



**HAL**  
open science

## Biotic interactions shape infection outcomes in *Arabidopsis*

Maryam Mahmoudi, Juliana Almario, Yiheng Hu, Lynn-Marie Tenzer, Kay Nieselt, Eric Kemen

► **To cite this version:**

Maryam Mahmoudi, Juliana Almario, Yiheng Hu, Lynn-Marie Tenzer, Kay Nieselt, et al.. Biotic interactions shape infection outcomes in *Arabidopsis*. 2024. hal-04773955

**HAL Id: hal-04773955**

**<https://hal.science/hal-04773955v1>**

Preprint submitted on 8 Nov 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Biotic interactions shape infection outcomes in *Arabidopsis*

Maryam Mahmoudi<sup>1</sup>, Juliana Almario<sup>2</sup>, Yiheng Hu<sup>1</sup>, Lynn-Marie Tenzer<sup>1</sup>, Kay Nieselt<sup>3</sup> and Eric Kemen<sup>1,\*</sup>

<sup>1</sup>Microbial Interactions in Plant Ecosystems, IMIT/ZMBP, Eberhard Karls University of Tübingen, Auf der Morgenstelle 32, 72076 Tübingen, Germany

<sup>2</sup>Université Claude Bernard Lyon 1, Laboratoire d'Ecologie Microbienne, UMR CNRS 5557, UMR INRAE 1418, VetAgro Sup, 69622, Villeurbanne, France

<sup>3</sup>Institute for Bioinformatics and Medical Informatics, Eberhard Karls University of Tübingen, Sand 14, 72076 Tübingen, Germany

\***Corresponding author:** Eric Kemen, Microbial Interactions in Plant Ecosystems, IMIT/ZMBP, Eberhard Karls University of Tübingen, Auf der Morgenstelle 32, 72076 Tübingen, Germany. Email: [eric.kemen@uni-tuebingen.de](mailto:eric.kemen@uni-tuebingen.de)

## Abstract

The plant microbiome protects plants from stresses, including pathogen attacks. However, identifying microbes that provide plant protection remains challenging in complex microbial communities. In this study, we analysed samples from natural *A. thaliana* populations, including both plants infected with the pathogenic oomycete *Albugo laibachii* and uninfected plants, over six years. Using machine learning classification models, we achieved high accuracy in distinguishing infected and uninfected plants based on microbiome abundance. We identified 80 key taxa associated with health and disease. Among the health-associated microbes (HCom), we selected bacteria, fungi, and cercozoa that effectively reduced pathogen presence in co-inoculation assays. In comparison, disease-associated microbes (DCom) were less effective in conferring protection. Our findings highlight the complexity of plant-microbe interactions and advance our understanding of microbial roles in plant disease ecology. By integrating ecological insights with machine learning, we take a significant step towards designing robust microbial consortia that enhance plant resilience against pathogens.

**Keywords:** Microbe-microbe interaction, infected and uninfected leaves, machine learning, plant pathogen, natural probiotic

## 35 Introduction

36 Similar to higher organisms such as humans, plant tissues are colonized by a wide  
37 range of microbes known as microbiota or microbiome. The microbiome associated  
38 with the plant leaves, i.e., the phyllosphere, is thought to play an important role in  
39 the physiology, fitness, and defense mechanism of the host against various biotic and  
40 abiotic perturbations [1]. Leaves are inhabited mainly by commensal species but can  
41 also harbor pathogenic bacteria, fungi, and oomycetes, which can significantly harm  
42 natural plants and crops [2], thereby causing annual crop yield losses and reducing food  
43 availability. Climate change is accelerating the spread of pathogens, thereby affecting  
44 forest health globally [3, 4]. Therefore, it is crucial to develop strategies that protect  
45 from pathogens in a changing environment.

46  
47 Biological control is an effective and environmentally friendly alternative to pesticides  
48 for combating microbial plant diseases [5]. For example, *Trichoderma*, an opportunistic  
49 fungal genus, is widely used as a biological control agent against phytopathogens and  
50 is studied for its role in helping plants manage biotic and abiotic stresses [6]. However,  
51 identifying and experimentally validating biocontrol microbes through traditional meth-  
52 ods can be slow and challenging [7]. Another approach to biocontrol involves identifying  
53 healthy microbiomes. While healthy plant communities are characterized by diverse and  
54 balanced microorganisms, as seen in comparisons between healthy and diseased plants  
55 [8]. However, it remains unclear what defines a healthy or beneficial microbiome. One  
56 promising solution to this challenge is the use of machine learning. Machine learning  
57 techniques have been used in microbiome research to accomplish various tasks, such as  
58 predicting host or environmental phenotypes and categorizing microbial properties, in-  
59 cluding monitoring changes in microbiome composition [9]. Machine learning classifiers  
60 were used to identify soil microbial patterns predicting the presence of *Fusarium oxyspo-*  
61 *rum*, the pathogen causing Fusarium wilt disease under field conditions [10]. Similarly,  
62 the random forest method accurately predicted productivity based on microbiome com-  
63 position at the order level. Significant differences in crop yield were associated with  
64 bulk soil microbiome composition, with many taxa contributing to nitrogen utilization  
65 [11]. Machine learning was used to identify bacterial strains important in reducing leaf  
66 infection with the pathogenic bacterium *Pseudomonas syringae* DC3000 [12]. However,  
67 few studies have investigated the different taxonomic groups of bacteria and eukaryotes  
68 in the microbiota of natural plants attacked by obligate biotrophic pathogens.

69  
70 The obligate biotrophic oomycete *Albugo laibachii* is a common pathogen of the *Bras-*  
71 *sicaceae* family and the causal agent of the white rust disease [13]. This pathogen was  
72 identified as a potential core and hub microbe in the leaf microbiome of *A. thaliana*  
73 since it showed persistence over several years and high interconnection in the microbial  
74 interaction network [14, 15]. *Albugo* infection was also shown to affect both epiphytic  
75 and endophytic bacterial colonization by reducing alpha diversity and secretion of an-  
76 timicrobial peptides [15, 16]. However, it is not clear how the microbiome of the leaf  
77 differs in plants infected or not with *Albugo*, and which microbial strains have the po-  
78 tential to promote or reduce infection with *Albugo*.

79  
80 In this study, we analysed the microbiome of *Arabidopsis thaliana* over a period of

81 six years as collected and described by Mahmoudi et al. [17]. High-throughput se-  
82 quencing analysis revealed differences in the microbiota composition associated with  
83 *Albugo*-infected and uninfected plants across host genotypes and sampling sites. Using  
84 statistical and machine learning classification algorithms, we identified candidate mi-  
85 crobes predictive of infected and uninfected states. Candidate microbes were shown to  
86 be distributed in different clusters in microbial interaction networks, highlighting their  
87 importance in the community's stability. Co-inoculation assays in *A. thaliana* confirmed  
88 the potential of the health-associated microbial communities (HCom) to reduce *Albugo*  
89 infection. In comparison, disease-associated microbial communities (DCom) exhibited a  
90 range of functions, from minimal effects to partial pathogen suppression, likely through  
91 niche competition. These findings highlight the functional redundancy of microbial  
92 communities from different phylogenetic groups in manipulating plant health outcomes  
93 and demonstrate the power of machine learning in informing biocontrol strategies.

94  
95

## 96 Results

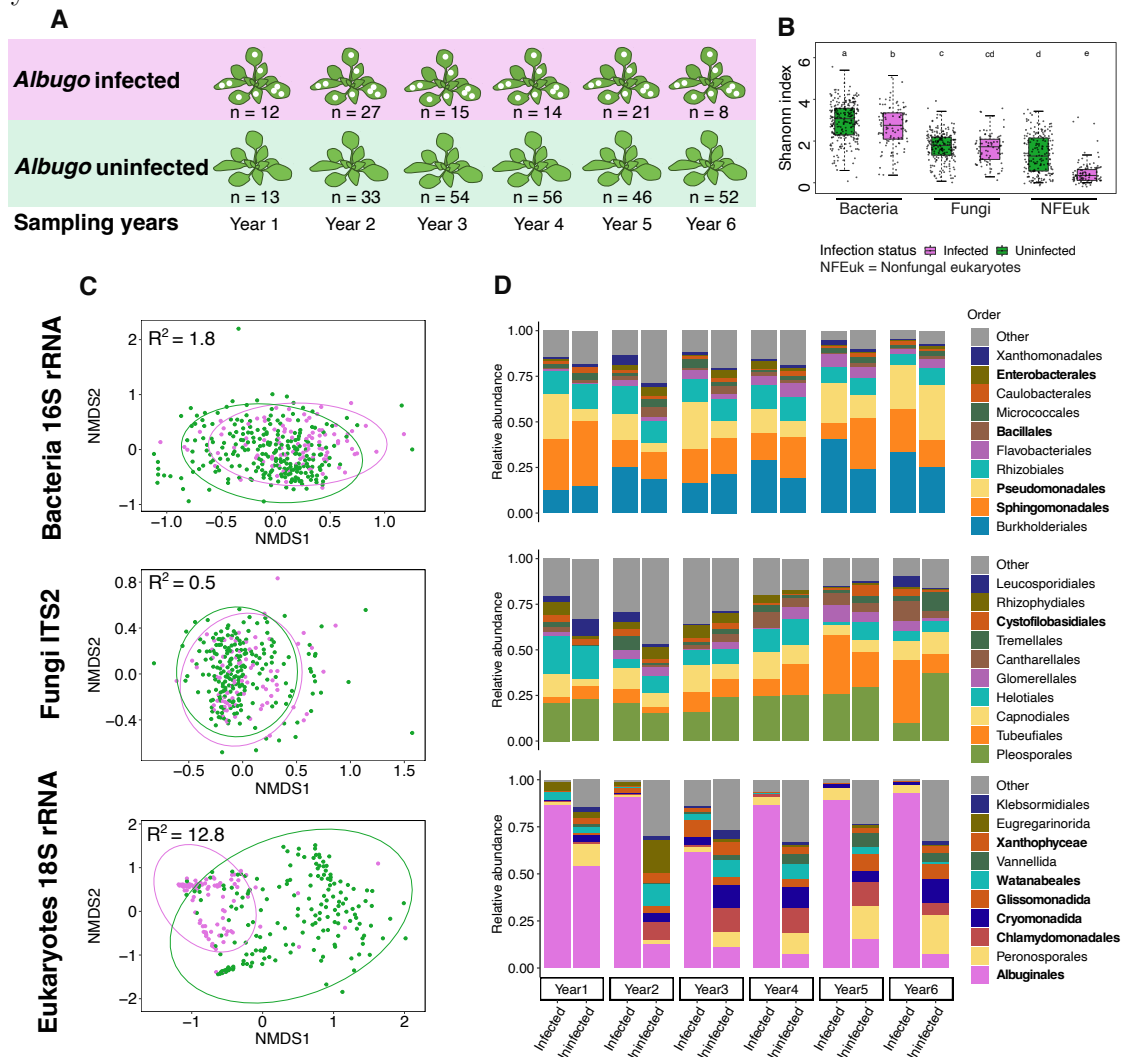
### 97 Comparison of phyllosphere microbiome in natural *A. thaliana* 98 populations: uninfected vs. infected with the obligate biotrophic 99 oomycete pathogen *Albugo*

100 To investigate the diversity and compositional dynamics of the phyllosphere micro-  
101 biome in the presence of the obligate biotrophic oomycete pathogen *Albugo*, we used  
102 a microbiome dataset described in Mahmoudi et al. [17]. In their study, *A. thaliana*  
103 samples were collected from six sites near Tübingen (southern Germany) with stable *A.*  
104 *thaliana* populations, with sampling repeated over six consecutive years (2014-2019).  
105 Genomic DNA was extracted from epiphytic and endophytic microbial communities,  
106 followed by amplicon sequencing for bacterial 16S rRNA, fungal ITS2, and eukaryotic  
107 18S rRNA. For the 18S eukaryotic data, fungal microbes were excluded, resulting in the  
108 nonfungal eukaryotes (NFEuk) dataset [17]. Here, we used the endophytic microbiomes  
109 for further analysis. Since *Albugo* was the major pathogen associated with *A. thaliana*  
110 at the time of sampling, the samples were categorized as infected or uninfected based  
111 on the presence or absence of white rust on the leaves (Fig. 1A).

