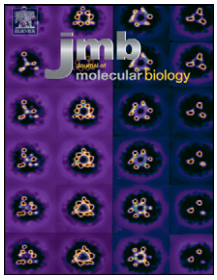

Towards Predictive Models of the Human Gut Microbiome

Vanni Bucci[1] **and Joao B. Xavier**[2]

*1 - **Department of Biology,** University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA*
*2 - **Program in Computational Biology,** Memorial Sloan–Kettering Cancer Center, 1275 York Avenue, Box 460, New York, NY 10065, USA*

Correspondence to Vanni Bucci and Joao B. Xavier: *vbucci@umassd.edu; xavierj@mskcc.org*
http://dx.doi.org/10.1016/j.jmb.2014.03.017
Edited by M. Eric

Abstract

The intestinal microbiota is an ecosystem susceptible to external perturbations such as dietary changes and antibiotic therapies. Mathematical models of microbial communities could be of great value in the rational design of microbiota-tailoring diets and therapies. Here, we discuss how advances in another field, engineering of microbial communities for wastewater treatment bioreactors, could inspire development of mechanistic mathematical models of the gut microbiota. We review the state of the art in bioreactor modeling and current efforts in modeling the intestinal microbiota. Mathematical modeling could benefit greatly from the deluge of data emerging from metagenomic studies, but data-driven approaches such as network inference that aim to predict microbiome dynamics without explicit mechanistic knowledge seem better suited to model these data. Finally, we discuss how the integration of microbiome shotgun sequencing and metabolic modeling approaches such as flux balance analysis may fulfill the promise of a mechanistic model.

Introduction

Mathematical models of multispecies microbial communities have a long tradition in design and control of environmental biotechnology processes. For more than 20 years, sophisticated mechanistic models have assisted engineers in understanding the relationship between operational parameters (e.g., flow rate, temperature and oxygenation) and microbial composition in wastewater treatment bioreactors [1]. Recent advances in metagenomics have reinforced the notion that the intestinal microbiome is composed of multiple microbial species in competition for limited nutrients and attachment sites and differentially susceptible to external perturbations, which is similar to bioreactors. Both the gut microbiota and wastewater have very high microbial biodiversity [2]. However, whereas the perturbations in bioreactors usually consist of changes in the operational variables such as flow rate of composition of the influent, in the microbiota, the perturbations are such as antibiotic treatment, changes in diet and exposure to external microbes (Fig. 1). There is a significant interest in optimal microbiome management due to its relevance to human health [3]. While most of our current insights come from experimental studies [4], it should be possible to develop mechanistically based mathematical models to assist in the design of intervention strategies, similarly to how engineers apply mathematical models in the design and management of bioreactors.

Mathematical Models in Environmental Biotechnology

Biological wastewater treatment is the process of clearing sewage water by converting the dissolved nutrients, which would otherwise cause eutrophication and bad water quality in the receiving water bodies, into microbial biomass, which can then be disposed or recycled. In these bioreactors, multispecies microbial communities grow by consuming carbon-, nitrogen- and phosphorus-rich organic compounds in the wastewater, effectively cleaning the water of these compounds. The microbial community can be very

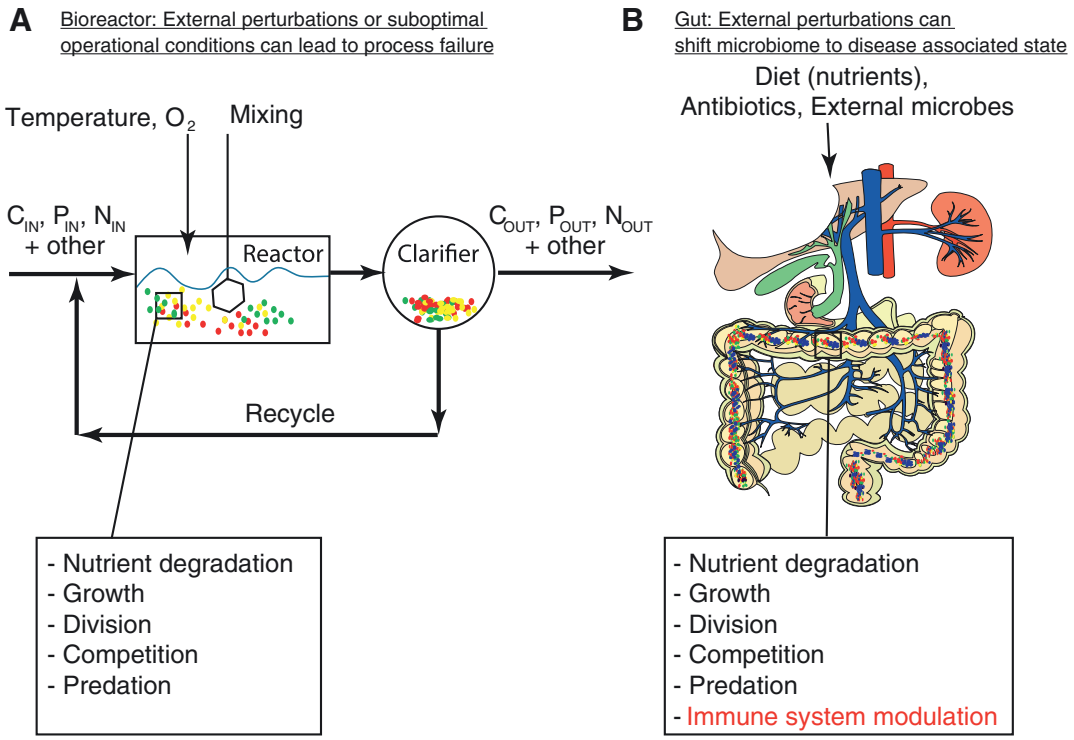

Fig. 1. Analogy between environmental engineering bioreactors and the human intestinal microbiome. Bioreactors, such as the activated sludge (A), use multispecies microbial communities to degrade waste (e.g., C = organic carbon) and other nutrients (e.g., N = nitrogen as ammonia, P = phosphorous) that if untreated can lead to environmental pollution problems. The optimal microbial community is selected by the bioengineer by tuning external parameters such as mixing, aeration and temperature. Composition shifts due to improper operational conditions can lead to system failure. (B) The intestinal microbiome is the multispecies microbial communities harbored in the human intestine. Similar to the environmental bioreactor communities, the microbiome degrades nutrients and is susceptible to external perturbations such as diet change and antibiotic application. Its composition shifts due to changes in diet or antibiotic use, and some community alterations have been associated with disease.

