A metabolic modeling-based framework for predicting trophic dependencies in native rhizobiomes of crop plants

Alon Avraham Ginatt*[1,2], Maria Berihu[1], Einam Castel[1], Shlomit Medina[1], Gon Carmi[1], Adi Faigenboim-Doron[3], Itai Sharon[4,5], Ofir Tal[6], Samir Droby[7], Tracey Somera[8], Mark Mazzola[9], Hanan Eizenberg[1], Shiri Freilich[1]

1. Department of Natural Resources, Newe Ya'ar Research Center, Agricultural Research Organization (Volcani institute), Ramat Yishay, Israel
2. Department of Plant Pathology and Microbiology, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel
3. Institute of Plant Sciences, Agricultural Research Organization (ARO), The Volcani Center, P. O Box 6, 5025001, Beit Dagan, Israel
4. Migal-Galilee Research Institute, P.O. Box 831, Kiryat Shmona 11016, Israel
5. Faculty of Sciences and Technology, Tel-Hai Academic College, Upper Galilee 1220800, Israel
6. Kinneret Limnological Laboratory, Israel Oceanographic and Limnological Research PO Box 447, Migdal 14950, Israel
7. Department of Postharvest Sciences, Agricultural Research Organization (ARO), the Volcani Center, 68 Ha Maccabim Road, Rishon LeZion 7505101, Israel
8. United States Department of Agriculture-Agricultural Research Service Tree Fruit Research Lab, 1104 N. Western Ave, Wenatchee, WA, 98801, USA
9. Department of Plant Pathology, Stellenbosch University, Private Bag X1, Matieland 7600, South Africa

Abstract

The exchange of metabolites (i.e., metabolic interactions) between bacteria in the rhizosphere determines various plant-associated functions. Systematically understanding the metabolic interactions in the rhizosphere, as well as in other types of microbial communities, would open the door to the optimization of specific pre-defined functions of interest, and therefore to the harnessing of the functionality of various types of microbiomes. However, mechanistic knowledge regarding the gathering and interpretation of these interactions is limited. Here, we present a framework utilizing genomics and constraint based modeling approaches, aiming to interpret the hierarchical trophic interactions in the soil environment. 243 genome-scale metabolic models of bacteria associated with a specific disease suppressive vs disease conducive apple rhizospheres were drafted based on genome resolved metagenomes, comprising an *in-silico* native microbial community. Iteratively simulating microbial community members' growth in a metabolomics-based apple root-like environment produced novel data on potential trophic successions, used to form a network of communal trophic dependencies. Network-based analyses have characterized interactions associated with beneficial vs non-beneficial microbiome functioning, pinpointing specific compounds and microbial species as potential disease supporting and suppressing agents. This framework provides a means for capturing trophic interactions and formulating a range of testable hypotheses regarding the metabolic capabilities of microbial communities within their natural environment. Essentially, it can be applied to different environments and biological landscapes, elucidating the conditions for the targeted manipulation of various microbiomes, and the execution of countless predefined functions.

Introduction

The rhizosphere serves as a hotspot for a diversity of interactions spanning from the secretion of organic compounds by plant roots to their uptake by the adjacent soil microbial community[1,2]. These interactions form a complex network of metabolic exchanges whose structure and function has a considerable impact on plant health[3]. Targeted secretion of exudates from plant roots' into the environment is fundamental to the recruitment of specific microbes and the assembly of a plant-selected community (i.e., the rhizobiome)[4–6]. Each plant has a unique profile of exudates guiding the formation of a specialized rhizobiome that is adapted to support its mineral absorption[7–9], secret plant growth supporting compounds[10–12], and provide protection against soil-borne microorganisms that are detrimental to its health[13–15]. A comprehensive understanding of the dynamics within the rhizosphere, considering both plant-microbe and microbe-microbe interactions, can guide the targeted assembly and maintenance of plant-beneficial soil microbial systems[16–19]. The development of such microbiome-based, plant-beneficial strategies presents ecologically sound alternatives to conventional, chemical-based solutions in supporting plant health and productivity. [20,21].

`Omics data in general and specifically metagenomics data analyses can potentially provide keys for unraveling the black box of plant-microbe and microbe-microbe interactions in complex ecosystems such as the soil[22–25]. Sequence-based analyses are, however, typically limited in terms of functional interpretation of community dynamics[26]. Constraint based modeling (CBM) is an approach that allows for the simulation of bacterial-metabolic activity in a given environment based on the constraints imposed by the annotated microbial genomes[22,27]. This approach has long been used for studying the physiology and growth of single cells, represented as Genome Scale Metabolic Models (GSMM), under varying conditions[28,29]. Applying CBM over a GSMM can be used to assess the uptake and secretion of metabolites in the environment under study[27]. Accordingly, when CBM is applied over multiple GSMMs, metabolite exchange profiles (secretion into and consumption from the environment) become interconnected. This sheds light on conditions supporting the growth of different bacterial groups within the community as well as their functional potential in the trophic network formed[16,30].

Advancement of sequencing technologies alongside the development of automatic pipelines for GSMM construction[31,32] has promoted an increased use of CBM for the modeling of communities with growing complexity[22–24,26,33,34]. The relevance of CBM to the study of microbiomes has been demonstrated in a variety of ecosystems and recent works have shown that CBM-based predictions can guide the development of strategies for microbiome management[34–37]. An accurate representation of microbial metabolic networks depends on the origin of the genomes analyzed. To date, most studies attempting to model the microbiome of specific ecosystems by GSMM represent native species using corresponding genome sequences from public depositories, a process which is usually referred to as 16S-based genome imputation[26,36,38]. Genome recovery, or genome-resolved metagenomes, often referred to as Metagenome Assembled Genomes (MAGs) allows one to obtain full genomes directly from metagenomes[39,40]. Constructing GSMMs based on MAGs derived from a specific biological sample[26] or directly from a native community[33], is fundamentally superior to using 16S-based genome imputation, for this approach enables a genuine view of the metabolic activities carried out *in situ*. Such an *in silico* representation of a native community (with respect to its environment) can be used to decrypt the myriad interconnected uptake and secretion exchange fluxes transpiring within the root-associated microbiome, spanning from root exudates to altered organic forms.

The current study describes the recovery of 395 unique MAGs from metagenomes constructed for the native rhizosphere community of apple rootstocks cultivated in orchard soil affected by apple replant disease (ARD)[41,42]. Soils were amended with *Brassicaceae* seed meal, or were not amended, supporting the development of either disease-suppressive or disease-conducive root microbiomes, respectively[41,42]. MAGs were recovered from metagenomics data collected from these apple rhizosphere microbiomes. GSMMs constructed for the MAGs provide an *in-silico* representation of highly abundant species in the native rhizosphere community. CBM simulations were then conducted in a rhizosphere-like environment, where microbial uptake-secretion fluxes were connected to form a directional trophic network. The aims of this study were twofold: first, to provide a general framework for delineating inter-species interactions occurring in the rhizosphere environment (Fig. 1) and second, to characterize the metabolic roles specific groups of bacteria fulfill in seed meal-amended (disease-suppressive) vs non-amended (disease-conducive) apple rhizobiome communities.

Results & Discussion

1. Assembly of a collection of MAGs representing a native microbial community from a soil agroecosystem

Metagenomic sequencing obtained from the rhizosphere of apple rootstocks grown in orchard soil with a documented history of replant disease resulted in a total of approximately 2 billion quality reads (after filtration) at a length of 150 bp, as described in Berihu et al[42]. Independent metagenomics assemblies of six different treatments yielded 1.4–2 million contigs longer than 2 kbp[42]. Here, assemblies were binned using MetaWRAP[40] into 296-433 high quality MAGs for each of the six treatments (Supp. Data 1; completion and contamination thresholds were set to 90/5, respectively[43]). De-replication (the process of combining highly similar genomes into a single representative genome) requiring 99% average nucleotide identity (ANI) of the overall 2233 MAGs yielded a collection of 395 unique MAGs (Fig. 2). Across samples, 30-36% of the raw reads were mapped to the MAG collection (Supp. Table 1), in comparison to 61–71% mapped to the non-binned contigs[42]. Using GTDB-Tk[44], taxonomic annotations were assigned at the phylum, order, genus and species level for 395, 394, 237 and three of the de-replicated MAGs, respectively, reflecting the genuine diversity of the rhizosphere community, which include many uncharacterized species[45]. Estimates of completion and contamination, total bin length and taxonomic affiliation for the MAG collections derived from each of the six assemblies, as well as the de-replicated MAGs, are provided in Supp. Data 1.

As in previous reports[1,46], Proteobacteria, Acidobacteria, Actinobacteria and Bacteroidetes were identified as the dominant phyla in the apple rhizosphere (Fig. 2A). Overall, the taxonomic distribution of the MAG collection corresponded with the profile reported for the same samples using alternative taxonomic classification approaches such as 16S rRNA amplicon sequencing and gene-based taxonomic annotations of the non-binned shotgun contigs (Fig. 2B). At the genus level, MAGs were classified into 143 genera in comparison to approximately 3000 genera that were identified for the same data based on gene-centric approaches[42] and approximately 1000 genera based on amplicon sequencing[41] of the same data.

The functional capabilities of the bacterial genomes in the apple rhizosphere were initially assessed based on KEGG functional annotations of their gene catalogue (Supp. Data 2, Supp. Fig. 1). The ten most frequent functional categories across the MAG collection were involved in primary metabolism, e.g., carbohydrate metabolism, and the biosynthesis of essential cellular building blocks such as amino acids and vitamins. Specialized functional categories included those associated with autotrophic nutrition such as carbon fixation and metabolism of nitrogen, sulfur and methane. Functional diversity was found to exist also when considering ubiquitous functions. For instance, though all bacteria are in need of the full set of amino acids, most genomes lack the full set of relevant biosynthesis pathways. The prevalence of biosynthetic pathways across MAGs (requiring at least a single relevant enzyme) ranges from 99.5% (e.g. glycine; missing in only two genomes) to 21% (e.g. tyrosine; detected in only 83 MAGs). Notably, the diversity of metabolic pathway completeness regarding amino acids and other essential cellular components suggests that the majority of bacterial soil species rely on an external supply of at least some of their obligatory nutritional demands.

2. Construction of a simulation system for exploring environment-dependent metabolic performances and growth of rhizosphere bacteria

Categorical classifications of discrete gene-entities, such as pathway completeness analyses, have several inherent limitations as an approach for the contextualization and functional interpretation of genomic information. First, categorical classifications may underestimate the completeness of robust pathways with multiple redundant routes resulting in low pathway completeness. Second, pathway completeness analyses do not take into consideration the directionality and continuity of a biosynthetic process whose full conductance requires the availability of a specific environmental resource together with the required successive series of genes/reactions. Although functional potential can be inferred based on the static set of genes and enzymes present, actual metabolic performances of bacterial species in soil are dynamic and reflect multiple factors, including the availability of different nutritional sources. Such sources can be environmental inputs like root exudates or downstream exchange metabolites secreted by cohabiting bacteria.

To better understand environment-dependent metabolic activity occurring in a native rhizosphere, a set of 395 GSMMs was constructed for the entire collection of the rhizosphere-bacterial community MAGs. All models were systematically subjected to validity and quality tests using MEMOTE[47], leaving a total of 243 GSMMs whose stoichiometric consistency was confirmed (Supp. Data 3). On average, GSMMs included 1924 reactions, from which 203 were exchange reactions (specific reactions carrying out extracellular import and secretion of metabolites) and 1312 metabolites (Supp. Data 4). Altogether, the GSMM set held 5152 unique metabolic reactions, 597 exchange reactions, and 2671 different compounds. The distribution of key model attributes (reactions, metabolites and exchanges) across phyla is shown in Supp. Fig. 2. Model features were scalable with those reported by Basile et al[26] and Heinken et al[34]. Additionally, comparison of GSMM scales indicated that the metabolic coverage (i.e., the number of reactions, which denote the potential of executing a metabolic function) of our data is within the same order of magnitude as described in recent large-scale automatic reconstructions[26,34].

Next, species-specific rich (optimal) and poor (suboptimal) media were defined for each model in order to broadly assess GSMM growth capacity. A rich environment was defined as a medium containing all metabolites for which the model encodes an exchange reaction. A poor environment was defined as the minimal set of compounds enabling growth (see Methods). Essentially, poor environments contained species-specific carbon, nitrogen and phosphorous sources together with other trace elements. In addition to the two automatic media, we aimed to design a realistic simulating environment to explore the impact of the root exudates on the native community. To this end, metabolomics data from a set of studies which characterized the root exudates of Geneva 935 (G935) or Malling 26 (M26) rootstock cultivars[48,49] - related to the G210 and M26 rootstocks whose rhizobiome was characterized here, were used to specify an array of apple root-derived compounds (Table 1). The list of secreted compounds was consistent with other reported profiles of plant root-exudates[1,50].

The growth of each of the 243 GSMMs was simulated in each of the three species-specific environments (rich, poor, poor + exudates; exudates were added to corresponding poor media to ensure the exudates have a feasible effect). As expected, growth performances were higher on the poor medium supplemented with exudates in comparison to growth on the poor medium alone but lower than the growth in rich medium (Supp. Fig. 3). Notably, GSMM growth patterns were not phylogenetically

conserved and were inconsistent between related taxa. Moreover, ranking of models' growth rate were inconsistent in the three different media (i.e., some models' growth rates were markedly affected by the simulated media whereas others did not; Supp. Fig. 3). This inconsistency indicates that the effect of exudates on community members is selective and differs between species (i.e., exudates increase the growth rates of some species more than others), as was previously reported[1,5,51]. The number of active exchange fluxes in each medium corresponds with the respective growth performances displaying noticably higher number of potentially active fluxes in the rich enviroenment (also when applying loopless FVA) (Supp. Fig. 4). Overall, Simulations confirmed the existence of a feasible solution space for all the 243 models as well as their capacity to predict growth in the respective environemnt (Supp. Data 5).

As a next step towards conducting simulations in a genuine natural-like environment, we aimed to define a single "rhizosphere environment" in which growth simulations for all models would take place. Unlike the species-specific root media (poor medium + exudates) which support growth of all models by artificially including multiple carbon sources that are derived from the automatic specifications of the poor medium, including such that are not provided by the root, this simulation environment was based on the root exudates (Table. 1) as the sole carbon sources. By avoiding the inclusion of non-exudate organic metabolites, the true-to-source rhizosphere environment was designed to reveal the hierarchical directionality of the trophic exchanges in soil, as rich media often mask various trophic interactions taking place in native communities[52]. Additionally, the rhizosphere environment also included an array of inorganic compounds used by the 243 GSMMs, which includes trace metals, ferric, phosphoric, and sulfuric compounds. Overall, the rhizosphere environment was composed of 60 inorganic compounds together with the 33 root exudates (Supp. Table 2). The rhizosphere environment supported the growth of only a subset of the GSMMs that were capable of using plant exudates (Supp. Fig. 3).

3. Simulating growth succession and hierarchical trophic exchanges in the rhizosphere community

To reflect the indirect effect of the root on the native community, (i.e., to capture the effect of root-supported bacteria on the growth of further community members), we

constructed the Microbial Community Succession Module (MCSM), a CBM-based algorithm aimed at predicting community-level trophic successions. Unlike tools designed for modelling microbial interactions[53,54], MCSM bypasses the need for defining a community objective function as the growth of each species is simulated individually. Trophic interactions are then inferred by the extent to which compounds secreted by bacteria could support the growth of other community members. The module iteratively grows the GSMMs in a defined environment, sums up their individual secretion profiles, and updates the initial simulation environment with those secreted compounds (Fig. 3A). Applying this algorithm to a microbiome in its native environment allows delineating the potential metabolic dependencies and interactions between bacterial species in a native community.

