Ecological competence *in the* rhizosphere

Genetic determinants, physiological plasticity and inference of root-associated bacteria from amplicon data



Juan José Sánchez Gil

Ecological competence in the rhizosphere

Genetic determinants, physiological plasticity,
and inference of root-associated bacteria
from amplicon data

Juan José Sánchez Gil

Ecological competence in the rhizosphere: genetic determinants, physiological plasticity, and inference of root-associated bacteria from amplicon data

PhD Thesis

Juan José Sánchez Gil
Utrecht University | Plant-Microbe Interactions | 2026

The work in this thesis was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) Groen II project (ALWGR.2017.002) and the Novo Nordisk Foundation (NNF19SA0059362).

Copyright: © 2026, Juan José Sánchez Gil

ISBN: 978-94-6534-562-8

Cover illustration: Ángel Luis Serrano Roldán

Printing: ProefschriftMaken

Ecological competence in the rhizosphere: genetic determinants, physiological plasticity, and inference of root-associated bacteria from amplicon data

Ecologische competentie in de rhizosfeer: genetische determinanten, fysiologische plasticiteit en inferentie van wortelgeassocieerde bacteriën op basis van amplicondata

(met een samenvatting in het Nederlands)

Competencia ecológica en la rizosfera: determinantes genéticos, plasticidad fisiológica e inferencia de bacterias asociadas a las raíces a partir de datos de amplicones

(con resumen en español)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 25 september 2026 des ochtends te 10.15 uur

door

Juan José Sánchez Gil

geboren op 13 juli 1993
te Algeciras, Spanje

Promotoren:

Prof. dr. ir. C.M.J. Pieterse

Prof. dr. B.E. Dutilh

Copromotor:

Dr. R. de Jonge

Beoordelingscommissie:

Prof. dr. C. Belzer

Prof. dr. J.M. Raaijmakers

Prof. dr. B. Snel

Prof. dr. S.J. Sørensen

Dr. K. Wippel

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) via het Groen II-project (ALWGR.2017.002) en van de Novo Nordisk Foundation (NNF19SA0059362).

(El señor Ujdt fi Allil tumbado boca abajo en el suelo con la cabeza apoyada en un almohadón. Nunca abre los ojos. Entra el amo de llaves)

AMO DE LLAVES. (Impasible). Señor.

SR. UJDT FI ALLIL. ¿Qué?

AMO DE LLAVES. El público.

SR. UJDT FI ALLIL. Pues que pase.

(Sale el amo de llaves parsimoniosamente. Entra el señor Farûq)

FARÛQ. (Saludando). El traje llamaba a la máscara abominable, y blenamiboá llamaba la máscara al traje. Ambos se repudian.

SR. UJDT FI ALLIL. Y repudian a la desnudez. La anulan. Quieren ser más que ella y a veces lo consiguen: la convencen y la acomplejan y ella se esconde. Pero cuando se une a la Voz puede salir victoriosa, lastimita da cuando no se tienen orejas para verla.

FARÛQ. Pero la desnudez les deja existir, porque ella es sabia y natural y les comprende. Ella les parió para sobrevivir. En el espejo la máscara y el traje mueren y se refleja la desnudez, pero solo si tú quieres oír.

(Abre la puerta el amo de llaves)

AMO DE LLAVES. (Impasible). Señor.

SR. UJDT FI ALLIL. ¿Qué?

AMO DE LLAVES. El público.

SR. UJDT FI ALLIL. ¿Quién?

AMO DE LLAVES. Su público.

SR. UJDT FI ALLIL. ¿Cuál?

AMO DE LLAVES. El suyo. El de debajo de la piel.

SR. UJDT FI ALLIL. Ah. Que pasen, estoy desnudo, no me harán nada.

(Sale el amo de llaves. Entran la Espiga, la Espiguilla y el Búho)

ESPIGUILLA. (Indignada). ¿Y para qué buscar la brisa si en las espaldas lleva ella siempre el cuchillo? Pues el día es solo el día y la tarde, pero la noche es todo noche, pero si está dañada... Durante el día llevas traje.

ESPIGA. (Enfadada, a la Espiguilla). Qué asco. Te repudio. Me repudio... Yo no tengo trajes.

BÚHO. ¿Todo para qué?

SR. UJDT FI ALLIL. Si al final las dos te corroen con el sol. Quémate la piel, al menos serás digno cuando venga la Muerte.

BÚHO. Porque te llevará desnudo en la clara noche, no en el trastornado sol.

FARÛQ. Ven, público, ven. Ven, que yo te entiendo, pero no entrarás ni saldrás de mí hasta que te procese. Hasta que en cueros te domine y así hacerte mío. Porque puedes ser una piedra, un terrón de tierra o un cuchillo, pero yo estaré desnudo y bajo mi piel, que se quema porque es piel y no máscara.

SR. UJDT FI ALLIL. Y contemplaré mis cicatrices desde dentro, orgulloso porque puedo tenerlas. Me las comeré y me haré más grande.

TODOS. Grande, grande. Grande. Grande.

VOZ. (Desde dentro). ¡Sí! ¡Grande!

SR. UJDT FI ALLIL. Muy grande. Y con moscas en la boca que escupiré siempre que mis entrañas hablen.

A mis sobrinos,
Leomar y Karolina

Table of Contents

Chapter 1. General introduction.....11

Chapter 2. Transcriptional plasticity of a rhizosphere bioinoculant reflects host phenotypic variability.....29

Chapter 3. The conserved *iol* gene cluster in *Pseudomonas* is involved in rhizosphere competence 87

Chapter 4. ASVNet: Inferring microbes from *16S rRNA* amplicon sequencing data 129

Chapter 5. General discussion.....153

References.....167

English summary 193

Nederlandse samenvatting.....199

Resumen en castellano 207

Acknowledgements 215

Curriculum vitae 222

List of publications.....223

Juan J. Sánchez-Gil[*1], Sanne W.M. Poppeliers[*1], and Ronnie de Jonge[1,2]

[1] Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

[2] AI Technology For Life, Department of Information and Computing Sciences, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

[*] These authors contributed equally to this manuscript.

Author contributions:

Juan J. Sánchez-Gil: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – Original Draft. **Sanne W.M. Poppeliers**: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – Original Draft. **Ronnie de Jonge**: Conceptualisation, Supervision, Project administration, Funding acquisition, Validation, Formal analysis, Writing – Review and Editing.

Adapted from: Sánchez-Gil, J.J., Poppeliers, S.W., de Jonge, R., 2023. **Microbes to support plant health: understanding bioinoculant success in complex conditions.** Current Opinion in Microbiology 73, 102286. DOI: 10.1016/j.mib.2023.102286.

CHAPTER ONE

General introduction

Abstract

A promising, sustainable way to enhance plant health and productivity is by leveraging beneficial microbes. Beneficial microbes are natural soil residents with proven benefits for plant performance and health. When applied in agriculture to improve crop yield and performance, these microbes are commonly referred to as bioinoculants. Yet, despite their promising properties, bioinoculant efficacy can vary dramatically in the field, hampering their applicability. Invasion of the rhizosphere microbiome is a critical determinant for bioinoculant success. Invasion is a complex phenomenon that is shaped by interactions with the local, resident microbiome and the host plant. Here, we explore all of these dimensions by cross-cutting ecological theory and molecular biology of microbial invasion in the rhizosphere. We refer to the famous Chinese philosopher and strategist Sun Tzu, who believed that solutions for problems require deep understanding of the problems themselves, to review the major biotic factors determining bioinoculant effectiveness.

Introduction: ponder and deliberate before you make a move

Past and present anthropogenic disturbances are diminishing soil quality world-wide and decreasing crop yields through an intense demand for food production. At the same time, there is a joint effort to develop sustainable agricultural techniques that can maintain or enhance productivity while reducing synthetic chemical inputs. One of the most promising methods to enhance plant health is by leveraging the benefits of the vast diversity of microbes that form the plant microbiome. Many pivotal features in plant fitness are dependent on the ensembled microbiome, such as nutrient homeostasis (Salas-González et al., 2021), resilience to drought and salt stress (He et al., 2022; H. Li et al., 2021), or natural protection against phytopathogens (Gu et al., 2022; Xu et al., 2022). Therefore, the scientific literature brims with examples of beneficial microbes and the responsible traits that can support plant health. These biological inoculants, or bioinoculants for short, are composed of single or multiple microbial species that are applied to soil and are the most common form of plant microbiota-engineering efforts.

The benefit associated with bioinoculants is commonly provided through their colonisation of the host plant roots, and most often, bioinoculant density in the rhizosphere environment greatly determines the eventual efficacy (Parnell et al., 2016; N. R. Wang et al., 2022). Whether the bioinoculant manages to engraft – that is, successfully invade a microbial community –, depends on the composition and interactions with and within the locally-adapted resident microbiota (Kurkjian et al., 2021; Qian and Akçay, 2020), edaphic factors (da Costa et al., 2020), host genotype (Wippel et al., 2021), root architecture and tissue-specific expression (Birt et al., 2022; Verbon et al., 2023), invader density and timing of invasion events (Dressler et al., 2019), and the inherent ability of the bioinoculant to circumvent these barriers and establish itself in the microbiome (Purtschert-Montenegro et al., 2022) (**Figure 1**). All of this commonly makes the design of bioinoculants strongly dependent on trial-and-error assays, and ultimately the most promising bioinoculants in the lab show variable efficacy in real-world agricultural conditions (O'Callaghan et al., 2022).

Some studies have already shown that global responses in the microbiome can be explained as an ensemble of the local interactions (Dal Co et al., 2020; Vliet et al., 2022). Therefore, optimising plant health through the design and application of beneficial microbes in agricultural conditions requires a deep mechanistic comprehension of all relevant interactions that take place during engraftment into the resident microbiome. In this review, by intertwining ecological theory with molecular research, we root for understanding of the following questions: (1) how is efficacy determined by interactions with other microbes, (2) how can we benefit from positive interactions with residential microbes to promote engraftment, (3) can we leverage bioinoculant-induced modifications on the community regardless

of invasion outcome, and (4) how does the interaction with the host determine effectiveness. We address these questions using Sun Tzu's philosophy from *The Art of War* to highlight the strategic nature behind optimisation of bioinoculant invasion and success.

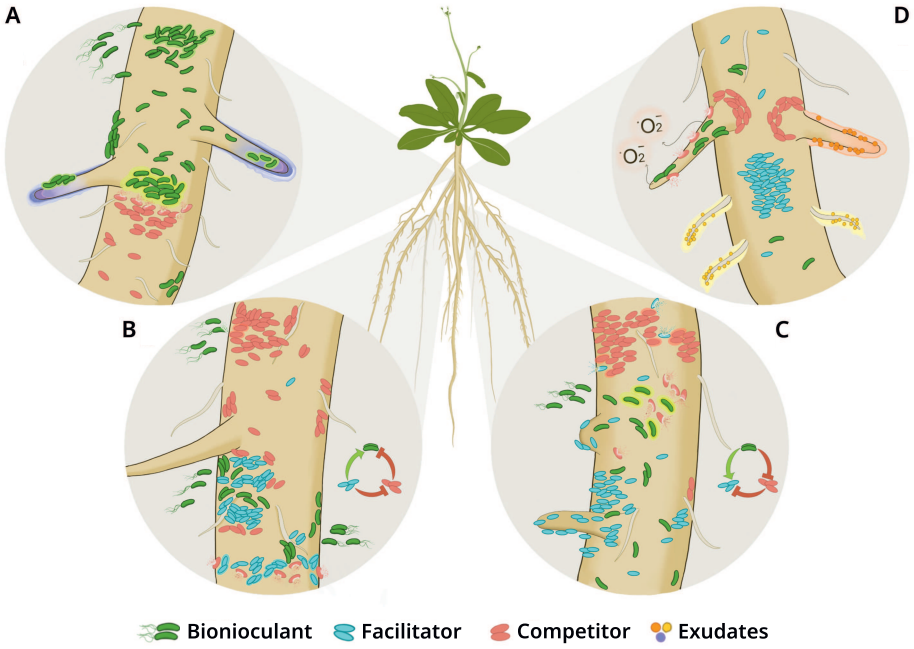


Figure 1 Processes and mechanisms influencing bioinoculant engraftment into the rhizosphere microbiome. (A) The activity of a bioinoculant depends on its physiological state and on interactions with other microbes. Top: relevant traits are only expressed in specific states and not in others. Middle: some traits are only expressed in interaction with specific microbes. Bottom: the interaction-triggered expression of traits is commonly of competitive nature, and is dependent on microbial relatedness. (B) Positive interactions with the residential microbiome can be used to favour invasion of an inoculated microbe. Top: a competitor is already present on the root, but without facilitator (or co-inoculated helper) invasion cannot occur. Middle: invasion can happen because a facilitator is present that benefits the invader directly. Bottom: the facilitator antagonises the competitor, leaving space for the invader to engraft. (C) Regardless of invasion success, the invader can induce effects on the resident microbiome that lead to positive results for plant health. Top: on the root, some beneficial residential microbes (blue) are restricted by the presence of competitors (red). Middle: invasion by a bioinoculant can benefit the lowly abundant microbes directly or kill the competitors. Bottom: the promoted residential can promote plant fitness regardless of what happens to the bioinoculant in future stages. (D) The plant host can control its microbiome through its immune system and by exuding different metabolites. The way the bioinoculant communicates with or is adapted to the host matters in predicting its success during invasion. Top: the plant host controls its microbiome through immune system activity, such as production of ROS, which benefit some microbes (green) but kill others (red). Middle: in other regions, different exudation profiles create different niches that select for different microbial communities. Bottom: the joint action of all processes, together with root architecture and soil properties, creates specific microhabitats for microbes.

Know thyself, know thy enemy: bioinoculant physiology and local interactions determine efficacy

Microbial traits such as growth rate, carbon utilisation, production of antibiotic compounds or the ability to form biofilms, are considered relevant for bioinoculant success. The expression of these traits, however, depends on the physiological state of the microbe. In the case of the *Pseudomonas* genus, a vastly divergent clade that comprises many potential bioinoculants (Passarelli-Araujo et al., 2022), most of the secondary metabolites that are relevant for their beneficial activity are produced during the biofilm state (Rieusset et al., 2020). At the same time, the production of these metabolites can be dramatically altered by other surrounding microbes, affecting bioinoculant efficacy. For example, production of the antimicrobials pyoluteorin and 2,4-diacetylphloroglucinol (DAPG) is crucial for the biocontrol activity of *Pseudomonas protegens*, and production of pyoluteorin can be enhanced upon sensing of a secreted metabolite from a related pseudomonad. In turn, antibiosis of pyoluteorin towards this related pseudomonad results in the release of intracellular compounds that inhibit production of DAPG in *P. protegens* (Hansen et al., 2022). In *Bacillus velezensis*, sensing pseudomonads' siderophores enhances the production of secondary metabolites, increasing its survival and triggering a broad antimicrobial activity (Andrić et al., 2022, 2021). Microbe-microbe interactions frequently enhance production of antimicrobial compounds across diverse environments, including the plant phyllosphere (Qi et al., 2021). Conversely, the very same study shows that isolates from the soil produced antimicrobials constitutively, probably as a result of adaptation to an environment with much higher microbial density. In a previous study, however, 42% of isolates from soil exhibit antimicrobial activity only when in interaction with other specific isolates (Tyc et al., 2014). These findings are in line with studies showing that most of microbial biosynthetic gene clusters (BGC) remain silent until their expression is induced by an external stimulus, commonly antimicrobial compounds (Seyedsayamdost, 2014).

Generally, direct competition is more frequent among closely related organisms and commonly leads to strain exclusion (Kehe et al., 2021). For instance, commensal pseudomonads were able to protect host plants against a closely related *Pseudomonas* pathogen via hitherto unknown mechanism (Shalev et al., 2022; N. R. Wang et al., 2022), and closely related *P. protegens* strains can inhibit each other via the release of tailocins, bactericidal protein complexes (Heiman et al., 2022). In contrast, closely related strains of *Bacillus subtilis* can essentially merge swarms and co-colonise surfaces, while kin discrimination leads to exclusion of distantly related strains from swarms (Kraigher et al., 2022), a phenomenon also observed during biofilm development (Kobayashi, 2021).

On the contrary, some microorganisms engraft in a fully independent manner, almost no matter what the resident community looks like. Such organisms carry the potential to be more broadly applicable in agriculture due to their reduced dependency on local conditions. Only few examples are described of such potentially superior bioinoculants, and one of them is *Pseudomonas putida* strain IsoF. This strain is able to aggressively invade biofilms, thanks to a horizontally-acquired type-IVB secretion system that confers a broad contact-dependent toxicity, ensuring *P. putida* engraftment into the rhizosphere community and enabling robust biocontrol activity (Purtschert-Montenegro et al., 2022).

The wise warrior avoids the battle: positive interactions with residential microbes to promote engraftment

Interactions between bioinoculants and other microorganisms do not necessarily have to be competitive, instead, they can also be of a facilitative nature. Specific members of the host microbiome can help the bioinoculant invade, both through direct and indirect mechanisms. Additionally, there are ample examples where bioinoculants that are co-inoculated with certain 'helper' strains are better able to establish and persist in a new environment, and consequently enhance plant performance. In bacilli, plant growth-promoting effects are dependent on biofilm formation in the rhizosphere, and this process is influenced by interactions with other microbes. For example, co-inoculation with helper strains that were not necessarily rhizosphere-competent themselves increased *B. subtilis* adherence on *Arabidopsis thaliana* roots (Eckshtain-Levi et al., 2020). In the cucumber rhizosphere, dual-species inoculation of *B. velezensis* and *Pseudomonas stutzeri* promoted plant growth more efficiently than individual species, and both inoculants reached higher abundances when co-inoculated in comparison with singular application. Intriguingly, such synergism was only evident under static nutrient-rich conditions, implying that spatial structure and nutrient availability are key to their collaborative nature (Sun et al., 2022). In general, compared with single-strain inoculation, consortia often have a larger effect on the resident microbial community (Hu et al., 2021; Z. Wang et al., 2021), and possibly therefore better stimulate plant growth (Emami et al., 2020) and protect plants against different stresses (Berendsen et al., 2018).

Understanding the mechanisms of pathogen invasion can help elucidate bioinoculant success or failure, since similar ecological principles apply. Many microorganisms might act as helpers of pathogens in the rhizosphere. Although the molecular mechanisms behind pathogen invasion remain elusive, the compatibility of siderophores (Gu et al., 2020) and tolerance to antimicrobials (Deng et al., 2022; Qi et al., 2021) produced by the resident community seem to play important roles. Facilitative interactions promote

invasion of the pathogen *Ralstonia solanacearum in vitro* and *in vivo* due to loss of pathogen inhibition (Li et al., 2019). Mathematical modelling of experimentally validated pairwise interactions predicts that inhibition of pathogen helper strains can reduce pathogen density in the rhizosphere more than direct targeting (Li et al., 2022). In general, more and more evidence confirms that the interactions between co-occurring species, as opposed to the presence or absence of a given species and its functional traits, are the most important determinant of disease suppression (Hu et al., 2020, 2016; Li et al., 2019).

Interactions between members of a community and an invader can transition from antagonistic to facilitative or vice-versa, when let to co-evolve. It has been recently reported that facilitation by a resident community enables invader survival and growth in a toxic medium where without these facilitative interactions the invader would not survive. However, the community became resistant to invasion after it co-evolved, possibly caused by increased growth of community members inhibiting invader survival (Piccardi et al., 2022). A transition from antagonism to coexistence was also observed in two well-known plant-beneficial microbes: *Pseudomonas chlororaphis* and *Bacillus amyloliquefaciens*. Mutual antagonism is conferred by the production of secondary metabolites with antimicrobial activity. Coevolution, however, drove the emergence of spontaneous mutants in both species that provide resistance against the competitor molecules by modifying the respective targets. Although the mutations came at the cost of pleiotropic effects, they facilitated the coexistence between both organisms that apparently outweigh these effects (Molina-Santiago et al., 2021). Together, both studies provide valuable insight into the adaptive mechanisms underlying simple and more complex microbial interactions that need to be considered when designing future bioinoculants.

All warfare is based on deception: leveraging invasion-induced effects on the resident community

Invasion by a bioinoculant has the ability to alter microbial communities, and these alterations may be leveraged to influence plant health, regardless of the outcome for the bioinoculant. The magnitude of this effect depends on the overlap between the ecological niche, in particular resource use, of the bioinoculant and that of the resident taxa. The presence of such overlap could affect competitive activities within the community and consequently alter its composition. For instance, transient invasion by bacilli induced changes in bacterial diversity in soil that persisted for long periods of time, regardless of invader persistence (Mawarda et al., 2022). Invasion primarily affected copiotrophs – fast-growing organisms found in nutrient-rich environments – during the early phases of community assembly that likely shared a niche with the invader, and as a result oligotrophs – slow-growing organisms found

in nutrient-sparse environments – appear to have benefitted at later stages of community assembly. *Escherichia coli* caused a similar effect in soil that persisted after its disappearance. Notably, changes in community composition following invasion correlate with community resource use (Mallon et al., 2018). A comparable shift in community composition occurred after the introduction of *B. subtilis* into a soil-derived synthetic community. As a result, the closest relatives in the community were repressed (Kiesewalter et al., 2020). These examples corroborate the idea that bioinoculants can affect the microbiome in a way that is dependent on bioinoculant and community functionality, but the question remains if such effects result in an alternative stable state, or if the community will convert back to its initial state with corresponding consequences for plant health. Alternative states, supported by changes in the abundance and/or presence of keystone microbial species or functionalities, could be responsible for the observed changes in plant health and the resistance against invasion by selected pathogens (Niu et al., 2017). Nevertheless, the legacy of invasion might facilitate future invasion attempts and recurrent inoculations, and might be one of the determinants of successful bioinoculant application (**Box 1**). Therefore, we conclude that it is important to track the effect of bioinoculant application over extended periods of time and the possible alternative states it might induce to know whether invasion causes long-lasting functional changes in the microbial community.

Indirect effects on resident communities can be used to tap into a 'new' resource of plant-beneficial microbes that are rare in soil and/or the rhizosphere under current conditions, via the addition of bioinoculants that do not necessarily need to be beneficial for the plant directly themselves. Introduction of consortia of plant-beneficial *Pseudomonas* strains in the tomato rhizosphere increased the frequency of rare taxa in the resident community, possibly through direct competition for resources with more abundant taxa. Notably, rather than invader consortia-specific beneficial traits, changes in community composition best explained the observed beneficial effects on plant growth, but how these changes resulted in improved plant growth remains to be answered. The effect was larger with more diverse inocula, suggesting synergistic effects between certain strains (Hu et al., 2021). Similarly, introduction of the beneficial microbe *B. amyloliquefaciens* into the tomato rhizosphere resulted in increased suppressiveness towards *R. solanacearum*, irrespective of inoculant abundance, possibly due to an enrichment of microbes carrying antimicrobial BGCs in the invaded communities (Deng et al., 2022).

In addition to invasion-induced effects on the resident microbial community, bioinoculants can also permanently alter host physiology, even after the bioinoculant is eliminated from the community. For example, bioinoculant-induced differential DNA methylation in pokeweed roots resulted in growth promotion, while functional-level variation in the rhizosphere microbiome induced by the inoculum was only detected at an early phase of microbial community assembly (C. Chen et al., 2022).

Knowing the battlefield ensures victory: invasion depends on interaction with the host

Bioinoculants ultimately need to interact with a host plant, which imposes a genotype-dependent selection that determines microbial assembly (Escudero-Martínez et al., 2022; Oyserman et al., 2022). Host-driven selection eventually enables emergence of compatibility and cooperation with its microbiome (Sharp and Foster, 2022), and this translates into microbial host preference (Wippel et al., 2021). Therefore, compatibility and interaction with the host plant is of high relevance in designing future effective bioinoculants.

Evolutionary studies show examples of how to optimise bioinoculant compatibility with the target host. In pseudomonads and bacilli in mono-association with *Arabidopsis*, cooperative and mutualistic microbial traits emerge rapidly (Blake et al., 2021; E. Li et al., 2021a). Evolved isolates showed improved root colonisation ability, in part driven by mutations in master regulators of bacterial behaviour, like the GacA/S two-component system that induced major physiological changes (E. Li et al., 2021a, 2021b; Nordgaard et al., 2022). Also another, master regulatory system in *Pseudomonas*, the ColR/S two-component system, has been linked to host adaptation at the strain level (Wiesmann et al., 2022), and is at the same time required for protection against other pathogenic pseudomonads (N. R. Wang et al., 2022). This host-driven evolution suggests that master regulatory systems are key targets for optimising bioinoculants compatibility and should be exploited in the future.

An important determinant of compatibility is the interaction with the host immune system. In some beneficial microbes like *B. velezensis*, host-derived reactive oxygen species (ROS) are required to stimulate bacterial secretion of auxin, which subsequently induces lateral root formation and protects against fungal pathogens (Tzipilevich et al., 2021). In contrast, ROS production is known to widely restrict colonisation in the rhizosphere (Song et al., 2021), and consequently, many successful root-colonising organisms, especially Actinobacteria and Gammaproteobacteria, have developed strategies to circumvent this by suppressing immune responses either locally or systemically (Ma et al., 2021; Teixeira et al., 2021; Yu et al., 2019).

Besides host immune activity, root exudate exploitation is a main driver of microbial compatibility. For instance, exuded inositol from *Arabidopsis* and tomato roots enhances chemotaxis and biofilm production of *Priestia megaterium* (formerly *Bacillus megaterium*) on their respective roots, resulting in growth promotion (Vílchez et al., 2020). In *B. subtilis* and in pseudomonads, sucrose exudation enhances colonisation, which leads to enhanced protection against the phytopathogens *Fusarium* and *Botrytis* (Tian et al., 2021).

Furthermore, immune activity and exudation profiles are not homogeneous throughout the root system. The diverse set of exuded metabolites accumulate in different regions of the root (Zhang et al., 2019), and the strength and scope of immune responses are cell-layer and cell-type-specific and dependent on local damage (Emonet et al., 2021; Fröschel et al., 2021; Rich-Griffin et al., 2020; Zhou et al., 2020). Moreover, experiments on the colonisation of Firmicutes and Proteobacteria isolates in leaves with different morphology showed that the spatial organisation of the colonised surface alone plays a key role in determining the success of an inoculant (Gupta et al., 2022). The combination of these three factors creates specific microhabitats on the root where the bioinoculant will eventually interact with its host (Birt et al., 2022; Kawa and Brady, 2022). For example, epidermal trichoblasts mediate the beneficial interaction with *Pseudomonas simiae* WCS417 through exudation of coumarins (Verbon et al., 2023), and many organisms colonise preferentially the root tip and elongation zone (Engelhardt et al., 2022).

These studies underline the importance of the bioinoculant-host compatibility, but also the relevance of the precisely invaded microhabitat for anticipating the local determinants for compatibility and eventual bioinoculant engraftment.

Whoever is first in the field will be fresh for the fight: timing in rhizosphere invasions

The rhizosphere microbiome and the functionalities it provides change with plant developmental stage. High biodiversity, as often seen in more mature rhizospheres, is often seen as a buffer to invasion (Case, 1990; Mallon et al., 2015; Tilman, 2004). Biodiversity can influence resistance to invasion through niche restriction/occupancy – that is, no available space for invader – (Herren, 2020; Hu et al., 2020), direct or indirect competition – for example, toxicity or activation of host immunity –, and niche displacement – phenotypic differentiation in space and time – (Mawarda et al., 2022). For example, in rice, *Enterobacter asburiae* was only able to promote rice shoot dry weight in the least diverse rich clay soils, likely because there were less keystone species in the most diluted microcosms that acted as generalist competitors of the bioinoculant (da Costa et al., 2020).

On young roots, pioneering opportunistic bacterial species occur that invest little in the production of toxic secondary metabolites, while at later stages bacteria with various life-strategies can coexist (Dini-Andreote and Raaijmakers, 2018), facilitated by niche differentiation (Nuccio et al., 2020). Stochasticity of microbiome assembly, mainly during the initial stages, leads to differential local configurations, which in turn leads to variability in the density of pathogen-suppressive members. This variability, although imperceptible at the global community scale, translates into differential antagonists densities and therefore an increase in community invasibility. This was

sufficient to explain invasion and disease onset by *R. solanacearum* (Gu et al., 2022; Lee et al., 2021; Wei et al., 2019). However, assembled communities from mature rhizospheres could prevent its invasion due to increased functional diversity, toxicity, and niche restriction (Hu et al., 2020).

The order of arrival of (invading) strains affects community assembly and determines the final microbial community composition. In *A. thaliana* and *Lotus japonica*, strains that arrived later in the resident communities reached lower proportions in the output communities, which could be due to niche pre-emption where early-arriving strains reduce the number of resources available. Nevertheless, strains that invaded communities on the host that they were isolated from were better able than strains from a different host plant, negating these priority effects. This effect was host-specific, i.e., strains were better able to invade communities on the host from which they were isolated, thereby overcoming these priority effects (Wippel et al., 2021). A similar effect was observed *in vitro* where early invaders were more successful than those trying to engraft in co-evolved communities at a later timepoint (Piccardi et al., 2022). In the maize rhizosphere, the correlation between AMF colonisation and soil biodiversity was time dependent, suggesting that once plants get established in the soil, positive interactions between AMF and the resident microbiota might not be as important anymore for AMF colonisation ability (Ferreira et al., 2021). Another example comes from the tomato phyllosphere. Pseudomonads belonging to the *fluorescens* clade might pre-empt as much niche space as possible on tomato leaves in an early stage of community assembly, and benefit from this later on even at the expense of a decreased maximal growth rate (Humphrey et al., 2020). In this case, competitiveness was optimised by shorter lag phases and lower growth rates, suggesting that prompter growth and dispersal are more determinant for bioinoculant efficacy than overall growth rate. However, recent estimations of growth rate across rhizosphere microbiome datasets show that the rhizosphere is dominated by genera with a high growth rate potential (López et al., 2023). Together, these studies suggest a calibrated balance between the time of exit from the lag phase and growth rate potential in order to succeed in complex communities, and they nuance the classical dichotomy of 'slow and stable' (*K*-strategist or oligotroph) vs. 'fast and fluctuating' (*r*-strategist or copiotroph) life histories.

While inoculation of different plant-beneficial strains or combinations thereof in soil resulted in only temporary changes in resident bacterial families, repeated inoculations permanently altered community succession in a strain-specific manner. Yet, beneficial effects of repeated inoculation on nutrient availability in the soil did decrease over time (Z. Wang et al., 2021). Repeated inoculations might thus be a solution for countering bioinoculant abundance decline and priority effects, although the efficacy of these measures can be time-dependent.

For a recent detailed account on the ecological mechanisms of invasion and the legacy on indigenous soil microbial communities in an agricultural context, we refer to (Liu et al., 2022; Mawarda et al., 2020), and for rele-

vant insights for bioinoculant application from macroecology please see (Albright et al., 2022).

In the midst of chaos, there is also opportunity: towards the bioinoculants of the future

In the previous sections, we discussed relevant sources of variability underlying bioinoculant engraftment and success. The current state of the topic would be probably defined by Sun Tzu as: *"one may know how to conquer without being able to do it"*. Yet, although we are technically able to apply bioinoculants in soil, conquering the complexity of microbial invasions requires more knowledge about the physiology of beneficial microbes in soil and the local-scale interactions that govern the microbiome (Vliet et al., 2022). For example, microbial biosynthetic capacities remain largely underexplored (Gavriilidou et al., 2022), and the role of known metabolites in microbial physiology are still being discovered, even in well-studied microbes (Anand et al., 2022; Tian et al., 2021). Motility in soil, for example, is another crucial trait for a soil bioinoculant, and is only now being characterised. In *B. subtilis*, Engelhardt and co-authors described recently a collective form of chemotactic motility in soil. When a community of *B. subtilis* cells sense the presence of a root in the vicinity, the cells move like flocks through the soil particles and arrange dynamically around the root tips (Engelhardt et al., 2022). This finding suggests that we need dedicated experiments to understand these mechanisms in current and future bioinoculants, as well as their relevance for successful colonisation.

Knowing the genetic and physiological characteristics from the bioinoculant and resident microbes allows for the prediction of genome-scale interactions, which in turn, allows for the engineering of the community in the long run. Recent advances, including single-cell sequencing techniques like Microbe-seq (Zheng et al., 2022) and Live-seq (W. Chen et al., 2022), could give more detailed information on the physiology of rhizosphere microbes, especially as improvements in sequencing and in sequence analysis software allow better re-construction of genomes from complex metagenomes (Pan et al., 2022) and more thorough retrieval of information from existing ones (Price et al., 2022). The interactions with the host and other microbes can benefit from new techniques in spatial metatranscriptomics (Saarenpää et al., 2022), together with the dissection of microbial transcriptional activity at single-cell resolution (Dar et al., 2021), and new combinations with expansion microscopy and sequencing (Alon et al., 2021). Other studies opt for overcoming the variability in efficacy with bold innovations like the use of encapsulated microbes that allow engineering of cooperative interactions (L. Wang et al., 2022).

However, sometimes, even "failed" bioinoculants might have beneficial effects and simply remain unseen in many experiments. A plant growth-pro-

moting *Bacillus* isolate that can promote *A. thaliana* growth *in vitro* did not have any effect on shoot biomass in artificial soil, but its application increased the total number of siliques (García-Cárdenas et al., 2022). This suggests that productivity-based parameters should be regarded in experimental conditions when evaluating bioinoculant effects.

For Sun Tzu, a deep knowledge of oneself and of the opponent was equivalent to avoiding war. Commonly, efforts in understanding and applying microbes as bioinoculants come from apparently opposite perspectives, ranging from studying specific genes in specific contexts that are difficult to generalise, to broad-scale ecological dynamics and concepts that have difficulties in explaining particular cases. However, relatively few studies integrate both worlds while trying to understand invasions as an ecological and mechanistic complex phenomenon. Ensuring bioinoculant benefits on plant fitness, either directly or indirectly, demands working in harmony with ecological principles and the molecular mechanisms that depend on, and give rise to, those principles. In this sense, future efforts should focus on defining the precise microhabitats where the bioinoculant can provide its most benefits, and study how its invasion depends on its own physiology under the conditions therein, and how interactions with the local community and the host determine its engraftment and its effects in the field.

Thesis outline

Microbial inoculants hold great promise for sustainable agriculture, yet their application in real-world conditions remains highly unpredictable and often inconsistent. As discussed in the previous sections, bioinoculant efficacy is shaped by a complex interplay of microbial ecophysiology, interactions with the resident microbiome, and host plant compatibility. The success or failure of microbial invasion is determined by multiple biotic and abiotic factors, including the functional traits of the inoculant – such as biofilm formation, secondary metabolite production, and competitive exclusion mechanisms –, the composition and diversity of the resident microbiota, and environmental conditions within the soil and rhizosphere.

Beyond microbe-microbe interactions, the host plant plays an active role in shaping its microbial community. Root exudation, immune responses, and spatial structuring of the rhizosphere create selective pressures that determine microbial colonisation success. Host genotype influences which microbes successfully establish on plant roots, while root architecture and localised immune responses create microhabitats that may promote or restrain bioinoculant engraftment. This dynamic relationship between microbial invaders, the resident microbiome, and the host plant underscores the importance of considering bioinoculant application within a broader ecological framework, rather than as isolated microbial additions to the soil.

Further complexity arises from the temporal and spatial dynamics of micro-

bial invasion. Invader density and timing of introduction play a crucial role in colonisation success, and microbial interactions are not static – they can shift from antagonistic to facilitative over time, particularly when communities are allowed to co-evolve. These dynamic factors complicate the prediction of bioinoculant success and their optimisation, as even microbes that initially struggle to establish can eventually influence rhizosphere composition and function in unforeseen ways.

As a result, while some introduced microbes successfully invade and persist in the rhizosphere, others fail due to competition, antagonistic interactions, or environmental constraints. Moreover, even a successful colonisation event does not necessarily translate into improved plant health or the intended beneficial effects, emphasising the need to understand the underlying microbial and host-related factors that drive variability in bioinoculant performance. In **Chapter 2**, we explore microbial factors that contribute to plant phenotypic variability following inoculation with a beneficial bacterium. For this, we take a metatranscriptomics approach to explore the transcriptional activity of the beneficial bacterium *Pseudomonas simiae* WCS417 in the natural rhizosphere of *Arabidopsis thaliana*. Our analysis reveals distinct gene expression patterns associated with larger versus smaller plants, suggesting microbial transcriptional responses that may influence host phenotypic differences. Notably, we observe a downregulation of genes related to biofilm formation and stress resistance in larger plants, alongside the upregulation of a gene potentially responsible for acidification via glucose degradation.

In **Chapter 3**, we take a broader genus-level perspective to investigate the genetic basis of rhizosphere competence in *Pseudomonas*. By comparing microbial densities of different *Pseudomonas* strains in the rhizosphere of *A. thaliana*, we identify a gene cluster involved in inositol metabolism – the *iol* locus – that is present only in the best colonisers. Through mutant competition experiments and *in vitro* assays, we demonstrate that the *iol* locus enhances root colonisation when in competition with a knock-out mutant and promotes swimming motility and siderophore production in response to inositol, a plant-derived compound. Additionally, we investigate the conservation of this genetic trait across the *Pseudomonas* genus and its role in shaping plant-microbe interactions. These findings highlight the importance of metabolic adaptation in microbial invasion and offer insights for optimising bioinoculant design.

However, despite significant progress in understanding the genetic and physiological determinants of microbial engraftment, bioinoculant applications remain largely trial-and-error, in part due to methodological constraints in microbiome research. While direct experimental approaches provide key insights into microbial traits and interactions, they cannot fully capture the complexity of microbial communities. In particular, many aspects of microbial competition, adaptation, and persistence remain difficult to quantify, especially when working with highly diverse and compositionally dynamic microbiomes. A major limitation arises from *16S rRNA* amplicon sequencing,

the most widely used technique for profiling plant-associated microbiomes. While valuable for taxonomic characterisation, its intrinsic complexity – such as intragenomic sequence heterogeneity, intergenomic sequence similarity, and variation in gene copy number – result in an inability to reliably assign amplicon sequences to their original microbial genomes. Therefore, microbial community compositions are commonly distorted estimates, where the proportion of different microbes are often over- or underestimated. Consequently, it is highly challenging to accurately infer microbial interactions and ecological dynamics. These limitations not only obscure the true structure of microbial communities but also hinder efforts to identify bioinoculant-compatible microbes and predict colonisation success.

In **Chapter 4**, we focus on overcoming the limitations of *16S rRNA* amplicon sequencing and we introduce a novel algorithmic approach, ASVNet. ASVNet integrates amplicon sequence similarity and co-abundance networks to reconstruct microbial community structures in a more accurate and biologically meaningful manner. By clustering sequences into empirical OTUs (eOTUs), ASVNet provides a biologically meaningful alternative to traditional OTU- and ASV-based clustering methods. Comparative analyses in synthetic and natural rhizosphere communities demonstrate that ASVNet-derived eOTUs better approximate true microbial composition, offering a more robust framework for microbiome studies and facilitating more precise ecological and functional inferences.

Finally, in **Chapter 5** I integrate the findings from the experimental and computational chapters within the broader context of microbial invasion, bioinoculant optimisation, and plant-microbiome interactions. I discuss how the insights gained from this work contribute to our understanding of the mechanisms governing microbial engraftment and how they can be applied to improve bioinoculant efficacy in agricultural settings.

By integrating microbial genomics and computational innovations, this thesis contributes to our current knowledge on the factors influencing microbial engraftment in plant-associated microbiomes. These insights help refine our ability to assess and predict bioinoculant performance, providing a step toward more reliable and effective applications in sustainable agriculture.

Juan J. Sánchez-Gil[1] and Ronnie de Jonge[1,2]

[1] Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

[2] AI Technology For Life, Department of Information and Computing Sciences, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

Author contributions:

Juan J. Sánchez-Gil: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – Original Draft. **Ronnie de Jonge:** Conceptualisation, Supervision, Project administration, Funding acquisition, Validation, Formal analysis, Writing – Review and Editing.

CHAPTER TWO

Transcriptional plasticity of a rhizosphere bioinoculant reflects host phenotypic variability

Abstract

Despite the increasing interest in microbial inoculants for sustainable agriculture, their inconsistent effects on plant performance remain poorly understood. A major challenge is the lack of insight into how microbial activity varies across individual plants under realistic environmental conditions. Here, we combine individual plant phenotyping with metatranscriptomic profiling to investigate the activity of the model bioinoculant *Pseudomonas simiae* WCS417 in the rhizosphere of *Arabidopsis thaliana* plants grown in natural soil. By stratifying plants by shoot biomass and comparing rhizosphere and bulk soil samples, we identify environment-specific transcriptional programs in both microbe and host. In bulk soil, WCS417 upregulates genes linked to stress adaptation, biofilm formation, and microbial competition – indicative of a persistence strategy under host-free conditions. In the rhizosphere, bacterial expression diverges according to plant phenotype: small plants associate with bacteria that express stress-related and biofilm genes, while big plants elicit expression of genes involved in translation, aerobic respiration, nutrient biosynthesis, and motility, consistent with a metabolically active and colonising state. Correspondingly, *Arabidopsis* roots of small plants display transcriptional signatures of nutrient limitation, oxidative stress, and growth constraint. In contrast, big plants exhibit enhanced expression of nutrient transporters, biosynthetic regulators, and coumarin-related genes, including *BGLU44*, suggesting a favourable root environment for microbial activity. Together, these results demonstrate that plant-to-plant heterogeneity is mirrored by microbial functional plasticity, and that successful colonisation by bioinoculants depends on local physiological context that includes plant growth and developmental processes. Our approach provides a path forward for dissecting microbe–host dynamics in complex systems and for improving the predictability of microbial inoculant performance in the field.

Introduction

The use of microbial bioinoculants – beneficial microorganisms applied to improve plant growth and resilience – offers a promising strategy for sustainable agriculture by reducing dependence on synthetic fertilisers and pesticides. However, despite their demonstrated effectiveness in controlled environments, bioinoculants often exhibit highly variable efficacy in field conditions. In some cases, inoculation leads to measurable improvements in crop yield or pathogen resistance, while in others, it results in no detectable benefit (O'Callaghan et al., 2022). This inconsistency presents a major limitation to their widespread application and underscores the need to understand the underlying causes of bioinoculant performance variability.

Several biotic and abiotic factors are known to influence bioinoculant success, including interactions with the resident microbial community (Kurkjian et al., 2021), soil properties (da Costa et al., 2020; Mosimann et al., 2017), environmental stressors (Lopes et al., 2021), and the physiological state of the host plant (Birt et al., 2022). These complex and dynamic variables shape the rhizosphere – the zone of intense biological activity surrounding and including plant roots – and ultimately determine whether an introduced microorganism can successfully colonise and function as an effective bioinoculant (Poppeliers et al., 2023). Yet, the precise mechanisms by which these factors modulate bioinoculant behaviour and efficacy remain poorly understood. In particular, it is unclear whether variability in plant response to inoculation reflects differential microbial performance across individual hosts, and if so, how such differences manifest at the physiological or transcriptional level. To improve bioinoculant efficacy, it is thus essential to understand what is happening at the level of individual plant–microbe interactions.

Studies that focus on the physiology and transcriptional responses of the inoculant itself, rather than community-wide shifts or averaged outcomes, are critical to uncover the precise mechanisms that govern colonisation success and functional engagement with the host, namely the activation of microbial genes and physiological traits that underlie plant-beneficial effects. This level of resolution would allow us to explore each plant–microbe interaction as a discrete ecological event, and to identify the conditions under which beneficial functions are actually expressed. Without this microbe-centred, mechanistic understanding, efforts to optimise bioinoculants will remain largely empirical and prone to failure in variable field settings.

To address this gap, we investigated the rhizosphere behaviour of the well-characterised plant growth-promoting bacterium *Pseudomonas simiae* WCS417 (hereafter WCS417; Pieterse et al., 2021) using *Arabidopsis thaliana* (hereafter *Arabidopsis*) grown in natural soil in the context of a complex microbial community. In an initial phenotypic assessment, we observed that WCS417 inoculation resulted in variable growth responses

across individual plants, with some showing clear promotion and others no benefit. This heterogeneity led us to hypothesise that WCS417 may adopt distinct physiological states depending on the host context.

To test this hypothesis, we employed a strain-specific approach combining metagenomic and metatranscriptomic analyses to track the abundance and transcriptional activity of WCS417 within the rhizosphere and soil microbial community. This strategy enabled us to directly monitor not only the presence of WCS417 but also its *in situ* gene expression across individual plants. We included bulk soil samples alongside rhizosphere samples to assess the transcriptional profile of WCS417 in the absence of plant hosts, providing a crucial baseline to interpret how its activity shifts in response to root-associated environments. To systematically examine this variability, we classified inoculated plants into two cohorts – small and big – based on their shoot size distribution. This allowed us to explore whether microbial behaviour diverges across host individuals and to assess the physiological status of each plant–microbe interaction independently. By comparing these subpopulations under otherwise uniform environmental conditions, we sought to determine whether differences in plant phenotype align with distinct microbial states. This approach provides a unique window into the real-time activity of an introduced strain in its ecological context, allowing us to uncover context-dependent microbial strategies that may explain variation in bioinoculant performance.

Our findings reveal that WCS417 exhibits distinct transcriptional profiles depending on the host plant phenotype and the surrounding environment. In bulk soil, WCS417 upregulates stress- and competition-related genes, indicating a survival-oriented state under challenging, host-free conditions. In small-plant rhizospheres, its expression suggests a multifaceted stress response likely driven by environmental constraints and host pressure, potentially limiting its growth-promoting potential. In contrast, in large-plant rhizospheres, WCS417 expression reflects a metabolically active and symbiotically engaged state, exemplified by an activation of genes required for protein synthesis, energy acquisition and motility.

Plant transcription mirrored these microbial states. Small plants showed redox and nutrient stress markers, while expression in big plants suggests a supportive but selective rhizosphere environment. Altogether, these results indicate that bioinoculant efficacy is not uniform but shaped by specific host–microbe interactions. Monitoring inoculant transcription in complex soils enables identification of the microbial states that underlie success or failure, providing a foundation for designing more effective and resilient bioinoculants for agriculture.

Results

The effect of WCS417 in the natural rhizosphere is variable

To assess variability in plant responses to WCS417 inoculation, we compared the effect of its inoculation on shoot growth across three different independent experiments, all measured at 21 days post-inoculation (dpi; **Figure 1A**). All experiments were conducted in topsoil collected from the Reijerscamp natural reserve in Wolfheze, The Netherlands, containing a native and complex microbial community. By using non-sterile natural soil rather than simplified substrates, these experiments capture the ecological constraints under which WCS417 must function, providing insight into its physiology and growth-promoting capacity under more realistic conditions. In two experiments, the effect of inoculation was only marginally significant (Experiment 1 *p*-value = 0.052 and Experiment 2 *p*-value = 0.059), while in the third, which included more replicates, a significant growth-promoting effect was observed (*p*-value = 0.023). When all data were combined, inoculated plants showed a robust and statistically significant 20% increase in shoot size compared to uninoculated controls (*p* = 0.00022; **Figure 1B**).

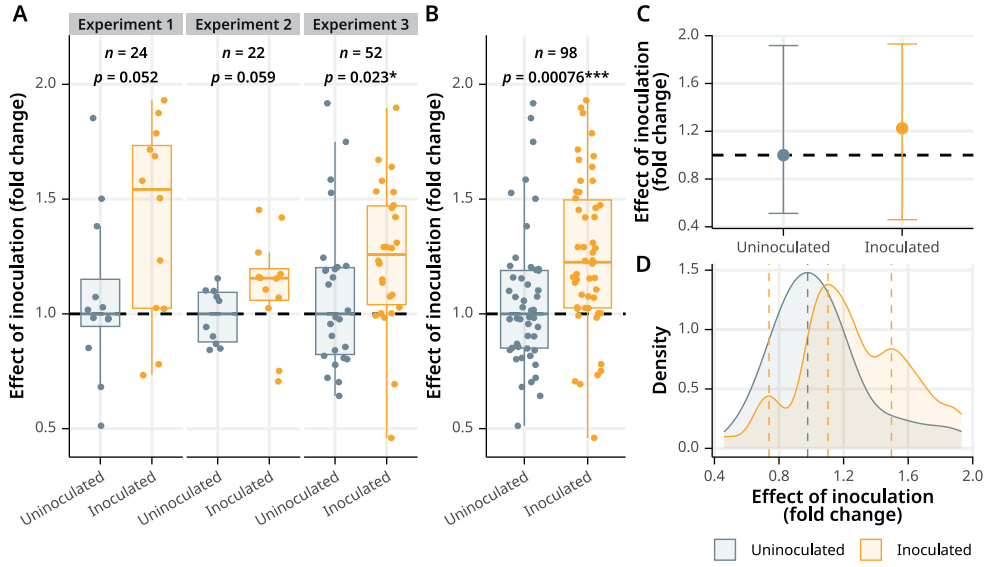


Figure 1. Host phenotypic variability upon WCS417 inoculation in natural soil. (A) Boxplots showing the result of WCS417 inoculation on three independent experiments. The data is shown as a fold-change of the inoculated versus uninoculated plants at 21 dpi. For each experiment, the result of a Wilcoxon test is shown above. The dashed line indicates the median value for uninoculated plants. (B) Same data in A for all experiments combined. (C) Plot depicting the min-max range of fold-changes in panel B. The dot represents the median value. (D) Density plot of the distribution of fold-changes. The dashed lines indicate the peaks of the distributions. It can be observed that the population of inoculated plants does not grow as a whole, but as three split groups.