112 A diversity analysis was conducted to compare the leaf-associated microbial commu-  
113 nities between infected and uninfected plants. Alpha diversity (within-sample diversity,  
114 measured by Shannon's index) demonstrated that, on average, infected plants exhibited  
115 a 1.1-fold and 2.8-fold reduction in bacterial and NFEuk community diversity, respec-  
116 tively, in comparison to uninfected plants (Tukey's HSD test,  $P < 0.05$ ). However, no  
117 significant differences were observed in fungal communities (Fig. 1B). A permutational  
118 multivariate analysis of variance (PERMANOVA) demonstrated that the 'infection sta-  
119 tus' of the plants explained 1.8% of the variation in bacteria, 0.5% in fungi, and 12.8%  
120 in nonfungal eukaryotes. These variations were further visualized using non-metric  
121 multidimensional scaling (NMDS), which revealed that the infected plants were most  
122 clearly separated from the uninfected groups in the NFEuk communities, followed by  
123 bacteria and less so in the fungal communities (Fig. 1C).



124 Differences between infected and uninfected communities were found to be associ-  
 125 ated with the enrichment of major microbial orders over sampling years (Wilcoxon test,  
 126  $P < 0.05$ ) (Fig. 1D, see also Fig. S1). Among bacteria, *Sphingomonadales*, which has  
 127 been demonstrated to be beneficial for plant health and productivity, [18] was 1.3 times  
 128 more abundant in uninfected plants, whereas *Pseudomonadales* was 1.5 times more  
 129 abundant in infected samples, based on a comparison of mean relative abundances.  
 130 Among the fungal orders, only the Basidiomycete yeast *Cystofilobasidiales* exhibited  
 131 a slightly higher significant abundance in infected plants (1.02 times more). Among  
 132 nonfungal eukaryotic orders, as expected, the order *Albuginales* (including *Albugo*)  
 133 showed a 6.7 times increase in infected plants. Interestingly, the most abundant orders  
 134 of green algae (*Watanabeales*, *Xanthophyceae* and *Chlamydomonadales*) were 5.9-14.5  
 135 times more abundance in uninfected plants, similar to *Cercozoa* (*Glissomonadida* and  
 136 *Cryomonadida*), with 3.0 and 7.6 times higher abundance in uninfected plants, respec-  
 137 tively.



138

139 **Figure 1. Diversity and composition of leaf microbial communities in *Al-***  
 140 ***bugo*-infected and uninfected plants.** (A) The number of annually sampled plants  
 141 per group (97 infected and 254 uninfected). (B) Alpha diversity is measured by Shan-  
 142 non's H index and represents the within-sample diversity of infected and uninfected

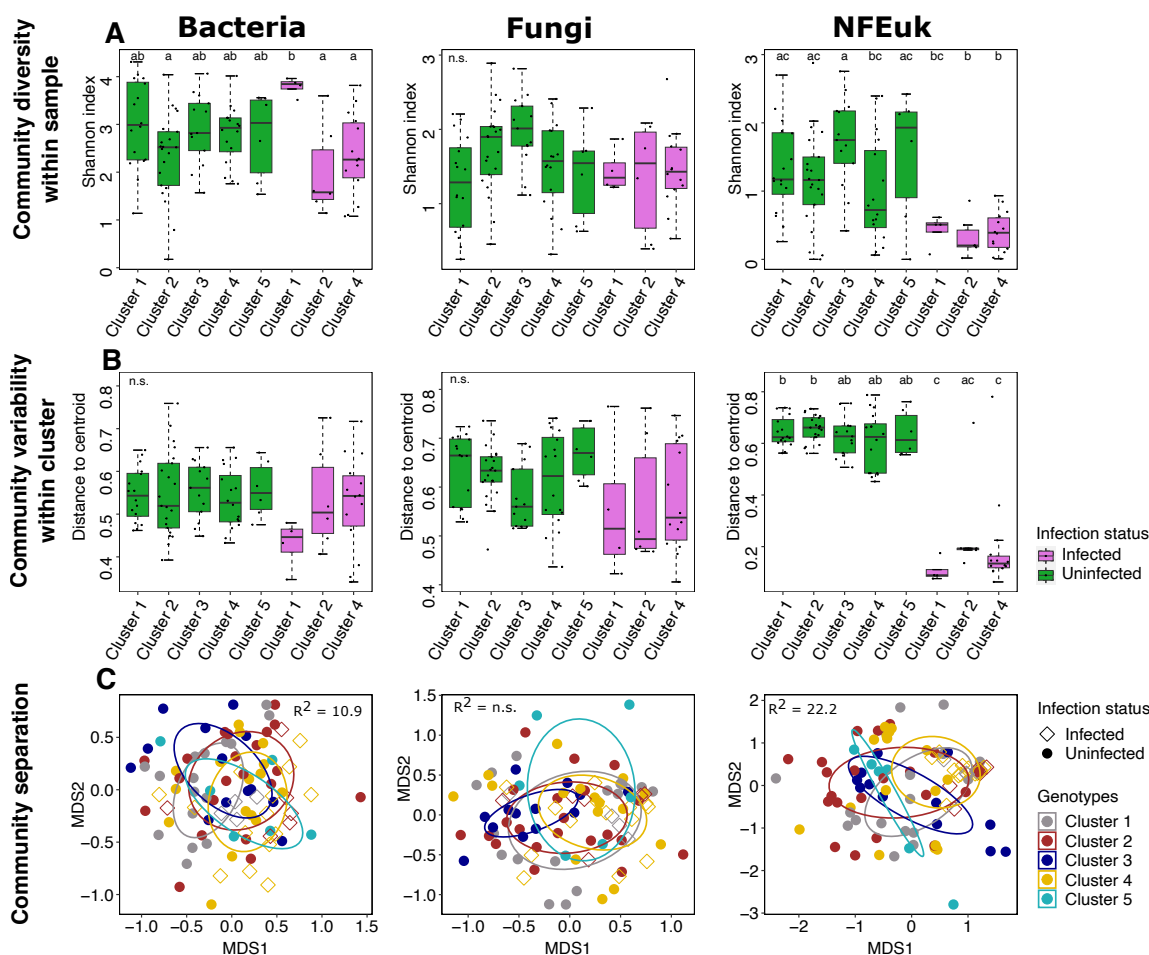
143 samples in bacteria, fungi, and nonfungal eukaryotic communities. Box plots display  
144 individual samples as dots. Different letters indicate statistically significant differences  
145 between groups (Tukey HSD's test,  $P < 0.05$ ). (C) Separation of infected (purple) and  
146 uninfected (green) samples using non-metric multidimensional scaling analysis (NMDS)  
147 based on Bray-Curtis dissimilarities. Each dot represents a single sample. (D) His-  
148 tograms show the relative abundance of bacteria, fungi, and nonfungal eukaryotic com-  
149 munities at the order level, categorized by plant infection status (infected/uninfected)  
150 over six years. Taxa with significant differences between infection stages are indicated  
151 in bold (see Fig. S1).

152

## 153 **Genotype variation in infected and uninfected *A. thaliana* cor-** 154 **relates with microbiome diversity**

155 To investigate whether leaf microbiome infection varies by host genotype, we used the  
156 whole genome sequencing data that was conducted to identify genotype clusters based  
157 on single nucleotide polymorphisms (SNPs) analysis [17]. Among the five identified  
158 clusters, three (clusters 1, 2, and 4) contain samples susceptible to *Albugo* infection.  
159 Clusters (2 and 4) exhibited significantly lower alpha diversity in NFEuk communities  
160 of infected plants as compared to uninfected clusters (Dunn test,  $P < 0.05$ ) (Fig. 2A). A  
161 similar pattern was observed in the variability within clusters (i.e., how far each sample  
162 was from the group's central point or centroid). Specifically, the clusters that were  
163 susceptible to infection showed less variability in this distance, meaning the samples in  
164 these groups were more similar to each other (Dunn test,  $P < 0.05$ ) (Fig. 2B). However,  
165 no significant differences were found in fungal communities. Microbiome compositional  
166 variation, visualized using NMDS, showed that samples from cluster 4 more clearly  
167 separated the infected plants, while cluster 5 distinguished uninfected samples in the  
168 NFEuk community (Fig. 2C, PERMANOVA,  $P < 0.05$ ). These patterns were less  
169 pronounced in bacterial communities (explaining 10.9% of the variation vs. 22.2%  
170 variation in NFEuk community) and were not significant in fungal communities (Fig.  
171 2C).

172 Interestingly, the susceptible clusters were distributed across different sampling sites,  
173 suggesting that, in addition to host genotypes, other abiotic factors might contribute  
174 to plant susceptibility to infection. Notably, genotype cluster 5, which contained only  
175 uninfected plants, was exclusively found at the ERG site (Table. S1). Patterns related  
176 to sampling sites were observed: two sites (K69 and PFN) had no infected plants, while  
177 four sites (EY, WH, JUG, and ERG) contained both infected and uninfected plants  
178 (Fig. S2). Site EY exhibited lower alpha diversity in bacterial communities of infected  
179 plants compared to uninfected plants in the same group, while in NFEuk communities,  
180 three sites (EY, WH, and JUG) showed lower alpha diversity in infected plants (Dunn  
181 test,  $P < 0.05$ ) (Fig. S2A). Regarding community variability, only site ERG showed  
182 reduced variability in infected plants within bacterial communities. In contrast, in  
183 NFEuk communities, the variability of infected plants was consistently lower than that  
184 of uninfected plants across all sites (Dunn test,  $P < 0.05$ ) (Fig. S2B). The combination  
185 of infection status and sampling site explained 8.4%, 7.8% and 21.2% of the variation  
186 in microbial communities for bacteria, fungi and NFEuk, respectively (Fig. S2C).



187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

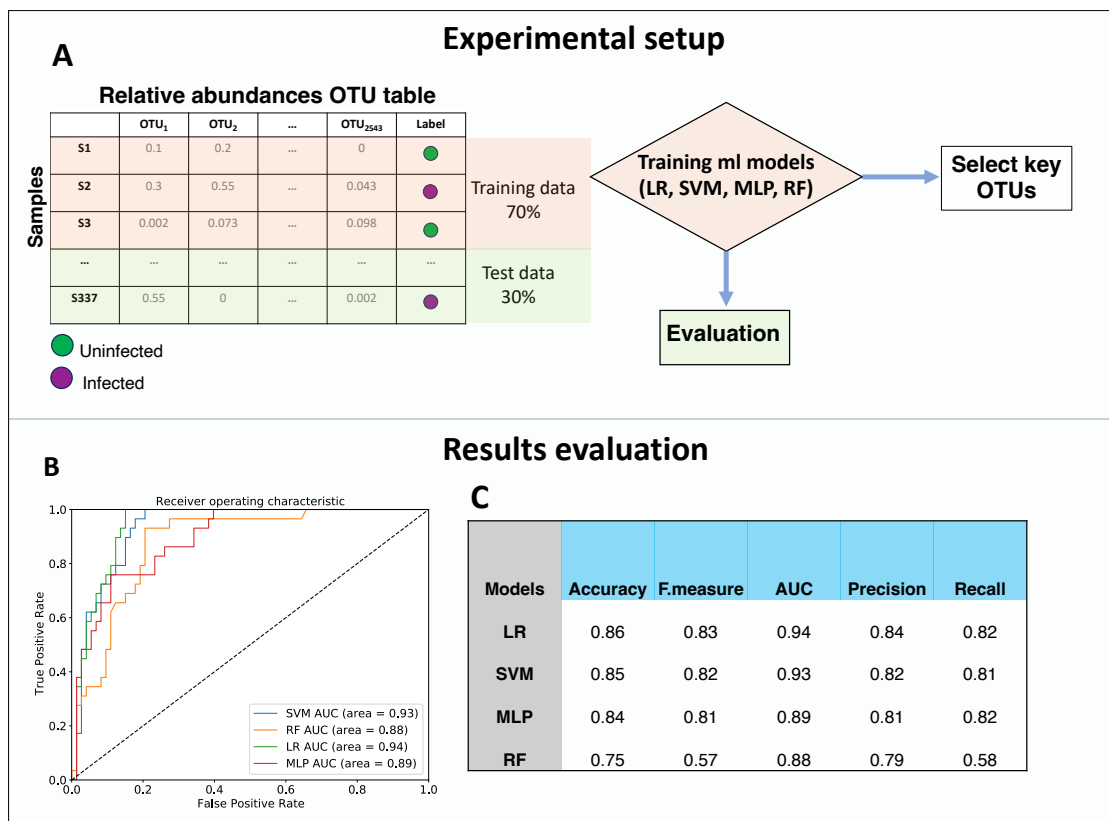
206

**Figure 2. Diversity and separation of leaf microbial communities of infected and uninfected plants among plant genotypes.** (A) Alpha diversity, measured by Shannon's H index, represents the within-sample diversity in microbial communities across infected and uninfected samples of different genotype clusters. (B) Within-cluster variability, quantified as the distance of individual samples from the centroid of their respective genotype cluster, illustrates the variation among microbial communities within each cluster. Different letters indicate statistically significant differences between groups (Dunn test,  $P < 0.05$ ). (C) Non-metric multidimensional scaling (NMDS) plots, based on Bray-Curtis dissimilarities, display the separation between infected and uninfected samples across genotype clusters. Explained variance ( $R^2$  values) from PERMANOVA models (Bray-Curtis dissimilarities), illustrating the impact of genotype clusters and infection status on the structure of leaf microbial communities.