Application of MCSM over the “rhizosphere environment” (i.e., 1st iteration, Root exudates and "inorganic compounds", Supp. Table 2) supported the growth of 27 GSMMs (Supp. Fig. 3, Fig. 3B). Then, the initial environment was updated with 145 additional compounds predicted to be secreted by the growing community members. The 2nd iteration supported the growth of 33 additional species whose growth was supported by compounds predicted to be secreted by bacteria that grew in the 1st iteration. Following the 2nd iteration, 25 new secreted compounds were added to the rhizosphere environment. The 3rd iteration supported the growth of 11 additional GSMMs, with one additional compound secreted. After the 3rd iteration, the updated environment did not support the growth of any new species. Overall, iterative growth simulations resulted in the successive growth of 71 species (Fig. 3B; iterations 1-3).

To enlarge the array of growing species, we tested the effect of the addition of organic phosphorous sources. Organic phosphorous is typically a limiting factor in soil[55] and its utilization varies greatly between microbial species (i.e., different P sources were shown to have a selective effect on different microbial groups[56]). The initial rhizosphere environment contained only inorganic phosphorous. During 1st -3rd MCSM simulations, nine organic P compounds were secreted to the simulation medium, which was updated accordingly. At the beginning of the 4th iteration, 31 additional organic P compounds were identified by screening the species-specific poor medium and were added to the medium (Supp. Table 2; organic Phosphorous compounds). The additional organic P compounds supported the growth of nine additional GSMMs (Fig. 3B) and led to the secretion of 13 new compounds, which were added to the environment. The 5th iteration supported the growth of four additional

GSMMs and two new compounds were secreted. Final simulations in the cumulative rhizosphere environment were composed of all secreted compounds and led to the same secretion and growth profile as the previous iteration. Therefore, no further growth iterations were conducted.

Overall, the successive iterations connected 84 out of 243 native members of the apple rhizosphere GSMM community via trophic exchanges. The inability of the remaining bacteria to grow, despite being part of the native root microbiome, possibly reflects the selectiveness of the root environment, which fully supports the nutritional demands of only part of the soil species, whereas specific compounds that might be essential to other species are less abundant[45]. It is important to note that the specific exudate profile used here represent a snapshot of the root metabolome as root secretion-profiles are highly dynamic, reflecting both environmental and plant developmental conditions. A possible complementary explanation to the observed selective growth might be the partiality of our simulation platform, which examined only plant-bacteria and bacteria-bacteria interactions while ignoring other critical components of the rhizosphere system such as fungi, archaea, protists and mesofauna, as well as less abundant bacterial species, components all known to metabolically interact[57]. Finally, the MAG collection, while relatively substantial, represents only part of the microbial community. Accordingly, the iterative growth simulations represent a subset of the overall hierarchical-trophic exchanges in the root environment, necessarily reflecting the partiality of the dataset.

In terms of the phylogenetic distribution of the models, 27 bacterial species grew on the 1st iteration (in which root exudates served as the sole organic sources). These bacteria represented 14 of the 17 phyla included in the initial model collection (consisting of 243 GSMMs) and maintained a distribution frequency similar to the original community. As in the full GSMM data set (Community bar, Fig. 3C), most of the species which grew in the 1st iteration belonged to the phyla *Acidobacteriota*, *Proteobacteria,* and *Bacteroidota*. This result concurred with findings from the work of Zhalnina et al, which reported that bacteria assigned to these phyla are the primary beneficiaries of root exudates[1]. Species from three out of the 17 phyla that did not grow in the first iteration - *Elusimicrobiota, Chlamydiota*, and *Fibrobacterota,* did grow on the 2nd iteration (Fig. 3C). Members of these phyla are known for their specialized metabolic dependencies. Such is the case for example with members of the

Elusimicrobiota phylum, which include mostly uncultured species whose nutritional preferences are likely to be selective[58].

At the order level, bacteria classified as *Sphingomonadales* (class *Alphaproteobacteria*), a group known to include typical inhabitants of the root environment[59], grew in the initial Root environment. In comparison, other root-inhabiting groups including the orders *Rhizobiales* and *Burkholderiales*[59], did not grow in the first iteration. *Rhizobiales* and *Burkholderiales* did, however, grow in the second and third iterations, respectively, indicating that in the simulations, the growth of these groups was dependent on exchange metabolites secreted by other community members (Supp. Fig. 5).

Overall, 158 organic compounds were secreted throughout the MCSM simulation (from which 12 compounds overlapped with the original exudate medium). These compounds varied in their distribution and were mapped into 12 biochemical categories (Fig. 3D). Whereas plant secretions are a source of various organic compounds, microbial secretions provide a source of multiple vitamins and co-factors not secreted by the plant. Microbial-secreted compounds included siderophores (staphyloferrin, salmochelin, pyoverdine, and enterochelin), vitamins (pyridoxine, pantothenate, and thiamin), and coenzymes (coenzyme A, flavin adenine dinucleotide, and flavin mononucleotide) – all known to be exchange compounds in microbial communities[60,61]. In addition, microbial secretions included 11 amino acids (arginine, lysine, threonine, alanine, serine, phenylalanine, tyrosine, leucine, glutamate, isoleucine, and methionine), also known as a common exchange currency in microbial communities[62]. Some microbial-secreted compounds, such as phenols and alkaloids, were reported to be produced by plants as secondary metabolites[63,64]. Additional information regarding mean uptake and secretion degrees of compounds classified to biochemical groups is found in Supp. Fig. 6.

Conceptually, the rhizosphere microbiota can be classified into two trophic groups: primary exudate consumers, comprising microbial species that are direct beneficiaries from the root exudates, and secondary consumers, comprising microbial species whose growth may be provided directly via the uptake of metabolites secreted by other members of the soil microbial community. In the iterative MCSM simulations, compounds secreted by some of the primary consumers largely sustained the growth of secondary consumers, which were not able to grow otherwise. The full information on the secretion profiles and models' growths is provided in Supp. Data 6.

To validate the ability of MCSM to capture trophic dependencies and succession, we further tested whether it can track the well-documented example of cellulose degradation - a multi-step process conducted by several bacterial strains that go through the conversion of cellulose and its oligosaccharide derivatives into ethanol, acetate and glucose, which are all eventually oxidized to CO_2[65]. Here, the simulation followed the trophic interactions in an environment provided with cellulose oligosaccharides (4 and 6 glucose units) on the 1st iteration (Supp. Table 3). The formed trophic successions detected along iterations captured the reported multi-step process (Supp. Fig. 7).

4. Associating trophic exchanges with soil health

The MAG collection analyzed in this study was constructed from shotgun libraries associated with apple rootstocks cultivated in orchard soil with a documented history of apple replant disease (ARD) and healthy/recovered (seed meal-amended) soils, providing a model system for disease-conducive vs disease-suppressive rhizosphere communities[41]. Briefly, rhizobiome communities obtained from apple rootstocks grown in replant orchard soil leading to symptomatic growth (non-amended samples) were termed ‘sick’, whereas samples in which disease symptoms were ameliorated following an established soil amendment treatment[66], were termed ‘healthy‘. Both sick and healthy plants were characterized by distinct differences in the structure and function of their rhizosphere microbial communities in the respective soil samples[15,41,42,66]. In order to correlate microbial metabolic interactions with soil performance, GSMMs were classified into one of three functional categories based on differential abundance (DA) patterns of their respective MAGs: predominantly associated with “healthy” soil (H), predominantly associated with “sick” soil (S) and none-associated (NA) (Supp. Data 7).

The functionally classified GSMMs (H, S, NA) were consolidated into a community network of metabolic interactions by linking their potential uptake and secretion exchange profiles (as predicted along growth iterations in Fig. 3). The network was built as a directed bipartite graph, in which the 84 feasible GSMM nodes and the 203 metabolite nodes (27 root exudates, 146 microbial-secreted compounds, and 30 additional organic-P compounds) were connected by 9773 directed edges,

representing the metabolic exchanges of organic compounds in the native apple rhizosphere community (Fig. 4A). Further information regarding node degrees is found in Supp. Data 8.

The directionality of the network enabled its untangling into sub-network motifs stemming from a root exudate to exchange interactions, and ending with an unconsumed end-metabolite. Two types of sub-networks were detected (Fig. 4C): 3-component (PM; Plant-Microbe) plant exudate – microbe – microbial secreted metabolite; and 5-component (PMM; Plant-Microbe-Microbe) plant exudate – microbe – intermediate microbial secreted metabolite – microbe – microbial secreted metabolite. Overall, the network included 45,972 unique PM paths and 571,605 unique PMM paths. Participation of GSMMs in PM paths ranged from 272 to 896 occurrences (Supp. Fig. 8A). GSSM participation in PMM paths ranged from 398 to 50,628 in the first microbe position (primary exudate consumer) and 1,388 to 19,738 occurrences in the second microbe position (secondary consumer). Frequency of GSMMs in the first position in PMM sub-network motifs was negatively correlated with the frequency of presence in second positions, possibly indicating species-specific preferences for a specific position/trophic level in the defined environment (Pearson=-0.279; p-value=0.009, Supp. Fig. 8B, C, D).

In order to explore the trophic preferences of bacteria associated with the different rhizosphere soil systems, the frequency of healthy (H), sick (S) or non-associated (NA) GSMMs in the PM and PMM sub-networks was compared (Supp. Data 9). GSMMs classified as S initiated a significantly higher number of PMM sub-networks (located in the first position) than GSMMs classified as NA and H (Fig. 5A). H classified PMM paths (first position) initiated a significantly higher number of sub-networks with GSMMs classified as NA compared to S-classified GSMMs, but no more than H-classified GSMMs (second position). Other PMM types did not show a significant effect at the second position. The higher number of trophic interactions formed by the S-classified primary exudate consumers in the PMM sub-network motifs suggests that non-beneficial bacteria may have a broader spectrum in terms of their utilization potential of root-secreted carbon sources compared to plant-beneficial bacteria. This might shed light on the dynamics of ARD, in which S-classified bacteria become increasingly dominant following long-term utilization of apple-root exudates, resulting in diminished capacity of the rhizosphere microbiome to suppress soil-borne pathogens[66,67].

In order to predict exchanges with potential to support/suppress dysbiosis, the frequency of DA GSMM types (i.e., H or S) associated with metabolites (either consumed or secreted) in the PM paths was assessed (Supp. Data 10). Considering consumed metabolites (root exudates), three and six compounds were found to be significantly more prevalent in H and S-classified PM paths, respectively (Fig. 5B). Notably, the S-classified root exudates included compounds reported to support dysbiosis and ARD progression. For example, the S-classified compounds gallic acid and caffeic acid (3,4-dihidroxy-trans-cinnamate) are phenylpropanoids – phenylalanine intermediate phenolic compounds secreted from plant roots following exposure to replant pathogens[68]. Though secretion of these compounds is considered a defense response, it is hypothesized that high levels of phenolic compounds can have autotoxic effects, potentially exacerbating ARD. Additionally, it was shown that genes associated with the production of caffeic acid were upregulated in ARD-infected apple roots, relative to those grown in γ-irradiated ARD soil[69,70], and that root and soil extracts from replant-diseased trees inhibited apple seedling growth and resulted in increased seedling root production of caffeic acid[71].

As to the microbial-secreted compounds, a total of 79 unique compounds were found to be significantly overrepresented in either S (42 compounds) or H (41 compounds) classified PM paths (Fig. 5C). Several secreted compounds classified as healthy exchanges (H) were reported to be potentially associated with beneficial functions. For instance, the compounds L-Sorbose (EX_srb__L_e) and Phenylacetaladehyde (EX_pacald_e), both over-represented in H paths (Fig. 5C), have been shown to inhibit the growth of fungal pathogens associated with replant disease[72,73]. Phenylacetaladehyde has also been reported to have nematicidal qualities[74].

Combining both exudate uptake data and metabolite secretion data, the full H-classified PM path 4-Hydroxybenzoate; GSMM_091; catechol (Fig. 4C; the consumed exudate, the GSMM, and the secreted compound, respectively) provides an exemplary model for how the proposed framework can be used to guide the design of strategies which support specific, advantageous exchanges within the rhizobiome. The root exudate 4-Hydroxybenzoate is metabolized by GSMM_091 (class *Verrucomicrobiae*, order *Pedosphaerales*) to catechol. Catechol is a precursor of a number of catecholamines, a group of compounds which was recently shown to increase apple tolerance to ARD symptoms when added to orchard[42,75]. This analysis (PM; Fig 4C), leads to formulating the testable prediction that 4-Hydroxybenzoate can serve as a

selective enhancer of catecholamine synthesizing bacteria associated with reduced ARD symptoms, and therefore serve as a potential source for indigenously produced beneficial compounds.

Conclusions

In this study, we present a framework combining metagenomics analyses with CBM, which can be used to gain a deeper understanding of the functionality, dynamics and division of labor among rhizosphere bacteria, and link their environment-dependent metabolism to biological significance. This exploratory framework aims to illuminate the black box of interactions occurring in the rhizospheres of crop plants and is based on the work of Berihu et al. 2022, in which a gene-centric analysis of metagenomics data from apple rhizospheres was conducted[42].

We recovered high-quality sets of environment-specific MAGs, constructed the corresponding GSMMs, and simulated community-level metabolic interactions. By including authentic apple root exudates in the models, we were able to begin untangling the highly complex plant-bacterial and bacterial-bacterial interactions occurring in the rhizosphere environment. More specifically**,** we used the framework to investigate a microbial community via examining its hierarchical secretion-uptake exchanges along multiple iterations (Fig. 3). These analyses, which linked community-derived secretion profiles with the growth of other community members, demonstrated the successive, trophic-dependent nature of microbial communities. These interactions were elucidated via construction of a community-exchange network (Fig. 4). Possible connections between root exudates, differentially abundant bacteria, their secreted end-products, and soil health were explored using the data derived from this network. From these analyses, we were able to associate different metabolic functionalities with diseased or healthy systems, and formulate new hypotheses regarding the general function of differentially abundant bacteria in the community.

The framework we present is currently conceptual. Dealing with a highly complex system such as the rhizobiome inevitably comes with limitations. These limitations include the usage of automatic GSMM reconstruction, inherent caveats of CBM and the use of single-species GSMMs, the lack of transcriptomic and spatial-chemo-physical data, and the exclusion of competition over all its forms. Furthermore, a

portion of the metabolomics data used in this framework was taken from a different source (different rootstock genotype), possibly introducing further bias to the analyses. This potential factor is due to the inherent discrepancy between the conditions from which genomics and metabolomics data were collected[1,4]. Also not considered in this framework is the role of eukaryotes in the microbial-metabolic interplay. Moreover, the use of an automatic GSMM reconstruction tool (CarveMe[31]), though increasingly used for depicting phenotypic landscapes, is generally less accurate than manual curation of metabolic models[32]. This approach typically neglects specialized functions involving secondary metabolism[16] and introduces additional biases such as the overestimation of auxotrophies[76,77]. Nevertheless, manual curation is practically non-realistic for hundreds of MAGs, an expected outcome considering the volume of sequencing projects nowadays. As the primary motivation of this framework is the development of a tool capable of transforming high-throughput, low-cost genomic information into testable predictions, the use of automatic metabolic network reconstruction tools was favored, despite their inherent limitations, in pursuit of addressing the necessity of pipelines systematically analyzing metagenomics data.

For these reasons, among others, the framework presented here is not intended to be used as a stand-alone tool for determining microbial function. The framework presented is designed to be used as a platform to generate educated hypotheses regarding bacterial function in a specific environment in conjunction with actual carbon substrates available in the particular ecosystem under study. The hypotheses generated provide a starting point for experimental testing required to gain actual, targeted and feasible applicable insights[38,42]. While recognizing its limitations, this framework is in fact highly versatile and can be used for the characterization of a variety of microbial communities and environments. Given a set of MAGs derived from a specific environment and environmental metabolomics data, this computational framework provides a generic simulation platform for a wide and diverse range of future applications.

