Keystone species determine the productivity of synthetic microbial biofilm communities

Xinli Sun[1, 2], Jiyu Xie[1], Daoyue Zheng[1], Wei Wang[1], Riyan Xia[1], Ruifu Zhang[1], Ákos T. Kovács[2*], Zhihui Xu[1*], Qirong Shen[1]

1 Jiangsu Provincial Key Lab of Solid Organic Waste Utilization, Jiangsu Collaborative Innovation Center of Solid Organic Wastes, Educational Ministry Engineering Center of Resource-Saving Fertilizers, The Key Laboratory of Plant Immunity, Nanjing Agricultural University, 210095 Nanjing, Jiangsu, Peoples R China

2 Bacterial Interactions and Evolution Group, DTU Bioengineering, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Xinli Sun and Jiyu Xie contributed equally to this work.

*Corresponding authors: Zhihui Xu and Ákos T. Kovács

E-mail: xzh2068@njau.edu.cn, atkovacs@dtu.dk

Abstract

Microbes typically reside in multi-species communities, whose interactions have considerable impacts on the robustness and functionality of such communities. To manage microbial communities, it is essential to understand the factors driving their assemblage and maintenance. Even though the community composition could be easily assessed, interspecies interactions during community establishment remain poorly understood. Here, we combined co-occurrence network analysis with quantitative PCR to examine the importance of each species within synthetic communities (SynComs) of pellicle biofilms. Genome-scale metabolic models and *in vivo* experiments indicated that the biomass of SynComs was primarily affected by keystone species that are acting either as metabolic

24 facilitators or as competitors. Our study suggests that a combination of co-occurrence network
25 analysis and metabolic modeling could explain the importance of keystone species in SynComs.

26 Introduction

27 In natural microbiome systems, hundreds or thousands of species occupy and interact with each other,
28 as well as the environment around them. To deal with this complexity, we can utilize synthetic
29 communities (SynComs) to identify common principles governing and structuring microbiomes
30 (Bengtsson-Palme, 2020). Several SynComs with moderate complexity and high controllability have
31 been developed to represent different natural environments (Blasche et al., 2017; Chan et al., 2017;
32 Fu et al., 2020; Lebeis et al., 2015; Weiss et al., 2021). Such knowledge is essential to disentangle
33 complex community interactions and to provide powerful tools for biotechnological, medical, or
34 agricultural purposes (Abram, 2015; Cavaliere et al., 2017; Cho and Blaser, 2012; Fitzpatrick et al.,
35 2020; Gómez-Godínez et al., 2021). To harness this potential, it is essential to preserve both the
36 microenvironments in which these beneficial processes can take place, as well as the microbial
37 populations that contribute to these processes. Based on the study of SynComs, researchers have
38 found that the assemblage and robustness of communities can be affected by several factors
39 including pH (Ortiz et al., 2021), spatial distribution (Liu et al., 2019), initial abundance (Gao et al.,
40 2020, 2019), niche specificity (Estrela et al., 2021), nutrient availability (Ratzke et al., 2020), keystone
41 species (Niu et al., 2017) and so on. Nevertheless, many questions remain about how diversity is
42 maintained, and ultimately, how the community function can be preserved and controlled.

43 Multiple species are widely distributed in nature as surface-associated communities also referred to
44 as biofilms (Sadiq et al., 2021). The biofilm formation process consists of the planktonic stage, initial
45 attachment to a surface, the building of micro-aggregates, biofilm maturation, and biofilm dispersal
46 (Lee and Yoon, 2017). These multi-species biofilm communities assemble as a result of interspecies
47 interactions, which determine not only the composition and distribution but also their biological
48 function. Compared with solitary bacteria, species residing in multispecies biofilms can gain fitness
49 advantages, including enhanced biomass production (Ren et al., 2015), stress tolerance (Lee et al.,
50 2014; Raghupathi et al., 2018), and access to complex nutrients (Breugelmans et al., 2008; Nielsen

et al., 2000) through cooperative interactions. In contrast, ecological competition is more pervasive in multispecies biofilms (Foster and Bell, 2012; Oliveira et al., 2015). Bacteria actively inhibit competitors by producing toxins, biosurfactants, and antimicrobials (Rendueles and Ghigo, 2012), or indirectly impair the fitness of neighbors through nutrient scavenging. The spatial structuring property of biofilms supports the intermixing of cooperating species and the partitioning of niches by competing species. Understanding how these interactions affect the formation of multi-species biofilms is therefore important for their utilization or inhibition.

Despite a wealth of information on microbiota composition gathered from sequencing techniques, the knowledge of microbial interactions that occur in the microbiota is severely lacking. Building a co-occurrence network from high-throughput sequence data is a common method of deriving hypotheses from these massive data (Faust et al., 2012). For instance, it has been applied to link taxa to biological functions of interest (Wei et al., 2019), to identify potential biotic interactions (Durán et al., 2018), and to explore habitat differentiation (Barberán et al., 2012). Another emerging method for exploring microbial interactions is genome-scale metabolic modeling, which can provide insights into metabolic interaction potential and metabolic resource overlap in multi-species communities (Zelezniak et al., 2015; Zorrilla et al., 2021). Microbiome network analysis and metabolic modeling could guide the study of SynComs for animal and plant health (Heinken et al., 2021; Poudel et al., 2016). However, the ecological relevance of predicted interactions remains poorly understood (Faust, 2021). The connectedness and strength of positive or negative interactions are not experimentally verified. The main challenge of studying microbial interactions in SynComs is fast and accurate quantification of the abundance of each species. This can be achieved by 16S amplicon sequencing, colony-forming unit counting, fluorescent labeling, and quantitative PCR. Both of the first two approaches are simple to employ, but they can be time-consuming. While the first two approaches are more convenient but time-consuming, the latter two approaches are more efficient but require extensive initial time and effort.

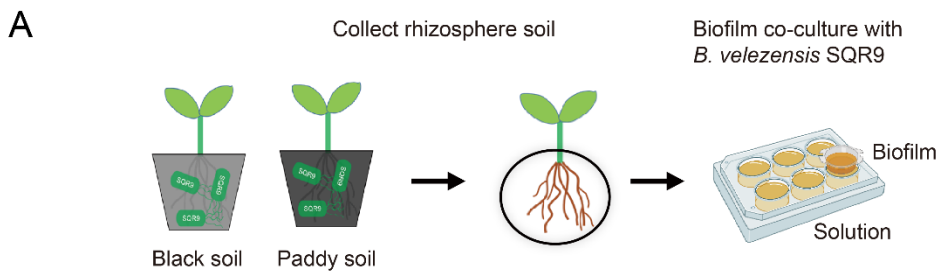
In this study, we constructed SynComs using isolates from a rhizosphere and predicted their interaction by analyzing co-occurrence networks. The network analysis was validated by quantitative

assays. The potential metabolic interactions were further investigated through experiments and metabolic modeling. The results suggested that keystone species determine community productivity through metabolic exchanges or resource competition. We propose that our study could inform the rational design of synthetic communities.

Results

C. rhizoplanae and *P. eucrina* were predicted to be keystone species in the eleven-species SynCom

In a previous study, we screened for cooperative bacteria for the plant-growth-promoting bacteria *Bacillus velezensis* SQR9, by studying the effect of strain SQR9 on rhizosphere microbiome composition (Sun et al., 2021). Parallel to those experiments, we collected and co-cultured the rhizosphere soil with strain SQR9 in a liquid medium to form pellicle biofilms (Figure 1—figure supplement 1A). Using an amplicon sequencing method, we determined the bacterial composition of the corresponding biofilm and the solution underneath. Consequently, 15 genera and 2 families were predicted to co-exist in the culture condition (Figure 1—figure supplement 1B). From our bacterial collection, we selected 11 isolates that corresponded with the predicted co-existing taxa and combined them into a synthetic biofilm community (Figure 1—figure supplement 2). These isolates are derived from three different phyla: Firmicutes, Proteobacteria, and Bacteroidetes.