## Identification of a microbial signature for predicting infected and uninfected leaves using machine learning models

We hypothesized that the plant microbiome is composed of distinct health-associated microbial communities (HCom) in uninfected plants and disease-associated microbial communities (DCom) in infected plants. These distinct microbial communities can serve as robust indicators of infection, enabling accurate discrimination between infected

207 and uninfected samples. To investigate this hypothesis, we used machine learning  
 208 classification models, including random forest (RF), support vector machine (SVM),  
 209 and logistic regression (LR), which are well-known for their interpretability and multi-  
 210 layer perceptron (MLP) (Fig. 3A). The training phase used 70% of the sample set,  
 211 consisting of 169 uninfected and 66 infected samples with 2,543 operational taxonomic  
 212 units (OTUs). The remaining 30% (73 uninfected and 29 infected samples) served as  
 213 the test set to assess predictive performance. Four different evaluation metrics were  
 214 employed, resulting in accuracies ranging from 75% to 86% (Fig. 3C). The SVM and  
 215 LR models achieved the highest accuracy of 85% and 86% respectively, with an area  
 216 under the curve (AUC) of 93% and 94%, outperforming the MLP and RF (Fig. 3B  
 217 and 3C). We then analysed the predictive role of each microbe by calculating and  
 218 comparing the feature importance of all the OTUs in the trained classification models  
 219 (SVM, RF, and LR) (Fig. 3A). Comparison of the three models revealed that they  
 220 shared 2,253 OTUs, indicating consistent microbial signatures associated with both  
 221 groups (infected/uninfected) that contributed significantly to the classification process  
 222 (Fig. S3A). Using recursive feature elimination with cross-validation, we identified the  
 223 most crucial OTUs for classification, referring to them as HCom and DCom. The  
 224 results showed that RF had the highest accuracy, achieving 91% accuracy (Fig. S3B)  
 225 with 40 selected OTUs (Fig. 4A). The LR model reached 87% accuracy with 4 selected  
 226 characteristics (Fig. 4B). In contrast, the SVM reached 86% accuracy with 53 selected  
 227 OTUs (Fig. 4C). It is interesting to note that four OTUs were shared by all three  
 228 models that originated from LR, and 13 OTUs were shared by SVM and RF (Fig. 4  
 229 and Fig. S3C).



230

231 **Figure 3. Classification of plant infection status and feature selection by**  
 232 **machine learning classifiers. (A) Workflow illustrating the methodology employed**

233 to distinguish infected samples from uninfected ones using machine learning models.  
234 The objective was to classify leaves uninfected and infected based on the observed  
235 symptoms of *Albugo* infection, utilizing the relative abundance OTU table of bacteria,  
236 fungi, and nonfungal eukaryotes. Four machine learning classifiers, namely support  
237 vector machine (SVM), random forest (RF), logistic regression (LR), and multilayer  
238 perceptron (MLP), were trained on 70% of the samples. The trained models were  
239 evaluated using the remaining 30% of the dataset. Feature importance (to select key  
240 important OTUs for classification) was extracted from the trained models and using  
241 the recursive feature elimination method (Figure. 4). (B) Receiver operating charac-  
242 teristic (ROC) curves. The area under the curve (AUC) values indicate the ability of  
243 each classifier to distinguish between infected and uninfected samples, with higher AUC  
244 values indicating better performance. (C) Additional performance metrics (accuracy,  
245 f-measure, precision, and recall) for each classifier on the test set.

246

247

248 Additionally, we hypothesized that HCom microbes can reduce the pathogenicity of  
249 *Albugo*, in contrast to DCom microbes. To test this hypothesis, we selected four can-  
250 didate microbes from bacterial, fungal, and non-fungal eukaryotic groups within each  
251 category (HCom and DCom) (Fig. 4, microbes in bold). The SVM and LR models  
252 provide both positive and negative coefficients to determine microbial importance in  
253 classifying infected versus uninfected leaves. We assigned scores based on these coeffi-  
254 cients, with negative scores indicating HCom OTUs as indicators of the uninfected class  
255 and positive scores representing DCom OTUs as indicators of the infected class (see Fig.  
256 4B and 4C). However, the RF model only provides positive scores, necessitating further  
257 examination of the relationship between the selected OTUs and their respective classes  
258 (Fig. 4A). Other selection criteria included diverse representations of bacteria and eu-  
259 karyotes, as well as laboratory availability. Among the HCom bacterial candidates,  
260 *Methylobacterium* OTU3 (*Methylobacterium goesingense*) and *Sphingomonas* OTU15  
261 (*Sphingomonas melonis*) were selected, while *Cystofilobasidium* OTU126 (*Cystofilo-*  
262 *basidium macerans*) and *Rhagostoma* OTU3 (*Rhagostoma epiphylla*) were selected for  
263 fungi and nonfungal eukaryotes, respectively. The Dcom candidates included *Duganella*  
264 OTU4 (*Duganella zooglooides*) and *Pseudomonas* OTU6 (*Pseudomonas viridiflava*)  
265 from the bacterial group, *Plectosphaerella* OTU16 (*Plectosphaerella niemeijerorum*) as  
266 fungal representative, and *Albugo* OTU5 (*Albugo*), representing nonfungal eukaryotes  
267 (Fig. 4A-C).



268

269 **Figure 4. Selected HCom and DCom microbes as indicators of uninfected and infected samples.** (A) Histogram showing the 40 OTUs selected by recursive  
 270 and infected samples. (A) Histogram showing the 40 OTUs selected by recursive feature elimination with k-fold cross-validation using random forest (RF). The x-axis  
 271

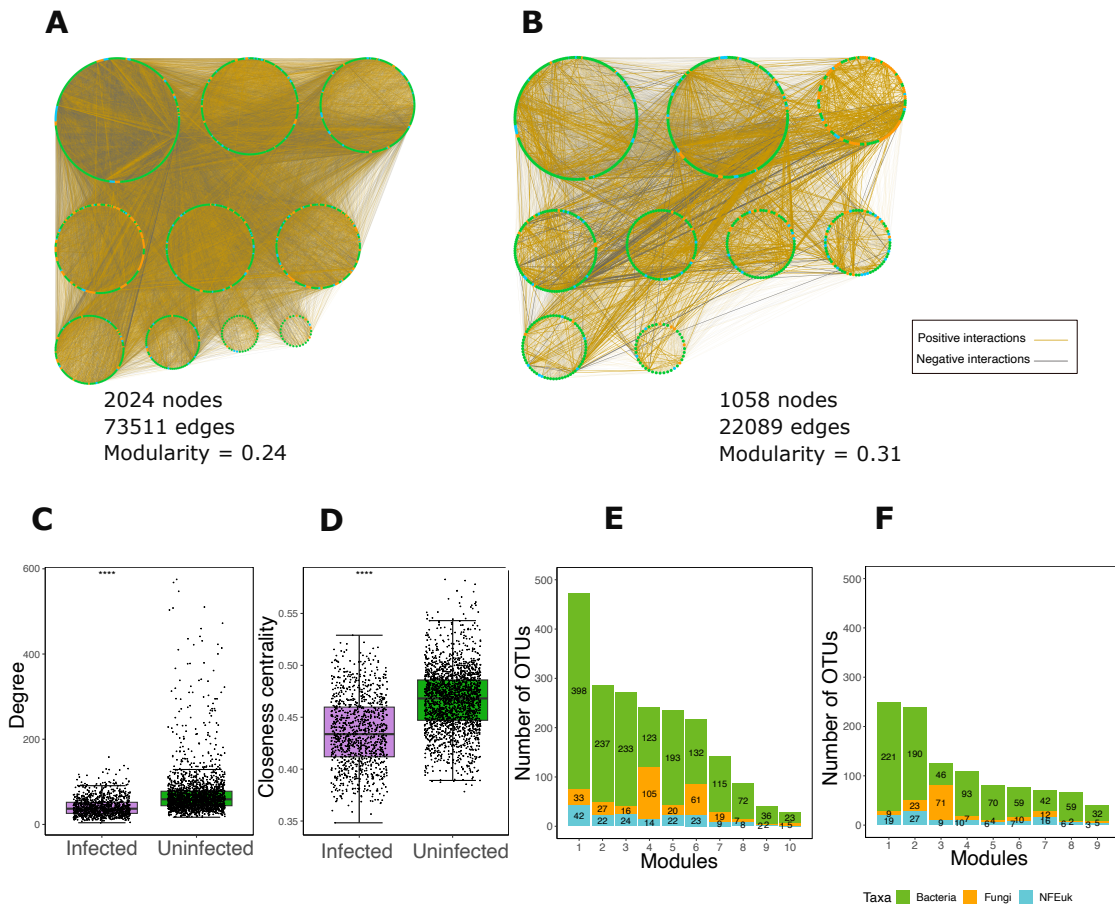


272 indicates the importance of each OTU for classification (values are normalized between  
273 0 and 1). (B) The 4 OTUs selected by recursive feature elimination with k-fold cross-  
274 validation using logistic regression (LR). (C) Normalized coefficient values of the 53  
275 OTUs selected by recursive feature elimination with k-fold cross-validation using sup-  
276 port vector machine (SVM). Negative values (B and C) indicate OTUs with high scores  
277 in discriminating uninfected leaves (HCom), while positive values indicate high scores  
278 in discriminating infected samples (DCom). (D) Bar plots illustrating the aggregated  
279 relative abundances of OTUs in infected and uninfected samples. The microbes high-  
280 lighted in bold (left) were selected for further experimental analysis.

281

## 282 **Infection reduces microbial network complexity and increases** 283 **compartmentalization in community structure**

284 Microbial networks are valuable for identifying potential interactions among microor-  
285 ganisms within a given community. This is achieved by correlating the abundances of  
286 different species. Therefore, we investigated the microbial networks in both infected  
287 and uninfected samples. The network resulting from the uninfected samples showed 1.9  
288 times greater number of nodes (OTUs) and 3.3 times more edges (connections between  
289 OTUs) when compared to the one generated from the infected samples (2,024 nodes  
290 and 73,511 edges vs. 1058 nodes and 22,089 edges, respectively) (Fig. 5A vs. Fig.  
291 5B). Notably, the topological characteristics of the uninfected network showed higher  
292 degree (more connections between microbes) and closeness centrality (microbes more  
293 closely connected to others) values ( $P < 0.001$ ) (Fig. 5C and 5D). To assess the ef-  
294 fect of infection on the microbial communities' compartmentalization, we compared the  
295 modularity of the constructed networks, a feature representing the degree of functional  
296 division and ecological niches within the microbial community [19]. The results showed  
297 that both networks contained a comparable number of modules, with the uninfected  
298 network having ten modules and the infected network having nine modules. However,  
299 the modularity value of the network derived from infected samples (0.31) was slightly  
300 higher than that of the network derived from uninfected samples (0.24) (Fig. 5B and  
301 5A), which indicates that infected generated network is more segmented, suggesting a  
302 stronger tendency to division of microbial communities into distinct functional groups  
303 or ecological niches. Notably, all the modules contained OTUs from different taxonomic  
304 categories of bacteria, fungi, and nonfungal eukaryotes (Fig. 5E and 5F). We examined  
305 the microbial interactions of HCom and DCom OTUs in both uninfected and infected  
306 networks. Results showed that these OTUs displayed distinct connectivity patterns. In  
307 the network generated from uninfected samples, the OTUs formed 72 OTUs with 493  
308 edges (Fig. S4A). In contrast, these OTUs exhibited fewer connections in the infected  
309 network, resulting in 64 nodes with 193 edges (Fig. S4B). Moreover, those OTUs are  
310 more sparsely distributed across different modules in the infected generated network  
311 compared to the uninfected network (Fig. S4D vs. Fig. S4C). The changes in connec-  
312 tivity patterns, particularly among the OTUs in the infected network, further highlight  
313 the reduced complexity and increased structural division in networks generated from  
314 infected samples.



315

316 **Figure 5. Changes in microbial co-abundance networks of infected and**  
 317 **uninfected plants.** Co-abundance networks for both uninfected (A) and infected  
 318 (B) samples, where nodes (circles) represent OTUs and edges (the connection between  
 319 OTUs) indicate correlations between these OTUs. Nodes are color-coded by micro-  
 320 bial taxa and grouped based on modularity clustering. Box plots show features per  
 321 node, i.e., degree (C) and closeness centralities (D) in infected and uninfected net-  
 322 works. Significance values indicate differences between groups based on the Wilcoxon  
 323 test ( $****P \leq 0.0001$ ). Histograms illustrate the distribution of OTUs within modules  
 324 for the network of uninfected (E) and infected (F) samples, respectively. These his-  
 325 tograms are further color-coded to distinguish microbial taxa, with green representing  
 326 bacteria, orange representing fungi, and blue representing NFEuk.

