



**HAL**  
open science

## Multi-genome metabolic modeling predicts functional inter-dependencies in the Arabidopsis root microbiome

Victor Mataigne, Nathan Vannier, Philippe Vandenkoornhuyse, Stéphane Hacquard

### ► To cite this version:

Victor Mataigne, Nathan Vannier, Philippe Vandenkoornhuyse, Stéphane Hacquard. Multi-genome metabolic modeling predicts functional inter-dependencies in the Arabidopsis root microbiome. *Microbiome*, 2022, 10 (1), pp.217. 10.1186/s40168-022-01383-z . hal-03923098

**HAL Id: hal-03923098**

**<https://hal.science/hal-03923098>**

Submitted on 4 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access



# Multi-genome metabolic modeling predicts functional inter-dependencies in the *Arabidopsis* root microbiome

Victor Mataigne<sup>1,2</sup>, Nathan Vannier<sup>2</sup>, Philippe Vandenkoornhuyse<sup>1\*</sup> and Stéphane Hacquard<sup>2\*</sup>

## Abstract

**Background:** From a theoretical ecology point of view, microbiomes are far more complex than expected. Besides competition and competitive exclusion, cooperative microbe-microbe interactions have to be carefully considered. Metabolic dependencies among microbes likely explain co-existence in microbiota.

**Methodology:** In this in silico study, we explored genome-scale metabolic models (GEMs) of 193 bacteria isolated from *Arabidopsis thaliana* roots. We analyzed their predicted producible metabolites under simulated nutritional constraints including “root exudate-mimicking growth media” and assessed the potential of putative metabolic exchanges of by- and end-products to avoid those constraints.

**Results:** We found that the genome-encoded metabolic potential is quantitatively and qualitatively clustered by phylogeny, highlighting metabolic differentiation between taxonomic groups. Random, synthetic combinations of increasing numbers of strains (SynComs) indicated that the number of producible compounds by GEMs increased with average phylogenetic distance, but that most SynComs were centered around an optimal phylogenetic distance. Moreover, relatively small SynComs could reflect the capacity of the whole community due to metabolic redundancy. Inspection of 30 specific end-product metabolites (i.e., target metabolites: amino acids, vitamins, phytohormones) indicated that the majority of the strains had the genetic potential to produce almost all the targeted compounds. Their production was predicted (1) to depend on external nutritional constraints and (2) to be facilitated by nutritional constraints mimicking root exudates, suggesting nutrient availability and root exudates play a key role in determining the number of producible metabolites. An answer set programming solver enabled the identification of numerous combinations of strains predicted to depend on each other to produce these targeted compounds under severe nutritional constraints thus indicating a putative sub-community level of functional redundancy.

**Conclusions:** This study predicts metabolic restrictions caused by available nutrients in the environment. By extension, it highlights the importance of the environment for niche potential, realization, partitioning, and overlap. Our results also suggest that metabolic dependencies and cooperation among root microbiota members compensate for environmental constraints and help maintain co-existence in complex microbial communities.

## Introduction

Microorganisms are known to colonize macroorganisms by establishing a dense network of interactions and contributing to essential functions that maintain their host homeostasis [1]. These functions are varied, ranging from protection against pathogens to nutrient uptake, and resistance to stresses such as heat or drought [2, 3].

\*Correspondence: philippe.vandenkoornhuyse@univ-rennes1.fr; hacquard@mpipz.mpg.de

<sup>1</sup> Université de Rennes 1, CNRS, UMR6553 ECOBIO, Campus Beaulieu, 35000 Rennes, France

<sup>2</sup> Max Planck Institute for Plant Breeding Research, Department of Plant Microbe Interactions, 50829 Cologne, Germany



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Together with other factors such as temperature, pH, oxygen, nutrients [4], or priority effects [5], the host itself is a niche that influences the composition of its microbiota [6, 7]. In plants, soil can be considered as a reservoir of microorganisms from which microorganisms are recruited to form the root microbiota [8, 9] and where root exudates play an active role [10, 11]. Last, microbe-microbe interactions are essential in shaping the structure and dynamics of microbiota [12–14], resulting in a dense network of interactions [15].

Understanding the diverse interactions between microbes is a critical step in achieving a holistic, community-level understanding of microbiota functioning. These microbe-microbe interactions exist on a spectrum from competition to cooperation [16], depending on time, space, the presence of other species, and energetic cost (etc.) [12]. Two main competing theories explain microbial assemblages. The niche differentiation theory states that phylogenetically similar species are more likely to compete due to their shared functional traits and overlapping resources, leading to less probable co-existence [17]. The habitat filtering theory suggests that dominant species exhibit similar functional traits because their presence is determined by environmental parameters [18]. The relative importance of competition and cooperation remains unclear [16, 19]. Different approaches have produced conflicting results: several computational approaches predict cross-feeding possibilities whereas several *in vitro* experiments rather pointed to competition [16]. Nevertheless, metabolic interactions and particularly metabolic dependencies are reported to play a major role in maintaining community diversity and stability and in explaining microbial co-existence [15, 20–24]. Extracellular metabolites can thus play a major role in microbial community assembly [25], and metabolic dependencies among strains may explain why some microbes cannot be cultured in standard laboratory conditions [26].

Microbial Systems Ecology is now regularly used to model complex systems such as ecological processes [27–31]. The acquisition and analysis of -omics data, coupled with modeling approaches, make it possible to computationally predict an organism's resource use, biosynthetic capabilities, deficiencies, and growth in different conditions, notably available nutrients [26, 28], hereafter referred to as “nutritional constraints” (Table 1). These models rely on the reconstruction of metabolic networks, genome-scale metabolic models (GEMs) from annotated genomes [31, 32]. Thus, making it possible to predict fundamental niche overlaps and competition between members of the same microbial community [33]. Studying the phylogenetic structure of microbial communities also enables the detection of correlations between the phylogenetic signal and metabolism [34–37].

We investigated *in silico* how phylogeny shapes GEMs at the scale of both individual strains (1) and small random combinations of strains (SynComs) (2), tested the strength of the effect of the constraint applied by available nutrients (including root exudates) on GEMs (3) and inspected whether metabolic cooperation among strains can alleviate these nutritional constraints (4). We tested four hypotheses: (i) unconstrained metabolism (see definition in Table 1) is highly clustered by phylogeny, meaning that predicted producible metabolites are differentiated or overlap between strains according to their taxonomy (H1); (ii) combinations of GEMs have more producible metabolites than single GEMs depending on the phylogenetic similarity between the corresponding bacteria (H2); (iii) available nutrients may have an impact on the metabolism of bacteria, leading to a reduction of producible metabolites from unconstrained to constrained metabolism (H3); and (iv) metabolic cooperation is common and likely compensates for nutritional constraints by allowing the production of specific key compounds (H4) under the strong hypothesis that every compound produced by a bacterium can be shared with others. This hypothesis is at least partially supported by several studies which analyzed or predicted bacteria and plant secretomes [38–40]. To test these hypotheses, we analyzed a collection of genomes of bacterial strains isolated from the roots of *Arabidopsis thaliana* [41] and used systems biology approaches to predict genome functioning *in silico*. The metabolism of each bacterium was predicted with GEMs reconstructed using genome annotation.

## Materials and methods

Table 1 lists the specific vocabulary and definitions used in this study. For a summary of genome sequence data processing and metrics acquisition, see Fig. 1.

### Genome data

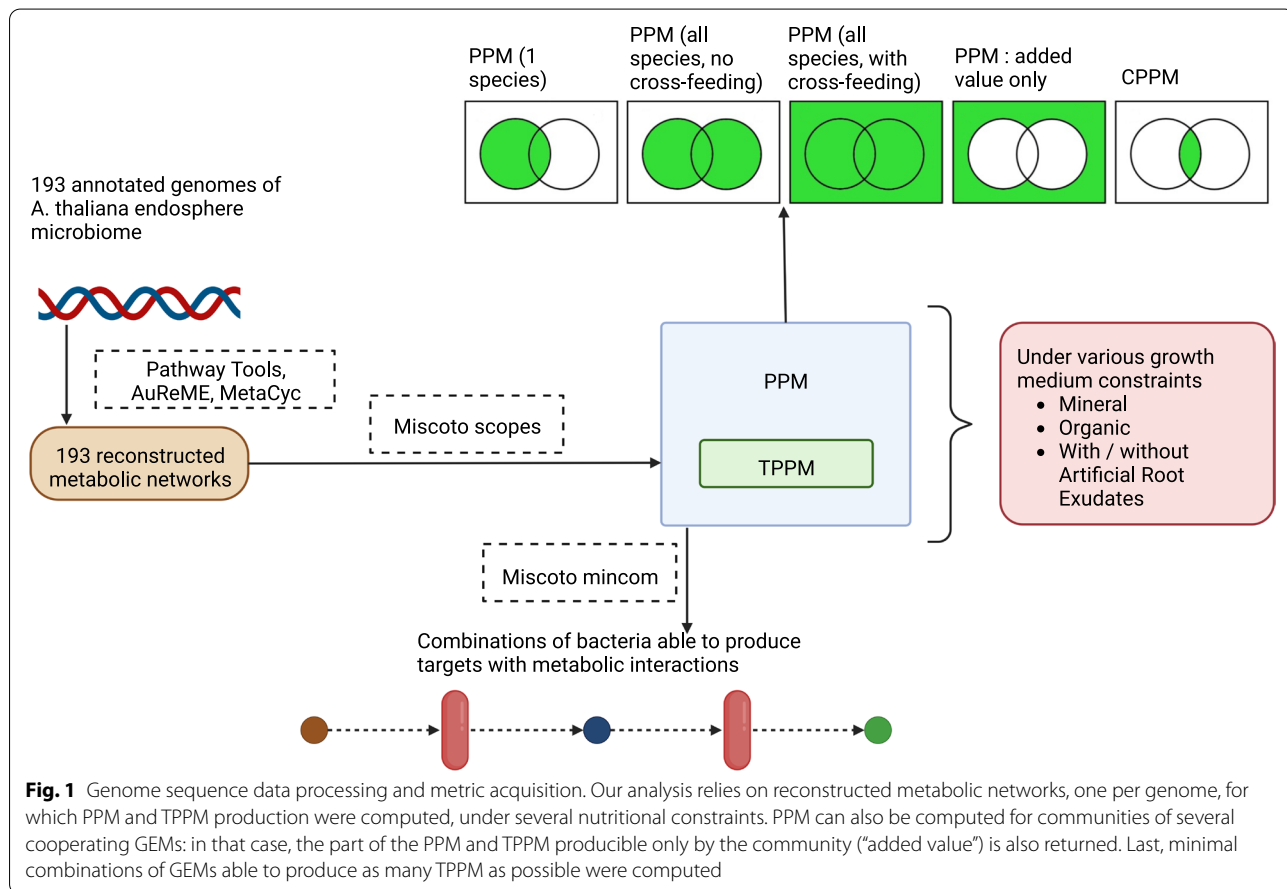
We used 193 annotated genomes [41] belonging to a culture collection of bacteria isolated from *A. thaliana* roots grown in the Cologne Agricultural Soil (Germany) and representing a taxonomically diverse core set of bacteria of the host plant [41]. Annotated genomes [41, 42] were downloaded from the At-SPHERE database (<http://www.at-sphere.com/>). Both the taxonomy and phylogeny of the whole set of genomes [41] were used. The phylogenetic tree was inferred by maximum likelihood [41] from a multi-alignment of 31 bacterial AMPHORA [43] genes obtained with Clustal Omega v1.2.1 [44] passed into FastTree v2.1 [45]. The genomes were sequenced on a HiSeq2500 (Illumina Inc, USA) and were considered as

**Table 1** Description of the metrics used. Here, the term “community” means either the *whole community* or a random subsample of strains (SynCom). The acronyms in bold in the table are used throughout the text