diverse both at the phylogenetic level and at the functional level [5]. While it is essential to maintain a proper microbial composition to assure efficient wastewater treatment [6,7], the bioengineer has only a limited number of operational handles on the system. The operational variables include aeration, mixing and flow rate, which if not optimally chosen can lead to bioreactor failure [8]. Mathematical models can be valuable tools to assist in the operation and control of these bioreactors, in order to enrich microbial composition in the right type of microbes [9].

Traditional mathematical models of wastewater treatment are based on differential equations that describe microbial population dynamics and the dynamics of chemical compounds in solution or in suspension. Most models adopt variations of the Activated Sludge Model (ASM) [1]. There are presently five versions of the ASM [10] that differ in their level of detail. In general, these models include bioreactions such as degradation of soluble, particulate and colloidal organic carbon sources and of nitrogen (as ammonia, nitrate and nitrite) and phosphorus sources. Microbial processes from mainly three functionally relevant microbial groups mediate these reactions. These microbial groups are ordinary heterotrophic organisms (i.e., organisms that use organic carbon for growth), autotrophic nitrifying organisms (organisms that use inorganic nitrogen sources such as ammonia) and phosphorus accumulating organisms (PAOs; which store phosphorous internally in the form of polyphosphates). This functional grouping was chosen because the goal of wastewater treatment is to remove nutrients before municipal or industrial effluents are released to the environment. In this functional grouping, the microbes do not necessarily belong to the same phylogenetic taxa. An example equation corresponding to PAOs net growth from the ASM2 model is displayed in Fig. 2A.

Even though many engineers rely on ASM to characterize bioreactor properties and designs [11], there is active research to introduce additional levels of detail [12]. Some of the limitations of ASM are due to oversimplified model representation (functional groups) and parameterization [13]. For example, ASM2 was modified to simulate phosphorous removal by PAOs and to include the presence of

