Deciphering the Functional Landscape of Microbiome-Mediated Invasion Resistance to Plant Pathogens

Wenlong Zuo[1,*], Lanxiang Wang[1,*], Yang Bai[2], Jonathan Friedman[3], Yang-Yu Liu[4,#], Lei Dai[1,#]

[1]*State Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.*

[2]*Peking-Tsinghua Center for Life Sciences, State Key Laboratory of Gene Function and Modulation Research, Peking-Tsinghua-NIBS Graduate Program, School of Life Sciences, Peking University, Beijing 100871, China*

[3]*Institute of Environmental Sciences, Hebrew University, Rehovot, Israel.*

[4]*Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.*

* These authors contributed equally to this work.

Corresponding authors: lei.dai@siat.ac.cn, yyl@channing.harvard.edu

ABSTRACT

Rational design of microbiomes requires a fundamental understanding of the community functional landscape, i.e., the mapping between community composition to function. The functional landscapes governing microbiome-mediated invasion resistance to pathogens, which is a critical function of host-associated microbiomes, are still poorly studied *in vivo*. In this study, we systematically assessed the invasion resistance of synthetic rhizosphere microbial communities (SynComs) to two distinct pathogens (*Pseudomonas syringae* DC3000*, Pseudomonas brassicacearum* Root401) *in planta*. Our experiments with two groups of SynComs revealed that communities with higher species richness exhibit significantly stronger resistance to pathogen invasion. We found that the invasion resistance can be accurately predicted by the community compositions using a linear regression model that account for additive effects of individual strains and pairwise epistasis. In particular, we inferred and validated that the invasion resistance is determined by a few species with strong inhibitory effects. Finally, using simulations of the generalized Lotka-Volterra model, we assessed the utility of linear regression in predicting invasion resistance of complex communities. In summary, our findings systematically profiled the functional landscape of invasion resistance and provided crucial insights for designing microbial communities to improve host health.

INTRODUCTION

Understanding and predicting the functional landscape of microbial communities—how community composition correlates with functional outcomes—is crucial and challenging[1-4], as the intricate interplay between microbes, hosts, and environmental factors often obscures clear predictive relationships[5]. While fully understanding the functional landscape remains complex, recent studies have shown the potential to address this challenge through data-driven approaches[2,6-8]. This holds significant potential for applications in health, agriculture, and environmental remediation[9-12].

Invasion resistance—the capacity of resident microbial communities to prevent the establishment of external pathogens—has received widespread attention in microbiome research[13-15]. The resident communities can defend invaders by different strategies, such as secreting antimicrobial compounds[16-18], competition for nutrients and physical space[19-21], and activating host immune response[22,23]. Such multi-layered defense strategies ensure that the stability of the microbiome is maintained while safeguarding the host from pathogen invasion. Importantly, the ecological principles of invasion resistance extend across diverse host-associated and environmental microbiomes, highlighting their broad relevance in maintaining ecosystem stability and host protection[14,24,25]. For example, diversity is an important ecological indicator of a community's resistance to colonization[26-28], as antibiotics treatment often leads to a loss of diversity and promotes the invasion of exogenous species[29].

Despite these advances, previous studies on invasion resistance of complex communities were mostly conducted *in vitro*[6,25]. In comparison to the costs of animal studies, gnotobiotic plants provide an ideal model system for low-cost and high-throughput experiments of microbial communities in the context of host environments[30]. Various gnotobiotic plant growth systems have been established to explore the effects of microbial communities on hosts under different scenarios, providing valuable insights into the mechanisms of microbial community assembly[31], pathogen suppression[32-34], and plant growth promotion[35] within a natural yet controlled setting.

In this study, we systematically investigated the invasion resistance of synthetic microbial communities against invading microbes using commensal isolates from Arabidopsis (*Arabidopsis thaliana*) root and pathogen strains *in planta*. By linking community composition to pathogen invasion outcomes, we aimed to understand the functional landscape of microbial communities that govern the pathogen invasion resistance. To do so, rhizosphere bacterial strains were assembled into synthetic communities (SynComs) and pre-inoculated on *Arabidopsis* seedlings. Subsequently, two distinct plant pathogen strains were introduced, and the invasion resistance of the

SynComs was evaluated by quantifying the pathogen colonization on plants. Our results demonstrate that communities with higher species richness generally exhibit stronger resistance to pathogen invasion. We found that certain strains play a critical role in invasion resistance, with their effects exhibiting epistatic interactions that partially diminish the additive effects. Finally, we performed simulations using the generalized Lotka-Volterra (gLV) model and demonstrated that invasion resistance is largely predictable by additive contributions of individual strains.

RESULTS

Profiling the resistance to pathogen invasion in synthetic rhizosphere microbial communities

We set up an experiment to assess the ability of synthetic communities (SynComs) to protect plants from pathogen invasion in a gnotobiotic model system (see **Fig. 1**). We selected 14 bacterial isolates from the rhizosphere of *Arabidopsis* to construct the bacteria pool for assembling SynComs (see **Methods**). These strains span four major classes—Actinobacteria, Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria—representing dominant root-associated bacteria in the rhizosphere of *Arabidopsis*[36,37] (**Fig. 1A**). We designed 67 SynComs consisting of 2 to 14 strains for testing (**Fig. 1B**, **Supplementary Table 1**). These SynComs covered approximately 0.5% of all possible combinations (2^{14} - 1=16,383). This design provides a diverse yet manageable subset of communities, enabling us to effectively investigate the relationship between community composition and invasion resistance.

Our *in planta* assay includes growing *Arabidopsis* seedlings, inoculating with desired SynComs, and subsequently challenging with pathogenic bacteria (**Fig. 1C**, **Methods**). We selected the well-established model pathogen *Pseudomonas syringae* pathovar *tomato* (*Pst*) strain (hereafter referred to as *Pst* DC3000)[38,39], as the invader strain. Specifically, we pre-inoculated 10-day-old *Arabidopsis* seedlings with SynComs. After a five-day incubation period, we introduced *Pst* DC3000 to simulate pathogen invasion.

Following an additional five days of cultivation, we measured the fresh weight of the seedlings and quantified the colony-forming units (CFUs) of pathogen colonization (**Fig. 1C**). We revealed a significant negative correlation between the pathogen colonization density (CFU/g) and the plant fresh weight (**Fig. S1**), highlighting the detrimental impact of pathogen invasion on plant growth.

To quantify the protective effect of the symbiotic microbial communities, we measured the reduction in pathogen colonization CFUs as the community's invasion resistance ability. As expected, we observed that the resistance ability progressively increased with the number of strains present in the community (**Fig. 1D**). Notably, the invasion resistance across different SynComs spanned four orders of magnitude, which is consistent with observations from previous studies[34].

Predicting the invasion resistance to plant pathogen *Pst* DC3000 by LASSO regression

Given the experimentally observed functional landscape, we constructed a linear regression model[2] based on the composition of SynComs to predict pathogen invasion resistance:

$$D = \beta_0 + \sum_i \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j \,. \quad (1)$$

Here, D represents the community's resistance to pathogen invasion, measured as the $\log_{10}$ reduction in pathogen colonization (CFUs) in plants inoculated with SynComs compared to germ-free plants. $x_i = +/-1$ indicates the presence/absence of strain i in the community[40]. The coefficient β_i represents the additive effect of including strain i, while β_{ij} captures the combined effect of including both strain i and j beyond their individual contributions (analogous to pairwise epistasis in the genetic fitness landscapes). Higher-order interactions are not included in the model.

We evaluated regression models with first-order and second-order approximations. According to Eq. (1), the first-order truncated model comprises 14 parameters, excluding the intercept, whereas the second-order truncated model comprises 105 parameters. We employed the L_1-regularized regression (LASSO)[41] with leave-one-out cross-validation to estimate the regression coefficients (see **Methods**). Our results show that the linear regression model effectively captures the relationship between community membership and resistance to pathogen invasion while only 6 and 14 parameters were retained in the first-order and the second-order model, respectively. As expected, incorporating second-order terms significantly improves prediction accuracy (**Figs. 2A**,**B**). To further evaluate the performance of LASSO regression, we repeated the analyses using 10-fold cross-validation with random data splits (see **Methods**). Across these splits, the first-order model achieved $R^2 \sim 0.75$ with only 6 nonzero parameters, while the second-order model improved R^2 to around 0.9 with up to 20 parameters (**Fig. S2**).

