-level metabolic modeling of the microbiome can be found in Table 1.

Systems-based analysis and *in silico* modeling of the microbiome are of course not limited to genomic or metagenomic data. Other types of data could potentially advance both the construction and the validation of such models. Specifically, metatranscriptomic and meta-metabolomic data will further improve modeling frameworks, providing a more precise characterization of the metabolic activity of the microbiome and the availability of metabolites in the environment [16, 93]. Additional data is also required for the application of specific modeling frameworks. For example, biomass composition and uptake rates should be characterized for each species in order to accurately construct species-specific constraint-based models. Such data are clearly challenging to obtain considering the difficulties associated with the efforts to culture many microbiome-related species. Ultimately, the research approach laid out in this article aims to provide system-level 'predictive' models of the human microbiome. Such models lay the foundation for numerous exciting applications, some of which are described below.

## DESIGNER MICROBIOMES AND ECOSYSTOMICS

A predictive system-level model of a complex system is often considered a touchstone of our understanding of that system. Indeed, a model of the human microbiome, capable of inferring the activity of the microbiome, its dynamics, and its impact on the host directly from species and gene composition data, would definitely indicate a principled understanding of the microbiome far beyond our current knowledge. Here, however, two potential applications that can be derived from such an ideal model are highlighted. These applications are obviously still out of reach and much work still lies ahead before they become a reality. Yet, they demonstrate the tremendous potential of systems-based microbiome research and some of the overarching goals of such research.

First, an accurate predictive model is an essential step in developing a framework for designing and directing bacteriotherapy. Bacteriotherapy, namely the modulation of one's microbiota via antibiotics and probiotics or the transplantation of a complete microbiota into a recipient, is an exciting clinical frontier [94]. Microbiome transplantation was recently shown, for example, to successfully resolve recurrent *Clostridium difficile* infection, re-establishing a normal and stable microbiota [95–97]. Related

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

**Table 1:** Useful online resources for systems biology and modeling of the human microbiome

| Resources | References |
|---|---|
| Microbial genomic data and analysis | |
| IMG | [80] |
| DACC | [81] |
| GOLD | [31] |
| Microbes online | [82] |
| RAST | [83] |
| Metagenomic data and analysis | |
| IMG/M | [84] |
| MG-RAST | [85] |
| METAREP | [86] |
| Metabolic databases | |
| KEGG | [23] |
| MetaCyc | [24] |
| Brenda | [87] |
| Metabolic model reconstruction, visualization and analysis | |
| The Model Seed | [34] |
| Systems Biology Research Group | [88] |
| iPath | [89] |
| Pathway Tools | [90] |
| Cytoscape | [91] |
| Cobra | [92] |
| Reverse ecology software | |
| NetSeed | [44] |

technologies, such as personalized microbiota collections and germ-free mice models, are also being developed [98, 99]. These efforts, however, are generally uninformed, utilizing a microbiota from a healthy donor and transplanting it, as is, into a sick recipient. Conversely, a predictive microbiome model can be integrated with an optimization framework to identify precise manipulations that could be applied to a given microbiota in order to derive a stable compositional shift and promote some predefined metabolic activity. Alternatively, such an integrated framework could be used for designing novel microbiomes, devising 'recipes' for reconstructing stable communities with some desired metabolic activity by mixing and matching available species at certain relative abundances. 'Designer' microbiomes can offer a therapeutic route for treating numerous diseases ranging from obesity, diabetes and inflammatory bowel disease to diarrhea and acute gastroenteritis or for promoting energy harvest in populations of undernourished children [100].

Second, taking a more theoretical perspective, a predictive microbiome model can be used to study the 'ecology of the possible' and to characterize the contours of the space of possible ecosystem configurations. Of specific interest is the mapping from microbiota composition to community-level metabolism and the regularities in this mapping [66]. A full characterization of this space is an important first step toward the development of 'ecosystomics'—a high-throughput systematic study of all realizable ecosystems in a given environment. Ecosystomic research can then provide a neutral model of ecology which is crucial for determining the significance of observed compositional patterns.