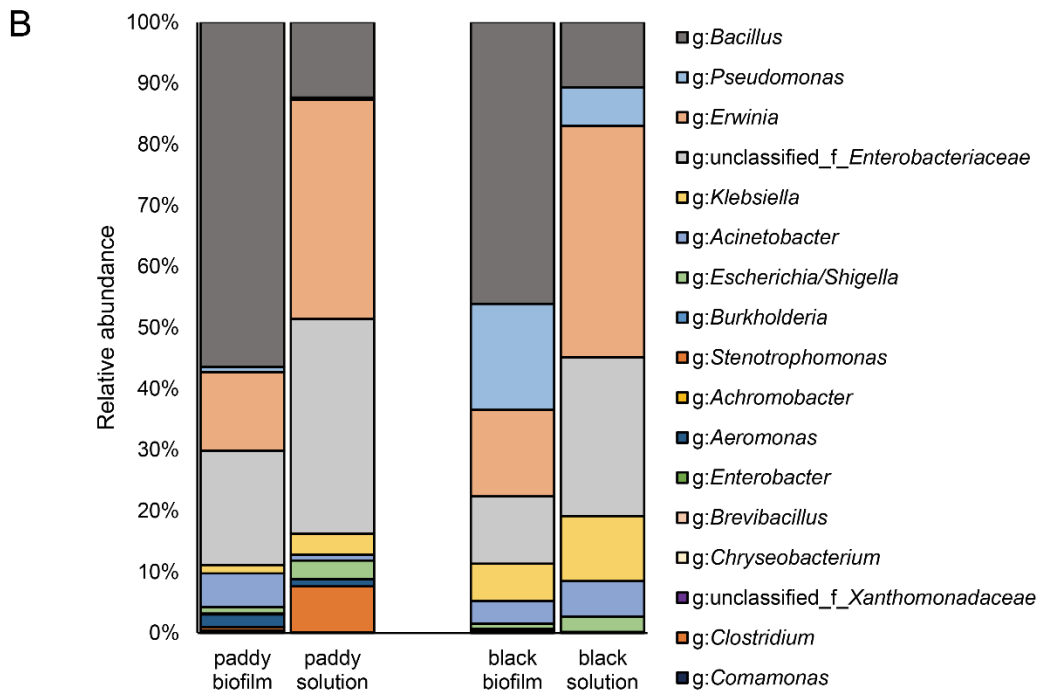


Figure 1—figure supplement 1. Origin of the initial 11 isolates. (**A**) Schematic diagram of the experimental setup. Rhizospheric black soil and paddy soil were collected from cucumber plants. The soil microbiota were co-cultivated with *B. velezensis* SQR9 at 30°C in TSB medium to form pellicle biofilms. After 24 hours of cultivation, the pellicle and the solution underneath were collected separately (n=3). The samples were sent for 16S rDNA amplicon sequencing. (**B**) Microbiome composition of the biofilm and solution. Based on OTU clustering and the RDP database taxonomic classification, 15 genera and 2 families were identified. Eleven matching isolates were selected from our laboratory bacteria collection (Sun et al., 2021).

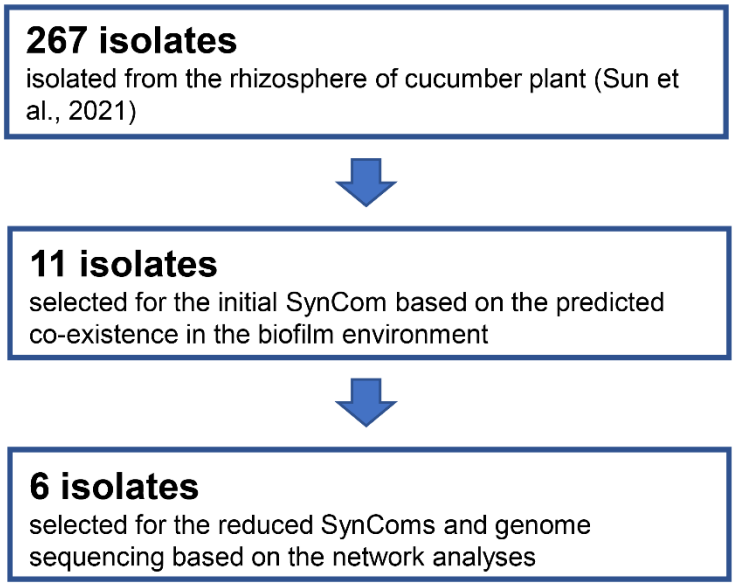
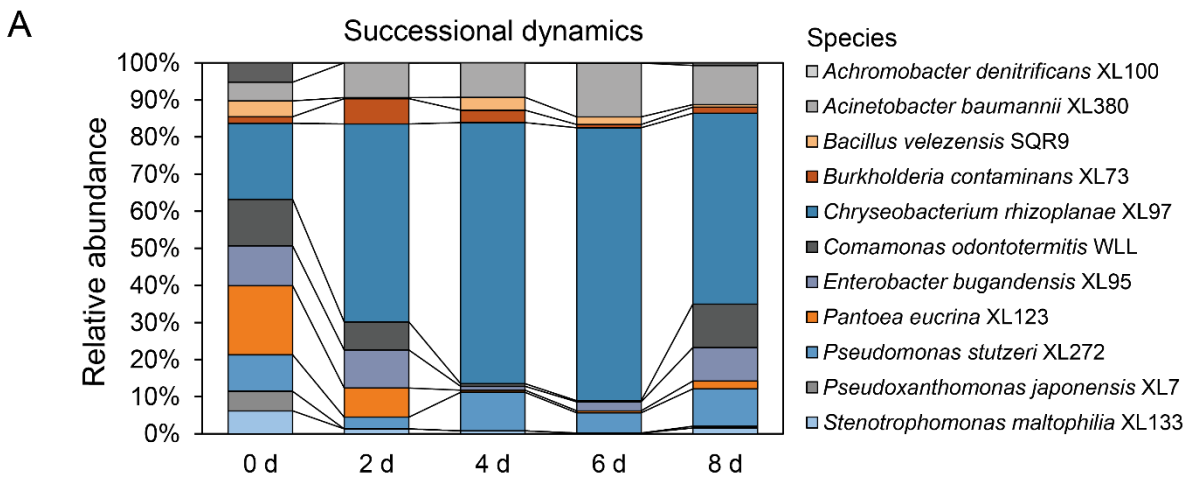


Figure 1—figure supplement 2. A flow chart of isolates used for the study, their origin and selection criteria.

101 We evaluated the stability of this SynCom by observing the compositional change of individual
102 members using 16S rRNA gene amplicon sequencing. Several time points were selected to represent
103 the development of a biofilm: initial aggregation (2d), establishment (4d), maturation (6d), and
104 dispersal (8d). Chr (*Chryseobacterium rhizoplanae*) was identified to be the predominant species
105 throughout the biofilm development (Figure 1A). Three species Ach (*Achromobacter denitrificans*),
106 Pxa (*Pseudoxanthomonas japonensis*), and Ste (*Stenotrophomonas maltophilia*) rapidly declined and
107 could not establish themselves in the community. Network co-occurrence analysis was employed to
108 examine the correlations among the species during succession. The correlations among the species
109 varied from less to more (Figure 1B). Pan (*Pantoea eucrina*) was consistently positively correlated
110 with other species, while Chr was consistently negatively correlated with other species. These two
111 species were predicted to act as hub taxa in this multi-species consortium. In the subsequent
112 experiments, we examined the interactions of six of the most abundant species, including Pan, Chr,
113 Com (*Comamonas odontotermitis*), Aci (*Acinetobacter baumannii*), Ent (*Enterobacter bugandensis*),
114 and Pse (*Pseudomonas stutzeri*). Other species were excluded either due to their low abundance at
115 the biofilm maturation stage, or due to the lack of correlation with the other species.

