Dynamic simulations of microbial communities under perturbations: opportunities for microbiome engineering

Beatriz García-Jiménez (beatriz.garcia@upm.es)
Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Campus de Montegancedo-UPM https://orcid.org/0000-0002-8129-6506

Jorge Carrasco
Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Campus de Montegancedo-UPM 28223 Pozuelo de Alarcón (Madrid) Spain

Joaquín Medina
Centro de Biotecnologia y Genomica de Plantas (CBGP, UPM-INIA) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Campus de Montegancedo-UPM 28223 Pozuelo de Alarcón (Madrid) Spain

Mark D Wilkinson
Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) Campus de Montegancedo-UPM 28223 Pozuelo de Alarcón (Madrid) Spain

---

Research

Keywords: microbiome, Artificial Intelligence, perturbation, microbial community, modeling, genome-scale metabolic model (GEM), Markov Decision Process (MDP), Crohn's disease, herbicide degradation, microbiome state

Posted Date: February 25th, 2020

DOI: https://doi.org/10.21203/rs.2.24431/v1

Dynamic simulations of microbial communities under perturbations: opportunities for microbiome engineering

Authors:
Beatriz García-Jiménez (1)
beatriz.garcia@upm.es

Jorge Carrasco (1)(2)
jorge.cmuriel@alumnos.upm.es

Joaquín Medina (1)
medina.joaquin@inia.es

Mark D. Wilkinson (1)(2)
markw@illuminae.com

Affiliations:
(1) Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA) Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus de Montegancedo-UPM 28223 Pozuelo de Alarcón (Madrid) Spain

(2) Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid (UPM) 28040 (Madrid) Spain

Corresponding author:
Beatriz García-Jiménez
beatriz.garcia@upm.es

RESEARCH

Dynamic simulations of microbial communities under perturbations: opportunities for microbiome engineering

Beatriz García-Jiménez[??*], Jorge Carrasco[??,??], Joaquín Medina[??] and Mark D. Wilkinson[??,??]

Abstract

Background: There are few large longitudinal microbiome studies, and fewer that include controlled, well-annotated perturbations between sampling-points. Thus, there are few opportunities to employ data-driven computational analyses of perturbed microbial communities over time.

Results: Our novel computational system simulates the dynamics of microbial communities under perturbations using genome-scale metabolic models (GEMs). Perturbations include modifications to a) the nutrients available in the medium, allowing modelling of prebiotics; and/or b) the microorganisms present in the community to model, for example, probiotics or pathogen infection. These simulations generate the quantity and types of information required by MDPbiome, an AI system which builds predictive models suggesting the perturbation(s) required to engineer microbial communities to a desired state. We call this novel combination of technologies "MDPbiomeGEM". We demonstrate, in a Crohn's disease microbiome, that MDPbiomeGEM correctly models the influence of both prebiotic fiber and a probiotic, resulting in a recommendation to consume inulin to recover from dysbiosis, consistent with prior biomedical knowledge. When used to model the soil microbiome's ability to degrade the herbicide Atrazine, differing recommendations arise depending on the highly variable state of the initial soil microbial composition, highlighting the relevance of both phosphate and microbes (i.e. *Halobacillus sp.* and *H.stevensii*) in a directed microbiome engineering strategy, consistent with previously published observations.

Conclusions: MDPbiomeGEM generates large volumes of longitudinal data of complex microbial communities experiencing perturbations. Machine learning on these data reveal patterns consistent with existing biological knowledge, supporting the validity of the approach. MDPbiomeGEM could save research resources by optimizing sample collection in metagenomics studies through identification of "informative" scenarios/time-points, or by predicting optimal in-vitro culture formulations for generating performant synthetic microbial communities. Finally, MDPbiomeGEM outputs include detailed information about the metabolic state of the community, which can be used to further interpret the impact of perturbations, and potentially could be used to predict novel metabolic biomarkers of a microbiome's state.

Keywords: microbiome; Artificial Intelligence; perturbation; microbial community; modeling; genome-scale metabolic model (GEM); Markov Decision Process (MDP); Crohn's disease; herbicide degradation; microbiome state

s
ología y
(CBGP,

ca de Madrid
acional de
ología Agraria

ncedo-UPM
arcón

ormation is
of the article

Introduction

Predicting individualized treatments based on microbiome analysis

Precision medicine and farming depends not only on the availability of individual patient/plant genomes, but also on their individual associated microbiomes [1, 2]. Thus, individualized treatments, in both medicine and agriculture, will include the ability to precisely engineer individual microbiomes towards healthier or more performant microbial compositions. Though in the most extreme cases this involves wholesale microbiome transplantation, less invasive treatments are feasible, entailing the use of external factors that alter the microbiome, such as diet, prebiotics, probiotics, antibiotics, nutrients, diet, etc.

MDPbiome is an Artificial Intelligence system, built using a Markov Decision Process, to provide in-silico perturbation recommendations (e.g., related to diet, pre/pro-biotics, drugs, fertilizers, etc.) that will guide the studied

microbiome through a path towards improved health or higher performance via predictions of microbial community responses to those perturbations.

Recent studies suggest that food-based interventions to modulate the gut microbiome are highly personalized, with different responses to the same food among individuals; concluding that the subject's existing gut microbiome should be taken into account to predict the effect of a particular nourishment [3]. Additionally, microbiome-based stratification - effectively, identification of microbiome "states" - has been emphasized when metabolic responses to diets are investigated in children [4]. Even more, novel molecules are beginning to be chemically designed specifically to remodel the microbiome, for example, causing a decrease in the growth of a particular microbe to promote the increase of other healthy and/or under-represented strains [5].

Beyond medicine, context-dependent, directed microbiome interventions could be applied to agriculture and the environment. External perturbations could modulate the soil microbiome to allow plants to grow in new conditions, such as those provoked by biotic or abiotic stresses such as climate change [6], salt-stress [7], fungal infection [8], and soil over-exploitation [9]. Industrial applications of microbiome engineering will increase the performance of added-value chemicals production, such as food (yogurt, beverages), probiotics, enzymes or biofuels [10, 11], by modulation of the strain ratio (community enrichment or reduction) or identification of the optimal medium composition [12].

Microbiome dynamics

The static microbial community composition in a multitude of niches have been extensively analyzed, mainly comparing healthy vs ill or compromised scenarios, in humans [13], plants [14], animals [15], etc. Although the microbiome is not stable over time [16], studies of microbiome dynamics are scarce, or focused on very few sampling points; few long-term metagenomics studies are available to illustrate the variability of microbial communities over time, and even fewer with sufficient metadata regarding external events that could influence microbial community changes [17].

There are multiple factors affecting microbiome dynamics. Alterations in microbiome composition have been observed in human gut due to changes in diet at a macro-level (e.g. Western diet versus Mediterranean or vegetarian diet) [18] or addition of specific nutrients such as resistant starch [19]. In plants, alterations in soil microbiomes happen in response to biotic and abiotic stresses, such as those listed above [20]. More than one hundred dysbioses have been related to human diseases [21], and disease-suppressing soils - defined by their microbial composition - are studied to promote plant resistance to pathogens [22].

If rich, longitudinal datasets were available, computational analyses could help elucidate microbial community dynamics and, indeed, provide insights into how to modify the microbiome in desirable ways. With sufficient samples, data-driven computational systems such as MDPbiome [17], could provide suggestions for potentially low-impact engineering of microbiomes, guiding medically or economically important microbial communities to healthier or more performant states.

Metabolic models for microbial community analysis

A genome scale metabolic model (GEM) [23] is a metabolic reconstruction of a particular organism. Although their initial application was focused on strain design for enhanced production of industrially-relevant biochemical products, their use was later extended to applications in a wide range of scenarios [24]. GEMs were originally designed to simulate the metabolism of one isolated strain or cell. Recent advances combine multiple GEMs to attempt to model microorganism communities [25]. There are, however, several technical and biochemical challenges to be addressed [26], including: a) lack of standardization, with models from different authors using different nomenclatures, hampering integration, b) accurately defining medium composition to mirror real-world scenarios, and c) accurately determining bounds of exchange rates in order to accurately model the fluxes between community members. Additionally, GEMs were originally designed to simulate a cell as a static snapshot, rather than to analyse microbial community dynamics. As such, this more challenging application will require substantial adaptation of GEM design [27].

Our proposal: MDPbiomeGEM

We propose a combination of MDPbiome and GEMs (MDPbiomeGEM), where multiple simulations of the dynamics of a (simplified) microbial community under perturbations will generate sufficient longitudinal data to enable an MDPbiome-based analysis. MDPbiomeGEM identifies different microbial community states arising in the simulated data, and the perturbation-caused transitions between states. It then computes a recommendation consisting of directed interventions that will move microbial communities from a disease or low-performance state to a more desirable state.

There are few studies combining microbiome analyses and metabolic models [28, 29, 30], and these are primarily static analyses. However, studies have indicated the importance of metabolites in shaping the microbiome [31] and have even proposed applying metabolic treatments to cure disrupted microbiota [32]. In addition, a recent review [33] suggests that specific microbial functions within a community (e.g. to signal or to nourish) could be replaced through the provision of significant known microbial metabolites (such as butyrate or lactate), which would act to replace the activities of the endogenous microbiota. Pursuit of such studies requires the availability of data reflecting microbial community evolution over time, including considerations of founding microbial population structure as well as explicit metabolite exchange and variability of the metabolite concentration in the medium over time. Such data, however, are rarely available, and even more rarely in sufficient quantities to support machine-learning studies. To generate the data required for such investigations, we propose GEMs as an appropriate means to simulate microbial communities experiencing specific types of perturbations (i.e. metabolically-oriented). Microbial communities are often stable until some limiting nutrient is exhausted, thus changes in the metabolites within the medium could provoke a disruption in the microbial community, that would then stabilize to a new steady state. We propose, therefore, that at least some stable states seen in real microbial populations, and the transitions between them are the result of dynamic changes in metabolite availability that act to perturb a given stable state. We further propose that perturbations of this type can be simulated in a microbial community modeled with GEMs.

Here we describe the design of a novel computational system, MDPbiomeGEM, which extends our established MDPbiome analytical system to utilize, as input, the results of GEM-based dynamic simulations of microbial communities, including features such as varying population structure, medium composition, and metabolic fluxes over time. We demonstrate the application of MDPbiomeGEM to a variety of different study types. We further show that simulated data, interpreted by MDPbiome, recapitulates existing biomedical knowledge, suggesting that such simulations are capable of generating data that reflects biomedically relevant real-world phenomena. MDPbiomeGEM's primary contribution is to provide an inexpensive, flexible approach to generate microbial population-structure and associated metabolic data, upon which early-stage hypotheses for minimally-invasive microbiome engineering studies may be based, using predictive tools such as MDPbiome.

Comparison with peer approaches

Given the diverse methods and areas of study that MDPbiomeGEM combines (microbiomes, metabolic modeling with GEMs, artificial intelligence models, dynamics in microbial communities, the influence of perturbations on them, etc.), comparison of MDPbiomeGEM with related approaches necessitates the exploration of distinct groups of slightly overlapping study designs and parameters (see Table 1)). In our overview, we include studies using GEMs to model microbiomes ('GEM' column = Yes in Table 1), studies related to the analysis of dynamics in microbial communities ('Static/Dynamic' column = $Dynamic$), and finally, studies of the effects of perturbations ('Perturbations' column = Yes).

As mentioned earlier, most studies combining metabolic models and microbiomes to date have been static. For example, Garza et al. [30] took metagenomics data as input to the GEMs (i.e. the initial population composition was defined by real relative abundance data), and predicted the nutrients in the medium as output. SteadyCom [29] designed a static model with the 9 most-abundant genera in the human gut to simulate the influence of an increase of fiber in the diet. This was done by changing the uptake rate of the fiber metabolite in each independent static run, rather than by modeling the more realistic scenario of constantly adjusting metabolite concentration (as our proposed method does). Steinway and colleagues [34] defined a static community model taking temporal metagenomics data

from mouse gut, perturbed with *C.difficile* and antibiotics. They used boolean models to define 'states' as all possible combinations of the presence or absence of individual strains. Pryor [35] evaluated the influence of metformin in gut microbiota with a GEM-based microbial community, defining a diet that included metformin in the medium composition. Their analysis was also executed as a static model, and disallowed interactions between microbial community members. Special attention must be given to the Thiele group's investigations, that combine hundreds or thousands of GEM models from human gut microbes, in a collection called AGORA [36]. They have also developed metabolic reconstructions for host cells (i.e. human or mouse cells), called Recon models [37]. AGORA and Recons are included in the Virtual Metabolic Human [38], a database collecting many kinds of metabolic-associated data related to nutrients, diseases, and other biomedically interesting features. In one application of their approach [39], they simulate medium composition alterations through modifications in the minimum and maximum allowed bounds in metabolic reactions, following multiple distinct cases over multiple associated static runs [28, 32]. This contrasts with the dynamic runs, with perturbations occurring between time-points, executed by MDPbiomeGEM. Section "**Comparing changes in the metabolome induced by perturbations**" includes a quantitative comparison of MDPbiomeGEM with a similar study from Bauer [32], who made some of their study data available which was then used as the basis for comparison.

In summary, most of the studies described above are static, but allow for large and potentially complex microbial communities (hundreds or thousands of strains) [34, 30, 35, 40]. In contrast, dynamic methods, including the MDPbiomeGEM approach described here, are only capable of modelling smaller, simpler population structures (approx. < 10) [25], but have the advantage of being capable of modelling microbial community changes and responses to perturbations.

With respect to habitat, most microbiome/GEM studies relate to the human gut [41]. This is likely due to the availability of GEMs for human gut bacteria, which have been more extensively studied than those in other habitats such as soil.

