Experimental frameworks for deciphering bacterial adaptation in the plant root microbiota

Doctoral thesis

for

the award of the doctoral degree

of the Faculty of Mathematics and Natural Sciences

of the University of Cologne

submitted by

Niklas Kiel

2026

Summary

The plant root microbiome consists of phylogenetically diverse bacteria that play a critical role in host health and development. Recent research has demonstrated that these microbial communities exhibit a strong degree of host preference, preferentially colonizing plant species with which they share a co-evolutionary history. This indicates that bacterial fitness in the rhizosphere is dependent on the host species. To date, research on bacterial adaptation within the rhizosphere has relied heavily on metagenomic surveys investigating the correlation between the abundance of specific functional genes and their corresponding strains. While experimental evolution approaches have been employed to shed light on rhizosphere adaptation in a time-resolved manner, most studies utilized single-strain approaches, thereby omitting the complexities of inter-species dynamics in communities. Consequently, the specific genetic determinants of bacterial host preference, and the evolutionary trajectories of these genes under imposed selective pressure, remain elusive. This thesis aims to bridge that knowledge gap by tracing bacterial adaptation to the roots of a novel host plant within a community context.

To address this question, we conducted a long-term evolution experiment utilizing bacterial communities composed of diverse phylogenetic clades. These synthetic communities (SynComs), consisting of 17 different strains, were successively propagated on the roots of their original and a novel host over a period of nearly two years. Analysis of these evolving communities through metagenomic shotgun sequencing revealed a remarkable level of parallel evolution. Cross-inoculation experiments demonstrated a significant increase in bacterial fitness over the course of the experiment. At the final time point, the fitness of the populations which evolved on the novel host plant was indistinguishable from the colonization fitness on the roots of the original host. Notably, bacteria evolving on the new host plant displayed a higher number of fixed mutations, suggesting a higher adaptive potential and stronger selective pressure under these conditions. During the evolution experiment, we established a comprehensive library of approximately 23,000 individual evolved isolates collected from different time points. Because the scale of this culture collection rendered sequencing-based strain verification impractical, we

developed a high-throughput method for automated strain identification. To this end, we present an autofluorescence-based framework for the taxonomic classification of known bacterial isolates. By applying this method to our collection of 23,000 evolved isolates, we were able to assign taxonomic labels to every individual isolate. This automatic indexing enabled the design of SynComs composed of evolved bacteria, allowing for a direct comparison of their root colonization capacity against their ancestral counterparts. Through these competition assays, we identified significant increases in fitness across several key taxa, demonstrating successful adaptation to the new host. Whole-genome sequencing of these isolates identified mutations in metabolic genes, regulators of biofilm formation and motility, and genes affecting cell-surface proteins. These findings suggest that bacterial adaptation in the rhizosphere is a substrate-driven process where the ability to adhere to the root surface is a primary selective advantage along with modulation of immunogenic cell-surface proteins.

Our results demonstrate that bacterial communities can be conditioned to achieve improved fitness on a novel host plant within reasonably relevant timescales. Consequently, the experimental framework presented in this thesis provides a robust foundation for the enhancement of bioinoculants. By utilizing this experimental evolution approach, researchers can optimize the rhizosphere competence of agriculturally relevant strains across diverse crop species, soil types, and fertilization regimes.

Zusammenfassung

Das Mikrobiom der Pflanzenwurzel besteht aus phylogenetisch diversen Bakterien, die eine entscheidende Rolle für die Gesundheit und Entwicklung der Wirtspflanze spielen. Jüngste Forschungsarbeiten haben gezeigt, dass diese bakteriellen Gemeinschaften eine ausgeprägte Wirtspräferenz aufweisen. Das heißt, dass sie bevorzugt die Wurzeln von Pflanzenarten besiedeln, mit denen sie eine gemeinsame Evolutionsgeschichte teilen. Dies deutet darauf hin, dass die bakterielle Fitness in der Rhizosphäre maßgeblich von der jeweiligen Wirtspflanze abhängt. Bisher stützte sich die Forschung zur bakteriellen Anpassung in der Rhizosphäre vorwiegend auf deskriptive Untersuchungen, welche die Korrelation zwischen der Abundanz spezifischer Genfunktionen und den entsprechenden taxonomischen Stämmen analysierten. Obwohl Ansätze der experimentellen Evolution eingesetzt wurden, um die Anpassung in der Rhizosphäre zeitaufgelöst zu untersuchen, beschränkten sich die meisten Studien auf einzelne Stämme. Dabei blieben die komplexen Dynamiken innerhalb einer Gemeinschaft weitgehend unberücksichtigt. Infolgedessen sind die spezifischen genetischen Determinanten der Wirtspräferenz sowie die evolutionären Verläufe dieser Gene unter selektivem Druck bisher nur unzureichend charakterisiert. Das Ziel dieser Arbeit war es, diese Wissenslücke zu schließen, indem die bakterielle Anpassung innerhalb eines komplexen Gemeinschaftskontexts direkt an den Wurzeln einer neuen Wirtspflanze verfolgt wurde.

Um dies zu untersuchen, führte ich eine langfristige Studie zur experimentellen Evolution mit bakteriellen Gemeinschaften durch, die sich aus verschiedenen phylogenetischen Kladen zusammensetzten. Diese synthetischen Gemeinschaften (SynComs) wurden über einen Zeitraum von fast zwei Jahren sukzessiv in der Rhizosphäre ihrer ursprünglichen sowie einer neuen Wirtspflanze propagiert. Die Analyse dieser evolvierenden Gemeinschaften mittels metagenomischer Sequenzierung offenbarte ein bemerkenswertes Maß an paralleler Evolution. Kreuzinokulationsexperimente zeigten eine signifikante Steigerung der bakteriellen Fitness im Verlauf des Experiments. Speziell zu den letzten Zeitpunkten war die Fitness der evolvierten Populationen auf dem neuen Wirt nicht mehr von der Fitness

auf dem ursprünglichen Wirt zu unterscheiden. Bemerkenswerterweise wiesen Bakterien, die auf der neuen Wirtspflanze evolvierten, eine höhere Anzahl fixierter Mutationen auf, was auf ein höheres adaptives Potenzial und einen stärkeren Selektionsdruck unter diesen neuartigen Bedingungen hindeutet. Um diese Anpassungen detaillierter untersuchen zu können, entwickelte ich eine hochskalierbare Methode zur taxonomischen Klassifizierung auf Basis von Autofluoreszenz. Dies ermöglichte das Screening einer umfassenden Bibliothek von etwa 23.000 evolvierten Bakterienisolaten. Unter Verwendung dieser Bibliothek stellte ich SynComs zusammen, um die Kompetenz für Wurzelbesiedlung gegenüber ihren jeweiligen Vorfahren zu bewerten. Durch diese Experimente konnte ich bei mehreren Schlüsseltaxa signifikante Fitnesssteigerungen feststellen, was eine erfolgreiche Anpassung an die neue Wirtsumgebung belegt. Die Genomsequenzierung dieser Isolate offenbarte Mutationen in Stoffwechselgenen, in Regulatoren der Biofilmbildung und Motilität sowie in immunogenen Oberflächenproteinen. Diese Ergebnisse legen nahe, dass die bakterielle Anpassung in der Rhizosphäre ein substratgesteuerter Prozess ist, bei dem die Fähigkeit zur Adhäsion an Wurzeloberflächen einen primären Selektionsvorteil darstellt. Letztlich bietet der in dieser Studie entwickelte Arbeitsablauf sowie die identifizierten Genkandidaten eine fundierte Basis für die Optimierung von Bioinkulanzien, um eine stabile Besiedlung durch nützliche Bakterien unter komplexen landwirtschaftlichen Bedingungen zu ermöglichen.

Table of Contents

List of abbreviations

Abbreviation	Definition
AAI	Average amino-acid identity
AF	Allele frequency
ANI	Average nucleotide identity
ANN	Artificial neural network
Arcsinh	Inverse hyperbolic sine
ASV	Amplicon sequencing variant
At	*Arabidopsis thaliana*
AUC	Area under the curve
CFU	Colony forming unit
CNN	Convolutional neural network
Cr	Chlamydomonas reinhardtii
CSV	Comma separated values
CV	Cross-validation
DP	Sequencing depth
EPS	Exopolysaccharides
ETI	Effector-triggered immunity
FDR	False discovery rate
flg22	Flagellin
FTIR	Fourier-transform infrared
GLM	Generalized linear model
HGT	Horizontal gene transfer
INDEL	Short insertions and deletions
ISR	Induced systemic resistance
Lj	*Lotus japonicus*
LOO	Leave-one-out
LPS	Lipopolysaccharides
LTEE	Long-term evolution experiment
MCC	Matthews Correlation Coefficient
ML	Machine learning
MS	Murashige and Skoog
MSE	Mean squared error
NADH	Nicotinamide adenine dinucleotide
OD	Optical density
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PRR	Pattern recognition receptors
PTI	Pattern triggered immunity
R	Coefficient of correlation
R^2	Coefficient of determination
RF	Random forest
RMSE	Root mean square error
SB	Strand bias
SNV	single-nucleotide variants
SVM	Support vector machines
T3SS	Type III Secretion System
TSB	Tryptic soy broth
tSNE	t-distributed stochastic neighbor embedding
TSV	Tab separated values
v/v	Volume by volume

Chapter 1

Introduction

"The role of the infinitely small in nature is infinitely large."

Louis Pasteur

1. Introduction

1.1. The plant root microbiome

In nature, plants do not exist axenically but rather participate in complex networks of multi-kingdom interactions (Agler et al., 2016). The microorganisms that colonize all plant tissues, from the leaves to the roots, are collectively known as the plant microbiota (Bulgarelli et al., 2013; Vorholt, 2012). The root microbiota is composed of bacteria, fungi, oomycetes, and archaea that occupy distinct ecological niches, including the root surface (rhizoplane), the surrounding soil (rhizosphere), and the internal tissues (endosphere) (Trivedi et al., 2020). Within these niches, microbial interactions with the host exist on a continuum ranging from mutualistic to pathogenic (Hirsch, 2004; Newton et al., 2010). Mutualists provide a measurable fitness benefit to the host, while commensals benefit from the host without inducing negative effects (Mendes et al., 2013). Conversely, pathogens induce a net fitness cost (Mendes et al., 2013). Crucially, the categorization of a specific microbe is not static but context-dependent, with outcomes shifting between beneficial and deleterious based on environmental factors such as abiotic stress, nutrient availability, and host developmental stage (Chaparro et al., 2014; Getzke et al., 2024; Johnson et al., 1997; Redman et al., 2002). The importance of a healthy microbiota cannot be overstated, given the substantial metabolic cost plants invest to recruit it (Tian et al., 2020). A healthy microbiota provides the plant host with numerous beneficial functions, including nitrogen fixation, phosphate solubilization, modulation of plant hormone levels, the suppression of soil-borne pathogens, and the induction of systemic resistance (ISR) (Hibbing et al., 2010; Oldroyd, 2013; Pieterse et al., 2014; Richardson & Simpson, 2011; Spaepen & Vanderleyden, 2011). Over time, this field of research has developed the concept of the plant holobiont, a theory considering the host plant and its associated microbiota as a single ecological and evolutionary unit (Trivedi et al., 2020; Vandenkoornhuyse et al., 2015).

The assembly of the root microbiota is not a random sampling of the soil but rather an active process driven by the host plant, which is mediated through the secretion of carbon-rich compounds into the soil, also known as rhizodeposition

(Kuzyakov & Domanski, 2000; Tian et al., 2020). Plants invest a significant portion of their produced carbon, estimated at up to 40% of all photoassimilates, into the secretion of these root exudates (Lynch & Whipps, 1990). Root exudates create a radial concentration gradient of high-energy metabolites, which is highest at the root and decreases towards the bulk soil (Kuzyakov & Razavi, 2019). Motile microorganisms detect these root exudates and migrate toward the root via chemotaxis (Sasse et al., 2018; Scharf et al., 2016). Consequently, absolute microbial abundance, i.e., the number of colony-forming units (CFUs), is positively correlated with root exudate concentration (Hacquard et al., 2015; Smalla et al., 2001). Conversely, microbial diversity, for instance measured as Shannon entropy, is negatively correlated with proximity to the root, as the nutrient-rich environment imposes a strong selective pressure and numerous species compete for limited resources (Bulgarelli et al., 2013). While the microbial load is highest at the root surface, the microbial diversity is significantly lower than that of the surrounding soil (Bulgarelli et al., 2012; Lundberg et al., 2012). This indicates that the root microbiota represents a subset of the surrounding soil community that has undergone a stringent filtering process mediated by plant exudates and the host immune system. With the rise of high-throughput amplicon sequencing such as 16S rRNA gene profiling, the taxonomic composition of these communities has been resolved. While soil communities are often rich in Acidobacteria and Actinobacteria (Fierer, 2017; Janssen, 2006; Trivedi et al., 2020), the root microbiome is consistently dominated by members of the phyla Proteobacteria, Firmicutes, and Bacteroidetes (Bulgarelli et al., 2012; Lundberg et al., 2012; Trivedi et al., 2020). These patterns appear robust across diverse plant species and geographical locations. For instance, Thiergart et al. characterized *Arabidopsis thaliana* populations across Europe and found that, despite significant variation in local soil communities, a small subset of bacterial taxa was consistently recruited to the root, illustrating the strength of host-mediated microbiota recruitment (Thiergart et al., 2020). Furthermore, this finding indicates that the root microbiota assembly is a deterministic process. However, it is important to emphasize that biotic interactions among microbes also impact community outcomes. Upon infection, plants often attempt to mitigate disease development through the selective recruitment of beneficial microbes (Berendsen et al., 2018; Y. Liu et al., 2024; Yuan et

al., 2018). The fungal pathogen *Verticillium dahliae* has been shown to counteract this plant response by secreting antimicrobial effector proteins that actively shape its bacterial environment (Kraege et al., 2026; Snelders et al., 2020, 2022).

While descriptive studies have been instrumental in defining the core microbiome and identifying correlations between microbial taxa and plant health, they face significant limitations (Mendes et al., 2011). Many studies provide only a snapshot of the community structure and lack the mechanistic insights seen in the *Verticillium* story, making it difficult to establish causality. Distinguishing between microbes that actively drive plant phenotypes and those that merely hitchhike on the root surface remains a major challenge (Vorholt et al., 2017). Consequently, to move from descriptive to mechanistic understanding of bacterial adaptation and host-microbe co-evolution, experimental frameworks that allow for the manipulation and tracking of these interactions are required.

1.2. Reductionist approaches to study plant-microbiota interactions

To overcome the limitations of descriptive studies, the biological complexity of natural communities must be reduced to an explainable level. This allows researchers to disentangle the dense web of multidirectional interactions and draw causative conclusions (Vorholt et al., 2017). Reductionist approaches serve as a powerful principle for experimental design, similar to the binary interaction studies that have historically defined microbial ecology and host-microbe interactions (Jessup et al., 2004). By utilizing gnotobiotic systems, researchers move from the complexity of nature to controlled environments. This allows the tweaking of individual experimental variables, and thus, shifts the focus from descriptive findings to mechanistic understanding (Chesneau et al., 2025; Marín et al., 2021; Rodriguez et al., 2019).

A cornerstone of this approach is the establishment of comprehensive culture collections derived from a host of interest. Historically, researchers have struggled with laboratory cultivation of the majority of bacterial taxa, both in terms of aggregated abundance and diversity (Amann et al., 1995). This phenomenon was known as the “great plate count anomaly” (Staley & Konopka, 1985). Recent developments of protocols for the establishment of culture collections have allowed researchers to

reisolate and cultivate the majority of host-associated members of the bacterial fraction of the microbiota (J. Zhang et al., 2021). Extensive bacterial libraries have been established from the roots of model organisms like *Arabidopsis thaliana* and *Lotus japonicus*, as well as major crops including rice, maize, sugar cane, and citrus (Armanhi et al., 2018; Bai et al., 2015; Niu et al., 2017; Wippel et al., 2021; J. Zhang et al., 2019; Y. Zhang et al., 2017). These collections, typically complemented with whole-genome sequences, enable the reconstruction of synthetic communities (SynComs) (Pfeilmeier et al., 2021, 2024; Vorholt et al., 2017; Wippel et al., 2021). SynComs, which are defined consortia of microbes, allow researchers to mimic the functional structure of natural microbiomes, but with adjustable complexity. Unlike single-strain inoculations, SynComs preserve the critical element of microbe-microbe interactions, allowing for the testing of ecological hypotheses in a controlled setting.

The power of SynComs lies in their ability to reveal phenotypes that only arise from community interactions, thus enabling synthetic ecology (Chesneau et al., 2025). For example, Durán et al. extended the culture collection concept beyond bacteria to include eukaryotic microorganisms, such as fungi and oomycetes, isolated from *Arabidopsis* roots (Durán et al., 2018). By designing multi-kingdom SynComs, they demonstrated that the plant survival rate was not dependent on a single taxon, but on the balance between prokaryotic and eukaryotic members. Specifically, bacteria were required to suppress the lethal effects of fungal members of the SynCom, a finding that would have been impossible in a bacterium-only or fungus-only experiment. Beyond survival, reductionist systems allow for the dissection of specific nutritional and developmental benefits. For instance, using culture collections of bacteria that colonize *Arabidopsis* roots in iron-limiting soil, Harbort et al. elucidated the mechanism of bacteria-mediated plant rescue. They were able to demonstrate that beneficial bacteria can mobilize iron for the plant host under limiting conditions via the secretion of coumarins (Harbort et al., 2020). Similarly, SynCom approaches using rice as a model crop have linked specific bacterial community members to tiller number, a key agronomic trait directly correlated with yield (J. Zhang et al., 2025).

Lastly, comparative SynCom approaches have been used to elucidate the host-driven assembly of the microbiota. Wippel et al. observed that *Arabidopsis* and *Lotus*

plants, grown side-by-side in natural soil, assembled distinct root communities despite the same starting inoculum (Wippel et al., 2021). To investigate the drivers of this differentiation, they constructed mixed SynComs comprised of taxonomically paired strains isolated from both host plants. When these mixed communities were applied to sterile plants in a gnotobiotic peat-based system, the bacteria displayed a clear host preference. This means that strains originally isolated from *Arabidopsis* reached higher aggregate abundances on *Arabidopsis* roots compared to *Lotus* roots. Similarly, bacteria originally recovered from *Lotus* roots showed a clear preference for *Lotus* roots. Importantly, this fitness advantage on the roots of the native host was a community trait, as it was only present when strains were in competition within a mixed community, rather than in mono-associations. However, the degree of host preference varied across taxa, where families like Phyllobacteriaceae showed strong host preference, while for instance Intrasporangiaceae showed weak or no host preference.

These findings highlight the importance of a shared co-evolutionary history, where the host plant selects for specific bacterial strains and possibly microbial functions (Lemanceau et al., 2017). Yet, despite the availability of genomic data for these SynComs, the specific genetic determinants driving this host preference, and the adaptive trajectories these bacteria took to acquire them, remain elusive.

1.3. Evolution based approaches to study microbial adaptation

Experimental evolution provides a powerful approach to studying the dynamics of adaptation by propagating populations under controlled conditions for hundreds or thousands of generations (Barrick & Lenski, 2013). This approach allows researchers to investigate the reproducibility of evolutionary outcomes and the specific genetic mechanisms underlying adaptation. The most iconic example is the *Escherichia coli* long-term evolution experiment (LTEE) initiated by Richard Lenski (Lenski et al., 1991). As of now, this experiment has been running for more than 30 years, corresponding to more than 60,000 bacterial generations (Good et al., 2017). Through this, researchers have learned that gains in bacterial fitness over time follow a power law (Wiser et al., 2013). That means that the rate of fitness gains decreases over time, while fitness continues to increase, suggesting that adaptive potential might be unbounded (Lenski et al., 2015). Furthermore, the LTEE has led to the observation

that beneficial mutations in bacteria fix sequentially or compete through clonal interference (Gerrish & Lenski, 1998; Good et al., 2017; Lang et al., 2013). These findings highlight the role of historical contingency, where rare innovations, such as aerobic citrate utilization, depend on a specific sequence of prior mutations (Blount et al., 2008; Good et al., 2017). However, these insights are largely derived from single-species populations evolving in simple, constant environments, which are conditions in strong contrast to the microbial life outside the lab.

In nature, biotic interactions create a dynamic fitness landscape that is fundamentally different from that of a monoculture. Theoretical and empirical work suggests that species interactions can alter both the speed and direction of evolution (Barraclough, 2015). For instance, the presence of competitors can inhibit adaptation by restricting access to ecological niches through niche occupation, effectively limiting the adaptive potential available for a species to diversify (Brockhurst et al., 2007). A major challenge in studying evolution within these communities is the presence of conflicting selective pressures. Organisms must balance adaptation to the abiotic environment, such as temperature or resource availability, with adaptation to biotic factors, such as competitors and predators. Studies have shown that evolution in diverse communities is often constrained in comparison to monocultures (Lawrence et al., 2012; Scheuerl et al., 2020). For example, adaptations that improve competitiveness against other species may come with a cost to growth rate or abiotic stress tolerance. Furthermore, the complexity of these interactions makes it difficult to identify the causal genetic changes driving adaptation, as the community composition might be constantly changing (Widder et al., 2016).

These community dynamics are particularly relevant in the context of the plant microbiome, as the plant host acts as a strong selective filter, shaping the diversity and composition of its associated microbiota (Bulgarelli et al., 2013). Microbes colonizing the root must adapt not only to the immune system of the plant and root exudate profile, but also to the intense competition from other soil-derived taxa. Experimental evolution studies in plant contexts have begun to disentangle these complex interactions. For example, research on legume symbiotic rhizobacteria has shown that the selection regime drastically alters evolutionary trajectories, leading to

either extra- or intracellular symbiotic adaptations (Marchetti et al., 2017). This effect was independent of the bacterial lifestyle and applied to free-living as well as host-associated bacteria. Beyond lifestyle switches, evolution experiments have pinpointed specific genetic targets of adaptation. For instance, Li et al. found that *Pseudomonas protegens* adapting to the *Arabidopsis* rhizosphere rapidly accumulated mutations in global regulators, resulting in a mutualistic relationship with the host (E. Li, de Jonge, et al., 2021). Furthermore, Batstone et al. demonstrated that these adaptive trajectories can shape the outcomes of evolution (Batstone et al., 2020). When the symbiont *Ensifer meliloti* was experimentally evolved on different *Medicago* genotypes, the bacteria became more cooperative with their native host, highlighting the strength of genotype-by-genotype interactions in shaping evolutionary outcomes (Batstone et al., 2020). However, most evolution experiments investigating plant-microbiota interactions rely on single-strain inocula, thus failing to capture the dynamics of community-driven adaptation in the rhizosphere.

Introducing a community context for host-adapted microbes raises fundamental questions regarding evolutionary outcomes. For instance, is the evolution of bacterial communities constrained in the presence of a host? It remains unclear whether the intense competition within a diverse microbiota narrows the available adaptive space or whether cross-feeding facilitates metabolic pathways that are inaccessible to monocultures. Furthermore, a community context prompts the question of whether the necessity and historical contingency of mutations are dependent on the surrounding community. A mutation might only provide a fitness advantage in the presence of a specific neighbour, making evolutionary outcomes highly dependent on the community composition. It remains to be determined whether epistatic interactions are restricted to a single genome. Alternatively, the fitness effect of a mutation in one bacterial genome may be dependent on the genetic background of the surrounding community members. Consequently, there is a need for experimental frameworks that investigate how communities shape the adaptive trajectories of the plant microbiota.

1.4. Thesis aims and outline

Current research has established which bacteria colonize plant roots, yet the mechanisms governing how these microbes adapt to the host environment over time remain largely elusive. Descriptive metagenomics offers insights but lacks mechanistic resolution, while traditional experimental evolution in monocultures or binary interactions lacks ecological complexity to resemble nature. The primary aim of this thesis is to bridge this knowledge gap by investigating host-driven bacterial adaptation within a community context. Rather than observing a static endpoint, this work seeks to disentangle the temporal dynamics of evolutionary changes that occur during the initial stages of host adaptation. To achieve this, we employed a reductionist yet ecologically relevant approach by utilizing defined SynCom inocula for an evolution experiment. This intermediate complexity allows researchers to account for ecological interactions, such as competitive exclusion and potential niche filling, while being simple enough to allow for precise identification of causal genetic mutations. However, a major bottleneck in validating evolutionary changes is the need to identify common and rare adaptive phenotypes. Therefore, evolved isolates need to be recovered and compared to their ancestral state. Standard molecular identification through amplicon sequencing is too laborious when screening thousands of isolates. Therefore, a secondary aim of this work is to establish a high-throughput framework for classification of re-isolated bacteria in liquid culture. By developing a novel spectral classification framework, we aimed to rapidly screen reisolated and evolved mutants from the frozen fossil record of the evolution experiment, thereby contributing to the understanding of bacterial root adaptation. The experimental concepts presented here can be utilized in other complementary systems such as the gut microbiota. Ultimately, this work provides a valuable framework for optimization of plant-growth-promoting rhizobacteria.

Chapter 2

Experimental evolution of root-associated microbial communities

"Now, here, you see, it takes all the running you can do, to keep in the same place."

Lewis Carroll (Through the Looking-Glass)

2. Experimental evolution of root-associated microbial communities

Authors: Niklas Kiel, Kathrin Wippel, Ruben Garrido-Oter

Own Contribution: Conceptualization, planning and conducting of the experiments; analysis of the data; writing of the manuscript.

Disclaimer: The material presented here is the same as a manuscript that has been submitted for peer review.

Therefore, this chapter represents original, unpublished research work by the author, and is included in this thesis as a component to the dissertation requirements. This inclusion does not constitute prior publication, and all rights remain with the author and the co-authors until a final decision is reached by the journal.

The co-authors of the submitted manuscript have consented to the inclusion of this work in this thesis.

2.1. Abstract

Microbial communities associated with plants are dynamic ecosystems whose structure and function influence host health and fitness. While the ecological roles of these communities are increasingly well understood, the evolutionary dynamics that shape them over time remain unclear. Experimental evolution constitutes a powerful tool to study microbial adaptation, but so far, its application has been largely focused on free-living or single-species populations. Here, we use experimental evolution to test whether synthetic bacterial communities can adapt to novel plant hosts. We show that replicate microbial communities undergo rapid and parallel genomic evolution, with measurable gains in colonization fitness. Adaptation involves mutations in regulatory processes, motility, and core metabolism, leading to significant fitness gains. Our findings demonstrate that the plant host can drive microbiome evolution without disrupting community stability.

2.2. Introduction

Plants live in close association with communities of diverse microbes that can impact their health, growth, and fitness (Bulgarelli et al., 2013; Fitzpatrick et al., 2020; Müller et al., 2016; Trivedi et al., 2020). These communities can provide beneficial functions such as nutrient mobilization (Harbort et al., 2020), protection against pathogens (Carrión et al., 2019; Durán et al., 2018; Raaijmakers & Mazzola, 2016; Spooren et al., 2024), and tolerance to environmental stresses (S. Liu et al., 2025). Root-associated microbiomes, in particular, have emerged as crucial determinants of plant performance and are increasingly viewed as integral components of plant adaptive strategies to environmental factors, such as nutrient limitation (Singh et al., 2025) or drought (De Vries et al., 2020).

Host factors exert selective pressure on members of these communities, potentially driving evolutionary processes that shape community composition, functionality, and host specificity over ecological and evolutionary timescales (Wilde et al., 2024). While the ecology and functions of plant microbiomes are increasingly well understood, their capacity to evolve and adapt to different host species remains largely unexplored. Recent work has shown that individual members of the plant

microbiome can adapt rapidly to novel host environments, leading to changes in colonization efficiency and functional outcomes (Batstone et al., 2020; E. Li, De Jonge, et al., 2021; E. Li, Zhang, et al., 2021). However, the mechanisms underlying these evolutionary processes in a community context are still largely unknown.

Experimental evolution offers a powerful framework to study microbial adaptation, enabling controlled tracking of evolutionary trajectories at high resolution over short timescales (Barrick & Lenski, 2013; Lenski, 2017). Long-term evolution experiments (LTEEs), in particular, have provided insights into the dynamics of microbial adaptation, showing that evolutionary responses can be highly reproducible across replicate populations exposed to the same selective pressures (Barrick et al., 2009; Good et al., 2017; Wiser et al., 2013). In bacteria, LTEEs have revealed convergent patterns of mutation and sustained fitness gains over tens of thousands of generations (Couce et al., 2024; Good et al., 2017; Tenaillon et al., 2016). Similar findings have been observed in LTEEs with eukaryotes such as *Saccharomyces cerevisiae*, where evolutionary outcomes appear to be the result of both chance and selection (Lang et al., 2013). These results point to a degree of determinism in evolution, shaped by selection, ecological context, and functional trade-offs.

However, in natural systems, evolutionary change rarely occurs in isolation. Microbes evolve in the context of complex communities, where interspecific interactions shape both ecological dynamics and adaptive trajectories. Several studies have begun to address this complexity by experimentally evolving simplified communities comprising two or more species, revealing that such interactions can constrain or redirect evolution, lead to changes in mutualism or competition, and alter functional outputs (Castledine et al., 2020; Evans et al., 2020; Fiegna et al., 2015; Goyal et al., 2022; Hall et al., 2018; Piccardi et al., 2024; Summers et al., 2010; Turkarslan et al., 2021). Yet most of these studies have focused on free-living communities with limited taxonomic complexity, and few have examined evolutionary dynamics in host-associated microbiomes.

Here, we build on this work by applying experimental evolution to synthetic bacterial communities associated with a plant host. Previously, we demonstrated that synthetic bacterial communities (SynComs), derived from root microbiomes of the

Brassica *Arabidopsis thaliana* (*At*) and the legume *Lotus japonicus* (*Lj*), exhibit clear host-preference phenotypes when inoculated onto their native versus non-native hosts (Wippel et al., 2021). We hypothesized that exposure to a novel plant species would impose selective pressures on these communities and their microbial members, driving adaptive genomic changes and leading to improved colonization fitness on the new host. To test this, we evolved root-derived SynComs across 16 plant generations under native and non-native host conditions and tracked genomic and phenotypic signatures of adaptation. We show that exposure to novel hosts rapidly induces reproducible genomic changes and measurable gains in colonization fitness. These findings demonstrate that plant hosts can drive microbiome evolution, with implications for the targeted improvement of microbial inoculants in agriculture.

2.3. Results

2.3.1. Changes in community composition during experimental evolution

To test whether plant hosts drive evolutionary change in their associated microbiota, we conducted an evolution experiment using two bacterial SynComs: *At*-SC, derived from *Arabidopsis thaliana* (Bai et al., 2015), and *Lj*-SC, derived from *Lotus japonicus* (Wippel et al., 2021). We repeatedly passaged them through roots of either their native or non-native plant host using a soil-based gnotobiotic system (Kremer et al., 2021) (**Figure 1A**). This yielded four experimental conditions representing all combinations of SynCom and host species (**Figure 1B)**. Each cycle comprised 5 weeks of plant growth, after which root bacterial communities were extracted and used to inoculate the axenic roots of the subsequent plant generation. This process was repeated for 16 plant cycles (**Figure 1A**), which corresponds to an estimated ~2,500 bacterial generations, on average across taxa (**Table 1**). At each harvesting step, we measured plant growth, extracted DNA samples for sequencing, and performed limiting dilutions to quantify absolute microbial abundances and isolate evolved bacterial strains (**Methods**).

Throughout the experiment, plant growth remained relatively constant, indicating stability in host-microbiota associations over time (**Fig. S 1A-B**). Neither pathogenic nor mutualistic interactions emerged from these repeated host passages

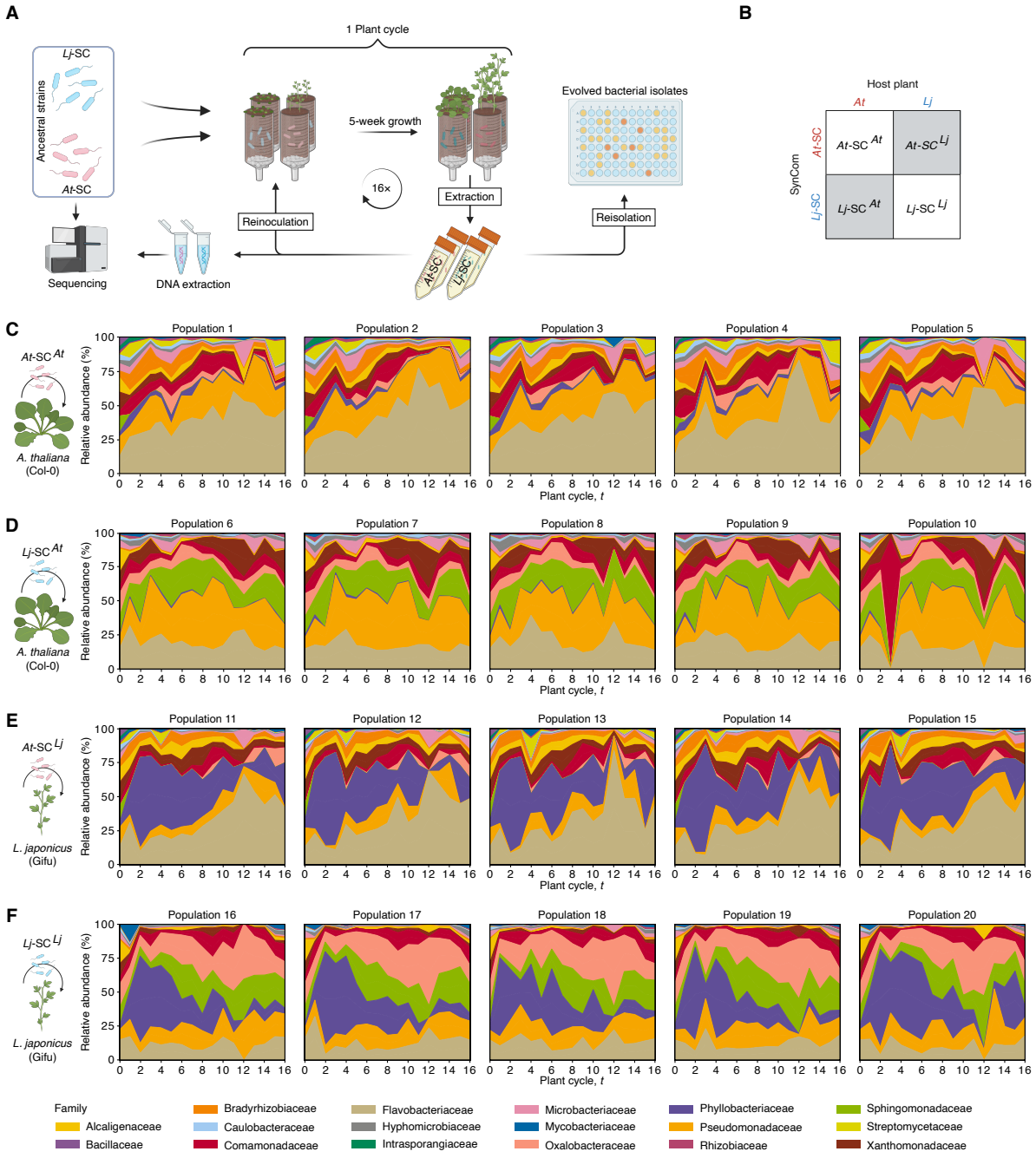

Figure 1. Experimental evolution of root-associated microbial communities. (**A**) Schematic overview of the conducted evolution experiment. (**B**) Summary of experimental conditions (2 plant species x 2 SynComs). Non-native conditions are indicated by dark grey. (**C**-**F**) Variation in community structure over time for SynComs *At*-SC^{At} (**C**), *Lj*-SC^{At} (**D**), *At*-SC^{Lj} (**E**), and *Lj*-SC^{Lj} (**F**). Bacterial abundances are shown as relative abundances (*y*-axis) plotted against time (*x*-axis). Different colours indicate the taxonomy of the respective strain.

and functional root nodule symbiosis remained intact in *Lotus* across all plant generations. Absolute bacterial abundances, assessed through colony-forming unit (CFU) counts from reisolated cultures (**Methods**), showed only a marginal increase throughout the experiment (**Fig. S 1C**). These observations suggest that community

adaptation occurred without substantial expansion of ecological niches or increases in bacterial carrying capacity on either host.

Community profiling *via* 16S rRNA amplicon sequencing showed that bacterial diversity remained high across all treatments and time points, displaying complex patterns of community composition (**Figure 1C-F**). Alpha diversity displayed no significant decline throughout the experiment (**Fig. S 2**). Most of the initial strains present in the start inocula (29 out of 34) persisted through all plant generations, highlighting remarkable community resilience. The exceptions included the two members of the Bacillaceae (Root131 and LjRoot5) and Intrasporangiaceae (Root101 and LjRoot24) families, and the *Lotus*-derived Streptomycetaceae (LjRoot303) strain, which were lost at different plant cycles (**Table 1**).

Comparison of community profiles across timepoints showed striking reproducibility among independent replicate populations, suggesting deterministic community assembly processes. The primary source of compositional variation was the host species, which accounted for a significant fraction of the total variance in microbial community diversity (28.43% of variance; $P = 0.0001$; **Figure 1C-F**). Despite overall stability, relative abundances of specific taxa fluctuated over time, leading to shifts in community composition (19.06% of variance; $P = 0.0001$; **Figure 1C-F**). Notably, one member of the Flavobacteriaceae family (Root935) increased in relative abundance (RA) by 90.1%, on average, from generation 1 to generation 16. Such dynamics could reflect evolutionary adaptation of individual strains to host-associated environments, resulting in fitness advantages over other community members. Alternatively, these changes over time may be the consequence of ecological interactions within the community, including competition and priority effects.

2.3.2. Accumulation and fixation of genomic variants in root-associated bacterial communities

To investigate if the changes in community composition were driven by microbial adaptation, we performed shotgun sequencing of bacterial populations at multiple timepoints throughout our evolution experiment. Following stringent quality control, variant calling, and allele frequency estimation (**Methods**), we identified

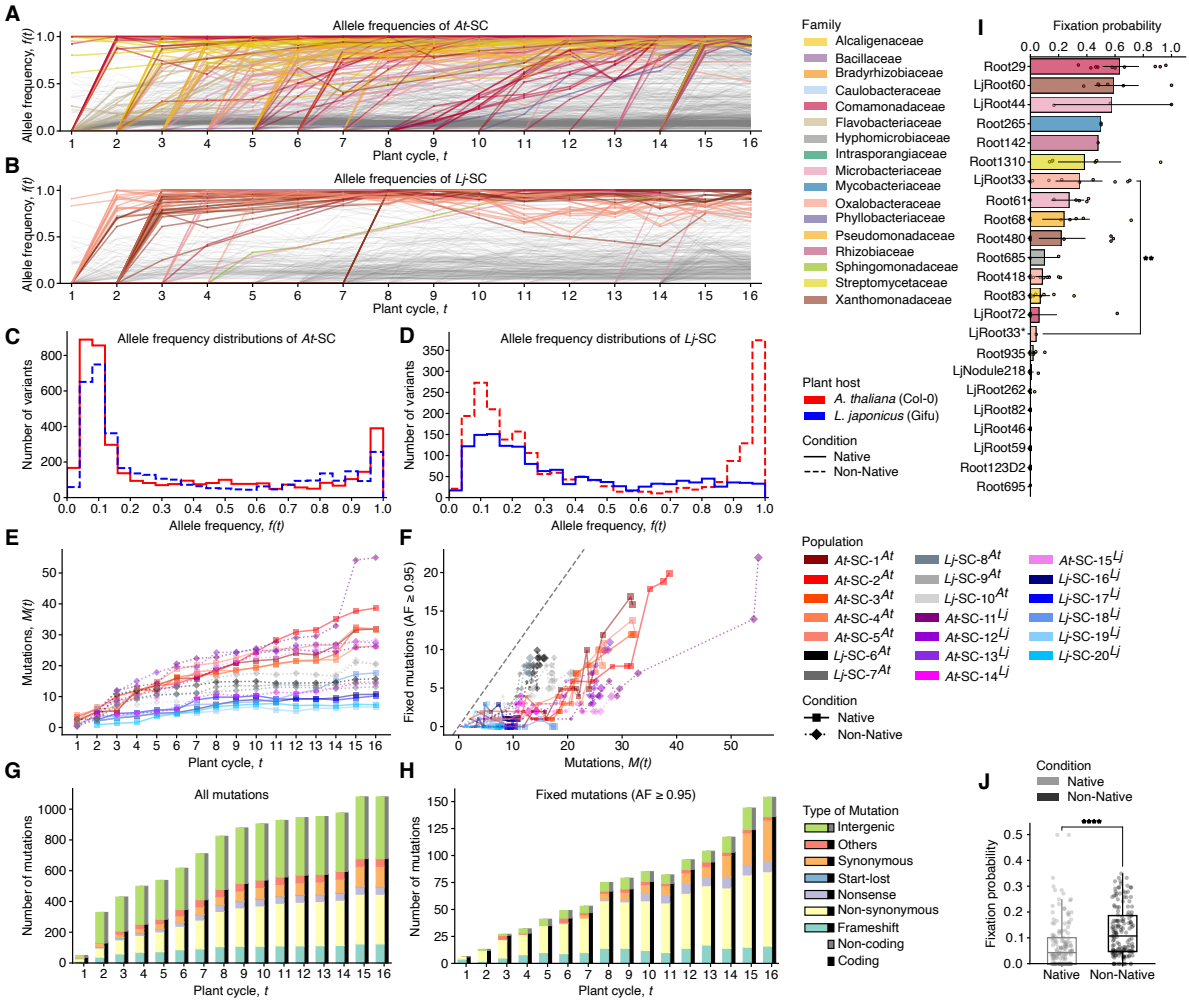

Figure 2. Population-level genetic diversity of evolving SynComs. (**A**-**B**) Individual variant trajectories over time that occurred in evolving *At*-SC (**A**; $n = 736$) and *Lj*-SC (**B**; $n = 344$). Each line represents the allele frequency as a function of time of mutations that arose during the evolution experiment. Mutations that were fixed in the population (AF ≥ 0.95) are coloured according to the taxonomy of the strain. Non-fixed variants are depicted in grey. (**C**-**D**) Histograms of allele frequency distributions for *At*-SC (**C**), and *Lj*-SC (**D**). (**E**) Total derived allele frequency $M(t)$ versus time coloured by replicate population. (**F**) Number of fixed mutations plotted against the $M(t)$. (**G**-**H**) Number of mutations by class plotted against the plant cycle (t) for all and fixed variants (allele frequency ≥ 0.95), respectively. Different colours indicate different classes of mutations. Features affecting coding sequences are additionally coloured in black. Intergenic mutations are coloured in grey. (**I**) Strain-specific fixation probabilities calculated from the fraction of fixed mutations and all mutations ($n = 151$). Colors indicate taxonomy and are the same as in A&B. Comparison of LjRoot33 and the hypermutator was performed using a one-sample t-test. (**J**) Fixation probabilities between native and non-natively evolving SynComs ($n = 313$). Groups were compared using a Mann-Whitney U test.

single-nucleotide variants (SNVs) and short insertions and deletions (indels) that arose and persisted in the genomes of all SynCom members over the course of the experiment. We observed consistent accumulation of these genetic variants in bacterial populations of both SynComs (*At*-SC and *Lj*-SC) in all experimental conditions, with a sustained increase over successive plant generations (**Figure 2A-**

B). Remarkably, both host environments (*Arabidopsis* and *Lotus* roots) drove comparable patterns of variant accumulation, which emerged consistently across independent replicate populations (**Figure 2C-D**).

Next, we calculated the rate at which mutations accumulate through the experiment, by calculating the total derived allele frequency $M(t)$, as the sum of allele frequencies for all mutations observed in a population at a given time t (**Methods**). These values, which correspond to the rate of molecular evolution in a population remained high but stable throughout the experiment (**Figure 2E**). A notable exception occurred in one replicate population, where the *Lotus*-derived Oxalobacteraceae strain (LjRoot33) acquired a mutation in the *mutS* mismatch repair gene, conferring a hypermutator phenotype (Montelone, 2006). This mutation drastically increased the rate of variant accumulation in this lineage by 158.57%, highlighting the emergence of hypermutability as a potential evolutionary strategy within host-associated communities.

However, despite overall similarity across conditions, considerable differences emerged when comparing variant accumulation rates among individual community members (**Fig. S 3**). For instance, strains from the Flavobacteriaceae family showed particularly high substitution rates (4.12×10^{-6} mutations x nucleotide^{-1} per generation, on average), whereas strains from other families such as Bradyrhizobiaceae accumulated significantly fewer mutations (4.19×10^{-7} mutations x nucleotide^{-1} x generation^{-1}, on average). Interestingly, we observed a significant correlation between changes in relative abundance during the evolution experiment (**Figure 1C-F**) and the total derived allele frequency $M(t)$ per strain (**Fig. S 4A**). This suggests that the observed changes in community composition could be at least in part explained by adaptive evolution.

Analysis of allele frequencies over time indicated that a subset of these variants consistently reached fixation in a population, suggesting their significance for adaptation to the plant root environment (**Figure 2A-B**). Conversely, another subset of variants persisted in replicate populations at lower abundances during multiple plant cycles, resulting in bi-modal distributions of allele frequencies (**Figure 2C-D**). These patterns are consistent with clonal interference, where beneficial mutations compete

within the population and only the most advantageous variants reach fixation. In this scenario, some beneficial mutations may not be fixed in the population if they are outcompeted by other bacterial lineages carrying alternative mutations with stronger fitness benefits. However, theory predicts that the lineage carrying the most beneficial variant always wins, causing a selective sweep and eliminating mutations present in other lineages. As a consequence, the total allele frequency at a given time point $M(t)$ should approximate the number of fixed variants in a population (Barrick & Lenski, 2013). Analysis of replicate lines from our evolution experiment displayed patterns consistent with this expectation (**Figure 2F**), indicating the importance of clonal interference as a driver of molecular evolution in bacterial populations, also in a community context.

We then tested if the observed genomic changes led to altered growth rates of individual community members. We employed a metagenomics-based method for estimating growth rates *in situ*, based on differences in read coverage distribution between the origin and terminus of replication in each bacterial genome (Joseph et al., 2022; Korem et al., 2015). We validated this approach through parallel measurements of *in vitro* growth rates of ancestral strains, showing that these two metrics were significantly correlated (**Fig. S 4B**). Throughout the experiment, growth rates at the strain level remained stable, suggesting that the observed genomic changes did not necessarily lead to increased growth at the sampled timepoints (**Fig. S 4C**). Given that samples were collected after five weeks of *in planta* bacterial growth in each cycle, it is possible that changes in growth rates are only relevant during the initial stages of community establishment. Alternatively, adaptive mutations may primarily enhance other traits such as competitiveness within the community, or host colonization efficiency.

Further examination revealed only a weak correlation between strain-specific growth rates and the number of detected variants (**Fig. S 4D**). For instance, the two *Pseudomonas* strains (Root68 and LjRoot59) showed a high growth rate *in planta* but comparatively low substitution rate, while the Flavobacteriaceae strains (Root935 and LjRoot82) displayed the opposite pattern. These contrasting observations suggest that variation in substitution rates is not solely explained by growth dynamics, and that

distinct selective pressures acting on different community members may also contribute to the observed patterns of variant accumulation.

2.3.3. Fixed variants are likely adaptive and vary in occurrence between community members

To better understand the mechanisms underpinning bacterial adaptation in our evolution experiment, we examined what types of mutations appear and are fixed in replicate bacterial populations. The genomic distribution of variants was strongly skewed toward coding regions, with 58.45% of all detected mutations located within protein-coding genes (**Figure 2G**). This enrichment was more pronounced among fixed variants (≥95% allele frequency), of which 88.68% were intragenic (**Figure 2H**). Further classification of these mutations also revealed that fixed variants exhibited a significantly elevated ratio of non-synonymous to synonymous substitutions compared to variants with lower allele frequencies (2.18-fold increase; $P = 0.04$), consistent with positive selection acting preferentially on non-synonymous, likely functional, mutations.

We also observed considerable variation in fixation probability across individual bacterial strains (**Figure 2I**), ranging from approximately 60-80% for members of the Comamonadaceae family (LjRoot72 and Root29) to less than 10% in other lineages, such as Flavobacteriaceae (Root935 and LjRoot82). Interestingly, we observed a significant difference between the fixation probabilities of the hypermutator population of strain LjRoot33 (*Oxalobacteraceae*) and their non-mutator counterparts (from 35.06 to 4.05%, $P = 0.01$). This pattern was expected given the higher rate at which hitchhiker mutations accumulate in hypermutator populations and is consistent with observations from previous LTEEs (Barnett et al., 2025; Good et al., 2017; Lind & Andersson, 2008; Sniegowski et al., 1997; Tenaillon et al., 2016; Wielgoss et al., 2013).

In general, the observed differences in fixation probabilities between bacterial strains are unlikely to be the consequence of varying sequencing coverage, which we have accounted for in our analyses (**Methods**). Instead, it likely reflects differences in effective population size, mutation rate, or in the strength of the selective pressure

imposed on each community member. Intriguingly, fixation probabilities were significantly higher in populations evolving on a non-native host species compared to those maintained on their native host (**Figure 2J**). This result supports the hypothesis of an increased evolutionary potential, or accelerated evolutionary change, when bacterial communities encounter new ecological contexts, such as a new host species.