116



117

118

119

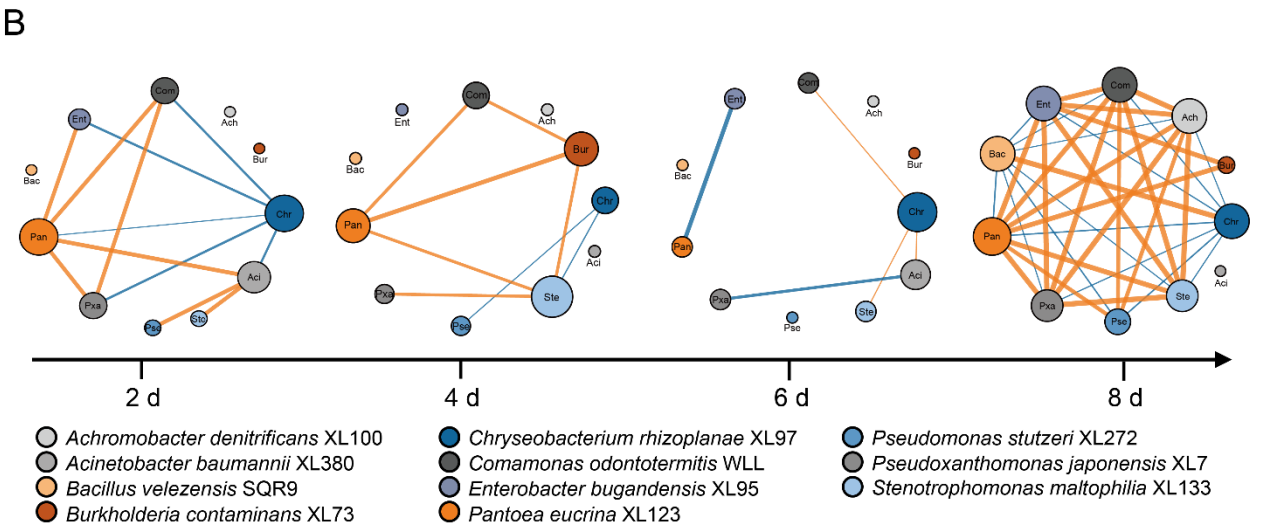


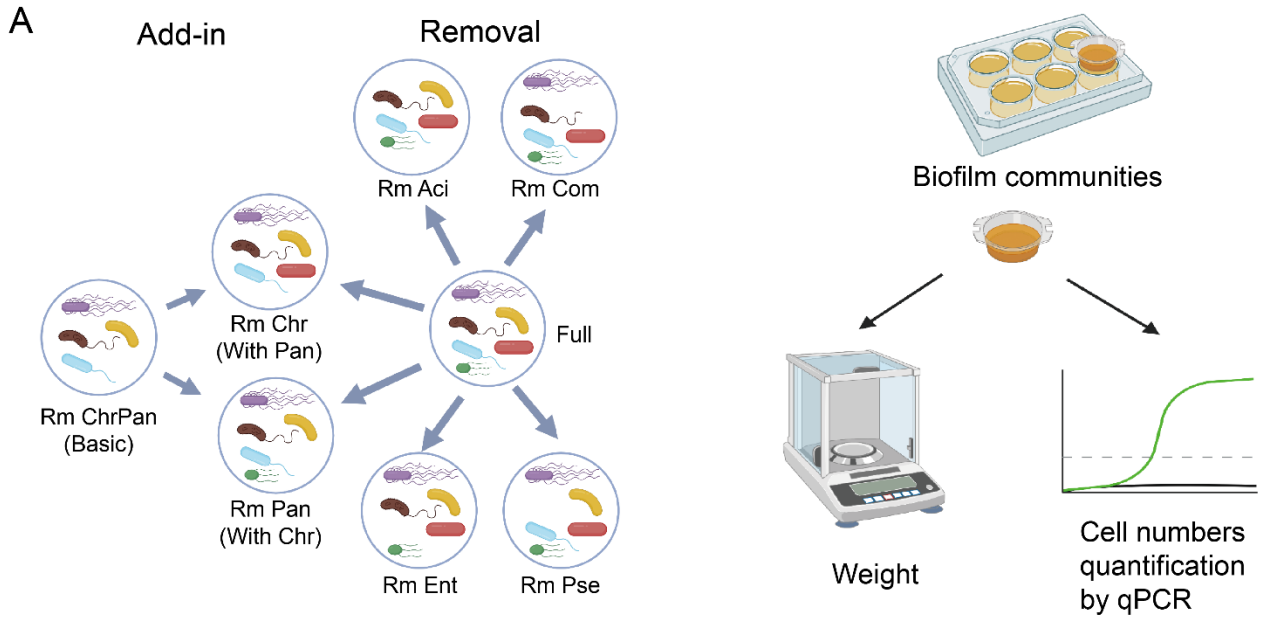
Figure 1. Dynamic changes of the initial 11-species consortium composition in the biofilm. (**A**) The relative abundance of isolates determined by 16S rRNA gene amplicon sequencing method (n=8). (**B**) Co-occurrence of bacterial clades within biofilm communities. Each node represents a bacterial species, node size is proportional to betweenness centrality. Line width indicates the interactive strength of interaction, and color indicates the sign of association (orange indicates positive, blue indicates negative).
Figure supplement 1. Origin of the initial 11 isolates.
Figure supplement 2. A flow chart of isolates used for the study, their origin and selection criteria.

Productivity of multi-species biofilm was affected by key members

To test the central role that Pan and Chr play in biofilm development, we evaluated the relative productivity of different combinations of the six strains. “Add-in” and “Removal” strategies were applied (Figure 2A). According to the “Add-in” strategy, the “Basic community” consisted of four members (Com, Aci, Ent, and Pse), upon which we added the predicted positive or negative member to the community and compared the total biomass. Based on our hypothesis, we anticipated that supplementation of the positive member would increase total biomass, while we anticipated the opposite for the negative member. On the other hand, the “Removal” strategy included all the six species as a full community, then one isolate was dropped out to obtain the five-species reduced communities. Biofilm productivity was assessed in two ways: fresh weight and total cell numbers (Figure 2A).

Compared to the "Full community", “Rm Chr” and “Rm Com” resulted in a higher biofilm weight, whereas “Rm Ent” and “Rm Pan” resulted in a lower biofilm weight (Figure 2B). This indicated that these four species are important to maintain the productivity of this synthetic community. However, the contribution of each species to the population productivity could not be determined using this approach. The previously used amplicon sequencing method was relatively slow and could only

137 provide information on the relative abundance. The main challenge in determining the composition is
138 to establish a fast and accurate method for quantifying the absolute abundance of each species. To
139 achieve this goal, we performed a genome-scale comparison and identified strain-specific genes in
140 each species. We designed multiple primers to amplify these genes and tested their specificity against
141 other species. Consequently, strain-specific primers were developed to selectively amplify only one
142 isolate from the six-members community (Table 1). Using these primers, we were able to determine
143 the contribution of each isolate to the community productivity. “Rm Chr” reached the highest number
144 of total cell numbers, while “Rm Pan” greatly decreased the total cell numbers (Figure 2C). Taken
145 together, we identified Pan and Chr as the keystone positive and negative species in terms of
146 community productivity, respectively.



147

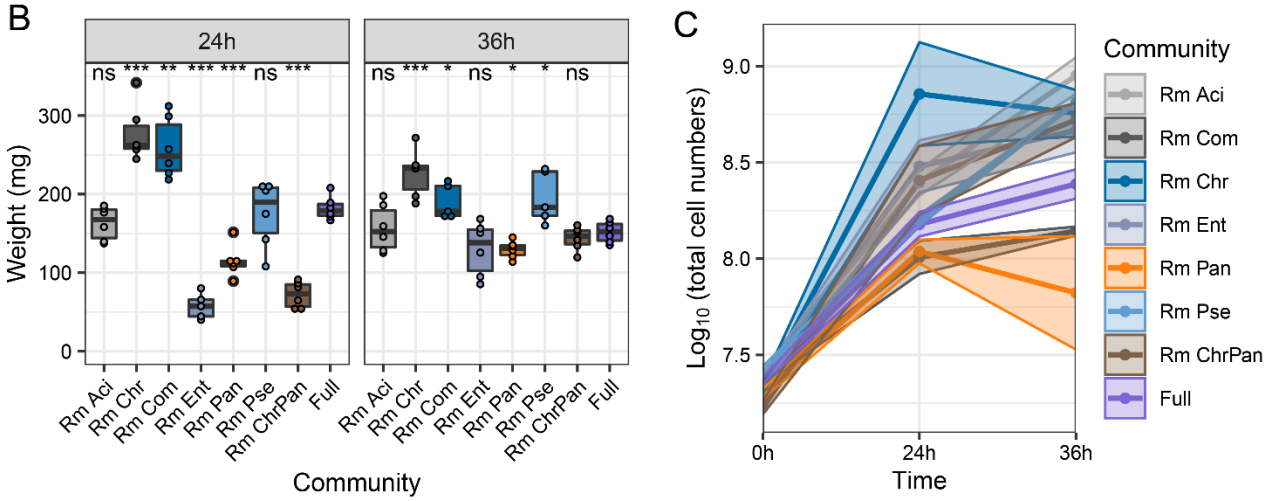


Fig 2 Productivity of the reduced SynComs. (**A**) Schematic diagram of experimental setup. Add in: positive or negative species were added in the four-species community (Rm ChrPan) to create five-species communities. Removal: each isolate was dropped out from the six-species community (Full) to create five-species communities. Biofilm biomass were quantified by fresh weight and cell numbers were quantified by qPCR. Diagrams were generated in Biorender. (**B**) Biofilm biomass quantified by fresh weight (n=5-6). The "Full community" was set as the control group. Asterisks indicate statistically significant according to unpaired student's *t* test via R (***$p<0.001$, **$p<0.01$, *$p<0.05$). (**C**) Total cell numbers in the communities (n=6). Color strip indicates standard deviation. All represents the full six-species communities, Rm Aci, Rm Chr, Rm Com, Rm Ent, Rm Pan, Rm Pse represent the five-species communities resulting from the removal of *Acinetobacter baumannii* XL380, *Chryseobacterium rhizoplanae* XL97, *Comamonas odontotermitis* WLL, *Enterobacter bugandensis* XL95, *Pantoea eucrina* XL123 and *Pseudomonas stutzeri* XL272, respectively. Rm ChrPan represents the four-species community resulting from the removal of *C. rhizoplanae* and *P. eucrina*.

C. rhizoplanae and *P. eucrina* were validated to be keystone species in the reduced community

The compositions of different communities were clustered based on the Bray-Curtis dissimilarity index (Figure 3A-C). After 24 hours, the communities clustered into two main groups: "Full community", "Basic community" (Rm ChrPan), "Rm Pan", and "Rm Chr" communities clustered together, whereas the other communities were grouped in one branch. The upper branch was dominated by Ent, the lower branch by Pan. Based on the hierarchical clustering, Pan or Chr played deterministic roles in species composition by influencing the relative abundance of Ent. Additionally, the random forest analysis revealed that Pan played a major role in determining the differences of community compositions (Figure 3D). After 36 hours, the communities were divided into two branches without any apparent disciplinarian. We found that the proportion of Pan cells was highest in the "Rm Ent" community, while the proportion of Ent cells was highest in the "Rm Pan" community (Figure 3C). Furthermore, random forest analysis revealed that Chr was the keystone species determining the compositional differences (Fig 3E).