In the current study, the root environment was represented by a single pool of resources (metabolites). As genuine root environments are highly dynamic and responsive to stimuli, a single environment can represent at best a temporary snapshot of the conditions. Conductance of simulations with several sets of resource pools (e.g., representing temporal variations in exudation profile) can add insights on their effect on trophic interactions and community dynamics. In parallel, confirming predictions

made in various environments will support an iterative process that will strengthen the predictive power of the framework and improve its accuracy as a tool for generating testable hypotheses. Similarly, complementing the genomics-based approaches done here with additional layers of 'omics information (mainly transcriptomics & metabolomics) can further constrain the solution space, deflate the number of potential metabolic routes and yield more accurate predictions of GSMMs' performances[33].

To summarize, we have constructed a framework enabling the analysis of interactions between microbes and between microbes and their hosts in their natural environment. Where recent studies begin to apply GSMM reconstruction and analysis starting from MAGs[33,78], this work applies the MAGs to GSMMs approach to conduct large-scale CBM analysis over high-quality MAGs derived from a native rhizosphere and explore the complex network of interactions in light of the functioning of the respective agro-ecosystem.

The application of this framework to the apple rhizobiome yielded a wealth of preliminary knowledge about the metabolic interactions occurring within it, including novel information on putative functions performed by bacteria in healthy vs. replant-diseased soil systems, and potential metabolic routes to control these functions. Overall, this framework aims to advance efforts seeking to unravel the intricate world of microbial interactions in complex environments including the plant rhizosphere. The framework is provided as a three stage-detailed pipeline in https://github.com/FreilichLab/Trophic_interactions_predicting_framework.

Methods

Recovery of Metagenomics Assembled Genomes MAGs from metagenomics data constructed for apple rhizosphere microbiomes

High coverage shotgun metagenomics sequence data were obtained from microbial DNA extracted from the rhizosphere of apple rootstocks cultivated in soil from a replant diseased orchard[42]. The experimental design included sampling of six different soil/apple rootstock treatments with five biological replicates each, as described in Somera et al. 2021[41]. Two different apple rootstocks (M26, ARD susceptible; G210, ARD tolerant) were grown in three different treatments: 1) orchard soil amended with *Brassica napus* seed meal, 2) orchard soil amended with *Brassica*

juncea/*Sinapis alba* seed meal (BjSa) and 3) no-treatment control soil (NTC) (see Supp. Table 1). Microbial DNA was extracted from rhizosphere soil and metagenomics data were assembled as described in[42]. In each assembly contigs were binned to recover MAGs using MetaWRAP pipeline (v1.3.1), which utilizes several independent binners[40]. The MAGs recovered by the different binners were collectively processed with the Bin_refinement module of metaWRAP, producing an output of a refined bin collection. A count table was constructed by mapping raw reads data of each assembly (e.g., BjSa; G210) to the bins, using BWA-MEM mapping software (version 0.7.17) with default parameters. Differential abundance (DA) of the reads associated with the respective bins in each assembly across the respective replicates was determined using the edgeR function implemented in R, requiring FDR adapted p value< 0.05. Read mapping information is shown in Supp. Table 1. Based on DA, MAGs were classified either as associated with healthy soil (H; BjSa differentially abundant), sick soil (S; NTC differentially abundant), or not associated with either soil type (NA; not differentially abundant at any treatment site).

Gene calling and annotation was performed with the Annotate_bins module of MetaWRAP. Pathway completeness was determined with KEGG Decoder v 1.0.8.2[79] based on the KO annotations extracted from Annotate_bins assignments. The quality of the genomes was determined with CheckM[80]. For phylogenomic analyses and taxonomic classification of each bacterial and archaeal genome, we searched for and aligned 120 bacterial marker genes of the MAGs using the identity and align commands of GTDB-Tk v1.5.0[44]. MAGs were dereplicated using dRep v2.3.2[81] using the default settings and MAGs from the six assemblies were clustered into a single non-redundant set. Phylogenomic trees were rooted by randomly selecting a genome from the sister lineage to the genus as determined from the topology of the bacterial and archaeal GTDB R06-RS202 reference trees. Closely related GTDB taxa identified with the "classify_wf" workflow were filtered using the taxa-filter option during the alignment step. Overall, a set of 395 high-quality genomes (≥90% completeness, <5% contamination) was used for downstream analyses.

Genome Scale Metabolic Model (GSMM) reconstruction, analysis and characterization of the MAG collection

GSMMs were constructed for each of the 395 MAGs using CarveMe v 1.5.1 [31], a python-based tool for GSMM reconstruction. Installation and usage of CarveMe was done as suggested in the original CarveMe webpage (https://carveme.readthedocs.io/en/latest/). The solver used for GSMM reconstruction is Cplex (v. 12.8.0.0). All GSMMs were drafted without gap filling as it might mask metabolic co-dependencies[52]. Stoichiometric consistency of all GSMMs was systematically assessed via the standardized MEMOTE test suite, a tool for GSMM quality and completion assessment[47]. GSMMs not stoichiometrically balanced were filtered out, as they might produce infeasible simulation results.

Analyses and simulations of GSMMs, as well as retrieval of model attributes (reactions, metabolites and exchanges, etc.), were conducted via the vast array of methods found in COBRApy[82], a python coding language package for analyzing constraint-based reconstructions. For each GSMM, initial growth simulations took place in three different model-specific environments: rich medium, poor medium and poor medium + exudates. The rich medium was composed of all the exchange reactions (i.e., exchange reaction for specific compounds) a model holds, gathered by the "exchanges" attribute found in each model. The poor medium was composed of the minimal set of compounds required for a specific GSMM/species to grow at a fixed rate. This set of compounds was identified using the minimal_medium module from COBRApy (minimize components=True, growth rate of 0.1 biomass increase $hour^{-1}$). Poor medium + exudates was defined as the poor medium with the addition of an array of apple root exudates. These compounds were retrieved from two metabolomics studies characterizing the exudates of apple rootstocks grown in Lane Mountain Sand (Valley, WA)[48,49]. Exudate compounds were aligned with the BIGG database[83] in order to format them for use in COBRApy. For each media type, GSMM growth rates were calculated by solving each model using the summary method in COBRApy, which utilizes Flux Balance Analysis (FBA) for maximizing biomass.

Construction of a common root environment medium and application of the Microbial Community Succession Module (MCSM)

In order to simulate the dynamics of the rhizosphere community in the root environment, a fourth growth medium representing a natural-like environment was defined and termed the "rhizosphere environment". The rhizosphere environment was

composed of two arrays of compounds: 1) the exudates (as described above) and 2) inorganic compounds essential for sustaining bacterial growth. This array was determined according to the minimal set of compounds identified for each GSMM (also described above). Rhizosphere environment components are provided in Supp. Table 2. Both sets of compounds were then consolidated into one array in which further simulations were conducted.

The MCSM (which is the first module out of three comprising this computational framework) simulates the growth of a microbial community by iteratively growing the GSMMs in the community and adding compounds "secreted" by the growing species to the simulation-environment (i.e. the medium), thus enriching the medium/environment and supporting further growth. Unlike Flux Balance Analysis (FBA), which is used for gathering an arbitrary solution regarding non-optimized fluxes, the MCSM uses Flux Variability Analysis (FVA) to determine exchange fluxes[84]. FVA gathers the full range of exchange fluxes (including both secretion and uptake) that satisfy the objective function of a GSMM (i.e., biomass increase). The FVA fraction of optimum was set to 0.9 (sustaining the objective function at 90% optimality, allowing a less restricted secretion profile). Secretion compounds added to the updated medium in each iteration were set to be given in optimal fluxes in next iteration, to ensure a metabolic effect based on the presence of specific metabolites in the environment, rather than their quantity. Flux boundaries of updated medium components were set to 1000 mmol/ gDW hour (for a specific exchange compound; gDW gram Dry Weight).

MCSM was initially simulated in the rhizosphere environment medium. After each growth iteration, GSMM growth-values and the compounds secreted by the species growing were collected, where the latter were added to the medium for the next iteration as described above. Growth iterations continued until no new compounds were secreted and no additional GSMMs had grown. After the 3rd iteration, a set of organic phosphorous compounds (containing both carbon and phosphorous) was added to the environment. These compounds were gathered from the pool of model-specific minimal compounds selected for use in the poor medium. Information regarding the chemical formula of these compounds was gathered with the formula attribute of each compound object in COBRApy. Along MCSM iterations, secreted metabolites were classified into biochemical categories based on BRITE annotations[85] or, in the absence of classification, manually. MCSM was further applied to inspect the framework's

ability to tracing cellulose degradation using cellulose medium (Supp. Table 3). The tutorial for the MCSM stage of the framework workflow is found at the GitHub page, along with the GSMMs and media files. Instructions for conducting the analysis are in the README.md file https://github.com/FreilichLab/Trophic_interactions_predicting_framework.

Construction of the exchange network and its untangling for screening sub-network motifs

For each GSMM, a directed bipartite network representing all potential metabolic exchanges occurring within the rhizosphere community was constructed based on uptake and secretion data derived from MCSM iterations. Edges in the network were connected between GSMM nodes and metabolite nodes, with edge directionality indicating either secretion or uptake of a metabolite by a specific GSMM. The network was constructed with the networkx package, a python language package for the exploration and analysis of networks and network algorithms. Network-specific topography was obtained using the Kamada-kawai layout[86]. Information regarding the degree and connectivity of the different node types was acquired from the graph object (G). Code for the network construction module in the framework is found at the project's GitHub page under the name NETWORK.py.

Untangling the exchange network into individual paths (i.e. sub-network motifs) was done using the all_shortest_paths function in networkx[86], applied over the exchanges network. Briefly, the algorithm screens for all possible shortest paths within the network, specifically screening for paths starting with an exudate node and ending secreted metabolite nodes (secreted by bacterial species but not consumed). This algorithm yielded two types of paths: 1) plant-microbe paths (PM) in which node positions one, two, and three represented exudate, microbe and secreted metabolite, respectively, and 2) plant-microbe-microbe paths (PMM). PMM paths (length of five nodes) were constructed based on PM paths (length of three nodes), in which positions four and five displayed unique (i.e. not found in PM paths) microbe and metabolite nodes, respectively. Code for the network untangling module in the framework is found at the project's GitHub page under the name PATHS.py.

Associating PM and PMM sub-network motifs features with soil health

Sub-network motifs (PM and PMM) were functionally classified as associated with healthy soil (H), sick soil (S), or not-associated with either soil type (NA) based on differential abundance of the corresponding MAGs (PM) or MAGs combination (PMM) in the sub-network. Next, the GSMMs in both PM and PMM sub-networks were characterized according functional classification. For PMMs, the distribution of counts of classified sub-networks, at the different positions, was compared using the ANOVA test, followed by a Tukey test to significantly distinguish the groups. GSMM classifications were further projected on uptake and secreted metabolites in the pathway motifs. For simplification, the analysis focused only on PM paths because PMM paths incorporate PM paths, and exchanges within a PMM path do not directly reflect the effect of an exudate on the secreted end-product (but over the intermediate compound). On that account, start/end metabolites in PM paths were associated with H/S/NA paths based on one-sided hypergeometric test, comparing the frequency of each compound in a functionally characterized path type (either H or S) vs. its frequency in NA classified paths and the reciprocal data set.

Data availability

A reproducible workflow for central computational analyses was deposited in https://github.com/FreilichLab/Trophic_interactions_predicting_framework, including all GSMMs. Raw metagenome sequences used in this study have been deposited in NCBI under BioProject accession number PRJNA779554 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA779554). MAGs FASTA sequences generated and used in this work are located in https://github.com/FreilichLab/Ginatt_MAGs.

Acknowledgements

We would like to express our gratitude to Asaf Sadeh and Roni Gafni for their advice as part of the Neve Yaar Stats and R support forum, and to Ariana Basile for her advice

regarding the conductance of quality control tests over the genome scale metabolic models. We would also like to thank Maya Bar Yehuda for graphical design.

This work was funded by the United States-Israel Binational Agricultural Research and Development Fund (BARD) [grant number US-5390-21].

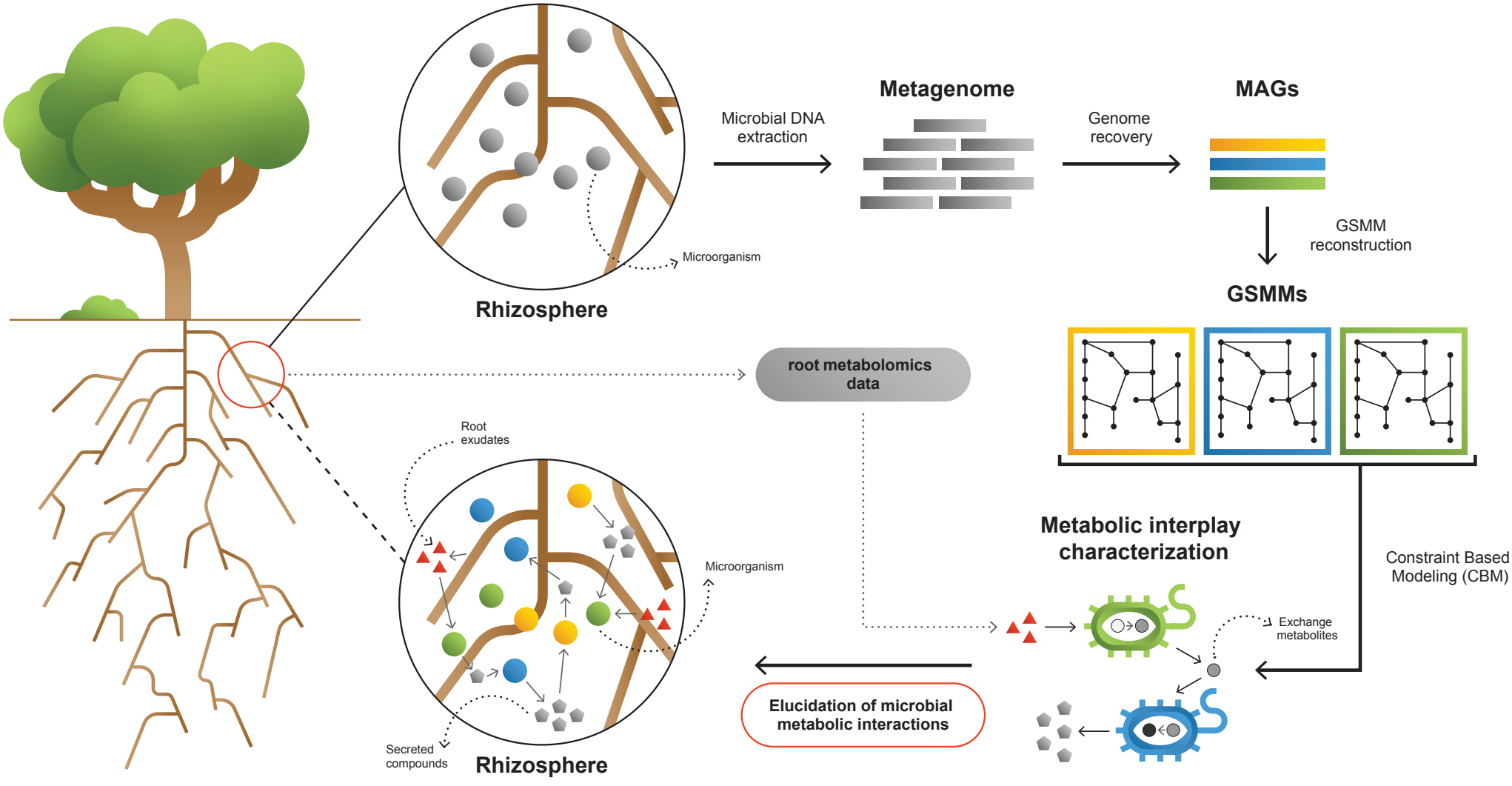

Microbial DNA extraction
Metagenome
Genome recovery
MAGs
GSMM reconstruction
GSMMs
Microorganism
Rhizosphere
root metabolomics data
Root exudates
Microorganism
Secreted compounds
Rhizosphere
Metabolic interplay characterization
Constraint Based Modeling (CBM)
Exchange metabolites
Elucidation of microbial metabolic interactions

Fig. 1 | A framework for interpreting and characterizing the network of metabolic interactions of root associated bacteria. Going from obscure rhizosphere (upper circle) to the elucidated rhizosphere (lower circle): Microbial genomic DNA extracted from rhizosphere soil serves the construction of a rhizosphere-derived metagenome and the recovery of rhizosphere community Metagenome Assembled Genomes (MAGs). For each MAG, a Genome Scale Metabolic Model (GSMM) is built, representing a specific species in the microbial community. Apple-root exudate profiles based on metabolomics data are used for the construction of a simulation environment. Via Constraint Based Modelling (CBM), applied on GSMMs, interactions are characterized in the simulated root environment, yielding predictive information regarding potential trophic exchanges within the native rhizosphere microbial community.