Moreover, we examined how the size of the training dataset affects the model's prediction performance. We generated 50 test sets, each consisting of 7 data points randomly selected from the full dataset, and then randomly sampled the training sets from the remaining data (see **Methods**). As the training set size increased, R^2 on the training data decreased slightly, while prediction accuracy on the test data improved, indicating enhanced model generalizability when approximately 50 SynComs were included in the training set (**Fig. 2C**).

Quantifying the contribution of individual strains to invasion resistance

The regression model allowed us to infer the contribution of individual strains to pathogen invasion resistance. By analyzing the coefficients of the first-order terms, we identified 4 key bacterial strains—specifically, *Pseudomonas* Root562 (PS), *Agrobacterium* Root491 (AG), *Achromobacter* Root170 (AM), and *Sinorhizobium* Root278 (SI)—as primary drivers of invasion resistance (**Fig. 3A**). These strains

consistently exhibited strong contributions to the community's resistance, regardless of whether the first-order or second-order approximation was applied. To validate these findings, we conducted an independent experiment to assess the protective effects of individual strains on plants. We found that the key strains inferred by the regression model ranked as the top 4 strains to resist pathogen invasion (**Fig. 3B**). While no direct test of these four strains against *Pst* DC3000 has been reported, species from the *Pseudomonas* and *Agrobacterium genera* have been shown to inhibit pathogens via type VI secretion systems[33], niche competition[42], or host-mediated mechanisms[43].

We also utilized the experimental data of individual strains resistance to pathogen invasion as a test set to evaluate the predictive performance of the regression model. The results indicate that the ability of individual strains to inhibit pathogen invasion can be effectively predicted using second-order LASSO regression model, with R^2=0.631 (**Fig. 3C**). Additionally, we split the data based on the number of strains present in the community (see **Methods**) and assessed the predictive accuracy of LASSO regression for independent test sets of different sub-communities. The LASSO analysis showed acceptable predictive performance when 6-bacteria or 8-bacteria communities were used as independent test sets (**Fig. S3**), indicating minimal bias toward any particular training set composition.

Identifying epistatic effects in invasion resistance

Next, we focused on pairwise interactions by analyzing the second-order term parameters β_{ij} in the regression model. These parameters capture the combined effect of strains i and j on predicting the pathogen invasion resistance. Using LASSO regression with leave-one-out cross-validation, only 8 out of 91 second-order parameters were non-zero, and most of the β_{ij} values were positive. Given that invasion resistance is defined as a negative value (i.e., decrease in pathogen colonization), a positive β_{ij} indicates antagonistic interactions—where the combined presence of two strains weakens resistance relative to the additive expectation—

suggesting that such sparse antagonistic interactions are more common in shaping community-level invasion resistance (**Fig. S4A**). Notably, these significant epistatic effects were predominantly associated with *Pseudomonas* Root562 and *Achromobacter* Root170. It is worth noting that the contributions of these strains to pathogen suppression (reflected in the first-order coefficients) changed significantly when second-order terms were included (**Fig. 3A**), underscoring the importance of accounting for epistatic interactions when predicting community function.

To further evaluate the combined effects of strains pairs, epistatic effect term δ_{ij} between two strains in the community to evaluate whether their combined effect deviates from a simple additive model[44]. We found that the δ_{ij} predicted by the model is directly related to β_{ij} (see **Methods, Fig. 3D, Fig. S4B**):

$$\delta_{ij} = 4\beta_{ij}. \quad (2)$$

In experimental measurement, the epistatic effects can be defined as:

$$\delta_{ij} = D_{ij} - (D_i + D_j), \quad (3)$$

where D_{ij} is the invasion resistance of the community containing strains i and j, D_i and D_j are the resistances contributed by strains i and j individually. We further focused on the most significant δ_{ij} values predicted by the regression model trained with all 67 communities. By comparing the model-predicted δ_{ij} values to the corresponding experimentally measured values (**Fig. S5**), we confirmed the model captured pairwise epistatic interactions within the synthetic communities. This consistency highlights the utility of the model in dissecting the contributions of specific strain interactions to pathogen invasion resistance.

Characterizing the invasion resistance to plant pathogen *Pseudomonas brassicacearum* Root401

To further investigate the pathogen invasion resistance of rhizosphere microbial communities, we conducted independent experiments with a second pathogen,

Pseudomonas brassicacearum Root401 (hereafter *P. brassicacearum* Root401)[42,45,46], isolated from the rhizosphere of *Arabidopsis*. We used the same pool of 14 bacterial strains to construct 105 distinct community subsets, encompassing single- strains communities (each of the 14 bacteria individually) and 91 SynComs (**Supplementary Table 2**). Consistent with the results observed for *Pst* DC3000, we found that the ability to inhibit the invasion of *P. brassicacearum* Root401 increased with species richness in the community (**Fig. 4A**). This suggests that a higher diversity of microbial species generally enhances the community's capacity to resist pathogen invasion. Furthermore, the inhibition capacity against *P. brassicacearum* Root401 was well-predicted by community composition using LASSO regression analysis, achieving an R^2 of 0.859 (**Fig. 4B, Fig. S7 and Fig. S8**).

We further analyzed the regression model parameters to understand the contributions of individual strains and their interactions. For the first-order terms, the model predicts that *Pseudomonas* Root562 (PS), and *Agrobacterium* Root491 (AG), are the most significant contributors to the inhibition of *P. brassicacearum* Root401 (**Fig. 4C**). These findings align with the results of single-strains tests (**Fig. S9**), reinforcing the key role of these strains in pathogen suppression. However, compared to the observations for *Pst* DC3000, the inhibitory effects of different strains on *P. brassicacearum* Root401 are relatively uniform.

To further explore the role of strains interactions in shaping invasion resistance, we predicted the epistatic effects of microbial strains jointly inhibiting pathogen invasion by using the second-order terms of the regression model. Notably, we found that most functional epistasis is negative (**Fig. 4D**). These findings highlight the importance of accounting for both individual contributions and interspecies interactions when evaluating the invasion resistance of microbial communities.

Taken together, our SynCom experiments with two different pathogens *Pst* DC3000 and *P. brassicacearum* Root401 revealed that the functional landscape of rhizosphere

252 microbial communities in pathogen invasion resistance is, to a large extent, smooth and
253 predictable. Our findings also emphasize the contribution of microbial interactions in
254 shaping community function.
255
256 **Understanding the functional landscape of invasion resistance using the gLV**
257 **model**
258 Population dynamic models such as the generalized Lotka–Volterra (gLV) and
259 consumer–resource (C–R) frameworks have been widely used to investigate invasion
260 resistance[47,48]. In particular, the gLV model has been applied to explore how community
261 diversity and interaction strength influence resistance to pathogen invasion[49-51]. Here,
262 we used the gLV model to investigate community-mediated resistance to pathogen
263 invasion, aiming to generalize our experimentally observed patterns and the utility of
264 data-driven approaches for identifying key strains and interactions (**Methods**).
265
266 We first investigated how the mean interaction strength (μ) influences invasion
267 resistance across different levels of strains richness. Consistent with our experimental
268 results, we observed that invasion resistance increases with community richness and
269 begins to plateau as additional strains are added to the communities. Furthermore,
270 communities with stronger mean interaction strengths exhibited greater suppression of
271 the invader, indicating that higher interspecies competition enhances resistance to
272 invasion (**Fig. 5A**).
273
274 Next, we assessed the performance of LASSO regression in simulated data based on
275 strain composition. Using first-order LASSO regression, which incorporates only
276 individual strain presence or absence, we observed strong predictive performance, with
277 R^2 values of 0.902, 0.950, and 0.920 for mean interaction strengths of –0.1, –0.3, and –
278 0.5, respectively **(Fig. S11A**). Incorporating second-order interaction terms further
279 improved prediction accuracy, yielding R^2 values of 0.990, 0.994, and 0.993 for the
280 same interaction strength levels (**Fig. S11B**). We further systematically explored the
281 performance of the second-order LASSO model across a broader range of interaction

means (μ = –0.5 to 0.1) and interaction variabilities (σ = 0.02 to 0.4). As high interaction variability (σ) often caused instability in the gLV simulation[52], the divergent data were excluded from further analysis. The LASSO regression achieved high accuracy when interaction variability (σ) is low, with R^2 values consistently above 0.9 (**Fig. 5B**). Assessing performance across training set sizes (μ = –0.3, σ = 0.1) showed that predictive accuracy improved with larger datasets and plateaued beyond 150 communities (**Fig. S12**).

