bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Full Title

From model to field: predictive design of microbial communities for resilient crop performance

Short Title

Designing SynComs to Enhance Crop Resilience

Author List

Shinichi Yamazaki[1,8,#], Masaru Nakayasu[2,#], Keiko Kanai[2], Rie Mizuno[2], Rumi Kaida[3], Sachiko Masuda[4], Arisa Shibata[4], Ken Shirasu[4], Atsushi J. Nagano[5,6], Yoshiharu Fujii[3], Akifumi Sugiyama[2] and Yuichi Aoki[1,7 *]

Author Affiliations

1. Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, 980-8573, Japan
2. Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan
3. Department of International Environmental and Agricultural Sciences, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, 183-8509, Japan

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

4. Plant Immunity Research Group, RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

5. Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, 464-8601, Japan

6. Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, 997-0017, Japan

7. Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, 980-8679, Japan

8. Current address: RIKEN Center for Sustainable Resource Science, Tsukuba, Ibaraki, 305-0074, Japan

# These authors contributed equally to this work.

***Corresponding Author:** Yuichi Aoki

Tohoku Medical Megabank Organization, Tohoku University

2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan

Phone: +81-22-274-6040

E-mail: yuichi.aoki.e3@tohoku.ac.jp

Author contributions

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

S.Y. and M.N. contributed equally. S.Y., M.N., A.S. and Y.A. conceived and designed the study. M.N. and A.S. isolated bacterial strains. M.N., K.K., R.M., and A.S. conducted SynCom inoculation experiments in the laboratory. R.K. and Y.F. conducted cultivation experiment in the field. M.N., S.M., A.S., and K.S. performed whole-genome sequencing. M.N. and A.J.N. performed RNA-seq experiment. M.N. performed 16S amplicon-seq experiment. S.Y., M.N., and Y.A. analyzed the data. S.Y., M.N, A.S., and Y.A. wrote the article with contributions from all the authors.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Abstract

Synthetic microbial communities (SynComs) represent a promising approach to enhance crop growth and stress resilience through microbiome engineering. However, the systematic design and field validation of SynComs remain limited. Here, we present a predictive framework for SynCom optimization, integrating plant phenotyping, microbial genomics, and machine learning. Using tomato as a model, we tested over 800 SynCom–temperature combinations consisting of root endophytic bacteria and rhizosphere metabolites. An Elastic Net regression model trained on plant biomass data accurately predicted the performance of unseen SynComs, with prediction accuracy plateauing at ~5% (301/6144) of all possible SynCom–temperature combinations. Incorporating genomic features significantly improved model performance, whereas microbiome compositional data alone were not informative. We applied the model to design novel SynComs, which were tested in both laboratory and field conditions using a commercial tomato cultivar. The model-informed SynCom enhanced plant growth in field trials and improved heat stress tolerance under controlled laboratory conditions. Multi-omics analyses and feature importance metrics identified specific microbial taxa, including *Sphingobium* sp., whose enrichment was linked to host plant metabolite (e.g., tomatine) and stress-responsive gene expression. Our results demonstrate a scalable strategy for the predictive design of beneficial microbiomes to improve resilient crop performance under real-world conditions.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Keywords

Synthetic microbial communities, SynCom, tomato, plant-microbe interactions, environmental stress resilience

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Introduction

The plant rhizosphere is a major site of plant–soil–microbe interactions and constitutes a unique chemical environment that contains a unique mix of various metabolites, organic compounds, and minerals. Root-associated microorganisms influence plant nutrient acquisition, pathogen resistance, and abiotic stress sensitivity [1]. These effects mean that plant rhizosphere microbiota is an important factor regulating the growth of host plants. Using microbes to improve plant nutrient use efficiency, alleviate environmental stress, and suppress plant pathogens has been proposed as a strategy to reduce the use of chemical fertilizers and pesticides in agriculture, thereby enhancing sustainable food production [2–4].

The rhizosphere contains a microbial community that differs from that present in bulk soil, and this is shaped by various metabolites secreted by plant roots. Root exudates contain over 10% of the carbon that plants fix via photosynthesis [5,6], and their primary metabolites, including sugars, organic acids, and amino acids, are essential carbon and nitrogen sources for soil microorganisms [7,8]. In recent years, researchers have found that specialized (secondary) metabolites are involved in the formation of various plant-specific rhizosphere microbiota. For example, tomatine, a saponin possessing antifungal, antibacterial, and insecticidal activity that is produced by tomato (*Solanum lycopersicum*), has been found to be involved in establishing and maintaining the rhizosphere microflora, where it increases the abundance of a specific bacterial group (e.g., genus

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Sphingobium [9]). Recently, our team revealed that another isolate from tomato roots, *Sphingobium* sp. RC1, participates in a tomatine degradation pathway, which suggests that it may be involved in interactions specific to tomato [10]. Other compounds, including the soybean isoflavonoid and maize flavanone [11,12], the sulfur-containing indole alkaloid camalexin produced by *Arabidopsis thaliana* [13], coumarin secreted by *A. thaliana* [14], and the benzoxazinoids produced by maize [15,16], are now thought to be involved in the shaping of the rhizosphere microbial community and the regulation of plant growth promotion by microbes. Furthermore, recent studies have shown that metabolite and microbial profiles within the rhizosphere are also altered by abiotic stress [17,18], suggesting that plants may modulate rhizosphere microbiota by varying root-secreted metabolite production to facilitate acclimation to various environmental changes. These findings highlight the potential to engineer rhizosphere microbial communities that support plant resilience under abiotic stress conditions.

Many plant growth-promoting (PGP) microorganisms have also been isolated from plant rhizospheres [19]. Known PGP traits include phytohormone production, increased nutrient availability via biological nitrogen fixation and solubilization of insoluble minerals, and biological control of pathogens via antibiotic and siderophore production [20–22]. Experimental work has shown that inoculating a single strain with a specific PGP trait is a simple method that is effective in specific environments; however, the effect may not always be consistent [23,24]. This

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

103 unpredictability can occur due to oversimplified models of the effects of plant-microbe
104 associations, and can result in inoculated microbes being quickly outcompeted by species in the
105 indigenous microbiome [23,25]. In addition, the soil microbiota creates a diverse and complex
106 network of interactions that can make it difficult to understand the detailed mechanisms at work.
107 One approach to resolving this issue involves the use of synthetic microbial communities
108 (SynCom). SynCom is an emerging technique involving the coculturing of multiple microbial
109 strains to mimic the structure and function of environmental microbial communities [26,27]. This
110 approach allows us to determine the function of complex microbial communities and their
111 biological interactions by isolating them as simple systems. For example, Castrillo *et al*. used
112 SynCom to demonstrate the association among phosphate starvation response, immune response,
113 and root microbiome in Arabidopsis [28], while Niu *et al.* elucidated the role-specific dynamics of
114 representative bacteria found in the maize root microbiome by using simplified SynCom as a
115 model ecosystem [29]. In addition, SynCom has been found to produce superior pathogen
116 suppression effects than single strain inoculation, indicating that constituent bacteria may exert
117 synergistic effects [30]. However, when designing a SynCom using different microbial compositions,
118 it may be costly or impossible to experimentally verify all possible combinations and thereby
119 identify the best SynCom. In such a case, it may be more helpful to make use of computational
120 approaches such as machine learning to optimize modeling and prediction. In recent years, some

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

121 studies have applied machine learning techniques to omics datasets (including the microbiome,
122 transcriptome, metabolome, and genome, among others) to predict soil health indices [31] and to
123 identify genetic and/or metabolic biomarkers associated with human disease [32,33].

124 In this study, we aimed to develop a model-based approach for optimizing SynComs to
125 promote plant growth and improve crop resilience under environmental stress conditions. We
126 designed a diverse SynCom composed of tomato root endophytic bacteria, including *Sphingobium*
127 sp. RC1, in combination with rhizosphere metabolites. Tomatine was added alongside the
128 SynCom, based on previous findings that microbial metabolic capacity can influence plant-
129 microbiota interactions in the rhizosphere [34–37]. Next, we constructed machine learning models by
130 combining various SynCom compositions with genomic and microbiome information to predict
131 plant growth and design optimal SynComs. Feature importance of constructed models and multi-
132 omics analysis of SynCom-inoculated plants were used to identify key microbes associated with
133 plant growth promotion. Finally, we investigated the molecular mechanisms underlying these
134 interactions and their role in supporting crop performance under stress conditions (Fig. 1).

135

136

137 **Results**

138 *Data collection of plant growth, root microbiome, and microbial genome.*

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

We first isolated 125 bacterial strains from the tomato root endophytic community. Taxonomic classification of these strains based on partial 16S rRNA gene sequences revealed that they belonged to four phyla (i.e., Actinobacteriota, Bacteroidota, Firmicutes, and Proteobacteria), representing 17 families and 26 genera (Supplemental Table S1). According to microbiome data produced by a previous study from our team [9] (DRA011415), most of these taxa were enriched in the tomato rhizosphere and root endosphere. We therefore selected nine phylogenetically diverse strains, including *Sphingobium* sp. RC1, from all strains that could be precultured from a stock (Fig. 2A).

We performed whole-genome sequencing analysis on all nine isolates. Overall we obtained nine high-quality circular genomes. We then attempted to identify these species by comparing sample genomes to known bacterial genomes. Two strains (T1-8 and T2-2) appeared as hits when querying the genomes of known species with an average nucleotide identity of more than 95%. For this analysis, four strains (T2-21, T2-23, T3-8, and T5-5) were marked as coming from the genomes of unspecified species, but other three strains (T2-11, T2-14, and RC1) did not hit any known genome and were considered to be new species (Supplemental Table S2). Gene prediction and functional annotation of the genome were then performed to characterize the genetic composition of each isolate. We compared the possession of genes related to PGP traits such as auxin production, the suppression of ethylene accumulation, phosphate solubilization,

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

157 siderophore production, and nitrogen fixation [20–22] among the genomes of the nine isolates
158 discussed here (Fig. 2B). With respect to plant hormones, T2-2 possessed the auxin biosynthetic
159 genes *iaaM* and *iaaH*, while T5-5 possessed the ACC deaminase gene *acdS*, which can degrade
160 ethylene precursor and suppress ethylene accumulation [38]. As for phosphate solubilization, all the
161 isolates except T2-23 possessed any of the *phn* genes that are involved in the degradation of
162 phosphonate [39]. In particular, two Rhizobiales isolates, T2-14 and T2-21, were found to possess
163 more *phn* genes than any others. T2-21 is also linked to a gluconic acid biosynthetic pathway that
164 is involved in phosphate solubilization. Regarding siderophore production, all nine isolates
165 possessed at least one of the genes involved in known siderophore biosynthesis. Moreover, four
166 isolates (i.e., T2-2, T5-5, T2-11, and T2-21) possessed multiple genes within these biosynthetic
167 pathways. Finally, none of the isolates possessed the *nif* gene required for biological nitrogen
168 fixation. Taken together, these results suggest that all nine isolates may be potential plant growth-
169 promoting bacteria.

170 The resulting SynCom, which consisted of nine isolates, was then inoculated into tomato
171 seedlings. These seedlings were grown under high temperature stress in a growth chamber, and
172 the effect of the SynCom on plant growth was determined by evaluating root and shoot fresh
173 weight. We also varied the SynCom composition, including various combinations of the nine
174 isolates, for example excluding 1 or 2 isolates, or including only 1 or 2. SynCom inoculation was

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

175 combined with addition of the plant metabolites tomatine or daidzein, and resulted in 2048
176 theoretical SynCom compositions. Next, we conducted 30 independent growing trials at 30°C
177 (optimal), 35°C, and 38°C (high temperature), thus obtaining growth data for a total of 834 tomato
178 plants accompanied by 218 SynCom combinations (Supplemental Table S3). In this series of trials,
179 bacterial contamination was suspected in some of the microbial stocks used for SynCom
180 preparation, so some of the strains used for the SynCom preparation were renewed for an 11th
181 trial. In addition, since we observed some seasonal changes in growth data even under a controlled
182 environment, a control condition (i.e., noninoculated, 30°C) was added from the 18th trial to
183 facilitate batch-to-batch correction (Supplemental Fig. S1). Overall, we obtained growth data for
184 301 plants of 102 different SynCom compositions to be used for subsequent modeling.