Metric	Description
<b>Nutritional constraint</b>	Available nutrients on which a GEM can rely on (i.e., the initial reactants of the whole network). Nutritional constraints are modeled with simulated growth media. In this paper, an “ <b>unconstrained</b> ” GEM represents its metabolic potential, i.e., all the metabolites it encodes and can theoretically produce.
Predicted Producing Metabolites ( <b>PPM</b> )	The list (number and composition) of all metabolites predicted to be producible by one or several GEMs simultaneously (also referred to as a <b>meta-GEM</b> ), under a nutritional constraint or in the absence of a constraint. This metric is used to summarize the <b>unconstrained and constrained (by available nutrients) metabolism inferred from genomes</b> .
Core Predicted Producing Metabolites ( <b>CPPM</b> )	The part (number and composition) of a community PPM which is individually producible by each GEM individually in a set of GEMs.
Targeted Predicted Producing Metabolites ( <b>TPPM</b> )	A set of 30 metabolites on which part of this study is focused. Their ability to be produced by one or several GEMs is analyzed (number and composition), under a nutritional constraint or in the absence of a constraint.
Community added value	The part of the PPM (number and composition) of several GEMs which is only producible by metabolic interactions within a community (i.e., not producible by a single GEM).
Average phylogenetic distance	The average of all pairwise phylogenetic distances between pairs of strains in a synthetic subsample of strains ( <b>SynCom</b> ). The whole community also has an average phylogenetic distance.

high-quality drafts genomes (see Bai et al. [41]). Paired-end reads were trimmed using Trimmomatic v0.33 [46]. Genomes were assembled using A5 [47] and SOAPdenovo [48]. In all cases, the assembly with the smaller number of contigs was selected. Functional annotation

was carried out using Prokka v1.11 [49] and the SEED subsystem, using the RAST server API [50]. Data assessing the quality of assembly (N50, N90, total length, number of contigs, number of reads) are available as supplementary data no. 3 and no. 4 in Bai et al. [41].



Elsewhere, a graphical summary of this genomic information is provided in Figure S1. We also checked the quality of the annotation of genomes with Busco v5 [51]. Overall, most of the core genes registered in the Busco database were identified in the genomes (i.e., in almost all the genomes, more than 95% of the searched Busco genes were predicted, Supplementary Figure S1C).

#### Reference database

The reference used to link genome annotation to metabolism was the MetaCyc database, a collection of organism-specific Pathway/Genome Databases (PGDBs). MetaCyc contains ~2 500 metabolic pathways from many organisms [52]. Two criteria motivated the choice of this database: first, it is manually curated; second, our tools, which are designed for the reconstruction of the metabolic network (*mpwt* and *AuReMe*) [53, 54] are pre-configured to work with this curated database.

#### Reconstruction of the metabolic networks (GEMs)

Metabolic networks of each genome were simulated with genome-scale metabolic models (GEMs) reconstructed with an automated command-line version of Pathway-Tools [55–57] using the *mpwt* program of the *meta-g2metabo* tool suite [52], then converted in *padmet* and *sbml* format with *AuReMe* and *padmet-utils* [54]. GEMs in *sbml* format were parsed with the Python *lxml* package when needed. All GEMs were drafts, used without gap-filling or manual curation. These steps are usually required to improve the quality of a GEM [58], but are likely to introduce false positives, particularly in the case of poorly known organisms, thereby masking potential metabolic dependencies. Consequently, we chose to rely on drafts of GEMs, i.e., we chose false negatives (due to flaws in genome annotation) over false positives.

#### Genomes and GEM metrics

In order to detect patterns between strains and metabolism, a set of metrics (the full definitions are listed in Table 1 above) were used and applied to a single GEM (i.e., single strains) and to random combinations (“SynComs”) comprising two to 20 GEMs. The Python API of *Miscoto scopes* [59] was used to compute all the predicted producible metabolites (PPM) for a single GEM, SynComs, and the whole community under simulated nutritional constraints. *AuReMe* reports were parsed to record constraint-free PPM. Phylogenetic distances were computed based on the phylogenetic tree with the Python package *ete3* [60]. Genome sizes were available in the annotation data. For each SynCom, the core predicted producible metabolites (CPPM) were also

computed using sets in Python 3 and *AuReMe* and *Miscoto scopes* outputs (Fig. 1). The added value of the full community was also computed with Python sets.

#### Targeted predicted producible metabolites (TPPM)

We studied the production capacity of a single GEM and of the whole community (meta-GEM, where all GEMs can leak and exchange any compound) to produce TPPM. The production of TPPM by a GEM can be computed under various nutritional constraints with simulated growth media (see the dedicated section below). We focused on 30 TPPM in this study: 17 amino acids (serine, alanine, and glutamic acid were excluded because they were present in the artificial root exudates, which were part of the growth media), eight B vitamins (thiamine diphosphate, riboflavin, nicotinate, (R)-pantothenate, pyridoxine, biotin, tetrahydro-folate, adenosylcobalamin), and five phytohormones (auxin, salicylic acid, abscisic acid, ethylene, jasmonic acid). Amino acids were chosen for their fundamental, ubiquitous biological importance. Vitamins were chosen according to their importance in metabolism. Phyto-hormones were chosen with respect to the root-associated trait of the microbial community under study. In addition, according to the reference database, these metabolite biosynthesis pathways and the genes encoding them are relatively well-known, easily predictable, and are expected in the studied taxa, thus reducing the risk of misses in the genome assembly, annotation errors, and false negatives in the reconstructed metabolic networks. When under a nutritional constraint (see below), TPPM production was computed for each GEM with the Python API of *Miscoto scopes* [59]. In the absence of a nutritional constraint, TPPM production was assessed with their absence/presence in *AuReMe* reports.

#### Modeling nutritional constraints (growth media)

Nutritional constraints were established by simulating different growth media. We used two online resources to choose the growth media: *MetaCyc* (previously mentioned), which contains few growth media with their detailed list of nutrients, and the *KOMODO* database [61], which contains a huge number of growth media, together with their composition, and their *MetaCyc* IDs. We modeled nine growth media (five poorly nutritive with mainly mineral nutrients and four highly nutritive, with more carbon sources, Table 2), to which a common list of cofactors [62] was added. Each medium has an “artificial root exudate” enriched version (giving a total of 22 media, Table 2), based on Baudoin et al. [63]. Growth media, except rich media, were chosen based on two conditions: (1) none of the TPPM should be included in a

**Table 2** Summary of the growth media used for nutritional constraints. Since rich media always contain some TPPM, the set of growth media studied varies with the analysis: TPPM were either analyzed all at once, without vitamins, or only phytohormones, depending on the content of the medium concerned. Details on the composition of the media can be found on gitlab

Growth media	Type	Contains TPPM	With mixture	Analyzed TPPM
M63 M9 Mineral medium Hydrogen oxidizing MBM	Poor	No	No	All
Basal phb pyruvate MMJS	Rich	Vitamins		Amino acids, phyto-hormones
LB-lennox enriched		Vitamins + amino acids	Tryptone (amino acids) + yeast extract	Phyto-hormones

medium's nutrients and (2) the nutrients should not contain any "mixture compounds," such as tryptone or yeast extract, whose composition is not described in MetaCyc. However, rich media did not fully meet these conditions because they contain vitamins. Among the rich media, LB broth (Lennox) contains tryptone and yeast extract. All the growth media were used to investigate a range of nutritional constraints on PPM, but some were not used for the analysis of the production of TPPM with which they were incompatible (Table 2).

#### Putative GEM combinations for metabolic interactions

For each nutritional constraint (i.e., each medium), Miscoto mincom [59] (a version implemented into the metage2metabo [53] package) was used with the whole microbial community as input (GEMs in *sbml* format). This tool uses answer set programming, a declarative approach oriented toward combinatorial problem-solving [64, 65]. It rapidly identified all the simplest combinations (called "solutions") of GEMs able to produce as many specified TPPM as possible under growth constraints. More complex solutions (with more GEMs) are ignored. TPPM were either all the aforementioned compounds, only amino acids, only vitamins, or only phytohormones, in accordance with the aforementioned condition "no TPPM in the growth media" (Table 2). Splitting TPPM was also important because results are strongly TPPM-dependent: for example, GEMs with an essential, but rare reaction for the production of a given TPPM will be very frequently represented, potentially masking alternative possible combinations of other TPPM. The results of each run were stored in a *json* format.

#### Quasi-Poisson GLMs

The correlations between the number of PPM/CPPM and SynCom size, genome size, and phylogenetic

distance were computed (in the absence of a nutritional constraint) on random SynComs. First, for each number of members (variable "SynCom size") between 2 and 20 (with a step of 1), 500 SynComs were built by randomly picking strains from the initial pool of 193 strains, without replacement. First, independence between explanatory variables was assessed (Figure S2D). Then, two generalized linear models (GLMs) were built based on a quasi-Poisson distribution (to counter overdispersion of the response metrics, Figure S2, C1 to C4), modeling the response of the number of PPM and CPPM of SynComs according to average phylogenetic distance, average genome size, and SynCom size. A supplemental polynomial (degree 2) term of phylogenetic distance was added to model the curve of the response metrics:

$$Y \sim P(\mu_i, \theta)$$

where  $i$  is a SynCom,  $Y$  is the number of PPM or CPPM it contains,  $P$  its average phylogenetic distance,  $G$  its average genome size, and  $S$  its number of strains (i.e., GEMS). The maximum SynCom size to include in the model was 12 strains. This size was chosen because of its position at the start of the plateau reached by all the metric values or variance (Figure S2, B1 to B4). This was also the size at which bigger SynComs started to show no significant difference from one SynCom size to the next when tested with many bootstrapped subsamples (for each SynCom size: 500 iterations of 50 strains each, one-sided Wilcoxon, Mann-Whitney tests, Figure S3). The model residuals were slightly biased and non-normal, caused by the uneven spreads of values in different-sized SynComs (Figures S4 and S5). Since R2 are not automatically returned with the models we used, they were computed with the following formula: 1—residual deviance/null deviance (which are available in the models' R summaries). The

added value of SynCom's PPM was disregarded for this analysis, because it was computed with Miscoto, which works only under an applied growth constraint.

### Other statistical analyses

Tests of the effect of taxonomy on metric distribution and of the growth media on the number of PPM were performed using non-parametric tests (Wilcoxon and Mann-Whitney rank-sum tests). Corresponding effect sizes were computed with Cliff's delta method. Principal Coordinate Analysis (PCoA) was performed on Jaccard distance matrices, with the *pcoa* and *vegdist* functions of the R packages *ape* and *Vegan* [66]. PERMANOVA was performed on the distance matrices with the *Adonis* function associated with a multivariate analog of Levene's test for homogeneity of variances (PERMDISP2 procedure), with the *betadisper* function of the same package. The growth media dendrogram was built by hierarchical clustering with the R base *hclust* function (with the default "complete" method), after computation of Bray-Curtis distances (with the *vegdist* function) based on the composition of the corresponding whole community PPM (qualitatively, i.e., which compounds are producible under which nutritional constraint, by how many GEMs). Tests involving a taxonomic effect excluded Bacteroidetes and Firmicutes phyla because of their small sample sizes (4 and 7 strains, respectively). The significance threshold was set at 0.01.

### Scripting

Data acquisition and links between tools inputs and outputs (Fig. 1) were organized using homemade Python 3 scripts. All Miscoto outputs were stored in *json* format, and the relevant data they contained (PPM and TPPM, number of genomes producing a TPPM under a given nutritional constraint, etc.) were parsed and stored as *csv* tables. Figures and data analysis were performed with R 4 with the *ggplot2* package [67] and Python 3 with the *matplotlib* and *seaborn* packages. Scripts and data are available at [https://gitlab.com/mataivic/article\\_metabolic\\_modelling\\_thaliana\\_microbiome](https://gitlab.com/mataivic/article_metabolic_modelling_thaliana_microbiome).