The demonstrated overall benefit, however, masks substantial variability among individual plants. As shown in **Figure 1C**, the spread of plant sizes (min–max range) was comparable between treatments, indicating that the inoculation did not shift the entire population uniformly. Instead, the growth distribution in inoculated plants exhibited increased frequency of particularly large individuals (**Figure 1D**), rather than a consistent shift across the whole population.

Density plots revealed that, while uninoculated plants followed a unimodal distribution centred near the median, the distribution of inoculated plants was trimodal. This included a group of underperformers with reduced growth, a central group comparable to uninoculated controls, and a third group of plants with markedly enhanced shoot size. These results indicate that WCS417 inoculation did not trigger a homogeneous response across the population, but instead induced divergent outcomes – strong growth promotion in some plants, neutral effects in others, and possible growth delay in a subset.

WCS417 is more active in the rhizosphere of bigger plants

Given the variability in plant responses to WCS417 inoculation, we hypothesised that the bacterium may not colonise or behave uniformly across all hosts. In particular, differences in WCS417 abundance or activity could underlie the observed variation in plant growth. To explore this hypothesis, we designed an experiment to track both the abundance and transcriptional activity of WCS417 by generating paired metagenomes and metatranscriptomes from plants grown in natural soil. Therefore, we grew a population of *Arabidopsis* plants in natural soil previously mixed with 10^8 CFU of WCS417 per gram soil, and then selected five rhizosphere samples from plants in the bottom 40% (hereafter *small plants*) and top 30% (hereafter *big plants*) of the shoot size distribution. In addition, we included five bulk soil samples that were also inoculated with WCS417, in order to assess its transcriptional profile in the absence of plant hosts and provide a baseline for interpreting root-associated responses (**Supplementary Figure 1**, **Supplementary Table 1**). The general pipeline to process all samples and a depiction of read fates are provided in **Supplementary Figures 2** and **3**.

Across all samples, we obtained a median of 96 million metagenomic reads per compartment (**Figure 2A**, **Supplementary Table 2**). As expected, bulk soil samples contained very few plant reads (~7,500, 0.01%). In rhizosphere samples, the number of *Arabidopsis* reads was significantly higher in big plant rhizospheres (~11 million, ~12%) compared to small plant rhizospheres (~8 million, ~8%), representing a 43% increase (**Figure 2B**). However, this difference in number of reads did not seem to arise from a difference in root biomass (**Supplementary Figure 4**). This could suggest a higher bacterial load on small plant roots, lowering the ratio of plant-to-microbial reads.

To quantify WCS417, we conservatively included only read pairs with a

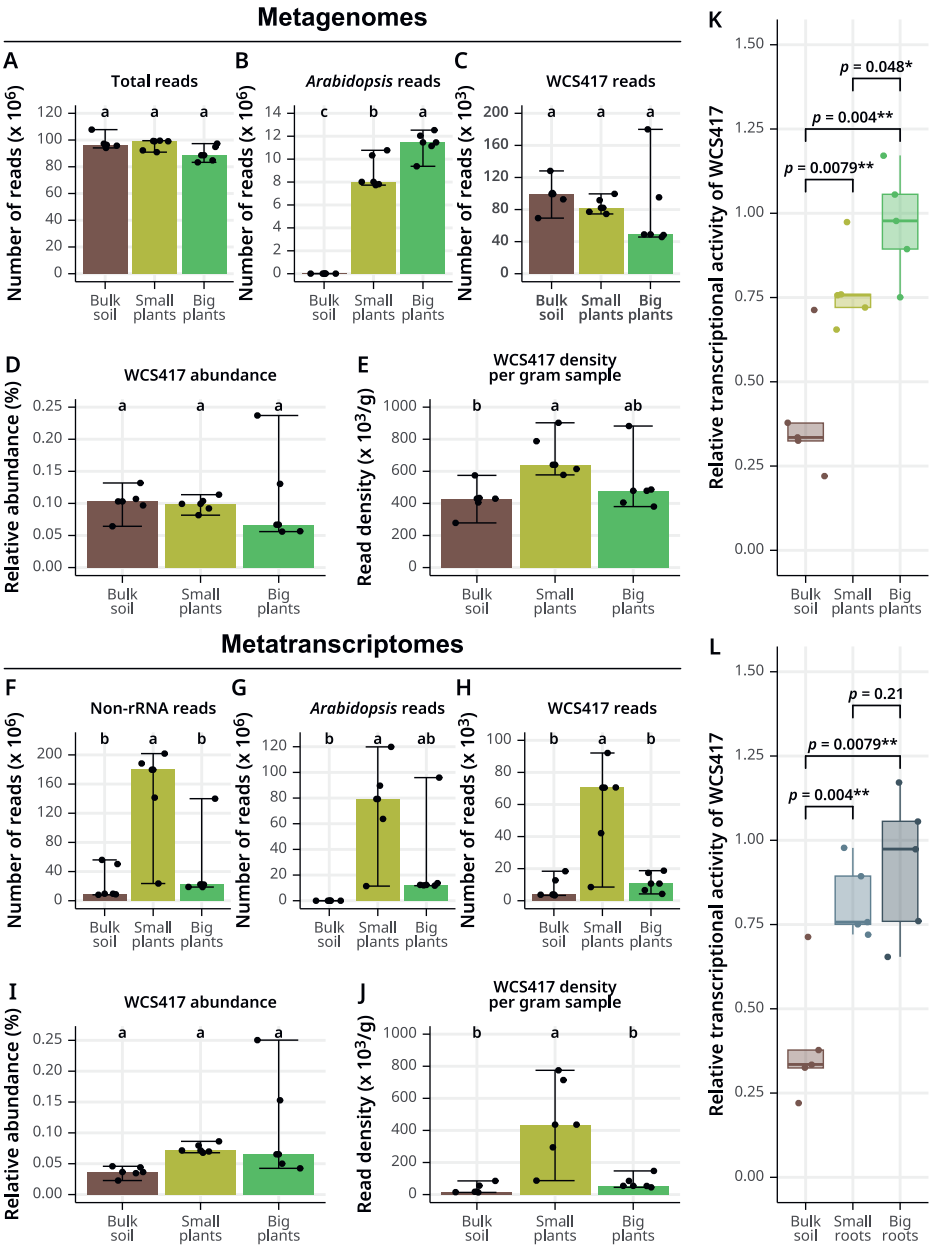


Figure 2. General depiction of sequencing results. Summary of the metagenomic (MG; A-E) and metatranscriptomic reads (MT; F-J). Each compartment is composed of five replicates. (A) Total number of metagenomic reads (in millions) per sample type. (B) Number of *Arabidopsis* reads in metagenomes (in millions). (C) Number of WCS417 reads in metagenomes (in thousands). (D) Density of WCS417 reads in the metagenome community per gram of root (or soil) sample. (E) Relative abundance of WCS417 in metagenome community. (F) Total number of metatranscriptomic reads. (G) Number of host plant reads in metatranscriptomes. (H) Number of WCS417 reads in metatranscriptomes. (I) Density of WCS417 metatranscriptomic reads per gram of root (or soil) sample. (J) Relative abundance of WCS417 transcriptional

reads in the microbial community. (K) Relative transcriptional activity of WCS417 in the different compartments, calculated as the ratio of transcriptional versus metagenomic abundances. The ratio of transcription over abundance is highest on the rhizosphere of big plants and lowest in bulk soil, while activity in small plants' rhizospheres is intermediate. (L) Same as in panel I, but re-classifying the plants according to root weight instead of shoot weight. Note that the overall transcriptional activity of WCS417 does not seem to differ between big and small roots. MG, metagenome; MT, metatranscriptome.

maximum of one mismatch to avoid misassignment from related pseudomonads in the microbial community. Absolute WCS417 counts varied and showed no statistically significant differences between compartments, perhaps due to the high initial inoculation density, although we observed a median 47% decrease in big plant samples (from ~92,000 in bulk soil and small plants to ~49,000 in big plants; **Figure 2C**). Similarly, relative abundance within the microbial community remained equal (**Figure 2D**), only again a non-significant downward trend was visible in big plants (from ~0.1% in bulk soil and small plants to ~0.07% in big plants). However, small plant roots had a 50% higher WCS417 read density per gram (**Figure 2E**), while the density in big plant roots was intermediate between the ones in bulk soil and small plants.

Intriguingly, metatranscriptomic read counts were markedly higher in small plant samples – roughly by 8-fold – compared to both bulk soil and big plants (**Figure 2F**), resulting in higher transcript counts for both *Arabidopsis* and WCS417 (**Figures 2G** and **2H**) and a higher density of WCS417 (**Figure 2J**), although we did not observe any strong disbalance in RNA concentrations during extraction or sample preparation **(Supplementary Figure 5).** Nevertheless, the relative abundance of WCS417 transcriptional reads remained unchanged across compartments and similar to its abundance in metagenomes (**Figure 2I**), indicating that the abundances themselves remained intact.

To better assess WCS417 activity, we calculated the ratio of cumulative transcriptome to genome abundance as a proxy for per-cell transcriptional activity. This analysis revealed that WCS417 was most active in the rhizospheres of big plants, followed by small plants, and least active in bulk soil (**Figure 2K**). To test whether this activity was influenced by root biomass rather than shoot size, we also regrouped samples by root weight. No difference in WCS417 activity was observed between small- and big-root plants (**Figure 2L**), indicating that the observed increase in transcriptional activity is specifically associated with shoot biomass and not confounded by soil amount.

Shoot size drives shifts in WCS417 gene expression patterns in the rhizosphere

To explore how the transcriptional profile of WCS417 varies across environments, we performed principal component analysis (PCA) on bacterial gene expression. Samples from rhizospheres clearly separated from bulk soil samples along the first principal component, indicating a strong

rhizosphere effect on WCS417 expression (*p*-value = 0.001 in a PERMANOVA test; **Figures 3A, 3B**).

Within the rhizosphere group, we further observed significant divergence in expression profiles between small- and big-plant roots (*p*-value = 0.012 in a PERMANOVA test; **Figure 3C**), suggesting that WCS417 responds transcriptionally not only to the presence of roots, but also to host size or condition.

A potential confounding factor underlying these expression differences might be the variable contribution of bulk soil to rhizosphere samples. Roots of different sizes and architectures can retain different amounts of adhering soil particles, which could artificially influence the apparent microbial transcriptional profile. Under this scenario, differences in microbial expression between samples from smaller and big plants would primarily reflect differences in soil carryover rather than a true plant-associated response.

To address this possibility, we quantified for each rhizosphere sample how transcriptionally distinct it was from bulk soil. Specifically, we calculated the

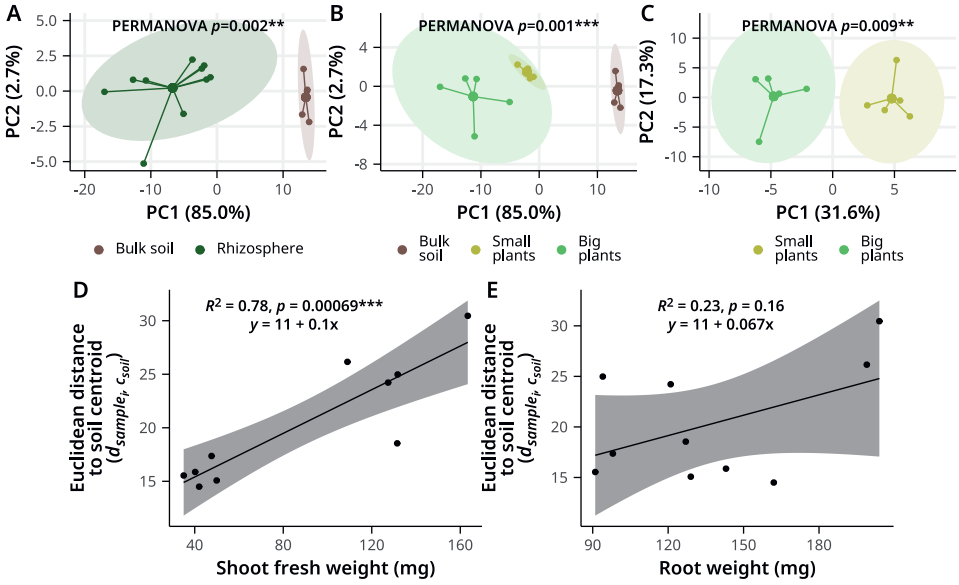


Figure 3. Transcriptional profiles of WCS417 shows different patterns in soil and rhizospheres and this is due to plant shoot biomass. (A-C) Scatterplots of PCA scores for WCS417 expression across samples on the first two principal components. In each plot, the significance of the differences between groups is assessed with a PERMANOVA test and shown above. (A) PCA scores for expression in rhizosphere versus bulk soil, regardless of plant size. (B) Same PCA scores as in A, but root samples are split into small and big plants. (C) PCA scores only on plant samples. Note that all differences between groups are significant, indicative of distinctive microbial expression patterns in soil and the root, but also between the rhizospheres of small and big plants. (D-E) Scatterplots of Euclidean distances between each plant sample in B and the centroid of bulk soil samples, with respect to shoot fresh weight (D) and root weight (E) Both plots show the result the R^2 and *p*-value of a linear model fitted for each relationship. The comparison indicates that the separation of root samples from soil samples is based on shoot fresh weight rather than root weight.

Euclidean distance of each rhizosphere sample to the centroid of bulk soil samples in the PCA space. This distance provides a quantitative measure of deviation from the soil-associated transcriptional state, with larger distances indicating a stronger rhizosphere-specific expression profile. If soil carryover were driving the observed separation, this distance would be expected to scale with root biomass rather than shoot biomass.

Instead, distances to the soil centroid were strongly correlated with shoot fresh weight (R^2 = 0.78, p < 0.001; **Figure 3D**), but showed no significant relationship with root weight (R^2 = 0.23, p = 0.16; **Figure 3E**). Consistently, analysis of the PCA loadings revealed that the first principal component was associated with shoot rather than root biomass (**Supplementary Figure 6**, **Supplementary Table 3**).

These findings indicate that the transcriptional shifts in WCS417 are not driven by differential soil carryover and reflect a genuine modulation of bacterial gene expression associated with plant shoot biomass.

WCS417 shows signs of stress and microbial competition in soil

In order to understand the behaviour of WCS417 in the absence of plant hosts, we compared its transcriptional profile in bulk soil to that in rhizosphere samples. This analysis revealed 600 differentially expressed genes (DEGs), of which only 19 were significantly enriched in soil after multiple testing correction (**Figure 4**, **Supplementary Table 4**), indicating that WCS417 differentially activates a relatively specific expression program under host-free conditions.

Several of the most strongly upregulated genes in soil are associated with general stress adaptation. For example, *rseP* (*PS417_RS21585*) encodes an intramembrane protease involved in the activation of the RpoE-mediated envelope stress response, also known as AlgU response in pseudomonads (Barchinger and Ades, 2013; H. Wang et al., 2021). Similarly, *nfuA* (*PS417_RS15715*) encodes a carrier of iron–sulphur clusters, which plays a key role in supporting enzymatic function under oxidative or iron-limiting conditions (Romsang et al., 2018). Another highly enriched gene was *PS417_RS06330*, whose homologue in *Pseudomonas syringae* (Psyr_3782) is a predicted hydrofilin linked to desiccation and osmotic stress tolerance (Freeman et al., 2013). In addition, *PS417_RS22705*, a putative ribosomal subunit interface protein, was upregulated. This gene is homologous to *PSPTO_1205*, a regulator of translation and growth under nutrient stress in *P. syringae* (Deng et al., 2014). Likewise, *aprA* (*PS417_RS14600*) encodes a Gac-Rsm/Fur-controlled extracellular metalloprotease implicated in the iron starvation response (Llamas et al., 2014). Interestingly, AprA is also associated to the interaction with the host in pathogens, particularly via evasion of host immunity (Acken and Li, 2023; Bardoel et al., 2011; Pel et al., 2014), although its expression in the rhizosphere compared to soil is rather low (**Supplementary Figure 7**).

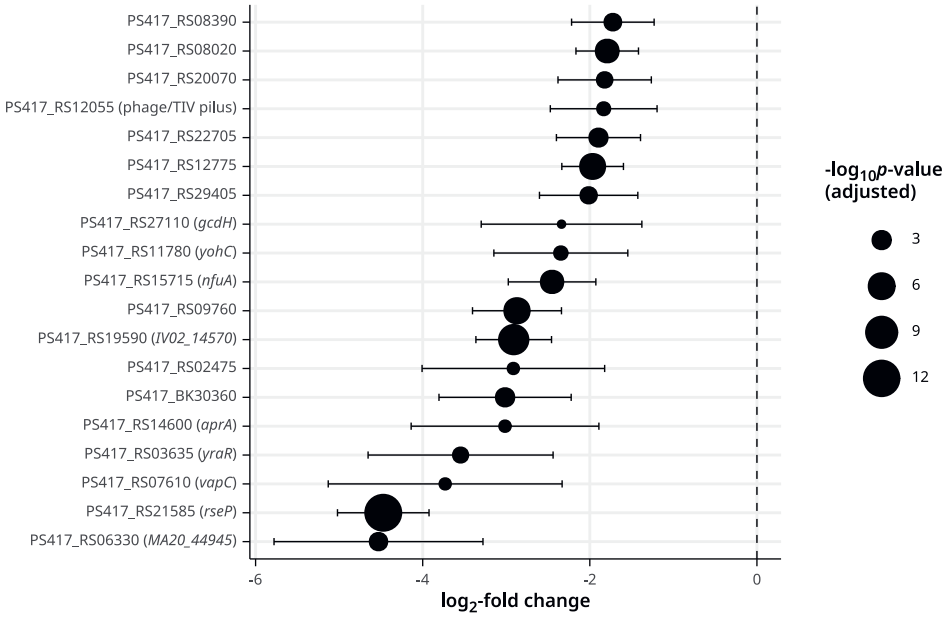


Figure 4. WCS417 activates stress and persistence-associated genes in bulk soil. Log_2-fold changes in gene expression for all upregulated genes in bulk soil compared to rhizosphere environments. Each row represents a DEG (Benjamin-Hochberg adjusted *p*-value < 0.05). Statistical significance for each gene is shown with different point sizes. Error bars indicate standard error of the fold-change estimate. Genes upregulated in soil include markers of general stress responses (e.g., *rseP*, *nfuA*, *PS417_RS06330*), nutrient limitation, and envelope stress, as well as genes involved in redox regulation, biofilm formation (*yohC*, *gcdH*, *aprA*), and microbial interactions, including prophage-associated loci.

Beyond stress adaptation, several soil-enriched genes are associated with early stages of biofilm development or surface colonisation. *yohC* (*PS417_RS11780*) is linked to redox stress, biofilm initiation, and pyomelanin production (Hunter and Newman, 2010). *gcdH* (*PS417_RS27110*), encoding a glutaryl-CoA dehydrogenase, is involved in biofilm development and is regulated by the Gac-Rsm system (Waite et al., 2005). These expression patterns suggest that in the soil environment, WCS417 may shift toward a sessile, matrix-producing lifestyle as a strategy to withstand environmental stress and persist in the absence of a host.

In parallel, several genes enriched in soil suggest that WCS417 is actively engaging with other microbes or viral elements. The genes *PS417_RS12775* and *PS417_RS12055* are both more highly expressed in soil and their homologues in *P. fluorescens* In5 (AL066_07090 and AL066_31290) have been associated with the transcriptional response to the fungal pathogen *Rhizoctonia solani* (Hennessy et al., 2017). Additionally, *PS417_RS09760* shows distant homology with a gene in the prophage ΦAH14a (WP_024075235) in *Pseudomonas gorinensis* TKP (Dziewit and Radlinska, 2016), a plant growth-promoting pseudomonad (Gonzalez et al., 2025).

Overall, the transcriptional profile in soil suggests that WCS417 activates

stress-response pathways and ecological interaction mechanisms to survive and persist outside the host root environment. This suggests a general reprogramming toward environmental sensing, resilience, and microbial competition.

WCS417 exhibits stress-associated gene expression in the rhizosphere of small plants

To understand how WCS417 responds to different root environments, we compared its transcriptional profiles between rhizospheres of small and big plants. A total of 181 genes were differentially expressed at a nominal *p*-value < 0.05, of which 72 were significantly upregulated in big-plant rhizospheres and 109 in small-plant rhizospheres (**Figure 5A**). After multiple testing correction, 31 genes remained differentially expressed significantly, from which 22 were upregulated in the rhizosphere of small plants and nine in big plants (**Figure 5B**, **Supplementary Table 5**).

Several of the genes more highly expressed in the rhizospheres of small plants were associated with redox imbalance and envelope stress responses. For instance, *nemR* (*PS417_RS14900*), a redox-sensitive transcriptional repressor responsive to reactive electrophilic species, was upregulated in small plants. Similarly, *cysI* (*PS417_RS15740*), encoding a subunit of sulphite reductase involved in oxidative stress defence and sulphur metabolism, and *dedA* (*PS417_RS03585*), implicated in membrane integrity and ion homeostasis, were also enriched in this group. The expression of these genes suggests that WCS417 experiences environmental or host-derived stress in the rhizosphere of small plants.

In line with this, multiple genes involved in biofilm formation and envelope stress were also significantly upregulated in WCS417 in the rhizosphere of small-plant roots. The *algU* sigma factor (*PS417_RS20735*), along with its partners *algB* (*PS417_RS27240*) and *algR* (*PS417_RS01005*), were upregulated, consistent with activation of the AlgU-controlled envelope stress response (Damron et al., 2009; Markel et al., 2016). This pathway supports alginate-based mucoid biofilm development and is often induced under hostile environmental conditions (Zhang et al., 2024). The biosynthesis gene *algD* (*PS417_RS22945*) was also upregulated in rhizospheres overall, though its expression did not significantly differ between small and big plant samples. Interestingly, the anti-sigma factor transcripts *rseA* (*PS417_RS20730*) and *rseB* (*PS417_RS20725*), which inhibit AlgU activity (Alba and Gross, 2004; Cezairliyan and Sauer, 2007), were also upregulated, suggesting a possible regulatory bottleneck or population-level heterogeneity in stress response activation.

Other genes enriched in the rhizosphere of small plant roots also supported a biofilm-associated state. These included *lrgA* (*PS417_RS04945*), an antiholin-like gene involved in regulating autolysis and eDNA release (Hynen et al., 2021), and *pgm* (*PS417_RS14655*), a phosphoglucomutase involved in

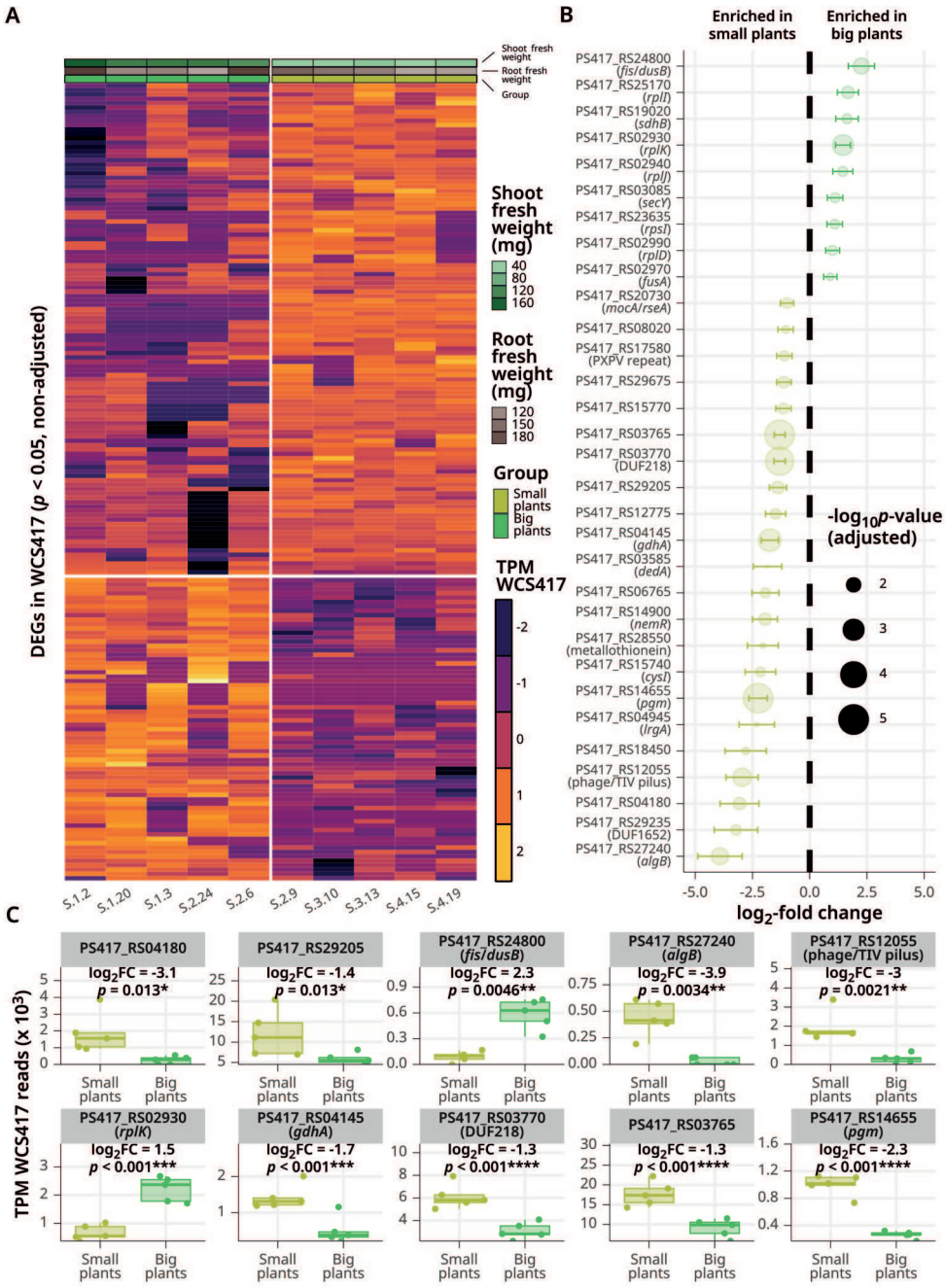


Figure 5. Transcriptional activity of WCS417 shows different patterns in the rhizospheres of big versus small plants. (A) Heatmap of all DEGs with a raw *p*-value < 0.05 in each sample. Cell colours indicate the transcripts per million (TPM) in WCS417 scaled per gene (rows). (B) Log_2-fold changes in gene expression for all DEGs with an adjusted *p*-value < 0.05 (Benjamin-Hochberg correction). Error bars indicate the standard error for the fold-changes. (C) Boxplots showing the TPM of WCS417 reads for the top ten significant DEGs in the rhizospheres of big versus small plants. All of them show the log_2-fold change and their adjusted *p*-value.

carbohydrate metabolism and exopolysaccharide synthesis (Ma et al., 2012). *PS417_RS04180*, a FleQ-regulated gene likely involved in exopolysaccharide export (Xiao et al., 2021), was also upregulated, as were two *cheV* homologs *PS417_RS06800* and *PS417_RS08700* associated with chemotaxis and surface adhesion (Fabian et al., 2024).

Collectively, these findings suggest that in the rhizosphere of small plant roots, WCS417 undergoes a shift toward a stress-associated, sessile lifestyle characterised by envelope stress signalling and biofilm-related gene expression. This likely reflects an adaptation to unfavourable or limiting conditions within these root environments.

WCS417 displays an active and motile transcriptional profile in the rhizosphere of big plants

In contrast to the stress- and biofilm-associated expression profile observed in small plant roots, WCS417 in big plant rhizospheres exhibited a distinct transcriptional program marked by upregulation of genes involved in protein synthesis, energy metabolism, and cellular growth. Of the 31 DEGs that remained significant after multiple testing correction, nine were enriched in the rhizosphere of big plant roots (**Figure 5B**), many of which are involved in core biosynthetic and respiratory processes (**Supplementary Table 5**) and leading to an overall significant enrichment of cell cycle-related genes (**Supplementary Figure 8**; **Supplementary Table 6**).

A strong signature of translational activity was evident, with numerous ribosomal protein genes significantly upregulated, including *rplC* (*PS417_RS02985*), *rplD* (*PS417_RS02990*), *rplI* (*PS417_RS25170*), *rplJ* (*PS417_RS02940*), *rplK* (*PS417_RS02930*), *rplS* (*PS417_RS05485*), *rpsH* (*PS417_RS03055*), *rpsI* (*PS417_RS23635*), *rpmE* (*PS417_RS25795*), *rpmB* (*PS417_RS00785*), and *yceD* (*PS417_RS06910*). These genes showed $\log_2$ fold changes ranging from ~0.7 to >2.7, with many exhibiting adjusted *p*-values < 0.05. Also upregulated were *rpoA* (*PS417_RS03105*), encoding the RNA polymerase α-subunit, and *fusA* (*PS417_RS02970*; EF-G), a key translation elongation factor. All these genes fall under the COG J category (translation, ribosomal structure, and biogenesis), and their coordinated upregulation points to a globally enhanced protein synthesis program.

The transcriptional regulator *fis* (*PS417_RS24800*) was also strongly enriched. Known for its role in nucleoid structuring and ribosomal RNA transcription, *fis* further supports the interpretation of a transcriptionally active, rapidly proliferating bacterial state in the rhizosphere of big plants (Deng et al., 2017; Dillon and Dorman, 2010).

Similarly, energy metabolism was upregulated, with elevated expression of multiple ATP synthase subunit genes – i.e., *atpA* (*PS417_RS00085*), *atpD* (*PS417_RS00095*), *atpG* (*PS417_RS00090*) and *atpH* (*PS417_RS00080*) – and electron transport chain components – i.e., *sdhA* (*PS417_RS19025*), *sdhB* (*PS417_RS19020*), *cyoB* (*PS417_RS05015*) and *cyoE* (*PS417_RS05030*).

The citric acid cycle gene *sucA* (*PS417_RS19015*) was also induced, consistent with increased aerobic respiration. In line with this heightened cellular activity, *secY* (*PS417_RS03085*), encoding a central component of the Sec-dependent protein secretion system, and the signal peptide peptidase gene *sppA* (*PS417_RS06900*) were also upregulated, reflecting increased protein export activity in the rhizosphere of big plants. These transcriptional changes reflect elevated energy demands and protein translocation activity in the rhizosphere medium.

Genes linked to nucleotide biosynthesis – including *purC* (*PS417_RS20675*), *purD* (*PS417_RS24815*), *purR* (*PS417_RS09860*), and *xpt* (*PS417_RS00870*) – and amino acid metabolism – including *glnS* (*PS417_RS11315*), *thiI* (*PS417_RS26070*), *gdhB* (*PS417_RS13080*), and *pepV* (*PS417_RS18785*) – were also more highly expressed in big plant roots. In addition, upregulation of envelope-related genes such as *lpxA* (*PS417_RS21560*), *murC* (*PS417_RS23140*), *lptB* (*PS417_RS23430*), and *oprI* (*PS417_RS19485*) indicates increased investment in cell envelope biosynthesis and remodelling.

Lastly, genes involved in motility and chemotaxis, including *flgD* (PS417_RS06820), *flgE* (PS417_RS06825), *fleS* (PS417_RS08500), *fleR* (PS417_RS08505), and *chpC* (PS417_RS01880), were more highly expressed in big plant roots. Although these increases did not remain significant after multiple testing correction, their consistent upregulation reflect enhanced surface exploration or navigation within a propitious root environment.

Together, these findings indicate that WCS417 experiences a metabolically favourable and growth-permissive niche in the rhizosphere of big plants, marked by high transcriptional activity, respiratory metabolism and biosynthetic investment.

Roots of small plants exhibit nutrient limitation and redox stress

To assess whether differences in plant size were associated with distinct physiological states of the host, we next examined *Arabidopsis* root gene expression profiles in small and big plants. *Arabidopsis* root transcriptomes did not show clear global separation between small and big plants in PCA space (**Figure 6A**), yet differential expression analysis identified 2,223 DEGs between big and small plants (**Figure 6B**). After multiple testing correction, 141 DEGs remained significant, with 94 and 47 genes upregulated in big and small plant roots, respectively (**Figure 6C**; **Supplementary Table 7**).

Among the genes more highly expressed in the roots of small plants, a dominant pattern was the upregulation of stress-responsive, nutrient-scavenging, and cell wall-related transcripts, suggesting a physiologically constrained state. The most significantly enriched transcript in small roots was *IRM1* (*AT5G65040*; **Figure 6D**), involved in root hair morphogenesis and iron deficiency responses. Similarly, *SULTR1;3* (*AT1G22150*), a high-affinity sulphate transporter responsive to sulphur starvation, was strongly induced.

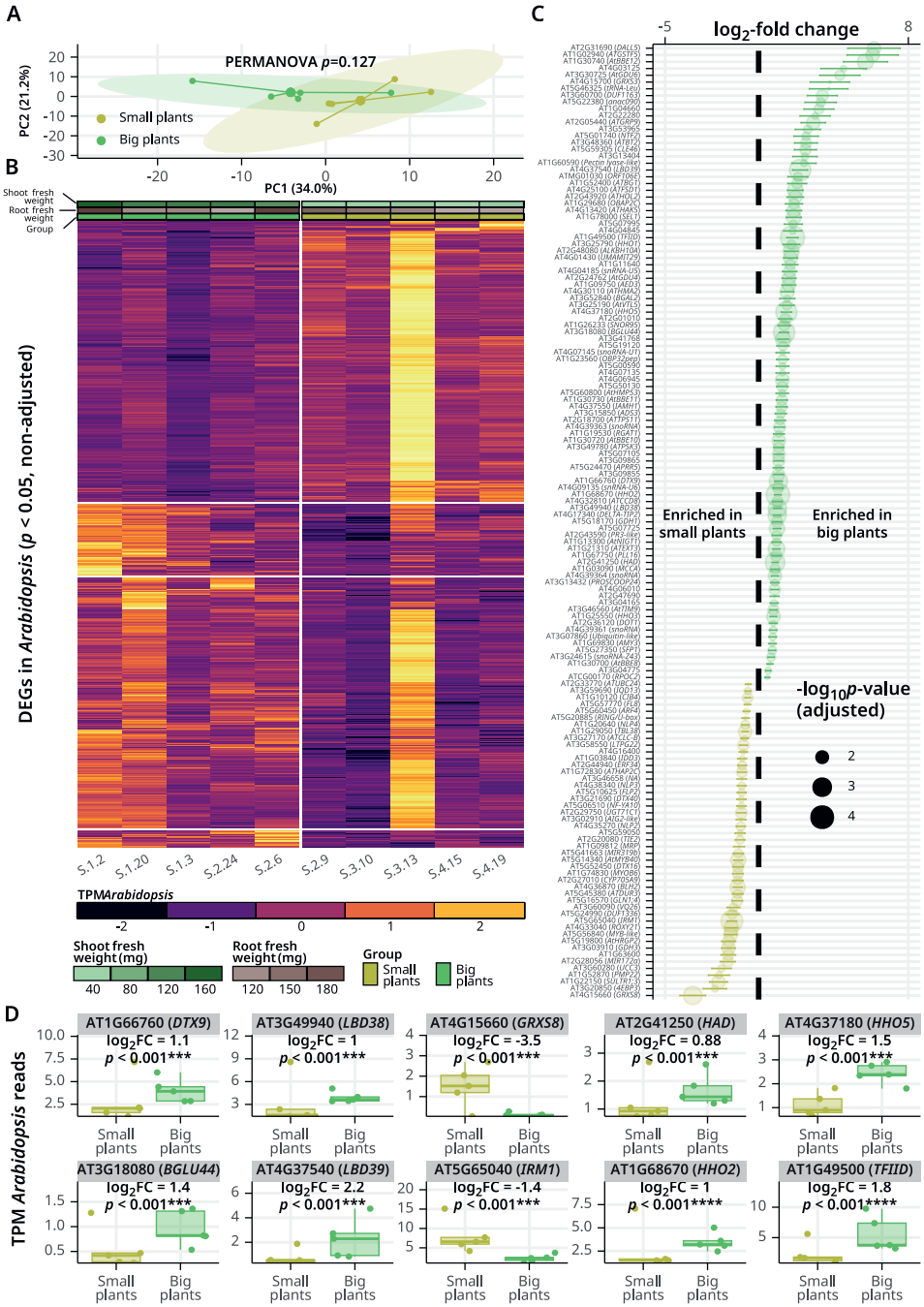


Figure 6. Transcriptional activity of WCS417-colonised *Arabidopsis* roots differs between big and small plants. (A) Scatterplot of PCA scores for *Arabidopsis* gene expression in WCS417-colonised roots of big and small plants. The significance of the differences between groups is assessed with a PERMANOVA test and shown in the plot. (B) Heatmap of all DEGs (rows) with a raw *p*-value < 0.05 in each sample (columns). Cell colours indicate the TPM in *Arabidopsis* scaled per gene. (C) Log_2-fold changes in gene expression for all DEGs with an

adjusted *p*-value < 0.05 (Benjamin-Hochberg correction). Error bars indicate the standard error for the fold-changes. (D) Boxplots showing the TPM of *Arabidopsis* reads for the top 10 significant DEGs in the roots of big versus small plants. All of them show the $\log_2$-fold change and their associated adjusted *p*-value. GO: Gene Ontology; GO:BP, biological process; GO:MF, molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes.

Nitrogen-related genes such as *GLN1;4* (*AT5G16570*) and *GDH3* (*AT3G03910*), which mediate ammonium assimilation and nitrogen recycling, were also upregulated. Transcriptional NIN-like regulators of nitrate uptake and signalling – *NLP2*, *NLP3*, and *NLP4* (*AT4G35270*, *AT4G38340*, *AT1G20640*) – alongside the urea transporter *ATDUR3* (*AT5G45380*), further support a nutrient acquisition response under limitation (**Supplementary Figure 9**).

Redox-related processes were also prominently enriched in small plants, mirroring bacterial stress responses in the same roots. *GRXS8* (*AT4G15660*), a glutaredoxin involved in iron–sulphur cluster biogenesis and redox balance, was among the most strongly upregulated genes. Other redox-related regulators included *ROXY21* (*AT4G33040*), a CC-type glutaredoxin regulating stress pathways, and *ATHAP2C* (*AT1G72830*), involved in MAP kinase signalling. The *VQ26* gene (*AT3G60090*), a WRKY co-regulator induced under oxidative stress, was also enriched. Members of the DTX detoxification transporter family – i.e., *DTX16* (*AT5G52450*), *DTX29* (*AT3G26590*), *DTX39* (*AT4G21910*), *DTX40* (*AT5G52450*) – were consistently overexpressed, highlighting active detoxification processes under stress.

A broad suite of transcriptional regulators was also more highly expressed in small roots, including *AtMYB40* (*AT5G14340*), *BLH2* (*AT4G36870*), and a MYB-like gene (*AT5G56840*), suggesting shifts in developmental and metabolic control. Hormone-related regulators were also differentially expressed, such as *ARF4* (*AT5G60450*), an auxin response factor, and transcription factors *ERF34* (*AT2G44940*), *NF-YA10* (*AT5G06510*), and *CIB4* (*AT1G10120*). Together, these patterns suggest a coordinated transcriptional response to stress, redox imbalance, and growth limitation.

Enrichment of cell wall biosynthesis and remodelling genes further supports developmental constraint in small plants. Genes such as *TBL38* (*AT1G29050*), *LTPG22* (*AT3G58550*), and *AtHRGP2* (*AT5G19800*), involved in lipid and glycoprotein modification of the wall matrix, were upregulated, potentially reflecting altered surface structure. Moreover, *IDD3* (*AT1G03840*), involved in developmental regulation; and *UCC3* (*AT3G60280*), implicated in lignin polymerisation and xylem integrity, also showed higher expression in small plants.

In summary, small-plant roots exhibit transcriptional signatures of nutritional- and redox-stress adaptation, and modifications of the root architecture. When considered alongside gene expression profiles of WCS417 in the same roots – which includes oxidative stress responses, envelope remodelling, and biofilm formation – these findings suggest a microenvironment that is both physiologically limiting and likely disruptive to optimal plant–microbe mutualistic interactions.

Roots of big plants display nutrient transport, metabolic activity, and reduced stress signalling

Contrasting the transcriptional stress signature observed in small plants, roots of big plants exhibited upregulation of genes associated with nutrient assimilation, intracellular transport, secondary metabolism, transcriptional regulation, and biosynthetic activity.

Genes related to nutrient exchange and solute transport were prominently enriched. These included *UMAMIT29* (*AT4G01430*), an amino acid transporter, and *DELTA-TIP2* (*AT4G17340*), a vacuolar intrinsic protein implicated in tonoplast-mediated transport. The glutamate export-related genes *AtGDU4* (*AT2G24762*) and *AtGDU6* (*AT3G30725*), along with *MCCA* (*AT1G03090*), involved in organic acid metabolism, support the interpretation of a metabolically active nutrient flux in big roots. The gene *GDH1* (*AT5G18170*), encoding glutamate dehydrogenase, was also upregulated, reflecting integration of nitrogen and carbon metabolism, contrasting with the ammonium-focused profile seen in small plants.

Big plant roots also displayed increased expression of genes involved in transcription, translation, and signalling. These included the chloroplast RNA polymerase gene *RPOC2* (*ATCG00170*), the nitrate-induced transcription factor *DOT1* (*AT2G36120*), and a range of small noncoding RNAs (e.g., *AT4G39364*, *AT4G04185*, *AT4G07145*), consistent with an active transcriptional and biosynthetic state. Upregulation of *ATPSK3* (*AT3G49780*), a phytosulfokine precursor associated with cell expansion, and *AtTIM9* (*AT3G46560*), involved in mitochondrial import, further support enhanced cellular activity.

Multiple transcriptional regulators were enriched, notably *homeodomain-leucine zipper* (HD-ZIP) transcription factors of the HHO family: *HHO1*, *HHO2*, *HHO3*, and *HHO5* (*AT3G25790*, *AT1G68670*, *AT1G25550*, *AT4G37180*). These are involved in nutrient and developmental regulation. In parallel, members of the LBD transcription factor family, *LBD38* (*AT3G49940*) and *LBD39* (*AT4G37540*), were also significantly upregulated, consistent with roles in root architecture and nitrogen response. Other enriched regulators included *IAMH1* (*AT4G37550*), associated with auxin biosynthesis; *AtNIGT1* (*AT1G13300*), a GARP-type nitrate-responsive factor; and *SEL1* (*AT1G78000*), involved in signalling and stress modulation.

Redox-associated metabolic genes were also enriched. These included genes for AtBBE-like oxidoreductases *AtBBE10*, *AtBBE11*, and *AtBBE12* (*AT1G30720*, *AT1G30730*, *AT1G30740*), along with the demethylase *ALKBH10A* (*AT2G48080*), potentially linked to epigenetic regulation. *GRXS3* (*AT4G15700*), a glutaredoxin, was enriched only in big roots, suggesting redox buffering in the absence of strong stress signals. Consistent with the downregulation of stress genes in WCS417 colonising big roots, few canonical plant stress markers were enriched in big roots. The presence of only *GRXS3*, in contrast to strong redox signalling in small plants, points to a less oxidative microenvironment and more favourable to microbial colonisation.

Interestingly, the β-glucosidase gene *BGLU44* (*AT3G18080*; **Figure 6D**; **Supplementary Table 8**) was among the most significantly upregulated transcripts. BGLU44 hydrolyses esculin into esculetin, a coumarin with well-established roles in iron mobilisation and microbial recruitment. The elevated expression of *BGLU44* supports a model in which big roots actively secrete coumarins that help shape a growth-promoting microbiome.

Together, these findings support a model in which big-plant roots display enhanced nutrient uptake, biosynthetic activity, transcriptional regulation, and coumarin-mediated microbiome modulation, alongside a reduction in stress signalling. This profile contrasts with the transcriptional constraints observed in small plants and likely reflects a microenvironment conducive to mutualistic plant–microbe interactions.

Discussion

Our combined transcriptomic profiling of WCS417 across bulk soil, small-plant, and big-plant rhizospheres reveals a strong environmental dependency in microbial physiology, shedding light on why bioinoculants often show inconsistent efficacy in agricultural settings.

In bulk soil, WCS417 expresses a suite of genes associated with oxidative stress response, envelope stress sensing, and microbial interaction mechanisms. In this environment, WCS417 strongly expresses *nfuA*, a gene encoding an iron–sulphur (Fe–S) cluster carrier protein that supports enzyme function under oxidative and metal stress. In *Escherichia coli* and *Pseudomonas aeruginosa*, *nfuA* is induced by iron starvation and enhances cell survival under such conditions (Angelini et al., 2008; Eng et al., 2020; Rincon-Enriquez et al., 2008; Romsang et al., 2018), potentially by inhibiting chemotactic responses and promoting a quiescent state (He et al., 2025). Furthermore, the *PS417_RS22705* gene encodes a putative ribosomal subunit interface protein that regulates expression of ribosomal genes under control of Fur in *P. syringae* pv. *tomato* DC3000 (Butcher et al., 2011). In *P. syringae*, expression of this gene represses growth in lowly nutritional conditions (Deng et al., 2014). Also in this organism, the product of *PS417_RS19590* is found in extracellular vesicles released under nutrient-limited conditions, suggesting that this gene is also functional in this context (McMillan and Kuehn, 2023). Overexpression of these genes indicates that WCS417 may encounter iron limitation in the soil environment and compensates this scenario by halting growth and upregulating components of Fe-S cluster repair.

Beyond general stress responses, several enriched genes suggest that WCS417 is actively responding to other organisms in soil. In *P. fluorescens* In5, homologs of the genes *PS417_RS12775* and *PS417_RS12055* are strongly upregulated during antagonistic interactions with *Rhizoctonia solani* (Hennessy et al., 2017), and may participate in competitive or defensive behaviours. Notably, *PS417_RS12055* is a predicted phage infection protein regulated by

MexT, a master regulator of the stringent response and quorum-sensing-regulated virulence (Figueroa et al., 2025), which further supports its view as related to interaction with other microbes. Another phage-related gene enriched in soil is *PS417_RS09760*, which is homologous to a gene in the prophage ΦAH14a in *Pseudomonas* sp. ANT_H14 (Dziewit and Radlinska, 2016), although potential advantages to the host are not known. Phage-derived genes have already been shown to act as competitive traits in pseudomonads (Vacheron et al., 2021) and their overexpression might confer competitive advantage over other organisms in soil.

An important WCS417 gene overexpressed in soil is *aprA*, which shows clear connections with microbe-microbe and host-microbe interactions. Its product is an extracellular metalloprotease, known as a serralysin, under control of the GacA-Rsm system in pseudomonads (Song et al., 2015). In *P. chlororaphis* 30–84, *aprA* is upregulated in the presence of a phenazine derivative and it stimulates the release of eDNA into the extracellular environment, where it allows cell adhesion and enhances biofilm structure (Wang et al., 2016, 2013). In *P. protegens* Pf-5, AprA is required to inhibit *Streptomyces scabies*, the etiological agent of potato scab (Pacheco-Moreno et al., 2021). In the gut of *Drosophila* , expression of the *aprA* gene in *P. entomophila* (*PSEEN1550*) is induced by the *Pseudomonas* virulence factor (Pvf) (Acken and Li, 2023), where it rapidly cleaves pro-monalysin, a pore-forming toxin, into the active form monalysin, leading to enhanced toxicity (Opota et al., 2011). Furthermore, *aprA* has been shown to be an important plant virulence factor in the pathogens *P. syringae* pv. *tomato* DC3000 and *P. aeruginosa* PAO25, a derived strain from PAO1 (Bardoel et al., 2011; Pel et al., 2014). When secreted, AprA degrades flagellin monomers in the surrounding environment, effectively avoiding the activation of FLS2 and the subsequent onset of innate immunity in the host (Pel et al., 2014)

Regulation of biofilm formation is further represented in soil. The *gcdH* gene encodes a gluraryl-CoA dehydrogenase also under control of the Gac-Rsm system and is upregulated during initial stages of biofilm development (Huertas-Rosales et al., 2021; Waite et al., 2005). Furthermore, the homolog of *yohC* in *P. fragi* 1793 (*CJU75_18560*) is overexpressed during biofilm initiation specifically, not during biofilm maturation nor dispersal (Wickramasinghe et al., 2021). In pseudomonads, YohC is required for the production of pyomelanin (Hunter and Newman, 2010) and its expression is controlled by the complex Las and Rhl quorum sensing regulons (Rasmussen et al., 2005b) and disturbed by antimicrobial compounds and quorum sensing inhibitors (Rasmussen et al., 2005a; Skindersoe et al., 2008). Pyomelanin is necessary for resistance against oxidative stress, resistance against toxic compounds and iron scavenging (Ahmad et al., 2016; Y. Li et al., 2021; Roca et al., 2008; Zheng et al., 2013), and prevention of phage replication (Mooij et al., 2007), suggesting physiological control of phage expression by WCS417 in soil.