327

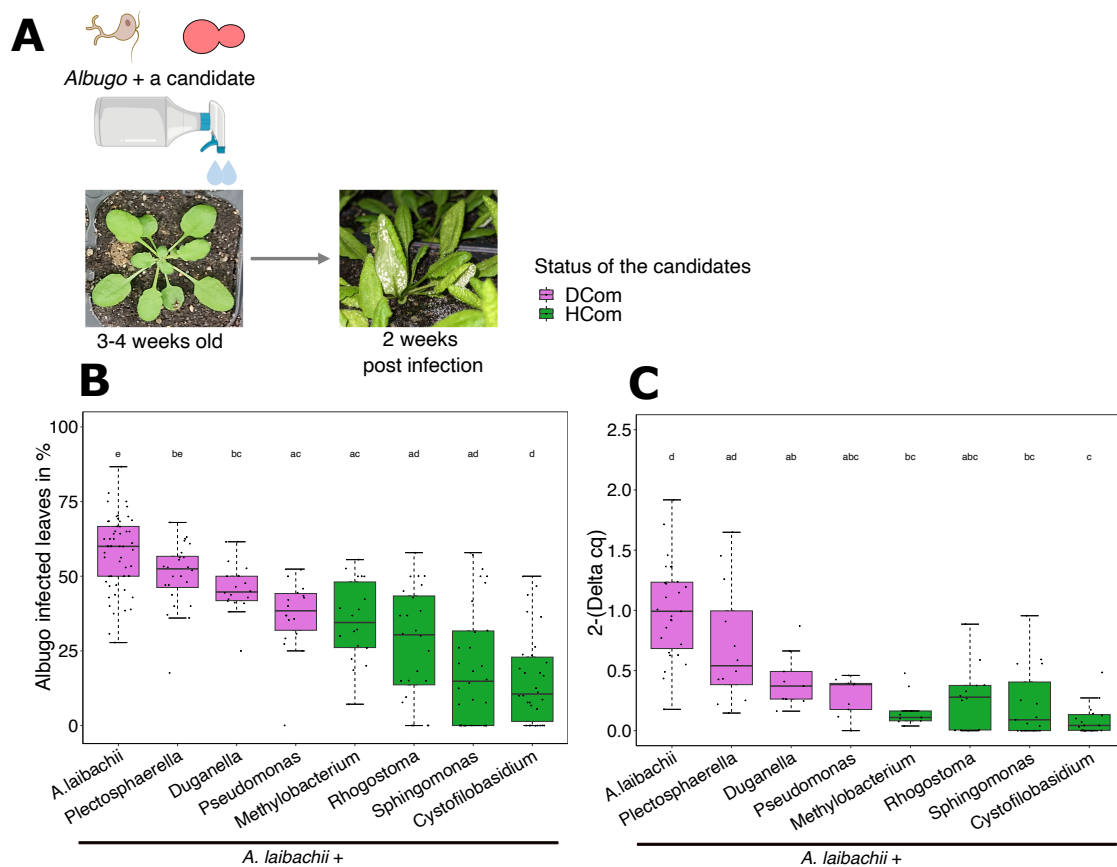
## 328 HCom confer protection against *Albugo* infection to varying 329 degrees

330 To investigate the protective effects of the selected microbes (Fig. 4) against the in-  
 331 fection caused by *Albugo*, first, a mixture of *Albugo* and each of the four microbes  
 332 from HCom was sprayed onto *Arabidopsis* leaves (Fig. 6A). The level of protection  
 333 was determined by measuring the percentage of infected leaves (Fig. 6B). All four  
 334 candidates significantly decreased the infection caused by *Albugo* (Dunn test,  $P <$   
 335  $0.05$ ). *Cystofilobasidium* exhibited the most pronounced effect, reducing *Albugo* levels  
 336 by an average of 73%. *Sphingomonas* caused a 66% reduction, followed by *Rhagostoma*

337 and *Methylobacterium*, which resulted in 53% and 40% decreases in *Albugo* infection,  
338 respectively (Fig. 6B). These observations were further confirmed by quantitative poly-  
339 merase chain reaction (qPCR) analysis, which demonstrated that samples exposed to  
340 the uninfected-associated microbes exhibited substantially lower amounts of *Albugo* as  
341 compared to control samples (Dunn test,  $P < 0.05$ ), with the average biomass of *Albugo*  
342 ranging from 72% to 90% (Fig. 6C). These results demonstrate that all the selected  
343 candidates associated with uninfected plants, namely, *Cystofilobasidium*, *Methylobac-*  
344 *terium*, *Rhogostoma*, and *Sphingomonas*, significantly decreased the infection levels of  
345 *Albugo*. The *Methylobacterium*-treated plants exhibited the highest plant biomass, with  
346 an average of 0.96 (g), compared to other treatments (Tukey's HSD test,  $P < 0.05$ ) (Fig.  
347 S5). Microscopy analysis revealed that *Rhogostoma* attached to the *Albugo* spores and  
348 feeds on free-living microbes in the environment (Fig. S6 and supplementary videos 1  
349 and 2). These findings highlight the potential of HCom microbes in protecting against  
350 *Albugo* infection, with varying levels of effectiveness across different microbial taxa.

351  
352 We then investigated the effect of DCom microbes on the pathogenicity of the *Albugo*.  
353 The *in-planta* infection assay demonstrated that *Plectosphaerella* had no significant  
354 effect on the infection level of *Albugo* (Dunn test,  $P > 0.05$ ). However, *Pseudomonas*  
355 and *Duganella* caused a decrease in infection of 36% and 20%, respectively (Dunn test,  
356  $P < 0.05$ ) (Fig. 6B). Likewise, qPCR outcomes supported the observed phenotype: the  
357 biomass of *Albugo* in the control group exhibited no significant changes in comparison  
358 with *Plectosphaerella* (Dunn test,  $P < 0.05$ ), whereas *Pseudomonas* and *Duganella*  
359 caused a 72% and 58% reduction in *Albugo* biomass, respectively (Fig. 6C). The find-  
360 ings reveal that while some DCom microbes are effective in reducing *Albugo* infection  
361 levels, their protective effects are less significant compared to those of HCom microbes.

362 To evaluate the specific effects of each microbe on plant health independently,  
363 we performed spray experiments on gnotobiotic plants, isolating the impact of in-  
364 dividual microbes without the influence of other microbial interactions (Fig. S7).  
365 Three distinct phenotypes were observed. Plants colonized with the DCom bacteria  
366 *Duganella* and *Pseudomonas* exhibited high mortality rates, with 10.0% and 11.3%  
367 survival, respectively, within three weeks post-colonization. As expected, the filamen-  
368 tous pathogenic *Plectosphaerella* and *Albugo* caused characteristic infection symptoms,  
369 namely, brownish leaves and roots and white rust disease, respectively. Forty-nine per-  
370 cent of the plants treated with HCom *Cystofilobasidium* survived three weeks post-  
371 colonization. The healthiest plants were those colonized by the remaining HCom mi-  
372 crobes—*Sphingomonas*, *Methylobacterium*, and *Rhogostoma*—with over 88% of the plants  
373 showing no discernible negative consequences. Overall, these results indicate that most  
374 HCom microbes lead to healthier plants than DCom microbes, underscoring the im-  
375 portance of specific microbial candidates in enhancing plant vitality and resistance to  
376 pathogen *Albugo*.



377

378 **Figure 6. Effects of HCom and DCom on infection caused by *Albugo*.**  
 379 (A) Three to four weeks old *Arabidopsis* plants were co-inoculated with *Albugo* and  
 380 each of the indicator microbes (HCom and DCom microbes) as identified in Figure 4.  
 381 Symptoms were recorded 2 weeks after infection. (B) Box plots showing the percent-  
 382 age of leaves infected with *Albugo* in the presence of HCom (green) and DCom strains  
 383 (purple). (C) Relative quantification of *Albugo* biomass in response to each indicator  
 384 microbe was conducted through qPCR targeting the *Albugo* EF1- $\alpha$  gene and normal-  
 385 izing to the *A. thaliana* EF1- $\alpha$  gene. The relative biomass was then calculated via the  
 386 ddcq method. Statistically significant differences between the groups were evaluated  
 387 using the Dunn test, with different letters indicating significant differences ( $P < 0.05$ ).

## 388 Discussion

389 In this study, we identify microbial signatures that distinguish between infected and un-  
 390 infected plants and explore their potential for developing effective probiotics to promote  
 391 plant health. The management of plant health through natural probiotics has gained  
 392 significant ecological and economic interest. Various microbes and synthetic microbial  
 393 communities have been found to increase plant resistance to pathogens under laboratory  
 394 conditions [12, 20]. Pathogens are known to impact the phyllosphere microbiome to es-  
 395 tablish their niche [21]. Recent studies show that pathogens, including *Verticillium* and  
 396 *Albugo*, release effector proteins to manipulate the microbial landscape, affecting micro-  
 397 biome composition and function [16, 22, 23]. However, there is a gap in understanding

398 these interactions under natural field conditions, where pathogens face a complex and  
399 heterogeneous host microbiome and abiotic stressors [17, 24]. Our study addresses this  
400 by investigating microbial communities associated with *A. thaliana* under natural field  
401 conditions over six years, focusing on changes in the presence or absence of *Albugo*  
402 infection.

403 Agler et al. [15] reported that the diversity of bacterial communities was lower  
404 in *Albugo*-infected plants. Extending this observation, our analysis of six-year time  
405 series data, along with 18S eukaryotic data, revealed that not only bacterial and fungal  
406 diversity affected by infection, but there was an even more pronounced loss of diversity  
407 within 18S nonfungal eukaryotic groups (Fig. 1). This loss was observed in 3 out of 5  
408 host genotypes that showed susceptibility to infection, leading to up to a 22% variation  
409 in the microbiome composition (Fig. 2). We attribute this reduction in diversity to a  
410 significant increase in the pathogen population, which can disrupt the balance of the  
411 established microbial community [10, 25]. In the case of *Albugo* infection, this imbalance  
412 may result from the pathogen's efficient suppression of host defenses, allowing non-  
413 host pathogens to proliferate [26], or from the release of microbial-modulating effector  
414 proteins [16]. Both mechanisms likely contribute to the establishment and maintenance  
415 of the pathogen's niche within the host.

416 To identify signatures, the pathogen *Albugo* imposes on the *A. thaliana* microbiome,  
417 we used machine learning prediction models to identify non-linear relationships and  
418 manage the complexity of the high-dimensional data [27]. Therefore, we conducted a  
419 systematic analysis of variations between uninfected and infected plant statuses, achiev-  
420 ing highly accurate classifications up to 91% (Fig. 3 and Fig. S3). These findings  
421 suggest the presence of predictive microbial signatures in these groups. Using the fea-  
422 ture selection technique, we pinpointed 3.1% of OTUs, including bacteria, fungi, and  
423 nonfungal eukaryotes, as key discriminators between infected and uninfected plants,  
424 corresponding to HCom and DCom microbes (Fig. 4 and Fig. S3). These results  
425 highlight that, despite the vast microbial diversity, only a small subset is significantly  
426 associated with plant health outcomes. Interestingly, we found that HCom and DCom  
427 microbes are distributed across various network modules (Fig. S4). Modularity in mi-  
428 crobial interaction networks can indicate diverse habitats, varying selective pressures,  
429 and phylogenetic clustering of related species [28], highlighting the crucial role of these  
430 microbes in the functionality of different community modules within the overall micro-  
431 bial community.

432 Building on these findings, we aimed to test whether microbial isolates matching the  
433 taxa identified by our machine learning prediction models could promote health or dis-  
434 ease status in field conditions. Here, we could show that the predicted HCom and DCom  
435 microbiome comprises distinct phylogenetic groups of bacteria, fungi, and nonfungal  
436 eukaryotes (Fig. 4). Interestingly, some taxa were represented in both infected and un-  
437 infected samples. These included known pathogens like the fungal Ascomycetes genus  
438 *Alternaria*, saprophytic fungi such as the Basidiomycete yeast genera *Dioszegia* and  
439 *Cystoflbasidium* and the Ascomycetes order *Heliales*. In addition, the bacterial gen-  
440 era *Pseudomonas* and *Sphingomonas* were also present in both groups. *Pseudomonas*  
441 is particularly notable for its ambivalent behavior, with a broad range of sublineages  
442 that can be either pathogenic or protective [29], either through microbe-microbe inter-  
443 actions or host interactions ([30]. This diversity was reflected in our dataset, where  
444 *Pseudomonas* OTU5 and OTU6 were predicted to be associated with infected sam-



445 ples, while *Pseudomonas* OTU110 was associated with health. A similar diversity has  
446 been reported for *Sphingomonas* [31], which, similar to *Pseudomonas*, can protect *A.*  
447 *thaliana* from pathogenic bacteria [32].