## Results

### A link between PPM composition and phylogeny at strain level resolution

We first tested how phylogeny structured the distributions of the different metrics under the hypothesis

that metabolic functions derived from unconstrained GEMs' differ among phylogeny groups (H1). Unconstrained GEMs corresponded to the situation where all the putative genes carried by a genome are considered as functioning (Table 1). Only Actinobacteria and Proteobacteria distributions were statistically tested because the number of strains was insufficient for other phyla. Smaller genome sizes and fewer PPM were observed in Actinobacteria than in Proteobacteria (Fig. 2B,  $p=0.0038$  and  $p < 0.0001$ , with effect sizes of  $-0.27$  and  $-0.48$ ). The bigger the genome, the bigger the number of PPM and TPPM (Fig. 2C). Bacteroidetes in the culture collection ( $n=4$ ) displayed small genomes and small numbers of PPM, while the number of PPM and the size of the genomes of Firmicutes ( $n=7$ ) were similar to those of the other phyla. Differentiation in PPM composition was also detected among phyla based on PERMANOVAs ( $p < 0.001$ ,  $R^2=0.213$ ,  $p$  (permdisp) = 0.005) with well-separated groups (including Bacteroidetes and Firmicutes) observed on PCoA (Fig. 2D). When only considering the 30 selected TPPM, this effect remained significant ( $p < 0.001$ ,  $R^2=0.175$ ,  $p$  (permdisp) = 0.2325), but between-group differentiation was reduced (Fig. 2D, Figure S6D). The same patterns were observed at the class level (Figure S6). This suggests strong metabolic differentiation between phyla at the whole GEM scale, but a more conserved metabolism at the TPPM scale.

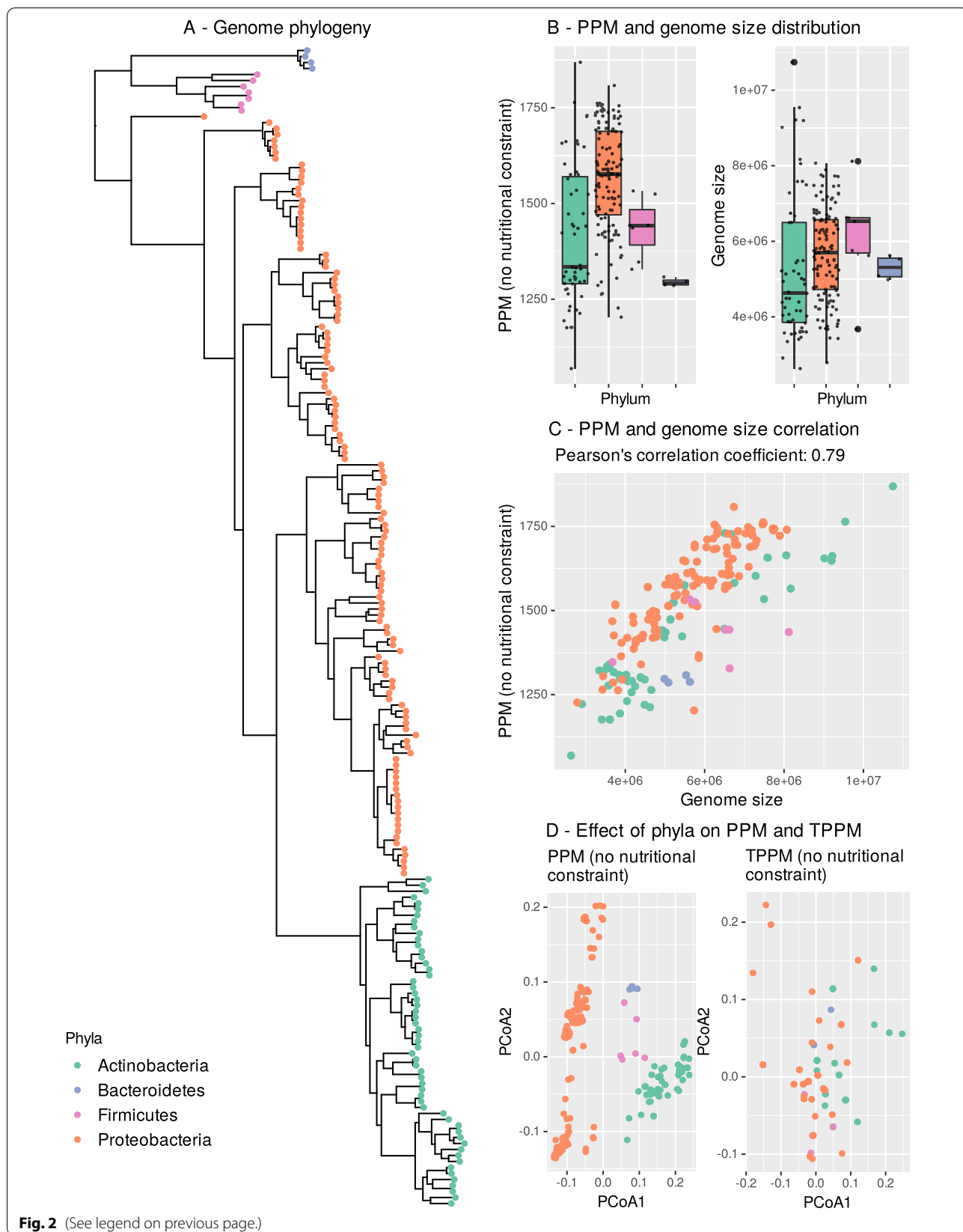
### SynCom PPM under unconstrained metabolism are more diverse than that of single strains and rapidly reach saturation

We extended the previous single-strain GEM approach to random SynComs to analyze the metabolic capacities (number of PPM and CPPM, without nutritional constraint) of merged GEMs (i.e., fully cooperative with all possible metabolic exchanges, H2). The correlations of the number of PPM and CPPM with SynCom size ( $n=2$  to 20 members on plots,  $n=2$  to 12 in GLMs), mean genome size, and mean phylogenetic distance were explored. The three explanatory variables were significantly correlated with both the number of PPM ( $R^2=0.86$ , diagnostic plots in Figure S4) and CPPM ( $R^2=0.82$ , diagnostic plots in Figure S5).

The size of the SynComs was positively correlated with their number of PPM (coefficient=0.36,  $p < 2e-16$ ) and negatively correlated with their number of CPPM (coefficient= $-0.65$ ,  $p < 0.001$ ), but this effect was more

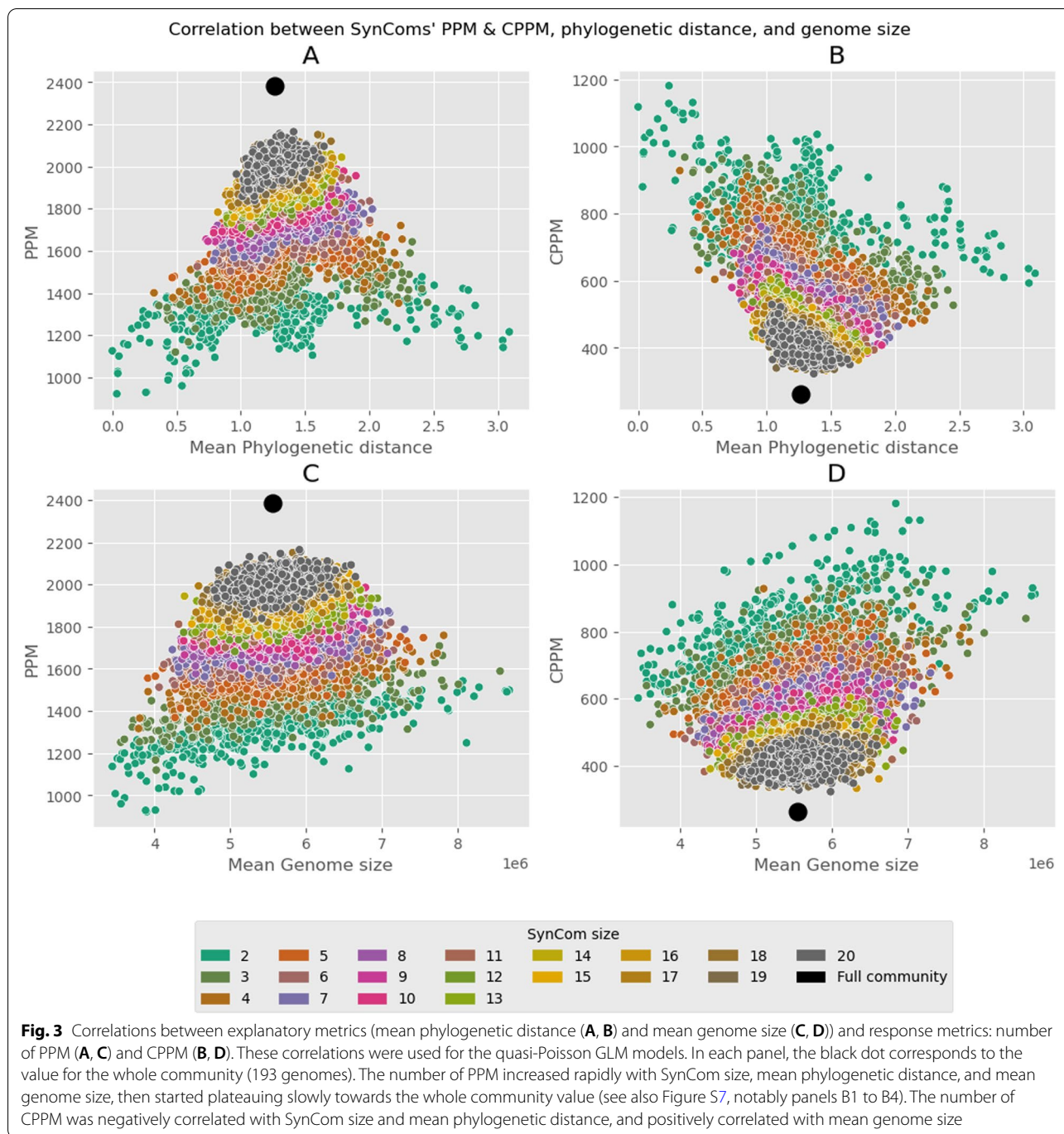
(See figure on next page.)

**Fig. 2** Description of the 193 genomes collected from *A. thaliana* root microbiota **A** Phylogenetic tree (maximum likelihood from a multi-alignment of AMPHORA genes in Clustal Omega). **B** Boxplots displaying the quantitative effect of phyla on genome size and on the distribution of the number of PPM. **C** Dot plots of the number of PPM and TPPM as a function of genome size. **D** PCoA displaying the qualitative effect of phyla on metabolite production (i.e., which compounds are produced by which taxa). Across the panels, colors match genome phyla



**Fig. 2** (See legend on previous page.)





pronounced in small SynComs. Notably, data from SynComs with many strains largely overlapped, plateauing at about 2000 producible metabolites and 400 core metabolites. Increasing SynCom to 193 strains resulted in increasingly fewer variations in SynComs, until the full community was reached with a PPM of 2 383 and a CPPM of 263. Interestingly, increasing the size of the SynComs rapidly returned numbers of PPM and CPPM

close to the whole microbial collection (Fig. 3) likely due to metabolic redundancy. Roughly, SynComs composed of more than ~12 GEMs displayed numbers of PPM and CPPM closer to the values of the whole 193-member community than the values of the smallest SynComs (Fig. 3).

Phylogenetic distance was positively correlated with SynComs PPM (polynomial coefficients=1.32 and -1.83,  $p <$

0.001) and negatively correlated SynComs CPPM (polynomial coefficients =  $-4.37$  and  $1.2$ ,  $p < 0.001$ , Fig. 3A, B). However, PPM reached a peak at a phylogenetic distance of  $\sim 1.3$  for small SynComs, then decreased (Figs. 3A and 4A). This decrease turned into a plateau when SynCom increased in size (Fig. 3A). CPPM among GEMs first decreased, then reached a plateau. This highlighted increasingly diverse metabolism and increasingly less shared metabolism among strains. Most SynComs, particularly big SynComs, were concentrated around this phylogenetic distance rather than spread equally along all possible distances (Fig. 3 and S4, A1 to A4). This corresponds to the values of most combinations of Proteobacteria and Actinobacteria (Figure S7 A to B) and is explained by the fact that these two phyla were the most frequent in the dataset.