These findings suggest that WCS417 in soil is engaged in a defensive and competitive program aimed at surviving a harsh, microbially crowded, and nutrient-limited environment, as suggested by the successful survival rates

of pseudomonads in soil (Molina et al., 2000). The expression of biofilm-related genes further supports the idea that WCS417 survives in fine pellicles around soil pores, all of which may enable the bacterium to persist in bulk soil until it encounters more favourable niches like plant roots.

In *Arabidopsis* roots, our data reveal two contrasting transcriptional programs depending on shoot size, accompanied by marked shifts in the transcriptome of WCS417. These coordinated host–microbe responses imply that small and big plants define distinct microenvironments in the rhizosphere, shaping both stress physiology and microbial behaviour.

In small plants, the transcriptome of *Arabidopsis* roots bears hallmarks of nutrient deprivation, redox imbalance, and stress-induced developmental arrest. Strong upregulation of *IRM1* and nutrient transporters (e.g., *SULTR1;3*, *GLN1;4*, *ATDUR3*) signals a compensatory effort to acquire limiting resources. Upregulation of redox-regulating genes such as *GRXS8* and *ROXY21* further suggests oxidative stress, a condition corroborated by the co-expression of stress-responsive transcription factors (e.g., *ATHAP2C*, *CIB4*, *BLH2*) and remodelling of the cell wall. Although canonical ROS-generating enzymes were not detected among the top hits, the coordinated upregulation of stress- and redox-associated genes is consistent with a condition of increased reactive oxygen species (ROS) levels and cellular efforts to mitigate oxidative damage.

Interestingly, roots of small plants exhibit a higher investment in cell wall strengthening, as reflected by the overexpression of genes such as *TBL38*, *LTPG22*, *AtHRGP2* and *UCC3,* which are involved in lipid modification, glycoprotein cross-linking, and lignin polymerisation of the wall matrix. This transcriptional profile points to reinforced cell wall structure and reduced surface plasticity in small plants. This observation is consistent with previous work showing that successful colonisation by WCS417 is associated with repression of cell wall biogenesis genes, a process that is hypothesised to facilitate root architectural changes that ultimately lead to growth promotion (Verbon et al., 2023).

Collectively, these changes point to a root system under physiological strain, characterized by limited nutrient flux, altered redox signalling, and modified surface architecture. The enhanced expression of cell wall-associated genes in small plants also aligns with the absence of a growth-promoting phenotype in this group, and supports the notion that cell wall remodelling is a key host component of effective WCS417-mediated growth promotion.

Upon colonisation of small plant roots, WCS417 transitions into a state characterised by partial but incomplete activation of the envelope stress response, likely in response to the same stimulus that triggered stress responses in the plant or by direct interaction with the plant itself. Specifically, we observed upregulation of genes related to the AlgU-mediated stress response. This response is typically linked to the development of a mucoid, alginate-based biofilm that protects embedded cells (H. Wang et al., 2021), although it has also been associated with plant immune suppression, increased production

of virulence factors, and enhanced motility in pseudomonads (H. Wang et al., 2021; Zhang et al., 2024).

In the roots of small plants, both *algU*, and *algR* are upregulated, consistent with the initiation of the AlgU response and suggesting that WCS417 enters a stress-associated colonisation state (Wang et al., 2024). However, this activation coincided with increased expression of the anti-sigma regulators *rseA* and *rseB*. This seemingly contradictory pattern may reflect temporal or spatial heterogeneity within the bacterial population, with subsets of cells initiating the stress response while others remain in recovery or inhibition phases. Alternatively, co-upregulation of both AlgU and its inhibitors may also indicate the presence of a negative feedback loop or regulatory bottleneck, as described in systems where RseP levels constrain the transition from stress sensing to full response execution (Delgado et al., 2018).

Importantly, stimulation of this system may be a consequence of immune pressure, as supported by the parallel induction of redox-related genes in *Arabidopsis* (e.g. *GRXS8*, *ROXY21*) alongside nutrient-stress markers. Consistent with this interpretation, in *Pseudomonas syringae* pv. *tomato* DC3000, prior activation of pattern-triggered immunity in the host by flagellin causes broad disruption of the AlgU-dependent transcriptional program *in planta*, and this misregulation is accompanied by reduced bacterial growth in the host (H. Wang et al., 2022). These observations suggest that successful host colonisation may require a properly coordinated and fully functional AlgU response, and that host immune priming or stress can disrupt this regulatory system. In this context, the partial and potentially unstable activation of the AlgU regulon in small plants may reflect a hostile rhizosphere niche in which immune pressure and nutrient stress constrain beneficial microbial activity, resulting in an intermediate and ultimately unsuccessful colonisation state.

Further evidence that WCS417 persists in the roots of small plants in a biofilm state is the overexpression of the anti-holin-like *lrgA* gene, which inhibits autolysis by controlling murein hydrolase activity, is required during biofilm maturation for control of eDNA release and stationary-phase survival (Ma et al., 2009, 2009; Sharma-Kuinkel et al., 2009; Tognon et al., 2019; Wang et al., 2016), and *cheV*, which is involved in adhesion (X. Xu et al., 2021) and coordinates complex chemosensory cues (Ferrández et al., 2002; Matilla et al., 2025).

The consistency between bacterial and plant stress responses reinforces the idea that small-plant roots are embedded in an unfavourable microenvironment in which WCS417 may struggle to engage in mutualistic colonisation, stalling the bacterium in a low-activity, stressed state. Collectively, in soil and roots of small plants, WCS417 enters a stress-resilient, low-activity state, investing in envelope repair, biofilm formation, and survival. Notably, a few genes enriched in small-plant roots (i.e., *PS417_RS12055*, *PS417_RS12775*, *PS417_RS08020*) were also significantly enriched in soil, suggesting that specific aspects of the bacterial response in small plants may resemble

stress-adaptive programs used in host-independent environments. While overall expression profiles remain distinct from soil, this overlap supports the idea that the environment around small-plant roots impose stress on microbial colonisers, potentially disrupting the establishment of a stable mutualistic interaction and limiting its efficacy as a plant-growth-promoting agent.

In contrast, roots of big plants show signs of active development and controlled plant-microbe interactions. Expression of several LBD and HHO transcription factors was enriched, aligning with developmental coordination and nutrient-responsive signalling. Interestingly, big plants overexpressed *BGLU44*, a β-glucosidase involved in the release of cytokinins and bioactive coumarins, particularly esculetin (Xu et al., 2004; Yang et al., 2021). Coumarins play key roles in shaping the rhizosphere microbiome through their antimicrobial properties and by promoting colonisation by beneficial microbes like WCS417 (Stringlis et al., 2019), which suggests a chemical environment that favours colonisation. Interestingly, inoculation of WCS417 on the roots of *A. thaliana* accession Col-0 induced a 5-6-fold increased exudation of esculin compared to uninoculated plants (Stringlis et al., 2018). Additionally, *BGLU44* is overexpressed in root endodermal cells upon interaction with WCS417, a tissue which has been shown to play critical roles in the interaction with the host plant (Verbon et al., 2023). This coordination highlights that big plants entered into a mutualistic state with the bacterium. Furthermore, the release of cytokinins by BGLU44 in the roots might contribute to the growth promotion induced by WCS417 inoculation (Ko et al., 2014; Wu et al., 2021), as well as influence microbial colonisation patterns via changes in root exudation (Yoneyama et al., 2020). The elevated expression of *BGLU44* in big plants is therefore consistent with a more dynamic and growth-supportive root environment, capable of actively engaging with a favourable microbial community.

A very relevant gene enriched in roots of big plants in terms of plant-microbe interactions is *RPP13L4* (*AT3G50950*). In potato, this gene provides disease resistance to *Streptomyces scabies*, etiological agent of potato scab (Yuan et al., 2025), and in *Arabidopsis* it has been reported to confer resistance to a broad range of pathogens, such as *Peronospora parasitica* (Bittner-Eddy et al., 2000), *Xanthomonas campestris* (Wang et al., 2015) and *P. syringae* via the identification of type III effectors (Lewis et al., 2010; Seto et al., 2017). The broad range activity of this gene suggests that its product is an important modulator of host responses, and likely plays a central role in stabilizing effective colonisation and promoting bioinoculant efficacy.

Contrary to small plants, WCS417 in big plant roots expresses genes associated with active growth, protein synthesis, transport, and symbiosis. The absence of stress-associated gene expression, together with the induction of ribosomal and redox-stabilising factors, supports a model of metabolic activation and mutualistic engagement. The bacterium appears to be experiencing a permissive and supportive niche. Indeed, big plants displayed transcriptional signatures of active growth, nutrient assimilation, and immune regulation. Upregulation of *GDH1*, *DELTA-TIP2*, and *UMAMIT29*

suggests sustained organic nitrogen turnover and robust intracellular transport. This interpretation is further supported by recent work showing that glutamine, a central transport form of organic nitrogen, can transiently leak from roots and act as a key driver of bacterial colonisation (Tsai et al., 2025). In that study, enhanced glutamine export and disruption of endodermal barriers increased glutamine availability in the rhizosphere, shaping microbial transcriptional responses and promoting bacterial establishment. The coordinated upregulation of glutamine transport genes in big plants is consistent with such an active, glutamine-centred nitrogen flux, reinforcing the view that successful WCS417 colonisation coincides with a metabolically favourable root niche, and contrasting with the ammonium-focused profile seen in small plants.

Importantly, this host-side metabolic state is also consistent with the bacterial transcriptional program. The recent study by Tsai and co-authors highlights chemotaxis toward glutamine-rich microenvironments as a key mechanism by which bacteria colonise permissive roots. In line with this, WCS417 shows coordinated upregulation of motility- and chemotaxis-associated genes, including *fleS*, *fleR*, *flgD*, *flgE*, and *chpC*, suggesting an enhanced capacity to sense and navigate host-derived nutrient gradients. Notably, *chpC*, which showed the strongest induction ($log_2FC = 3.95$), is a chemotaxis-related protein and plays a key role in controlling biofilm expansion in *Pseudomonas aeruginosa* (Nolan et al., 2020), has been implicated in plant colonisation by *Clavibacter michiganensis* (Stork et al., 2008) and is highly expressed in the rhizosphere by *P. putida* KT2440 (Matilla et al., 2007). Although some genes did not all pass multiple-testing correction (i.e., *flg* genes), perhaps due to the local nature of these chemotactic signals, their concerted induction supports a model in which big roots provide structured chemical cues that promote active bacterial exploration and stable colonisation.

Expression of WCS417 in the roots of big plants was also characterised by a general upregulation of genes for ribosomal proteins and nucleotide and amino acid biosynthesis, suggesting a general activation of WCS417 metabolism and likely reflecting greater access to nutrients in the rhizosphere. Furthermore, WCS417 shows an active remodelling of the outer membrane. Notably, upregulation of *lpxA*, responsible for lipopolysaccharide (LPS) biosynthesis, potentially reflects an adaptation to plant-derived stress or a strategy to stabilise the outer membrane. Also enriched were genes involved in peptidoglycan and outer membrane component assembly. For example, *murC* participates in the early steps of peptidoglycan synthesis (Garde et al., 2021) and the *lptB* gene encodes an ATPase component of the Lpt (lipopolysaccharide transport) system, which delivers LPS to the outer membrane, further supporting enhanced envelope biogenesis (Sperandeo et al., 2007). Together, these findings suggest that WCS417 invests in strengthening or modifying its outer membrane in big plants, which may contribute to environmental resilience, immune evasion, or competitive interactions in the rhizosphere.

Interestingly, although WCS417 *aprA* expression in the rhizosphere was highest in soil and comparable among plants, its expression in the rhizosphere of big plants was nearly abolished (**Supplementary Figure 7**). This raises the question of whether this repression in the rhizosphere of big plants reflects a trade-off with increased motility and flagellin demand, or instead indicates that recognition by the host immune system is required for the growth-promoting phenotype.

Nevertheless, these features are hallmarks of a symbiotically engaged and metabolically active microbial population, reinforcing the idea that the roots of big plants offer a favourable niche that promotes mutualistic bacterial behaviours. This stands in sharp contrast to the stress-oriented expression profile observed in small plant roots and bulk soil. However, whether the distinct transcriptomic states of WCS417 are a cause or consequence of the different plant phenotypes remains unresolved. It is also still unclear whether these differences reflect points along a dynamic continuum of plant-microbe feedbacks or instead represent stable, diverging states. One possibility is that small plants are simply early-stage individuals in which growth promotion has not yet manifested – implying that, over time, their phenotype and microbial activity could resemble that of big plants. Alternatively, small plants may represent stalled or suboptimal developmental trajectories, potentially due to failed or ineffective microbial interactions.

Our results reveal that the behaviour of *P. simiae* WCS417 in the rhizosphere is highly plastic, shaped by environmental context: from stress-adapted survival in soil, to potential immune restriction and metabolic restraint in the rhizosphere of small plants, to a transcriptional program consistent with symbiotic mutualism in the rhizosphere of big plants. This continuum of microbial states offers a valuable framework for interpreting the context-dependent efficacy of bioinoculants. It highlights that variability in field performance is not solely dictated by microbial genetics or formulation, but also by the physiological state of the host plant and the local environment. Identifying the key drivers of microbial behaviour, particularly in the roots of small or underperforming plants, will be critical for designing more reliable and responsive plant-microbe systems in agriculture.

Materials and methods

Arabidopsis culture

Seeds of *Arabidopsis thaliana* accession Col-0 were surface-sterilised with chlorine gas in a glass chamber and subsequently aerated in a flow cabinet for 30 min to remove the residual chlorine (Van Wees et al., 2013). Sterile seeds were sown in agar plates containing 1x Murashige and Skoog (MS) medium supplemented with 0.5% sucrose. Plates were sealed twice with parafilm, let stratify for 48 h at 4°C, and positioned vertically in a growth chamber under short-day conditions (22 °C; 10 h light/14 h dark; light intensity 100 µmol/m^2/s) for two weeks.

Soil treatments, transplantations, and microbiome sampling

A single colony of *P. simiae* WCS417 (Pieterse et al., 2021) was spread onto five King's B (KB) medium agar plates and incubated overnight at 28 °C. Per plate, 5 ml of $MgSO_4$ 10 mM was used to scrape the microbial cell layer off the surface, and the resulting suspension was collected into a 50-ml tube. The cell suspension was washed three times by centrifugation at 3000 x g for 10 min, resuspending in 40-50 ml $MgSO_4$ 10 mM every step. The final cell pellet was resuspended in 20 ml of $MgSO_4$ 10 mM, and the bacterial density was measured by determining the optical density (OD) at 600 nm (equivalence: 1 OD_{600} unit, $8{\cdot}10^8$ CFU/ml).

In order to have a complex natural microbial community, we used topsoil collected from the Reijerscamp natural reverse in Wolfheze, The Netherlands (52°01'02.55", 5°77'99.83"), were natural *Arabidopsis* populations occur (Berendsen et al., 2018). After drying the soil, we added the microbial suspension to reach an inoculum density of 10^8 CFU/g of dry soil with a final volume of 100 ml/kg of soil. For the uninoculated treatment, the soil was moistened with $MgSO_4$ 10 mM at 100 ml/kg of soil.

Eighty two-weeks-old *Arabidopsis* seedlings were transferred from plates to individual pots (diameter = 5.5 cm and height = 5 cm; MXC5,5 Pöppelmann), each of them containing 100 g of inoculated soil. From these 80 plants, 20 were sown in uninoculated soil and 60 in inoculated soil. Five pots were left unplanted. All pots were placed on trays on individual Petri dishes to prevent cross-contaminations and kept in the growth chamber under the same conditions as mentioned above. The trays were kept closed with a lid for the first two days post-transplantation. For the remaining time of the cultivation period, the lids were removed, and the plants were watered every two days with 5-10 ml of tap water. Roots were eventually harvested at 21 dpi (5 weeks-old plants) following the *less-soil* method as described elsewhere (Poppeliers et al., 2024) – i.e., shaking the roots until no loosely-adhered soil particles remained – and rhizosphere and shoot fresh weight was

recorded. For soil samples, we collected approximately 0.25 g of soil from the unplanted pots. All samples were placed in 2-ml tubes containing glass beads, shaken twice at 20 Hz in a TissueLyser, snap frozen in liquid nitrogen, and stored at -80°C until nucleic acid extractions.

Nucleic acid extractions and library preparation

After classifying plant samples into the big and small categories, we selected five rhizospheres of big plants, five of small plants, and five bulk soil samples for extractions of nucleic acids and metagenomic and metatranscriptomic library preparations.

Genomic DNA and total RNA extractions were performed by Vertis Biotechnologie AG (Germany). Briefly, the samples were ground under liquid nitrogen and genomic DNA and total RNA were isolated using the RNeasy PowerSoil Total RNA Isolation Kit (Qiagen) from the tissue and soil powder.

Metagenomic library preparation was performed by the Utrecht sequencing facility. Briefly, the total DNA was mechanically fragmented by ultrasonication and libraries were prepared using the TruSeq DNA Nano Kit (Illumina) following manufacturer's instructions. Libraries were size-selected to enrich for fragments of approximately 350–450 bp prior to sequencing.

Metatranscriptomic preparations were also performed by Vertis Biotechnologie. Briefly, rRNA molecules (plant plus bacteria) and polyadenylated host mRNA were depleted from rhizosphere samples using an in-house developed proprietary protocol using homology probes. The ribodepleted RNA samples were first fragmented using ultrasound (1 pulse of 30s at 4°C). Then, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer. The first-strand cDNA was purified and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA.

For bulk soil samples, the total RNA was first fragmented using ultrasound (4 pulses of 30 s each at 4°C). Then, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as primer. The first strand cDNA was purified and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. The resulting cDNA was PCR-amplified using a high-fidelity DNA polymerase. From the pre-amplified cDNA libraries, single-strands were generated by digestion with Lambda Exonuclease (NEB). Ribosomal cDNA molecules (plant plus bacteria) were depleted using the same in-house developed protocol.

The resulting rRNA-depleted libraries from rhizosphere and bulk soil samples were then reamplified to about 10-20 ng/µl using a high-fidelity DNA polymerase. The cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and was analysed by capillary electrophoresis.

Sequencing of both metagenomics and metatranscriptomics libraries was performed by the Utrecht sequencing facility on an Illumina NovaSeq 6000 platform at S1 configuration with paired-end 150-bp reads.

Data processing and gene expression quantifications

Quality checks for both metagenomic and metatranscriptomic reads were performed with FastQC v0.11.9 and MultiQC v1.23 (Andrews, 2010; Ewels et al., 2016) and reads were subsequently filtered by quality, residual poly-G artifacts and length with cutadapt v5.0 (Martin, 2011). Adapter contamination was minimal due to tight size selection during library preparation.

Metagenomic reads were dehosted with Bowtie2 v2.5.4 (Langmead and Salzberg, 2012) using the latest TAIR10 *A. thaliana* genome release (Reiser et al., 2024). Resulting dehosted reads were mapped against the latest WCS417 genome (GCF_000698265.1) also with Bowtie2. Subsequently, we used a custom script to retrieve final WCS417 reads from Bowtie2 mappings, considering only end-to-end mappings with maximum one mismatch per read pair, regardless of number of mapping loci (Max1MM reads). Remaining reads were considered community reads (**Supplementary Figure 2**).

From metatranscriptomic reads, rRNA-derived reads were removed with SortMeRNA v4.3.7 (Kopylova et al., 2012) using the default v4.3 database. Non-rRNA reads were subsequently dehosted by mapping against the TAIR10 release of the *Arabidopsis thaliana* genome with HISAT2 v2.2.1 (Kim et al., 2019). And the resulting dehosted reads were mapped to the WCS417 genome with Bowtie2. As with metagenomic reads, end-to-end mappings with maximum one mismatch per read pair, regardless of number of mapping loci, were retrieved using a custom script. Reads that failed to map with this high confidence to the WCS417 genome were considered community reads.

Prior to calculating expression for WCS417, we updated the previous annotations by retrieving new ones with Bakta (Schwengers et al., 2021). From these new annotations, we added all of them that were not found in the previous GFF annotations within a vicinity of 5 nt from any other gene. Expression data from both *Arabidopsis* and WCS417 was obtained by mapping with HTSeq v2.0 (Putri et al., 2022) to the latest Oct 2024 Araport11 GFF annotations (Cheng et al., 2017) and the newly updated annotations for WCS417, respectively. All differential expression analyses were performed with DESeq2 v1.38.3 (Love et al., 2014).

Functional annotations were retrieved with PANTHER v16.0 (Mi et al., 2021) for *Arabidopsis* genes and with eggNOG-mapper v2.1.12 (Cantalapiedra et al., 2021) with the eggNOG v5.0 database (Huerta-Cepas et al., 2019) for WCS417. Pathway enrichment analyses were performed with *clusterProfiler* package v4.6.2 in R v4.2.1 (R Core Team, 2022).

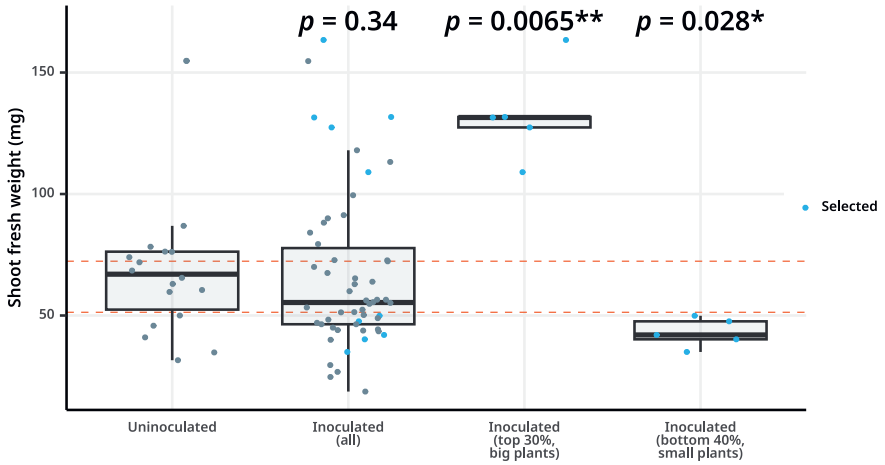
Data availability

All metagenomic and metatranscriptomic reads are accessible under the BioProject PRJNA1108915. Scripts used to process, analyse data and generate figures are deposited in a GitHub repository (https://github.com/jjsanchezgil/Metatranscriptomes_WCS417).

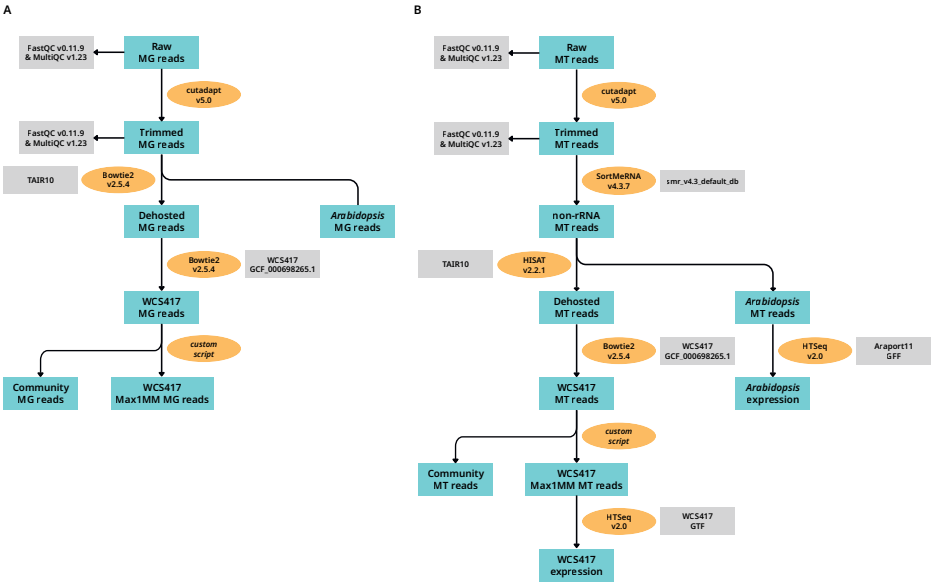
Acknowledgements

The authors want to thank Bas E. Dutilh, Corné M.J. Pieterse and members of the Plant-Microbe Interactions lab for helpful discussions.

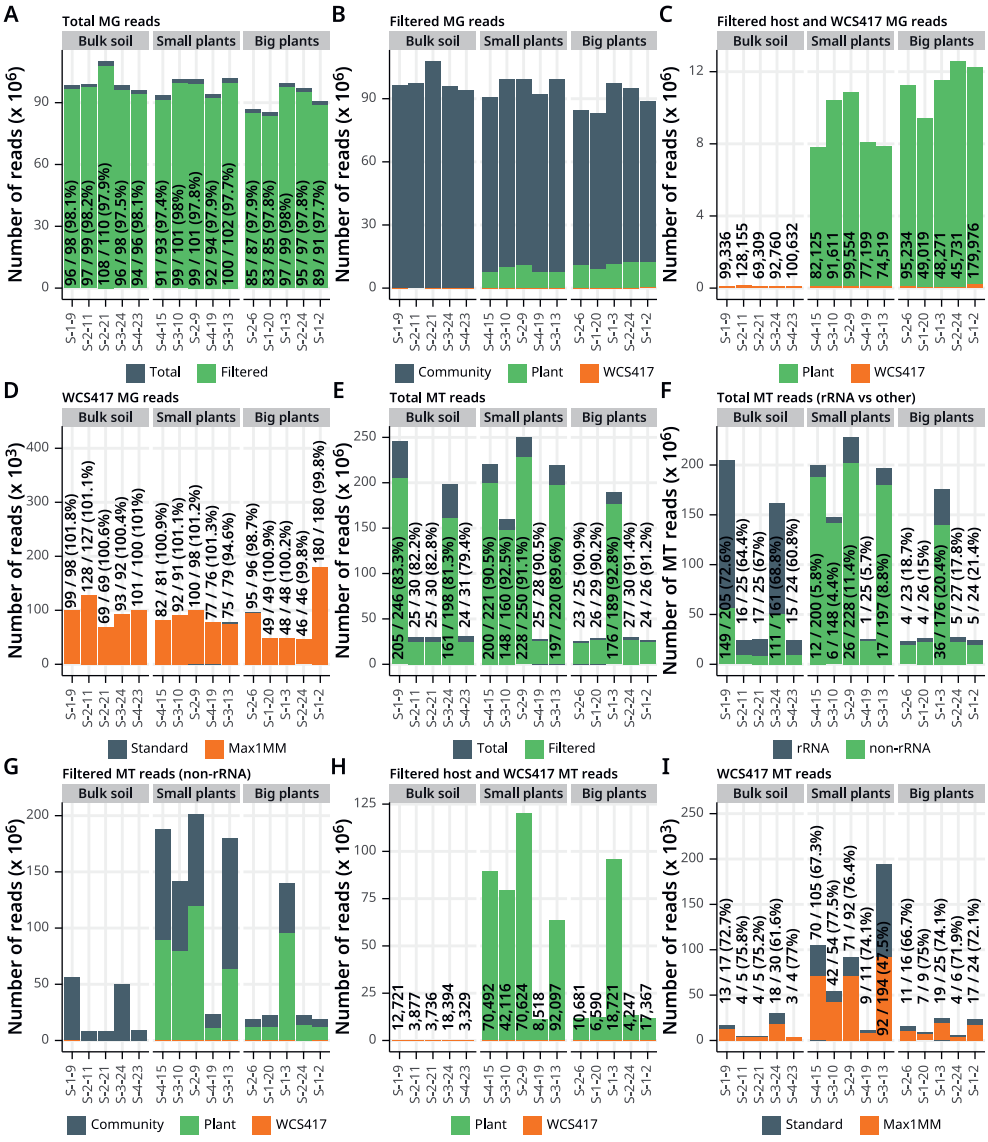
Supplementary Figures



Supplementary Figure 1. Selection of plants in this study. Boxplots showing the distribution of shoot fresh weight for plants inoculated with WCS417 or uninoculated. Dashed lines show the 40th and 70th percentile in the data, which were used as cutoffs to select plants to study differential activity in big versus small plants. Significance above boxplots show Wilcoxon tests comparing with the uninoculated group of plants.

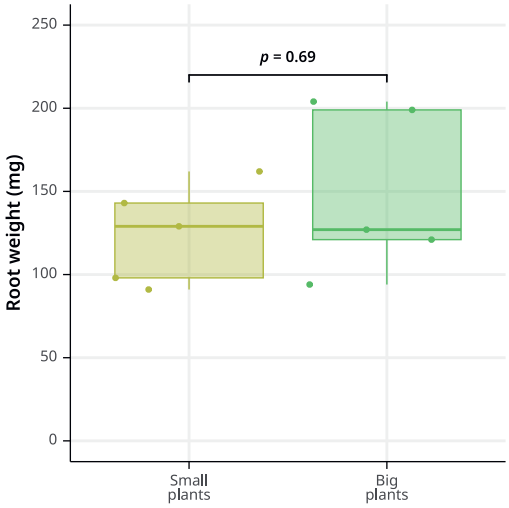


Supplementary Figure 2. Workflow for the analysis of metagenomes and metatranscriptomes. Pipeline for the processing of metagenomic (A) and metatranscriptomic (B) sequencing reads.

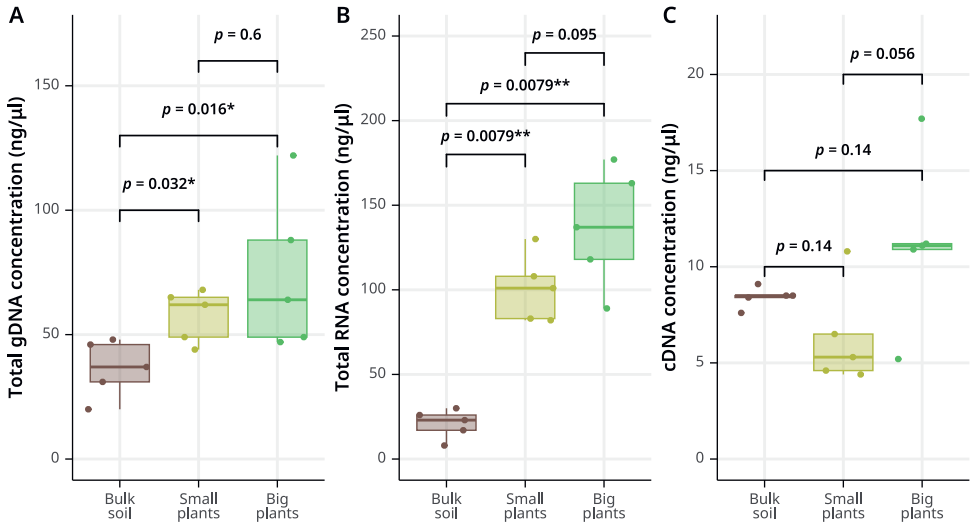


Supplementary Figure 3. General depiction of read fates. (A) Barplot of total and filtered MG (metagenomic) reads. Above each bar, the plot shows the rounded number of filtered reads, total reads and the percent of filtered reads. (B) Stacked barplot depicting the total number of filtered MG reads for *Arabidopsis*, WCS417 and community reads, counted as all reads that did not map to host or WCS417. (C) Same as in B, but only showing *Arabidopsis* and WCS417 reads. Numbers indicate the total number of WCS417 reads found. (D) Total number of WCS417 MG reads as output by Bowtie2 (standard) and after filtering for maximum one mismatch per read pair (Max1MM). The plot also shows the number of Max1MM reads, standard, and the percent of Max-1MM reads. Values above 100% indicate that there were more Max1MM reads

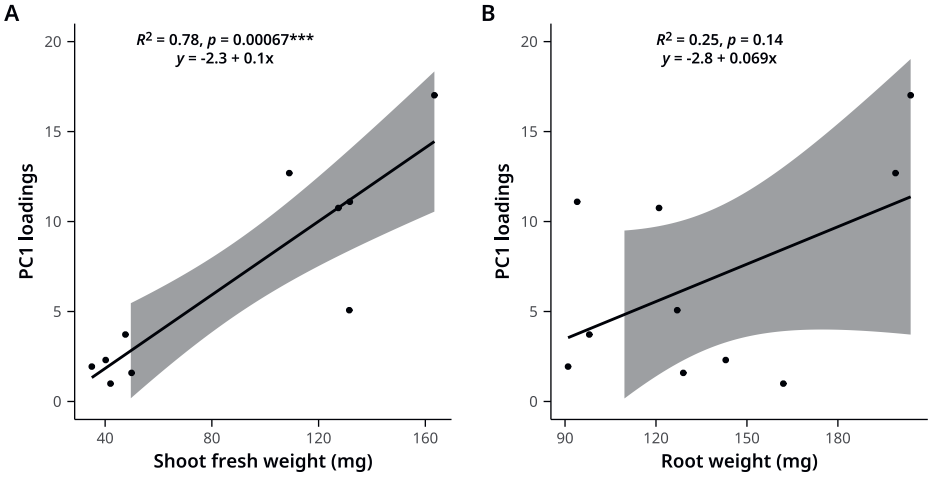
than output by Bowtie2 using the standard parameters. This is because those reads were discarded by Bowtie2 due to mapping multiple times in the WCS417 genome even though their alignment was perfect or near perfect. Note the change of scale from millions to thousands. (E) Same as in A, but for MT (metatranscriptomic) reads. (F) Number of MT reads classified as rRNA and non-rRNA by SortMeRNA (Kopylova et al., 2012). Values above each bar indicate the number of rRNA reads, number of total filtered reads and the percent of rRNA reads in the sample. (G) Same as in B, but for non-rRNA MT reads. (H) Same as in C, but for non-rRNA MT reads. (I) Same as in D, but for non-rRNA MT reads.



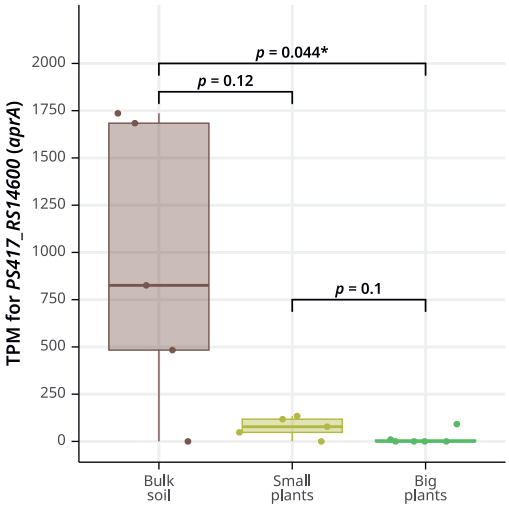
Supplementary Figure 4. Big and small plants have comparable root biomass. Boxplots showing the root biomass for big and small plants. No significant difference is observed between the two groups.



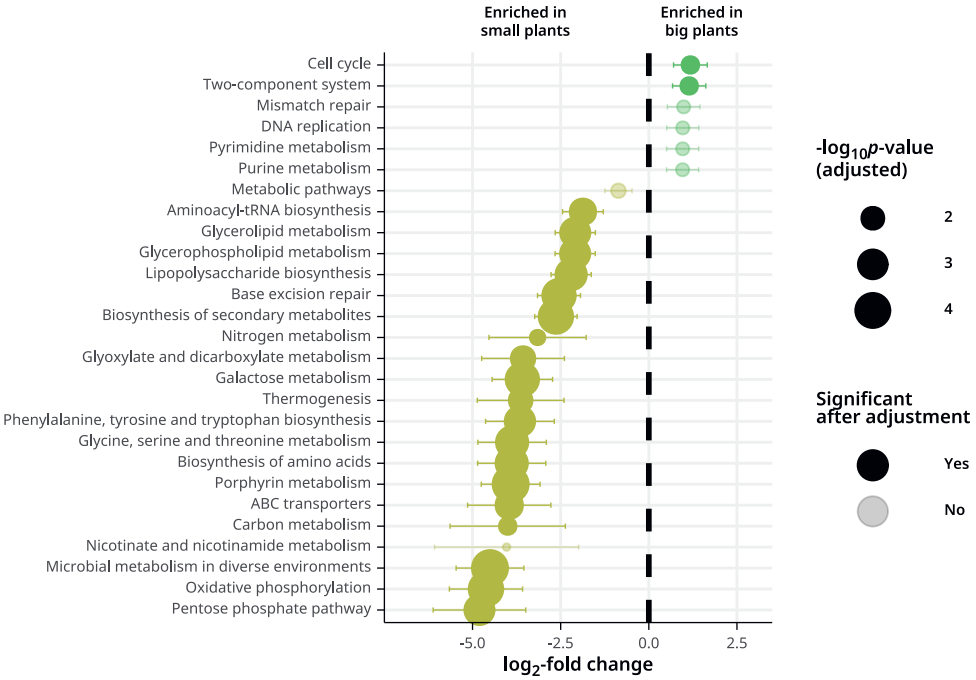
Supplementary Figure 5. Boxplots of nucleic acid extractions and cDNA preparations. (A) Total genomic DNA. (B) Total raw RNA. (C) Total processed cDNA.



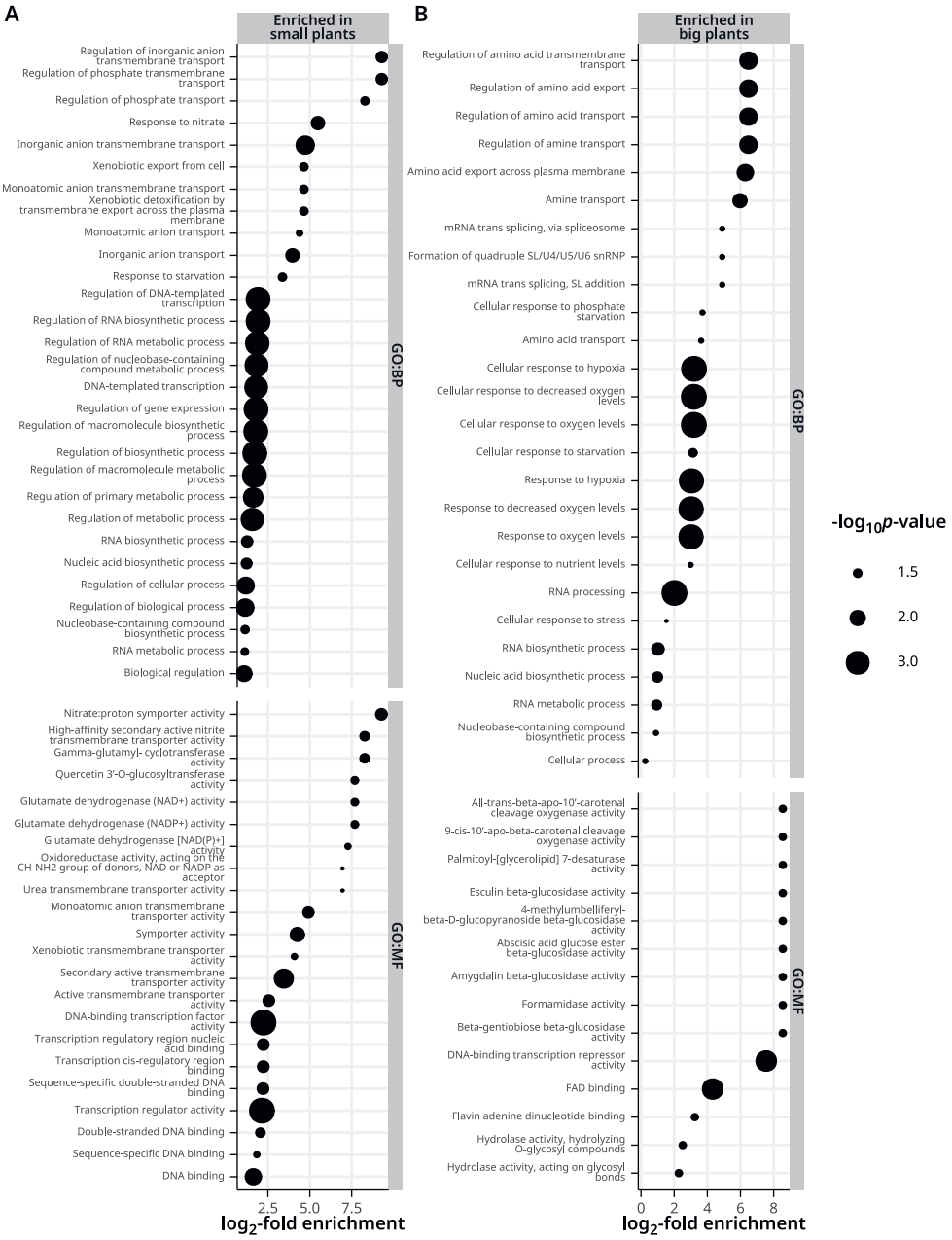
Supplementary Figure 6. Principal component correlations with shoot and root weight. Loadings of the first principal component of the PCA in Figure 3B as a function of shoot fresh weight (A) and root weight (B).



Supplementary Figure 7. Expression of the metalloprotease gene *aprA* is highest in soil and least in the rhizosphere of big plants. Boxplot showing the TPM values for the *aprA* gene in the different experimental groups. The difference in expression levels in small versus big plants is statistically similar. However, note that in comparison with bulk soil, expression in big plants is significantly lower and close to zero.



Supplementary Figure 8. Differentially enriched pathways in WCS417 in the rhizosphere of big versus small plants. Solid colours denote differentially enriched pathways with an adjusted *p*-value < 0.05 (Benjamin-Hochberg correction), while significant pathways that did not pass the threshold after adjustment are shown in transparency. Error bars indicate the standard error for the fold-changes.



Supplementary Figure 9. Differentially enriched plant pathways in the roots of big versus small *Arabidopsis* plants. (A) Differentially enriched pathways in small plants. (B) Differentially enriched pathways in big plants. All *p*-values are FDR-corrected.

Supplementary Tables

Supplementary Table 1. Samples included in this study.

Sample	Compartment	Group	Shoot weight (g)	Shoot weight FC	Sample weight (g)
S-1-2	Rhizosphere	Big	0,1634	2,438806	0,204
S-2-24	Rhizosphere	Big	0,1317	1,965672	0,094
S-1-3	Rhizosphere	Big	0,1315	1,962687	0,127
S-1-20	Rhizosphere	Big	0,1274	1,901493	0,121
S-2-6	Rhizosphere	Big	0,109	1,626866	0,199
S-3-13	Rhizosphere	Small	0,0499	0,744776	0,129
S-4-19	Rhizosphere	Small	0,0476	0,710448	0,098
S-2-9	Rhizosphere	Small	0,042	0,626866	0,162
S-3-10	Rhizosphere	Small	0,0402	0,6	0,143
S-4-15	Rhizosphere	Small	0,035	0,522388	0,091
S-1-9	Bulk soil	Soil	-	-	0,229
S-2-11	Bulk soil	Soil	-	-	0,223
S-2-21	Bulk soil	Soil	-	-	0,249
S-3-24	Bulk soil	Soil	-	-	0,216
S-4-23	Bulk soil	Soil	-	-	0,248

Supplementary Table 2. Summary of metagenomic and metatranscriptomic reads.

Sample	Group	MG Total	MG Filtered	MG Plant	MG WCS417	MG Commu-nity	MT Total	MT Filtered	MT No-rRNA	MT Plant	MT WCS417	MT Commu-nity
S-1-20	Big	85160921	83312454	9375461	49019	73887974	29219530	26363343	22412555	12316531	6590	10089434
S-2-6	Big	86601823	84789018	11144479	95234	73549305	25344286	23032417	18736142	11741824	10681	6983637
S-1-2	Big	90719630	88628170	12045620	179976	76402574	26383312	24071042	18928191	11983617	17367	6927207
S-2-24	Big	96951925	94864293	12530607	45731	82287955	29502112	26968972	22180528	13698251	4247	8478030
S-1-3	Big	99310261	97317719	11460034	48271	85809414	189438093	175822639	139952433	95931310	18721	44002402
S-2-21	Soil	110088401	107781618	9500	69309	107702809	29868810	24729176	8151738	1262	3736	8146740
S-4-23	Soil	95959308	94105946	7283	100632	93998031	30839642	24486399	9594249	378	3329	9590542
S-3-24	Soil	98177076	95770183	6877	92760	95670546	198165654	161103002	50305609	5231	18394	50281984
S-1-9	Soil	98196771	96365965	14493	99336	96252136	245629964	204676762	55985024	30171	12721	55942132
S-2-11	Soil	98975339	97238796	7563	128155	97103078	29981669	24650874	8785413	1041	3877	8780495
S-2-9	Small	101255468	99072485	10771372	99554	88201559	249885789	227638972	201784332	119878297	70624	81835411
S-3-10	Small	101320720	99287369	10331362	91611	88864396	159848902	147882345	141400475	79277616	42116	62080743
S-3-13	Small	101836119	99513292	7814296	74519	91624477	219754849	196979935	179675523	63739713	92097	115843713
S-4-15	Small	93385629	90939374	7730728	82125	83126521	220732111	199734531	188228740	89654195	70492	98504053
S-4-19	Small	94262825	92254450	8014168	77199	84163083	27649817	25023776	23601976	11394075	8518	12199383

Supplementary Table 3. PCA loadings and distances to centroid.

Sample	Group	Compart-ment	PC1	PC2	PC1 centre (comp)	PC2 centre (comp)	PC1 centre (group)	PC2 centre (group)	Dis-tance to centroid (group)	Shoot weight (g)	Sample weight (g)
S-1-2	Big	Rhizosphere	-17,02	-0,05	-6,72	0,22	-11,3	-1,05	30,46	0,1634	0,204
S-2-24	Big	Rhizosphere	-11,10	-5,14	-6,72	0,22	-11,33	-1,05	24,98	0,1317	0,094
S-1-3	Big	Rhizosphere	-5,07	-1,61	-6,72	0,22	-11,33	-1,05	18,55	0,1315	0,127
S-1-20	Big	Rhizosphere	-10,75	0,75	-6,72	0,22	-11,33	-1,05	24,22	0,1274	0,121
S-2-6	Big	Rhizosphere	-12,69	0,81	-6,72	0,22	-11,33	-1,05	26,16	0,109	0,199
S-4-15	Small	Rhizosphere	-1,94	1,81	-6,72	0,22	-2,11	1,49	15,54	0,035	0,091
S-4-19	Small	Rhizosphere	-3,72	2,23	-6,72	0,22	-2,11	1,49	17,36	0,0476	0,098
S-3-13	Small	Rhizosphere	-1,59	0,82	-6,72	0,22	-2,11	1,49	15,08	0,0499	0,129
S-2-9	Small	Rhizosphere	-0,99	0,98	-6,72	0,22	-2,11	1,49	14,50	0,042	0,162
S-3-10	Small	Rhizosphere	-2,30	1,60	-6,72	0,22	-2,11	1,49	15,87	0,0402	0,143
S-1-9	Soil	Soil	13,93	-2,19	13,44	-0,44	13,44	-0,44	1,82	-	0,229
S-2-11	Soil	Soil	13,65	0,10	13,44	-0,44	13,44	-0,44	0,58	-	0,223
S-2-21	Soil	Soil	13,09	1,57	13,44	-0,44	13,44	-0,44	2,04	-	0,249
S-3-24	Soil	Soil	13,62	0,003	13,44	-0,44	13,44	-0,44	0,48	-	0,216
S-4-23	Soil	Soil	12,89	-1,66	13,44	-0,44	13,44	-0,44	1,34	-	0,248

Supplementary Table 4. DEGs of WCS417 in bulk soil.