185 We performed microbiome analysis via 16S rRNA amplicon sequencing of the endophytes
186 of all tomato roots after taking root growth measurements. The relative abundance of amplicon
187 sequence variants (ASVs) that corresponded to each of the nine inoculants was found to increase
188 following inoculation, confirming that all inoculants had colonized their respective tomato roots
189 (Supplemental Fig. S2). We also observed that plant metabolites exerted an effect, with
190 simultaneous inoculation with tomatine significantly increasing the relative abundance of RC1
191 (Fig. 3A). In contrast no such change was observed for the other eight bacterial strains or during
192 simultaneous inoculation with daidzein (Supplemental Fig. S3 and S4). Next, diversity analysis

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

193 was performed to examine the effects of temperature and plant metabolite on the endophytic
194 community. Overall, high temperature did not affect species richness and evenness, but the
195 addition of plant metabolites tended to increase both diversity indices (Fig. 3B). Principal
196 coordinate analysis (PCoA) based on weighted UniFrac distance showed a significant effect of
197 temperature on endophytic community (PERMANOVA: $p < 0.001$, $R^2 = 0.021$), but we did not
198 observe a clear separation by temperature alone (Supplemental Fig. S5). Taken together, these
199 results suggest that the effect of high temperature on the endophytic community is not a
200 determinant factor and can be controlled by SynCom inoculation.
201
202 *Modeling and application*
203 Based on the data collected, we constructed three different models to predict tomato growth.
204 Specifically, we used SynCom inoculant composition (SynCom-inoculant model; S-I model),
205 SynCom gene composition (SynCom-Genome model; S-G model), and microbial composition of
206 the roots (Microbiome model; MB model) as explanatory variables, respectively. The algorithm
207 we chose was an Elastic Net regression model, and this was then used to evaluate and compare
208 the performance of the three models above. First, for the S-I model explanatory variables were
209 divided into three categories: microbes, chemicals (i.e., tomatine and daidzein), and environment
210 (temperature). Each of their respective contributions to overall plant growth was examined based

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

211 on model performance. We found that environmental factors alone had the largest effect on plant
212 growth (E: Mean R^2 = 0.402), while microbial and chemical factors alone each had very small
213 effects (M: Mean R^2 = 0.046; C: Mean R^2 = 0.010; Fig. 4A). However, the model containing all
214 three factors as well as the interaction of two variables (MCEX) showed the highest performance
215 (Mean R^2 = 0.510). This suggested that microbial and chemical factors can both affect plant
216 growth, and so we used all of these as explanatory variables in subsequent models.

217 The S-G model using the SynCom genome information showed significantly better
218 prediction performance (Mean R^2 = 0.600) than the S-I model, and showed the highest
219 performance among all models developed in this study (Fig. 4B). This result means that the gene
220 composition of SynCom and especially the presence of specific functional genes strongly
221 influence tomato growth. In contrast, using root microbiome data did not improve the predictive
222 performance of the MB model (Mean R^2 = 0.467) relative to the S-I model. Instead, models using
223 phylum-level profiles and inferred metabolic pathway profiles showed significantly lower
224 performance (Mean R^2 = 0.407; Mean R^2 = 0.384). These results suggest that microbiome
225 information sampled after cultivation is insufficiently able to explain the biomass of host plants
226 and that the composition of SynCom inoculants and/or genomic data is crucial.

227 We also that, in addition to the Elastic Net algorithm used in most models, we also tested
228 Lasso and Random Forest algorithms. However, in all three models (i.e., S-I, S-G, and MB) the

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

229 Elastic Net algorithm consistently showed the best performance (Supplemental Fig. S6). We also
230 examined the effect of sample size by stepwise subsampling of training data. In doing so we found
231 that the Elastic Net model was mostly saturated in predictive performance using the full size of
232 the current dataset, indicating that the sample size of our dataset was sufficient for modeling
233 (Supplemental Fig. S7).

234 To verify the versatility of constructed models, we designed new SynComs based on the
235 S-I model from the top, middle, and bottom predicted values. We then acquired new growth data
236 and evaluated the predictive performance of the models (Fig. 4C). Overall, a significant
237 correlation was observed between measured and predicted values for the S-I model ($p = 8.704e-09$, $r = 0.619$), and an even higher correlation was obtained for the S-G model ($p = 2.815e-13$, $r = 0.735$). These results suggest that it may be useful to design a new SynCom using the machine
240 learning model, and that more effective SynCom design and modification can be achieved by
241 using genetic information from constituent bacteria.

242 Based on the feature importance of the S-I model, we identified specific candidate microbe-
243 microbe and microbe-plant metabolite interactions that may promote plant growth. The top 15
244 variables deemed most important for positive association with tomato growth were microbial
245 interactions, either with each other or with metabolites (Fig. 5A,B). Of these, the simultaneous
246 inoculation of two isolates—e.g., T2-2 and RC1 or T3-8 and T5-5—was found to significantly

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

247 improve tomato growth when cultivated at either optimal (30ºC) or high (35°C) temperature (Fig.
248 5B,C). In addition, the simultaneous addition of RC1 and tomatine microbe-metabolite pair was
249 found to significantly improve heat stress tolerance (Fig. 5C). Taken together, these results show
250 that such interactions have a synergistic effect on plant growth promotion.

251 Although we also attempted to functionally characterize the important genes of the S-G
252 model by performing GO enrichment analysis, this was discarded since only about 10% of all
253 genes were annotated using GO terms, which was not sufficiently informative.

254

255 *Multi-omics analysis of tomato plants inoculated with SynCom*

256 The top five SynComs designed based on the S-I model, hereafter termed Good SynCom G1–G5,
257 were evaluated again using a field cultivation experiment to determine their effects on plant
258 growth. Most of the Good SynComs tended to increase plant biomass. In particular, SynCom G2
259 showed a statistically significant growth-promoting effect relative to non-inoculated plants (Fig.
260 6A). Overall, SynCom G2 consisted of six bacterial strains (i.e., T2-2, T2-11, T2-23, T3-8, T5-5,
261 and RC1) combined with tomatine. We subsequently performed a detailed multi-omics analysis
262 of SynCom G2 using laboratory experiments. These results were then compared to non-inoculated
263 and Bad SynCom B5 preparations, and inoculated tomatoes were grown under optimal and high
264 temperature conditions to generate plant biomass, leaf and root transcriptome, and root

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

265 microbiome data. SynCom B5 consisted of tomatine and five bacterial strains (i.e., T1-8, T2-14,
266 T2-21, T2-23, and T5-5) and was among the bottom five SynComs generated based on the S-I
267 model. Overall, SynCom G2 inoculation was associated with a significant improvement in plant
268 tolerance to heat stress, whereas SynCom B5 inoculation resulted in no differences relative to
269 non-inoculated plants (Fig. 6B–D).

270 In transcriptome analysis, principal component analysis and clustering analysis showed that
271 tomato gene expression was clearly different based on tissue and temperature (Supplemental Figs.
272 S8 and S9). Moreover, the effect of SynCom inoculation was remarkable in the roots exposed to
273 heat stress conditions (Supplemental Fig. S8). Next, we conducted a differential gene expression
274 analysis between SynCom-inoculated and non-inoculated tomatoes. We found that the number of
275 differentially expressed genes (DEGs) was highest in roots inoculated with SynCom G2 under
276 heat stress conditions and was the lowest in roots inoculated with SynCom B5 at the same
277 temperature (Supplemental Fig. S10). Next, DEGs were clustered based on their coexpression
278 patterns and functionally characterized by gene ontology (GO) enrichment analysis. These results
279 showed that, for root tissue subjected to heat stress conditions, G2 SynCom inoculation
280 upregulated the expression of genes involved in “response to heat” and “protein folding” (Fig. 7).
281 For example, we found significant up-regulation of heat shock transcription factors (HSFs) and
282 heat shock proteins (HSPs), two important regulators of heat stress response and protein quality,

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

283 respectively (Fig. 8). Such changes in gene expression may have improved the heat tolerance of
284 the tomato plants. Moreover, we also found that SynCom B5 inoculation increased the expression
285 of genes involved in "response to biotic stimulus" in leaves (Fig. 7), thereby suggesting that this
286 SynCom exerts an unfavorable effect on tomato plants.

287 In the microbiome analysis, a principal coordinates analysis (PCoA) based on Bray-Curtis
288 dissimilarity distance and PERMANOVA revealed that the endophytic community of tomato roots
289 was significantly affected by SynCom inoculation ($p < 0.001$, $R^2 = 0.410$; Fig. 9A). In contrast,
290 temperature showed a limited effect ($p = 0.082$, $R^2 = 0.051$). Moreover, the alpha diversity of the
291 endophytic community was decreased by SynCom G2 inoculation (Fig. 9B). In general, the
292 composition of bacterial taxa on the genus level differed greatly among treatments, and SynCom
293 members increased in abundance following inoculation (Supplemental Fig. S11). Especially in
294 SynCom G2-inoculated tomatoes, *Pseudoxanthomonas* (same genus as T2-2) was the most
295 dominant, accounting for 30% of the total (Fig. 9C). Differential abundance analysis using
296 ALDEx2 showed significant increases in *Achromobacter* (same as T5-5), Micrococcaceae (same
297 as T2-23) and *Pseudoxanthomonas* in SynCom G2 relative to the control (Fig. 9D). For SynCom
298 B5, we observed significant increases in *Phyllobacterium* (same as T2-21) and *Achromobacter*.
299 Of these, *Achromobacter* (T5-5) was twofold higher in SynCom G2-inoculated tomatoes than
300 SynCom B5 inoculation. In addition, *Achromobacter* was the dominant genus (i.e., ranking in the

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

301 top 15) only in the SynCom G2-inoculated tomatoes even though it was a common constituent in
302 both SynCom G2 and B5.

303 Finally, integrated omics analyses were performed using sparse canonical correlation
304 analysis (sCCA) and co-occurrence network analysis. We extracted relevant components among
305 different omics data for plant biomass, leaf and root transcriptome, and root microbiomes by
306 sCCA. Next, co-occurrence network analysis using WGCNA was used to analyze the relationship
307 between pairs of components and to examine the characteristic co-occurrence modules associated
308 with plant biomass. Overall, sCCA was able to extract components of 4,938 leaf genes, 7,089 root
309 genes, and 176 root bacterial taxa (genus level). Network analysis allowed the classification of
310 the 12,205 components into 29 modules based on co-occurrence patterns. When considering the
311 specific module associated with plant biomass (i.e., the fresh weight of shoots and roots), the
312 network analysis identified four genes in leaves, 85 genes in roots, and one bacterial taxon,
313 *Sphingobium*. Although we observed no significant enrichment of specific GO terms in this
314 module, the gene Solyc05g053080.2, which encodes endomembrane-type CA-ATPase 4, was the
315 hub and was found to be associated with *Sphingobium* and plant biomass (Fig. 10A). In addition,
316 the relative abundance of *Sphingobium* and the gene expression of Solyc05g053080.2 were both
317 strongly correlated with plant biomass (Fig. 10B). Taken together, these results indicate that
318 *Sphingobium* spp. in the root microbiota are closely related to tomato growth and that specific

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

319 plant genes may be involved in this relationship.

320

321

322 **Discussion**

323 To increase sustainability of crop production in an era of climate change, it is important to
324 maximize crop production potential by making better use of soilborne microorganisms and
325 metabolites produced by plants. One approach to achieve this is to mimic microbial communities
326 using SynCom techniques [26]. In this study, we used machine learning to optimize SynCom design,
327 and followed this process from model construction to experimental validation (Fig. 1).

328 Specifically, we constructed models to predict plant growth based on SynCom structural
329 information and achieved high prediction performance. These results show that it is feasible to
330 design an arbitrary SynCom based on machine learning models. Here, we used relatively simple
331 data and models (e.g., binary data describing inoculum presence/absence and linear models). We
332 built the model to represent 50% of the plant biomass (Mean $R^2 = 0.510$) from real data for about
333 5% of all SynCom combinations. Moreover, performance improvement was found to be almost
334 saturated in the linear model (Supplemental Fig. S7), and we therefore conclude that there was
335 little benefit in further increasing the sample size. This model may serve as a useful example for
336 future studies applying similar approaches.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

337 The SynCom-Genome model, which incorporates information regarding SynCom gene
338 composition, showed the highest predictive performance in this study (Fig. 4B). This suggests
339 that SynCom genetic composition, or the metabolic functions within the community, can strongly
340 influence plant growth and abiotic stress tolerance. Although the metabolic activities of the
341 community remain poorly understood, it has recently been reported that the exchange of
342 extracellular metabolites, or metabolic interactions, can affect the community phenotype, e.g.,
343 drug resistance [40]. In addition, recent studies have used mutant libraries to identify genetic
344 requirements for the establishment of microbial interactions [41] and have attempted to understand
345 metabolic systems through modeling using genomic information from the community [42,43]. Such
346 genetic information may be important not only for improving our biological understanding of
347 plant-microbe interactions but also for bioengineering applications. For example, DNA sequence
348 motifs that are characteristic of important gene groups may be used as biomarkers, and SynComs
349 can be further improved by functional modification via genetic engineering and/or by more
350 efficient SynCom design.