2.3.4. Parallel evolution of SynComs and reproducible genomic signatures of host adaptation

To further characterize genomic changes associated with bacterial adaptation to plant roots, we analysed the variant profiles using principal component analysis (PCA) based on allele frequencies of detected variants across experimental conditions and timepoints. Notably, replicate populations evolving independently under the same conditions (i.e., host species) exhibited remarkable similarity in their genomic trajectories, evidenced by their clustering in the PCA plots (**Figure 3A-B**). Within each plant host species, samples also grouped by their plant generation, further illustrating the cumulative nature of genetic divergence and variant accumulation over time. These patterns suggest that repeated exposure to the same selective environment drives reproducible evolutionary trajectories, even in complex multispecies communities. We hypothesized that such reproducible evolutionary outcomes could be explained by specific genetic variants appearing independently in multiple replicate lines. Supporting this hypothesis, we identified 197 bacterial genes that acquired high-confidence mutations in multiple populations compared to 307 genes with mutations in only one replicate line. Employing a binomial statistical test, we identified 255 genes (corresponding to 50.9% of all genes with detected mutations) showing significant patterns of non-random emergence across independent populations (**Table 2**; **Methods**). While some of these mutated genes were identified in either SynCom independently of the host species (77 in *At*-SC; **Figure 3C**, and 28 in *Lj*-SC; **Figure 3D**), we also identified mutated genes found exclusively in multiple replicate bacterial populations growing on their non-native host (17 for *At*-SC growing on *L. japonicus* roots; **Figure 3C,** and 28 for *Lj*-SC growing on *A. thaliana*; **Figure 3D**). These findings further support the hypothesis that repeated exposure to a new host species induces reproducible genetic changes in independently evolving bacterial communities. Next,

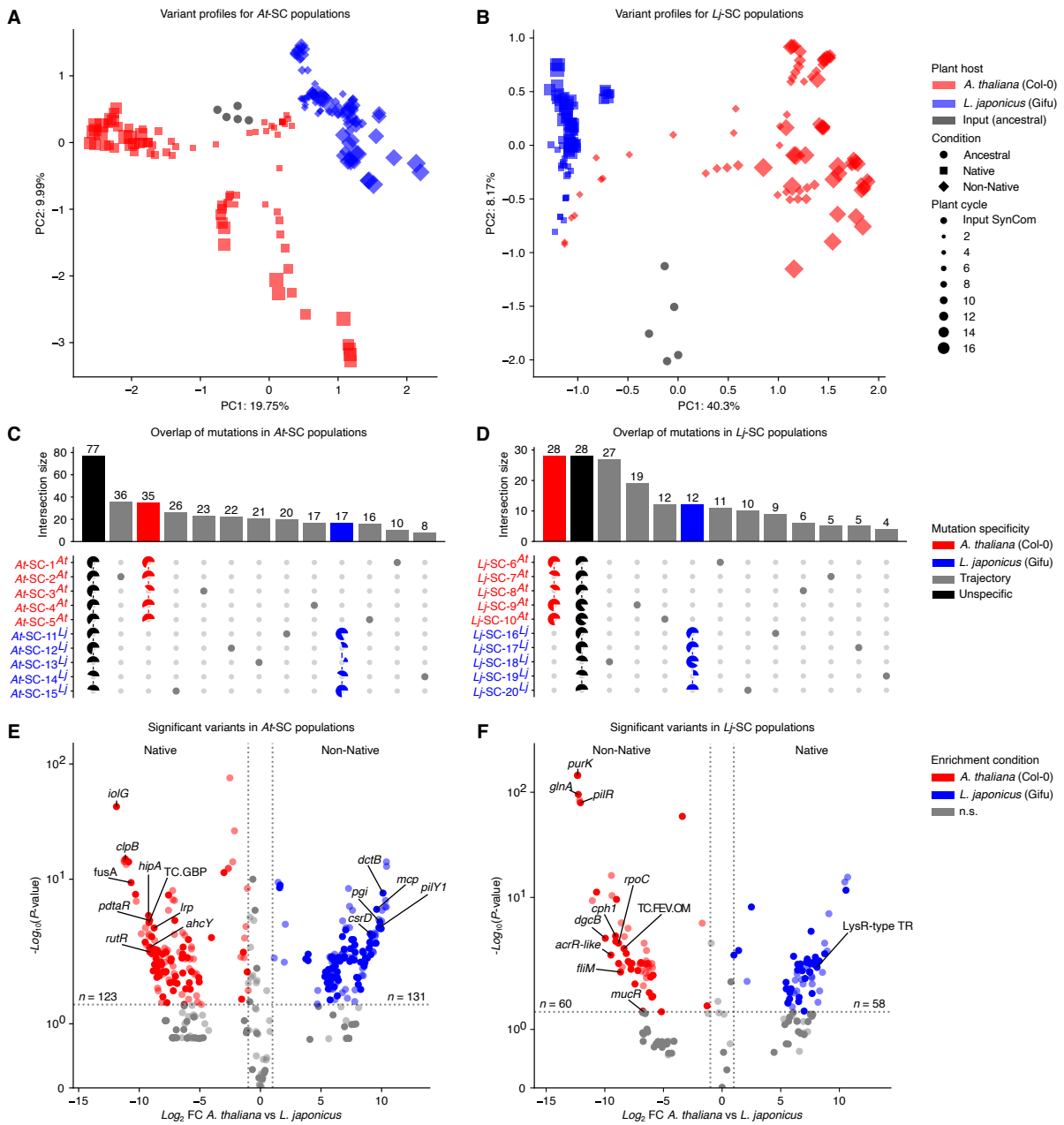

Figure 3. Host-specific variant profiles. (**A**-**B**) Principal component analysis using a matrix of allele frequency values for SynComs *At*-SC (**A**; *n* = 163) and *Lj*-SC (**B**; *n* = 160). Different colours indicate the host plant the SynComs were evolving on during the evolution experiment. Point sizes represent the timepoint (*t*). (**C**-**D**) Upset plots showing the mutational overlap for the *At*-SC (**C**) and *Lj*-SC (**D**) SynComs on the level of open reading frames. Black bars and pie charts indicate mutated open reading frames that were found in both hosts. Red and blue elements represent ORFs that were hit by mutations exclusively during evolution either on *Arabidopsis* or *Lotus* roots. Grey elements were specific to a replicate population. Pie charts summarize the fraction of mutated ORFs that were found in a given replicate population. (**E**-**F**) Mutations significantly enriched on either *Arabidopsis* (red) or *Lotus* (blue) within *At*-SC (**E**) and *Lj*-SC (**F**). Enrichments were determined by fitting a generalized linear model per detected mutation (*n* = 543). *P*-values of enrichments were adjusted using false discovery rate (α = 0.05).

we performed a generalized linear model (GLM) analysis aimed at identifying genetic variants significantly enriched in either SynCom evolving on *Arabidopsis* or *Lotus*

roots. We found 372 out of 543 (or 68.51%) variants that were enriched in bacterial populations associated with one host, indicating differential selection pressures operating on these communities (**Figure 3E-F**; **Table 3**). This is consistent with the observed differences in fixation probabilities (**Figure 2J**), and the clear host-driven clustering patterns observed in the PCA plots (**Figure 3A-B**). Collectively, these results suggest intensified selective pressures and greater adaptive potential when bacterial communities encounter a new host species.

To better understand the functional relevance of host species-enriched variants, we analysed their distribution across different gene categories. A notable proportion of these variants affected genes encoding transcriptional regulators (e.g., LysR-type transcriptional regulators), and two-component systems (e.g., *pilR*, *dctB*). Such mutations have been observed in other experimental evolution systems, where they are thought to mediate rapid adaptation through modulation of transcriptional outputs, rather than innovation by acquisition of new functionality (Barnett et al., 2025; Brown et al., 2011; Conrad et al., 2011; Couce et al., 2024; Deatherage et al., 2017; Hindré et al., 2012; Lescat et al., 2017; E. Li, De Jonge, et al., 2021; E. Li, Zhang, et al., 2021; Saxer et al., 2014; Tenaillon et al., 2016). Besides mutations in genes related to transcriptional regulation, we also found a significant number of host species-enriched variants in genes related to motility and pilus assembly, and thus likely relevant for root colonization (e.g., *fliM*, *fliE*, *fliF*, *pilN*), or involved in core metabolic processes in bacteria (e.g., *glnA*, *purK*, *bglB*, *icd*), which could play a role in adaptation to the distinct root exudation profile of a new host. Collectively, these data support the hypothesis that repeated exposure of SynComs to a new plant species result in reproducible patterns of genomic adaptation.

2.3.5. Increased fitness during colonization of a novel host species after experimental evolution

Next, we sought to experimentally validate whether the genomic changes observed during our evolution experiment translated into increased bacterial fitness during root colonization of a new host species. To achieve this, we conducted a series of competition experiments at multiple timepoints, assessing the capacity of evolved populations to compete for colonization on both plant hosts (**Methods**).

In these experiments, we co-inoculated axenic *Arabidopsis* and *Lotus* plants with independently evolving populations extracted from the previous generation of plants. Specifically, we matched populations evolving on their non-native host against each other (*At*-SC^{Lj} vs. *Lj*-SC^{At}), forcing them to compete for colonization of either plant species. As controls, we performed parallel competition experiments using populations evolving on their native host (*At*-SC^{At} vs. *Lj*-SC^{Lj}). After 5 weeks of plant growth, we assessed community composition through amplicon sequencing and estimated host preference as a proxy for bacterial fitness (**Methods**).

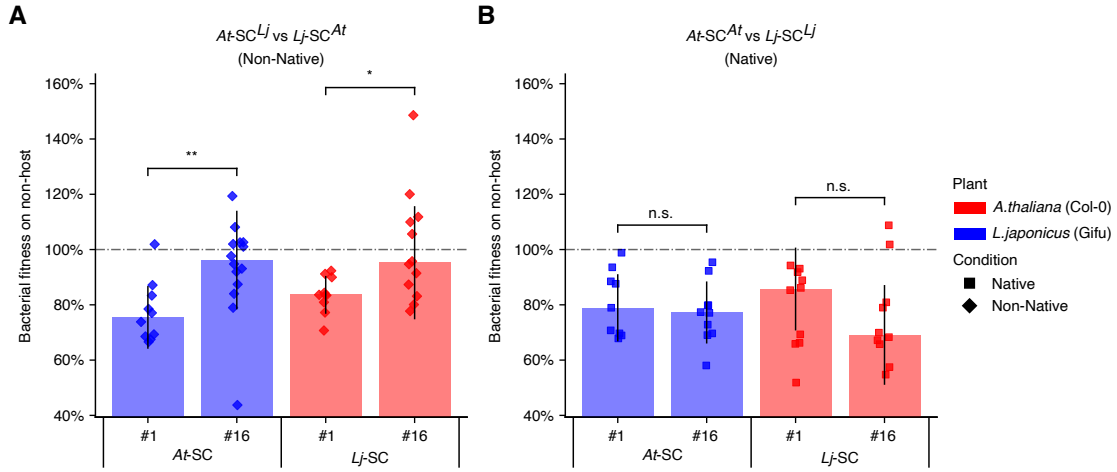

Figure 4. Bacterial fitness changes during experimental evolution. (**A**) Fitness changes of non-natively evolving *At*-SC^{Lj}, and *Lj*-SC^{At} (n = 22-24) at plant cycle the beginning (t = 1) and the end (t = 16) of the evolution experiment. Comparison was performed with a Mann-Whitney U test, and P-values were adjusted by calculating the false discovery rate (α=0.05). (**B**) Fitness changes of natively evolving *At*-SC^{At}, and *Lj*-SC^{Lj} (n = 19-20).

For the *Lj*-derived SynComs evolving on their non-native host (*Lj*-SC^{At}) we observed a significant increase in bacterial fitness during colonization of *A. thaliana* roots of 13.9% by the end of the experiment (generation 16) compared to the start (generation 1; **Figure 4A**). A similar pattern was observed for the *At*-derived SynComs evolving on *L. japonicus* (*At*-SC^{Lj}), which experienced an increase in fitness throughout the experiment of 27.5% (**Figure 4A**). This improvement in fitness emerged gradually and became statistically significant from generation 9 onwards, remaining stable until generation 16 (**Fig. S 5**). In contrast, for populations from both SynComs evolving in their native host (*At*-SC^{At} vs. *Lj*-SC^{Lj}) we observed no significant increase in fitness during colonization of the non-native host (**Figure 4B**), as expected given their lack of exposure to the new plant species.

These results demonstrate that repeated exposure to a novel host species for 16 plant generations (~20 months) is sufficient to induce measurable increases in bacterial fitness. This results in colonization efficiencies exhibited by evolved SynComs that reach, or even surpass, the levels observed in native, host-adapted communities. Importantly, this phenotypic outcome aligns with the reproducible genomic signatures identified in our sequencing analyses (**Figure 3**), providing strong evidence that these genetic changes underpin enhanced fitness on novel hosts. Collectively, our results illustrate the rapid and reproducible adaptive evolution of synthetic bacterial communities to new host environments, driven by host-specific selective pressures.

2.3.6. Identification and validation of adaptive mutations in evolved bacterial isolates

To independently confirm our findings and identify genetic determinants underlying the observed changes in fitness, we isolated individual bacterial strains throughout the experiment (**Figure 1A**, **Figure 5A-B**, **Fig. S 5A**, and **Fig. S 6**). We focused on isolates derived from the final plant generation (16), from which we validated and characterized a collection of 34 evolved clonal bacterial isolates. These isolates were subjected to whole-genome sequencing using long-read PacBio technology.

Comparative analyses between the PacBio genome assemblies and earlier shotgun-sequencing data revealed that the majority (90.48%) of fixed variants identified through population-level sequencing were reliably found in the high-quality genome assemblies of the clonal isolates (**Figure 5A-B**). We also analysed these genomes for large-scale structural rearrangements and signatures of horizontal gene transfer (HGT) events within our evolved bacterial strains. While structural rearrangements occurred sporadically (e.g., a large plasmid loss in strain LjRoot80; **Table 4**), we found no clear evidence of HGT events, suggesting limited genetic exchange within these synthetic communities during the evolution experiment.

To experimentally validate fitness changes observed at the population level (**Figure 4A-B**), we reconstituted SynComs partially composed of evolved clonal

isolates (**Figure 5A-B**). We then performed competition experiments by matching non-natively (*At*-SC^{Lj} and *Lj*-SC^{At}) and natively evolved (*At*-SC^{At} and *Lj*-SC^{Lj}) SynComs with their respective ancestral counterparts (denoted *At*-SC* and *Lj*-SC*) and inoculated them onto axenic *Arabidopsis* and *Lotus* roots. This experimental setup allowed us to assess changes in fitness of evolved clonal isolates compared to the ancestral strains in a community context. These experiments showed that SynComs evolved on a non-native host exhibited increased host preference compared to the ancestral SynComs (**Fig. S 7**), in line with previous results obtained from competition experiments of evolving populations (**Figure 4A**).

We detected four evolved strains with significant fitness increases during colonization of the non-native host (**Figure 5C**). Specifically, the strains Root480^{Lj} (Xanthomonadaceae), and Root142^{Lj} (Rhizobiaceae), originally derived from *A. thaliana* roots, increased their fitness in colonization of *L. japonicus* roots by 376.62% ($P = 1.22 \times 10^{-4}$) and 94.99% ($P = 3.05 \times 10^{-4}$), respectively. Similarly, strains LjRoot60^{At} (Xanthomonadaceae), and LjRoot262^{At} (Sphingomonadaceae) increased their fitness when colonizing their non-native host, *A. thaliana* by 116.85% ($P = 1.52 \times 10^{-4}$), and 106.19% ($P = 1.53 \times 10^{-4}$), respectively. Notably, these taxa were among the most abundant community members throughout the evolution experiment (**Figure 1C-F**), suggesting that their high relative abundance and central ecological role may be related to their adaptive potential to new host environments.

Consistent with this increased fitness, PacBio sequencing confirmed the presence of adaptive mutations in several key genes within the genomes of these evolved strains, which were also previously identified in our population-level analyses via shotgun sequencing (**Table 5**). For instance, Root480^{Lj} carried two mutations in genes related to glycogen metabolism (i.e., glycogen synthase and glycosyltransferase), which may alter carbon and energy storage in this strain, leading to adaptation to a new host environment with higher nutrient availability. Another example was found in LjRoot262^{At}, which carried multiple mutations in the *cysJ* gene, encoding the sulphite reductase enzyme. We speculate that this could be the result of

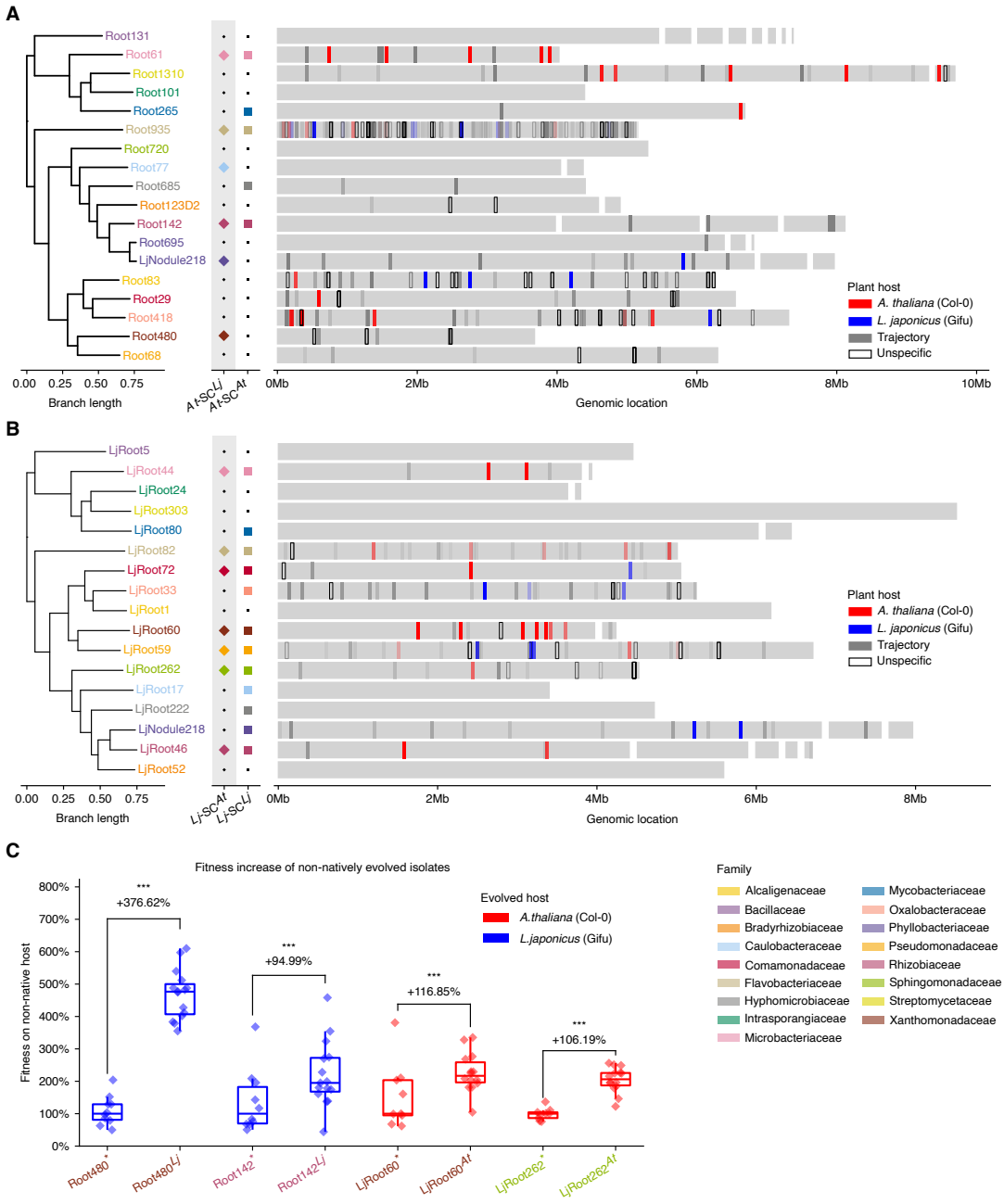

Figure 5. Genetic landscape and phenotype of evolved isolates. (**A**-**B**) Genomic context of detected mutations in the genomes of SynComs *At*-SC (**A**) and *Lj*-SC (**B**). Phylogenetic trees were inferred from all open reading frames. Coloured squares indicate strains for which an evolved isolate was obtained. Colouring of strain names and markers is according to taxonomy. Grey bars indicate the genomic structure of the respective strain. Red, blue, grey and transparent boxes mark candidate mutations that were determined from population level sequencing during the evolution experiment. (**C**) Bacterial fitness during root colonization of four evolved and reisolated strains that showed significantly increased competitiveness under non-native conditions (n = 9-15). Comparisons were performed between the ancestral and evolved strain using a Mann-Whitney U test, and P-values were adjusted by calculating the false discovery rate (α = 0.05).

metabolic streamlining of this strain, originally isolated from roots of *L. japonicus*, in response to *A. thaliana* root exudates, which are rich in sulphur-containing

glucosinolates (Halkier & Gershenzon, 2006; Kopriva et al., 2024) and provide an alternative sulphur source that could be exploited by this strain. Together, these experiments validate both the changes in fitness (**Figure 4**) and the reproducible genomic signatures of adaptation (**Figure 3**) observed during the evolution experiment. Collectively, our findings indicate that the observed genomic and phenotypic changes are driven primarily by adaptation to the new host species, rather than artifacts of adaptation to our reductionist experimental setup.

2.4. Discussion

Microbial communities associated with plants are dynamic ecosystems whose structure and functions influence host health and fitness. While the ecological roles of these communities are increasingly well understood, the evolutionary dynamics that shape them over time have remained largely unexplored. Here, we report the results of an evolution experiment that directly addresses this question. By passaging replicated synthetic bacterial communities through two plant host species for over 16 plant generations, we demonstrate that complex, multispecies microbiomes can undergo rapid and reproducible adaptation to novel hosts, resulting in measurable gains in colonization fitness and the fixation of host-specific genetic changes.

This work builds on the LTEE framework, which has been applied mostly to single-strain (Barrick et al., 2009; Conrad et al., 2011; Couce et al., 2024; Lang et al., 2013; Tenaillon et al., 2016) populations or relatively low-complexity, free-living microbial communities (Castledine et al., 2020; Evans et al., 2020; Fiegna et al., 2015; Hall et al., 2018; Summers et al., 2010; Turkarslan et al., 2021), by extending it to a taxonomically-diverse, host-associated microbiota. Unlike classical LTEEs, where adaptation is often traced to individual clonal lineages in well-mixed environments, our approach captures how selection unfolds across multiple interacting taxa in the structured context of the plant root. Despite this complexity, we observe a surprising degree of evolutionary convergence. Replicate communities exposed to a new host display parallel genomic trajectories (**Figure 3A-B**), overlapping mutation targets (**Figure 3C-D**), and consistent gains in host species-specific colonization fitness (**Figure 4**).

The reproducibility of these outcomes suggests that selection imposed by the plant host constitutes a strong and predictable force in shaping microbiome evolution. This is further supported by the observation that community members evolved on their non-native host accumulated mutations more rapidly and fixed a greater proportion of them compared to populations maintained on their native host (**Figure 2J**). Notably, these patterns were not accompanied by major ecological disruptions. Community composition remained diverse and stable over time, with most founding strains persisting through all 16 plant generations (**Figure 1C-F**). This ecological stability, paired with sustained genetic change, suggests that evolutionary adaptation can proceed without drastic community turnover.

At the genomic level, adaptive evolution manifested in the repeated fixation of mutations in genes belonging to specific functional categories, particularly genes involved in transcriptional regulation, two-component signalling systems, motility, and core metabolism (**Figure 3E-F**). Many of these categories have also been implicated in previous experimental evolution studies (Conrad et al., 2011; Couce et al., 2024; E. Li, De Jonge, et al., 2021; Tenaillon et al., 2016), suggesting that similar genetic routes to adaptation may be shared across systems. However, the recurrence of these mutations across multiple independent populations, and their enrichment under non-native host conditions, highlights the role of the host species in shaping evolutionary outcomes. In particular, adaptation to a novel host appears to involve fine-tuning of regulatory networks and streamlining of metabolic capabilities, rather than acquisition of entirely new functions, a pattern consistent with a model of adaptation through optimization rather than innovation (Barrick & Lenski, 2013).

Our experimental design further allowed us to directly link genetic change to phenotypic outcome. Through a series of competition assays, we show that SynComs evolving on a new plant species consistently outperform both ancestral communities (**Figure 5C**), and those evolved on their native host as controls (**Figure 4**). In several cases, we could trace these gains to putative adaptive mutations in genes related to nutrient assimilation and stress tolerance. For example, mutations in glycogen metabolism and sulphur reduction pathways may reflect selective responses to host-specific root exudates. These strains were also among the most abundant community

members throughout the evolution experiment (**Figure 1C-F**), suggesting a link between ecological success and adaptive potential. Such findings point to the potential for rapid functional tuning of bacterial strains within a community context, with implications for engineering or selecting beneficial microbiota.

Beyond the specific context of plant-microbiota interactions, our study contributes to a broader understanding of how evolution operates in host-associated microbial ecosystems. Unlike many free-living microbial populations, host-associated microbial communities experience repeated population bottlenecks, spatially structured environments, and fluctuating nutrient conditions, all factors that may constrain or direct evolutionary trajectories. Our data show that under such conditions, microbial communities can still evolve rapidly and predictably, even in a community setting. These results have conceptual and practical implications. Conceptually, they suggest that plant microbiomes are not only ecologically dynamic, but also evolutionarily plastic, and that hosts may exert selective forces that shape their microbiota over short timescales. This opens avenues for studying their co-evolutionary dynamics and raises questions about the extent to which host specificity and microbiome function are evolutionarily linked. Practically, our findings inform strategies for microbiome-based interventions in agriculture. They imply that microbial inoculants may be improved through directed evolution approaches that harness host-imposed selection, rather than relying solely on static strain collections or functional screenings.

While our work provides a controlled model for studying microbiome evolution, it also raises new challenges. For example, it remains difficult to disentangle the contributions of interspecies interactions, spatial structure, and host immunity to the observed evolutionary outcomes. Similarly, although we identify candidate adaptive mutations, functional validation remains labour-intensive, particularly in a community context. Finally, our use of synthetic communities, while necessary for tractability, does not fully capture the complexity of natural microbiomes. For example, our initial SynComs include single representative strains of the most abundant bacterial families present in the root microbiota, and do not include species-level variation at the start of the experiment. This could explain why we do not observe clear signatures of HGT

during the evolution experiment, given that the likelihood of exchange of genetic material is lower between bacteria from different taxonomic groups (Soucy et al., 2015; Thomas & Nielsen, 2005). In addition, some microbial taxa observed in natural microbiomes have not yet been cultured (Bai et al., 2015; Wippel et al., 2021) and were therefore excluded from our experiment. Future work should extend this framework to more complex communities and additional hosts, and include transcriptomic, metabolomic, and integrated analyses of host responses to evolved communities.

Together, our results provide experimental evidence that root microbiomes can evolve rapidly, reproducibly, and lead to significant increases in fitness in response to host selection. Our work also provides a number of candidate genes and functions that putatively mediate adaptation to a new plant species, which will be functionally validated in future studies. In addition, our experimental setup establishes a framework for studying the evolution of complex communities in associations with a eukaryotic host, allowing us to probe the rules that govern microbiome evolution.

2.5. Material and Methods

2.5.1. Bacterial material and growth conditions

Bacterial SynComs were reconstituted from previously published culture collections (Bai et al., 2015; Wippel et al., 2021). The two SynComs, *At*-SC and *Lj*-SC, were designed to maximize taxonomic coverage while ensuring that every member had a unique 16S rRNA gene sequence and could be differentiated using amplicon sequencing. Each SynCom comprised 17 strains, representing 17 distinct bacterial families, with the same families represented in both *At*-SC and *Lj*-SC. The design also preserved compositional similarity to SynComs used in a previous study where they were demonstrated to show robust host preference phenotypes (Wippel et al., 2021), thus, facilitating comparative analyses. Bacteria were cultured in TY-medium (5 g l^{-1} tryptone, 3 g l^{-1} yeast extract, 10 mM $CaCl_2$), either in liquid culture or on agar plates with 15 g l^{-1} Bacto agar (Difco). Cultures were incubated at 25 °C; liquid cultures were shaken at 180 rpm. We also determined the maximum *in vitro* growth rate of each strain by adjusting cultures to an initial OD_{600} of 0.02, incubating them in TY-medium at 25 °C for 72 h, and recording absorbance every 10 minutes using a plate reader (Tecan Infinite 200 PRO). Maximum growth rates were estimated using the Growthcurver R package (Sprouffske & Wagner, 2016). The ancestral genomes of all SynCom members were sequenced using PacBio HiFi and served as references for all downstream analyses (16S amplicon, shotgun metagenome, and whole-genome sequencing). Exact SynCom compositions and strain-specific physiological parameters are summarized in **Table 1**.

2.5.2. Plant material and growth conditions

As host plants, *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) and *Lotus japonicus* ecotype Gifu B-129 were used. Plant material and seed sterilization followed established protocols (Wippel et al., 2021). Briefly, *Arabidopsis* seeds were surface-sterilized with 70% ethanol for 5 minutes, washed twice in 100% ethanol for 1 minute, rinsed five times with sterile Millipore-filtered water (ddH_2O), and stratified in darkness at 4 °C. *Lotus* seeds were scarified with sandpaper, incubated in 1% bleach for 20 minutes, and rinsed five times with sterile water. The sterilized *Lotus* seeds were

germinated on sterile Whatman paper moistened with sterile ddH_2O in square Petri dishes under short-day photoperiods (10 h light at $21\,^{\circ}C$ and 14 h dark at $21\,^{\circ}C$).

2.5.3. Gnotobiotic plant experiments

All gnotobiotic plant experiments were conducted using FlowPots (Kremer et al., 2021). Five days prior to the start of the experiment, peat was mixed with vermiculite in a 2:1 ratio. The mixture was then moistened with 200 ml ddH_2O per litre to ensure efficient sterilization during autoclaving. FlowPots were assembled by filling each pot with 65 ml of the soil mixture, with a nylon mesh placed at the bottom to prevent clogging of the system. The assembled FlowPots were autoclaved once daily until the start of the experiment, totalling four rounds of autoclaving.

At the start of the experiment, each FlowPot was flushed with 50 ml of sterilized ddH_2O and 50 ml of 0.2x Murashige and Skoog (MS) medium (0.88 g l^{-1} MS salts) via vacuum infiltration using a QIAGEN manifold (QIAvac 24 Plus, Qiagen). Subsequently, either three *Arabidopsis* or *Lotus* seeds were planted per pot. All procedures were conducted in a clean bench to ensure sterility throughout. Axenic plants were grown for two weeks under short-day conditions.

Next, FlowPots were inoculated with either 20 ml of sterile 0.2x MS medium (axenic control) or 20 ml of 0.2x MS medium supplemented with the SynCom (evolution condition) inoculum. Inoculations were performed under sterile conditions in a clean bench. The pots were then maintained under short-day photoperiods (10 h light at $21\,^{\circ}C$ and 14 h dark at $21\,^{\circ}C$) for additional five weeks until harvest (totalling seven weeks of plant growth), at which point shoot fresh weight was recorded.

2.5.4. Artificial selection of root-derived SynComs

The evolution experiment was conducted from April 2022 until December 2023 and comprised a total of 16 plant cycles (*t*), with each replicate line consisting of two FlowPots, and each FlowPot containing three plant individuals, on average. At the start, bacterial strains were streaked on individual TY-agar plates 14 days prior to inoculation. Taxonomy was confirmed via 16S rRNA gene amplification followed by Sanger sequencing of the obtained amplicons (sequencing performed at Eurofins;

primers 799F and 1192R) (Bai et al., 2015). Bacterial cultures were inoculated as liquid cultures (10 ml of sterile TY-medium with 10 mM $CaCl_2$) seven days prior to the FlowPot inoculation and grown at $25\,^{\circ}C$ while shaking at 180 rpm for seven days. Before the FlowPot inoculation, each liquid culture was washed twice in 10 mM $MgSO_4$ and the optical density was adjusted. For the final inocula, bacterial strains were diluted in 0.2x MS medium to reach a final OD_{600} of 0.02. For all subsequent cycles, extracted root-associated microbial communities containing living bacteria were used as inocula for the next plant cycle, ensuring a continuous selection regime on the bacterial populations.

For the extraction of live root-microbial communities, plant roots were dissected at the end of each plant cycle using sterilized forceps, tweezers, and a brush. After careful removal of all soil particles, roots were stored in phosphate-buffered saline (pH = 7) until all samples were processed. Roots were then dried on sterile Whatman paper and transferred into 250 µl of sterilized 10 mM $MgSO_4$. Root-associated bacteria were released by grinding the roots with sterilized pestles. The resulting root slurry was split into four portions: (a) 10 µl or 5 µl were used for limiting dilutions of *Arabidopsis*- or *Lotus*-derived communities, respectively, with the purpose of counting Colony Forming Units (CFUs) and for re-isolation of evolved isolates; (b) 70 µl were used for inoculation of the next plant cycle; (c) 70 µl from *Arabidopsis* or 35 µl from *Lotus* roots were used for competition experiments; (d) the remaining 100 µl were used for DNA extraction and subsequent amplicon or shotgun metagenome sequencing.

For inoculation of the next plant cycle, 70 µl of root microbial community was diluted into 40 ml 0.2x MS medium, and inoculation of two FlowPots was performed via vacuum infiltration (QIAvac 24 Plus, Qiagen). All steps involved in the harvesting procedure were carried out in sterile conditions. The 100 µl DNA samples were treated with 10 µl (30 Kunitz units) RNase-free DNase I (QIAGEN) for 10 minutes at $37\,^{\circ}C$ to remove free-floating plant DNA. Samples were then snap-frozen in liquid nitrogen and stored at $-80\,^{\circ}C$ until further processing. DNA extractions were performed using the QIAamp DNA Micro Kit (QIAGEN; Protocol: Isolation of Genomic DNA from Tissues). Purified DNA was eluted in 30 µl nuclease-free water and stored at $-20\,^{\circ}C$ until further

processing. DNA samples from the evolution experiment were subjected to amplicon sequencing (in-house MiSeq instrument) and shotgun sequencing (Novogene).

2.5.5. Competition experiments

To assess changes in bacterial fitness, we performed competition experiments between evolving populations throughout the evolution experiment. Evolving replicate populations of *At*-SC and *Lj*-SC were combined at selected plant cycles (1, 3, 5, 7, 9, 11, 12, 13, 14, 15, and 16) to yield 34-member communities. For these experiments, we mixed the natively evolving *At*-SC (*At*-SC^{At}) with the natively evolving *Lj*-SC (*Lj*-SC^{Lj}), as well as the non-natively evolving *At*-SC (*At*-SC^{Lj}) with the non-natively evolving *Lj*-SC (*Lj*-SC^{At}). Pairings of replicate populations (i.e., which *At*-SC and *Lj*-SC were matched in a replicate of the competition experiment) were determined using a pseudo-random number generator in Python.

During a pilot experiment, we observed that CFUs derived from *Lotus* root samples were approximately two-fold higher than those from *Arabidopsis* roots. To establish an equivalent starting point for the competition experiments, 70 µl of the *Arabidopsis*-derived root microbial community was mixed with 35 µl of the *Lotus*-derived community. Importantly, the host preference phenotype, and thus our fitness assessments of differential colonization across plant roots, remained robust against such variations in input ratios (Wippel et al., 2021).

FlowPots for the competition experiments were assembled, inoculated, and SynCom-infiltrated on the same day as the harvest of the evolving populations. Harvesting was performed as previously described (Wippel et al., 2021). In brief, plant roots were dissected using sterile forceps, tweezers, and a brush while submerged in sterile ddH_2O. Soil particles were carefully removed, and roots were blotted on sterile Whatman paper to remove excess liquid. Dissected roots were transferred into Lysing Matrix E tubes (MP Biomedicals), snap-frozen in liquid nitrogen, and stored at $-80\,^{\circ}\mathrm{C}$ until further processing. DNA extraction was carried out using the FastDNA Spin Kit for Soil (MP Biomedicals) following the manufacturer's instructions, and DNA was eluted in 50 µl nuclease-free water. Samples were stored at $-20\,^{\circ}\mathrm{C}$ until further processing.

2.5.6. Re-isolation of evolved isolates

For re-isolation of evolved SynCom members, limiting dilutions were performed as previously described (J. Zhang et al., 2021). Briefly, 10 µl of homogenized root slurry from *Arabidopsis* or 5 µl from *Lotus* were first diluted in 1.25 ml of 10 mM $MgSO_4$. This was followed by further dilution in 50 ml of 0.5× TY-medium (supplemented with 10 mM $CaCl_2$) to reach the final concentration. For the 16th plant cycle, re-isolation efforts were doubled. Dilution factors in the second step were adjusted as needed across the experiment to ensure appropriate final concentrations (summarized in **Table 6**). For culture collection establishment, 160 µl of 0.5× TY-medium containing evolved strains was distributed into three 96-well plates (six plates for the 16th cycle). Plates were sealed with micropore tape (3M) and incubated at root temperature for two weeks. This procedure yielded 60 plates from plant cycles 1, 3, 5, 7, 9, 11, 13, and 15, and 120 plates from cycle 16. After incubation, wells showing visible growth were identified by scanning the plates with a plate reader (Tecan Infinite 200 PRO) to obtain fluorescence spectra and optical density data. Next, 150 µl of 80% (vol/vol) sterilized glycerol was added to each well. Plates were sealed with an aluminium seal (ThermoFisher) and stored at $-80\,^{\circ}\mathrm{C}$ until further processing. Fluorescence spectra were used to classify re-isolated taxa, enabling high-throughput screening without the need to sequence all plate × well combinations.

First, we collected training data ranging from 144 to 216 replicates per strain, as well as 432 replicates for medium-only controls (**Table 7**), yielding a comprehensive dataset for strain-specific classification. Spectral data were processed using custom Python scripts. Briefly, raw spectra were filtered to retain values within the linear range of the measurement device. To account for plate-to-plate variation, we performed medium-only measurements for each plate; these medium-only spectra were used to normalize the fluorescence data in a plate-specific manner. Standardization was performed by applying a z-score transformation followed by a hyperbolic arcsine transformation. The resulting normalized spectra were then corrected using a Python reimplementation of the ComBat function from the sva package (Leek et al., 2014) to account for day-to-day measurement differences. The adjusted spectra were renormalized using a z-score transformation. The final spectra were used to train a

random forest classifier with 500 estimators (scikit-learn v1.5.2). The resulting classifier was therefore capable of distinguishing wells containing only medium from wells exhibiting bacterial growth, and of discriminating among different taxa. We then applied the classifier to the fluorescence spectra obtained from the collection of evolved and re-isolated bacterial isolates (**Fig. S 6**). From these results, we estimated the number of wells lacking bacterial growth. Accounting for the applied dilution factors (**Table 6**), we inferred absolute bacterial abundances (i.e., colony-forming units) of root-associated microbial communities at the population level (**Fig. S 1C**).

For recovery of specific strains of interest from the evolution experiment, glycerol stock plates were retrieved from the $-80\,^{\circ}\mathrm{C}$ freezer, and TY agar plates were spot-inoculated and streaked using the three-way dilution method. After one week of growth, single colonies were picked, and taxonomic identity was confirmed via 16S rRNA Sanger sequencing. Verified colonies were inoculated into 5 ml liquid TY-medium supplemented with $CaCl_2$ and incubated for one week at $25\,^{\circ}\mathrm{C}$ with shaking at 180 rpm. Glycerol stocks were prepared by mixing 800 μl of 80% sterilized glycerol (vol/vol) with 800 μl of culture and stored at $-80\,^{\circ}\mathrm{C}$ for long-term preservation.

The resulting culture collection of evolved strains comprised approximately 23,000 CFUs across all plant cycles. From this collection, 309 strains were re-isolated, re-streaked, and validated via 16S rRNA sequencing. Of these high-quality isolates, 79 were used for reconstitution experiments (**Table 8**), and 32 were subjected to whole-genome sequencing using PacBio HiFi sequencing (Max Planck Genome Centre).

2.5.7. Reconstitution experiments with evolved isolates

Gnotobiotic reconstitution experiments were performed to independently validate the results of the competition experiments conducted during the evolution experiment. Axenic plants were grown for two weeks in FlowPots and inoculated with 0.2x MS medium (control) or with 0.2x MS medium containing different SynComs (**Table 8**), as described previously. Mixed communities consisting of ancestral and evolved strains were prepared using the same protocol used for initiating the evolution experiment. The design of the evolved communities (**Table 8**) was largely determined

by the availability of sequence-verified evolved isolates at the time of the experiment. Our intention was to perform competition experiments with SynComs composed exclusively of evolved strains; however, two practical constraints prevented complete recovery of all evolved members. First, bacterial abundances on plant roots follow strongly skewed, log-shifted abundance distributions, which makes it challenging to recover low-abundance isolates. Second, evolved isolates can only be obtained from strains that persisted until the final plant cycle. Thus, taxa that went extinct during the evolution experiment are necessarily excluded. As a consequence, the competition experiments using evolved isolates, particularly experiment 1 (exp_1; **Table 8**), were carried out with SynComs consisting only partially of evolved strains. For taxa where no evolved representative could be recovered, we substituted with the ancestral strain, yielding SynComs that were evolved but partially ancestral. This strategy was chosen to eliminate SynCom diversity as a confounding factor. Thus, the competition experiments comprised 19 of 34 (56%) evolved strains. When accounting for extinction events (i.e., excluding taxa that went extinct during the experiment), these proportions increase to 19 of 29 (66%), representing the majority of community members. Harvesting was performed five weeks post-inoculation and followed the same procedure as used for the competition experiments during the evolution experiment.

2.5.8. 16S rRNA gene amplicon sequencing

At each plant cycle, we performed amplicon sequencing of the bacterial 16S rRNA gene in order to characterize community structures. DNA samples were first adjusted to 1 ng μl^{-1} in nuclease-free water (QIAGEN). The 16S rRNA v5-v7 region was then amplified using primers 799F and 1192R with the following parameters: $94\,^{\circ}C$/2 min, $94\,^{\circ}C$/30 s, $55\,^{\circ}C$/30 s, $72\,^{\circ}C$/30 s, $72\,^{\circ}C$/10 min for 25 cycles (Durán et al., 2018). Following amplification, PCR products were treated with 1 μl Antarctic phosphatase, 1 μl Exonuclease I, and 2.44 μl Antarctic Phosphatase buffer (New England BioLabs GmbH, Frankfurt, Germany). Next, 3 μl of the digested PCR product was used as template for a second PCR to add sample-specific barcodes ($94\,^{\circ}C$/2 min, $94\,^{\circ}C$/30 s, $55\,^{\circ}C$/30 s, $72\,^{\circ}C$/30 s, $72\,^{\circ}C$/10 min for 10 cycles). The correct size of the amplicons from both PCR steps was confirmed using gel electrophoresis. After the second PCR, all samples were pooled to generate the final library. Gel electrophoresis

and subsequent gel extraction using QIAquick (QIAGEN) were performed to remove chloroplast-derived 16S rRNA amplicons. The final library was purified with AMPure XP (Beckman Coulter) and eluted in 30 µl nuclease-free water. DNA concentration was quantified using QuantiFluor (Promega), and sequencing was performed on the in-house Illumina MiSeq platform.

2.5.9. Processing of 16S rRNA gene amplicon data

Raw sequencing reads obtained from the MiSeq platform were demultiplexed using QIIME2 v2021.11.0 (Bolyen et al., 2018). Forward and reverse reads were subsequently merged with FLASH v2.2.00 (Magoč & Salzberg, 2011). Sequences containing undetermined nucleotides (i.e., "N" bases) were identified with usearch v10.0.240 (Edgar, 2010) and removed from the dataset. Reference-based error correction of high-quality sequences was performed using Rbec v1.1.4 (P. Zhang et al., 2021) with a subsampling depth of 500 and a Phred-scaled quality threshold of 33. The resulting count table was corrected for 16S rRNA gene copy-number variation and filtered for a minimal depth of 1000 reads. Subsequent downstream analyses were conducted in Python. Counts were normalized by calculating strain-specific relative abundances (RA). Stacked area plots (**Fig. 1C-F**) were plotted using ggplot2 and data preprocessing was performed using the *get_Muller_df* function of the ggmuller R package (Noble, 2019). Multivariate analysis (i.e., the effect of host plant on community structure) were performed using the R package vegan (Oksanen et al., 2007).

2.5.10. Host preference and fitness indices

We used a modified version of the host preference index (Wippel et al., 2021) to assess bacterial fitness. Specifically, fitness at the SynCom level (At-SC and Lj-SC) was estimated from differences in relative abundances observed in the competition experiments during the evolution experiment. For each sample i, we first calculated the aggregated relative abundance, $a_{sc,plant}^{(i)}$, of all strains in SynCom sc on host $plant$. These values were then normalized by dividing by the median aggregated abundance of the same SynCom on its native host, $\tilde{a}_{sc,\text{native}}$, yielding the fitness score $f_{sc,plant}^{(i)}$ as

defined in Equation 1. Here, $sc \in$ {At-SC, Lj-SC} denotes the SynCom, $plant \in$ {*Arabidopsis, Lotus*} indicates the host plant, i indexes individual samples, and "native" refers to the host on which the SynCom naturally evolved and were originally isolated from.

$$f_{sc,plant}^{(i)} = \frac{a_{sc,plant}^{(i)}}{\tilde{a}_{sc,\text{native}}} \quad (1)$$

Differences in bacterial fitness were then investigated by comparing $f_{sc,plant}^{(i)}$ across different plant cycles (t) for a given condition. In particular, we focused on comparing fitness between the first and the last (sixteenth) plant cycle (plant cycle 1 vs. plant cycle 16; **Fig. 4**).

For competition experiments with individually re-isolated evolved strains, fitness was calculated analogously to the SynCom-level fitness (Equation 1). Here, abundances of evolved isolates were normalized to the median abundance of their ancestral counterparts (plant cycle $t = 0$), yielding the fitness score $f_{strain}^{(i,t)}$ defined in Equation 2.

$$f_{strain}^{(i,t)} = \frac{a_{strain}^{(i,t)}}{\tilde{a}_{strain}^{(t=0)}} \quad (2)$$

In this equation, $a_{strain}^{(i,t)}$ denotes the relative abundance of isolate $strain$ (recovered at plant cycle t) in sample i, and $\tilde{a}_{strain}^{(t=0)}$ represents the median abundance of the ancestral, non-evolved strain. Here, $strain \in$ {*At*-SC strains*, Lj*-SC strains}, t indicates the plant cycle from which the evolved strain was isolated, and i indexes individual samples. In practice, competition experiments with re-isolated strains were restricted to isolates from the final plant cycle ($t = 16$). To identify candidate evolved strains, we focused on non-natively evolved isolates that exhibited a statistically significant increase in fitness, i.e., for which $f_{strain}^{(i,16)} > f_{strain}^{(i,0)}$ under the applied significance threshold, as determined by an adjusted *Mann-Whitney U* test.

2.5.11. Processing of whole genome sequencing data

Raw HiFi sequencing reads were first filtered to remove the internal PacBio sequencing control by mapping the reads using blasr v1.3.1 (Chaisson & Tesler, 2012). Only unmapped reads were retained for downstream analyses. Assemblies

were then performed using Hifiasm v0.25.0-r726 with default parameters (Cheng et al., 2021) and Flye v2.9.6-b1802 (Kolmogorov et al., 2020) with coverage set to 70 and genome size set to 5 Mb. Both assembly graphs were visualized using Bandage v0.8.1 (Wick et al., 2015) and manually inspected for structural issues such as bubbles. The quality of the assemblies was validated by re-aligning the filtered reads against the newly obtained assemblies using minimap2 v2.28-r1209 (-ax map-hifi) (H. Li, 2018). Contigs with a mapping quality lower than 30 were considered assembly artifacts and removed. Circlator v1.5.5 (Hunt et al., 2015) was then used to scaffold circular, chromosome-level assemblies, and the remaining contigs were ordered by size in descending order. Assembly statistics were computed using assembly-stats v1.0.1 (Challis, 2014), and the assembly with the highest *N50* was selected for downstream processing. The final assemblies were further evaluated for (a) contamination using CheckM2 v1.0.1 (Chklovski et al., 2023), (b) completeness using BUSCO v5.8.2 (Seppey et al., 2019), and (c) correct taxonomy using GTDB-Tk v2.1.1 (Chaumeil et al., 2022). Genome annotation was performed with Prokka v1.14.5 (Seemann, 2014) using an in-house curated KEGG-ontology database (Kanehisa & Goto, 2000).

The same procedure was applied to evolved strains isolated from generation 16 of the evolution experiment. Mutations in evolved isolates were identified using two approaches: (a) assembly-based variant calling, where the assembly of the evolved genome was mapped against the assembly of the ancestral strain with Minimap2 (-x asm5) and calling variants using paftools.js (H. Li, 2018); and (b) read-based variant calling by mapping the HiFi reads against the ancestral assembly with Minimap2 (-ax map-hifi) and calling variants using DeepVariant v1.9.0 (-model_type="PACBIO"; (Poplin et al., 2018). The resulting variant call sets were compared using custom Python scripts. Because the two sets were highly similar, only the read-based variant calls from DeepVariant were retained for downstream analysis, as this approach also allowed estimation of allele frequencies for identified variants.

Lastly, we also investigated the occurrence of structural variation (e.g., horizontal gene transfer, duplications or larger insertions or deletions). Therefore, we aligned the reads obtained from evolved isolates against the ancestral genome using

Minimap2. The resulting alignments were sorted by coordinates and used for structural variant calling using sniffles v2.6.3 (Smolka et al., 2024) with default parameters. Only structural variants with a homozygous genotype call (GT 1/1) were considered for downstream analysis. The majority of structural variants were deletions (12/13); however, we also detected one insertion. In order to identify the origin of the inserted DNA sequence, we used blast v2.14.1 (Altschul et al., 1990) to search the inserted DNA fragment in a concatenated reference containing all genome assemblies of the two SynComs. As the inserted sequence seemed to originate from the same genome, we concluded that these structural variants were not the consequence of horizontal gene transfer. Notably, the inserted sequence affected the coding sequence of a transposase, strongly suggesting an intra-genome rearrangement rather than horizontal gene transfer.

2.5.12. Processing of shotgun sequencing data

Raw sequencing reads were subjected to quality control and host decontamination using KneadData v0.12.0 (https://huttenhower.sph.harvard.edu/kneaddata). Briefly, KneadData performed adapter trimming with Trimmomatic v0.39-2 (Bolger et al., 2014) (parameters: MINLEN:60, SLIDINGWINDOW:4:20), removal of tandem repeat sequences using TRF v4.09.1 (Benson, 1999), host DNA removal with Bowtie2 v2.5.3 (Langmead & Salzberg, 2012), and quality assessment with FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). As host reference genomes, we used the *Arabidopsis* TAIR10 (Phytozome genome ID: 167; NCBI taxonomy ID: 3702) and *Lotus* Lj1.0v1 (Phytozome genome ID: 571; NCBI taxonomy ID: 34305) genome assemblies.

Next, the decontaminated reads were mapped against a concatenated SynCom reference using Bowtie2. The resulting alignments were sorted with *samtools sort* v1.6 (H. Li, 2009), mate coordinates were added using *samtools fixmate*, and duplicate alignments were removed using *samtools markdup*. Potential mapping errors were corrected by performing probabilistic realignments with *lofreq viterbi* v2.1.5 (Wilm et al., 2012). Alignment and base-calling quality scores were then refined using *lofreq indelqual* (-dindel) and *lofreq alnqual*. The resulting alignments were used for all

downstream analyses. Strain-specific sequencing depth and coverage were determined with *samtools coverage*.

In situ growth rates were estimated using coPTR v1.1.6 (Joseph et al., 2022), which leverages uneven coverage distributions between the bacterial origin of replication and the terminus. First, decontaminated sequencing reads were mapped against the concatenated SynCom reference genomes using *coptr map* with default parameters. Growth rates were then calculated with *coptr extract* and *coptr estimate* with default parameters. Strain specific *in situ* growth rates were then parsed and analysed using custom Python scripts.