Genome sizes were positively correlated with both PPM (coefficient =  $0.23$ ,  $p < 0.001$ ) and CPPM (coefficient =  $0.46$ ,  $p < 0.001$ , Fig. 3C, D). Thus, genome sizes compensated for the negative effect of phylogenetic distance on metabolic redundancy. At a fixed phylogenetic distance, SynComs with bigger genomes displayed both higher metabolic diversity and a bigger core metabolism than SynComs with smaller genomes (Figure S7 C1&C2). SynComs were also aggregated around an optimal value linked to the corresponding optimal phylogenetic distance and the taxonomic composition of SynComs.

Our results validated that GEM combination from phylogenetically distant strains promotes metabolic diversity, restricts metabolic redundancy, and confers additional metabolic capabilities than single-strain GEMs. Beyond these anticipated results, our results also demonstrated that most meta-GEMs were aggregated around the same mean phylogenetic distance, that a few GEMs were sufficient to approximate whole community metabolism, and that big genomes increased both metabolic diversity and redundancy

#### Number and composition of PPM and TPPM are modulated by nutritional constraints

In this analysis, we predicted shifts in GEMs metabolism depending on whether nutritional constraints were applied or not to the model (exemplified by the composition of the simulated growth media, Table 1, H3). “Poor media” refers to severe nutritional constraints (i.e., with mainly mineral nutrients), while “rich media” refers to more permissive constraints (i.e., with more carbon sources, see the “Methods” section).

At the scale of a single GEM, PPM under nutritional constraints were dramatically reduced compared to PPM of unconstrained GEMs (all  $p < 0.001$  with Holm correction, Fig. 4A, right plot). PPM under poor media were significantly fewer in number than PPM under rich media ( $p < 0.001$ , Cliff’s delta effect size =  $-0.8978$  without ARE and  $p < 0.001$ , Cliff’s delta effect size =  $-0.5102$  with ARE, Fig. 4A). Supplementing media with compounds that artificially mimicked the exudate composition of plant roots (ARE) was predicted to significantly increase the number of PPM compared to non-supplemented media (green vs. orange in Fig. 4A, Table 3,  $p < 0.001$ ). Notably, the composition of PPM was more similar (i.e., similar sets of producible compounds) across media containing ARE, irrespective of whether the original media were poor or rich (Fig. 4B). Hence, the addition of ARE is predicted to unlock the production of the same metabolites across media. Most of the GEMs have the potential to produce most of the TPPM in the absence of nutritional constraints (27 out of 30 being producible by a single GEM in the full dataset) but cannot complete the entire pathways under most of the nutritional constraints, Fig. 4D). The supplementation of growth media with ARE was predicted to increase the number of TPPM producible by single GEM (Fig. 4D).

At the whole community scale (i.e., all 193 genomes), the added value provided by a metabolic exchange between all GEMs increased the number of community PPM to similar values regardless of the nutritional constraints (Fig. 4C). Depending on the type of medium (poor/rich) and the absence/presence of ARE, the community-added value increased the number of PPM from 22 to 140% (Fig. 4C) of the number of PPM of the community without metabolic exchange. Thus, GEM functioning is limited by available nutrients that determine which reactions can be activated, but metabolic exchanges between all GEMs of a community can largely compensate for growth constraints, whether they are severe or not.

#### The simplest SynComs are predicted to produce TPPM through metabolic exchanges

After exploring the capacity of the whole community to compensate for nutritional constraints, we explored how smaller assemblages can avoid the same constraints. We used “Miscoto mincom” [59], an answer set programming

(See figure on next page.)

**Fig. 4** Effect of nutritional constraints on PPM and TPPM. **A** Boxplots of the number of PPM per GEM according to the growth medium, compared to no constraints. **B** Hierarchical clustering (“complete” method, Bray-Curtis distances) of growth media according to the composition of the PPM of the whole community (without community-added value). **C** Number of PPM of the complete community according to the growth media, with the value added by metabolic cross-feeding. **D** TPPM production per growth medium. Decimal numbers indicate the fraction of GEMs capable of producing the compound. Gray cells correspond to growth media already containing the TPPM, which are thus irrelevant. On all axes of the figures, poor media are labeled in black, rich media in red, ARE-enriched media in bold, and standard media in plain text

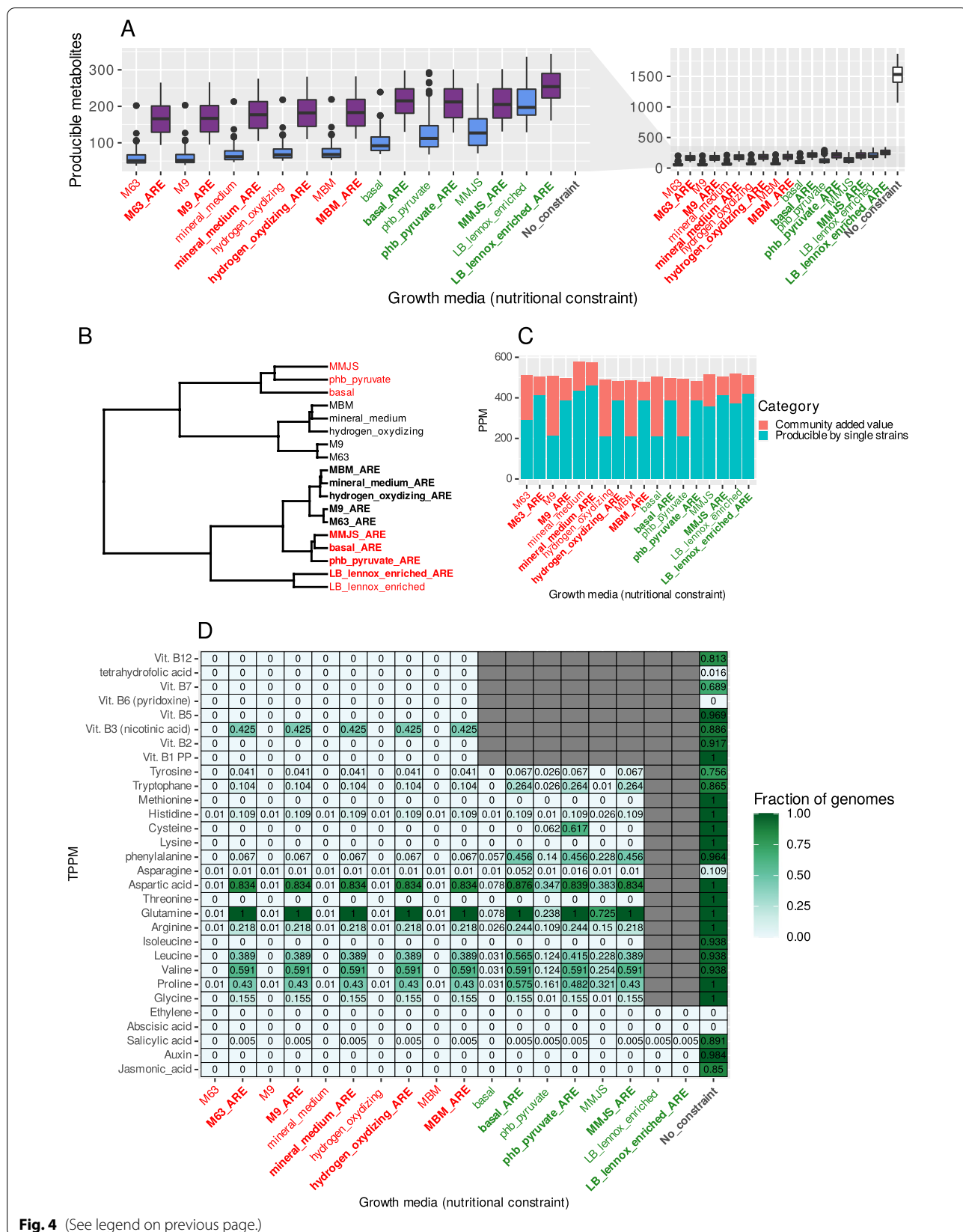


Fig. 4 (See legend on previous page.)

**Table 3** Wilcoxon rank-sum tests on the effect of ARE on the number of TPPM

Media (with and without ARE comparison)	Mann-Whitney <i>p</i> value	Effect size (Cliffs' delta)
M63	$p < 0.001^*$	-0.9786
M9	$p < 0.001^*$	-0.9787
Mineral medium	$p < 0.001^*$	-0.9788
Hydrogen oxidizing	$p < 0.001^*$	-0.9787
MBM	$p < 0.001^*$	-0.9788
Basal	$p < 0.001^*$	-0.9723
phb_pyruvate	$p < 0.001^*$	-0.785
MMJS	$p < 0.001^*$	-0.7081
LB lennox enriched	$p < 0.001^*$	-0.523

solver designed to automatically find (under a simulated nutritional constraint) all the simplest (i.e., smallest) combinations of GEMs that complete the metabolic pathways to produce the 30 specified TPPM (H4). Combinations of GEMs are considered as a meta-GEM when an incomplete pathway in a single GEM can be completed by another, thanks to exchange of intermediate products.

There were many possible—always small—combinations of two or three GEMs sufficient to produce some of the the TPPM. The sets of combinations were also largely identical from one medium to another (Fig. 5A), suggesting an important effect of identical nutrients among growth media and shared reactions among GEMs. The total number of GEMs involved in combinations varied depending on the TPPM included in the analysis. For example, 67 GEMs were returned when all TPPM were considered in poor media (11 in rich media), while the whole set of 193 GEMs (for a few media) was returned when only amino acids or vitamins were considered (Fig. 5A).

Globally, rich media (amino-acid and phytohormone TPPM) produced very few combinations, meaning they allow more GEMs to produce TPPM without predictable mandatory metabolic dependencies or cooperation (Fig. 5A). For example, there were less than 10 minimal combinations of two different GEMs predicted to produce 12 amino acids under rich media, while simulations using poor media returned dozens to thousands of

combinations of two or three GEMs. The addition of ARE also reduced the number of simplest combinations (for example from 2288 to 82 for rich media with vitamins as TPPM), except for the situation “all TPPM + poor media” (Fig. 5A, first row). In such a situation, the ARE-enriched poor media contained more combinations (724 for ARE-enriched poor media, 448 for standard poor media, for 14 producible TPPM).

The combinations of GEMs increased the number of TPPM compared to single-GEM capacities under several growth constraints (Fig. 5B). In poor media, six to seven supplemental amino acids were predicted to be producible by thousands of combinations of two or three GEMs compared to individual GEMs (often six amino acids). Interestingly, for rich media, each amino acid was predicted to be producible by at least one GEM, but no GEM could produce by itself all the 17 targeted amino acids, resulting in a few combinations of strains predicted to exchange end products instead of intermediate metabolites (Fig. 5B). Among vitamins and phytohormones, only nicotinic acid (vitamin B3) and salicylic acid pathways were predicted as complete, with metabolic exchanges required between two GEMs under severe nutritional constraints.

The frequency of GEMs in the solutions was highly varied with a few GEMs occurring in several hundred solutions (Fig. 6A). The other GEMs were much less frequent with only a few occurrences. Thus, for all TPPM, most solutions can be aggregated with a set of seven GEMs (Supplementary Table S1, seven first rows), belonging to strains of the phyla Proteobacteria and, surprisingly, Firmicutes (despite being very underrepresented in the dataset). When TPPM are split according to the category, the majority of solutions can be aggregated with a set of 15 GEMs (Supplementary Table S1). We found no clear correlations between the frequency of a GEM in the solutions and the size of its corresponding genome (Fig. 6B). In fact, the high frequency of these particular GEMs was driven by their strong contribution to the production of a few particular TPPM (Fig. 6A), notably with the case of salicylic acid (Figure S8). In MetaCyc, the bacterial **salicylic acid pathway** is composed of two reactions depending on the availability of chorismate (which has a longer pathway), but other reactions that take place outside

(See figure on next page.)