Gene	Name	Base mean	log_2FC	SE(FC)	stat	p-value	BH adj. p-value	COG
PS417_ RS21585	*rseP/mucP*	41,142	-4,472	0,548	-8,163	3,28E-16	4,62E-13	M
PS417_ RS19590	IV02_14570	16,095	-2,911	0,453	-6,432	1,26E-10	1,37E-08	S
PS417_ RS09760	-	19,767	-2,872	0,532	-5,394	6,88E-08	2,11E-06	S
PS417_ RS12775	-	37,268	-1,966	0,369	-5,332	9,70E-08	2,80E-06	S
PS417_ RS08020	-	123,838	-1,792	0,374	-4,789	1,68E-06	2,63E-05	S
PS417_ RS15715	*nfuA*	13,909	-2,452	0,524	-4,679	2,88E-06	3,84E-05	C
PS417_ BK30360	-	10,259	-3,014	0,791	-3,811	1,39E-04	9,89E-04	-
PS417_ RS22705	-	20,872	-1,897	0,503	-3,769	1,64E-04	1,12E-03	J
PS417_ RS06330	MA20_44945	5,796	-4,529	1,250	-3,622	2,92E-04	1,77E-03	K
PS417_ RS08390	-	17,677	-1,724	0,494	-3,490	4,82E-04	2,71E-03	S
PS417_ RS29405	-	10,267	-2,014	0,589	-3,422	6,22E-04	3,35E-03	S
PS417_ RS20070	-	14,307	-1,822	0,559	-3,262	1,11E-03	5,19E-03	S
PS417_ RS03635	*yraR*	5,045	-3,547	1,107	-3,204	1,36E-03	6,14E-03	GM
PS417_ RS11780	*yohC*	5,893	-2,347	0,801	-2,928	3,41E-03	1,28E-02	S
PS417_ RS12055	-	13,773	-1,834	0,639	-2,871	4,10E-03	1,46E-02	NU
PS417_ RS14600	*aprA*	4,667	-3,015	1,123	-2,683	7,29E-03	2,30E-02	Q
PS417_ RS02475	-	8,825	-2,915	1,093	-2,668	7,63E-03	2,38E-02	T
PS417_ RS07610	*vapC*	1,981	-3,731	1,400	-2,666	7,68E-03	2,39E-02	D
PS417_ RS27110	*caiA/gcdH*	3,794	-2,338	0,961	-2,432	1,50E-02	3,87E-02	I

Supplementary Table 5. DEGs of WCS417 in the rhizosphere of big versus small plants. Genes with adjusted *p*-value < 0.05 are shown in bold.

Group	Gene	Name	Base mean	log_2FC	SE(FC)	stat	p-value	BH adj. p-value	COG
Small	**PS417_ RS03765**	**-**	**96,091**	**-1,309**	**0,235**	**-5,565**	**2,62E-08**	**6,18E-06**	**S**
Small	**PS417_ RS14655**	***pgm***	**24,128**	**-2,251**	**0,401**	**-5,618**	**1,93E-08**	**6,18E-06**	**G**
Small	**PS417_ RS03770**	**-**	**79,632**	**-1,311**	**0,246**	**-5,329**	**9,90E-08**	**1,56E-05**	**S**
Small	**PS417_ RS04145**	***gdhA***	**40,838**	**-1,739**	**0,372**	**-4,670**	**3,01E-06**	**3,55E-04**	**J**
Small	**PS417_ RS12055**	**-**	**10,556**	**-2,953**	**0,703**	**-4,200**	**2,67E-05**	**2,10E-03**	**NU**
Small	**PS417_ RS27240**	***algB***	**7,5879**	**-3,919**	**0,966**	**-4,056**	**4,99E-05**	**3,37E-03**	**T**
Small	**PS417_ RS29205**	**-**	**35,611**	**-1,385**	**0,377**	**-3,669**	**2,44E-04**	**1,28E-02**	**S**
Small	**PS417_ RS04180**	**-**	**8,0015**	**-3,061**	**0,848**	**-3,611**	**3,05E-04**	**1,31E-02**	**S**
Small	**PS417_ RS14900**	***nemR***	**12,392**	**-1,952**	**0,545**	**-3,580**	**3,43E-04**	**1,35E-02**	**K**
Small	**PS417_ RS29675**	**-**	**61,648**	**-1,123**	**0,320**	**-3,507**	**4,53E-04**	**1,65E-02**	**S**
Small	**PS417_ RS15770**	**-**	**31,523**	**-1,146**	**0,332**	**-3,446**	**5,69E-04**	**1,92E-02**	**S**
Small	**PS417_ RS20730**	***mocA/ rseA***	**111,65**	**-0,988**	**0,289**	**-3,416**	**6,36E-04**	**2,00E-02**	**T**
Small	**PS417_ RS12775**	**-**	**25,634**	**-1,490**	**0,441**	**-3,378**	**7,29E-04**	**2,02E-02**	**S**
Small	**PS417_ RS29235**	**-**	**7,5315**	**-3,219**	**0,948**	**-3,394**	**6,89E-04**	**2,02E-02**	**S**
Small	**PS417_ RS06765**	**-**	**13,551**	**-1,932**	**0,590**	**-3,272**	**1,07E-03**	**2,19E-02**	**S**
Small	**PS417_ RS17580**	**-**	**78,048**	**-1,105**	**0,335**	**-3,296**	**9,81E-04**	**2,19E-02**	**S**
Small	**PS417_ RS15740**	***cysI***	**11,663**	**-2,150**	**0,665**	**-3,231**	**1,23E-03**	**2,35E-02**	**S**
Small	**PS417_ RS08020**	**-**	**88,944**	**-1,052**	**0,335**	**-3,136**	**1,71E-03**	**3,11E-02**	**S**
Small	**PS417_ RS18450**	**-**	**6,0282**	**-2,792**	**0,900**	**-3,102**	**1,92E-03**	**3,24E-02**	**S**
Small	**PS417_ RS28550**	**-**	**10,535**	**-2,044**	**0,675**	**-3,029**	**2,46E-03**	**4,00E-02**	**S**
Small	**PS417_ RS04945**	***lrgA***	**8,5157**	**-2,311**	**0,773**	**-2,991**	**2,78E-03**	**4,38E-02**	**M**

Small	**PS417_ RS03585**	***dedA***	**12,989**	**-1,843**	**0,622**	**-2,962**	**3,06E-03**	**4,66E-02**	**I**
Small	PS417_ RS19575	-	7,646	-2,128	0,727	-2,926	3,44E-03	5,07E-02	L
Small	PS417_ RS00195	-	6,2883	-2,667	0,923	-2,889	3,86E-03	5,53E-02	S
Small	PS417_ RS06300	-	12,116	-1,910	0,664	-2,874	4,05E-03	5,62E-02	S
Small	PS417_ RS19395	-	6,3525	-2,823	0,990	-2,853	4,34E-03	5,85E-02	S
Small	PS417_ RS08075	-	11,863	-1,553	0,557	-2,787	5,32E-03	6,44E-02	S
Small	PS417_ RS11535	-	10,886	-1,589	0,572	-2,779	5,46E-03	6,44E-02	S
Small	PS417_ RS12515	-	10,678	-1,947	0,696	-2,797	5,16E-03	6,44E-02	S
Small	PS417_ RS14670	-	30,143	-1,054	0,378	-2,788	5,30E-03	6,44E-02	C
Small	PS417_ RS13860	-	18,512	-1,346	0,488	-2,758	5,81E-03	6,69E-02	S
Small	PS417_ RS22705	-	14,048	-1,341	0,492	-2,727	6,40E-03	7,19E-02	J
Small	PS417_ RS27160	-	6,3269	-2,149	0,794	-2,707	6,79E-03	7,46E-02	S
Small	PS417_ RS08390	-	14,029	-1,516	0,574	-2,642	8,25E-03	8,45E-02	S
Small	PS417_ RS18380	*vanB*	7,3455	-1,972	0,748	-2,637	8,37E-03	8,45E-02	S
Small	PS417_ RS20735	*algU*	88,229	-0,683	0,260	-2,628	8,59E-03	8,45E-02	K
Small	PS417_ RS25255	*hflX*	26,447	-1,081	0,411	-2,630	8,54E-03	8,45E-02	S
Small	PS417_ RS21925	*tetR*	9,6033	-1,873	0,728	-2,574	1,00E-02	9,68E-02	K
Small	PS417_ RS04345	*dnaK*	16,680	-1,259	0,493	-2,554	1,06E-02	9,73E-02	O
Small	PS417_ RS13630	*glgA*	6,6484	-2,102	0,824	-2,552	1,07E-02	9,73E-02	G
Small	PS417_ RS06085	*spy*	10,451	-1,611	0,639	-2,523	1,17E-02	1,02E-01	NPTU
Small	PS417_ RS06755	*amrZ*	17,266	-1,197	0,478	-2,505	1,23E-02	1,02E-01	S
Small	PS417_ RS22725	-	8,8605	-1,686	0,672	-2,508	1,21E-02	1,02E-01	S
Small	PS417_ RS27405	*cynT*	12,021	-1,334	0,531	-2,513	1,20E-02	1,02E-01	P

Small	PS417_ RS14280	*arnC*	29,598	-0,952	0,392	-2,432	1,50E-02	1,16E-01	M
Small	PS417_ RS29200	-	19,578	-1,854	0,768	-2,413	1,58E-02	1,18E-01	S
Small	PS417_ RS04155	-	29,748	-1,634	0,690	-2,368	1,79E-02	1,28E-01	S
Small	PS417_ RS10460	*ais*	43,834	-0,693	0,293	-2,366	1,80E-02	1,28E-01	G
Small	PS417_ RS01005	*algR*	11,775	-1,525	0,647	-2,357	1,84E-02	1,28E-01	KT
Small	PS417_ RS08970	-	5,9967	-2,040	0,871	-2,343	1,92E-02	1,29E-01	S
Small	PS417_ RS20725	*rseB*	58,996	-0,798	0,343	-2,328	1,99E-02	1,31E-01	T
Small	PS417_ RS04245	-	7,7294	-1,576	0,685	-2,302	2,13E-02	1,35E-01	L
Small	PS417_ RS06010	*cspC*	12,529	-1,197	0,519	-2,304	2,12E-02	1,35E-01	K
Small	PS417_ RS07155	*corA*	7,8861	-1,631	0,709	-2,301	2,14E-02	1,35E-01	P
Small	PS417_ RS25815	*hslV*	7,3119	-1,624	0,715	-2,271	2,32E-02	1,44E-01	O
Small	PS417_ RS02225	*glxA*	6,8703	-1,967	0,875	-2,249	2,45E-02	1,50E-01	K
Small	PS417_ RS05620	-	74,041	-0,622	0,280	-2,216	2,67E-02	1,59E-01	Q
Small	PS417_ RS05190	*iscR*	16,063	-0,988	0,461	-2,146	3,18E-02	1,75E-01	K
Small	PS417_ RS05920	*aspS*	8,8693	-1,488	0,696	-2,139	3,25E-02	1,75E-01	J
Small	PS417_ RS08250	*nlpC*	6,7104	-1,939	0,907	-2,137	3,26E-02	1,75E-01	M
Small	PS417_ RS08420	*phhA*	19,556	-0,987	0,457	-2,159	3,09E-02	1,75E-01	E
Small	PS417_ RS29405	-	6,6836	-1,680	0,802	-2,096	3,61E-02	1,85E-01	S
Small	PS417_ RS26030	*blc*	13,141	-1,177	0,567	-2,077	3,78E-02	1,92E-01	M
Small	PS417_ RS11690	*clpA*	24,732	-0,887	0,430	-2,063	3,91E-02	1,96E-01	O
Small	PS417_ RS15245	*glgB*	6,8616	-1,396	0,680	-2,053	4,01E-02	1,99E-01	G
Small	PS417_ BK30340	-	136,98	-0,889	0,440	-2,020	4,34E-02	2,08E-01	-
Small	PS417_ RS01300	-	23,445	-0,771	0,384	-2,007	4,48E-02	2,08E-01	S

Small	PS417_ RS04340	*grpE*	14,028	-1,052	0,523	-2,012	4,42E-02	2,08E-01	O
Small	PS417_ RS25190	*rlmB*	16,056	-1,048	0,522	-2,009	4,45E-02	2,08E-01	J
Small	PS417_ RS03610	*aldB*	11,461	-1,075	0,542	-1,986	4,71E-02	2,14E-01	C
Small	PS417_ RS22195	*grxD*	6,3092	-1,632	0,822	-1,985	4,72E-02	2,14E-01	C
Small	PS417_ RS00880	-	2,7213	-3,451	1,595	-2,164	3,05E-02	-	T
Small	PS417_ RS01440	-	2,8869	-2,915	1,242	-2,348	1,89E-02	-	O
Small	PS417_ RS02115	*fba*	4,5473	-3,681	1,170	-3,145	1,66E-03	-	G
Small	PS417_ RS03185	-	2,897	-3,368	1,468	-2,294	2,18E-02	-	L
Small	PS417_ RS03890	-	4,382	-2,564	1,045	-2,454	1,41E-02	-	S
Small	PS417_ RS04250	-	3,019	-2,582	1,295	-1,994	4,62E-02	-	S
Small	PS417_ RS05175	*suhB*	4,552	-4,197	1,240	-3,384	7,14E-04	-	G
Small	PS417_ RS05275	*ybdL*	2,384	-3,289	1,502	-2,190	2,85E-02	-	S
Small	PS417_ RS05600	*groS*	5,612	-1,744	0,796	-2,190	2,85E-02	-	O
Small	PS417_ RS06800	*cheV*	3,489	-3,127	1,399	-2,236	2,54E-02	-	T
Small	PS417_ RS07085	-	2,901	-2,847	1,395	-2,040	4,13E-02	-	M
Small	PS417_ RS07475	*xdhB*	3,378	-3,733	1,428	-2,615	8,93E-03	-	F
Small	PS417_ RS07495	*zipA*	5,245	-1,951	0,971	-2,008	4,46E-02	-	D
Small	PS417_ RS08700	*chev-2*	2,523	-3,310	1,451	-2,281	2,25E-02	-	T
Small	PS417_ RS09485	-	1,796	-3,506	1,768	-1,984	4,73E-02	-	S
Small	PS417_ RS11795	*wrbA*	3,553	-2,396	1,184	-2,024	4,29E-02	-	S
Small	PS417_ RS12235	*yhhW_1*	2,356	-3,290	1,527	-2,155	3,12E-02	-	S
Small	PS417_ RS12945	*alkB1*	3,903	-3,977	1,244	-3,197	1,39E-03	-	I
Small	PS417_ RS13405	*nirD*	2,140	-3,037	1,519	-1,999	4,56E-02	-	P

Small	PS417_ RS15535	-	5,721	-2,983	0,994	-3,002	2,69E-03	-	S
Small	PS417_ RS16265	*acoR*	4,444	-3,099	1,206	-2,568	1,02E-02	-	KQ
Small	PS417_ RS16385	*pcaB*	4,319	-4,130	1,210	-3,412	6,44E-04	-	F
Small	PS417_ RS17280	*uvrY*	5,654	-2,808	0,978	-2,871	4,09E-03	-	K
Small	PS417_ RS17340	-	2,780	-3,450	1,313	-2,627	8,62E-03	-	T
Small	PS417_ RS18830	-	4,017	-2,236	1,074	-2,082	3,73E-02	-	S
Small	PS417_ RS19475	*aroG1*	5,409	-2,000	0,855	-2,339	1,93E-02	-	E
Small	PS417_ RS19490	*ycfS*	2,057	-3,722	1,798	-2,070	3,85E-02	-	M
Small	PS417_ RS20170	*parA*	5,203	-2,125	1,055	-2,014	4,40E-02	-	S
Small	PS417_ RS21315	-	2,307	-3,250	1,540	-2,110	3,48E-02	-	G
Small	PS417_ RS21775	-	4,883	-3,245	1,078	-3,010	2,62E-03	-	S
Small	PS417_ RS22425	*padR*	2,164	-3,787	1,529	-2,476	1,33E-02	-	P
Small	PS417_ RS23090	-	2,300	-3,190	1,527	-2,090	3,67E-02	-	V
Small	PS417_ RS23955	*pgeF*	4,722	-2,795	1,014	-2,755	5,86E-03	-	J
Small	PS417_ RS24335	*avrE*	2,482	-3,306	1,398	-2,365	1,80E-02	-	S
Small	PS417_ RS24485	*pqiA2*	2,102	-3,650	1,545	-2,362	1,82E-02	-	S
Small	PS417_ RS25015	-	3,293	-2,538	1,230	-2,063	3,91E-02	-	P
Small	PS417_ RS25880	*hisI*	2,811	-3,017	1,304	-2,314	2,07E-02	-	E
Small	PS417_ RS27465	-	2,354	-3,861	1,631	-2,367	1,79E-02	-	S
Big	**PS417_ RS02930**	***rplK***	**33,841**	**1,456**	**0,327**	**4,447**	**8,70E-06**	**8,21E-04**	**J**
Big	**PS417_ RS24800**	***fis/dusB***	**9,3063**	**2,261**	**0,572**	**3,951**	**7,77E-05**	**4,59E-03**	**K**
Big	**PS417_ RS25170**	***rplI***	**17,437**	**1,670**	**0,462**	**3,615**	**3,00E-04**	**1,31E-02**	**J**
Big	**PS417_ RS02940**	***rplJ***	**16,600**	**1,447**	**0,437**	**3,308**	**9,40E-04**	**2,19E-02**	**J**

Big	**PS417_ RS02990**	***rplD***	**37,386**	**1,003**	**0,305**	**3,291**	**1,00E-03**	**2,19E-02**	**J**
Big	**PS417_ RS03085**	***secY***	**32,400**	**1,111**	**0,338**	**3,283**	**1,03E-03**	**2,19E-02**	**U**
Big	**PS417_ RS19020**	***sdhB***	**11,563**	**1,631**	**0,492**	**3,318**	**9,05E-04**	**2,19E-02**	**C**
Big	**PS417_ RS23635**	***rpsI***	**46,430**	**1,093**	**0,339**	**3,228**	**1,24E-03**	**2,35E-02**	**J**
Big	**PS417_ RS02970**	***fusA***	**37,383**	**0,901**	**0,290**	**3,108**	**1,88E-03**	**3,24E-02**	**J**
Big	PS417_ RS02960	*rpsL*	10,612	1,889	0,669	2,823	4,75E-03	6,23E-02	-
Big	PS417_ RS00080	*atpH*	13,636	1,425	0,540	2,640	8,30E-03	8,45E-02	C
Big	PS417_ RS25795	*rpmE*	37,144	0,831	0,324	2,561	1,04E-02	9,73E-02	J
Big	PS417_ RS03105	*rpoA*	36,566	0,773	0,308	2,506	1,22E-02	1,02E-01	K
Big	PS417_ RS06910	*yceD*	59,444	0,609	0,245	2,487	1,29E-02	1,04E-01	S
Big	PS417_ RS19015	*sucA*	22,735	0,930	0,374	2,483	1,30E-02	1,04E-01	C
Big	PS417_ RS06930	*fabG*	6,8717	2,136	0,875	2,443	1,46E-02	1,15E-01	IQ
Big	PS417_ RS23140	*murC*	9,3499	1,402	0,580	2,416	1,57E-02	1,18E-01	M
Big	PS417_ RS05485	*rplS*	7,4246	1,685	0,707	2,384	1,71E-02	1,26E-01	J
Big	PS417_ RS23430	*lptB*	7,1335	1,967	0,832	2,363	1,81E-02	1,28E-01	S
Big	PS417_ RS26070	*thiI*	13,689	1,081	0,462	2,342	1,92E-02	1,29E-01	H
Big	PS417_ RS00095	*atpD*	9,411	1,383	0,594	2,327	2,00E-02	1,31E-01	C
Big	PS417_ RS19025	*sdhA*	12,843	1,033	0,463	2,232	2,56E-02	1,55E-01	C
Big	PS417_ RS03050	*rpsN*	18,973	0,825	0,378	2,185	2,89E-02	1,70E-01	J
Big	PS417_ RS00785	*rpmB*	7,5972	1,459	0,677	2,154	3,12E-02	1,75E-01	J
Big	PS417_ RS02945	*rplL*	16,385	1,014	0,467	2,170	3,00E-02	1,75E-01	J
Big	PS417_ RS05015	*cyoB*	12,323	1,037	0,484	2,143	3,21E-02	1,75E-01	C
Big	PS417_ RS19485	*oprI*	49,036	0,740	0,343	2,158	3,10E-02	1,75E-01	S

Big	PS417_ RS05480	*trmD*	31,926	0,662	0,311	2,130	3,31E-01	1,76E-01	J
Big	PS417_ RS06825	*flgE*	8,3487	1,382	0,653	2,117	3,43E-02	1,80E-01	N
Big	PS417_ RS03000	*rplB*	27,974	0,764	0,362	2,109	3,50E-02	1,81E-01	J
Big	PS417_ RS00085	*atpA*	17,722	0,818	0,400	2,044	4,10E-02	2,01E-01	C
Big	PS417_ RS07370	*mscS*	9,6305	1,108	0,553	2,006	4,49E-02	2,08E-01	M
Big	PS417_ RS13080	*gdhB*	13,632	1,181	0,583	2,024	4,29E-02	2,08E-01	E
Big	PS417_ RS02985	*rplC*	25,388	0,693	0,352	1,968	4,91E-02	2,19E-01	J
Big	PS417_ RS03055	*rpsH*	29,734	0,757	0,385	1,968	4,91E-02	2,19E-01	J
Big	PS417_ RS00090	*atpG*	5,8675	2,641	0,801	3,295	9,84E-04	-	C
Big	PS417_ RS00870	*xpt*	1,496	4,926	2,274	2,166	3,03E-02	-	F
Big	PS417_ RS01855	*gshB*	1,535	4,930	2,464	2,001	4,54E-02	-	H
Big	PS417_ RS01880	*chpC*	2,142	3,945	1,568	2,516	1,19E-02	-	NT
Big	PS417_ RS01995	*yceJ*	3,459	2,688	1,196	2,247	2,46E-02	-	C
Big	PS417_ RS02760	*potA*	1,583	5,169	2,026	2,551	1,07E-02	-	P
Big	PS417_ RS05030	*cyoE*	3,157	2,746	1,138	2,414	1,58E-02	-	O
Big	PS417_ RS06190	*icmP*	1,446	4,446	1,875	2,371	1,78E-02	-	P
Big	PS417_ RS06820	*flgD*	5,868	1,380	0,698	1,976	4,81E-02	-	N
Big	PS417_ RS06900	*sppA*	1,099	4,662	2,286	2,040	4,14E-02	-	OU
Big	PS417_ RS08260	*cobO*	1,197	4,589	2,266	2,026	4,28E-02	-	H
Big	PS417_ RS08500	*fleS*	3,752	2,643	0,951	2,778	5,47E-03	-	T
Big	PS417_ RS08505	*fleR*	3,232	2,746	1,210	2,270	2,32E-02	-	T
Big	PS417_ RS08825	-	2,234	5,487	1,684	3,259	1,12E-03	-	M
Big	PS417_ RS08830	*apeB*	1,137	4,710	2,246	2,097	3,60E-02	-	E
Big	PS417_ RS09860	*purR*	1,363	4,914	2,130	2,306	2,11E-02	-	K

Big	PS417_ RS09865	*rbsK*	1,840	3,940	1,603	2,458	1,40E-02	-	G
Big	PS417_ RS11315	*glnS*	2,927	2,681	1,277	2,100	3,57E-02	-	J
Big	PS417_ RS11365	*phaD2*	1,400	4,913	2,310	2,127	3,34E-02	-	CP
Big	PS417_ RS12785	-	3,536	3,123	1,279	2,442	1,46E-02	-	S
Big	PS417_ RS13155	*sdaC*	2,171	3,188	1,507	2,116	3,43E-02	-	E
Big	PS417_ RS17025	*fadH*	1,198	4,602	2,265	2,031	4,22E-02	-	C
Big	PS417_ RS18785	*pepV*	1,849	4,527	1,962	2,307	2,11E-02	-	E
Big	PS417_ RS19255	-	1,600	5,209	2,043	2,549	1,08E-02	-	S
Big	PS417_ RS20675	*purC*	2,063	3,410	1,621	2,104	3,54E-02	-	F
Big	PS417_ RS21560	*lpxA*	2,085	5,572	1,717	3,244	1,18E-03	-	I
Big	PS417_ RS21575	*ompH*	3,325	3,193	1,268	2,517	1,18E-02	-	M
Big	PS417_ RS22480	-	1,692	5,143	2,007	2,563	1,04E-02	-	C
Big	PS417_ RS23075	-	1,044	4,597	2,284	2,012	4,42E-02	-	S
Big	PS417_ RS24115	-	1,886	4,421	1,906	2,319	2,04E-02	-	S
Big	PS417_ RS24210	*ychF*	2,858	5,328	1,564	3,405	6,61E-04	-	J
Big	PS417_ RS24780	*accB*	3,113	2,713	1,201	2,259	2,39E-02	-	I
Big	PS417_ RS24815	*purD*	4,007	2,268	1,011	2,244	2,49E-02	-	F
Big	PS417_ RS25575	-	1,081	4,642	2,271	2,044	4,09E-02	-	S
Big	PS417_ RS26970	*aaeX*	1,560	5,038	1,909	2,640	8,30E-03	-	S
Big	PS417_ RS27605	*trkA*	1,117	4,673	2,302	2,030	4,23E-02	-	P
Big	PS417_ RS29980	-	2,175	3,880	1,615	2,403	1,63E-02	-	NT

Supplementary Table 6. Differentially expressed pathways of WCS417 in the rhizosphere of big versus small plants. Differentially expressed pathways with adjusted *p*-value < 0.05 are shown in bold.

Group	Pathway	Base mean	log_2FC	SE(FC)	stat	p-value	BH adj. p-value
Small	**Microbial metabolism in diverse environments**	**22,79147**	**-4,50458**	**0,96089**	**-4,68792**	**2,76E-06**	**6,07E-05**
Small	**Porphyrin metabolism**	**26,13238**	**-3,91884**	**0,833914**	**-4,69934**	**2,61E-06**	**6,07E-05**
Small	**Oxidative phosphorylation**	**14,38271**	**-4,61865**	**1,038297**	**-4,4483**	**8,66E-06**	**0,000127**
Small	**Biosynthesis of secondary metabolites**	**113,3684**	**-2,63524**	**0,601659**	**-4,37996**	**1,19E-05**	**0,000131**
Small	**Base excision repair**	**45,03589**	**-2,54755**	**0,609557**	**-4,17935**	**2,92E-05**	**0,000214**
Small	**Galactose metabolism**	**31,52857**	**-3,58734**	**0,857552**	**-4,18323**	**2,87E-05**	**0,000214**
Small	**Biosynthesis of amino acids**	**11,85068**	**-3,88903**	**0,966273**	**-4,02477**	**5,7E-05**	**0,000356**
Small	**Glycine, serine and threonine metabolism**	**11,75904**	**-3,87762**	**0,9706**	**-3,99507**	**6,47E-05**	**0,000356**
Small	**Lipopolysaccharide biosynthesis**	**107,6818**	**-2,20268**	**0,569538**	**-3,86748**	**0,00011**	**0,000538**
Small	**Phenylalanine, tyrosine and tryptophan biosynthesis**	**9,313516**	**-3,6558**	**0,97425**	**-3,75243**	**0,000175**	**0,000771**
Small	**Glycerolipid metabolism**	**68,05183**	**-2,08612**	**0,569633**	**-3,66222**	**0,00025**	**0,000865**
Small	**Glycerophospholipid metabolism**	**68,14347**	**-2,08854**	**0,569462**	**-3,66756**	**0,000245**	**0,000865**
Small	**Pentose phosphate pathway**	**7,697548**	**-4,803**	**1,313481**	**-3,6567**	**0,000255**	**0,000865**
Small	**ABC transporters**	**25,41206**	**-3,95908**	**1,181628**	**-3,35053**	**0,000807**	**0,002535**
Small	**Aminoacyl-tRNA biosynthesis**	**55,49617**	**-1,86926**	**0,577541**	**-3,23658**	**0,00121**	**0,003549**
Small	**Glyoxylate and dicarboxylate metabolism**	**6,58912**	**-3,56692**	**1,173426**	**-3,03975**	**0,002368**	**0,006511**
Small	**Thermogenesis**	**5,410971**	**-3,63474**	**1,229555**	**-2,95614**	**0,003115**	**0,008063**
Small	**Carbon metabolism**	**2,940145**	**-4,00178**	**1,634595**	**-2,44818**	**0,014358**	**0,03325**
Small	**Nitrogen metabolism**	**3,103843**	**-3,15464**	**1,378175**	**-2,289**	**0,02208**	**0,046262**
Small	Metabolic pathways	461,951	-0,86237	0,386345	-2,23213	0,025606	0,051213
Small	Nicotinate and nicotinamide metabolism	1,303923	-4,03353	2,043595	-1,97374	0,048411	0,078892
Big	**Cell cycle**	**316,9886**	**1,179701**	**0,477568**	**2,470227**	**0,013503**	**0,033007**
Big	**Two-component system**	**319,587**	**1,145009**	**0,471403**	**2,428941**	**0,015143**	**0,033315**
Big	DNA replication	190,6396	0,963163	0,456341	2,110622	0,034805	0,059615
Big	Mismatch repair	189,2931	0,988113	0,462928	2,134484	0,032803	0,059615
Big	Purine metabolism	190,8745	0,957934	0,454916	2,105738	0,035227	0,059615
Big	Pyrimidine metabolism	190,7828	0,95998	0,454908	2,110275	0,034835	0,059615

Supplementary Table 7. DEGs of *Arabidopsis* roots of big versus small plants.

Group	Gene	Name	Base mean	log2FC	SE(FC)	stat	p-value	BH adj. p-value
Small	AT5G65040	*IRM1*	355,634	-1,433	0,25346	-5,655	1,56E-08	1,25E-04
Small	AT4G15660	*GRXS8*	56,842	-3,521	0,685736	-5,135	2,82E-07	8,44E-04
Small	AT4G36870	*BLH2*	202,758	-1,120	0,237922	-4,707	2,51E-06	3,34E-03
Small	AT5G14340	*AtMYB40*	134,656	-1,093	0,233424	-4,684	2,81E-06	3,54E-03
Small	AT4G33040	*ROXY21*	243,180	-1,460	0,314346	-4,645	3,41E-06	4,08E-03
Small	AT3G03910	*GDH3*	221,302	-1,616	0,360439	-4,483	7,35E-06	6,33E-03
Small	AT5G56840	MYB-like	161,396	-1,492	0,330559	-4,514	6,35E-06	6,33E-03
Small	AT1G22150	*SULTR1;3*	40,277	-2,146	0,483114	-4,443	8,89E-06	6,49E-03
Small	AT1G29050	*TBL38*	709,156	-0,723	0,162944	-4,435	9,22E-06	6,49E-03
Small	AT5G19800	*AtHRGP2*	61,978	-1,601	0,375806	-4,260	2,05E-05	1,12E-02
Small	AT2G29750	*UGT71C1*	1486,630	-0,926	0,219153	-4,226	2,38E-05	1,24E-02
Small	AT3G58550	*LTPG22*	1547,870	-0,813	0,19429	-4,186	2,84E-05	1,33E-02
Small	AT1G63600	-	64,259	-1,623	0,390713	-4,154	3,26E-05	1,47E-02
Small	AT5G52450	*DTX16*	162,172	-1,098	0,264313	-4,155	3,26E-05	1,47E-02
Small	AT5G16570	*GLN1;4*	778,355	-1,146	0,276813	-4,140	3,48E-05	1,49E-02
Small	AT2G28056	*MIR172a*	39,811	-1,649	0,400562	-4,116	3,85E-05	1,53E-02
Small	AT1G72830	*ATHAP2C*	603,596	-0,897	0,218692	-4,100	4,14E-05	1,56E-02
Small	AT1G10120	*CIB4*	411,555	-0,590	0,144634	-4,078	4,54E-05	1,62E-02
Small	AT1G74830	*MYOB6*	94,855	-1,111	0,272928	-4,071	4,69E-05	1,62E-02
Small	AT5G24990	*DUF1336*	36,278	-1,410	0,353858	-3,986	6,73E-05	2,12E-02
Small	AT4G35270	*NLP2*	395,567	-0,955	0,240619	-3,971	7,17E-05	2,19E-02
Small	AT5G45380	*ATDUR3*	493,500	-1,121	0,283549	-3,954	7,68E-05	2,24E-02
Small	AT1G20640	*NLP4*	558,492	-0,719	0,182308	-3,944	8,03E-05	2,25E-02
Small	AT5G41663	*MIR319b*	328,630	-1,086	0,276575	-3,928	8,57E-05	2,36E-02
Small	AT2G27010	*CYP705A9*	258,529	-1,114	0,284502	-3,915	9,04E-05	2,41E-02
Small	AT2G44940	*ERF34*	491,063	-0,885	0,225874	-3,917	8,97E-05	2,41E-02
Small	AT3G60090	*VQ26*	98,453	-1,350	0,34474	-3,915	9,03E-05	2,41E-02
Small	AT2G20080	*TIE2*	264,735	-0,995	0,255804	-3,891	9,99E-05	2,55E-02
Small	AT3G59690	*IQD13*	547,312	-0,570	0,149715	-3,807	1,41E-04	3,21E-02
Small	AT3G60280	*UCC3*	27,283	-1,985	0,52065	-3,812	1,38E-04	3,21E-02
Small	AT4G38340	*NLP3*	159,893	-0,906	0,237542	-3,815	1,36E-04	3,21E-02
Small	AT5G10625	*FLP2*	443,514	-0,910	0,239555	-3,798	1,46E-04	3,29E-02
Small	AT1G52870	*PMP22*	13,184	-2,122	0,561657	-3,779	1,58E-04	3,50E-02
Small	AT3G20850	*4EBP3*	20,547	-2,342	0,619382	-3,781	1,56E-04	3,50E-02
Small	AT5G06510	*NF-YA10*	349,825	-0,922	0,245877	-3,749	1,78E-04	3,61E-02
Small	AT3G21690	*DTX40*	186,336	-0,921	0,247594	-3,719	2,00E-04	3,93E-02
Small	AT5G59050	-	176,734	-0,994	0,267084	-3,721	1,99E-04	3,93E-02
Small	AT1G03840	*IDD3*	148,952	-0,885	0,23857	-3,708	2,09E-04	4,03E-02
Small	AT5G60450	*ARF4*	284,392	-0,651	0,175825	-3,701	2,15E-04	4,09E-02
Small	AT3G27170	*ATCLC-B*	415,664	-0,749	0,20293	-3,691	2,23E-04	4,18E-02
Small	AT4G16400	-	117,338	-0,821	0,22291	-3,685	2,29E-04	4,21E-02
Small	AT5G20885	RING/U-box	994,006	-0,699	0,191577	-3,651	2,62E-04	4,71E-02
Small	AT3G46658	-	49,209	-0,902	0,248045	-3,637	2,76E-04	4,94E-02
Small	AT1G09812	*MRP*	75,003	-1,050	0,289928	-3,620	2,94E-04	5,00E-02
Small	AT2G33770	*ATUBC24*	399,844	-0,556	0,153122	-3,630	2,84E-04	5,00E-02
Small	AT3G02910	AIG2-like	869,468	-0,941	0,259414	-3,628	2,86E-04	5,00E-02
Small	AT5G57770	*FL8*	292,531	-0,600	0,165508	-3,623	2,91E-04	5,00E-02

Big	AT1G49500	*TFIID*	638,108	1,807	0,297484	6,076	1,24E-09	2,96E-05
Big	AT1G68670	*HHO2*	539,842	1,029	0,174566	5,896	3,74E-09	4,47E-05
Big	AT3G18080	*BGLU44*	183,347	1,360	0,246376	5,522	3,35E-08	1,61E-04
Big	AT4G37540	*LBD39*	195,134	2,215	0,399838	5,540	3,02E-08	1,61E-04
Big	AT4G37180	*HHO5*	320,386	1,482	0,27162	5,455	4,89E-08	1,95E-04
Big	AT2G41250	*HAD*	254,105	0,879	0,166527	5,276	1,32E-07	4,53E-04
Big	AT1G66760	*DTX9*	684,778	1,057	0,207943	5,084	3,69E-07	8,44E-04
Big	AT3G49940	*LBD38*	502,558	1,002	0,196393	5,102	3,36E-07	8,44E-04
Big	AT4G17340	*DELTA-TIP2*	2850,790	0,989	0,194869	5,075	3,87E-07	8,44E-04
Big	AT1G21310	*ATEXT3*	2826,560	0,952	0,193691	4,915	8,88E-07	1,77E-03
Big	AT1G30740	*AtBBE12*	4,037	5,706	1,180388	4,834	1,34E-06	2,41E-03
Big	AT2G48080	*ALKBH10A*	40,534	1,738	0,360306	4,824	1,41E-06	2,41E-03
Big	AT4G25100	*ATFSD1*	390,415	1,957	0,411875	4,751	2,02E-06	3,03E-03
Big	AT5G07725	-	4809,530	0,983	0,206453	4,760	1,94E-06	3,03E-03
Big	AT3G41768	-	14504,700	1,305	0,276634	4,719	2,37E-06	3,34E-03
Big	AT1G78000	*SEL1*	452,551	1,873	0,407251	4,600	4,22E-06	4,82E-03
Big	AT4G30110	*ATHMA2*	64,728	1,530	0,333747	4,584	4,57E-06	4,98E-03
Big	AT4G01430	*UMAMIT29*	71,893	1,705	0,374163	4,556	5,22E-06	5,44E-03
Big	AT1G25550	*HHO3*	931,154	0,788	0,175621	4,486	7,26E-06	6,33E-03
Big	AT2G24762	*AtGDU4*	58,941	1,623	0,362032	4,482	7,40E-06	6,33E-03
Big	AT4G06945	-	215,083	1,266	0,281002	4,504	6,66E-06	6,33E-03
Big	AT1G02940	*ATGSTF5*	3,151	6,090	1,369152	4,448	8,68E-06	6,49E-03
Big	AT2G05440	*ATGRP9*	581,523	2,920	0,656049	4,451	8,54E-06	6,49E-03
Big	AT2G43920	*ATHOL2*	67,069	1,955	0,440745	4,436	9,16E-06	6,49E-03
Big	AT4G13420	*ATHAK5*	150,033	1,880	0,421031	4,466	7,96E-06	6,49E-03
Big	AT1G09750	*AED3*	299,250	1,614	0,365759	4,413	1,02E-05	6,86E-03
Big	AT1G30720	*AtBBE10*	218,717	1,079	0,244607	4,410	1,03E-05	6,86E-03
Big	AT2G31690	*DALL5*	3,564	6,162	1,40139	4,397	1,10E-05	7,06E-03
Big	AT4G39364	snoRNA	48981,900	0,871	0,198263	4,393	1,12E-05	7,06E-03
Big	AT5G07105	-	2330,750	1,073	0,247375	4,338	1,44E-05	8,81E-03
Big	AT3G09855	-	1336,550	1,057	0,24484	4,319	1,57E-05	9,40E-03
Big	AT4G04185	snRNA-U5	1407,940	1,658	0,385822	4,296	1,74E-05	1,01E-02
Big	AT3G09865	-	1309,730	1,069	0,251009	4,259	2,05E-05	1,12E-02
Big	AT4G39363	snoRNA	13288,800	1,113	0,261264	4,259	2,06E-05	1,12E-02
Big	AT3G60700	DUF1163	11,740	3,403	0,801103	4,247	2,16E-05	1,15E-02
Big	AT5G59305	*CLE46*	15,751	2,499	0,594037	4,207	2,59E-05	1,32E-02
Big	AT1G67750	*PLL16*	403,885	0,941	0,22473	4,189	2,80E-05	1,33E-02
Big	AT4G06010	-	409,101	0,856	0,204289	4,192	2,76E-05	1,33E-02
Big	AT5G18170	*GDH1*	928,991	0,985	0,234923	4,194	2,74E-05	1,33E-02
Big	AT1G30730	*AtBBE11*	252,131	1,240	0,299091	4,145	3,39E-05	1,49E-02
Big	AT4G32810	*ATCCD8*	243,740	1,005	0,242711	4,139	3,49E-05	1,49E-02
Big	AT4G09135	snRNA-U6	179,799	1,036	0,250856	4,130	3,63E-05	1,53E-02
Big	AT1G29680	*OBAP2C*	19,546	1,887	0,458361	4,117	3,84E-05	1,53E-02
Big	AT3G48360	*ATBT2*	1348,110	2,532	0,614156	4,123	3,75E-05	1,53E-02
Big	AT4G37550	*IAMH1*	89,757	1,229	0,298711	4,113	3,91E-05	1,53E-02
Big	AT5G50130	-	100,253	1,263	0,307413	4,109	3,97E-05	1,53E-02
Big	AT1G26233	*SNOR95*	474,728	1,367	0,334081	4,091	4,29E-05	1,56E-02
Big	AT3G46560	*AtTIM9*	200,743	0,789	0,192606	4,094	4,24E-05	1,56E-02
Big	AT3G49780	*ATPSK3*	82,993	1,074	0,262334	4,095	4,23E-05	1,56E-02
Big	AT1G11640	-	54,448	1,682	0,412887	4,074	4,63E-05	1,62E-02
Big	AT1G52400	*ATBG1*	43,037	2,059	0,506644	4,065	4,81E-05	1,62E-02
Big	AT2G01010	-	13037,000	1,428	0,351004	4,067	4,76E-05	1,62E-02
Big	AT1G03090	*MCCA*	328,239	0,874	0,215642	4,055	5,01E-05	1,67E-02

Big	AT3G30725	*AtGDU6*	5,554	4,339	1,075367	4,034	5,47E-05	1,80E-02
Big	AT1G04660	-	7,354	3,219	0,799979	4,024	5,72E-05	1,85E-02
Big	AT2G43590	PR3-like	369,112	0,965	0,241746	3,991	6,59E-05	2,10E-02
Big	AT1G23560	*OBP32pep*	41,922	1,276	0,320934	3,977	6,97E-05	2,17E-02
Big	AT3G25190	*AtVTL5*	95,956	1,512	0,380954	3,969	7,22E-05	2,19E-02
Big	AT5G19120	-	258,855	1,297	0,327159	3,965	7,35E-05	2,20E-02
Big	AT3G15850	*ADS3*	44,998	1,182	0,298636	3,959	7,52E-05	2,22E-02
Big	AT3G25790	*HHO1*	65,669	1,754	0,444195	3,949	7,86E-05	2,25E-02
Big	ATMG01030	*ORF106E*	16,416	2,096	0,530817	3,948	7,89E-05	2,25E-02
Big	AT5G22380	*anac090*	11,604	3,314	0,840808	3,942	8,09E-05	2,25E-02
Big	AT2G18700	*ATTPS11*	246,782	1,146	0,293775	3,900	9,63E-05	2,51E-02
Big	AT5G27350	*SFP1*	611,092	0,723	0,185357	3,900	9,63E-05	2,51E-02
Big	AT3G24615	*snoRNA-Z43*	1353,980	0,664	0,170334	3,896	9,79E-05	2,52E-02
Big	AT3G53965	-	8,750	2,629	0,676457	3,886	1,02E-04	2,57E-02
Big	AT1G13300	*AtNIGT1*	349,144	0,955	0,246419	3,877	1,06E-04	2,63E-02
Big	AT5G01740	*NTF2*	15,062	2,536	0,654637	3,873	1,07E-04	2,65E-02
Big	AT5G07995	-	13,699	1,852	0,480391	3,855	1,16E-04	2,83E-02
Big	AT5G60800	*AtHMP53*	49,524	1,261	0,329826	3,823	1,32E-04	3,18E-02
Big	AT3G07860	Ubiquitin-like	586,157	0,748	0,196442	3,807	1,41E-04	3,21E-02
Big	AT3G13404	-	15,861	2,474	0,649063	3,811	1,38E-04	3,21E-02
Big	AT4G39361	snoRNA	15217,500	0,775	0,203478	3,808	1,40E-04	3,21E-02
Big	AT1G19530	*RGAT1*	69,560	1,081	0,287411	3,761	1,69E-04	3,53E-02
Big	AT1G60590	Pectin lyase-like	47,644	2,425	0,644102	3,765	1,66E-04	3,53E-02
Big	AT3G04165	-	32109,500	0,826	0,219612	3,761	1,69E-04	3,53E-02
Big	AT3G13432	*PROSCOOP24*	93,379	0,866	0,229903	3,767	1,65E-04	3,53E-02
Big	AT3G52840	*BGAL2*	381,604	1,528	0,406188	3,762	1,69E-04	3,53E-02
Big	AT5G00590	-	736,926	1,274	0,338331	3,765	1,66E-04	3,53E-02
Big	AT5G24470	*APRR5*	168,511	1,057	0,281179	3,761	1,70E-04	3,53E-02
Big	AT3G04775	-	1754,640	0,512	0,136204	3,758	1,71E-04	3,53E-02
Big	AT4G04845	-	24,159	1,820	0,485029	3,752	1,76E-04	3,59E-02
Big	AT2G22280	-	22,909	3,014	0,80488	3,745	1,80E-04	3,63E-02
Big	AT4G03125	-	3,757	4,541	1,217916	3,728	1,93E-04	3,85E-02
Big	AT2G36120	*DOT1*	2354,530	0,782	0,210965	3,706	2,10E-04	4,03E-02
Big	AT4G07135	-	415,007	1,273	0,343339	3,707	2,10E-04	4,03E-02
Big	AT1G69830	*AMY3*	209,224	0,741	0,200618	3,696	2,19E-04	4,14E-02
Big	AT4G07145	*snoRNA-U1*	413,156	1,280	0,346986	3,688	2,26E-04	4,20E-02
Big	ATCG00170	*RPOC2*	970,408	0,465	0,127076	3,663	2,49E-04	4,55E-02
Big	AT1G30700	*AtBBE8*	305,355	0,600	0,164319	3,652	2,60E-04	4,71E-02
Big	AT2G47690	-	181,187	0,856	0,236023	3,626	2,88E-04	5,00E-02
Big	AT4G15700	*GRXS3*	5,700	3,807	1,05029	3,625	2,89E-04	5,00E-02
Big	AT5G46325	tRNA-Leu	6,042	3,740	1,032811	3,621	2,94E-04	5,00E-02

Supplementary Table 8. Differentially expressed pathways of *Arabidopsis* roots of big versus small plants. Cat, GO category; P, precision; R, recall; GR, gene ratio; FC, enrichment; BP, biological process; MF, molecular function.