2.5.13. Variant calling and analysis of shotgun sequencing data

Variant calling was performed using LoFreq, considering only mappings with a mapping quality ≥ 20 (-min-mq 20), using a *P*-value cutoff of 0.05 (-sig 0.05) and enabling indel calling (-call-indels). Potential multi-allelic variant calls were decomposed using *vt decompose* v0.5 (Tan et al., 2015) (smart mode, -s), followed by normalization with *vt normalize* and removal of duplicate variants using *vt uniq*. Annotation of the candidate variants and effect predictions were performed with snpEff v4.3t (Cingolani et al., 2012) (settings: -o gatk -noLog -noStats -no-downstream -no-upstream -no-utr -v -c). We further examined whether intergenic variants occurred in flanking regions of coding sequences to account for variants with potential effects on gene expression with a set of custom Python scripts. Finally, VCF files were parsed using custom Python scripts to obtain tab separated tabular files (tsv).

Variant calling from population-level sequencing poses several challenges. In particular, the detectability of a variant is constrained by the sequencing depth of the respective genome relative to the total community depth. Consequently, variants from low-abundance taxa are easily underestimated when depth is limiting. To account for this effect, we performed variant-level rarefaction as follows. We generate $N = 100$ evenly spaced subsampling depths $depth_s$

$$depth_s = depth_{\min} + (i-1)\frac{depth_{\max} - depth_{\min}}{N-1} \quad (i = 1, \ldots, N).$$

where $depth_{\min}$ corresponds to the number of reads for the shotgun sample with the lowest number of reads and $depth_{\max}$ to the highest observed number of reads. For sample i and subsampling depth s we compute the scaling factor $\alpha_{s,i}$

$$\alpha_{s,i} = \frac{depth_s}{reads_i}.$$

If $\alpha_{s,i} > 1$ (i.e., the requested subsampling depth exceeds the raw reads of sample i), we skip that (s, i) pair (no up-sampling). For each variant v in sample i we denote the raw read support by

$$f_{\text{ref}},\ r_{\text{ref}},\ f_{\text{alt}},\ r_{\text{alt}}$$

(forward / reverse reads supporting reference and alternative alleles). The *in silico* subsampled counts are then obtained by scaling:

$$\tilde{f}_{\text{ref}} = \left(\alpha_{s,i}\, f_{\text{ref}}\right), \quad \tilde{r}_{\text{ref}} = \left(\alpha_{s,i}\, r_{\text{ref}}\right),$$

$$\tilde{f}_{\text{alt}} = \left(\alpha_{s,i}\, f_{\text{alt}}\right), \quad \tilde{r}_{\text{alt}} = \left(\alpha_{s,i}\, r_{\text{alt}}\right).$$

We then compute the subsampled depth (DP) and strand balance (SB):

$$\text{DP} = \tilde{f}_{\text{ref}} + \tilde{r}_{\text{ref}} + \tilde{f}_{\text{alt}} + \tilde{r}_{\text{alt}}, \qquad \text{SB} = \frac{\tilde{f}_{\text{ref}} + \tilde{f}_{\text{alt}}}{\tilde{r}_{\text{ref}} + \tilde{r}_{\text{alt}}}.$$

A variant passes the subsampled filter if

$$\text{DP} > 10 \quad \text{and} \quad 0.2 < \text{SB} < 5.$$

We repeat this procedure for all N subsampling depths in $depth_s$, discarding variants for which $\alpha_{s,i} > 1$. The pass/fail outcomes across subsampling depths provide a depth-aware measure of variant robustness. We visualized these results as rarefaction curves (**Fig. S 8**), plotting the number of detected variants per sample as a function of subsampling depth. Samples that reach a clear plateau (asymptote) at higher subsampling depths were considered to have sufficient coverage for robust variant calling. In contrast, samples that show an approximately linear increase in detected variants with depth are indicative of incomplete saturation and insufficient sequencing depth. Therefore, samples exhibiting no clear plateau were prioritized for additional sequencing to reach saturation. Additionally, we sought to independently

validate the accuracy of the shotgun sequencing data. To this end, we compared the strain-specific relative abundances obtained from amplicon sequencing and shotgun sequencing. We observed a strong and highly significant correlation (R^2 = 0.844; $P < 10^{-308}$) between the two methods, indicating that the metagenomic shotgun sequencing data are highly accurate.

Allele frequencies (af_i) at a given genomic locus i, were calculated as the fraction of reads supporting a variant ($reads_{\text{alt},i}$) relative to the total number of reads covering that locus ($reads_{\text{total},i}$; **Equation 3**).

$$af_i = \frac{reads_{\text{alt},i}}{reads_{\text{total},i}} \tag{3}$$

The variant call set was further filtered to include only variants supported by at least three mutant reads, with a strand bias between 0.2 and 5, a minimum Phred-scaled quality score of 30, a minimum total sequencing depth (i.e., the number of reads at the respective loci) of 10. In addition, variants had to be detected in at least four distinct timepoints within the same replicate population, and exhibit an allele frequency (AF) ≥ 0.1 in at least one plant cycle. To capture variants that emerged late in the experiment, the requirement for detection across plant cycles was relaxed: variants in generations 2 and 15 were required to be detected in at least three timepoints, whereas variants in generations 1 and 16 were required to be detected in at least two timepoints. Lastly, for downstream analysis we removed all variants that were already fixed in the starting inoculum to exclusively focus on variants that were being selected for during the experiment. For visualization of variant allele frequencies over time (**Figure 2A-B**), missing values were imputed by linear interpolation where applicable; otherwise, values were set to zero for early plant cycles and carried forward from the last detected value for later plant cycles. Variant classes (**Figure 2G-H**) obtained from snpEff and post-processed with custom Python scripts. To simplify analysis, related snpEff variant types were grouped into broader categories:

Others:	INTRAGENIC, CODON_CHANGE_PLUS_CODON_INSERTION, CODON_INSERTION, CODON_CHANGE_PLUS_CODON_DELETION, TRANSCRIPT, CODON_DELETION
Non-synonymous:	NON_SYNONYMOUS_CODING
Synonymous:	SYNONYMOUS_CODING
Nonsense:	STOP_GAINED
Frameshift:	FRAME_SHIFT

Start-lost:	START_LOST
Intergenic:	INTERGENIC

For each replicate population and timepoint t, we calculated the total derived allele frequency, *M(t)*, as the sum of all observed allele frequencies (Good et al., 2017). Fixation probabilities were calculated as the fraction of fixed variants (AF ≥ 0.95) relative to the total number of detected variants (Good et al., 2017). The resulting values were either aggregated at the strain level to obtain strain-specific fixation probabilities (**Figure 2I**) or at the condition level to obtain differential fixation probabilities for natively and non-natively evolved SynComs (**Figure 2J**). To estimate the proportion of variance in the variant call set explained by experimental conditions (host plant, SynCom, and plant cycle), redundancy analysis (RDA) was performed using scikit-bio (Rideout et al., 2025). Euclidean distances were calculated from the matrices of allele frequencies, followed by principal coordinate analysis (**Figure 3A-B**).

For the generation of the Upset plots (**Figure 3C-D**), we first determined which replicate populations harboured candidate mutations in any given open reading frame (ORF). Next, we assessed whether a particular ORF was mutated exclusively in SynComs evolving on the same host and assigned a colour accordingly (red for SynComs evolving on *Arabidopsis* and blue for *Lotus*). ORFs that showed no pattern of host specificity were coloured black, while ORFs mutated in only a single replicate population were coloured grey. By convention, upset plots display all possible groups (in our case, all possible combinations of replicate populations with the same mutated ORFs). However, we chose to aggregate counts at the level of *Arabidopsis*-specific, *Lotus*-specific, line-specific, and unspecific ORFs. We summarized this information as pie charts (**Figure 3C-D**). In this modified version of the Upset plot, each pie chart indicates the relative fraction of mutated ORFs occurring in the corresponding replicate population.

To test for host-specific enrichment of candidate variants, we implemented generalized linear models (GLMs) using statsmodels v0.14.4 (Skipper & Josef, 2010). Each GLM was trained to predict allele frequency enrichment as a function of host plant. *P*-values were corrected for multiple testing across variants within each SynCom

using false discovery rate (FDR) control. Data visualization was performed using Seaborn and Matplotlib using Python. To assess non-random patterns of mutation occurrence, we applied a binomial test in a strain-specific and ORF-specific manner. Specifically, the probability $P(X \geq k)$ of observing k or more mutations in a given ORF was calculated using the Equation 4.

$$P(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \quad (4)$$

The total number of trials (n) was defined as the total number of mutations observed in the respective strain across all replicate populations. For each ORF, k (the number of successes) was defined as the number of replicates in which a mutation was detected within the respective ORF. The expected success probability, p, was calculated as the ratio of the added length (in base-pairs) of the ORF ($length_{\text{ORF}}$) and its upstream region ($length_{\text{upstream}}$) to the total genome length in nucleotides ($length_{\text{genome}}$), as shown in Equation 5.

$$p = \frac{length_{\text{ORF}} + length_{\text{upstream}}}{length_{\text{genome}}} \quad (5)$$

To reduce false positives arising from singleton events, only ORFs with $k \geq 2$ were considered in the analysis. *P*-values, from the binomial test were corrected for multiple comparisons across ORFs using the false discovery rate (FDR) control.

2.5.14. Data visualization

Plotting was, with the exception of **Figure 1C-F**, performed using the Matplotlib and Seaborn ecosystem under Python v3.12.10. Phylogenetic trees (**Figure 5A-B**) were plotted using Biopython v1.74 (Cock et al., 2009). As tree we used the species tree obtained from Orthofinder with default parameters v2.5.5 (Emms & Kelly, 2019) performed in the two SynComs (*At*-SC and *Lj*-SC). Whenever boxplots were used in figures, data were represented as median values (horizontal line), Q1 - 1.5× interquartile range (boxes) and Q3 + 1.5× interquartile range (whiskers), whereas small circles or triangles displayed with jitter corresponded to individual datapoints.

Schematic images (i.e., **Figure 1A-B**; **Fig. S 5A**) were generated with Biorender. Final figures were assembled using Inkscape.

2.5.15. Statistical analysis and reproducibility

For statistical inference, non-parametric tests were applied where appropriate. Two-group comparisons (including fitness changes during and after the evolution experiment, plant phenotypes, alpha diversity, and fixation probabilities) were evaluated using the Mann-Whitney *U*-test. Fitness indices between ancestral and evolved isolates (**Figure 5C**) were assessed using a one-sided test. The fixation probability of LjRoot33 and its hypermutator (**Figure 2I**) was evaluated via a one-sample *t*-test. Correlations (e.g., growth rates *in vitro* vs. *in situ*, growth rate vs. *M(t)*, relative abundance vs. *M(t)*, and CFUs vs. time) were quantified using Spearman's *r*. Variant enrichments were determined using generalized linear models (GLMs), and non-randomness was evaluated via binomial tests. *P*-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure, with significance set at $\alpha = 0.05$ unless stated otherwise.

2.5.16. Code and data availability

Raw sequencing reads have been deposited at the European Nucleotide Archive (ENA) under accession numbers PRJEB98833 (amplicon sequencing), PRJEB98838 (shotgun sequencing) and PRJEB98839 (PacBio sequencing). Scripts for data analysis and plotting are available on GitHub: https://github.com/bdn-227/exp_evo_rootmicrobiota.

2.6. Supplementary Figures

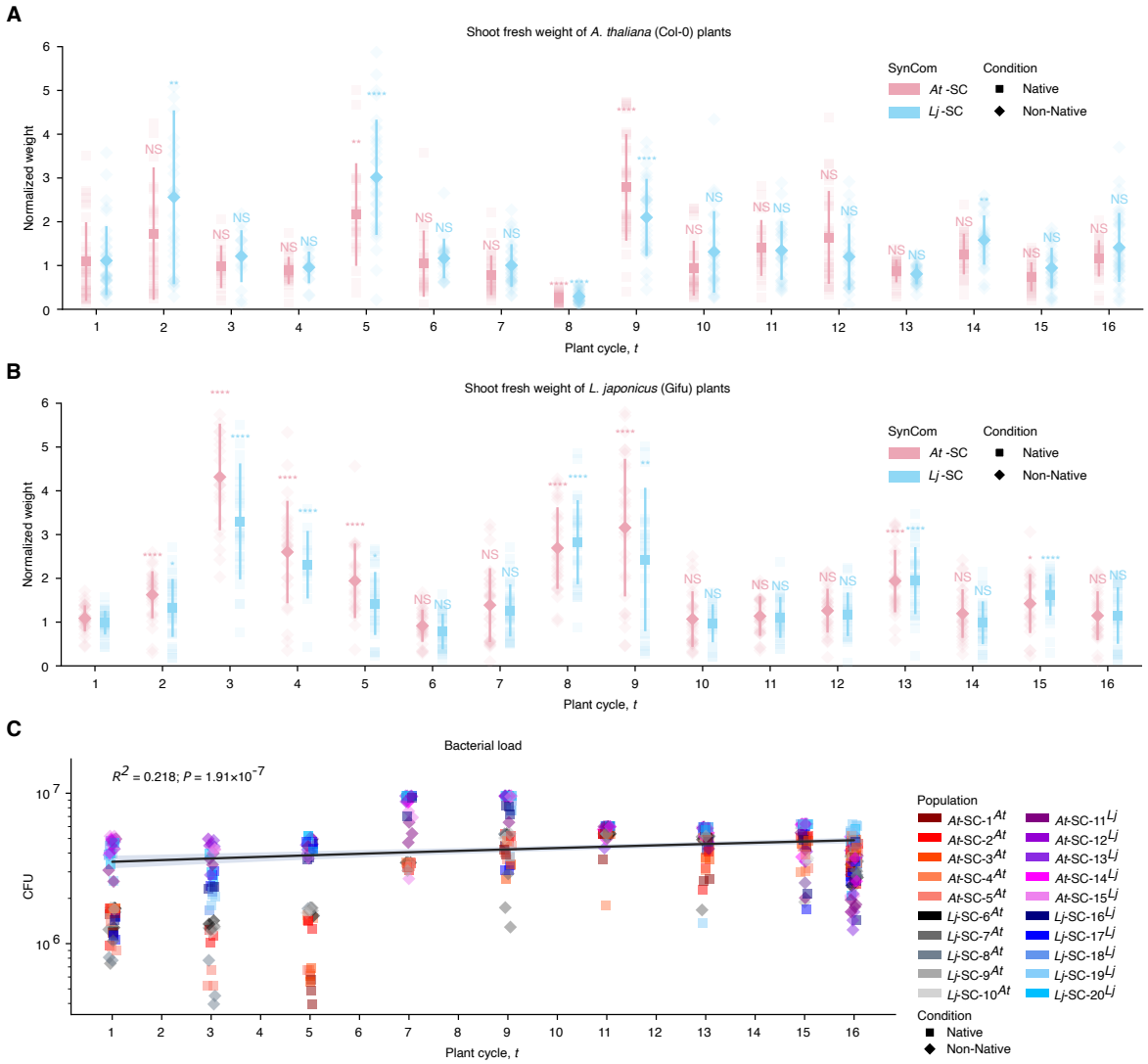

Fig. S 1. Plant phenotypes. Plant fresh weight (**A**-**B**) and bacterial CFU counts over time (**C**). Plant weight (n = 875-897) was normalized to axenic controls. Statistical comparisons were made against the first plant cycle, with false discovery rate correction ($\alpha = 0.05$). CFU values ($n = 560$) were estimated from reisolated strain cultures by adjusting the number of wells with visible growth for the dilution factor. Temporal trends were assessed using Spearman's rank correlation between plant cycle and CFU count.

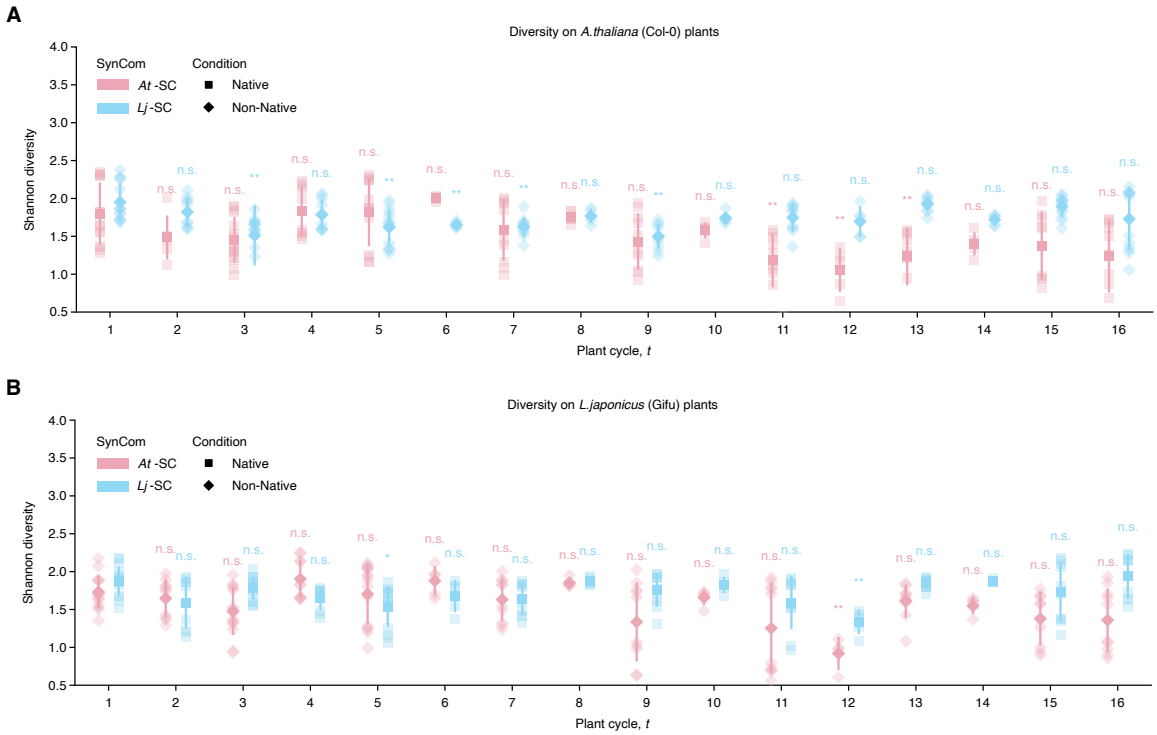

Fig. S 2. Bacterial α-diversity of evolving populations. On *Arabidopsis thaliana* (**A**) and *Lotus japonicus* (**B**) obtained from amplicon data. Shannon entropy was calculated for SynComs evolving on both host plants (n = 294-299). Statistical comparisons were made against the first plant cycle, with P-values adjusted for multiple testing using false discovery rate control (α = 0.05).

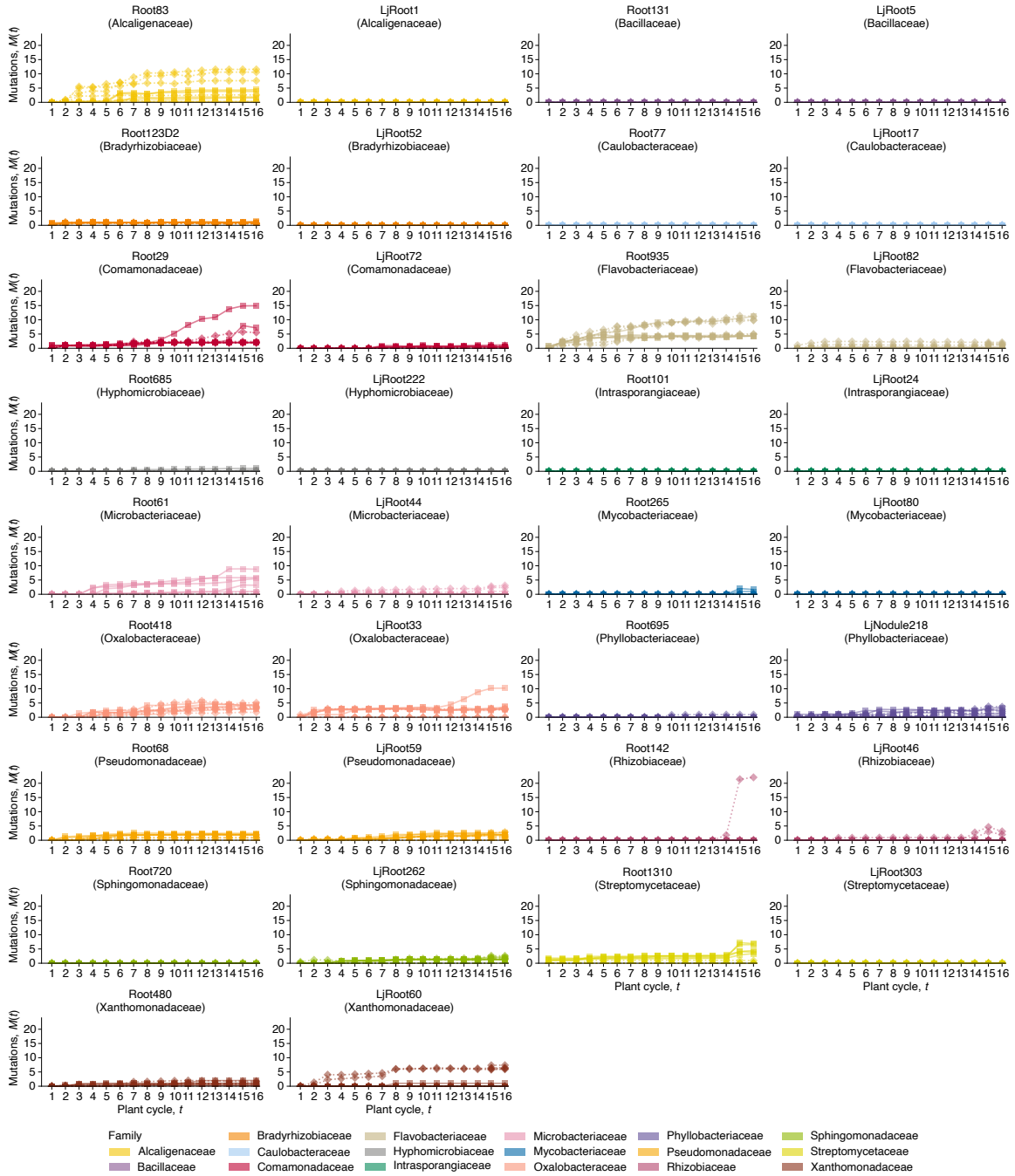

Fig. S 3. Strain-specific patterns of variant accumulation over time. Total derived allele frequencies *M(t)* are plotted against plant cycle for each strain. Lines represent the temporal trajectory of mutation accumulation within individual genomes.

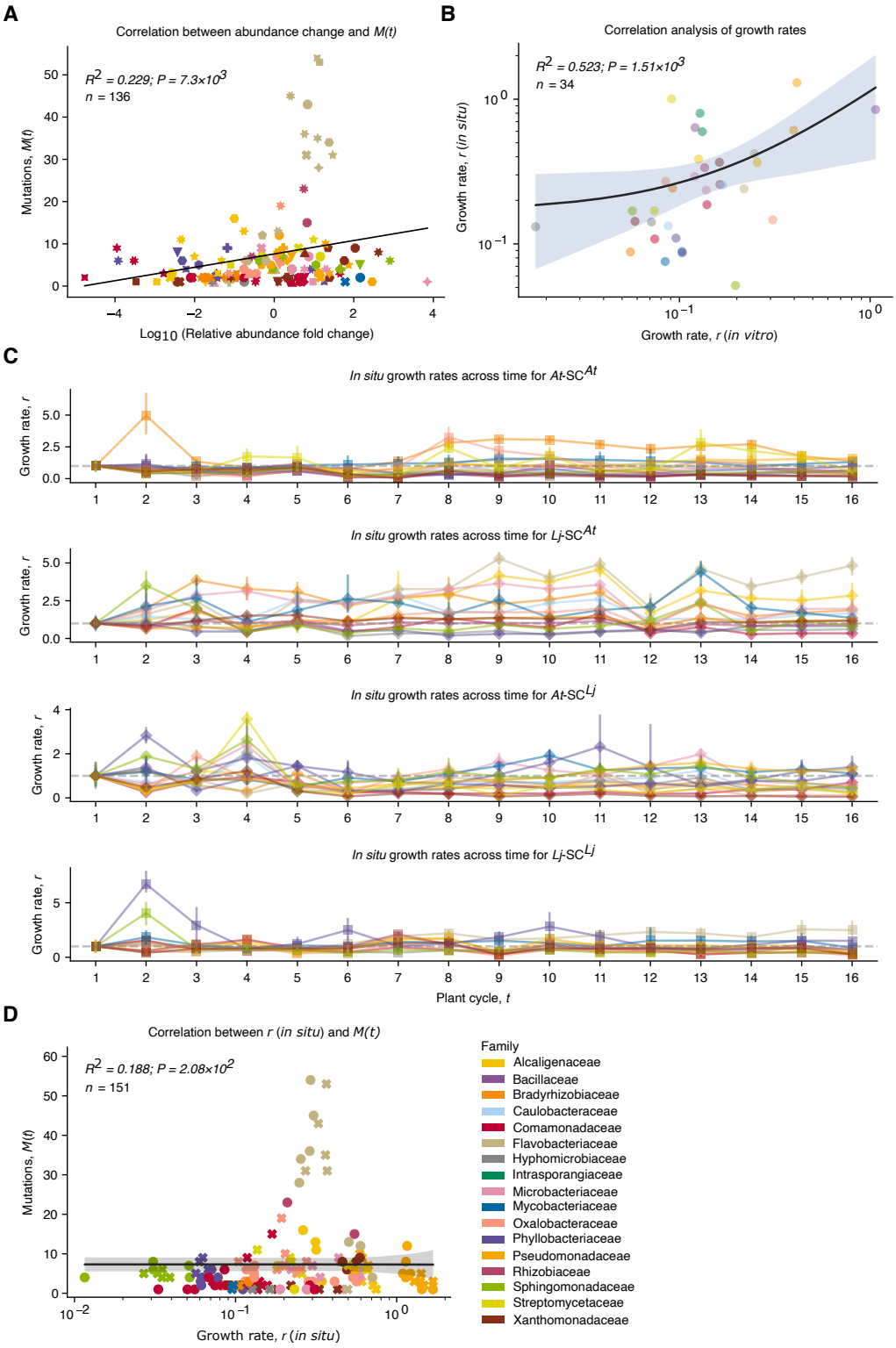

Fig. S 4. Bacterial growth rates during the evolution experiment. (**A**) Correlation between fold change in relative abundance and total derived allele frequency *M(t)* (*n* = 136). (**B**) Correlation between *in vitro* and *in situ* growth rates (*n* = 34). (**C**) *In situ* growth rates across plant cycles (*n* = 4508). (**D**) Correlation between *in situ* growth rates and *M(t)* (*n* = 151).

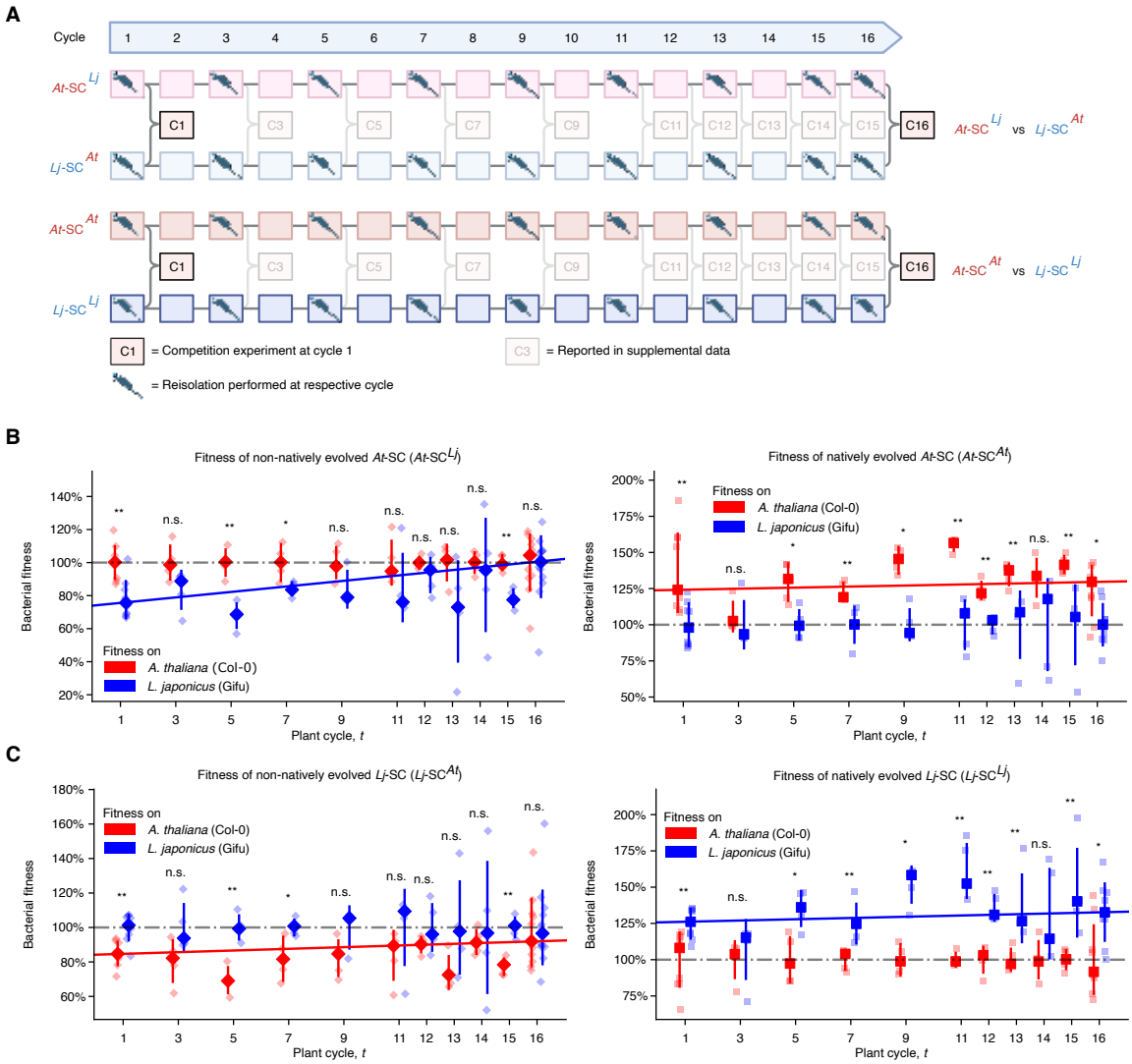

Fig. S 5. Changes in host preference of evolving populations. (**A**) Schematic of competition experiments indicating timepoints and mixed conditions. Pipette symbols mark collection of evolved isolates via limiting dilution. (**B-C**) Fitness trajectories of *At*-SC and *Lj*-SC evolving under native and non-native conditions (n = 128-134). Statistical comparisons assessed SynCom fitness on *Arabidopsis* versus *Lotus* roots (colours). Upper panels show fitness on non-native hosts, lower panels on native hosts. Coloured lines depict fitted trajectories of bacterial fitness on the experimental evolution host over time.

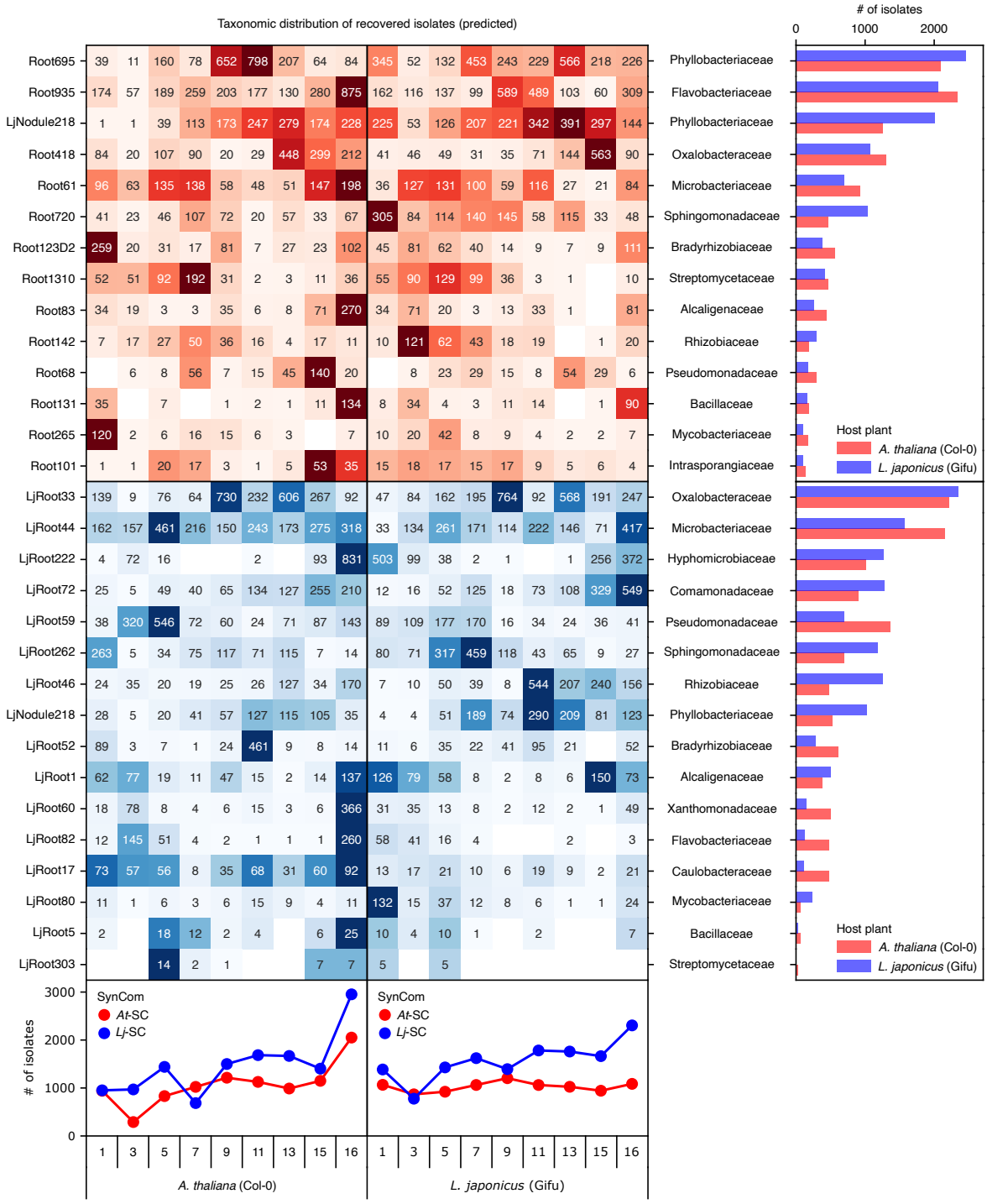

Fig. S 6. Taxonomic distribution of evolved and reisolated strains. Heatmap showing the predicted number of isolates per taxon at each timepoint. *At*-SC strains are red, *Lj*-SC strains are blue. Aggregated counts per SynCom and host are shown below the heatmap, and total counts per taxon across all timepoints are shown on the right.

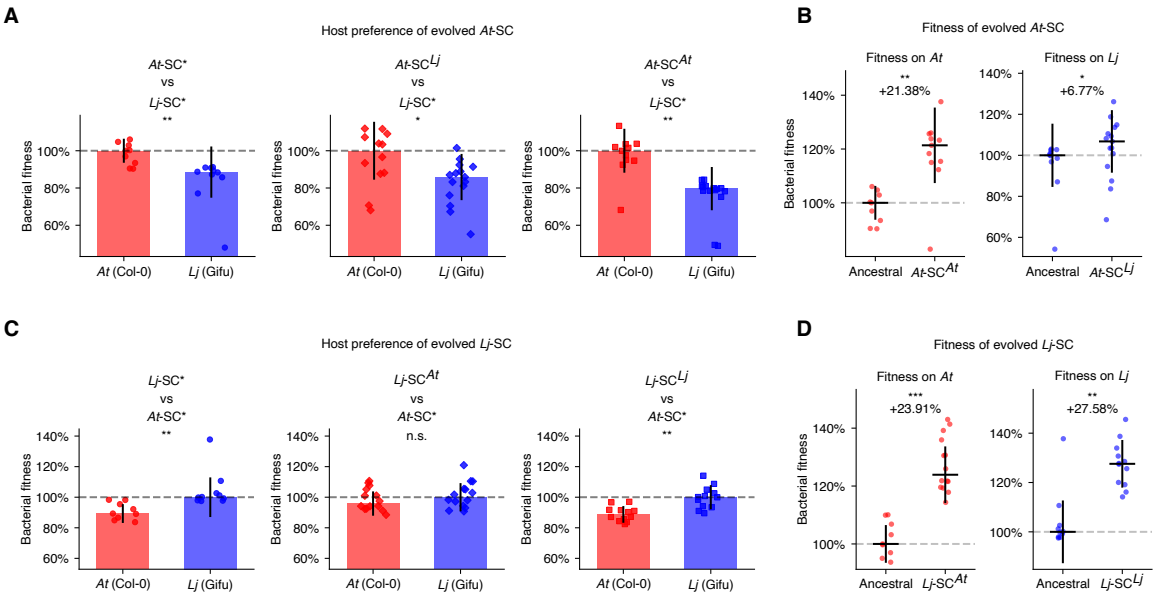

Fig. S 7. Fitness changes in SynComs containing evolved isolates. (**A** and **C**) Host preference of evolved isolates from *At*-SC and *Lj*-SC (n = 10-15). The first panel shows host preference of the ancestral strains; the second shows competitions between non-natively evolved isolates and the ancestral counterpart of the other SynCom; the third shows competitions between natively evolved isolates and the corresponding ancestral SynCom. Fitness values were normalized to the native host. (**B** and **D**) Fitness of evolved *At*-SC and *Lj*-SC on both hosts, normalized to their ancestral fitness. Multiple testing correction was applied using false discovery rate control (α = 0.05).

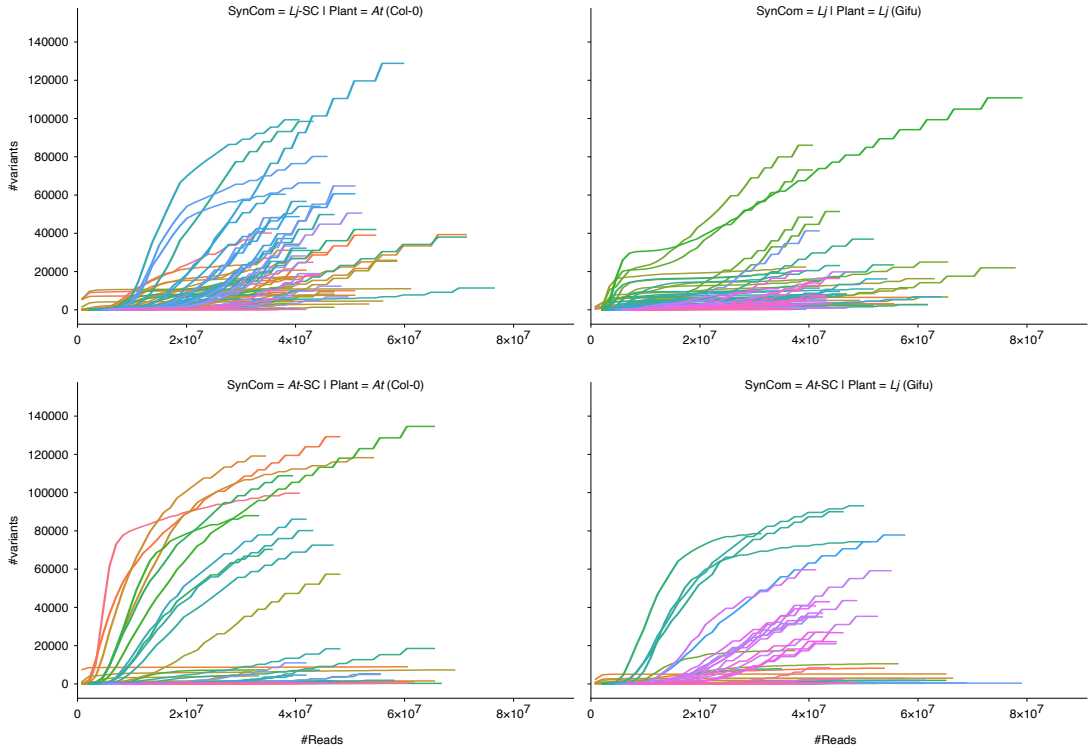

Fig. S 8. Rarefaction curves of variant calls. The number of detected variants (*y*-axis) is plotted against subsampling depth. Each line represents a shotgun sequencing sample. Analyses were performed for both SynComs (*At*-SC and *Lj*-SC) evolving on both hosts, resulting in four facets.

2.7. Supplementary tables

Table 1. Summary of the SynComs used in this study. Strain name, native host, and taxonomic family are listed ($n = 34$). If extinction events occurred, the generation at which they occurred is indicated in the respective column. Strain-specific *in vitro* growth rates were inferred from real-time OD measurements of clonal strains in microtiter plates. *In situ* growth rates were calculated from uneven coverage distributions observed in population-level shotgun sequencing data. From these rates, the estimated number of bacterial doublings throughout the evolution experiment were calculated.

Strain	Collection	Family	Extinction	r (in vitro) [1/h]	r (in situ) [1/h]	Doublings
LjNodule218	Lj-PSHERE	Phyllobacteriaceae	-	0,10	0,08	1564
LjRoot1	Lj-PSHERE	Alcaligenaceae	-	0,26	0,42	3871
LjRoot17	Lj-PSHERE	Caulobacteraceae	-	0,17	0,28	2520
LjRoot222	Lj-PSHERE	Hyphomicrobiaceae	-	0,07	0,14	1063
LjRoot24	Lj-PSHERE	Intrasporangiaceae	3	0,13		1937
LjRoot262	Lj-PSHERE	Sphingomonadaceae	-	0,20	0,04	2968
LjRoot303	Lj-PSHERE	Streptomycetaceae	1	0,09		1376
LjRoot33	Lj-PSHERE	Oxalobacteraceae	-	0,08	0,36	1282
LjRoot44	Lj-PSHERE	Microbacteriaceae	-	0,12	0,39	1820
LjRoot46	Lj-PSHERE	Rhizobiaceae	-	0,14	0,43	2055
LjRoot5	Lj-PSHERE	Bacillaceae	1	0,12		1822
LjRoot52	Lj-PSHERE	Bradyrhizobiaceae	-	0,09	0,30	1358
LjRoot59	Lj-PSHERE	Pseudomonadaceae	-	0,40	0,85	6002
LjRoot60	Lj-PSHERE	Xanthomonadaceae	-	0,16	0,46	2455
LjRoot72	Lj-PSHERE	Comamonadaceae	-	0,14	0,21	1520
LjRoot80	Lj-PSHERE	Mycobacteriaceae	-	0,08	0,08	1244
LjRoot82	Lj-PSHERE	Flavobacteriaceae	-	0,25	0,50	3748
Root101	At-SPHERE	Intrasporangiaceae	15	0,13	0,61	1995
Root123D2	At-SPHERE	Bradyrhizobiaceae	-	0,06	0,10	834
Root131	At-SPHERE	Bacillaceae	11	1,07		15926
Root1310	At-SPHERE	Streptomycetaceae	-	0,07	0,21	1070
Root142	At-SPHERE	Rhizobiaceae	-	0,16	0,26	2469
Root265	At-SPHERE	Mycobacteriaceae	-	0,10	0,09	1558
Root29	At-SPHERE	Comamonadaceae	-	0,07	0,10	1111
Root418	At-SPHERE	Oxalobacteraceae	-	0,31	0,16	4406
Root480	At-SPHERE	Xanthomonadaceae	-	0,06	0,18	885
Root61	At-SPHERE	Microbacteriaceae	-	0,14	0,31	2091
Root68	At-SPHERE	Pseudomonadaceae	-	0,41	1,49	6259
Root685	At-SPHERE	Hyphomicrobiaceae	-	0,02	0,14	265
Root695	At-SPHERE	Phyllobacteriaceae	-	0,10	0,10	1453
Root720	At-SPHERE	Sphingomonadaceae	-	0,06	0,16	857
Root77	At-SPHERE	Caulobacteraceae	-	0,09	0,15	1315
Root83	At-SPHERE	Alcaligenaceae	-	0,13	0,47	1908
Root935	At-SPHERE	Flavobacteriaceae	-	0,22	0,31	3305

Table 2. Genes exhibiting non-random patterns of mutation across replicate populations. The table lists the gene ID, strain, genomic coordinates, evolution host plant, and corresponding *P*-values of genes showing patterns of non-randomness (n = 259). Non-randomness was assessed at the ORF level using binomial tests, and *P*-values were adjusted for multiple testing using false discovery rate (FDR) control (α = 0.05). In addition, the table indicates the replicate populations in which mutations for each ORF were found. KEGG ontology identifiers and gene names are also provided.