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

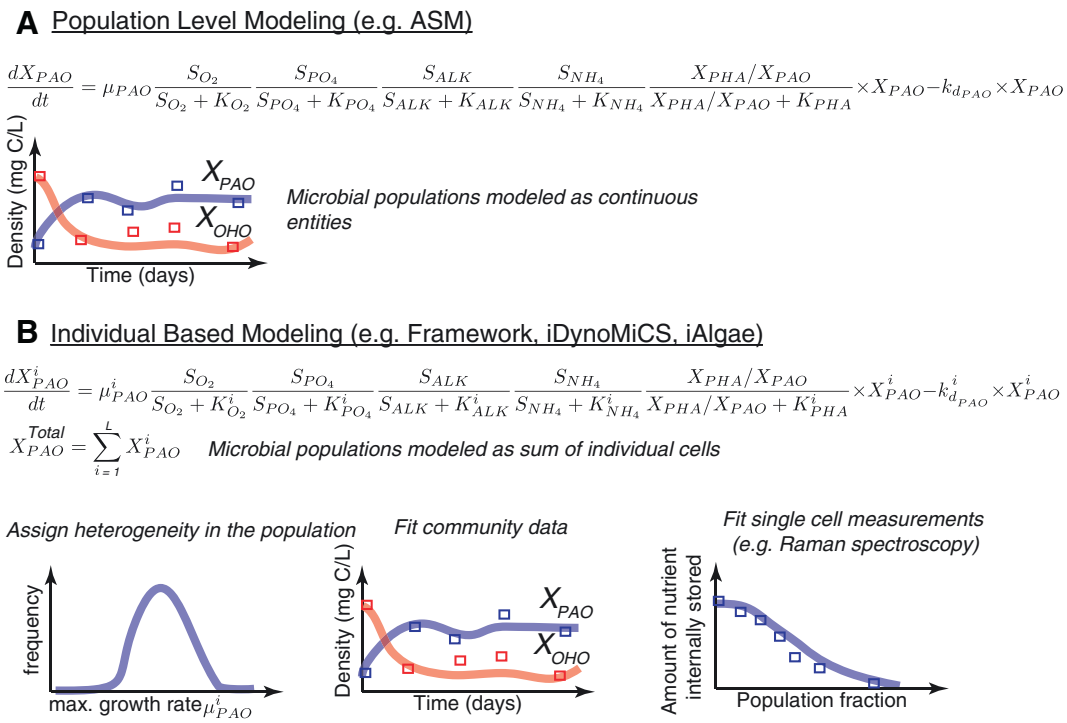

Fig. 2. Population- and individual-level approaches used to model environmental bioreactor microbial communities. (A) Population-level modeling, such as the ASM, considers microbial population as continuous entities. These models do not account for population heterogeneity in parameters and internally stored nutrients and are used to reproduce bulk community measurements. μ_{PAO} is the maximum growth rate of phosphate accumulating organisms (PAOs), S_{O_2} represents the oxygen bulk concentration, S_{PO_4} represents the bulk phosphate concentrations, S_{ALK} is the alkalinity of the wastewater, S_{NH_4} represents the bulk ammonia concentrations, X_{PHA} is the poly-hydroxy-alkanoates internally stored concentration for PAOs and X_{PAO} is the biomass density of PAOs in the wastewater. K_{O_2}, K_{PS}, K_{ALK}, K_{NH_4} and K_{PHA} are the half-saturation constants for the respective saturation kinetics. (B) IbM has been used to respond to the increasing number of single-cell measurements (e.g., Raman spectroscopy [14,15]). These measurements show that microbial functionally relevant groups in bioreactors are heterogeneous, such that they harbor different level of internally stored nutrients. Fitting single-cell observations and bulk community data with IbM lead to a deeper understanding of the system under investigation. In the IbM approach, total PAOs growth is an emergent property of the summation over each individual cell contribution to growth which is dependent on the current state of the individual cell, its parameters and the local concentrations of solutes. With superscript *i*, we indicate the parameters corresponding to individual cell *i* = *1…L* and *L* is the total number of individuals. Examples of IbMs are the Framework [16], iDynoMiCS [17] and iAlgae [15].

another microbial subgroup, the glycogen accumulating organisms, which compete with the PAOs for carbon uptake therefore limiting the bioreactor performance in phosphorous [18]. Nevertheless, a recent review still argues that the metabolic differences among microbial subgroups within the PAOs (or glycogen accumulating organisms), which results from lumping phylogenetically different species into the same category, are responsible for some of the inconsistencies of the application of these models [19].

Individual-Based Models

Most ASMs ignore the fact that microbial populations, even isogenic ones, can be highly heterogeneous phenotypically [20]. Individual bacterial cells in the same population can differ, for example, in their growth rate or the level of internally stored nutrients. The latter has been recently shown experimentally using single-cell Raman spectroscopy to measure intracellular metabolites coupled with fluorescence *in situ* hybridization microscopy and may have strong implications in the biology of environmental bioreactor microbial communities [14]. Population-level models such as the ASM typically neglect this information. However, parameters obtained by bulk population measurements can be poor estimations of the actual mechanistic parameters and therefore limit the applicability of a calibrated model to different scenarios, which is a consequence of the mathematical concept known as Jensen's inequality explaining that the weighted average of a given property of the individuals in the population can be greater than the simple average value in the population [21]. This

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

shortcoming can be to some extent resolved using individual-based modeling (IbM) that accounts for cell heterogeneity by explicitly simulating individual bacterial cells within the population [21]. The global population-level behavior results from the cumulative behavior of the individuals and their interactions (Fig. 2B). A recent study showed the importance of accounting for heterogeneity in enhanced biological phosphorus removal systems by showing that IbMs are able to predict the distribution of internal stored nutrients obtained by single-cell Raman spectroscopy methods [15].

Another important source of individual variability comes from environmental heterogeneities. In microbial communities such as thick biofilms, bacterial growth will be limited by diffusion of nutrients [22]. Because diffusion of nutrients is a slow process compared to their consumption by bacteria, spatial gradients will form in a biofilm. Gradients basically mean that bacteria in different locations of the biofilm will experience different local concentrations. The spatial heterogeneity has consequences to the way of bacteria interaction within a biofilm. IbMs that take diffusion into account have been of great importance to unveil the role of gradients in biofilm processes [17,22–28].

Although IbMs are more computationally demanding [18] than equivalent population models, these models are increasingly being adopted to study microbial processes, particularly in surface-attached (or biofilm) communities [29,30]. IbMs allow a greater level of detail at microscopic scales with implications for the overall bioreactor functioning [21]. One notable example of a successful IbM is the multiyear, multigroup project to develop a general IbM for microbial dynamics in surface-attached reactors called iDynoMiCS [17]. iDynoMiCS was already applied to evaluate reactor performance as a function of the relative ratio between biomass limitation and diffusion limitation in a membrane biofilm reactor [31]. Similar IbMs were used to design optimal strategies for nitrogen removal in granular bed sludge reactors [16] and in other areas of biofilm biology [27,32,33].

Metagenomics: Unveiling the Intestinal Microbiome

Like wastewater treatment communities, the intestinal microbiota is a highly diverse ecosystem that performs important nutrient degradation functions [34]. The gut microbiota has the additional roles of providing resistance against colonization by enteropathogens [35–38] and modulating immunity [38–40]. Recent studies, supported by next-generation DNA sequencing, have established novel connections between the intestinal microbiota composition and disease [41,42]. Compositional shifts in the intestinal microbiota have been linked to chronic diseases such as obesity [43], diabetes [44] and Crohn's disease [45]. Interestingly, even drug-induced transient changes increase the risk of developing acute intestinal infections [46] or pulmonary viral infections in mammalian hosts [47].

Two main methods are currently used to investigate this community with high-throughput DNA sequencing [48]. The first one, amplicon or targeted sequencing, uses pooled sequencing of the PCR product of 16S ribosomal RNA (16S rRNA) genes followed by mapping the resulting uniquely identified sequences to a taxonomic database (see Ref. [49] as example). The second method, shotgun sequencing, aims to characterize the whole metagenomic content in a sample (the microbiome) and identifying function by mapping sequencing reads to a database of genes with presumably known function (KEGG, COG) [50]. While 16S analysis provides mostly demographic information (which microbes are present and how abundant they are), shotgun sequencing can provide information on the processes available in this ecosystem (e.g., pathways) [51].

Current microbiome analyses provide a unique view into this, still little known, microbial ecosystem. Nonetheless, the quantitative analyses presented so far are mostly descriptive, such as comparisons of sample composition using quantitative indices and correspondence analysis [52] or cross-sectional statistical tests [53,54]. Top-down modeling frameworks such as Singular Value Decomposition [55] or Mixture Models Engines [56] that aim to identify patterns in the data have been employed to individuate stereotypical microbial responses to external perturbations but provide little information on community interactions. Correlation analyses have been used to evaluate interactions among species and the effect of external perturbations [57,58], but these methods lack in mechanistic insight. The temporal predictability of a microbial ecosystem was recently quantified using time-decay similarity indices [59] but this method cannot be used to predict the intestinal microbiota dynamics [60] (Fig. 3A).