## CONCLUDING REMARKS

Systems biology research has already revolutionized genomics and could similarly transform metagenomic research and particularly research of the human microbiome [16]. Specifically, *in silico* models of the microbiome, aiming to capture its activity, organization and ecology, will allow us to go beyond comparative analysis and to study the microbiome as a complex, multiscale and hierarchical biological system. The statistician George E. P. Box once remarked [101]: 'All models are wrong, but some are useful.' Clearly, the models described in this article and, for that matter, any model of the microbiome that will be developed in the foreseeable future, are certain to be inaccurate and to capture only a simplified subset of this microbial system. Yet, as Box stated, some of these modeling efforts are extremely useful and hold great promise. Systems based research represents a unique opportunity for addressing several of the most pressing questions concerning the human microbiome: what determines the assembly of the microbiome and what role do interspecific interactions play in its composition? Which factors govern the response of the microbiome to various perturbations? How does the microbiome, as a whole, interact with the human host and how does it impact human health? These are fundamentally system-level questions that can be addressed only by considering system-level attributes of the microbiome and acknowledging the many interactions between the various components that comprise this complex system.

### KEY POINTS

- The human microbiome is a complex biological system—interactions between numerous genes and between the various species comprising the microbiome markedly affect its function, dynamics and impact on the host.

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

- Studying the human microbiome calls for a systems-based research and for system-level modeling, ultimately leading to a better and more profound understanding of the microbiome.
- Computational systems biology of *in silico* metabolic models proved extremely useful in studying microbial metabolism.
- Two fundamentally different approaches can be used to model microbiome metabolism: genome-based multi-species models and metagenomics-based supra-organism models.
- Preliminary studies of these modeling approaches demonstrate tremendous potential but several challenges should be addressed before a comprehensive modeling framework can be introduced.

## FUNDING

E.B. is an Alfred P. Sloan Research Fellow.