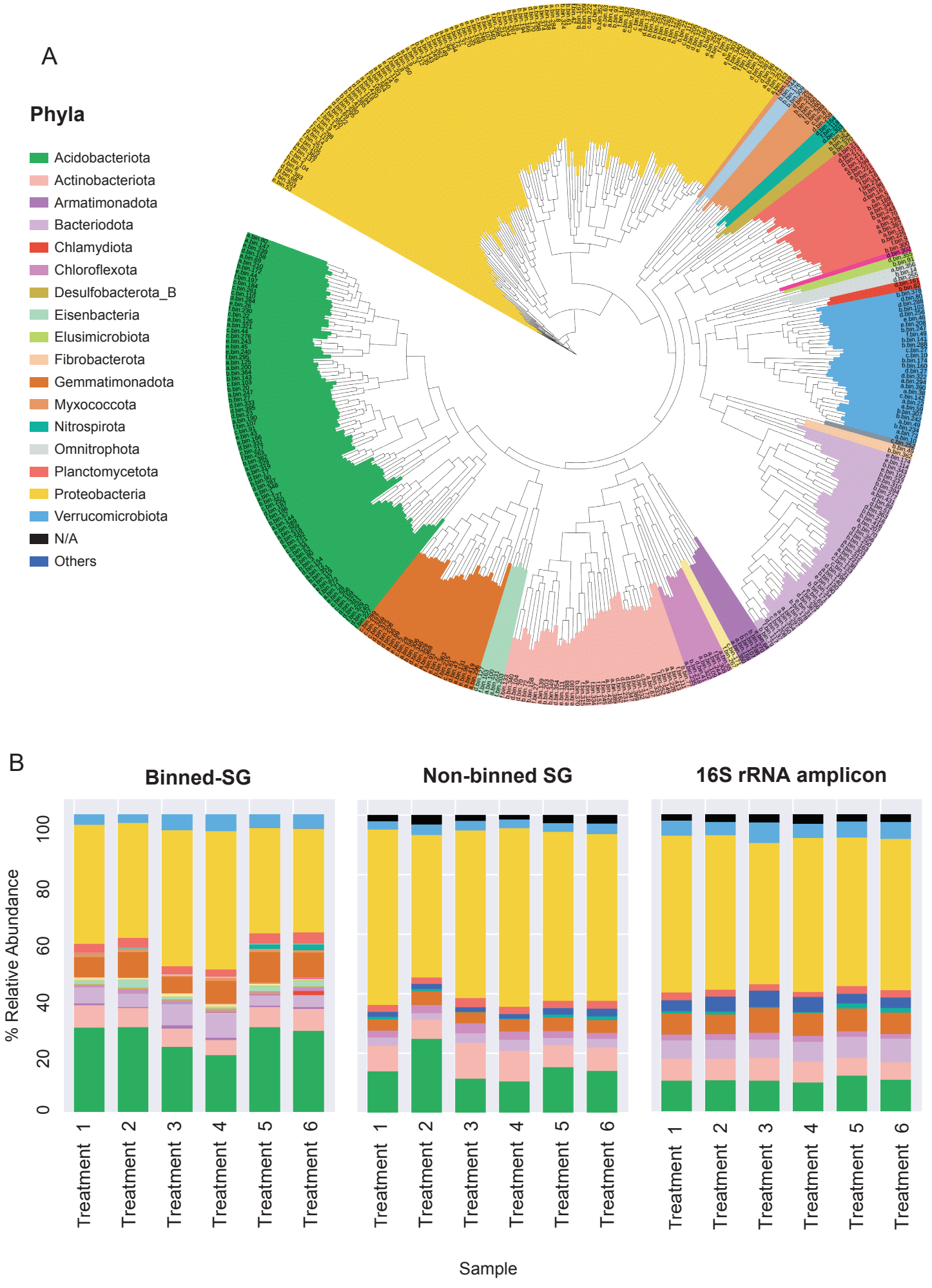

A
Phyla
Acidobacteriota
Actinobacteriota
Armatimonadota
Bacteriodota
Chlamydiota
Chloroflexota
Desulfobacterota_B
Eisenbacteria
Elusimicrobiota
Fibrobacterota
Gemmatimonadota
Myxococcota
Nitrospirota
Omnitrophota
Planctomycetota
Proteobacteria
Verrucomicrobiota
N/A
Others
B
Binned-SG
Non-binned SG
16S rRNA amplicon
% Relative Abundance
0
20
40
60
80
100
Treatment 1
Treatment 2
Treatment 3
Treatment 4
Treatment 5
Treatment 6
Sample

Fig. 2 | Phylogenomic surveys of the genomic collection assembled from sequence data derived from apple rhizosphere samples. A. A phylogenomic tree constructed from 395 de-replicated MAGs extracted from the metagenomes. The tree is based on concatenated marker proteins according to GTDB-Tk[44]. **B.** Relative abundance of bacterial taxa at the phylum level as inferred from different phylogenomic classification approaches applied for the same samples: Binned-SG (MAGs derived from shotgun metagenomics); Non-binned SG (contigs derived from shotgun metagenomics); 16S rRNA amplicon derived from the same samples. Taxonomic classifications of rRNA amplicon data and non-binned-SG sequences are taken from Somera et al.[41] and Berihu et al.[42], respectively. Shotgun metagenomics data from Berihu et. al was used here for construction of MAGs catalogue. "Treatments 1 – 6" relate to six specific combinations of rootstock and soil amendment as detailed in Supp. Table 1. Relative abundance was calculated as the average value in five replicates conducted for each treatment.

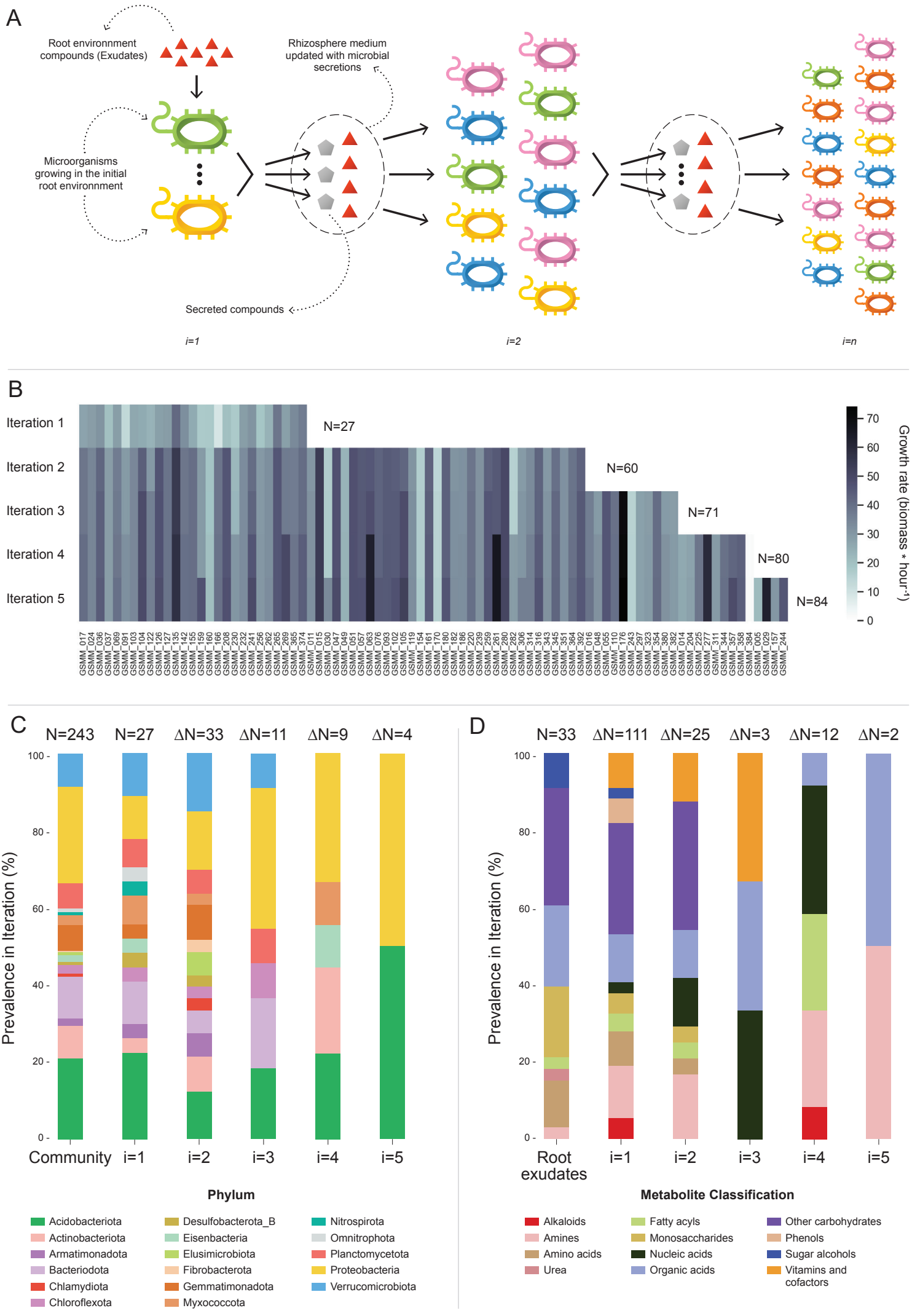

A
Root environnment compounds (Exudates)
Rhizosphere medium updated with microbial secretions
Microorganisms growing in the initial root environnment
Secreted compounds
i=1
i=2
i=n
B
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
N=27
N=60
N=71
N=80
N=84
Growth rate (biomass * hour⁻¹)
C
N=243
N=27
ΔN=33
ΔN=11
ΔN=9
ΔN=4
Prevalence in Iteration (%)
Community
i=1
i=2
i=3
i=4
i=5
Phylum
Acidobacteriota
Actinobacteriota
Armatimonadota
Bacteriodota
Chlamydiota
Chloroflexota
Desulfobacterota_B
Eisenbacteria
Elusimicrobiota
Fibrobacterota
Gemmatimonadota
Myxococcota
Nitrospirota
Omnitrophota
Planctomycetota
Proteobacteria
Verrucomicrobiota
D
N=33
ΔN=111
ΔN=25
ΔN=3
ΔN=12
ΔN=2
Prevalence in Iteration (%)
Root exudates
i=1
i=2
i=3
i=4
i=5
Metabolite Classification
Alkaloids
Amines
Amino acids
Urea
Fatty acyls
Monosaccharides
Nucleic acids
Organic acids
Other carbohydrates
Phenols
Sugar alcohols
Vitamins and cofactors

Fig. 3 | Microbial Community Succession Module (MCSM), and characterization of trophic dynamics in the community along iterations. A. An illustration of the iterative microbial community growth module, representing the growth of community members along iterations starting in the "Rhizosphere environment" and updated with microbial secretion outputs. **B.** Growth rate characterization of GSMMs in the community along iterations. Each column in the heat-map represents a different GSMM (only models which have grown in the rhizosphere environment are presented); growth rate is indicated by the color bar. Iterations are represented by rows. Blank spaces indicate models not growing at that specific iteration. N is the total number of species that grew after each iteration. **C.** Distribution of GSMMs growing along iterations at the phylum level. **D.** Distribution of organic compounds secreted along iterations, classified into biochemical groups. Root exudates bar (far left) represents the classification of organic compounds in the initial "Rhizosphere environment". Numbers on top of bars in both C, and D (designated by N) denotes the number of new entries in a specific iteration, with respect to the previous iteration.

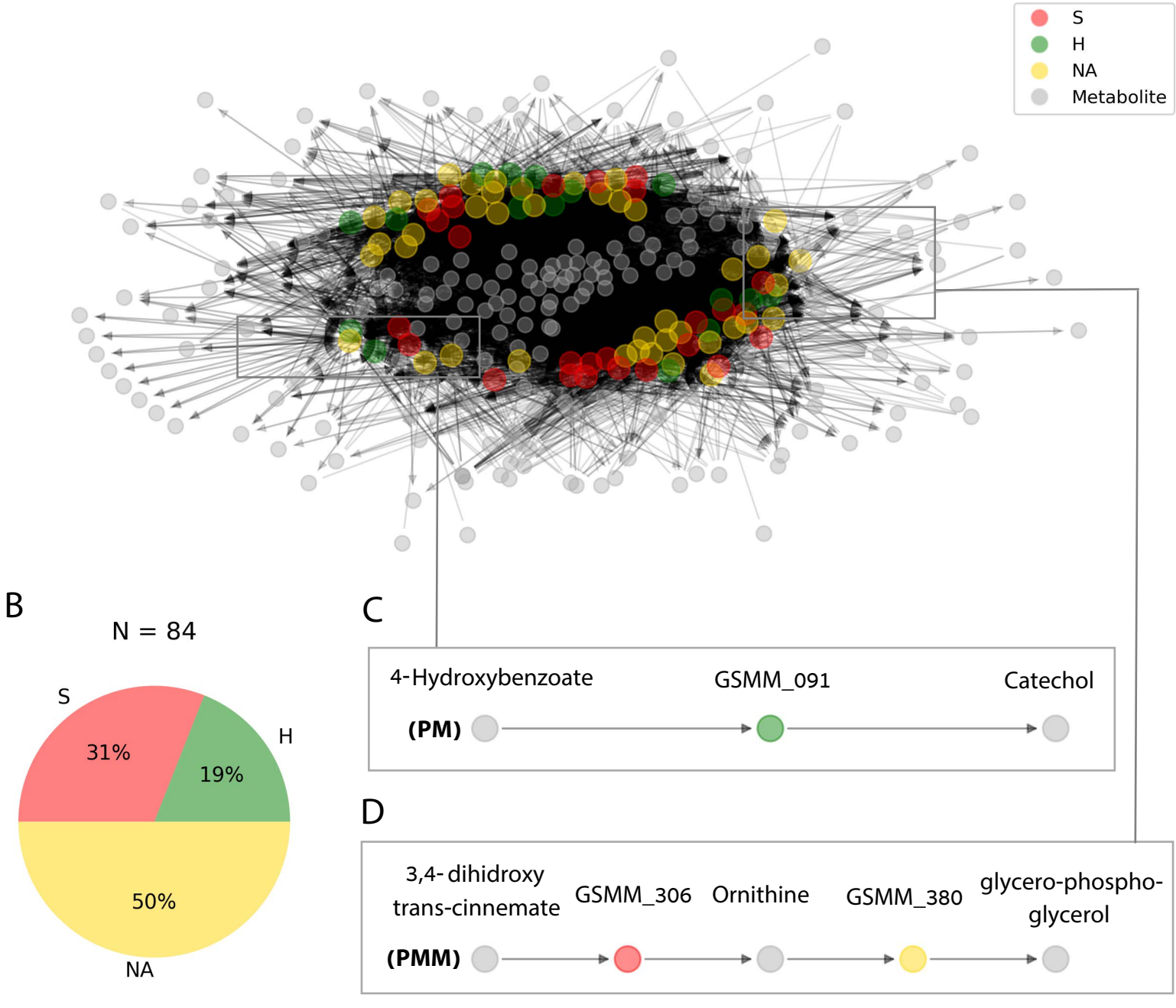

S
H
NA
Metabolite
B
N = 84
S
31%
H
19%
50%
NA
C
4-Hydroxybenzoate
GSMM_091
Catechol
(PM)
D
3,4- dihidroxy trans-cinnemate
GSMM_306
Ornithine
GSMM_380
glycero-phospho-glycerol
(PMM)

Fig. 4 | Trophic interactions based on community exchange fluxes predicted along iterations in the simulated rhizosphere environment. A. Network representation of potential metabolite exchanges between rhizosphere community members. Edges in the network are directional; arrowhead from a grey node (metabolite) pointing towards a colorful node (GSMM) indicates uptake; arrowhead from a colorful node (GSMM) pointing towards a grey node (metabolite) indicates secretion. Node colors correspond to differential abundance classification of GSMMs in the different plots; H, S, NA are Healthy, Sick, Not-Associated, respectively. Metabolites found in the center of the network are of a higher connectivity degree (i.e., are involved in more exchanges). Only organic compounds are included in the network. Grey rectangles illustrate a zoom-in to specific 5 and 3 partite sub-networks (C, D). **B.** Pie chart distribution of GSMMs classified according to differential abundance of reads mapped to the respective MAGs; N is the total number of GSMMs in the network. **C, D.** Examples of specific sub-network motifs derived from the community network; PM (a three component sub-network motif, upper); PMM (a five component sub-network motif, lower), respectively. The full list of PM and PMM sub-network motifs is found in Supplementary data 9.