Bauer et al. [32] and DiMucci et al. [42] apply two dynamic GEM modeling methods - BacArena [43] and COMETS [44], respectively; however, they nevertheless analyse only static data by only taking into account the final time point. BacArena, using an agent based approach, is able to manage hundreds of GEMs, while COMETS is limited to only a few. Conversely, Popp et al. [45] recently proposed a dynamic approach able to simulate a community with hundreds of GEMs, and follow features over time such as biomass and metabolite concentration. Unfortunately, none of these three dynamic approaches allow for the inclusion of perturbations over the time course of the simulation. Borer et al. [46] provide a dynamic GEM-based approach, focused on model soil habitats, with the distinct feature of modeling spatial properties. However, to model space required simplification of other aspects of the model, for example, avoiding changes in relative biomass between individual species, integrating all bacterial biomass per spatial unit in an unique GEM (running one Flux Balance Analysis (FBA) per unit), avoiding independent functionality per-microbe, and not modelling metabolite exchange between microbes. In addition, their dynamic implementation includes fixed time steps, is not configurable, and is more coarsely-grained than the approach we propose here, which is based on Ordinary Differential Equations. Moreover, they reduce the GEM size (in terms of number of metabolic reactions) while our alternative approach - MMODES (Metabolic Models based Ordinary Differential Equations Simulation; described in Sections '**The MDPbiomeGEM system**' and '**MMODES**') - is able to work with full-size GEMs.

Of the remaining dynamic studies, most are not GEM-based, but rather Bayes [47, 48], Lotka-Volterra [49] or Machine Learning approaches [50]. Shaw et al. [47] models microbiomes experiencing perturbations, but considers only 2 states ('healthy' vs an alternative stable state). This contrasts with the MDPbiomeGEM system which can model $>= 2$ states undergoing perturbations. Venturelli et al. [49] predict microbial community parameters, such as growth rate, building a model to infer time-series data using high-throughput sequencing data as input. Through our use of GEMs, we are additionally able to predict metabolite concentration, and its fluctuations over time, which is the primary goal of other non GEM-based approaches [51, 52, 50].

Finally, with the exception of Bauer's "static" personalized diet recommendations, only MDPbiomeGEM creates an output that includes individualized recommendations regarding the potentially step-wise series of perturbations that would be required to engineer any microbiome from one state to a more desirable state, while avoiding less desirable states.

Results

The MDPbiomeGEM system

MDPbiomeGEM is a novel combination of computational methods, beginning with simulated microbiome dynamics data using GEMs, followed by prediction of the interventions required to direct the engineering of such a microbial community into a desirable state.

Fig. 1 provides a schematic overview of the MDPbiomeGEM system. It consists of two primary components. MMODES is a novel tool that utilizes GEMs to simulate microbial community dynamics, including directed perturbations. Numerous simulation runs, in which distinct, specified perturbations happen randomly over time, generate many unique time series describing the microbial dynamics in each run. These simulated time-series data are then used as input to our previously described MDPbiome system [17] to build a Markov Decision Process (MDP), which uses these to generate state-specific and goal-specific intervention recommendation(s). Detailed descriptions of all components are available in Materials and Methods section.

We will now show examples of MDPbiomeGEM being applied to several distinct scenarios, where we are able to compare its output against well-established biological knowledge in order to judge its utility and/or accuracy.

Suggesting personalized treatments for recovery from gut microbiome-related diseases, such as Crohn's

The relevance of fiber in the diet is highlighted in several studies. For example, a fiber deprivation diet has been reported to alter the mouse gut microbiome, leading to disease risk because microbes consume the gut mucus barrier as their nutrient in place of fiber [53]. Another study analyzes the influence of inulin as a prebiotic in human, varying the amount of dietary fiber intake [54]. A study assessing the impact of separate food groups on microbiome composition was undertaken in a recent publication about dietary influence on the longitudinal microbiome [3]. They conclude that fiber from fruits and grains tend to generate similar microbiome profiles compared to other fibres. Additionally, Johnson et al. [3] highlight the molecular differences between fiber compounds, and the metabolically distinct use of fiber by bacterial species with different metabolic enzymes. These studies highlight the utility of studying the diet/microbiome relationship at the molecular level. In [55] *Faecalibacterium* is suggested as a probiotic and starch as a prebiotic, due to their positive influence in human gut microbiome health. This well-described, straightforward, but multi-faceted model was selected as a validating case-study for the utility of MDPbiomeGEM, as we will now describe.

We selected the microbial community and GEMs described by [56], composed of representatives of two of the three dominant phyla in the human gut (Actinobacteria and Firmicutes): *Bifidobacterium adolescentis L2-32* (iBif452) and *Faecalibacterium prausnitzii A2-165* (iFap484). Both strains have a cross-feeding relation through acetate. Our goal is to model the dynamics of the human gut microbiome associated with Crohn's disease (CD), such as a depletion of *F.prausnitzii* as is symptomatic of this disease [57].

Using MMODES, we generated 94 simulated time-series; an example of a single output time-series is shown in Fig. 2a. Each run differs in its initial relative abundances of the two microbes in the community, covering the full range of possibilities (from healthy to dysbiosis; i.e. higher to lower proportions of *F.prausnitzii*), and each simulation experiences several perturbations randomly selected among a defined set, and randomly occurring (or not) over a series of fixed time-points. The perturbations we defined in this case study involve, first, different fiber sources (considered as prebiotics) that feed one or both strains, specifically: a) inulin (only fermented by *F.prausnitzii*), b) starch (only fermented by *B.adolescentis*) or c) FOS (Fructo-oligo-saccharides) (consumed by both strains). Published GEM models were adjusted to match the fermentation capabilities determined experimentally by [58] (see Materials and Methods).

A second defined perturbation involved the input of *F.prausnitzii* as a probiotic, simulating a sudden small or large increase in the biomass of that strain. *F.prausnitzii* has been reported as a potential next-generation probiotic against Crohn's disease [59, 57].

Taking these MMODES simulated time-series as input, we then use MDPbiome to automatically identify statistically significant, discrete microbiome states in the simulated data. MDPbiome identified three robust states (Fig.2b), which we labelled "healthy", "risky", and "dysbiosis" (corresponding to *F.prausnitzii* concentrations greater

than/˜equal to/less than *B.adolescentis*, respectively). Fig.2c shows the state transitions over time from the simulated time-series, based on these three state classifications (by color). High Short Chain Fatty Acid concentrations (e.g. butyrate) are indicative of a healthier microbiome because they promote bowel homeostasis. Selecting this as the desired outcome, we use MDPbiome to generate a state transition network (Fig.2d), which is then solved - that is, the optimal path to the goal-state is selected - using maximization of butyrate production as the MDPbiome utility function. As shown by the red arrows in Fig.2d, MDPbiome's recommended policy to recover from dysbiosis is to consume inulin as a prebiotic, based on the model's prediction that this will move the community to a state with a higher proportion of *F.prausnitzii* than *B.adolescentis*, and therefore a higher butyrate production. This recommendation is true for both the "dysbiosis" state, and the "risky" state.

Assessment of the Inulin recommendation

MDPbiome defines a stability ratio that measures the reliability of its optimal recommendation. This involves simulating one thousand input transition tables using a Dirichlet distribution, and then measuring the ratio of recommendations that remain the same as the original optimal [17]. The output is a quantitative measure of confidence in the recommendation, and is typically visualized as a bar plot. The stability assessment for the Inulin case study is shown in Fig. 3, where the proportion of the bar taken-up by a given perturbation is indicative of the stability of that recommendation. In this case, the recommendation of Inulin is extremely strong (large salmon-colored areas in Fig. 3a).

Additionally, MDPbiome specifies a metric to determine if the policy is general enough to prescribe it to new subjects. Using a Leave-One-Out Cross Validation, and given that the subjects do not necessarily follow the MDP recommendations, we measure whether the microbiome will transit to a better or worse state when the subject is following versus not following the recommendation. Fig. 3b shows a higher tendency to move to a better or equal state (green plus blue bars) when following (row F) the policy than not following it (row nF), with the tendency to move to a better microbiome state (green bar) being much higher when following the recommendation. Transition to a worse state (red bar) is observed when not following the recommendation, but is almost entirely absent when following the recommendation.

Comparing changes in the metabolome induced by perturbations

In this section we compare our results with those provided by Bauer et al. [32], related to microbial abundances and metabolite concentrations in Crohn's disease (CD) vs healthy subjects. We executed 3 key comparisons: (A) with experimental data (their Figure 3a); (B) with simulated data (and treatment) by Bauer et al. (their Figure 2e-5e); and (C) an additional analysis that is uniquely enabled by MDPbiomeGEM.

Comparison of synthetic versus experimental data

We took experimental microbiome data related to Crohn's disease from [60], that also are shown in [32]-Figure.3a. MMODES synthetic data are in-agreement with those experimental data with respect to relative abundances of the microbes studied: *Bifidobacterium* is higher in CD versus control, and *Faecalibacterium* is higher in control vs. Crohn's. In contrast, Bauer's simulated data shows *Bifidobacterium* higher in the control population - in opposition to both experimental observations, and our simulated data.

Regarding metabolites, both simulated datasets - Bauer et al. and ours - are consistent with experimental data: the butyrate concentration is higher in healthy patients than in CD patients (see Additional file 1). The MDPbiomeGEM ratio more closely matches the experimental data than that of Bauer et al. with respect to butyrate. However, in the case of acetate, the Bauer et al. simulation produces results that more closely match experimental data.

Comparison between MMODES and Bauer et al. simulated datasets

To assist in the comparing MMODES and Bauer et al. synthetic data, we have constructed a set of figures that mirror those of [32]-Figure 2e and 5e, which compares the concentration of multiple metabolites in the two states considered (healthy and disease) and an additional predicted state after a treatment has been applied. Our Additional file 2

represents the same comparison of metabolites in two of our states (healthy and dysbiosis) with the equivalent states in Bauer's study. To mirror Bauer's "treatment" state, we define as our treatment state those samples where the inulin recommendation from MDPbiome has been applied to samples that were previously in the dysbiosis state. Ideally, one would expect that the treatment would move the metabolite concentrations closer to those seen in healthy states. This is observed in the Short Chain Fatty Acids (SCFA) butyrate and formate, and somewhat so in H and H_2O, while acetate shows the opposite behaviour, where the predicted treatment state is farther from the healthy state. In the remaining metabolites there are no remarkable differences. Although our system has the capacity to apply successive perturbations, this scenario cannot be compared to Bauer's data. Bauer et al. concluded that transport reactions for fiber discriminate between healthy and CD subjects. This is in-agreement with our results showing state discrimination with SCFA metabolites (i.e. degradation products from fiber). In conclusion, the MMODES component of MDPbiomeGEM is able to generate comparable simulated data to those obtained by Bauer et al.

Additional analysis contributed by MDPbiomeGEM: Changes in the metabolome induced by perturbations

Beyond the prediction of microbiome state-dependent metabolite concentrations, our system is additionally capable of exploring changes in metabolite concentrations that result from perturbations.

Fig. 4 shows how the concentration of various metabolites in the medium changes over time depending on the perturbation applied in the simulation. From Fig. 4, we can conclude that, as expected, perturbations influence metabolites, and often in distinct ways depending on the nature of the perturbation. For example, the highest butyrate increase happens when using inulin as a perturbation, followed by FOS. This observation is in agreement with the recommendation of inulin by MDPbiome, where butyrate concentration was maximized in the utility function. Inulin also promotes the highest increase in isoleucine concentration. Formate follows a similar pattern to butyrate, with a high increase after inulin perturbation, followed by FOS and finally starch. Starch produces the highest increase in certain other metabolites in the medium, such as ethanol, glycolaldehyde and succinate. Inulin, when measured as a metabolite (see the Inulin column), as expected, increased after addition of inulin as a perturbation, but also notably increases with *F.prausnitzii* as a probiotic. Some metabolites increase notably following all (tested) perturbations, such as maltose and arabinose, while galactose decreases following all perturbations (likely because it is the main carbon source in the simulated media).

Metabolite concentration changes are dependent not only on the perturbation applied, but also on the starting state of the microbiome, as Additional file 3 shows, where panels A, B, and C show the effect of perturbations on Dysbiosis, Healthy, and Risky states, respectively. For example, addition of inulin decreases the concentration of hydrogen sufide, kestose and starch only when applied to the dysbiosis state. Starch promotes a high increase of proline in the dysbiosis state, but less so in risky and healthy states. Similarly, there are clear differences between microbiome states in the concentration ratio changes of certain metabolites, as seen in the L.Proline, L.Isoleucine, L.glutamate and hydrogen sulfide columns in panels A, B, and C.

Therefore, we conclude that the different perturbations induce distinct changes in metabolites in the simulated medium. Our system captures changes in metabolites beyond their dependence on microbiome state, but including those induced by perturbations, thus overcoming a limitation of the Bauer et al. approach. MDPbiomeGEM, therefore, provides a metabolic interpretation of perturbation recommendations.

In conclusion, this section has examined the capabilities of MDPbiomeGEM compared to the most similar peer approach [32] which similarly combines microbiomes and GEMs, and applies them to a common objective: Crohn's disease recovery and dietary interventions. We find that MDPbiomeGEM metabolite concentration prediction correlates well with Bauer's simulation, despite using only two GEMs (versus hundreds); MDPbiomeGEM predicts how to recover from Crohn's disease by focusing on distinct sources of fiber, versus all metabolites as in Bauer et al., and using a dynamic simulation rather than a static one. Finally, MDPbiome additionally provides information related to metabolome changes induced by perturbations.

Retrieving knowledge from metabolic fluxes with Machine Learning

Apart from biomass and metabolite concentration, solving GEMs with mathematical methods (FBA) also returns data about fluxes through metabolic reactions, known as "fluxomics" data. Simulated fluxomics data from GEM

simulations have been proposed as the input to supervised Machine Learning (ML) classifiers to investigate, for example, the prediction of ecological niches (endosphere or rhizosphere) for isolated microbes (one GEM) [61] or the prediction of ecological interactions between pairs of microbes in communities (multiple GEMs) [42]. Similarly, unsupervised ML methods (mainly clustering) have been applied to simulated fluxomics data from GEMs. In these proposals, the input to the clustering becomes the ratio of change of the biomass reaction under differing conditions such as medium composition or gene knock-out/s in the participating organism [62], availability of nutrients such as oxygen, or changes in diet [36]. As such, existing ML approaches primarily utilize one metabolic reaction - the biomass objective function - rather than using all available metabolic reactions.