Environmental engineering provides mathematical frameworks to predict temporal dynamics of microbial communities as a function of multiple external stimuli and perturbations. The same approach in principle can be applied to the intestinal microbiome. However, there is a large gap between sequencing data and the level of mechanistic detail used in biochemically realistic models. Can we bridge this gap in order to develop models of intestinal microbiota dynamics and its relation to health and disease?

Mechanistic Models of the Intestinal Microbiota

A recent model proposed for the human intestinal flora [61,62] aimed to simulate the microbial digestion

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

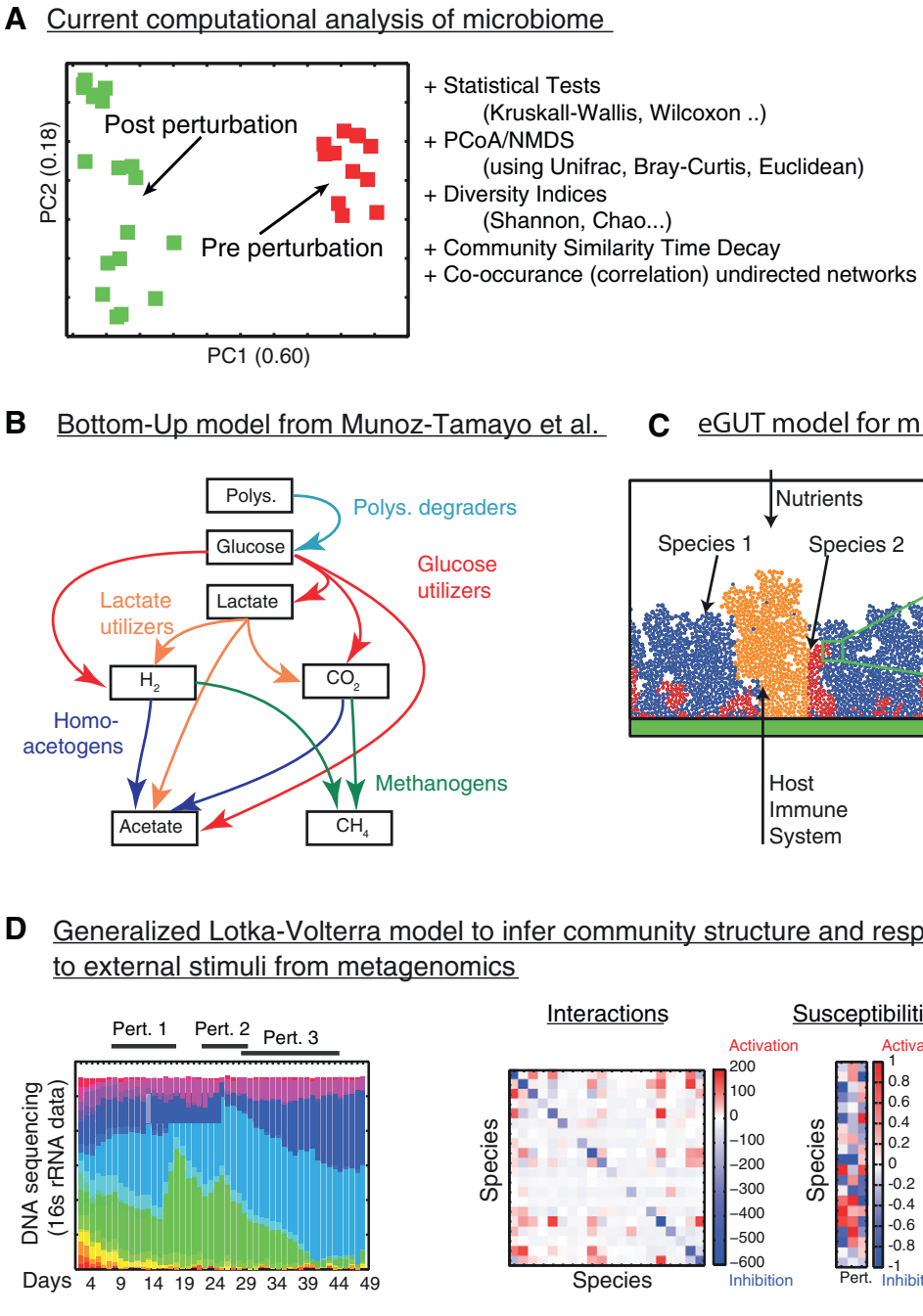

Fig. 3. Computational analysis and mathematical models of the intestinal microbiome. (A) Example and list of current computational approaches used to analyze community data for microbiome (16S rRNA) studies. (B) Schematic of a model to simulate polysaccharide digestion in the intestine [61]. This model is built in a bottom-up approach, which aims to determine population-level behaviors from detailed biochemical reactions. Parameters are often taken from literature. No comparison with data is provided in the original paper. (C) Conceptual scheme of the approach used in the eGUT simulator under development by Jan Kreft. (D) Generalized Lotka–Volterra system of equation with time-variable perturbations approach. The method uses ecological modeling and machine learning to infer network of microbial interactions, susceptibilities to external perturbations and growth rates. The parameters inferred are used in an ecological community model that can then be used to predict ecosystem dynamics by numerical simulations or to identify steady states. The data shown is an example obtained by simulation to validate the inference method and shows microbiota dynamics in the presence of three distinct perturbations representing antibiotic administration [63].