351 The feature importance analysis of the model constructed in this study provides notable
352 insight into interactions between organisms within the tomato rhizosphere. The most important
353 features of the S-I model are microbial interactions, which suggests that the impact of inoculation
354 with multiple microorganisms is significantly greater than the impact of inoculation with a single

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

355 microorganism. In particular, the top two combinations—i.e., T2-2 with RC1 and T3-8 with T5-
356 5—showed significantly higher growth when simultaneously inoculated (Fig. 5B). Subsequent
357 multi-omics analyses suggested that RC1 was closely related to plant biomass, while T2-2 and
358 T5-5 were microbial taxa that were characteristically dominant in the root microbiota following
359 SynCom G2 inoculation (Fig. 9C). Although T5-5 was present in both SynCom G2 and B5, it was
360 relatively more abundant in SynCom G2. This may be due to interactions between T5-5 and T3-
361 8 in SynCom G2, which may in turn promote the colonization of T5-5 in the roots. Genome
362 analysis revealed that both T2-2 and T5-5 possess genes involved in plant hormone metabolism
363 (Fig. 2B). For example, the auxin biosynthesis gene possessed by T2-2 may be a representative
364 PGP trait, since auxin-producing bacteria such as *Aeromonas punctata* and *Azospirillum*
365 *brasilense* are known to stimulate plant growth and induce morphological alteration in
366 Arabidopsis [44–46]. Furthermore, the ACC deaminase gene possessed by T5-5 functions to degrade
367 ethylene precursors and limit ethylene levels in plants exposed to stress conditions [38]. For instance,
368 ACC deaminase-producing bacteria such as *Bacillus* sp., *Brevibacterium linens*, and
369 *Streptomyces venezuelae* have been found to enhance rice plant growth under salinity and heat
370 stress [47–50]. Considering these studies, T2-2 and T5-5 isolated here may play a central function in
371 the promotion of tomato growth by SynCom G2.

372 Next we considered the interaction between microbes and metabolites, and the feature

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

373 importance of the S-I model suggested that there was a crucial relationship between RC1 and
374 tomatine. In a previous study from our team, we showed that tomatine, which is secreted by
375 tomato roots into the rhizosphere, is involved in assembling the tomato rhizosphere microbiota
376 and increases the abundance of *Sphingobium* spp. in the soil [9]. More recently, we also found that
377 RC1 has a tomatine degradation pathway [10]. In this study, we found that tomatine enhanced the
378 colonization of inoculum RC1 (Fig. 3A). This finding confirms these earlier reports, and the fact
379 that simultaneous inoculation enhanced heat stress tolerance of tomato plants (Fig. 5C) suggested
380 that the tomatine-mediated relationship between *Sphingobium* spp. and tomato is important for
381 tomato stress tolerance. In contrast, such effects were not observed with the soybean-derived
382 metabolite daidzein, implying that impact of plant-derived metabolites on the rhizosphere
383 microbiota may be compound-specific. Multi-omics analysis also showed that among the many
384 microbes present in the tomato rhizosphere, *Sphingobium* spp. are the most closely related to
385 tomato growth (Fig. 10). Therefore, *Sphingobium* spp. may be a potential indicator of tomato
386 growth. However, we also found that *Sphingobium* sp. RC1 did not have as many representative
387 PGP trait-related genes as other SynCom members (Fig. 2B). Consequently, it may indirectly
388 contribute as a helper microbe. For example, it may suppress the antimicrobial effect of tomatine,
389 thereby promoting the colonization of other beneficial microbes and the development of a healthy
390 microbial community within the rhizosphere. Moreover, the expression

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

391 of Solyc05g053080 showed a strong positive correlation with both *Sphingobium* abundance and
392 plant biomass (Fig. 10). Although its precise physiological function remains unclear, it may be
393 involved in tomatine biosynthesis, potentially promoting tomato-*Sphingobium* interactions and
394 contributing to enhanced plant growth.

395 Next, we note that characteristic changes in gene expression in tomatoes inoculated with
396 SynCom G2 included an increase in the expression levels of HSF and HSP genes under high
397 temperature conditions (Fig. 8) relative to the corresponding levels observed in non-inoculated
398 and SynCom B5-inoculated tomatoes. This result indicates that SynCom G2 inoculation
399 augmented the heat stress response of tomato plants, resulting in improved growth under heat
400 stress conditions. Similar observations have been reported in other studies, where inoculations of
401 ACC deaminase-producing *Brevibacterium* and thermotolerant *Bacillus* were found to increase
402 HSP gene expression under heat stress and enhance the heat tolerance of rice and soybean,
403 respectively. [50,51]. In general, HSFs are key transcription factors regulating the heat stress response
404 [52], while HSPs are molecular chaperones that prevent the aggregation of denaturing proteins and
405 maintain protein homeostasis under heat and other forms of stress [53]. In addition, in the plant
406 biomass-associated network module (Fig. 10A), the hub gene Solyc05g053080 was found to be
407 homologous to endomembrane-type CA-ATPase 4 (ECA4), a gene that has been found to be
408 crucial for reactive oxygen species (ROS) production by RBOHD in *Arabidopsis thaliana* under

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

409 salinity stress conditions [54]. In general, RBOHD is critical for rapid systemic ROS signaling in
410 response to diverse abiotic and biotic stimuli in plants [55,56]. Moreover, ROS signaling leads to
411 downstream HSF activation and HSP accumulation in the heat stress response [57,58]. For the
412 SynCom G2 inoculation designed in this study, it is possible that nonpathogenic or symbiotic
413 bacterial stimuli may have activated these signaling pathways and thereby enhanced the heat
414 stress response.

415 It should be noted that since this study analyzed gene expression only one month after
416 SynCom inoculation, it is possible that some direct physiological responses to SynCom
417 inoculation may not have been observed. Further detailed analysis of time-series changes after
418 SynCom inoculation is expected to further elucidate the molecular mechanisms involved in
419 SynCom colonization and plant response.

420 In conclusion, we demonstrate that a simple linear model can effectively predict and guide
421 the design of synthetic microbial communities that enhance plant growth and stress resilience. We
422 show that incorporating genomic information from SynCom members further improves prediction
423 accuracy. In addition, we validate the effectiveness of model-designed SynComs in actual field
424 cultivation. Finally, feature importance analysis of our model and multi-omics analysis suggests
425 that specific microbial interactions within the rhizosphere microbial community play a key role
426 in conferring abiotic stress tolerance in tomato plants.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Materials and Methods

Isolation and culturing of tomato endophytic bacteria

To collect root endophytes, bacterial strains were isolated from the roots of tomato plants (*S. lycopersicum* cv. Micro-Tom). Tomato seeds were sown in pots filled with a 1:1 mixture of vermiculite and field soil collected at the Kyoto University of Advanced Science, Kameoka, Kyoto, Japan (34°99'38"N, 135°55'14"E). Seedlings were then grown in the laboratory for seven weeks at 25°C under a 16h/8h light/dark cycle. At seven weeks we obtained samples of tomato roots, then carefully removed the soil adhering to the roots, including the rhizosphere and rhizoplane, following a previously described protocol [59]. Next, endosphere compartments were washed with tap water before being homogenized in a 10 mM $MgCl_2$ solution by mortar and pestle. The resulting homogenates were then diluted and distributed into agar plates containing six different media, i.e., Tryptic Soy Broth (TSB), Tryptone Yeast extract Glucose (TYG), Yeast Extract Mannitol (YEM), M408, M715, and Tap Water Yeast Extract (TWYE) media [60]. All plates were incubated for 2–3 days at 28°C. Next, bacterial genomic DNA was extracted from each colony using the following method. First, colonies were suspended in 10 µL of an extraction buffer containing 25 mM NaOH and 0.2 mM EDTA heated at 95°C for 30 min. Next, 10 µL of 40 mM Tris-HCl solution (pH 6.8) was added. The *16S rRNA* genes in their respective isolates were

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

445 then PCR-amplified using the following primer set: 10F (5′-GTTTGATCCTGGCTCA-3′) and
446 800R (5′-TACCAGGGTATCTAATCC-3′). PCR amplification used KOD FX Neo (TOYOBO,
447 Osaka, Japan) and extracted DNA as the template sequence. The thermocycle program was as
448 follows: 94°C for 2 min followed by 35 cycles of 98°C for 10 sec, 50°C for 30 sec, and 68°C for
449 1 min. The generated PCR products were then purified using a Wizard Genomic DNA Purification
450 Kit (Promega, Madison, WI, USA) with all procedures performed as per the manufacturer's
451 instructions. After amplification, *16S rRNA* gene sequences were sequenced using the 10F primer,
452 after which each bacterial isolate was annotated using BLAST search. All isolates were cultured
453 on a growth medium (pH 7.0) containing 10 g L^{-1} peptone, 10 g L^{-1} beef extract, and 5 g L^{-1}
454 NaCl, and were stored in 20% glycerol at −80°C. Isolates were passaged on agar plates containing
455 the above growth medium.
456
457 *Genome analyses of SynCom member bacteria*
458 Each bacterial member of a given SynCom was streaked onto a growth medium agar plate as
459 described above. Next, a single colony was cultivated in 10–50 mL of liquid medium for three
460 days at 28°C following preculture in 2 mL of the same medium. After incubation, these cultures
461 were centrifuged at 4,000 × *g* for 5 min before being washed in 5 mL of TE buffer (i.e., 10 mM
462 Tris-HCl (pH 8.0) and 1 mM EDTA). Cell pellets were then stored at −30°C for genomic DNA

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

463 extraction. Genomic DNA of T2-2, T2-11, T2-21, and T5-5 was extracted using a Genomic DNA
464 Buffer Set and a QIAGEN Genomic-tip 100/G column (QIAGEN, Hilden, Germany), with all
465 procedures following the manufacturer's instructions. Genomic DNA of T1-8, T2-14, T2-23, and
466 T3-8 was extracted as per a previously described protocol [61] with minor modifications. Briefly,
467 cell pellets were lysed with TE buffer containing lysozyme, sodium dodecyl sulfate, and
468 proteinase K before being incubated at 37°C for 1 hour. Next, cleared lysate was pushed 10 times
469 through a syringe (38 mm × 0.8 mm) before being deproteinized using TE-saturated
470 phenol:chloroform:isoamyl alcohol (25:24:1). Genomic DNA was then collected via ethanol
471 precipitation before being dissolved in 100 μL of TE buffer. DNA concentrations were measured
472 using a BioSpec-nano (Shimadzu, Kyoto, Japan) and a Qubit 2.0 Fluorometer (Thermo Fisher
473 Scientific, Waltham, MA, USA). After quantification, all genomic DNA samples were kept at 4°C
474 until genome analysis.

475 Library construction and whole-genome sequencing were carried out as per a previously
476 described protocol [34] with minor modifications. Briefly, a DNA library was cut off at 15 kbp (for
477 T2-2, T2-11, T2-21, T2-23, and T5-5) or 10 kbp (for T1-8, T2-14, and T3-8) using a Blue Pippin
478 size selection system (Sage Science, Beverly, MA, USA). Genomic libraries were then sequenced
479 on a single PacBio sequel II system 2.0 cell. Using the resulting long-read data, genomes for T2-
480 2, T2-11, T2-21, T2-23, and T5-5 were assembled using the Hierarchical Genome Assembly

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Process (HGAP) version 4 as implemented within SMRTlink version 10.0. Genomes of T1-8, T2-14, and T3-8 were assembled using the Improved Phased Assembler (IPA) implemented within SMRTlink version 10.2 according to their expected genome sizes. Circlator version 1.5.5 [62] was used to assess whether genome assemblies were circularizable and to predict the starting position.