In this study, we wish to find more informative patterns in the metabolic fluxes resulting from the MMODES simulations. To achieve this, we apply clustering algorithms to the metabolic fluxes of not just one but many reactions, spanning a microbial community (versus a single microbe) and observing fluxes over time rather than at a single point. The goal is to find patterns over time between metabolic fluxes occurring in multiple organisms sharing a common medium. Clustering and retrieval of the most relevant reactions are described in section '**Clustering and identifying relevant reactions**'.

When metabolic reactions from the gut microbiome case study are grouped according to their flux values 5 clusters emerge (Additional file 4). The five medoids correspond to different growth rates of the two microbes: 1) both growing, 2) both on stationary phase, 3 and 4) one or the other growing and 5) *B.adolescentis* growing faster than *F.prausnitzii* (due to a starch perturbation). The most relevant reactions to discriminate between those clusters are shown on the metabolic map in Fig. 5. The most important distinguishing reaction was MTHFD (Methylenetetrahydrofolate dehydrogenase NADP) (Additional file 5 shows the detailed list). Interestingly, most of the reactions are gathered around the butyrate pathway, consistent with the important metabolites identified in the earlier analyses. The metabolic map (Fig. 5) shows that acetate (see ACtex reaction), secreted by *B.adolescentis*, enhances butyrate production (see Ex_Butyrate reactions).

In conclusion, MDPbiomeGEM provides a novel approach to identifying key metabolic factors underlying the population dynamics of microbial communities as they respond, in a coordinated way, to perturbations. This kind of knowledge can be used to, for example, identify druggable targets, or even to predict easily-measured biomarkers of microbial community state in the absence of sequence data or prior biological knowledge.

Soil microbiome pre-/pro-biotics to optimize herbicide degradation

Here we undertake a second case study to demonstrate the flexibility of MDPbiomeGEM by applying it to a completely distinct environment - that of the soil microbiome. Additional contributions from this second case study include: a) extending the number of heterogeneous GEM models MMODES uses in its simulation; b) increasing the complexity of the MDPbiomeGEM output models with more microbiome states and varying recommendations; and c) comparison with available (albeit limited) experimental data.

The goal in this case study is to determine how to perturb the microbiome to accelerate soil detoxification of the herbicide atrazine. The presence of this herbicide in drinking water is linked to hormone disrupting disorders such as Alzheimer, Autism, Parkinson's, and developmental problems in offspring [63, 64]. As such, it is banned in many countries, however not in the United States, where atrazine detoxification is therefore an important issue.

We take a representative of a soil microbial community of three members (*P. aurescens, H. stevensii, Halobacillus sp.*) that degrades atrazine, with GEMs designed and described by [65]. *P. aurescens* is the atrazine degrader, so its presence in the microbial community is mandatory for the degradation to occur. The other two strains influence atrazine degradation according to the experimental results [65]. Using MMODES, we simulate 116 microbial growth time series with different initial composition of the three strains.

We applied several perturbations to the microbial community over time, modifying both the nutrients and microbe biomass; in particular, adding phosphate to the medium, or incrementing the biomass of *H. stevensii* and/or *Halobacillus sp.* - both of which are strains that interact with *P. aurescens* to enhance its degradation activity. MDPbiomeGEM identified 8 microbiome states (see Fig. 6a) with statistically distinct community structures. It then computed the relative probability to move between these states in a perturbation-dependent manner (see Fig. 6b).

We found distinct ratios of atrazine degradation in the eight microbiome states (see black dots in Fig. 6a) following MDPbiome's automated clustering procedure [66]. The ratio of atrazine degradation calculated from our simulated data is not directly correlated with the relative abundance of the main degrader *P. aurescens* - an observation shared with that in original paper. Regarding state transitions, Fig. 6b highlights that phosphate is only a relevant driver in the case of transition from state 7 (with the lowest relative abundance of *P. aurescens*) to state 3 (reaching around 75% of that strain), and is the only starting-state where phosphate addition is the recommendation to move to a higher degradation state. This recommendation also shows a very high stability (see long green bar in row 7 of Fig. 6c). Additional file 7 shows the (large) state-transition diagram arising from this more complex case study.

These results show that MDPbiomeGEM is capable of providing contextually-sensitive predictions (that is, depending on starting microbiome state) for much more complex scenarios than the first case study. In this case, the perturbations that should be applied to improve herbicide degradation: phosphate supplementation (as 'prebiotics') or increasing the biomasses of *Halobacillus sp.* and/or *H. stevensii* (as 'probiotics'). Our system's suggestions are very heterogeneous (different colors per row/state in Fig. 6c) and stable in general (long bars), with a high dependency on the current state of the microbial community, revealing how individualized the recommendations of MDPbiomeGEM can be.

Xu et al. [65] executed in-vitro experiments mirroring the simulation modelled here using MDPbiomeGEM, and concluded that "*Halobacillus sp.* best supported the atrazine degradation followed by the combination *Halobacillus sp.* and *H. stevensii*, and then *H. stevensii* "(see Figure 3c from [65]). The results from MDPbiomeGEM, therefore, are largely in-agreement with that, given that the highest ratios of atrazine degradation (see black dots in Fig. 6a) correspond to the microbiome states 4, 2 and 6, where *P.aurescens* is assisted by both *H. stevensii* and *Halobacillus sp.*; almost exclusively by *Halobacillus sp.*; or primarily by *H. stevensii*, respectively.

Materials and Methods

Genome Scale Metabolic models (GEM)

A Genome Scale Metabolic Model (GEM) is a metabolic reconstruction of a particular organism, that gathers together its genome sequence and annotations from databases through a combination of automatic and manual workflows [67]. A GEM is usually represented in a stoichiometric network with reactions and metabolites. This can be transformed into a mathematical model, enabling the quantitative prediction of phenotype, in terms of fluxes through individual reactions. Flux Balance Analysis (FBA) [68] is the most widely applied method to solve the model, with the typical goal of optimizing cell growth, with linear programming, at a given time step. Constraints Based Reconstruction and Analysis (COBRA) methods [69, 70] are focused on analyzing single metabolic models; however, our work is distinct in that we combine several GEMs to model a microbial community.

There is an increasing interest in combining GEMs to simulate microbial communities [12], with the objectives of understanding community interactions and dynamics in experimentally hard-to-reproduce/manipulate environments (e.g. the innate gut microbiome [71]) and developing comprehensive computational tools to engineer synthetic consortia (e.g. [72]).

MMODES

With the goal of generating dynamic, versus static, models we developed a novel multi-strain dynamic FBA (dFBA) tool to perform dynamic simulations of microbial communities, using GEMs, and taking perturbations over time into account. We called this tool MMODES (Metabolic Models based Ordinary Differential Equations Simulation). The simulated community can be characterized as follows: a) multi-strain; b) dynamic (over time); c) hybrid, updating both biomass and metabolite concentration at each time-point; and d) inclusive of perturbations between time-points. MMODES is based on the approach of Daphne [73] which exhibited many of the properties we desired. Our main distinguishing contributions are: 1) the inclusion of perturbations; 2) concentration updates of specific, rather than *ad hoc*, metabolites; 3) a simplified human interface.

As a formal mathematical postulation, dFBA is a system of Ordinary Differential Equations (ODE), described by equations 1 and 2.

$$\frac{dB_i}{dt} = v_{BM_i} - \delta B_i \tag{1}$$

$$\frac{dC_j}{dt} = \sum_{i=1}^{n} v_j B_i + \varepsilon_j \tag{2}$$

Where B_i is the biomass of model i (g); v_j is the flux of the exchange reaction j or biomass (BM) (mmol/gDW/h); δ is the death rate; C_j is the amount (mmol) of j in the medium; and ε_j is the refreshment (mmol); i.e., the constant addition of a metabolite to the medium over time. Medium and biomass are initialised in the system as concentration units (mmol). Compounds j are automatically computed as the union of all extracellular metabolites of all the models. v_j is constrained at each time step for each model based on C_j of the previous time step.

If the kinetic parameters of uptake rate j are known, Michaelis-Menten kinetics is computed as a lower bound lb_j of v_j. Otherwise, the lower bound is not changed unless it surpasses $C_j B_i$, as shown in Equation 3. Upper bounds up_j are preserved as reported by the original corresponding GEM model.

$$\begin{aligned} & lb_j \geq v_j B_i \geq up_j \\ & lb_j = -\frac{C_j V_{max}}{C_j + K_m} \\ & lb_j \leq C_j B_i \end{aligned} \tag{3}$$

Overall, at each time point, three steps are performed on each model in the community. First, their exchange lower bounds are computed as per equation 3. Second, FBA or parsimonious FBA (pFBA) is performed, retrieving a solution of fluxes. Third, this solution is used to compute equation 1 for the biomass differential and equation 2 for the medium differential. MMODES manages the single GEMs with the COBRApy implementation [69] and solves them with (p)FBA. The order of models is randomized at every time point. The dynamics is implemented in a similar way to that of dynamicFBA, by alternating a phase of solving individual models with a phase of updating the metabolite concentrations in the medium and the strain biomasses.

The choice among different numerical methods was important for a stiff ODE problem, such as the one presented here. In this context, 'stiffness' means that the numerical methods to solve the differential equation are numerically unstable. Hence, in some areas of the solution, the methods approximate to the analytical solution worse than the expected error. In MMODES, we implemented both variable-step and fixed-step solutions. Variable-step solvers were integrated with the FORTRAN wrapper in the SciPy package to enhance performance. The adaptive methods used are 5th order Runge-Kutta and Backward Differentiation Formula (BDF, [74]), the latter being better suited for a stiff problem. In addition, two fixed-step methods were implemented in native Python: Forward Euler Approach (FEA; [75] and 4th order Runge-Kutta (RK4; [75]). Though poorer in reflecting the analytic solution compared to BDF for stiff ODE problems, these final fixed-step methods were considered to be the only option when the ODE system was computationally highly-demanding, as in the atrazine case.

Comparison with GEM community dynamic methods

There are several categories of studies that use GEMs for microbial community modeling [25]. Briefly, these classifications can be simplified into static and dynamic approaches. Static methods make up the majority of studies (e.g. MO-FBA, (d)OptCom, cFBA, CASINO, Mminte, CarveMe, Microbiome modeling toolbox, etc.), with dynamic approaches being rarer and not widely used (e.g. COMETS, BacArena, Daphne, etc.) [27]. In the static methods, the metabolic matrices of individual strains are unified in just one matrix, or alternately, the individual matrices are preserved but are directly interconnected by exchange reactions so they neither allow free metabolite exchange with the medium nor accumulation of metabolites in that medium. Table 2 shows a comparison of the dynamic methods that are able to simulate a multi-strain consortium, using more than one GEM model.

The three main contributions of MMODES are, therefore: (i) the ability to model perturbations over time; (ii) versatility of setting integrator and kinetics parameters (fine-tuning of the Michaelis-Mentens parameters for specific reactions, death rates, refreshment, single-strain and community growth limitations, integrator, and linear programming solver); and (iii) easily extensible, with the ability to run new simulations with just few lines of code. Although

DAPHNE [73] also accepts fine-tuning of some parameters (kinetic, refreshment, lineal programming solver), it does not provide an interface to modify them easily. Compared to BacArena or COMETS, MMODES does not take spatial distribution into account; space it is out of the scope of this work, which exclusively focuses on time series with perturbations.

Beyond the dFBA implementation discussed above, MMODES provides a user-interface to instantiate the medium, to assign known kinetic parameters to specific reactions, to add assorted perturbations, to represent simulations graphically and to provide the output in different formats (COMETS compatible).

MDPbiomeGEM

MDPbiomeGEM is a combination of MMODES with MDPbiome. Technically, to link the output of MMODES to the input of MDPbiomeGEM, MMODES output is transformed into a microbiome standard R phyloseq object [76] - the input for MDPbiome. To generate the OTU matrix of that object, OTU abundances are populated with microbe biomasses arising from the various MMODES time-series (based on the presumption that these two kinds of data are functionally equivalent). The multitude of time points provided by MMODES are filtered out, preserving only one time point corresponding to each simulated perturbation - equivalent to a sampling point in a microbiome longitudinal experiment - usually the step just before or after one of the randomly-occurring perturbations; this assumes that the remaining time-points are associated with community stabilization post-perturbation.

Once the simulated data have been transformed, MDPbiome is then applied in the usual way to identify stable and statistically robust microbiome states using the robust clustering procedure described in [66].

Finally, a utility function should be defined, and provided to MDPbiome such that it knows the goal. The utility function required by MDPbiome, for each microbiome state, is a numeric value representing the "goodness" of that state according to the selected goal [17]. For example, in a disease to healthy time-series, the samples associated to a disease state should have a lower utility value than the samples associated to a healthier state. Usually, the utility value per state is computed as the average of the utility value per samples within that state. For the synthetic data arising from MMODES, the utility function must be defined in terms of the available synthetic data (biomasses, nutrient concentration and metabolic fluxes), most often using expert knowledge about the simulated microbiome. For example, lower vs higher nutrient concentration in a dysbiosis vs healthy reported study, or knowledge about differential metabolic functions activated over time. The concentration increase of a helpful nutrient or the decrease of a harmful one in the medium between two sampling points is often a good utility function.

One final issue relates the number of simulated time-series to run with MMODES to generate sufficient coverage of all possible state-transitions, and their associated perturbations, such that MDPbiome can generate a reliable state-transition map. This, clearly, depends on the number of distinct perturbations, and the number of states identified by MDPbiome's clustering algorithm. As a consequence, in practice, it may require several iterations for MDPbiomeGEM to achieve a robust set of recommendations.