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

of complex polysaccharides into sugars and end products such as short chain fatty acids, methane and carbon dioxide. The model uses an engineering modeling framework similar to the ASM and models the intestinal tract as multiple interconnected completely mixed reactors, each representing a section of the human intestine. Each compartment is subdivided in two additional subcompartments corresponding to the lumen and the mucosa. Nine chemical compounds and four functional microbial groups are modeled using mass balance equations. As is common in environmental engineering, the bioconversions of microbial growth and chemical substrate transformations are modeled with saturation kinetics using parameters obtained from the literature or previous experiments. The model was applied to investigate effect of diets and host secretion of mucins on microbial composition and function (Fig. 3B).

There have been efforts to apply IbM as well. A recent example investigated how the social interaction between microbial species can be modulated by a third species [26]. Such a systematic analysis of multispecies interactions may eventually be extended to large numbers of species to model biodiverse microbiomes. Another study used a two-species model but included interaction with the host [64]. This individual-based model, perhaps the first to describe host–microbiome interactions, revealed that host secretions could be important in maintaining microbiome diversity. Members of the group that created iDynoMiCS [17] are developing an individual-based model of the mammalian gut microbiota. The project, called eGUT, is still in early stages of development and aims to produce a software tool for the broad scientific community, with the hope that *in silico* investigations may reduce the need for *in vivo* experimentation with animals in the future. Compared to the iDynoMiCS framework, which focuses in inter-species interactions and nutrient consumption, eGUT will also include a description of the intestinal wall, processes simulating immunity and inflammation and a sequence of larger compartments with three regions per compartment corresponding to mucosa, lumen and particles in the lumen (Fig. 3C).

Inferring Community Structure from Metagenomic Data

Mechanistic models such as the ones described above typically rely on a course-grained functional definition of microbial groups (e.g., glucose utilizers, methanogens) that is difficult to conciliate with most microbiota sequencing data. 16S amplicon sequencing metagenomics primarily provide information on the phylogeny of the bacteria compositing the consortium, rather than providing information on their function. Nonetheless, methods to integrate these two levels of information are now under development [51,65]. Such methods can be highly valuable to assist in model development using data from high-throughput sequencing, which have already demonstrated that the presence or absence of a specific bacterium may change ecosystem health and development significantly [66]. Similarly, spatially structured IbMs describe processes at a much finer scale compared with metagenomics and may not benefit from the metagenomics revolution directly.

Modeling population dynamics using the generalized Lotka–Volterra framework is a possible solution for predictive modeling of the intestinal microbiota that can leverage current metagenomic data. This concept has been first suggested by White [67] and very recently led to independent implementations by three distinct research groups [63,68,69]. This type of modeling derives from the classical Lotka–Volterra dynamics used in mathematical ecology to model simple systems such as two-species predator–prey dynamics. The generalized Lotka–Volterra extends the concept to an arbitrary number of species (see the box). The recent papers [63,68,69] propose methods to infer the species–species interaction parameters directly from time series of absolute abundances of microbial groups, which may be obtained from longitudinal metagenomics studies. This approach is similar in concept to methods of inference of molecular networks such as metabolic modeling using S systems [70].

Importantly, like in other network inference methods, the data-driven inference of Lotka–Volterra dynamics from metagenomics time series allows measuring parameters in the natural setting where the microbes reside. This is important for two key reasons. First, it enables measuring growth parameters of unculturable microbes, which would be impossible to measure by traditional methods requiring microbial isolation. Second, even when microbes are culturable, growth rates measured *in vitro* are unlikely to match *in vivo* growth rates because it is difficult to recreate the exact conditions experienced by microbes in the gut. Perhaps another key feature is that the generalized Lotka–Volterra can be extended to model time-dependent external perturbations [61]. Under this extension, the temporal variation in total abundance of species can be computed as the balance between growth, interaction with itself and all the other present in the ecosystem but also a response to multiple time-varying external perturbations (see Box 1). In the case of an antibiotic perturbation, the method provides a quantitative estimation of the sensitivity to antibiotic of every microbial group [61]. The method can also be used to assess the effect of other perturbations such as changes in diet.

The novelty of generalized Lotka–Volterra approaches lies in the fact that the inferred ecological structure of the intestinal microbiota and its response

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

Box 1
Intestinal microbiome community structure inference and dynamic predictions using generalized Lotka–Volterra with time-dependent perturbations.

The time derivative of the microbial species x_i is computed as a balance between growth, interactions and response to time-varying external stimuli (e.g., antibiotic, diet) and is computed as

$$\frac{dx_i(t)}{dt} = x_i(t)\left(\mu_i + \sum_{j=1}^{L} M_{ij}x_j(t) + \sum_{k=1}^{P}\int_{ik} u_k(t)\right).$$

Here μ_i denotes the unlimited growth rate of unit i in the absence of any competition, the interaction term M_{ij} characterizes the effect of unit j on i. In particular, $M_{ij} > 0$ stands for activation and $M_{ij} < 0$ stands for repression. (No interaction is accordingly indicated by $M_{ij} = 0$). Effect of external perturbations is described by $\sum_{k=1}^{P}\int_{ik} u_k(t)$ where u_k represents an external, time-variable stimulus of a perturbation $k = 1\ldots P$ whose relative susceptibility for each unit i is represented by $\in_{ik}$. This model can be discretized and rewritten in matrix formulation such that the parameters corresponding to growth, interaction and susceptibilities can be inferred from temporal high-throughput microbiota observations using cross-validation techniques and Tikhonov regularization-based optimization approaches [61].

to external perturbations can be used for temporal predictions across short and long time scales (Fig. 3D). These models can generate hyphotheses that, once properly validated, can lead to mechanistic insight. Possible applications of these models for the rational development of microbiota-tailoring therapies include the optimization of antibiotic dosage and timing to reduce impact on beneficial gut microorganisms.