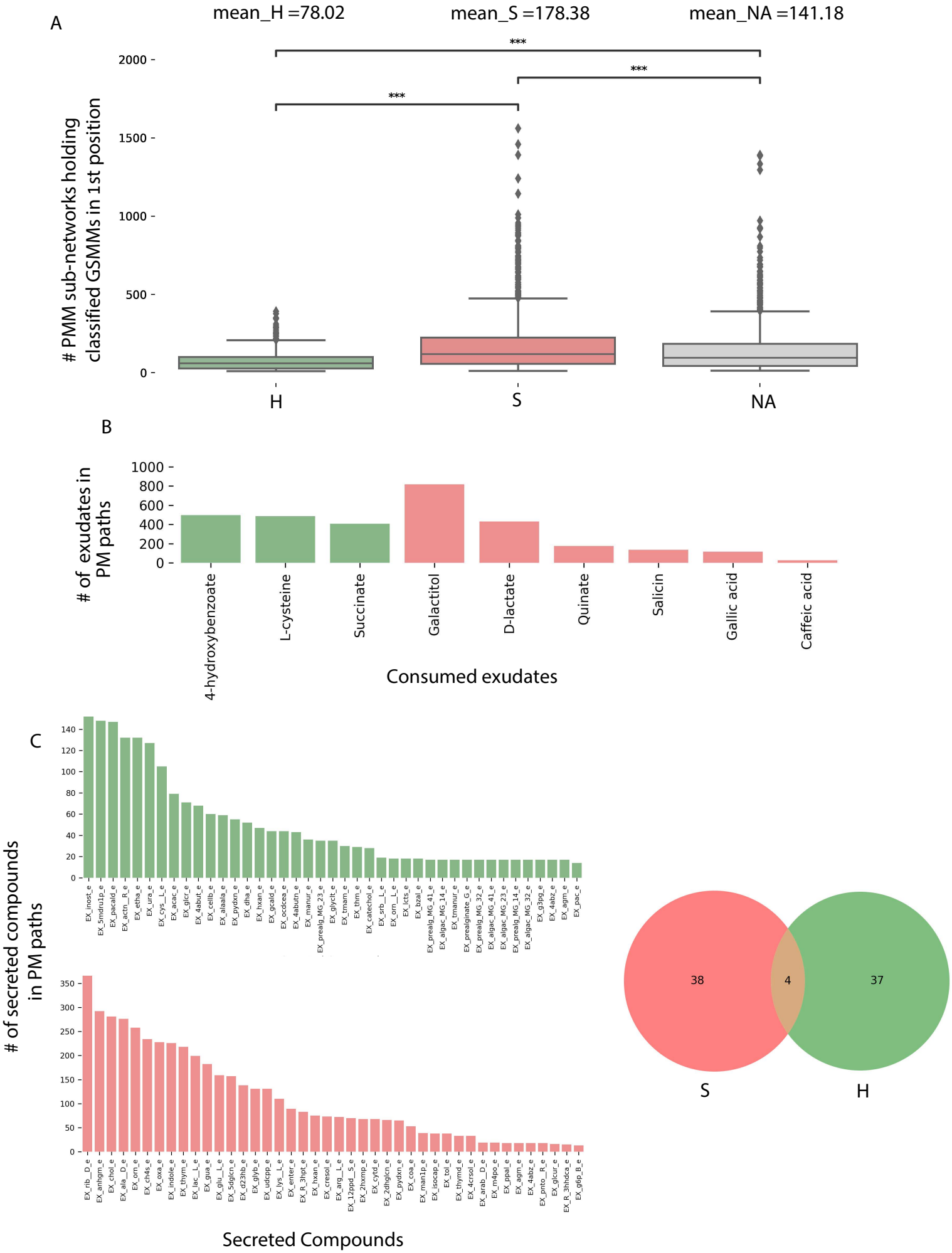

A
mean_H =78.02
mean_S =178.38
mean_NA =141.18
***
***
***
PMM sub-networks holding classified GSMMs in 1st position
2000
1500
1000
500
0
H
S
NA
B
# of exudates in PM paths
1000
800
600
400
200
0
4-hydroxybenzoate
L-cysteine
Succinate
Galactitol
D-lactate
Quinate
Salicin
Gallic acid
Caffeic acid
Consumed exudates
C
# of secreted compounds in PM paths
Secreted Compounds
38
4
37
S
H

Fig. 5 | Characterization of PM and PMM sub-network motifs' features associated with differentially abundant (DA) GSMMs. A. Count distribution of PMM sub-networks determined by GSMMs in the first microbial position classified as either healthy (H), sick (S), or non-associated (NA). Boxes extend from first quantile to the third quantile, middle line represents the median; dots outside whiskers indicate outliers. Distinction of groups was determined using ANOVA followed by the Tukey post-hoc test. Asterisks indicate significance of test (*** <= 0.005 ** <= 0.01 * <= 0.05) **B.** Bar plot indicating the number of exudates significantly associated with H or S-classified PM sub-networks (Hypergeometric test; FDR <= 0.05; green: healthy-H, red: sick-S). **C**. Bar plots indicate the number of secreted compounds in PM sub-networks, which are significantly associated with H-classified (upper, colored green), or S-classified (lower, colored red) (Hypergeometric test; FDR <= 0.05). Compound name identifiers were taken from the BIGG database (X-axes). Venn diagram on the right represents the intersection of secreted compounds derived from both S and H classified PM sub-networks.

Table. 1 | Apple root exudates adopted from the works of Leisso et al.[48,49], included in the rhizosphere environment.	
Amino acids	*L-asparagine*
	L-aspartate
	L-cysteine
	L-valine
	Beta-Alanine
Monosaccharides	*Rhamnose*
	Glycerate
	Ribose
	Galactose
	Xylose
	Erythrose
Sugar-alcohols	*Sorbitol*
	Galactitol
	Glycerol
Other carbohydrates	*3,4-Dihydroxybenzoate*
	3,4-Dihydroxy-trans-cinnamate
	4-Hydroxybenzoate
	Benzoate
	Esculin
	Ferulate
	Gallic acid
	Quinate
	Salicin
	Trans-Cinnamate
Organic acids	*D-Galacturonate*
	D-Lactate
	D-Malate
	Malonate
	Oxalate
	Pyruvate
	Succinate
Carbamides	*Urea*
Fatty acyls	*Octadecanoic acid*

Supplementary information

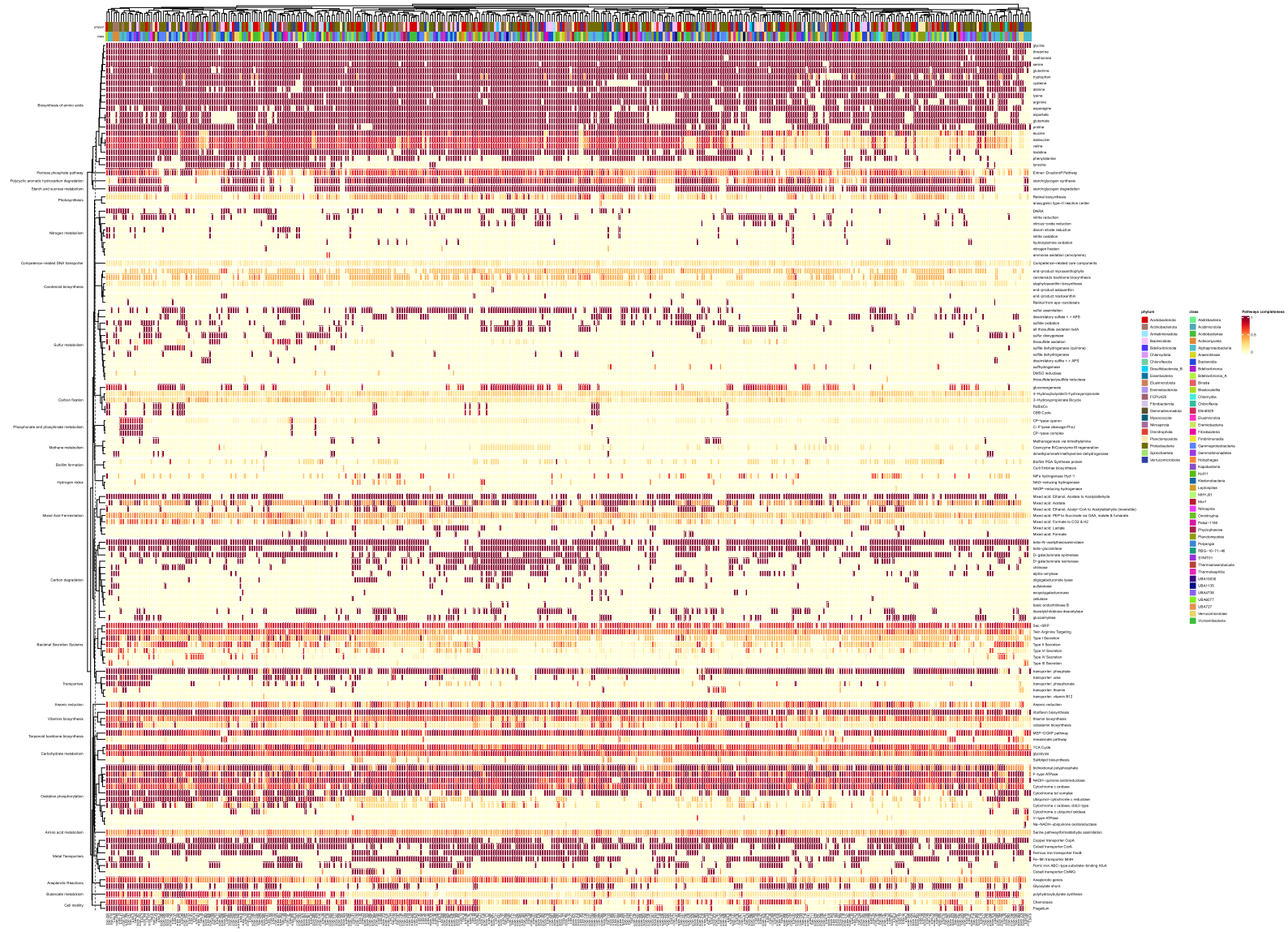
A metabolic modeling-based framework for predicting trophic dependencies in native rhizobiomes of crop plants

Ginatt et al.