448 Given their relevance to *A. thaliana* health, we selected *Pseudomonas* and *Sphin-*  
449 *gomonas* isolates for testing. Our tests confirmed that the *Pseudomonas* isolate was  
450 pathogenic, while the *Sphingomonas* isolate did not harm the plant when colonizing  
451 the host alone (Fig. S7). From our HCom microbes, *Methylobacterium* and *Sphin-*  
452 *gomonas* showed further robust suppressive effects (Fig. 6). This is consistent with  
453 previous observations that *Sphingomonas* is able to suppress *Pseudomonas syringae*  
454 disease symptoms [32]. Further to the genera of *Sphingomonas* and *Pseudomonas*,  
455 members of the *Methylobacterium* genus and the family *Oxalobacteriaceae* are highly  
456 adapted to live on plants with high abundance and diversity [33, 34]. While *Methy-*  
457 *lobacterium* was predicted to be HCom in our analyses, the *Oxalobacteriaceae* genus  
458 *Duganella* was strongly associated with *Albugo* infection. Our experiment validated  
459 that the *Duganella* isolate exhibited pathogenic behavior (Fig. S7).

460 Among the HCom selected microbes, the basidiomycete yeast *Cystofilobasidium ex-*  
461 *hibited* the strongest protective effects (Fig. 6). *Cystofilobasidium* is particularly note-  
462 worthy, as several basidiomycete yeasts have been empirically identified as biocontrol  
463 agents against postharvest diseases in various fruits. This includes our computationally  
464 identified yeasts *Vishniacozyma* [35] and *Leucosporidium* [36], as well as our experimen-  
465 tally validated yeast, *Cystofilobasidium* [37]. Beyond yeast, another crucial group influ-  
466 encing plant health is the *Cercozoa*, which plays a significant role in shaping bacterial  
467 and fungal communities through selective predation [38, 39]. Our study demonstrates  
468 that the HCom *Rhogostoma* can suppress infection. As primary microbial predators,  
469 *Rhogostoma* directly targets bacterial and fungal pathogens, exerting a consumptive ef-  
470 fect on various pathogenic strains [40, 41]. The reduction in infection levels may occur  
471 through direct consumption of *Albugo* zoospores or by preying on bacterial and fun-  
472 gal species that facilitate *Albugo* infection. Additionally, *Rhogostoma* may indirectly  
473 contribute to plant health by promoting beneficial microbes and enhancing interac-  
474 tive activities, such as biofilm formation, as has been recently shown for the protist  
475 *Cercomonas lenta* in the rhizosphere [42].

476 In summary, our findings support the concept of functional similarity [43], where  
477 diverse microbial taxa, including bacteria, fungi, and nonfungal eukaryotes, share over-  
478 lapping ecological roles that contribute to plant health. This concept suggests that  
479 while different microbial species can perform similar functions, their effectiveness may  
480 vary depending on environmental conditions. Therefore, if abiotic factors (e.g., radi-  
481 ation and humidity) become unfavorable for one group, others can compensate, ensuring  
482 continued protection against pathogens [17]. Leveraging functional similarity presents  
483 promising opportunities for biocontrol methods to reduce pathogen pressure across  
484 diverse environmental conditions. As climate change increases the frequency of envi-  
485 ronmental fluctuations, leading to greater stress on plants, enhancing environmental  
486 resilience through the use of robust probiotics will be essential for future food security.



## 487 Method

### 488 Diversity analysis

489 The OTU tables for bacteria, fungi, and nonfungal eukaryotes, along with the corre-  
490 sponding genotype clusters for each sample, were obtained from the study by Mah-  
491 moudi et al. [17]. The OTU tables were modified by excluding epiphytic samples and  
492 those with fewer than 50 reads. Subsequently, OTU abundance tables were utilized to  
493 compute Shannon’s H diversity index using the ‘estimate-richness’ function in the Phy-  
494 loseq R package [44] for estimating alpha-diversity. To assess between-sample diversity,  
495 relative abundance OTU tables were computed and transformed using  $\log_{10}(x + 1)$   
496 before calculating Bray-Curtis dissimilarities, which were then employed for nonmetric  
497 multidimensional scaling ordination (NMDS) using the ‘ordinate’ function in Phyloseq  
498 [44] R package. Three PERMANOVA analysis on Bray-Curtis dissimilarities were con-  
499 ducted to identify the primary factors (‘infection stages’, ‘combined infection stages and  
500 genotype’ and ‘combined infection stages and sampling sites’) influencing the leaf micro-  
501 biome’s structure, utilizing the ‘adonis2’ function in the Vegan package [45] with 10,000  
502 permutations ( $P < 0.05$ ). A beta-dispersion analysis on Bray-Curtis dissimilarities was  
503 conducted to compare sample-to-sample variation within each group (genotype clusters  
504 and sampling sites) (multivariate homogeneity of group dispersion analysis, “betadis-  
505 per”; Vegan package). The means were compared using the nonparametric multivariate  
506 test for multiple groups (‘dunnTest’ function in the FSA package [46], with Benjamini-  
507 Hochberg adjusted P-values  $< 0.05$ ), and the nonparametric ranked test for two groups  
508 (‘wilcox.test’ function in the stats package [47],  $P < 0.05$ ). All analyses were conducted  
509 in R (version 4.1.2) [48].

### 510 Machine learning analysis

511 The OTU tables of bacteria, fungi, and nonfungal eukaryotes were converted into rel-  
512 ative abundance tables and merged into a single table. The OTU tables were further  
513 filtered to retain only those OTUs present in at least five samples (2543 OTUs and  
514 337 samples). A processed dataset was created using two binary labels, ‘infected’ and  
515 ‘uninfected,’ corresponding to the phenotype of the collected plants. Plants were la-  
516 beled ‘infected’ if white rust disease caused by *Albugo* was observed and ‘uninfected’ if it  
517 was not. The scikit-learn package in the Python programming environment was used to  
518 train and validate machine learning classifiers [49]. First, samples were divided into two  
519 parts: training (70%) and testing (30%) (train\_test\_split function, test\_size=0.3, shuf-  
520 fle=True). Subsequently, machine learning models were trained. The trained models  
521 included Support Vector Machine (svm.SVC function, kernel=‘linear’), Random For-  
522 est (RandomForestClassifier function, n\_estimators=1000, min\_samples\_split=2), multi-  
523 layer perceptron (MLPClassifier function, solver=‘adam’, alpha=1e-5, random\_state=1,  
524 learning\_rate=“adaptive”, max\_iter=500, hidden\_layer\_sizes=(100,100,100)), and Lo-  
525 gistic Regression (LogisticRegression function, default parameters). The labels of the  
526 test sets were predicted (model.predict function). The prediction results were compared  
527 with the actual labels of the samples to calculate the models’ performance in terms  
528 of accuracy (accuracy\_score function), f-measure (f1\_score function, average=‘macro’),  
529 precision (precision\_score, average=‘macro’), and recall (recall\_score function). The

530 area under the curve (AUC) was calculated using false positive and true positive rates  
531 (roc\_curve function) and plotted using the matplotlib package. The importance of each  
532 OTU in the trained models was obtained using model.coef\_ (for Support Vector Machine  
533 and Logistic Regression) and model.feature\_importances\_ (for Random Forest). Recur-  
534 sive feature elimination was used to identify the most important OTUs (HCom and  
535 DCom) for classifying groups (RFECV function, step=1, cv=total number of samples,  
536 shuffle=True, scoring='accuracy', min\_features\_to\_select=1).

## 537 **Microbial network analysis and properties**

538 To construct microbial correlation networks, samples of infected and uninfected plants  
539 were separated, and the OTU tables encompassing bacteria, fungi, and nonfungal eu-  
540 karyotes were merged to conduct a comprehensive examination of microbial interac-  
541 tions. The OTU tables were then filtered to retain only OTUs present in at least  
542 five samples, resulting in 2,543 OTUs across 242 samples for uninfected plants and  
543 1,058 OTUs across 95 samples for infected plants. These filtered OTU tables were  
544 used to calculate correlations using the SparCC algorithm[50], which uses Aitchison's  
545 log-ratio analysis and is specifically designed to handle compositional data with high  
546 sparsity. The SparCC correlation scores were computed on the FastSpar platform [51]  
547 with default parameters. Pseudo P-values were generated through 1000 bootstraps  
548 to assess statistical significance. For further analysis, only correlations meeting the  
549 criteria of  $P < 0.01$  were included. Modularity analysis was performed using Python's  
550 networkx package (version 3.1) [52]. Community detection was applied using the Lou-  
551 vain Community Detection Algorithm (community.community\_louvain.best\_partition  
552 function)) [53] the modularity score was calculated on detected modules (and com-  
553 munity.community\_louvain.modularity functions). We used Cytoscape (version 3.7.1)  
554 [54] to visualize and analyze the microbial interaction networks.

## 555 **Infections of *A. thaliana* leaves and quantification of *Albugo* 556 biomass by qPCR**

557 Overnight liquid cultures of bacteria and yeast were diluted in the fresh medium until  
558 they reached an  $OD_{600}$  of 0.2. The resulting cultures were then centrifuged at 1200 g  
559 for 5 minutes, and the resulting pellets were resuspended in  $MgCl_2$ . A spore and cell  
560 concentration of  $25 \times 10^4$  spores/mL or cells/ml was prepared for *Plectospherella* and  
561 *Rhogostoma*. A spore solution of *Albugo* was prepared by collecting *Albugo laibachii*  
562 Nc14, which had previously been cultured on *A. thaliana* Ws-0. Water was added to  
563 the leaves, and the samples were kept on ice for 1 hour before filtering. The num-  
564 ber of cells (*Rhogostoma*) or spores (*Plectospherella* and *Albugo*) was measured by  
565 taking 50  $\mu$ L of the solution, placing it on a hemocytometer, and examining it un-  
566 der an epifluorescence Axiophot microscope. Approximately 4-5 mL of each sample  
567 were carefully combined with 5-6 mL of *Albugo* solution ( $25 \times 10^4$  spores/mL) and  
568 evenly applied to 4-5 weeks old *A. thaliana* seedlings (Ws-0 accessions) using airbrush  
569 guns. Following two weeks, leaf disease symptoms were assessed, differentiating be-  
570 tween infected and uninfected leaves, and quantified as a percentage, and plants' fresh  
571 weights were measured. The leaves were then stored at  $-80^\circ C$ . DNA extraction was

572 performed using FastDNATM Spin Kit for Soil (MP Bio) as described in the manu-  
573 facturer's protocol. For qPCR, a mixture was prepared consisting of 7.5  $\mu$ L of SYBR  
574 Green supermix, 5  $\mu$ L of DNA (approximately 50 ng), 1.9  $\mu$ L of NFW, and 0.3  $\mu$ L of  
575 forward and reverse primers (10  $\mu$ M each), resulting in a total reaction volume of 15  
576  $\mu$ L. The sample measurements were triplicated using a Bio-Rad CFX Connect real-time  
577 PCR detection system. The quantification of *Albugo* DNA relative to the plant DNA  
578 was determined using the following oligonucleotide sequences: *A. thaliana* EF1- $\alpha$ : 5'-  
579 AAGGAGGCTGCTGAGATGAA-3', 5'-TGGTGGTCTCGAACTTCCAG-3'; *Albugo*  
580 EF1- $\alpha$ : 5'-GTGTTCTGCACATCCACACC-3', 5'-GACCTTGACGGATGAAAGGA-  
581 3'. Cq values obtained during the amplification of oomycete DNA were subtracted  
582 from DNA amplicons of *A. thaliana* (called ddCq). Then, the relative biomass of  
583 *Albugo* for each group was determined using the formula  $2^{-\text{ddCq}}$ . These results were  
584 further normalized by comparing the *Albugo* biomass in each experiment (x) to the con-  
585 trol. Specifically, we calculated the ratio of the ddCq for any treatment in experiment  
586 x to the average ddCq of the *Albugo* control in the same experiment (experiment x).