Next, genome-based species identification of SynCom members was performed using GTDB-Tk version 2.1.0 using the default parameters [63]. Gene prediction within genomes was performed using Prokka version 1.14.6 [64]. The functional annotation of predicted gene sets was performed using eggNOG-mapper version 2.1.9 using the default parameters [65,66]. For plant growth-promoting traits, pathways related to auxin production, suppression of ethylene accumulation, phosphate solubilization, siderophore production, and nitrogen fixation were retrieved from the MetaCyc [67] and KEGG pathway databases [68]. The names and KEGG Orthology IDs of genes involved in these pathways were queried in the functional annotations of the respective genome. Using this process, we created a plant growth-promoting (PGP)-related gene possession profile from combined sets of genes. Finally, orthologous gene groups of SynCom member genomes were inferred using OrthoFinder version 2.5.4 using the default parameters [69].

Laboratory growth experiments of SynCom-inoculated tomato plants

We first sterilized tomato seeds before sowing them on Murashige–Skoog medium supplemented

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

499 with 3% sucrose and 0.3% gellan gum as per a previously described protocol [9]. Next tomato
500 seedlings were grown at 25°C for two weeks before they were subjected to SynCom inoculation.
501 For each SynCom preparation, a particular configuration was precultured in 2–3 mL liquid growth
502 medium before being cultivated in 75–500 mL of the same medium for three days at 28°C.
503 Cultures were then centrifuged at 4,000 × *g* for 5 min before being washed with 20 mL of
504 autoclave-sterilized tap water. Cell pellets were then resuspended in 20 mL sterile tap water. Next,
505 two-week-old tomato seedlings were transferred to pots containing 40 g vermiculite. The
506 inoculation mixture contained 100mL sterile tap water containing 0.2% (w/v) Otsuka House No.
507 1 and No. 2 standard nutrients (Otsuka Chemical Co., Osaka, Japan) as well as a cell suspension
508 of each SynCom member at OD_{600} = 0.05; these solutions were inoculated into tomato seedlings.
509 Plants grew at 25°C for 14 days before being assigned one of three different growth temperatures
510 (i.e., 30°C, 35°C, or 38°C) in which they grew for another two weeks. Throughout the cultivation
511 period all tomato plants were grown under a 16h/8h light/dark cycle. After cultivation, tomato
512 plants were separated into aerial parts and roots. The roots were rinsed with tap water, and any
513 residual water was removed using Kimwipes. Finally, we measured the fresh weight of the above
514 ground parts and roots of the tomato plants.
515
516 *Microbiome analysis of endophytic bacteria in tomato roots*

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

After fresh weight measurements were taken, SynCom-inoculated tomato roots were homogenized using a BioMasher® II (Nippi. Inc., Tokyo, Japan). Next, DNA was extracted from the resulting homogenate using a DNeasy PowerSoil Pro Kit (QIAGEN) with all procedures performed as per the manufacturer's instructions. Next, amplicons of the 16S rRNA V4 region were prepared from each root DNA sample as per a previously described protocol [9]. In brief, the V4 region was PCR-amplified using a 515F/806R primer set with mitochondrial- and chloroplast-specific peptide nucleic acids (mPNA [N-term-GGCAAGTGTTCTTCGGA-C-term] and pPNA [N-term-GGCTCAACCCTGGACAG-C-term], respectively [PANAGENE Inc, Daejeon, Republic of Korea]). This was done to prevent the amplification of tomato mitochondrial and plastid sequences [70]. After PCR, the resulting amplicons were purified using an AMPure XP kit (Beckman Colter, Danvers, MA, USA) before being attached to a MiSeq (Illumina, San Diego, CA, USA) adapter via PCR. After purification of PCR products, the prepared 16S amplicon mixture was sequenced on a MiSeq (Illumina) platform by FASMAC Co. Ltd to obtain 2 × 250-bp paired-end sequences.

Raw sequences were processed using the QIIME2 version 2021.2 pipeline [71]. Raw paired-end sequences were first trimmed before the first 20 bases and after 200 bases. Next, denoised amplicon sequence variants (ASVs) were generated using DADA2 [72] using the q2-dada2 plugin in QIIME2. Taxonomic classification of ASVs was performed using the q2-feature-classifier

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

plugin and a Naive Bayes classifier [73], which has been pretrained on operational taxonomic units (99% identity) from the 515F/806R region of sequences in the SILVA rRNA gene database release 138 [74]. After filtering out mitochondrial and chloroplastic sequences, we obtained 20,141–374,296 reads per sample, representing 8,603 ASVs, and these were rarefied to 20,000 reads per sample for subsequent diversity analysis. Next, alpha and beta diversity indices were computed using the core-metrics-phylogenetic pipeline of the q2-diversity plugin. This plugin also performed principal coordinate analysis (PCoA) using weighted and unweighted UniFrac distances. Multiple alignment of ASVs was performed using MAFFT [75] and a phylogenetic tree was generated by FastTree [76], both implemented within the q2-phylogeny plugin. Finally, metagenomic functional abundance profiles were estimated using PICRUSt2 [77] and the KEGG pathway database.

Microbiome analysis of SynCom G2- and SynCom B5-inoculated tomato plants was performed using the same sequencing protocol described above. Here, we obtained 66,497–103,566 reads per sample representing 1,290 ASVs. These were then rarefied to 66,000 reads per sample for subsequent diversity analysis.

All statistical analyses of microbiome datasets were performed using R. The R package *stats* was used for Wilcoxon tests, the package *vegan* for PERMANOVA tests [78], and *ALDEx2* for differential abundance analysis [79,80].

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Model construction

As the response variable, plant shoot biomass values were first converted to ratios with control plants (i.e., noninoculated, 30°C) for each cultivation trial. These values were then box-cox transformed. For all models, the 301 samples were split 7:3 between training and test datasets for model construction and subsequent performance evaluation. This step was randomly repeated 50 times to evaluate average performance relative to variation in the training and test data split. Also, the hyperparameters of each model were determined using the five-fold cross-validation method.

For SynCom-Inoculant models, the presence or absence of microbes and metabolites in each SynCom was expressed as 1/0 and the cultivation temperatures (i.e., 30°C, 35°C, and 38°C) were scaled from 0 to 1. These variables and the interaction terms between each variable pair were used as the set of explanatory variables. For the SynCom-Genome model, the gene composition of each SynCom was estimated based on orthologous gene possession profile of constituent SynCom bacteria, and the presence or absence of orthologous gene groups was expressed as 1/0. These variables, the presence or absence of metabolites, and scaled cultivation temperature values were used as explanatory variables. For the Microbiome model, taxonomy profiles at the Phylum and Family levels and KEGG pathway profiles were filtered to the taxon and pathway detected in at least 5% (15 samples) of the 301 samples. The filtered taxon and pathway count data were centered log-ratio (CLR) transformed. These variables, the presence or

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

571 absence of metabolites, and scaled cultivation temperature values were used as explanatory
572 variables.

573 The following R packages: *ROCR*, *caret*, *glmnet*, *lme4*, and *randomForest* were used for
574 machine learning [81–84].

575

576 *Validation of SynCom model-based design*

577 The S-I model was used to predict plant growth for all SynCom compositions and growing
578 temperatures (i.e., 30°C, 35°C, and 38°C). We selected those SynCom compositions predicted to
579 be in the top 5% of predicted growth outcomes for each growing temperature, classified them into
580 eight groups using the k-means method, and then selected one from each group to validate the
581 eight unexplored SynCom compositions. We used a similar procedure to identify "bottom 5%"
582 and "10% around the median" SynCom compositions. These were also classified into eight groups
583 and a representative was then selected, yielding a total of 24 newly designed SynCom
584 compositions. Laboratory growth experiments were conducted as described above.

585

586 *Field trials of SynCom-inoculated tomato plants*

587 Tomato seeds (*S. lycopersicum* cv. Aiko) were sown on commercial soil (Nihonhiryo com.) and
588 grown for 2 weeks in a growth chamber (25°C, 16h/8h light/dark photoperiod). On the day of

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

589 inoculation, seedlings were transplanted to pots (diameter 9 cm, height 7 cm) filled with 200 g of
590 soil containing a tomatine aqueous solution (25 nmol/g soil); specifically, we added 80 ml tap
591 water containing 5 μmol tomatine (Tokyo Chemical Industry) to each measured quantity of 200
592 g soil. Each pot was then placed in a plastic bag to prevent water loss. Next, prepared (see above)
593 SynCom bacterial suspension solutions were inoculated onto the seedlings. The inoculation was
594 performed such that 12.5 μL aqueous solution of each SynCom (OD_{600} = 0.5) was included in 1
595 gram of soil. Seedlings were then grown for 2 weeks of acclimatization in the greenhouse in early
596 summer (i.e., June 24–July 8, 2022) before being moved to a field site at the Tokyo University of
597 Agriculture and Technology, Fuchu, Tokyo, Japan (35°68'54"N, 139°48'04"E). Each plot was one
598 square meter in size, and contained four plants of the same treatment. Plots were separated by 50
599 cm of empty space, and four plots were created for each SynCom group, including the control. In
600 total, 96 plants were cultivated for one month in the field. Fresh weights of aerial/aboveground
601 and root portions of plants were measured immediately after sampling. Plant dry weight was then
602 estimated after oven-drying at 80°C until reaching a constant weight.
603
604 *Transcriptomic analysis of SynCom-inoculated tomato plants*
605 Fully-expanded upper leaves and roots were sampled from SynCom-inoculated tomato plants,
606 then were ground in liquid nitrogen using a mortar and pestle. Total RNA was extracted using

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

607 RNeasy Plant Mini Kits (QIAGEN) and a RNase-Free DNase Set (QIAGEN), with all procedures
608 following the manufacturer's protocol. Next, RNA-seq libraries were constructed using the Lasy-
609 Seq version 1.1 method [85]. Finally, prepared libraries were sequenced on a HiSeqX (Illumina)
610 platform to obtain 2 × 150-bp paired-end read datasets.

611 Raw reads were first subjected to quality control by removing low-quality bases using
612 Trimmomatic [86] using the default parameters. Trimmed reads were then aligned to the tomato
613 genome (Solanum_lycopersicum. SL3.0) using STAR version 2.7.6a [87] and the Ensembl Plants
614 release 50 [88] gene annotation framework. Gene expression levels were estimated as transcripts per
615 million (TPM) using RSEM version 1.3.3 [89] using the default parameters.

616 All statistical analyses were performed using R. Of the 25,366 genes detected, 14,948 genes
617 in the leaf and 16,727 genes in the root were kept for downstream analysis following the exclusion
618 of low expression genes (i.e., mean TPM < 1). The *DESeq2* R package was used for differential
619 expression analysis [90]. GO annotation of tomato genes was obtained from Ensembl Plants using
620 the *biomaRt* R package [91], and enrichment analysis was performed using the *GOstats* and
621 *GSEABase* R packages [92]. Finally, gene expression patterns were clustered and visualized using
622 the *pheatmap* R package.

623

624 *Integrated multiomics analyses*

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Sparse multiple canonical correlation analysis was performed on plant growth (i.e., shoot and root biomass) data, gene expression profiles (i.e., TPM value, including 14,948 genes from leaves and 16,727 genes from roots), and root endophytic bacteria profiles (CLR value, 361 taxa at genus level), using the *PMA* R package [93]. Parameter tuning was performed using 100 permutations and five iterations. Based on the sparse canonical variates of the first and second components, we extracted 4,938 genes in the leaves and 7,089 genes in the roots, representing 176 taxa.

Next, a multi-omics co-occurrence network was created using WGCNA [94] using a previously described protocol with some modification [59]. Briefly, a signed network was created using the Spearman correlation matrix as well as the soft thresholding power selected from a scale-free topology fit index R-squared value of 0.75. Module identification was then performed using a dynamic tree cut based on a hierarchical clustering tree. Network connections with a topological overlap above a threshold of 0.08 were extracted, and the resulting subnetwork was visualized using the *igraph* R package [95].

Data availability

The genome assemblies of SynCom member bacteria were deposited in DDBJ/ENA/GenBank under accession numbers AP027908–AP027921. Raw sequence data for genome analysis were deposited at the DDBJ Sequence Read Archive (DRA) under accession numbers DRR453250–

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

643 DRR453257 (BioProject accession number PRJDB15212). Raw data describing SynCom
644 composition and plant growth used for modeling are provided in Supplemental Table S3.