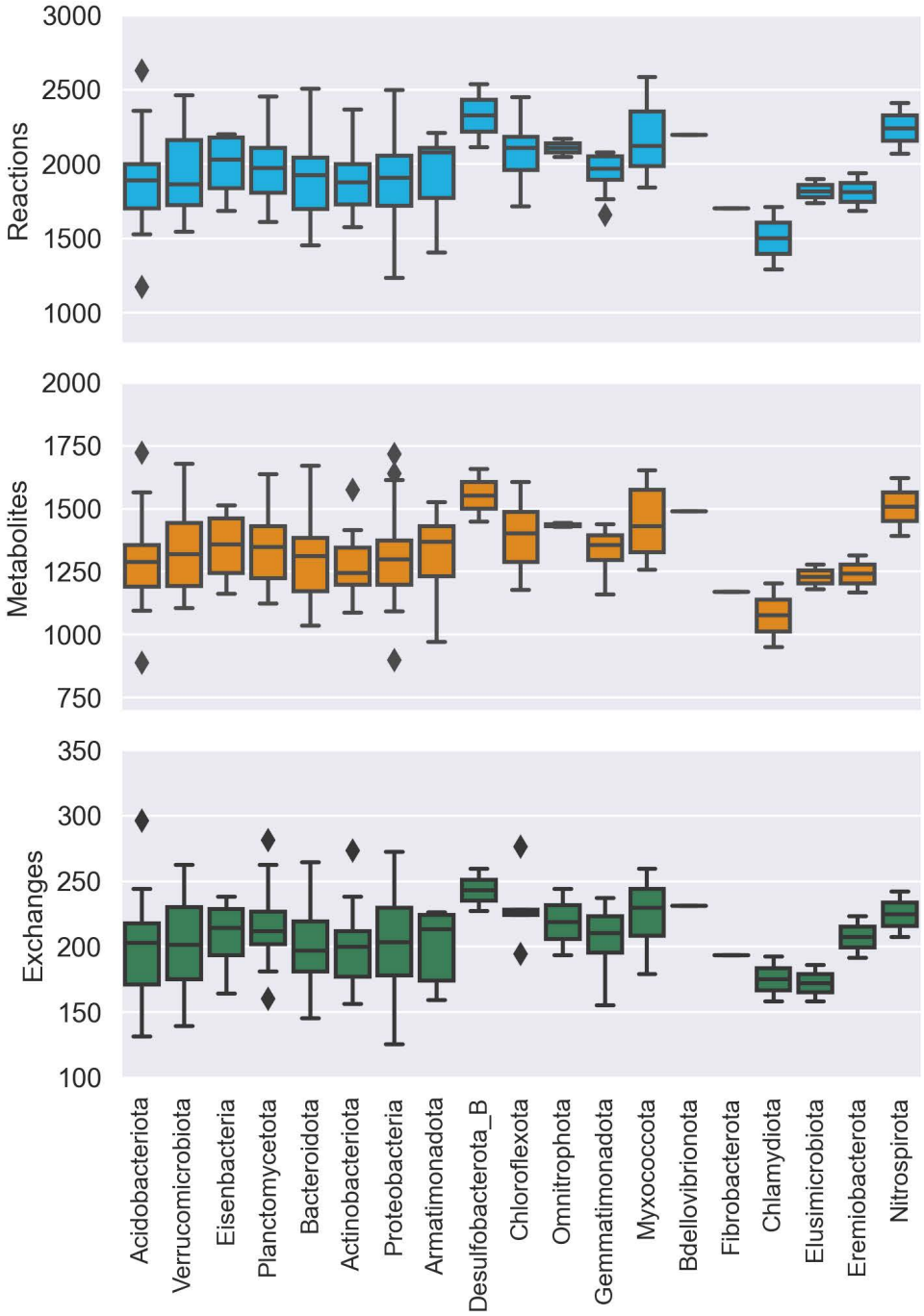
List of Content

Supplementary Figures 1-8

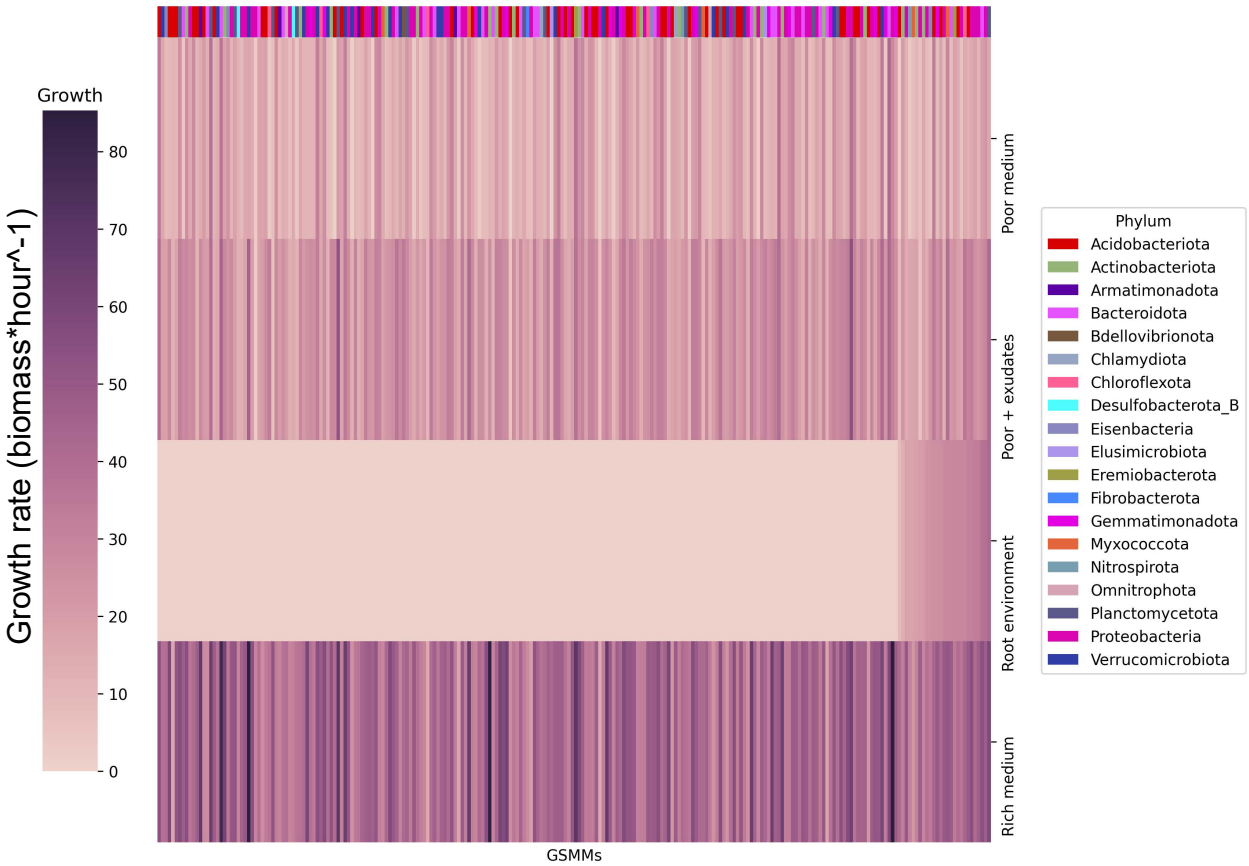
Supplementary Tables 1-3



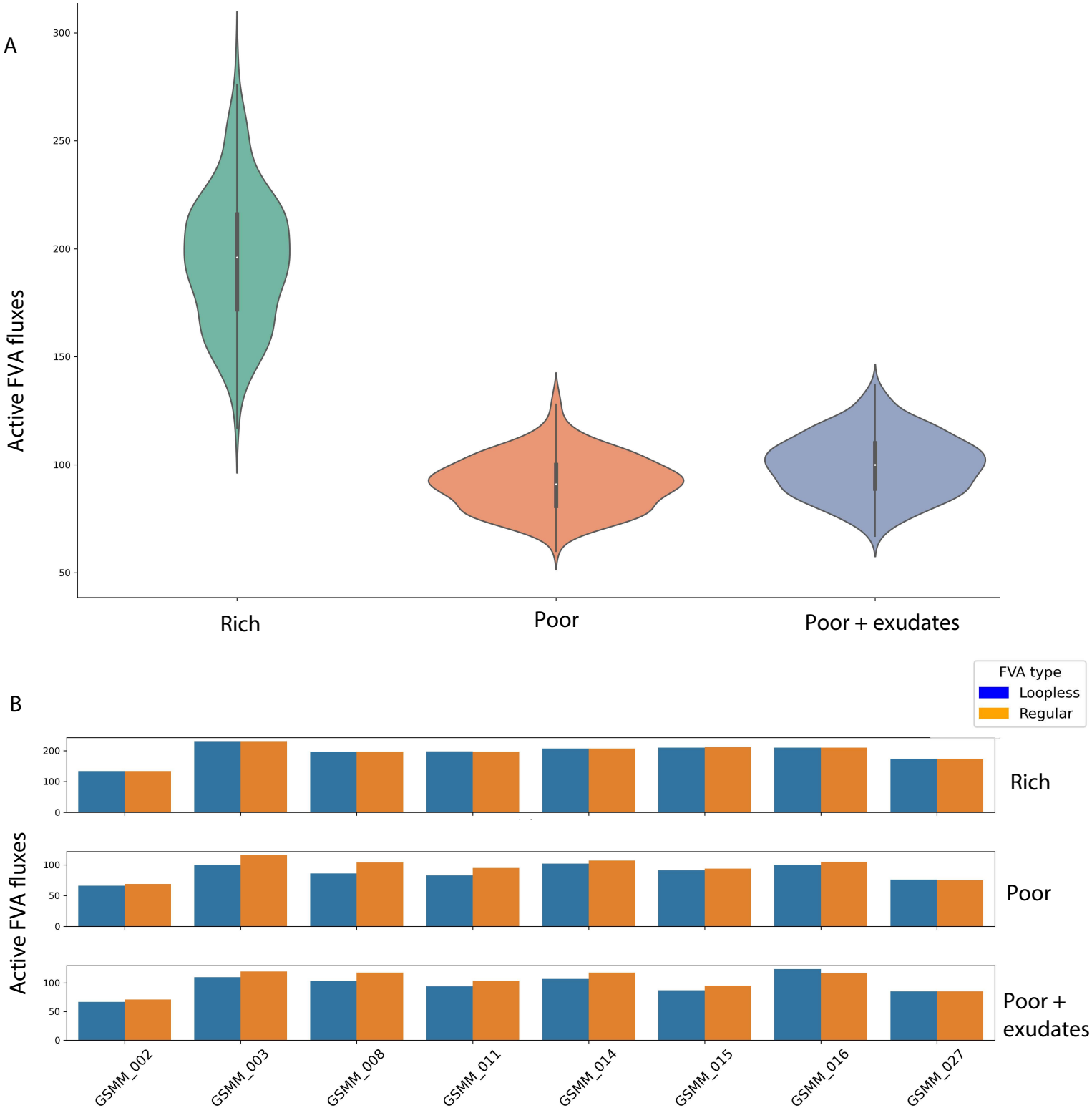
Supplementary Figure 1. Metabolic pathway completion analysis of community de-replicated MAGs. The presented cluster map links the MAGs with different metabolic pathways based on completion values of a specific metabolic pathway. Values are concluded based on the ratio between the count of EC numbers present in a MAG, which correspond to a specific pathway (derived from annotated genes), and the total EC numbers in that specific KEGG pathway. Left axis represent the KEGG KO of a set of specific functions, which are indicated on the right axis. Top and bottom color bars on the upper axis represent the taxonomic annotation of the MAGs (phylum and class, respectively), which are indicated on the bottom axis.



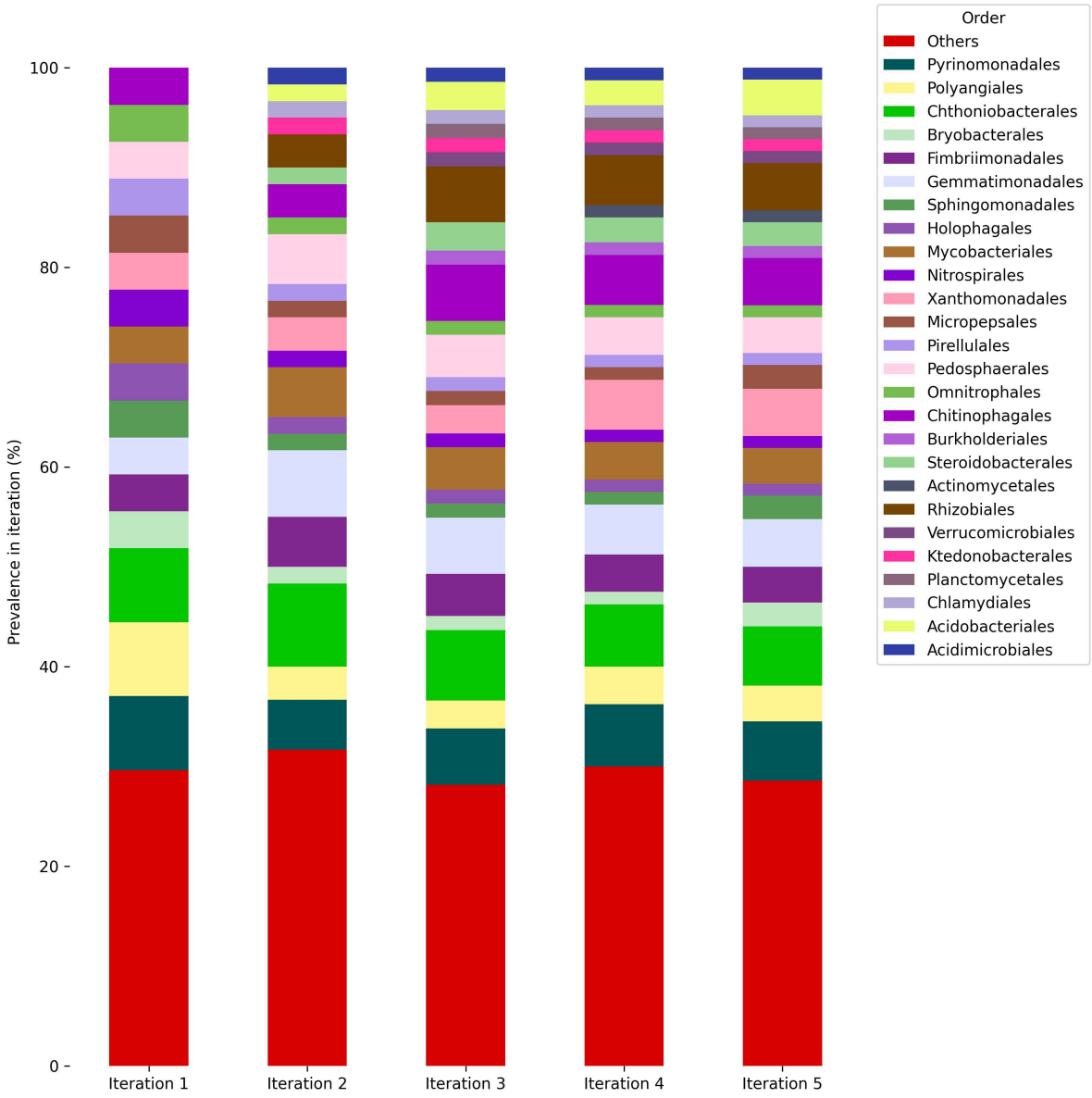
Supplementary Figure 2. Distribution of GSMM attributes at the phylum level. Subplots of reactions, metabolites and exchanges presenting the distribution of the respective attributes for each group of bacteria, affiliated with a specific phylum.



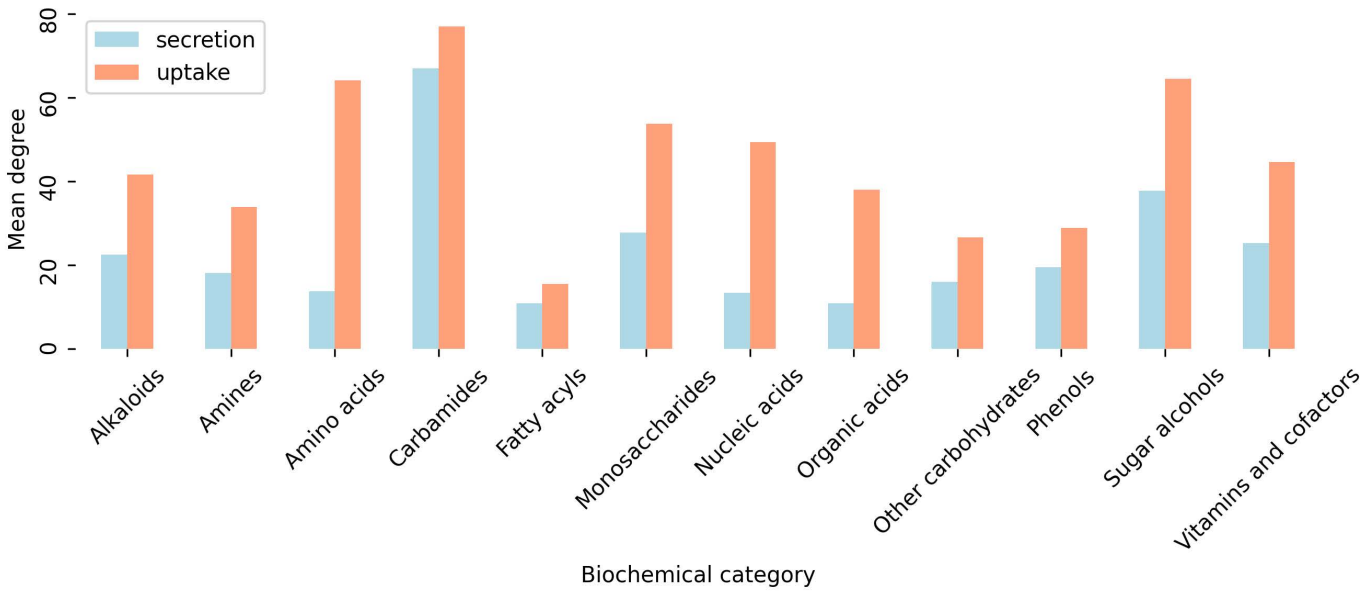
Supplementary Figure 3. Bacterial growth rates in different environments. Each row represents the growth rate (biomass increase hour-1) of species in a specialized medium; from top to bottom: poor medium, poor medium supplemented with root exudates, root environment medium and rich medium. Each column represents a GSMM. Models are sorted by their growth score in the root environment medium. Poor, Rich, and Poor + exudates media are model specific, thus sustain growth of all models. The horizontal color bar on top of the plot represents the phyla of the corresponding GSMM. Actual growth values are provided in Supp. Data 5.