## 587 Infections of *A. thaliana* leaves with individual strains in sterile 588 conditions

589 Following the protocol of Eitzen et al., [23], sterilized *A. thaliana* seeds (Ws-0 acces-  
590 sions) were sown on 1/2 strength Murashige Skoog (MS) medium and incubated for  
591 3-4 days in a cold room (4C, darkness). The MS plates were then placed in growth  
592 chambers set to 22°C with a short-day light cycle (8 hours of light) and 33-40% humid-  
593 ity. The seedlings were grown under these conditions for 4-5 weeks before inoculation.  
594 Overnight liquid cultures of bacteria and yeast were diluted in the fresh medium until  
595 they reached an OD<sub>600</sub> of 0.2. The resulting cultures were then centrifuged at 1200  
596 g for 5 minutes, and the resulting pellets were resuspended in MgCl<sub>2</sub>. A spore con-  
597 centration of  $25 \times 10^4$  spores/mL was meticulously prepared for *Plectospherella* and  
598 *Rhogostoma*. Five hundred  $\mu$ L of each culture was evenly sprayed onto 4-5-week-old  
599 *A. thaliana* seedlings using airbrush guns. Phenotypes of plants were scored after 2-3  
600 weeks post-infection.

## 601 Microscopical observations

602 A mixture of *Albugo* spores and *Rhogostoma* cells, each at a concentration of approxi-  
603 mately 100,000 cells/ml, was prepared and placed in a plastic petri dish. Observations  
604 were made under an epifluorescence Axiophot microscope (Zeiss, up to 64x magnifica-  
605 tion) and/or inverted microscope (Zeiss LSM880) 2 to 8 days post-mix.

## 606 Availability of data and material

607 OTU tables and scripts are available here

608 [https://gitlab.plantmicrobe.de/maryam\\_mahmoudi/HealthMarkers](https://gitlab.plantmicrobe.de/maryam_mahmoudi/HealthMarkers)

## 609 Competing interests

610 None to declare

## 611 Author Contributions

612 MM, KN, and EK devised the study. MM and LMT performed the experiments. MM,  
613 JA and YH interpreted the data. MM wrote the scripts and visualized the data. MM,  
614 JA, YH, KN and EK contributed to writing and preparing the manuscript. All authors  
615 read and approved the final manuscript.

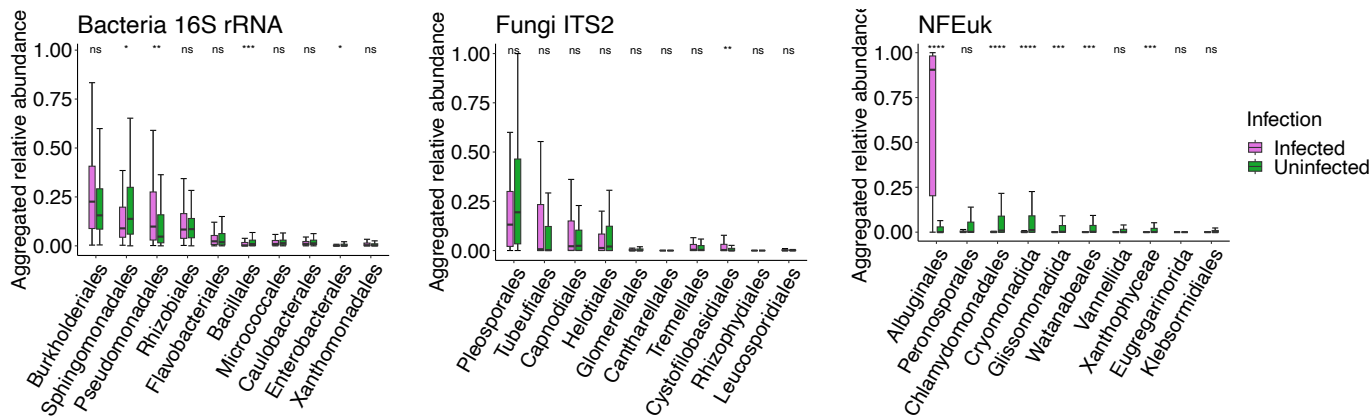
## 616 Acknowledgements

617 We thank Dr. Libera Lo Presti for the critical reading of the manuscript. We further  
618 thank the de.NBI Cloud Storage and bwForCluster BinAC Tübingen provide resources  
619 for storing and analyzing the data.

## 620 Funding

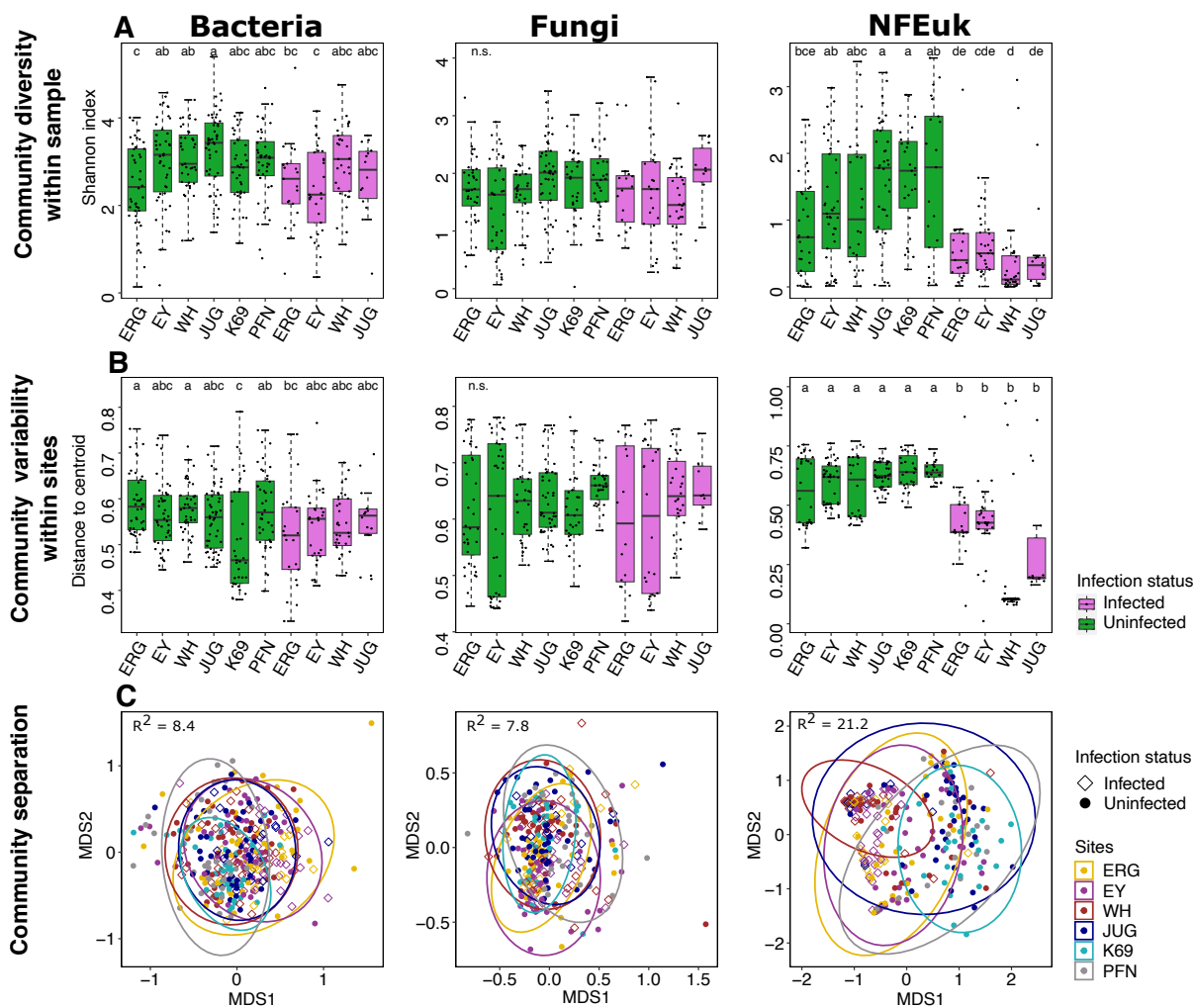
621 This project has been funded by the European Research Council (ERC) under the  
622 DeCoCt research program (grant agreement: ERC-2018-COG 820124), the Deutsche  
623 Forschungsgemeinschaft under Germany's excellence strategy-EXC 2124-390838134,  
624 and the SPP 2125 DECRyPT program from the DFG.

## 625 Supplementary figures



626

627 **Supplementary figure 1. Changes in highly abundant microbial taxa col-**  
628 **onizing *A. thaliana*'s in infected and uninfected leaves.** Box plots (green =  
629 uninfected, purple = infected) show the relative abundance of the orders of bacteria,  
630 fungi and nonfungal eukaryotes in individual samples aggregated by 'infection status'.  
631 Significance values between groups are based on Wilcoxon's test: n.s. ( $P > 0.05$ ), \* ( $P$   
632  $\leq 0.05$ ), \*\* ( $P \leq 0.01$ ), \*\*\* ( $P \leq 0.001$ ), and \*\*\*\* ( $P \leq 0.0001$ ).  
633



634

635

636

637

638

639

640

641

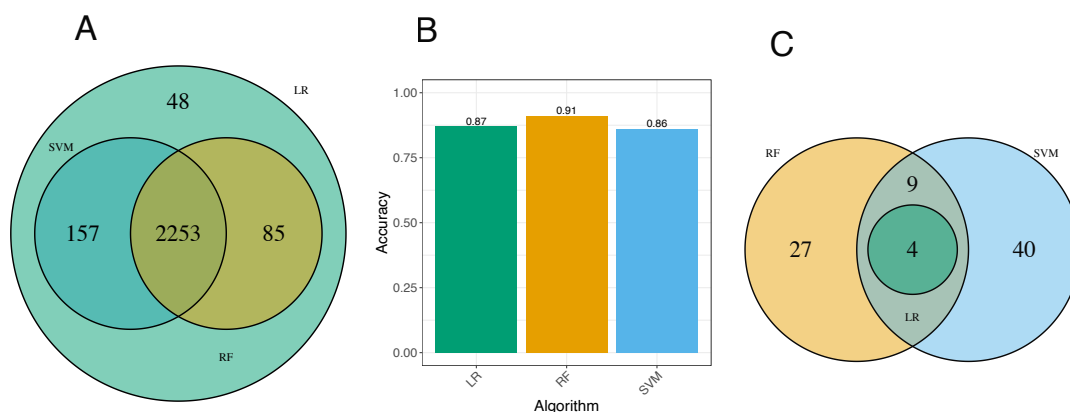
642

643

644

645

**Supplementary figure 2. Diversity and variability of leaf microbial communities of infected and uninfected plants among sampling sites.** (A) Alpha diversity, measured by Shannon's H index and representing the within-sample diversity of infected and uninfected samples among sampling sites in microbial communities. (B) Within-site variability (distance to the group centroid). Different letters indicate statistically significant differences between groups (Dunn test,  $P < 0.05$ ). (C) Non-metric multidimensional scaling (NMDS) plots, based on Bray-Curtis dissimilarities, display the separation between infected and uninfected samples across sampling sites. Explained variance ( $R^2$  values) from PERMANOVA models (Bray-Curtis dissimilarities), illustrating the impact of sampling sites and infection status on the structure of leaf microbial communities.



646

647

648

649

650

651

652

653

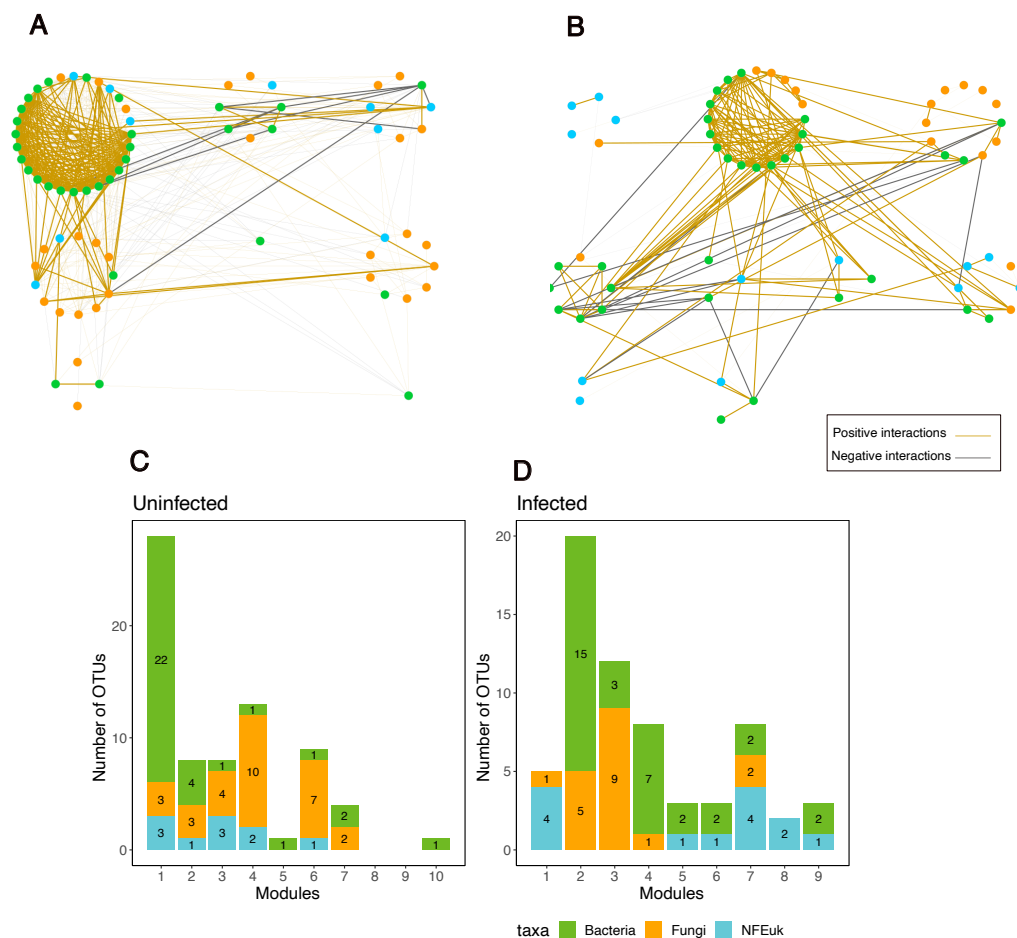
654

655

656

**Supplementary Figure 3. Comparing the importance of microbes for classifying leaves in uninfected and infected using machine learning models.**

To discriminate infected from uninfected samples based on microbial signatures, four classification models were trained (Figure. 3). (A) The Venn diagram represents the number of common microbes with absolute scores greater than 0 among different models during the training phase. (B) Bar plots represent the accuracy of models trained using recursive feature elimination. (C) The Venn diagram represents the number of shared microbes with absolute scores higher than 0 after performing recursive feature elimination.