To understand why LASSO regression can approximate the full gLV model (with up to 210 parameters, **Fig. S13**), we analyzed how individual gLV parameters affect pathogen suppression and interspecies epistatic effects. We found that any strain with a negative interaction coefficient toward the invader ($a_{0i} < 0$ in gLV model), reduces invader abundance, even when interactions are symmetric (e.g., $a_{i0} = a_{0i} = -0.3$, **Fig. S14A**). Regarding pairwise epistatic effects, we observed that within a certain range of interaction variability, the epistatic effect term δ_{ij} is near 0, with a slight bias toward the positive side. Since invasion resistance is defined as a decrease in invader abundance, positive δ_{ij} indicates that the combined presence of two strains slightly reduces the additive resistance effect (**Fig. S14B, C**). These findings indicate that within a certain range of variability, invasion resistance arises from predominantly additive contributions of individual strains, explaining why regression models with sparse parameters can predict the inhibition of invading pathogens.

Discussion

In this study, we profiled invasion resistance of synthetic rhizosphere microbial communities to two distinct pathogens *in planta*. We demonstrated that the invasion resistance conferred by synthetic microbial communities can be accurately modeled by LASSO regression based on the community composition profiles. Notably, the regression model only uses information about resident species' presence or absence at their inoculation, not at the time of invasion, nor does it use quantitative data about their abundance. To assess the generality of this approach, we simulated community-

mediated resistance using the gLV model. Furthermore, the LASSO regression accurately predicted invasion resistance across gLV-generated communities, showing strong agreement with the estimated contributions of individual strains (**Fig. S12B**). These findings underscore that statistical models can yield interpretable and biologically meaningful insights into microbial community functions.

Despite the inherent complexity of microbial interactions, our results reveal that species composition and their interactions are key factors influencing pathogen resistance[33,34]. This highlights the potential for simplifying analyses and improving predictions of community function[2]. Our findings identified *Pseudomonas* Root562 and *Agrobacterium* Root491 as the strains contributing most significantly to resistance against both pathogens. Interestingly, other strains exhibited varying inhibitory effects depending on the pathogen, reflecting the nuanced role of species interactions. Notably, although *Agrobacterium* and *Pseudomonas* spp. are reported to engage in interspecies interactions through type VI secretion systems[53-55], our strains showed no apparent inhibitory effects against pathogen *in vitro* (**Fig. S10**), suggesting that invasion resistance *in planta* is largely mediated by host-dependent mechanisms[45,56].

Characterizing the landscape of invasion resistance provide deeper insights into how microbial communities protect plants from pathogens. While additive effects generally explain why communities with higher species richness show stronger pathogen inhibition, weak negative interactions can limit these benefits. Notably, the function of a randomly selected community with a given species richness can vary greatly, highlighting that not all communities of the same size are equally effective[34]. This underscores the importance of rational design: communities selected based on model can achieve the same level of invasion resistance as full 14-strain community with only a few members (**Fig. S15**), consistent with prior work showing that well designed consortia of five strains can effectively suppress pathogen invasion[34]. These findings further support that understanding the functional landscape through modeling can guide the design of synthetic communities tailored for specific protective functions[57].

Modern agriculture faces a multitude of biotic and abiotic stresses, leading to significant crop yield losses. As a sustainable strategy for agricultural development, synthetic microbial communities have demonstrated promising potential in enhancing nutrient utilization and improving resistance against diseases in crops[33,58,59]. Rationally designed synthetic communities, informed by these principles and advanced modeling techniques, offering an environmentally friendly approach to plant protection and further applications. For instance, leveraging the synergistic interactions between a core strain and helper strains, researchers constructed a simplified yet functional synthetic microbial community. This SynCom significantly enhanced soybean growth by improving colonization, nodulation, and nitrogen fixation efficiency, achieving results comparable to those of a more complex synthetic community[60]. Moreover, SynComs have also demonstrated remarkable efficacy in enhancing crop disease resistance. For example, a simplified *Pseudomonas*-centric SynCom, colonized the rhizosphere through microbial synergistic interactions, significantly promoting watermelon growth and bolstering its resistance to soil-borne diseases[61]. Our findings emphasize the critical role of ecological principles, including species diversity and interactions, in shaping community resistance to pathogen invasion[51,62,63].

Methods

Plant materials and bacterial strains

Seeds of *Arabidopsis* (ecotype Columbia, Col-0) were sourced from laboratory stock. Fourteen bacterial commensals isolated from *Arabidopsis* roots, along with the detrimental strain *Pseudomonas brassicacearum* Root401, were generously provided by Prof. Paul Schulze-Lefert (Max Planck Institute for Plant Breeding Research, Cologne, Germany). Detailed information on the individual strains is available at At-SPHERE (http://www.at-sphere.com/) [64]. The pathogen strain *Pseudomonas syringae* pv. tomato DC3000 was generously provided by Prof. Alberto Macho (Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences, China). Kanamycin resistance was introduced into both *Pst* DC3000 and *P. brassicacearum* Root401 by conjugative transfer of plasmid pMRE132 (Addgene plasmid # 118486) from the donor strain *Escherichia coli* S17.

Phylogenetic tree construction

Phylogenetic relationships among the bacterial strains used in this study were inferred from their 16S rRNA gene sequences. Multiple sequence alignment was performed, and phylogenetic trees were constructed using the MEGA[65]. The resulting trees were subsequently edited and visually refined with iTOL[66].

Plant growth and microbial inoculation

Mono-association or synthetic community inoculation experiments were performed as previously described with minor modifications[30]. Individual bacterial strains from glycerol stock were incubated on half-strength tryptic soy broth (TSB, Sigma-Aldrich, USA) agar plates at 25°C for 5 days. A single colony of each strain was then cultured in half-strength TSB liquid medium at 25°C with shaking at 200 rpm for 2 days. On the day of inoculation, the bacterial cultures were subcultured at a 1:1 ratio for an additional 5 hours. A 500 µL aliquot of each bacterial culture was centrifuged at 6,500 × g for 2 minutes, washed twice with 10 mM $MgCl_2$, and resuspended in 1 mL of $MgCl_2$. Optical density at 600 nm (OD_{600}) was measured using SynergyTM H1 microplate reader (BioTek, USA). Bacterial suspensions were then diluted with $MgCl_2$ supplemented with 0.005% Silwet-77 (Shanghai Yuanye, China) to a final OD_{600} of 0.01. For synthetic bacterial communities of 2, 4, 6, 8, 10, or 14 members, equal volumes of the prepared

bacterial suspensions were pooled together prior to inoculation.

Arabidopsis seeds were surface-sterilized with 75% ethanol for 1 minute, followed by 20% bleach for 15 minutes, and rinsed 5 times with sterile distilled water. Seeds were evenly sown on half-strength Murashige and Skoog (MS) (Sigma-Aldrich, USA) agar plates containing 0.5% agar and 1% sucrose. After 2 days of stratification at 4°C in the dark, the seeds were vertically grown in a growth chamber under a 16-h dark/8-h light cycle at 22°C for 4 days. Ten uniformly growing seedlings were then transferred to fresh half-strength MS plates. After an additional 6 days of growth, six uniformly growing 10-day-old seedlings were selected for microbial inoculation.

A 10 mL aliquot of bacterial suspension was spread on the plates, ensuring that each seedling was completely submerged. After 5-min immersion, the bacterial suspension was discarded, and the plates were carefully dried. The seedlings were then cultured for another 5 days in the growth chamber. Subsequently, the seedlings were challenged with the kanamycin-labeled pathogen following the same procedure. Pathogen colonization on the seedlings was assessed 5 days post-invasion.