geneID	strain	chrom	location	plant	syncom	padj	pct_line	line	ko	alias
Root68_gene_19465	Root68	Root68_contig_1	4317231	col0	at	3,14E-15	1	1, 2, 3, 4, 5		hypothetical protein
Root123D2_gene_11745	Root123D2	Root123D2_contig_1	2471397	col0	at	2,80E-14	1	1, 2, 3, 4, 5	K12506	ispDF
Root29_gene_25385	Root29	Root29_contig_1	5667789	col0	at	2,45E-13	1	1, 2, 3, 4, 5	K01784	galE, GALE
Root29_gene_25260	Root29	Root29_contig_1	5642802	col0	at	2,45E-13	1	1, 2, 3, 4, 5	K03100	lepB
Root123D2_gene_15050	Root123D2	Root123D2_contig_1	3121371	col0	at	2,79E-13	1	1, 2, 3, 4, 5	K02014	TC.FEV.OM
Root68_gene_22920	Root68	Root68_contig_1	5104081	col0	at	1,82E-12	1	1, 2, 3, 4, 4, 5, 5	K21024	bifA
Root61_gene_13160	Root61	Root61_contig_1	2741764	col0	at	2,20E-12	1	1, 2, 3, 4, 5	K07243	FTR, FTH1, efeU
Root61_gene_13165	Root61	Root61_contig_1	2742976	col0	at	2,29E-12	1	1, 2, 3, 4, 5	K02028	ABC.PA.A
Root1310_gene_19875	Root1310	Root1310_contig_1	4630701	col0	at	6,79E-12	0,8	1, 3, 4, 4, 5, 5	K07119	K07119
Root1310_gene_19860	Root1310	Root1310_contig_1	4628374	col0	at	4,81E-11	0,8	1, 3, 4, 4, 5, 5		hypothetical protein
Root418_gene_820	Root418	Root418_contig_1	192082	col0	at	6,01E-11	0,8	1, 3, 4, 5	K02483	K02483
Root61_gene_3270	Root61	Root61_contig_1	729374	col0	at	8,86E-11	0,8	1, 3, 4, 5	K09017	rutR
Root935_gene_4275	Root935	Root935_contig_1	934313	col0	at	9,13E-11	1	1, 2, 3, 4, 5		hypothetical protein
Root935_gene_7035	Root935	Root935_contig_1	1571453	col0	at	9,88E-11	1	1, 1, 2, 2, 3, 4, 5	K11991	tadA
Root83_gene_740	Root83	Root83_contig_1	163003	col0	at	1,08E-10	0,6	3, 3, 3, 4, 5, 5, 5, 5		tRNA-Asn(gtt)
Root935_gene_7610	Root935	Root935_contig_1	1704272	col0	at	1,34E-10	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5		hypothetical protein
Root935_gene_16795	Root935	Root935_contig_1	3789670	col0	at	1,34E-10	1	1, 1, 2, 2, 3, 3, 4, 4, 5		hypothetical protein
Root1310_gene_35700	Root1310	Root1310_contig_1	8125312	col0	at	2,20E-10	0,8	1, 3, 4, 5	K00010	iolG
Root935_gene_5150	Root935	Root935_contig_1	1139487	col0	at	2,33E-10	1	1, 1, 2, 3, 4, 5		hypothetical protein
Root935_gene_20640	Root935	Root935_contig_1	4795347	col0	at	4,25E-10	1	1, 2, 3, 4, 4, 5	K00954	E2.7.7.3A, coaD, kdtB
Root935_gene_15180	Root935	Root935_contig_1	3448197	col0	at	4,60E-10	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5	K16137	nemR
Root418_gene_27355	Root418	Root418_contig_1	6319508	col0	at	5,19E-10	0,8	2, 3, 4, 5	K05962	E2.7.13.1
Root935_gene_5460	Root935	Root935_contig_1	1217195	col0	at	5,88E-10	1	1, 2, 3, 3, 4, 5		hypothetical protein
Root935_gene_20235	Root935	Root935_contig_1	4707455	col0	at	5,88E-10	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5	K07095	K07095
Root1310_gene_28320	Root1310	Root1310_contig_1	6474670	col0	at	1,56E-09	0,8	1, 3, 4, 5	K02355	fusA, GFM, EFG
Root935_gene_7010	Root935	Root935_contig_1	1566832	col0	at	2,29E-09	1	1, 1, 2, 2, 3, 4, 5		hypothetical protein
Root935_gene_15100	Root935	Root935_contig_1	3425034	col0	at	2,52E-09	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5	K12243	pchR
Root935_gene_5475	Root935	Root935_contig_1	1219606	col0	at	4,93E-09	1	1, 2, 3, 3, 4, 5	K07004	K07004
Root935_gene_21490	Root935	Root935_contig_1	5001977	col0	at	7,74E-09	1	1, 2, 2, 3, 4, 4, 5	K01218	gmuG
Root935_gene_7685	Root935	Root935_contig_1	1732119	col0	at	8,44E-09	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5	K07182	K07182
Root1310_gene_41560	Root1310	Root1310_contig_2	138190	col0	at	9,11E-09	1	1, 2, 3, 4, 5		hypothetical protein
Root61_gene_7285	Root61	Root61_contig_1	1554691	col0	at	9,49E-09	0,6	1, 1, 4, 4, 5, 5		hypothetical protein
Root418_gene_21770	Root418	Root418_contig_1	5090865	col0	at	1,24E-08	0,6	1, 2, 3	K03624	greA
Root935_gene_20585	Root935	Root935_contig_1	4783665	col0	at	1,42E-08	1	1, 2, 3, 4, 4, 5		hypothetical protein
Root935_gene_21425	Root935	Root935_contig_1	4983093	col0	at	1,52E-08	1	1, 2, 2, 3, 4, 4, 5	K01881	PARS, proS
Root418_gene_1545	Root418	Root418_contig_1	345158	col0	at	1,87E-08	0,6	1, 4, 5, 5	K03719	lrp
Root935_gene_16810	Root935	Root935_contig_1	3793597	col0	at	1,88E-08	1	1, 1, 2, 2, 3, 3, 4, 4, 5	K01193	INV, sacA
Root83_gene_1140	Root83	Root83_contig_1	246389	col0	at	2,48E-08	0,6	1, 2, 5	K07154	hipA
Root935_gene_5835	Root935	Root935_contig_1	1296518	col0	at	3,85E-08	1	1, 2, 3, 3, 4, 4, 4, 5		hypothetical protein
Root1310_gene_20780	Root1310	Root1310_contig_1	4821360	col0	at	3,86E-08	0,6	1, 2, 5	K18476	tetR
Root265_gene_32235	Root265	Root265_contig_1	6621428	col0	at	4,23E-08	0,4	1, 4	K03098	APOD
Root935_gene_11660	Root935	Root935_contig_1	2622677	col0	at	6,12E-08	0,8	2, 3, 4, 5		hypothetical protein
Root935_gene_645	Root935	Root935_contig_1	131455	col0	at	6,42E-08	0,8	1, 2, 3, 4		hypothetical protein
Root935_gene_4250	Root935	Root935_contig_1	928419	col0	at	7,75E-08	1	1, 2, 3, 4, 5	K07386	pepO
Root935_gene_660	Root935	Root935_contig_1	133948	col0	at	8,80E-08	0,8	1, 2, 3, 4		hypothetical protein
Root935_gene_6235	Root935	Root935_contig_1	1383380	col0	at	1,31E-07	0,6	1, 3, 5		tRNA-Leu(tag)
Root265_gene_32215	Root265	Root265_contig_1	6617973	col0	at	1,34E-07	0,4	1, 4	K20074	prpC, phpP
Root935_gene_2460	Root935	Root935_contig_1	520629	col0	at	1,34E-07	0,8	1, 3, 4, 5	K05595	marC
Root1310_gene_20775	Root1310	Root1310_contig_1	4819608	col0	at	1,36E-07	0,6	1, 2, 5	K13653	K13653
Root935_gene_20255	Root935	Root935_contig_1	4712519	col0	at	1,58E-07	1	1, 1, 2, 2, 3, 3, 4, 4, 5, 5		hypothetical protein
Root935_gene_8110	Root935	Root935_contig_1	1820563	col0	at	2,28E-07	0,8	1, 2, 2, 3, 4, 4	K06911	K06911
Root935_gene_5110	Root935	Root935_contig_1	1128926	col0	at	2,58E-07	1	1, 1, 2, 3, 4, 5	K21449	ata, sadA, emaA
Root935_gene_11680	Root935	Root935_contig_1	2626072	col0	at	4,51E-07	0,8	2, 3, 4, 5	K07720	yesN
Root61_gene_18250	Root61	Root61_contig_1	3890678	col0	at	4,52E-07	0,6	1, 4, 5	K06148	ABCC-BAC
Root935_gene_2360	Root935	Root935_contig_1	497618	col0	at	6,47E-07	0,8	1, 3, 4, 5	K07568	queA
Root935_gene_9780	Root935	Root935_contig_1	2212326	col0	at	6,57E-07	0,8	2, 3, 4, 5	K09704	K09704
Root61_gene_17685	Root61	Root61_contig_1	3758589	col0	at	9,55E-07	0,6	1, 4, 5	K03695	clpB
Root935_gene_8125	Root935	Root935_contig_1	1823627	col0	at	9,55E-07	0,8	1, 2, 2, 3, 4, 4	K01638	aceB, glcB
Root480_gene_5575	Root480	Root480_contig_1	1277366	col0	at	1,93E-06	0,4	1, 3	K00786	E2.4.-.-
Root418_gene_21115	Root418	Root418_contig_1	4954758	col0	at	3,03E-06	0,4	1, 2		hypothetical protein
Root935_gene_20195	Root935	Root935_contig_1	4699741	col0	at	4,26E-06	0,6	1, 2, 5	K03718	asnC
Root83_gene_3385	Root83	Root83_contig_1	729930	col0	at	4,96E-06	0,4	4, 5	K01710	E4.2.1.46, rfbB, rffG
Root935_gene_2255	Root935	Root935_contig_1	472611	col0	at	6,43E-06	0,6	1, 2, 4		hypothetical protein
Root418_gene_6055	Root418	Root418_contig_1	1380318	col0	at	9,62E-06	0,4	1, 1, 4	K22010	pdtaR
Root935_gene_9870	Root935	Root935_contig_1	2237628	col0	at	1,56E-05	0,6	1, 2, 3	K12132	prkC, stkP
Root418_gene_18155	Root418	Root418_contig_1	4277658	col0	at	2,09E-05	0,4	1, 4	K21645	hypT, qseD
Root418_gene_1535	Root418	Root418_contig_1	342459	col0	at	2,66E-05	0,4	3, 4	K18297	mexT
Root935_gene_20185	Root935	Root935_contig_1	4696971	col0	at	2,66E-05	0,6	1, 2, 5	K00786	E2.4.-.-
Root418_gene_17110	Root418	Root418_contig_1	4027604	col0	at	2,78E-05	0,4	3, 5	K03477	ulaR
Root935_gene_16255	Root935	Root935_contig_1	3674931	col0	at	2,96E-05	0,6	1, 4, 5	K04084	dsbD
Root418_gene_23005	Root418	Root418_contig_1	5356367	col0	at	2,96E-05	0,4	1, 4	K03285	TC.GBP
Root935_gene_6245	Root935	Root935_contig_1	1385362	col0	at	4,00E-05	0,6	1, 3, 5		hypothetical protein
Root29_gene_2630	Root29	Root29_contig_1	580152	col0	at	4,28E-05	0,4	2, 2, 2, 2, 3, 3	K01251	E3.3.1.1, ahcY
Root935_gene_16235	Root935	Root935_contig_1	3670020	col0	at	4,99E-05	0,6	1, 4, 5	K01661	menB
Root935_gene_3555	Root935	Root935_contig_1	765070	col0	at	1,05E-04	0,6	1, 2, 3	K07683	nreB
Root935_gene_9860	Root935	Root935_contig_1	2233434	col0	at	1,38E-04	0,6	1, 1, 2, 3, 3	K15836	fhlA

Root935_gene_13985	Root935	Root935_contig_1	3166029	col0	at	1,67E-04	0,4	2, 5		hypothetical protein
Root935_gene_8035	Root935	Root935_contig_1	1808159	col0	at	1,78E-04	0,6	2, 4, 5	K03555	mutS
Root418_gene_19605	Root418	Root418_contig_1	4611907	col0	at	1,95E-04	0,4	1, 5	K12132	prkC, stkP
Root935_gene_17380	Root935	Root935_contig_1	3974718	col0	at	1,96E-04	0,4	1, 2		hypothetical protein
Root935_gene_13990	Root935	Root935_contig_1	3166471	col0	at	2,20E-04	0,4	2, 5		hypothetical protein
Root29_gene_4140	Root29	Root29_contig_1	880350	col0	at	2,26E-04	0,4	2, 2, 2, 3, 3, 3, 3	K03046	rpoC
Root935_gene_6660	Root935	Root935_contig_1	1481213	col0	at	2,98E-04	0,4	2, 3		hypothetical protein
Root935_gene_9875	Root935	Root935_contig_1	2238159	col0	at	4,77E-04	0,4	1, 3	K01252	entB, dhbB, vibB, mxcF
Root83_gene_23890	Root83	Root83_contig_1	5256008	col0	at	5,05E-04	0,4	1, 3	K15662	ituB, mycB, bmyB
Root935_gene_370	Root935	Root935_contig_1	72950	col0	at	5,17E-04	0,4	2, 3		hypothetical protein
Root935_gene_2195	Root935	Root935_contig_1	460016	col0	at	5,34E-04	0,4	1, 4	K03561	exbB
Root935_gene_19910	Root935	Root935_contig_1	4638367	col0	at	6,03E-04	0,6	2, 4, 5	K13611	pksJ
Root935_gene_10375	Root935	Root935_contig_1	2346829	col0	at	6,06E-04	0,4	1, 4		hypothetical protein
Root935_gene_6590	Root935	Root935_contig_1	1463972	col0	at	8,04E-04	0,4	2, 3	K01783	rpe, RPE
Root1310_gene_41130	Root1310	Root1310_contig_2	35951	col0	at	8,11E-04	0,6	1, 1, 2, 4	K01061	E3.1.1.45
Root935_gene_390	Root935	Root935_contig_1	76399	col0	at	8,56E-04	0,4	2, 3		hypothetical protein
Root1310_gene_41150	Root1310	Root1310_contig_2	41269	col0	at	9,69E-04	0,6	1, 1, 2, 4		hypothetical protein
Root935_gene_4845	Root935	Root935_contig_1	1077745	col0	at	1,04E-03	0,4	2, 4	K00721	DPM1
Root935_gene_6890	Root935	Root935_contig_1	1537640	col0	at	1,24E-03	0,4	2, 5	K10764	metZ
Root935_gene_19350	Root935	Root935_contig_1	4504871	col0	at	1,28E-03	0,4	4, 5	K02343	dnaX
Root935_gene_18825	Root935	Root935_contig_1	4357082	col0	at	1,90E-03	0,4	2, 3		hypothetical protein
Root935_gene_6845	Root935	Root935_contig_1	1524468	col0	at	2,21E-03	0,4	2, 5	K00789	metK
Root935_gene_18770	Root935	Root935_contig_1	4347065	col0	at	3,69E-03	0,4	2, 3	K01205	NAGLU
Root935_gene_770	Root935	Root935_contig_1	152408	col0	at	5,00E-03	0,4	1, 2, 2	K21573	susC
Root935_gene_1955	Root935	Root935_contig_1	406198	col0	at	7,07E-03	0,4	4, 4, 5, 5	K02014	TC.FEV.OM
Root935_gene_10370	Root935	Root935_contig_1	2343282	col0	at	8,23E-03	0,4	1, 4	K21573	susC
Root935_gene_17350	Root935	Root935_contig_1	3960805	col0	at	1,07E-02	0,4	1, 2	K21573	susC
LjRoot33_gene_19090	LjRoot33	LjRoot33_contig_1	4201337	col0	lj	9,91E-16	1	10, 6.2, 7.2, 8.2, 9	K02970	RP-S21, MRPS21, rpsU
LjRoot59_gene_510	LjRoot59	LjRoot59_contig_1	110800	col0	lj	2,68E-15	1	10, 6.2, 7.2, 8.2, 9		hypothetical protein
LjRoot262_gene_20675	LjRoot262	LjRoot262_contig_1	4469365	col0	lj	3,87E-14	1	10, 6.2, 7.2, 8.2, 9	K20971	ladS
LjRoot33_gene_2900	LjRoot33	LjRoot33_contig_1	662020	col0	lj	2,45E-13	1	10, 6.2, 7.2, 8.2, 9	K03285	TC.GBP
LjRoot33_gene_2905	LjRoot33	LjRoot33_contig_1	663633	col0	lj	3,31E-13	1	10, 6.2, 7.2, 8.2, 9	K18334	fucD
LjRoot82_gene_890	LjRoot82	LjRoot82_contig_1	175022	col0	lj	4,02E-13	1	6.2, 6.2, 7.2, 7.2, 8.2, 8.2,	K03642	rlpA
LjRoot60_gene_15760	LjRoot60	LjRoot60_contig_1	3353015	col0	lj	2,08E-12	1	10, 6.2, 7.2, 8.2, 9	K02663	pilN
LjRoot59_gene_480	LjRoot59	LjRoot59_contig_1	105264	col0	lj	2,18E-12	1	10, 6.2, 7.2, 8.2, 9		hypothetical protein
LjRoot82_gene_880	LjRoot82	LjRoot82_contig_1	173576	col0	lj	4,17E-12	1	6.2, 6.2, 7.2, 7.2, 8.2, 8.2,	K00384	trxB, TRR
LjRoot59_gene_15155	LjRoot59	LjRoot59_contig_1	3498801	col0	lj	5,08E-12	1	10, 6.2, 7.2, 8.2, 9	K01895	ACSS, acs
LjRoot60_gene_13125	LjRoot60	LjRoot60_contig_1	2794540	col0	lj	7,81E-12	1	10, 6.2, 7.2, 8.2, 9	K03284	corA
LjRoot60_gene_14340	LjRoot60	LjRoot60_contig_1	3055796	col0	lj	1,08E-11	1	10, 6.2, 7.2, 8.2, 9	K01589	purK
LjRoot60_gene_8205	LjRoot60	LjRoot60_contig_1	1738422	col0	lj	1,17E-11	1	10, 6.2, 7.2, 8.2, 9	K01915	glnA, GLUL
LjRoot33_gene_23015	LjRoot33	LjRoot33_contig_1	5028318	col0	lj	1,51E-11	1	10, 6.2, 7.2, 8.2, 9	K03043	rpoB
LjRoot82_gene_5505	LjRoot82	LjRoot82_contig_1	1185160	col0	lj	1,58E-11	0,8	10, 6.2, 8.2, 9		tRNA-Leu(caa)
LjRoot60_gene_10810	LjRoot60	LjRoot60_contig_1	2283746	col0	lj	2,62E-11	1	10, 6.2, 6.2, 7.2, 8.2, 9	K02416	fliM
LjRoot60_gene_15210	LjRoot60	LjRoot60_contig_1	3230051	col0	lj	4,47E-11	1	10, 6.2, 7.2, 8.2, 9	K02667	pilR, pehR
LjRoot59_gene_10585	LjRoot59	LjRoot59_contig_1	2404281	col0	lj	4,81E-11	0,8	6.2, 7.2, 8.2, 8.2, 9	K07689	uvrY, gacA, varA
LjRoot262_gene_17410	LjRoot262	LjRoot262_contig_1	3745419	col0	lj	8,66E-11	1	10, 6.2, 7.2, 8.2, 9	K02014	TC.FEV.OM
LjRoot262_gene_17440	LjRoot262	LjRoot262_contig_1	3755756	col0	lj	8,66E-11	1	10, 6.2, 7.2, 8.2, 9	K03320	amt, AMT, MEP
LjRoot262_gene_13355	LjRoot262	LjRoot262_contig_1	2884523	col0	lj	1,60E-10	1	10, 6.2, 7.2, 8.2, 9		16S ribosomal RNA
LjRoot262_gene_13360	LjRoot262	LjRoot262_contig_1	2888083	col0	lj	2,09E-10	1	10, 6.2, 7.2, 8.2, 9		hypothetical protein
LjRoot82_gene_5540	LjRoot82	LjRoot82_contig_1	1192723	col0	lj	2,94E-10	0,8	10, 6.2, 8.2, 9		hypothetical protein
LjRoot60_gene_15770	LjRoot60	LjRoot60_contig_1	3355051	col0	lj	3,43E-10	1	10, 6.2, 7.2, 8.2, 9	K05366	mrcA
LjRoot44_gene_15235	LjRoot44	LjRoot44_contig_1	3109147	col0	lj	1,26E-08	0,6	10, 7.2, 9	K05350	bglB
LjRoot262_gene_11030	LjRoot262	LjRoot262_contig_1	2428727	col0	lj	1,87E-08	0,8	6.2, 7.2, 8.2, 9	K02014	TC.FEV.OM
LjRoot44_gene_15270	LjRoot44	LjRoot44_contig_1	3117994	col0	lj	4,89E-08	0,6	10, 7.2, 9	K00031	IDH1, IDH2, icd
LjRoot72_gene_290	LjRoot72	LjRoot72_contig_1	63537	col0	lj	5,05E-08	0,6	10, 8.2, 9	K21023	mucR
LjRoot59_gene_22290	LjRoot59	LjRoot59_contig_1	5048722	col0	lj	9,55E-07	0,6	10, 6.2, 9, 9	K07678	barA, gacS, varS
LjRoot72_gene_11025	LjRoot72	LjRoot72_contig_1	2407198	col0	lj	1,21E-06	0,4	6.2, 7.2	K13590	dgcB
LjRoot60_gene_16945	LjRoot60	LjRoot60_contig_1	3600954	col0	lj	1,21E-06	0,6	10, 6.2, 9	K00404	ccoN
LjRoot60_gene_16940	LjRoot60	LjRoot60_contig_1	3599171	col0	lj	1,50E-06	0,6	10, 6.2, 9	K00425	cydA
LjRoot82_gene_10640	LjRoot82	LjRoot82_contig_1	2410707	col0	lj	2,20E-06	0,6	10, 6.2, 9	K00031	IDH1, IDH2, icd
LjRoot59_gene_19055	LjRoot59	LjRoot59_contig_1	4396299	col0	lj	2,35E-06	0,4	10, 6.2	K21089	mlrA
LjRoot82_gene_14760	LjRoot82	LjRoot82_contig_1	3330101	col0	lj	3,15E-06	0,4	6.2, 7.2		hypothetical protein
LjRoot59_gene_19470	LjRoot59	LjRoot59_contig_1	4486750	col0	lj	3,92E-06	0,4	10, 9		hypothetical protein
LjRoot82_gene_14735	LjRoot82	LjRoot82_contig_1	3325788	col0	lj	4,77E-06	0,4	6.2, 7.2		hypothetical protein
LjRoot44_gene_13065	LjRoot44	LjRoot44_contig_1	2633517	col0	lj	7,77E-06	0,4	10, 10, 9, 9	K03577	acrR, smeT
LjRoot59_gene_19460	LjRoot59	LjRoot59_contig_1	4484661	col0	lj	7,92E-06	0,4	10, 9		hypothetical protein
LjRoot59_gene_22035	LjRoot59	LjRoot59_contig_1	4997932	col0	lj	1,81E-05	0,4	10, 9	K03773	fklB
LjRoot59_gene_22070	LjRoot59	LjRoot59_contig_1	5002812	col0	lj	2,16E-05	0,4	10, 9	K03554	rdgC
LjRoot59_gene_24405	LjRoot59	LjRoot59_contig_1	5530014	col0	lj	2,49E-05	0,4	10, 10, 9	K13590	dgcB
LjRoot59_gene_24395	LjRoot59	LjRoot59_contig_1	5527672	col0	lj	2,89E-05	0,4	10, 10, 9	K00821	argD
LjRoot82_gene_18690	LjRoot82	LjRoot82_contig_1	4348268	col0	lj	6,03E-05	0,4	10, 6.2	K00800	aroA
LjRoot82_gene_18695	LjRoot82	LjRoot82_contig_1	4349444	col0	lj	8,58E-05	0,4	10, 6.2	K22757	QCT, qpcT
LjRoot59_gene_6600	LjRoot59	LjRoot59_contig_1	1495101	col0	lj	8,81E-05	0,4	10, 8.2	K21024	bifA
LjRoot262_gene_20665	LjRoot262	LjRoot262_contig_1	4465398	col0	lj	9,03E-05	0,4	10, 7.2, 7.2, 7.2	K21023	mucR
LjRoot46_gene_16205	LjRoot46	LjRoot46_contig_1	3361846	col0	lj	1,15E-04	0,4	10, 10, 9, 9	K21023	mucR
LjRoot60_gene_16050	LjRoot60	LjRoot60_contig_1	3408833	col0	lj	1,39E-04	0,4	6.2, 9	K02558	mpl
LjRoot82_gene_20970	LjRoot82	LjRoot82_contig_1	4902042	col0	lj	1,43E-04	0,4	8.2, 9	K11354	cph1
LjRoot262_gene_20670	LjRoot262	LjRoot262_contig_1	4466846	col0	lj	1,46E-04	0,4	8.2, 9	K20971	ladS
LjRoot46_gene_7540	LjRoot46	LjRoot46_contig_1	1565129	col0	lj	2,62E-04	0,4	, 10, 10, 10, 10, 10, 10,	K03046	rpoC
LjRoot262_gene_18735	LjRoot262	LjRoot262_contig_1	4040312	col0	lj	6,47E-03	0,4	7.2, 9	K01406	prtC
LjNodule218_gene_28430	LjNodule218	LjNodule218_contig_1	5796999	gifu	at	5,95E-23	1	1, 12, 12, 13, 13, 14, 14, 15, 15		hypothetical protein
Root29_gene_25385	Root29	Root29_contig_1	5667789	gifu	at	1,37E-14	1	11, 12, 13, 14, 15	K01784	galE, GALE
Root29_gene_25260	Root29	Root29_contig_1	5642802	gifu	at	1,37E-14	1	11, 12, 13, 14, 15	K03100	lepB
Root123D2_gene_11745	Root123D2	Root123D2_contig_1	2471397	gifu	at	3,99E-14	1	11, 12, 13, 14, 15	K12506	ispDF
Root123D2_gene_15050	Root123D2	Root123D2_contig_1	3121371	gifu	at	4,02E-13	1	11, 12, 13, 14, 15	K02014	TC.FEV.OM
Root83_gene_740	Root83	Root83_contig_1	163003	gifu	at	1,18E-12	0,8	, 13, 13, 13, 13, 13, 15, 15, 15, 15, 15, 15		tRNA-Asn(gtt)
Root935_gene_4275	Root935	Root935_contig_1	934313	gifu	at	9,88E-11	1	11, 12, 13, 14, 15		hypothetical protein
Root935_gene_7035	Root935	Root935_contig_1	1571453	gifu	at	1,07E-10	1	11, 11, 12, 13, 14, 14, 15	K11991	tadA
Root935_gene_7610	Root935	Root935_contig_1	1704272	gifu	at	1,44E-10	1	11, 12, 12, 13, 13, 14, 15		hypothetical protein
Root935_gene_5150	Root935	Root935_contig_1	1139487	gifu	at	2,54E-10	1	11, 11, 12, 12, 13, 14, 15		hypothetical protein
Root935_gene_20640	Root935	Root935_contig_1	4795347	gifu	at	4,60E-10	1	11, 11, 12, 12, 13, 14, 15	K00954	E2.7.7.3A, coaD, kdtB
Root935_gene_6235	Root935	Root935_contig_1	1383380	gifu	at	4,60E-10	0,8	12, 13, 14, 15		tRNA-Leu(tag)

Root935_gene_15180	Root935	Root935_contig_1	3448197	gifu	at	4,95E-10	1	.1, 11, 12, 12, 13, 14, 14, 1	K16137	nemR
Root935_gene_20235	Root935	Root935_contig_1	4707455	gifu	at	6,43E-10	1	.1, 12, 12, 12, 13, 13, 14, 1	K07095	K07095
Root68_gene_22920	Root68	Root68_contig_1	5104081	gifu	at	1,75E-09	0,6	11, 12, 15	K21024	bifA
Root935_gene_7010	Root935	Root935_contig_1	1566832	gifu	at	2,51E-09	1	11, 11, 12, 13, 14, 14, 15		hypothetical protein
Root935_gene_15100	Root935	Root935_contig_1	3425034	gifu	at	2,75E-09	1	.1, 11, 12, 12, 13, 14, 14, 1	K12243	pchR
Root935_gene_7685	Root935	Root935_contig_1	1732119	gifu	at	9,14E-09	1	11, 12, 12, 13, 13, 14, 15	K07182	K07182
Root935_gene_9780	Root935	Root935_contig_1	2212326	gifu	at	1,24E-08	1	11, 12, 13, 14, 15	K09704	K09704
Root935_gene_20585	Root935	Root935_contig_1	4783665	gifu	at	1,54E-08	1	11, 11, 12, 12, 13, 14, 15		hypothetical protein
Root935_gene_16795	Root935	Root935_contig_1	3789670	gifu	at	1,97E-08	0,8	11, 11, 12, 12, 13, 14		hypothetical protein
Root83_gene_12605	Root83	Root83_contig_1	2742935	gifu	at	2,19E-08	0,6	11, 13, 15		hypothetical protein
Root418_gene_1535	Root418	Root418_contig_1	342459	gifu	at	2,23E-08	0,6	11, 12, 13	K18297	mexT
Root935_gene_5835	Root935	Root935_contig_1	1296518	gifu	at	4,18E-08	1	, 12, 12, 13, 13, 14, 14, 15, 15, 15, 15, 15		hypothetical protein
Root418_gene_27355	Root418	Root418_contig_1	6319508	gifu	at	4,51E-08	0,6	11, 12, 13	K05962	E2.7.13.1
Root935_gene_11660	Root935	Root935_contig_1	2622677	gifu	at	6,52E-08	0,8	11, 12, 14, 15		hypothetical protein
Root935_gene_5460	Root935	Root935_contig_1	1217195	gifu	at	6,88E-08	0,8	11, 12, 14, 14, 15		hypothetical protein
Root935_gene_4250	Root935	Root935_contig_1	928419	gifu	at	8,51E-08	1	11, 12, 12, 13, 14, 15	K07386	pepO
Root418_gene_26825	Root418	Root418_contig_1	6179549	gifu	at	8,87E-08	0,6	11, 14, 15	K03406	mcp
Root935_gene_8035	Root935	Root935_contig_1	1808159	gifu	at	1,56E-07	1	11, 12, 13, 14, 15	K03555	mutS
Root935_gene_20255	Root935	Root935_contig_1	4712519	gifu	at	1,74E-07	1	.1, 12, 12, 13, 13, 14, 14, 15, 15		hypothetical protein
Root935_gene_5110	Root935	Root935_contig_1	1128926	gifu	at	2,84E-07	1	11, 11, 12, 12, 13, 14, 15	K21449	ata, sadA, emaA
Root935_gene_5475	Root935	Root935_contig_1	1219606	gifu	at	3,65E-07	0,8	11, 12, 14, 14, 15	K07004	K07004
Root83_gene_9680	Root83	Root83_contig_1	2111763	gifu	at	3,73E-07	0,6	12, 14, 15	K10125	dctB
Root935_gene_11680	Root935	Root935_contig_1	2626072	gifu	at	4,82E-07	0,8	11, 11, 12, 14, 14, 15	K07720	yesN
Root935_gene_6245	Root935	Root935_contig_1	1385362	gifu	at	7,95E-07	0,8	12, 13, 14, 15		hypothetical protein
Root935_gene_16810	Root935	Root935_contig_1	3793597	gifu	at	1,12E-06	0,8	11, 11, 12, 12, 13, 14	K01193	INV, sacA
Root935_gene_19910	Root935	Root935_contig_1	4638367	gifu	at	1,17E-06	1	11, 12, 13, 14, 15	K13611	pksJ
Root418_gene_21770	Root418	Root418_contig_1	5090865	gifu	at	3,03E-06	0,4	12, 13	K03624	greA
Root480_gene_5575	Root480	Root480_contig_1	1277366	gifu	at	3,47E-06	0,4	11, 12	K00786	E2.4.-.-
Root935_gene_645	Root935	Root935_contig_1	131455	gifu	at	6,56E-06	0,6	11, 12, 12, 13		hypothetical protein
Root935_gene_660	Root935	Root935_contig_1	133948	gifu	at	8,43E-06	0,6	11, 12, 12, 13		hypothetical protein
Root83_gene_3385	Root83	Root83_contig_1	729930	gifu	at	2,55E-05	0,4	11, 13	K01710	E4.2.1.46, rfbB, rffG
Root935_gene_21490	Root935	Root935_contig_1	5001977	gifu	at	3,07E-05	0,6	11, 12, 13	K01218	gmuG
Root935_gene_13870	Root935	Root935_contig_1	3141457	gifu	at	3,74E-05	0,6	11, 12, 15	K21023	mucR
Root935_gene_21425	Root935	Root935_contig_1	4983093	gifu	at	4,77E-05	0,6	11, 12, 13	K01881	PARS, proS
LjNodule218_gene_13790	LjNodule218	LjNodule218_contig_1	2891344	gifu	at	6,09E-05	0,2	15	K00001	E1.1.1.1, adh
Root480_gene_11095	Root480	Root480_contig_1	2454939	gifu	at	8,48E-05	0,4	14, 15	K02674	pilY1
Root83_gene_19085	Root83	Root83_contig_1	4193334	gifu	at	8,81E-05	0,4	13, 15	K01810	GPI, pgi
LjNodule218_gene_690	LjNodule218	LjNodule218_contig_1	143166	gifu	at	1,25E-04	0,2	15	K12368	dppA
Root935_gene_9860	Root935	Root935_contig_1	2233434	gifu	at	1,44E-04	0,6	11, 11, 12, 15	K15836	fhlA
LjNodule218_gene_695	LjNodule218	LjNodule218_contig_1	145234	gifu	at	1,74E-04	0,2	15	K01897	ACSL, fadD
Root83_gene_9685	Root83	Root83_contig_1	2113295	gifu	at	1,82E-04	0,4	11, 15	K18765	csrD
Root935_gene_9875	Root935	Root935_contig_1	2238159	gifu	at	4,94E-04	0,4	11, 12	K01252	entB, dhbB, vibB, mxcF
Root935_gene_11675	Root935	Root935_contig_1	2625257	gifu	at	5,42E-04	0,4	11, 14		hypothetical protein
Root935_gene_20360	Root935	Root935_contig_1	4736386	gifu	at	5,51E-04	0,4	11, 12		hypothetical protein
Root935_gene_4235	Root935	Root935_contig_1	925875	gifu	at	6,65E-04	0,4	11, 15		hypothetical protein
Root935_gene_8630	Root935	Root935_contig_1	1926816	gifu	at	8,41E-04	0,4	11, 15		hypothetical protein
Root935_gene_9870	Root935	Root935_contig_1	2237628	gifu	at	8,83E-04	0,4	11, 15	K12132	prkC, stkP
Root935_gene_8110	Root935	Root935_contig_1	1820563	gifu	at	9,35E-04	0,4	12, 12, 13	K06911	K06911
Root935_gene_4240	Root935	Root935_contig_1	926787	gifu	at	1,23E-03	0,4	11, 15	K01698	hemB, ALAD
Root935_gene_16255	Root935	Root935_contig_1	3674931	gifu	at	1,37E-03	0,4	11, 12	K04084	dsbD
Root935_gene_2425	Root935	Root935_contig_1	512770	gifu	at	1,87E-03	0,4	11, 14		hypothetical protein
Root935_gene_8125	Root935	Root935_contig_1	1823627	gifu	at	1,94E-03	0,4	12, 12, 13	K01638	aceB, glcB
Root935_gene_16235	Root935	Root935_contig_1	3670020	gifu	at	1,98E-03	0,4	11, 12	K01661	menB
Root935_gene_8800	Root935	Root935_contig_1	1958966	gifu	at	2,14E-03	0,6	11, 14, 15		hypothetical protein
Root935_gene_890	Root935	Root935_contig_1	191182	gifu	at	1,01E-02	0,4	11, 12	K21573	susC
Root935_gene_14485	Root935	Root935_contig_1	3274704	gifu	at	3,17E-02	0,4	11, 15	K11904	vgrG
LjNodule218_gene_28430	LjNodule218	LjNodule218_contig_1	5796999	gifu	lj	5,95E-23	0,8	.6, 16, 17, 17, 18, 18, 20, 20		hypothetical protein
LjRoot33_gene_19090	LjRoot33	LjRoot33_contig_1	4201337	gifu	lj	4,36E-14	1	16, 17, 18, 19, 20	K02970	RP-S21, MRPS21, rpsU
LjRoot72_gene_20060	LjRoot72	LjRoot72_contig_1	4410772	gifu	lj	2,79E-13	0,8	16, 17, 18, 20	K02408	fliE
LjRoot72_gene_290	LjRoot72	LjRoot72_contig_1	63537	gifu	lj	2,96E-12	1	16, 17, 18, 19, 20	K21023	mucR
LjRoot33_gene_2900	LjRoot33	LjRoot33_contig_1	662020	gifu	lj	1,98E-11	1	16, 17, 18, 19, 20	K03285	TC.GBP
LjRoot72_gene_20055	LjRoot72	LjRoot72_contig_1	4408902	gifu	lj	2,79E-11	0,8	16, 17, 18, 20	K02409	fliF
LjRoot33_gene_2905	LjRoot33	LjRoot33_contig_1	663633	gifu	lj	3,13E-11	1	16, 17, 18, 19, 20	K18334	fucD
LjRoot262_gene_17410	LjRoot262	LjRoot262_contig_1	3745419	gifu	lj	3,62E-11	1	16, 17, 18, 19, 20	K02014	TC.FEV.OM
LjRoot262_gene_17440	LjRoot262	LjRoot262_contig_1	3755756	gifu	lj	3,62E-11	1	16, 17, 18, 19, 20	K03320	amt, AMT, MEP
LjRoot59_gene_10585	LjRoot59	LjRoot59_contig_1	2404281	gifu	lj	8,66E-11	0,8	17, 18, 19, 20	K07689	uvrY, gacA, varA
LjRoot262_gene_20665	LjRoot262	LjRoot262_contig_1	4465398	gifu	lj	1,13E-09	0,8	16, 18, 19, 19, 20, 20	K21023	mucR
LjRoot33_gene_23015	LjRoot33	LjRoot33_contig_1	5028318	gifu	lj	1,43E-09	1	16, 17, 18, 19, 20	K03043	rpoB
LjRoot59_gene_510	LjRoot59	LjRoot59_contig_1	110800	gifu	lj	2,25E-09	0,6	16, 17, 20		hypothetical protein
LjRoot59_gene_13620	LjRoot59	LjRoot59_contig_1	3159779	gifu	lj	6,06E-09	0,8	16, 17, 18, 18, 20	K20975	K20975
LjRoot262_gene_13355	LjRoot262	LjRoot262_contig_1	2884523	gifu	lj	1,16E-08	0,8	16, 17, 19, 20		16S ribosomal RNA
LjRoot262_gene_13360	LjRoot262	LjRoot262_contig_1	2888083	gifu	lj	1,40E-08	0,8	16, 17, 19, 20		hypothetical protein
LjNodule218_gene_25505	LjNodule218	LjNodule218_contig_1	5215623	gifu	lj	4,44E-08	0,6	.6, 17, 17, 17, 17, 17, 18, 1	K08234	yaeR
LjRoot59_gene_10905	LjRoot59	LjRoot59_contig_1	2485308	gifu	lj	1,31E-07	0,6	16, 18, 19	K14657	nodD
LjRoot33_gene_14235	LjRoot33	LjRoot33_contig_1	3136198	gifu	lj	1,90E-07	0,6	16, 18, 20	K02556	motA
LjRoot59_gene_480	LjRoot59	LjRoot59_contig_1	105264	gifu	lj	2,44E-07	0,6	16, 17, 20		hypothetical protein
LjRoot59_gene_15155	LjRoot59	LjRoot59_contig_1	3498801	gifu	lj	4,41E-07	0,6	17, 19, 20	K01895	ACSS, acs
LjRoot59_gene_13775	LjRoot59	LjRoot59_contig_1	3195024	gifu	lj	5,23E-07	0,6	17, 18, 19	K02584	nifA
LjRoot262_gene_20670	LjRoot262	LjRoot262_contig_1	4466846	gifu	lj	5,83E-07	0,6	16, 17, 17, 18	K20971	ladS
LjRoot33_gene_11725	LjRoot33	LjRoot33_contig_1	2581544	gifu	lj	2,68E-05	0,4	17, 18, 18		hypothetical protein
LjRoot59_gene_24405	LjRoot59	LjRoot59_contig_1	5530014	gifu	lj	3,27E-05	0,4	16, 17	K13590	dgcB
LjRoot59_gene_24395	LjRoot59	LjRoot59_contig_1	5527672	gifu	lj	3,88E-05	0,4	16, 17	K00821	argD
LjRoot33_gene_19375	LjRoot33	LjRoot33_contig_1	4269610	gifu	lj	4,00E-05	0,4	17, 18	K14682	argAB
LjNodule218_gene_13790	LjNodule218	LjNodule218_contig_1	2891344	gifu	lj	6,09E-05	0,2	20	K00001	E1.1.1.1, adh
LjRoot33_gene_11730	LjRoot33	LjRoot33_contig_1	2582295	gifu	lj	6,23E-05	0,4	17, 18, 18	K07467	rstA1
LjRoot59_gene_10925	LjRoot59	LjRoot59_contig_1	2489293	gifu	lj	7,43E-05	0,4	18, 19		hypothetical protein
LjNodule218_gene_690	LjNodule218	LjNodule218_contig_1	143166	gifu	lj	1,25E-04	0,2	16	K12368	dppA
LjRoot33_gene_19695	LjRoot33	LjRoot33_contig_1	4327744	gifu	lj	1,66E-04	0,4	16, 20	K14393	actP
LjNodule218_gene_695	LjNodule218	LjNodule218_contig_1	145234	gifu	lj	1,74E-04	0,2	16	K01897	ACSL, fadD
LjRoot59_gene_22290	LjRoot59	LjRoot59_contig_1	5048722	gifu	lj	1,89E-04	0,4	16, 17, 17	K07678	barA, gacS, varS
LjRoot33_gene_19370	LjRoot33	LjRoot33_contig_1	4267312	gifu	lj	2,04E-04	0,4	17, 18	K03406	mcp

Table 3. Genetic variants significantly enriched in SynCom populations evolved on *Arabidopsis* or *Lotus* roots. The table lists the variant ID, strain, genomic position, evolution host plant, and corresponding *P*-value (n = 372). Enrichment was assessed using generalized linear models (GLMs), and *P*-values were corrected for multiple testing by controlling the false discovery rate at (α = 0.05). KEGG ontologies are provided for mutations that affect coding regions.

ID	chrom	plant	syncom	padj	type	ko
Root1310_contig_1_409069_T_TC	Root1310_contig_1	col0	at	1,61E-04	INTERGENIC	
Root1310_contig_1_4630356_CGG_C	Root1310_contig_1	col0	at	2,06E-02	INTERGENIC	
Root1310_contig_1_4630356_CG_C	Root1310_contig_1	col0	at	4,32E-04	INTERGENIC	
Root1310_contig_1_6475531_G_A	Root1310_contig_1	col0	at	4,53E-10	NON_SYNONYMOUS_CODING	K02355
Root1310_contig_1_8125810_G_C	Root1310_contig_1	col0	at	2,26E-43	SYNONYMOUS_CODING	K00010
Root1310_contig_2_138462_C_CG	Root1310_contig_2	col0	at	1,93E-76	FRAME_SHIFT	
Root1310_contig_2_172761_G_GC	Root1310_contig_2	col0	at	9,05E-03	INTERGENIC	
Root1310_contig_2_39566_CGG_C	Root1310_contig_2	col0	at	1,69E-02	INTERGENIC	
Root265_contig_1_6620217_A_C	Root265_contig_1	col0	at	4,64E-02	INTERGENIC	
Root29_contig_1_135577_C_T	Root29_contig_1	col0	at	1,55E-02	SYNONYMOUS_CODING	K02058
Root29_contig_1_135601_T_C	Root29_contig_1	col0	at	1,53E-02	SYNONYMOUS_CODING	K02058
Root29_contig_1_135616_A_G	Root29_contig_1	col0	at	4,67E-02	SYNONYMOUS_CODING	K02058
Root29_contig_1_5717333_C_G	Root29_contig_1	col0	at	3,79E-02	SYNONYMOUS_CODING	K02274
Root29_contig_1_5717358_C_G	Root29_contig_1	col0	at	1,18E-02	NON_SYNONYMOUS_CODING	K02274
Root29_contig_1_5717360_G_C	Root29_contig_1	col0	at	1,18E-02	SYNONYMOUS_CODING	K02274
Root29_contig_1_580946_C_T	Root29_contig_1	col0	at	5,54E-03	SYNONYMOUS_CODING	K01251
Root29_contig_1_580976_C_A	Root29_contig_1	col0	at	2,69E-03	SYNONYMOUS_CODING	K01251
Root29_contig_1_580985_G_A	Root29_contig_1	col0	at	1,88E-02	SYNONYMOUS_CODING	K01251
Root418_contig_1_1274670_C_T	Root418_contig_1	col0	at	5,72E-03	SYNONYMOUS_CODING	K08483
Root418_contig_1_127474_G_T	Root418_contig_1	col0	at	1,17E-02	STOP_GAINED	K21023
Root418_contig_1_1380568_T_C	Root418_contig_1	col0	at	6,74E-04	NON_SYNONYMOUS_CODING	K22010
Root418_contig_1_1380807_C_G	Root418_contig_1	col0	at	4,04E-05	NON_SYNONYMOUS_CODING	K22010
Root418_contig_1_192314_A_C	Root418_contig_1	col0	at	4,69E-03	NON_SYNONYMOUS_CODING	K02483
Root418_contig_1_192540_C_A	Root418_contig_1	col0	at	3,78E-03	NON_SYNONYMOUS_CODING	K02483
Root418_contig_1_192558_G_A	Root418_contig_1	col0	at	1,09E-02	STOP_GAINED	K02483
Root418_contig_1_193146_T_C	Root418_contig_1	col0	at	1,09E-02	NON_SYNONYMOUS_CODING	K07652
Root418_contig_1_342586_G_A	Root418_contig_1	col0	at	2,22E-02	NON_SYNONYMOUS_CODING	K18297
Root418_contig_1_342592_G_A	Root418_contig_1	col0	at	4,76E-03	NON_SYNONYMOUS_CODING	K18297
Root418_contig_1_345326_G_T	Root418_contig_1	col0	at	1,25E-02	STOP_GAINED	K03719
Root418_contig_1_345607_C_T	Root418_contig_1	col0	at	1,67E-04	NON_SYNONYMOUS_CODING	K03719
Root418_contig_1_4329085_A_T	Root418_contig_1	col0	at	5,89E-04	INTERGENIC	
Root418_contig_1_4374095_G_A	Root418_contig_1	col0	at	1,18E-02	NON_SYNONYMOUS_CODING	K18297
Root418_contig_1_4954676_A_G	Root418_contig_1	col0	at	1,14E-03	INTERGENIC	
Root418_contig_1_4954759_T_C	Root418_contig_1	col0	at	1,64E-02	START_LOST	
Root418_contig_1_5316220_T_TCGGCCA	Root418_contig_1	col0	at	3,04E-02	CODON_INSERTION	K21023
Root418_contig_1_5356656_G_A	Root418_contig_1	col0	at	6,20E-05	NON_SYNONYMOUS_CODING	K03285
Root418_contig_1_6320055_GCGAGCGCCAGGTGAT_G	Root418_contig_1	col0	at	4,05E-08	ON_CHANGE_PLUS_CODON_DELE	K05962
Root418_contig_1_6320471_CA_C	Root418_contig_1	col0	at	4,72E-03	FRAME_SHIFT	K05962
Root418_contig_1_742612_G_A	Root418_contig_1	col0	at	3,48E-02	INTERGENIC	
Root480_contig_1_2453978_C_T	Root480_contig_1	col0	at	1,49E-02	NON_SYNONYMOUS_CODING	K08084
Root480_contig_1_2478176_C_T	Root480_contig_1	col0	at	3,04E-02	NON_SYNONYMOUS_CODING	
Root480_contig_1_526338_CG_C	Root480_contig_1	col0	at	7,66E-04	FRAME_SHIFT	K02664
Root480_contig_1_528054_T_C	Root480_contig_1	col0	at	2,29E-02	NON_SYNONYMOUS_CODING	K02666
Root61_contig_1_1554757_G_A	Root61_contig_1	col0	at	7,18E-15	NON_SYNONYMOUS_CODING	
Root61_contig_1_1554759_C_A	Root61_contig_1	col0	at	2,84E-15	SYNONYMOUS_CODING	
Root61_contig_1_1957227_G_T	Root61_contig_1	col0	at	2,77E-02	INTERGENIC	
Root61_contig_1_2742851_G_T	Root61_contig_1	col0	at	3,66E-14	INTERGENIC	
Root61_contig_1_2742855_C_G	Root61_contig_1	col0	at	3,70E-07	INTERGENIC	
Root61_contig_1_3760068_G_A	Root61_contig_1	col0	at	1,17E-15	SYNONYMOUS_CODING	K03695
Root61_contig_1_3891187_T_TC	Root61_contig_1	col0	at	7,18E-15	FRAME_SHIFT	K06148
Root61_contig_1_729846_AGC_A	Root61_contig_1	col0	at	4,60E-03	FRAME_SHIFT	K09017
Root685_contig_1_2544770_A_C	Root685_contig_1	col0	at	3,22E-03	INTERGENIC	
Root68_contig_1_384095_C_CA	Root68_contig_1	col0	at	1,37E-02	INTERGENIC	
Root68_contig_1_4317644_GT_G	Root68_contig_1	col0	at	4,36E-13	FRAME_SHIFT	
Root68_contig_1_5105855_G_A	Root68_contig_1	col0	at	1,79E-02	NON_SYNONYMOUS_CODING	K21024
Root68_contig_1_5105910_TGAC_T	Root68_contig_1	col0	at	3,82E-12	ON_CHANGE_PLUS_CODON_DELE	K21024
Root68_contig_1_5455787_T_C	Root68_contig_1	col0	at	1,55E-02	SYNONYMOUS_CODING	K03979
Root68_contig_1_780670_G_T	Root68_contig_1	col0	at	2,34E-02	NON_SYNONYMOUS_CODING	K21699
Root83_contig_1_1057795_T_C	Root83_contig_1	col0	at	1,53E-02	NON_SYNONYMOUS_CODING	K16048
Root83_contig_1_247123_C_T	Root83_contig_1	col0	at	1,44E-05	SYNONYMOUS_CODING	K07154
Root83_contig_1_5256123_GC_G	Root83_contig_1	col0	at	4,13E-02	FRAME_SHIFT	K15662
Root83_contig_1_534289_C_T	Root83_contig_1	col0	at	1,53E-02	NON_SYNONYMOUS_CODING	K13243
Root83_contig_1_5714254_G_GC	Root83_contig_1	col0	at	7,25E-03	INTRAGENIC	
Root83_contig_1_682313_C_T	Root83_contig_1	col0	at	4,72E-03	NON_SYNONYMOUS_CODING	K03230
Root935_contig_1_1078167_CT_C	Root935_contig_1	col0	at	5,72E-03	FRAME_SHIFT	K00721
Root935_contig_1_1078167_C_CT	Root935_contig_1	col0	at	8,49E-03	FRAME_SHIFT	K00721
Root935_contig_1_1217856_CT_C	Root935_contig_1	col0	at	4,73E-09	INTERGENIC	
Root935_contig_1_1297331_A_T	Root935_contig_1	col0	at	1,26E-02	STOP_GAINED	