645

646 **Acknowledgments**

647 This work was supported by JST CREST (Grant Number JPMJCR15O2 and JPMJCR17O2) and
648 GteX Program (JPMJGX23B2), and RIKEN TRIP initiative to K.S., Japan.

649

650 **Figure legends**

651 **Figure 1.** Overview of this study. **A.** Data collection. Endophytic bacteria were isolated from
652 tomato roots, then various SynCom compositions were prepared and inoculated into tomato plants
653 to examine their effects on plant growth. We also performed genome analysis of isolated
654 endophytic bacteria and microbiome analysis of SynCom-inoculated tomato plants. **B.** Modeling
655 and application. Based on data collected, we created three models to predict plant growth. The
656 first model used the SynCom bacterial composition as an explanatory variable, the second model
657 used the SynCom genetic composition, and the third model used the actual microbiome
658 composition. New SynComs were then designed by constructing predictive models; these were
659 then validated in laboratory and field cultivation experiments. **C.** Multi-omics analysis of
660 SynCom-inoculated tomato plants. Phenotype, transcriptome, and microbiome data of inoculated

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

661 tomato plants were obtained for newly-designed “Good” and “Bad” SynCom to validate their
662 predicted effects on plant growth, changes in gene expression, and effects on microbial
663 community structure. We also estimated the important genetic and microbial factors related to
664 tomato growth using integrated multi-omics network analyses.
665
666 **Figure 2.** Phylogenetic and genetic characteristics of bacterial isolates from the tomato root
667 endophytic community. **A.** Taxonomic tree of bacterial isolates from the tomato root endophytic
668 community. Each isolate was assigned to a genus-level taxon by BLAST query of partial 16S
669 rRNA sequences. The taxonomic hierarchical tree was then drawn using GraPhlAn. Strain names
670 of selected SynCom members are indicated in the tree. Circular heatmaps outside the tree show
671 differences in relative abundance of family-level taxa between bulk soil and rhizosphere or bulk
672 soil and root endosphere, as indicated in previous data. The effect sizes estimated by ALDEx2 are
673 shown as differences. Red indicates enrichment in the rhizosphere or root endosphere. **B.** Profiles
674 of plant growth-promoting (PGP) trait-related genes carried by SynCom members. Genes for PGP
675 trait-related pathways were determined by querying the MetaCyc and KEGG databases, and their
676 presence in each genome was estimated.
677
678 **Figure 3.** Microbiome analysis of SynCom-inoculated tomato roots used for model construction.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

679 **A.** Effect of metabolite addition on RC1 colonization on roots. Asterisk (*) indicates a significant
680 difference in group means ($p < 0.05$, Wilcoxon rank-sum test). **B.** Effect of growing temperature
681 and metabolite addition on the alpha diversity of root-associated bacterial communities. Different
682 letters indicate significant differences between their mean values ($p < 0.05$, Wilcoxon rank-sum
683 test with Bonferroni correction).
684
685 **Figure 4**. Summary of model performance. **A.** Effects of microbial, chemical, and environmental
686 factors on plant growth prediction using the S-I model. M, microbial factor; C, chemical factor;
687 E, environmental factor; X, interaction of two variables. **B.** Comparison of the predictive
688 performance of the S-I, S-G, and MB models. Splitting of training and test data (i.e., in a 7:3 ratio)
689 was randomly repeated 50 times to evaluate average model performance. Different letters indicate
690 significant differences between group means ($p < 0.05$, Welch's t-test with Bonferroni correction).
691 **C.** Verification of the versatility of the S-I and S-G models. Correlation between predicted and
692 measured values of Good (top 5%), Neutral (median) and Bad (bottom 5%) SynComs designed
693 using the S-I model.
694
695 **Figure 5.** Feature importance of the SynCom-Inoculant model. **A.** Distribution of regression
696 coefficients for the S-I model. Here, coefficients are ranked in ascending order, and the

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

697 distribution and median of rankings over 50 iterations are shown in box plots and a histogram. **B.**
698 Top 15 positive important variables for tomato growth and their effect sizes. They are all two
699 variable interactions, and the heatmap shows the effect sizes of single or double variables on plant
700 growth. Values indicate logarithms of median fold change relative to the minus:minus group.
701 Asterisks (*) indicate significant difference relative to the minus:minus group ($p < 0.05$, rank-
702 based Dunnett test). Var, variable; FC, fold change. **C.** Plant growth enhancement by microbe-
703 microbe (T2-2 and RC1 strains) and microbe-chemical (RC1 and tomatine) interactions.
704
705 **Figure 6.** Validation of the predicted beneficial effect of Good SynCom on tomato plant growth.
706 **A.** Field growth of Good SynCom G1-G5-inoculated tomato plants. Orange empty circles
707 represent predicted values for the S-I model. The field was divided into four blocks, including
708 four plots per treatment and four plants per plot ($n = 16$). Asterisk (*) indicates a significant
709 difference relative to the uninoculated control ($p < 0.05$, Dunnett test). **B.** Representative growth
710 of SynCom G2- and B5-inoculated tomato plants in the laboratory experiment. **C, D.** Shoot and
711 root biomass of SynCom G2- and B5-inoculated tomato plants in the laboratory experiment.
712 SynCom-inoculated tomato plants were grown under optimum (30°C) and high temperature
713 (38°C) conditions ($n = 5$). Asterisks (*) indicate a significant difference relative to an uninoculated
714 control ($p < 0.05$, Dunnett test).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

715

716 **Figure 7.** Expression pattern clustering and GO-term enrichment of differentially expressed genes
717 (DEGs) responsive to SynCom inoculation. DEGs related to SynCom G2 and B5 inoculation were
718 identified for each tissue and temperature (FDR < 0.01, DESeq2). Next, these DEGs were
719 clustered using the k-means method. The optimal k was determined based on the Bayesian
720 information criterion. Representative gene expression patterns within each cluster were revealed
721 by the heatmap. GO enrichment analysis was performed for each gene cluster (FDR < 0.05,
722 hypergeometric test).

723

724 **Figure 8.** Expression profiles of HSF and HSP genes. Among all DEGs identified following
725 SynCom inoculation, those annotated with the gene symbol “HSF” and the GO-term “protein
726 folding” were extracted and their expression patterns are shown in the heatmap. Asterisks (*)
727 indicate significant differences relative to an uninoculated control (FDR < 0.01, DESeq2).

728

729 **Figure 9.** Root-associated microbial community structure following SynCom inoculation. **A.**
730 Principal coordinate analysis (PCoA) plot of Bray–Curtis distance among samples. **B.** Shannon
731 index representing the alpha diversity of the microbial community. Different letters indicate a
732 significant difference between SynCom group means ($p < 0.05$, Welch’s t-test with Bonferroni

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

733 correction; n = 3 for G2 and B5 at 30ºC, n = 4 for the others). **C.** Relative abundance of the
734 dominant top 15 taxa within the community resolved to the genus level. Values indicate the means
735 of 3-4 biological replicates. **D.** Differentially abundant taxa on the genus level by SynCom
736 inoculation or high temperature status. An ALDEx2 generalized linear model (aldex.glm) was
737 used to test the effect of each factor, and effect sizes are shown by the heatmap. Asterisks (*)
738 indicate significant differences relative to the control (FDR < 0.1, ALDEx2).
739
740 **Figure 10.** Plant biomass-associated modules in a multi-omics co-occurrence network. **A.** Sub-
741 network of modules related to plant biomass, created using WGCNA. Nodes represent tomato
742 genes, bacterial taxa on the genus level, and plant phenotypic traits. Edges were drawn based on
743 the topological overlap matrix with a threshold of 0.08. **B.** Correlations between plant biomass
744 and relative abundance of *Sphingobium* sp. as well as between plant biomass and the expression
745 levels of the hub gene Solyc05g053080. SCC, Spearman correlation coefficient; FW, fresh weight.
746
747

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

748 **References**

749 1. Mahmud, K. *et al.* Rhizosphere microbiome manipulation for sustainable crop
750 production. *Current Plant Biology* **27**, 100210 (2021).
751 2. de Vries, F. T., Griffiths, R. I., Knight, C. G., Nicolitch, O. & Williams, A.
752 Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* **368**,
753 270–274 (2020).
754 3. Babalola, O. O. *et al.* Rhizosphere Microbiome Cooperations: Strategies for
755 Sustainable Crop Production. *Curr Microbiol* **78**, 1069–1085 (2021).
756 4. Osayande, I. S., Han, X. & Tsuda, K. Dynamic shifts in plant-microbe
757 relationships. *Plant Biotechnology* **42**, 271–277 (2025).
758 5. Jones, D. L., Nguyen, C. & Finlay, R. D. Carbon flow in the rhizosphere: carbon
759 trading at the soil–root interface. *Plant Soil* **321**, 5–33 (2009).
760 6. Pausch, J. & Kuzyakov, Y. Carbon input by roots into the soil: Quantification of
761 rhizodeposition from root to ecosystem scale. *Glob Change Biol* **24**, 1–12 (2018).
762 7. Korenblum, E., Massalha, H. & Aharoni, A. Plant–microbe interactions in the
763 rhizosphere via a circular metabolic economy. *The Plant Cell* **34**, 3168–3182 (2022).
764 8. Hemelda, N. M. & Noutoshi, Y. Root-exuded sugars as drivers of rhizosphere
765 microbiome assembly. *Plant Biotechnology* **42**, 215–227 (2025).
766 9. Nakayasu, M. *et al.* Tomato roots secrete tomatine to modulate the bacterial
767 assemblage of the rhizosphere. *Plant Physiology* **186**, 270–284 (2021).
768 10. Nakayasu, M. *et al.* Tomato root-associated *Sphingobium* harbors genes for
769 catabolizing toxic steroidal glycoalkaloids. *mBio* **14**, e00599-23 (2023).
770 11. Okutani, F. *et al.* Rhizosphere modelling reveals spatiotemporal distribution of
771 daidzein shaping soybean rhizosphere bacterial community. *Plant Cell Environ* **43**, 1036–
772 1046 (2020).
773 12. Yu, P. *et al.* Plant flavones enrich rhizosphere Oxalobacteraceae to improve
774 maize performance under nitrogen deprivation. *Nat. Plants* **7**, 481–499 (2021).
775 13. Koprivova, A. *et al.* Root-specific camalexin biosynthesis controls the plant
776 growth-promoting effects of multiple bacterial strains. *Proc. Natl. Acad. Sci. U.S.A.* **116**,
777 15735–15744 (2019).
778 14. Stringlis, I. A. *et al.* MYB72-dependent coumarin exudation shapes root
779 microbiome assembly to promote plant health. *Proc. Natl. Acad. Sci. U.S.A.* **115**, (2018).
780 15. Hu, L. *et al.* Root exudate metabolites drive plant-soil feedbacks on growth and
781 defense by shaping the rhizosphere microbiota. *Nat Commun* **9**, (2018).
782 16. Cotton, T. E. A. *et al.* Metabolic regulation of the maize rhizobiome by