Input Genome Scale Models and adjustments

Human gut microbiome: the 2 models were *i*Bif452 (*Bifidobacterium adolescentis L2-32*) and *i*Fap484 (*Faecalibacterium prausnitizii A2-135*) from [56]. Uptake reactions of glucose for both models were adjusted to 10 mmol/gDW/h. This value represents uptake rate in YCFAG medium (Yeast extract-Casitone-Fatty Acids), supplemented with glucose for *F. prausnitzii* [77] and for minimum uptake rate tested in MCB medim (Medium for Colon Bacteria) for various strains of *Bifidobacterium* [78]. Other carbohydrate source uptakes were scaled accordingly to the number of carbon atoms in glucose. In addition, reactions to enable different documented uptakes of dietary fibers [58] were added: inulin (PubChem CID:16219508) consumption was enabled for *i*Fap484; starch (PubChem CID:24836924) consumption for *i*Bif452, as non-soluble starch that reaches the colon; and kestose (PubChem CID:12304825) consumption in both models, as a representative fructo-oligo-saccharide. Perturbations were added every 8 hours, emulating food intake. Quantitative details of initial medim and perturbations can be found in Additional files 8 and 9. MMODES simulation of this microbial community was computed with pFBA and BDF.

Soil microbiome: the GEMs used in this consortium were kindly provided by the authors of [65], since with the exception of *P. aurescens TC1* [79] they had not yet been made publicly available. The minimum mineral

medium described in the original publication supplemented with atrazine (PubChem CID: 2256) was used in this case. To provide input for MDPbiome, 116 runs of 5 hours with randomized initial consortia and perturbations were performed. Quantitative details of initial medium (minimum or root exudate-like simulating proximity to the plant) and perturbations can be found in Additional files 10 and 11. MMODES simulation of this microbial community was computed with FBA and FEA.

Clustering and identifying relevant reactions

Given that GEMs have hundreds or thousands of reactions per model, and we have several models and multiple time points, preprocessing steps to filter-out some reactions should be done before clustering fluxes. PCA for dimensionality reduction of features has been applied previously to metabolic fluxes [80]. However, we are interested in the interpretation of the results based on the original metabolic reactions, which is not possible with PCA due to its change of feature representation to a new latent space. As such, a different approach was necessary.

First, before clustering fluxes, we select one hundred equidistant points, uniformly distributed over the time-series for every run. Reactions with a constant value over time are removed. We remove reactions that provide redundant metabolic knowledge as follows: Transport and exchange reactions are eliminated, preserving the reactions of the inner metabolism for which the proposed interpretations will have a stronger grounding. We then preserve only one representative of coupled reactions, i.e., those with the same value among all samples and belonging to the same pathway. This allows us to reduce data complexity with a minimal loss of information. Flux values of the resulting subset of reactions are rescaled between -1 and 1.

Second, metabolic reactions are grouped based on flux values over time and samples using unsupervised machine learning (clustering), following the criteria in [66] to select the optimal number of clusters.

Third, predicted flux-states are used as labels for a Random Forest classifier [81] and feature importance is computed; i.e., the reactions that best described the differences between flux-states. The same protocol has been previously used in metabolic flux analysis, for guiding the refinement of GEMs [82]. Variable importance was measured using the Mean Decrease in Accuracy (MDA) [83]. The number of trees constructed was 1000 and the minimum terminal node size was 42.

Ratio of metabolite increase

In our comparison with Bauer et al. [32] regarding metabolome changes (section '**Comparing changes in the metabolome induced by perturbations**'), we manually extracted data from Bauer et al. figures, text and their references to compare with our results because machine-readable data are not available. Regarding the metabolite concentration in different states, or after a perturbation, we chose to compute the increase (or associated ratio of increase) rather than the absolute concentration in mM as [32] does, since that would depend on the initial concentration and biomass of strains. We compute the ratio of metabolite increase as the concentration at the beginning of the subsequent state, divided by the concentration at the beginning of the current state. Thus, a ratio equals to 1 means no modification in the concentration during that period, less than 1 means a decrease in the concentration and greater than 1 means an increase in the concentration. This computation is primarily applied to comparisons of butyrate concentration, in our study.

Discussion

Our system MMODES is able to simulate a variety of guided changes in a microbial community over time, that can be grouped in two main kinds of perturbations: a) increase/decrease in a metabolite's concentration in the medium, b) increase/decrease the biomass of a specific microbe. The former could be considered as a simulation of the use of prebiotics, and the latter as a simulation of probiotics or a knock-out. In addition, synbiotic (pre and pro-biotics at the same time) can also easily be modeled; for example, in the human gut microbiome case study, increasing the amount of inulin and *F.prausnitzii* simultaneously. Similarly, targeted antibiotics could be considered perturbations that reduce the biomass of specific community members. Different combinations of basic perturbations, such as different metabolites in different amounts, could be also modeled with our system (e.g. the distinct fiber types in the

proportion they are in a particular food, as Wong et al. [84] claims). The trade-off when considering these possibilities is that the greater the number of different combinations, the greater the number of time series needed to adequately fill the MDP transition table in MDPbiome. Finally, our system is distinct from all other approaches in its ability to model successive perturbations, while the few peer approaches that model perturbations (Table 1) allow a maximum of one.

Our first case study was based on the distinct chemical composition of different fibers that can be metabolized by distinct gut microbes. Adjustment in the concentration of these fibers affects the relative biomass of the bacteria in the community, which in turn affects the biochemistry of the gut environment that can result in inflammatory bowel diseases (IBD) such as Crohn's. Our results are consistent with what is known about Crohn's disease evolution. For example, a depletion in *F.prausnitzii* is correlated with Crohn's disease [57, 59] and an increment in *F.prausnitzii* is associated with an improvement in IBD symptoms [58]. Biological studies support our system's prediction regarding a positive influence of fiber on the gut microbiome. For example, modulating the gut microbiome with dietary supplements [85], demonstrated that fiber is a promotor of the increase of butyrate and *Faecalibacterium* in the mouse caecum. It would appear, then, that in our first case study MDPbiomeGEMs predictions are supported by prior biomedical knowledge regarding how different amounts and types of fiber influence the microbial community. Nevertheless, the specific scenario modeled in that case study - where distinct fiber sources are administered, in differing concentrations - has never been undertaken experimentally [84], because food and even prebiotics mix different types of fiber. As such, we propose that MDPbiomeGEM can assist with the planning and design of novel intervention studies through enabling precisely controlled, accurate simulations that would be difficult to mimic through *in vivo* experiments. This will enable better tuning of real-world intervention experiments prior to testing them, reducing risk, and saving both time and money.

The second, more complex case study demonstrates additional properties of MDPbiomeGEM, including its ability to model more complex microbial populations, identify numerous stable states in a dynamic microbiome, predict the metabolic capacity of each state, and generate state-specific intervention recommendations (in contrast with the single recommendation for every situation that arose from Crohn's use case). In addition, this case study differs from the Crohn's case in the structure of the input data - that is, the use of more but shorter time series; i.e. fewer sampling points over more participants. This reveals the versatility of MMODES, which is capable of simulating a wide variety of experimental designs.

MDPbiomeGEM is flexible in the kinds of intervention outcomes that can be discovered because of flexibility in, for example, the definition of the perturbations, the utility function (e.g. butyrate concentration or atrazine degradation), and the microbes within the community. The primary limitation of MDPbiomeGEM is the availability (and accuracy) of GEMs for the selected microbes of interest. The metabolic networks for commonly-studied microbes, such as those in the human gut, have often been constructed as a GEM, however bacteria from the soil microbiome are scarce in the global GEM databases such as BiGG [86]. In addition, these metabolic networks often require additional case-specific fine-tuning, as was done with the GEMs provided by [65], which in turn requires real-world data and/or metabolic knowledge that may or may not be available. This affects the accuracy/reliability of the MMODES simulation outputs, and thus of the final MDPbiomeGEM output.

Our study was also hampered by a more global curatorial problem, that needs to be solved quickly by the microbial research community. Microbial community analyses using metabolic modeling requires that all models be sharing the identifiers for common compounds in order to enable cross-talk among them. This lack of standardization limits the rapid and accurate synthesis of the desired configurations of microorganisms in a simulated community, even when models are available. Community-efforts such as Memote [87], a quality validator of GEMs, are assisting with GEM nomenclature standardization. Moreover, the National Microbiome Data Collaborative (https://microbiomedata.org/) recently established a GO FAIR Implementation Network (https://www.go-fair.org/implementation-networks/) to act as an umbrella organization, coordinating global efforts to bring microbiome data in-line with the requirements of the FAIR Data Principles [88].

Modeling dynamics of a microbial community, using GEM approaches, is inherently difficult. GEMs were originally designed to study steady-state events, although GEMs have been extended to be applied in dynamic contexts, such

as this work, following the assumptions of dFBA and ODE. However, using GEMs in dynamic simulations requires reliable experimental data (i.e. enzymatic) such that the models can be reliably parameterized, and such data is often lacking [89]. In addition, dynamic simulations naturally invoke a higher computational cost than static simulations. As such, simplifying offsets and counterbalances are usually implemented, which in-turn simplifies the overall model of the microbial community. Many of the systems collected in Table 1 claim to study dynamic events, however in practice they simplify this to one static point under distinct environmental conditions. Other studies reduce the reactions and/or metabolites in the GEM to only the core metabolic network. Other approaches unify various GEMs, avoiding nutrient exchange among them or with the medium. At the cost of high computational expense, our approach preserves all of these dynamic elements to help ensure models are as close as possible to physiological conditions. The main simplification in MMODES is, therefore, related to community size. MMODES is, as a practical matter, limited to a small number of community members (<=10 in practice, a number considered reasonable by [25]), where the cost of increasing the community size is not only computational, but also inherent in the complexity that each additional GEM adds to the overall experiment; increasing size requires GEM discovery/creation, curation, nomenclature harmonization, medium definition and integration with the other GEMs in the community. For reference, we evaluated the performance of MMODES running a simulation of a microbial community with up to 10 homogeneous strains of *E.coli* - a fairly large and complex GEM - and we report our results in Additional file 6. Nevertheless, the practical scale of an MMODES simulation will depend on configuration parameters such as the number of strains, the length of the time series, the size of each GEM and the method chosen to solve the ODE (see section '**MMODES**'). Thus, MMODES would be optimally used for experiments involving communities where there is a well-understood interaction between 2-5 microbes, in well-curated GEMs, with well-defined perturbations, and where a well-defined and quantitatively measurable biomedical/biochemical objective.

Given the limited availability of in-vivo fluxomics data [90], the in-silico fluxomics simulated data provided by GEMs are very valuable and provided at lower cost than experimental techniques [91]. However, although FBA returns quantitative data, it is very important to highlight that the certainty/reliability of these predicted values will depend on the quality of the GEM that generates fluxes, and on the biological knowledge on which the model was based (that may be incomplete). Given these limitations, we believe that qualitative analyses of fluxomics data from GEMs are likely more reliable than quantitative ones. Our system attempts to avoid limits in precision through finding clusters of metabolic behaviour (flux-states), and comparing fluxome data under different conditions or at different time points, rather than attempting to quantitatively define a single individual phenotype.

Finally, MMODES simulated data provides numerous advantages versus executing *in vitro* or *in vivo* microbiome experiments, such as: homogeneous sampling points; time series of arbitrary length; no limits on the number of subjects; well-defined and quantitative perturbations of many kinds (nutrient changes, increase in abundance of strains, addition of new strains, antibiotic application, pathogen infection, etc); allowing any combination of those perturbations; and the ability to target a subset of strains of interest or focus on known metabolic functions. Many of these features are difficult or impossible to do in non-simulated scenarios, and thus provide a unique opportunity to address detailed questions about microbiome community evolution, as a step towards planning real-world microbiome engineering efforts.

Conclusions

Cao et al. [92] highlights the necessity for 'predictive modeling frameworks and tools for precise manipulation of microbiome behaviors. MDPbiomeGEM addresses this challenge.

The main contributions of MDPbiomeGEM are: 1) integration of distinct GEMs into a dynamic microbial community simulation (MMODES); 2) The inclusion of iterative, combinatorial perturbations during a microbial community dynamics simulation; 3) novel description of distinct microbiome states based on their metabolic fluxes; 4) Analysis/visualization of perturbed community growth dynamics; and 5) further validation of the ability of MDPbiome to generate accurate intervention predictions.

We found that the output model and policy from MDPbiome was consistent with both manual examination of the simulated data, and with biomedical knowledge. Thus we suggest that MDPbiomeGEM may be an effective tool

for filtering non-optimal solutions, or making predictions during the planning phases of a microbiome engineering initiative in an industrial setting, providing the ability to minimize resource usage on failed experiments. Moreover, it provides a method for generating contextually-sensitive intervention plans (e.g. for personalized medicine or precision agriculture) in the absence of sufficient experimentally-derived metagenomics data.

Finally, MDPbiomeGEM provides the ability to predict novel biomarkers, by enabling iterative simulation/state-identification/disease-association until a strong predicted association is identified. These associations may then be studied in more depth at the metabolic level, potentially revealing novel biomarkers based on metabolic activities and/or the presence of metabolites in the medium. Such cheap and low-intervention biomarkers extremely useful for precision medicine, as exemplified by the recently published urinary biomarker for irritable bowel syndrome [93]). Thus, application of MDPbiomeGEM could eventually lead to less invasive or environmentally destructive practices in medicine and agriculture respectively.

Declarations

Acknowledgements
Thanks to the authors of [65] for kindly providing us with the GEMs of soil microbiome case study.

Authors' contributions
BGJ conceived the idea and designed the study; BGJ, MDW and JM evaluated the outcomes; BGJ, MDW, JC and JM wrote/edited the manuscript. JC (MMODES) and BGJ wrote the software, executed the experiments, and analysed the data. All authors read and approved the final manuscript.

Funding
This research was supported by the Isaac Peral cofund award to MDW from the Universidad Politécnica de Madrid; and by the"Severo Ochoa Program for Centres of Excellence in R&D"from the Agencia Estatal de Investigación of Spain (grant SEV-2016-0672 (2017-2021)) to the CBGP, and the National Institute for Agriculture and Food Research and Technology (INIA) (RTA2015-00014-c02-01 to JM) and UE Prima (PCI2019-103610 to JM). BGJ was supported by a Postdoctoral contract associated to the Severo Ochoa Program.