Root935_contig_1_1297518_C_CA	Root935_contig_1	col0	at	2,00E-03	FRAME_SHIFT	
Root935_contig_1_1297768_C_CA	Root935_contig_1	col0	at	3,74E-03	FRAME_SHIFT	
Root935_contig_1_1297962_AT_A	Root935_contig_1	col0	at	9,22E-03	FRAME_SHIFT	
Root935_contig_1_1298222_G_A	Root935_contig_1	col0	at	4,32E-04	NON_SYNONYMOUS_CODING	
Root935_contig_1_1316988_A_G	Root935_contig_1	col0	at	1,42E-03	NON_SYNONYMOUS_CODING	K00059
Root935_contig_1_132172_T_TA	Root935_contig_1	col0	at	1,29E-04	INTERGENIC	
Root935_contig_1_1401964_T_G	Root935_contig_1	col0	at	4,58E-02	NON_SYNONYMOUS_CODING	K07027
Root935_contig_1_1454665_A_G	Root935_contig_1	col0	at	3,42E-03	INTERGENIC	
Root935_contig_1_1464651_G_GT	Root935_contig_1	col0	at	1,06E-08	INTERGENIC	
Root935_contig_1_152722_CA_C	Root935_contig_1	col0	at	2,64E-03	FRAME_SHIFT	K21573
Root935_contig_1_152722_C_CA	Root935_contig_1	col0	at	5,13E-08	FRAME_SHIFT	K21573
Root935_contig_1_1530432_C_T	Root935_contig_1	col0	at	1,61E-02	SYNONYMOUS_CODING	K01740
Root935_contig_1_1537615_C_CA	Root935_contig_1	col0	at	2,29E-07	INTERGENIC	
Root935_contig_1_1611700_C_T	Root935_contig_1	col0	at	2,88E-02	NON_SYNONYMOUS_CODING	K11912
Root935_contig_1_1731437_AT_A	Root935_contig_1	col0	at	5,19E-27	INTERGENIC	
Root935_contig_1_1749195_T_C	Root935_contig_1	col0	at	1,53E-02	NON_SYNONYMOUS_CODING	K02276
Root935_contig_1_1785807_T_TA	Root935_contig_1	col0	at	4,81E-04	INTERGENIC	
Root935_contig_1_1821643_TA_T	Root935_contig_1	col0	at	3,01E-02	INTERGENIC	
Root935_contig_1_2127209_C_T	Root935_contig_1	col0	at	8,12E-03	NON_SYNONYMOUS_CODING	K01235
Root935_contig_1_2346790_G_A	Root935_contig_1	col0	at	7,64E-03	INTERGENIC	
Root935_contig_1_2596130_A_AT	Root935_contig_1	col0	at	2,29E-02	FRAME_SHIFT	K17734
Root935_contig_1_269954_T_C	Root935_contig_1	col0	at	3,43E-03	SYNONYMOUS_CODING	K02014
Root935_contig_1_2853309_C_T	Root935_contig_1	col0	at	4,13E-02	SYNONYMOUS_CODING	
Root935_contig_1_3165168_T_A	Root935_contig_1	col0	at	2,99E-03	NON_SYNONYMOUS_CODING	K08995
Root935_contig_1_3166289_A_AT	Root935_contig_1	col0	at	2,92E-07	INTERGENIC	
Root935_contig_1_3181438_G_A	Root935_contig_1	col0	at	1,90E-02	NON_SYNONYMOUS_CODING	K03407
Root935_contig_1_3258209_G_GT	Root935_contig_1	col0	at	3,63E-04	FRAME_SHIFT	
Root935_contig_1_3310667_C_T	Root935_contig_1	col0	at	2,84E-02	NON_SYNONYMOUS_CODING	K02014
Root935_contig_1_3322174_G_GT	Root935_contig_1	col0	at	5,43E-04	INTERGENIC	
Root935_contig_1_3821960_C_T	Root935_contig_1	col0	at	4,69E-03	SYNONYMOUS_CODING	K03657
Root935_contig_1_3824103_A_AT	Root935_contig_1	col0	at	1,70E-02	INTERGENIC	
Root935_contig_1_3964735_A_AT	Root935_contig_1	col0	at	4,21E-06	INTERGENIC	
Root935_contig_1_406899_A_AG	Root935_contig_1	col0	at	1,01E-03	FRAME_SHIFT	K02014
Root935_contig_1_408008_GA_G	Root935_contig_1	col0	at	4,06E-05	FRAME_SHIFT	K02014
Root935_contig_1_4208505_A_C	Root935_contig_1	col0	at	2,22E-02	NON_SYNONYMOUS_CODING	K00227
Root935_contig_1_4374759_G_GT	Root935_contig_1	col0	at	9,22E-03	INTERGENIC	
Root935_contig_1_4505930_G_GA	Root935_contig_1	col0	at	1,55E-02	FRAME_SHIFT	K02343
Root935_contig_1_4698200_T_TA	Root935_contig_1	col0	at	8,14E-04	INTERGENIC	
Root935_contig_1_4900559_G_GA	Root935_contig_1	col0	at	2,58E-03	INTERGENIC	
Root935_contig_1_4981665_GA_G	Root935_contig_1	col0	at	1,05E-02	FRAME_SHIFT	K11354
Root935_contig_1_4981952_G_A	Root935_contig_1	col0	at	4,97E-02	NON_SYNONYMOUS_CODING	K02490
Root935_contig_1_498825_G_GT	Root935_contig_1	col0	at	3,55E-13	INTERGENIC	
Root935_contig_1_4990575_CA_C	Root935_contig_1	col0	at	8,24E-03	INTERGENIC	
Root935_contig_1_4990575_C_CA	Root935_contig_1	col0	at	7,18E-15	INTERGENIC	
Root935_contig_1_5125176_A_AT	Root935_contig_1	col0	at	9,22E-03	INTERGENIC	
Root935_contig_1_54389_T_G	Root935_contig_1	col0	at	1,64E-02	INTERGENIC	
Root935_contig_1_674876_T_C	Root935_contig_1	col0	at	4,99E-02	NON_SYNONYMOUS_CODING	
Root935_contig_1_697897_T_A	Root935_contig_1	col0	at	1,81E-02	NON_SYNONYMOUS_CODING	K00457
Root935_contig_1_73621_GA_G	Root935_contig_1	col0	at	3,30E-03	INTERGENIC	
Root935_contig_1_73621_G_GA	Root935_contig_1	col0	at	4,68E-03	INTERGENIC	
Root935_contig_1_764462_GT_G	Root935_contig_1	col0	at	1,25E-02	FRAME_SHIFT	K02479
Root935_contig_1_765821_T_TA	Root935_contig_1	col0	at	3,85E-03	FRAME_SHIFT	K07683
Root935_contig_1_766617_C_CA	Root935_contig_1	col0	at	8,81E-03	FRAME_SHIFT	K07683
Root935_contig_1_766963_A_T	Root935_contig_1	col0	at	9,22E-03	STOP_GAINED	K07683
Root935_contig_1_973800_G_GT	Root935_contig_1	col0	at	3,64E-02	INTERGENIC	
LjRoot262_contig_1_1862391_GGC_G	LjRoot262_contig_1	col0	lj	7,41E-04	INTERGENIC	
LjRoot262_contig_1_2430990_C_T	LjRoot262_contig_1	col0	lj	5,42E-04	NON_SYNONYMOUS_CODING	K02014
LjRoot262_contig_1_4047593_GGC_G	LjRoot262_contig_1	col0	lj	3,95E-02	FRAME_SHIFT	K01406
LjRoot262_contig_1_4466588_C_A	LjRoot262_contig_1	col0	lj	4,38E-03	NON_SYNONYMOUS_CODING	K21023
LjRoot262_contig_1_4469648_GCC_G	LjRoot262_contig_1	col0	lj	5,97E-12	FRAME_SHIFT	K20971
LjRoot44_contig_1_1626806_G_A	LjRoot44_contig_1	col0	lj	4,38E-03	SYNONYMOUS_CODING	K00130
LjRoot44_contig_1_2633782_A_C	LjRoot44_contig_1	col0	lj	3,46E-03	NON_SYNONYMOUS_CODING	K03577
LjRoot44_contig_1_2633785_A_G	LjRoot44_contig_1	col0	lj	1,47E-03	NON_SYNONYMOUS_CODING	K03577
LjRoot44_contig_1_3110344_TCG_T	LjRoot44_contig_1	col0	lj	4,68E-10	INTERGENIC	
LjRoot46_contig_1_1567249_C_T	LjRoot46_contig_1	col0	lj	2,10E-04	SYNONYMOUS_CODING	K03046
LjRoot46_contig_1_3361966_G_GT	LjRoot46_contig_1	col0	lj	4,86E-02	FRAME_SHIFT	K21023
LjRoot46_contig_1_3362048_G_C	LjRoot46_contig_1	col0	lj	1,52E-04	STOP_GAINED	K21023
LjRoot59_contig_1_1495960_G_A	LjRoot59_contig_1	col0	lj	1,28E-02	NON_SYNONYMOUS_CODING	K21024
LjRoot59_contig_1_2404599_G_T	LjRoot59_contig_1	col0	lj	2,81E-02	NON_SYNONYMOUS_CODING	K07689
LjRoot59_contig_1_2404816_C_T	LjRoot59_contig_1	col0	lj	8,39E-03	NON_SYNONYMOUS_CODING	K07689
LjRoot59_contig_1_2539572_T_TG	LjRoot59_contig_1	col0	lj	7,04E-03	INTERGENIC	
LjRoot59_contig_1_5002046_C_CA	LjRoot59_contig_1	col0	lj	2,10E-04	INTERGENIC	
LjRoot59_contig_1_5050547_T_C	LjRoot59_contig_1	col0	lj	1,18E-02	NON_SYNONYMOUS_CODING	K07678
LjRoot59_contig_1_5051315_AC_A	LjRoot59_contig_1	col0	lj	9,59E-03	FRAME_SHIFT	K07678
LjRoot59_contig_1_5051344_C_T	LjRoot59_contig_1	col0	lj	4,55E-03	STOP_GAINED	K07678
LjRoot59_contig_1_5416778_G_GT	LjRoot59_contig_1	col0	lj	1,12E-02	INTERGENIC	
LjRoot59_contig_1_5529884_A_C	LjRoot59_contig_1	col0	lj	6,68E-03	INTERGENIC	
LjRoot60_contig_1_1739115_T_A	LjRoot60_contig_1	col0	lj	5,53E-96	NON_SYNONYMOUS_CODING	K01915
LjRoot60_contig_1_2201884_G_T	LjRoot60_contig_1	col0	lj	1,67E-02	NON_SYNONYMOUS_CODING	K10536
LjRoot60_contig_1_2284474_C_CAG	LjRoot60_contig_1	col0	lj	1,04E-02	FRAME_SHIFT	K02416

LjRoot60_contig_1_2284474_C_CGACAG	LjRoot60_contig_1	col0	lj	5,51E-145	FRAME_SHIFT	K02416
LjRoot60_contig_1_2370124_G_A	LjRoot60_contig_1	col0	lj	2,35E-02	NON_SYNONYMOUS_CODING	K13587
LjRoot60_contig_1_2795286_C_T	LjRoot60_contig_1	col0	lj	1,14E-59	NON_SYNONYMOUS_CODING	K03284
LjRoot60_contig_1_3056787_T_C	LjRoot60_contig_1	col0	lj	1,35E-145	NON_SYNONYMOUS_CODING	K01589
LjRoot60_contig_1_3230440_G_GC	LjRoot60_contig_1	col0	lj	4,47E-80	FRAME_SHIFT	K02667
LjRoot60_contig_1_3353762_T_TA	LjRoot60_contig_1	col0	lj	2,01E-84	INTERGENIC	
LjRoot60_contig_1_3409212_G_A	LjRoot60_contig_1	col0	lj	1,16E-03	NON_SYNONYMOUS_CODING	K02558
LjRoot60_contig_1_3409735_G_GC	LjRoot60_contig_1	col0	lj	4,98E-02	FRAME_SHIFT	K02558
LjRoot60_contig_1_3600869_CG_C	LjRoot60_contig_1	col0	lj	1,87E-06	INTERGENIC	
LjRoot60_contig_2_79500_G_A	LjRoot60_contig_2	col0	lj	1,31E-02	NON_SYNONYMOUS_CODING	
LjRoot60_contig_2_91274_C_A	LjRoot60_contig_2	col0	lj	2,72E-02	NON_SYNONYMOUS_CODING	K12282
LjRoot72_contig_1_2408189_A_G	LjRoot72_contig_1	col0	lj	8,11E-05	NON_SYNONYMOUS_CODING	K13590
LjRoot82_contig_1_1189654_CA_C	LjRoot82_contig_1	col0	lj	3,85E-11	INTERGENIC	
LjRoot82_contig_1_1336619_A_AT	LjRoot82_contig_1	col0	lj	4,55E-03	INTERGENIC	
LjRoot82_contig_1_1551518_AT_A	LjRoot82_contig_1	col0	lj	4,65E-03	INTERGENIC	
LjRoot82_contig_1_1635349_T_TA	LjRoot82_contig_1	col0	lj	6,68E-03	INTERGENIC	
LjRoot82_contig_1_174756_T_C	LjRoot82_contig_1	col0	lj	2,05E-06	INTERGENIC	
LjRoot82_contig_1_174881_TA_T	LjRoot82_contig_1	col0	lj	6,86E-17	INTERGENIC	
LjRoot82_contig_1_1996935_AT_A	LjRoot82_contig_1	col0	lj	1,01E-02	INTERGENIC	
LjRoot82_contig_1_2034972_G_GT	LjRoot82_contig_1	col0	lj	4,38E-03	INTERGENIC	
LjRoot82_contig_1_2151422_G_GA	LjRoot82_contig_1	col0	lj	4,38E-03	INTERGENIC	
LjRoot82_contig_1_2412928_AT_A	LjRoot82_contig_1	col0	lj	2,71E-10	FRAME_SHIFT	K00031
LjRoot82_contig_1_2445414_AT_A	LjRoot82_contig_1	col0	lj	4,66E-03	INTERGENIC	
LjRoot82_contig_1_2476063_T_TA	LjRoot82_contig_1	col0	lj	3,23E-03	INTERGENIC	
LjRoot82_contig_1_2930589_T_TA	LjRoot82_contig_1	col0	lj	4,65E-03	INTERGENIC	
LjRoot82_contig_1_3299145_T_TA	LjRoot82_contig_1	col0	lj	4,38E-03	INTERGENIC	
LjRoot82_contig_1_3326028_AT_A	LjRoot82_contig_1	col0	lj	8,17E-06	INTERGENIC	
LjRoot82_contig_1_3537247_C_CA	LjRoot82_contig_1	col0	lj	6,93E-03	INTERGENIC	
LjRoot82_contig_1_3605497_A_AT	LjRoot82_contig_1	col0	lj	3,94E-03	FRAME_SHIFT	K02014
LjRoot82_contig_1_3765347_C_CA	LjRoot82_contig_1	col0	lj	1,79E-02	INTERGENIC	
LjRoot82_contig_1_4349246_TA_T	LjRoot82_contig_1	col0	lj	5,67E-05	INTERGENIC	
LjRoot82_contig_1_4876963_G_GT	LjRoot82_contig_1	col0	lj	1,09E-02	INTERGENIC	
LjRoot82_contig_1_4902044_G_A	LjRoot82_contig_1	col0	lj	5,01E-05	START_LOST	K11354
LjRoot82_contig_1_4912599_CA_C	LjRoot82_contig_1	col0	lj	4,65E-03	INTERGENIC	
LjRoot82_contig_1_707836_C_CA	LjRoot82_contig_1	col0	lj	4,38E-03	INTERGENIC	
LjNodule218_contig_1_145111_C_G	LjNodule218_contig_1	gifu	at	1,75E-02	INTERGENIC	
LjNodule218_contig_1_1610265_C_T	LjNodule218_contig_1	gifu	at	4,49E-02	NON_SYNONYMOUS_CODING	
LjNodule218_contig_1_2892223_C_A	LjNodule218_contig_1	gifu	at	2,90E-02	NON_SYNONYMOUS_CODING	K00001
LjNodule218_contig_1_4969536_A_C	LjNodule218_contig_1	gifu	at	4,77E-03	NON_SYNONYMOUS_CODING	K18900
LjNodule218_contig_1_5058279_G_C	LjNodule218_contig_1	gifu	at	7,97E-03	NON_SYNONYMOUS_CODING	
LjNodule218_contig_1_5797012_A_AT	LjNodule218_contig_1	gifu	at	7,18E-15	FRAME_SHIFT	
LjNodule218_contig_1_5797434_G_C	LjNodule218_contig_1	gifu	at	1,19E-13	STOP_GAINED	
LjNodule218_contig_1_5940509_C_T	LjNodule218_contig_1	gifu	at	1,46E-03	SYNONYMOUS_CODING	K00862
LjNodule218_contig_1_6431150_GA_G	LjNodule218_contig_1	gifu	at	1,37E-02	INTERGENIC	
LjNodule218_contig_1_649805_T_G	LjNodule218_contig_1	gifu	at	4,39E-03	NON_SYNONYMOUS_CODING	K18900
Root123D2_contig_1_1345332_G_GC	Root123D2_contig_1	gifu	at	1,02E-03	FRAME_SHIFT	K05562
Root29_contig_1_1210506_G_A	Root29_contig_1	gifu	at	3,27E-02	INTERGENIC	
Root29_contig_1_3971888_A_T	Root29_contig_1	gifu	at	1,02E-03	NON_SYNONYMOUS_CODING	K02485
Root418_contig_1_122369_A_G	Root418_contig_1	gifu	at	8,57E-03	NON_SYNONYMOUS_CODING	K03548
Root418_contig_1_1292705_T_G	Root418_contig_1	gifu	at	2,43E-02	NON_SYNONYMOUS_CODING	K05515
Root418_contig_1_2513745_AGCG_A	Root418_contig_1	gifu	at	6,74E-04	ON_CHANGE_PLUS_CODON_DELETION	
Root418_contig_1_272252_G_A	Root418_contig_1	gifu	at	9,05E-03	NON_SYNONYMOUS_CODING	K02495
Root418_contig_1_2850980_A_C	Root418_contig_1	gifu	at	7,85E-03	INTERGENIC	
Root418_contig_1_342461_G_C	Root418_contig_1	gifu	at	8,57E-03	START_LOST	K18297
Root418_contig_1_342613_T_G	Root418_contig_1	gifu	at	1,09E-02	NON_SYNONYMOUS_CODING	K18297
Root418_contig_1_342788_G_C	Root418_contig_1	gifu	at	1,65E-03	NON_SYNONYMOUS_CODING	K18297
Root418_contig_1_3740163_C_A	Root418_contig_1	gifu	at	1,92E-02	NON_SYNONYMOUS_CODING	K01838
Root418_contig_1_4028048_G_A	Root418_contig_1	gifu	at	9,22E-03	NON_SYNONYMOUS_CODING	K03477
Root418_contig_1_4611165_AGC_A	Root418_contig_1	gifu	at	4,66E-04	FRAME_SHIFT	
Root418_contig_1_4911334_G_A	Root418_contig_1	gifu	at	1,03E-02	SYNONYMOUS_CODING	K15125
Root418_contig_1_558788_C_T	Root418_contig_1	gifu	at	5,00E-03	STOP_GAINED	K10125
Root418_contig_1_6180100_A_C	Root418_contig_1	gifu	at	3,31E-06	NON_SYNONYMOUS_CODING	K03406
Root418_contig_1_6320573_G_A	Root418_contig_1	gifu	at	1,60E-03	NON_SYNONYMOUS_CODING	K05962
Root418_contig_1_6320781_A_C	Root418_contig_1	gifu	at	1,11E-02	NON_SYNONYMOUS_CODING	K05962
Root480_contig_1_1422844_T_TG	Root480_contig_1	gifu	at	1,01E-02	FRAME_SHIFT	K02014
Root480_contig_1_2458234_GC_G	Root480_contig_1	gifu	at	1,59E-04	FRAME_SHIFT	K02674
Root480_contig_1_2478388_G_A	Root480_contig_1	gifu	at	4,76E-03	NON_SYNONYMOUS_CODING	
Root480_contig_1_589665_C_A	Root480_contig_1	gifu	at	4,11E-03	NON_SYNONYMOUS_CODING	K16150
Root68_contig_1_5105765_C_T	Root68_contig_1	gifu	at	5,81E-03	NON_SYNONYMOUS_CODING	K21024
Root83_contig_1_1336183_T_C	Root83_contig_1	gifu	at	6,08E-03	NON_SYNONYMOUS_CODING	K01434
Root83_contig_1_149112_GT_G	Root83_contig_1	gifu	at	1,81E-06	INTRAGENIC	
Root83_contig_1_1909548_C_CG	Root83_contig_1	gifu	at	3,99E-02	INTRAGENIC	
Root83_contig_1_2112524_T_TG	Root83_contig_1	gifu	at	2,63E-08	FRAME_SHIFT	K10125
Root83_contig_1_2114467_G_GC	Root83_contig_1	gifu	at	4,32E-04	FRAME_SHIFT	K18765
Root83_contig_1_2289291_GT_G	Root83_contig_1	gifu	at	4,29E-03	INTRAGENIC	
Root83_contig_1_2499682_CA_C	Root83_contig_1	gifu	at	4,57E-05	INTRAGENIC	
Root83_contig_1_2596089_C_T	Root83_contig_1	gifu	at	1,26E-02	SYNONYMOUS_CODING	K11754
Root83_contig_1_2743044_A_AGTTCGCGCACCACCG	Root83_contig_1	gifu	at	2,85E-07	CODON_INSERTION	
Root83_contig_1_3551252_CA_C	Root83_contig_1	gifu	at	1,61E-04	INTRAGENIC	
Root83_contig_1_3619425_G_GC	Root83_contig_1	gifu	at	1,09E-02	TRANSCRIPT	

Root83_contig_1_3930419_G_GC	Root83_contig_1	gifu	at	2,96E-02	INTRAGENIC	
Root83_contig_1_3932491_CA_C	Root83_contig_1	gifu	at	1,81E-06	INTRAGENIC	
Root83_contig_1_4193522_A_G	Root83_contig_1	gifu	at	5,70E-05	NON_SYNONYMOUS_CODING	K01810
Root83_contig_1_4494043_A_G	Root83_contig_1	gifu	at	1,77E-02	NON_SYNONYMOUS_CODING	
Root83_contig_1_5360520_G_GC	Root83_contig_1	gifu	at	1,88E-02	FRAME_SHIFT	K01953
Root83_contig_1_5416987_G_GC	Root83_contig_1	gifu	at	1,38E-03	INTRAGENIC	
Root83_contig_1_5549433_C_T	Root83_contig_1	gifu	at	2,90E-02	NON_SYNONYMOUS_CODING	K02477
Root83_contig_1_6239225_CA_C	Root83_contig_1	gifu	at	8,65E-05	INTRAGENIC	
Root935_contig_1_1133350_GT_G	Root935_contig_1	gifu	at	1,57E-04	INTERGENIC	
Root935_contig_1_1296667_T_A	Root935_contig_1	gifu	at	3,78E-03	STOP_GAINED	
Root935_contig_1_1296766_T_A	Root935_contig_1	gifu	at	3,30E-03	STOP_GAINED	
Root935_contig_1_1297043_A_AT	Root935_contig_1	gifu	at	2,88E-02	FRAME_SHIFT	
Root935_contig_1_1297211_TA_T	Root935_contig_1	gifu	at	6,00E-03	FRAME_SHIFT	
Root935_contig_1_1297457_G_GA	Root935_contig_1	gifu	at	8,49E-03	FRAME_SHIFT	
Root935_contig_1_1297667_GA_G	Root935_contig_1	gifu	at	1,07E-02	FRAME_SHIFT	
Root935_contig_1_1297922_A_T	Root935_contig_1	gifu	at	8,57E-03	STOP_GAINED	
Root935_contig_1_1297942_TA_T	Root935_contig_1	gifu	at	7,52E-03	FRAME_SHIFT	
Root935_contig_1_1298035_T_A	Root935_contig_1	gifu	at	3,01E-02	STOP_GAINED	
Root935_contig_1_1298172_A_G	Root935_contig_1	gifu	at	3,86E-04	NON_SYNONYMOUS_CODING	
Root935_contig_1_132172_TA_T	Root935_contig_1	gifu	at	1,52E-06	INTERGENIC	
Root935_contig_1_1332066_C_T	Root935_contig_1	gifu	at	2,65E-02	NON_SYNONYMOUS_CODING	K01775
Root935_contig_1_1360017_C_CT	Root935_contig_1	gifu	at	1,86E-04	FRAME_SHIFT	K07710
Root935_contig_1_1612582_G_A	Root935_contig_1	gifu	at	1,42E-02	NON_SYNONYMOUS_CODING	K11912
Root935_contig_1_1810305_C_T	Root935_contig_1	gifu	at	1,31E-09	NON_SYNONYMOUS_CODING	K03555
Root935_contig_1_1818721_T_C	Root935_contig_1	gifu	at	1,72E-03	NON_SYNONYMOUS_CODING	K18954
Root935_contig_1_1831048_G_A	Root935_contig_1	gifu	at	1,35E-02	NON_SYNONYMOUS_CODING	K08303
Root935_contig_1_1851687_A_G	Root935_contig_1	gifu	at	1,83E-02	NON_SYNONYMOUS_CODING	K21470
Root935_contig_1_191603_C_T	Root935_contig_1	gifu	at	1,26E-02	NON_SYNONYMOUS_CODING	K21573
Root935_contig_1_1927794_T_C	Root935_contig_1	gifu	at	1,57E-04	NON_SYNONYMOUS_CODING	
Root935_contig_1_194808_T_TA	Root935_contig_1	gifu	at	1,61E-02	INTERGENIC	
Root935_contig_1_1961396_A_G	Root935_contig_1	gifu	at	4,21E-06	NON_SYNONYMOUS_CODING	
Root935_contig_1_2001240_A_G	Root935_contig_1	gifu	at	2,29E-02	NON_SYNONYMOUS_CODING	K06147
Root935_contig_1_2161811_T_C	Root935_contig_1	gifu	at	4,24E-02	SYNONYMOUS_CODING	K01400
Root935_contig_1_2212383_G_A	Root935_contig_1	gifu	at	3,93E-10	NON_SYNONYMOUS_CODING	K09704
Root935_contig_1_2318366_A_G	Root935_contig_1	gifu	at	4,19E-02	SYNONYMOUS_CODING	
Root935_contig_1_2467231_A_G	Root935_contig_1	gifu	at	2,28E-02	SYNONYMOUS_CODING	K19802
Root935_contig_1_2481186_CT_C	Root935_contig_1	gifu	at	2,90E-02	INTERGENIC	
Root935_contig_1_2625923_G_GA	Root935_contig_1	gifu	at	3,37E-05	INTERGENIC	
Root935_contig_1_2784442_A_T	Root935_contig_1	gifu	at	8,00E-03	STOP_GAINED	K01126
Root935_contig_1_2802407_T_C	Root935_contig_1	gifu	at	7,38E-03	SYNONYMOUS_CODING	K06607
Root935_contig_1_2900558_G_A	Root935_contig_1	gifu	at	3,18E-03	SYNONYMOUS_CODING	K03306
Root935_contig_1_2926333_G_A	Root935_contig_1	gifu	at	7,53E-03	NON_SYNONYMOUS_CODING	K01176
Root935_contig_1_2974990_C_T	Root935_contig_1	gifu	at	1,25E-02	NON_SYNONYMOUS_CODING	K02529
Root935_contig_1_2975790_A_G	Root935_contig_1	gifu	at	1,87E-02	NON_SYNONYMOUS_CODING	K02529
Root935_contig_1_2998136_G_A	Root935_contig_1	gifu	at	2,39E-02	INTERGENIC	
Root935_contig_1_3037344_C_T	Root935_contig_1	gifu	at	3,38E-03	NON_SYNONYMOUS_CODING	K01488
Root935_contig_1_3142761_C_T	Root935_contig_1	gifu	at	2,35E-04	NON_SYNONYMOUS_CODING	K21023
Root935_contig_1_3195796_C_T	Root935_contig_1	gifu	at	3,03E-03	NON_SYNONYMOUS_CODING	K01534
Root935_contig_1_3276641_G_A	Root935_contig_1	gifu	at	6,57E-04	SYNONYMOUS_CODING	K11904
Root935_contig_1_3412560_G_A	Root935_contig_1	gifu	at	1,55E-02	NON_SYNONYMOUS_CODING	K21573
Root935_contig_1_346656_A_G	Root935_contig_1	gifu	at	8,49E-03	SYNONYMOUS_CODING	K19430
Root935_contig_1_3713994_G_A	Root935_contig_1	gifu	at	1,05E-02	NON_SYNONYMOUS_CODING	K20276
Root935_contig_1_3784595_T_C	Root935_contig_1	gifu	at	2,08E-02	SYNONYMOUS_CODING	
Root935_contig_1_3832483_G_A	Root935_contig_1	gifu	at	1,09E-02	INTERGENIC	
Root935_contig_1_3850836_A_G	Root935_contig_1	gifu	at	1,86E-04	SYNONYMOUS_CODING	K21573
Root935_contig_1_3912126_A_G	Root935_contig_1	gifu	at	1,53E-02	NON_SYNONYMOUS_CODING	K02477
Root935_contig_1_3912619_G_GT	Root935_contig_1	gifu	at	1,55E-02	INTERGENIC	
Root935_contig_1_3933964_C_T	Root935_contig_1	gifu	at	9,22E-03	NON_SYNONYMOUS_CODING	K05989
Root935_contig_1_4032327_A_G	Root935_contig_1	gifu	at	2,22E-02	NON_SYNONYMOUS_CODING	K21573
Root935_contig_1_4047504_T_G	Root935_contig_1	gifu	at	1,37E-02	SYNONYMOUS_CODING	K03306
Root935_contig_1_4085547_T_C	Root935_contig_1	gifu	at	8,92E-03	NON_SYNONYMOUS_CODING	K00782
Root935_contig_1_4220861_G_A	Root935_contig_1	gifu	at	9,05E-03	SYNONYMOUS_CODING	
Root935_contig_1_4347401_CT_C	Root935_contig_1	gifu	at	1,05E-02	FRAME_SHIFT	K01205
Root935_contig_1_4513695_T_C	Root935_contig_1	gifu	at	1,83E-02	SYNONYMOUS_CODING	K20276
Root935_contig_1_4545572_A_G	Root935_contig_1	gifu	at	1,77E-02	SYNONYMOUS_CODING	
Root935_contig_1_4546121_T_C	Root935_contig_1	gifu	at	1,05E-02	NON_SYNONYMOUS_CODING	K02014
Root935_contig_1_4641132_T_C	Root935_contig_1	gifu	at	3,64E-09	SYNONYMOUS_CODING	K13611
Root935_contig_1_4736301_G_GT	Root935_contig_1	gifu	at	4,87E-09	INTERGENIC	
Root935_contig_1_4782185_G_A	Root935_contig_1	gifu	at	6,04E-04	NON_SYNONYMOUS_CODING	K01897
Root935_contig_1_4816946_T_TA	Root935_contig_1	gifu	at	4,90E-02	FRAME_SHIFT	
Root935_contig_1_4835112_A_G	Root935_contig_1	gifu	at	5,43E-04	NON_SYNONYMOUS_CODING	K01286
Root935_contig_1_4933058_G_A	Root935_contig_1	gifu	at	2,09E-02	NON_SYNONYMOUS_CODING	K01657
Root935_contig_1_5014185_G_A	Root935_contig_1	gifu	at	8,57E-03	NON_SYNONYMOUS_CODING	K05349
Root935_contig_1_5027432_T_A	Root935_contig_1	gifu	at	1,05E-02	NON_SYNONYMOUS_CODING	K15531
Root935_contig_1_5058783_CA_C	Root935_contig_1	gifu	at	1,55E-02	INTERGENIC	
Root935_contig_1_5058783_C_CA	Root935_contig_1	gifu	at	5,81E-03	INTERGENIC	
Root935_contig_1_5069765_A_T	Root935_contig_1	gifu	at	2,00E-03	STOP_GAINED	K15738
Root935_contig_1_5108675_T_C	Root935_contig_1	gifu	at	2,29E-02	NON_SYNONYMOUS_CODING	
Root935_contig_1_514114_A_G	Root935_contig_1	gifu	at	4,21E-06	SYNONYMOUS_CODING	
Root935_contig_1_61760_G_A	Root935_contig_1	gifu	at	5,27E-03	INTERGENIC	
Root935_contig_1_667891_T_C	Root935_contig_1	gifu	at	7,38E-03	NON_SYNONYMOUS_CODING	

Root935_contig_1_765221_GT_G	Root935_contig_1	gifu	at	3,66E-02	FRAME_SHIFT	K07683
Root935_contig_1_766085_G_A	Root935_contig_1	gifu	at	1,53E-02	NON_SYNONYMOUS_CODING	K07683
Root935_contig_1_926689_C_CA	Root935_contig_1	gifu	at	4,41E-05	INTERGENIC	
Root935_contig_1_929391_T_C	Root935_contig_1	gifu	at	1,88E-02	SYNONYMOUS_CODING	K07386
Root935_contig_1_988016_C_T	Root935_contig_1	gifu	at	2,84E-02	STOP_GAINED	K00568
LjNodule218_contig_1_1200563_T_C	LjNodule218_contig_1	gifu	lj	1,34E-02	NON_SYNONYMOUS_CODING	K02112
LjNodule218_contig_1_14367_G_T	LjNodule218_contig_1	gifu	lj	4,65E-03	NON_SYNONYMOUS_CODING	K00564
LjNodule218_contig_1_145111_C_G	LjNodule218_contig_1	gifu	lj	1,06E-02	INTERGENIC	
LjNodule218_contig_1_1917987_A_G	LjNodule218_contig_1	gifu	lj	1,67E-02	SYNONYMOUS_CODING	
LjNodule218_contig_1_5215851_G_A	LjNodule218_contig_1	gifu	lj	4,38E-03	SYNONYMOUS_CODING	K08234
LjNodule218_contig_1_5215854_G_C	LjNodule218_contig_1	gifu	lj	4,07E-03	SYNONYMOUS_CODING	K08234
LjNodule218_contig_1_5215865_G_T	LjNodule218_contig_1	gifu	lj	4,07E-03	NON_SYNONYMOUS_CODING	K08234
LjNodule218_contig_1_5797012_A_AT	LjNodule218_contig_1	gifu	lj	9,74E-15	FRAME_SHIFT	
LjNodule218_contig_1_5797434_G_C	LjNodule218_contig_1	gifu	lj	2,58E-16	STOP_GAINED	
LjNodule218_contig_1_6097160_C_G	LjNodule218_contig_1	gifu	lj	4,86E-02	NON_SYNONYMOUS_CODING	K01998
LjNodule218_contig_2_453842_C_A	LjNodule218_contig_2	gifu	lj	8,06E-04	INTERGENIC	
LjRoot262_contig_1_1454890_A_G	LjRoot262_contig_1	gifu	lj	3,34E-02	NON_SYNONYMOUS_CODING	K05799
LjRoot262_contig_1_3119111_AGC_A	LjRoot262_contig_1	gifu	lj	1,75E-03	FRAME_SHIFT	K03744
LjRoot262_contig_1_4466405_G_T	LjRoot262_contig_1	gifu	lj	1,98E-12	NON_SYNONYMOUS_CODING	K21023
LjRoot262_contig_1_4467079_G_C	LjRoot262_contig_1	gifu	lj	7,41E-04	NON_SYNONYMOUS_CODING	K20971
LjRoot262_contig_1_4467102_G_C	LjRoot262_contig_1	gifu	lj	1,21E-03	NON_SYNONYMOUS_CODING	K20971
LjRoot262_contig_1_4469688_G_C	LjRoot262_contig_1	gifu	lj	9,20E-03	NON_SYNONYMOUS_CODING	K20971
LjRoot33_contig_1_1237766_A_G	LjRoot33_contig_1	gifu	lj	1,67E-02	INTERGENIC	
LjRoot33_contig_1_2414495_TCAG_T	LjRoot33_contig_1	gifu	lj	1,43E-02	CODON_DELETION	
LjRoot33_contig_1_2581958_G_T	LjRoot33_contig_1	gifu	lj	3,43E-02	INTERGENIC	
LjRoot33_contig_1_2581987_G_A	LjRoot33_contig_1	gifu	lj	2,20E-02	INTERGENIC	
LjRoot33_contig_1_3136903_CTTG_C	LjRoot33_contig_1	gifu	lj	1,73E-05	ON_CHANGE_PLUS_CODON_DELE	K02556
LjRoot33_contig_1_3210097_G_C	LjRoot33_contig_1	gifu	lj	2,02E-02	NON_SYNONYMOUS_CODING	K02412
LjRoot33_contig_1_3214383_T_G	LjRoot33_contig_1	gifu	lj	3,01E-02	NON_SYNONYMOUS_CODING	K02417
LjRoot33_contig_1_3240739_A_C	LjRoot33_contig_1	gifu	lj	2,02E-02	NON_SYNONYMOUS_CODING	K03411
LjRoot33_contig_1_3661832_G_GCAGGCGTT	LjRoot33_contig_1	gifu	lj	5,39E-03	FRAME_SHIFT	K03592
LjRoot33_contig_1_3978348_C_T	LjRoot33_contig_1	gifu	lj	2,38E-02	INTERGENIC	
LjRoot33_contig_1_4269566_GCACGACTTTGCGCCC_G	LjRoot33_contig_1	gifu	lj	1,51E-02	INTERGENIC	
LjRoot33_contig_1_4329412_G_C	LjRoot33_contig_1	gifu	lj	4,65E-03	NON_SYNONYMOUS_CODING	K14393
LjRoot33_contig_1_4907943_C_T	LjRoot33_contig_1	gifu	lj	1,40E-02	INTERGENIC	
LjRoot33_contig_1_5228069_A_T	LjRoot33_contig_1	gifu	lj	2,82E-02	NON_SYNONYMOUS_CODING	K01995
LjRoot59_contig_1_1318840_C_T	LjRoot59_contig_1	gifu	lj	3,54E-02	NON_SYNONYMOUS_CODING	K06194
LjRoot59_contig_1_2034368_G_C	LjRoot59_contig_1	gifu	lj	2,35E-02	NON_SYNONYMOUS_CODING	K00372
LjRoot59_contig_1_2404453_C_T	LjRoot59_contig_1	gifu	lj	7,72E-09	NON_SYNONYMOUS_CODING	K07689
LjRoot59_contig_1_2485667_T_A	LjRoot59_contig_1	gifu	lj	6,84E-03	NON_SYNONYMOUS_CODING	K14657
LjRoot59_contig_1_2487898_C_A	LjRoot59_contig_1	gifu	lj	6,94E-03	INTERGENIC	
LjRoot59_contig_1_2490400_G_A	LjRoot59_contig_1	gifu	lj	2,10E-04	NON_SYNONYMOUS_CODING	
LjRoot59_contig_1_2509521_T_A	LjRoot59_contig_1	gifu	lj	2,52E-03	NON_SYNONYMOUS_CODING	K16839
LjRoot59_contig_1_3159170_C_T	LjRoot59_contig_1	gifu	lj	1,25E-02	NON_SYNONYMOUS_CODING	K20975
LjRoot59_contig_1_3160449_C_A	LjRoot59_contig_1	gifu	lj	4,11E-02	NON_SYNONYMOUS_CODING	K20975
LjRoot59_contig_1_3160868_T_A	LjRoot59_contig_1	gifu	lj	1,31E-02	NON_SYNONYMOUS_CODING	K20975
LjRoot59_contig_1_3160981_C_A	LjRoot59_contig_1	gifu	lj	2,02E-02	NON_SYNONYMOUS_CODING	K20975
LjRoot59_contig_1_3161472_C_T	LjRoot59_contig_1	gifu	lj	1,11E-02	NON_SYNONYMOUS_CODING	K20975
LjRoot59_contig_1_3195626_C_T	LjRoot59_contig_1	gifu	lj	1,03E-02	NON_SYNONYMOUS_CODING	K02584
LjRoot59_contig_1_3195977_T_G	LjRoot59_contig_1	gifu	lj	6,03E-03	NON_SYNONYMOUS_CODING	K02584
LjRoot59_contig_1_3196024_C_T	LjRoot59_contig_1	gifu	lj	9,75E-03	NON_SYNONYMOUS_CODING	K02584
LjRoot59_contig_1_3500373_G_T	LjRoot59_contig_1	gifu	lj	1,48E-03	STOP_GAINED	K01895
LjRoot59_contig_1_3603194_G_C	LjRoot59_contig_1	gifu	lj	3,80E-03	NON_SYNONYMOUS_CODING	K00131
LjRoot59_contig_1_4922906_G_C	LjRoot59_contig_1	gifu	lj	2,20E-02	INTERGENIC	
LjRoot59_contig_1_5048823_G_GC	LjRoot59_contig_1	gifu	lj	3,94E-03	FRAME_SHIFT	K07678
LjRoot59_contig_1_5049620_T_C	LjRoot59_contig_1	gifu	lj	1,05E-02	NON_SYNONYMOUS_CODING	K07678
LjRoot59_contig_1_5050133_C_CG	LjRoot59_contig_1	gifu	lj	1,57E-02	FRAME_SHIFT	K07678
LjRoot59_contig_1_5121022_G_GC	LjRoot59_contig_1	gifu	lj	3,37E-02	INTERGENIC	
LjRoot59_contig_1_6086385_G_A	LjRoot59_contig_1	gifu	lj	1,96E-03	NON_SYNONYMOUS_CODING	K07675
LjRoot59_contig_1_899570_C_T	LjRoot59_contig_1	gifu	lj	3,23E-03	SYNONYMOUS_CODING	K02032
LjRoot59_contig_1_958826_C_T	LjRoot59_contig_1	gifu	lj	4,86E-03	NON_SYNONYMOUS_CODING	K01803
LjRoot72_contig_1_418349_G_A	LjRoot72_contig_1	gifu	lj	2,43E-02	NON_SYNONYMOUS_CODING	K03043
LjRoot72_contig_1_4410629_A_C	LjRoot72_contig_1	gifu	lj	1,00E-07	INTERGENIC	

Table 4. Structural genome rearrangements detected in the whole-genome sequencing of evolved isolates. The table lists the strain, genomic regions, and type of structural variant (SV; *n* = 12). In addition, the variation type (insertion or deletion), length, and corresponding evolution host are provided. For insertions, the likely origin of the inserted DNA fragment is indicated.

strain	family	chrom	pos	type	SVLEN	plant	syncom	condition	insertion_origin	insertion_qlen
LjRoot262	Sphingomonadaceae	LjRoot262_contig_1	2703013	DEL	-108	gifu	lj(gifu)-1	native		
LjRoot46	Rhizobiaceae	LjRoot46_contig_5	50198	INS	1417	gifu	lj(gifu)-1	native	LjRoot46_contig_4	1418
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	5734594	DEL	-620	gifu	lj(gifu)-1	native		
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	5859699	DEL	-596	gifu	lj(gifu)-1	native		
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	4453767	DEL	-604	gifu	lj(gifu)-1	native		
LjRoot60	Xanthomonadaceae	LjRoot60_contig_1	3887682	DEL	-247	gifu	lj(gifu)-1	native		
LjRoot80	Mycobacteriaceae	LjRoot80_contig_1	820540	DEL	-702	gifu	lj(gifu)-1	native		
LjRoot262	Sphingomonadaceae	LjRoot262_contig_1	2703013	DEL	-108	col0	lj(col0)-1	non-native		
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	5859725	DEL	-596	col0	lj(col0)-1	non-native		
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	5734568	DEL	-620	col0	lj(col0)-1	non-native		
LjRoot59	Pseudomonadaceae	LjRoot59_contig_1	4453767	DEL	-604	col0	lj(col0)-1	non-native		
LjRoot60	Xanthomonadaceae	LjRoot60_contig_1	3887682	DEL	-247	col0	lj(col0)-1	non-native		

Table 5. Mutations identified in highly adapted evolved isolates. For each variant (*n* = 67), the strain background, percent fitness change, *P*-value, genomic region, distance to the nearest downstream ORF, and corresponding evolution host are provided. For mutations within coding sequences, the distance was set to zero. The associated KEGG ontology and gene alias are listed.

strain	fitness_change	padj	chrom	feature	location	distance	syncom	plant	ko	alias	ref	alt
Root480	376,62	1,22E-04	Root480_contig_1	CDS	1277616	0	at	gifu	K00786	E2.4.-.-	A	G
Root480	376,62	1,22E-04	Root480_contig_1	CDS	589665	0	at	gifu	K16150	K16150	C	A
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	3600869	85	lj	col0	K00404	ccoN	C	CG
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	3353762	52	lj	col0	K02663	pilN	T	TA
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	3230440	0	lj	col0	K02667	pilR, pehR	G	GC
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	3056787	0	lj	col0	K01589	purK	T	C
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	2795286	0	lj	col0	K03284	corA	C	T
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	2284474	0	lj	col0	K02416	fliM	C	CGACAG
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	1739115	0	lj	col0	K01915	glnA, GLUL	T	A
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	1284472	0	lj	col0		hypothetical proteir	C	CATCGGTTGCGCGATCGGCCGGCGGCCACGCTGGTAG
LjRoot60	116,85	1,53E-04	LjRoot60_contig_1	CDS	1284458	0	lj	col0		hypothetical proteir	TG	T
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	4469648	0	lj	col0	K20971	tadS	GCC	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	tRNA	938901	32	lj	col0		tRNA-Glu(ctc)	GC	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938835	0	lj	col0	K00380	cysJ	G	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938833	0	lj	col0	K00380	cysJ	CG	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938828	0	lj	col0	K00380	cysJ	G	T
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938826	0	lj	col0	K00380	cysJ	T	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938823	0	lj	col0	K00380	cysJ	A	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938821	0	lj	col0	K00380	cysJ	G	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938818	0	lj	col0	K00380	cysJ	CAA	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938813	0	lj	col0	K00380	cysJ	G	T
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938812	0	lj	col0	K00380	cysJ	A	T
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938807	0	lj	col0	K00380	cysJ	G	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938804	0	lj	col0	K00380	cysJ	T	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938800	0	lj	col0	K00380	cysJ	T	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938799	0	lj	col0	K00380	cysJ	C	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938798	0	lj	col0	K00380	cysJ	C	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938791	0	lj	col0	K00380	cysJ	T	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938790	0	lj	col0	K00380	cysJ	A	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938786	0	lj	col0	K00380	cysJ	G	T
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938780	0	lj	col0	K00380	cysJ	A	ATGC
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938779	0	lj	col0	K00380	cysJ	T	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938778	0	lj	col0	K00380	cysJ	G	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938776	0	lj	col0	K00380	cysJ	A	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938773	0	lj	col0	K00380	cysJ	C	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938768	0	lj	col0	K00380	cysJ	A	C
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938767	0	lj	col0	K00380	cysJ	C	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938764	0	lj	col0	K00380	cysJ	C	G
LjRoot262	106,19	1,53E-04	LjRoot262_contig_1	CDS	938760	0	lj	col0	K00380	cysJ	TTTG	T
Root142	94,99	3,05E-04	Root142_contig_1	CDS	1967516	0	at	gifu		hypothetical proteir	T	G
Root142	94,99	3,05E-04	Root142_contig_2	CDS	1657570	0	at	gifu	K13590	dgcB	T	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	1604478	0	at	gifu	K02986	RP-S4, rpsD	A	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815368	0	at	gifu		hypothetical proteir	G	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815228	0	at	gifu		hypothetical proteir	T	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815155	0	at	gifu		hypothetical proteir	G	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815153	0	at	gifu		hypothetical proteir	G	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815152	0	at	gifu		hypothetical proteir	G	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815151	0	at	gifu		hypothetical proteir	C	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815148	0	at	gifu		hypothetical proteir	A	T
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815144	0	at	gifu		hypothetical proteir	A	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815141	0	at	gifu		hypothetical proteir	C	CGACGGT
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815137	0	at	gifu		hypothetical proteir	T	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815136	0	at	gifu		hypothetical proteir	C	G
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815135	0	at	gifu		hypothetical proteir	C	T
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815130	0	at	gifu		hypothetical proteir	G	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815129	0	at	gifu		hypothetical proteir	A	G
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815128	0	at	gifu		hypothetical proteir	A	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815111	0	at	gifu		hypothetical proteir	C	G
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815109	0	at	gifu		hypothetical proteir	G	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815108	0	at	gifu		hypothetical proteir	T	G
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815106	0	at	gifu		hypothetical proteir	G	T
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815105	0	at	gifu		hypothetical proteir	T	A
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815102	0	at	gifu		hypothetical proteir	C	G
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815098	0	at	gifu		hypothetical proteir	G	C
Root142	94,99	3,05E-04	Root142_contig_1	CDS	815097	0	at	gifu		hypothetical proteir	C	T
Root142	94,99	3,05E-04	Root142_contig_4	CDS	712228	0	at	gifu	K02051	ABC.SN.S	A	AG
Root142	94,99	3,05E-04	Root142_contig_2	CDS	81818	0	at	gifu	K22293	rspR	C	A

Table 6. Dilution factors used for the culture collection of evolved isolates. For each reisolation, the experimental condition (SynCom and plant), plant generation, dilution factor and number of plates are provided.

plant	syncom	generation	dilution	#plates
col0	at	1	18000	15
col0	lj	1	18000	15
gifu	at	1	54000	15
gifu	lj	1	54000	15
col0	at	3	18000	15
col0	lj	3	18000	15
gifu	at	3	54000	15
gifu	lj	3	54000	15
col0	at	5	18000	15
col0	lj	5	18000	15
gifu	at	5	54000	15
gifu	lj	5	54000	15
col0	at	7	36000	15
col0	lj	7	36000	15
gifu	at	7	100000	15
gifu	lj	7	100000	15
col0	at	9	56000	15
col0	lj	9	56000	15
gifu	at	9	100000	15
gifu	lj	9	100000	15
col0	at	11	56000	15
col0	lj	11	56000	15
gifu	at	11	62500	15
gifu	lj	11	62500	15
col0	at	13	54000	15
col0	lj	13	54000	15
gifu	at	13	62000	15
gifu	lj	13	62000	15
col0	at	15	54000	15
col0	lj	15	54000	15
gifu	at	15	65000	15
gifu	lj	15	65000	15
col0	at	16	54000	30
col0	lj	16	54000	30
gifu	at	16	65000	30
gifu	lj	16	65000	30

Table 7. Summary of the dataset used for training the random forest classifier. The table reports the number of individual training instances for each strain and for the TY-medium control, comprising a total of 35 classes used to train the classifier on fluorescence spectral data.

strainID	replicates_in_training_data
Root131	144
LjRoot44	144
Root935	144
Root61	144
LjRoot60	144
LjRoot303	144
Root77	144
Root83	144
LjRoot33	144
LjRoot80	144
Root123D2	144
LjRoot17	144
LjRoot24	216
Root1310	216
LjRoot5	216
Root142	216
Root685	216
LjRoot262	216
LjRoot52	216
LjRoot59	216
LjRoot82	216
LjRoot222	216
Root720	216
Root265	216
LjRoot46	216
LjNodule218	216
LjRoot72	216
Root68	216
Root480	216
Root418	216
LjRoot1	216
Root29	216
Root101	216
Root695	216
medium	432

Table 8. Evolved isolates used for reconstitution experiments. Reisolated strains from the evolution experiment are listed along with their respective metadata ($n = 34$). Columns include a unique identifier, original strain name, taxonomic family, SynCom, evolution host (i.e., the host on which the strain evolved during the experiment), the experimental evolution condition, and the competition experiments in which the strain was used.

reisolD	strain	family	syncom	plant	condition	experiment
53	LjNodule218	Phyllobacteriaceae	lj	gifu	native	exp_1
2-23	LjNodule218	Phyllobacteriaceae	at	gifu	non-native	exp_1
62	LjRoot17	Caulobacteraceae	lj	gifu	native	exp_1
80	LjRoot222	Hyphomicrobiaceae	lj	gifu	native	exp_1
2-43	LjRoot222	Hyphomicrobiaceae	lj	gifu	native	exp_1
2-44	LjRoot262	Sphingomonadaceae	lj	gifu	native	exp_1
2-3	LjRoot262	Sphingomonadaceae	lj	col0	non-native	exp_1
72	LjRoot33	Oxalobacteraceae	lj	gifu	native	exp_1
2-43	LjRoot33	Oxalobacteraceae	lj	gifu	native	exp_1
74	LjRoot44	Microbacteriaceae	lj	gifu	native	exp_1
167	LjRoot44	Microbacteriaceae	lj	col0	non-native	exp_1
2-45	LjRoot46	Rhizobiaceae	lj	gifu	native	exp_1
81	LjRoot59	Pseudomonadaceae	lj	gifu	native	exp_1
140	LjRoot59	Pseudomonadaceae	lj	col0	non-native	exp_1
72	LjRoot60	Xanthomonadaceae	lj	gifu	native	exp_1
2-94	LjRoot60	Xanthomonadaceae	lj	col0	non-native	exp_1
5	LjRoot60	Xanthomonadaceae	lj	col0	non-native	exp_1
54	LjRoot72	Comamonadaceae	lj	gifu	native	exp_1
138	LjRoot72	Comamonadaceae	lj	col0	non-native	exp_1
61	LjRoot80	Mycobacteriaceae	lj	gifu	native	exp_1
2-68	LjRoot80	Mycobacteriaceae	at	col0	native	exp_1
71	LjRoot82	Flavobacteriaceae	lj	gifu	native	exp_1
2-94	LjRoot82	Flavobacteriaceae	lj	col0	non-native	exp_1
90	Root142	Rhizobiaceae	at	col0	native	exp_1
44	Root142	Rhizobiaceae	at	gifu	non-native	exp_1
2-74	Root265	Mycobacteriaceae	at	col0	native	exp_1
21	Root480	Xanthomonadaceae	at	gifu	non-native	exp_1
93	Root61	Microbacteriaceae	at	col0	native	exp_1
2-34	Root61	Microbacteriaceae	at	gifu	non-native	exp_1
2-68	Root685	Hyphomicrobiaceae	at	col0	native	exp_1
11	Root77	Caulobacteraceae	at	gifu	non-native	exp_1
2-77	Root935	Flavobacteriaceae	at	col0	native	exp_1
134	Root935	Flavobacteriaceae	lj	col0	non-native	exp_1
18	Root935	Flavobacteriaceae	at	gifu	non-native	exp_1

Acknowledgements

We acknowledge C. Duran for the assistance during the evolution experiment, G. Altay for assistance in reisolating individual evolved isolates and the members of the lab for fruitful discussions.

Funding:

German Research Foundation under the German Excellence Strategy (EXC number 2048/1 project 390686111)

European Research Council (project PHYCOSPHERES).

Author Contributions:

Conceptualization: NK, KW, RGO

Methodology: NK, KW, RGO

Investigation: NK

Visualization: NK

Funding acquisition: RGO

Project administration: RGO

Supervision: KW, RGO

Writing – original draft: NK, RGO

Writing – review & editing: NK, KW, RGO

Competing interests:

Authors declare that they have no competing interest.

Chapter 3
Heterogeneity of Bacterial Fluorescent Spectra

"Diversity is not a nuisance to be eliminated; it is the raw material of evolution."
Theodosius Dobzhansky

3. Heterogeneity of bacterial fluorescent spectra

3.1. Introduction

As researchers move beyond descriptive community-level surveys, there is an increasing demand for large, comprehensive culture collections (Cross et al., 2019). The goal is to maximize the recovery of bacterial diversity to capture not just the abundant taxa, but also rare species that may play important ecological roles (Jousset et al., 2017). To address this need, strategies to isolate novel bacteria range from low-throughput methods that mimic natural micro-environments, to high-throughput dilution-to-extinction approaches. However, while isolation protocols have advanced significantly, the downstream capacity to taxonomically classify these isolates has largely failed to keep pace. This bottleneck represents a fundamental challenge in the field of culturomics (Lagier et al., 2018).