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

783 benzoxazinoids. *ISME J* **13**, 1647–1658 (2019).
784 17. Oppenheimer-Shaanan, Y. *et al.* A dynamic rhizosphere interplay between tree
785 roots and soil bacteria under drought stress. *eLife* **11**, e79679 (2022).
786 18. Li, G. *et al.* Integrated Microbiome and Metabolomic Analysis Reveal
787 Responses of Rhizosphere Bacterial Communities and Root exudate Composition to
788 Drought and Genotype in Rice (Oryza sativa L.). *Rice* **16**, 19 (2023).
789 19. Kong, Z. & Liu, H. Modification of Rhizosphere Microbial Communities: A
790 Possible Mechanism of Plant Growth Promoting Rhizobacteria Enhancing Plant Growth
791 and Fitness. *Frontiers in Plant Science* **13**, (2022).
792 20. Souza, R. D., Ambrosini, A. & Passaglia, L. M. P. Plant growth-promoting
793 bacteria as inoculants in agricultural soils. *Genet. Mol. Biol.* **38**, 401–419 (2015).
794 21. Kumar, R., Swapnil, P., Meena, M., Selpair, S. & Yadav, B. G. Plant Growth-
795 Promoting Rhizobacteria (PGPR): Approaches to Alleviate Abiotic Stresses for
796 Enhancement of Growth and Development of Medicinal Plants. *Sustainability* **14**, 15514
797 (2022).
798 22. Ansabayeva, A. *et al.* Plant Growth-Promoting Microbes for Resilient Farming
799 Systems: Mitigating Environmental Stressors and Boosting Crops Productivity—A
800 Review. *Horticulturae* **11**, 260 (2025).
801 23. Tosi, M., Mitter, E. K., Gaiero, J. & Dunfield, K. It takes three to tango: the
802 importance of microbes, host plant, and soil management to elucidate manipulation
803 strategies for the plant microbiome. *Can. J. Microbiol.* **66**, 413–433 (2020).
804 24. Bano, S., Wu, X. & Zhang, X. Towards sustainable agriculture: rhizosphere
805 microbiome engineering. *Appl Microbiol Biotechnol* **105**, 7141–7160 (2021).
806 25. Thomas, P. & Sekhar, A. C. Effects Due to Rhizospheric Soil Application of an
807 Antagonistic Bacterial Endophyte on Native Bacterial Community and Its Survival in
808 Soil: A Case Study with Pseudomonas aeruginosa from Banana. *Front. Microbiol.* **7**,
809 (2016).
810 26. Shayanthan, A., Ordoñez, P. A. C. & Oresnik, I. J. The Role of Synthetic
811 Microbial Communities (SynCom) in Sustainable Agriculture. *Front. Agron.* **4**, 896307
812 (2022).
813 27. Martins, S. J. *et al.* The Use of Synthetic Microbial Communities to Improve
814 Plant Health. *Phytopathology®* **113**, 1369–1379 (2023).
815 28. Castrillo, G. *et al.* Root microbiota drive direct integration of phosphate stress
816 and immunity. *Nature* **543**, 513–518 (2017).
817 29. Niu, B., Paulson, J. N., Zheng, X. & Kolter, R. Simplified and representative
818 bacterial community of maize roots. *Proc. Natl. Acad. Sci. U.S.A.* **114**, (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

30. Li, Z. *et al.* A simplified synthetic community rescues Astragalus mongholicus from root rot disease by activating plant-induced systemic resistance. *Microbiome* **9**, 217 (2021).
31. Wilhelm, R. C., Es, H. M. van & Buckley, D. H. Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biology and Biochemistry* **164**, 108472 (2022).
32. Lewis, J. E. & Kemp, M. L. Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. *Nat Commun* **12**, 2700 (2021).
33. Wu, L.-D. *et al.* Analysis of potential genetic biomarkers using machine learning methods and immune infiltration regulatory mechanisms underlying atrial fibrillation. *BMC Med Genomics* **15**, 64 (2022).
34. Shimasaki, T. *et al.* Tobacco Root Endophytic *Arthrobacter* Harbors Genomic Features Enabling the Catabolism of Host-Specific Plant Specialized Metabolites. *mBio* **12**, (2021).
35. Thoenen, L. *et al.* Bacterial tolerance to host-exuded specialized metabolites structures the maize root microbiome. *Proc. Natl. Acad. Sci. U.S.A.* **120**, (2023).
36. Thoenen, L. *et al.* The lactonase BxdA mediates metabolic specialisation of maize root bacteria to benzoxazinoids. *Nat Commun* **15**, (2024).
37. Aoki, N. *et al.* An isoflavone catabolism gene cluster underlying interkingdom interactions in the soybean rhizosphere. *ISME Communications* **4**, (2024).
38. Shahid, M. *et al.* Bacterial ACC deaminase: Insights into enzymology, biochemistry, genetics, and potential role in amelioration of environmental stress in crop plants. *Front. Microbiol.* **14**, 1132770 (2023).
39. Suleimanova, A. *et al.* Phosphate Solubilization and Plant Growth Promotion by Pantoea brenneri Soil Isolates. *Microorganisms* **11**, 1136 (2023).
40. Yu, J. S. L. *et al.* Microbial communities form rich extracellular metabolomes that foster metabolic interactions and promote drug tolerance. *Nat Microbiol* **7**, 542–555 (2022).
41. Morin, M., Pierce, E. C. & Dutton, R. J. Changes in the genetic requirements for microbial interactions with increasing community complexity. *eLife* **7**, (2018).
42. Colarusso, A. V., Goodchild-Michelman, I., Rayle, M. & Zomorrodi, A. R. Computational modeling of metabolism in microbial communities on a genome-scale. *Current Opinion in Systems Biology* **26**, 46–57 (2021).
43. Dillard, L. R., Payne, D. D. & Papin, J. A. Mechanistic models of microbial community metabolism. *Mol. Omics* **17**, 365–375 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

855 44. Iqbal, A. & Hasnain, S. Aeromonas punctata PNS-1: a promising candidate to
856 change the root morphogenesis of Arabidopsis thaliana in MS and sand system. *Acta*
857 *Physiol Plant* **35**, 657–665 (2013).
858 45. Spaepen, S., Bossuyt, S., Engelen, K., Marchal, K. & Vanderleyden, J.
859 Phenotypical and molecular responses of *A rabidopsis thaliana* roots as a result of
860 inoculation with the auxin-producing bacterium *A zospirillum brasilense* . *New*
861 *Phytologist* **201**, 850–861 (2014).
862 46. Tsukanova, K. A., Chebotar, V. K., Meyer, J. J. M. & Bibikova, T. N. Effect of
863 plant growth-promoting Rhizobacteria on plant hormone homeostasis. *South African*
864 *Journal of Botany* **113**, 91–102 (2017).
865 47. Bal, H. B., Nayak, L., Das, S. & Adhya, T. K. Isolation of ACC deaminase
866 producing PGPR from rice rhizosphere and evaluating their plant growth promoting
867 activity under salt stress. *Plant Soil* **366**, 93–105 (2013).
868 48. Chatterjee, P., Samaddar, S., Niinemets, Ü. & Sa, T.-M. Brevibacterium linens
869 RS16 confers salt tolerance to Oryza sativa genotypes by regulating antioxidant defense
870 and H+ ATPase activity. *Microbiological Research* **215**, 89–101 (2018).
871 49. Yoolong, S. *et al.* Modulation of salt tolerance in Thai jasmine rice (Oryza sativa
872 L. cv. KDML105) by Streptomyces venezuelae ATCC 10712 expressing ACC deaminase.
873 *Sci Rep* **9**, 1275 (2019).
874 50. Choi, J., Roy Choudhury, A., Walitang, D. I., Lee, Y. & Sa, T. ACC deaminase-
875 producing Brevibacterium linens RS16 enhances heat-stress tolerance of rice (Oryza
876 sativa L.). *Physiologia Plantarum* **174**, e13584 (2022).
877 51. Khan, M. A. *et al.* Thermotolerance effect of plant growth-promoting Bacillus
878 cereus SA1 on soybean during heat stress. *BMC Microbiol* **20**, 175 (2020).
879 52. Hoshikawa, K., Pham, D., Ezura, H., Schafleitner, R. & Nakashima, K. Genetic
880 and Molecular Mechanisms Conferring Heat Stress Tolerance in Tomato Plants. *Front.*
881 *Plant Sci.* **12**, 786688 (2021).
882 53. Haslbeck, M. & Vierling, E. A First Line of Stress Defense: Small Heat Shock
883 Proteins and Their Function in Protein Homeostasis. *Journal of Molecular Biology* **427**,
884 1537–1548 (2015).
885 54. Lee, J., Hanh Nguyen, H., Park, Y., Lin, J. & Hwang, I. Spatial regulation of
886 RBOHD via AtECA4-mediated recycling and clathrin-mediated endocytosis contributes
887 to ROS accumulation during salt stress response but not flg22-induced immune response.
888 *The Plant Journal* **109**, 816–830 (2022).
889 55. Miller, G. *et al.* The Plant NADPH Oxidase RBOHD Mediates Rapid Systemic
890 Signaling in Response to Diverse Stimuli. *Sci. Signal.* **2**, (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

891 56. Kadota, Y., Shirasu, K. & Zipfel, C. Regulation of the NADPH Oxidase RBOHD
892 During Plant Immunity. *Plant Cell Physiol* **56**, 1472–1480 (2015).
893 57. Wang, L. *et al.* Hydrogen Peroxide Acts Upstream of Nitric Oxide in the Heat
894 Shock Pathway in Arabidopsis Seedlings. *Plant Physiol.* **164**, 2184–2196 (2014).
895 58. Fortunato, S., Lasorella, C., Dipierro, N., Vita, F. & De Pinto, M. C. Redox
896 Signaling in Plant Heat Stress Response. *Antioxidants* **12**, 605 (2023).
897 59. Yamazaki, S. *et al.* Field multi-omics analysis reveals a close association
898 between bacterial communities and mineral properties in the soybean rhizosphere. *Sci*
899 *Rep* **11**, (2021).
900 60. Bai, Y. *et al.* Functional overlap of the Arabidopsis leaf and root microbiota.
901 *Nature* **528**, 364–369 (2015).
902 61. Hahn, M. & Hennecke, H. Localized mutagenesis in Rhizobium japonicum.
903 *Molec. Gen. Genet.* **193**, 46–52 (1984).
904 62. Hunt, M. *et al.* Circlator: automated circularization of genome assemblies using
905 long sequencing reads. *Genome Biol* **16**, (2015).
906 63. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2:
907 memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**,
908 5315–5316 (2022).
909 64. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**,
910 2068–2069 (2014).
911 65. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and
912 phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.
913 *Nucleic Acids Research* **47**, D309–D314 (2019).
914 66. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas,
915 J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
916 Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829
917 (2021).
918 67. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and
919 the BioCyc collection of Pathway/Genome Databases. *Nucl. Acids Res.* **42**, D459–D471
920 (2014).
921 68. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M.
922 KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*
923 **53**, D672–D677 (2025).
924 69. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for
925 comparative genomics. *Genome Biol* **20**, (2019).
926 70. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

927 Practical innovations for high-throughput amplicon sequencing. *Nat Methods* **10**, 999–
928 1002 (2013).
929 71. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome
930 data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).
931 72. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina
932 amplicon data. *Nat Methods* **13**, 581–583 (2016).
933 73. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene
934 amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 1–17
935 (2018).
936 74. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved
937 data processing and web-based tools. *Nucleic Acids Research* **41**, 590–596 (2013).
938 75. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software
939 Version 7: Improvements in Performance and Usability. *Molecular Biology and*
940 *Evolution* **30**, 772–780 (2013).
941 76. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-
942 Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
943 77. Douglas, G. M. *et al.* PICRUSt2 for prediction of metagenome functions. *Nature*
944 *Biotechnology* **38**, 685–688 (2020).
945 78. Oksanen, J. *et al.* Vegan: community ecology package. Ordination methods,
946 diversity analysis and other functions for community and vegetation ecologists. *R*
947 *package ver* 2–3 (2015).
948 79. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B.
949 ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-
950 Seq. *PLoS ONE* **8**, e67019 (2013).
951 80. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing
952 datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth
953 experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
954 81. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing
955 classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
956 82. Kuhn, M. Building Predictive Models in*R*Using the**caret**Package. *J. Stat. Soft.*
957 **28**, (2008).
958 83. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized
959 Linear Models via Coordinate Descent. *J. Stat. Soft.* **33**, (2010).
960 84. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects
961 Models Using**lme4**. *J. Stat. Soft.* **67**, (2015).
962 85. Kamitani, M., Kashima, M., Tezuka, A. & Nagano, A. J. Lasy-Seq: a high-

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

963 throughput library preparation method for RNA-Seq and its application in the analysis of
964 plant responses to fluctuating temperatures. *Sci Rep* **9**, (2019).
965 86. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
966 Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
967 87. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
968 15–21 (2013).
969 88. Yates, A. D. *et al.* Ensembl Genomes 2022: an expanding genome resource for
970 non-vertebrates. *Nucleic Acids Research* **50**, D996–D1003 (2022).
971 89. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq
972 data with or without a reference genome. (2011).
973 90. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
974 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).
975 91. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the
976 integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**,
977 1184–1191 (2009).
978 92. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term
979 association. *Bioinformatics* **23**, 257–258 (2007).
980 93. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition,
981 with applications to sparse principal components and canonical correlation analysis.
982 *Biostatistics* **10**, 515–534 (2009).
983 94. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation
984 network analysis. *BMC Bioinformatics* **9**, (2008).
985 95. Csárdi, G. & Nepusz, T. The igraph software package for complex network
986 research. in (2006).