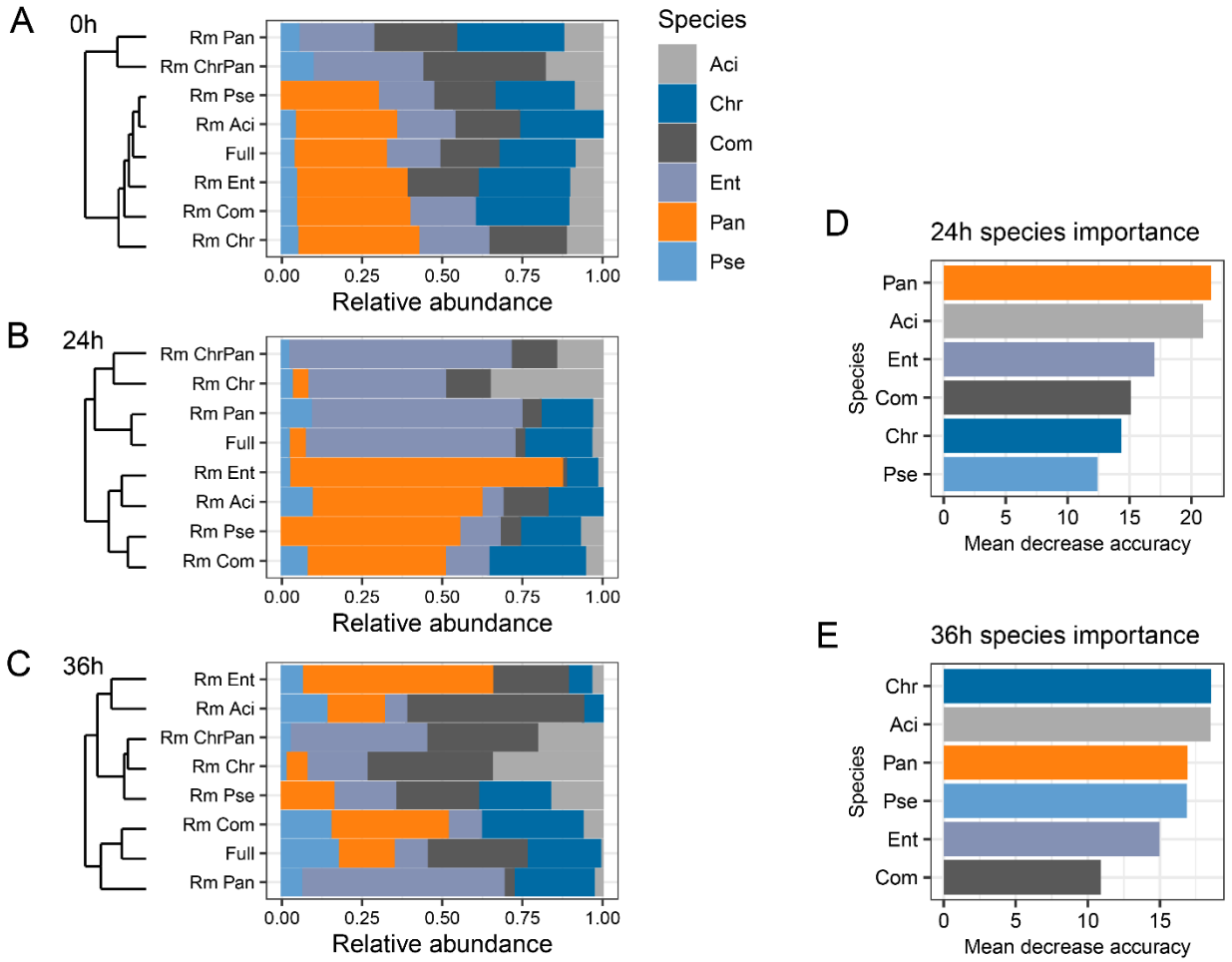


FIG 3 Species composition and importance. (**A-C**) Hierarchical clustering analysis of the synthetic community composition (n=6). Samples include a six-species communities (Full), a four-species community (Rm ChrPan) and six five-species communities (Rm communities). Rm Aci, Rm Chr, Rm Com, Rm Ent, Rm Pan, Rm Pse represent the five-species communities resulting from the removal of *Acinetobacter baumannii* XL380, *Chryseobacterium rhizoplanae* XL97, *Comamonas odontotermitis* WLL, *Enterobacter ludwigii* XL95, *Pantoea eucrina* XL123 and *Pseudomonas stutzeri* XL272, respectively. The samples were collected at the initial inoculation (A), 24h (B) and 36h (C). (**D-E**) Species importance in the SynComs as predicted by the random forest analysis.

163

164 ## Metabolic facilitation explained the positive interaction

165 To gain insights into the metabolic interaction potential of the SynCom members, we reconstructed
166 genome-scale metabolic models, using the M9 glucose minimal medium as the input medium. We
167 derived the likely exchanged metabolites across communities and the strength of metabolic coupling
168 - SMETANA score (Zelezniak et al., 2015). The “Full community” and the “Add Pan” community
169 displayed higher cross-feeding potential than the “Add Chr community” and the “Basic community”
170 (Figure 4A-D). Pan was the principal donor in the communities with Pan, whereas Ent was the
171 principal donor in the communities without Pan. This result explained the dominance of these two
172 species in the SynComs. In contrast, adding Chr to the SynComs had little effect on the global

173 metabolic exchanges. Chr acted more as a receiver rather than a donor in the “Full community”
174 (Figure 4A). In addition, the add-in or removal of keystone species also affected the proportion of
175 exchanged compounds (Figure 4E-H). Amino acids and phosphates were the major categories of
176 metabolites exchanged across communities. When the positive member Pan was removed, we
177 observed higher reliance on amino acid cross-feeding and lower dependence on organoheterocyclic
178 compounds exchange. Moreover, we highlighted the interactions involving Pan as the donor in the
179 “Full community” (Figure 4I) and in the “Add Pan” (Rm Chr) community (Figure 4J). We observed that
180 the metabolic contribution of Pan benefited all community members.

181 Growth assays were used to determine the accuracy of the metabolic models. Firstly, all species
182 except Chr were able to grow individually in the M9 glucose minimal medium (data not shown), thus
183 five of them have the potential to serve as donors. Secondly, growth in the spent culture medium was
184 utilized to determine metabolic interaction potential. Each isolate was cultured in the M9 glucose
185 medium until the glucose was undetectable. We obtained sterile spent medium and used it to cultivate
186 every other member of the community and themselves. The maximum growing capacity was
187 determined and defined as growth facilitation (Figure 4K). Results showed that the spent medium
188 from Pan and Ent could support the growth of all other isolates. Pse's metabolic by-products were
189 also capable of supporting the growth of three isolates. Aci, Com, and Chr acted only as receivers but
190 not donors. These results fully supported the metabolic models, suggesting Pan and Ent contributed
191 to the overall productivity by cross-feeding, whereas Chr acted as a cheater.