Group	Cat.	GO function	*p*-value	P	R	FC	Odds	GR
Small	BP	Regulation of RNA biosynthetic process	0,001	0,371	0,006	3,770	6,021	0,371
Small	BP	Regulation of DNA-templated transcription	0,001	0,371	0,006	3,770	6,021	0,371
Small	BP	Regulation of biosynthetic process	0,001	0,429	0,005	3,292	5,777	0,429
Small	BP	Regulation of gene expression	0,001	0,429	0,006	3,465	6,084	0,429
Small	BP	Regulation of macromolecule biosynthetic process	0,001	0,429	0,005	3,425	6,012	0,429
Small	BP	Regulation of macromolecule metabolic process	0,001	0,429	0,005	3,237	5,681	0,429
Small	BP	Regulation of RNA metabolic process	0,001	0,371	0,006	3,649	5,826	0,371
Small	BP	Regulation of nucleobase-containing compound metabolic process	0,001	0,371	0,006	3,533	5,640	0,371
Small	BP	DNA-templated transcription	0,001	0,371	0,006	3,474	5,546	0,371
Small	BP	Regulation of metabolic process	0,001	0,429	0,005	2,969	5,209	0,429
Small	BP	Regulation of primary metabolic process	0,003	0,371	0,005	3,080	4,914	0,371
Small	BP	Inorganic anion transmembrane transport	0,004	0,086	0,042	26,481	30,189	0,086
Small	BP	Regulation of biological process	0,006	0,514	0,004	2,211	4,557	0,514
Small	BP	Regulation of cellular process	0,006	0,486	0,004	2,271	4,421	0,486
Small	BP	Biological regulation	0,009	0,514	0,003	2,116	4,362	0,514
Small	BP	Inorganic anion transport	0,015	0,086	0,025	15,668	17,546	0,086
Small	BP	Response to nitrate	0,015	0,057	0,071	44,765	51,044	0,057
Small	BP	Regulation of phosphate transmembrane transport	0,022	0,029	1,000	626,714	-	0,029
Small	BP	Regulation of inorganic anion transmembrane transport	0,022	0,029	1,000	626,714	-	0,029
Small	BP	RNA biosynthetic process	0,022	0,371	0,004	2,399	3,823	0,371
Small	BP	Nucleic acid biosynthetic process	0,023	0,371	0,004	2,347	3,740	0,371
Small	BP	Xenobiotic detoxification by transmembrane export across the plasma membrane	0,032	0,057	0,040	25,069	27,649	0,057
Small	BP	Monoatomic anion transmembrane transport	0,032	0,057	0,040	25,069	27,649	0,057
Small	BP	Xenobiotic export from cell	0,032	0,057	0,040	25,069	27,649	0,057
Small	BP	Response to starvation	0,032	0,086	0,016	10,330	11,468	0,086
Small	BP	Nucleobase-containing compound biosynthetic process	0,032	0,371	0,004	2,204	3,512	0,371
Small	BP	Regulation of phosphate transport	0,032	0,029	0,500	313,357	644,088	0,029
Small	BP	RNA metabolic process	0,035	0,371	0,003	2,177	3,468	0,371
Small	BP	Monoatomic anion transport	0,039	0,057	0,033	20,890	22,882	0,057
Small	MF	DNA-binding transcription factor activity	0,000	0,324	0,008	4,705	7,002	0,324
Small	MF	Transcription regulator activity	0,000	0,324	0,007	4,425	6,581	0,324
Small	MF	Secondary active transmembrane transporter activity	0,004	0,135	0,018	10,924	12,835	0,135
Small	MF	DNA binding	0,008	0,324	0,005	3,099	4,600	0,324

Small	MF	Symporter activity	0,013	0,081	0,031	19,123	21,437	0,081
Small	MF	Transcription cis-regulatory region binding	0,021	0,162	0,008	4,672	5,609	0,162
Small	MF	Sequence-specific double-stranded DNA binding	0,021	0,162	0,007	4,614	5,538	0,162
Small	MF	Active transmembrane transporter activity	0,021	0,135	0,009	5,866	6,835	0,135
Small	MF	Transcription regulatory region nucleic acid binding	0,021	0,162	0,008	4,672	5,609	0,162
Small	MF	Nitrate:proton symporter activity	0,021	0,027	1,000	618,324	-	0,027
Small	MF	Monoatomic anion transmembrane transporter activity	0,023	0,054	0,049	30,162	33,464	0,054
Small	MF	Gamma-glutamylaminecyclotransferase activity	0,029	0,027	0,500	309,162	634,444	0,027
Small	MF	Double-stranded DNA binding	0,029	0,162	0,007	4,150	4,977	0,162
Small	MF	High-affinity secondary active nitrite transmembrane transporter activity	0,029	0,027	0,500	309,162	634,444	0,027
Small	MF	Glutamate dehydrogenase (NADP+) activity	0,035	0,027	0,333	206,108	317,222	0,027
Small	MF	Glutamate dehydrogenase (NAD+) activity	0,035	0,027	0,333	206,108	317,222	0,027
Small	MF	Quercetin 3'-O-glucosyltransferase activity	0,035	0,027	0,333	206,108	317,222	0,027
Small	MF	Glutamate dehydrogenase [NAD(P)+] activity	0,040	0,027	0,250	154,581	211,481	0,027
Small	MF	Xenobiotic transmembrane transporter activity	0,040	0,054	0,027	16,940	18,381	0,054
Small	MF	Sequence-specific DNA binding	0,040	0,162	0,006	3,574	4,283	0,162
Small	MF	Urea transmembrane transporter activity	0,045	0,027	0,200	123,665	158,611	0,027
Small	MF	Oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor	0,045	0,027	0,200	123,665	158,611	0,027
Big	BP	Cellular response to hypoxia	0,001	0,100	0,029	9,178	10,468	0,100
Big	BP	Cellular response to oxygen levels	0,001	0,100	0,029	9,102	10,379	0,100
Big	BP	Cellular response to decreased oxygen levels	0,001	0,100	0,029	9,102	10,379	0,100
Big	BP	RNA processing	0,001	0,200	0,013	4,040	5,096	0,200
Big	BP	Response to decreased oxygen levels	0,002	0,100	0,026	8,124	9,234	0,100
Big	BP	Response to oxygen levels	0,002	0,100	0,026	8,094	9,199	0,100
Big	BP	Response to hypoxia	0,002	0,100	0,026	8,246	9,377	0,100
Big	BP	Regulation of amine transport	0,010	0,029	0,286	89,531	128,606	0,029
Big	BP	Regulation of amino acid transport	0,010	0,029	0,286	89,531	128,606	0,029
Big	BP	Regulation of amino acid export	0,010	0,029	0,286	89,531	128,606	0,029
Big	BP	Regulation of amino acid transmembrane transport	0,010	0,029	0,286	89,531	128,606	0,029
Big	BP	Amino acid export across plasma membrane	0,012	0,029	0,250	78,339	107,172	0,029
Big	BP	Amine transport	0,017	0,029	0,200	62,671	80,379	0,029
Big	BP	RNA biosynthetic process	0,023	0,314	0,006	2,030	2,967	0,314
Big	BP	Nucleic acid biosynthetic process	0,028	0,314	0,006	1,986	2,903	0,314
Big	BP	RNA metabolic process	0,030	0,329	0,006	1,926	2,874	0,329
Big	BP	Cellular response to starvation	0,034	0,057	0,028	8,765	9,532	0,057

Big	BP	Nucleobase-containing compound biosynthetic process	0,042	0,314	0,006	1,865	2,725	0,314
Big	BP	Amino acid transport	0,042	0,043	0,039	12,369	13,410	0,043
Big	BP	mRNA trans splicing, SL addition	0,042	0,029	0,095	29,844	33,844	0,029
Big	BP	Cellular response to nutrient levels	0,042	0,057	0,025	7,933	8,603	0,057
Big	BP	Formation of quadruple SL/U4/U5/U6 snRNP	0,042	0,029	0,095	29,844	33,844	0,029
Big	BP	Cellular response to phosphate starvation	0,042	0,043	0,042	13,057	14,187	0,043
Big	BP	mRNA trans splicing, via spliceosome	0,042	0,029	0,095	29,844	33,844	0,029
Big	BP	Cellular process	0,042	0,900	0,004	1,189	11,865	0,900
Big	BP	Cellular response to stress	0,044	0,143	0,009	2,885	3,385	0,143
Big	MF	FAD binding	0,005	0,066	0,053	20,003	22,548	0,066
Big	MF	DNA-binding transcription repressor activity	0,005	0,033	0,500	187,525	386,695	0,033
Big	MF	Beta-gentiobiose beta-glucosidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	Formamidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	Hydrolase activity, hydrolysing O-glycosyl compounds	0,037	0,098	0,015	5,741	6,447	0,098
Big	MF	Hydrolase activity, acting on glycosyl bonds	0,037	0,098	0,013	4,881	5,469	0,098
Big	MF	Amygdalin beta-glucosidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	Flavin adenine dinucleotide binding	0,037	0,066	0,025	9,495	10,396	0,066
Big	MF	Abscisic acid glucose ester beta-glucosidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	4-methylumbelliferyl-beta-D-glucopyranoside beta-glucosidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	Esculin beta-glucosidase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	Palmitoyl-[glycerolipid] 7-desaturase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	9-cis-10'-apo-beta-carotenal cleavage oxygenase activity	0,037	0,016	1,000	375,049	-	0,016
Big	MF	All-trans-beta-apo-10'-carotenal cleavage oxygenase activity	0,037	0,016	1,000	375,049	-	0,016

Juan J. Sánchez-Gil[1], Sanne W. M. Poppeliers[1], Jordan Vacheron[2], Hao Zhang[1], Bart Odijk[1], Christoph Keel[2], Ronnie de Jonge[1,3]

[1] Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

[2] Department of Fundamental Microbiology, University of Lausanne, Lausanne, CH-1015, Switzerland

[3] AI Technology For Life, Department of Information and Computing Sciences, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

Author contributions:

Juan J. Sánchez-Gil: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – Original Draft. **Sanne W. M. Poppeliers, Bart Odijk:** Investigation. **Jordan Vacheron and Christoph Keel:** Methodology, Resources, Writing – Review and Editing. **Hao Zhang:** Methodology, Resources. **Ronnie de Jonge:** Supervision, Project administration, Funding acquisition, Validation, Formal analysis, Writing – Review and Editing.

Published as: Sánchez-Gil, J.J., Poppeliers, S.W.M., Vacheron, J., Zhang, H., Odijk, B., Keel, C., de Jonge, R., 2023. **The conserved *iol* gene cluster in *Pseudomonas* is involved in rhizosphere competence**. Current Biology 33. DOI: 10.1016/j.cub.2023.05.057

CHAPTER THREE

The conserved *iol* gene cluster in *Pseudomonas* is involved in rhizosphere competence

Abstract

The *Pseudomonas* genus has shown great potential as a sustainable solution to support agriculture through its plant-growth-promoting and biocontrol activities. However, their efficacy as bioinoculants is limited by unpredictable colonisation in natural conditions. Our study identifies the *iol* locus, a gene cluster in *Pseudomonas* involved in inositol catabolism, as a feature enriched among superior root colonisers in natural soil. Further characterisation revealed that the *iol* locus increases competitiveness, potentially caused by an observed induction of swimming motility and the production of fluorescent siderophore in response to inositol, a plant-derived compound. Public data analyses indicate that the *iol* locus is broadly conserved in the *Pseudomonas* genus and linked to diverse host-microbe interactions. Together, our findings suggest the *iol* locus as a potential target for developing more effective bioinoculants for sustainable agriculture.

Introduction

With a human population surpassing the 8 billion mark, the modern agricultural industry faces the challenge of increasing crop yields while reducing synthetic inputs. A promising solution leverages plant-beneficial microbes as soil amendments, collectively known as bioinoculants, which have shown positive effects such as promotion of plant growth, competition against phytopathogens, and activation of host immune defenses (Bakker et al., 2020). Typically, bioinoculants need to successfully colonise the rhizosphere, which comprises the roots and the region of the soil under their direct influence, to provide such benefits. Numerous candidates for bioinoculants have been identified and studied *in vitro* or in semi-controlled conditions, and the molecular mechanisms behind their positive effects are an actively studied topic. However, the natural rhizosphere environment presents significant differences that hinder the effective application of bioinoculants in agriculture. Successful colonisation of the rhizosphere depends on edaphic factors (factors such as soil chemistry and structure), host genotype, and in particular, the complex web of microbial interactions in the rhizosphere microbiome that results in a strong resistance to invasion (Poppeliers et al., 2023). Consequently, many promising bioinoculants fail to exert their positive effects or show variable efficacy in the field. Nonetheless, many organisms commonly colonise and persist in the rhizosphere driven by their intrinsic capabilities, some reaching higher abundances than others. For example, Proteobacteria and Bacteroidetes are repeatedly found to be enriched within the rhizosphere of diverse plant species (Ling et al., 2022). Particularly, the *Pseudomonas* genus comprises many representatives of rhizosphere-competent organisms. These pseudomonads display a vast diversity of competence traits and functional plasticity that grants them the ability to inhabit the rhizosphere of many plant species, such as secretion systems, evasion of plant immune defenses, or production of antimicrobial compounds (Bernal et al., 2017; Liu et al., 2018; Passarelli-Araujo et al., 2022; Pieterse et al., 2021; Yu et al., 2019; Zboralski and Filion, 2020). Characterising the underlying genetic determinants of such rhizosphere-competent microbes can help build a mechanistic understanding of the requirements of enhanced colonisation and in turn, such information can be used to design future effective bioinoculants to be applied in the field.

With the aim of finding genetic features for rhizosphere competence in natural conditions, we here compared a set of diverse *Pseudomonas* isolates from our in-lab collection for their ability to colonise the rhizosphere and root surface of *Arabidopsis thaliana* in natural soil. Genomic analysis of the best colonisers highlighted metabolism of *myo*-inositol (inositol), enabled by the *iol* locus, as an enriched explanatory feature for enhanced root colonisation. To determine the contribution of the *iol* locus to root colonisation, we generated an *iol*-deletion mutant and determined its competitiveness in root colonisation assays in soil. We show that the *iol* mutant is less competitive and that, as expected, the *iol* locus is required for consumption of inositol.

We further provide evidence suggesting that, besides acting as a nutrient source, inositol acts as a signal that induces swimming motility and the production of fluorescent siderophores. Genomic analyses show that the *iol* locus is broadly conserved in the *Pseudomonas* genus, and we discuss evidence of its expression and relevance in interaction with diverse hosts and successful colonisation. Together, our findings suggest a relevant role for the *iol* locus and plant-derived inositol in driving rhizosphere competence in pseudomonads.

Results

Competitive root colonisers are enriched for inositol phosphate metabolism

To identify genetic traits among root-associated microorganisms for competitive root colonisation in natural soil conditions, we introduced a set of plant root-associated *Pseudomonas* isolates in a natural soil and subsequently compared their abundances (**Figure S1**). To aid in isolate selection, we sequenced and compared the genomes of a group of root-associated pseudomonads from our collection, isolated from the same natural soil, and several well-known root-associated isolates, and we chose a group of six isolates that varied in phylogenetic spectrum, source environments, and effects on host plants (**Figure S2, Table S1**): *Pseudomonas simiae* WCS417 (WCS417), *Pseudomonas capeferrum* WCS358 (WCS358), *Pseudomonas protegens* CHA0 (CHA0), *Pseudomonas fluorescens* RS158 (RS158), *Pseudomonas* sp. WCS317 (WCS317) and *Pseudomonas* sp. WCS134 (WCS134). We introduced these isolates individually into natural soil reaching an inoculum density of 10^8 colony forming units (CFUs)/g of soil and grew *Arabidopsis thaliana* ecotype Col-0 seedlings on this soil for 21 days. Recognising that inferences about the colonising ability of root-associated microorganisms could be influenced by the method used to harvest the plant roots, we separated samples into two fractions: the rhizosphere, which includes soil attached to the root and that is under the influence of the root, and the rhizoplane or root fraction, which consists of the root without adhering soil. Following extraction of these fractions, we determined the absolute abundance of each isolate in all samples by amplicon-based sequencing of the *16S rRNA* gene, using an internal control as a reference for normalisation.

Isolates displayed differential population densities across fractions (**Figure 1**, **Figure S3A**). With the exception of WCS134, isolates survived similarly in bulk soil, although WCS358 and CHA0 showed slightly lower populations. Colonisation in rhizosphere samples is in general lower than in soil, although the relative differences seen among isolates is very similar to the one in soil. Root samples proved to be the most effective wash for assessing differential

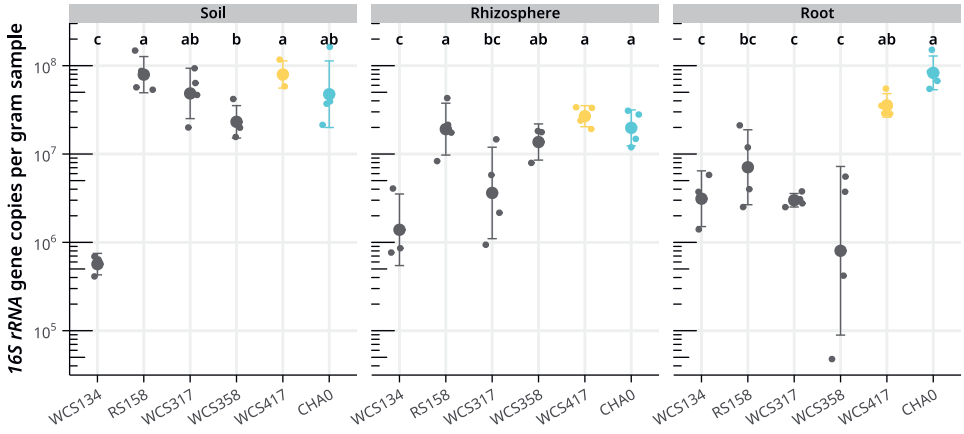


Figure 1. *P. protegens* CHA0 and *P. simiae* WCS417 are the best root colonisers. Absolute inoculum densities after 21 days post inoculation (dpi) measured as the number of *16S rRNA* copies per gram of sample found in soil (left), rhizosphere (middle), and root samples (right). The letters indicate groups of statistically different densities in each compartment according to a Sidak *post hoc* test on negative binomial tests. *P. protegens* CHA0 and *P. simiae* WCS417 are highlighted in cyan and yellow, respectively, to differentiate them from the remaining isolates. See also **Figures S1-S3** and **Table S1**.

colonisation among the isolates (**Figure 1**). Specifically, WCS417 and CHA0 exhibited approximately 8.6- and 20-fold higher colonisation levels, respectively, compared with the rest of the isolates. We expected each isolate to show intermediate abundances in the rhizosphere fraction compared with soil and roots. However, CHA0 and WCS134 reached higher populations in root than in rhizosphere samples; WCS417 colonises the rhizosphere and root fractions equally; and RS158, WCS317, and WCS358 exhibit higher abundances in the rhizosphere fraction, which suggests differing lifestyles among these pseudomonads (**Figure S3B**).

Based on these results, we selected WCS417 and CHA0 as the best root colonisers and proceeded to investigate the genetic traits underlying their superior root colonisation.

To investigate the genetic traits that distinguish these organisms from others, we extracted the orthologous protein families shared by WCS417 and CHA0. In total, we identified 203 such families exclusively found in the best colonisers (**Figure 2A**), comprising 204 and 206 orthologues in WCS417 and CHA0, respectively. Biosynthesis of antibiotics and biofilm formation were the most common annotations in these genes, accounting for approximately 5% of the orthologues each (12 genes in WCS417 and 10 genes in CHA0 for biosynthesis of antibiotics, and 10 genes in each microbe for biofilm formation). Notably, enrichment analysis on the annotations identified metabolism of inositol phosphate, the fourth most common annotation, as the most enriched function, followed by the tenth most common annotation for biosynthesis of siderophores (**Figure 2B**). The strong enrichment for inositol phosphate metabolism is caused by the presence of a

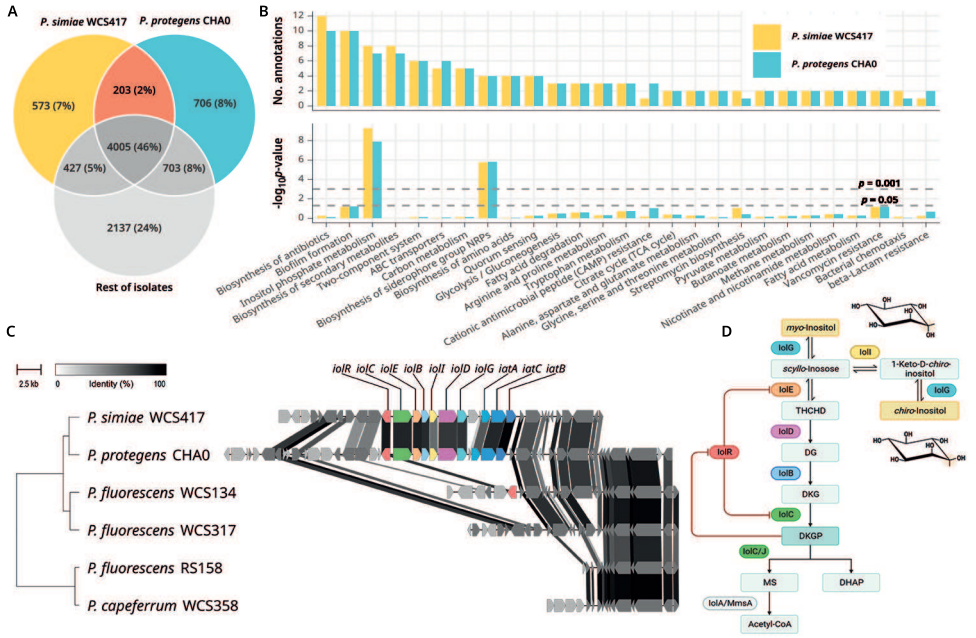


Figure 2. Inositol phosphate metabolism is the most enriched function in the best root colonisers. (A) Venn diagram depicting the genetic relationships between the best root colonisers and the rest of isolates. The orange area shows the set of orthologous protein families present in the genomes of WCS417 and CHA0 but absent in the other pseudomonads. (B) Bar plots showing the counts for KEGG pathway annotations (left) and their FDR-adjusted *p*-value in a Fisher's enrichment test for the set of shared orthologues between WCS417 and CHA0, versus the other pseudomonads in this study. Only the annotations that are found more than once are shown. Note that the most enriched function is inositol phosphate metabolism. (C) Phylogenetic tree of the isolates in this study and the *iol* locus responsible for the enrichment in the annotations for inositol phosphate metabolism. (D) Predicted pathway for inositol metabolism in *Pseudomonas*. The proteins in the pathway are coloured as the genes in (C). The chiral positions (C1, C2) that distinguish the *myo*, *scyllo*, and *chiro* enantiomers of inositol are shown as a shaded area on the 3D structural representations. THCHD: 3,5/4-trihydroxycyclohexa-1,2-dione; DG: 5-deoxy-glucuronate; DKG: 2-deoxy-5-keto-D-gluconate; DKGP: 2-deoxy-5-keto-D-gluconate-6-phosphate; MS: malonic semialdehyde (3-oxopropanoate); DHAP: dihydroxyacetone phosphate. Pathway in panel D was created with BioRender.com.

locus (hereafter *iol* locus) in WCS417 and CHA0 that is absent in the other isolates (**Figure 2C**). The locus spans 14 and 13 kbp in WCS417 and CHA0, respectively, and comprises 10 genes: a transcription factor (*iolR*), six genes responsible for inositol catabolism (*iolCEBIDG*), and three genes that encode an ABC transporter responsible for inositol intake (*iatABC*). In WCS417, an extra gene, *mocA*, encoding an oxidoreductase, is located upstream of the ABC transporters.

The IolCEBIDG enzymes potentially catalyse five consecutive reactions that transform the sugar alcohol *myo*-inositol successively into *scyllo*-inosose (IolG), trihydroxycyclohexadione (IolE), deoxyglucuronate (IolD), deoxy-keto-glucuronate (IolB), deoxy-keto-gluconate-6-phosphate (DKGP; IolC), and malonate semialdehyde and dihydroxyacetone phosphate (DHAP; IolC)

(**Figure 2D**), which could theoretically enter other pathways or be directed towards anabolism as acetyl-coenzyme A (CoA) in a reaction catalysed by MmsA/IolA, which is codified elsewhere in both pseudomonad genomes. The predicted transcriptional repressor of the locus, IolR, is inhibited by the interaction with the intermediate product DKGP, allowing the expression of the entire locus (Dong et al., 2020; Hellinckx et al., 2017; Kohler et al., 2011).

The *iol* locus enhances competence in natural soil

To investigate the contribution of the *iol* locus to pseudomonad competence in soil, we created the multi-gene Δ*iolRCEBIDG* mutant (Δ*iol*) in CHA0 and subsequently assessed its plant root colonisation. We inoculated GFP-tagged CHA0 wild type (WT) cells and mCherry-tagged Δ*iol* mutant cells individually and a 1:1 mixture of both genotypes in soil as before, reaching a inoculum density of 10^8 CFU/gr of soil, and then grew *Arabidopsis* seedlings for 21 days on this soil. In individual inoculations, the colonisation ability of both genotypes in soil and on the plant root was equivalent (**Figure 3A**). However, when co-inoculated, WT cells dramatically outcompeted mutant cells, achieving a 5.3-fold higher median density on the roots (**Figure 3B**).

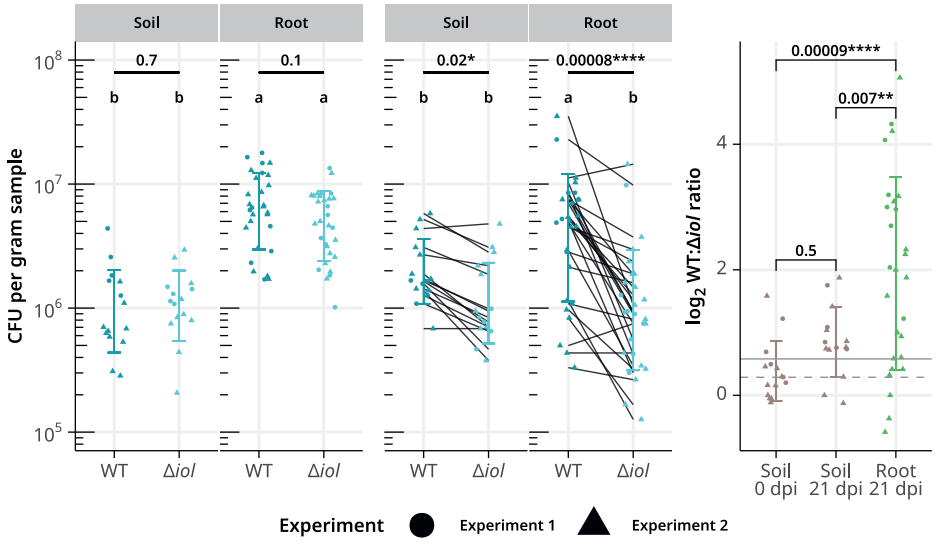


Figure 3. The *iol* locus enhances root colonising competence in *P. protegens* CHA0. (A) CFU counts per gram of sample of WT and mutant cells in individual inoculations in soil (left) and root samples (right) at 21 dpi. Significance values derive from a negative binomial test for count data, adjusting for variations between experiments. (B) Paired CFU counts of each genotype in the mixed 1:1 co-inoculation at 21 dpi in soil (left) and roots (right). Significance is calculated using a paired *t*-test on all pair-wise comparisons. The letters in (A) and (B) indicate groupings according to a negative binomial model adjusted for experimental variation. Notice that the number of Δ*iol* cells in co-inoculation experiments is similar on roots and in soil. (C) Log_2-transformed ratio of WT and Δ*iol* mutant cells in the mixed treatment. Large shapes indicate the median ratio per experiment, whereas lines show the mean initial ratio per experiment (solid line: experiment 1; dashed line: experiment 2). On the roots, WT cells achieve a significantly higher population at 21 dpi, reaching an approximate ratio of 5:1.

Interestingly, although the mutant achieves comparable densities when inoculated individually, its density on the root in competition was the same as in soil, suggesting that the root was primarily colonised by WT cells. As expected, the WT:Δ*iol* ratio on the root indicates a clear plant-driven advantage of WT cells (**Figure 3C**).

The *iol* locus is required for catabolism of inositol

To confirm the contribution of the *iol* locus to inositol catabolism, we performed growth curve measurements using inositol as the sole carbon source in minimal M9 medium. Our results showed that Δ*iol* mutant cells did not exhibit any growth after 80 h, in contrast to WT cells, which could utilise inositol to grow, albeit to only half the density reached with glucose (**Figure 4A** and **B).**

In non-minimal CC medium supplemented with inositol, we observed a comparable growth curve dynamic in WT cells after 36 h (**Figure 4C-D**), although the growth benefit was transient and variable (**Figure S4**). Surprisingly, we observed that inositol induced a decline in the density of Δ*iol* mutant cells after the exponential phase, at around 12 h, suggesting that inositol may have a toxic or bacteriostatic effect on these cells.

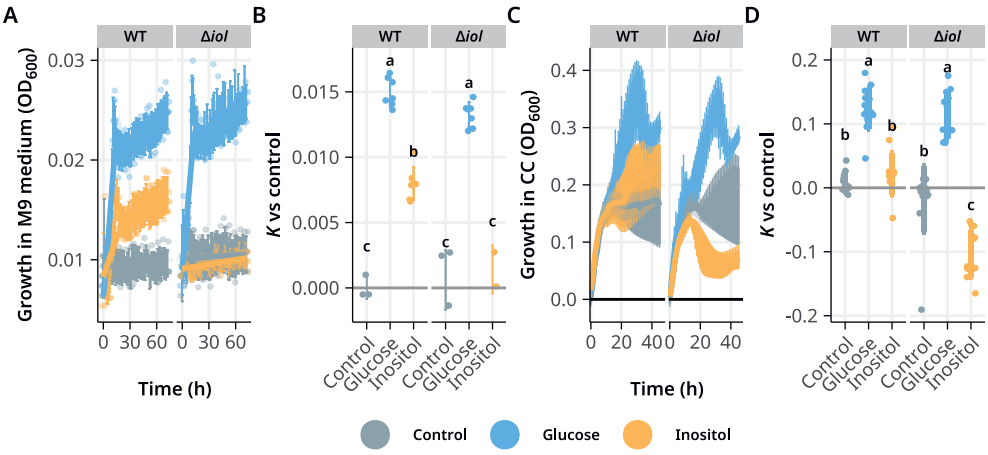


Figure 4. The *iol* locus is required for using inositol as a sole carbon source and affects growth of Δ*iol* cells on inositol. (A) Growth curves of WT and Δ*iol* mutant cell populations in non-supplemented minimal M9 medium (control) or in M9 medium supplemented with 100 mM glucose or inositol measured as OD at 600 nm. Mutant cells are unable to grow with inositol, indicating that the *iol* locus is required to catabolise inositol. (B) Fitted maximum population size (carrying capacity, *K*) of the growth data in (A). (C) Growth curves in non-supplemented CC medium (control), or in CC medium supplemented with 10 mM glucose or inositol. The data summarises two independent growth curve assays. Despite the variability within treatments across experiments, the curves show that both genotypes stabilise in the stationary phase in control medium. Mutant cells growing in inositol-supplemented medium follow a similar pattern during the stationary phase (0-12 h), but then the optical density drops noticeably. (D) Fitted carrying capacity (*K*, maximum population size) of the growth data in (C). For both assays, every time point consists of measurements for seven technical replicates. See also **Figure S4**.

Inositol enhances swimming motility through the *iol* locus

Inositol metabolism has been linked to motility and other relevant root colonisation traits in previous studies (Fry et al., 2001; Hamilton et al., 2021; Kohler et al., 2010; Vílchez et al., 2020). Based on this, we hypothesised that the enhanced competitiveness of CHA0 in soil may also be due to an induction of motility by inositol. To test this hypothesis, we conducted typical bacterial swimming and swarming assays as we did previously for spontaneous root-competitive *P. protegens* CHA0 mutants (E. Li et al., 2021b) in the absence and presence of inositol. These experiments revealed that inositol improved swimming motility of WT cells by 30% after 48 h, whereas it decreased swimming ability of mutant cells by 30% (**Figure 5**). Intriguingly, we observed that, although WT cell swarming motility is unaffected by the addition of inositol, mutant cell swarming is totally abolished in this condition (**Figure S5**). Glucose supplementation only promoted swimming at a 10-fold higher concentration, although the mean radius was still 10% smaller than that observed with inositol. This finding, together with the reduced growth benefit of inositol, suggests that inositol likely functions as a signal in addition to, or rather than, that of an advantageous carbon source.

Besides the effect on swimming, we noticed that mutant colonies exhibited a fainter color, analogous to the decrease in optical density during plank-

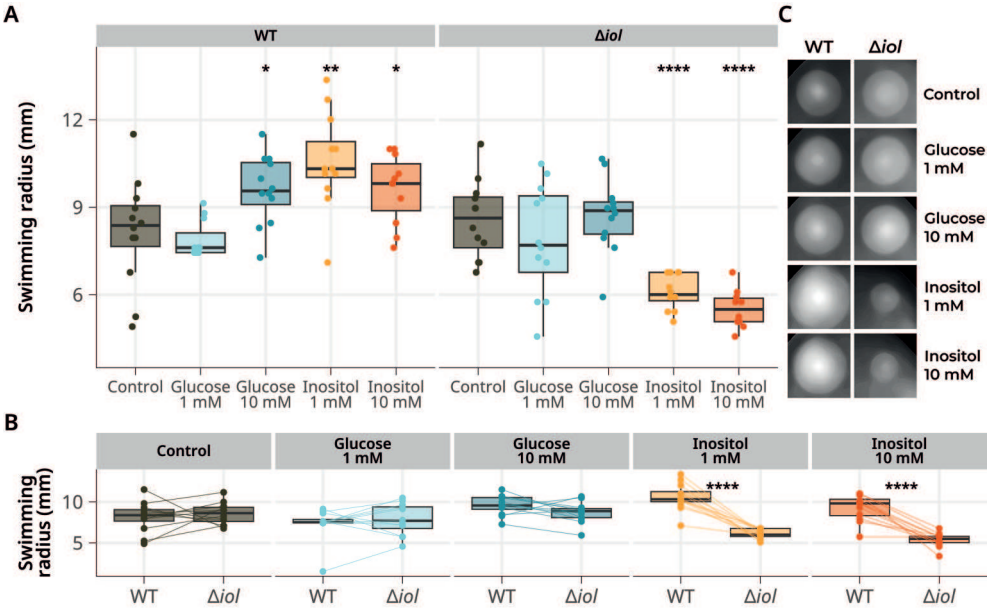


Figure 5. Inositol enhances swimming motility through the *iol* locus in *P. protegens* CHA0. (A) Swimming radius of WT (left) and Δ*iol* mutant cells (right) in CC medium either unsupplemented or supplemented with glucose or inositol at 1 or 10 mM. Inositol induces swimming motility at 1 mM in WT cells, whereas motility is suppressed in the mutant. (B) Same data as in (A) but shown per treatment and pairing the swimming halo for both genotypes per plate, highlighting the genotype-dependent differences in motility. (C) Representative images of WT and Δ*iol* mutant cell swimming behaviour in each treatment. See also **Figure S5**.

tonic growth at the onset of the stationary phase in growth curve assays. Therefore, we decided to monitor the changes in its phenotype over time in a dedicated assay. Here, we performed multiple inoculations again in swimming plates, inoculating both genotypes either individually or in combination. Mutant cells growing with inositol were visibly similar to the WT at 1 day post inoculation (dpi), but the color intensity of the colonies faded after reaching the end of the plate, approximately at the end of the second day (**Figure S6**). Intriguingly, we observed scattered mutant colonies reappearing in plates with inositol at 6 dpi, but only in plates where the mutant grew individually (**Figure 6A**). Moreover, WT colonies at this time exhibited visibly brighter yellow colonies, indicative of an increased production of fluorescent siderophores (**Figure 6**). Altogether, these results suggest a role for the *iol* locus and inositol metabolism in quorum behavior and competence in the rhizosphere.

The *iol* locus is broadly conserved among pseudomonads and a determinant for diverse host–microbe interactions

Considering the potential of the *iol* locus in determining competitiveness of pseudomonads in the rhizosphere, we hypothesised that it would be a conserved trait in the *Pseudomonas* genus (**Figure 7**). Exploratory analysis of all 15,742 *Pseudomonas* genomes in the NCBI RefSeq database revealed

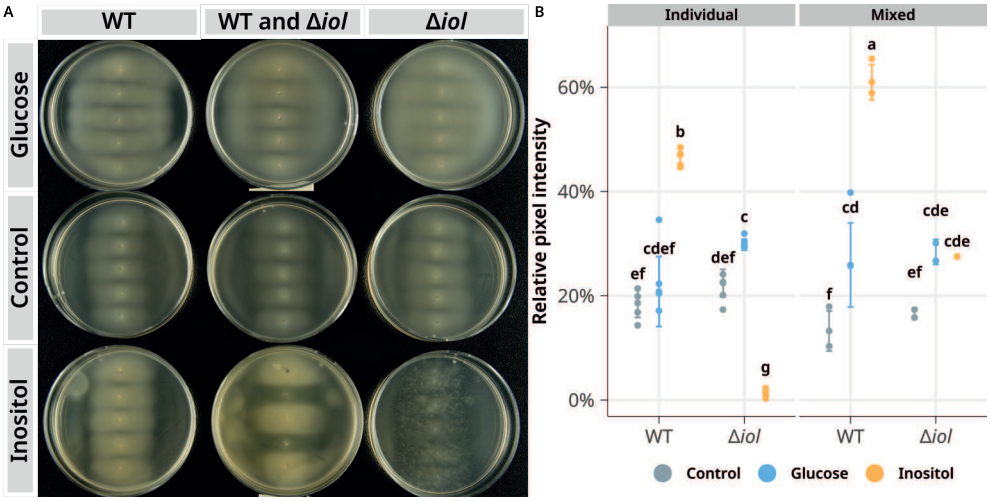


Figure 6. Inositol enhances the production of fluorescent siderophores and affects Δ*iol* mutant physiology. Qualitative assessment of the effect of inositol and glucose on WT and Δ*iol* mutant cells. (A) Representative images of colonies at 6 dpi in swimming plates supplemented with 10 mM glucose (first row), unsupplemented (second row), or supplemented with 10 mM inositol (third row). The first and third column show 5 inoculations of WT and mutant cells, respectively. The second column shows a combined inoculation of mutant cells surrounded by WT cells, used to evaluate possible interactions between both genotypes. Note how WT colonies show a brighter yellow color, characteristic of fluorescent siderophore production. Mutant colonies disintegrate at 3 dpi, and colonies re-appear at 6 dpi as seemingly independent and scattered populations. (B) Qualitative measurement of siderophores produced in the plates shown in (A) using the intensity of yellow pixels as a proxy. See also **Figure S6**.

that the *iol* locus is found in the majority of *Pseudomonas* species. The most notable exception is *Pseudomonas aeruginosa*, none of which encoded the *iol* locus (8,781 genomes, ~56% of all genomes). From the remaining 6,962 genomes, approximately 67% contain six or more *iol* genes spaced less than 20 kbp from each other or distributed over a limited number of genomic loci, yielding a total of 4,642 *iol*+ pseudomonads. Genomes in which the *iol* genes could not be found in a single locus typically had a low assembly quality (N50 < 50 kbp) and were considered *iol*+ if they met the requirement of the minimum number of six genes in a small (2-3) number of loci. In general, pseudomonads exhibit a binary *iol* locus presence-absence distribution at the species level. The frequency of *iol*+ strains is over 90% in 11 out of 23 species and over 80% in 14 out of 23 species, not including the umbrella species *Pseudomonas fluorescens* and *Pseudomonas* spp. The remaining nine species contain at most 10% (*Pseudomonas koreensis*) *iol*+ representatives (**Figure 7B**).

Among *iol*+ genomes, the core *iolRCEBDG* genes are found in a very high frequency, between 97% and 100% of genomes, indicative of co-evolution as a locus and suggesting that all genes are required to perform their biological function in the proposed catabolic pathway. Interestingly, around 10% of *iol*+ genomes lack *iolI* (454 pseudomonads), and the distribution of these 454 over the phylogenetic tree suggests that *iolI* loss occurred multiple times throughout the evolution of the genus. *IolI* encodes an inosose isomerase, which is responsible for the conversion of 1-keto-D-*chiro*-inositol into *scyllo*-inosose. The absence of *iolI* could therefore indicate the lack of *chiro*-inositol-derived compounds in the environment of these specific organisms. Other common structures of the locus contain a copy of a methylmalonate semialdehyde dehydrogenase (*iolA*/*mmsA*) between *iolB* and *iolD*, and/or a sugar phosphate isomerase/epimerase between *iolG* and the *iatACB* genes, potentially *iolH* (**Figure 7C**). Neither of these genes was found in WCS417 or CHA0. Lastl, a gene encoding a transcription factor, *dksA*, and a putative chemotaxis protein, *mcp*, are located downstream of the *iatACB* genes, suggesting a direct connection between the *iol* locus and motility.

In order to search for more evidence of the relationship between root colonisation and the *iol* locus, we explored publicly available transposon-insertion sequencing (Tn-seq) data in WCS417 (Cole et al., 2017). In this study, a collection of random insertion WCS417 mutants was let to colonise *Arabidopsis* roots in plates for seven days, after which the abundance of each mutant was used to assign a fitness (or performance) value for root colonisation for each gene. According to our analysis, root colonisation appears to be supported by *iolR*, the *iatABC* transporters, *iolD*, and the *traR/dksA* transcription activator. However, the remaining genes in the locus (*iolCEBIG*) did not show a significant contribution to colonisation, suggesting that consumption of inositol is not *per se* the most determinant feature for successful colonisation *in vitro* (**Figure 7D**). To further investigate the relationship between inositol and colonisation in the context of the *iol* locus, we made use of publicly available data from additional genome-wide, transposon mutant-based fitness

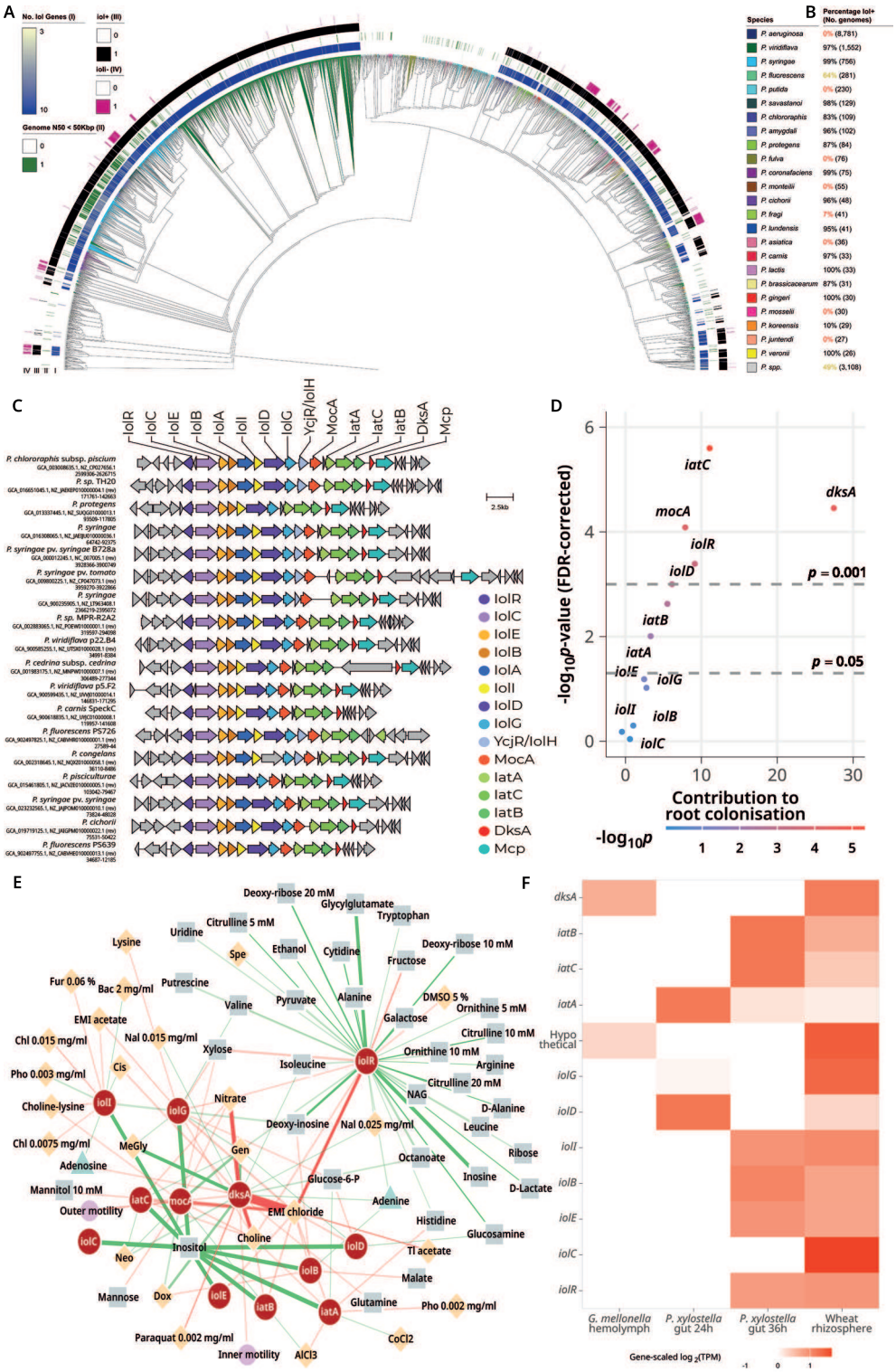

A
No. Iol Genes (I)
3
10
Iol+ (III)
0
1
Iol- (IV)
0
1
Genome N50 < 50Kbp (II)
0
1
IV III II I
B
Species
Percentage Iol+ (No. genomes)
P. aeruginosa 0% (8,781)
P. viridiflava 97% (1,552)
P. syringae 99% (756)
P. fluorescens 64% (281)
P. putida 0% (230)
P. savastanoi 98% (129)
P. chlororaphis 83% (109)
P. amygdali 96% (102)
P. protegens 87% (84)
P. fulva 0% (76)
P. coronafaciens 99% (75)
P. monteilii 0% (55)
P. cichorii 96% (48)
P. fragi 7% (41)
P. lundensis 95% (41)
P. asiatica 0% (36)
P. carnis 97% (33)
P. lactis 100% (33)
P. brassicacearum 87% (31)
P. gingeri 100% (30)
P. mosselii 0% (30)
P. koreensis 10% (29)
P. juntendi 0% (27)
P. veronii 100% (26)
P. spp. 49% (3,108)
C
IolR
IolC
IolE
IolB
IolA
IolI
IolD
IolG
YcjR/IolH
MocA
IatA
IatC
IatB
DksA
Mcp
2.5kb
D
-log10p-value (FDR-corrected)
iatC
dksA
mocA
iolR
iolD
p = 0.001
iatB
iatA
iolE
iolG
p = 0.05
iolI
iolB
iolC
Contribution to root colonisation
-log10p
E
Deoxy-ribose 20 mM
Glycylglutamate
Tryptophan
Citrulline 5 mM
Uridine
Lysine
Spe
Ethanol
Cytidine
Deoxy-ribose 10 mM
Fructose
Fur 0.06 %
Bac 2 mg/ml
Putrescine
Valine
Pyruvate
Alanine
DMSO 5 %
Ornithine 5 mM
EMI acetate
Galactose
Chl 0.015 mg/ml
Nal 0.015 mg/ml
Xylose
Citrulline 10 mM
Ornithine 10 mM
Cis
Isoleucine
Arginine
Pho 0.003 mg/ml
Citrulline 20 mM
Choline-lysine
NAG
Nitrate
D-Alanine
Deoxy-inosine
Nal 0.025 mg/ml
Leucine
Chl 0.0075 mg/ml
MeGly
Gen
Ribose
Adenosine
Octanoate
Inosine
Glucose-6-P
D-Lactate
Mannitol 10 mM
Adenine
Outer motility
EMI chloride
Histidine
Choline
Glucosamine
Inositol
Tl acetate
Neo
Malate
Mannose
Dox
Glutamine
Pho 0.002 mg/ml
Paraquat 0.002 mg/ml
CoCl2
Inner motility
AlCl3
F
dksA
iatB
iatC
iatA
Hypothetical
iolG
iolD
iolI
iolB
iolE
iolC
iolR
G. mellonella hemolymph
P. xylostella gut 24h
P. xylostella gut 36h
Wheat rhizosphere
Gene-scaled log2(TPM)

Figure 7. The *iol* locus is conserved in the *Pseudomonas* genus and contributes to diverse host-microbe interactions. (A) Phylogenetic tree of 15,743 *Pseudomonas* genomes available in the NCBI RefSeq database showing the distribution of the *iol* locus in the genus. 8,781 *P. aeruginosa* genomes, evolutionary distinct from the other 6,962 *Pseudomonas* genomes and all lacking the *iol* locus, were pruned from the tree. Ring I: number of *iol* genes in the highest scoring locus on the respective genome based on cblaster (Gilchrist et al., 2021). Ring II: factor indicating whether the N50 value (log_2-scaled) for a given genome is below 50 kbp as a measure for genome assembly integrity. Ring III: summary of all significantly *iol*+-scoring regions throughout the genome (minimal 6 out of 10 genes), confirming that genomes with fewer *iol* genes in the best scoring locus (ring I, light-colored) typically have a low genome N50 (ring II, green) but still encode the full locus. Ring IV: factor indicating the absence of *iolI* in the locus, showing that the loss of *iolI* follows a non-phylogenetic pattern in the genus. (B) Summary of the frequency of *iol*+ genomes in each *Pseudomonas* species, according to the annotated species names. (C) Locus alignment from a representative set of genomes in the tree, showing locus conservation in the genus. Some loci harbor an extra *iol* gene, *iolA*/*mmsA*, downstream of *iolB*, and two other genes, *mocA* and *ycjR*, encoding the sugar phosphate isomerase/epimerase *iolH*, upstream of the genes encoding the ABC transporters. (D) Scatterplot showing the relative contribution of the genes in the *iol* locus in *P. simiae* WCS417 to *Arabidopsis* root colonisation *in vitro* based on publicly available mutant fitness data (Cole et al., 2017). The x axis represents the contribution of each gene to colonisation, using the *t*-like effect-size statistic versus initial mutant populations, as previously reported (Cole et al., 2017). The y axis represents the FDR-corrected *p*-value for the contribution of each gene, as inverse logarithm of the raw value for ease of interpretation. (E) Bipartite network depicting the contribution of each gene in the *iol* locus in *P. simiae* WCS417 to survival in different environments included in the Fitness Browser (Price et al., 2018). Genes are shown as red nodes, and links connect these with conditions in which they show differential fitness across genome-wide fitness assays. Green and red links indicate a positive or negative contribution to survival, respectively. Link width and opacity indicate link weight (absolute contribution). Conditions are colored by their nature: carbon sources (grey), stressors (yellow), or nitrogen sources (light teal). Notice how *iolR* is not directly related to survival in inositol but in a myriad of other carbon sources and specific stressors. Notably, *iol* genes show many links to antimicrobial compounds. EMI, ethylmethylimidazolium; Gen, gentamicin; Neo, neomycin; Pho, phosphomycin; Dox, doxycycline; Bac, bacitracin; Chl, chloramphenicol; Cis, cisplatin; Fur, furfuraldehyde; NaI, nalidixate; Spe, spectinomycin; DMSO, dimethylsulfoxide; CoCl2, cobalt (II) chloride; AlCl3, aluminium (III) chloride; Tl acetate, thallium acetate; MeGly, methylglyoxal. (F) Heatmap of gene expression data from Vesga et al. (2020), showing expression of the *P. protegens* CHA0 *iol* locus during insect and plant colonisation/infection. Other relevant colonisation genes are also included. Expression data is shown as gene-scaled log_2-scaled transcripts per million.

experiments in WCS417, in which mutants were exposed to many different conditions, including stresses, carbon and nitrogen sources, and motility assays (Price et al., 2018). We used this data to construct a bipartite network connecting the genes in the *iol* locus to the conditions where they showed a significant fitness contribution (**Figure 7E**). Unexpectedly, while most of the genes seemed related to metabolism of inositol, neither *iolR* nor *dksA* appear to contribute directly to fitness in inositol as a sole carbon source. Instead, our analysis revealed that *iolR* is primarily related to consumption of other carbohydrates, while *dksA* plays a role in survival when cultured with antimicrobial compounds. Both transcription factors contribute to susceptibility to stressors like nitrate, choline, and imidazolium chloride, and promote survival in gentamicin. Interestingly, we observed a negative relationship between *dksA* and swimming motility, suggesting that the induction of motility by inositol in WT cells might be due to an inositol-induced inhibition of DksA activity. We did not find any obvious connection to the production of fluorescent siderophores.