Another important application of network inference methods is to identify keystone members of the microbial ecosystem [69] and key interactions within the microbiota [63]. Candidate species and interactions could be properly validated with scalable *in vivo* mouse models with well-defined communities [71–73] as another step towards clinical application.

Moving towards Whole-Cell Community Models

Because they rely strictly on 16S rRNA amplicon DNA sequencing, the present generalized Lotka–Volterra models do not consider the whole genomic content and therefore neglect important aspects of selection and evolution such as antibiotic resistance. A species from a specific phylogenetic group that mutates by acquiring antibiotic resistance will still be placed in that same phylogenetic group. This creates inconsistencies in the model, since the model presently assumes that the properties of the group are static. In principle, one can extend the framework to allow for time-dependent parameters and time-dependent number of interacting entities such as *Enterococcus* and vancomycin-resistant *Enterococcus*, which are relevant in clinical applications [4,46,74]. However, this requires knowledge on the community beyond that obtained by amplicon sequencing. Also, very importantly, composition shifts observed from community profiling may not necessarily correlate with shifts in ecosystem functionality.

The increasing affordability of nucleic acid sequencing is enabling the acquisition of shotgun metagenomics and even metatranscriptomics data [75,76]. Parallel advances in mass spectrometry technologies allow quantifying abundance of metabolic products such as proteins and metabolites [77]. Bioinformatics analysis on these data enables investigating the microbiome at the level of specific pathways [51] and reconstructing *de novo* the genome of (at least) the most abundant bacteria [78–80]. Combined metatranscriptome and metabolome analysis of microbial communities [81] can help identify which genes are actively transcribed [82] and what proteins and metabolites are produced and degraded. This information will be very valuable in next-generation modeling efforts [83].

Recent studies have focused in extending the widely used flux balance analysis (FBA) [84], a mathematical approach for analyzing metabolite fluxes through metabolic networks, to community flux balance analysis (cFBA) [85,86] and dynamic flux balance analysis (dFBA) [87]. As deducible from its name,

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

cFBA is focused on determining the metabolic fluxes across multiple interacting microbial strains. Compared to traditional FBA, which constrains the single strain metabolic models under the assumption of growth rate maximization, cFBA relies on a multilevel optimization description. For example, OptCom [86], a recently developed computational platform for cFBA, constrains the community metabolic model by considering potential fitness tradeoffs between the individual and the whole community. dFBA instead assumes that the intracellular solute concentrations are in equilibrium with the extracellular environment and, similarly to FBA and cFBA, the resulting underdetermined stoichiometric model is solved under the assumption of a biochemical objective such as growth rate maximization. In dFBA, the metabolic model is instead coupled with metabolite mass balance equations in the extracellular environment. It assumes a pseudo-steady state where the reactions of the internal metabolites are on a shorter time scale compared to the extracellular nutrients and metabolites. The coupling between cellular metabolism and extracellular metabolites is achieved through the constraint of rates of extracellular metabolite uptake and excretion from the cell. Similarly to traditional FBA, due to the underdetermined nature of the systems of equations, optimization algorithms are essential to compute a solution.

The microbial ecology field is rapidly moving towards integrating multiple level of information to evaluate microbial ecosystem functionality. Nevertheless, metabolic modeling at the current state still treats the community as a single entity interacting with the surrounding environment. Agent-based models capable of going beyond the present oversimplified scenarios by including cellular processes in each microbial cell such as DNA replication, transcription and translation of a few key genes and linking the single-cell behavior to that of the whole population could move the field forward [88]. These types of models can provide a scaffold for model calibration and validation using multi-omics data towards a predictive model of the intestinal microbiome.

Acknowledgments

This work was supported by the Office of the Director, National Institutes of Health of the National Institutes of Health under Award Number DP2OD008440 to J.B.X. and the Integrated Cancer Biology Program of the National Cancer Institute under Grant U54 CA14896704, a seed grant from the Lucille Castori Center for Microbes, Inflammation and Cancer. We thank Joana S. Torres for help with figure preparation.

Received 15 February 2014;
Received in revised form 24 March 2014;
Accepted 30 March 2014
Available online xxxx

Keywords:
mathematical modeling;
network inference;
metagenomics;
antibiotic;
biofilm

Abbreviations used:
ASM, Activated Sludge Model; PAO, phosphorus accumulating organism; IbM, individual-based modeling; FBA, flux balance analysis; cFBA, community flux balance analysis; dFBA, dynamic flux balance analysis.