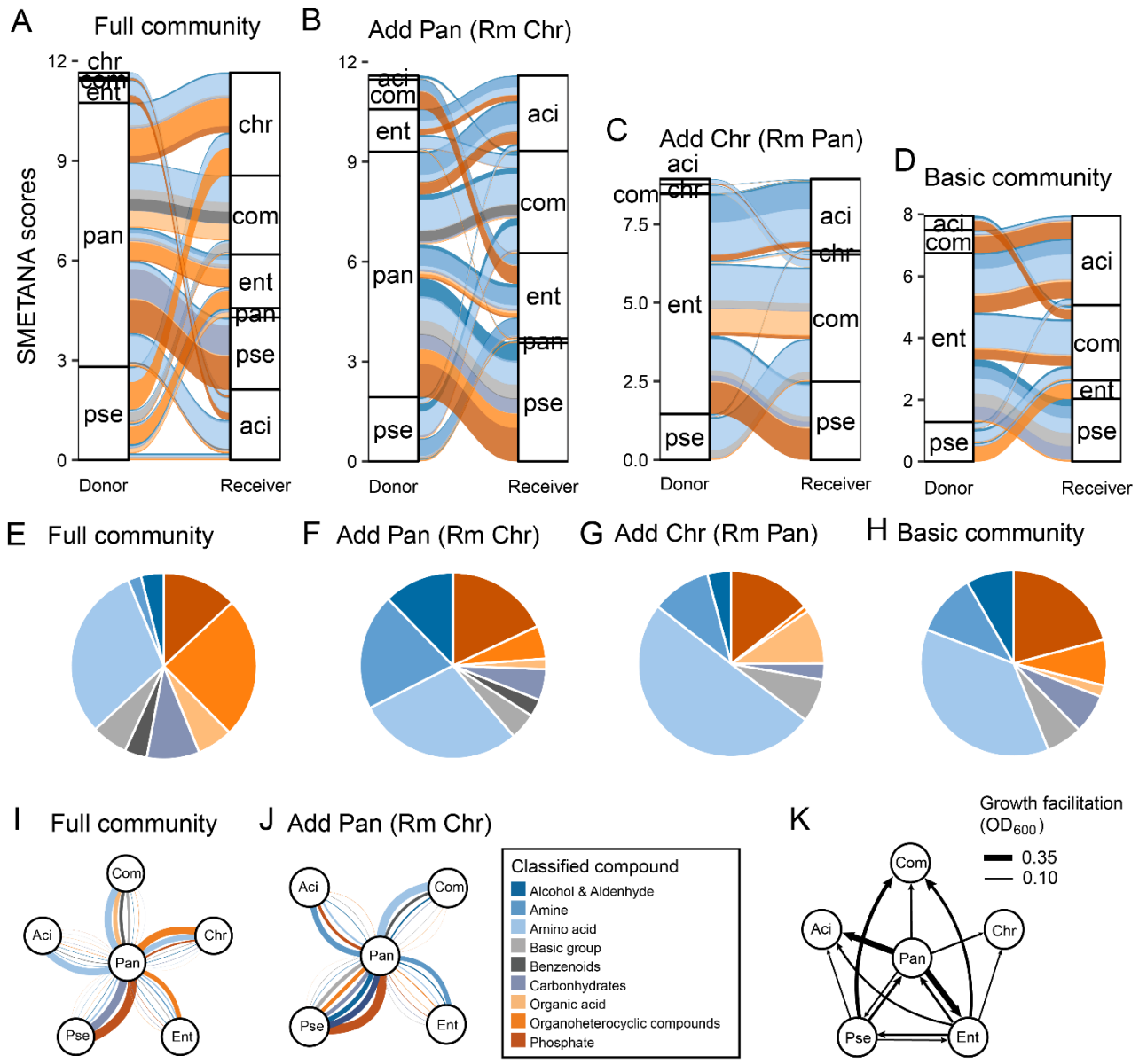
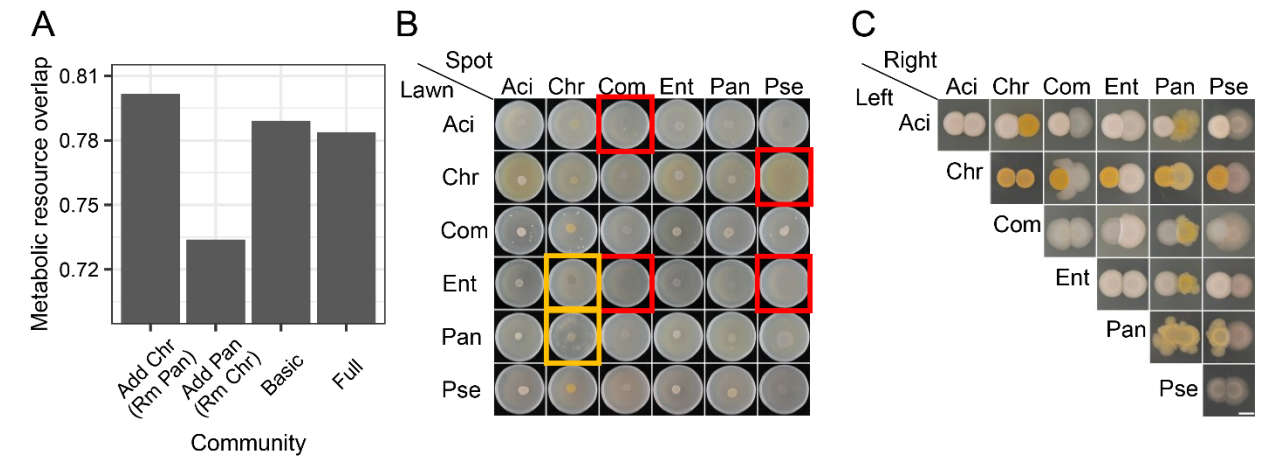


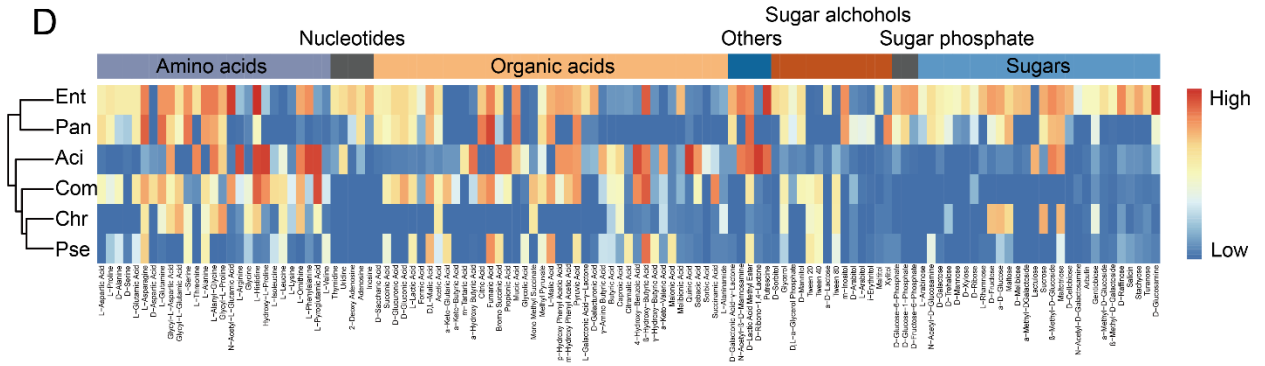
Fig 4 Metabolic facilitation. (**A-J**) Metabolic interaction potential simulated by genome-scale metabolic modelling. (**A-D**) Alluvial diagrams showing compounds exchanged within communities. Thickness of lines are SMETANA scores. (**E-H**) Pie charts showing the proportion of compounds exchanged within communities. (**I-J**) Flower plots showing interactions involving positive species *P. eucrina* XL123 (centered in each panel) as a donor across communities. Thickness of lines are proportional to magnitude of SMETANA score. Colors of all plots are the classification of compounds. (**K**) Growth facilitation assessed by growth in spent medium. The spent medium was prepared by growing the corresponding species in M9 medium with 1% glucose till the glucose was under detection then filter sterilized. The species were grown in the filtered spent medium for 24 h and the growth were measured as the OD_{600}. Line width indicates the OD_{600}.

Resource competition explained the negative interaction

Higher-order interactions may further be explained by metabolic competition. Using metabolic modeling, the metabolic resource overlap was simulated (Figure 5A). We observed high metabolic resource overlap across all communities, indicating intense resource competition among SynCom members. The addition of the negative member Chr resulted in the highest metabolic resource overlap while adding the positive member Pan resulted in a reduction of resource competition. To test the

199 possibility of direct competition, we performed spot-on-lawn and pair-wise spot assays on TSB agar
200 plates (Figure 5B-C). Both Aci and Ent completely inhibited the growth of Com. Similarly, Chr and Ent
201 inhibited Pse growth. However, these inhibitions were not caused by direct antagonism as no
202 inhibition zones were observed in the respective spot-on-lawn assay plates (Figure 5B, red square).
203 In light of these results, no clear boundary was observed between two strains when they were spotted
204 adjacent to one another (Figure 5C). Interestingly, Pan and Ent could change the colony color of Chr
205 from light yellow to white (Figure 5B, yellow square). We also assessed the carbon source metabolic
206 ability by high-throughput phenotypic microarrays (Figure 5D). Ent and Pan were generalists that can
207 utilize a wide variety of carbon sources. Other community members were specialized in using amino
208 acids and organic acids as carbon sources while having a limited ability to utilize nucleotides, sugar
209 alcohols, and sugars. Collectively, resource competition was widespread among the SynCom
210 members.





211

Fig 5 Resource competition. (**A**) Bar chart showing the metabolic resource overlap simulated by the metabolic modeling. (**B**) Spot-on-lawn assays. The lawn species were spread on TSB agar plates at an OD_{600} of 0.02 and dried, 5 μl of the spot species were spotted on the center at an OD_{600} of 0.4. The plate diameter is 6 cm. Photos were taken after 48 h incubation at 30°C. (**C**) Pair-wise spot assays. 5 μl of the two species were plot next to each other at an OD_{600} of 1. Scale bar represents 3 mm. Photos were taken after 48 h incubation at 30°C. (**D**) Carbon source metabolic ability measured by phenotype microarrays.

Discussion

A key concern in SynCom research is to understand how microbial interactions affect community composition and productivity. In our study, we evaluated the contribution of each individual to community productivity and revealed the metabolic interactions. Our SynCom system identified Pan (*P. eucrina*) and *Chr* (*C. rhizoplanae*) as important drivers of community interaction networks through metabolic cross-feeding and resource competition.

Manipulating microbial consortia has a variety of applications; however, attempts to engineer microbial consortia often fail to achieve the expected results owing to the unexpected effects of interactions within the community. Although the practical application of this SynCom is not yet clear, accurate prediction of the species composition is the first step toward manipulating microbial consortia (Widder et al., 2016). In this study, we assessed the SynCom composition using the amplicon sequencing method and qPCR. In the initial eleven-species biofilm community, Chr was the dominant species across all the timepoints. In the reduced six-species biofilm community, however, Ent and Pan became the most abundant species. These findings indicated that other species, although less abundant or not associated with community members, also affect the final composition of the SynCom. This result was consistent with a previous study that showed rare species were indispensable to the synergistic effects in multi-species biofilms (Ren et al., 2015). Furthermore, higher-order interactions only occur in the presence of additional species, emphasizing the importance of studying microbial interactions in the context of communities (Friedman et al., 2017).