## *References*

1. Koenig JE, Spor A, Scalfone N, *et al*. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* 2011;**108(Suppl)**:4578–4585.
2. Grice EA, Kong HH, Conlan S, *et al*. Topographical and temporal diversity of the human skin microbiome. *Science* 2009;**324**:1190–1192.
3. Costello EK, Lauber CL, Hamady M, *et al*. Bacterial community variation in human body habitats across space and time. *Science* 2009;**326**:1694–1697.
4. Turnbaugh PJ, Ley RE, Hamady M, *et al*. The human microbiome project. *Nature* 2007;**449**:804–810.
5. Eckburg PB, Bik EM, Bernstein CN, *et al*. Diversity of the human intestinal microbial flora. *Science* 2005;**308**: 1635–1638.
6. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *PNAS* 2011;**108**: 4554–4561.
7. Turnbaugh PJ, Hamady M, Yatsunenko T, *et al*. A core gut microbiome in obese and lean twins. *Nature* 2009;**457**: 480–484.
8. Ley RE. Obesity and the human microbiome. *Curr Opin Gastroenterol* 2010;**26**:5–11.
9. Manichanh C, Rigottier L.-Gois, Bonnaud E, *et al*. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;**55**: 205–211.
10. Trosvik P, Rudi K, Straetkvern KO, *et al*. Web of ecological interactions in an experimental gut microbiota. *Environ Microbiol* 2010;**12**:2677–2687.
11. Westerhoff HV, Palsson B. The evolution of molecular biology into systems biology. *Nat Biotechnol* 2004;**22**: 1249–1252.
12. Noble D. The music of life. *Nature* 2006;**435**:2005–2005.
13. Kitano H. Systems biology: a brief overview. *Science* 2002; **295**:1662–1664.
14. Stelling J. Mathematical models in microbial systems biology. *Curr Opin Microbiol* 2004;**7**:513–518.
15. Reed JL, Palsson BØ. Thirteen years of building constraint-based in silico models of Escherichia coli. *J Bacteriol* 2003;**185**:2692–2699.
16. Raes J, Bork P. Molecular eco-systems biology: towards an understanding of community function. *Nat Rev Microbiol* 2008;**6**:693–699.
17. Röling WF, Ferrer M, Golyshin PN. Systems approaches to microbial communities and their functioning. *Curr Opin Biotechnol* 2010;**21**:532–538.
18. Durot M, Bourguignon P-Y, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev* 2009;**33**:164–190.
19. Ruppin E, Papin AJ, de Figueiredo L, *et al*. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr Opin Biotechnol* 2010;**21**:502–510.
20. Cloots L, Marchal K. Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. *Curr Opin Microbiol* 2011;**14**:599–607.
21. Oberhardt AM, Palsson BØ, Papin AJ. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 2009;**5**:320.
22. Alon U. Biological networks: the tinkerer as an engineer. *Science* 2003;**301**:1866–1867.
23. Kanehisa M, Goto S, Hattori M, *et al*. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006;**34**:D354–D357.
24. Caspi R, Altman T, Dale JM, *et al*. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 2010; **38**:D473–D479.
25. Deville Y, Gilbert D, Van Helden J, *et al*. An overview of data models for the analysis of biochemical pathways. *Brief Bioinform* 2003;**4**:246–259.
26. Jeong H, Tombor B, Albert R, *et al*. The large-scale organization of metabolic networks. *Nature* 2000;**407**: 651–654.
27. Stelling J, Klamt S, Bettenbrock K, *et al*. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 2002;**420**:190–193.
28. Kreimer A, Borenstein E, Gophna U, *et al*. The evolution of modularity in bacterial metabolic networks. *Proc Natl Acad Sci USA* 2008;**105**:6976–6981.
29. Wunderlich Z, Mirny LA. Using the topology of metabolic networks to predict viability of mutant strains. *Biophys J* 2006;**91**:2304–2311.
30. Freilich S, Kreimer A, Borenstein E, *et al*. Decoupling environment-dependent and independent genetic robustness across bacterial species. *PLoS Comput Biol* 2010; **6**:10.
31. Liolios K, Chen IM, Mavromatis K, *et al*. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010;**38**:D346–D354.
32. Oberhardt MA, Puchalka J, Fryer KE, *et al*. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J Bacteriol* 2008;**190**: 2790–2803.
33. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 2010;**5**:93–121.
34. Henry CS, DeJongh M, Best AA, *et al*. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 2010;**28**:969–974.

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

35. Papin JA, Stelling J, Price ND, *et al*. Comparison of network-based pathway analysis methods. *Trends Biotechnol* 2004;**22**:400–405.

36. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 2010;**28**:245–248.

37. Parter M, Kashtan N, Alon U. Environmental variability and modularity of bacterial metabolic networks. *BMC Evol Biol* 2007;**7**:169.

38. Borenstein E, Kupiec M, Feldman MW, *et al*. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci USA* 2008;**105**: 14482–14487.

39. Janga SC, Babu MM. Network-based approaches for linking metabolism with environment. *Genome Biol* 2008; **9**:239.

40. Levy R, Borenstein E. Reverse ecology: from systems to environments and back. *Adv Exp Med Biol* 2012;**751**.

41. Borenstein E, Feldman MW. Topological signatures of species interactions in metabolic networks. *J Comput Biol* 2009; **16**:191–200.

42. Freilich S, Kreimer A, Borenstein E, *et al*. Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol* 2009;**10**:R61.

43. Cottret L, Milreu PV, Acuña V, *et al*. Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, baumannia cicadellinicola and sulcia muelleri, with their insect host, Homalodisca coagulate. *PLoS Comput Biol* 2010;**6**:13.

44. Carr R, Borenstein E. NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. *Bioinformatics* 2012;**28**: 734–735.

45. Nelson KE, Weinstock GM, Highlander SK, *et al*. A catalog of reference genomes from the human microbiome. *Science* 2010;**328**:994–999.