**Fig. 5** Minimal combinations of GEMs needed to produce TPPM. **A** Summary of the number of combinations of GEMs able to produce TPPM, for all TPPM together, amino acids only, vitamins only, and phytohormones only. **B** Details of producible and unproducible TPPM with single GEM capacities (red) and added values of combinations with metabolic completions (green). TPPM and nutritional constraints are ordered in the same way as in panel **A**. Results are given for each nutritional constraint (black: poor, red: rich), without (plain) or with (bold) ARE. The list of nutritional constraints varies depending on the TPPM considered, in accordance with the no-TPPM-in-media condition (i.e., growth media containing TPPM in their composition were excluded when necessary). On all axes in all the figures, poor media are labeled in black, rich media in red, ARE-enriched media in bold, and standard media in plain text

A - Minimal combinations : summaries

All TPPM

MBM_ARE	16	3	11
MBM	16	3	11
hydrogen_oxidizing_ARE	16	3	11
hydrogen_oxidizing	16	3	11
mineral_medium_ARE	724	3	67
mineral_medium	448	3	67
M9_ARE	724	3	67
M9	448	3	67
M63_ARE	724	3	67
M63	448	3	67

Combinations  
Bacteria per combination  
Total number of bacteria

Amino acids

MMJS_ARE	6	2	5
MMJS	6	2	5
phb_pyruvate_ARE	2	2	3
phb_pyruvate	2	2	3
basal_ARE	6	2	5
basal	7734	3	193
MBM_ARE	60	3	30
MBM	58	3	30
hydrogen_oxidizing_ARE	60	3	30
hydrogen_oxidizing	58	3	30
mineral_medium_ARE	2	2	3
mineral_medium	4518	3	193
M9_ARE	2	2	3
M9	4518	3	193
M63_ARE	2	2	3
M63	4518	3	193

Combinations  
Bacteria per combination  
Total number of bacteria

Vitamins

MBM_ARE	82	1	82
MBM	2288	2	193
hydrogen_oxidizing_ARE	82	1	82
hydrogen_oxidizing	2288	2	193
mineral_medium_ARE	82	1	82
mineral_medium	2288	2	193
M9_ARE	82	1	82
M9	2288	2	193
M63_ARE	82	1	82
M63	2288	2	193

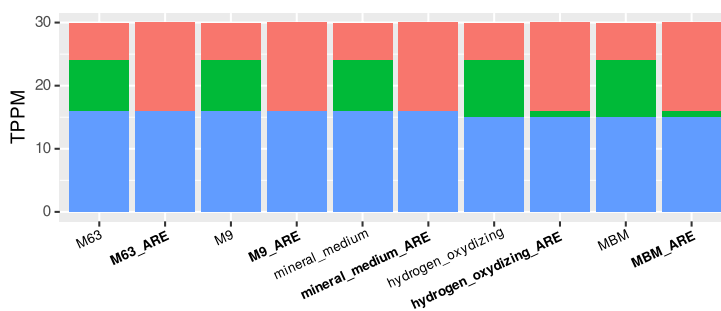
Combinations  
Bacteria per combination  
Total number of bacteria

Phytohormones

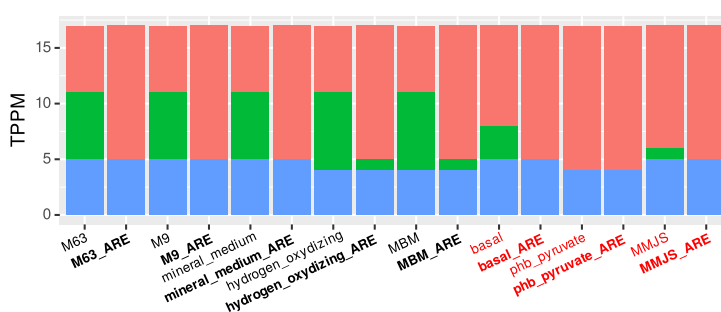
LB_tennox_enriched_ARE	1	1	1
LB_tennox_enriched	1	1	1
MMJS_ARE	1	1	1
MMJS	306	2	143
phb_pyruvate_ARE	1	1	1
phb_pyruvate	1	1	1
basal_ARE	1	1	1
basal	132	2	83
MBM_ARE	111	2	74
MBM	111	2	74
hydrogen_oxidizing_ARE	111	2	74
hydrogen_oxidizing	111	2	74
mineral_medium_ARE	1	1	1
mineral_medium	111	2	74
M9_ARE	111	2	74
M9	111	2	74
M63_ARE	1	1	1
M63	111	2	74

Combinations  
Bacteria per combination  
Total number of bacteria

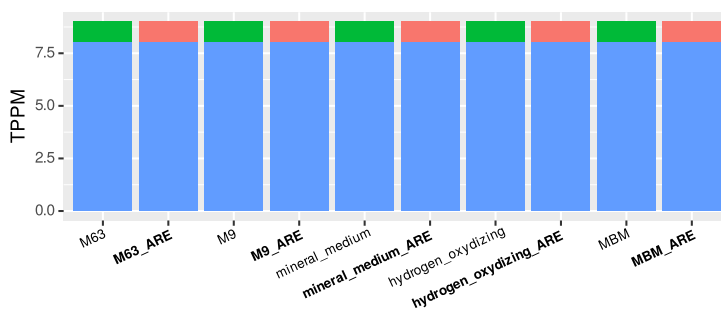
B - Target production



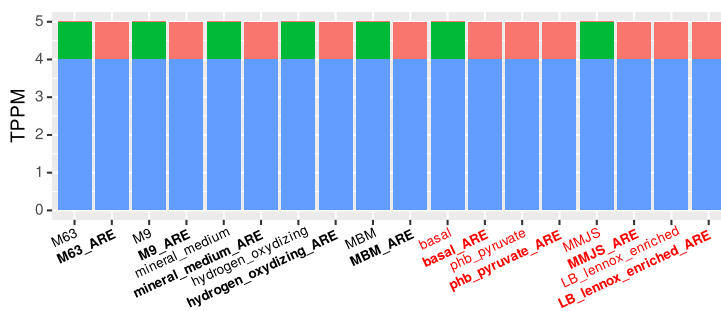
Growth media



Growth media



Growth media



Growth media

Targets Producing\_by\_single\_strains Newly\_producible Still\_unproductive

Fig. 5 (See legend on previous page.)

this pathway are also recorded in the database. The first reaction is encoded into five GEMs only, which are part of the aforementioned seven main GEMs. The second reaction is encoded into 47 GEMs. Most of the combinations for the completion of this pathway of salicylic acid are then built with these GEMs. We noted that 47 GEMs differ from the total of 74 GEMs returned by the solver, which highlights the production of salicylic acid by other means than the pathways described above (as described in Lefevre et al. [68] in plants and Mishra and Baek [69] in plants and bacteria). One GEM was remarkable in that it was the only one predicted to encode the complete salicylic acid pathway: the *Pseudomonas* identified as “Root569” and could putatively produce salicylic acid when growth conditions are optimal.

Hence, the answer set programming solver predicts that many small combinations of GEMs could compensate for diverse nutritional constraints by exchanging metabolites, unlocking the production of several TPPM inaccessible by a single GEM.

## Discussion

### Fundamental niche signature in GEMs

The analysis of the PPM of the 193 GEMs of the *A. thaliana* culture collection demonstrated that phylogenetically related bacterial species share more similar metabolism than distant species. Our first hypothesis (H1) was thus validated in line with the literature [70]. Previous research attempted to predict ecological traits from genomic and metabolic information [33, 71], highlighting differentiation between taxonomic groups and hierarchical conservation within groups [72]. However, finding fundamental niche signatures in GEMs remains a challenging task [28], and complementary trait-based approaches were used to determine ecological attributes or correlate overlapping niches with phylogeny [71, 73].

### Phylogenetic distance, similarity and complementarity, antagonism, and cooperation in SynComs

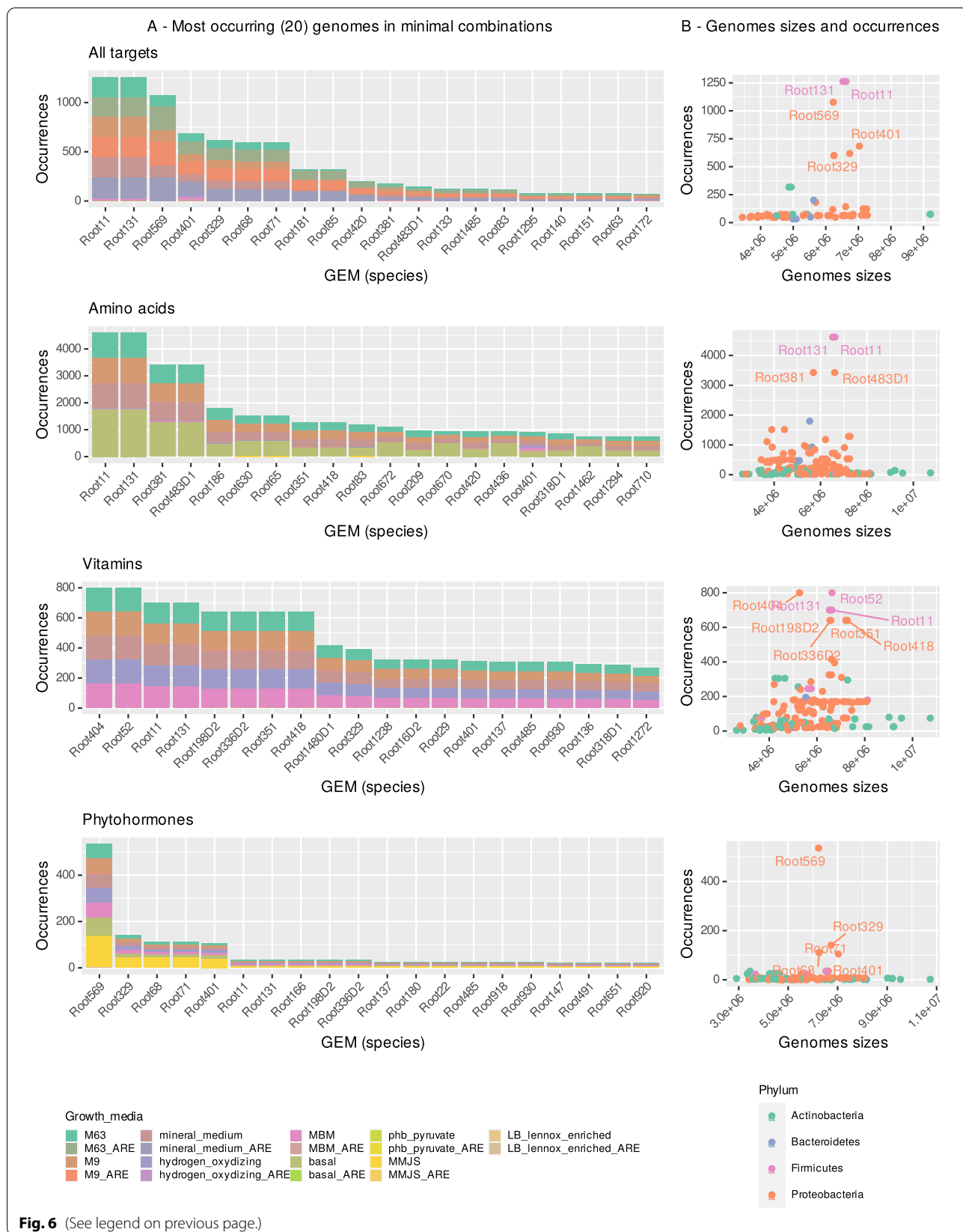
SynComs allowed more diverse metabolism when they contain phylogenetically distant strains, our second hypothesis (H2) was thus validated, along with many unpredicted results (discussed below). Several studies used metrics similar to ours and obtained comparable results [17, 74–76], highlighting a correlation between metabolic similarity/dissimilarity and phylogenetic

distance. Phylogenetically distant bacteria can be predicted to have less metabolic resource overlap and a higher potential to cooperate [76], while phylogenetically closer taxa are expected to compete [17]. Based on phylogenetic similarity and dissimilarity, other studies attempted to go further and interpret whether niche differentiation or habitat filtering was at play. Results are conflicting and research is currently far from a consensus. For example, some authors found that species in the gut microbiome tend to co-occur more frequently with their competitors, thus highlighting habitat filtering [18], while others showed that increased phylogenetic relatedness was correlated with competitive exclusion among bacterivorous protist species [77], thus favoring niche differentiation.