Pathogen colonization measurement

Colony-forming units (CFUs) were quantified as previously described with minor modifications [30,67]. Briefly, after 5 days of pathogen invasion, individual seedlings were placed in pre-weighed sterile tubes containing metal beads. The tubes were reweighed to determine the fresh weight of the seedling. The samples were then homogenized using a TissueLyzer (Shanghai Cebo, China) at 30 Hz for 30 seconds. A 500 µL aliquot of 1 × PBS buffer was added to each tube, and the samples were serially diluted in a sterile 96-well plate. A 5 uL aliquot of the liquid sample was placed on the far left side of a half-strength TSB agar plate containing 50 µg/mL kanamycin. The plate was then slowly tilted to the right to allow the liquid to flow evenly, and once the liquid reached a certain point, the plate was slowly tilted to the left to redistribute any remaining liquid. Finally, the plates were incubated at 25°C for two days until single colonies appeared. Pathogen colonization on the seedlings was determined based on the CFU count.

Binary inhibition assay

The assay was performed as previously described[68] with minor modifications. Individual bacterial colonies were cultured in half-strength TSB liquid medium at 25°C for 48 h with shaking (200 rpm) for 2 days. Bacterial suspensions were adjusted to an optical density at 600 nm (OD_{600}) of 0.1 using fresh medium. For pathogen lawn preparation, bacterial suspensions were standardized to OD_{600} = 1.0, and 2.5 mL of the suspension was mixed with 25 mL of half-strength TSB agar (pre-cooled to 42°C) and

poured into sterile Petri dishes. Subsequently, 2.5 μL aliquots of attacker strains were spotted onto the agar surface. Plates were incubated at ambient temperature (25°C) for 72 h, after which inhibition zones were documented photographically.

LASSO Regression

All datasets were fitted to the equation (1) truncated at the first and second order using LASSO regression. Both leave-one-out and random ten-fold cross-validation methods were applied to estimate regression parameters and ensure robust model performance. Models were implemented using the lasso function in MATLAB 2023b and executed on a personal computer. The regularization parameter λ was selected using the minimum 1-SE criterion across cross-validations[69]. To fit the datasets, all available data points were initially used to train the linear regression model. The resulting R^2 values and regression coefficients were analyzed to evaluate overall model performance.

To assess potential overfitting, 50 independent test sets were randomly sampled from the full datasets. Data were first divided into seven groups based on inhibition strength, and one data point from each group was randomly selected to form a test set of seven points. For each test set, 50 training sets of varying sizes were randomly selected from the remaining data. We randomly generated 50 test sets and perform LASSO regression with training sets with various set sizes. LASSO regression was then performed on each training set with 10-fold cross-validation, and model performance was evaluated using R^2 values on both the corresponding training and test sets. Model performance was evaluated by calculating R^2 values on both the training and test sets. To further assess potential prediction bias related to community composition, training and test sets were also split according to SynComs groupings.

Calculation of epistatic effects

To evaluate whether the combined effect of two species deviates from a simple additive model, we define the epistatic effects δ_{ij} as follows:

$$\delta_{ij} = D_{ij} - \left(D_i + D_j - D_0\right)$$

Where D_{ij} is the invasion resistance of the community containing both strain i and j, D_i and D_j are the invasion resistances contributed by strain i and j individually, and D_0 is the residual resistance predicted by the LASSO model when $x_i = -1$, representing the resistance to pathogen invasion in the absence of any species. Using the coefficients obtained from the LASSO regression model, D_0 can be calculated as:

$$D_0 = \beta_0 - \sum \beta_i + \sum \beta_{ij}$$

For communities consisting of one or two strains, the invasion resistance is calculated as follows:

For a single strain i:

$$D_i = D_0 + 2\beta_i - 2\sum_k \beta_{ik}$$

For a pair of strains i and j:

$$D_{ij} = D_0 + 2\beta_i + 2\beta_j - 2\sum_k \beta_{ik} - 2\sum_k \beta_{jk} + 4\beta_{ij} = D_i + D_j - D_0 + 4\beta_{ij}$$

Finally, the epistatic effect δ_{ij} is calculated as:

$$\delta_{ij} = D_{ij} - \left(D_i + D_j - D_0\right) = 4\beta_{ij}.$$

This approach allows us to assess whether the interaction between strains i and j in a community deviate from an additive model of pathogen resistance, using the coefficients derived from the LASSO regression model.

Generalized Lotka-Volterra (gLV) Model Simulation

To assess whether a generalized Lotka-Volterra (gLV) model can capture the experimentally observed invasion resistance patterns, we simulated a community of 15 bacterial species, where species $i = 0$ represents the pathogen, and species $i = 1$ to 14 represent the symbiotic (non-pathogenic) community members. The population dynamics were governed by the following gLV equations:

$$\frac{dx_i}{dt} = x_i\left(\gamma_i + \sum_j a_{ij}x_j\right),$$

where x_i is the abundance of species i, γ_i is the intrinsic growth rate of species i, a_{ij} (when $i \neq j$) accounts for the direct impact that species j has on the population change of species i. We set $\gamma_i \sim \mathcal{N}(1, 0.2^2)$, with the invader's growth rate fixed at $\gamma_0 = 1$.

To explore different ecological interaction regimes, we systematically varied the mean $\eta \in [-0.5, 0.1]$ and standard deviation $\sigma \in [0.02, 0.4]$ of the pairwise interaction where $a_{ij} \sim -|\mathcal{N}(\eta, \sigma^2)|$. Self-interaction terms $a_{ii} = -1$ were fixed to prevent unbounded growth. The effects of pathogen on other species were fixed as $a_{0i} = -0.3$, For each (η, σ) pair, simulations were repeated three times with different random seeds.

We simulated pathogen abundance under various scenarios to evaluate invasion resistance across different levels of species richness. In total, 345 community compositions were examined, including all 14 single-species drop-out communities (n = 14), the full 14-member community (n = 1), and 30 randomly sampled communities

for each richness levels of 2-12 species (n=330).

The log-transformed final abundance of the pathogen was regressed using a 2-order LASSO model. Model was trained using 10-fold cross-validation, and the performance was evaluated by R^2 between predicted and simulated log-abundances. All simulations and analyses were performed using MATLAB R2023b on a standard personal computer.

Data availability

All source data of this study are available along with the code on GitHub: https://github.com/JacobZuo/Invasion-Resistance.

Code availability

All code for simulation and scripts used to generate figures are available on GitHub https://github.com/JacobZuo/Invasion-Resistance.

Author Contribution

L.D. and L.W. conceived the study. L.W., W.Z., L.D., and Y.Y.L. designed the experiments, interpreted the results, and wrote the manuscript. W.Z. implemented the regression models and performed the *in silico* simulations. L.W. conducted the *in planta* experiments. L.W. and W.Z. analyzed the experimental data. J.F. and Y.B. contributed to data interpretation, discussed the results, and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Acknowledgements

We thank Prof. Paul Schulze-Lefert for sharing the strains. We thank Dr. Xuwen Wang for insightful discussions on modeling. This work was supported by National Natural Science Foundation of China 32061143023 (L.D.), National Natural Science Foundation of China 32471555 (W.Z.) Guangdong Basic and Applied Basic Research Foundation 2023A1515012006 (L.W.), National Natural Science Foundation of China 32400231 (L.W.)