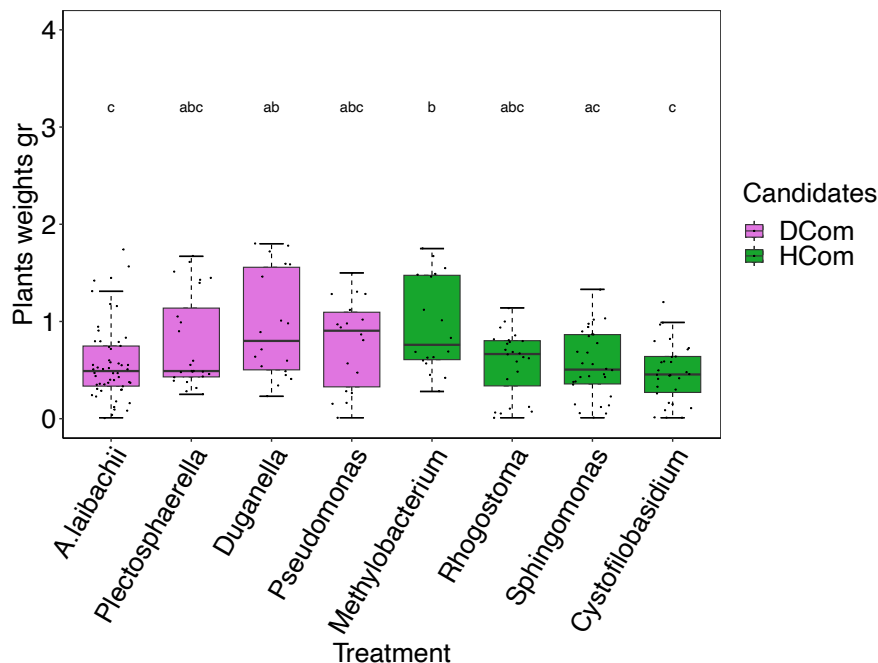


657



658 **Supplementary Figure 4. Changes in the co-abundance interactions of**  
659 **HCom and DCom OTUs.** Interactions between HCom and DCom OTUs in the  
660 network constructed for uninfected (A) and infected (B) samples. OTUs are clustered  
661 by module and colored by taxa. Edges are colored according to the correlations: pos-  
662 itive correlations are colored in orange, and negative correlations are colored in gray.  
663 Histograms (C and D) show the OTU distribution within modules for the networks of  
664 uninfected and infected samples, respectively. These histograms are further color-coded  
665 to distinguish microbial taxa: green represents bacteria, orange represents fungi, and  
666 blue represents nonfungal eukaryotes.

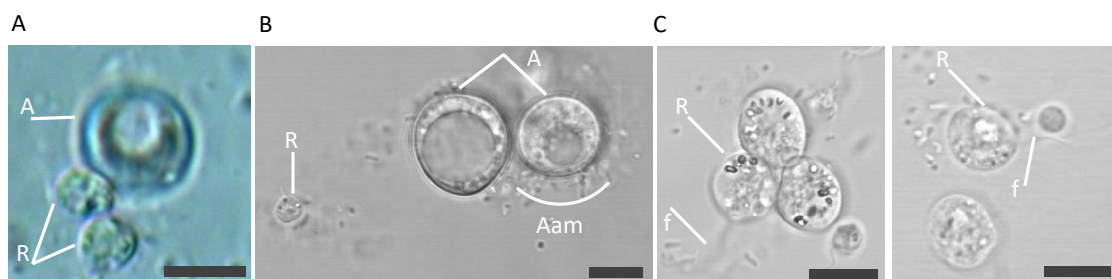
667



668

669 **Supplementary Figure 5. Fresh weight of plants inoculated with selected**  
670 **strains (see Figure. 6).** Box plots showing the weights of leaves infected with *Al-*  
671 *bugo* in the presence of HCom strains (green) and DCom strains (purple). Statistically  
672 significant differences between the two groups were evaluated using Tukey's HSD test,  
673 with different letters indicating significant differences ( $P < 0.05$ ).

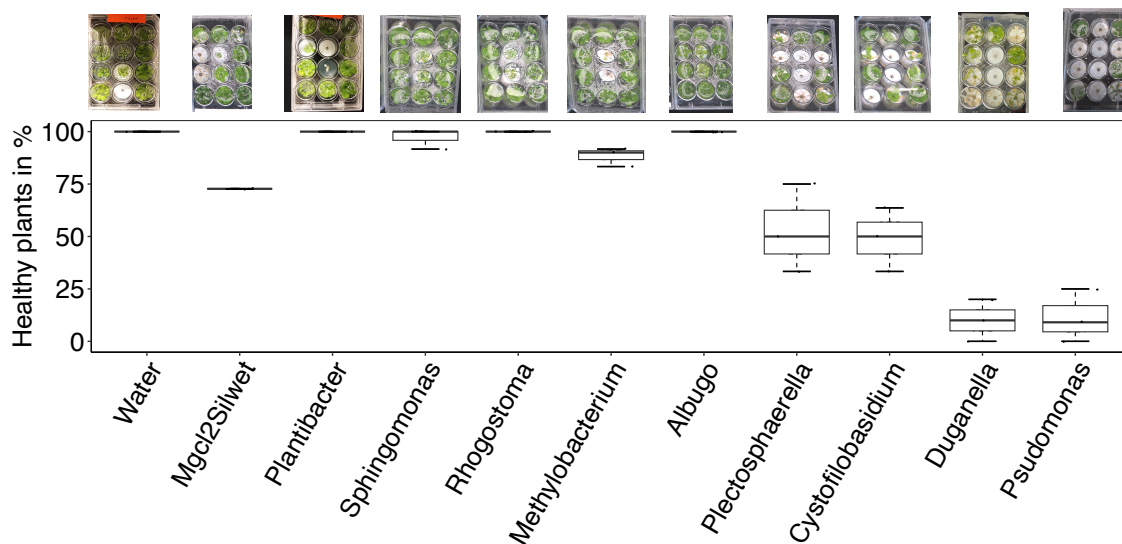
674



675

676 **Supplementary Figure 6. Possible interactions of *R. epiphylla* with *Al-***  
677 ***bugo*.** (A) Single cells of *Rhogostoma* (R) attached to the spore of *Albugo* (A). (B)  
678 *Albugo*'s spores associated with different microbes (Aam). (C) *Rhogostoma* cells feed-  
679 ing on other microbes or *Albugo*'s zoospores via filopodia (f) (See also S. Video1 and  
680 S. Video2). Measure bar indicates 10  $\mu$ m.

681



682

683 **Supplementary Figure 7. Effect of different HCom and Dcom microbes**  
684 **on WS-0 *Arabidopsis* plants under gnotobiotic conditions.** Each microbe was  
685 sprayed on 3-4 weeks-old *Arabidopsis* plants under gnotobiotic conditions. Box plots  
686 show the percentage of healthy plants three weeks post-inoculation.

687

## 688 Supplementary tables legends

689 **Table. S1.** Detailing number of samples and sampling locations.

690 **Table. S2.** Results of blasting the sequences of key OTUs used in Figure 4  
691 with NCBI databases.

## 692 Supplementary videos legends

693 **Supplementary Video S1 and S2.** *Rhogostoma* grazing with their filopodia  
694 in the solution containing *Albugo*'s spores.

## 695 References

696 [1] Julia A Vorholt. Microbial life in the phyllosphere. *Nature Reviews Microbiology*,  
697 10(12):828–840, 2012.

698 [2] Reza Sohrabi, Bradley C Paasch, Julian A Liber, and Sheng Yang He. Phyllosphere  
699 microbiome. *Annual review of plant biology*, 74:539–568, 2023.

700 [3] Brajesh K Singh, Manuel Delgado-Baquerizo, Eleonora Egidi, Emilio Guirado,  
701 Jan E Leach, Hongwei Liu, and Pankaj Trivedi. Climate change impacts on plant  
702 pathogens, food security and paths forward. *Nature Reviews Microbiology*, pages  
703 1–17, 2023.

704 [4] Susan Trumbore, Paulo Brando, and Henrik Hartmann. Forest health and global  
705 change. *Science*, 349(6250):814–818, 2015.

- 706 [5] Anna Bonaterra, Esther Badosa, Núria Daranas, Jesús Francés, Gemma Roselló,  
707 and Emilio Montesinos. Bacteria as biological control agents of plant diseases.  
708 *Microorganisms*, 10(9):1759, 2022.
- 709 [6] Sheridan L Woo, Rosa Hermosa, Matteo Lorito, and Enrique Monte. Tricho-  
710 derma: a multipurpose, plant-beneficial microorganism for eco-sustainable agri-  
711 culture, 2023.
- 712 [7] Zhenyan Zhang, Qi Zhang, Hengzheng Cui, Yan Li, Nuohan Xu, Tao Lu, Jian  
713 Chen, Josep Penuelas, Baolan Hu, and Haifeng Qian. Composition identification  
714 and functional verification of bacterial community in disease-suppressive soils by  
715 machine learning. *Environmental Microbiology*, 24(8):3405–3419, 2022.
- 716 [8] Michael J Sweet and Mark T Bulling. On the importance of the microbiome and  
717 pathobiome in coral health and disease. *Frontiers in Marine Science*, 4:9, 2017.
- 718 [9] Ricardo Hernández Medina, Svetlana Kutuzova, Knud Nor Nielsen, Joachim Jo-  
719 hansen, Lars Hestbjerg Hansen, Mads Nielsen, and Simon Rasmussen. Machine  
720 learning and deep learning applications in microbiome research. *ISME Communi-  
721 cations*, 2(1):98, 2022.
- 722 [10] Jun Yuan, Tao Wen, He Zhang, Mengli Zhao, C Ryan Penton, Linda S Thomashow,  
723 and Qirong Shen. Predicting disease occurrence with high accuracy based on soil  
724 macroecological patterns of fusarium wilt. *The ISME Journal*, 14(12):2936–2950,  
725 2020.
- 726 [11] Hao-Xun Chang, James S Haudenschild, Charles R Bowen, and Glen L Hartman.  
727 Metagenome-wide association study and machine learning prediction of bulk soil  
728 microbiome and crop productivity. *Frontiers in Microbiology*, 8:519, 2017.
- 729 [12] Barbara Emmenegger, Julien Massoni, Christine M Pestalozzi, Miriam Bortfeld-  
730 Miller, Benjamin A Maier, and Julia A Vorholt. Identifying microbiota community  
731 patterns important for plant protection using synthetic communities and machine  
732 learning. *Nature Communications*, 14(1):7983, 2023.
- 733 [13] Eric Kemen, Anastasia Gardiner, Torsten Schultz-Larsen, Ariane C Kemen,  
734 Alexi L Balmuth, Alexandre Robert-Seilaniantz, Kate Bailey, Eric Holub, David J  
735 Studholme, Dan MacLean, et al. Gene gain and loss during evolution of obligate  
736 parasitism in the white rust pathogen of arabidopsis thaliana. *PLoS biology*, 9(7):  
737 e1001094, 2011.
- 738 [14] Juliana Almario, Maryam Mahmoudi, Samuel Kroll, Mathew Agler, Aleksandra  
739 Placzek, Alfredo Mari, and Eric Kemen. The leaf microbiome of arabidopsis dis-  
740 plays reproducible dynamics and patterns throughout the growing season. *Mbio*,  
741 13(3):e02825–21, 2022.
- 742 [15] Matthew T Agler, Jonas Ruhe, Samuel Kroll, Constanze Morhenn, Sang-Tae Kim,  
743 Detlef Weigel, and Eric M Kemen. Microbial hub taxa link host and abiotic factors  
744 to plant microbiome variation. *PLoS biology*, 14(1):e1002352, 2016.