Rather than favoring niche differentiation or habitat filtering, SynComs unconstrained metabolism question the aggregation of most SynComs around a putative optimal phylogenetic distance, along with a metabolic diversity peak for smaller SynComs (Fig. 3A). The peak is probably due to the taxonomic composition of SynComs with a high phylogenetic distance. These SynComs all incorporate a Firmicutes or a Bacteroidetes (Figure S7 A&B), which both display lower metabolic capacities than the other phyla, causing an inevitable decrease in PPM compared to other combinations. However, the aggregation of SynComs around the same phylogenetic distance is more interesting. Closely related bacteria have similar needs and hence face strong competition for resources [15] despite many cross-feeding opportunities. Conversely, very distant bacteria avoid competition, i.e., are adapted to colonize different niches, and their needs only overlap to a limited extent. The observed optimal phylogenetic distance in our results could be driven by the taxonomic composition of the dataset (mostly composed of Proteobacteria and Actinobacteria), and the composition is itself partially determined by the host plant [78, 79]. More importantly, this optimum can be interpreted as a trade-off between metabolic similarity and dissimilarity. Optimum niche overlap would limit competition for resources among closely related species while being sufficient to promote exchanges of metabolites. However, the existence of microbial communities at both ends of the competitive-cooperation spectrum has also been highlighted, mostly dependent on the environment (soil, free-living, or host-associated environments) [80].

(See figure on next page.)

**Fig. 6** Frequencies of GEMs in minimal combinations needed to produce TPPM. **A** The 20 most frequently occurring GEMs in all possible minimal combinations for all TPPM, for amino acids only, for vitamins only, and for phytohormones only. **B** Occurrences of GEMs in minimal combinations as a function of the size of their genome. Text-annotated dots indicate the most frequently occurring GEMs. No relation was found between the size of a genome and its frequency in the combinations



**Fig. 6** (See legend on previous page.)

### Metabolism is nutritional-constraint dependent

Sets of constraints were modeled using growth media with different compositions. Our third working hypothesis (H3) stating that the available nutrients have a significant impact on the metabolism of bacteria was thus validated.

GEMs displayed low metabolic capabilities under severe growth constraints, alleviated by ARE in line with the nature of the genome collection used, i.e., isolated from *A. thaliana* roots. This result suggests that exudates from plant roots are important determinants of the range of producible metabolites in the root microbiome. The effect of ARE was particularly visible for TPPM: the same set of TPPM was produced with the addition of ARE, regardless of the growth medium. This could either reflect the fact that our TPPM are core compounds on which most bacteria rely and are expected to metabolize. TPPM are hypothesized to create similar metabolic patterns among GEMs, activated by ARE.

These findings highlight the importance of nutrient availability for the metabolic profile of the bacterial community [81]. However, there is currently no consensus on the effects of all nutrients on microbial communities. For example, despite nitrogen being a strong predictor of metabolism, its enrichment has different impacts on the diversity and composition of soil microbiota [82]. Root exudates are known to modify microbial communities, but only the effect of a few compounds was recently elucidated [83–85].

The marked difference between unconstrained and constrained metabolism could also be due to the difficulty to correctly model an environment using simulated nutritional constraints. Indeed, some TPPM were impossible to reproduce in any medium, despite full completion rates within several GEMs (such as auxin and B12 vitamin). However, it could also reflect the fact that most growth media are not adapted to all organisms, many of which are known to be difficult to grow in vitro [86]. Hence, metabolic cooperation would be especially relevant with such microorganisms. Reverse ecology has been attempted to avoid in vitro culture problems, for example, by computing the set of nutrients required by a metabolic network to produce biomass [64, 75, 76, 87]. Such approaches allowed the computation of overlapped and differentiated growth requirements of several organisms, thereby advancing our knowledge of the ecological niche and metabolic interactions.

### Metabolic dependencies are predicted to be major drivers of microbial community structure

Metabolic exchanges were found to be essential to improve the metabolic capacities of GEMs, both at

the scale of the whole community and at the scale of a combination of a few GEMs, thus validating our fourth hypothesis (H4). However, we did not expect such big differences between poor and rich media, nor the counterintuitive effect of ARE on very poor media with many TPPM.

### Metabolic exchanges depend on nutritional constraints and compensate for severe growth constraints

Rich growth media and ARE unlocked more PPM and TPPM for single GEMs, with little metabolic cooperation required (Fig. 5). Indeed, growth constraints are likely weaker in rich media since most nutrients are available, thereby unlocking many reactions and their associated metabolic pathways. Reciprocally under poor media, i.e., severe growth constraints, single GEMs were not self-sufficient and more metabolic exchanges were required to produce some TPPM. This was reflected by the marked difference in the number of possible combinations of GEMs able to produce TPPM. However, when considering all TPPM at once in very poor media (M9, M63, mineral medium), supplementation with ARE had a different effect and increased the number of predicted combinations of GEMs (Fig. 5A), which was surprising as they are supposed to improve single GEM autonomy. We concluded that under very severe nutritional constraints, an extremely small subset of GEMs can produce intermediate products and cooperate. In this case, ARE unlocked enough chemical reactions in other GEMs to compensate for the constraints, making several of them self-sufficient for some TPPM. Hence, there was no need for cross-feeding of intermediate compounds, but no GEM was self-sufficient for all the TPPM, suggesting a possible exchange of finished products (hence more combinations). The combined effect of the availability of nutrients in the soil and nutrients secreted by the host might then be a powerful driver of metabolic interactions. In support of this hypothesis, Klitgord and Segré [88] found that there is always a way to predict a growth medium inducing metabolic interactions between pairs of seven species. However, they failed to predict a viable medium for individual species, again highlighting the importance of cooperation. Finally, at the scale of the whole community, the major compensation predicted by cooperation between all GEMs (Fig. 4C) also underlines the importance of metabolic cooperation to counter strong nutritional constraints.

### Minimal combinations of GEMs reflect functional redundancy for the targeted compounds

The results allowed us to predict putative cooperation between GEMs able to produce relevant TPPM of the



root microbiome (amino acids, vitamins, phytohormones). In most TPPM categories, a number of putative combinations were returned when strong nutritive constraints and were applied with or without ARE. These results are in agreement with those obtained by Frioux et al. [59] and Thommes et al. [89]. The number of solutions predicted echoed recent research which predicted a wide range of metabolites that can be secreted without cost, generating countless cross-feeding opportunities [90]. Even if these results are tightly linked to the TPPM and constraints considered, combinations of bacteria are predicted to be able to co-metabolize to complete core, ubiquitous metabolic pathways. This observation can be interpreted as community-level functional redundancy [59] and as an insurance of the completion of metabolic processes for nutrition or interaction with host plants under a range of environmental constraints [91]. These putative redundancies in metabolic completions could play a key role in maintaining stability under variable environmental constraints. Beyond these core metabolic functions, it would be interesting to extend the analysis of minimal communities to secondary, more specific metabolism.

The software used, “Miscoto mincom” [59], worked in a way such that only the most parsimonious solutions are calculated, the simplest combinations of GEMs that fit the applied constraints. However, these putative solutions do not mean more complex combinations do not exist to produce the TPPM. Finally, we did not explore all possible solutions to distinguish mutualistic and unidirectional cross-feeding. One strain could be the final producer of a TPPM by taking advantage of the secretion of another strain, i.e., with no mutualistic exchange, or alternatively, several exchanges may be required. Both interacting behaviors likely coexist, even at the level of a single bacterium, depending on the other bacteria. To date, knowledge is lacking on these behaviors among co-existing bacteria. Other hypotheses concerning the bacterial secretomes may better explain what actually happens in living systems.

#### **The effect of genome size remains unclear**

Bacteria with big genomes are usually considered as generalist species with wider niches [92]. They indeed have a higher unconstrained metabolism (higher PPM) thus likely a higher probability to possess uncommon, important reactions involved in the production of the chosen TPPM. Antagonistic bacteria are also more likely to have larger genomes, linking antagonistic, and generalist strategies [74]. Conversely, bacteria with small genomes are more likely to be involved in metabolic interactions due to their reduced set of reactions [93, 94]. Our results showed that bacteria with larger genomes exhibited both

higher metabolic similarity and complementarity than others (Figure S7 C1&C2), pointing to a putative reservoir of functions which compensated for low phylogenetic distances, but without providing many clues about their orientation towards cooperation or antagonism. The absence of any correlation between GEMs' corresponding genome size and their frequencies in putative metabolic interactions also prevented us from establishing a link between generalist/specialist behavior and cooperation or auxotrophies provider.

#### **Only a few strains are needed to reach the community potential**

As the size of SynComs increased, their metabolism quickly became similar (Fig. 3 and S4). According to these results, the whole community's unconstrained metabolism can be approximated using only a few dozen GEMs (linked with H2). In addition, most combinations for the production of TPPM involved a reduced set of GEMs, recalling previous studies which identified core microbiota composed of a reduced pool of species [5] to perform and/or optimize a biological function [95], analyze the impact of core strains on the whole microbiome [91, 96], or to study host colonization processes [41]. Our results suggest the existence of core functions, echoing other studies reported that the functional stability of the microbiota is maintained regardless of the strains chosen, as long as each functional group is chosen [97].

#### **Conclusions and prospects**

Metabolic diversity and similarity were detected according to the genome taxonomy. The multi-genome metabolic modeling analyses we performed enabled us to predict functional inter-dependencies and revealed a long-lasting ecological paradigm, a trade-off between competition and cooperation. We also found that putative metabolic interactions are common and constraint-dependent, thereby revealing community-level interlinkages and cooperation that make it possible to buffer nutritional constraints. The large number of interactions underlines the importance of richness and diversity in microbial communities for community-level functioning. Taken together, these results provide clues to the best way to decipher microbial interactions in a microbiota beyond the limits of the set of genomes used in the present study.

In this perspective, deeper and more realistic genome-based modeling approaches based on the cost of exchanges, flux analysis, and the use of continuous nutrient depletion over time could provide a closer look at the community-level genomic toolbox used to respond to constraints and to decipher evolutionary and behavioral responses to these constraints in either fluctuating

or constant environments. All approaches leading to putative responses will need to be tested experimentally. With this aim in view, multi-omics data, including secretome [25], a current research frontier, could be used to overcome the challenges to data interoperability. For instance, single isolates could be cultivated on various growth media, and analysis of their secretome would be a way to test the predictions on PPM. A deeper analysis could involve co-cultures of synthetic communities made of two (or more) isolates, predicted to compete or to be metabolically interdependent by the exchange of compounds. Also, cultures of isolates on a minimal medium enriched with the secretome of other isolates would be an interesting way to study the dependency of an isolate on another while limiting any other interaction, such as competition.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01383-z>.