Reference

1. Sanchez, A., Bajic, D., Diaz-Colunga, J., Skwara, A., Vila, J.C.C., and Kuehn, S. (2023). The community-function landscape of microbial consortia. Cell Syst *14*, 122–134. 10.1016/j.cels.2022.12.011.
2. Skwara, A., Gowda, K., Yousef, M., Diaz-Colunga, J., Raman, A.S., Sanchez, A., Tikhonov, M., and Kuehn, S. (2023). Statistically learning the functional landscape of microbial communities. Nat Ecol Evol *7*, 1823–1833. 10.1038/s41559-023-02197-4.
3. Gopalakrishnappa, C., Gowda, K., Prabhakara, K.H., and Kuehn, S. (2022). An ensemble approach to the structure-function problem in microbial communities. iScience *25*, 103761. 10.1016/j.isci.2022.103761.
4. Thomas, J.L., Rowland-Chandler, J., and Shou, W. (2024). Artificial selection of microbial communities: what have we learnt and how can we improve? Current Opinion in Microbiology *77*. 10.1016/j.mib.2023.102400.
5. Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol *15*, 579–590. 10.1038/nrmicro.2017.87.
6. Wu, L., Wang, X.W., Tao, Z., Wang, T., Zuo, W., Zeng, Y., Liu, Y.Y., and Dai, L. (2024). Data-driven prediction of colonization outcomes for complex microbial communities. Nat Commun *15*, 2406. 10.1038/s41467-024-46766-y.
7. Michel-Mata, S., Wang, X.W., Liu, Y.Y., and Angulo, M.T. (2022). Predicting microbiome compositions from species assemblages through deep learning. Imeta *1*. 10.1002/imt2.3.
8. Yuan, A.E., and Shou, W. (2022). Data-driven causal analysis of observational biological time series. Elife *11*. 10.7554/eLife.72518.
9. Sorbara, M.T., and Pamer, E.G. (2022). Microbiome-based therapeutics. Nature Reviews Microbiology *20*, 365–380. 10.1038/s41579-021-00667-9.
10. Albright, M.B.N., Louca, S., Winkler, D.E., Feeser, K.L., Haig, S.J., Whiteson, K.L., Emerson, J.B., and Dunbar, J. (2022). Solutions in microbiome engineering: prioritizing barriers to organism establishment. ISME J *16*, 331–338. 10.1038/s41396-021-01088-5.
11. Foo, J.L., Ling, H., Lee, Y.S., and Chang, M.W. (2017). Microbiome engineering: Current applications and its future. Biotechnology Journal *12*. 10.1002/biot.201600099.
12. Northen, T.R., Kleiner, M., Torres, M., Kovacs, A.T., Nicolaisen, M.H., Krzyzanowska, D.M., Sharma, S., Lund, G., Jelsbak, L., Baars, O., et al. (2024). Community standards and future opportunities for synthetic communities in plant-microbiota research. Nat Microbiol *9*, 2774–2784. 10.1038/s41564-024-01833-4.
13. Buffie, C.G., and Pamer, E.G. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol *13*, 790–801. 10.1038/nri3535.
14. Kinnunen, M., Dechesne, A., Proctor, C., Hammes, F., Johnson, D., Quintela-Baluja, M., Graham, D., Daffonchio, D., Fodelianakis, S., Hahn, N., et al. (2016). A conceptual framework for invasion in microbial communities. ISME J *10*, 2773–2775. 10.1038/ismej.2016.75.
15. Mendes, R., Garbeva, P., and Raaijmakers, J.M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev *37*, 634–663. 10.1111/1574-6976.12028.

16. Chang, P.V. (2020). Chemical Mechanisms of Colonization Resistance by the Gut Microbial Metabolome. ACS Chem Biol *15*, 1119–1126. 10.1021/acschembio.9b00813.
17. Chassaing, B., and Cascales, E. (2018). Antibacterial Weapons: Targeted Destruction in the Microbiota. Trends Microbiol *26*, 329–338. 10.1016/j.tim.2018.01.006.
18. Booth, S.C., Smith, W.P.J., and Foster, K.R. (2023). The evolution of short- and long-range weapons for bacterial competition. Nat Ecol Evol *7*, 2080–2091. 10.1038/s41559-023-02234-2.
19. Lloyd, D.P., and Allen, R.J. (2015). Competition for space during bacterial colonization of a surface. J R Soc Interface *12*, 0608. 10.1098/rsif.2015.0608.
20. Woelfel, S., Silva, M.S., and Stecher, B. (2024). Intestinal colonization resistance in the context of environmental, host, and microbial determinants. Cell Host Microbe *32*, 820–836. 10.1016/j.chom.2024.05.002.
21. Bakkeren, E., Piskovsky, V., and Foster, K.R. (2025). Metabolic ecology of microbiomes: Nutrient competition, host benefits, and community engineering. Cell Host Microbe *33*, 790–807. 10.1016/j.chom.2025.05.013.
22. Ivanov, II, Atarashi, K., Manel, N., Brodie, E.L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K.C., Santee, C.A., Lynch, S.V., et al. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. Cell *139*, 485–498. 10.1016/j.cell.2009.09.033.
23. Paasch, B.C., Sohrabi, R., Kremer, J.M., Nomura, K., Cheng, Y.T., Martz, J., Kvitko, B., Tiedje, J.M., and He, S.Y. (2023). A critical role of a eubiotic microbiota in gating proper immunocompetence in Arabidopsis. Nat Plants *9*, 1468–1480. 10.1038/s41477-023-01501-1.
24. Mallon, C.A., Elsas, J.D.V., and Salles, J.F. (2015). Microbial invasions: the process, patterns, and mechanisms. Trends Microbiol *23*, 719–729. 10.1016/j.tim.2015.07.013.
25. Jones, M.L., Rivett, D.W., Pascual-Garcia, A., and Bell, T. (2021). Relationships between community composition, productivity and invasion resistance in semi-natural bacterial microcosms. Elife *10*. 10.7554/eLife.71811.
26. Kennedy, T.A., Naeem, S., Howe, K.M., Knops, J.M., Tilman, D., and Reich, P. (2002). Biodiversity as a barrier to ecological invasion. Nature *417*, 636–638. 10.1038/nature00776.
27. Spragge, F., Bakkeren, E., Jahn, M.T., E, B.N.A., Pearson, C.F., Wang, X., Pankhurst, L., Cunrath, O., and Foster, K.R. (2023). Microbiome diversity protects against pathogens by nutrient blocking. Science *382*, eadj3502. 10.1126/science.adj3502.
28. Stachowicz, J.J., Whitlatch, R.B., and Osman, R.W. (1999). Species Diversity and Invasion Resistance in a Marine Ecosystem. Science *286*, 1577–1579. 10.1126/science.286.5444.1577.
29. Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J., and Pamer, E.G. (2012). Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infect Immun *80*, 62–73. 10.1128/IAI.05496-11.
30. Ma, K.W., Ordon, J., and Schulze-Lefert, P. (2022). Gnotobiotic Plant Systems for Reconstitution and Functional Studies of the Root Microbiota. Curr Protoc *2*, e362. 10.1002/cpz1.362.

31. Carlstrom, C.I., Field, C.M., Bortfeld-Miller, M., Muller, B., Sunagawa, S., and Vorholt, J.A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. Nat Ecol Evol *3*, 1445–1454. 10.1038/s41559-019-0994-z.

32. Hu, J., Wei, Z., Friman, V.P., Gu, S.H., Wang, X.F., Eisenhauer, N., Yang, T.J., Ma, J., Shen, Q.R., Xu, Y.C., and Jousset, A. (2016). Probiotic Diversity Enhances Rhizosphere Microbiome Function and Plant Disease Suppression. mBio *7*. 10.1128/mBio.01790-16.

33. Vogel, C.M., Potthoff, D.B., Schafer, M., Barandun, N., and Vorholt, J.A. (2021). Protective role of the Arabidopsis leaf microbiota against a bacterial pathogen. Nat Microbiol *6*, 1537–1548. 10.1038/s41564-021-00997-7.

34. Emmenegger, B., Massoni, J., Pestalozzi, C.M., Bortfeld-Miller, M., Maier, B.A., and Vorholt, J.A. (2023). Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. Nat Commun *14*, 7983. 10.1038/s41467-023-43793-z.

35. Finkel, O.M., Salas-Gonzalez, I., Castrillo, G., Conway, J.M., Law, T.F., Teixeira, P., Wilson, E.D., Fitzpatrick, C.R., Jones, C.D., and Dangl, J.L. (2020). A single bacterial genus maintains root growth in a complex microbiome. Nature *587*, 103–108. 10.1038/s41586-020-2778-7.

36. Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., Del Rio, T.G., et al. (2012). Defining the core Arabidopsis thaliana root microbiome. Nature *488*, 86–90. 10.1038/nature11237.

37. Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., et al. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature *488*, 91–95. 10.1038/nature11336.

38. Innerebner, G., Knief, C., and Vorholt, J.A. (2011). Protection of Arabidopsis thaliana against leaf-pathogenic Pseudomonas syringae by Sphingomonas strains in a controlled model system. Appl Environ Microbiol *77*, 3202–3210. 10.1128/AEM.00133-11.

39. Buell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T., Gwinn, M.L., Dodson, R.J., Deboy, R.T., Durkin, A.S., Kolonay, J.F., et al. (2003). The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. Proceedings of the National Academy of Sciences *100*, 10181–10186. 10.1073/pnas.1731982100.