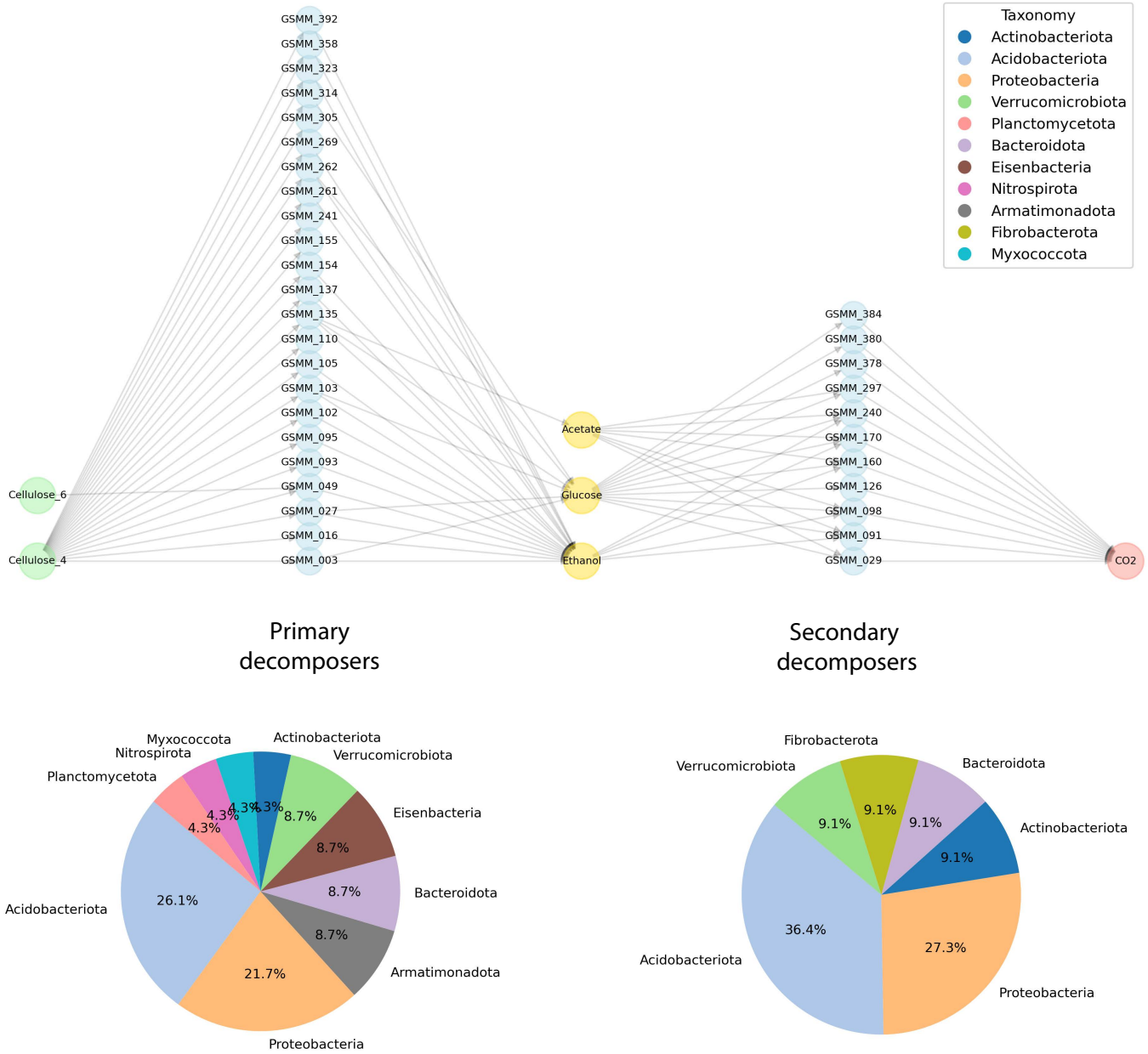
Supplementary Figure 4. FVA performances of GSMMs in different environments (Supp. Fig. 3, Supp. Data 5). **A**. Distribution of potentially active exchange reactions (non-zero minimum FVA flux) in the different environments . Solid line inside each violin indicates the interquartile range (IQR). White point in IQR indicates the median value. Whiskers extending from the IQR indicate the range within 1.5 times the IQR from the quartiles. Violin width at a given value represents the density of data points at that value. **B**. Loopless FVA scores compared to regular FVA for models in the 3 different environments. Bars indicate the count of active fluxes (non-zero minimum FVA flux). Only a subset of models was used for this analysis.



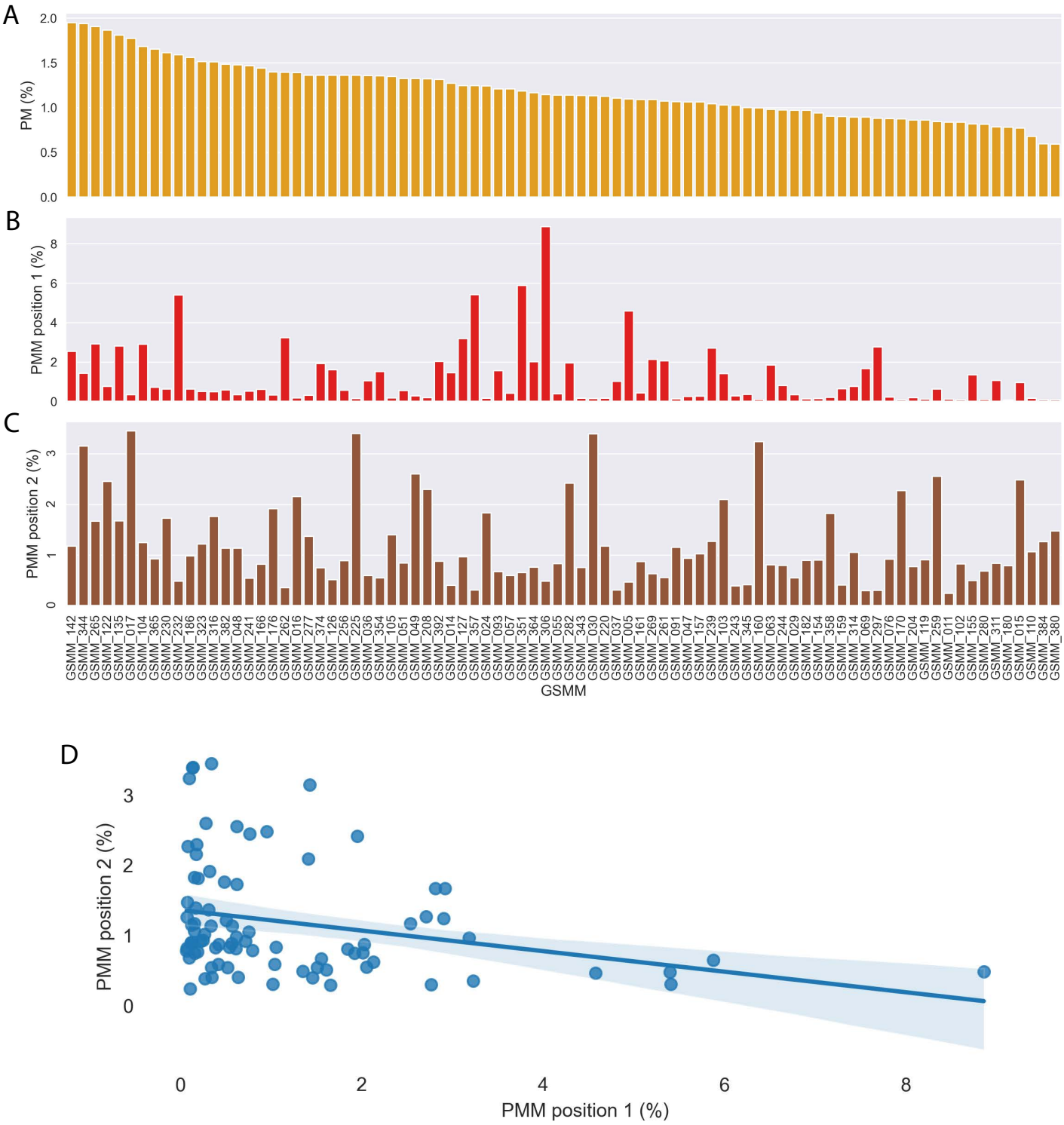
Supplementary Figure 5. **Phylogenetic distribution of models growing along iterations at the order level.** Bars represent the taxonomic distribution of the community (i.e., overall models that grew along iteration, with respect to the corresponding iteration). Colors are indicative of different orders. Orders with insufficient nomenclature were classified as "Others".



Supplementary Figure 6**. Uptake and secretion degrees of biochemically-classified metabolite nodes, as inferred from the communal interaction network.** Values indicate mean amount of links between community GSMMs and nodes classified to the corresponding category. Biochemical classification was conducted both based on BRITE database and manually.



Supplementary Figure 7. Application of MCSM over the process of cellulose decomposition as described by Kato et al[1]. 5-partite network exhibiting the uptake of cellulose oligomers (4 and 6 units of connected D-glucose) by primary decomposers, through secretion of intermediate compounds and their metabolization by secondary decomposers to CO2. Distribution of phyla of primary and secondary decomposers is denoted by pie charts. Though MAGs were notconstructed for the original species as in Kato et al., among the primary consumers, species corresponding to the Acidobacteria (Acidobacteriales)[2], Actinobacteria[3], Bacteriodetes[4], Proteobacteria (Xanthomonadales)[5] and Verrucobacteria[6] groups are found to be capable of degrading cellulose compounds via enzymatic mechanisms.



Supplementary Figure 8**. Community GSMM-centric characterization of plant-microbe** P**M**) **and plant- microbe-microbe** (P**MM**) **motifs. A, B, C**. Frequency of community GSMMs in) PM, PMM (first/second position) in subnetwork motifs, respectively. **D.** correlation of GSMMs position frequency in PMM subnetworks; Pearson, P value=0.009.

Supplementary Tables

Supplementary Table 1. Summary table of the mapping of reads to MAGs across treatments. Treatment types are fully descriebd in Berihu *et al*[7].

Treatment	Sample	#reads	#mapped reads vs. Bins	mapping rates
1	BjSa_G210_1	318,022,594	112,624,412	35.41
1	BjSa_G210_2	281,761,476	99,921,545	35.46
1	BjSa_G210_3	416,038,372	145,599,092	35
1	BjSa_G210_4	282,301,036	95,304,805	33.76
1	BjSa_G210_5	447,141,566	160,309,647	35.85
2	BjSa_M26_1	376,878,959	135,763,911	36.02
2	BjSa_M26_2	347,182,908	122,895,292	35.4
2	BjSa_M26_3	416,564,165	144,390,674	34.66
2	BjSa_M26_4	665,109,915	236,939,650	35.62
2	BjSa_M26_5	467,544,898	162,091,275	34.67
3	Can_G210_1	337,813,971	109,814,240	32.51
3	Can_G210_2	326,790,059	105,115,075	32.17
3	Can_G210_3	290,305,045	94,566,244	32.57
3	Can_G210_4	467,206,263	153,295,356	32.81
3	Can_G210_5	317,179,883	98,665,507	31.11
4	Can_M26_1	346,489,491	116,381,526	33.59
4	Can_M26_2	332,022,731	107,875,006	32.49
4	Can_M26_3	281,669,999	93,377,935	33.15
4	Can_M26_4	352,892,914	114,191,672	32.36
4	Can_M26_5	423,521,175	142,376,677	33.62
5	NTC_G210_1	322,470,447	101,793,678	31.57
5	NTC_G210_2	381,163,976	116,560,703	30.58
5	NTC_G210_3	394,506,571	128,257,851	32.51
5	NTC_G210_4	289,809,167	88,881,030	30.67
5	NTC_G210_5	440,239,346	132,706,841	30.14
6	NTC_M26_1	359,395,622	106,970,231	29.76
6	NTC_M26_2	567,592,079	177,868,308	31.34
6	NTC_M26_3	619,940,510	189,976,498	30.64
6	NTC_M26_4	371,318,086	108,134,325	29.12
6	NTC_M26_5	340,890,336	102,616,038	30.1

Supplementray Table 2. Table of compounds used throughout iterations in the MCSM

inorganic compounds	Root exudates	Organic phosphoroic compounds
Sodium	Pyruvate	AMP C10H12N5O7P
Hydrogen peroxide	Benzoate	CMP C9H12N3O8P
Inorganic triphosphate	L-Cysteine	DTMP C10H13N2O8P
Chromate	Urea	D-Glucose 1-phosphate
Iron (Fe3+)	Gallic acid	Glycerophosphoserine
Diphosphate	D-Galacturonate	UDPglucose
Cadmium	L-Asparagine	Coenzyme A
2 Hydroxylamino 4 6 dinitrotoluene	Quinate	D-Mannose 1-phosphate
Copper	L-Valine	2(alpha-D-Mannosyl-6-phosphate)-D-glycerate
Nitrite	L-Aspartate	UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine
Copper	Succinate	2',3'-Cyclic AMP
Arsenite	Galactitol	2',3'-Cyclic CMP
Nickel	4-Hydroxybenzoate	2',3'-Cyclic GMP
Mercury	Beta-Alanine	Guanosine 3 phosphate C10H12N5O8P
Hydrogen	D-Galactose	D-Fructose 6-phosphate
Ammonia	Trans-Cinnamate	Glycerophosphoglycerol
Ammonium	Ferulate	NMN C11H14N2O8P
Potassium	D-Sorbitol	Phosphate
Silver	D-Malate	Diphosphate
Reduced ferredoxin	D-Ribose	Alpha,alpha'-Trehalose 6-phosphate
Oxidized ferredoxin	D-Lactate	3 AMP C10H12N5O7P
Polyphosphate 50	3,4-Dihydroxy-trans-cinnamate	Sn-Glycero-3-phosphocholine
Hydrogen sulfide	D-Xylose	Sn-Glycero-3-phosphoethanolamine
Zinc	Glycerate	Sn-Glycero-3-phospho-1-inositol
Arsenate	Malonate	D-Glucuronate 1-phosphate
Phosphonate	Oxalate	Glycerol 2-phosphate
Magnesium	D-Erythrose	O-Phospho-L-serine
Manganese	Glycerol	Alpha-D-Ribose 5-phosphate
Nitrous oxide	Octadecanoic acid	UMP C9H11N2O9P
Sulfur	Salicin	Undecaprenyl diphosphate
Fe2+ mitochondria	Esculin	2-Phosphoglycolate
H2O	Rhamnose	Propanoyl phosphate
3Fe-4S] damaged iron-sulfur cluster	3,4-Dihydroxybenzoate	Beta D glucose 6 phosphate C6H11O9P
Molybdate		3 CMP C9H12N3O8P
4Fe-4S] iron-sulfur cluster		3 UMP C9H11N2O9P
Chloride		Glycerol 3-phosphate
Superoxide anion		GMP C10H12N5O8P
Sulfate		UDP-N-acetyl-D-glucosamine
Nitrogen		2-Aminoethylphosphonate
Cobalt		Undecaprenyl phosphate
Tetrathionate		
2 Hydroxylamino 4 6 dinitrotoluene		
4 Hydroxylamino 2 6 dinitrotoluene		
Trithionate		
Selenate		
Calcium		
Nitric oxide		
Phosphate		
Thiosulfate		

Supplementary Table 3. Initial medium compounds used in MCSM Cellulose degradation process

BIGG_identifier	Name	elements
EX_cell4_e	Cellulose - 4 glucose units	['C', 'H', 'O']
EX_cell6_e	Cellulose - 6 glucose units	['C', 'H', 'O']
EX_3amp_e	3-Adenosine-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_3cmp_e	3-Cytosine-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_3gmp_e	3-Guanine-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_ade_e	Adenine	['C', 'H', 'N']
EX_agm_e	Agmitine	['C', 'H', 'N']
EX_amp_e	Adenosine-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_arg__L_e	Arginine	['C', 'H', 'N', 'O']
EX_asn__L_e	Aspargine	['C', 'H', 'N', 'O']
EX_ca2_e	Calcium	['Ca']
EX_cl_e	Cholrine	['Cl']
EX_cmp_e	Cytidine-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_coa_e	Co-enzyme A	['C', 'H', 'N', 'O', 'P', 'S']
EX_cobalt2_e	Cobalt	['Co']
EX_cu2_e	Copper	['Cu']
EX_cyst__L_e	Cysteine	['C', 'H', 'N', 'O', 'S']
EX_fe2_e	Fe2+	['Fe']
EX_fe3_e	Fe3+	['Fe']
EX_fe3dcit_e	Fe3+ di-citrate	['C', 'Fe', 'H', 'O']
EX_fe3pyovd_kt_e	Pyoverdine	['C', 'Fe', 'H', 'N', 'O']
EX_feenter_e	Enterobactin	['C', 'Fe', 'H', 'N', 'O']
EX_gln__L_e	Glutamine	['C', 'H', 'N', 'O']
EX_his__L_e	Histidine	['C', 'H', 'N', 'O']
EX_hom__L_e	Homoserine	['C', 'H', 'N', 'O']
EX_ile__L_e	Isoleucine	['C', 'H', 'N', 'O']
EX_k_e	Potassium	['K']
EX_leu__L_e	Leucine	['C', 'H', 'N', 'O']
EX_lys__L_e	Lysine	['C', 'H', 'N', 'O']
EX_mg2_e	Magnesium	['Mg']
EX_mn2_e	Manganese	['Mn']
EX_nh4_e	Ammonium	['H', 'N']
EX_nmn_e	Nicotine-mononucleotide	['C', 'H', 'N', 'O', 'P']
EX_no2_e	Nitrite	['N', 'O']
EX_no3_e	Nitrate	['N', 'O']
EX_o2_e	Oxygen	['O']
EX_orn__L_e	Ornithine	['C', 'H', 'N', 'O']
EX_phe__L_e	Phenylalanine	['C', 'H', 'N', 'O']
EX_pi_e	Phosphate	['H', 'O', 'P']
EX_ppi_e	Diphosphate	['H', 'O', 'P']
EX_pro__L_e	Proline	['C', 'H', 'N', 'O']
EX_salchs4fe_e	Salmochelin	['C', 'Fe', 'H', 'N', 'O']
EX_so4_e	Sulafte	['O', 'S']
EX_thm_e	Thiamin	['C', 'H', 'N', 'O', 'S']
EX_thr__L_e	Threonine	['C', 'H', 'N', 'O']
EX_tol_e	Toluene	['C', 'H']
EX_trp__L_e	Tryptophan	['C', 'H', 'N', 'O']
EX_tsul_e	Thiosulfate	['O', 'S']
EX_tyr__L_e	Tyrosine	['C', 'H', 'N', 'O']
EX_ump_e	Uracil-monophosphate	['C', 'H', 'N', 'O', 'P']
EX_val__L_e	Valine	['C', 'H', 'N', 'O']
EX_zn2_e	Zinc	['Zn']
EX_tet_e	Tetrathionate	['O', 'S']
EX_fe3dhbzs3_e	2-3-dihydroxybenzoylserine	['C', 'Fe', 'H', 'N', 'O']
EX_cu_e	Copper	['Cu']
EX_fcmcbtt_e	Carboxymyobactin	['C', 'Fe', 'H', 'N', 'O']
EX_tton_e	Trithionate	['H', 'O', 'S']
EX_fol_e	Folate	['C', 'H', 'N', 'O']

References

1. Kato, S., Haruta, S., Cui, Z. J., Ishii, M. & Igarashi, Y. Stable coexistence of five bacterial strains as a cellulose-degrading community. *Appl. Environ. Microbiol.* **71**, 7099–7106 (2005).