Resource competition plays a major role in shaping bacterial communities, while metabolic exchanges promote group survival (Goldford et al., 2018; Zelezniak et al., 2015). In the present study, we simulated the metabolic interactions using metabolic models and assessed their predictability using phenotypic assays. Specifically, communities with negative species exhibited higher resource

236 competition, whereas communities with positive species displayed greater metabolite exchange.
237 These results indicate that metabolic interactions play a key role in determining the composition of
238 communities. Other mechanisms, such as niche partitioning created by spatial and temporal
239 heterogeneity, and trade-offs between nutrient acquisition and environmental tolerance, can also
240 promote coexistence between species (Louca et al., 2018). Our research provides insights into how
241 metabolic competition and cooperation simultaneously shape the community composition. The
242 methodology of network co-occurrence analysis combined with qPCR quantification, metabolic
243 modeling, and pair-wise interaction can be applied to various SynComs studies. Ultimately, such
244 studies should translate to a deeper understanding of how microbial communities behave in their
245 native environments, and this knowledge may be applied to wastewater treatment, disease
246 suppression, and crop yield enhancement.

247 Materials and methods

248 Strains and genome sequencing

249 As shown in Table 1, eleven bacterial isolates derived from the cucumber rhizosphere (Sun et al.,
250 2021) were subjected to this study. These isolates were selected based on their co-existence in
251 biofilms (see Results).

Table 1. Strains used in this study

Strain	Abbreviation
Achromobacter denitrificans XL100	Ach
Acinetobacter baumannii XL380	Aci
Bacillus velenzensis SQR9	Bac
Burkholderia contaminans XL73	Bur
Chryseobacterium rhizoplanae XL97	Chr
Comamonas odontotermitis WLL	Com
Enterobacter bugandensis XL95	Ent
Pantoea eucrina XL123	Pan

Pseudomonas stutzeri XL272	Pse
Pseudoxanthomonas japonensis XL7	Pxa
Stenotrophomonas maltophilia XL133	Ste

252

253 The whole genomes of the selected isolates were sequenced by different companies at different times.
254 The genomes of Aci, Com, Pxa and Ste were sequenced using a combination of PacBio RS II and
255 Illumina HiSeq 4000 sequencing platforms. The genome of Aci, Pxa and Ste were sequenced at the
256 Beijing Genomics Institute (BGI, Shenzhen, China). The Illumina data were used to evaluate the
257 complexity of the genome. Raw sequencing data and the assembled genome have been deposited
258 to the National Center for Biotechnology Information (NCBI) under the BioProject accession number
259 PRJNA593376, PRJNA762936 and PRJNA762715. The genome of Com was sequenced at Majorbio
260 Bio-Pharm Technology Co., Ltd. Raw sequencing data and the assembled genome have been
261 deposited to the NCBI under the BioProject accession number PRJNA762695.

262 The genomes of Bur, Chr, Ent, and Pan were sequenced using PacBio Sequel platform and Illumina
263 NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. Raw sequencing data
264 and the assembled genomes have been deposited to the NCBI under the BioProject accession
265 number PRJNA593683, PRJNA721858, PRJNA761942, and PRJNA762676. Pse was sequenced by .
266 Genomes were automatically annotated by NCBI PGAP.

267 Co-existence prediction

268 To create a multi-species biofilm community, we adopted co-existence criteria. In parallel with our
269 previous study (Sun et al., 2021), rhizosphere soil of cucumber soil was collected. Two types of soil
270 were chosen: black soil and paddy soil. Bacterial abundance in the rhizosphere soil was roughly
271 estimated by CFU counting on TSB agar plates. Equal volumes of soil suspension were mixed with
272 *B. velezensis* SQR9 at an optical density at 600 nm (OD_{600}) of 1 and inoculated in 10 ml of TSB liquid
273 medium at a 1:100 ratio. Biofilms formed at the air-liquid interface and the solution underneath were
274 collected separately after 24 h of incubation at 30°C. Each treatment had three biological replicates.

275 Genomic DNA of the samples was extracted using an E.Z.N.A. Bacterial DNA Kit (Omega Bio-tek,
276 Inc.) following the manufacturer's instructions. Universal primers targeting the V3-V4 regions of the
277 16S rRNA gene were used to construct the DNA library for sequencing. Paired-end sequencing of
278 bacterial amplicons was performed on the Illumina MiSeq instrument (300 bp paired-end reads). Raw
279 sequencing data have been deposited to the NCBI SRA database under BioProject accession number
280 PRJNA739098. Reads were processed using the USEARCH pipeline. The paired-end reads were
281 merged using the "fastq_mergepairs" command. High-quality sequences were then selected using
282 the "fastq_filter" command and dereplicated using the "derep_fulllength" command. The singletons
283 were removed using "unoise3" algorithm and chimeric sequences were removed using "uchime_ref"
284 command. The remaining sequences were clustered to OTUs based on 97% sequence similarity. The
285 OTUs were classified at the genus and family level using the RDP database. The bacterial
286 composition in the biofilm was visualized in Microsoft Office Excel 2019 to predict co-existence.

287 ## Initial 11-species SynCom growth condition

288 The start inoculum was prepared by mixing equal volumes of species at an OD_{600} of 1. The eleven-
289 species biofilm was cultivated by mixing 4ml of the mixed-species community (1%) with 400 ml of
290 tryptic soy broth (TSB) and incubating at 30°C for two to eight days.

291 ## Initial 11-species SynCom composition analysis

292 Biofilms formed at the air-liquid interface were collected on days 2, 4, 6, and 8. Each time point had
293 eight biological replicates. Genomic DNA of the biofilm samples was extracted as previously
294 described. 16S rRNA gene amplicon sequencing was conducted and analyzed as described before.
295 The difference in the analysis is that the generated sequences were not clustered and directly used
296 to create ASV table. Taxonomy of the ASVs was assigned to the species with a reference database
297 consisting of the full 16S rRNA gene sequences of the eleven species. The compositional changes
298 were visualized in Microsoft Office Excel 2019. Microbial co-occurrence networks were constructed
299 to show the interactions among species during biofilm development. Spearman correlations among
300 all taxa were calculated using the R *psych* package. Only edges with correlation scores > 0.6 were

301 kept ($p < 0.05$, FDR-adjusted). Correlation networks were visualized via Gephi using the Fruchterman
302 Reingold layout (Bastian et al., 2009).

303 Reduced SynComs growth condition

304 The four-, five- and six-species biofilms were grown in 6-well microtiter plates (VWR) insert with 100
305 μm sterile nylon mesh cell strainers (Biologix Cat #15-1100). 10 ml of TSB liquid medium and 100 μl
306 of start inoculum were added. The plates were incubated for 24 h at 30 °C to allow the biofilm to grow
307 on top of the nylon mesh cell strainer.

308 Reduced SynComs biomass quantification

309 Biomass was defined by the fresh weight of biofilm. To measure the biomass, the cell strainer was
310 taken out, removed visible drops with paper, and weighed. The fresh weight was total weight minus
311 the weight of the nylon mesh. Each treatment had six biological replicates.

312 Reduced SynComs cell numbers quantification by qPCR

313 To quantify the cell numbers of each isolate within the multi-species biofilm, strain-specific primers
314 were designed for the selected six isolates based on the whole-genome comparison. Genome
315 comparison was performed using Roary (Page et al., 2015). The primers target the single-copy gene
316 of the corresponding species. Conventional PCR and qPCR melt curves were used to evaluate the
317 specificity of the primers obtained. Primer sequences and target genes were listed in Table 2.
318 Standard curves were generated using plasmids containing corresponding fragments. qPCR was
319 performed with Applied Biosystems Real-Time PCR Instrument. Reaction components are as follow
320 7.2 μl H_2O, 10 μl 2× ChamQ SYBR qPCR Master Mix (Vazyme), 0.4 μl 10 μM of each primer and 2
321 μl template DNA. The PCR programs were carried out under the following conditions: 95 °C for 10
322 min, 40 cycles of 95 °C for 30 s, 60 °C for 45 s, followed by a standard melting curve segment. Each
323 treatment had six biological replicates, and each sample was run in triplicates (technical replicates).