46. Little AEF, Robinson CJ, Peterson SB, *et al*. Rules of engagement: interspecies interactions that regulate microbial communities. *Annu Rev Microbiol* 2008;**62**:375.

47. McInerney MJ, Sieber JR, Gunsalus RP. Syntrophy in anaerobic global carbon cycles. *Curr Opin Biotechnol* 2009;**20**: 623–632.

48. Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol Lett* 2010; **309**:1–7.

49. Dunne JA, Williams RJ, Martinez ND. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol Lett* 2002;**5**:558–567.

50. Dunne JA, Williams RJ, Martinez ND. Food-web structure and network theory: the role of connectance and size. *Proc Natl Acad Sci USA* 2002;**99**:12917–12922.

51. Shou W, Ram S, Vilar JMG. Synthetic cooperation in engineered yeast populations. *Proc Natl Acad Sci USA* 2007;**104**: 1877–1882.

52. Kolenbrander PE. Multispecies communities: interspecies interactions influence growth on saliva as sole nutritional source. *Int J Oral Sci* 2011;**3**:49–54.

53. Arumugam M, Raes J, Pelletier E, *et al*. Enterotypes of the human gut microbiome. *Nature* 2011;**473**:174–180.

54. Chaffron S, Rehrauer H, Pernthaler J, *et al*. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res* 2010;**20**: 947–959.

55. Barberán A, Bates ST, Casamayor EO, *et al*. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* 2012;**6**:343–351.

56. Horner-Devine M, Silver JM, Leibold AM, *et al*. A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 2007;**88**:1345–1353.

57. Cody ML, Diamond JM. *Ecology and Evolution of Communities*. Cambridge, MA: Harvard Univ Press, 1975.

58. Pascual M, Dunne JA. *Ecological Networks: Linking Structure to Dynamics in Food Webs*. New York, NY: Oxford University Press, 2006.

59. Christian N, Handorf T, Ebenhöh O. Metabolic synergy: increasing biosynthetic capabilities by network cooperation. *Genome Inform Int Conf Genome Inform* 2007;**18**:320–329.

60. Handorf T, Ebenhöh O, Heinrich R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J Mol Evol* 2005;**61**:498–512.

61. Freilich S, Kreimer A, Meilijson I, *et al*. The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res* 2010;**38**: 3857–3868.

62. Stolyar S, Van Dien S, Hillesland KL, *et al*. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* 2007;**3**:92.

63. Bordbar A, Lewis NE, Schellenberger J, *et al*. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol* 2010;**6**:422.

64. Zhuang K, Izallalen M, Mouser P, *et al*. Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments. *ISME J* 2011;**5**:305–316.

65. Wintermute EH. Silver PA, Emergent cooperation in microbial metabolism. *Mol Syst Biol* 2010;**6**:407.

66. Klitgord N, Segrè D. Environments that induce synthetic microbial ecosystems. *Ecosystems* 2010;**6**:e1001002.

67. Klitgord N, Segrè D. Ecosystems biology of microbial metabolism. *Curr Opin Biotechnol* 2011;**22**:541–546.

68. Klitgord N, Segrè D. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Inform Int Conf Genome Inform* 2010;**22**: 41–55.

69. Taffs R, Aston JE, Brileya K, *et al*. In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Syst Biol* 2009;**3**:114.

70. Tzamali E, Poirazi P, Tollis IG. A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities. *BMC Syst Biol* 2011;**5**:167.

71. Gordon JI, Klaenhammer TR. A rendezvous with our microbes. *Proc Natl Acad Sci USA* 2011;**108(Suppl)**: 4513–4515.

72. Lederberg J. Infectious history. *Science* 2000;**288**:287–293.

73. Tringe SG, von Mering C, Kobayashi A, *et al*. Comparative metagenomics of microbial communities. *Science* 2005;**308**: 554–557.