Based on the manifest connection between root colonisation and the locus, we set out to explore its expression during host colonisation. For this, we examined publicly available data from a study performed by Pilar Vesga and co-authors (2020). Here, the transcriptional activity of *P. protegens* CHA0 was assayed during the colonisation of the hemolymph of an insect host (*Galleria mellonella*), of the insect gut (*Plutella xylostella*), and of wheat roots (*Triticum aestivum*) (Vesga et al., 2020). We found that the *iol* locus is expressed during colonisation on wheat roots and in the insect gut, but not in the insect hemolymph. Considering that hemolymph data were collected following direct injection, these data suggest that the *iol* locus is related to host perception and/or colonisation (**Figure 7F**).

Discussion

In our study, we compared the colonisation ability of a set of pseudomonads in non-sterile natural soil by amplicon-based sequencing of the *16S rRNA* gene. The use of spike-in internal controls for absolute quantification from complex environments has been applied in recent studies (Kim et al., 2021; Tkacz et al., 2018) and permits high-throughput quantification that does not depend on isolate-specific resistance markers and/or dilution series of the samples. Using this approach, we found that the compared colonization ability in our set of isolates varied substantially between rhizosphere and root samples (**Figure 1**), which suggests that these pseudomonads exhibit different lifestyles and environmental preferences: microbes enriched on the root will be firmly attached to the host and embedded in a biofilm on the surface, i.e., WCS417 and CHA0, whereas organisms preferring the rhizosphere might benefit from diffusing nutrients and potentially avoid stress from an active immune system. In any case, our findings indicate that careful consideration of the method selection should be taken in microbial colonisation studies, as different treatments might select for microbes with different lifestyles and lead to disparate conclusions about their ecological competence.

The two best root colonisers, WCS417 and CHA0, are well-known pseudomonads that have been widely studied for their plant-beneficial effects (Haas and Défago, 2005; Kupferschmied et al., 2013; Pieterse et al., 2021). The adaptation to a plant-associated lifestyle requires optimised consumption of plant-derived resources, among which inositol and its derivatives play an important role. In plants, inositol polyphosphates are main constituents of the plasma membrane and have key roles in signaling, immunity, and phosphorus storage (Castrillo et al., 2017; Dindas et al., 2022; Gomes et al., 2005; Tan et al., 2007), and inositol exudation through root tissues has been reported in many species (Chaparro et al., 2014; Pini et al., 2017). In fact, the ability to metabolise inositol emerged as a potentially distinctive trait in plant-associated pseudomonads in a genomic meta-analysis (Levy et al., 2018), supporting the idea that the presence of the *iol* locus in WCS417 and

CHA0 arises, indeed, from optimised adaptation to plant hosts. In general, expression of the *iol* genes *in planta* or in response to root exudates has also been observed in previous studies, although often indirectly or outside of the experimental focus. Expression of the *iol* genes in isolates from the genera *Cupriavidus, Burkholderia*, and *Rhizobium* was upregulated in response to root exudates of the legume *Mimosa pudica* (Klonowska et al., 2018). In pseudomonads, expression of the locus and its relation to competence have also been observed in isolates with divergent lifestyles. The pathogen *Pseudomonas syringae* pv. *tomato* DC3000 expresses the *iol* genes when grown in tomato apoplast extracts (Rico and Preston, 2008) and during growth in the *Arabidopsis* apoplast (Nobori et al., 2018). Likewise, *Pseudomonas syringae* pv. *syringae* B728a requires a functional *iol* locus for full pathogenicity in bean leaves (*Phaseolus vulgaris*), especially during apoplastic colonisation (Helmann et al., 2019; Yu et al., 2013). Notably, we found that the vast majority of *Pseudomonas syringae* genomes are *iol*+, highlighting its relevance for this plant pathogenic bacterium (**Figure 7B**). These findings are further corroborated by a recent comprehensive analysis of the distribution of the *iol* locus by Weber and Fuchs (2022). Their analysis revealed that hundreds of bacteria from diverse families and disparate ecological niches encode the capacity to catabolise inositol via the *iol* locus. These include animal and plant pathogens, soil commensals, and rhizosphere-associated bacteria similar to the pseudomonads discussed here (Weber and Fuchs, 2022).

However, the precise physiological function of the *iol* locus seems to be variable and isolate-specific. In a study comparing the expression patterns of a selection of eight pseudomonads in response to root exudates from *Brachypodium distachyon*, six out of eight strains showed upregulation of *iol* genes (Mavrodi et al., 2021). *Pseudomonas protegens* Pf-5, the closest relative of CHA0, increased the expression of the entire locus and yet did not exhibit any growth when grown in inositol as a sole carbon source, suggesting again that utilisation of inositol for growth is not the only function of the locus. In fact, upregulation of the *iol* locus in the other pseudomonads in their study, either fully or partially, did not seem to be a requirement for growth in inositol at all.

In our root colonisation assays, Δ*iol* mutant CHA0 cells can colonise equally with WT cells when inoculated individually, whereas in competition with WT cells, they fail to reach a higher population density than in soil. We also observed growth defects in mutant cells when grown in inositol. Interestingly, a similar phenomenon has been described in the Alphaproteobacteria *Sinorhizobium meliloti*, *Sinorhizobium fredii*, and *Rhizobium leguminosarum*, where the *iol* genes play an important role in defining competence in root nodules (Fry et al., 2001). In *Sinorhizobium*, *iolG* mutants form abnormal bacteroids, are impaired in nitrogen fixation, and are unable to utilise rhizopines, *scyllo*-inositol derivatives produced in nodules, which ultimately results in a loss of competitiveness against WT cells (Galbraith et al., 1998; Jiang et al., 2001; Kohler et al., 2010; Murphy et al.,

1995). Yet these growth defects do not seem to pose a disadvantage in our individual inoculations because mutant density on the root is not different from WT populations. The stimulation of swimming motility in WT cells by inositol, however, might promote quicker arrival during initial stages of colonisation and therefore enable an earlier occupation of the root surface. In fact, expression of the *iol* genes is restricted to the initial phase of infection in the plant pathogen *Ralstonia solanacearum* via a quorum-sensing mechanism. Moreover, knockout mutants for *iolG* in this pathogen are impaired in their capacity to colonise tomato roots during the initial stages of infection (Hamilton et al., 2021), and remarkably, swimming motility during this phase has also been linked to successful pathogenicity in an independent study (Corral et al., 2020). A direct link between inositol and bacterial motility was previously observed in *Priestia megaterium* (formerly *Bacillus megaterium*). In this organism, inositol exuded by the roots of *Arabidopsis* and tomato plants (*Solanum lycopersicum*) induced a strong chemotactic response and enhanced biofilm formation (Vílchez et al., 2020). Noteworthy, quorum-dependent responses, such as the one described in *Ralstonia* and *Priestia*, commonly take place in the transition to the stationary phase (Goo et al., 2012). Furthermore, rhizopines, which act as quorum-sensing signals in *Sinorhizobium* species (Zuniga-Soto et al., 2019), affect motility, growth, and biofilm formation and also act during the stationary phase and not before (Krysciak et al., 2014). The connection between inositol metabolism, the *iol* locus, and quorum behavior could explain the decrease in optical density in the Δ*iol* CHA0 at the onset of the stationary phase that we observe and, additionally, the enhanced production of siderophores that we observed after 6 days on plates. This is further supported by our results in swarming motility, for which quorum-based coordination is required (Akbari Kiarood et al., 2020; Daniels et al., 2004), where inositol does not seem to affect swarming of WT cells, whereas it abolishes swarming in the *iol* mutant.

Our findings imply an important role for the *iol* locus during *Pseudomonas* root colonisation beyond the utilization of inositol as a carbon source. Our analyses of publicly available genome-wide fitness data show that root colonisation *in vitro* depends primarily on the *iolR*, *iatABC*, and *dksA* genes while being seemingly independent of inositol consumption. Additionally, IolR does not affect inositol consumption, contrasting with previous results (Dong et al., 2020; Hellinckx et al., 2017; Kohler et al., 2011), but is required for root colonisation and exhibits many links to the metabolism of diverse carbon sources. Moreover, swimming motility appears to be inhibited by DksA, suggesting that the induction of swimming by inositol in our experiments might be due to an inositol-induced repression of DksA. Together with our findings on the overproduction of fluorescent siderophores, this information supports a more complex mechanism in response to the presence of host-derived inositol in the environment. Indeed, IolR regulates behavioral responses toward inositol in other organisms, such as the expression of type III secretion system in *Salmonella typhimurium* (Cordero-Alba et al., 2012), xylose uptake in *Corynebacterium glutamicum* (Emami et al., 2020), or cell

autoaggregation and biofilm formation in *Aeromonas hydrophila* (Dong et al., 2020). Interestingly, cell aggregation by the Pel and Psl exopolysaccharides (EPS) increases siderophore production in *Pseudomonas aeruginosa* by raising internal levels of cyclic di-GMP, which, in turn, triggers a Gac-mediated activation of genes related to siderophores production (Visaggio et al., 2015). At the same time, survival in the stationary phase and the production of Pel and Psl EPS are enabled by the sigma factor RpoS (Irie et al., 2010). It seems, therefore, plausible to hypothesise that our observed increase in production of fluorescent siderophores in WT cells is due to an interaction with RpoS in the stationary phase, resulting in an inositol-induced change in EPS composition and an eventual increase of siderophore production. This would translate into enhanced iron-scavenging capabilities in WT cells in response to host clues, with host-derived inositol as a proxy, whereas in mutant cells, a misregulation of RpoS activity could be the cause of aberrant growth in the stationary phase.

In general, the *iol* locus has been associated to a myriad of different phenotypes in other bacteria, and its physiological and ecological functions also seem to go beyond the mere obtention of energy from inositol. Instead, the locus appears to be embedded in a complex mechanism of sensing and responding to host-related environmental cues. For example, in *Caulobacter crescentus*, the iron-deficiency response regulator Fur represses expression of the *iol* locus in an iron-independent manner (da Silva Neto et al., 2013), and in *Salmonella typhimurium*, the locus is linked to antibiotic resistance (Lynch and Kierstead, 1985). In salmonellae, the small RNA *RssR* controls expression of the *iol* locus and inositol metabolism (Kröger et al., 2018), leading to complex expression dynamics controlled at the bacterial population level (Kröger et al., 2011).

The specificity of the phenotype associated with the *iol* locus in pseudomonads may be reflected in the diversity of locus decorations. Our analysis of the *iol* locus among available *Pseudomonas* genomes shows that, besides the core locus (*iolRCEBDG*), specific conformations with additional genes can be found. In particular, we observed that *iolI* has been lost in ~10% of *iol*+ pseudomonads. The inosose isomerase IolI enables the uptake of *chiro*-inositol through the conversion of 1-keto-D-*chiro*-inositol into *scyllo*-inosose, which can be readily incorporated into the main pathway. Although derivatives of inositol can be found in soil in diverging amounts, the *chiro* conformation is rare in soil. However, it is produced by plants of many species in differing amounts and forms (Gomes et al., 2005; Siracusa et al., 2022; Turner et al., 2002; Xia and Wang, 2006), which further supports the link between the presence of *iolI* and a plant-associated lifestyle. Together with the lack of well-defined taxon-specific loss of *iolI*, this suggests that the different structures of the *iol* locus that we described here might correspond to host-specific locus conformations according to the composition of host exudates. Divergent *iol* loci in *Pseudomonas* can include several other genes, such as *iolA*, *mocA*, and *dksA*. The methylmalonate semialdehyde dehydrogenase IolA/MmsA produces acetyl-CoA from the last digestion

product in the predicted pathway, malonate semialdehyde, or malonyl-CoA from methylmalonate semialdehyde. This gene is frequently found elsewhere in the genome, suggesting that this conversion can be regulated in both inositol-dependent and -independent ways. The *mocA* gene was originally annotated as a Gfo/Idh/MocA-like oxidoreductase, which is a large protein family with many disparate functions described, among them the metabolism of rhizopines in *Sinorhizobium* species (Taberman et al., 2016). In parallel, annotations of this gene in some pseudomonads suggest that this gene might encode a homologue of IolW, a *scyllo*-inositol dehydrogenase that catalyzes the conversion of this inositol enantiomer into *scyllo*-inosose and allows its incorporation in the main pathway. Both annotations, MocA and IolW, point toward the use of *scyllo*-inositol or rhizopine derivatives and possible associations with nodulating bacteria. Further experiments are required to clarify whether pseudomonads can in fact use rhizopines or other *scyllo*-inositol-derived exuded compounds in the rhizosphere. Lastly, DksA, encoded in the 3'-end of the locus, appears involved in survival in stress induced by methylglyoxal in our analysis of the publicly available genome-wide fitness data. This is one of the few reported phenotypes for IolS in *Bacillus* species, an aldo-keto reductase with an otherwise unknown function (Chandrangsu et al., 2014; Ehrensberger and Wilson, 2004). Similar to DksA in pseudomonads, IolS does not seem to be required for inositol metabolism, yet it is conserved in the *iol* locus in bacilli and upregulated in response to maize exudates (Fan et al., 2012).

In conclusion, we show that the *iol* locus determines rhizosphere competence in pseudomonads. Detailed characterisation further revealed that its function entails more complex mechanisms than the mere consumption of the sugar alcohol *myo*-inositol, and that these mechanisms are taxon- and, possibly, host-specific. In parallel with our study, Bridget O'Banion and co-authors also identified inositol as an important plant-derived compound determining root colonisation phenotypes in *Pantoea*, including an enhancement of bacterial motility among them (O'Banion et al., 2023). In our pseudomonad, plant-derived inositol might induce the repression of DksA and enable an enhancement of swimming motility, subsequently promoting early arrival in the rhizosphere, prompter colonisation, and therefore enhanced competitiveness. However, the aberrant growth in mutant cells during the switch to the stationary phase implies an additional mechanism during this phase that results in increased capability of iron scavenging of WT cells. Future research should dig deeper into the mechanistic processes by which competence is optimised by the *iol* locus in the *Pseudomonas* genus, including the potential role of RpoS and EPS production and composition. The findings in our study provide evidence of the fundamental basis of rhizosphere colonisation and signify an important step toward the development of robust bioinoculants in sustainable agriculture.

Methods

Growth of *Arabidopsis thaliana*

Seeds of *Arabidopsis* accession Col-0 were surface-sterilised with chlorine gas in a glass chamber and subsequently aerated in a flow cabinet to remove the residual chlorine for 30 min. Sterile seeds were sown in agar plates containing 1x Murashige and Skoog (MS) medium (Duchefa Biochemie) supplemented with 0.5% sucrose. Plates were sealed twice with parafilm, let stratify for 48 h at 4°C, and positioned vertically in a growth chamber under short-day conditions for two weeks (22 °C; 10 h light/14 h dark; light intensity 100 µmol/m^2/s).

Bacterial cultures

All pseudomonads are cultured on King's B (KB) medium plates (2% proteose peptone, 8.6 mM K_2HPO_4, 6.1 mM $MgSO_4 \cdot 7H_20$, 1% v/v glycerol) agar plates and incubated overnight at 28 °C. A live aliquot of *Salinibacter ruber* M31 was grown for 14 days in liquid Medium 936 for *S. ruber* as prescribed by DSMZ (3.3 M NaCl, 200 mM $MgSO_4$ x 7 H_2O, 170 mM $MgCl_2$ x 6 H_2O, 8.5 mM $CaCl_2$ x 2 H_2O, 67 mM KCl, 3 mM $NaHCO_3$, 6 mM NaBr, 1 g/l yeast extract, pH 7.2) at 37 °C and 100 rpm.

Soil treatments, transplantations, and microbiome sampling

A schematic representation of the experimental design can be seen in **Figure S1**. Per isolate, single colonies were spread onto five KB plates and incubated overnight at 28 °C. Per plate, 5 ml of $MgSO_4$ 10 mM was used to scrape the microbial cell layer off the surface, and the resulting suspension was collected into 50-ml tubes for every isolate. The cell suspension was washed thrice by centrifugation at 3000 x g for 10 min, resuspending in 40-50 ml $MgSO_4$ 10 mM every step. The final cell pellet was resuspended in 20 ml of $MgSO_4$ 10 mM, and the bacterial density was measured by determining the optical density (OD) at 600 nm (equivalence: 1 OD_{600} unit, $8 \cdot 10^8$ CFU/ml).

In this study we used top soil collected from the Reijerscamp natural reserve (52°01'02.55", 5°77'99.83", Wolfheze, Netherlands), a gleyic placic podzol consisting of coarse sand and gravel where natural *Arabidopsis* populations occur and that has been previously used for microbiome studies in *Arabidopsis* (Berendsen et al., 2018). The soil was dried, sieved to remove large gravel, and inoculated with each microbial suspension to reach an inoculum density of 10^8 CFU/g of soil in a final volume of 100 ml/kg of soil. For the mock treatment, the soil was moistened with $MgSO_4$ 10 mM at 100 ml/kg of soil. This initial inoculum density was decided based on the detection limit of our spike-in method in preliminary assessments of the technique. Given a total *Salinibacter* spike-in abundance of $9.84 \cdot 10^5$ copies per sample, we

calculated the density detection limit by dividing this abundance by sample weight. In our experiment, this resulted in an average detection limit across samples of $2.1 \cdot 10^4 \pm 2.5 \cdot 10^3$ CFU/g, with certain samples showing a detection limit up to 10^6 CFU/g (**Figure S7**), indicating that the initial inoculation density had to be higher than 10^6 CFU/g in order to ensure reliable quantifications of our inoculants. Therefore, and also to account for the possibility that the density of some microbes could significantly decrease in time, we decided to use an initial inoculum density of 10^8 CFU/g as previously reported in other assays (Van Wees et al., 2013).

Per inoculation treatment, 12 two-weeks-old *Arabidopsis* seedlings were transferred to individual pots (diameter = 5.5 cm and height = 5 cm; MXC5,5 Pöppelmann) containing 100 g of inoculated Reijerscamp soil each. All pots were placed on trays on individual Petri dishes to prevent cross-contaminations and kept in the growth chamber under the same conditions as mentioned above. The trays were kept closed with a lid for the first two days post-transplantation. For the remaining time of the cultivation period, the lids were removed, and the plants were watered every two days with 5-10 ml of tap water. Roots were eventually harvested at 21 dpi (5 weeks-old) and processed in two different ways to obtain the different plant fractions. For rhizosphere samples, roots were washed in 50-ml tubes with 30 ml sterile 10 mM $MgSO_4$ by inverting the tube 10 times. For root samples, roots were instead washed twice in 50-ml tubes by vortexing 15 s in 30 ml sterile phosphate-Silwet buffer (PBS-S; 137 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 KCl, 0.02% Silwet). Roots cleaned with both methods, representing different fraction samples, were retrieved with sterile tweezers, tapped dry on clean paper, and transferred to 2-ml tubes containing 1 ml $MgSO_4$ 10 mM with two glass beads. For soil samples, we collected approximately 0.25 g of soil from unplanted pots. Every fraction (soil, rhizosphere, and root) consisted of four replicates. Sample weight was obtained by weighing the tubes before and after collecting the sample. The tubes were shaken twice at 20 Hz in a TissueLyser, snap frozen in liquid nitrogen, and stored at -80°C until DNA extractions. This experiment to screen for superior root colonisation was performed once.

DNA extractions, *16S rRNA* amplicon library preparation, and data normalisation

Nucleic acid extractions from all sample types were performed on a KingFisher Flex (Thermo Fisher Scientific) using a protocol based on the MagAttract PowerSoil DNA KF Kit. The resulting DNA was quantified with Qubit fluorometric quantitation (Invitrogen) with the BR dsDNA assay kit (Thermo Fisher Scientific).

In order to obtain absolute counts of each isolate, we used DNA from *Salinibacter ruber* (*Salinibacter*), an extreme halophile found in saltern crystalliser ponds that is not found in our samples and that encodes a single *16S rRNA* gene copy in its genome. Based on the genome size of 3.6 Mbp

(GenBank accession ID GCA_000013045.1), we calculated an expected copy number of $2.46 \cdot 10^5$ chromosomes per nanogram of genomic DNA, with an equivalent number of copies of the *16S rRNA* gene. Cell lysis buffer was spiked with 4 ng of *S. ruber* DNA per sample, aiming at a final concentration of 0.1-1% of the total amount of DNA. Library preparation was conducted according to Illumina's 16S Sequencing Library Preparation protocol, using phasing primers for the V3-V4 region (**Table S2**). The final libraries were quantified using Qubit fluorometric quantitation with the BR dsDNA assay kit, pooled at a sample concentration of 4 nM, and sequenced on Illumina MiSeq to obtain 2x300 bp paired-end reads.

The resulting reads were processed using cutadapt (Martin, 2011) and the DADA2 pipeline (Callahan et al., 2016). To increase sensitivity of the DADA2 pipeline for the introduced isolates and their respective *16S rRNA* sequences, we used the known V3-V4 region of the isolates, plant mitochondrion, plant chloroplast, and *Salinibacter* as priors in the amplicon sequence variant (ASV) inference step. For every inoculation treatment, ASVs were inferred from all treated samples and mock samples by prioring with the sequence of the respective isolate, plant, and *Salinibacter*. Sequences used as priors are listed in **Table S3**.

After removal of chimeric sequences, the counts of every isolate were normalised with respect to the counts of *Salinibacter* by taking the ratio of isolate counts to the observed *Salinibacter* counts, and multiplying this ratio by the total number of added *Salinibacter* copies in the sample preparation (4 ng/sample · $2.46 \cdot 10^5$ 16S gene copies/ng, $9.84 \cdot 10^5$ copies/sample). Final microbial densities are calculated by normalising estimated *16S rRNA* copy numbers with plant weight and subtracting counts from uninoculated samples when applicable, i.e., for WCS417 and WCS358. **Figure S3** shows the raw densities used to generate **Figure 1** before subtraction of the density in uninoculated samples.

Genomic analysis

To identify genes that could potentially explain the enhanced colonisation ability in *P. simiae* WCS417 and *P. protegens* CHA0, we subjected the genomes of all isolates to a comparative genetic analysis based on presence/absence of orthologues (**Figure S2**). First, proteomes were inferred from the genome sequences using Prokka (Seemann, 2014). The proteomes were then used as input for OrthoFinder (Emms and Kelly, 2019) in order to infer phylogenetic relationships among all genes and group them into orthologous protein families or orthogroups. The result is a table of all orthogroups inferred and a list of its orthologues found in each individual genome. In parallel, all proteomes were annotated using the eggNOG-mapper (Huerta-Cepas et al., 2019) and the KEGG database (Kanehisa and Goto, 2000). Enriched functions in the set of orthologues shared by only WCS417 and CHA0 were identified by a Fisher's exact test by comparing the functions of these orthologues with those in the complete proteomes.

Mutant construction

Deletion of the entire *iol* gene cluster (from PPRCHA0_2627 to PPRCHA0_2633, i.e., from *iolR* to *iolG*) was generated by homologous recombination using the suicide vector pEMG and the I-SceI system adapted to *P. protegens* (Kupferschmied et al., 2014). Briefly, the pEMG::Δ*iol* vector for homologous recombination was constructed by fusing a fragment upstream with a fragment downstream of the *iol* locus, amplified with primers *del-iol-1*/*del-iol-2* and *del-iol-3*/*del-iol-4*, respectively, and cloning the resulting fused fragment into pEMG. Competent CHA0 cells were electroporated with pEMG::Δ*iol* and successful insertional mutants were selected on Luria-Bertani (LB) plates supplemented with 25 µg/ml kanamycin. Competent mutant cells were subsequently transformed with the expression plasmid pSW-2 by electroporation and selected for gentamicin resistance on LB plates supplemented with 10 µg/ml gentamicin. Individual colonies were grown overnight at 30°C in liquid LB supplemented with 10 µg/ml gentamicin. Two milliliters of the overnight culture were transferred to 10 ml of fresh LB supplemented with 2 mM m-toluate and 10 µg/ml gentamicin and incubated for 7 h at 30°C to allow a second homologous recombination event. Dilutions of bacterial cultures were plated on LB plates without antibiotics and isolated colonies were screened for kanamycin sensitivity. Mutants were identified by specific PCR and sequencing of the respective genomic region with primers *check-iol-F* and *check-iol-R*. A list of all plasmids and oligonucleotides used in this study can be seen in **Table S2** and **S4**.

Root colonisation and competition assays

Experiments with *gfp*-tagged CHA0 WT and *mCherry*-tagged mutant were performed as described previously for the isolates comparisons, with minor differences. Individual inoculations were performed at the same microbial density as described, 10^8 CFU/g soil, while mixed inoculations contained $0.5{\cdot}10^8$ CFU/g of each isolate, in order to obtain a comparable total microbial density as in individual inoculations. At 21 dpi, only root and not rhizosphere samples were collected. Instead of freezing, 10-fold dilutions of the shaken supernatant were plated on KB agar plates supplemented with 10 µg/ml gentamicin and incubated overnight at 28°C. Root microbial density was calculated by normalising CFU counts by root fresh weight.

Growth assays

For planktonic growth, overnight cultures of CHA0 WT and the Δ*iol* mutant were diluted 1/10 in fresh, liquid KB medium and grown 4 h in order to synchronise both genotypes in the exponential phase. The cell suspension was washed thrice by centrifugation at 3,000 x g, decanting of the supernatant, and resuspension in the respective medium. Cells were inoculated individually in flat-bottom 96-well plates at a final OD_{600} equal to 0.01 in either M9 for the assessment of sole carbon sources (26 mM Na_2HPO_4,·$7H_20$, 11 mM KH_2PO_4, 4.3 mM NaCl, 9.3 mM NH_4Cl) or 0.1X Cook's Cytophaga (CC) to study

the effect of inositol addition on growth (0.2% tryptone). For assessment of inositol as a carbon source, the medium was left unsupplemented or supplemented with 100 mM of glucose or inositol. For assaying effects of inositol on growth, medium was left unsupplemented or supplemented with 10 mM glucose or inositol.

Growth curves were obtained by measuring OD_{600} every 30 min in a plate reader (Synergy H1, BioTek) for 72 h (inositol usage as carbon source) or 48 h (inositol effects on growth). Growth curve analyses were done with the *growthcurver* package in R (Sprouffske and Wagner, 2016).

Motility assays and siderophore plates

Swimming and swarming motility were assayed in 0.1X Cook's Cytophaga-agar (CCA) minimal medium at 0.3% or 0.5%, respectively. For all assays, the medium was supplemented with inositol or glucose as a control carbon source at 1 or 10 mM or left unsupplemented. Cells were prepared as for growth experiments. For the swimming assay, sterile pipette tips were dipped in the cell suspension and the agar was gently pierced until reaching the middle of the agar, and without reaching the bottom. Pictures were taken after 48 h of growth at 21 ºC and analyzed with ImageJ (Schindelin et al., 2015). The swimming halo radii and diameters were used to assess swimming motility. For swarming, 1 µl of bacterial suspension was carefully inoculated on the agar surface and avoiding damage of the agar. Pictures were taken at the same time points as for swimming, and swarming ability was measured as the surface area covered by bacteria. In both assays, we inoculated CHA0 WT, CHA0 Δ*iol*, and two previously reported spontaneous CHA0 mutant that exhibit promoted swimming motility as positive control (*sadB*), and impaired swimming as negative control (*fleQ*) (E. Li et al., 2021b). Every treatment consisted of 20 replicate plates and all experiments were repeated at least twice.

Siderophore plates followed the same procedure as swimming plates, using 1X CC medium. In order to ensure that all inoculations were performed at the same distance from each other, they were all done by dipping five pipette tips in a cell suspension with a multichannel pipette and inserting the pipette tips until the center of the agar, as described for swimming inoculations. Every treatment consisted of three replicates and was repeated twice.

Analyses on publicly available data

For the analysis of *in vitro* genome-wide fitness data of root colonisation in WCS417, we used supplementary data from Cole et al. (2017) and selected the genes in the *iol* locus in WCS417 (*iolR*, PS417_11845; *iolC*, PS417_11850; *iolD,* PS417_11855; *iolE,* PS417_11860; *iolB*, PS417_11865; *iolI*, PS417_11870; *iolD*, PS417_11875; *iolG,* PS417_11880; *mocA*, PS417_11885; *iatA,* PS417_11890; *iatC*, PS417_11895; *iatB* PS417_11900). For the analysis of genome-wide fitness data of WCS417 in different conditions, we retrieved the data from Price et al. (2018) available in Fitness Browser (https://fit.genomics.lbl.gov/cgi-bin/

help.cgi). We kept all experimental conditions in which any of the genes in the *iol* locus of WCS417 were affected with a *t*-like score and a fitness value greater than 0.5. For the remaining conditions with more than one repeat, we took the median fitness value. We built a bipartite network of genes and conditions according to the fitness values using the *network* (Butts, 2008) and *ggnet* (Briatte, 2023), R packages for construction and visualisation, respectively. For the network representation, we used the Fruchterman-Reingold algorithm with 1,000 iterations and an area and repulsion radius of 30,000,000. For the analysis of transcriptional activity of CHA0 in the host, we retrieved sequenced data from Vesga et al. (2020) from the NCBI database (BioProject accession PRJNA595077) and mapped the reads to the CHA0 genome (assembly GCF_900560965.1) with kallisto v0.45.0 (Bray et al., 2016). Transcripts-per-million values for each gene were $\log_2$-transformed and scaled per gene.

Data and code availability

The raw *16S rRNA* amplicon sequencing data and genomes in this study have been deposited at NCBI Sequence Read Archive (BioProject: PRJNA960651) and are publicly available as of the date of publication. The R code and data, both the data generated here and the public data used to generate **Figure 7**, have been deposited at Zenodo (https://doi.org/10.5281/zenodo.7855318) and are publicly available as of the date of publication. All software and algorithms used in this study are shown in **Table S5**.

Acknowledgements

The authors want to thank Bas E. Dutilh, Corné M.J. Pieterse and members of the Plant-Microbe Interactions lab for helpful discussions. We acknowledge the Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidised by the University Medical Center Utrecht and The Netherlands X-omics Initiative (NWO project 184.034.019). This study was supported by the NWO Green II Grant no. ALWGR.2017.002 (S.W.M.P., J.J.S.G., R.d.J.), the Novo Nordisk Foundation Grant no. NNF19SA0059362 (R.d.J.), and the China Scholarship Council scholarship no. 201406300090 (H.Z.).

Supplementary Figures

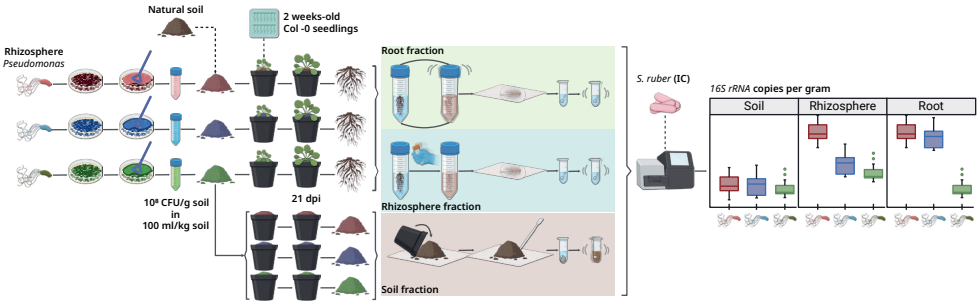


Figure S1. Schematic representation of the experimental design. Root-associated pseudomonads are mixed through natural soil and let to colonise Arabidopsis roots for 21 days. Rhizosphere and root samples are collected and mixed with DNA of *Salinibacter ruber* as internal control. Absolute population density of each pseudomonad is calculated by sequencing the *16S rRNA* gene and normalising the counts with the internal control and the sample weight. Created with BioRender.com.

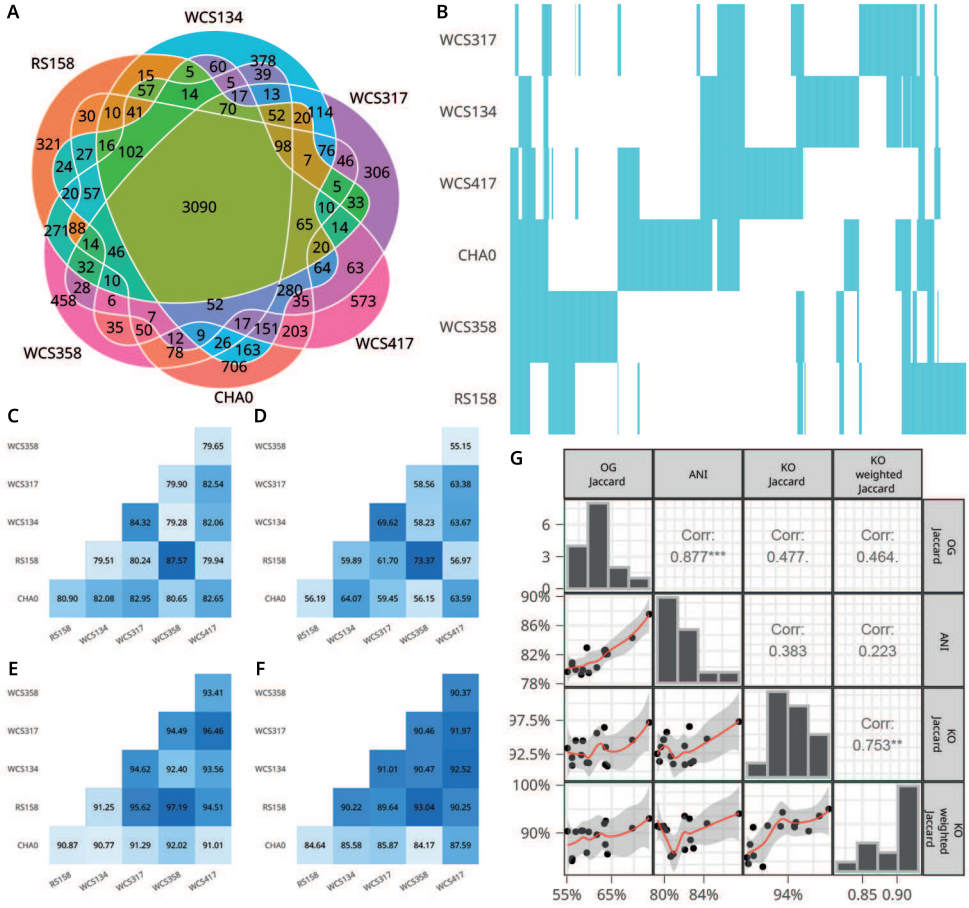


Figure S2. The isolates in this study are a diverse set of root-associated pseudomonads. (A) Venn diagram showing the relationship of orthologue group presence-absence among isolates' genomes. Out of 8,752 orthogroups, 3,090 (35%) are shared by all isolates. (B) Distribution of the non-core genes. (C) Pair-wise average nucleotide identity (ANI) among isolates. (D) Pair-wise unweighted Jaccard similarity of the orthogroup presence/absence per genome (i.e., percentage of shared genes between isolate pairs). (E) Pair-wise unweighted Jaccard similarity of KEGG pathway annotations per genome. The relatively high values suggest that the functional capabilities of the isolates are very similar. (F) Same as E, but showing the weighted Jaccard similarity between the compositions of pathway annotations. The differences in the weighted vs unweighted metric indicate that certain functions encoded in the genomes do not distribute evenly between them, highlighting the functional diversity in the genomes. (G) Pair-wise correlation plots of the previously shown metrics in C-F.

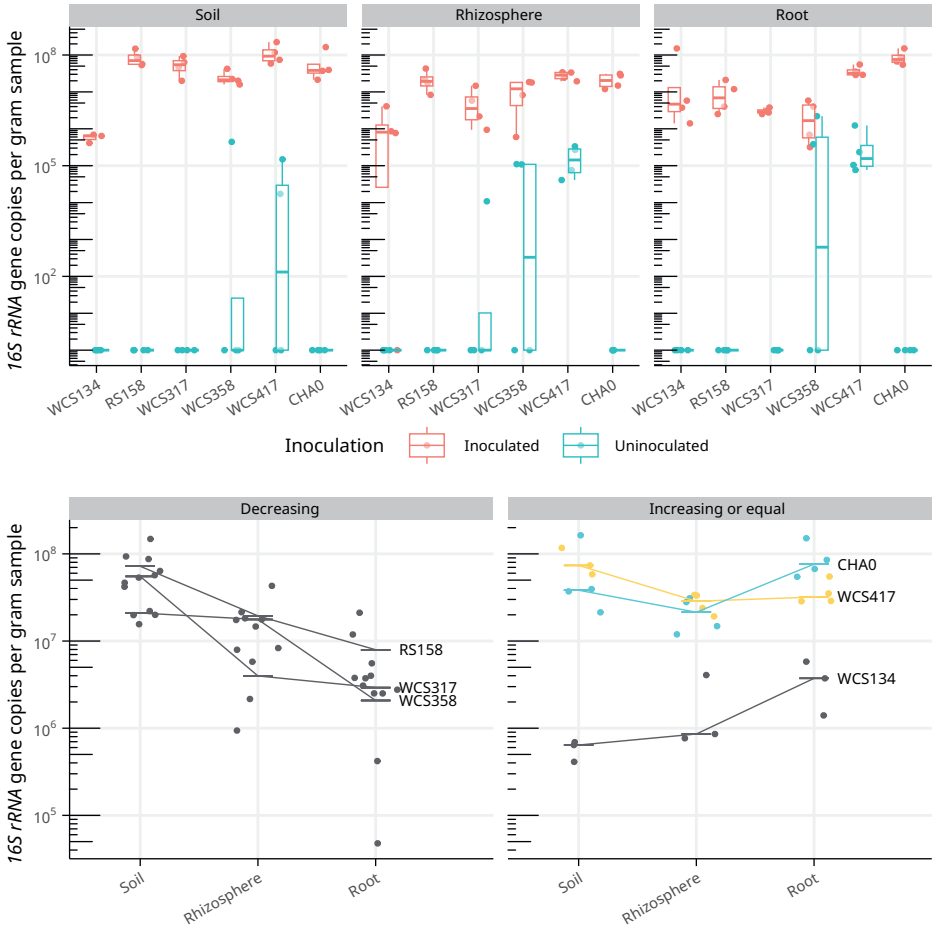


Figure S3. Raw microbial densities and isolate densities across fractions. (A) Absolute inoculum densities at 21 dpi, for both inoculated and uninoculated samples, previous to the subtraction of counts in uninoculated samples as in Figure 1. (B) Absolute inoculum densities after 21 dpi measured as the number of *16S rRNA* copies per gram of sample found in bulk soil (left), rhizosphere (middle), and root samples (right). *P. protegens* CHA0 and *P. simiae* WCS417 are highlighted in cyan and yellow, respectively, to differentiate them from the remaining isolates. The density of some isolates decreases in the root (RS158, WCS317, and WCS358), while others increase or remain equal (WCS417, CHA0, and WCS134).

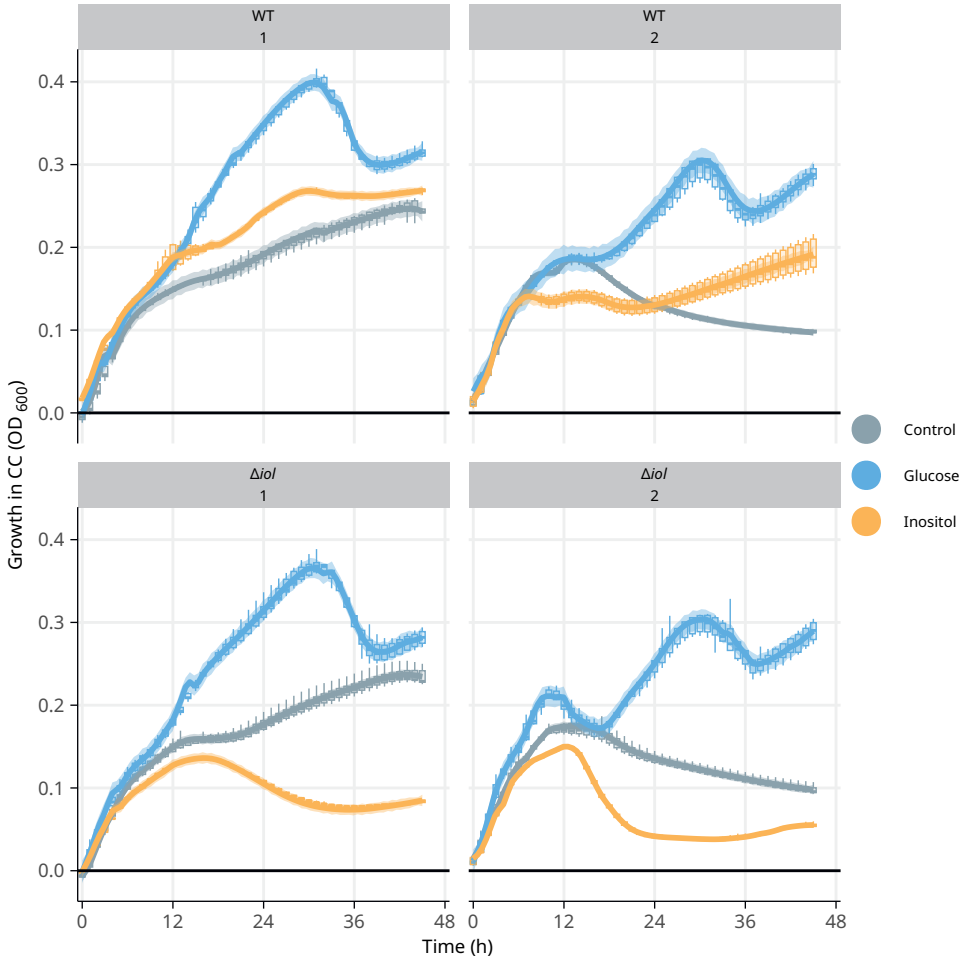


Figure S4. Inositol effects on WT cells are highly variable. Growth curves of WT and Δ*iol* cells in CC medium unsupplemented or supplemented with 10 mM glucose or inositol, measured as OD at 600 nm. The figure shows two examples of independent experiments. The effects of inositol in the WT are clearly variable per experiment, and the benefit of inositol on growth is only apparent in some cases. Every time point consists of measurements for seven technical replicates.

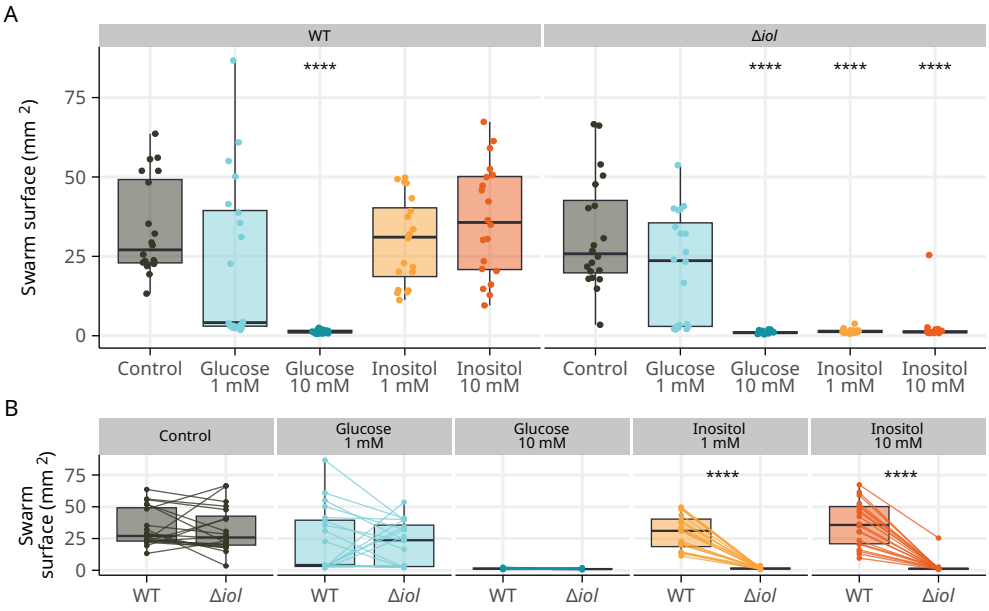


Figure S5. Inositol represses swarming motility in mutant cells. (A) Swarming surface of WT (left) and Δ*iol* cells (right) in CC medium either unsupplemented or supplemented with glucose or inositol at 1 or 10 mM. Inositol does not seem to affect swarming motility in WT cells, while swarming is abolished in the mutant. (B) Same data as in A shown per treatment and pairing the swarming surface for both genotypes per plate, showing the difference in motility.

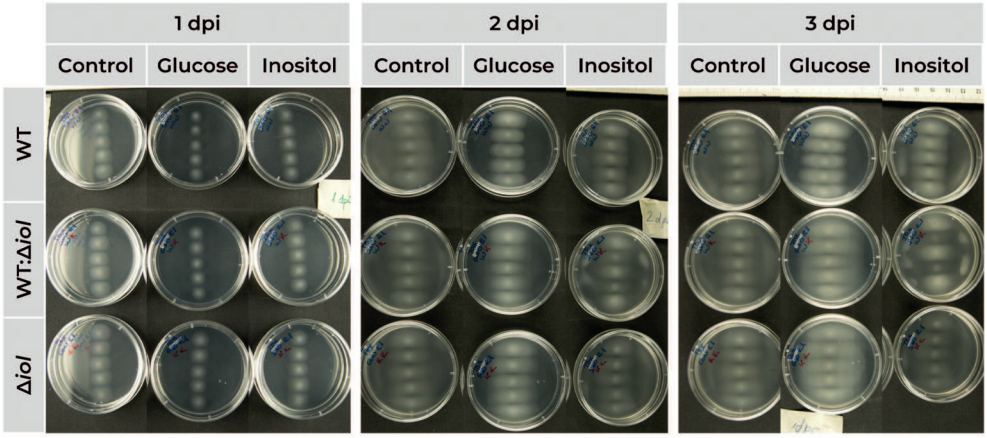


Figure S6. Inositol affects mutant physiology. Qualitative assessment of the effect of inositol and glucose on WT and mutant cells. Images show the same plates as in Figure 6 at 1, 2, and 3 dpi. Per day, the figure shows the plates organised as follows. First column: control conditions; second column: glucose 10 mM; third column: inositol 10 mM; first row: WT cells only; second row: combination of WT and mutant cells (WT, Δ*iol*, WT, Δ*iol*, WT); third row: mutant cells only. Mutant colonies start fading after reaching the end of the plate at approximately the end of the second day.