- 745 [16] Daniel Gómez-Pérez, Monja Schmid, Vasvi Chaudhry, Yiheng Hu, Ana Velic,  
746 Boris Maček, Jonas Ruhe, Ariane Kemen, and Eric Kemen. Proteins released  
747 into the plant apoplast by the obligate parasitic protist *albugo* selectively repress  
748 phyllosphere-associated bacteria. *New Phytologist*, 2022.
- 749 [17] Maryam Mahmoudi, Juliana Almario, Katrina Lutap, Kay Nieselt, and Eric  
750 Kemen. Microbial communities living inside plant leaves or on the leaf sur-  
751 face are differently shaped by environmental cues. *ISME Communications*, 4  
752 (1):ycae103, 08 2024. ISSN 2730-6151. doi: 10.1093/ismeco/ycae103. URL  
753 <https://doi.org/10.1093/ismeco/ycae103>.
- 754 [18] Sajjad Asaf, Muhammad Numan, Abdul Latif Khan, and Ahmed Al-Harrasi. Sph-  
755 ingomonas: from diversity and genomics to functional role in environmental re-  
756 mediatioin and plant growth. *Critical Reviews in Biotechnology*, 40(2):138–152,  
757 2020.
- 758 [19] Lisa Röttjers and Karoline Faust. From hairballs to hypotheses—biological insights  
759 from microbial networks. *FEMS microbiology reviews*, 42(6):761–780, 2018.
- 760 [20] Christine M Vogel, Daniel B Potthoff, Martin Schäfer, Niculò Barandun, and Ju-  
761 lia A Vorholt. Protective role of the arabidopsis leaf microbiota against a bacterial  
762 pathogen. *Nature microbiology*, 6(12):1537–1548, 2021.
- 763 [21] Yong-Guan Zhu, Chao Xiong, Zhong Wei, Qing-Lin Chen, Bin Ma, Shu-Yi-Dan  
764 Zhou, Jiaqi Tan, Li-Mei Zhang, Hui-Ling Cui, and Gui-Lan Duan. Impacts of  
765 global change on the phyllosphere microbiome. *New Phytologist*, 2022.
- 766 [22] Nick C Snelders, Hanna Rovenich, Gabriella C Petti, Mercedes Rocafort,  
767 Grady CM van den Berg, Julia A Vorholt, Jeroen R Mesters, Michael F Seidl,  
768 Reindert Nijland, and Bart PHJ Thomma. Microbiome manipulation by a soil-  
769 borne fungal plant pathogen using effector proteins. *Nature Plants*, 6(11):1365–  
770 1374, 2020.
- 771 [23] Katharina Eitzen, Priyamedha Sengupta, Samuel Kroll, Eric Kemen, and Gunther  
772 Doehlemann. A fungal member of the arabidopsis thaliana phyllosphere antago-  
773 nizes *albugo laibachii* via a *gh25* lysozyme. *Elife*, 10:e65306, 2021.
- 774 [24] Talia L Karasov, Manuela Neumann, Laura Leventhal, Efthymia Symeonidi, Gau-  
775 tam Shirsekar, Aubrey Hawks, Grey Monroe, Moisés Exposito-Alonso, Joy Bergel-  
776 son, et al. Continental-scale associations of arabidopsis thaliana phyllosphere mem-  
777 bers with host genotype and drought. *Nature Microbiology*, pages 1–11, 2024.
- 778 [25] Wu Xiong, Rong Li, Yi Ren, Chen Liu, Qingyun Zhao, Huasong Wu, Alexandre  
779 Jousset, and Qirong Shen. Distinct roles for soil fungal and bacterial communities  
780 associated with the suppression of vanilla fusarium wilt disease. *Soil Biology and*  
781 *Biochemistry*, 107:198–207, 2017.
- 782 [26] Khaoula Belhaj, Liliana M Cano, David C Prince, Ariane Kemen, Kentaro  
783 Yoshida, Yasin F Dagdas, Graham J Etherington, Henk-jan Schoonbeek, H Pe-  
784 ter van Esse, Jonathan DG Jones, et al. Arabidopsis late blight: infection of a

- 785 nonhost plant by *albugo laibachii* enables full colonization by phytophthora infes-  
786 tans. *Cellular Microbiology*, 19(1):e12628, 2017.
- 787 [27] Dan Knights, Laura Wegener Parfrey, Jesse Zaneveld, Catherine Lozupone, and  
788 Rob Knight. Human-associated microbial signatures: examining their predictive  
789 value. *Cell host & microbe*, 10(4):292–296, 2011.
- 790 [28] Jens M Olesen, Jordi Bascompte, Yoko L Dupont, and Pedro Jordano. The mod-  
791 ularity of pollination networks. *Proceedings of the National Academy of Sciences*,  
792 104(50):19891–19896, 2007.
- 793 [29] Talia L Karasov, Juliana Almario, Claudia Friedemann, Wei Ding, Michael Giolai,  
794 Darren Heavens, Sonja Kersten, Derek S Lundberg, Manuela Neumann, Julian  
795 Regalado, et al. Arabidopsis thaliana and pseudomonas pathogens exhibit stable  
796 associations over evolutionary timescales. *Cell host & microbe*, 24(1):168–179,  
797 2018.
- 798 [30] Or Shalev, Talia L Karasov, Derek S Lundberg, Haim Ashkenazy, Pratchaya  
799 Pramoj Na Ayutthaya, and Detlef Weigel. Commensal pseudomonas strains fa-  
800 cilitate protective response against pathogens in the host plant. *Nature ecology &*  
801 *evolution*, 6(4):383–396, 2022.
- 802 [31] Derek S Lundberg, Roger de Pedro Jové, Pratchaya Pramoj Na Ayutthaya,  
803 Talia L Karasov, Or Shalev, Karin Poersch, Wei Ding, Anita Bollmann-Giolai,  
804 Ilya Bezrukov, and Detlef Weigel. Contrasting patterns of microbial dominance  
805 in the arabidopsis thaliana phyllosphere. *Proceedings of the National Academy of*  
806 *Sciences*, 119(52):e2211881119, 2022.
- 807 [32] Gerd Innerebner, Claudia Knief, and Julia A Vorholt. Protection of arabidopsis  
808 thaliana against leaf-pathogenic pseudomonas syringae by sphingomonas strains  
809 in a controlled model system. *Applied and environmental microbiology*, 77(10):  
810 3202–3210, 2011.
- 811 [33] Claudia Knief, Lisa Frances, and Julia A Vorholt. Competitiveness of diverse  
812 methylobacterium strains in the phyllosphere of arabidopsis thaliana and identifi-  
813 cation of representative models, including *m. extorquens* pa1. *Microbial ecology*,  
814 60:440–452, 2010.
- 815 [34] Claudia Bartoli, Léa Frachon, Matthieu Barret, Mylène Rigal, Carine Huard-  
816 Chauveau, Baptiste Mayjonade, Catherine Zanchetta, Olivier Bouchez, Dominique  
817 Roby, Sébastien Carrère, et al. In situ relationships between microbiota and po-  
818 tential pathobiota in arabidopsis thaliana. *The ISME journal*, 12(8):2024–2038,  
819 2018.
- 820 [35] María Florencia Gorordo, María Ester Lucca, and Marcela Paula Sangorrín. Bio-  
821 control efficacy of the *vishniacozyma victoriae* in semi-commercial assays for the  
822 control of postharvest fungal diseases of organic pears. *Current Microbiology*, 79  
823 (9):259, 2022.



- 824 [36] Maria Florencia Perez, Luciana Contreras, Nydia Mercedes Garnica,  
825 María Verónica Fernández-Zenoff, María Eugenia Farías, Milena Sepulveda,  
826 Jacqueline Ramallo, and Julián Rafael Dib. Native killer yeasts as biocontrol  
827 agents of postharvest fungal diseases in lemons. *PloS one*, 11(10):e0165590, 2016.
- 828 [37] Jia Liu, Michael Wisniewski, Samir Droby, Silvana Vero, Shiping Tian, and Vera  
829 Hershkovitz. Glycine betaine improves oxidative stress tolerance and biocontrol  
830 efficacy of the antagonistic yeast *Cystofilobasidium infirmominiatum*. *International*  
831 *Journal of Food Microbiology*, 146(1):76–83, 2011.
- 832 [38] Kenneth Dumack, Sebastian Flues, Karoline Hermanns, and Michael Bonkowski.  
833 Rhogostomidae (cercozoa) from soils, roots and plant leaves (*Arabidopsis thaliana*):  
834 Description of *Rhogostoma epiphylla* sp. nov. and *R. cylindrica* sp. nov. *European*  
835 *journal of protistology*, 60:76–86, 2017.
- 836 [39] Michael Bonkowski. Protozoa and plant growth: the microbial loop in soil revisited.  
837 *New Phytologist*, 162(3):617–631, 2004.
- 838 [40] Kenneth Dumack, Christina Baumann, and Michael Bonkowski. A bowl with  
839 marbles: revision of the thecate amoeba genus *Lecythium* (Chlamydropyridae,  
840 Tectofilosida, Cercozoa, Rhizaria) including a description of four new species and  
841 an identification key. *Protist*, 167(5):440–459, 2016.
- 842 [41] Bao-Anh Thi Nguyen, Kenneth Dumack, Pankaj Trivedi, Zahra Islam, and Hang-  
843 Wei Hu. Plant associated protists—untapped promising candidates for agrifood  
844 tools. *Environmental Microbiology*, 25(2):229–240, 2023.
- 845 [42] Sai Guo, Stefan Geisen, Yani Mo, Xinyue Yan, Ruoling Huang, Hongjun Liu, Zhilei  
846 Gao, Chengyuan Tao, Xuhui Deng, Wu Xiong, et al. Predatory protists impact  
847 plant performance by promoting plant growth-promoting rhizobacterial consortia.  
848 *The ISME Journal*, page wrae180, 2024.
- 849 [43] Nico Eisenhauer, Jes Hines, Fernando T Maestre, and Matthias C Rillig. Recon-  
850 sidering functional redundancy in biodiversity research. *npj Biodiversity*, 2(1):9,  
851 2023.
- 852 [44] Paul J McMurdie and Susan Holmes. phyloseq: an R package for reproducible  
853 interactive analysis and graphics of microbiome census data. *PloS one*, 8(4):e61217,  
854 2013.
- 855 [45] Jari Oksanen, F Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R  
856 Minchin, RB O’Hara, Gavin L Simpson, Peter Solymos, M Henry H Stevens, Helene  
857 Wagner, et al. Package ‘vegan’. *Community ecology package, version*, 2(9):1–295,  
858 2013.
- 859 [46] Derek Ogle and Maintainer Derek Ogle. Package ‘fsa’. *Cran Repos*, 1:206, 2017.
- 860 [47] R Core Team, Maintainer R Core Team, MASS Suggests, and S Matrix. Package  
861 stats. *The R Stats Package*, 2018.



- 862 [48] R R Core Team et al. R: A language and environment for statistical computing.  
863 2013.
- 864 [49] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel,  
865 Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss,  
866 Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of*  
867 *machine Learning research*, 12:2825–2830, 2011.
- 868 [50] Jonathan Friedman and Eric J Alm. Inferring correlation networks from genomic  
869 survey data. *PLoS computational biology*, 8(9):e1002687, 2012.
- 870 [51] Stephen C Watts, Scott C Ritchie, Michael Inouye, and Kathryn E Holt. Fastspar:  
871 rapid and scalable correlation estimation for compositional data. *Bioinformatics*,  
872 35(6):1064–1066, 2019.
- 873 [52] Aric Hagberg, Pieter J Swart, and Daniel A Schult. Exploring network structure,  
874 dynamics, and function using networkx. Technical report, Los Alamos National  
875 Laboratory (LANL), Los Alamos, NM (United States), 2008.
- 876 [53] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefeb-  
877 vre. Fast unfolding of communities in large networks. *Journal of statistical me-*  
878 *chanics: theory and experiment*, 2008(10):P10008, 2008.
- 879 [54] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang,  
880 Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a  
881 software environment for integrated models of biomolecular interaction networks.  
882 *Genome research*, 13(11):2498–2504, 2003.