The evolution experiment described in the previous chapter generated an extensive culture collection, offering a frozen fossil record of evolved isolates. Through the large-scale sampling of adapting populations across multiple time points and host species, approximately 23,000 individual bacterial isolates were obtained. This collection represents an extensive dataset, capable of elucidating the genetic determinants of host-driven adaptation in a community context. However, the potential of this collection of bacteria is highly dependent on knowing the identity of each isolate. To confirm bacterial phenotypes, such as changes in colonization fitness, it was necessary to retrieve the correct target strain from these mixed populations. Yet, the size of the culture collection renders standard verification protocols, such as Sanger or amplicon sequencing of marker genes, logistically impractical.

3.1.1. Progress and challenges in establishing culture collections

Over the last decade, significant efforts have been made to establish comprehensive and taxonomically diverse culture collections that mirror the natural complexity of the plant microbiome. These culture collections serve as the starting point for mechanistic studies by enabling the reconstruction of SynComs that closely resemble natural communities (Vorholt et al., 2017). Strategies for the establishment

of these libraries range from techniques that prioritize physiological relevance to high-throughput methods designed for large-scale efforts.

Methods focusing on physiological relevance often employ *in situ* cultivation approaches to recover unculturable bacteria by mimicking the chemical and structural properties of the natural environment (Kaeberlein et al., 2002; Vartoukian et al., 2010). To combine this environmental mimicking with higher resolution, single-cell encapsulation in gel microdroplets allows for the exchange of metabolites while maintaining axenic conditions (Zengler et al., 2005).

To achieve the scale required for large libraries, however, researchers typically rely on high-throughput dilution-to-extinction strategies (Connon & Giovannoni, 2002). Here, microbial suspensions are diluted until aliquots contain a single microbial cell. This approach eliminates competition from fast-growing species and improves the isolation success of slower-growing strains (J. Zhang et al., 2021). The success of these isolation strategies is highly dependent on media composition. Many bacteria fail to grow in isolation because they lack the capacity to produce essential metabolites. For instance, supplementing media with specific siderophores normally produced by neighboring bacteria can facilitate the growth of otherwise dependent isolates (D'Onofrio et al., 2010). By optimizing these cultivation protocols, extensive libraries such as the *At*-SPHERE and *Lj*-SPHERE collections have been established for the model plants *Arabidopsis thaliana* and *Lotus japonicus* (Bai et al., 2015; Wippel et al., 2021). These collections have been instrumental in identifying host-specific commensals and revealing the phenomenon of bacterial host preference. Recent work has extended these efforts to aquatic systems, establishing the *Cr*-SPHERE collection from the phycosphere of the green alga *Chlamydomonas reinhardtii*, highlighting the broad applicability of these methods (Durán et al., 2022).

The primary challenge in culturomics is the increasing number of isolates recovered. While the current gold standard of taxonomic identification via 16S rRNA gene sequencing is effective for collections with a few thousand isolates, the workflow scales linearly in terms of cost and labor (J. Zhang et al., 2021). This scaling becomes a prohibitive barrier when applied to large collections, such as the 23,000 isolates presented in this thesis, rendering standard sequencing impractical. Furthermore,

these molecular methods are destructive endpoint measurements as they do not allow for the sorting of living cells based on their identity in real-time. Consequently, a proxy for taxonomic identity is required to bridge the gap between what can be recovered and what can be verified. Ideally, such a proxy would be rapid, non-destructive, and free of reagent costs.

3.1.2. High-throughput phenotyping

To resolve the discrepancy between isolation capacity and identification throughput, a shift from genotypic verification to phenotypic fingerprinting is required. Whole-organism fingerprinting techniques, such as Fourier-transform infrared (FTIR) spectroscopy, Raman spectroscopy, and fluorescence spectroscopy, measure the biochemical composition of the cell (Braga et al., 2013; Jarvis & Goodacre, 2008; Leblanc & Dufour, 2002; K. S. Lee et al., 2024; van Baar, 2000). Among these techniques, intrinsic fluorescence spectroscopy is particularly suited for high-throughput screening because it is non-destructive, rapid, and requires no external reagents or staining.

As reviewed by Ammor, living bacterial cells act as complex mixtures of endogenous fluorophores (Ammor, 2007). These intracellular molecules absorb light at specific excitation wavelengths and emit it at longer emission wavelengths. The resulting spectral fingerprint is primarily shaped by aromatic amino acids (~280 nm) and visible-range metabolic cofactors like NADH (excitation ~340/450 nm). Together, these features serve as proxies for protein composition and cellular energetics (Ammor, 2007). Because the ratio and quantum yield of these fluorophores vary between species, the complete spectrum could serve as a unique identifier. To extract biological meaning from this noisy and complex data, machine learning (ML) is required. ML algorithms, such as support vector machines (SVM), random forests (RF), and artificial neural networks (ANN), are designed to identify patterns in high-dimensional datasets. In the context of biology, ML has traditionally been applied to genomics, for example in gene annotation (Holst et al., 2025; Stanke & Waack, 2003). More recently, ANN-based methods have been developed for classification and segmentation of cells in microscopy images (Archit et al., 2025; R. Liu et al., 2022).

We propose that coupling high-throughput isolation, following a dilution-to-extinction approach in microplates, with autofluorescence-based classification of these cultures represents a novel approach to overcome the current limitations in culturomics. By training an ML-model on the spectral features of known reference strains, the algorithm can learn the non-linear patterns that distinguish one taxon from another. Once trained, the model can instantly classify unknown isolates based on their fluorescent spectra, effectively converting a complex physical measurement into a taxonomic label. This combination of intrinsic fluorescence spectroscopy and ML forms the basis of the high-throughput classification framework developed in this chapter.

3.2. Results

3.2.1. Overview of the framework

In liquid cultures, different bacterial species can have vastly distinct phenotypes. We hypothesized that these macroscopically visible differences are the result of distinct protein compositions, metabolic byproducts, and structural cellular components, all of which contribute to unique fluorescent signals. These qualitative observations suggested that different taxa possessed unique optical properties that could be systematically characterized.

3.2.2. Different bacterial taxa show different fluorescent fingerprints

To characterize the optical properties of the community members, we acquired spectral fingerprints for all 34 strains comprising the SynComs, which were used in the evolution experiment described in the previous chapter. Utilizing a standardized approach (**Figure 6A**), we obtained a dataset that highlights the distinct spectral profiles of these taxa (**Figure 6B-G**). Upon excitation at 330 nm, distinct spectral responses among different strains were evident. Flavobacteriaceae and Caulobacteraceae exhibited emission maxima at 390 nm, whereas Comamonadaceae showed a blue-shifted peak at 375 nm (**Figure 6B**). These variations likely captured differences in the cellular pyridoxine pool (Pöhlker et al., 2012). Alternatively, this could indicate a stronger relative contribution from tryptophan residues in the protein scaffold (Losi et al., 2004). Oxalobacteraceae were further distinguished by a shoulder at 420 nm, potentially representing the accumulation of 4-pyridoxic acid, a byproduct of rapid vitamin turnover (Bueno et al., 2004). Potential differences in metabolic state were further captured at 345 nm excitation, where Rhizobiaceae and Comamonadaceae displayed intensity profiles distinct from those of Flavobacteriaceae and Caulobacteraceae (**Figure 6C**). These shifts likely reflect variations in cytosolic NADH/NAD^+ ratios, offering a spectral snapshot of family-specific metabolic rates (Ince et al., 1992). In the visible spectrum, unique signatures that could be related to tetrapyrrole biosynthesis separated Streptomycetes and Rhizobiaceae from other taxa (Masahiko Taniguchi & Jonathan S. Lindsey, 2023). Specifically, Streptomycetes exhibited a distinct 615 nm emission upon 375 nm

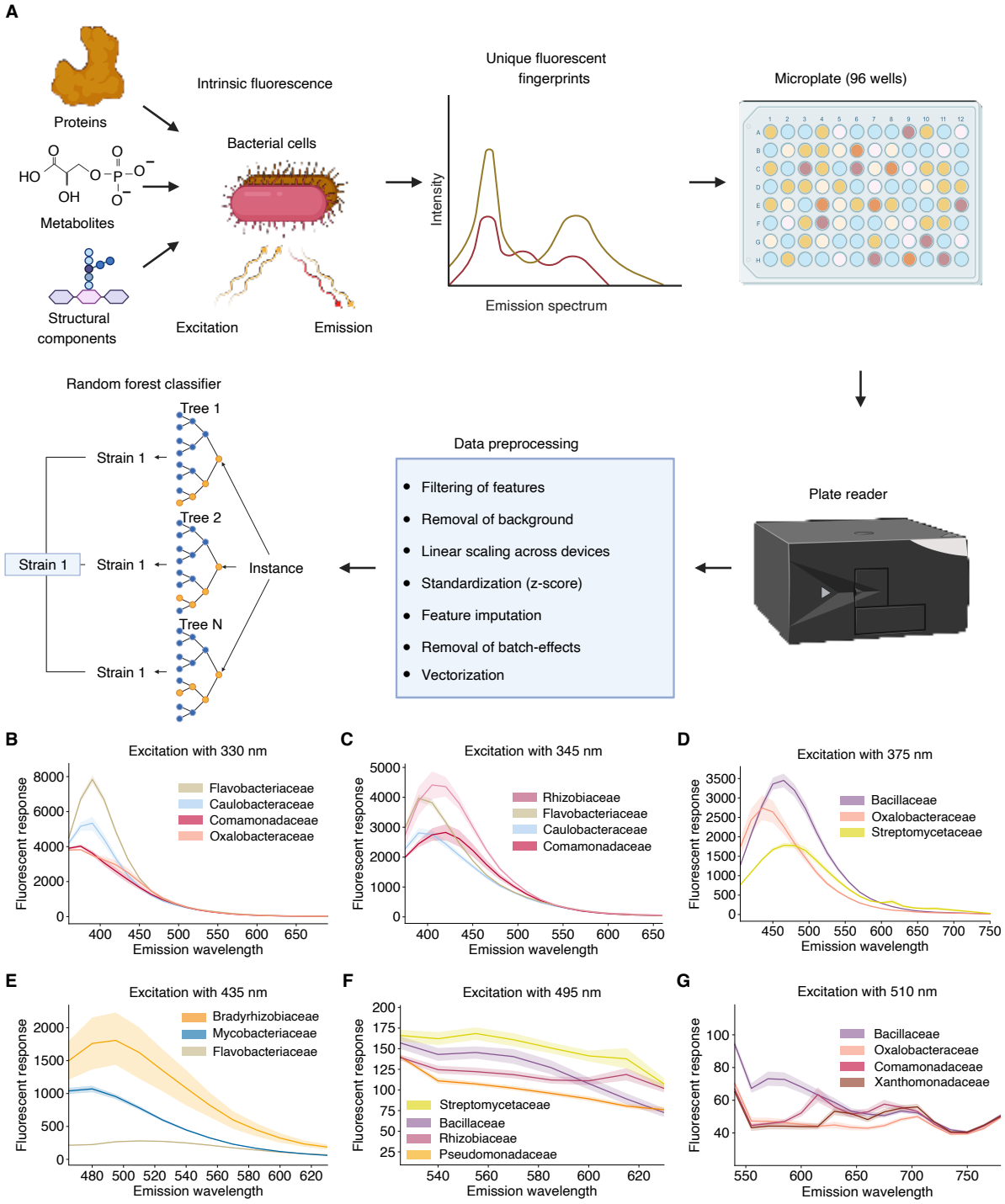

Figure 6. Differential spectral fingerprints. (**A**) Schematic overview of the workflow. Proteins, metabolites and structural components each contain unique fluorescent signatures. These distinct features could be used to scan newly established culture collections in microplates using a plate reader. A data processing pipeline is then applied to ensure proper data format and scale. Finally, the normalized data is used to perform predictions with machine learning models of different modality. (**B-G**) Example spectra of different strains highlighting differences between strains (n = 20-30).

excitation, which could be a consequence of coproporphyrin III-related compounds (Perez-Ortiz et al., 2023) (**Figure 6D**). While both Streptomycetes and Rhizobiaceae

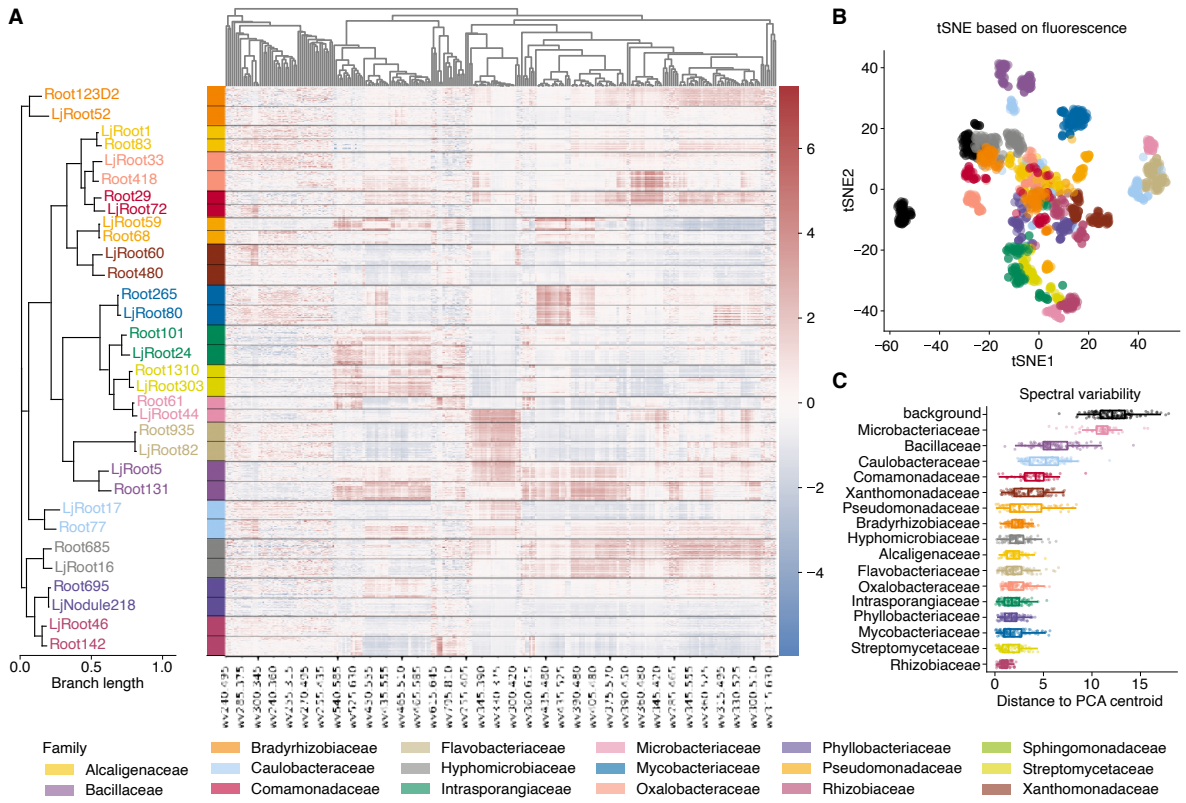

Figure 7. Diversity of intrinsic fluorescence profiles. (**A**) Heatmap of normalized spectral features (*z*-score; n_*features* = 343; *n* = 887), ordered by a maximum-likelihood phylogenetic tree based on concatenated AMPHORA genes. Tip labels are coloured by taxonomic family. (**B**) tSNE of spectral profiles (*n* = 887). Each point represents an event, coloured by family, showing robust clustering of taxonomic groups. (**C**) Intra-family spectral variability (*n* = 887). Boxplots display the variance of spectral profiles within each family, highlighting differences in phenotypic plasticity between taxa.

showed porphyrin-related Q-band signals at 495 nm excitation (**Figure 6F**), this signature was notably absent in Pseudomonas. Likely, the dominance of blue-green fluorescent pyoverdines in Pseudomonas masks the porphyrin intermediates (Ringel & Brüser, 2018). Finally, Bradyrhizobiaceae could be distinguished by a specific emission at 495 nm under 435 nm excitation (**Figure 6E**). This signal, completely undetected in groups like Flavobacteriaceae, is indicative of the presence of Coenzyme F420 (Eirich et al., 1979) or the riboflavin precursor lumazine (J. Lee et al., 1989). This finding is consistent with the enrichment of specialized cofactors essential for the Bradyrhizobiaceae symbiotic lifestyle (Dakora et al., 2015; Wunderer et al., 2024).

Integrating all acquired spectral data revealed differences among the 34 strains, even between closely related isolates belonging to the same family (**Figure 7A**). The heatmap analysis highlights unique, strain-specific spectral features, suggesting that

fluorescence fingerprinting may allow for discrimination at the strain level rather than being limited to the family level. This finding is further strengthened by the tSNE visualization (**Figure 7B**), where several families (color-coded) form distinct, separated clusters, indicative of significant intra-family variation. Despite this heterogeneity, strains from the same family generally retained a higher degree of similarity compared to those from different taxonomic groups, explaining their broad co-localization within the tSNE space. To quantify this observation, we calculated the spectral variability, defined as the distance of a measured fluorescent event from the centroid of its respective family following PCA transformation (**Figure 7C**). We found that families such as Microbacteriaceae and Bacillaceae exhibited higher intra-family variability, whereas others, including Streptomycetaceae and Rhizobiaceae, appeared more spectrally homogeneous. From these analyses, we concluded that bacterial strains can be effectively classified based on their fluorescent features. However, the predictive accuracy of such classification likely depends on the inherent intra-family homogeneity and the specific metabolic status of the populations at the time of measurement.

3.2.3. Time-efficient classification via information-rich spectral features

While full-spectrum fluorescence scanning yielded a high-dimensional dataset containing over 800 features, the measurement time is substantial, totaling approximately 5 hours per plate. To overcome this temporal bottleneck and enable high-throughput screening, we aimed to reduce the feature space without compromising classification accuracy. Following our initial broad-spectrum scans, we trained a random forest classifier to systematically evaluate feature contribution and identify the spectral variables holding the highest predictive power (**Figure 8A**). We chose a random forest (with 500 trees) as an initial model as they can be easily interpreted, are robust to overfitting, and work well out-of-the-box without extensive hyperparameter tuning. Based on this analysis, we selected the 50 most informative features for subsequent measurements (**Figure 8B**). To ensure robust downstream statistical processing, this informative set was padded with supplementary features to maintain a sufficient dataset size (**Figure 8B**). Analysis of feature importance revealed that the excitation wavelengths driving classification are biologically distinct. Excitation

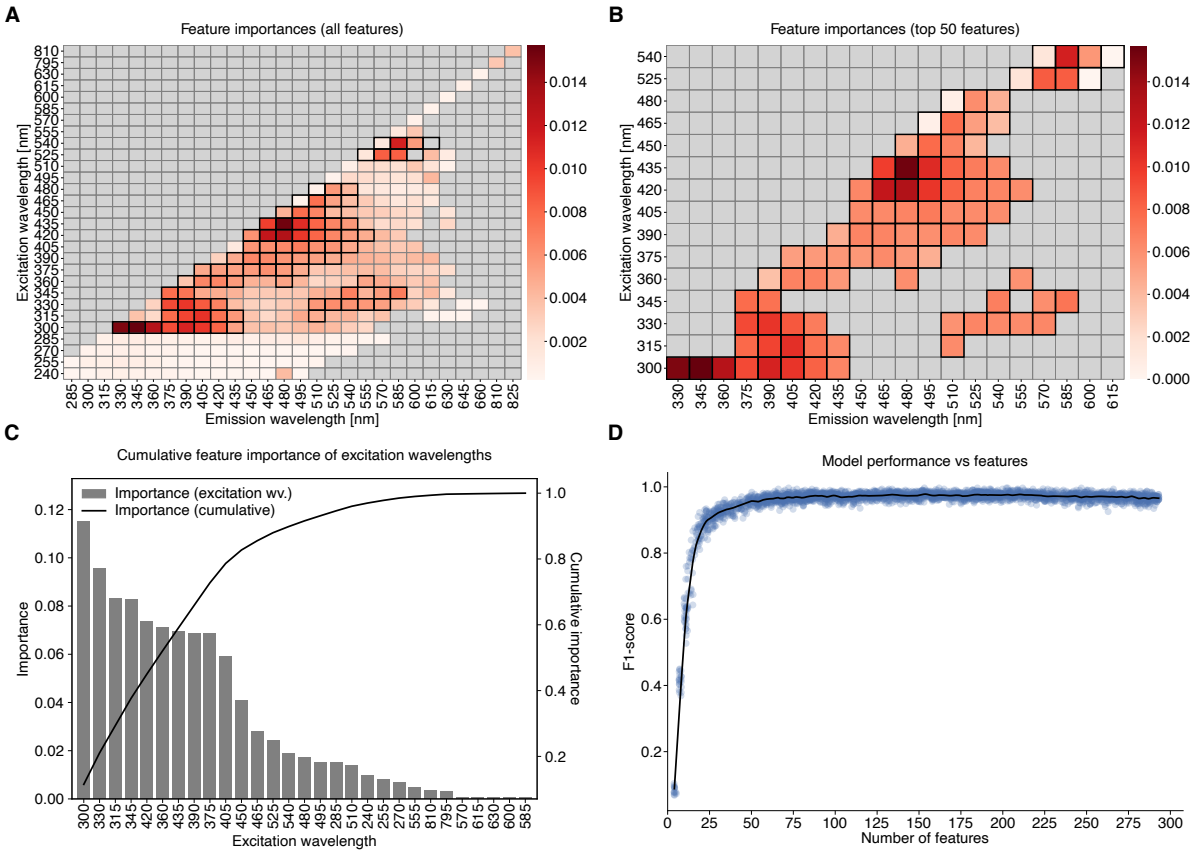

Figure 8. Feature importances of fluorescent spectra. (A-B) Heatmaps visualizing random forest feature importance mapped onto the excitation-emission matrix for the complete feature space (**A**) and the top 50 discriminative features (**B**). Colour intensity indicates the relative importance of each excitation-emission pair. (**C**) Feature importance aggregated by excitation wavelength. Grey bars represent the summed importance for each specific excitation wavelength, while the black line traces the cumulative importance across the ranked feature set. (**D**) Feature saturation curve illustrating the relationship between classification accuracy and feature dimensionality ($n = 2760$). Blue dots represent individual model evaluations using progressively larger feature subsets, fitted with a polynomial regression (black line).

at 300 nm provided the highest information gain, followed by 330 nm, 315 nm, and 345 nm. Cumulatively, these four excitation ranges account for approximately 50% of the total feature importance (**Figure 8C**). Biochemically, this suggests that the model primarily discriminates strains based on the ratio between structural components and metabolic cofactors. While the absorption maximum of tryptophan is 280 nm, excitation at 300 nm targets the red-shifted end of its spectrum (Lakowicz & Masters, 2008). This wavelength is often superior for whole-cell measurements as it mitigates the non-fluorescent absorption of UV light by DNA and RNA (Schmid, 2001). Consequently, the fluorescent response at 300 nm might be a better proxy for total protein content, ribosomal density, and cell size. As mentioned in previous sections,

the 315 nm - 345 nm range corresponds to metabolites such as vitamins and NADH/NAD^+ (Bueno et al., 2004; Chance et al., 1979; Ince et al., 1992; Pöhlker et al., 2012; Richards-Kortum & Sevick-Muraca, 1996). The high importance of these features suggests that the model effectively discriminates between strains based on their biosynthetic capabilities, intracellular storage, and respiratory activity. Strikingly, when feature importances are aggregated by excitation wavelength (**Figure 8C**), they follow a curve of logarithmic growth. This indicates that robust classification does not require the full spectrum but can be achieved with a reduced set of highly informative markers. To validate this, we performed an *in silico* feature inclusion analysis, successively adding features to the model starting from a single variable (**Figure 8D**). We observed a rapid increase in predictive accuracy that plateaued after approximately 50-75 features, after which the inclusion of additional spectral data yielded negligible improvements. Although it remains to be tested whether these specific features are universally applicable or specific to the strains tested here, these results confirm that a reduced spectral scan is sufficient for robust classification. This optimization reduces acquisition time significantly, facilitating the scaling of phenotypic screening to high-throughput formats.

3.2.4. Classification accuracy depends on model architecture and taxonomic diversity

While our initial analyses used a standard random forest classifier (scikit-learn), we wondered if alternative machine learning architectures could yield superior classification performance. To this end, we implemented a randomized grid search across five distinct model types: random forest, extra trees, a fully connected neural network, CatBoost, and LightGBM. For each architecture, we defined a specific hyperparameter search space (detailed in **3.4.7**) and evaluated 100 distinct configurations using 5-fold cross-validation, resulting in a total of 2,500 trained and validated models. We found that, on average, gradient boosting frameworks outperformed other architectures, with LightGBM yielding the highest accuracy, followed closely by CatBoost, random forests, and extra trees (**Figure 9A**). Strikingly, the fully connected neural network exhibited the lowest prediction accuracy. This underperformance may be the result of the sensitivity of neural networks to feature

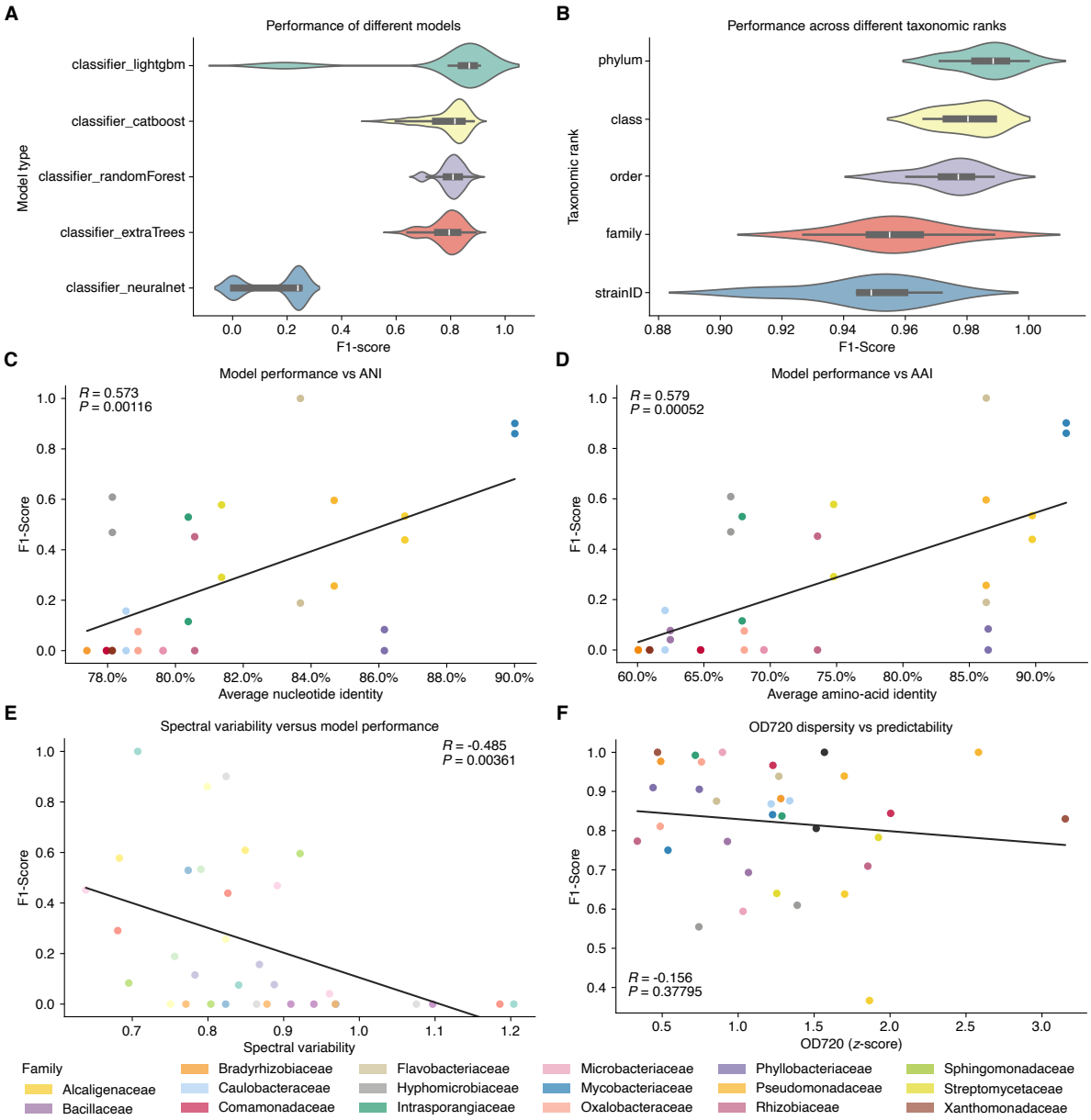

Figure 9. Benchmarking model performance and robustness. (**A**) Comparison of F1-scores across different machine learning architectures (n = 2500). (**B**) Classification performance of the optimized LightGBM model across taxonomic ranks (phylum to strain-level; n = 50). (**C-D**) Model generalization versus genomic distance. Classification F1-scores of the optimized LightGBM model are plotted against the average nucleotide identity (**C**; ANI; n = 29) and average amino acid identity (**D**; AAI; n = 32) between the training and testing sets. Higher identity values indicate closer genomic relatedness between the reference and query strains. (**E**) Impact of spectral variability on model performance (n = 34). The F1-score is plotted against the intra-family spectral variance. (**F**) Impact of growth heterogeneity on model performance (n = 34). The F1-score is plotted against the family-wise dispersity in optical density measurements. Note: For (C-F), each point represents an individual strain while the colour represents the taxonomic family. Performance metrics were derived from LOO-CV. All statistical associations represent Spearman rank correlations.

scaling, whereas tree-based ensemble methods are generally scale-agnostic. Furthermore, the architecture employed here consisted exclusively of dense layers.

State-of-the-art deep learning approaches typically utilize convolutional or transformer-based layers for improved capture of underlying patterns in complex data (LeCun et al., 2015; Vaswani et al., 2017). Given its consistently superior performance, the LightGBM framework was selected for all subsequent analyses. Using this optimized model, we next evaluated how classification performance varies across taxonomic ranks. For this, we tested the model at the strain, family, order, class, and phylum level. Unsurprisingly, the F1-score improved progressively at higher taxonomic ranks (**Figure 9B**). This trend was expected, as higher ranks contain fewer classes, thereby increasing the baseline probability of correct classification by random chance. Additionally, taxonomic conservation of specific biosynthetic pathways and growth dynamics could explain this improved performance. To investigate how the phylogenetic diversity of the training data impacts predictive capability, we performed leave-one-out (LOO) validation. In this setup, the dataset was split in a way that the test set contained a specific taxon, while the training set contained all other taxa. We then calculated the average nucleotide identity (ANI) and average amino-acid identity (AAI) between the test and training strains to determine if genetic relatedness determines classification success. We observed a significant positive correlation between the F1-score and both ANI (R = 0.573, P = 0.00116; **Figure 9C**) and AAI (R = 0.579, P = 0.00052; **Figure 9D**). This indicates that the model performs significantly better when a closely related strain is present in the training data, suggesting that the spectral features driving classification are taxonomically conserved. Consequently, the deployment of this method as a broad classification strategy would require a diverse training set that maximizes phylogenetic coverage. Conversely, we identified a significant negative correlation between the spectral variability of a family and its corresponding F1-score (R = –0.485, P = 0.00361; **Figure 9E**). This finding confirms that taxa exhibiting higher spectral heterogeneity are inherently more difficult to classify accurately. Finally, we assessed the robustness of the classifier against variations in bacterial biomass. Although the data processing pipeline includes normalization and standardization steps to mitigate artifacts arising from optical density differences, it was crucial to verify that the same strain could be accurately identified across varying optical densities. To evaluate this, we analysed the fluorescence profiles of bacterial strains cultured across a range of optical densities.

For this, we partitioned the dataset based on the per-taxon dispersity of OD measurements. The training set comprised the 90% of events closest to the mean OD_{720}, while the test set consisted of the 10% of events exhibiting the highest deviation (90th quantile of *z*-score transformed OD_{720}). Using this approach, we observed only a slightly negative, non-significant relationship between OD dispersity and the F1-score ($R = -0.156$, $P = 0.3779$; **Figure 9F**). This demonstrates that model performance is largely robust to variations in biomass, confirming the effectiveness of the data processing pipeline in removing OD-related artifacts. Collectively, these findings demonstrate that fluorescence-based spectral classification is a robust approach for bacterial identification. Importantly, for the 34-member synthetic community analysed here, the optimized taxon-specific classifier achieved an F1-score of approximately 0.95 (**Figure 9B**).

3.2.5. Fluorescent spectra are additive for multi-strain mixtures

Having established the utility of fluorescence fingerprinting for monocultures, we next investigated whether this framework could be adapted to decompose multi-strain mixtures. We hypothesized that the fluorescence spectra mixtures are additive. Mathematically, this means that when corrected for optical density, the spectra of a mixture correspond to the weighted mean of the individual monoculture spectra based on their relative abundances. To test this hypothesis, we generated binary co-cultures with varying mixing ratios (**Figure 10A-B**). A PCA of the resulting spectral data revealed that shifting the ratio between two strains generally resulted in a linear trajectory of samples within the spectral space. However, contrary to the expectation that this shift would align exclusively with the first principal component, we observed that a significant proportion of the variance was driven by replicate variability. This indicates that experimental noise introduces deviations from the idealized mixing gradient. To further validate the additivity hypothesis, we compared the empirical spectra of defined mixtures against theoretical spectra calculated by averaging the monoculture fluorescent responses (**Figure 10C-D**; grey line: expected, black line: measured). Generally, the measured spectra closely approximated the expected additive response, supporting the hypothesis that fluorescence signals are additive (**Figure 10C-D)**. However, notable deviations were observed in specific combinations.

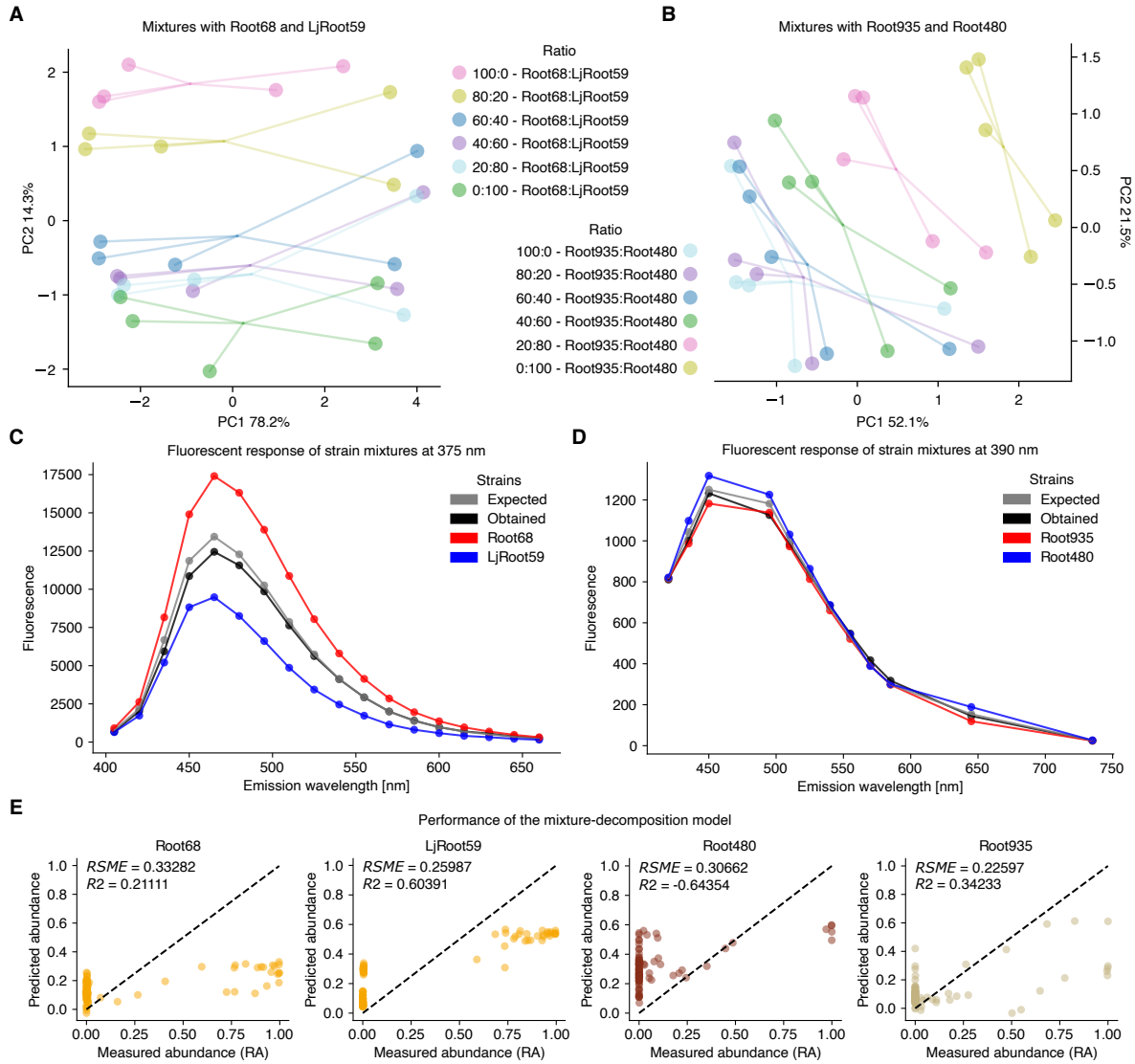

Figure 10. Assessment of spectral additivity in 2-strain scenarios. (**A-B**) PCA of trajectories of binary mixtures titrated at mixing ratios ranging from 100:0 to 0:100 in 20% increments (n = 30). (**C-D**) Representative spectra of binary mixtures. Blue and red lines denote the spectra of the individual monocultures. The grey line represents the theoretical spectrum calculated assuming strict linear additivity, while the black line shows the experimentally measured spectrum. (**E**) Accuracy of spectral decomposition of mixtures (n = 25-30). The scatter plot compares the relative abundance of strains in binary mixtures determined by 16S rRNA amplicon sequencing (observed) against the composition predicted by linear spectral decomposition (predicted). The black diagonal line indicates a perfect correlation.

For instance, in a mixture of Flavobacteriaceae (Root935) and Xanthomonadaceae (Root480), the spectral profile was disproportionately dominated by the Root935 signal (**Figure 10D**). This discrepancy suggests that while the additivity hypothesis holds up to a degree, specific inter-strain interactions may induce metabolic changes, such as the differential production of secondary metabolites, that alter the fluorescence phenotype in co-culture compared to isolation. Despite these biological complexities,

we proceeded under the assumption of spectral additivity to develop a decomposition model. For this, we generated a synthetic training dataset consisting of 10,000 *in silico* mixtures by randomly assigning relative abundances to pairs of strains and calculating their expected spectra under the assumption of perfect additivity. This dataset was used to train a multi-output regressor designed to predict strain abundances from complex spectral inputs. The trained regressor was then applied to the physically generated laboratory mixtures, utilizing 16S rRNA amplicon sequencing as the ground truth for relative abundance (**Figure 10E)**. Contrary to expectations, the model failed to accurately predict the community composition (**Figure 10E**). Performance metrics indicated poor predictive power, with the root mean square error (RMSE) ranging between 0.22 and 0.33, and the coefficient of determination (R^2) frequently being low or becoming negative.

Taken together, these results demonstrate that the fluorescence-based framework is currently insufficient for the reliable decomposition of microbial mixtures. This limitation likely stems from two factors. First, the issue of spectral redundancy, where distinct community compositions yield indistinguishable spectral profiles. Second, the physiological plasticity of bacteria, where inter-species interactions alter metabolic states and fluorescence outputs relative to monocultures (Conway et al., 2012; Morris et al., 2013). Consequently, accurate community decomposition will require alternative approaches or the integration of additional sensing modalities.

3.2.6. Case study: large-scale classification of evolved bacterial populations

Next, we aimed to use the framework for rapid identification of isolates recovered from the evolution experiment. The reisolation effort generated an extensive collection comprising 600 96-well plates (batches of 60 and 120 plates from specific time points). To accommodate this throughput, we restricted spectral scanning to the three most informative excitation wavelengths identified in our feature selection analysis: 300 nm, 315 nm, and 330 nm. Furthermore, to maximize precision, we trained separate, SynCom-specific models for the *At*-SC and *Lj*-SC, respectively. Since each SynCom contained only a single representative per bacterial family, separating the models by community effectively reduced the complexity of the classification problem. Model performance was evaluated using confusion matrices

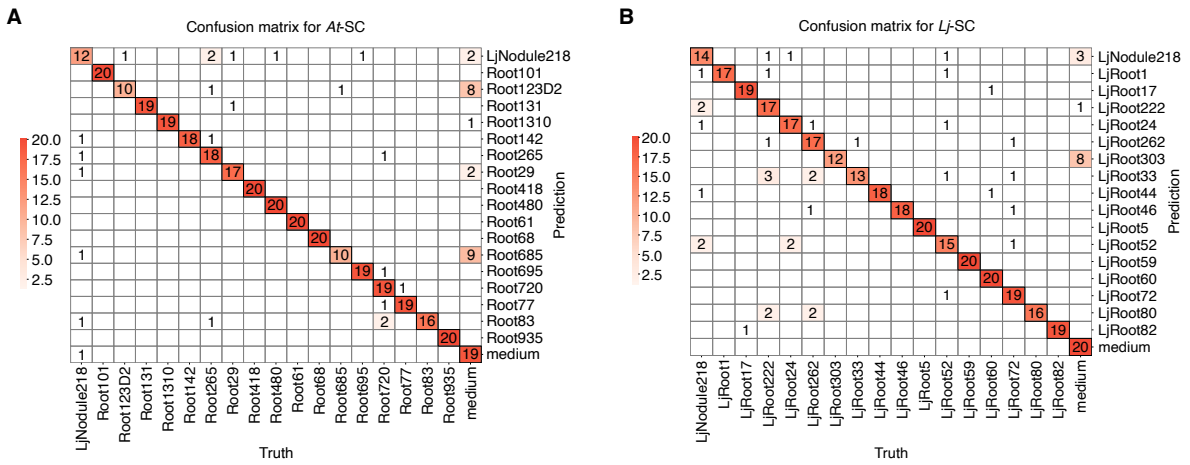

Figure 11. Strain-level classification performance across host-specific SynComs. Confusion matrices displaying the classification accuracy for two independent models trained on the *Arabidopsis thaliana* SynCom (*At*-SC; *n* = 380) (**A**) and the *Lotus japonicus* SynCom (*Lj*-SC; *n* = 360) (**B**). Rows represent the true strain identities, while columns represent the predicted labels. Colour intensity corresponds to the count of classification events, with the diagonal elements indicating correct predictions.

derived from a balanced validation set (**Figure 11A-B**). The classifiers demonstrated overall high accuracy, although a few instances of misclassification were observed. Notably, errors were prevalent among slow-growing strains, which were occasionally misclassified as sterile media. These false negatives likely result from low biomass accumulation, leading to weak fluorescent signals that are confounded by the background scattering of the culture medium. Next, we used these models to predict the identities of the strains in the culture collection of evolved isolates. This allowed us to virtually index the entire collection with strain identities. In total, we assigned taxonomic classifications to approximately 23,000 wells exhibiting bacterial growth (**Fig. S 6**). The efficiency of this spectral screening pipeline cannot be overstated, as it eliminated the need for exhaustive sequencing at the screening stage, allowing us to immediately target and recover specific taxa of interest for downstream characterization. To validate the biological relevance of this pipeline, we utilized the predicted labels to select specific isolates from the evolution experiment for phenotypic and genomic analysis. We were able to show an increase in colonization capacity of the evolved and reisolated strains compared to their ancestors (**Figure 5C**). Furthermore, we subjected 34 of these identified isolates to whole-genome sequencing (PacBio), enabling the identification of specific mutations, which are likely responsible for the observed fitness gains (**Table 9**).

3.2.7. Genomic signatures of host adaptation in evolved isolates

The genomic analysis of evolved isolates revealed distinct mutational patterns that likely underpin the observed increases in root colonization fitness (**Table 9**). These genetic changes appeared to converge on three primary adaptive strategies including surface remodeling for immune evasion, metabolic fine-tuning, and the modulation of lifestyle transitions between motility and sessile biofilm formation.

Root480 (Xanthomonadaceae), which exhibited the most substantial fitness gain among all evolved strains (**Figure 5C**), acquired two non-synonymous substitutions in central carbon metabolism, one in a glycosyltransferase and another in a glycogen synthase (**Table 9**). The glycosyltransferase mutation likely modifies surface structures such as lipopolysaccharides or exopolysaccharides, potentially camouflaging the bacterium from the *Lotus* immune system, thereby preventing pattern-triggered immunity (Newman et al., 2013; Ranf et al., 2015; Whitfield & Trent, 2014). The mutation in glycogen synthase points to a metabolic remodeling, potentially reflecting a reduced need for glycogen storage due to the specific carbon availability in the *Lotus* exudate profile, or a modification of the carbon flux towards these surface polymers.

Root142 acquired non-synonymous mutations in a diguanylate cyclase and the transcriptional repressor *rpsR* (**Table 9**). These modifications likely optimize the balance between physical adherence and host interaction. Specifically, the diguanylate cyclase mutation may alter intracellular cyclic di-GMP levels, promoting a stronger biofilm phenotype (Römling et al., 2023). Simultaneously, the mutation in the *rspAB* repressor suggests a modulated regulation of the Type III secretion system (T3SS). Potentially, this might reduce the secretion of immunogenic effectors to avoid triggering host defenses (Zamioudis & Pieterse, 2012). Together, these genetic changes allow the bacterium to enhance its persistence on the root surface while remaining undetected by the plant immune system.

Similarly, LjRoot60 (Xanthomonadaceae) demonstrated a shift towards a sessile lifestyle through frameshift mutations in *pilR* (Type IV pili) and *fliM* (flagellar motor switch) (**Table 9**). By abolishing these highly immunogenic motility structures, the bacterium likely avoids recognition by the pattern recognition receptors of the plant

immune system (Boller & Felix, 2009). This strategy is complemented by a mutation in glutamine synthase, which likely tunes the nitrogen assimilation to match the specific amino acid profile of the *Lotus* rhizosphere (Lesuffleur et al., 2007).

LjRoot262 featured mutations in *cysJ* (sulfite reductase) and the sensor kinase *ladS* (**Table 9**). Strikingly, these *cysJ* mutations were observed during evolution in the *Arabidopsis* rhizosphere, which is known to be rich in sulfur compounds (Bednarek et al., 2009; Strehmel et al., 2014). The disruption of *cysJ* points towards reductive evolution, as the loss of the energetically costly reduction of inorganic sulfate may confer a fitness advantage (Cole et al., 2017). Simultaneously, the loss of *ladS* indicates a shift in virulence and biofilm regulation via the *gacA*/*gacS* axis (E. Li, De Jonge, et al., 2021; Song et al., 2023).

Taken together, resolving these adaptations at the single-isolate level confirms that host-driven evolution is primarily characterized by immune evasion and metabolic optimization. Ultimately, the fluorescence-based screening of the culture collection not only facilitated mechanistic insights into the evolution experiment, but also served as a robust proof-of-concept, demonstrating the feasibility of characterizing microbial culture collections at an unprecedented scale.

Table 9: Genetic variation detected in evolved isolates. Strain name, host from the evolution experiment, gene name, description of gene function and mutational effect are provided. These strains were recovered and taxonomically classified based on fluorescent fingerprints. The presented strains represent a subset of the 34 evolved isolates that were selected based on their changes in colonization fitness. Genes without annotations were removed from the table and genes containing multiple mutations were summarized. The complete table can be accessed on https://github.com/bdn-227/flu_clf_thesis.

Strain	Evolution Host	Gene	Description	Effect
Root480	*L. japonicus*	K00786	Glycosyltransferase	Non-Synonymous
Root480	*L. japonicus*	K16150	Glycogen synthase	Non-Synonymous
LjRoot60	*A. thaliana*	*ccoN*	Cytochrome c oxidase cbb3-type subunit I	Upstream
LjRoot60	*A. thaliana*	*pilN*	Type IV pilus assembly protein PilN	Upstream
LjRoot60	*A. thaliana*	*pilR*,	Two-component system, NtrC family, response regulator PilR	Frame shift
LjRoot60	*A. thaliana*	*purK*	5-(carboxyamino)imidazole ribonucleotide synthase	Non-Synonymous
LjRoot60	*A. thaliana*	*corA*	Magnesium transporter	Non-Synonymous

LjRoot60	*A. thaliana*	*fliM*	Flagellar motor switch protein FliM	Frame shift
LjRoot60	*A. thaliana*	*glnA*	Glutamine synthetase	Non-Synonymous
LjRoot262	*A. thaliana*	*ladS*	Two-component system, sensor histidine kinase LadS	Frame shift
LjRoot262	*A. thaliana*	*cysJ*	Sulfite reductase	Many mutations
Root142	*L. japonicus*	*dgcB*	Diguanylate cyclase	Non-Synonymous
Root142	*L. japonicus*	*rpsD*	Small subunit ribosomal protein S4	Synonymous
Root142	*L. japonicus*	K02051	Transport system substrate-binding protein	Frame shift
Root142	*L. japonicus*	*rspR*	rspAB operon transcriptional repressor	Non-Synonymous

3.3. Discussion

In this chapter, we explored the feasibility of intrinsic fluorescence for the classification of bacterial cultures. We found differences among bacterial cultures within and across different taxonomic families. This led to the development of a classifier that could accurately predict the taxonomic identity of bacterial cultures (F1-score approximately 0.95). By applying this classifier to the culture collection of reisolated bacteria from the evolution experiment, we successfully classified over 23,000 wells with bacterial growth - a throughput unfeasible with conventional sequencing approaches. This large-scale indexing of the collection facilitated insights into bacterial adaptation within the rhizosphere, revealing that adaptation to a novel host plant is primarily driven by selection for immune evasion and metabolic fine-tuning.