40. Poelwijk, F.J., Krishna, V., and Ranganathan, R. (2016). The Context-Dependence of Mutations: A Linkage of Formalisms. PLoS Comput Biol *12*, e1004771. 10.1371/journal.pcbi.1004771.

41. Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology *58*, 267–288. 10.1111/j.2517-6161.1996.tb02080.x.

42. Amrhein, A., Zhang, M., Hacquard, S., Heintz-Buschart, A., and Wippel, K. (2025). Pseudomonas intra-genus competition determines the protective function of synthetic bacterial communities in Arabidopsis thaliana. PLoS Biol *23*, e3002882. 10.1371/journal.pbio.3002882.

43. Ma, K.-W., Niu, Y., Jia, Y., Ordon, J., Copeland, C., Emonet, A., Geldner, N., Guan, R.,

659 Stolze, S.C., Nakagami, H., et al. (2021). Coordination of microbe–host homeostasis
660 by crosstalk with plant innate immunity. Nature Plants *7*, 814–825. 10.1038/s41477-
661 021-00920-2.
662 44. Diaz-Colunga, J., Skwara, A., Vila, J.C.C., Bajic, D., and Sanchez, A. (2024). Global
663 epistasis and the emergence of function in microbial consortia. Cell *187*, 3108–3119
664 e3130. 10.1016/j.cell.2024.04.016.
665 45. Ma, K.W., Niu, Y., Jia, Y., Ordon, J., Copeland, C., Emonet, A., Geldner, N., Guan, R.,
666 Stolze, S.C., Nakagami, H., et al. (2021). Coordination of microbe-host homeostasis
667 by crosstalk with plant innate immunity. Nat Plants *7*, 814–825. 10.1038/s41477-021-
668 00920-2.
669 46. Getzke, F., Hassani, M.A., Crusemann, M., Malisic, M., Zhang, P., Ishigaki, Y.,
670 Bohringer, N., Jimenez Fernandez, A., Wang, L., Ordon, J., et al. (2023). Cofunctioning
671 of bacterial exometabolites drives root microbiota establishment. Proc Natl Acad Sci U
672 S A *120*, e2221508120. 10.1073/pnas.2221508120.
673 47. Ke, S., Xiao, Y., Weiss, S.T., Chen, X., Kelly, C.P., and Liu, Y.Y. (2023). A
674 computational method to dissect colonization resistance of the gut microbiota against
675 pathogens. Cell Rep Methods *3*, 100576. 10.1016/j.crmeth.2023.100576.
676 48. Kurkjian, H.M., Akbari, M.J., and Momeni, B. (2021). The impact of interactions on
677 invasion and colonization resistance in microbial communities. PLoS Comput Biol *17*,
678 e1008643. 10.1371/journal.pcbi.1008643.
679 49. Case, T.J. (1990). Invasion resistance arises in strongly interacting species-rich model
680 competition communities. Proceedings of the National Academy of Sciences *87*, 9610–
681 9614. 10.1073/pnas.87.24.9610.
682 50. Hu, J., Barbier, M., Bunin, G., and Gore, J. (2025). Collective dynamical regimes
683 predict invasion success and impacts in microbial communities. Nat Ecol Evol *9*, 406–
684 416. 10.1038/s41559-024-02618-y.
685 51. Hromada, S., Qian, Y., Jacobson, T.B., Clark, R.L., Watson, L., Safdar, N., Amador-
686 Noguez, D., and Venturelli, O.S. (2021). Negative interactions determine Clostridioides
687 difficile growth in synthetic human gut communities. Mol Syst Biol *17*, e10355.
688 10.15252/msb.202110355.
689 52. Bunin, G. (2017). Ecological communities with Lotka-Volterra dynamics. Phys Rev E
690 *95*, 042414. 10.1103/PhysRevE.95.042414.
691 53. Bernal, P., Allsopp, L.P., Filloux, A., and Llamas, M.A. (2017). The Pseudomonas
692 putida T6SS is a plant warden against phytopathogens. The ISME Journal *11*, 972–987.
693 10.1038/ismej.2016.169.
694 54. Ma, L.-S., Hachani, A., Lin, J.-S., Filloux, A., and Lai, E.-M. (2014). Agrobacterium
695 tumefaciens Deploys a Superfamily of Type VI Secretion DNase Effectors as Weapons
696 for Interbacterial Competition In Planta. Cell Host & Microbe *16*, 94–104.
697 10.1016/j.chom.2014.06.002.
698 55. Decoin, V., Gallique, M., Barbey, C., Le Mauff, F., Poc, C.D., Feuilloley, M.G.J.,
699 Orange, N., and Merieau, A. (2015). A Pseudomonas fluorescens type 6 secretion
700 system is related to mucoidy, motility and bacterial competition. BMC Microbiology
701 *15*. 10.1186/s12866-015-0405-9.
702 56. Wang, Y., Pruitt, R.N., Nürnberger, T., and Wang, Y. (2022). Evasion of plant immunity

by microbial pathogens. Nature Reviews Microbiology *20*, 449–464. 10.1038/s41579-022-00710-3.

57. Vrancken, G., Gregory, A.C., Huys, G.R.B., Faust, K., and Raes, J. (2019). Synthetic ecology of the human gut microbiota. Nat Rev Microbiol *17*, 754–763. 10.1038/s41579-019-0264-8.

58. Wen, T., Zhao, M., Yuan, J., Kowalchuk, G.A., and Shen, Q. (2020). Root exudates mediate plant defense against foliar pathogens by recruiting beneficial microbes. Soil Ecology Letters *3*, 42–51. 10.1007/s42832-020-0057-z.

59. Niu, B., Paulson, J.N., Zheng, X., and Kolter, R. (2017). Simplified and representative bacterial community of maize roots. Proc Natl Acad Sci U S A *114*, E2450–E2459. 10.1073/pnas.1616148114.

60. Li, Y., Li, R., Liu, R., Shi, J., Qiu, X., Lei, J., Zhao, X., Wang, C., Ge, M., Xu, H., et al. (2025). A simplified SynCom based on core-helper strain interactions enhances symbiotic nitrogen fixation in soybean. J Integr Plant Biol *67*, 1582–1598. 10.1111/jipb.13881.

61. Qiao, Y., Wang, Z., Sun, H., Guo, H., Song, Y., Zhang, H., Ruan, Y., Xu, Q., Huang, Q., Shen, Q., and Ling, N. (2024). Synthetic community derived from grafted watermelon rhizosphere provides protection for ungrafted watermelon against Fusarium oxysporum via microbial synergistic effects. Microbiome *12*. 10.1186/s40168-024-01814-z.

62. Li, M., Wei, Z., Wang, J., Jousset, A., Friman, V.P., Xu, Y., Shen, Q., and Pommier, T. (2019). Facilitation promotes invasions in plant-associated microbial communities. Ecol Lett *22*, 149–158. 10.1111/ele.13177.

63. Li, M., Pommier, T., Yin, Y., Wang, J., Gu, S., Jousset, A., Keuskamp, J., Wang, H., Wei, Z., Xu, Y., et al. (2022). Indirect reduction of Ralstonia solanacearum via pathogen helper inhibition. ISME J *16*, 868–875. 10.1038/s41396-021-01126-2.

64. Bai, Y., Muller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Munch, P.C., Spaepen, S., Remus-Emsermann, M., et al. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. Nature *528*, 364–369. 10.1038/nature16192.