2. Kulichevskaya, I. S. *et al.* Acidicapsa borealis gen. nov., sp. nov. and Acidicapsa ligni sp. nov., subdivision 1 Acidobacteria from Sphagnum peat and decaying wood. *Int. J. Syst. Evol. Microbiol.* **62**, 1512–1520 (2012).

3. Depart-, M. & Building, L. S. Lignocellulose-degrading actinomycetes. **46**, 145–163 (1987).

4. Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut Bacteroidetes: The food connection. *Front. Microbiol.* **2**, 1–16 (2011).

5. Dow, J. M. & Daniels, M. J. Pathogenicity determinants and global regulation of pathogenicity of Xanthomonas campestris pv. campestris. *Curr. Top. Microbiol. Immunol.* **192**, 29–41 (1994).

6. Bergmann, G. T. *et al.* The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol. Biochem.* **43**, 1450–1455 (2011).

7. Berihu, M. *et al.* A framework for the targeted recruitment of crop - beneficial soil taxa based on network analysis of metagenomics data. *Microbiome* 1–21 (2023) doi:10.1186/s40168-022-01438-1.

References

1. Zhalnina, K. *et al.* Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* **3**, 470–480 (2018).

2. Gomariz, M. *et al.* From community approaches to single-cell genomics: The discovery of ubiquitous hyperhalophilic Bacteroidetes generalists. *ISME J.* **9**, 16–31 (2015).

3. Singh, B. K., Millard, P., Whiteley, A. S. & Murrell, J. C. Unravelling rhizosphere-microbial interactions: Opportunities and limitations. *Trends Microbiol.* **12**, 386–393 (2004).

4. Korenblum, E. *et al.* Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. **117**, (2020).

5. Sasse, J., Martinoia, E. & Northen, T. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends Plant Sci.* **23**, 25–41 (2018).

6. Venturi, V. & Keel, C. Signaling in the Rhizosphere. *Trends Plant Sci.* **21**, 187–198 (2016).

7. Olanrewaju, O. S., Glick, B. R. & Babalola, O. O. Mechanisms of action of plant growth promoting bacteria. *World J. Microbiol. Biotechnol.* **33**, 1–16 (2017).

8. Rawat, P., Das, S., Shankhdhar, D. & Shankhdhar, S. C. Phosphate-Solubilizing Microorganisms: Mechanism and Their Role in Phosphate Solubilization and Uptake. *J. Soil Sci. Plant Nutr.* **21**, 49–68 (2021).

9. Compant, S., Clément, C. & Sessitsch, A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biol. Biochem.* **42**, 669–678 (2010).

10. Finkel, O. M. *et al.* A single bacterial genus maintains root growth in a complex microbiome. *Nature* **587**, 103–108 (2020).

11. Ghosh, S. K., Bera, T. & Chakrabarty, A. M. Microbial siderophore – A boon

to agricultural sciences. *Biol. Control* **144**, 104214 (2020).

12. Kudoyarova, G. *et al.* Phytohormone Mediation of Interactions Between Plants and Non-Symbiotic Growth Promoting Bacteria Under Edaphic Stresses. *Front. Plant Sci.* **10**, 1–11 (2019).

13. Berendsen, R. L., Pieterse, C. M. J. & Bakker, P. A. H. M. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478–486 (2012).

14. Ngalimat, M. S. *et al.* Plant Growth-Promoting Bacteria as an Emerging Tool to Manage Bacterial Rice Pathogens. *Microorg. 2021, Vol. 9, Page 682* **9**, 682 (2021).

15. Mazzola, M. & Freilich, S. Prospects for biological soilborne disease control: Application of indigenous versus synthetic microbiomes. *Phytopathology* **107**, 256–263 (2017).

16. Freilich, S. *et al.* Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* **2**, (2011).

17. Embree, M., Liu, J. K., Al-Bassam, M. M. & Zengler, K. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15450–15455 (2015).

18. Tsoi, R. *et al.* Metabolic division of labor in microbial systems. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2526–2531 (2018).

19. Zengler, K. & Zaramela, L. S. The social network of microorganisms - How auxotrophies shape complex communities. *Nat. Rev. Microbiol.* **16**, 383–390 (2018).

20. Vessey, J. K. Plant growth promoting rhizobacteria as biofertilizers. *Plant Soil* **255**, 571–586 (2003).

21. Toju, H. *et al.* Core microbiomes for sustainable agroecosystems. *Nat. Plants* **4**, 247–257 (2018).

22. Faust, K. & Raes, J. Microbial interactions : from networks to models. **10**, (2012).

23. Widder, S. *et al.* MINI REVIEW Challenges in microbial ecology: building

predictive understanding of community function and dynamics. *ISME J.* **10**, 2557–2568 (2016).

24. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).

25. San León, D. & Nogales, J. Toward merging bottom–up and top–down model-based designing of synthetic microbial communities. *Curr. Opin. Microbiol.* **69**, 102169 (2022).

26. Basile, A. *et al.* Revealing metabolic mechanisms of interaction in the anaerobic digestion microbiome by flux balance analysis. *Metab. Eng.* **62**, 138–149 (2020).

27. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).

28. Cuevas, D. A. *et al.* From DNA to FBA: How to build your own genome-scale metabolic model. *Front. Microbiol.* **7**, 1–12 (2016).

29. Price, N. D., Reed, J. L. & Palsson, B. Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).

30. Stolyar, S. *et al.* Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* **3**, 1–14 (2007).

31. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).

32. Henry, C. S. *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* **28**, 977–982 (2010).

33. Zampieri, G., Campanaro, S., Angione, C. & Treu, L. Metatranscriptomics-guided genome-scale metabolic modeling of microbial communities. *Cell Reports Methods* **3**, 100383 (2023).

34. Heinken, A. *et al.* Genome-scale metabolic reconstruction of 7,302 human

microorganisms for personalized medicine. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-022-01628-0.

35. Faust, K. Microbial Consortium Design Benefits from Metabolic Modeling. *Trends Biotechnol.* **37**, 123–125 (2019).

36. Xu, X. *et al.* Modeling microbial communities from atrazine contaminated soils promotes the development of biostimulation solutions. *ISME J.* **13**, 494–508 (2019).

37. Ruan, Z. *et al.* Interspecies Metabolic Interactions in a Synergistic Consortium Drive Efficient Degradation of the Herbicide Bromoxynil Octanoate. *J. Agric. Food Chem.* **70**, 11613–11622 (2022).

38. Dhakar, K. *et al.* Modeling-Guided Amendments Lead to Enhanced Biodegradation in Soil. *mSystems* **7**, (2022).

39. Taş, N. *et al.* Metagenomic tools in microbial ecology research. *Curr. Opin. Biotechnol.* **67**, 184–191 (2021).

40. Uritskiy, G. V, Diruggiero, J. & Taylor, J. MetaWRAP - A flexible pipeline for genome-resolved metagenomic data analysis 08 Information and Computing Sciences 0803 Computer Software 08 Information and Computing Sciences 0806 Information Systems. *Microbiome* **6**, (2018).

41. Somera, T. S., Freilich, S. & Mazzola, M. Comprehensive analysis of the apple rhizobiome as influenced by different Brassica seed meals and rootstocks in the same soil/plant system. *Appl. Soil Ecol.* **157**, 103766 (2021).

42. Berihu, M. *et al.* A framework for the targeted recruitment of crop‐beneficial soil taxa based on network analysis of metagenomics data. *Microbiome* 1–21 (2023) doi:10.1186/s40168-022-01438-1.

43. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

44. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database.

Bioinformatics **36**, 1925–1927 (2020).

45. Buée, M., de Boer, W., Martin, F., van Overbeek, L. & Jurkevitch, E. The rhizosphere zoo: An overview of plant-associated communities of microorganisms, including phages, bacteria, archaea, and fungi, and of some of their structuring factors. *Plant Soil* **321**, 189–212 (2009).

46. Xu, J. *et al.* The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* **9**, (2018).

47. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).

48. Leisso, R., Rudell, D. & Mazzola, M. Soil Biology & Biochemistry Metabolic composition of apple rootstock rhizodeposits differs in a genotype-speci fi c manner and affects growth of subsequent plantings. *Soil Biol. Biochem.* **113**, 201–214 (2017).

49. Leisso, R., Rudell, D. & Mazzola, M. Targeted Metabolic Profiling Indicates Apple Rootstock Genotype-Specific Differences in Primary and Secondary Metabolite Production and Validate Quantitative Contribution From Vegetative Growth. **9**, (2018).

50. Naveed, M. *et al.* Plant exudates may stabilize or weaken soil depending on species, origin and time. *Eur. J. Soil Sci.* **68**, 806–816 (2017).

51. Stringlis, I. A. *et al.* MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5213–E5222 (2018).

52. Opatovsky, I. *et al.* Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment. *BMC Genomics* **19**, 1–14 (2018).

53. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**, (2020).

54. Dukovski, I. *et al.* A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). *Nat. Protoc.* **16**, 5030–

5082 (2021).

55. Huang, L. M., Jia, X. X., Zhang, G. L. & Shao, M. A. Soil organic phosphorus transformation during ecosystem development: A review. *Plant Soil* **417**, 17–42 (2017).

56. Zheng, L. *et al.* Roles of phosphorus sources in microbial community assembly for the removal of organic matters and ammonia in activated sludge. *Front. Microbiol.* **10**, 1–13 (2019).

57. Bardgett, R. D. & Van Der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).

58. Uzun, M. *et al.* Recovery and genome reconstruction of novel magnetotactic Elusimicrobiota from bog soil. *ISME J.* 1–11 (2022) doi:10.1038/s41396-022-01339-z.

59. Lei, S. *et al.* Analysis of the community composition and bacterial diversity of the rhizosphere microbiome across different plant taxa. *Microbiologyopen* **8**, 1–10 (2019).

60. Ghosh, S. K., Banerjee, S. & Sengupta, C. Bioassay, characterization and estimation of siderophores from some important antagonistic fungi. *J. Biopestic.* **10**, 105–112 (2017).

61. Lu, X., Heal, K. R., Ingalls, A. E., Doxey, A. C. & Neufeld, J. D. Metagenomic and chemical characterization of soil cobalamin production. *ISME J.* **14**, 53–66 (2020).

62. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

63. Justin, K., Edmond, S., Ally, M. & Xin, H. Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Properties. *J. Pharm. Pharmacol.* **2**, 377–392 (2014).

64. Yang, W. *et al.* A Genomic Analysis of Bacillus megaterium HT517 Reveals the Genetic Basis of Its Abilities to Promote Growth and Control Disease in Greenhouse Tomato. *Genet. Res. (Camb).* **2022**, (2022).

65. Kato, S., Haruta, S., Cui, Z. J., Ishii, M. & Igarashi, Y. Stable coexistence of five bacterial strains as a cellulose-degrading community. *Appl. Environ. Microbiol.* **71**, 7099–7106 (2005).

66. Mazzola, M., Hewavitharana, S. S. & Strauss, S. L. Brassica seed meal soil amendments transform the rhizosphere microbiome and improve apple production through resistance to pathogen reinfestation. *Phytopathology* **105**, 460–469 (2015).

67. Mazzola, M. Transformation of soil microbial community structure and Rhizoctonia-suppressive potential in response to apple roots. *Phytopathology* **89**, 920–927 (1999).

68. Balbín-Suárez, A. *et al.* Root exposure to apple replant disease soil triggers local defense response and rhizoplane microbiome dysbiosis. *FEMS Microbiol. Ecol.* **97**, 1–14 (2021).

69. Weiß, S., Liu, B., Reckwell, D., Beerhues, L. & Winkelmann, T. Impaired defense reactions in apple replant disease-Affected roots of Malus domestica ‘M26’. *Tree Physiol.* **37**, 1672–1685 (2017).

70. Weiß, S., Bartsch, M. & Winkelmann, T. Transcriptomic analysis of molecular responses in Malus domestica ‘M26’ roots affected by apple replant disease. *Plant Mol. Biol.* **94**, 303–318 (2017).

71. Sun, N. *et al.* Effects of Organic Acid Root Exudates of Malus hupehensis Rehd. Derived from Soil and Root Leaching Liquor from Orchards with Apple Replant Disease. *Plants* **11**, (2022).

72. Howell, C. R. Seed Treatment with L-Sorbose to Control Damping-Off or Cotton Seedlings by Rhizoctonia solani. *Phytopathology* **68**, 1096 (1978).

73. Zou, C. S., Mo, M. H., Gu, Y. Q., Zhou, J. P. & Zhang, K. Q. Possible contributions of volatile-producing bacteria to soil fungistasis. *Soil Biol. Biochem.* **39**, 2371–2379 (2007).

74. Gomes, V. A. *et al.* Activity of papaya seeds (Carica papaya) against Meloidogyne incognita as a soil biofumigant. *J. Pest Sci. (2004).* **93**, 783–792 (2020).

75. Gao, T. *et al.* Exogenous dopamine and overexpression of the dopamine synthase gene MdTYDC alleviated apple replant disease. *Tree Physiol.* **41**, 1524–1541 (2021).

76. Price, M. Erroneous predictions of auxotrophies by CarveMe. *Nat. Ecol. Evol.* **7**, 194–195 (2023).

77. Machado, D. & Patil, K. R. Reply to: Erroneous predictions of auxotrophies by CarveMe. *Nat. Ecol. Evol.* **7**, 196–197 (2023).

78. Zorrilla, F., Buric, F., Patil, K. R. & Zelezniak, A. MetaGEM: Reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.* **49**, (2021).

79. Graham, E. D. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J.* 1861–1866 (2018) doi:10.1038/s41396-018-0091-3.

80. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

81. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).

82. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, (2013).

83. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522 (2016).

84. Mahadevan, R. Ã. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models $. **5**, 264–276 (2003).

85. Aoki-Kinoshita, K. F. & Kanehisa, M. Gene annotation and pathway mapping in KEGG. *Methods Mol. Biol.* **396**, 71–91 (2007).

86. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure,

dynamics, and function using NetworkX. *7th Python Sci. Conf. (SciPy 2008)* 11–15 (2008).