Availability of Data and Materials
The datasets used and generated during the current study are available in the Zenodo repository, 10.5281/zenodo.3667601
MDPbiomeGEM software is available at Github (https://github.com/beatrizgj/MDPbiomeGEM) and at Docker hub. MMODES is a novel Python3 package available at Github (https://github.com/carrascomj/mmodes) and at pip (https://pypi.org/project/mmodes/).

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
aff1[0]Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA)
Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Campus de Montegancedo-UPM 28223 Pozuelo de Alarcón (Madrid) Spain. aff2[0]Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid (UPM) 28040 Madrid Spain.

References

1. Kuntz, T.M., Gilbert, J.A.: Introducing the Microbiome into Precision Medicine. Trends in Pharmacological Sciences **38**(1), 81–91 (2017). doi:10.1016/j.tips.2016.10.001
2. Andreote, F.D., Pereira e Silva, M.d.C.: Microbial communities associated with plants: learning from nature to apply it in agriculture **37**, 29–34 (2017). doi:10.1016/j.mib.2017.03.011
3. Johnson, A.J., Vangay, P., Al-Ghalith, G.A., Hillmann, B.M., Ward, T.L., Shields-Cutler, R.R., Kim, A.D., Shmagel, A.K., Syed, A.N., Walter, J., Menon, R., Koecher, K., Knights, D., Knights, D.: Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. Cell Host & Microbe **25**(6), 789–8025 (2019). doi:10.1016/j.chom.2019.05.005
4. Zhong, H., Penders, J., Shi, Z., Ren, H., Cai, K., Fang, C., Ding, Q., Thijs, C., Blaak, E.E., Stehouwer, C.D.A., Xu, X., Yang, H., Wang, J., Wang, J., Jonkers, D.M.A.E., Masclee, A.A.M., Brix, S., Li, J., Arts, I.C.W., Kristiansen, K.: Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. Microbiome **7**(1), 2 (2019). doi:10.1186/s40168-018-0608-z
5. Toward Novel Therapeutics Via Directed Modeling of the Gut Microbiome. In: American Chemical Society (ACS) Fall 2019 National Meeting & Exposition (2019). `https://plan.core-apps.com/acs_sd2019/abstract/79fd34f0-461e-452f-9bc8-3930f37975a0`
6. Valencia, E., Gross, N., Quero, J.L., Carmona, C.P., Ochoa, V., Gozalo, B., Delgado-Baquerizo, M., Dumack, K., Hamonts, K., Singh, B.K., Bonkowski, M., Maestre, F.T.: Cascading effects from plants to soil microorganisms explain how plant species richness and simulated climate change affect soil multifunctionality. Global Change Biology **24**(12), 5642–5654 (2018). doi:10.1111/gcb.14440
7. Yuan, Y., Brunel, C., van Kleunen, M., Li, J., Jin, Z.: Salinity-induced changes in the rhizosphere microbiome improve salt tolerance of Hibiscus hamabo. Plant and Soil **443**(1-2), 525–537 (2019). doi:10.1007/s11104-019-04258-9
8. Vergnes, S., Gayrard, D., Veyssière, M., Toulotte, J., Martinez, Y., Dumont, V., Bouchez, O., Rey, T., Dumas, B.: Phyllosphere Colonization by a Soil Streptomyces sp. Promotes Plant Defense Responses Against Fungal Infection. Molecular Plant-Microbe Interactions®, 05–190142 (2019). doi:10.1094/MPMI-05-19-0142-R
9. Yanardağ, I.H., Zornoza, R., Bastida, F., Büyükkiliç-Yanardağ, A., García, C., Faz, A., Mermut, A.R.: Native soil organic matter conditions the response of microbial communities to organic inputs with different stability. Geoderma **295**, 1–9 (2017). doi:10.1016/J.GEODERMA.2017.02.008

10. van den Bogert, B., Boekhorst, J., Pirovano, W., May, A.: On the Role of Bioinformatics and Data Science in Industrial Microbiome Applications. Frontiers in Genetics **10**, 721 (2019). doi:10.3389/fgene.2019.00721
11. Jiang, L.-L., Zhou, J.-J., Quan, C.-S., Xiu, Z.-L.: Advances in industrial microbiome based on microbial consortium for biorefinery. Bioresources and bioprocessing **4**(1), 11 (2017). doi:10.1186/s40643-017-0141-0
12. Eng, A., Borenstein, E.: Microbial community design: methods, applications, and opportunities. Current Opinion in Biotechnology **58**, 117–128 (2019). doi:10.1016/j.copbio.2019.03.002
13. Young, V.B.: The role of the microbiome in human health and disease: An introduction for clinicians. BMJ (Online) **356** (2017). doi:10.1136/bmj.j831
14. Zhang, C., Zhang, Y., Ding, Z., Bai, Y.: Contribution of Microbial Inter-kingdom Balance to Plant Health. Molecular Plant **12**(2), 148–149 (2019). doi:10.1016/j.molp.2019.01.016
15. Rist, V.T.S., Weiss, E., Eklund, M., Mosenthin, R.: Impact of dietary protein on microbiota composition and activity in the gastrointestinal tract of piglets in relation to gut health: A review. Animal **7**(7), 1067–1078 (2013). doi:10.1017/S1751731113000062
16. Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J.: Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology **25**, 56–66 (2015). doi:10.1016/j.mib.2015.04.004
17. García-Jiménez, B., de la Rosa, T., Wilkinson, M.D.: MDPbiome: microbiome engineering through prescriptive perturbations. Bioinformatics **34**(17), 838–847 (2018). doi:10.1093/bioinformatics/bty562
18. Singh, R.K., Chang, H.W., Yan, D., Lee, K.M., Ucmak, D., Wong, K., Abrouk, M., Farahnik, B., Nakamura, M., Zhu, T.H., Bhutani, T., Liao, W.: Influence of diet on the gut microbiome and implications for human health. Journal of Translational Medicine **15**(1) (2017). doi:10.1186/s12967-017-1175-y
19. Maier, T.V., Lucio, M., Lee, L.H., VerBerkmoes, N.C., Brislawn, C.J., Bernhardt, J., Lamendella, R., McDermott, J.E., Bergeron, N., Heinzmann, S.S., Morton, J.T., González, A., Ackermann, G., Knight, R., Riedel, K., Krauss, R.M., Schmitt-Kopplin, P., Jansson, J.K.: Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome. mBio **8**(5) (2017). doi:10.1128/mBio.01343-17
20. Hartman, K., Tringe, S.G.: Interactions between plants and soil shaping the root microbiome under abiotic stress. Biochemical Journal **476**(19), 2705–2724 (2019). doi:10.1042/BCJ20180615
21. Rojo, D., Méndez-García, C., Raczkowska, B.A., Bargiela, R., Moya, A., Ferrer, M., Barbas, C.: Exploring the human microbiome from multiple perspectives: factors altering its composition and function. FEMS Microbiology Reviews **41**(4), 453–478 (2017). doi:10.1093/femsre/fuw046
22. Berendsen, R.L., Pieterse, C.M.J., Bakker, P.A.H.M.: The rhizosphere microbiome and plant health. Trends in Plant Science **17**(8), 478–486 (2012). doi:10.1016/j.tplants.2012.04.001
23. Thiele, I., Palsson, B.Ø.: A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols **5**(1), 93–121 (2010). doi:10.1038/nprot.2009.203
24. Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y.: Current status and applications of genome-scale metabolic models. Genome Biology **20**(1), 121 (2019). doi:10.1186/s13059-019-1730-3
25. Perez-Garcia, O., Lear, G., Singhal, N.: Metabolic Network Modeling of Microbial Interactions in Natural and Engineered Environmental Systems. Frontiers in Microbiology **7**, 673 (2016). doi:10.3389/fmicb.2016.00673
26. Gottstein, W., Olivier, B.G., Bruggeman, F.J., Teusink, B.: Constraint-based stoichiometric modelling from single organisms to microbial communities. Journal of The Royal Society Interface **13**(124) (2016)
27. Kim, O.D., Rocha, M., Maia, P.: A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering. Frontiers in Microbiology **9**, 1690 (2018). doi:10.3389/fmicb.2018.01690
28. Heinken, A., Thiele, I.: Anoxic conditions promote species-specific mutualism between gut microbes in silico. Applied and Environmental Microbiology **81**(12), 4049–4061 (2015). doi:10.1128/AEM.00101-15. https://aem.asm.org/content/81/12/4049.full.pdf
29. Chan, S.H.J., Simons, M.N., Maranas, C.D.: SteadyCom: Predicting microbial abundances while ensuring community stability. PLoS computational biology **13**(5), 1005539 (2017). doi:10.1371/journal.pcbi.1005539
30. Garza, D.R., van Verk, M.C., Huynen, M.A., Dutilh, B.E.: Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nature Microbiology **3**(4), 456–460 (2018). doi:10.1038/s41564-018-0124-8
31. Schluter, J., Foster, K.R.: The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. PLoS Biology **10**(11), 1001424 (2012). doi:10.1371/journal.pbio.1001424
32. Bauer, E., Thiele, I.: From metagenomic data to personalized in silico microbiotas: predicting dietary supplements for crohn's disease. npj Systems Biology and Applications **4**(1), 27 (2018). doi:10.1038/s41540-018-0063-2
33. Dominguez Bello, M.G., Knight, R., Gilbert, J.A., Blaser, M.J.: Preserving microbial diversity: Microbiota from humans of all cultures are needed to ensure the health of future generations. Science **362**(6410), 33–34 (2018). doi:10.1126/science.aau8816
34. Steinway, S.N., Biggs, M.B., Loughran, T.P., Papin, J.A., Albert, R.: Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. PLOS Computational Biology **11**(6), 1004338 (2015). doi:10.1371/journal.pcbi.1004338
35. Pryor, R., Norvaisas, P., Marinos, G., Best, L., Thingholm, L.B., Quintaneiro, L.M., De Haes, W., Esser, D., Waschina, S., Lujan, C., Smith, R.L., Scott, T.A., Martinez-Martinez, D., Woodward, O., Bryson, K., Laudes, M., Lieb, W., Houtkooper, R.H., Franke, A., Temmerman, L., Bjedov, I., Cochemé, H.M., Kaleta, C., Cabreiro, F.: Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy. Cell **178**(6), 1299–131229 (2019). doi:10.1016/j.cell.2019.08.003
36. Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D.A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R.M.T., Thiele, I.: Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nature biotechnology **35**(1), 81–89 (2017). doi:10.1038/nbt.3703
37. Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., Thorleifsson, S.G., Agren, R., Bölling, C., Bordel, S., Chavali, A.K., Dobson, P., Dunn, W.B., Endler, L., Hala, D., Hucka, M., Hull, D., Jameson, D., Jamshidi, N., Jonsson, J.J., Juty, N., Keating, S., Nookaew, I., Le Novère, N., Malys, N., Mazein, A., Papin, J.A., Price, N.D., Selkov, E., Sigurdsson, M.I., Simeonidis, E., Sonnenschein, N., Smallbone, K., Sorokin, A., van Beek, J.H.G.M., Weichart, D., Goryanin, I., Nielsen, J., Westerhoff, H.V., Kell, D.B., Mendes, P., Palsson, B.Ø.: A community-driven global reconstruction of human metabolism. Nature Biotechnology **31**(5), 419–425 (2013). doi:10.1038/nbt.2488
38. Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D.A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., Puente, A., Rodrigues, M., Roy, A., Rouquaya, M., Wiltgen, L., Žagare, A., John, E., Krueger, M., Kuperstein, I., Zinovyev, A., Schneider, R., Fleming, R.M.T., Thiele, I.: The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Research **47**(D1), 614–624 (2018). doi:10.1093/nar/gky992
39. Baldini, F., Heinken, A., Heirendt, L., Magnusdottir, S., Fleming, R.M.T., Thiele, I.: The Microbiome Modeling Toolbox: from microbial interactions to personalized microbial communities. Bioinformatics **35**(13), 2332–2334 (2019). doi:10.1093/bioinformatics/bty941
40. Diener, C., Gibbons, S.M., Resendis-Antonio, O.: MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. mSystems **5**(1) (2020). doi:10.1128/mSystems.00606-19
41. Sen, P., Orešič, M.: Metabolic modeling of human gut microbiota on a genome scale: An overview. Metabolites **9**(2) (2019). doi:10.3390/metabo9020022
42. DiMucci, D., Kon, M., Segrè, D.: Machine Learning Reveals Missing Edges and Putative Interaction Mechanisms in Microbial Ecosystem Networks. mSystems **3**(5) (2018). doi:10.1128/msystems.00181-18
43. Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., Kaleta, C.: BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex

communities. PLOS Computational Biology **13**(5), 1005544 (2017). doi:10.1371/journal.pcbi.1005544
44. Harcombe, W., Riehl, W., Dukovski, I., Granger, B., Betts, A., Lang, A., Bonilla, G., Kar, A., Leiby, N., Mehta, P., Marx, C., Segrè, D.: Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial Dynamics. Cell Reports **7**(4), 1104–1115 (2014). doi:10.1016/j.celrep.2014.03.070
45. Popp, D., Centler, F.: µbialSim: constraint-based dynamic simulation of complex microbiomes. bioRxiv, 716126 (2019). doi:10.1101/716126
46. Borer, B., Ataman, M., Hatzimanikatis, V., Or, D.: Modeling metabolic networks of individual bacterial agents in heterogeneous and dynamic soil habitats (IndiMeSH). PLoS Computational Biology **15**(6) (2019). doi:10.1371/journal.pcbi.1007127
47. Shaw, L.P., Bassam, H., Barnes, C.P., Walker, A.S., Klein, N., Balloux, F.: Modelling microbiome recovery after antibiotics using a stability landscape framework. ISME Journal **13**(7), 1845–1856 (2019). doi:10.1038/s41396-019-0392-1
48. Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., Deng, L., Yeliseyev, V., Delaney, M.L., Liu, Q., Olle, B., Stein, R.R., Honda, K., Bry, L., Gerber, G.K.: MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome Biology **17**(1), 121 (2016). doi:10.1186/s13059-016-0980-6
49. Venturelli, O.S., Carr, A.V., Fisher, G., Hsu, R.H., Lau, R., Bowen, B.P., Hromada, S., Northen, T., Arkin, A.P.: Deciphering microbial interactions in synthetic human gut microbiome communities. Molecular Systems Biology **14**(6) (2018). doi:10.15252/msb.20178157
50. Costello, Z., Martin, H.G.: A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. npj Systems Biology and Applications **4**(1), 19 (2018). doi:10.1038/s41540-018-0054-3
51. Mallick, H., Franzosa, E.A., Mclver, L.J., Banerjee, S., Sirota-Madi, A., Kostic, A.D., Clish, C.B., Vlamakis, H., Xavier, R.J., Huttenhower, C.: Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature Communications **10**(1) (2019). doi:10.1038/s41467-019-10927-1
52. Morton, J.T., Aksenov, A.A., Nothias, L.F., Foulds, J.R., Quinn, R.A., Badri, M.H., Swenson, T.L., Van Goethem, M.W., Northen, T.R., Vazquez-Baeza, Y., Wang, M., Bokulich, N.A., Watters, A., Song, S.J., Bonneau, R., Dorrestein, P.C., Knight, R.: Learning representations of microbe–metabolite interactions. Nature Methods (2019). doi:10.1038/s41592-019-0616-3
53. Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., Young, V.B., Henrissat, B., Wilmes, P., Stappenbeck, T.S., Núñez, G., Martens, E.C.: A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell **167**(5), 1339–135321 (2016). doi:10.1016/j.cell.2016.10.043
54. Healey, G., Murphy, R., Butts, C., Brough, L., Whelan, K., Coad, J.: Habitual dietary fibre intake influences gut microbiota response to an inulin-type fructan prebiotic: A randomised, double-blind, placebo-controlled, cross-over, human intervention study. British Journal of Nutrition **119**(2), 176–189 (2018). doi:10.1017/S0007114517003440
55. Sanders, M.E., Merenstein, D.J., Reid, G., Gibson, G.R., Rastall, R.A.: Probiotics and prebiotics in intestinal health and disease: from biology to the clinic. Nature reviews. Gastroenterology & hepatology **16**(10), 605–616 (2019). doi:10.1038/s41575-019-0173-3
56. El-Semman, I.E., Karlsson, F.H., Shoaie, S., Nookaew, I., Soliman, T.H., Nielsen, J.: Genome-scale metabolic reconstructions of bifidobacterium adolescentis l2-32 and faecalibacterium prausnitzii a2-165 and their interaction. BMC Systems Biology **8**(1), 41 (2014). doi:10.1186/1752-0509-8-41
57. Martín, R., Miquel, S., Benevides, L., Bridonneau, C., Robert, V., Hudault, S., Chain, F., Berteau, O., Azevedo, V., Chatel, J.M., Sokol, H., Bermúdez-Humarán, L.G., Thomas, M., Langella, P.: Functional characterization of novel faecalibacterium prausnitzii strains isolated from healthy volunteers: A step forward in the use of f. prausnitzii as a next-generation probiotic. Frontiers in Microbiology **8** (2017)
58. Rios-Covian, D., *et al.*: Enhanced butyrate formation by cross-feeding betweenFaecalibacterium prausnitziiandBifidobacterium adolescentis. FEMS Microbiology Letters **362**(21), 176 (2015). doi:10.1093/femsle/fnv176
59. Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottiere, H.M., Dore, J., Marteau, P., Seksik, P., Langella, P.: Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of crohn disease patients. Proceedings of the National Academy of Sciences **105**(43), 16731–16736 (2008). doi:10.1073/pnas.0804812105
60. Lewis, J.D., Chen, E.Z., Baldassano, R.N., Otley, A.R., Griffiths, A.M., Lee, D., Bittinger, K., Bailey, A., Friedman, E.S., Hoffmann, C., Albenberg, L., Sinha, R., Compher, C., Gilroy, E., Nessel, L., Grant, A., Chehoud, C., Li, H., Wu, G.D., Bushman, F.D.: Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host and Microbe **18**(4), 489–500 (2015). doi:10.1016/j.chom.2015.09.008
61. Chien, J., Larsen, P.: Predicting the Plant Root-Associated Ecological Niche of 21 Pseudomonas Species Using Machine Learning and Metabolic Modeling (2017). 1701.03220
62. Segrè, D., DeLuna, A., Church, G.M., Kishony, R.: Modular epistasis in yeast metabolism. Nature Genetics **37**(1), 77–83 (2005). doi:10.1038/ng1489
63. Gammon, D.W., Aldous, C.N., Carr, W.C., Sanborn, J.R., Pfeifer, K.F.: A risk assessment of atrazine use in California: Human health and ecological aspects. Pest Management Science **61**(4), 331–355 (2005). doi:10.1002/ps.1000
64. Wirbisky, S.E., Weber, G.J., Sepúlveda, M.S., Lin, T.L., Jannasch, A.S., Freeman, J.L.: An embryonic atrazine exposure results in reproductive dysfunction in adult zebrafish and morphological alterations in their offspring. Scientific Reports **6** (2016). doi:10.1038/srep21337
65. Xu, X., Zarecki, R., Medina, S., Ofaim, S., Liu, X., Chen, C., Hu, S., Brom, D., Gat, D., Porob, S., Eizenberg, H., Ronen, Z., Jiang, J., Freilich, S.: Modeling microbial communities from atrazine contaminated soils promotes the development of biostimulation solutions. The ISME journal **13**(2), 494–508 (2019). doi:10.1038/s41396-018-0288-5
66. García-Jiménez, B., Wilkinson, M.D.: Robust and automatic definition of microbiome states. PeerJ **2019**(3) (2019). doi:10.7717/peerj.6657
67. Bordbar, A., Monk, J.M., King, Z.A., Palsson, B.O.: Constraint-based models predict metabolic and associated cellular functions. Nature reviews. Genetics **15**(2), 107–20 (2014). doi:10.1038/nrg3643
68. Orth, J.D., Thiele, I., Palsson, B.Ø.: What is flux balance analysis? Nature biotechnology **28**(3), 245–8 (2010). doi:10.1038/nbt.1614
69. Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R.: COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology **7**(1), 74 (2013). doi:10.1186/1752-0509-7-74
70. Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Žagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., El Assal, D.C., Valcarcel, L.V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., El Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H.S., Bryant, W.A., Aragón Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B., Thiele, I., Fleming, R.M.T.: Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nature Protocols **14**(3), 639–702 (2019). doi:10.1038/s41596-018-0098-2. 1710.04038
71. Shoaie, S., Nielsen, J.: Elucidating the interactions between the human gut microbiota and its host through metabolic modeling. Frontiers in Genetics **5**(APR) (2014). doi:10.3389/fgene.2014.00086
72. García-Jiménez, B., García, J.L., Nogales, J.: FLYCOP: Metabolic modeling-based analysis and engineering microbial communities. Bioinformatics **34**(17), 954–963 (2018). doi:10.1093/bioinformatics/bty561
73. Succurro, A., Segrè, D., Ebenhöh, O.: Emergent subpopulation behavior uncovered with a community dynamic metabolic model of escherichia coli diauxic growth. mSystems **4**(1) (2019)
74. Curtiss, C.F., Hirschfelder, J.O.: Integration of Stiff Equations. Proceedings of the National Academy of Sciences **38**(3), 235–243 (1952). doi:10.1073/pnas.38.3.235
75. Butcher, J.C.: Numerical Methods for Ordinary Differential Equations, Second Edition, pp. 1–463. John Wiley and Sons, ??? (2008).

doi:10.1002/9780470753767
76. McMurdie, P.J., Holmes, S.: phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE **8**(4), 61217 (2013)
77. Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., Flint, H.J.: Acetate utilization and butyryl coenzyme A (CoA): Acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. Applied and Environmental Microbiology **68**(10), 5186–5190 (2002). doi:10.1128/AEM.68.10.5186-5190.2002
78. Van Der Meulen, R., Adriany, T., Verbrugghe, K., De Vuyst, L.: Kinetic analysis of bifidobacterial metabolism reveals a minor role for succinic acid in the regeneration of NAD+ through its growth-associated production. Applied and Environmental Microbiology **72**(8), 5204–5210 (2006). doi:10.1128/AEM.00146-06
79. Ofaim, S., Zarecki, R., Porob, S., Gat, D., Lahav, T., Xu, X., Kashi, Y., Aly, R., Jiandong, J., Eizenberg, H., Ronen, Z., Freilich, S.: Genome-Scale reconstruction of Paenarthrobacter aurescens TC1 metabolic model towards the study of atrazine bioremediation. bioRxiv, 536011 (2019). doi:10.1101/536011
80. Barrett, C.L., Herrgard, M.J., Palsson, B.O.: Decomposing complex reaction networks using random sampling, principal component analysis, and basis rotation. BMC Systems Biology **3**(1), 30 (2009). doi:10.1186/1752-0509-3-30
81. Breiman, L.: Random Forests. Mach.Learn. **45**(1), 5–32 (2001)
82. Medlock, G.L., Papin, J.A.: Guiding the Refinement of Biochemical Knowledgebases with Ensembles of Metabolic Networks and Machine Learning. Cell Systems (2020). doi:10.1016/j.cels.2019.11.006
83. Louppe, G., Wehenkel, L., Sutera, A., Geurts, P.: Understanding variable importances in forests of randomized trees. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26, pp. 431–439. Curran Associates, Inc., ??? (2013). http://papers.nips.cc/paper/4928-understanding-variable-importances-in-forests-of-randomized-trees.pdf
84. Wong, C., Harris, P.J., Ferguson, L.R.: Potential benefits of dietary fibre intervention in inflammatory bowel disease. International Journal of Molecular Sciences **17**(6) (2016)
85. Cherbuy, C., Bellet, D., Robert, V., Mayeur, C., Schwiertz, A., Langella, P.: Modulation of the caecal gut microbiota of mice by dietary supplement containing resistant starch: Impact is donor-dependent. Frontiers in Microbiology **10**(JUN) (2019). doi:10.3389/fmicb.2019.01234
86. Norsigian, C.J., Pusarla, N., McConn, J.L., Yurkovich, J.T., Dräger, A., Palsson, B.O., King, Z.: BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Research (2020). doi:10.1093/nar/gkz1054
87. Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N., Fleming, M.T., García-Jiménez, B., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Uk Kim, H., King, Z., Klamt, S., Klipp, E., Lakshmanan, M., Le Novère, N., Lee, D.-Y., Yup Lee, S., Lee, S., Lewis, N.E., Ma, H.: Memote: A community driven effort towards a standardized genome-scale metabolic model test suite. bioRxiv. doi:10.1101/350991
88. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Van Der Lei, J., Van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data **3** (2016). doi:10.1038/sdata.2016.18
89. Tummler, K., Klipp, E.: The discrepancy between data for and expectations on metabolic models: How to match experiments and computational efforts to arrive at quantitative predictions? Current Opinion in Systems Biology **8**, 1–6 (2018). doi:10.1016/j.coisb.2017.11.003
90. Zhang, Z., Shen, T., Rui, B., Zhou, W., Zhou, X., Shang, C., Xin, C., Liu, X., Li, G., Jiang, J., Li, C., Li, R., Han, M., You, S., Yu, G., Yi, Y., Wen, H., Liu, Z., Xie, X.: CeCaFDB: A curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by 13C-fluxomics. Nucleic Acids Research **43**(D1), 549–557 (2015). doi:10.1093/nar/gku1137
91. Zampieri, G., Vijayakumar, S., Yaneske, E., Angione, C.: Machine and deep learning meet genome-scale metabolic modeling. PLoS Computational Biology **15**(7) (2019). doi:10.1371/journal.pcbi.1007084
92. Cao, X., Hamilton, J.J., Venturelli, O.S.: Understanding and Engineering Distributed Biochemical Pathways in Microbial Communities. Biochemistry **58**(2), 94–107 (2019). doi:10.1021/acs.biochem.8b01006
93. Yamamoto, M., Pinto-Sanchez, M.I., Bercik, P., Britz-McKibbin, P.: Metabolomics reveals elevated urinary excretion of collagen degradation and epithelial cell turnover products in irritable bowel syndrome patients. Metabolomics **15**(6) (2019). doi:10.1007/s11306-019-1543-0
94. King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., Palsson, B.O.: Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLOS Computational Biology **11**(8), 1004321 (2015). doi:10.1371/journal.pcbi.1004321
95. Hove, H., Mortensen, P.B.: Influence of intestinal inflammation (IBD) and small and large bowel length on fecal short-chain fatty acids and lactate. Digestive Diseases and Sciences **40**(6), 1372–1380 (1995). doi:10.1007/BF02065554
96. Monk, J.M., Charusanti, P., Aziz, R.K., Lerman, J.A., Premyodhin, N., Orth, J.D., Feist, A.M., Palsson, B.: Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences of the United States of America **110**(50), 20338–20343 (2013). doi:10.1073/pnas.1307797110

Figures

Figure 1 MDPbiomeGEM general schema. MMODES uses GEM models (left) to generate microbial community synthetic, dynamic data (center, for example Fig. 2a), that are in turn consumed by MDPbiome (right), and used to generate a directed intervention recommendation.

Figure 2 MDPbiomeGEM modeling Crohn's disease recovery. a Example of a single run of MMODES, generating a microbial community and media component time-series with perturbations (Probiotic *F.prausnitzii* and/or Prebiotic Inulin added at the indicated time points). **b** MDPbiome's identification of three robustly distinguishable microbiome states - labelled "dysbiosis", "risky", and "healthy". **c** Sampling of microbiome state transitions over time from multiple runs of MMODES, with each row being a single time-series output, and the colors matching the colors of the state labels in panel B. **d** MDPbiome state-transition diagram, with recommended interventions to avoid, or recover from, Crohn's disease dysbiosis highlighted in red. 'Pro-high/low' means *F.prausnitzii* as a probiotic in high/low concentration.