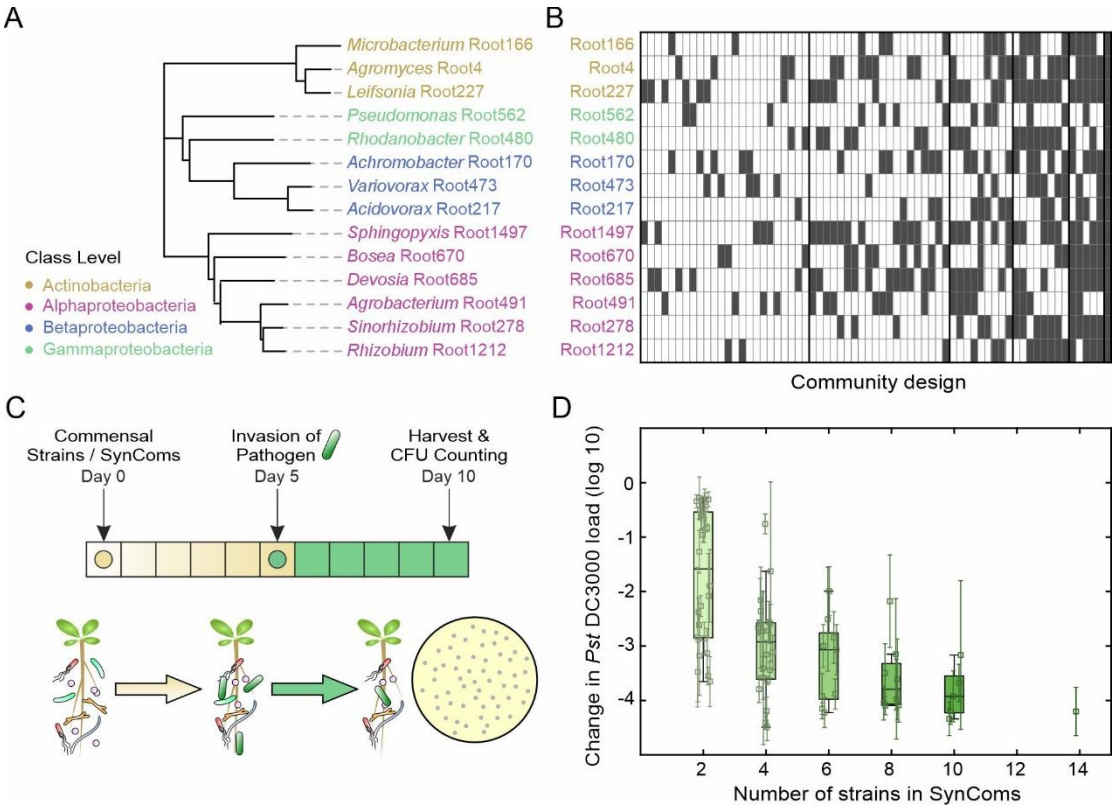
65. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol *35*, 1547–1549. 10.1093/molbev/msy096.

66. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res *47*, W256–W259. 10.1093/nar/gkz239.

67. Wang, L., Liu, Y., Ni, H., Zuo, W., Shi, H., Liao, W., Liu, H., Chen, J., Bai, Y., Yue, H., et al. (2024). Systematic characterization of plant-associated bacteria that can degrade indole-3-acetic acid. PLoS Biol *22*, e3002921. 10.1371/journal.pbio.3002921.

68. Chen, T., Nomura, K., Wang, X., Sohrabi, R., Xu, J., Yao, L., Paasch, B.C., Ma, L., Kremer, J., Cheng, Y., et al. (2020). A plant genetic network for preventing dysbiosis in the phyllosphere. Nature *580*, 653–657. 10.1038/s41586-020-2185-0.

69. The Elements of Statistical Learning. (2009). 10.1007/978-0-387-84858-7.



747

748 **Figure 1. Profiling the functional landscape of synthetic communities (SynComs)**
749 **in protecting plants from pathogen invasion *in planta*.**

750 **A.** Phylogenetic tree of the strains selected for constructing SynComs.

751 **B.** Each row represents a strain, and each column represents a community design. The
752 blocks indicate the presence (1) or absence (0) of the corresponding strain in each
753 community. Synthetic communities (S_2–S_{14}) consisting of 2 to 14 strains were
754 constructed. A total of 67 distinct communities were evaluated.

755 **C.** Overview of the experimental workflow. 10-day-old *Arabidopsis* seedlings were
756 first inoculated with different synthetic bacterial communities. Five days later,
757 pathogens were introduced to simulate invasion. After an additional five days of
758 cultivation, seedling fresh weight and pathogen colonization, measured as colony-
759 forming units (CFU), were quantified.

760 **D.** Resistance to *Pst* DC3000 invasion across synthetic communities of varying
761 diversity. Each data point represents the mean and the standard deviation ($n \geq 3$
762 replicates) of a specific community. The boxplot displays the median and interquartile
763 range of invasion resistance among communities with the same number of species.

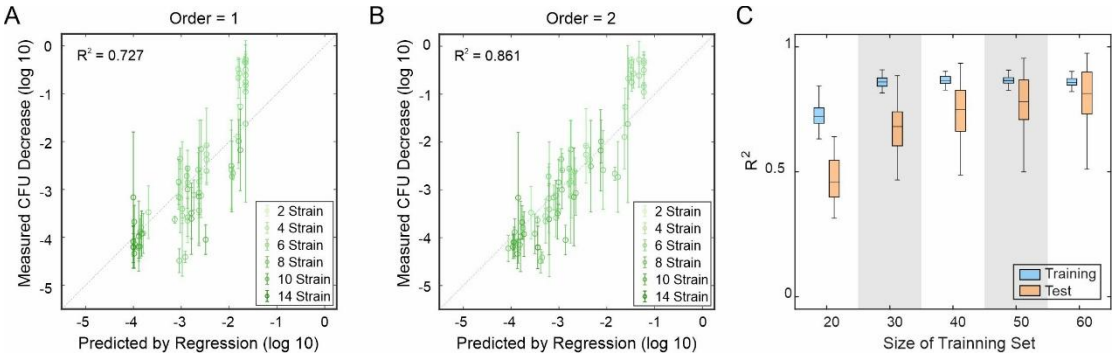


Figure 2. Predicting the invasion resistance of SynComs to plant pathogen *Pst* DC3000 by LASSO regression.

A,B. Scatter plots showing the predicted values for the first-order (A) and second-order (B) regression fits obtained via LASSO regression with leave-one-out cross-validation. Vertical error bars indicate the standard deviation across experimental replicates.

C. The predictive performance of the LASSO regression model increases with the size of the training set. Fifty test sets were generated by randomly selecting 7 data points from the full dataset. For each test set, 50 random training sets of varying sizes were drawn from the remaining data, and a second-order LASSO regression was fitted with 10-fold cross-validation (see Methods). The mean R^2 was calculated separately for the training and test sets. Box plots show the distribution of these mean R^2 values across the 50 test sets.

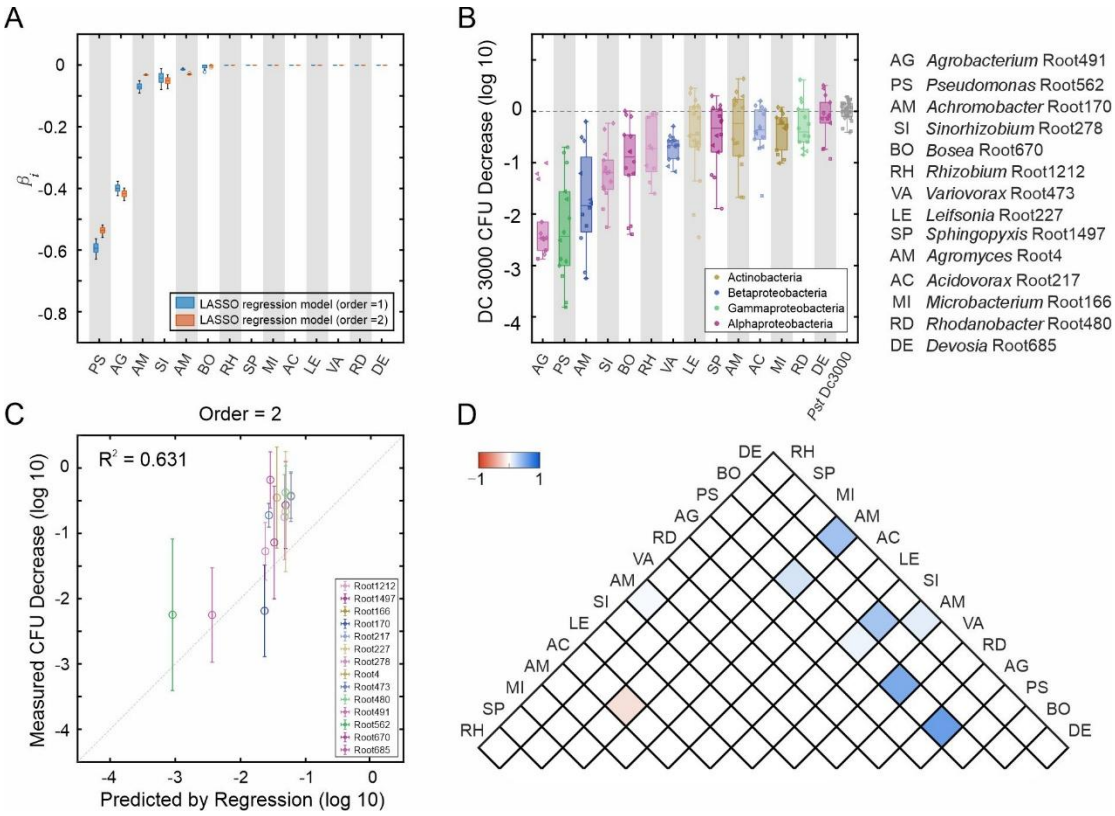


Figure 3. The strength of individual strains to resist *Pst* DC3000 invasion can be inferred by the first-order coefficients in LASSO regression.