Table 2. Primers used in this study

Primer	Sequence	Product size (bp)	Target gene locus_tag (Genbank)
Aci_F	ATTTAGTATCTGGTGAAGTCATCCGTA	92	GOD87_RS00290
Aci_R	CCGACAAATAAAGCTTGAGTAACTCC		
Com_F	CTCAAAACCAGTGTGATCGTGGAA	109	LAD35_RS00960
Com_R	TATTGCCCATCAGCAGAGTGTAGC		
Chr_F	GAACATCAGTTATCTTGTGAGCGGTA	94	KB553_RS00055
Chr_R	CATACAGGCTCCCATTCCTATTGTG		
Ent_F	AGCGTTACAGCAGCTACAGGATATTCACC	96	K9O83_RS00400
Ent_R	CTTTTCACCATCACCCCATCCCTCGGTA		
Pan_F	TTAACATCGAAAAGCCTTCCCACCGTA	101	LAC65_RS01220
Pan_R	ATTCATCAGAAGCGCATGTATTACACT		
Pse_F	GAAATTCATCTTCGAACACAGCACAC	124	GOM96_RS02845
Pse_R	CTAGCTAACGGGGTTAAGTGCTTC		

Pair-wise assay

The direct competition of these isolates against each other was evaluated using the spot-on-lawn assay and the pair-wise spot assay. Spot-on-lawn assay: 5 ml of lawn species (OD_{600} ~ 0.02) grown in TSB medium was spread onto a 25 ml TSB plate (1.5% agar) and removed by pipetting. Plates were dried for 20 min. 5 µl of spot species (OD_{600}~ 0.4) grown in TSB medium was spot on the center of the plates. Pair-wise spot assay: 5 µl of the dual-species (OD_{600}~ 1) grown in TSB medium were spot on TSB plate (1.5% agar) at 5 mm between the center of each colony. Plates were grown at 30 °C and imaged at 48 h. The experiments were performed twice.

Growth promotion assay

334 The potential growth promotion of the bacterial metabolites to another species was evaluated using
335 a spent medium growth curve assay(Sun et al., 2021). Donor bacteria were grown in the M9 medium
336 with 1% glucose till the glucose was under detection. The cell culture was spun down, then the spent
337 medium was filter-sterilized and directly used as the medium of growth curve assay. 2 μl of recipient
338 species (OD_{600}~ 1) was inoculated to 200 μl spent medium or M9 glucose medium in a 10×10 well
339 Honeycomb Microplate. OD_{600} was measured every 30 minutes at 30 °C with Bioscreen C Automated
340 Microbiology Growth Curve Analysis System. Each treatment has 3 replicates. The experiments were
341 performed twice. The carrying capacity (maximum population size) was compared.

342 Reduced SynComs genome-scale metabolic modeling

343 Metabolic models were reconstructed using the CarveMe pipeline (Machado et al., 2018). The quality
344 of the metabolic models was validated using MEMOTE (Lieven et al., 2020). The metabolic interaction
345 potential and metabolic resource overlap for each community were analyzed using SMETANA
346 (Zelezniak et al., 2015; Zorrilla et al., 2021). The simulated cross-feeding results were summarized
347 as SMETANA score, which estimates the strength of metabolic exchanging (Zelezniak et al., 2015).

348 Carbon source metabolic activity measurement

349 We used PM1 (BIOLOG Cat #13101) and PM2 (BIOLOG Cat #13102) phenotypic microarrays to
350 assess the carbon source utilization ability of the community members (Bochner et al., 2001). The
351 assays were performed following the manufacturer's instructions. Briefly, 100 μl of diluted cell
352 suspension of each species mixed with the BiOLOG redox dyes were added to each well of the PM
353 plates, and the plates inoculated with water and dyes were used as negative controls. All of the plates
354 were then incubated at 30 °C for up to 48 h. If the species could utilize the carbon source in a well,
355 the colorless tetrazolium dye will be reduced to purple formazan by cell respiration. The color changes
356 were measured by an endpoint absorbance at 590 nm with a microplate reader. The variable level of
357 color changes indicates the carbon source metabolic activity.

358 Data analysis and figures

All the data needed to evaluate the conclusions in this paper are provided in the figures. Source data related to this paper would be available online upon publication. Data were analyzed using R studio. Statistical analysis methods were described in figure legends. A random forest approach was used to identify the contribution of the six species responsible for the differences in community composition. The analysis was performed using the R *randomForest* package (Andy Liaw et al., 2018). Plots were generated using Microsoft Office Excel 2019 (stacked bar plot), Graphpad prism 8 (heatmap), R *ggplot2*, *ggpubr, ggalluvial, pheatmap* packages, and Adobe Illustrator CC 2020 (Adobe Inc.). Schematic diagrams were generated using BioRender.

Acknowledgments

This work was financially supported by the National Nature Science Foundation of China (31972512 and 42090064), the Innovative Research Team Development Plan of the Ministry of Education of China (Grant No. IRT_17R56), the Fundamental Research Funds for the Central Universities (KYXK202009). XS was supported by a Chinese Scholarship Council fellowship. ÁTK was supported by the Danish National Research Foundation (DNRF137) for the Center for Microbial Secondary Metabolites and the Novo Nordisk Foundation via the INTERACT project (grant number NNF19SA0059360). During her unexpected trap in Denmark caused by the pandemic of Omicron, author XS is extremely grateful to ZX for his financial support, and to Danish guitarist Niklas Johansen for his weekly lessons and constant spiritual support.

Competing Interests

The authors declare that there are no competing financial interests related to the work described.

Author contributions

XS, JX, ZX, ÁTK designed the study, XS, JX, RX performed the experiments. DZ and WW performed the metabolic modeling. XS and JX analyzed the data and created the figures. XS wrote the first draft of the manuscript, ZX, ÁTK, QS revised the manuscript.

383 ## References

384 Abram F. 2015. Systems-based approaches to unravel multi-species microbial community
385 functioning. *Computational and Structural Biotechnology Journal* **13**:24–32.
386 doi:10.1016/J.CSBJ.2014.11.009

387 Barberán A, Bates ST, Casamayor EO, Fierer N. 2012. Using network analysis to explore co-
388 occurrence patterns in soil microbial communities. *ISME Journal* **6**:343–351.
389 doi:10.1038/ismej.2011.119

390 Bengtsson-Palme J. 2020. Microbial model communities: to understand complexity, harness the
391 power of simplicity. *Computational and Structural Biotechnology Journal* **18**:3987–4001.
392 doi:10.1016/j.csbj.2020.11.043

393 Blasche S, Kim Y, Oliveira AP, Patil KR. 2017. Model microbial communities for ecosystems biology.
394 *Current Opinion in Systems Biology* **6**:51–57. doi:10.1016/J.COISB.2017.09.002

395 Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput
396 phenotypic testing and assay of gene function. *Genome Research* **11**:1246–1255.
397 doi:10.1101/gr.186501

398 Breugelmans P, Barken KB, Tolker-Nielsen T, Hofkens J, Dejonghe W, Springael D. 2008.
399 Architecture and spatial organization in a triple-species bacterial biofilm synergistically
400 degrading the phenylurea herbicide linuron. *FEMS Microbiology Ecology* **64**:271–282.
401 doi:10.1111/j.1574-6941.2008.00470.x

402 Cavaliere M, Feng S, Soyer OS, Jiménez JI. 2017. Cooperation in microbial communities and their
403 biotechnological applications. *Environmental Microbiology* **19**:2949–2963. doi:10.1111/1462-
404 2920.13767

405 Chan SHJ, Simons MN, Maranas CD. 2017. SteadyCom: Predicting microbial abundances while
406 ensuring community stability. *PLOS Computational Biology* **13**:e1005539.
407 doi:10.1371/JOURNAL.PCBI.1005539

408 Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. *Nature*
409 *Reviews Genetics 2012 13:4* **13**:260–270. doi:10.1038/nrg3182

410 Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, Hacquard S. 2018.
411 Microbial interkingdom interactions in roots promote <i>Arabidopsis<i> survival. *Cell* **175**:973-
412 983.e14. doi:10.1016/j.cell.2018.10.020

413 Estrela S, Sanchez-Gorostiaga A, Vila JCC, Sanchez A. 2021. Nutrient dominance governs the
414 assembly of microbial communities in mixed nutrient environments. *eLife* **10**.
415 doi:10.7554/ELIFE.65948

416 Faust K. 2021. Open challenges for microbial network construction and analysis. *The ISME Journal*
417 1–8. doi:10.1038/s41396-021-01027-4

418 Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. 2012. Microbial
419 co-occurrence relationships in the human microbiome. *PLoS Computational Biology* **8**.
420 doi:10.1371/journal.pcbi.1002606

421 Fitzpatrick CR, Salas-González I, Conway JM, Finkel OM, Gilbert S, Russ D, Teixeira PJPL, Dangl JL.
422 2020. The Plant Microbiome: From Ecology to Reductionism and beyond. *Annual Review of*
423 *Microbiology* **74**:81–100. doi:10.1146/annurev-micro-022620-014327

424 Foster KR, Bell T. 2012. Competition, not cooperation, dominates interactions among culturable
425 microbial species. *Current Biology* **22**:1845–1850. doi:10.1016/j.cub.2012.08.005

426 Friedman J, Higgins LM, Gore J. 2017. Community structure follows simple assembly rules in
427 microbial microcosms. *Nature Ecology and Evolution* **1**:1–7. doi:10.1038/s41559-017-0109

428 Fu H, Uchimiya M, Gore J, Moran MA. 2020. Ecological drivers of bacterial community assembly in
429 synthetic phycospheres. *Proceedings of the National Academy of Sciences of the United States*
430 *of America* **117**:3656–3662. doi:10.1073/pnas.1917265117

431 Gao C-H, Cao H, Cai P, Sørensen SJ. 2020. The initial inoculation ratio regulates bacterial coculture
432 interactions and metabolic capacity. *The ISME Journal* 1–12. doi:10.1038/s41396-020-00751-7