Figure 3 Stability and generality of the MDPbiome recommendations to recover from Crohn's disease. a Stability assessment: rows represent the three microbiome states. Within each row, the color fragments represent each of the defined perturbations. The darker color fragment represents the optimal recommendation for each state (one per row). **b** Generality assessment: frequency of transitions when following (row F), or not following (row nF), the optimal recommendation. Better (green), equal (blue) and worse (red) state-transition are defined based on the utility function for sorting states - in this case, higher butyrate concentration.

Figure 4 Metabolites change in their concentration ratio after perturbation. Average of samples from all microbiome states after a particular perturbation was applied. Salmon indicates no change in concentration; yellow and light orange indicate a decrease; blue colors indicate an increase. 2.5 was established as maximum possible relative concentration ratio.

Figure 5 Metabolic map showing the most discriminant reactions. The most important reactions visualized on a map of *F.prausnitzii* metabolism. Flux values used in the analysis were taken from one of the five medoids (only *F.prausnitzii* growing). The relevant reactions are circled in green. Numbers labeling the reaction are fluxes (mmol/gDW/h). The legend on the right shows a longer list of relevant reactions, where some were omitted from the map to improve readability and redundancy. The metabolic map is taken under the MIT license from Escher [94].

Figure 6 Atrazine degradation results. a Atrazine degradation ratio (black dots) is computed as the percentage of atrazine that is degradaded between two sampling points. It is computed by sample (several samples per time series), and may be aggregated by state (mean of samples within that state) or by perturbation (mean of ratio of samples after a specific perturbation is applied). **b** Relative probability to move between the states. **c** Stability assessment (described in Fig.3a). Hb=*Halobacillus sp.*, Hm=*H. stevensii*, Hb-Hm=both.

Table 1 Comparison of MDPbiomeGEM with similar studies. Explanations of the columns are as follows: **Static/Dynamic**: The community modeling approach. BacArena and COMETS are dynamic approaches, but they are used as static ones in the case studies listed in this table. **GEM**: If GEM models were generated or used in the publication, with links to the models, if available. **# strains**: Number of strains in the microbial community. **Habitat**: Habitat used in the study. **Omics data required**: If abundances were used to define the microbial proportions in the community model, and if so, of what type. **Validation**: If the validation of the approach is *In-vivo* or *In-silico*. **Microbiome states**:The number of states that were examined in the study ("no" means the idea of a state was not addressed); N=2 usually means healthy vs dysbiosis. **Temporal analysis**: Whether the analysis has a temporal component. **Perturbations**: Whether the method allows perturbations to the microbial community. **Software available**: If the software is available, and if it is applicable to new cases versus being applicable only to the published scenario. **Study goal**: Brief description of the objective of the approach.

Table 2 Comparison between dynamic approaches to community metabolic modeling

Tables

Additional Files

Additional file 1 — Figure S1

(pdf) **Comparison of MDPbiomeGEM vs Bauer et al. [32]: Experimental data**. Adapted from [32]-Figure 3b. Experimental data taken from Table 6-column 2 in [95] and Bauer data from the original figure. With respect to the remaining metabolites in the original figure, lactate concentration is constant, while propionate and isobutyrate could not be computed because our microbial community does not include the Bacteroidia class strains that generate those metabolites.

Additional file 2 — Figure S2

(pdf) **Comparison of MDPbiomeGEM vs Bauer et al. [32]: Simulated data and treatment**. Replicated from [32]-Figure 2e and 5e. Limited to metabolites with log(concentration)+1 lower than 1mM. "Treatment" means the sample immediately subsequent to one identified as being in the state "dysbiosis" by MDPbiome, and where the perturbation applied was Inulin (i.e. the recommended treatment according to MDPbiomeGEM).

Additional file 3 — Figure S3

(pdf) **Metabolites change in their concentration ratio after perturbation.** Samples split by state: A) Dysbiosis, B) Healthy, C) Risky. Average of samples after a particular perturbation was applied. Salmon indicates no change in concentration; yellow and light orange indicate a decrease; blue colors indicate an increase. 2.5 was established as maximum possible relative concentration ratio.

Additional file 4 — Figure S4

(pdf) **Flux clustering results.** Five clusters were identified as the k best number of clusters following the robust clustering criteria of [66], which takes three clustering assessment metrics into account: Silhouette width (SI), Prediction Strength (PS) and Jaccard Index. k = 5 reaches the best SI and Jaccard scores, with a local maximum for the PS score.

Additional file 5 — Figure S5

(pdf) **25 most important reactions in the gut microbiome case study.** The reactions are sorted by Mean Decrease in Accuracy for each independent cluster and on average. The ID of the reaction corresponds to the ID in the models. FAEPRAA2165 corresponds to *i*Fap484 and BIFADO to *i*Bif452. The five clusters correspond to different growth rates of the two microbes: 4 - both growing; 3 - both in stationary phase; 1 - *i*Bif452 growing; 5 - *i*Fap484 growing; 2 - *i*Bif452 growing faster than *i*Fap484 due to a starch perturbation.

Additional file 6 — Figure S6
(pdf) **MMODES performance with a simulation of 10 homogeneous E.coli in a community.** We used a large GEM, including 2782 reactions and 1973 metabolites (*i*WFL_1372 [96]) MMODES took 3h 24 min running on a single Intel Xeon(R) CPU E5-1620 3.60GHz processor to compute a 12 hour simulation.

Additional file 7 — Figure S7
(pdf) **MDPbiome state-transition diagram for herbicide degradation in the soil microbiome.** Each node represents one of the microbiome states identified by MDPbiome in the soil microbiome case study. The labels on the arrows indicate the perturbation (H3PO4: phosphate, Hb:*Halobacillus sp.*, Hm:*H. stevensii*, Hb-Hm: both) and the probability to move to the indicated state.

Additional file 8 — Table S1
(pdf/odt) **Initial medium in the gut microbial community.** Ticks indicate the metabolite belongs to the corresponding model. Metabolite IDs are taken from the *i*Bif452 model, if present, otherwise the ID is taken from the *i*Fap484 model. A concentration of 4 mmol/L was assigned to non-limiting substrates.

Additional file 9 — Table S2
(pdf/odt) **Perturbations of the gut microbial community.** One perturbation from the table (randomly selected) was applied to the community every 8 hours. *Model ID. **Concentration of prebiotics is in mmol/L, concentration of probiotics is in g/L.

Additional file 10 — Table S3
(pdf/odt) **Initial medium in the soil microbial consortium.** Metabolites in black were in both the minimal medium and the root exudate-like medium. Metabolites in blue were only added to the latter. Ticks indicate the metabolite belongs to the corresponding model. A concentration of 10 mmol/L was assigned to non-limiting metabolites. Concentration was set at to 0.1 mmol/L to reflect phosphate-limiting conditions.

Additional file 11 — Table S4
(pdf/odt) **Perturbations of the soil microbial community.** One perturbation from the table (randomly selected) was applied to the community every hour. *Model ID. **Concentration of prebiotics is in mmol/L, concentration of probiotics is in g/L.

Figures

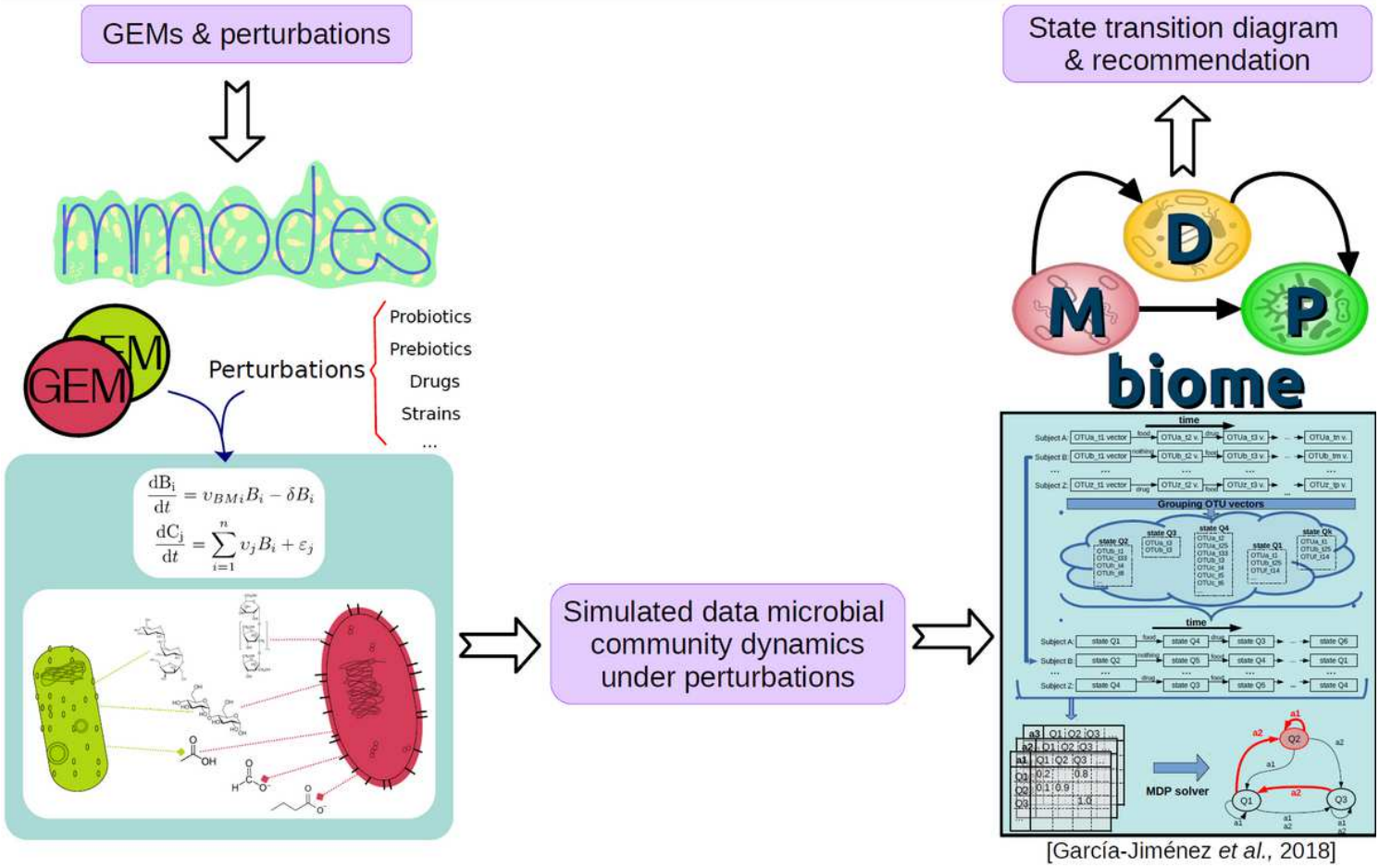

Figure 1

MDPbiomeGEM general schema. MMODES uses GEM models (left) to generate microbial community synthetic, dynamic data (center, for example Fig. 2a), that are in turn consumed by MDPbiome (right), and used to generate a directed intervention recommendation.

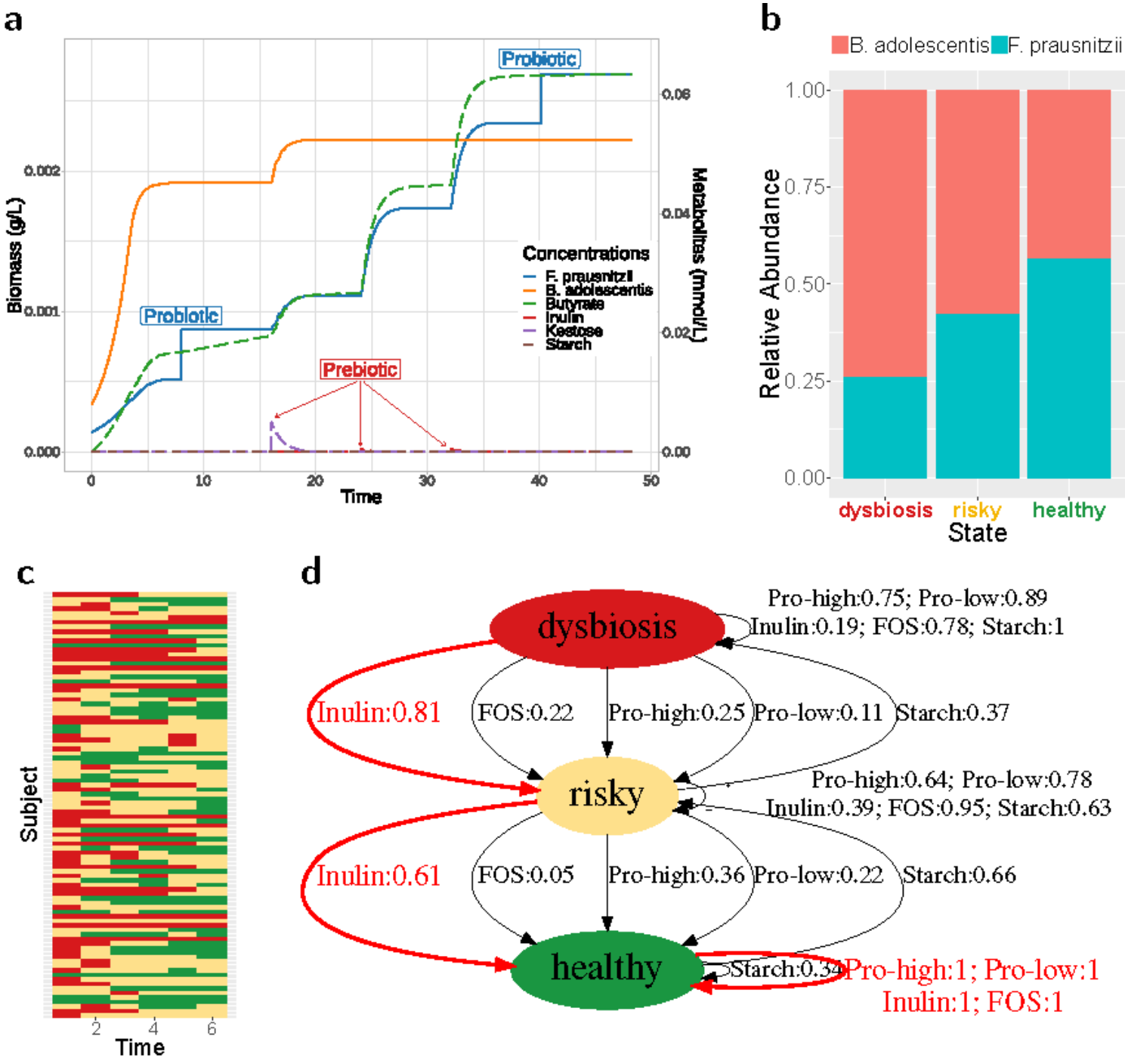

Figure 2

MDPbiomeGEM modeling Crohn's disease recovery. a Example of a single run of MMODES, generating a microbial community and media component time-series with perturbations (Probiotic F.prausnitzii and/or Prebiotic Inulin added at the indicated time points). b MDPbiome's identification of three robustly distinguishable microbiome states - labelled \dysbiosis", \risky", and \healthy". c Sampling of microbiome state transitions over time from multiple runs of MMODES, with each row being a single time-series output, and the colors matching the colors of the state labels in panel B. d MDPbiome state-transition diagram, with recommended interventions to avoid, or recover from, Crohn's disease dysbiosis highlighted in red. `Pro-high/low' means F.prausnitzii as a probiotic in high/low concentration.