The fact that the bacteria could be accurately classified in monoculture indicates that, at least for a 34-member community, there are sufficient metabolic differences to allow for robust classification (De Paulo et al., 2024; Giana et al., 2003; Sohn et al., 2009). Despite this accuracy, the framework faced limitations when applied to microbial mixtures (RMSE 0.22-0.33). This indicates that the hypothesis of spectral additivity of strain mixtures was insufficient. This discrepancy suggests that bacterial fluorescence is a dynamic phenotype influenced by inter-species interactions and nutritional status (Momeni et al., 2025; J. Porter et al., 1995). For instance, the spectral dominance of Flavobacteriaceae observed in co-cultures implies that bacteria alter their metabolic output, and consequently their spectral properties. This phenomenon recapitulates findings where metabolic profiles of co-cultures differ significantly from monocultures due to competition or cross-feeding (Ponomarova & Patil, 2015). Alternatively, spectral redundancy among mixtures could lead to confounding effects that impair spectral decomposition. Furthermore, we currently face limitations regarding sensitivity and resolution, as spectral data are inherently sensitive to experimental conditions such as media composition, pH, and growth phase (Allen et al., 2022; Hanson et al., 2019; Laflamme et al., 2011; Luan et al., 2026; Mukherjee et al., 2020).

The integration of spectral classification with high-throughput dilution-to-extinction isolation is, to our knowledge, a novel approach that represents a significant methodological advancement for large-scale culturomics. Furthermore, this framework can function as an anomaly detector. By establishing a database of known fluorescent fingerprints, wells exhibiting spectra that are distinct from the reference entries can be flagged as candidates for novel taxa. This enables the targeted recovery of previously uncultured strains, mirroring outlier-detection strategies used in flow cytometry to identify rare cell populations (Azad et al., 2010; Monaco et al., 2016).

The logical next step is to leverage non-linear models, such as ANNs, to robustly differentiate between taxa and their diverse, interaction-dependent metabolic states (H. Liu et al., 2023). Furthermore, the method could be extended to operate on the single cell level. We have begun to explore this possibility with promising preliminary results (F1-score of 0.66; **Appendix I**). Future efforts should focus on coupling spectral sorting with single-cell sequencing to link distinct fluorescent phenotypes directly to metagenomic bins. This reverse genomics approach could guide the design of strain-specific growth media (Cross et al., 2019). The democratization of this technology relies on the establishment of a global, open-access database of bacterial spectral signatures. Analogous to the SILVA or Greengenes repositories for 16S rRNA, such a resource would enable researchers to identify isolates across laboratories (DeSantis et al., 2006; Quast et al., 2013). Ultimately, establishment of strict metadata standards, which have already been established for Raman spectroscopy, could convert this framework to a robust classification platform for bacterial cultures at scale (K. S. Lee et al., 2024).

3.4. Material and methods

3.4.1. Data

Data for the training, validation, and testing were derived from a combination of existing culture collections. This dataset comprises previously described isolates from *Arabidopsis thaliana* (roots) (Bai et al., 2015), *Lotus japonicus* (roots) (Wippel et al., 2021), and the *Chlamydomonas reinhardtii* phycosphere (Durán et al., 2022). A complete compilation of the dataset is summarized in **Table 10**.

Table 10. Summary of the data used for training, validation and testing. Bacterial origin, measurement strategy, number of taxonomic families, features and devices are stated.

Host	Strategy	Families	Strains	Features	Devices
Arabidopsis (root)	flow cytometer	15	15	377	1
Chlamydomonas	flow cytometer	12	19	377	1
Lotus (root)	flow cytometer	14	14	377	1
Mixtures	flow cytometer	-	2	377	1
Arabidopsis (root)	plate reader	20	65	821	2
Lotus (root)	plate reader	18	66	821	2
Mixtures	plate reader	-	24	817	1

3.4.2. Bacterial growth conditions

To generate data for the training, validation, and testing of the classifier, bacterial strains were first recovered from frozen glycerol stocks by streaking onto TY-agar plates (5 g L^{-1} tryptone, 3 g L^{-1} yeast extract, 15 g L^{-1} agar, supplemented with 10 mM $CaCl_2$). The plates were incubated at room temperature for 7 days. Subsequently, single colonies were inoculated into liquid 0.5× TY medium (2.5 g L^{-1} tryptone, 1.5 g L^{-1} yeast extract, 10 mM $CaCl_2$) and incubated at 25°C with orbital shaking at 180 rpm for 7 days. For the preparation of the 96-well plates for spectral scanning, the liquid cultures were adjusted to an optical density (OD_{600}) of 0.1 in fresh 0.5× TY-medium. Each well was inoculated with 160 μL of the standardized cell suspension and sealed with micropore tape (3M). Following an initial 7-day incubation period, the plates were scanned using a TECAN Infinite 200 PRO plate reader. Upon completion of the

spectral scanning, a 10 μL aliquot was sampled from each well for sequencing and stored at -20°C until further processing.

3.4.3. Plate reader measurements

Spectral measurements were conducted using four distinct protocols, categorized by their software interface and spectral resolution. Individual plates were processed using the Tecan iControl software, while batches of two or more plates were measured using the EvoWare framework. The latter utilized an automated plate-stacking robot to allow for high-throughput, unsupervised data acquisition. For both the manual and automated modes, two distinct measurement protocols were used. Measurements were performed across the entire available excitation range, with all corresponding emission wavelengths recorded. The measurement space was defined to strictly avoid primary light scattering. For this, emission wavelengths (λ_{em}) were recorded in 15 nm increments, beginning at the first discrete step greater than $\lambda_{ex} + 15\,nm$ (effectively $\lambda_{ex} + 30\,nm$) and extending to a maximum of 840 nm. Consequently, all emission data satisfying the condition $\lambda_{em} \leq \lambda_{ex} + 15nm$ were excluded from the analysis. From this, we derived the total set of emission wavelengths (E) for any excitation wavelength λ_{ex} as detailed in Equation 6.

$$E_{\lambda_{ex}} = \{\lambda_{ex} + 15n \mid \in \{2, 3, \ldots\}, \lambda_{ex} + 15n \leq 840nm\} \quad (6)$$

Due to the high number of data points, a full scan required approximately 5 hours per plate. Reduced feature scans were performed to accommodate large sample volumes when measurement time was a limiting factor. These reduced protocols were developed to target a subset of the most informative spectral features, thereby allowing a measurement time of approximately 20 minutes per plate. Specifically, fluorescence was recorded across the full emission spectrum using excitation wavelengths of 300 nm, 315 nm, and 330 nm. The data output format was dependent on the software interface. Measurements performed via Tecan iControl were exported as Excel files, whereas data obtained using the EvoWare framework were exported as plain ASCII files for downstream computational processing.

3.4.4. Processing of plate reader data

Raw outputs from the plate reader were parsed using a custom pipeline implemented in Python. Data manipulation was performed primarily using the Polars library v1.12.0[1] for high-performance dataframe operations. Excel files generated by the Tecan iControl software were converted into a tidy data format while preserving essential metadata, including filename, timestamp, device ID, measurement mode (absorbance or fluorescence), excitation/emission wavelengths, temperature, kinetic time points, and well coordinates. For datasets acquired via the EvoWare framework, individual ASCII files, which are generated for each discrete measurement, were parsed chronologically and concatenated to form a dataframe containing the same metadata. The tidy dataframes were subjected to a multi-step preprocessing pipeline to ensure compatibility with machine learning frameworks. For this, a library of specialized functions was implemented to standardize the spectral features, thus ensuring signal quality and comparability across samples. First, feature filtering was applied to restrict the dataset to the intersection of excitation and emission wavelength pairs recorded across all individual events. To ensure signal integrity, measurements falling outside the linear dynamic range of the instrument were removed. Specifically, negative fluorescent responses and values flagged as saturation events ("OVER") were excluded from the analysis. To account for background noise and media auto-fluorescence, background standardization was performed using linear scaling. For this, an optical density threshold, $OD_{720} \leq 0.09$, was empirically determined to identify medium-only control wells. For each excitation-emission pair, raw fluorescence values were standardized by dividing them by the mean intensity of these corresponding medium controls. Subsequently, to reduce data skewness and improve the stability of predictive ML-models across several orders of magnitude, an inverse hyperbolic sine (arcsinh) transformation was applied to all standardized values (Finak et al., 2010). To ensure the classification pipeline remained agnostic to the underlying bacterial cell density (OD_{720}), the data were normalized. While both area under the curve (AUC) and *z*-score normalization were identified as effective strategies, *z*-score normalization

[1] https://github.com/pola-rs/polars

was prioritized for large-scale datasets due to its significantly lower computational overhead. Finally, datasets were refined to handle missing values and technical variation. To address gaps resulting from the initial filtering steps, a per-event imputation strategy was implemented. For this, missing values were substituted using statistical descriptors (mean, median, minimum, maximum, zero) or a specific flag value (10^{10}) depending on the downstream application. To mitigate systematic technical bias, batch effect removal was performed using a Python re-implementation of the ComBat algorithm (originally from the sva package), which corrects for variations introduced by measurement date or the specific instrument used (Leek et al., 2014). The training dataset was subjected to stringent quality control to ensure high-level model performance. Outliers were identified in a taxa-specific manner by grouping spectral events sharing an identical 16S rRNA sequence (V5-V7 region). Following dimensionality reduction via principal component analysis (PCA), the resulting principal component scores were *z*-score transformed. Any event with a transformed score greater than three was classified as a biological or technical outlier and removed from the final training set (Saha et al., 2016). Finally, data types were standardized, with feature columns formatted as floating-point numbers and dependent variables (taxonomic labels) as strings.

3.4.5. DNA extraction and library preparation

Taxonomic verification of the bacterial isolates, and the determination of microbial composition in mixed-strain experiments, were achieved through 16S rRNA gene amplicon sequencing. DNA was extracted from the microplate cultures using an alkaline lysis protocol adapted from previously established protocols (J. Zhang et al., 2021). For this procedure, a 10 μL aliquot of each bacterial culture was transferred to a 96-well PCR plate and mixed with 17 μL of alkaline lysis buffer containing 25 mM NaOH and 0.2 mM Na_2-EDTA at pH 12. After incubation for 30 minutes at 95°C, 17 μl of a neutralization buffer containing 40 mM Tris-HCl at pH 7.5 was added. The sequencing libraries were generated via a two-step PCR approach targeting the V5-V7 hypervariable region of the 16S rRNA gene. During the primary PCR, the target region was amplified using the degenerate primers 799F (5'-AACMGGATTAGATACCCKG-3') and 1192R (5'-ACGTCATCCCCACCTTCC-3') (Bai

et al., 2015; Wippel et al., 2021; J. Zhang et al., 2021). The reaction began with an initial denaturation at 94°C for 2 minutes, followed by 25 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds, and extension at 72°C for 30 seconds, finishing with a final extension at 72°C for 10 minutes (Durán et al., 2018; J. Zhang et al., 2021). The resulting amplicons were verified for correct size using gel electrophoresis. Prior to the second PCR step, the PCR products were enzymatically digested by incubation with 1 μL of Antarctic Phosphatase, 1 μL of Exonuclease I, and 2.44 μL of the corresponding Antarctic Phosphatase buffer. Next, 3 μL of the digested PCR-product served as the template for the second PCR step, which incorporated sample-specific barcodes and Illumina-compatible adapters. This amplification phase utilized a shortened protocol of 10 cycles consisting of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, following an initial 2-minute denaturation at 94°C and ending with a 10-minute extension at 72°C. Following a final verification of the indexed amplicons by gel electrophoresis, the samples were pooled and the library was purified via gel extraction using the QIAquick Gel Extraction Kit (QIAGEN). Lastly, the library was purified using AMPure XP beads according to the manufacturer's instructions. DNA concentrations were quantified using the QuantiFluor (Promega), and final sequencing was conducted on an in-house Illumina MiSeq platform.

3.4.6. Processing of amplicon sequencing data

The raw sequencing data obtained from the MiSeq platform were initially demultiplexed using QIIME2 v2021.11.0 (Bolyen et al., 2018). To reconstruct the full-length hypervariable region, forward and reverse reads were merged using FLASH v2.2.00 (Magoč & Salzberg, 2011) allowing for a maximum overlap of 250 nucleotides. Following merging, any sequences containing undefined nucleotides were filtered out using Usearch v10.0.240 (Edgar, 2010) to ensure high data quality. The resulting sequences were then subjected to denoising and error correction using the QIIME2 implementations of DADA2 and vsearch (Bolyen et al., 2018; Callahan et al., 2016; Rognes et al., 2016). Taxonomic classification of these ASVs was performed using the pre-trained Naive bayes classifier integrated within the QIIME2 framework (Bolyen et al., 2018). In the final step of the pipeline, sequences originating from non-target DNA, such as chloroplasts and mitochondria, were excluded. The refined feature table

was then exported from the QIIME2 environment for downstream statistical analysis. Only wells with a purity exceeding 0.95, defined as the dominant ASV representing at least 95% of the total relative abundance, were retained for downstream analysis.

3.4.7. Training and optimization of predictive models

The normalized spectral data, paired with the taxonomic identities confirmed via amplicon sequencing, served as the foundation for training several ML models. To facilitate rapid prototyping and ensure seamless comparability across different computational frameworks, a custom BaseClassifier class was implemented in Python. This modular architecture allowed for the systematic evaluation of various algorithms, including multi-output random forests (Pedregosa et al., 2011), extra trees (Pedregosa et al., 2011), support vector machines (Pedregosa et al., 2011), XGBoost (Chen & Guestrin, 2016), CatBoost (Dorogush et al., 2018), and LightGBM (Shi et al., 2025), as well as a single-output neural network (Pedregosa et al., 2011). To streamline the workflow, specialized methods were developed to handle variable input and output dimensions natively, alongside a comprehensive set of evaluation functions. This modular approach streamlined the extensive testing of both model architectures and preprocessing schemes by calculating key performance metrics, specifically the accuracy score, F1-score, Matthews Correlation Coefficient (MCC), as well as confusion matrices. Model evaluations were primarily performed using the F1-score, as it provides a balanced metric of precision and recall. It is calculated based on the number of true positives (TP), false positives (FP), and false negatives (FN), as defined in Equation 7 (Pedregosa et al., 2011).

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP - FN)} \tag{7}$$

Identifying informative features was essential to streamline the scanning process, thereby increasing measurement throughput while maintaining high classification accuracies. Therefore, a random forest classifier was trained using repeated 5-fold cross-validation (CV) for 10 iterations. From these trained classifiers the feature importances were inferred and aggregated across all folds and repeats. This yielded a comprehensive set of highly-informative features that was used for

spectral scans with a reduced feature set (**Figure 8A-C**). To prove accurate classification performance of the framework, iterative feature addition was performed (**Figure 8D**)**.** The model was initially fitted using the single most informative feature. In subsequent iterations, the next most informative feature was added sequentially until the model was trained on the complete feature set.

The optimization of the data preprocessing pipeline was implemented with an exhaustive grid search, where the spectral data were processed using every possible combination of parameters. The performance of each configuration was validated using 5-fold CV to determine the resulting accuracy, F1-score, and MCC. To quantify the specific contribution of each preprocessing parameter to the classification F1-score, a linear least-squares regressor with L2 regularization was employed (Pedregosa et al., 2011). In this meta-analysis, the preprocessing parameters functioned as predictors, while the model was trained to predict the F1-score of the resulting taxonomic classifier. This methodology enabled the extraction of regression coefficients for each parameter, providing a clear statistical measure of their relative impact on overall classification performance. For the downstream analysis, we used the parameter configuration that had the highest F1-score (summarized in **Table 11**).

Table 11. Summary of parameters for data preprocessing. Parameter name, description and the parameter setting used for the presented results.

Parameter name	Parameter description	Setting
filter_common_features	Subset features that are present in all individual events	True
include_negative	Retain negative responses	False
include_over	Retain values outside linear space	False
scale_by_medium	Strategy to normalize with medium background	“estimated”
arcsinh	Inverse hyperbolic sine	False
normalize_by	Strategy for standardization	“scale” (*z*-score)
impute_strategy	Strategy for filling of missing values after standardization	False
batches	Metadata used for batch effect correction	["date", "device"]

Additionally, a screen across different model architectures was performed to find the best performing model and a corresponding set of optimized hyperparameters. Therefore, a search was performed across 5 model types and 100 hyperparameter

configurations using 5-fold CV. The model types and hyperparameter space that was tested is summarized in **Table 12**.

Table 12. Hyperparameters space for different model types. Model type, name of the hyperparameter and the search space are provided.

Model type	Hyperparameter	Search space
Random forest	n_estimators	X~U (100, 1501)
Random forest	criterion	["gini", "entropy"]
Random forest	max_features	[None , "sqrt", "log2", 0.2, 0.5, 0.8]
Random forest	max_depth	None or X~U (3, 101)
Random forest	min_samples_split	X~U (2, 51)
Random forest	min_samples_leaf	X~U (1, 21)
Random forest	bootstrap	[True, False]
SVM	kernel	["linear", "poly", "rbf", "sigmoid"]
SVM	C	[10^{-3}, 10^{-2}, …, 10^{3}]
SVM	gamma	["scale", "auto", 10^{-4}, 10^{-2}, .., 10^{1}]
SVM	class_weight	[None, "balanced"]
Xgboost	max_depth	[3, 5, 7, 9]
Xgboost	min_child_weight	[1, 3, 5, 7]
Xgboost	subsample	[0.6, 0.8, 1.0]
Xgboost	colsample_bytree	[0.6, 0.8, 1.0]
Xgboost	gamma	[0, 0.1, 0.5, 1]
Xgboost	reg_lambda	[0.5, 1, 5]
Xgboost	reg_alpha	[0, 0.1, 1]
Xgboost	learning_rate	[0.01, 0.05, 0.1]
Xgboost	n_estimators	[200, 500, 1000]
Neural net	hidden_layer_sizes	[50, 100, (100,50), (100,100,50)]
Neural net	activation	["relu", "tanh"]
Neural net	alpha	[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}]
Neural net	solver	["adam", "lbfgs"]
Neural net	learning_rate_init	[10^{-5}, 10^{-4}, 10^{-3}]
Neural net	max_iter	[300, 500, 800]
lightGBM	learning_rate	[0.01, 0.03, 0.05, 0.1]
lightGBM	n_estimators	[100, 200, 500, 1000]
lightGBM	num_leaves	[31, 63, 127]
lightGBM	min_child_samples	[10, 30, 100]
lightGBM	feature_fraction	[0.6, 0.8, 1.0]
lightGBM	bagging_fraction	[0.6, 0.8, 1.0]
catBoost	learning_rate	[None, 0.01, 0.03, 0.05, 0.1]
catBoost	depth	[None, 4, 6, 8]
catBoost	l2_leaf_reg	[None, 1, 3, 10]
catBoost	bagging_temperature	[None, 0, 0.5, 1.0]
catBoost	iterations	[None, 100, 200, 500]

catBoost	early_stopping_rounds	[None, 20, 50, 100]

3.4.8. Evaluation of predictive models

To evaluate model performance across different taxonomic levels, the classifier was trained and tested specifically at each rank. To assess performance relative to taxonomic novelty, the average nucleotide identity (ANI) using FastANI v1.34 (Jain et al., 2018) and average amino-acid identity (AAI) using EzAAI v1.2.4 (Kim et al., 2021) were calculated. These tools generated a pairwise distance matrix containing ANI/AAI values for all strain pairs. Model performance was tested using leave-one-out (LOO) cross-validation. For this, the distance between the test strain and its closest relative in the training set was derived from the ANI/AAI matrix. Finally, Spearman's rank correlation was calculated to quantify the relationship between these genomic distances and the model's predictive performance. To evaluate the influence of OD on model performance, the data were partitioned according to OD deviation. For this, OD values measured at 720 nm (od720.720) were *z*-score normalized to quantify the distance of each observation from the mean in standard deviation units. For each strain, the 90th percentile (q_{90}) of the *z*-scores was calculated. The dataset was then split based on this threshold, where observations with an absolute *z*-score ≥ q_{90} were assigned to the test set, while the remaining observations (*z*-score < q_{90}) constituted the training and validation set. This approach ensured that the test set contained samples that diverged most significantly from the mean OD_{720} in a strain-specific manner, allowing us to assess whether taxonomic classification performance is robust to strong gradients in cell density. The relationship between model performance and mean OD dispersion was assessed using Spearman's rank correlation.

3.4.9. Decomposition of mixtures

Monocultures were grown to saturation and mixed by volume (v/v), creating a concentration gradient ranging from 100:0 to 0:100 in 20% increments. These mixtures were subjected to full-spectral scanning using a TECAN plate reader. Initial analysis indicated that mixed spectra were not perfectly additive across the entire measured range. To optimize the input for the mixture-decomposition models, the downstream

analysis was restricted to excitation wavelengths that exhibited robust linear additivity: 345, 360, 375, 390, 405, 420, and 450 nm. The additivity of these features was verified via PCA of the processed spectra with data processing identical to **3.4.7** (**Figure 10A-B**). Training data for the decomposition models was generated from the available spectra of the monocultures. For this, 10,000 random binary strain pairings were generated. Mixing ratios were sampled from a random uniform distribution such that the sum of the relative abundances equaled 1. Next, fluorescent spectra were obtained by calculating the relative abundance-weighted mean of the constituent monocultures at the selected wavelengths (**Figure 10C-D**). As regression framework, CatBoost was utilized due to its state-of-the-art performance with default settings. To enable the simultaneous prediction of abundances for multiple strains, the CatBoost regressor was wrapped in a MultiOutputRegressor framework (scikit-learn). The decomposition model was trained using the *in silico* generated mixtures and tested on the experimentally generated two-strain mixtures described above. Performance was benchmarked against amplicon sequencing data, which served as the ground truth for relative abundance. The model accuracy was evaluated using mean squared error (MSE), root mean square error (RMSE), and the coefficient of determination (R^2), calculated both globally and on a per-class basis.

3.4.10. Data visualization

The schematic overviews presented in this chapter were created using BioRender. All data visualizations were exclusively generated within the Python environment, primarily utilizing the Matplotlib v3.9.2 (Hunter, 2007) and Seaborn v0.13.2 (Waskom, 2021) libraries. For the integration of phylogenetic and spectral data in **Figure 7A**, a phylogenetic tree was first constructed using PhyloPhlAn v3.1.1 (Asnicar et al., 2020) and FastTree v2.2 (Price et al., 2009). The resulting tree was visualized and rendered using the Phylo module of the Biopython package v1.85 (Cock et al., 2009). To explore the high-dimensional spectral space, t-distributed stochastic neighbor embedding (tSNE) was performed as shown in **Figure 7B**. This dimensionality reduction was calculated using the scikit-learn implementation, with the perplexity parameter dynamically set to $perplexity = round(\sqrt{N_{features}})$, where $N_{features}$ represents the total number of features.

3.4.11. Code and data availability

The parsed spectral datasets and the corresponding ASV tables generated during this study have been made publicly available to ensure reproducibility. These data, along with the associated metadata, can be accessed via the dedicated GitHub repository at https://github.com/bdn-227/flu_clf_thesis. The computational framework developed for this thesis, including the custom BaseClassifier architecture, data processing pipelines, and auxiliary functions, is implemented in the BacPy2 Python package. The complete source code and documentation for this package are available at https://github.com/bdn-227/bacpy2.

3.4.12. Declaration regarding the use of generative AI

For the preparation of this dissertation, the generative AI model Gemini (Google) was utilized exclusively for the purposes of linguistic optimization, proofreading of grammar and spelling, and improving sentence structure. I hereby declare that the scientific content, the development of hypotheses, the interpretation of data, and the formulation of all conclusions are entirely my own intellectual work. I have verified all AI-generated suggestions and accept full responsibility for the final content of this thesis.

Chapter 4

Discussion

"Essentially, all models are wrong, but some are useful."

George E. P. Box

4. Discussion

Microbial communities residing at the root-soil interface are dynamic ecosystems that are essential for plant health. While the ecological processes driving the microbiota assembly are increasingly well understood, the evolutionary dynamics that shape these communities remain elusive. In this thesis, we present a study that combines experimental evolution with a novel high-throughput screening framework to address this gap. Specifically, we experimentally evolved taxonomically diverse synthetic bacterial communities by repeated passaging through the rhizosphere of two plant host species, and tracked evolutionary changes using whole-population sequencing and competition-based fitness assays. This approach allowed us to link host-specific selection to reproducible genetic changes and measurable gains in community-level colonization performance.

4.1. Evolutionary plasticity of the root microbiota

This work was motivated by previous evolution experiments conducted with single strains or low-complexity communities in unstructured environments (Barrick et al., 2009; Castledine et al., 2020; Conrad et al., 2011; Couce et al., 2024; Evans et al., 2020; Fiegna et al., 2015; Hall et al., 2018; Lang et al., 2013; Summers et al., 2010; Tenaillon et al., 2016; Turkarslan et al., 2021). We extended this approach by employing taxonomically diverse synthetic communities (SynComs) within the structured environment of the plant root. This experimental framework allowed us to examine not only whether root-associated microbial communities evolve, but also how host species and community context shape the repeatability and genetic basis of that evolution. The results presented in this work highlight the remarkable evolutionary plasticity of the root microbiota, challenging the view of the microbiome as a static assembly of soil-derived taxa. Instead, our findings suggest that the root microbiota is a dynamic entity capable of rapid adaptation, driven by a complex interplay of host selection and community dynamics. The observation that communities displayed similar genomic trajectories with overlapping mutation targets provides strong evidence that the host plant acts as a dominant and deterministic evolutionary force. This phenomenon, often termed parallel evolution, suggests that the selective

pressures exerted by the root environment, such as specific immune responses or exudate profiles, are uniform and strong enough to drive independent bacterial lineages toward similar genotypic solutions (Bailey et al., 2015, 2017). This aligns with recent experimental evolution studies showing that host-mediated selection can rapidly reshape the genetic architecture of commensal bacteria, often targeting genes involved in motility, metabolism, and cell surface regulation to enhance root colonization (Conrad et al., 2011; Couce et al., 2024; E. Li, Zhang, et al., 2021; Morella et al., 2020; Tenaillon et al., 2016). The presented data indicate that adaptation is primarily characterized by the fine-tuning of existing gene networks rather than the acquisition of novel functions through horizontal gene transfer or large-scale genomic rearrangements. This pattern of optimization suggests that the ancestral bacteria already possess the fundamental metabolic repertoire required for root survival, but lack the regulatory fine-tuning for optimal fitness (Barrick & Lenski, 2013). By tweaking global regulators and catalytic enzymes, bacteria can rapidly optimize nutrient uptake or immune evasion without the high metabolic cost of maintaining new genomic islands (Levy et al., 2018). Additionally, evolutionary time is a critical factor here, as the emergence of novel biochemical functions is a rare event that typically requires timescales significantly longer (Blount et al., 2008). A striking finding was that bacterial populations evolving on non-native hosts accumulated and fixed mutations more rapidly than those on native hosts. This can be interpreted using Fisher's Geometric Model of adaptation (Matuszewski et al., 2014). Bacteria on non-native hosts are likely further from their fitness optimum, resulting in a stronger selection pressure that favours adaptive mutations to bridge the fitness gap (Tenaillon, 2014). In contrast, strains on native hosts, presumably having a long history of co-adaptation, are closer to their fitness peak, where stabilizing selection prevails.

Our experimental design allowed us to link genetic changes to phenotypic outcomes. Using a series of competition assays, we demonstrate that SynComs evolved on a novel plant species consistently outperform ancestral communities, indicating increased bacterial fitness. Notably, strains with high initial abundance exhibited the largest gains in fitness over the course of the experiment. This link between ecological abundance and adaptive trajectories suggests a winner-takes-it-all (or winner-takes-more-than-the-rest) dynamic in the root microbiome. Strains that

can initially colonize at high densities have a larger population size, which could increase the supply of beneficial mutations and thereby increase the efficiency of selection (Lanfear et al., 2014). This creates a positive feedback loop, further solidifying the dominance of specific taxa in the rhizosphere. The initial competition for the root surface allows early adapting strains to exclude competitors, effectively monopolizing the niche. As demonstrated in recent studies using SynComs, the order of arrival and the speed of initial adaptation can determine the final community structure, which would give early dominant strains an evolutionary head start (Fukami, 2015; Wippel et al., 2021). Finally, it is crucial to consider the potential cost of specialization. The observed fine-tuning to the root environment may come with a trade-off, reducing the fitness of the respective strain in the bulk soil or on alternative hosts. Rapid adaptation to the root niche often involves the downregulation or loss of gene function required for survival in the harsh, nutrient-poor soil environment (Finkel et al., 2020; Levy et al., 2018; E. Li, De Jonge, et al., 2021; Zhalnina et al., 2018). This trade-off is central to the bacterial life cycle, suggesting that while adaptation enhances competitiveness on the root, it may impact community resilience and composition in the absence of a host plant.

4.2. The role of community context in shaping evolutionary outcomes

Building on the experimental evolution results presented in Chapter 2 and their phenotypic validation, an important next question is how strongly these evolutionary outcomes depend on the surrounding microbial community context. In a monoculture setting, bacteria experience the absence of interspecific competitors. Under these conditions, evolutionary theory suggests that the population would evolve towards a generalist phenotype, maximizing the uptake of all available host exudates to reach the highest possible carrying capacity (Kassen, 2002). In contrast, within a community of bacteria, the presence of competitors likely forces strains into an existing niche, driving ecological displacement. Here, selection favours mutations that minimize niche overlap, such as specializing on specific substrates that competitors cannot utilize efficiently (Lawrence et al., 2012). Consequently, we speculate that in the absence of a community, there would have been a broader range of metabolic adaptations, whereas the community context likely channelled evolution towards metabolic

specialization to avoid resource competition. Furthermore, in mixed communities, bacteria must allocate resources not only for growth but also for competition such as secretion systems, antibiotic production, and resistance, to defend their niche (Foster & Bell, 2012). In a monoculture, these selective pressures are absent. Consequently, bacteria evolving in isolation often downregulate or lose these costly aggressive traits to redirect energy toward faster growth, a phenomenon observed in experimental evolution studies where cheaters arise in the absence of enemies (Foster & Bell, 2012; Frank, 2010; E. Li, De Jonge, et al., 2021). Thus, the genomic stability of defence genes observed in this study might be a direct consequence of the surrounding community, which acted as a stabilizing selection force and thereby prevented the loss of costly but necessary competitive traits (Granato et al., 2019; Scheuerl et al., 2020). Finally, the surrounding community fundamentally alters the fitness landscape on which these bacteria evolve. This concept implies that fitness peaks and valleys are not static host features but dynamic targets dependent on the community structure (Scheuring & Yu, 2012). Consequently, the parallel evolution we observed across replicates may not just reflect host determinism, but also a consistent set of biotic constraints imposed by the specific composition of our SynCom. To elucidate the effect of the surrounding community in more detail, further research that utilizes many different communities is required.

4.3. Limitations of the study

While this study provides significant insights into the short-term evolutionary dynamics of the root microbiota, several limitations in the experimental design must be considered. A notable observation in our data was the lack of horizontal gene transfer (HGT). This is likely a consequence of the phylogenetic structure of our SynComs, which were composed of a single representative from each bacterial family. In natural soil, HGT is frequently driven by homologous recombination between closely related strains (Thomas & Nielsen, 2005). The phylogenetic distance between members of our SynCom likely acted as a barrier to genetic exchange, effectively limiting the evolutionary potential to vertical inheritance. Consequently, the presented study may underestimate the rate of adaptation in natural settings, where mobile genetic elements facilitate rapid trait acquisition (Smillie et al., 2011). Although the use

of SynComs is necessary for experimental tractability, these communities lack the enormous biological complexity of natural microbiomes. Furthermore, our gnotobiotic system excluded different ecological interaction partners commonly found in soil such as bacteriophages, protists, and nematodes (Armon, 2010; Geisen et al., 2018; Martins et al., 2022). In natural ecosystems, these predators exert strong evolutionary pressures, often referred to as kill-the-winner dynamics, which can prevent any single taxon from dominating (Gao et al., 2019; Koskella & Brockhurst, 2014; Winter et al., 2010). Despite the controlled environment, disentangling the specific drivers of the observed evolution remains a challenge. In the rhizosphere, selective pressures are a composite of host immune responses (Lebeis et al., 2015), nutrient availability (Sasse et al., 2018), and interspecies competition (Hassani et al., 2018; Mesny et al., 2024; Snelders et al., 2020). It remains difficult to distinguish whether a specific mutation was fixed to evade the host immune system, to utilize a specific resource, or to outcompete other bacteria. Additionally, the gnotobiotic system may induce lab artifacts, where bacteria adapt to the experimental setup itself rather than to the host. Taken together, while we observed rapid adaptation in a host-specific manner, we must be cautious in assuming these exact genomic trajectories would play out identically in the field.

4.4. Future Perspectives

The central conceptual takeaway of this work is that host-associated microbiomes are not static entities but evolutionarily flexible and capable of rapid adaptation over short timescales. This realization suggests a paradigm shift in how we approach microbiome engineering. Rather than relying solely on genetic modification using transgenes and thereby producing genetically modified organisms (GMOs), we can harness experimental evolution to train microbial communities. By utilizing the framework described in this thesis, beneficial bacteria can be experimentally evolved to optimize their compatibility with specific crop cultivars. Through this, one could effectively generate non-GMO bioinoculants tailored for modern agriculture (Mueller & Sachs, 2015). Future efforts should focus on identifying a core set of strains that are already approved for field use and possess proven plant growth-promoting traits. These strains could then be trained through experimental evolution in order to increase

their colonization fitness for a particular crop variety, soil type, or even fertilizer regime (Kaminsky et al., 2019). However, to achieve this, the current experimental framework must be streamlined to allow for high scalability and standardization.

A critical biological challenge remains, as we must rigorously prove that adaptation to the host does not come at the cost of beneficial functions. Evolutionary theory suggests that cheating behaviour is widespread (Sandoz et al., 2007). Ultimately, this could result in strains evolving to maximize their own fitness by sacrificing costly cooperative traits, such as nitrogen fixation or nutrient delivery (Kiers et al., 2003; Sachs et al., 2004). Hence, it is plausible that bioinoculants could evolve away from their intended function over time. Therefore, future work must determine if improved root and rhizosphere competence comes with a cost for plant growth-promoting traits. If this were the case, maintenance of these traits might require co-selection regimes ensuring that colonization success remains coupled to beneficial service (S. S. Porter & Simms, 2014; Sachs et al., 2004). While the findings presented in this thesis are promising, the ultimate test lies in translation to the field. Future research must validate that lab-evolved host preference confers a competitive advantage in agricultural soil. Additionally, functional validation of the candidate genes under field conditions is a necessary step to strengthen the findings of this study.

Finally, for these findings to impact real-world agriculture, industrial constraints must be considered early in the development of bioinoculants. Strains selected for commercialization must be fermentable at ultra-high densities to allow production at industrial scale. Consequently, some taxa that perform well in the rhizosphere may be unsuitable candidates if they become genetically unstable or lose viability under the stressful conditions of industrial bioreactors (Herrmann & Lesueur, 2013; Rugbjerg & Sommer, 2019). Bridging the gap between evolutionary biology and bioprocess engineering will be essential in translating these adapted microbes into viable agricultural products.

5. Conclusions

Together, our results provide experimental evidence that root microbiomes can evolve rapidly and reproducibly in response to host selection, leading to substantial gains in fitness. Our study further identifies a set of candidate genes and functions that may mediate adaptation to a new plant species and will be the focus of future functional validation. Finally, our experimental approach establishes a framework for investigating the evolution of complex microbial communities in association with a eukaryotic host, enabling systematic tests of the principles driving microbiome evolution.

6. References

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D., & Kemen, E. M. (2016). Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology*, *14*(1), e1002352.

Allen, D. M., Einarsson, G. G., Tunney, M. M., & Bell, S. E. J. (2022). Characterization of Bacteria Using Surface-Enhanced Raman Spectroscopy (SERS): Influence of Microbiological Factors on the SERS Spectra. *Analytical Chemistry*, *94*(26), 9327–9335. https://doi.org/10.1021/acs.analchem.2c00817

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Amann, R. I., Ludwig, W., & Schleifer, K.-H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), 143–169.

Ammor, M. S. (2007). Recent Advances in the Use of Intrinsic Fluorescence for Bacterial Identification and Characterization. *Journal of Fluorescence*, *17*(5), 455–459. https://doi.org/10.1007/s10895-007-0180-6

Archit, A., Freckmann, L., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Teuber, C., Spitzner, M., Tapia Contreras, C., Buckley, G., von Haaren, S., Gupta, S., Grade, M., Wirth, M., Schneider, G., Dengel, A., Ahmed, S., & Pape, C. (2025). Segment Anything for Microscopy. *Nature Methods*, *22*(3), 579–591. https://doi.org/10.1038/s41592-024-02580-4

Armanhi, J. S. L., de Souza, R. S. C., Damasceno, N. de B., de Araújo, L. M., Imperial, J., & Arruda, P. (2018). A Community-Based Culture Collection for Targeting Novel Plant Growth-Promoting Bacteria from the Sugarcane Microbiome. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2017.02191/full

Armon, R. (2010). Soil bacteria and bacteriophages. In *Biocommunication in soil microorganisms* (pp. 67–112). Springer.

Asnicar, F., Thomas, A. M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., & May, U. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications*, *11*(1), 2500.

Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology*, *4*(8), 577–587. https://doi.org/10.1038/nrmicro1460

Azad, A., Langguth, J., Fang, Y., Qi, A., & Pothen, A. (2010). Identifying Rare Cell Populations in Comparative Flow Cytometry. In V. Moulton & M. Singh (Eds.), *Algorithms in Bioinformatics* (pp. 162–175). Springer. https://doi.org/10.1007/978-3-642-15294-8_14

Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P. C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A. C., Vorholt, J. A., & Schulze-Lefert, P. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*, *528*(7582), 364–369. https://doi.org/10.1038/nature16192

Bailey, S. F., Blanquart, F., Bataillon, T., & Kassen, R. (2017). What drives parallel evolution? *BioEssays*, *39*(1), e201600176. https://doi.org/10.1002/bies.201600176

Bailey, S. F., Rodrigue, N., & Kassen, R. (2015). The Effect of Selection Environment on the Probability of Parallel Evolution. *Molecular Biology and Evolution*, *32*(6), 1436–1448. https://doi.org/10.1093/molbev/msv033

Barnett, M., Meister, L., & Rainey, P. B. (2025). Experimental evolution of evolvability. *Science*, *387*(6736), eadr2756. https://doi.org/10.1126/science.adr2756

Barraclough, T. G. (2015). How do species interactions affect evolutionary dynamics across whole communities? *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 25–48.

Barrick, J. E., & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics*, *14*(12), 827–839. https://doi.org/10.1038/nrg3564

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, *461*(7268), 1243–1247. https://doi.org/10.1038/nature08480

Batstone, R. T., O'Brien, A. M., Harrison, T. L., & Frederickson, M. E. (2020). Experimental evolution makes microbes more cooperative with their local host genotype. *Science*, *370*(6515), 476–478. https://doi.org/10.1126/science.abb7222

Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubsky, J., Mansurova, M., Humphry, M., Consonni, C., Panstruga, R., & Sanchez-Vallet, A. (2009). A glucosinolate

metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*, *323*(5910), 101–106.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.

Berendsen, R. L., Vismans, G., Yu, K., Song, Y., de Jonge, R., Burgman, W. P., Burmølle, M., Herschend, J., Bakker, P. A. H. M., & Pieterse, C. M. J. (2018). Disease-induced assemblage of a plant-beneficial bacterial consortium. *The ISME Journal*, *12*(6), 1496–1507. https://doi.org/10.1038/s41396-018-0093-1

Blount, Z. D., Borland, C. Z., & Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proceedings of the National Academy of Sciences*, *105*(23), 7899–7906.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Boller, T., & Felix, G. (2009). A renaissance of elicitors: Perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology*, *60*, 379–406.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., & others. (2018). *QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science*.

Braga, P. A. C., Tata, A., dos Santos, V. G., Barreiro, J. R., Schwab, N. V., dos Santos, M. V., Eberlin, M. N., & Ferreira, C. R. (2013). Bacterial identification: From the agar plate to the mass spectrometer. *Rsc Advances*, *3*(4), 994–1008.

Brockhurst, M. A., Colegrave, N., Hodgson, D. J., & Buckling, A. (2007). Niche occupation limits adaptive radiation in experimental microcosms. *PLoS One*, *2*(2), e193.

Brown, C. T., Fishwick, L. K., Chokshi, B. M., Cuff, M. A., Jackson, J. M., Oglesby, T., Rioux, A. T., Rodriguez, E., Stupp, G. S., Trupp, A. H., Woollcombe-Clarke, J. S., Wright, T. N., Zaragoza, W. J., Drew, J. C., Triplett, E. W., & Nicholson, W. L. (2011). Whole-Genome Sequencing and Phenotypic Analysis of Bacillus subtilis Mutants following Evolution under Conditions of

Relaxed Selection for Sporulation. *Applied and Environmental Microbiology*, *77*(19), 6867–6877. https://doi.org/10.1128/AEM.05272-11

Bueno, C., Guerrero, J., & Encinas, M. V. (2004). Spectroscopic Properties of 4-Pyridoxic Acid as a Function of pH and Solvent. *Helvetica Chimica Acta*, *87*(4), 940–948. https://doi.org/10.1002/hlca.200490087

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., & Schmelzer, E. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature*, *488*(7409), 91–95.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology*, *64*(1), 807–838. https://doi.org/10.1146/annurev-arplant-050312-120106

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583.

Carrión, V. J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., Hollander, M. de, Ruiz-Buck, D., Mendes, L. W., Ijcken, W. F. J. van, Gomez-Exposito, R., Elsayed, S. S., Mohanraju, P., Arifah, A., Oost, J. van der, Paulson, J. N., Mendes, R., Wezel, G. P. van, Medema, M. H., & Raaijmakers, J. M. (2019). Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science*, *366*(6465), 606–612. https://doi.org/10.1126/science.aaw9285

Castledine, M., Padfield, D., & Buckling, A. (2020). Experimental (co)evolution in a multi-species microbial community results in local maladaptation. *Ecology Letters*, *23*(11), 1673–1681. https://doi.org/10.1111/ele.13599

Chaisson, M. J., & Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics*, *13*(1), 238.

Challis, R. (2014). *Rjchallis/assembly-stats 17.02*. https://doi.org/10.5281/zenodo.322347

Chance, B., Schoener, B., Oshino, R., Itshak, F., & Nakase, Y. (1979). Oxidation-reduction ratio studies of mitochondria in freeze-trapped samples. NADH and flavoprotein fluorescence signals. *The Journal of Biological Chemistry*, *254*(11), 4764–4771.

Chaparro, J. M., Badri, D. V., & Vivanco, J. M. (2014). Rhizosphere microbiome assemblage is affected by plant development. *The ISME Journal*, *8*(4), 790–803.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics*, *38*(23), 5315–5316.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170–175.

Chesneau, G., Herpell, J., Garrido-Oter, R., & Hacquard, S. (2025). From synthetic communities to synthetic ecosystems: Exploring causalities in plant–microbe–environment interactions. *New Phytologist*, *245*(2), 496–502. https://doi.org/10.1111/nph.20250

Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods*, *20*(8), 1203–1212.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118}; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & others. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422.

Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R. R., Deutschbauer, A. M., Dangl, J. L., & Visel, A. (2017). Genome-wide identification of bacterial plant colonization genes. *PLOS Biology*, *15*(9), e2002860. https://doi.org/10.1371/journal.pbio.2002860

Connon, S. A., & Giovannoni, S. J. (2002). High-Throughput Methods for Culturing Microorganisms in Very-Low-Nutrient Media Yield Diverse New Marine Isolates. *Applied and Environmental Microbiology*.

Conrad, T. M., Lewis, N. E., & Palsson, B. Ø. (2011). Microbial laboratory evolution in the era of genome-scale science. *Molecular Systems Biology*, *7*(1), 509. https://doi.org/10.1038/msb.2011.42

Conway, C. A., Esiobu, N., & Lopez, J. V. (2012). Co-Cultures of Pseudomonas aeruginosa and Roseobacter denitrificans Reveal Shifts in Gene Expression Levels Compared to Solo Cultures. *The Scientific World Journal*, *2012*(1), 120108. https://doi.org/10.1100/2012/120108

Couce, A., Limdi, A., Magnan, M., Owen, S. V., Herren, C. M., Lenski, R. E., Tenaillon, O., & Baym, M. (2024). Changing fitness effects of mutations through long-term bacterial evolution. *Science*, *383*(6681), eadd1417. https://doi.org/10.1126/science.add1417

Cross, K. L., Campbell, J. H., Balachandran, M., Campbell, A. G., Cooper, C. J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., Yang, Z., Parks, J. M., & Podar, M. (2019). Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nature Biotechnology*, *37*(11), 1314–1321. https://doi.org/10.1038/s41587-019-0260-6

Dakora, F. D., Matiru, V., & Kanu, A. S. (2015). Rhizosphere ecology of lumichrome and riboflavin, two bacterial signal molecules eliciting developmental changes in plants. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2015.00700/full

Davis, K. M., & Isberg, R. R. (2016). Defining heterogeneity within bacterial populations via single cell approaches. *BioEssays*, *38*(8), 782–790. https://doi.org/10.1002/bies.201500121

De Paulo, E. H., Magalhães, G. B., Moreira, M. P. B., Nascimento, M. H. C., Heringer, O. A., Filgueiras, P. R., & Ferrão, M. F. (2024). Classification of water by bacterial presence using chemometrics associated with excitation-emission matrix fluorescence spectroscopy. *Microchemical Journal*, *197*, 109804. https://doi.org/10.1016/j.microc.2023.109804

De Vries, F. T., Griffiths, R. I., Knight, C. G., Nicolitch, O., & Williams, A. (2020). Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science*, *368*(6488), 270–274. https://doi.org/10.1126/science.aaz5192

Deatherage, D. E., Kepner, J. L., Bennett, A. F., Lenski, R. E., & Barrick, J. E. (2017). Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. *Proceedings of the National Academy of Sciences*, *114*(10). https://doi.org/10.1073/pnas.1616132114

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*.

D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., & Lewis, K. (2010). Siderophores from Neighboring Organisms Promote the Growth of Uncultured Bacteria. *Chemistry & Biology*, *17*(3), 254–264. https://doi.org/10.1016/j.chembiol.2010.02.010

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). *CatBoost: Gradient boosting with categorical features support* (No. arXiv:1810.11363). arXiv. https://doi.org/10.48550/arXiv.1810.11363

Durán, P., Flores-Uribe, J., Wippel, K., Zhang, P., Guan, R., Melkonian, B., Melkonian, M., & Garrido-Oter, R. (2022). Shared features and reciprocal complementation of the Chlamydomonas and Arabidopsis microbiota. *Nature Communications*, *13*(1), 406. https://doi.org/10.1038/s41467-022-28055-8

Durán, P., Thiergart, T., Garrido-Oter, R., Agler, M., Kemen, E., Schulze-Lefert, P., & Hacquard, S. (2018). Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. *Cell*, *175*(4), 973-983.e14. https://doi.org/10.1016/j.cell.2018.10.020

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461.

Eirich, L. D., Vogels, G. D., & Wolfe, R. S. (1979). Distribution of coenzyme F420 and properties of its hydrolytic fragments. *Journal of Bacteriology*, *140*(1), 20–27. https://doi.org/10.1128/jb.140.1.20-27.1979

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, *297*(5584), 1183–1186.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.

Evans, R., Beckerman, A. P., Wright, R. C. T., McQueen-Mason, S., Bruce, N. C., & Brockhurst, M. A. (2020). Eco-evolutionary Dynamics Set the Tempo and Trajectory of Metabolic Evolution in Multispecies Communities. *Current Biology*, *30*(24), 4984-4988.e4. https://doi.org/10.1016/j.cub.2020.09.028

Fiegna, F., Moreno-Letelier, A., Bell, T., & Barraclough, T. G. (2015). Evolution of species interactions determines microbial community productivity in new environments. *The ISME Journal*, *9*(5), 1235–1245. https://doi.org/10.1038/ismej.2014.215

Fierer, N. (2017). Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579–590.

Finak, G., Perez, J.-M., Weng, A., & Gottardo, R. (2010). Optimizing transformations for automated, high throughput analysis of flow cytometry data. *BMC Bioinformatics*, *11*(1), 546. https://doi.org/10.1186/1471-2105-11-546

Finkel, O. M., Salas-González, I., Castrillo, G., Conway, J. M., Law, T. F., Teixeira, P. J. P. L., Wilson, E. D., Fitzpatrick, C. R., Jones, C. D., & Dangl, J. L. (2020). A single bacterial genus maintains root growth in a complex microbiome. *Nature*, *587*(7832), Article 7832. https://doi.org/10.1038/s41586-020-2778-7

Fitzpatrick, C. R., Salas-González, I., Conway, J. M., Finkel, O. M., Gilbert, S., Russ, D., Teixeira, P. J. P. L., & Dangl, J. L. (2020). The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annual Review of Microbiology*, *74*(1), 81–100. https://doi.org/10.1146/annurev-micro-022620-014327

Foster, K. R., & Bell, T. (2012). Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species. *Current Biology*, *22*(19), 1845–1850. https://doi.org/10.1016/j.cub.2012.08.005

Frank, S. A. (2010). Microbial secretor–cheater dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2515–2522. https://doi.org/10.1098/rstb.2010.0003

Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, *46*(Volume 46, 2015), 1–23. https://doi.org/10.1146/annurev-ecolsys-110411-160340

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., & Jousset, A. (2019). Protists: Puppet Masters of the Rhizosphere Microbiome. *Trends in Plant Science*, *24*(2), 165–176. https://doi.org/10.1016/j.tplants.2018.10.011

Geisen, S., Mitchell, E. A., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L. D., Jousset, A., Krashevska, V., & Singer, D. (2018). Soil protists: A fertile frontier in soil biology research. *FEMS Microbiology Reviews*, *42*(3), 293–323.

Gerrish, P. J., & Lenski, R. E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, *102*(0), 127–144.

Getzke, F., Wang, L., Chesneau, G., Böhringer, N., Mesny, F., Denissen, N., Wesseler, H., Adisa, P. T., Marner, M., & Schulze-Lefert, P. (2024). Physiochemical interaction between osmotic stress and a bacterial exometabolite promotes plant disease. *Nature Communications*, *15*(1), 4438.

Giana, H. E., Silveira, L., Zângaro, R. A., & Pacheco, M. T. T. (2003). Rapid Identification of Bacterial Species by Fluorescence Spectroscopy and Classification Through Principal Components Analysis. *Journal of Fluorescence*, *13*(6), 489–493. https://doi.org/10.1023/B:JOFL.0000008059.74052.3c

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., & Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature*, *551*(7678), 45–50. https://doi.org/10.1038/nature24287

Goyal, A., Bittleston, L. S., Leventhal, G. E., Lu, L., & Cordero, O. X. (2022). Interactions between strains govern the eco-evolutionary dynamics of microbial communities. *eLife*, *11*, e74987. https://doi.org/10.7554/eLife.74987

Granato, E. T., Meiller-Legrand, T. A., & Foster, K. R. (2019). The Evolution and Ecology of Bacterial Warfare. *Current Biology*, *29*(11), R521–R537. https://doi.org/10.1016/j.cub.2019.04.024

Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., McHardy, A. C., Dangl, J. L., Knight, R., & Ley, R. (2015). Microbiota and host nutrition across plant and animal kingdoms. *Cell Host & Microbe*, *17*(5), 603–616.