433 Gao CH, Zhang M, Wu Y, Huang Q, Cai P. 2019. Divergent Influence to a Pathogen Invader by
434 Resident Bacteria with Different Social Interactions. *Microbial Ecology* **77**:76–86.
435 doi:10.1007/s00248-018-1207-z

436 Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P, Sanchez
437 A. 2018. Emergent simplicity in microbial community assembly. *Science* **361**:469–474.
438 doi:10.1126/science.aat1168

439 Gómez-Godínez LJ, Martínez-Romero E, Trejo JB, Arteaga-Garibay RI. 2021. Tools and challenges to
440 exploit microbial communities in agriculture. *Current Research in Microbial Sciences* **2**:100062.
441 doi:10.1016/j.crmicr.2021.100062

442 Heinken A, Basile A, Hertel J, Thinnes C, Thiele I. 2021. Genome-scale metabolic modeling of the
443 human microbiome in the era of personalized medicine. *https://doi.org/101146/annurev-*
444 *micro-060221-012134* **75**:199–222. doi:10.1146/ANNUREV-MICRO-060221-012134

445 Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, del Rio TG,
446 Jones CD, Tringe SG, Dangl JL. 2015. Salicylic acid modulates colonization of the root
447 microbiome by specific bacterial taxa. *Science* **349**:860–864. doi:10.1126/science.aaa8764

448 Lee K, Yoon SS. 2017. <i>Pseudomonas aeruginosa<i> biofilm, a programmed bacterial life for
449 fitness. *Journal of Microbiology and Biotechnology* **27**:1053–1064.
450 doi:10.4014/jmb.1611.11056

451 Lee KWK, Periasamy S, Mukherjee M, Xie C, Kjelleberg S, Rice SA. 2014. Biofilm development and
452 enhanced stress resistance of a model, mixed-species community biofilm. *ISME Journal* **8**:894–
453 907. doi:10.1038/ismej.2013.194

454 Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S,
455 Correia K, Diener C, Dräger A, Ebert BE, Edirisinghe JN, Faria JP, Feist AM, Fengos G, Fleming
456 RMT, García-Jiménez B, Hatzimanikatis V, van Helvoirt W, Henry CS, Hermjakob H, Herrgård
457 MJ, Kaafarani A, Kim HU, King Z, Klamt S, Klipp E, Koehorst JJ, König M, Lakshmanan M, Lee D-
458 Y, Lee SY, Lee S, Lewis NE, Liu F, Ma H, Machado D, Mahadevan R, Maia P, Mardinoglu A,
459 Medlock GL, Monk JM, Nielsen J, Nielsen LK, Nogales J, Nookaew I, Palsson BO, Papin JA, Patil
460 KR, Poolman M, Price ND, Resendis-Antonio O, Richelle A, Rocha I, Sánchez BJ, Schaap PJ,
461 Malik Sheriff RS, Shoaie S, Sonnenschein N, Teusink B, Vilaça P, Vik JO, Wodke JAH, Xavier JC,
462 Yuan Q, Zakhartsev M, Zhang C. 2020. MEMOTE for standardized genome-scale metabolic
463 model testing. *Nature Biotechnology 2020 38:3* **38**:272–276. doi:10.1038/s41587-020-0446-y

464 Liu W, Jacquiod S, Brejnrod A, Russel J, Burmølle M, Sørensen SJ. 2019. Deciphering links between
465 bacterial interactions and spatial organization in multispecies biofilms. *ISME Journal* **13**:3054–
466 3066. doi:10.1038/s41396-019-0494-9

467 Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS,
468 Srivastava DS, Crowe SA, Doebeli M, Parfrey LW. 2018. Function and functional redundancy in
469 microbial systems. *Nature Ecology & Evolution 2018 2:6* **2**:936–943. doi:10.1038/s41559-018-
470 0519-1

471 Machado D, Andrejev S, Tramontano M, Patil KR. 2018. Fast automated reconstruction of genome-
472 scale metabolic models for microbial species and communities. *Nucleic Acids Research*
473 **46**:7542–7553. doi:10.1093/NAR/GKY537

474 Nielsen AT, Tolker-Nielsen T, Barken KB, Molin S. 2000. Role of commensal relationships on the
475 spatial structure of a surface-attached microbial consortium. *Environmental Microbiology*
476 **2**:59–68. doi:10.1046/j.1462-2920.2000.00084.x

477 Niu B, Paulson JN, Zheng X, Kolter R. 2017. Simplified and representative bacterial community of
478 maize roots. *Proceedings of the National Academy of Sciences of the United States of America*
479 **114**:E2450–E2459. doi:10.1073/pnas.1616148114

480 Oliveira NM, Martinez-Garcia E, Xavier J, Durham WM, Kolter R, Kim W, Foster KR. 2015. Biofilm
481 formation as a response to ecological competition. *PLoS Biology* **13**:1–23.
482 doi:10.1371/journal.pbio.1002191

483 Ortiz A, Vega NM, Ratzke C, Gore J. 2021. Interspecies bacterial competition regulates community
484 assembly in the C. elegans intestine. *ISME Journal* 1–12. doi:10.1038/s41396-021-00910-4

485 Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA,
486 Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*
487 **31**:3691–3693. doi:10.1093/BIOINFORMATICS/BTV421

488 Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, McSpadden Gardener BB, Kinkel LL, Garrett KA.
489 2016. Microbiome networks: A systems framework for identifying candidate microbial
490 assemblages for disease management. *Phytopathology* **106**:1083–1096. doi:10.1094/PHYTO-
491 02-16-0058-FI

Raghupathi PK, Liu W, Sabbe K, Houf K, Burmølle M, Sørensen SJ. 2018. Synergistic Interactions within a Multispecies Biofilm Enhance Individual Species Protection against Grazing by a Pelagic Protozoan. *Frontiers in Microbiology* **8**:2649. doi:10.3389/fmicb.2017.02649

Ratzke C, Barrere J, Gore J. 2020. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology and Evolution* **4**:376–383. doi:10.1038/s41559-020-1099-4

Ren D, Madsen JS, Sørensen SJ, Burmølle M. 2015. High prevalence of biofilm synergy among bacterial soil isolates in cocultures indicates bacterial interspecific cooperation. *ISME Journal* **9**:81–89. doi:10.1038/ismej.2014.96

Rendueles O, Ghigo JM. 2012. Multi-species biofilms: how to avoid unfriendly neighbors. *FEMS Microbiology Reviews* **36**:972–989. doi:10.1111/j.1574-6976.2012.00328.x

Sadiq FA, Burmølle M, Heyndrickx M, Flint S, Lu W, Chen W, Zhao J, Zhang H. 2021. Community-wide changes reflecting bacterial interspecific interactions in multispecies biofilms. *Critical Reviews in Microbiology* 1–21. doi:10.1080/1040841x.2021.1887079

Sun X, Xu Z, Xie J, Hesselberg-Thomsen V, Tan T, Zheng D, Strube ML, Dragoš A, Shen Q, Zhang R, Kovács ÁT. 2021. Bacillus velezensis stimulates resident rhizosphere Pseudomonas stutzeri for plant health through metabolic interactions. *ISME Journal* 1–14. doi:10.1038/s41396-021-01125-3

Wei Z, Gu Y, Friman VP, Kowalchuk GA, Xu Y, Shen Q, Jousset A. 2019. Initial soil microbiome composition and functioning predetermine future plant health. *Science Advances* **5**:759–784. doi:10.1126/sciadv.aaw0759

513 Weiss AS, Burrichter AG, Chakravarthy A, Raj D, von Strempel A, Meng C, Kleigrewe K, Münch PC,
514 Rössler L, Huber C, Eisenreich W, Jochum LM, Göing S, Jung K, Lincetto C, Hübner J, Marinos G,
515 Zimmermann J, Kaleta C, Sanchez A, Stecher B. 2021. In vitro interaction network of a synthetic
516 gut bacterial community. *The ISME Journal 2021* 1–15. doi:10.1038/s41396-021-01153-z

517 Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, Cordero OX, Brown SP, Momeni B, Shou
518 W, Kettle H, Flint HJ, Haas AF, Laroche B, Kreft J-U, Rainey PB, Freilich S, Schuster S, Milferstedt
519 K, van der Meer JR, Groβkopf T, Huisman J, Free A, Picioreanu C, Quince C, Klapper I, Labarthe
520 S, Smets BF, Wang H, Soyer OS. 2016. Challenges in microbial ecology: building predictive
521 understanding of community function and dynamics. *The ISME Journal 2016 10:11* **10**:2557–
522 2568. doi:10.1038/ismej.2016.45

523 Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. 2015. Metabolic dependencies
524 drive species co-occurrence in diverse microbial communities. *Proceedings of the National*
525 *Academy of Sciences* **112**:6449–6454. doi:10.1073/PNAS.1421834112

526 Zorrilla F, Buric F, Patil KR, Zelezniak A. 2021. metaGEM: reconstruction of genome scale metabolic
527 models directly from metagenomes. *Nucleic Acids Research* **49**:e126–e126.
528 doi:10.1093/NAR/GKAB815

529

530