A. Coefficients of the first-order terms in regression analysis for both first-order and second-order approximations. The boxplot displays the median and interquartile range (1st and 3rd quartiles) across 100 repeat LASSO regressions with random 10-fold cross-validation splits.

B. Ability of a single species to resist invasion by the pathogen *Pst* DC3000. The boxplot shows the median and interquartile range (1st and 3rd quartiles) based on n ≥ 6 replicates from two experimental batches.

C Scatter plots showing the prediction of the ability of individual species to resist pathogen invasion for the second-order obtained via LASSO regression, respectively. Vertical error bars indicate the standard deviation across experimental replicates.

D. Difference in invasion resistance of co-cultures predicted by LASSO regression versus additive expectations. Blue indicates that invasion resistance is lower than the additive expectation, meaning that co-inoculation of the two strains results in higher invasion than expected. Orange indicates the opposite, where invasion resistance is higher than the additive expectation.

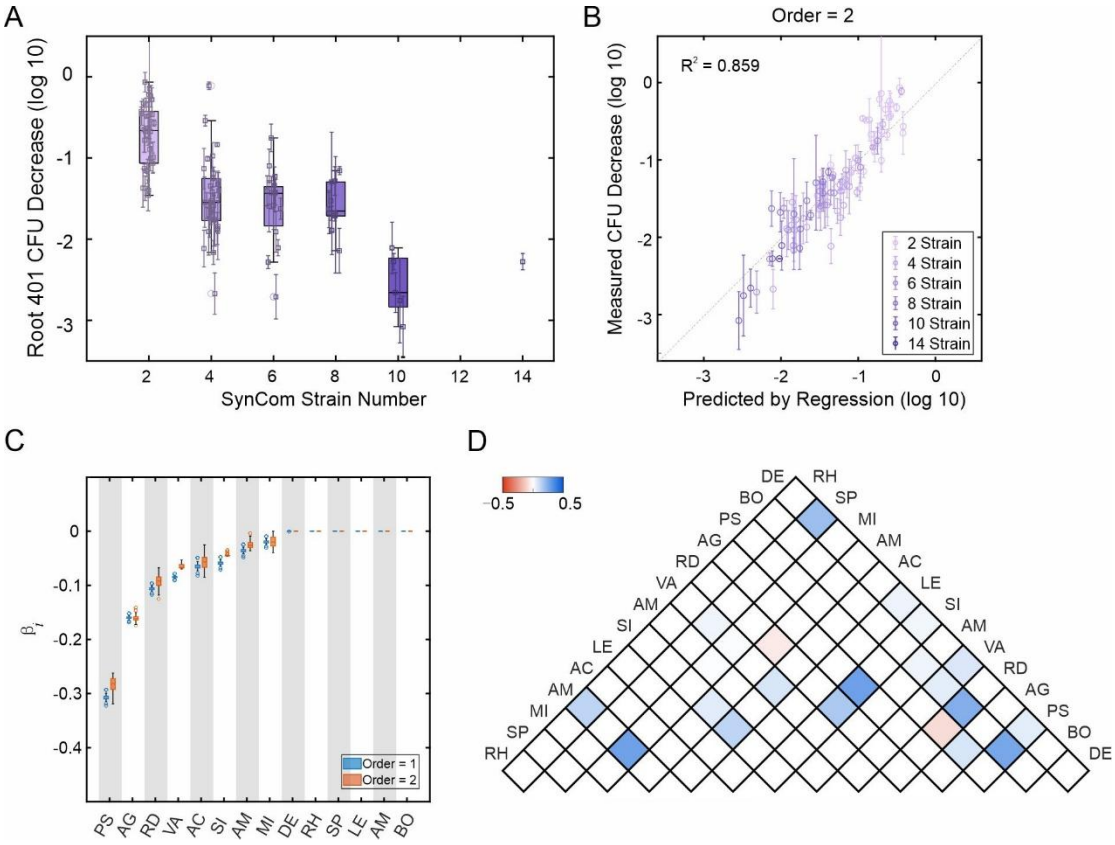


Figure 4. Profiling the functional landscape of invasion resistance to plant pathogen *Pseudomonas brassicacearum* Root401.

A. Resistance to *P. brassicacearum* Root401 invasion across 91 synthetic communities of varying sizes. Each data point represents the mean resistance ability of a specific community, with error bars showing the standard deviation across replicates. The boxplot displays the median and interquartile range (1st and 3rd quartiles) in resistance among communities with the same species number.

B. Scatter plots showing the predicted values obtained via second-order LASSO regression. Vertical error bars indicate the standard deviation across experimental replicates.

C. Coefficients of the first-order terms in regression analysis for both first-order and second-order approximations. The boxplot displays the median and interquartile range (1st and 3rd quartiles) across 100 repeat LASSO regressions with random 10-fold cross-validation splits.

D. Difference in invasion resistance of co-cultures predicted by LASSO regression versus additive expectations. Blue indicates that invasion resistance is lower than the additive expectation, meaning that co-inoculation of the two strains results in higher invasion than expected. Orange indicates the opposite, where invasion resistance is higher than the additive expectation.

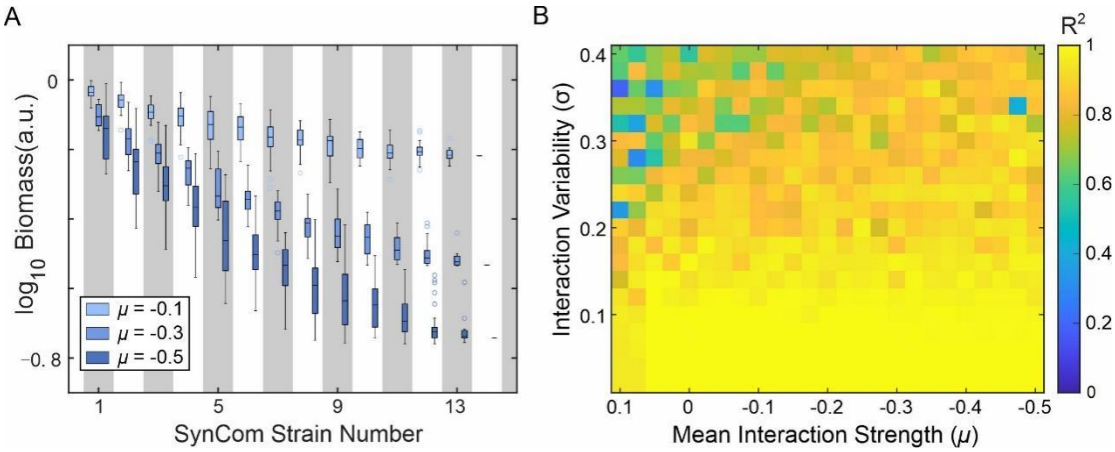


Figure 5. Predicting invasion resistance of simulated communities in the generalized Lotka-Volterra model.

A. Boxplot showing the resistance to pathogen invasion across communities of varying species richness, as predicted by gLV model simulations (Methods). Different colors indicates different mean interactions strength (μ).The boxplot represents the median and interquartile range (1st and 3rd quartiles) of resistance values among communities with the same species number.

B. Heatmap displaying the mean R^2 values of second-order LASSO regression applied to simulation data generated from the gLV model under varying mean interaction strengths (μ) and interaction variabilities (σ). Each value reflects the average R^2 from three independent gLV simulations. Diverged data points were excluded from the LASSO analysis.