Halkier, B. A., & Gershenzon, J. (2006). BIOLOGY AND BIOCHEMISTRY OF GLUCOSINOLATES. *Annual Review of Plant Biology*, *57*(1), 303–333. https://doi.org/10.1146/annurev.arplant.57.032905.105228

Hall, J. P. J., Harrison, E., & Brockhurst, M. A. (2018). Competitive species interactions constrain abiotic adaptation in a bacterial soil community. *Evolution Letters*, *2*(6), 580–589. https://doi.org/10.1002/evl3.83

Hanson, C., Bishop, M. M., Barney, J. T., & Vargis, E. (2019). Effect of growth media and phase on Raman spectra and discrimination of mycobacteria. *Journal of Biophotonics*, *12*(11), e201900150. https://doi.org/10.1002/jbio.201900150

Harbort, C. J., Hashimoto, M., Inoue, H., Niu, Y., Guan, R., Rombolà, A. D., Kopriva, S., Voges, M. J. E. E. E., Sattely, E. S., Garrido-Oter, R., & Schulze-Lefert, P. (2020). Root-Secreted Coumarins and the Microbiota Interact to Improve Iron Nutrition in Arabidopsis. *Cell Host & Microbe*, *28*(6), 825-837.e6. https://doi.org/10.1016/j.chom.2020.09.006

Hassani, M. A., Durán, P., & Hacquard, S. (2018). Microbial interactions within the plant holobiont. *Microbiome*, *6*(1), 58.

Herrmann, L., & Lesueur, D. (2013). Challenges of formulation and quality of biofertilizers for successful inoculation. *Applied Microbiology and Biotechnology*, *97*(20), 8859–8873. https://doi.org/10.1007/s00253-013-5228-8

Hibbing, M. E., Fuqua, C., Parsek, M. R., & Peterson, S. B. (2010). Bacterial competition: Surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, *8*(1), 15–25.

Hindré, T., Knibbe, C., Beslon, G., & Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, *10*(5), 352–365. https://doi.org/10.1038/nrmicro2750

Hirsch, A. M. (2004). Plant-microbe symbioses: A continuum from commensalism to parasitism. *Symbiosis*, *37*(1–3), 345–363.

Holst, F., Bolger, A. M., Kindel, F., Günther, C., Maß, J., Triesch, S., Kiel, N., Saadat, N., Ebenhöh, O., Usadel, B., Schwacke, R., Weber, A. P. M., Bolger, M. E., & Denton, A. K. (2025). Helixer: Ab initio prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. *Nature Methods*, 1–8. https://doi.org/10.1038/s41592-025-02939-1

Hunt, M., Silva, N. D., Otto, T. D., Parkhill, J., Keane, J. A., & Harris, S. R. (2015). Circlator: Automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, *16*(1), 294.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Ince, C., Coremans, J. M. C. C., & Bruining, H. A. (1992). In Vivo NADH Fluorescence. In W. Erdmann & D. F. Bruley (Eds.), *Oxygen Transport to Tissue XIV* (pp. 277–296). Springer US. https://doi.org/10.1007/978-1-4615-3428-0_30

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114.

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, *72*(3), 1719–1728.

Jarvis, R. M., & Goodacre, R. (2008). Characterisation and identification of bacteria using SERS. *Chemical Society Reviews*, *37*(5), 931–936.

Jessup, C. M., Kassen, R., Forde, S. E., Kerr, B., Buckling, A., Rainey, P. B., & Bohannan, B. J. (2004). Big questions, small worlds: Microbial model systems in ecology. *Trends in Ecology & Evolution*, *19*(4), 189–197.

Johnson, N. C., Graham, J. H., & Smith, F. A. (1997). Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *The New Phytologist*, *135*(4), 575–585.

Joseph, T. A., Chlenski, P., Litman, A., Korem, T., & Pe'er, I. (2022). Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. *Genome Research*, *32*(3), 558–568. https://doi.org/10.1101/gr.275533.121

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Küsel, K., Rillig, M. C., Rivett, D. W., Salles, J. F., van der Heijden, M. G. A., Youssef, N. H., Zhang, X., Wei, Z., & Hol, W. H. G. (2017). Where less may be more: How the rare biosphere pulls ecosystems strings. *The ISME Journal*, *11*(4), 853–862. https://doi.org/10.1038/ismej.2016.174

Kaeberlein, T., Lewis, K., & Epstein, S. S. (2002). Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science*, *296*(5570), 1127–1129. https://doi.org/10.1126/science.1070633

Kaminsky, L. M., Trexler, R. V., Malik, R. J., Hockett, K. L., & Bell, T. H. (2019). The Inherent Conflicts in Developing Soil Microbial Inoculants. *Trends in Biotechnology*, *37*(2), 140–151. https://doi.org/10.1016/j.tibtech.2018.11.011

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.

Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology*, *15*(2), 173–190. https://doi.org/10.1046/j.1420-9101.2002.00377.x

Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume–rhizobium mutualism. *Nature*, *425*(6953), 78–81.

Kim, D., Park, S., & Chun, J. (2021). Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. *Journal of Microbiology*, *59*(5), 476–480.

Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T. P. L., & Pevzner, P. A. (2020). metaFlye: Scalable long-read metagenome assembly using repeat graphs. *Nature Methods*, *17*(11), Article 11. https://doi.org/10.1038/s41592-020-00971-x

Kopriva, S., Rahimzadeh Karvansara, P., & Takahashi, H. (2024). Adaptive modifications in plant sulfur metabolism over evolutionary time. *Journal of Experimental Botany*, *75*(16), 4697–4711. https://doi.org/10.1093/jxb/erae252

Korem, T., Zeevi, D., Suez, J., Weinberger, A., Avnit-Sagi, T., Pompan-Lotan, M., Matot, E., Jona, G., Harmelin, A., Cohen, N., Sirota-Madi, A., Thaiss, C. A., Pevsner-Fischer, M., Sorek, R., Xavier, R. J., Elinav, E., & Segal, E. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science*, *349*(6252), 1101–1106. https://doi.org/10.1126/science.aac4812

Koskella, B., & Brockhurst, M. A. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews*, *38*(5), 916–931.

Kraege, A., Punt, W., Doddi, A., Zhu, J., Schmitz, N., Snelders, N. C., & Thomma, B. P. H. J. (2026). Undermining the cry for help: The phytopathogenic fungus Verticillium dahliae secretes an

antimicrobial effector protein to undermine host recruitment of antagonistic Pseudomonas bacteria. *New Phytologist*, *249*(1), 406–417. https://doi.org/10.1111/nph.70686

Kremer, J. M., Sohrabi, R., Paasch, B. C., Rhodes, D., Thireault, C., Schulze-Lefert, P., Tiedje, J. M., & He, S. Y. (2021). Peat-based gnotobiotic plant growth systems for Arabidopsis microbiome research. *Nature Protocols*, *16*(5), Article 5. https://doi.org/10.1038/s41596-021-00504-6

Kuzyakov, Y., & Domanski, G. (2000). Carbon input by plants into the soil. Review. *Journal of Plant Nutrition and Soil Science*, *163*(4), 421–431. https://doi.org/10.1002/1522-2624(200008)163:4%253C421::AID-JPLN421%253E3.0.CO;2-R

Kuzyakov, Y., & Razavi, B. S. (2019). Rhizosphere size and shape: Temporal dynamics and spatial stationarity. *Soil Biology and Biochemistry*, *135*, 343–360.

Laflamme, C., Simard, J.-R., Buteau, S., Lahaie, P., Nadeau, D., Déry, B., Houle, O., Mathieu, P., Roy, G., Ho, J., & Duchaine, C. (2011). Effect of growth media and washing on the spectral signatures of aerosolized biological simulants. *Applied Optics*, *50*(6), 788–796. https://doi.org/10.1364/AO.50.000788

Lagier, J.-C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J.-M., Fournier, P.-E., & Raoult, D. (2018). Culturing the human microbiota and culturomics. *Nature Reviews Microbiology*, *16*(9), 540–550. https://doi.org/10.1038/s41579-018-0041-0

Lakowicz, J. R., & Masters, B. R. (2008). Principles of Fluorescence Spectroscopy, Third Edition. *Journal of Biomedical Optics*, *13*(2), 029901. https://doi.org/10.1117/1.2904580

Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution*, *29*(1), 33–41. https://doi.org/10.1016/j.tree.2013.09.009

Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D., & Desai, M. M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, *500*(7464), 571–574. https://doi.org/10.1038/nature12344

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.

Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., & Barraclough, T. G. (2012). Species Interactions Alter Evolutionary Responses to a Novel Environment. *PLOS Biology*, *10*(5), e1001330. https://doi.org/10.1371/journal.pbio.1001330

Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., Malfatti, S., Glavina del Rio, T., Jones, C. D., Tringe, S. G., & Dangl, J. L. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*, *349*(6250), 860–864. https://doi.org/10.1126/science.aaa8764

Leblanc, L., & Dufour, É. (2002). Monitoring the identity of bacteria using their intrinsic fluorescence. *FEMS Microbiology Letters*, *211*(2), 147–153. https://doi.org/10.1111/j.1574-6968.2002.tb11217.x

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lee, J., O'Kane, D. J., & Gibson, B. G. (1989). Dynamic fluorescence study of the interaction of lumazine protein with bacterial luciferases. *Biophysical Chemistry*, *33*(1), 99–111.

Lee, K. S., Landry, Z., Athar, A., Alcolombri, U., Pramoj Na Ayutthaya, P., Berry, D., De Bettignies, P., Cheng, J.-X., Csucs, G., Cui, L., Deckert, V., Dieing, T., Dionne, J., Doskocil, O., D'Souza, G., García-Timermans, C., Gierlinger, N., Goda, K., Hatzenpichler, R., … Stocker, R. (2024). MicrobioRaman: An open-access web repository for microbiological Raman spectroscopy data. *Nature Microbiology*, *9*(5), 1152–1156. https://doi.org/10.1038/s41564-024-01656-3

Leek, J. T., Johnson, W. E., Parker, H. S., Fertig, E. J., Jaffe, A. E., Storey, J. D., Zhang, Y., & Torres, L. C. (2014). *Package "sva."*

Lemanceau, P., Blouin, M., Muller, D., & Moënne-Loccoz, Y. (2017). Let the Core Microbiota Be Functional. *Trends in Plant Science*, *22*(7), 583–595. https://doi.org/10.1016/j.tplants.2017.04.008

Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal*, *11*(10), 2181–2194. https://doi.org/10.1038/ismej.2017.69

Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, *138*(6), 1315–1341. https://doi.org/10.1086/285289

Lenski, R. E., Wiser, M. J., Ribeck, N., Blount, Z. D., Nahum, J. R., Morris, J. J., Zaman, L., Turner, C. B., Wade, B. D., & Maddamsetti, R. (2015). Sustained fitness gains and variability in fitness

trajectories in the long-term evolution experiment with Escherichia coli. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1821), 20152292.

Lescat, M., Launay, A., Ghalayini, M., Magnan, M., Glodt, J., Pintard, C., Dion, S., Denamur, E., & Tenaillon, O. (2017). Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an *Escherichia coli* natural isolate in the streptomycin-treated mouse gut. *Molecular Ecology*, *26*(7), 1802–1817. https://doi.org/10.1111/mec.13851

Lesuffleur, F., Paynel, F., Bataillé, M.-P., Le Deunff, E., & Cliquet, J.-B. (2007). Root amino acid exudation: Measurement of high efflux rates of glycine and serine from six different plant species. *Plant and Soil*, *294*(1), 235–246.

Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., Rangel Alvarez, B., Lundberg, D. S., Lu, T.-Y., Lebeis, S., Jin, Z., McDonald, M., Klein, A. P., Feltcher, M. E., Rio, T. G., … Dangl, J. L. (2018). Genomic features of bacterial adaptation to plants. *Nature Genetics*, *50*(1), 138–150. https://doi.org/10.1038/s41588-017-0012-9

Li, E., de Jonge, R., Liu, C., Jiang, H., Friman, V.-P., Pieterse, C. M. J., Bakker, P. A. H. M., & Jousset, A. (2021). Rapid evolution of bacterial mutualism in the plant rhizosphere. *Nature Communications*, *12*(1), 3829. https://doi.org/10.1038/s41467-021-24005-y

Li, E., De Jonge, R., Liu, C., Jiang, H., Friman, V.-P., Pieterse, C. M. J., Bakker, P. A. H. M., & Jousset, A. (2021). Rapid evolution of bacterial mutualism in the plant rhizosphere. *Nature Communications*, *12*(1), 3829. https://doi.org/10.1038/s41467-021-24005-y

Li, E., Zhang, H., Jiang, H., Pieterse, C. M. J., Jousset, A., Bakker, P. A. H. M., & De Jonge, R. (2021). Experimental-Evolution-Driven Identification of *Arabidopsis* Rhizosphere Competence Genes in Pseudomonas protegens. *mBio*, *12*(3), e00927-21. https://doi.org/10.1128/mBio.00927-21

Li, H. (2009). The sequence alignment/map (SAM) format and SAMtools 1000 genome project data processing subgroup. *Bioinformatics*, *25*, 1.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.

Lidstrom, M. E., & Konopka, M. C. (2010). The role of physiological heterogeneity in microbial population behavior. *Nature Chemical Biology*, *6*(10), 705–712. https://doi.org/10.1038/nchembio.436

Lind, P. A., & Andersson, D. I. (2008). Whole-genome mutational biases in bacteria. *Proceedings of the National Academy of Sciences*, *105*(46), 17878–17883. https://doi.org/10.1073/pnas.0804445105

Liu, H., Xia, J., Xin, J., Zhao, H., & Ni, Y. (2023). Wavelength detection technique of overlapping spectra in the serial WDM FBGs by convolutional neural network. *Optical Fiber Technology*, *75*, 103206. https://doi.org/10.1016/j.yofte.2022.103206

Liu, R., Dai, W., Wu, T., Wang, M., Wan, S., & Liu, J. (2022). AIMIC: Deep Learning for Microscopic Image Classification. *Computer Methods and Programs in Biomedicine*, *226*, 107162. https://doi.org/10.1016/j.cmpb.2022.107162

Liu, S., Wu, J., Cheng, Z., Wang, H., Jin, Z., Zhang, X., Zhang, D., & Xie, J. (2025). Microbe-mediated stress resistance in plants: The roles played by core and stress-specific microbiota. *Microbiome*, *13*(1), 111. https://doi.org/10.1186/s40168-025-02103-z

Liu, Y., Zhang, H., Wang, J., Gao, W., Sun, X., Xiong, Q., Shu, X., Miao, Y., Shen, Q., Xun, W., & Zhang, R. (2024). Nonpathogenic Pseudomonas syringae derivatives and its metabolites trigger the plant “cry for help” response to assemble disease suppressing and growth promoting rhizomicrobiome. *Nature Communications*, *15*(1), 1907. https://doi.org/10.1038/s41467-024-46254-3

Losi, A., Ternelli, E., & Gärtner, W. (2004). Tryptophan Fluorescence in the Bacillus subtilis Phototropin-related Protein YtvA as a Marker of Interdomain Interaction¶. *Photochemistry and Photobiology*, *80*(1), 150. https://doi.org/10.1562/2004-03-17-RC-116.1

Luan, L., Tang, J.-W., Luo, Y.-F., Ru, X.-J., Luo, Y., Yao, Y.-N., Liu, S.-L., Gu, B., & Wang, L. (2026). Influences of growth media and culture time on the SERS spectral patterns of three common pathogenic bacteria: A comparative analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *348*, 127078. https://doi.org/10.1016/j.saa.2025.127078

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., & Rio, T. G. del. (2012). Defining the core Arabidopsis thaliana root microbiome. *Nature*, *488*(7409), 86–90.

Lynch, J. M., & Whipps, J. M. (1990). Substrate flow in the rhizosphere. *Plant and Soil*, *129*(1), 1–10.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963.

Marchetti, M., Clerissi, C., Yousfi, Y., Gris, C., Bouchez, O., Rocha, E., Cruveiller, S., Jauneau, A., Capela, D., & Masson, C. (2017). Experimental evolution of rhizobia may lead to either extra- or intracellular symbiotic adaptation depending on the selection regime. In *Molecular Ecology* (Vol. 26, Issue 7, pp. 1818–1831). Wiley.

Marín, O., González, B., & Poupin, M. J. (2021). From Microbial Dynamics to Functionality in the Rhizosphere: A Systematic Review of the Opportunities With Synthetic Microbial Communities. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.650609

Martins, S. J., Taerum, S. J., Triplett, L., Emerson, J. B., Zasada, I., de Toledo, B. F., Kovac, J., Martin, K., & Bull, C. T. (2022). Predators of soil bacteria in plant and human health. *Phytobiomes Journal*, *6*(3), 184–200.

Masahiko Taniguchi & Jonathan S. Lindsey. (2023). Absorption and fluorescence spectra of open-chain tetrapyrrole pigments–bilirubins, biliverdins, phycobilins, and synthetic analogues. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, *55*, 100585. https://doi.org/10.1016/j.jphotochemrev.2023.100585

Matuszewski, S., Hermisson, J., & Kopp, M. (2014). FISHER'S GEOMETRIC MODEL WITH A MOVING OPTIMUM. *Evolution*, *68*(9), 2571–2588. https://doi.org/10.1111/evo.12465

Mendes, R., Garbeva, P., & Raaijmakers, J. M. (2013). The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, *37*(5), 634–663.

Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., Van Der Voort, M., Schneider, J. H., Piceno, Y. M., DeSantis, T. Z., Andersen, G. L., & Bakker, P. A. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, *332*(6033), 1097–1100.

Mesny, F., Bauer, M., Zhu, J., & Thomma, B. P. H. J. (2024). Meddling with the microbiota: Fungal tricks to infect plant hosts. *Current Opinion in Plant Biology*, *82*, 102622. https://doi.org/10.1016/j.pbi.2024.102622

Momeni, M. K., Mohammadi, M. A., & Farhadian, S. (2025). Investigating the influence of diverse treatments on the autofluorescence properties of Bacillus subtilis. *Scientific Reports*, *15*(1), 38906. https://doi.org/10.1038/s41598-025-22746-0

Monaco, G., Chen, H., Poidinger, M., Chen, J., de Magalhães, J. P., & Larbi, A. (2016). flowAI: Automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*, *32*(16), 2473–2480. https://doi.org/10.1093/bioinformatics/btw191

Montelone, B. A. (2006). *DNA Repair and Mutagenesis.* Second Edition. *By* Errol C Friedberg, Graham C Walker, Wolfram Siede, Richard D Wood, Roger A Schultz, and, Tom Ellenberger. *Washington (DC): ASM Press.* $179.95. Xxix + 1118 p + 11 pl; ill.; index. ISBN: 1-55581-319-4. 2006. *The Quarterly Review of Biology*, *81*(3), 273–273. https://doi.org/10.1086/509407

Morella, N. M., Weng, F. C.-H., Joubert, P. M., Metcalf, C. J. E., Lindow, S., & Koskella, B. (2020). Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *Proceedings of the National Academy of Sciences*, *117*(2), 1148–1159. https://doi.org/10.1073/pnas.1908600116

Morris, B. E. L., Henneberger, R., Huber, H., & Moissl-Eichinger, C. (2013). Microbial syntrophy: Interaction for the common good. *FEMS Microbiology Reviews*, *37*(3), 384–406. https://doi.org/10.1111/1574-6976.12019

Mueller, U. G., & Sachs, J. L. (2015). Engineering Microbiomes to Improve Plant and Animal Health. *Trends in Microbiology*, *23*(10), 606–617. https://doi.org/10.1016/j.tim.2015.07.009

Mukherjee, R., Verma, T., Nandi, D., & Umapathy, S. (2020). Understanding the effects of culture conditions in bacterial growth: A biochemical perspective using Raman microscopy. *Journal of Biophotonics*, *13*(1), e201900233. https://doi.org/10.1002/jbio.201900233

Müller, D. B., Vogel, C., Bai, Y., & Vorholt, J. A. (2016). The Plant Microbiota: Systems-Level Insights and Perspectives. *Annual Review of Genetics*, *50*(1), 211–234. https://doi.org/10.1146/annurev-genet-120215-034952

Newman, M.-A., Sundelin, T., Nielsen, J. T., & Erbs, G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers in Plant Science*, *4*, 139.

Newton, A. C., Fitt, B. D., Atkins, S. D., Walters, D. R., & Daniell, T. J. (2010). Pathogenesis, parasitism and mutualism in the trophic space of microbe–plant interactions. *Trends in Microbiology*, *18*(8), 365–373.

Niu, B., Paulson, J. N., Zheng, X., & Kolter, R. (2017). Simplified and representative bacterial community of maize roots. *Proceedings of the National Academy of Sciences*, *114*(12), E2450–E2459.

Noble, R. (2019). *ggmuller: Create Muller Plots of Evolutionary Dynamics*. https://doi.org/10.5281/zenodo.591304

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. *Community Ecology Package*, *10*(631–637), 719.

Oldroyd, G. E. (2013). Speak, friend, and enter: Signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology*, *11*(4), 252–263.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Perez-Ortiz, G., Sidda, J. D., Peate, J., Ciccarelli, D., Ding, Y., & Barry, S. M. (2023). Production of copropophyrin III, biliverdin and bilirubin by the rufomycin producer, Streptomyces atratus. *Frontiers in Microbiology*, *14*, 1092166. https://doi.org/10.3389/fmicb.2023.1092166

Pfeilmeier, S., Petti, G. C., Bortfeld-Miller, M., Daniel, B., Field, C. M., Sunagawa, S., & Vorholt, J. A. (2021). The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves. *Nature Microbiology*, *6*(7), 852–864. https://doi.org/10.1038/s41564-021-00929-5

Pfeilmeier, S., Werz, A., Ote, M., Bortfeld-Miller, M., Kirner, P., Keppler, A., Hemmerle, L., Gäbelein, C. G., Petti, G. C., Wolf, S., Pestalozzi, C. M., & Vorholt, J. A. (2024). Leaf microbiome dysbiosis triggered by T2SS-dependent enzyme secretion from opportunistic Xanthomonas pathogens. *Nature Microbiology*, *9*(1), 136–149. https://doi.org/10.1038/s41564-023-01555-z

Piccardi, P., Ulrich, E., Garcia-Garcerà, M., Martino, R. D., Testa, S. E. A., & Mitri, S. (2024). The evolution of reduced facilitation in a four-species bacterial community. *Evolution Letters*, *8*(6), 828–840. https://doi.org/10.1093/evlett/qrae036

Pieterse, C. M., Zamioudis, C., Berendsen, R. L., Weller, D. M., Van Wees, S. C., & Bakker, P. A. (2014). Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology*, *52*(1), 347–375.

Pöhlker, C., Huffman, J. A., & Pöschl, U. (2012). Autofluorescence of atmospheric bioaerosols – fluorescent biomolecules and potential interferences. *Atmospheric Measurement Techniques*, *5*(1), 37–71. https://doi.org/10.5194/amt-5-37-2012

Ponomarova, O., & Patil, K. R. (2015). Metabolic interactions in microbial communities: Untangling the Gordian knot. *Current Opinion in Microbiology*, *27*, 37–44. https://doi.org/10.1016/j.mib.2015.06.014

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., & others. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983–987.

Porter, J., Edwards, C., & Pickup, R. w. (1995). Rapid assessment of physiological status in Escherichia coli using fluorescent probes. *Journal of Applied Bacteriology*, *79*(4), 399–408. https://doi.org/10.1111/j.1365-2672.1995.tb03154.x

Porter, S. S., & Simms, E. L. (2014). Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecology Letters*, *17*(9), 1121–1129.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Raaijmakers, J. M., & Mazzola, M. (2016). Soil immune responses. *Science*, *352*(6292), 1392–1393. https://doi.org/10.1126/science.aaf3252

Ranf, S., Gisch, N., Schäffer, M., Illig, T., Westphal, L., Knirel, Y. A., Sánchez-Carballo, P. M., Zähringer, U., Hückelhoven, R., & Lee, J. (2015). A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in Arabidopsis thaliana. *Nature Immunology*, *16*(4), 426–433.

Redman, R. S., Sheehan, K. B., Stout, R. G., Rodriguez, R. J., & Henson, J. M. (2002). Thermotolerance generated by plant/fungal symbiosis. *Science*, *298*(5598), 1581–1581.

Richards-Kortum, R., & Sevick-Muraca, E. (1996). Quantitative optical spectroscopy for tissue diagnosis. *Annual Review of Physical Chemistry*, *47*, 555–606. https://doi.org/10.1146/annurev.physchem.47.1.555

Richardson, A. E., & Simpson, R. J. (2011). Soil microorganisms mediating phosphorus availability update on microbial phosphorus. *Plant Physiology*, *156*(3), 989–996.

Rideout, J. R., Bolyen, E., McDonald, D., Baeza, Y. V., Alastuey, J. C., Pitman, A., Morton, J., Zhu, Q., Navas, J., Gorlick, K., Aton, M., Debelius, J., Xu, Z., llcooljohn, Shorenstein, J., Luce, L., Treuren, W. V., Chase, J., charudatta-navare, … Tapo, C. (2025). *scikit-bio/scikit-bio: Scikit-bio 0.7.0* (Version 0.7.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.15988672

Ringel, M. T., & Brüser, T. (2018). The biosynthesis of pyoverdines. *Microbial Cell*, *5*(10), 424–437. https://doi.org/10.15698/mic2018.10.649

Rodriguez, P. A., Rothballer, M., Chowdhury, S. P., Nussbaumer, T., Gutjahr, C., & Falter-Braun, P. (2019). Systems Biology of Plant-Microbiome Interactions. *Molecular Plant*, *12*(6), 804–821. https://doi.org/10.1016/j.molp.2019.05.006

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

Römling, U., Cao, L.-Y., & Bai, F.-W. (2023). Evolution of cyclic di-GMP signalling on a short and long term time scale. *Microbiology*, *169*(6), 001354. https://doi.org/10.1099/mic.0.001354

Rubbens, P., Props, R., Boon, N., & Waegeman, W. (2017). Flow Cytometric Single-Cell Identification of Populations in Synthetic Bacterial Communities. *PLOS ONE*, *12*(1), e0169754. https://doi.org/10.1371/journal.pone.0169754

Rugbjerg, P., & Sommer, M. O. (2019). Overcoming genetic heterogeneity in industrial fermentations. *Nature Biotechnology*, *37*(8), 869–876.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., & Bull, J. J. (2004). The evolution of cooperation. *The Quarterly Review of Biology*, *79*(2), 135–160.

Saha, P., Roy, N., Mukherjee, D., & Sarkar, A. K. (2016). Application of Principal Component Analysis for Outlier Detection in Heterogeneous Traffic Data. *Procedia Computer Science*, *83*, 107–114. https://doi.org/10.1016/j.procs.2016.04.105

Sandoz, K. M., Mitzimberg, S. M., & Schuster, M. (2007). Social cheating in Pseudomonas aeruginosa quorum sensing. *Proceedings of the National Academy of Sciences*, *104*(40), 15876–15881. https://doi.org/10.1073/pnas.0705653104

Sasse, J., Martinoia, E., & Northen, T. (2018). Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends in Plant Science*, *23*(1), 25–41. https://doi.org/10.1016/j.tplants.2017.09.003

Saxer, G., Krepps, M. D., Merkley, E. D., Ansong, C., Deatherage Kaiser, B. L., Valovska, M.-T., Ristic, N., Yeh, P. T., Prakash, V. P., Leiser, O. P., Nakhleh, L., Gibbons, H. S., Kreuzer, H. W., & Shamoo, Y. (2014). Mutations in Global Regulators Lead to Metabolic Selection during Adaptation to Complex Environments. *PLoS Genetics*, *10*(12), e1004872. https://doi.org/10.1371/journal.pgen.1004872

Scharf, B. E., Hynes, M. F., & Alexandre, G. M. (2016). Chemotaxis signaling systems in model beneficial plant–bacteria associations. *Plant Molecular Biology*, *90*(6), 549–559.

Scheuerl, T., Hopkins, M., Nowell, R. W., Rivett, D. W., Barraclough, T. G., & Bell, T. (2020). Bacterial adaptation is constrained in complex communities. *Nature Communications*, *11*(1), 754. https://doi.org/10.1038/s41467-020-14570-z

Scheuring, I., & Yu, D. W. (2012). How to assemble a beneficial microbiome in three easy steps. *Ecology Letters*, *15*(11), 1300–1307. https://doi.org/10.1111/j.1461-0248.2012.01853.x

Schmid, F. (2001). Biological Macromolecules: UV-visible Spectrophotometry. In Wiley, *Encyclopedia of Life Sciences* (1st ed.). Wiley. https://doi.org/10.1038/npg.els.0003142

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069.

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. In *Gene prediction: Methods and protocols* (pp. 227–245). Springer.

Shi, Y., Ke, G., Soukhavong, D., Lamb, J., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., Titov, N., & Cortes, D. (2025). *lightgbm: Light Gradient Boosting Machine*. https://github.com/Microsoft/LightGBM

Singh, B. K., Hu, H.-W., Macdonald, C. A., & Xiong, C. (2025). Microbiome-facilitated plant nutrient acquisition. *Cell Host & Microbe*, *33*(6), 869–881. https://doi.org/10.1016/j.chom.2025.05.005

Skipper, S., & Josef, P. (2010). statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H., & Berg, G. (2001). Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: Plant-dependent enrichment and seasonal shifts revealed. *Applied and Environmental Microbiology*, *67*(10), 4742–4751.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, *480*(7376), 241–244. https://doi.org/10.1038/nature10571

Smolka, M., Paulin, L. F., Grochowski, C. M., Horner, D. W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., & others. (2024). Detection of mosaic and population-level structural variants with Sniffles2. *Nature Biotechnology*, *42*(10), 1571–1580.

Snelders, N. C., Rovenich, H., Petti, G. C., Rocafort, M., van den Berg, G. C. M., Vorholt, J. A., Mesters, J. R., Seidl, M. F., Nijland, R., & Thomma, B. P. H. J. (2020). Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nature Plants*, *6*(11), 1365–1374. https://doi.org/10.1038/s41477-020-00799-5

Snelders, N. C., Rovenich, H., & Thomma, B. P. H. J. (2022). Microbiota manipulation through the secretion of effector proteins is fundamental to the wealth of lifestyles in the fungal kingdom. *FEMS Microbiology Reviews*, *46*(5), fuac022. https://doi.org/10.1093/femsre/fuac022

Sniegowski, P. D., Gerrish, P. J., & Lenski, R. E. (1997). Evolution of high mutation rates in experimental populations of E. coli. *Nature*, *387*(6634), 703–705. https://doi.org/10.1038/42701

Sohn, M., Himmelsbach, D. S., Barton, F. E., & Fedorka-Cray, P. J. (2009). Fluorescence Spectroscopy for Rapid Detection and Classification of Bacterial Pathogens. *Applied Spectroscopy*, *63*(11), 1251–1255.

Song, H., Li, Y., & Wang, Y. (2023). Two-component system GacS/GacA, a global response regulator of bacterial physiological behaviors. *Engineering Microbiology*, *3*(1), 100051. https://doi.org/10.1016/j.engmic.2022.100051

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, *16*(8), 472–482. https://doi.org/10.1038/nrg3962

Spaepen, S., & Vanderleyden, J. (2011). Auxin and plant-microbe interactions. *Cold Spring Harbor Perspectives in Biology*, *3*(4), a001438.

Spooren, J., Van Bentum, S., Thomashow, L. S., Pieterse, C. M. J., Weller, D. M., & Berendsen, R. L. (2024). Plant-Driven Assembly of Disease-Suppressive Soil Microbiomes. *Annual Review of Phytopathology*, *62*(1), 1–30. https://doi.org/10.1146/annurev-phyto-021622-100127

Sprouffske, K., & Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, *17*(1), 172.

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*(1), 321–346.

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*(suppl_2), ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Strehmel, N., Böttcher, C., Schmidt, S., & Scheel, D. (2014). Profiling of secondary metabolites in root exudates of Arabidopsis thaliana. *Phytochemistry*, *108*, 35–46.

Summers, Z. M., Fogarty, H. E., Leang, C., Franks, A. E., Malvankar, N. S., & Lovley, D. R. (2010). Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria. *Science*, *330*(6009), 1413–1415. https://doi.org/10.1126/science.1196526

Sun, J., Tárnok, A., & Su, X. (2020). Deep Learning-Based Single-Cell Optical Image Studies. *Cytometry Part A*, *97*(3), 226–240. https://doi.org/10.1002/cyto.a.23973

Tan, A., Abecasis, G. R., & Kang, H. M. (2015). Unified representation of genetic variants. *Bioinformatics*, *31*(13), 2202–2204.

Tenaillon, O. (2014). The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics*, *45*(Volume 45, 2014), 179–201. https://doi.org/10.1146/annurev-ecolsys-120213-091846

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Médigue, C., Schneider, D., & Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, *536*(7615), 165–170. https://doi.org/10.1038/nature18959

Thiergart, T., Durán, P., Ellis, T., Vannier, N., Garrido-Oter, R., Kemen, E., Roux, F., Alonso-Blanco, C., Ågren, J., Schulze-Lefert, P., & Hacquard, S. (2020). Root microbiota assembly and adaptive differentiation among European Arabidopsis populations. *Nature Ecology & Evolution*, *4*(1), 122–131. https://doi.org/10.1038/s41559-019-1063-3

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, *3*(9), 711–721. https://doi.org/10.1038/nrmicro1234

Tian, T., Reverdy, A., She, Q., Sun, B., & Chai, Y. (2020). The role of rhizodeposits in shaping rhizomicrobiome. *Environmental Microbiology Reports*, *12*(2), 160–172. https://doi.org/10.1111/1758-2229.12816

Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., & Singh, B. K. (2020). Plant–microbiome interactions: From community assembly to plant health. *Nature Reviews Microbiology*, *18*(11), 607–621. https://doi.org/10.1038/s41579-020-0412-1

Turkarslan, S., Stopnisek, N., Thompson, A. W., Arens, C. E., Valenzuela, J. J., Wilson, J., Hunt, K. A., Hardwicke, J., de Lomana, A. L. G., Lim, S., Seah, Y. M., Fu, Y., Wu, L., Zhou, J., Hillesland, K. L., Stahl, D. A., & Baliga, N. S. (2021). Synergistic epistasis enhances the co-operativity of mutualistic interspecies interactions. *The ISME Journal*, *15*(8), 2233–2247. https://doi.org/10.1038/s41396-021-00919-9

van Baar, B. L. (2000). Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiology Reviews*, *24*(2), 193–219.

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., & Dufresne, A. (2015). The importance of the microbiome of the plant holobiont. *New Phytologist*, *206*(4), 1196–1206. https://doi.org/10.1111/nph.13312

Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, *309*(1), 1–7. https://doi.org/10.1111/j.1574-6968.2010.02000.x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, *10*(12), 828–840.

Vorholt, J. A., Vogel, C., Carlström, C. I., & Müller, D. B. (2017). Establishing causality: Opportunities of synthetic communities for plant microbiome research. *Cell Host & Microbe*, *22*(2), 142–155.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Weigel, W. A., & Dersch, P. (2018). Phenotypic heterogeneity: A bacterial virulence strategy. *Microbes and Infection*, *20*(9), 570–577. https://doi.org/10.1016/j.micinf.2018.01.008

Whitfield, C., & Trent, M. S. (2014). Biosynthesis and export of bacterial lipopolysaccharides. *Annual Review of Biochemistry*, *83*(1), 99–128.

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*(20), 3350–3352.

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., Cordero, O. X., Brown, S. P., Momeni, B., & Shou, W. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *The ISME Journal*, *10*(11), 2557–2568.

Wielgoss, S., Barrick, J. E., Tenaillon, O., Wiser, M. J., Dittmar, W. J., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R. E., & Schneider, D. (2013). Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proceedings of the National Academy of Sciences*, *110*(1), 222–227. https://doi.org/10.1073/pnas.1219574110

Wilde, J., Slack, E., & Foster, K. R. (2024). Host control of the microbiome: Mechanisms, evolution, and disease. *Science*, *385*(6706), eadi3338. https://doi.org/10.1126/science.adi3338

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., & Nagarajan, N. (2012). LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201.

Winter, C., Bouvier, T., Weinbauer, M. G., & Thingstad, T. F. (2010). Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: The "Killing the Winner" Hypothesis Revisited. *Microbiology and Molecular Biology Reviews*. https://doi.org/10.1128/mmbr.00034-09

Wippel, K., Tao, K., Niu, Y., Zgadzaj, R., Kiel, N., Guan, R., Dahms, E., Zhang, P., Jensen, D. B., Logemann, E., & others. (2021). Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nature Microbiology*, *6*(9), 1150–1162.

Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, *342*(6164), 1364–1367. https://doi.org/10.1126/science.1243357

Wunderer, M., Markt, R., Prem, E. M., Peer, N., Mullaymeri, A., & Wagner, A. O. (2024). Cofactor F420 tail length distribution in different environmental samples. *Heliyon*, *10*(20), e39127. https://doi.org/10.1016/j.heliyon.2024.e39127

Yuan, J., Zhao, J., Wen, T., Zhao, M., Li, R., Goossens, P., Huang, Q., Bai, Y., Vivanco, J. M., Kowalchuk, G. A., Berendsen, R. L., & Shen, Q. (2018). Root exudates drive the soil-borne legacy of aboveground pathogen infection. *Microbiome*, *6*(1), 156. https://doi.org/10.1186/s40168-018-0537-x

Zamioudis, C., & Pieterse, C. M. (2012). Modulation of host immunity by beneficial microbes. *Molecular Plant-Microbe Interactions*, *25*(2), 139–150.

Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E. J., Woodnutt, G., Short, J. M., & Keller, M. (2005). High-Throughput Cultivation of Microorganisms Using Microcapsules. In *Methods in Enzymology* (Vol. 397, pp. 124–130). Elsevier. https://doi.org/10.1016/S0076-6879(05)97007-9

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., Cho, H., Karaoz, U., Loqué, D., Bowen, B. P., Firestone, M. K., Northen, T. R., & Brodie, E. L. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology*, *3*(4), 470–480. https://doi.org/10.1038/s41564-018-0129-3

Zhang, J., Liu, Y.-X., Guo, X., Qin, Y., Garrido-Oter, R., Schulze-Lefert, P., & Bai, Y. (2021). High-throughput cultivation and identification of bacteria from the plant root microbiota. *Nature Protocols*, *16*(2), 988–1012.

Zhang, J., Liu, Y.-X., Zhang, N., Hu, B., Jin, T., Xu, H., Qin, Y., Yan, P., Zhang, X., Guo, X., Hui, J., Cao, S., Wang, X., Wang, C., Wang, H., Qu, B., Fan, G., Yuan, L., Garrido-Oter, R., … Bai, Y. (2019). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nature Biotechnology*, *37*(6), 676–684. https://doi.org/10.1038/s41587-019-0104-4

Zhang, J., Wang, B., Xu, H., Liu, W., Yu, J., Wang, Q., Yu, H., Wei, J.-W., Dai, R., Zhou, J., He, Y., Zou, D., Yang, J., Ban, X., Hu, Q., Meng, X., Liu, Y.-X., Wang, B., Hu, B., … Bai, Y. (2025). Root microbiota regulates tiller number in rice. *Cell*, *188*(12), 3152-3166.e16. https://doi.org/10.1016/j.cell.2025.03.033

Zhang, P., Spaepen, S., Bai, Y., Hacquard, S., & Garrido-Oter, R. (2021). Rbec: A tool for analysis of amplicon sequencing data from synthetic microbial communities. *ISME Communications*, *1*(1), 73.

Zhang, Y., Xu, J., Riera, N., Jin, T., Li, J., & Wang, N. (2017). Huanglongbing impairs the rhizosphere-to-rhizoplane enrichment process of the citrus root-associated microbiome. *Microbiome*, *5*(1), 97.

7. Appendix I: Single-Cell Classification via Spectral Flow Cytometry

In this section, I briefly describe the preliminary yet promising results of extending the fluorescence classification framework to the single-cell level using spectral flow cytometry.

7.1. Methods

7.1.1. Preparation of bacteria

Growth conditions were the same as used for plate reader-based measurements, extensively described in **3.4.2**. Bacteria were washed and resuspended in phosphate-buffered saline (PBS; pH = 7.0). For this, 5 ml dense cultures were centrifuged at 4000 x g for 10 minutes. The supernatant was discarded and bacterial cultures were resuspended in 5 ml PBS.

7.1.2. Flow cytometer measurements

Single-cell measurements were performed using the BD FACSDiscover S8 platform. 30,000 individual fluorescent events were generated per sample using default settings. Importantly, the samples were not treated with any fluorescent dye. Hence, the observed spectra are expected to be purely a consequence of the intrinsic fluorescence. To account for potential background effects, we also performed assays using only tryptic soy broth (TSB) medium and phosphate-buffered saline (PBS). Raw outputs were parsed using FlowJo 11 and exported as comma-separated-values (CSV) file.

7.1.3. Processing of flow cytometry data

CSV files containing the fluorescent events obtained from the flow cytometer were parsed using Polars. In contrast to the plate reader outputs, the CSV files were already in a machine-learning accessible format. Hence, the individual measurements (one file per sample) were concatenated and the corresponding metadata were added. For standardization of the flow cytometer-derived data, we transformed the data by

calculating the inverse hyperbolic sine (arcsinh). After this, the data were normalized by *z*-score transformation. Area under the curve (AUC) normalization was not implemented, because the dataset size obtained from the flow cytometer is orders of magnitude larger. For comparison, a canonical measurement using the plate reader using a full scan would yield 96 × 821 = 78,816 data points, while the flow cytometer yielded 30,000 × 377 = 11,310,000 data points per sample. Hence, complicated mathematical operations such as calculation of the AUC were slow and have a huge memory footprint. Missing features were imputed using one of the following options: mean, median, min, max, zero or 10^{10} (acting as a flag for missing data). Prior to PCA, the data were standardized using *z*-score transformation using the scikit-learn's StandardScaler transformer.

7.1.4. Training of models

For the flow cytometer-derived data, we used the LightGBM framework for all machine-learning based approaches. The reason for this is that LightGBM has proven out-of-the-box performance, allows for easy parallelization, is extremely memory efficient and very fast during training and inference for large datasets (Shi et al., 2025).

7.2. Results

To bridge the gap between bulk population measurements and single-cell heterogeneity, we aimed to adapt our fluorescence classification framework to single-cell resolution. We utilized a BD FACSDiscover S8 spectral flow cytometer to acquire full-spectrum profiles of the 34-member SynCom. In contrast to standard cytometry, which uses discrete bandpass filters, this system captures the full emission spectrum ranging from ultraviolet to infrared, yielding a high-dimensional dataset of 377 spectral features per cell. Initial unsupervised analysis using PCA revealed significant overlap between taxa, in marked contrast to the distinct clustering observed in plate-reader measurements (**Figure 12A**). This reduced separability likely reflects the high stochasticity of gene expression and metabolic states at the single-cell level, a phenomenon often masked in bulk population averages (Elowitz et al., 2002). Despite this overlap, we trained a tree-based classifier to extract discriminative features. Analysis of feature importance revealed that spectral channels in the UV range

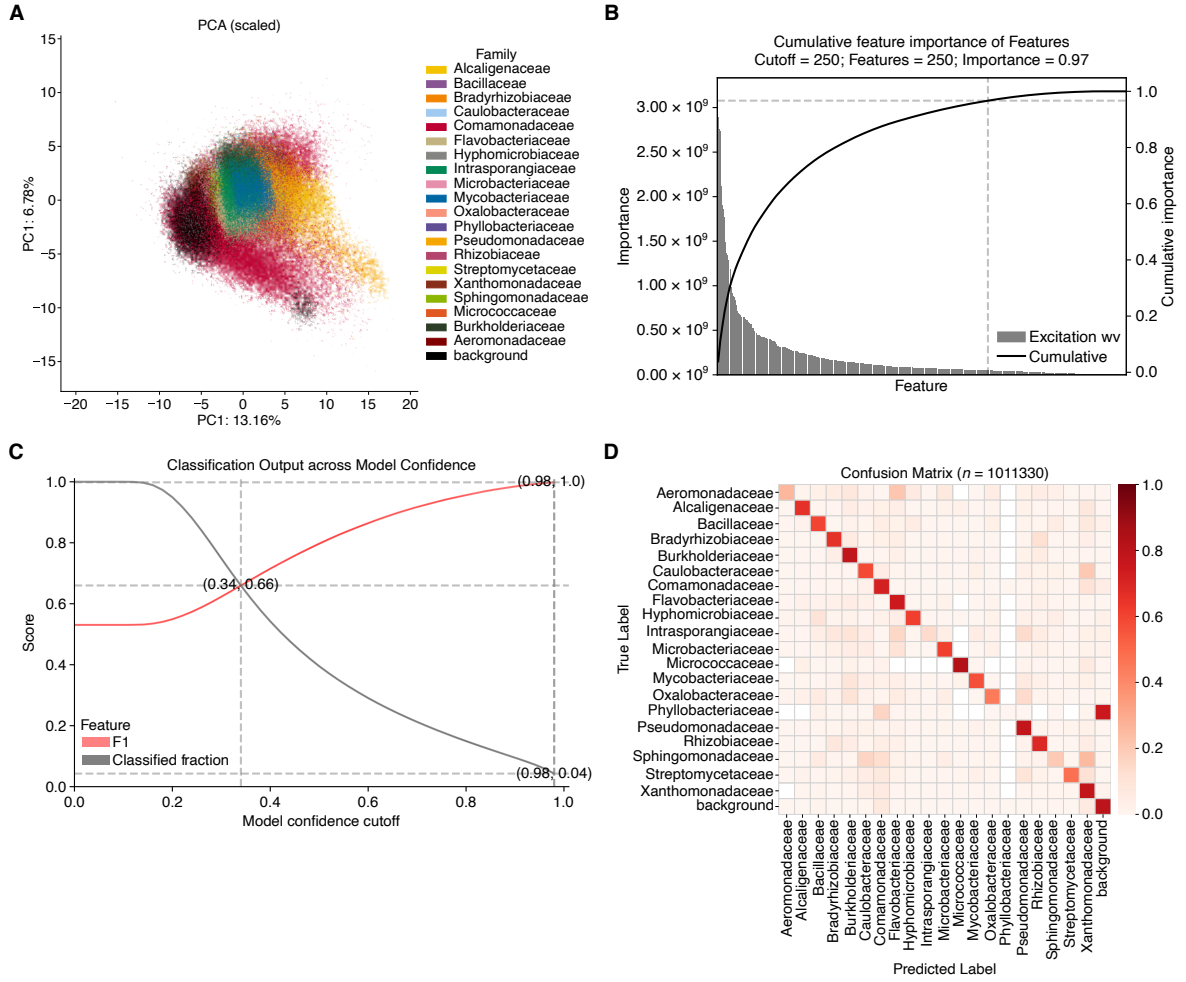

Figure 12. Single-cell taxonomic classification using high-dimensional spectral flow cytometry. (**A**) Principal component analysis of standardized single-cell fluorescent profiles. Each dot represents an individual event (*n* = 1,386,179), colored by taxonomic family. (**B**) Feature importance analysis. Grey bars display the importance of individual spectral features derived from a random forest classifier. The black line depicts the cumulative importance, with the dashed vertical line indicating the selected threshold of the top 250 features used for downstream analysis. (**C**) Precision-yield trade-off in single-cell classification. The red line tracks the F1-score as a function of the model's confidence threshold (probability cutoff), while the grey line indicates the corresponding fraction of retained (classified) events. Three key points are highlighted. Maximum F1-score, maximum classification yield, and the optimal balance point between accuracy and event retention. (**D**) Confusion matrix of the optimized single-cell classifier (*n* = 1,011,330). Rows represent true strain identities and columns represent predicted labels, with colour intensity proportional to the number of classified events.

(associated with intrinsic fluorophores like tryptophan and NADH) and morphological parameters (forward and side scatter representing size and granularity, respectively) were the most informative predictors (**Figure 12B**) (Chance et al., 1979; Losi et al., 2004). A cumulative importance analysis indicated that the top 250 features accounted for 97% of the total information. Using this subset of features, an initial classifier yielded a modest global F1-score of 0.55 (**Figure 12C**). A deeper inspection of the

predictions revealed that a significant proportion of cells were classified with low confidence (probability < 0.30), indicating high uncertainty of the decision boundaries of the model. To address this, we implemented a confidence-based strategy, effectively allowing the model to reject ambiguous or low-confidence events. This approach introduced a trade-off between classification precision and yield (the fraction of the total number of events). As the confidence cutoff increased, the F1-score improved significantly, but the number of classified events dropped (**Figure 12C**). We identified an optimal sweet spot at a confidence cutoff of 0.34. At this threshold, the model retained and classified 66.6% of all events with a significantly improved F1-score of 0.66. This equilibrium suggests that the model is well-calibrated, as the rejection of the lower third of uncertain predictions yielded a proportional performance gain in the remaining two-thirds of the dataset, effectively balancing the precision-recall trade-off. The confusion matrix for this optimized model (**Figure 12D**) demonstrates robust classification for most taxa, with the notable exception of Phyllobacteriaceae, which showed consistent misclassification, suggesting a potential lack of distinct fluorescent markers or similarities to residual particles originating from the medium.

7.3. Discussion

The transition from bulk to single-cell spectral classification presents unique challenges, primarily with respect to the signal-to-noise ratio. While plate readers integrate the signal from millions of cells, smoothing out individual variations, flow cytometry captures the noisy reality of biological heterogeneity (Avery, 2006; Davis & Isberg, 2016; Lidstrom & Konopka, 2010; Weigel & Dersch, 2018). The lower initial accuracy (F1 ~ 0.55) highlights the difficulty of distinguishing taxonomically distinct but metabolically similar cells without the use of exogenous labels (Rubbens et al., 2017). The presented optimization of classification output via feature selection and confidence cutoffs suggests that this workflow can be further optimized at several parts of the framework. Experimentally, increasing the signal-to-noise ratio could be achieved by optimizing the media that are used to reduce background scattering. Computationally, the pipeline could be refined by incorporating deep learning architectures, such as convolutional neural networks (CNNs), which have shown

superior performance in handling spectral data (Sun et al., 2020). Despite current limitations, achieving an F1-score of 0.66 in a label-free manner is a promising proof-of-concept. The ability to classify two-thirds of a population with decent confidence opens the door to reverse genomics strategies. By coupling this spectral sorting with downstream single-cell sequencing, researchers could physically isolate specific spectral phenotypes corresponding to uncultured taxa, bypassing the need for cultivation prior to identification (Cross et al., 2019). Furthermore, if connected to an automated sampling system, this framework could function as a live community tracker, enabling real-time monitoring of microbial dynamics in bioreactors or hydroponic systems without the lag time of sequencing.