987

988

989

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

Main figures

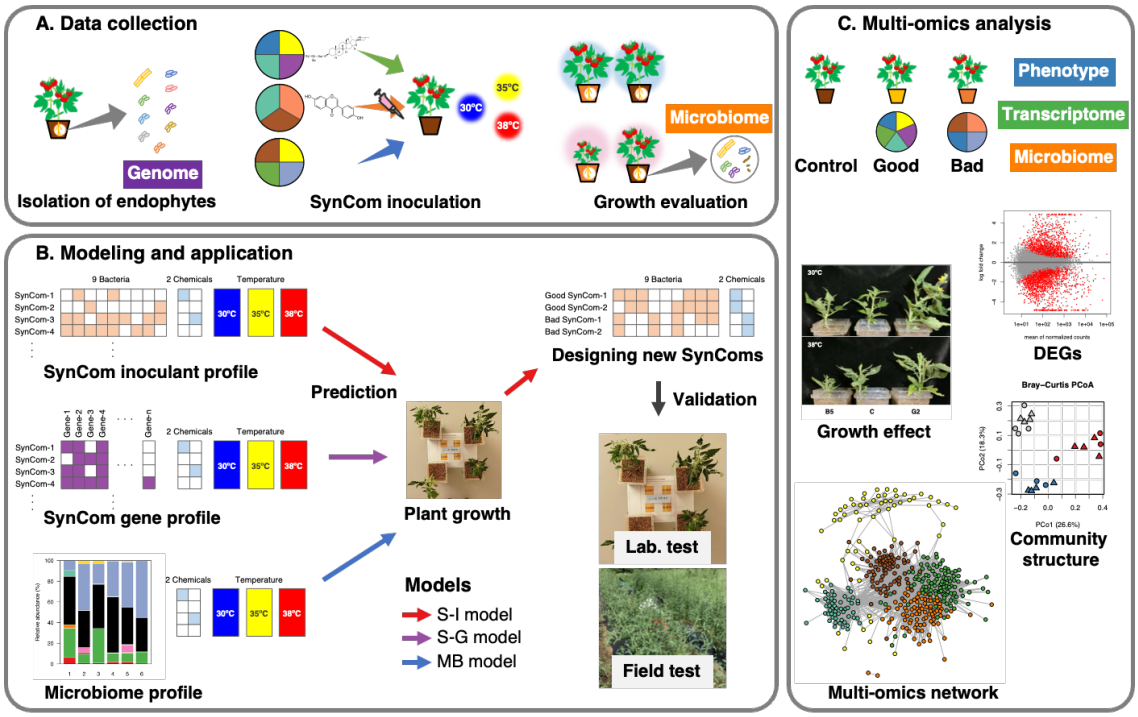

Figure 1. Overview of this study. **A.** Data collection. Endophytic bacteria were isolated from tomato roots, then various SynCom compositions were prepared and inoculated into tomato plants to examine their effects on plant growth. We also performed genome analysis of isolated endophytic bacteria and microbiome analysis of SynCom-inoculated tomato plants. **B.** Modeling and application. Based on data collected, we created three models to predict plant growth. The first model used the SynCom bacterial composition as an explanatory variable, the second model used the SynCom genetic composition, and the third model used the actual microbiome composition. New SynComs were then designed by constructing predictive models; these were then validated in laboratory and field cultivation experiments. **C.** Multi-omics analysis of SynCom-inoculated tomato plants. Phenotype, transcriptome, and microbiome data of inoculated tomato plants were obtained for newly-designed "Good" and "Bad" SynCom to validate their predicted effects on plant growth, changes in gene expression, and effects on microbial community structure. We also estimated the important genetic and microbial factors related to tomato growth using integrated multi-omics network analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

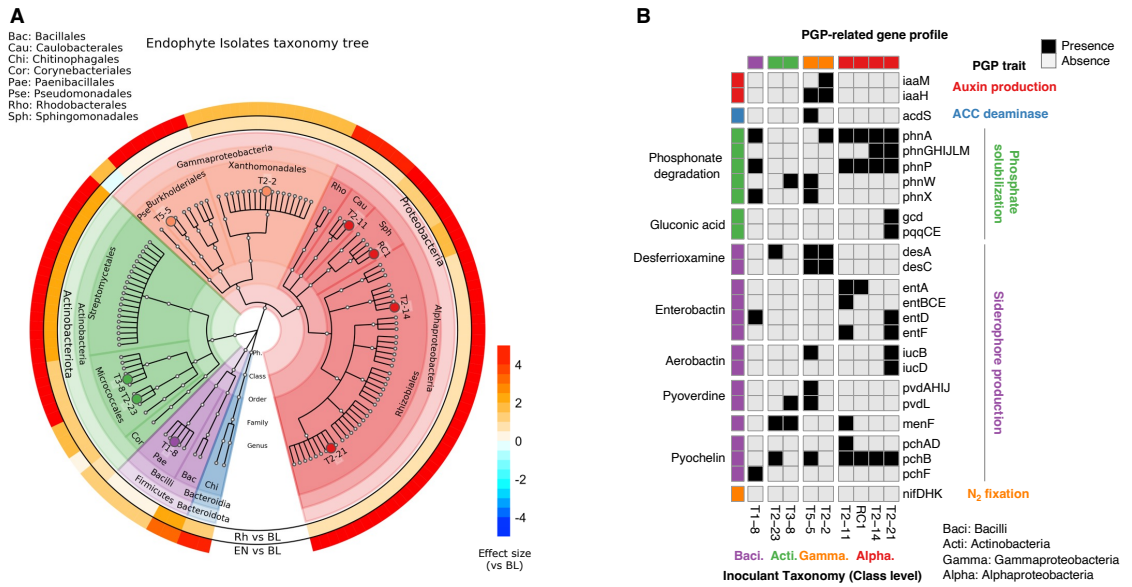

Figure 2. Phylogenetic and genetic characteristics of bacterial isolates from the tomato root endophytic community. **A.** Taxonomic tree of bacterial isolates from the tomato root endophytic community. Each isolate was assigned to a genus-level taxon by BLAST query of partial 16S rRNA sequences. The taxonomic hierarchical tree was then drawn using GraPhlAn. Strain names of selected SynCom members are indicated in the tree. Circular heatmaps outside the tree show differences in relative abundance of family-level taxa between bulk soil and rhizosphere or bulk soil and root endosphere, as indicated in previous data. The effect sizes estimated by ALDEx2 are shown as differences. Red indicates enrichment in the rhizosphere or root endosphere. **B.** Profiles of plant growth-promoting (PGP) trait-related genes carried by SynCom members. Genes for PGP trait-related pathways were determined by querying the MetaCyc and KEGG databases, and their presence in each genome was estimated.

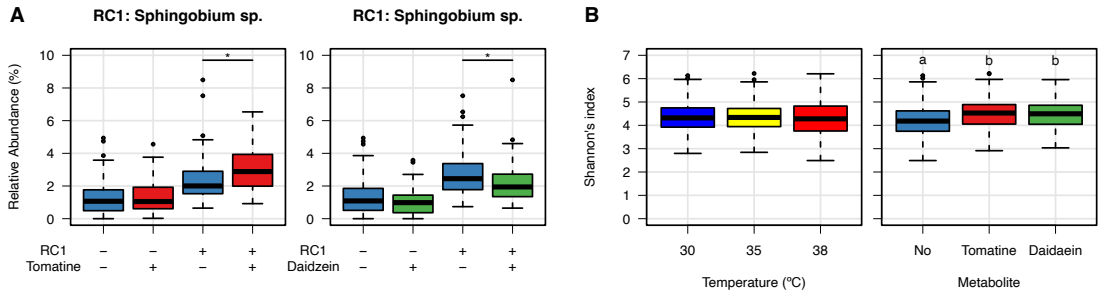

Figure 3. Microbiome analysis of SynCom-inoculated tomato roots used for model construction. **A.** Effect of metabolite addition on RC1 colonization on roots. Asterisk (*) indicates a significant difference in group means ($p < 0.05$, Wilcoxon rank-sum test). **B.** Effect of growing temperature and

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

metabolite addition on the alpha diversity of root-associated bacterial communities. Different letters indicate significant differences between their mean values ($p < 0.05$, Wilcoxon rank-sum test with Bonferroni correction).

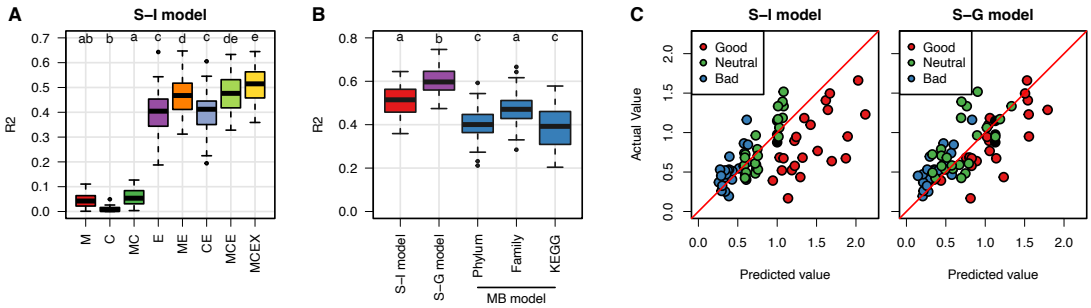

Figure 4. Summary of model performance. **A.** Effects of microbial, chemical, and environmental factors on plant growth prediction using the S-I model. M, microbial factor; C, chemical factor; E, environmental factor; X, interaction of two variables. **B.** Comparison of the predictive performance of the S-I, S-G, and MB models. Splitting of training and test data (i.e., in a 7:3 ratio) was randomly repeated 50 times to evaluate average model performance. Different letters indicate significant differences between group means ($p < 0.05$, Welch's t-test with Bonferroni correction). **C.** Verification of the versatility of the S-I and S-G models. Correlation between predicted and measured values of Good (top 5%), Neutral (median) and Bad (bottom 5%) SynComs designed using the S-I model.

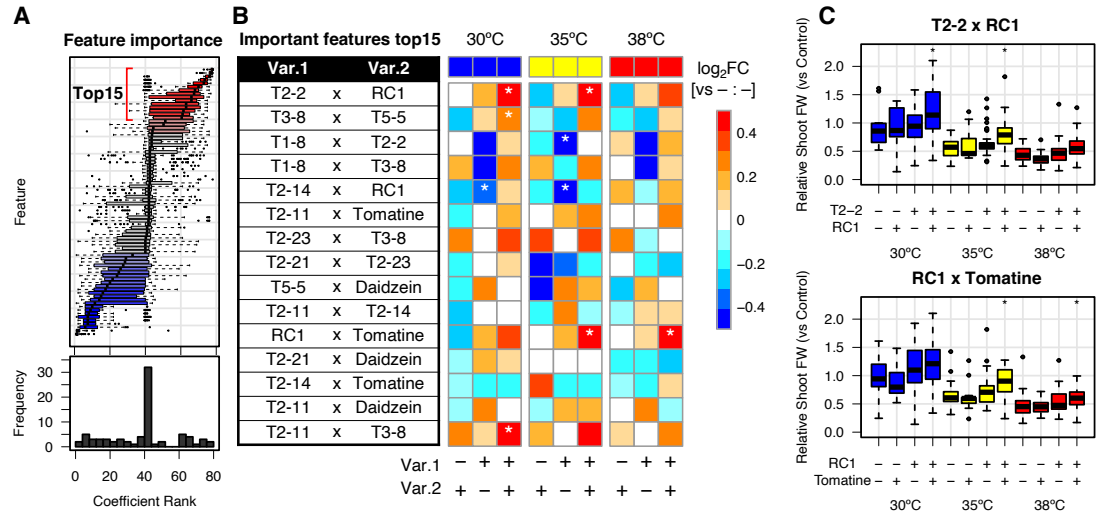

Figure 5. Feature importance of the SynCom-Inoculant model. **A.** Distribution of regression coefficients for the S-I model. Here, coefficients are ranked in ascending order, and the distribution

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

and median of rankings over 50 iterations are shown in box plots and a histogram. **B.** Top 15 positive important variables for tomato growth and their effect sizes. They are all two variable interactions, and the heatmap shows the effect sizes of single or double variables on plant growth. Values indicate logarithms of median fold change relative to the minus:minus group. Asterisks (*) indicate significant difference relative to the minus:minus group ($p < 0.05$, rank-based Dunnett test). Var, variable; FC, fold change. **C.** Plant growth enhancement by microbe-microbe (T2-2 and RC1 strains) and microbe-chemical (RC1 and tomatine) interactions.