74. Greenblum S, Turnbaugh PJ, Borenstein E. Metagenomic systems biology of the human gut microbiome reveals

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012

topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci USA* 2012;**109**:594–599.

75. Duarte NC, Becker SA, Jamshidi N, *et al*. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* 2007;**104**: 1777–1782.
76. Shlomi T, Cabili MN, Herrgård MJ, *et al*. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008;**26**:1003–1010.
77. Libralato S, Christensen V, Pauly D. A method for identifying keystone species in food web models. *Ecol Model* 2006;**195**:153–171.
78. Scholz MB, Lo C-Chi, Chain PSG. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol* 2012;9–15.
79. Iverson V, Morris RM, Frazar CD, *et al*. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science* 2012;**587**.
80. Markowitz VM, Chen I-Min A, Palaniappan K, *et al*. IMG: the integrated microbial genomes database and comparative analysis system. *Database* 2012;**40**:115–122.
81. Peterson J, Garges S, Giovanni M, *et al*. The NIH Human Microbiome Project. *Genome Res* 2009;**19**:2317–2323.
82. Dehal PS, Joachimiak MP, Price MN, *et al*. MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucl Acids Res* 2010;**38**:D396–D400.
83. Aziz RK, Bartels D, Best AA, *et al*. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 2008;**9**:75.
84. Markowitz VM, Ivanova NN, Szeto E, *et al*. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 2008;**36**:D534–D538.
85. Meyer F, Paarmann D, D'Souza M, *et al*. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* 2008;**9**:386.
86. Goll J, Rusch DB, Tanenbaum DM, *et al*. METAREP: JCVI metagenomics reports—an open source tool for high-performance comparative metagenomics. *Bioinformatics* 2010;**26**:2631–2632.
87. Scheer M, Grote A, Chang A, *et al*. BRENDA, the enzyme information system in 2011. *Database* 2011;**39**:670–676.
88. Feist AM, Herrgård MJ, Thiele I, *et al*. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 2009;**7**:129–143.
89. Letunic I, Yamada T, Kanehisa M, *et al*. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* 2008;**33**:101–103.
90. Karp PD, Paley SM, Krummenacker M, *et al*. Pathway Tools version 13. 0: integrated software for pathway/genome informatics and systems biology. *Access* 2009;**11**: 40–79.
91. Smoot ME, Ono K, Ruscheinski J, *et al*. Cytoscape 2. 8: new features for data integration and network visualization. *Bioinformatics* 2011;**27**:431–432.
92. Schellenberger J, Que R, Fleming RMT, *et al*. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature Protocols* 2011;**6**:1290–1307.
93. Warnecke F, Hess M. A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *J Biotechnol* 2009;**142**:91–95.
94. Collins MD, Gibson GR. Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *Am J Clin Nutr* 1999;**69**:1052S–1057S.
95. Khoruts A, Sadowsky MJ. Therapeutic transplantation of the distal gut microbiota. *Mucosal Immunol* 2011;**4**:4–7.
96. Khoruts A, Dicksved J, Jansson JK, *et al*. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol* 2010;**44**:354–360.
97. Grehan MJ, Borody TJ, Leis SM, *et al*. Durable alteration of the colonic microbiota by the administration of donor fecal flora. *J Clin Gastroenterol* 2010;**44**:551–561.
98. Goodman AL, Kallstrom G, Faith JJ, *et al*. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *PNAS* 2011;**108**:6252–6257.
99. Faith JJ, Rey FE, O'Donnell D, *et al*. Creating and characterizing communities of human gut microbes in gnotobiotic mice. *ISME J* 2010;**4**:1094–1098.
100. Preidis GA, Versalovic J. Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era. *Gastroenterology* 2009;**136**: 2015–2031.
101. Box GEP, Draper NR. *Empirical Model-Building and Response Surfaces*. New York: Wiley, 1987.

Downloaded from http://bib.oxfordjournals.org/ at University of Washington on December 21, 2012