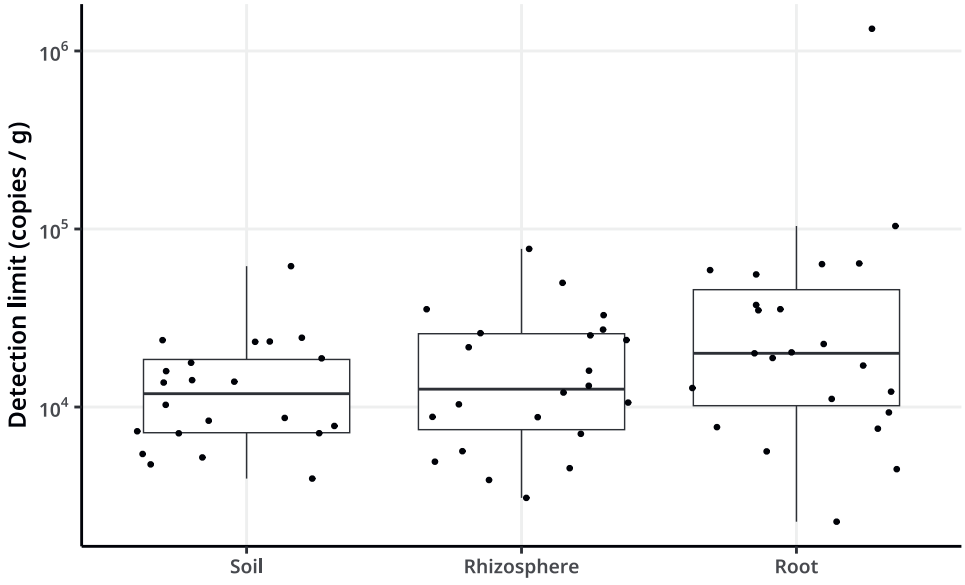


Figure S7. Detection limit of microbial density. Boxplot depicting the detection limit of microbial density in our spike-in method, using the minimum detectable *16S rRNA* gene copies per gram of sample.

Supplementary Tables

Table S1. Isolates used in this study. *NL, Netherlands; SZ, Switzerland. **GP, growth promotion; ISR, induced systemic resistance; BA, biocontrol agent; ND, not determined.

Name	Origin	When	Where*	Effect on host**	References
P. simiae WCS417	Wheat rhizosphere	1983	NL	GP, ISR	Pieterse et al., 2021
P. protegens CHA0	Tobacco rhizosphere	1986	SZ	ISR, BA	Stutz et al., 1986
P. capeferrum WCS358	Potato rhizosphere	1981	NL	ISR, BA	Berendsen et al., 2015; de Boer et al., 1999
P. sp. WCS134	Wheat rhizosphere	1983	NL	BA	de Weger et al, 1986
P. fluorescens RS158	Radish rhizosphere	1987	NL	BA	de Boer et al., 1999
P. sp. WCS317	Potato rhizosphere	1981	NL	ND	de Weger et al., 1986; Geels and Schippers, 1983

Table S2. List of all oligonucleotides used in this study.

Sequence	Source	Identifier
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCGCCTTACCTGTGGCCTACGGGNG-GCWGCAG	This study	50-NGS1-16s-N701
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTAGTACGGAGTGGCCTACGGGNGG-CWGCAG	This study	51-NGS1-16s-N702
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTGCCTTGCGACCTACGGGNGGCW-GCAG	This study	52-NGS1-16s-N703
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCAGGAATGACCTACGGGNGGCW-GCAG	This study	53-NGS1-16s-N704
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGAGTCCCGACCTACGGGNGGCW-GCAG	This study	54-NGS1-16s-N705
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGCCTACGACCTACGGGNGGCW-GCAG	This study	55-NGS1-16s-N706
TCGTCGGCAGCGTCAGATGTGKTATAAGAGACAGGTAGAGAGGTCCTACGGGNGGCW-GCAG	This study	56-NGS1-16s-N707
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTCTCTGGTCCTACGGGNGGCWG-CAG	This study	57-NGS1-16s-N708
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCGTAGCTCCTACGGGNGGCWGCAG	This study	58-NGS1-16s-N709
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGCCTCGTCCTACGGGNGGCWGCAG	This study	59-NGS1-16s-N710
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCCTCTTCCTACGGGNGGCWGCAG	This study	60-NGS1-16s-N711
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTCTACCCTACGGGNGGCWGCAG	This study	61-NGS1-16s-N712
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGATCGCCACTTCTGACTACHVGGG-TATCTAATCC	This study	62-NGS1-16s-N501
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCTCTATTTCTCTGACTACHVGG-GTATCTAATCC	This study	63-NGS1-16s-N502
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATCCTCTACTCAGACTACHVGGG-TATCTAATCC	This study	64-NGS1-16s-N503
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGTAGAGATAGACTACHVGGG-TATCTAATCC	This study	65-NGS1-16s-N504
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAAGGAGCTA GACTACHVGGG-TATCTAATCC	This study	66-NGS1-16s-N505
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTGCATATC GACTACHVGGG-TATCTAATCC	This study	67-NGS1-16s-N506
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAGGAGTAA GACTACHVGGG-TATCTAATCC	This study	68-NGS1-16s-N507
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTAAGCCT GACTACHVGGGTATCTA-ATCC	This study	69-NGS1-16s-N508
cgGAATTCctgctcaaggacctgaaa	This study	del-iol-1
cggGGTACCttcgacgactagcccttc	This study	del-iol-2
cggGGTACCgcctatgacgtcgagctg	This study	del-iol-3
cgGGATCCcgccactcggttctgaat	This study	del-iol-4
cttgagcgatgtgcagat	This study	check-iol-F
ttctttgtgcagaagccg	This study	check-iol-R

Table S3. Sequences for the V3-V4 regions used as priors in the DADA2 pipeline.

Name	Sequence
Salinibacter-V3V4	TGGGGAATCTTGCACAATGGGGTCACCCCTGATGCAGCCATGCCGCGTGGAG-GAAGACACCCCTATGGGGCGTAAACTCCTTTTCTGAATGAAGAAACCCCTG-TAGCTTCAGGGCGCGACGGTAGTTCAGGAATAAGCACCGGCTAACTCCGTGC-CAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTGTCCGGAATCACTGGGTGTA-AAGGGTGTGCAGGCGGGGCAGCAAGTCGGATGTGAAACCCCATGGCTTAACCAT-GGAGGTGCATTCGAAACTGTTGCTCTTGAGTCCCGGAGAGGCTGTCGGAATTCGT-GGTGTAGCGGTGAAATGCGTAGATATCACGAGGAACACCAGAGGCGAAAGCGGA-CAGCTGGACGGGTACTGACGCTCAGGCACGAAAGCGTGGGGAGCAAACA
Mitochondrion-V3V4	TGGGGAATCTTGGACAATGGGCGAAAGCCCGATCCAGCAATATCGCGTGAGT-GAAGAAAGGCAATGCCGCTTGTAAAGCTCTTTCGTCGAGTGCGCGATCATGACAG-GACTCGAGGAAGAAGCCCCGGCTAACTCCGTGCCAGCAGCCGCGGTAAAACG-GGGGGGGCAAGTGTTCTTCGGAATGACTGGGCGTAAAGGGCACGTAGGCGGT-GAATCGGGTTGAAAGTGAAAGTCGCCAAAAAGTGGCGGAATGCTTTCGAAAC-CAATTCACTTGAGTGAGACAGAGGAGAGTGGAATTTCGTGTGGAGGGGTGAAATC-TACAGATCTACGAAGGAACGCCAAAAGCGAAGGCAGCTCTCTGGGTCCCTAC-CGACGCTGGGGTGCGAAAGCATGGGGAGCGAACG
Chloroplast-V3V4	TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATGCCGCGTGGAG-GTAGAAGGCCTACGGGTCCTGAACTTCTTTTCCCAGAGAAGAAGCAATGACGG-TATCTGGGGAATAAGCATCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAG-GATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCTGTAGGTGGCTTTTTA-AGTCCGCCGTCAAATCCCAGGGCTCAACCCTGGACAGGCGGTGGAAACTAC-CAAGCTTGAGTACGGTAGGGGCAGAGGGAATTTCCGGTGGAGCGGTGAAATGCG-TAGAGATCGGAAAGAACACCAACGGCGAAAGCACTCTGCTGGGCCGACACTGA-CACTGAGAGACGAAAGCTAGGGGAGCGAATG
RS158-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTA-AGTTAATACCTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCT-GTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGG-GCGTAAAGCGCGCGTAGGTGGTTCGTTAAGTTGGATGTGAAAGCCCCGGGCT-CAACCTGGGAACTGCATCCAAAACTGGCGAGCTAGAGTACGGTAGAGGGTGGT-GGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGC-GAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAG-CAAACA
WCS134-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGTTGTAGAT-TAATACTCTGCAATTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGC-CAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA-AAGCGCGCGTAGGTGGTTCGTTAAGTTGGATGTGAAATCCCCGGGCTCAACCTGG-GAACTGCATTCAAAACTGTCGAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCT-GTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGAC-CACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACA
WCS317-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGTTGTAGAT-TAATACTCTGCAATTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGC-CAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA-AAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAAGCCCCGGGCTCAACCTGG-GAACTGCATTCAAAACTGACAAGCTAGAGTATGGTAGAGGGTGGTGGAATTTCCT-GTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGAC-CACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACA

WCS358-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGTAAGC-GAATACCTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTGC-CAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA-AAGCGCGCGTAGGTGGTTTGTTAAGTTGAATGTGAAAGCCCCGGGCTCAACCTGG-GAACTGCATCCAAAACTGGCAAGCTAGAGTACGGTAGAGGGTGGTGGAATTTCCT-GTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGAC-CACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAGCAAACA
WCS417-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGT-TACCTAATACGTGATTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCT-GTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGG-GCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAATCCCCGGGCT-CAACCTGGGAACTGCATTCAAAACTGACTGACTAGAGTATGGTAGAGGGTGGT-GGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGC-GAAGGCGACCACCTGGACTAATACTGACACTGAGGTGCGAAAGCGTGGGGAG-CAAACA
CHA0-V3V4	TGGGGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGTGT-GAAGAAGGTCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAGGGCAGT-TACCTAATACGTGATTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCT-GTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGG-GCGTAAAGCGCGCGTAGGTGGTTTGTTAAGTTGGATGTGAAAGCCCCGGGCT-CAACCTGGGAACTGCATCCAAAACTGGCAAGCTAGAGTATGGTAGAGGGTGGT-GGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGC-GAAGGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGGAG-CAAACA

Table S4. Plasmids used in this study.

Name	Source	Identifier
pSEVA212S; *ori*R6K, *lacZα* with two flanking I-SceI sites; Km^R, Ap^R	Martínez-García and de Lorenzo, 2011	pEMG
*ori*RK2, *xylS*, P_m::*I-sce*I; Gm^R	Martínez-García and de Lorenzo, 2011	pSW-2
pME11149; suicide plasmid for the in-frame deletion of the *iol* cluster (PPRCHA0_2627 to PPRCHA0_2633)	This study	pEMG::Δ*iol*

Table S5. List of all software and algorithms used in this study.

Software	Source	Version
R	R Core Team	4.2.2
ggplot2	Wickham, 2016	3.4.1
ggpubr	Kassambara, 2020	0.5.0
ggVennDiagram	Gao, 2022	1.2.2
KEGGREST	Tenenbraum, 2022	1.38.0
dada2	Callahan et al., 2016	1.26.0
multcomp	Hothorn et al., 2008	1.4.22
dplyr	Wickham et al., 2022	1.0.10
MASS	Venables and Ripley, 2002	7.3.57
emmeans	Length, 2023	1.8.4
magrittr	Bache and Wickham, 2022	2.0.3
ggtext	Wilke and Wiernik, 2022	0.1.2
patchwork	Pedersen, 2022	1.1.2
ggsci	Xiao, 2018	2.9
ggnet	Briatte, 2023	0.1.0
network	Butts, 2008	1.13.0.1
growthcurver	Sprouffske, 2016	0.3.1
Hmisc	Harrell, 2022	4.7-2
pheatmap	Kolde, 2019	1.0.12
scales	Wickham and Seidel, 2022	1.2.1
extrafont	Chang, 2023	0.19
assertthat	Wickham, 2019	0.2.1
Biostrings	Pagès et al., 2023	2.66.0
PerformanceAnalytics	Peterson and Carl, 2020	2.0.4
stringr	Wickham and Rstudio, 2022	1.5.0
magick	Ooms, 2021	2.7.3
readxl	Wickham and Bryan, 2022	1.4.1
ggrepel	Slowikowski, 2022	0.9.2
tidyr	Wickham et al., 2023	1.3.0
rstatix	Kassambara, 2022	0.7.1
reshape2	Wickham, 2007	4.2.2
OrthoFinder	Emms and Kelly, 2019	2.5.4
Prokka	Seeman, 2014	1.14.6

fastANI	Jain et al., 2018	1.33
eggNOG-mapper	Huerta-Cepas et al., 2019	2.1.8
cblaster	Gilchrist et al., 2021	1.3.17
mashtree	Katz et al., 2019	1.2.0
ImageJ	Schindelin et al., 2015	1.53
kallisto	Bray et al., 2016	0.45.0
cutadapt	Martin, 2011	2.8

Juan J. Sánchez-Gil[1] and Ronnie de Jonge[1,2]

[1] Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

[2] AI Technology For Life, Department of Information and Computing Sciences, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

Author contributions:

Juan J. Sánchez-Gil: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualisation, Writing – Original Draft. **Ronnie de Jonge**: Conceptualisation, Supervision, Project administration, Funding acquisition, Validation, Writing – Review and Editing.

Preprint: Sánchez-Gil, J.J. and de Jonge, R., 2025. **ASVNet: Inferring microbes from *16S rRNA* amplicon sequencing data**. bioRxiv. DOI: 10.1101/2025.11.21.689717.

CHAPTER

FOUR

ASVNet: Inferring microbes from *16S rRNA* amplicon sequencing data

Abstract

Amplicon sequencing of the variable regions of the *16S rRNA* gene is one of the most common techniques in microbiome research. A crucial step in analysing this data is to cluster raw sequences into biologically interpretable groups that approximate the actual organisms in a community, which reduces data complexity, facilitates diversity analyses, and enables clearer ecological and functional interpretations. Most approaches rely on fixed sequence similarity thresholds, typically operational taxonomic units (OTUs) at 97% or amplicon sequence variants (ASVs) at 99-100%. However, finding the correct correspondence between amplicons and the original organisms in a complex community remains challenging, primarily due to the presence of varying copy numbers within genomes and the high similarity of amplicon sequences among species. While the stringent threshold in ASVs prevents merging sequences from different species, it often splits sequences from the same genome into multiple ASVs, inflating species counts. Conversely, the 97% similarity threshold in OTUs tends to frequently collapse multiple distinct organisms into a single unit regardless of their functional or ecological (dis)similarity, and thereby underestimating true diversity. To address these limitations, we introduce ASVNet, an algorithmic framework that leverages amplicon sequence similarity and co-abundance networks to better reflect the underlying organisms in a community. ASVNet builds on the principle that amplicon counts from the same organism co-vary across samples and it thus attempts to cluster these co-occurring, similar sequences as empirical OTUs (eOTUs). Comparative analyses in synthetic and natural root-associated microbial communities with genome-sequenced isolates show that ASVNet-derived eOTUs more accurately represent the original microbes and possess greater biological relevance than OTUs defined at any fixed threshold. By providing a biologically grounded framework to amplicon sequence data interpretation, ASVNet extends our ability to extract meaningful insights from complex microbiome datasets.

Main

Challenges in amplicon sequencing for microbial community analysis

In prokaryotes, the genes responsible for the production of ribosomal RNA (rRNA) are commonly contained within the *rrn* operon, although it is frequent to find individual genes located elsewhere in the genome (Brewer et al., 2020). Due to the high amounts of rRNA required for cell survival and reproduction, the *rrn* operon is found in multiple copies in 92% of known bacteria, where it generally ranges from 1 to 37 copies per genome (Pan et al., 2023; Větrovský and Baldrian, 2013). The *rrn* copy number varies according to taxonomy and determines key (eco)physiological traits, such as maximum growth rate, overall metabolic capacities, and ability to respond to fluctuating and complex environments (Klappenbach et al., 2000; Lin et al., 2023; Roller et al., 2016). Nevertheless, this essential role of *rrn* genes in bacterial metabolism imposes strong evolutionary restrictions, which translates in high conservation of both its sequence and structure (Pei et al., 2010), a property that has enabled their extensive use as a phylogenetic marker for decades (Goossens et al., 2023; Johnson et al., 2019; Winker and Woese, 1991).

Particularly, the *rrnB* gene, also commonly referred to as *16S rRNA* (hereafter, *16S* gene), encodes the 1,5 kbp long *16S rRNA* that conforms the nucleotidic portion of the small ribosomal subunit. Despite having a highly conserved sequence, the *16S* gene harbours nine hypervariable regions (V1-V9) that exhibit taxon-specific variability throughout prokaryotic genomes. Consequently, amplicon sequencing of these hypervariable regions is the most common technique for microbial identification in microbiome research, although sequencing of the entire *16S* gene is becoming increasingly more popular as it enables enhanced taxonomic resolution.

In the archetypal microbiome study, the desired region of the *16S* gene is PCR-amplified with primers located at the conserved flanking regions. Subsequently, the amplicons are sequenced and the read counts for every sequence are recorded and processed to obtain a representation of the microbial community. However, the amplification and sequencing steps are prone to errors and the raw data accrues many spurious reads that emerge from technical imprecisions. Consequently, algorithmic efforts in microbiome data processing focus on removal and/or correction of these erroneous sequences, a procedure known as denoising (Callahan et al., 2016).

The complex nature of the *16S* gene introduces significant biological and technical challenges when analysing amplicon microbiome data. As mentioned above, bacterial genomes often carry multiple copies of the *16S* gene, causing an overrepresentation of organisms with higher copy number (Větrovský and Baldrian, 2013). Moreover, about 60% of prokaryotic genomes harbour multiple copies of the gene, which can differ in their sequence

(intragenomic heterogeneity) (Bodilis et al., 2012; Coenye and Vandamme, 2003; Pan et al., 2023; Pei et al., 2010). Furthermore, identical or highly similar copies are commonly shared across different species or genera (intergenomic conservation), resulting in copies being more similar across genomes than within a single genome (Pan et al., 2023). Consequently, some amplicon sequences represent multiple organisms simultaneously, while others

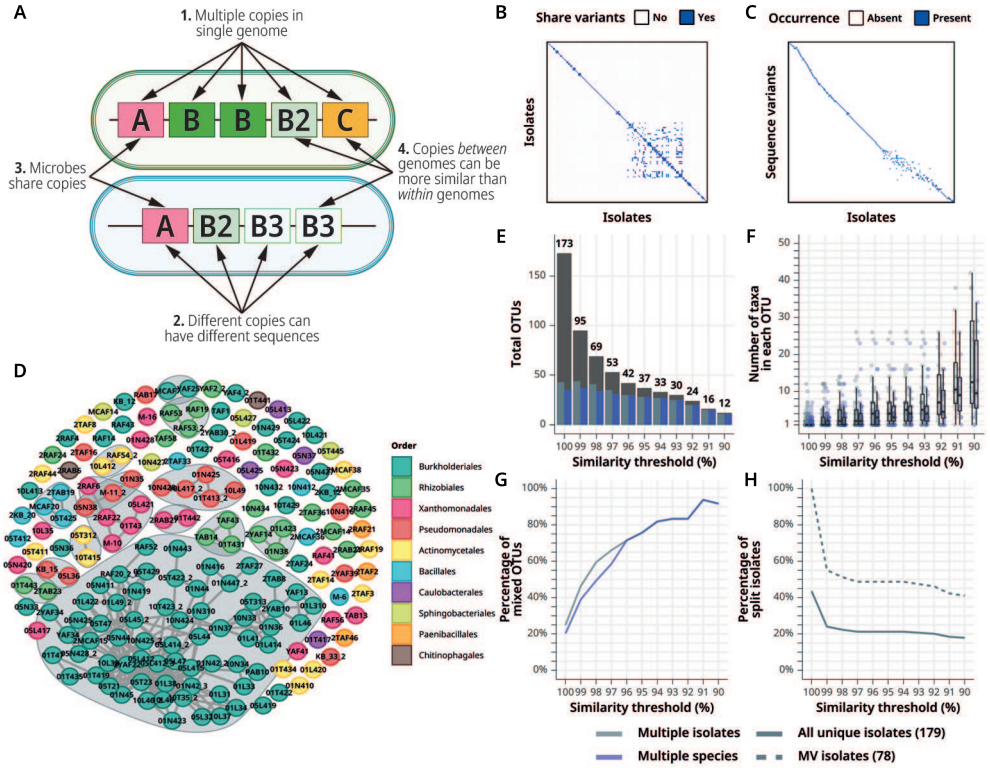


Figure 1. Single similarity thresholds alone cannot correctly cluster microbes. (A) Sources of complexity in *16S* gene amplicon sequencing datasets. All these phenomena increase the complexity in the analysis of this type of data, making it impossible to classify sequences in a biologically meaningful way based only on a fixed threshold. (B) Incidence matrix of the 134 isolates included in the SynCom and all *16S* gene sequence variants in their genomes. Cells in blue denote presence of a variant in an isolate genome. Note how the relationship variant-isolate is relatively simple in many isolates, but very complex in a group of them located at the bottom-right of the matrix. (C) Adjacency matrix of the isolates, depicting pair-wise co-occurrence of variants in their genomes. Blue cells denote events in which the same sequence variant occurs in both isolates in the comparison. The intricate set of interactions, seen towards the bottom-right of the matrix, indicate complex relationships between these isolates in terms of the variants that they share. (D) Network visualisation of the adjacency matrix in C, showing the isolates connected when they share V3-V4 sequences. The shaded areas enclose groups of isolates that share sequences between them but not with other isolates outside. (E) Total number of *k*OTUs found by clustering sequences from SynCom isolates at decreasing similarity thresholds. The blue-grey and blue bars show the number of those *k*OTUs that are chimeric at the isolate and species level, respectively. (F) Number of chimeric *k*OTUs containing either different isolates (blue-grey) and species (blue). (G) Percentage of chimeric *k*OTUs at the same similarity thresholds, depicting OTUs merging multiple isolates (blue-grey) and species (blue). (H) Percentage of isolates that are split into different *k*OTUs at different similarity thresholds, shown as a fraction of all isolates (full line) and of MV isolates (dashed line).

remain organism-specific, leading to skewed microbial abundance estimates. Additionally, the average *16S* copy number and the degree of intra- and intergenomic sequence similarity vary across taxa, further distorting the final community representation (**Figure 1A**; Schloss, 2021; Větrovský and Baldrian, 2013).

To date, numerous approaches have been developed to process and analyse *16S* amplicon data (Prodan et al., 2020). The two predominant strategies differ primarily in the similarity thresholds used for clustering. Amplicon sequence variants (ASVs) aim for high sequence resolution (99-100% similarity) and try to group identical sequences after denoising. While this prevents merging sequences from distinct species, ASVs often split multiple variants of the *16S* gene from the same genome into separate clusters (Schloss, 2021), artificially elevating diversity estimates. Consequently, ASV-based analyses tend to overestimate taxa with many distinct copies and underestimate those with fewer or shared copies (Pan et al., 2023). Alternatively, operational taxonomic units (OTUs) group sequences at a fixed similarity threshold (typically at 97% similarity), designed to encapsulate species-level variation (Konstantinidis and Tiedje, 2005). In contrast to ASVs, OTUs underestimate diversity by merging distinct organisms and species into chimeric clusters, regardless of their functional similarity or evolutionary relationships (Nguyen et al., 2016; Pan et al., 2023; Schloss, 2021). However, at the same time a threshold as low as 94.75% is required to merge all seven *16S* copies in *Escherichia coli* genomes into a single OTU (Schloss, 2021), highlighting the inherent limitations of fixed sequence similarity thresholds for cluster definition.

In summary, no single fixed similarity threshold fully captures the complex relationship between individual *16S* copies and their source organisms. Consequently, fixed thresholds negatively impact statistical analyses and biological interpretations by artificially inflating or underestimating diversity, splitting variants from single organisms, or merging distinct organisms into inaccurate clusters. We reasoned that these problems would be alleviated by correctly clustering amplicon sequences into the original organisms from which they arise. With this goal in mind, we first examined the effect of fixed thresholds on clustering accuracy in a controlled dataset with known genome-sequence relationships. Subsequently, we introduce a data-driven approach aimed at more accurately clustering amplicon sequences to reflect their true source organisms.

Clusters based solely on sequence similarity do not resemble original genomes in a synthetic community

In order to study the relationship between genomes and their corresponding *16S* gene sequences using fixed thresholds, we utilise amplicon sequencing data from an experiment that uses a synthetic community (SynCom) of 134 microbes, all of which were isolated from plant roots grown in natural soil and with sequenced genomes (Selten et al., 2024a). This dataset originates from a plant root colonisation experiment in which

the SynCom members are exposed to the roots of different plant hosts and varying nutritional conditions. First, we assessed the number of different V3-V4 *16S* variable region sequences associated with these isolates. Among them, 65 isolates contain more than one V3-V4 sequence and were therefore named multiple-variant (MV) isolates. The remaining isolates are considered single-variant (SV) isolates. Overall, approximately 63% of genomes share at least one *16S* variant with any other. Many of these variants appear in few genomes and are shared by small and isolated groups of microbes, although a considerable number of them show a fuzzy distribution across genomes (**Figure 1B**), which translates into a large group of isolates sharing variants in a complex pattern (**Figure 1C, D**). To further study the impact of different clustering cut-offs on the identity of the resulting clusters, we calculated all pair-wise sequence similarities and clustered all sequences into OTUs from 100% to 90% similarity (**Figure 1E**). For this, we used the *k*-mer-based sequence similarity (*k* = 5) instead of classical alignments, as the former are faster to compute and represent local alignments accurately (**Supplementary Figure 1**). We named the resulting clusters *k*-mer-based OTUs, or *k*OTUs. As expected, the total number of *k*OTUs decreased with lower thresholds, and the number of chimeric *k*OTUs that merged multiple isolates, and even multiple species, increased (**Figure 1F, G**). At 100%, around 25% of all *k*OTUs already contains more than one single isolate and 20% of them contain sequences from more than one species (**Figure 1F**). The precise number of isolates and species merged into one single *k*OTU ranges from 1-10 for 100% similarity to 1-42 at 90% (**Figure 1G**). At the same time, around 60% of MV isolates are split into multiple *k*OTUs at 99%, and this is still 50% at 97% similarity (**Figure 1H**). In conclusion, and in agreement with previous studies, we did not find any "one-size-fits-all" threshold that could meaningfully cluster amplicons into units resembling the original genomes.

Correlation in sequence abundances reflects latent intragenomic co-occurrence

Given the above-mentioned limitations, it seems highly advantageous to incorporate additional biologically-relevant information that guides the clustering process to accurately assemble genome-like clusters. In metagenomics, for example, commonly used algorithms leverage sequence co-abundances to facilitate the assignment of contigs to potential genomes (Alneberg et al., 2014; Kang et al., 2019) These strategies have the potential to enhance clustering decisions and align the results more closely with biological reality. For co-abundance to be useful at grouping variants into genome-like structures, correlations between similar variants should reflect their co-occurrence within genomes (intragenomic co-occurrence). That is, co-abundance should exhibit a measurable relationship with intragenomic co-occurrence (**Figure 2**). In the SynCom data, all variants show at least some degree of co-abundance with any other sequence, and groups of correlated variants emerge visibly (**Figure 2A**). Among these correlations, many might arise from biological interactions between the different under-

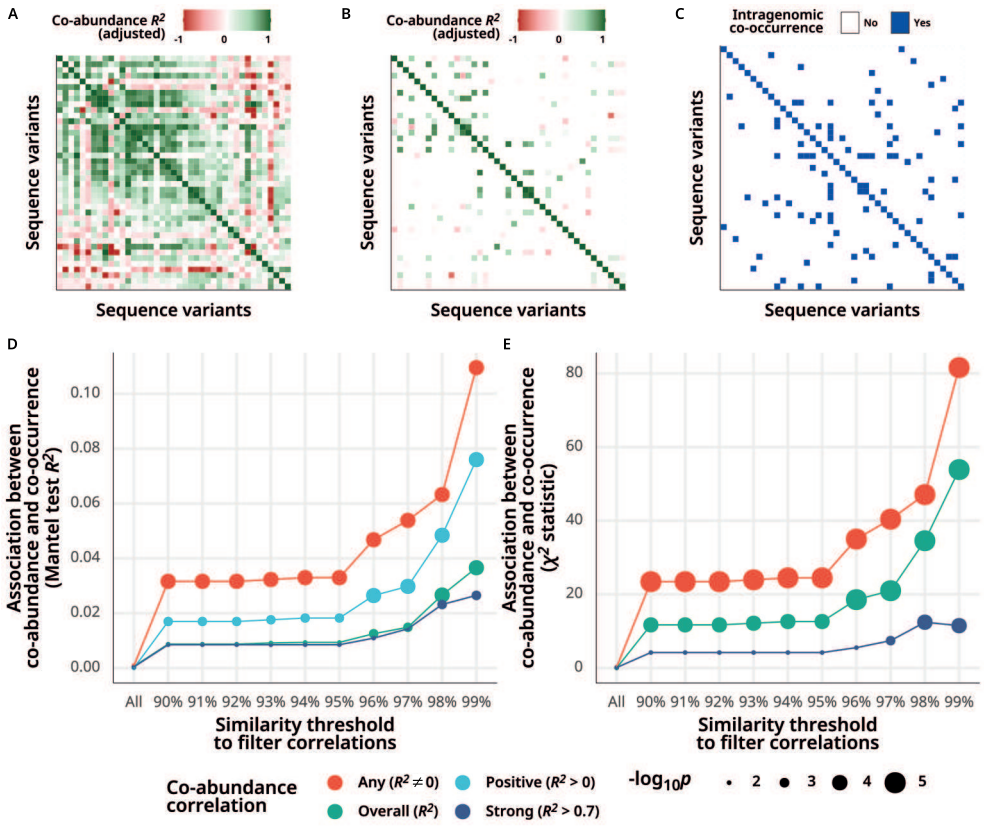


Figure 2 Using sequence similarity and co-abundance between variants can help enhance clustering. (A) Pearson co-abundance correlations (R^2) of all unique sequences in the isolates. Rows and columns are ordered using Isomap seriation (Hahsler et al., 2025) applied to the co-abundance distance matrix. This procedure embeds the pairwise distances into a low-dimensional manifold and finds a 1-dimensional ordering that places similar variants adjacent to each other. (B) Same co-abundance matrix as in A, but only showing correlations for pairs of sequences at least 90% similar according to their *k*-mer similarity. Note how most of correlations in A came from distant variants, yet the matrix still depicts groups of somewhat similar sequences with strong correlations in the data, which could be used to cluster sequences into similarly-behaving groups of variants. (C) Intragenomic co-occurrence of all sequences in at least one genome, in the same seriation order as in A and B. (D) Line plot showing the result of a Mantel test comparing pair-wise abundance correlations in the SynCom data with their intragenomic co-occurrence pattern. The graph depicts the R^2 of the tests performed at increasingly higher thresholds. As the similarity between sequences increases, the test shows a higher association with their co-occurrence. (E) Line plot following the same strategy as in D in a χ^2 test. Similar to the Mantel approach, the χ^2 tests also reveal an increasing association between correlation status (correlated vs non-correlated) and the co-occurrence of variants within a genome as more similar sequences are considered. Any: all non-zero correlations are considered; overall: all correlation values are considered, unfiltered matrix; positive: positive correlations only; strong: only positive correlations above 0.7.

lying organisms and, consequently, confound the statistical association to physical co-occurrence within individual genomes. Therefore, we decided to discard all correlations between very distant sequences and to only consider co-abundances between sequences at least 90% similar (**Figure 2B**). Then, to gradually reduce the effect of confounding correlations, we

applied increasingly higher similarity thresholds to examine the relationship between co-abundances and intragenomic co-occurrence patterns (**Figure 2C**) using Mantel tests with 9,999 permutations. The analysis revealed a significant association between co-abundance and co-occurrence among variants, and the strength of this association increased as the sequences considered were more similar (**Figure 2D**), potentially due to the increasing removal of confounding comparisons. Interestingly, when considering both positive and negative correlations (any non-zero correlation), the association of co-abundance with co-occurrence was around 1.4 to 1.8-fold higher than when only positive correlations are considered. This suggests that negative correlations from similar sequences also carry information about their co-occurrence in genomes. These negative correlations might represent cases of competitive exclusion between similar strains, in which shared variants co-occur with other strain-specific variants in the genomes of competing microbes, rendering the strain-specific ones negatively correlated. In a similar manner, we used a χ^2-test to study whether correlation status of variants – e.g., variants correlated vs not correlated – was related to their co-occurrence status – e.g., variants co-occur vs do not co-occur – (**Figure 2E**). Consistently, we observed a significant enrichment of correlated variants when they also co-occur, and an emphasised statistical signal – i.e., χ^2-statistic – as similarity between variants increased. Notably, both the Mantel and χ^2 analyses show a clear transition around the 95% similarity threshold. Below this value, many of the remaining sequence pairs still represent comparisons across different genomes, where correlations are largely shaped by ecological interactions between strains rather than by physical linkage. These ecology-driven correlations dilute the association with true intragenomic co-occurrence. Around 95%, most of these non-informative comparisons have been filtered out, and the remaining pairs are increasingly dominated by variants from the same genome or closely related strains. This shift leads to a sudden strengthening of the association between co-abundance and intragenomic co-occurrence, producing the upward inflection observed at this threshold.

In conclusion, correlations of abundance across samples indeed carry useful information about the occurrence of those sequences within the genomes, and therefore, abundances can be leveraged to guide the clustering of *16S* gene sequences into biologically meaningful clusters.

ASVNet: empirical OTUs from sequence similarity and co-abundance networks

We applied the principle described above into a novel algorithmic approach, which we named ASVNet – after amplicon sequence variants networks –, aimed at reconstructing the correspondence between amplicon sequences and the organisms in a microbial community. ASVNet operates on the premises that different *16S* variants from the same genome will (1) be somewhat similar, and (2) strongly correlate across all samples. ASVNet follows these

principles to integrate both sequence similarity and co-abundance information to infer empirical operational taxonomic units (eOTUs; **Supplementary Figure 2**). The concept of eOTUs and calling clusters based on both taxonomy and co-abundances was already followed by Probst and co-authors in the development of the PhyloChip technology (Probst et al., 2014). After hybridisation of environmental DNA with the probes in a PhyloChip, eOTUs were inferred by finding correlated groups of taxonomically-related probes according to the fluorescence score across samples. Both for ASVNet and for PhyloChip eOTUs, the term 'empirical' signifies that the resulting eOTUs are *ad hoc* and depend on the variation induced by or present in the experimental settings. Consequently, ASVNet requires sample-to-sample variation and variability induced by experimental conditions to increase the probability that any two individual microbes will exhibit a differential behaviour across samples, and therefore facilitate the clustering variants from an individual organism into a singular eOTU.

The algorithm starts by building a global sequence similarity network (SSN) of all variants, which is then partitioned into groups of similar sequences (SSN clusters). Subsequently, for each SSN cluster, ASVNet computes a sequence correlation network (SCN) based on co-abundances. Lastly, eOTU inference occurs in an iterative process that involves clustering each SCN into groups of sequences that are consistently correlated across samples. During the construction of the SCN, weak and noisy correlations are handled via either of two distinct modes. In the first and default mode, *empirical,* potentially spurious correlations are discarded by applying an empirically-determined heuristic threshold. This threshold is obtained by randomly sub-sampling pairwise correlations and selecting a specific percentile from the resulting distribution. The second mode, *regularised*, uses a regularised approach – elastic net regression – aimed at removing as many noisy and irrelevant correlations as possible while leaving the most meaningful ones, therefore enhancing resolution and inducing sparsity – the number of correlations that are turned into zeroes, and thus discarded. This mode requires setting a parameter, α – i.e., penalty mixture –, which controls the stringency of the method. When α=0, weak correlations are shrunk towards (but not made equal to) zero, whereas they are totally turned into zeroes and discarded when α=1.

ASVNet eOTUs preserve isolate integrity better than *k*OTUs

To test ASVNet and search for the optimal configuration, we continued using the SynCom dataset as a ground truth, as the variability across nutritional conditions and plant hosts makes these data ideal to assess the quality of the algorithm. We searched for the best initial SSN threshold (*k*) and SCN construction mode, and we benchmarked the results from ASVNet against *k*OTUs at 97% and 99% similarity thresholds (**Figure 3**). We ran ASVNet with *k* = 90%, 92.5%, and 95%, in both *empirical* and *regularised* modes. For *regularised* runs, we tested multiple decreasing levels of stringency while still maintaining sparsity in the data, and therefore selected *α* values

of 1, 0.9, 0.8, and 0.5. First, we explored the total number of clusters found by each method. As expected, *k*OTUs at 99% show the highest number of clusters (**Figure 3A**). Interestingly, ASVNet with an SSN threshold of $k = 90\%$ reported a total number of clusters that is very similar to the actual number of organisms in the SynCom, while applying higher SSN thresholds – i.e., $k = 92.5\%$ and $k = 95\%$ – increased the number of clusters approximately 2-fold with respect to the previous cut-off, but no dramatic differences are observed between different ASVNet modes. This shows that the selected SSN threshold strongly affects the total number of final clusters, with low values enabling abundance comparisons between very distant variants, and potentially resulting in more chimeric eOTUs.

Next, we compared the different methods for their ability to effectively cluster sequence variants from the SynCom isolates. We labelled clusters as correct if (1) it pointed to only one isolate and (2) all the observed sequences from that isolate were within that same cluster. Among all variant-isolate configurations, particularly resolving MV isolates poses a significant challenge, as the different variants should end up in the same eOTU. These cases are crucial for assessing whether the algorithm could effectively group sequences into groups resembling the original genomes. Notably, although *k*OTUs at 99% contained more correct clusters, those were mostly coming from SV isolates, meaning that, expectably, *k*OTUs at 99% splits better into individual sequences. However, the highest number of correctly-grouped MV isolates is achieved by ASVNet after SSN cutoffs $k = 92.5\%$ and 95% (**Figure 3B**), accounting for approximately 40% and 22% of all correct clusters, respectively (**Figure 3C**). In comparison, *k*OTUs had a much lower frequency of correct MV isolates, i.e., 9% for *k*OTUs at 97% and 10% for *k*OTUs at 99%. In general, the *regularised* mode did not perform better than the *empirical* counterpart, and thus we removed this mode for subsequent analyses.

Beyond perfect clusters, we also interrogated ASVNet and *k*OTUs for their tendency to generate chimeric clusters (**Figure 3D**; number of isolates within a cluster) and to split sequences from MV isolates into different clusters (**Figure 3E**; number of clusters per MV isolate). In general, chimeric ASVNet eOTUs agglomerate fewer isolates than *k*OTUs, and there are fewer cases of isolates split into multiple eOTUs than with *k*OTUs at 99%. Notably, chimeric *k*OTUs are based solely on similarity among sequences, while chimeric eOTUs may arise from isolates that behave similarly across samples, and therefore are a result of ecological similarity. To test whether chimeric eOTUs merge isolates in a more biologically meaningful manner, we evaluated the functional similarity among the organisms within chimeric OTUs. Specifically, we predicted the set of KEGG pathways for all organisms in each chimeric OTU and computed the Jaccard index for each pair-wise comparison. After this, we defined the median functional similarity of the chimeric cluster as the median of Jaccard index values for all comparisons. Both eOTUs and *k*OTUs at 97% show a median functional similarity within clusters of more than 90%, contrasting less than 85% in the case of chimeric *k*OTUs at 99% (**Figure 3F**). This suggest that chimeric eOTUs, although technically not pointing at

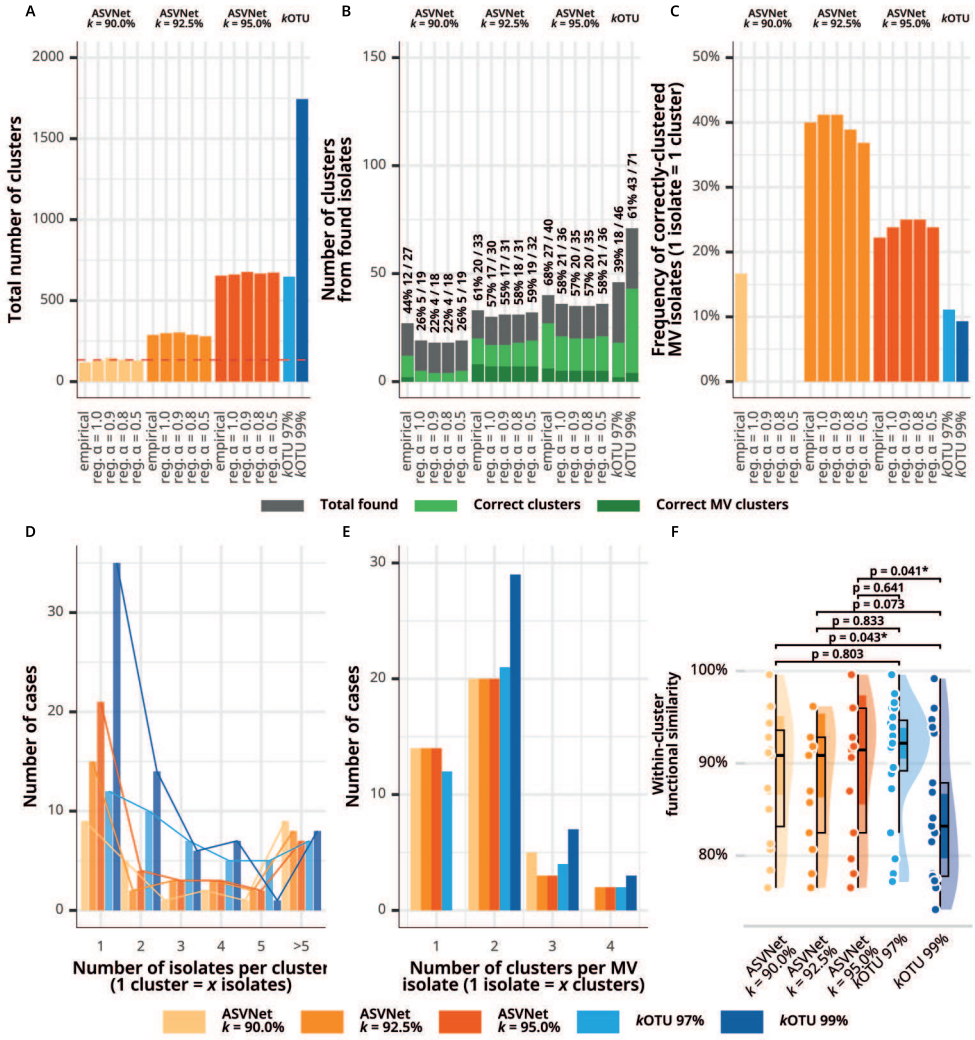


Figure 3. ASVNet resolves the SynCom dataset into more meaningful eOTUs. (A) Total number of clusters found by ASVNet and *k*OTUs at 97% and 99% sequence similarity. The dashed line marks the number of isolates in the experiment. (B) Number of clusters containing *rrn* sequences from the isolates in the microbial collection. The overlayed clear green bar depicts the total number of correctly-called isolates (both SV and MV isolates), and the dark green bar the total number of correctly-called MV isolates. Note that ASVNet finds the highest number of correct MV isolates. (C) Percentage of correctly-called MV isolates versus all correct clusters, showing that ASVNet after a cut-off of 92.5% has the highest ratio of correctly-called MV isolates. (D) Histogram of the number of isolates merged per cluster as a measure of the merging tendency of the algorithms. Note that ASVNet merges on average fewer isolates in every cluster. (E) Histogram of the number of clusters with sequences from MV isolates as a measure of the splitting tendency of the algorithms. ASVNet clusters MV isolates mostly as one or two eOTUs, while *k*OTU methods tend to split them more often into different *k*OTUs. (F) Median within-cluster functional similarity for incorrectly-called clusters, showing that clusters that do not correspond directly to an original *rrn* configuration are still biologically meaningful and have a higher within-cluster functional similarity than *k*OTUs at 99%. The darker shade on top of each boxplot shows the 95% confidence interval around the median value (notches). *Emp.*, empirical; *reg.*, regularised

single organisms, indeed merge isolates in a biologically more meaningful manner than *k*OTUs at 99% similarity.

ASVNet eOTUs better resemble original isolates in complex datasets

We observed that ASVNet can successfully cluster sequence variants from MV SynCom isolates, and consequently displays enhanced performance compared to *k*OTUs across various SSN thresholds. However, despite the high complexity of this particular SynCom, this dataset is still relatively simple compared with real-world communities, and derives from a controlled environment with few organisms and a small number of different variants. Yet, the main objective of ASVNet is to be applied to natural community microbiome datasets. Such datasets typically exhibit high diversity and encompass numerous closely-related organisms, in which sequence variants co-occur in a more intricate pattern, and where many variants show significant sequence similarity to others. One of such complex environments is the plant rhizosphere, a dynamic environment formed by plant roots and the narrow region of surrounding soil under their direct influence (Poppeliers et al., 2023). We used two separate rhizosphere microbiome datasets to test ASVNet performance. The '*rhizosphere gradient experiment*' data contains 96 samples consisting of microbial communities in different compartments of the natural rhizosphere, achieved through increasingly stronger washes aimed to enrich for the microbes on the plant roots (Poppeliers et al., 2024). The '*root time series experiment*' data also contains 96 samples, all of them from the same rhizosphere compartment, but in this case the data represents a longitudinal study that follows the development of the microbial community in time (unpublished data). Both experiments were performed with natural soil extracted from the same location, from which a microbial collection had been previously established and all its genomes made available (Berendsen et al., 2018; Selten et al., 2024a, 2024b; Stringlis et al., 2018). Furthermore, this collection captured approximately 53% of all original microbes found in the *Arabidopsis* rhizosphere (Selten et al., 2024a), which makes this microbial collection a valuable benchmarking tool to assess the comparison of eOTUs with *k*OTUs. Following the same principles as with the SynCom dataset, we called eOTUs with ASVNet in *empirical* mode after applying the same three SSN thresholds and clustered all sequences into *k*OTUs at 97% and 99% for comparison (**Figure 4**). In contrast with the simpler SynCom data set, ASVNet reported many more eOTUs than *k*OTU methods in both rhizosphere sets for all SSN thresholds (**Figure 4A**). This suggests that ASVNet is affected by sample complexity, potentially due to a difficulty to resolve the clustering of the SCN with an increasing number of low-abundant microbes with few observations. Indeed, eOTUs tend to contain fewer sequences and with overall less abundance than *k*OTUs (**Supplementary Figure 3**), as previously observed with the SynCom data (**Figure 3D**), which may indicate that eOTUs more closely reflect the expected biological structure.

For both datasets, we assessed the number of clusters containing *16S* gene sequence variants from the expected microbes. ASVNet eOTUs comprised the highest number of correctly classified isolates, approximately 70% of eOTUs, compared to 34% and 55% for *k*OTUs at 97% and 99%, respectively (**Figure 4B**). Importantly, eOTUs also contained the highest number of correctly classified MV microbes compared to *k*OTUs. These results show that integrating sequence similarity and co-abundance information in the clustering process helps construct eOTUs that better resemble the organisms in the ecosystem under study.

Altogether, our results show that co-abundance across samples can be used to solve the problem of clustering *16S* gene amplicon sequences into biologically-meaningful units. Moving forward, we aim to further refine ASVNet to address some of the important sources of complexity mentioned above, such as the assignment of sequences to more than one eOTU and the calculation of eOTU-adjusted abundances. So far, recent advances are achieving this when reference genomes are available (Selten et al., 2025), yet reconstructing genomes in reference-free settings remains a major challenge. In this sense, ASVNet could further benefit from incorporating advanced fuzzy clustering strategies and deep learning techniques to optimise the correspondence between variant co-abundance and the most-probable intragenomic co-occurrence structure. In addition, recent frameworks such as SyFi (Selten et al., 2025), together with the large body of publicly available microbial genomes and raw sequencing data, offer the possibility to build genome-informed, microbiome-independent eOTU references. These genome-derived eOTUs could serve as external priors to guide ASVNet in reconstructing organisms, particularly in settings where no sample-specific

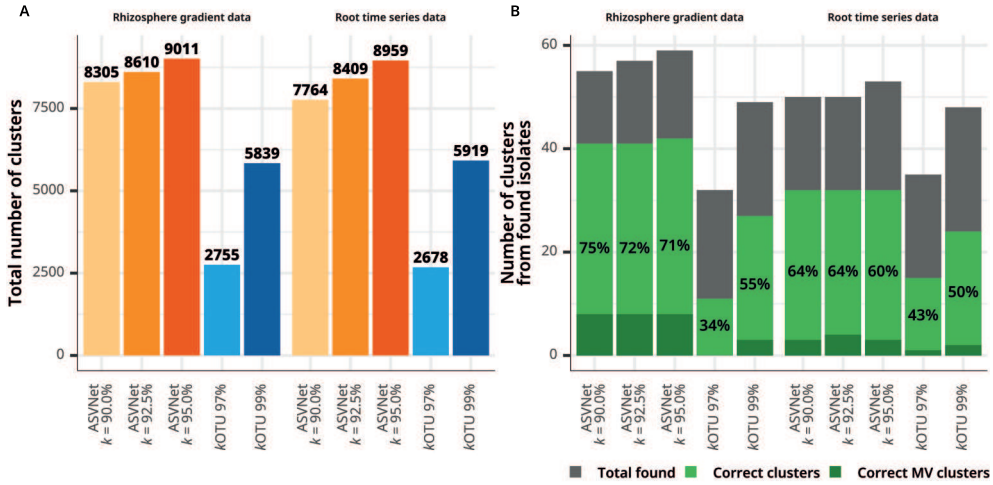


Figure 4. ASVNet finds more expected microbes in complex datasets. (A) Total number of clusters reported by ASVNet and *k*OTUs at 97% and 99% similarity for both studies in this analysis. (B) Number of clusters containing *16S* gene sequences from the isolates in the microbial collection constructed from the same soil that was used in both studies. Note that the number of correct clusters, both for SV and MV isolates, is higher for ASVNet at any initial SSN threshold used.

reference genomes are available. Ultimately, ASVNet lays the groundwork for future methodologies that aim to unveil the true composition of microbial communities through amplicon sequencing. By doing so, these future approaches have the potential to unlock deeper biological insights currently masked by multiple layers of complexity and embedded in both past and future amplicon sequence endeavours.