References

[1] Henze M. Activated sludge models: ASM1, ASM2, ASM2d and ASM3. London, UK: IWA Publishing; 2000 9.
[2] McLellan SL, Huse SM, Mueller-Spitz SR, Andreishcheva EN, Sogin ML. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. Environ Microbiol 2010;12:378–92.
[3] Delzenne NM, Neyrinck AM, Cani PD. Modulation of the gut microbiota by nutrients with prebiotic properties: consequences for host health in the context of obesity and metabolic syndrome. Microb Cell Factories 2011;10:S10.
[4] Ubeda C, Bucci V, Caballero S, Djukovic A, Toussaint NC, Equinda M, et al. Intestinal microbiota containing *Barnesiella* species cures vancomycin-resistant *Enterococcus faecium* colonization. Infect Immun 2013;81:965–73.
[5] Bassin JP, Kleerebezem R, Muyzer G, Rosado AS, van Loosdrecht MCM, Dezotti M. Effect of different salt adaptation strategies on the microbial diversity, activity, and settling of nitrifying sludge in sequencing batch reactors. Appl Microbiol Biotechnol 2012;93:1281–94.
[6] Moralejo-Gárate H, Mar'atusalihat E, Kleerebezem R, van Loosdrecht MCM. Microbial community engineering for biopolymer production from glycerol. Appl Microbiol Biotechnol 2011;92:631–9.
[7] Shimada T, Li X, Zilles JL, Morgenroth E, Raskin L. Effects of the antimicrobial tylosin on the microbial community structure of an anaerobic sequencing batch reactor. Biotechnol Bioeng 2011;108:296–305.
[8] Leitão RC, van Haandel AC, Zeeman G, Lettinga G. The effects of operational and environmental variations on anaerobic wastewater treatment systems: a review. Bioresour Technol 2006;97:1105–18.
[9] Volcke EIP, Picioreanu C, De Baets B, van Loosdrecht MCM. The granule size distribution in an anammox-based granular sludge reactor affects the conversion—implications for modeling. Biotechnol Bioeng 2012;109:1629–36.
[10] Hauduc H, Rieger L, Oehmen A, van Loosdrecht M, Comeau Y, Héduit A, et al. Critical review of activated sludge modeling: state of process knowledge, modeling concepts, and limitations. Biotechnol Bioeng 2013;110:24–46.
[11] Sharma KR, Corrie S, Yuan Z. Integrated modelling of sewer system and wastewater treatment plant for investigating the impacts of chemical dosing in sewers.

Water Sci Technol J Int Assoc Water Pollut Res 2012;65:1399–405.
[12] Trutnau M, Petzold M, Mehlig L, Eschenhagen M, Geipel K, Müller S, et al. Using a carbon-based ASM3 EAWAG Bio-P for modelling the enhanced biological phosphorus removal in anaerobic/aerobic activated sludge systems. Bioprocess Biosyst Eng 2011;34:287–95.
[13] Le Moullec Y, Potier O, Gentric C, Leclerc JP. Activated sludge pilot plant: comparison between experimental and predicted concentration profiles using three different modelling approaches. Water Res 2011;45:3085–97.
[14] Majed N, Chernenko T, Diem M, Gu AZ. Identification of functionally relevant populations in enhanced biological phosphorus removal processes based on intracellular polymers profiles and insights into the metabolic diversity and heterogeneity. Environ Sci Technol 2012;46:5010–7.
[15] Bucci V, Majed N, Hellweger FL, Gu AZ. Heterogeneity of intracellular polymer storage states in enhanced biological phosphorus removal (EBPR)—observation and modeling. Environ Sci Technol 2012;46:3244–52.
[16] Xavier JB, De Kreuk MK, Picioreanu C, Van Loosdrecht MCM. Multi-scale individual-based model of microbial and bioconversion dynamics in aerobic granular sludge. Environ Sci Technol 2007;41:6410–7.
[17] Lardon LA, Merkey BV, Martins S, Dötsch A, Picioreanu C, Kreft J-U, et al. iDynoMiCS: next-generation individual-based modelling of biofilms. Environ Microbiol 2011;13:2416–34.
[18] Whang LM, Filipe CDM, Park JK. Model-based evaluation of competition between polyphosphate- and glycogen-accumulating organisms. Water Res 2007;41:1312–24.
[19] Oehmen A, Carvalho G, Lopez-Vazquez CM, Van Loosdrecht MCM, Reis MAM. Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. Water Res 2010;44:4992–5004.
[20] Lidstrom ME, Konopka MC. The role of physiological heterogeneity in microbial population behavior. Nat Chem Biol 2010;6:705–12.
[21] Hellweger FL, Bucci V. A bunch of tiny individuals—individual-based modeling for microbes. Ecol Model 2009;220:8–22.
[22] Stewart PS. Diffusion in biofilms. J Bacteriol 2003;185:1485–91.
[23] Picioreanu C, Kreft J-U, Van Loosdrecht MCM. Particle-based multidimensional multispecies biofilm model. Appl Environ Microbiol 2004;70:3024–40.
[24] Chambless JD, Hunt SM, Stewart PS. A three-dimensional computer model of four hypothetical mechanisms protecting biofilms from antimicrobials. Appl Environ Microbiol 2006;72:2005–13.
[25] Xavier JB, Foster KR. Cooperation and conflict in microbial biofilms. Proc Natl Acad Sci 2007;104:876–81.
[26] Mitri S, Xavier JB, Foster KR. Social evolution in multispecies biofilms. Proc Natl Acad Sci 2011. http://dx.doi.org/10.1073/pnas.1100292108.
[27] Bucci V, Nadell CD, Xavier JB. The evolution of bacteriocin production in bacterial biofilms. Am Nat 2011;178:E162–73.
[28] Nadell CD, Bucci V, Drescher K, Levin SA, Bassler BL, Xavier JB. Cutting through the complexity of cell collectives. Proc R Soc B Biol Sci 2013;280:20122770.
[29] Park S, Bae W, Rittmann BE. Multi-species nitrifying biofilm model (MSNBM) including free ammonia and free nitrous acid inhibition and oxygen limitation. Biotechnol Bioeng 2010;105:1115–30.
[30] Skybová T, Přibyl M, Pocedič J, Hasal P. Mathematical modeling of wastewater decolorization in a trickle-bed bioreactor. J Biotechnol 2012;157:512–23.
[31] Martin KJ, Picioreanu C, Nerenberg R. Multidimensional modeling of biofilm development and fluid dynamics in a hydrogen-based, membrane biofilm reactor (MBfR). Water Res 2013;47:4739–51.
[32] Chambless JD, Stewart PS. A three-dimensional computer model analysis of three hypothetical biofilm detachment mechanisms. Biotechnol Bioeng 2007;97:1573–84.
[33] Merkey BV, Lardon LA, Seoane JM, Kreft J-U, Smets BF. Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. Environ Microbiol 2011;13:2435–52.
[34] Walter J, Ley R. The human gut microbiome: ecology and recent evolutionary changes. Annu Rev Microbiol 2011;65:411–29.
[35] Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal *Salmonella* diarrhea. PLoS Pathog 2010;6:e1001097.
[36] Fukuda S, Toh H, Taylor TD, Ohno H, Hattori M. Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. Gut Microbes 2012;3:449–54.
[37] Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature 2011;469:543–7.
[38] Stecher B, Hardt WD. The role of microbiota in infectious disease. Trends Microbiol 2008;16:107–14.
[39] Atarashi K, Tanoue T, Oshima K, Suda W, Nagano Y, Nishikawa H, et al. Treg induction by a rationally selected mixture of *Clostridia* strains from the human microbiota. Nature 2013;500:232–6.
[40] Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell 2009;139:485–98.
[41] Blaser M, Bork P, Fraser C, Knight R, Wang J. The microbiome explored: recent insights and future challenges. Nat Rev Microbiol 2013;11:213–7.
[42] Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet 2012;13:260–70.
[43] Flint HJ. Obesity and the gut microbiota. J Clin Gastroenterol 2011;45:S128–32.
[44] Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS One 2010;5:e9085.
[45] Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13:R79.
[46] Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, et al. Vancomycin-resistant *Enterococcus* domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest 2010;120:4332–41.
[47] Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS, et al. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc Natl Acad Sci 2011. http://dx.doi.org/10.1073/pnas.1019378108.
[48] Consortium, T.H.M.P. A framework for human microbiome research. Nature 2012;486:215–21.