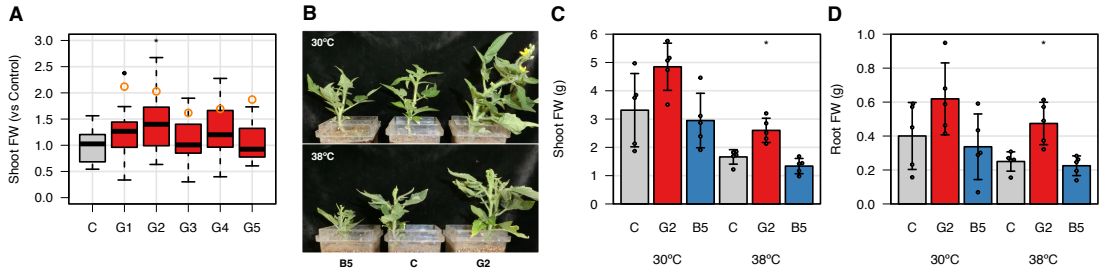

Figure 6. Validation of the predicted beneficial effect of Good SynCom on tomato plant growth. **A.** Field growth of Good SynCom G1-G5-inoculated tomato plants. Orange empty circles represent predicted values for the S-I model. The field was divided into four blocks, including four plots per treatment and four plants per plot (n = 16). Asterisk (*) indicates a significant difference relative to the uninoculated control ($p < 0.05$, Dunnett test). **B.** Representative growth of SynCom G2- and B5-inoculated tomato plants in the laboratory experiment. **C, D.** Shoot and root biomass of SynCom G2- and B5-inoculated tomato plants in the laboratory experiment. SynCom-inoculated tomato plants were grown under optimum (30°C) and high temperature (38°C) conditions (n = 5). Asterisks (*) indicate a significant difference relative to an uninoculated control ($p < 0.05$, Dunnett test).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

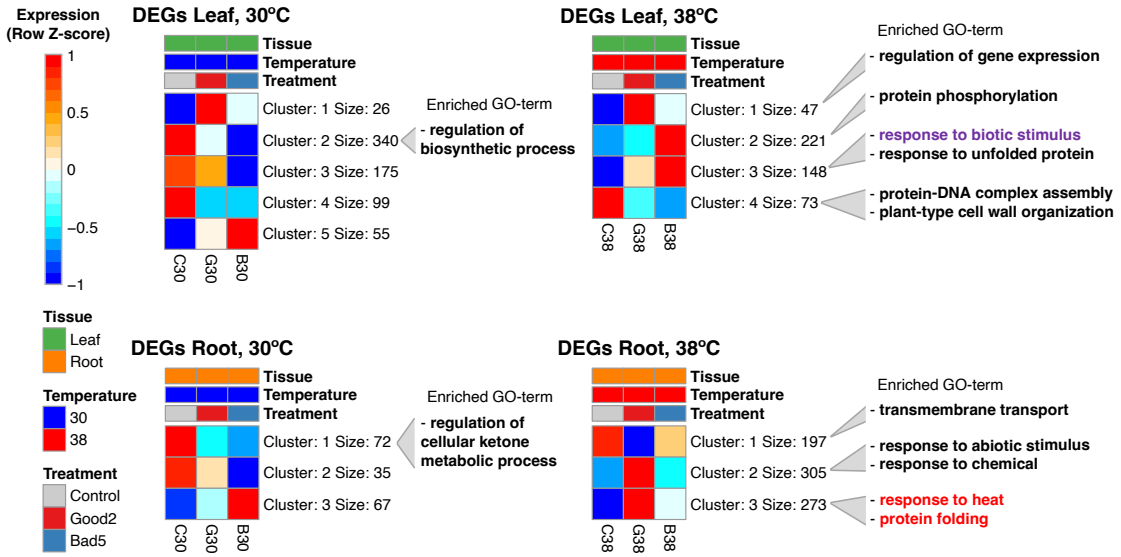

Figure 7. Expression pattern clustering and GO-term enrichment of differentially expressed genes (DEGs) responsive to SynCom inoculation. DEGs related to SynCom G2 and B5 inoculation were identified for each tissue and temperature (FDR < 0.01, DESeq2). Next, these DEGs were clustered using the k-means method. The optimal k was determined based on the Bayesian information criterion. Representative gene expression patterns within each cluster were revealed by the heatmap. GO enrichment analysis was performed for each gene cluster (FDR < 0.05, hypergeometric test).

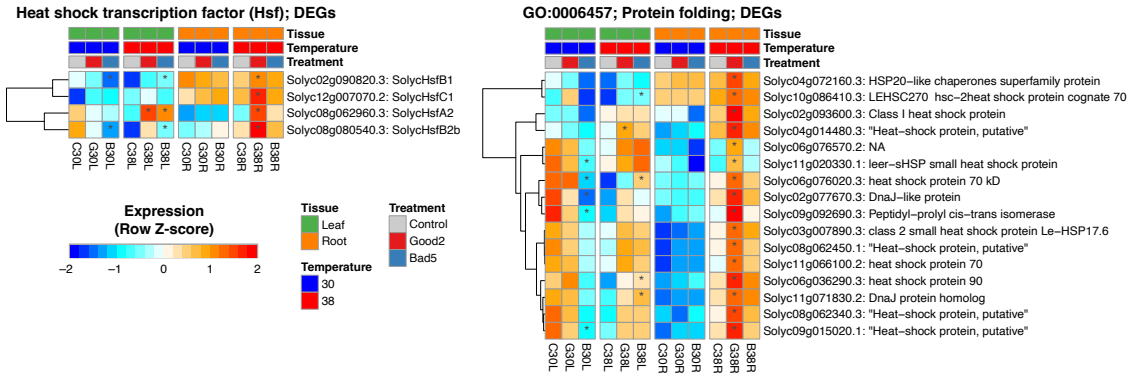

Figure 8. Expression profiles of HSF and HSP genes. Among all DEGs identified following SynCom inoculation, those annotated with the gene symbol “HSF” and the GO-term “protein folding” were extracted and their expression patterns are shown in the heatmap. Asterisks (*) indicate significant differences relative to an uninoculated control (FDR < 0.01, DESeq2).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

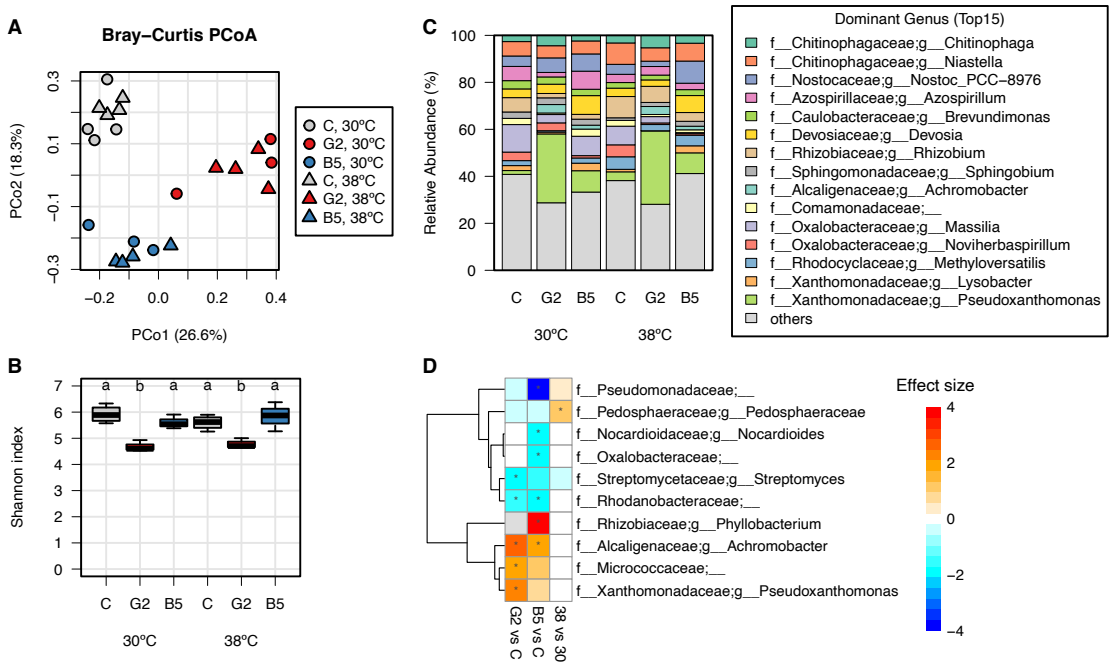

Figure 9. Root-associated microbial community structure following SynCom inoculation. **A.** Principal coordinate analysis (PCoA) plot of Bray–Curtis distance among samples. **B.** Shannon index representing the alpha diversity of the microbial community. Different letters indicate a significant difference between SynCom group means ($p < 0.05$, Welch's t-test with Bonferroni correction; n = 3 for G2 and B5 at 30ºC, n = 4 for the others). **C.** Relative abundance of the dominant top 15 taxa within the community resolved to the genus level. Values indicate the means of 3-4 biological replicates. **D.** Differentially abundant taxa on the genus level by SynCom inoculation or high temperature status. An ALDEx2 generalized linear model (aldex.glm) was used to test the effect of each factor, and effect sizes are shown by the heatmap. Asterisks (*) indicate significant differences relative to the control (FDR < 0.1, ALDEx2).

bioRxiv preprint doi: https://doi.org/10.1101/2025.10.21.683643; this version posted October 22, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

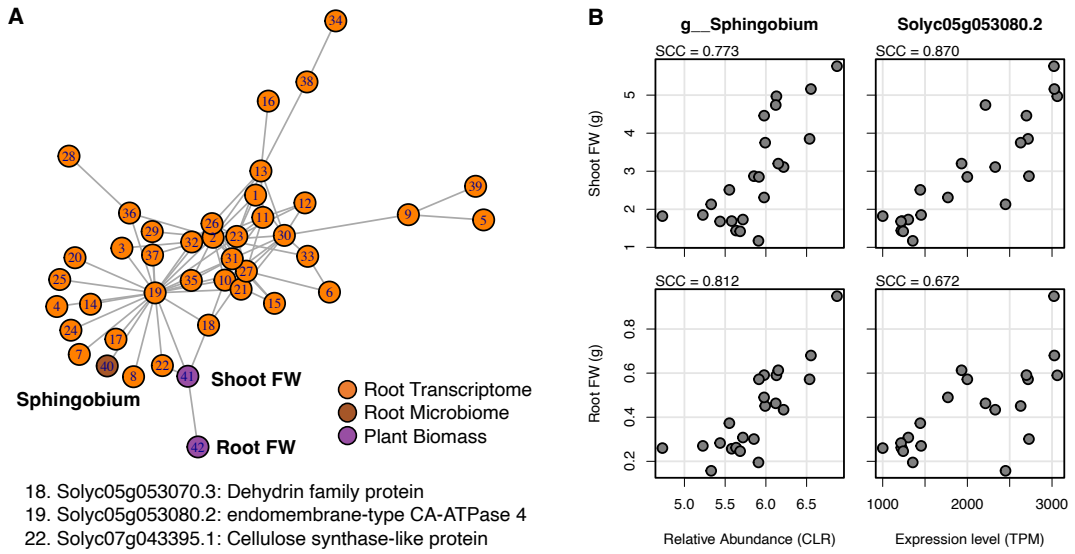

Figure 10. Plant biomass-associated modules in a multi-omics co-occurrence network. **A.** Sub-network of modules related to plant biomass, created using WGCNA. Nodes represent tomato genes, bacterial taxa on the genus level, and plant phenotypic traits. Edges were drawn based on the topological overlap matrix with a threshold of 0.08. **B.** Correlations between plant biomass and relative abundance of *Sphingobium* sp. as well as between plant biomass and the expression levels of the hub gene Solyc05g053080. SCC, Spearman correlation coefficient; FW, fresh weight.