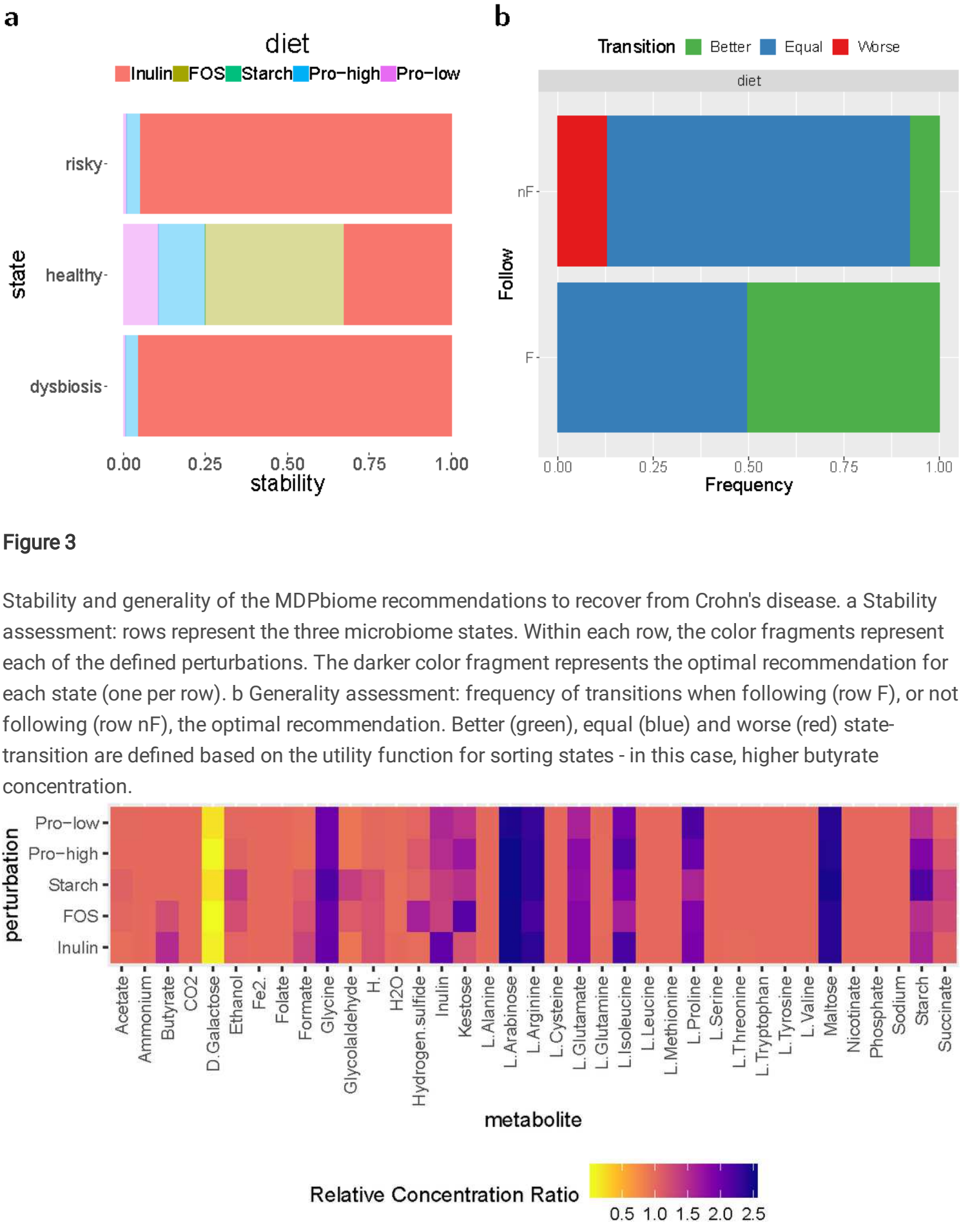

Figure 3

Stability and generality of the MDPbiome recommendations to recover from Crohn's disease. a Stability assessment: rows represent the three microbiome states. Within each row, the color fragments represent each of the defined perturbations. The darker color fragment represents the optimal recommendation for each state (one per row). b Generality assessment: frequency of transitions when following (row F), or not following (row nF), the optimal recommendation. Better (green), equal (blue) and worse (red) state-transition are defined based on the utility function for sorting states - in this case, higher butyrate concentration.

Figure 4

Metabolites change in their concentration ratio after perturbation. Average of samples from all microbiome states after a particular perturbation was applied. Salmon indicates no change in concentration; yellow and light orange indicate a decrease; blue colors indicate an increase. 2.5 was established as maximum possible relative concentration ratio

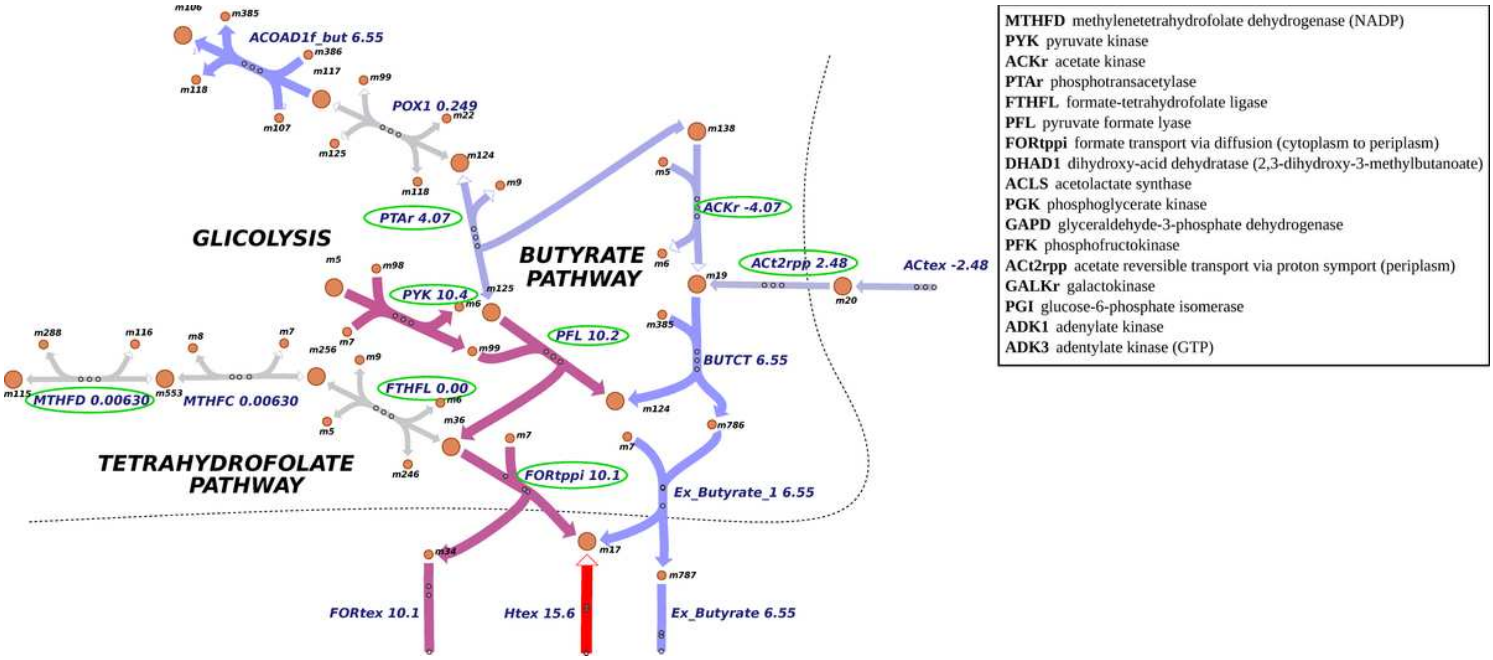

Figure 5

Metabolic map showing the most discriminant reactions. The most important reactions visualized on a map of F.prausnitzii metabolism. Flux values used in the analysis were taken from one of the five medoids (only F.prausnitzii growing). The relevant reactions are circled in green. Numbers labeling the reaction are uxes (mmol/gDW/h). The legend on the right shows a longer list of relevant reactions, where some were omitted from the map to improve readability and redundancy. The metabolic map is taken under the MIT license from Escher [94].

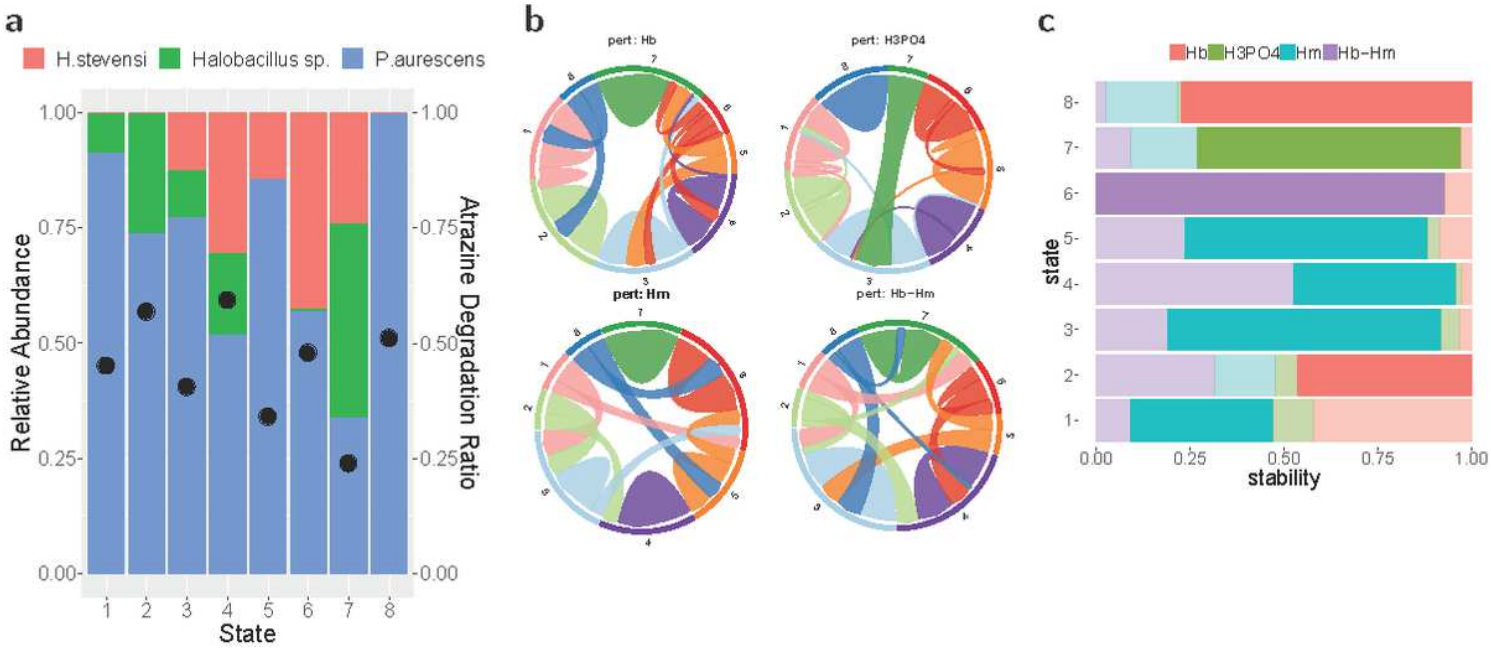

Figure 6

Atrazine degradation results. a Atrazine degradation ratio (black dots) is computed as the percentage of atrazine that is degradaded between two sampling points. It is computed by sample (several samples per time series), and may be aggregated by state (mean of samples within that state) or by perturbation (mean of ratio of samples after a specific perturbation is applied). b Relative probability to move between the states. c Stability assessment (described in Fig.3a). Hb=Halobacillus sp., Hm=H. stevensii, Hb-Hm=both.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile10TableS3.pdf
- AdditionalFile6FigureS6.pdf
- AdditionalFile5FigureS5.pdf
- AdditionalFile7FigureS7.pdf
- AdditionalFile8TableS1.pdf
- AdditionalFile9TableS2.pdf
- Table2.pdf
- AdditionalFile10TableS3.odt
- AdditionalFile11TableS4.odt
- AdditionalFile2FigureS2.pdf
- AdditionalFile9TableS2.odt

- AdditionalFile11TableS4.pdf
- Table1.pdf
- AdditionalFile3FigureS3.pdf
- AdditionalFile4FigureS4.pdf
- AdditionalFile1FigureS1.pdf
- AdditionalFile8TableS1.odt