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017

[49] Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol 2008;6:e280.
[50] Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature 2012;486:222–7.
[51] Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol 2012;8:e1002358.
[52] Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J 2011;5:169–72.
[53] Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12:R60.
[54] Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res 2012;22:292–8.
[55] Bucci V, Bradde S, Biroli G, Xavier JB. Social interaction, noise and antibiotic-mediated switches in the intestinal microbiota. PLoS Comput Biol 2012;8:e1002497.
[56] Gerber GK, Onderdonk AB, Bry L. Inferring dynamic signatures of microbes in complex host ecosystems. PLoS Comput Biol 2012;8:e1002624.
[57] Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol 2012;8:e1002606.
[58] Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol 2012;8:e1002687.
[59] Shade A, Gregory Caporaso J, Handelsman J, Knight R, Fierer N. A meta-analysis of changes in bacterial and archaeal communities with time. ISME J 2013;7:1493–506.
[60] Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol 2012;10:538–50.
[61] Muñoz-Tamayo R, Laroche B, Walter É, Doré J, Leclerc M. Mathematical modelling of carbohydrate degradation by human colonic microbiota. J Theor Biol 2010;266:189–201.
[62] Muñoz-Tamayo R, Laroche B, Walter É, Doré J, Duncan SH, Flint HJ, et al. Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. FEMS Microbiol Ecol 2011;76:615–24.
[63] Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol 2013;9:e1003388.
[64] Schluter J, Foster KR. The evolution of mutualism in gut microbiota via host epithelial selection. PLoS Biol 2012;10:e1001424.
[65] Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc Natl Acad Sci U S A 2013;110:12804–9.
[66] Wu H-J, Ivanov II, Darce J, Hattori K, Shima T, Umesaki Y, et al. Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. Immunity 2010;32:815–27.
[67] White JR. Novel methods for metagenomic. Analysis 2010 [http://hdl.handle.net/1903/10292].
[68] Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD. Mathematical modeling of primary succession of murine intestinal microbiota. Proc Natl Acad Sci 2014;111:439–44.
[69] Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. http://arxiv.org/abs/1402.0511; 2014.
[70] Sorribas A, Savageau MA. Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. Math Biosci 1989;94:239–69.
[71] Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, Wollam A, et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proc Natl Acad Sci U S A 2009;106:5859–64.
[72] Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science 2011;333:101–4.
[73] Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI. Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. Sci Transl Med 2014;6:220ra11.
[74] Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis 2012;55:905–14.
[75] Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature 2013;493:45–50.
[76] Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell 2013;152:39–50.
[77] Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. Chem Soc Rev 2009;38:1882–96.
[78] Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J. Ray Meta: scalable *de novo* metagenome assembly and profiling. Genome Biol 2012;13:R122.
[79] Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. Nucleic Acids Res 2012;40:e155.
[80] Peng Y, Leung HCM, Yiu SM, Chin FYL. Meta-IDBA: a *de novo* assembler for metagenomic data. Bioinformatics 2011;27:i94–i101.
[81] Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. Sci Rep 2013;3. http://dx.doi.org/10.1038/srep02532.
[82] Leimena MM, Ramiro-Garcia J, Davids M, van den Bogert B, Smidt H, Smid EJ, et al. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. BMC Genomics 2013;14:530.
[83] Greenblum S, Chiu H-C, Levy R, Carr R, Borenstein E. Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities. Curr Opin Biotechnol 2013;24:810–20.
[84] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8.
[85] Khandelwal RA, Olivier BG, Röling WFM, Teusink B, Bruggeman FJ. Community flux balance analysis for microbial consortia at balanced growth. PLoS One 2013;8:e64567.
[86] Zomorrodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 2012;8:e1002363.
[87] Höffner K, Harwood SM, Barton PI. A reliable simulator for dynamic flux balance analysis. Biotechnol Bioeng 2013;110:792–802.
[88] Hellweger FL. Escherichia coli adapts to tetracycline resistance plasmid (pBR322) by mutating endogenous potassium transport: *in silico* hypothesis testing. FEMS Microbiol Ecol 2013;83:622–31.

Please cite this article as: Bucci Vanni, Xavier Joao B., Towards Predictive Models of the Human Gut Microbiome, J Mol Biol (2014), http://dx.doi.org/10.1016/j.jmb.2014.03.017