**Additional file 1: Figure S1.** General information about genomic data and genomes annotations of the culture collection used. (A) Distribution of the ratios of N50 over genome length. N50 cannot be directly compared because genomes have various length. Using such a ratio allows a comparison between N50. The closer to 1, the lesser small contigs are needed to cover 50% of the genome. Most genomes have a N50 of approximately 10% of the total genome length, and a few genomes have very high N50. (B) Distribution of the number of contigs in the genomes assemblies. (C) Distribution of the annotation completeness of Busco's core genes. The closer to 100, the more all core genes in Busco's database are present in the annotation. **Figure S2.** In order to fix the maximum Syncom size to inject into Poisson GLMs, we compared every ( $s, s+1$ ) pair of SynComs' PPM and CPPM,  $s$  being a SynCom size in [2, 20]. For each ( $s, s+1$ ) pair, 200 pairs of random subsets of size  $n=50$  SynComs were taken, and their PPM and CPPM were tested with Wilcoxon, Mann & Whitney tests. Boxplots of the 200 p values are displayed for all size comparisons and for PPMs (top) and CPPMs (bottom). Red lines are  $p=0.05$  and green crosses are 1st quantiles. SynComs of 12 strains were chosen as a limit because it was the minimum size at which less than 10% of the p values were under 0.05 for PPM and CPPM. **Figure S3.** Description of the 193 genomes collected from *A. thaliana* root microbiota (A) Phylogenetic tree (maximum likelihood on a multi-alignment of AMPHORA genes). (B) Boxplots displaying the quantitative effect of class on genome sizes and PPM distributions. (C) Plots of the producible metabolites or TPPM as a function of genome sizes. (D) PCoA displaying the qualitative effect of phyla on metabolite production (i.e. which compounds are produced by which taxa). Colours match classes of strains. **Figure S4.** Details of the correlations between explanatory metrics and response metrics for SynComs with two strains only. (A-B) Patterns of SynComs taxonomic composition in the correlation between PPM and CPPM and phylogenetic distance. Proteobacteria and Actinobacteria combinations are responsible for the PPM peak. (C-D) there is also a taxonomic signal among the PPM and CPPM responses to genome size. (E-F) Effect of genome sizes on SynComs' PPM and CPPM. SynComs with a bigger average genome size have both bigger PPMs and CPPMs than SynComs with a smaller average genome size. For each plot, only SynComs with 2 strains are shown. **Figure S5.** Density plots show that for each SynCom size, most values are concentrated around a narrow range. (B1 to B4) Boxplots showing the reach of a plateau (in terms of values and/or variances) for each metric. Only data for SynComs' size below the plateaus were kept in the regression models (size 2 to 12 strains). (C1 to C4) distributions of the different metrics, split by SynCom size. (D) phylogenetic distance and genome size are not correlated, making their use as independent variables valid in the

quasi-Poisson regression. **Figure S6.** Diagnostic plots of the quasi-poisson GLM modelling the response of PPM in SynComs. **Figure S7.** Diagnostic plots of the quasi-poisson GLM modelling the response of the CPPM in SynComs. **Figure S8.** A schematic view of the salicylic acid biosynthesis pathway. There are only two reactions, possessed respectively by 47 and 6 GEMs, Root569 being the only GEM with a complete pathway. Under not constraining growth media, Root569 has the capacity to produce salicylic acid by itself but lost this ability under severe nutritional constraints. In such cases, the set of strains has to exchange intermediate metabolites to produce salicylic acid. A possible hypothesis is that Root569 is incapable of producing chorismate under severe growth constraints, whereas other strains are. **Table S1.** Taxonomy of the most frequently occurring GEMs in the putative combinations of GEMs permitting TPPM production through metabolic exchanges. The TPPM categories in which each GEM is the most involved are mentioned.

## Acknowledgements

We acknowledge Daphne Goodfellow for the English editing. We also acknowledge Europe, CNRS, and DFG for funding.

## Authors' contributions

SH and PV conceived the project and research. VM did the research. SH, PV, and NV mentored VM. VM wrote the first draft of the manuscript. The authors contributed to the final version of this manuscript. The authors read and approved the final manuscript.

## Authors' information

VM was a Ph.D. student (France & Germany). NV was a post-doctoral researcher at the Max Planck Institute of Cologne (Germany). PV is a full professor at the University of Rennes (France). SH is a group leader at the Max Planck Institute of Cologne (Germany).

## Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by a grant from the French Ministry for Research and Innovation, by a grant from the CNRS (EC2CO), and also by a starting grant to S. Hacquard from the European Research Council (MICRORULES 758003), the Max Planck Institute, the Cluster of Excellence on Plant Sciences (CEPLAS), and the "Priority Programme: Deconstruction and Reconstruction of the Plant Microbiota (SPP DECRYPT 2125)," both funded by the Deutsche Forschungsgemeinschaft.

## Availability of data and materials

All the datasets are available online and all the scripts are available on the GitLab [https://gitlab.com/mataivic/article\\_metabolic\\_modelling\\_thaliana\\_microbiome](https://gitlab.com/mataivic/article_metabolic_modelling_thaliana_microbiome). Sequenced genomes are available on the At-SPHERE database (<http://www.at-sphere.com/>)

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

All the authors have validated the submitted manuscript.

### Competing interests

The authors declare that they have no competing interests.

Received: 23 January 2022 Accepted: 23 September 2022

Published online: 09 December 2022

## References

1. Turner TR, James EK, Poole PS. The plant microbiome. *Genome Biol.* 2013;14:209. <https://doi.org/10.1186/gb-2013-14-6-209>.

2. Rodriguez PA, Rothballer M, Chowdhury SP, Nussbaumer T, Gutjahr C, Falter-Braun P. Systems biology of plant-microbiome interactions. *Mol Plant*. 2019;12:804–21. <https://doi.org/10.1016/j.molp.2019.05.006>.
3. Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. *Trends Plant Sci*. 2012;17:478–86. <https://doi.org/10.1016/j.tplants.2012.04.001>.
4. Hacquard S, Garrido-Oter R, González A, Spaepen S, Ackermann G, Lebeis S, et al. Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe*. 2015;17:603–16. <https://doi.org/10.1016/j.chom.2015.04.009>.
5. Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K. Core microbiomes for sustainable agroecosystems. *Nat Plants*. 2018;4:247–57. <https://doi.org/10.1038/s41477-018-0139-4>.
6. Griggs RG, Steenwerth KL, Mills DA, Bokulich NA. Sources and assembly of microbial communities in vineyards as a functional component of winegrowing. *Front Microbiol*. 2021;12:Article 673810. <https://doi.org/10.3389/fmicb.2021.673810>.
7. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: ecology, functions, and emerging trends in microbial application. *J Adv Res*. 2019;19:29–37. <https://doi.org/10.1016/j.jare.2019.03.004>.
8. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol*. 2020;18:607–21. <https://doi.org/10.1038/s41579-020-0412-1>.
9. Xiong C, Zhu YG, Wang JT, Singh B, Han LL, Shen JP, et al. Host selection shapes crop microbiome assembly and network complexity. *New Phytol*. 2021;229:1091–104. <https://doi.org/10.1111/nph.16890>.
10. Haichar FEZ, Marol C, Berge O, Rangel-Castro JI, Prosser JJ, Balesdent J, et al. Plant host habitat and root exudates shape soil bacterial community structure. *ISME J*. 2008;2:1221–30. <https://doi.org/10.1038/ismej.2008.80>.
11. Badri DV, Vivanco JM. Regulation and function of root exudates. *Plant Cell Environ*. 2009;32:666–81. <https://doi.org/10.1111/j.1365-3040.2009.01926.x>.
12. Pacheco AR, Segrè D. A multidimensional perspective on microbial interactions. *FEMS Microb Lett*. 2019;366:fnz125. <https://doi.org/10.1093/femsle/fnz125>.
13. Konopka A. What is microbial community ecology. *ISME J*. 2009;3:1223–30. <https://doi.org/10.1038/ismej.2009.88>.
14. Hassani MA, Durán P, Hacquard S. Microbial interactions within the plant holobiont. *Microbiome*. 2018;58. <https://doi.org/10.1186/s40168-018-0445-0>.
15. Sung J, Kim S, Cabatbat JJT, Jang S, Jin YS, Jung GY, et al. Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat Commun*. 2017;8:15393. <https://doi.org/10.1038/ncomms15393>.
16. Coyte KZ, Rakoff-Nahoum S. Understanding competition and cooperation within the mammalian gut microbiome. *Curr Biol*. 2019;29:R538–44. <https://doi.org/10.1016/j.cub.2019.04.017>.
17. Lam TJ, Stambouliau M, Han W, Ye Y. Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. *PLoS Comput Biol*. 2020;16:e1007951. <https://doi.org/10.1371/journal.pcbi.1007951>.
18. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci U S A*. 2013;110:12804–9. <https://doi.org/10.1073/pnas.1300926110>.
19. Johnson DR, Goldschmidt F, Lilja EE, Ackermann M. Metabolic specialization and the assembly of microbial communities. *ISME J*. 2012;6:1985–91. <https://doi.org/10.1038/ismej.2012.46>.
20. Zomorodi AR, Segrè D. Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nat Commun*. 2017;8:1563. <https://doi.org/10.1038/s41467-017-01407-5>.
21. Morris JJ, Lenski RE, Zinser ER. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio*. 2012;3:e00036–e00012. <https://doi.org/10.1128/mBio.00036-12>.
22. Lovley DR. Happy together: microbial communities that hook up to swap electrons. *ISME J*. 2017;11:327–36. <https://doi.org/10.1038/ismej.2016.136>.
23. Seif Y, Choudhary KS, Hefner Y, Anand A, Yang L, Palsson BO. Metabolic and genetic basis for auxotrophies in Gram-negative species. *Proc Natl Acad Sci U S A*. 2020;117:6264–73. <https://doi.org/10.1073/pnas.1910499117>.
24. Mee MT, Collins JJ, Church GM, Wang HH. Syntrophic exchange in synthetic microbial communities. *Proc Natl Acad Sci U S A*. 2014;21:149–56. <https://doi.org/10.1073/pnas.1405641111>.
25. Fritts RK, McCully AL, McKinlay JB. Extracellular metabolism sets the table for microbial cross-feeding. *Microbiol Mol Biol Rev*. 2021;85:1–29. <https://doi.org/10.1128/mmbbr.00135-20>.
26. Bernstein DB, Dewhirst FE, Segrè D. Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. *eLife*. 2019;8:e39733. <https://doi.org/10.7554/eLife.39733.001>.
27. Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, et al. Signatures of ecological processes in microbial community time series. *Microbiome*. 2018;6:120. <https://doi.org/10.1186/s40168-018-0496-2>.
28. Muller EEL, Faust K, Widder S, Herold M, Martínez Arbas S, Wilmes P. Using metabolic networks to resolve ecological properties of microbiomes. *Curr Opin Syst Biol*. 2018;8:73–80. <https://doi.org/10.1016/j.coisb.2017.12.004>.
29. Kumar M, Ji B, Zengler K, Nielsen J. Modelling approaches for studying the microbiome. *Nat Microbiol*. 2019;4:1253–67. <https://doi.org/10.1038/s41564-019-0491-9>.
30. Mataigne V, Vannier V, Vandenkoornhuysse P, Hacquard S. Microbial systems ecology to understand cross-feeding in microbiomes. *Front Microbiol*. 12:Article 780469. <https://doi.org/10.3389/fmicb.2021.780469>.
31. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson B. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*. 2009;7:129–43. <https://doi.org/10.1038/nrmicro1949>.
32. Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*. 2010;28:977–82. <https://doi.org/10.1038/nbt.1672>.
33. Alneberg J, Bennis C, Beier S, Bunse C, Quince C, Ininbergs K, et al. Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun Biol*. 2020;3:119. <https://doi.org/10.1038/s42003-020-0856-x>.
34. Easson CG, Thacker RW. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front Microbiol*. 2014;5:Article 532. <https://doi.org/10.3389/fmicb.2014.00532>.
35. Aguirre De Cárcer D. A conceptual framework for the phylogenetically constrained assembly of microbial communities. *Microbiome*. 2019;7:142. <https://doi.org/10.1186/s40168-019-0754-y>.
36. Li T, Long M, Li H, Gatesoupe FJ, Zhang X, Zhang Q, et al. Multi-omics analysis reveals a correlation between the host phylogeny, gut microbiota and metabolite profiles in cyprinid fishes. *Front Microbiol*. 2017;8:Article 454. <https://doi.org/10.3389/fmicb.2017.00454>.
37. Goberna M, Verdú M. Predicting microbial traits with phylogenies. *ISME J*. 2016;10:959–67. <https://doi.org/10.1038/ismej.2015.171>.
38. Orsi WD, Richards TA, Francis WR. Predicted microbial secretomes and their target substrates in marine sediment. *Nat Microbiol*. 2017;3:32–7. <https://doi.org/10.1038/s41564-017-0047-9>.
39. Bednarek P, Kwon C, Schulze-Lefert P. Not a peripheral issue: secretion in plant-microbe interactions. *Curr Opin Plant Biol*. 2010;13:378–87. <https://doi.org/10.1016/j.pbi.2010.05.002>.
40. Baran R, Brodie EL, Mayberry-Lewis J, Hummel E, Da Rocha UN, Chakraborty R, et al. Exometabolite niche partitioning among sympatric soil bacteria. *Nat Commun*. 2015;6:8289. <https://doi.org/10.1038/ncomms9289>.
41. Bai Y, Müller DB, Srinivas G, Garrido-oter R, Potthoff E, Rott M, et al. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*. 2015;528:364–9. <https://doi.org/10.1038/nature16192>.
42. Wippel K, Tao K, Niu Y, Zgadzaj R, Kiel N, Guan R, et al. Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nat Microbiol*. 2021;6:1150–62. <https://doi.org/10.1038/s41564-021-00941-9>.
43. Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol*. 2008;9:R151. <https://doi.org/10.1186/gb-2008-9-10-r151>.
44. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539. <https://doi.org/10.1038/msb.2011.75>.

45. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
47. Tritt A, Eisen JA, Facciotti MT, Darling AE. An integrated pipeline for de novo assembly of microbial genomes. *PLoS One*. 2012;7:e42304.
48. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20:265–72.
49. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
50. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*. 2005;33:5691–702.
51. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. busco update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–54.
52. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 2008;36:623–31. <https://doi.org/10.1093/nar/gkm900>.
53. Belcour A, Frioux C, Aite M, Hildebrand F, Siegel A. Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife*. 2020;9:e61968. <https://doi.org/10.7554/eLife.61968>.
54. Aite M, Chevallier M, Frioux C, Trottier C, Got J, Cortés MP, et al. Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Comput Biol*. 2018;14:e1006146. <https://doi.org/10.1371/journal.pcbi.1006146>.
55. Karp PD, Paley S, Romero P. The pathway tools software. *Bioinformatics*. 2002;18:S225–32. [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.S225](https://doi.org/10.1093/bioinformatics/18.suppl_1.S225).
56. Karp PD, Paley SM, Midford PE, Krummenacker M, Billington R, Kothari A, et al. Pathway tools version 24.0: integrated software for pathway/genome informatics and systems biology. *arXiv*. 2020; <https://arxiv.org/pdf/1510.03964.pdf>
57. Karp PD, Latendresse M, Caspi R. The pathway tools pathway prediction algorithm. *Stand Genomic Sci*. 2011;5:424–9. <https://doi.org/10.4056/sigs.1794338>.
58. Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*. 2010;5:93–121. <https://doi.org/10.1038/nprot.2009.203>.
59. Frioux C, Fremy E, Trottier C, Siegel A. Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*. 2018;34:i934–43. <https://doi.org/10.1093/bioinformatics/bty588>.
60. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 2016;33:1635–8. <https://doi.org/10.1093/molbev/msw046>.
61. Oberhardt MA, Zarecki R, Gronow S, Lang E, Klenk HP, Gophna U, et al. Harnessing the landscape of microbial culture media to predict new organism-media pairings. *Nat Commun*. 2015;6:8493. <https://doi.org/10.1038/ncomms9493>.
62. Kim T, Dreher K, Nilo-Poyanco R, Lee I, Fiehn O, Lange BM, et al. Patterns of metabolite changes identified from large-scale gene perturbations in arabidopsis using a genome-scale metabolic network. *Plant Physiol*. 2015;167:1685–98. <https://doi.org/10.1104/pp.114.252361>.
63. Baudoin E, Benizri E, Guckert A. Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. *Soil Biol Biochem*. 2003;35:1183–92. [https://doi.org/10.1016/S0038-0717\(03\)00179-2](https://doi.org/10.1016/S0038-0717(03)00179-2).
64. Laniau J, Frioux C, Nicolas J, Baroukh C, Cortes M, Got J, et al. Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks. *PeerJ*. 5:e3860. <https://doi.org/10.7717/peerj.3860>.
65. Collet G, Eveillard D, Gebser M, Prigent S. Extending the metabolic network of *Ectocarpus Siliculosus* using answer set programming. In: Cabalar P, editor. *Log Program Nonmonotonic Reason*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 245–56. [https://doi.org/10.1007/978-3-642-40564-8\\_25](https://doi.org/10.1007/978-3-642-40564-8_25).
66. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. *vegan: community ecology package*. 2020. <https://cran.r-project.org/package=vegan>
67. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag; 2016. <https://ggplot2.tidyverse.org>. ISBN 978-3-319-24277-4
68. Lefeuvre H, Bauters L, Gheysen G. Salicylic acid biosynthesis in plants. *Front Plant Sci*. 2020;11:article338. <https://doi.org/10.3389/fpls.2020.00338>.
69. Mishra AK, Baek KH. Salicylic acid biosynthesis and metabolism: a divergent pathway for plants and bacteria. *Biomolecules*. 2021;11:705. <https://doi.org/10.3390/biom11050705>.
70. Jaffe AL, Castelle CJ, Matheus Carnevali PB, Gribaldo S, Banfield JF. The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC Biol*. 2020;18:69. <https://doi.org/10.1186/s12915-020-00804-5>.
71. Barberán A, Ramirez KS, Leff JW, Bradford MA, Wall DH, Fierer N. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol Lett*. 2014;17:794–802. <https://doi.org/10.1111/ele.12282>.
72. Martiny JBH, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. *Science*. 2015;350:aac9323. <https://doi.org/10.1126/science.aac9323>.
73. Fahimipour AK, Gross T. Mapping the bacterial metabolic niche space. *Nat Commun*. 2020;11:1–8. <https://doi.org/10.1038/s41467-020-18695-z>.
74. Russel J, Røder HL, Madsen JS, Burmølle M, Sørensen SJ. Antagonism correlates with metabolic similarity in diverse bacteria. *Proc Natl Acad Sci U S A*. 2017;114:10684–8. <https://doi.org/10.1073/pnas.1706016114>.
75. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc Natl Acad Sci U S A*. 2015;112:6449–54. <https://doi.org/10.1073/pnas.1522642113>.
76. Hester ER, Jetten MSM, Welte CU, Lucker S. Metabolic overlap in environmentally diverse microbial communities. *Front Genet*. 2019;10:Article 989. <https://doi.org/10.3389/fgene.2019.00989>.
77. Violle C, Nemergut DR, Pu Z, Jiang L. Phylogenetic limiting similarity and competitive exclusion. *Ecol Lett*. 2011;14:782–7. <https://doi.org/10.1111/j.1461-0248.2011.01644.x>.
78. Choi K, Khan R, Lee SW. Dissection of plant microbiota and plant-microbiome interactions. *J Microbiol*. 2021;59:281–91. <https://doi.org/10.1007/s12275-021-0619-5>.
79. Sun H, Jiang S, Jiang C, Wu C, Gao M, Wang Q. A review of root exudates and rhizosphere microbiome for crop production. *Environ Sci Pollut Res*. 2021;28:54497–510. <https://doi.org/10.1007/s11356-021-15838-7>.
80. Machado D, Maistrenko OM, Andrejev S, Kim Y, Bork P, Patil KR, et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nat Ecol Evol*. 2021;5:195–203. <https://doi.org/10.1038/s41559-020-01353-4>.
81. Erlandson S, Wei X, Savage J, Cavender-Bares J, Peay K. Soil abiotic variables are more important than Salicaceae phylogeny or habitat specialization in determining soil microbial community structure. *Mol Ecol*. 2018;27:2007–24. <https://doi.org/10.1111/mec.14576>.
82. Cui J, Yuan X, Zhang Q, Zhou J, Lin K, Xu J, et al. Nutrient availability is a dominant predictor of soil bacterial and fungal community composition after nitrogen addition in subtropical acidic forests. *PLoS One*. 2021;16:e0246263. <https://doi.org/10.1371/journal.pone.0246263>.
83. Musilova L, Ridl J, Polivkova M, Macek T, Uhlík O. Effects of secondary plant metabolites on microbial populations: changes in community structure and metabolic activity in contaminated environments. *Int J Mol Sci*. 2016;17:1205. <https://doi.org/10.3390/ijms17081205>.
84. Voges MJEE, Bai Y, Schulze-Lefert P, Sattely ES. Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. *Proc Natl Acad Sci U S A*. 2019;116:12558–65. <https://doi.org/10.1073/pnas.1820691116>.
85. Jacoby RP, Koprivova A, Kopriva S. Pinpointing secondary metabolites that shape the composition and function of the plant microbiome. *J Exp Bot*. 2021;72:57–69. <https://doi.org/10.1093/jxb/eraa424>.
86. Mas A, Jamshidi S, Lagadeuc Y, Eveillard D, Vandenkoornhuyse P. Beyond the black queen hypothesis. *ISME J*. 2016;10:2085–91. <https://doi.org/10.1038/ismej.2016.22>.

87. Borenstein E, Kupiec M, Feldman MW, Ruppin E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci U S A*. 2008;105:14482–7. <https://doi.org/10.1073/pnas.0806162105>.
88. Klitgord N, Segrè D. Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol*. 2010;6:e1001002. <https://doi.org/10.1371/journal.pcbi.1001002>.
89. Thommes M, Wang T, Zhao Q, Paschalidis IC, Segrè D. Designing metabolic division of labor in microbial communities. *mSystems*. 2019;4:e00263–18. <https://doi.org/10.1128/mSystems.00263-18>.
90. Pacheco AR, Moel M, Segrè D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat Commun*. 2019;10:103. <https://doi.org/10.1038/s41467-018-07946-9>.
91. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, et al. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*. 2015;349:860–4. <https://doi.org/10.1126/science.aaa8764>.
92. Sriswasdi S, Yang CC, Iwasaki W. Generalist species drive microbial dispersion and evolution. *Nat Commun*. 2017;8:1162. <https://doi.org/10.1038/s41467-017-01265-1>.
93. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. *ISME J*. 2014;8:1553–65. <https://doi.org/10.1038/ismej.2014.60>.
94. Gil R, Peretó J. Small genomes and the difficulty to define minimal translation and metabolic machineries. *Front Ecol Evol*. 2015;3:Article123. <https://doi.org/10.3389/fevo.2015.00123>.
95. Niu B, Paulson NJ, Zheng X, Kolter R. Simplified and representative bacterial community of maize roots. *Proc Natl Acad Sci U S A*. 2017;24–6. <https://doi.org/10.1073/pnas.1616148114>.
96. Bodenhausen N, Bortfeld-Miller M, Ackermann M, Vorholt JA. A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet*. 2014;10:e1004283. <https://doi.org/10.1371/journal.pgen.1004283>.
97. Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW, et al. High taxonomic variability despite stable functional structure across microbial communities. *Nat Ecol Evol*. 2016;1:0015. <https://doi.org/10.1038/s41559-016-0015>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