Methods

The ASVNet pipeline

Phase I: SSN clusters from original sequences

The input for ASVNet can be either raw FASTQ files from *16S* gene amplicon sequencing experiments or a sequence table with counts – e.g., as result of denoisers like DADA2 (Callahan et al., 2016). The first step in ASVNet is to cluster the sequences by sequence similarity. This avoids intensive calculations of all pair-wise correlations when sequences are not expected to come from the same genome. Pair-wise sequence comparisons are performed using the *k*-mer similarity (by default, $k = 5$), as it allows fast calculations. This similarity is based on the proportions of *k*-mers shared by each pair of sequences (Zielezinski et al., 2017) and aligns well with (local) alignment-based similarity (**Supplementary Figure 1**). After sequence comparisons, a low and permissive threshold is applied. The thresholds used in this study are based on the required threshold reported to collapse all sequences from *E. coli* into a single OTU, i.e., 94.75% (Schloss, 2021). Based on our results, the default value is set at 92.5%. After applying the threshold, the resulting SSN matrix is partitioned in clusters of broadly similar sequences with MCL v14.137 (Dongen, 2000), which uses Markov clustering for graph-based clustering. The MCL algorithm is used in various commonly used bioinformatics tools for protein homology clustering (to find orthologues) and to cluster genes from co-expression data. The resulting SSN clusters are subsequently input into the second phase of ASVNet.

Phase II: finding eOTUs

Due to technical imprecisions during amplification and sequencing, amplicon datasets accumulate many erroneous sequences that derive from the original true sequences. It is therefore expected that the initial correlations do not reflect biological reality in the samples. To account for this, eOTUs are inferred in each SSN cluster in a recursive process that starts by finding SCN clusters. The abundances for all sequences in a SCN cluster are merged into a single cluster and the process is repeated until no new strong correlations emerge or all sequences are merged (**Figure 5**). In a more detailed manner, the eOTU finding pipeline follows the next main steps:

Step 1. Compute pair-wise linear correlations. The initial counts are transformed with the centred log ratio (CLR) before calculating correlations as it has been previously suggested (Gloor et al., 2017). Both Pearson and Spearman correlations can be calculated, although Spearman is the default option.

Step 2. Filter correlations. Low correlations are removed using a rectifying function based on a logistic-adjusted rectified linear unit (ReLU). Briefly, the logistic-ReLU is constructed from bootstrapped correlations of the original matrix. These random correlations are used to calculate the correlation p at the αth percentile (by default, 99th percentile). After this, a ReLU function with an inflection point at p is modified by a logistic-like curve that slowly moves into a slope of 1. This logistic-ReLU can then be used to turn all correlation values that can be observed by chance to zeroes, whilst it leaves the remaining correlation values unaffected. The construction of the logistic-ReLU is done only once at the beginning of phase II but it is used throughout the pipeline in every iteration of the correlations to filter low values. The result is a sparser matrix that represents the SCN.

Step 3. Find and evaluate clusters. Clusters in the SCN are identified with MCL. In this step, ASVNet can be configured to find increasingly more fragmented clusters, and will select the most compact one with the lowest sum of internal negative correlations, assuming that eOTUs that better represent genomes will contain the fewest number of such internal negative correlations and best compactness.

Step 4. Merge sequences in same cluster. When the optimal clustering is found, the abundances inside each cluster are summed and a matrix with a reduced number of columns is returned. If there are still multiple clusters, the process is repeated from step 1 and will stop when no new correlations are found after filtering or all sequences are merged into a single cluster. When stopped, the final eOTUs are saved and ASVNet continues with the next SSN cluster.

In the *regularised* mode, ASVNet follows a slightly stricter path in which the linear correlations in step 1 are replaced by regularised regression (elastic net regression) with either an user-specified or a globally-optimised regularisation penalty parameter λ, and a mixing parameter α (by default 1 for lasso regression). Elastic net regression continues iteratively until no more clusters are found, after which ASVNet continues the pipeline in the *empirical* mode using the previous clusters as initial clusters.

Processing reads and SSN construction

The amplicon sequencing samples from all three studies included here – i.e., '*SynCom experiment*', '*rhizosphere gradient experiment*', and '*root time series experiment*' – were generated by amplifying the V3-V4 hypervariable region of the *rrn* gene (unpublished data). Briefly, raw reads were quality-filtered, adapter-trimmed, and merged using bbmerge from the BBMap v38.96 suite (Bushnell et al., 2017). Merged reads were dereplicated using

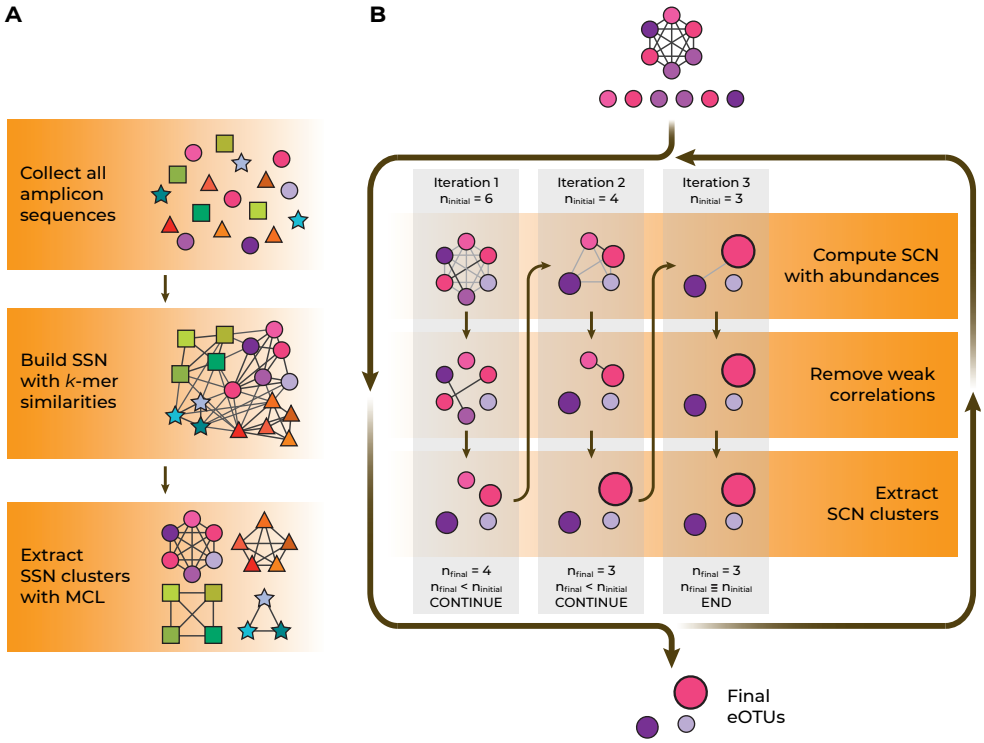


Figure 5. The ASVNet pipeline to infer eOTUs. (A) Phase I of the ASVNet pipeline. ASVNet collects amplicon sequences and builds a SSN using all pair-wise *k*-mer similarities, after which it partitions it into SCN clusters with the MCL algorithm. (B) Phase II of the pipeline. Per cluster of similar sequences, ASVNet builds the SCN using the correlations based on sequence abundances, filters low correlations, and finds clusters with MCL. ASVNet continues repeating this process with the clusters until no new correlations are found and eventually reports the resulting eOTUs.

the *derepFastq()* function from the DADA2 R package v1.21.0 (Callahan et al., 2016), which is transformed into a matrix of counts of unique reads. For each dataset, sequences with sufficient value variation (at least five different values) and prevalence (found in at least five samples), and samples with more than 1,000 counts, were kept for further processing. The SSN from the resulting sequences is constructed by calculating all pair-wise *k*-mer similarities with a parallelised version of the *kdistance()* function from the *kmer* R package (Wilkinson, 2018) and included in ASVNet as *kDistance()*. The result is the SSN matrix, which is further filtered with the different minimum similarity *k* thresholds, after which the SSN is partitioned into the SSN clusters with the MCL algorithm wrapper function *ssnMCL()* in ASVNet.

Code availability

ASVNet is available on a GitHub repository and can be accessed from https://github.com/jjsanchezgil/ASVNet. The code for figures in this manuscript is available at https://github.com/jjsanchezgil/ASVNet_figures_scripts.

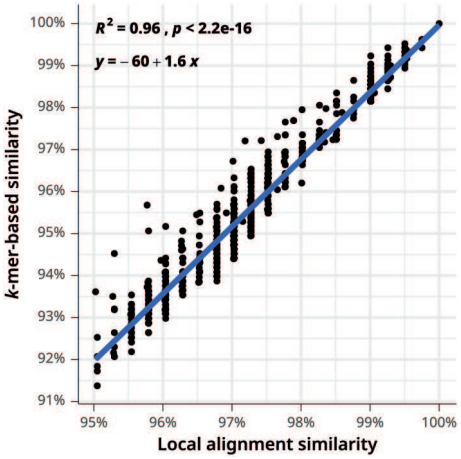
Acknowledgements

The authors want to thank Sanne Poppeliers, Gijs Selten, Bas E. Dutilh, Corné M.J. Pieterse, and rest of members of the Plant-Microbe Interactions lab for helpful discussions. This study was supported by the NWO Green II Grant no. ALWGR.2017.002 (S.W.M.P., J.J.S.G., R.d.J.) and the Novo Nordisk Foundation Grant no. NNF19SA0059362 (R.d.J.).

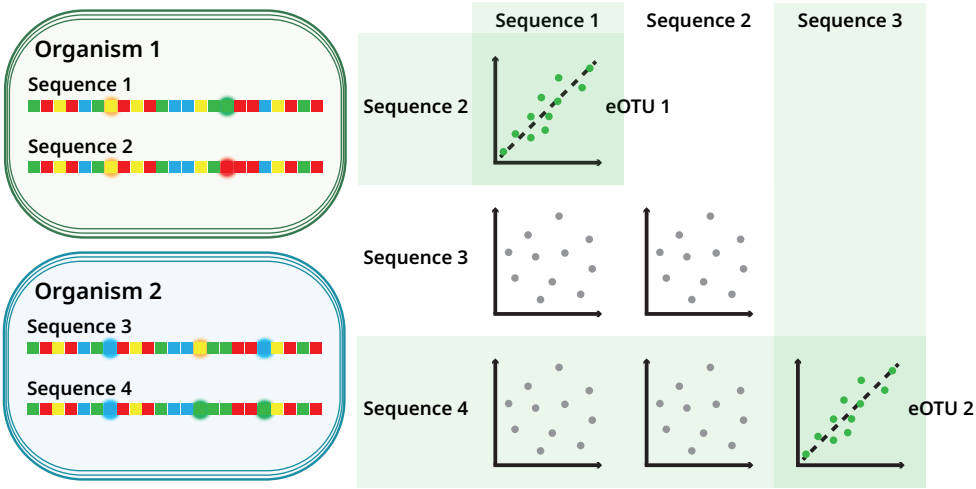
Competing interests

This project received partial financial support from the private company Novozymes A/S, Denmark.

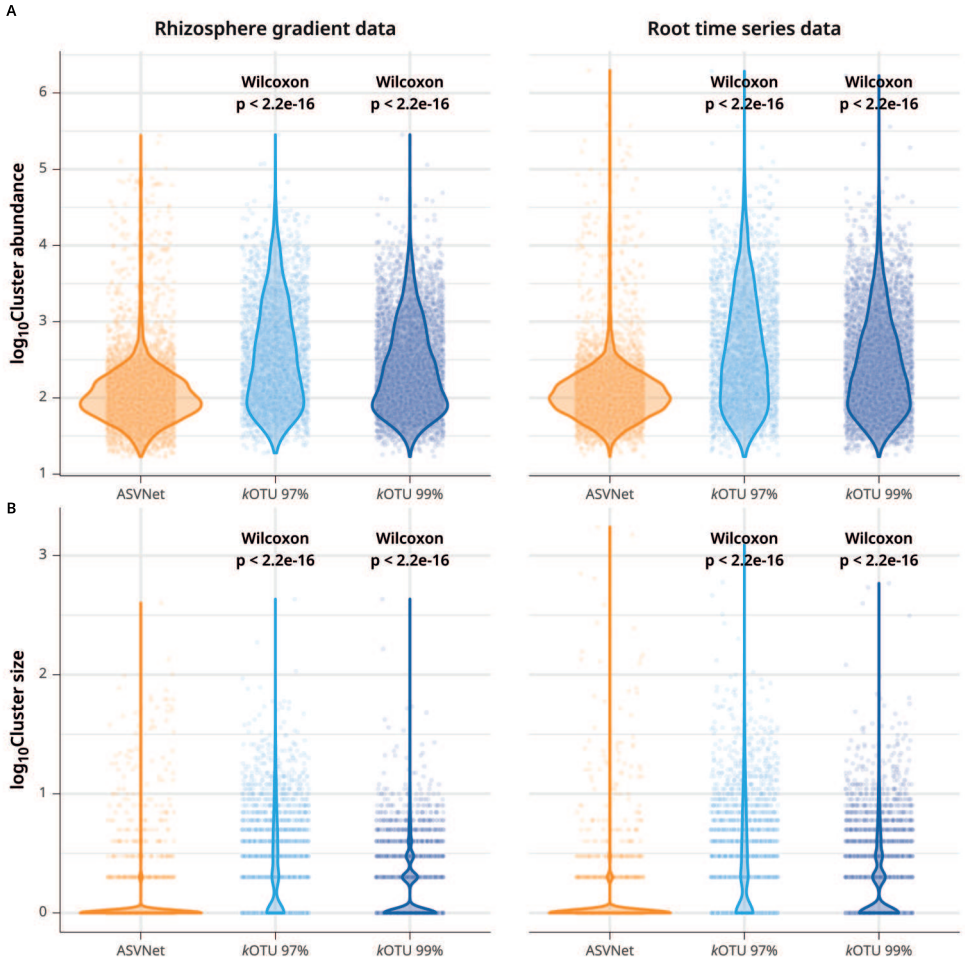
Supplementary Figures



Supplementary Figure 1. *K*-mer based similarity is very similar to local alignments. Scatter plot depicting the relationship between *k*-mer based (*k* = 5) and local alignment similarities of sequences in the SynCom dataset. The percent identity for local alignments was calculated as the proportion of identical positions divided by the length of the shorter sequence. The plot shows all pair-wise comparisons with an alignment similarity above 95%, extracted from a larger sample of 5,000 bootstrapped sequence pairs. Pearson's R^2 for the correlation is shown in the graph, together with its *p*-value and the equation for the linear regression. The high R^2 and the low *p*-value denote a strong correlation between both similarity metrics.



Supplementary Figure 2. Conceptual overview of the rationale behind ASVNet. The left panel shows two example organisms, each containing two distinct *16S rRNA* gene copies (Sequences 1–4), illustrating intragenomic heterogeneity. Sequences from the same genome differ slightly but are more similar to each other than to sequences from other organisms. The right panel presents a toy correlogram of pairwise correlations in sequence abundances across samples. Variants from the same organism (Sequences 1 and 2 from Organism 1) show strong co-abundance correlations and are grouped into the same empirical OTU (eOTU 1). In contrast, sequences from different organisms (e.g., Sequence 3 and 4 from Organism 2) are less correlated and fall outside this eOTU. ASVNet integrates both sequence similarity and co-abundance patterns to group amplicon variants into eOTUs, aiming to reconstruct the underlying organismal composition of a microbial community.



Supplementary Figure 3. ASVNet eOTUs are less abundant and smaller clusters. Violin plots showing the log distribution of cluster abundances (A) and sizes (B) for ASVNet and *k*OTUs for the '*rhizosphere gradient*' data (left panels) and the '*root time series data*' (right panels). In every panel, the result of a Wilcoxon test comparing ASVNet to *k*OTU distributions is shown on top of each *k*OTU violin. For both datasets, ASVNet eOTUs tend to be less abundant and group fewer sequences than *k*OTUs at both similarity thresholds, as indicated by Wilcoxon tests.

Juan J. Sánchez-Gil[1]

[1] Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, 3584 CH, Utrecht, The Netherlands

Author contributions:

Juan J. Sánchez-Gil: Conceptualisation, Investigation, Writing – Original Draft.

CHAPTER FIVE

General discussion

Microbial inoculants hold immense promise for sustainable agriculture, offering sustainable alternatives to chemical fertilisers and pesticides by harnessing beneficial plant-microbe interactions. Yet, despite decades of research and countless laboratory successes, translation to field settings remains inconsistent. In this final chapter, I weave together all the chapters of this thesis to develop a unified conceptual framework for why and how microbial colonisation and function vary so widely in real soils, and how methodological advances can sharpen our ability to anticipate and guide these outcomes.

While many studies have focused on community-level patterns or broad taxonomic shifts, improving bioinoculant efficacy ultimately requires an understanding of the precise physiological state of the inoculant itself. These strain-level dynamics are what dictates whether beneficial traits are expressed and whether the inoculant can function as intended. Without directly measuring microbial behaviour *in situ* at the individual level, we risk overlooking key failure points in colonisation, activity, or compatibility. Thus, identifying how the physiology of a single strain varies across microenvironments is essential for resolving the causes of inconsistent field performance and advancing toward predictive, mechanism-based approaches in microbiome-informed agriculture.

In this thesis, we set out to elucidate the multifaceted determinants of microbial engraftment, function, and inference in plant-associated microbiomes, with an eye toward improving the reliability of bioinoculants in sustainable agriculture. Across three empirical and computational studies, we have addressed: (1) the transcriptional plasticity of a model bioinoculant in natural soil (**Chapter 2**); (2) genetic determinants of enhanced rhizosphere competence in *Pseudomonas* via the *iol* locus (**Chapter 3**); (3) algorithmic advances to overcome *16S*-based clustering limitations (**Chapter 4**); and (4) their synthesis in the broader ecological and methodological framework that shapes microbial invasion outcomes. Here, I integrate these findings to develop a coherent understanding of how microbe, host, and method converge to determine both the ecological reality and our ability to observe it.

Soil as the context of plant–microbe interactions

Interactions between plants and microbes do not emerge in a vacuum. These are scaffolded by the particular biotic and abiotic properties of the surrounding soil, which define the initial conditions for colonisation and the primary selective pressures that microbes must navigate.

Soil is, in fact, the ecological stage of plant–microbe interactions and one of the most complex ecosystems on Earth. As the foundational component of terrestrial life, it simultaneously supports plant growth and harbours an extraordinary diversity of microorganisms. At the scales relevant for microbes, soil is a dynamic matrix of microaggregates and pores. Aggregates are clusters of minerals and organic carbon that function as hotspots of microbial abundance, decomposition, nutrient cycling, and cross-kingdom interactions (Vos et al., 2013; Wilpiszeski et al., 2019). Among the aggregates, pores act as fragmented aqueous habitats where the thickness of water films regulate microbial motility, restrict diffusion of solutes and offer refuge from predation and competition (Bickel and Or, 2023; Tecon and Or, 2016). Within this architecture, minerals, organic matter, water, gases, and biota interact to create a constantly shifting patchwork of microenvironments (Bach et al., 2018; Vos et al., 2013; Wilpiszeski et al., 2019). Steep gradients of pH, redox potential, oxygen, and nutrients structure these niches (Wilpiszeski et al., 2019), while their accessibility changes with soil moisture and physical connectivity (Bickel and Or, 2023).

Moisture fluctuations, seasonal temperature shifts, and root growth continually reshape its chemical and biological architecture (Bardgett and van der Putten, 2014). Plant roots themselves act as powerful ecosystem engineers, releasing exudates that alter local nutrient availability, oxygen status, and microbial assembly (Philippot et al., 2013). These rhizosphere-driven changes occur together with intrinsic soil processes, such as organic matter turnover, mineral weathering, and biotic activity, to continuously modify the structure and chemistry of the soil. The result is the generation of countless temporally fluctuating microenvironments where microbial habitats are in constant flux and colonisation opportunities are transient and highly localised.

These microscale heterogeneities ensure that even neighbouring microbes experience markedly different physicochemical conditions, underscoring the ecological individuality of soil microhabitats. Consequently, pores and aggregates function as evolutionary incubators (Rillig et al., 2017), fostering phenotypic diversity and generating a vast genetic pool from which adaptation and innovation emerge (Vos et al., 2013). As a result, soils represent one of the largest reservoirs of microbial diversity known, harbouring billions of cells and tens of thousands of taxa per gram (Anthony et al., 2023; Bahram et al., 2018). This extraordinary richness fosters a dense ecological network of interactions that regulates both stability and invasibility, imposing strong biotic resistance to newcomers (Mallon et al., 2015). Indeed, as **Chapter 2**

showed, *Pseudomonas simiae* WCS417 (hereafter, WCS417) in bulk soil upregulates genes linked to stress adaptation and microbial antagonism, highlighting how even competent root colonisers must first endure the competitive intensity of soil before establishing beneficial partnerships with plants.

In sum, soil functions simultaneously as a dynamic *reservoir* and *filter* for the eventual configuration of the root-associated microbiome, and therefore for plant–microbe interactions. Its structural heterogeneity generates microhabitats with divergent selective pressures, its microbial richness enforces community-level resistance, and its temporal dynamism reshuffles ecological opportunities. In fact, comparative studies across soil types have repeatedly shown that soil characteristics and microbiota exert a much stronger influence on root-associated microbial composition than plant genotype or developmental stage (Durán et al., 2022; Liu et al., 2019; C. Yang et al., 2022) and that bioinoculant efficiency depends on soil type and nutritional content (da Costa et al., 2020). This observation highlights the role of soil as the main microbial bank, as the root-associated microbiome can almost exclusively be assembled from the pool of organisms available in the surrounding soil. Thus, while the host plant may exert an important layer of control over its associated microbiome, it does so within the constraints imposed by the soil environment, which co-determines compatibility, colonisation, and mutualism. Consequently, any attempt to predict or engineer bioinoculant success must account for the ecological complexity of soil as the dominant factor shaping root-associated microbial communities.

Host control in plant-microbe interactions

Within the constraints of the soil and its microbial residents, the plant imposes another strong layer of selectivity. Hosts exert active control over their associated microbiomes, shaping microbial communities through immune regulation, metabolic filtering, and structural compartmentalisation. Rather than merely excluding pathogens, effective host control maintains a constant underlying pressure that forces the community to adapt to the host and other residential microbes. This phenomenon eventually fosters plant-microbe compatibility by aligning microbial behaviour with host needs, enabling the emergence of cooperation and mutualism (Sharp and Foster, 2022).

A key component of host control over the microbiome is constitutive immune surveillance, in which plants continuously perceive microbial-associated molecular patterns (MAMPs) from resident microbes in the environment. This basal immune activity constrains microbial proliferation while allowing stable associations to form, thereby contributing to microbial homeostasis (Zhang, 2023). Recent work has proposed that such continuous, low-level immune signalling may operate as a distinct mode of immune regulation, tentatively termed the latent defence response (LDR), which differs

from damage- or danger-induced immunity in that it is maintained in the absence of disease symptoms or tissue damage (Zhang, 2023).

Recent studies show that the strength and outcome of this background immune surveillance varies across host genotypes. In *Arabidopsis thaliana*, natural accessions differ in microbiome responsiveness due to variation in immune receptors such as TNLs (Rensburg et al., 2025), and the receptor-like kinase FERONIA modulates reactive oxygen species (ROS) production to restrict pseudomonad colonisation (Song et al., 2021). Similarly, host-derived ROS induced by commensal microbes can function as a preventative mechanism against commensal-to-pathogen transitions, thereby stabilising symbiotic associations (Entila et al., 2024).

Additional support for this mode of host control comes from the general non-self response (GNSR) genes identified in the phyllosphere. The GNSR genes are consistently induced in response to the resident leaf microbiota, and their loss leads to marked dysbiosis characterised by overgrowth of opportunistic pathogens and severe consequences for the host (Keppler et al., 2025). Together, these findings support a model in which continuous immune perception of non-self cues, rather than episodic defence activation alone, plays a central role in shaping and stabilising plant-associated microbial communities.

Comparable genotype-dependent effects have been found in other crops. In potato (*Solanum tuberosum*), disparities in microbiome effects and modulation of microbial colonisation across cultivars – due to variations in PTI5 ERF transcription factors – have been recently described (Lukan et al., 2025; Zhao et al., 2025). Moreover, specific *NLR* genes of barley (*Hordeum vulgare*) have also been shown to strongly shape rhizosphere microbial community composition (Escudero-Martínez et al., 2022). These genotype-specific effects illustrate how host immunity impacts microbial colonisation, even under similar environmental conditions.

This immune surveillance does not operate unilaterally. Microbes are active participants in this dialogue, and some beneficials and commensals secrete non-pathogenic factors (NPFs) that subvert or dampen the host's immune activity, tipping the balance toward successful colonisation (Zhang, 2023). In fact, it has been recently proposed that host responsiveness to such NPFs likely determines the compatibility of particular host-microbe interactions, especially with commensals and beneficials (Y. Yang et al., 2022). Others appear to require an intact host immune system for successful colonisation (Tzipilevich et al., 2021), potentially using it as a cue for host health in symbiotic mutualistic interactions. This interpretation is supported by our findings in **Chapter 2**, where WCS417 exhibited the strongest downregulation of the gene encoding the extracellular protease AprA specifically big plants. In pathogenic bacteria, AprA is a well-characterised factor that degrades flagellin monomers to dampen host immune activation and facilitate immune evasion (Pel et al., 2014). The repression of *aprA* in WCS417 during successful plant association therefore suggests that, in a permis-

sive host context, immune evasion through flagellin degradation is neither required nor advantageous. Instead, WCS417 appears tuned to colonise hosts in which immune surveillance is active yet permissive, rather than fully suppressed. This pattern is consistent with a model of bidirectional sensing, in which microbial transcriptional states adjust to host immune context and colonisation outcomes depend on both host genotype and microbial responsiveness.

When the immune system is disrupted either by stress, mutation or other pathological factors, microbial communities shift toward metabolic independence, breaking cross-feeding interactions and increasing susceptibility to invasion (Case, 1990; Herren, 2020; Veseli et al., 2025). In this scenario, loss of microbiota homeostasis leads to a dysbiotic state in which commensals and opportunistic pathogens are allowed to exploit new previously restricted niches, ultimately degrading host tissues and causing disease (Pfeilmeier et al., 2021).

Host plants also influence microbial compatibility through exudate composition, which can selectively inhibit or promote specific microbes. Compounds like camalexin (Koprivova et al., 2019), putrescine (Kuiper et al., 2001) and coumarins (Stringlis et al., 2018) exemplify the selective power of plant-derived metabolites. Although the effects of each individual compound in full exudate profiles remain poorly understood, mostly due to the sheer complexity of the combinatorial effects of all exudates on microbial physiology (Estrela et al., 2021), it is clear that root-secreted compounds create chemical niches that fluctuate with developmental stage, nutrient status, and stress.

The spatial organisation of roots adds an additional layer of selectivity. Root zones differ in anatomy, cell identity, exudation patterns, and immune activity, generating heterogeneous microhabitats with distinct selective pressures (Birt et al., 2022; Kawa and Brady, 2022). High-resolution spatial profiling has confirmed that microbial communities are non-randomly distributed along and across roots, with distinct taxa occupying specific niches that correlate with local chemical and immune landscapes, including the spatial accumulation of defence-related metabolites such as the phytoalexin camalexin and coumarins like fraxetin (Cao et al., 2023). As a result, spatially localised selection pressures within a single root system shape microbial community composition and functional states, with direct consequences for colonisation outcomes.

Altogether, these examples highlight how host-mediated mechanisms shape the boundaries of microbial compatibility and define preferences across plant species and genotypes (Wippel et al., 2021). Moreover, many of these traits vary even among genetically identical plants grown in the same soil, as shown here in **Chapter 2**. The resulting rhizosphere is a unique patchy mosaic of microhabitats, shaped globally and locally, in space and time, by multiple factors – from immune responses to microhabitat structure –, in which microbial colonisers must thrive while battling for survival,

proliferation, and compatibility with the host. Therefore, to truly understand and improve bioinoculant performance, it is essential to move beyond static snapshots of host–microbe interactions and examine the ecophysiology of microbes in detail and in their natural, fluctuating environments. This need for context-aware, mechanistic insight is the focus of the main experimental chapters in this thesis.

Host variability and bioinoculant performance

Successfully deploying microbial inoculants is a challenge that goes beyond finding beneficial strains. As discussed above, host plants try to exert tight control over their microbiomes through immune surveillance, metabolic filtering, and spatial structuring. These controls do not simply act as static barriers; rather, they form a dynamic and competitive environment shaped by immune activity and resource availability that together influence microbial behaviours *in situ*. Our findings from **Chapter 2** reveal how host-mediated heterogeneity correlates with distinct microbial physiological states, with direct consequences for the efficacy of bioinoculants.

Using metatranscriptomic profiling of WCS417 in the rhizospheres of *Arabidopsis thaliana* individuals grown in natural soil, we observed notable transcriptional plasticity. On small plants, which showed signs of nutritional and oxidative stress, WCS417 adopted a stress-oriented program characterised by the activation of envelope stress responses (*algU*, *rseP*) and biofilm-associated factors (*gcdH*, *aprA*), indicating a persistence strategy in a less hospitable niche. In contrast, on larger, vigorous hosts – likely richer in labile carbon exudates and with more permissive immune activity – the bacterium expressed ribosomal proteins, respiratory complexes, and motility factors, reflecting a shift to a growth-active, exploratory state.

These divergent states exemplify how host heterogeneity and microbial function are tightly coupled, with microbial transcriptional programmes both responding to, and potentially reinforcing, distinct host outcome states. In **Chapter 2**, plant performance in natural soil resolved into a trimodal distribution following inoculation, indicating that individual hosts can occupy qualitatively different physiological states within the same experimental context. Under such conditions, WCS417 activity may reflect the host state it encounters while also feeding back onto that state through immune modulation and resource exchange. These patterns highlight that (1) mutualism is a conditional outcome of particular host–microbe interactions, and (2) microbial traits are finely tuned responses embedded in feedback loops with the host. In fact, it has been shown how even well-characterised bioinoculants can have no effect on the host or even turn into pathogens when confronted with a restrictive or dysregulated host (Beskrovnaya et al., 2020).

Our findings in **Chapter 2** thus illustrate that microbial functional traits such as biofilm formation, motility, or stress resistance are actively modulated in

response to individual host cues. This adaptive behaviour likely reflects feedback loops between host immune surveillance (e.g. LDR), metabolite secretion, and microbial sensing mechanisms. For example, traits like chemotaxis and quorum-regulated biofilm formation may allow microbes to remain within the bounds of compatibility, avoiding host sanctions while exploiting exudates. Yet, as our data suggests, the same microbe can be redirected toward persistence or competition under stress, leading to shifts in efficacy. These context-dependent dialogues explain why traits that succeed in controlled trials can fail when exposed to real soils with divergent conditions.

Summarising, although the findings in this chapter cannot fully clarify whether this differential activity of the inoculant is cause or consequence of the phenotype and activity of the host, the variable performance of WCS417 across *Arabidopsis* individuals underscores a broader principle: inoculant efficacy is inseparable from *in situ* physiology, and goes beyond mere genetic potential. Understanding and harnessing this variability is key to the development of robust and predictable bioinoculant technologies.

Genetic determinants of competitiveness in the rhizosphere

Extending beyond the transcriptional responses described in **Chapter 2**, **Chapter 3** turns to the genetic foundations of rhizosphere competence. The focus shifted from how a single inoculant adapts its physiology to how different pseudomonads vary in their inherent capacity to colonise roots in natural soil. By comparing superior and less successful colonisers, this chapter aimed to identify genetic features that distinguish strains able to consistently establish on roots with a larger population in complex rhizosphere environments.

Competitiveness in the rhizosphere often arises from traits that combine resource capture, persistence and interference against neighbours. For iron, siderophore biosynthesis genes (*pvd* genes) secure iron acquisition under limiting conditions but also deprive competitors of this resource, directly shaping community structure (Contreras-Moreno et al., 2024; Gu et al., 2025; Joshi and Khan, 2024; Shao et al., 2024). Antagonistic interactions are further enhanced by secretion systems, such as the type VI secretion system in *Pseudomonas putida* KT2440, which provides an advantage through contact-dependent inhibition of rival bacteria (Bernal et al., 2017; Vázquez et al., 2025). Similarly, secondary metabolite biosynthetic gene clusters like the *phl* operon for 2,4-diacetylphloroglucinol or the *plt* cluster for pyoluteorin production are critical not only for biocontrol but also for competitive exclusion of related pseudomonads (Haas and Défago, 2005; Zhang et al., 2025). Traits enabling persistence, such as biofilm formation through *pel* and *psl* operons, general stress responses orchestrated by *rpoS*, and flagellar/chemotaxis loci that ensure rapid occupation of nutrient-rich microhabitats,

likewise demonstrate how competitiveness arises from a tight coupling of survival, interference, and dispersal strategies (Colin et al., 2021; de Weert et al., 2002; Fleming et al., 2022; Matilla and Krell, 2024; Muriel-Millán et al., 2024).

Within this context, **Chapter 3** highlighted the *iol* locus, a 10-gene cluster for *myo*-inositol catabolism, as a determinant of competitive root colonisation, revealing multifaceted roles of inositol in rhizosphere competence. Comparative genomics showed that *iol* is enriched in top-performing *Pseudomonas* strains but absent from less-competent relatives, pointing to a key adaptive trait. Functional assays demonstrated that the locus enables inositol consumption and mediates physiological responses to this compound, including enhanced swimming motility and siderophore production. Additionally, we observed an intriguing phenotype in Δ*iol* cells during swimming assays in the presence of inositol. After the expanding front reached the edge of the plate, the culture appeared to lose its coherent swimming pattern, followed by the emergence of scattered and punctate growth across the agar surface. While the mechanistic basis of this behaviour remains unresolved, it suggests that inositol elicits behavioural responses beyond its role as a nutrient. Notably, a qualitatively similar disruption of swimming patterns is visible when WCS417 is exposed to camalexin, where uniform expansion is replaced by sparse, discontinuous growth (Zhou et al., 2025). These findings position inositol, a common component of root exudates, as a signalling molecule that shapes bacterial behaviour during colonisation. In line with this, expression data from diverse host contexts show that the *iol* locus is activated during plant and insect colonisation, underscoring its importance in host-associated environments.

Inositol has emerged more broadly as a host-derived cue in plant–microbe interactions. Recent work from O'Banion and co-authors has shown that inositol availability influences microbial fitness and colonisation dynamics in the root environment, contributing to host-dependent outcomes of microbial association (O'Banion et al., 2023). Similarly, work by Vílchez and colleagues demonstrated that plant-derived inositol acts as a chemotactic signal and biofilm inducer in *Priestia megaterium*, linking inositol perception directly to root colonisation behaviour (Vílchez et al., 2020). Together, these studies support the view that inositol is not merely a carbon source, but a conserved signal through which plants shape microbial recruitment and activity in the rhizosphere.

Beyond the *iol* locus, several metabolic pathways illustrate how resource acquisition is interwoven with broader ecological functions. Sucrose availability can trigger a signalling cascade that promotes soil-surface motility and enhances root colonisation by *Bacillus subtilis*, illustrating how a catabolic module couples to colonisation (Tian et al., 2021). Purines released by wild soybean roots during salt stress enrich *Pseudomonas* populations and induce motility, linking stress-induced exudates to microbial behaviour (Zheng et al., 2024). Likewise, rhizopine utilisation genes in *Sinorhizobium* species not only enable catabolism of these compounds but also regulate

motility and biofilm formation, proving essential for competitive nodulation (Chávez-Jacobo et al., 2023). In all these cases, catabolism is embedded in regulatory systems that connect perception of host-derived molecules to behavioural or symbiotic outcomes. Viewed in this context, the *iol* locus highlights a growing paradigm: plant-derived compounds like inositol, sucrose, purines and rhizopines act both as nutrients and as signals that drive physiology, competition, and ultimately colonisation success in natural soils. Mining or engineering loci responsible for exploitation and response towards such exudated compounds from environmental isolates, offers a promising path toward customisable inoculants.

Current challenges and future perspectives in microbiome studies

The recurring mismatch between lab promises and field unpredictability underscores the need for predictive models that capture the dynamic interplay of host, microbe, and environment. Despite major advances in understanding the ecological and genetic bases of rhizosphere competence, predictive bioinoculant design remains constrained by methodological bottlenecks. Chief among these is our limited capacity to reliably infer the true microbial entities from community sequencing data. Errors at this stage propagate directly into ecological interpretation, because our ability to detect, interpret, and reason about host–microbe–environment interactions depends critically on how microbial entities are inferred from sequence data. Amplicon sequencing of the *16S rRNA* gene remains the most widely used tool in plant microbiome research, yet it is inherently limited by fundamental biological properties such as intragenomic heterogeneity, intergenomic conservation and copy number variation. These biological properties distort the correspondence between *16S rRNA* gene sequence variation and the source organisms, inflating or collapsing diversity estimates and obscuring ecological interactions. For example, a collection of 78 isolates of *Myxococcus xanthus*, all of which bearing the same *16S rRNA* allele and isolated from the same soil patch, exhibited marked phenotypic differences in motility, physiology, predation, secondary metabolite production and compatibilities during swarm encounters, resulting in tens of differently behaving groups (Vos et al., 2013).

In the recent years, important algorithmic attempts have been proposed to model this complexity. In the case of SynComs, where genomes are typically known, the complexity of the *16S rRNA* gene assignment can be addressed by matching observed allele proportions in the community with the expected proportions in the original genomes (Selten et al., 2025). For communities with unknown genomes, approaches have focused on more efficient sequence-based clustering (Luan et al., 2023), incorporating covariates such as taxonomy (Furneaux et al., 2025), or predicting and correcting for *16S rRNA* gene copy number variation (Gao and Wu, 2023; Miao et al.,

2024). However, none of the current strategies fully resolves the central amplicon-to-genome assignment problem, particularly in environments as complex as soils, where intragenomic heterogeneity and high strain diversity remain major obstacles.

Chapter 4 introduced ASVNet as a biologically informed algorithmic advance that addresses this problem by integrating sequence similarity with co-abundance patterns across samples to reconstruct empirical OTUs (eOTUs) that faithfully approximate original microbes. By moving beyond fixed thresholds and leveraging ecological covariation, ASVNet yields clusters that better preserve genome integrity and align data analysis more closely with ecological principles. In doing so, it reframes inference from static sequence similarity toward dynamic ecological behaviour.

Looking forward, innovations like ASVNet open new opportunities for high-resolution ecological inference, collapsing technical noise while retaining biologically coherent clusters. Yet they also highlight key challenges: the empirical framework in ASVNet depends on condition-rich datasets with sufficient sample-to-sample variation, meaning that experimental designs must embrace rather than minimise natural heterogeneity. Future advances of ASVNet to solve its current limitations, such as the need to assign a single sequence to multiple overlapping eOTUs, will require more advanced techniques such as fuzzy clustering or artificial intelligence approaches such as graph-based neural networks to aid to solve the amplicon-to-genome problem. Nevertheless, the ASVNet algorithm proposed here represents a proof-of-concept that highlights the need to inform the clustering process with biological information that guides it towards a meaningful set of OTUs.

This problem, importantly, is not merely technical. It underpins our ability to monitor inoculant strains in mixed communities, track colonisation and assembly dynamics over time, and detect subtle microbiome shifts that signal stress, disease, or nutritional imbalances. Even genome-resolved metagenomics, while more direct, still captures only a modest fraction of soil diversity, with typical assemblies recruiting less than 10% of metagenomic reads back to metagenome-assembled genomes (L. Xu et al., 2021). Together, these observations reinforce the need for hybrid frameworks that integrate amplicon clustering, genome-resolved metagenomics, and multi-omics profiling to link sequences with real ecological entities. Within this trajectory, ASVNet represents a proof-of-concept that highlights that embedding ecological information into sequence inference is both feasible and necessary for linking sequence data to real ecological insights and advancing microbiome-informed agriculture.

Concluding remarks

The application of plant–microbe interactions in the field has long oscillated between the promise of harnessing microbial allies for sustainable agriculture and the disappointment when controlled successes falter in the field. As echoed in the opening metaphor of this thesis, successful deployment of bioinoculants requires knowing the battlefield, the allies, and the adversaries. The battlefield is the soil, structured into countless microhabitats that both constrain and enable microbial life. The allies are the inoculants themselves, whose genetic toolkits and physiological plasticity determine their strategies for survival and cooperation. The adversaries are the resident microbes, whose sheer diversity and competitive defences impose formidable resistance. To know all three – soil, inoculant, and community – is to transform uncertainty into foresight, and only then can we design robust inoculants that succeed not just in the lab, but in the living soil.

This thesis has shown that, to achieve this goal, microbes must not be treated as static additives but as active and adaptive ecological actors whose behaviours emerge from the interplay of soil, host, and community. Across the chapters, I have shown that inoculant physiology is plastic and context-dependent (**Chapter 2**), that specific genetic modules such as the *iol* locus act as signal receivers as much as nutrient catabolisers (**Chapter 3**), and that biologically-informed algorithms such as ASVNet can reshape the way we interpret marked gene sequence-based microbiome data (**Chapter 4**). Together, these findings converge on a single message: to understand how microbes live and persist in the rhizosphere, we need to embrace its ecological complexity rather than reduce it.

By combining ecological theory with innovative computational approaches, the next generation of plant microbiome research can move from promise to practice, establishing microbial inoculants a reliable foundation for resilient, sustainable food systems.

Acknowledgements

The author thanks Ronnie de Jonge for the supervision and revision of this manuscript. The authors want to thank as well Bas E. Dutilh, Corné M.J. Pieterse and members of the Plant-Microbe Interactions lab for helpful discussions. This study was supported by the NWO Green II Grant no. ALWGR.2017.002 (J.J.S.G., R.d.J.), and the Novo Nordisk Foundation Grant no. NNF19SA0059362 (R.d.J.).

REFERENCES

CITED WORKS

Acken, K.A., Li, B., 2023. ***Pseudomonas* virulence factor controls expression of virulence genes in *Pseudomonas entomophila***. *PLOS ONE* 18, e0284907. https://doi.org/10.1371/journal.pone.0284907

Ahmad, S., Lee, S.Y., Kong, H.G., Jo, E.J., Choi, H.K., Khan, R., Lee, S.-W., 2016. **Genetic determinants for pyomelanin production and its protective effect against oxidative stress in *Ralstonia solanacearum***. *PLOS ONE* 11, e0160845. https://doi.org/10.1371/journal.pone.0160845

Akbari Kiarood, S.L., Rahnama, K., Golmohammadi, M., Nasrollanejad, S., 2020. **Quorum-quenching endophytic bacteria inhibit disease caused by *Pseudomonas syringae* pv. *syringae* in *Citrus* cultivars**. *Journal of Basic Microbiology* **60**, 746–757. https://doi.org/10.1002/jobm.202000038

Alba, B.M., Gross, C.A., 2004. **Regulation of the *Escherichia coli* σ^E-dependent envelope stress response**. *Molecular Microbiology* 52, 613–619. https://doi.org/10.1111/j.1365-2958.2003.03982.x

Albright, M.B.N., Louca, S., Winkler, D.E., Feeser, K.L., Haig, S.-J., Whiteson, K.L., Emerson, J.B., Dunbar, J., 2022. **Solutions in microbiome engineering: prioritizing barriers to organism establishment**. *The ISME Journal* **16**, 331–338. https://doi.org/10.1038/s41396-021-01088-5

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. **Binning metagenomic contigs by coverage and composition**. *Nature Methods* **11**, 1144–1146. https://doi.org/10.1038/nmeth.3103

Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., Prior, R., Cui, Y., Payne, A.C., Yao, C.-C., Suk, H.-J., Wang, R., Yu, C.-C. (Jay), Tillberg, P., Reginato, P., Pak, N., Liu, S., Punthambaker, S., Iyer, E.P.R., Kohman, R.E., Miller, J.A., Lein, E.S., Lako, A., Cullen, N., Rodig, S., Helvie, K., Abravanel, D.L., Wagle, N., Johnson, B.E., Klughammer, J., Slyper, M., Waldman, J., Jané-Valbuena, J., Rozenblatt-Rosen, O., Regev, A., IMAXT CONSORTIUM, Church, G.M., Marblestone, A.H., Boyden, E.S., 2021. **Expansion sequencing: spatially precise *in situ* transcriptomics in intact biological systems**. *Science* **371**, eaax2656. https://doi.org/10.1126/science.aax2656

Anand, A., Falquet, L., Abou-Mansour, E., L'Haridon, F., Keel, C., Weisskopf, L., 2022. **Much more than a toxin: hydrogen cyanide is a volatile modulator of bacterial behaviour**. *bioRxiv* 2021.09.29.462390-2021.09.29.462390. https://doi.org/10.1101/2021.09.29.462390

Andrews, S., 2010. **FastQC a quality control tool for high throughput sequence data**.

Andrić, S., Meyer, T., Rigolet, A., Prigent-Combaret, C., Höfte, M., Balleux, G., Steels, S., Hoff, G., De Mot, R., McCann, A., De Pauw, E., Argüelles Arias, A., Ongena, M., 2021. **Lipopeptide interplay mediates molecular interactions between soil bacilli and pseudomonads**. *Microbiology Spectrum* **9**, e02038-21. https://doi.org/10.1128/spectrum.02038-21

Andrić, S., Rigolet, A., Argüelles Arias, A., Steels, S., Hoff, G., Balleux, G., Ongena, L., Höfte, M., Meyer, T., Ongena, M., 2022. **Plant-associated *Bacillus* mobilizes its secondary metabolites upon perception of the siderophore pyochelin produced by a *Pseudomonas* competitor**. *The ISME journal*. https://doi.org/10.1038/s41396-022-01337-1

Angelini, S., Gerez, C., Ollagnier-de Choudens, S., Sanakis, Y., Fontecave, M., Barras, F., Py, B., 2008. **NfuA, a new factor required for maturing Fe/S proteins in *Escherichia coli* under oxidative stress and iron starvation conditions**. *The Journal of Biological Chemistry* 283, 14084–14091. https://doi.org/10.1074/jbc.M709405200

Anthony, M.A., Bender, S.F., van der Heijden, M.G.A., 2023. **Enumerating soil biodiversity**. *Proceedings of the National Academy of Sciences* **120**, e2304663120. https://doi.org/10.1073/pnas.2304663120

Bach, E.M., Williams, R.J., Hargreaves, S.K., Yang, F., Hofmockel, K.S., 2018. **Greatest soil microbial diversity found in micro-habitats**. *Soil Biology and Biochemistry* **118**, 217–226. https://doi.org/10.1016/j.soilbio.2017.12.018

Bache, S., Wickham, H., 2022. **magrittr: a forward-pipe operator for R**.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M.,

Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. **Structure and function of the global topsoil microbiome**. *Nature* **560**, 233–237. https://doi.org/10.1038/s41586-018-0386-6

Bakker, P.A.H.M., Berendsen, R.L., Van Pelt, J.A., Vismans, G., Yu, K., Li, E., Van Bentum, S., Poppeliers, S.W.M., Sanchez Gil, J.J., Zhang, H., Goossens, P., Stringlis, I.A., Song, Y., de Jonge, R., Pieterse, C.M.J., 2020. **The soil-borne identity and microbiome-assisted agriculture: looking back to the future**. *Molecular Plant*, Special Issue on Plant-Microbe Interactions **13**, 1394–1401. https://doi.org/10.1016/j.molp.2020.09.017

Barchinger, S.E., Ades, S.E., 2013. **Regulated proteolysis: control of the *Escherichia coli* σ^E-dependent cell envelope stress response**. *Sub-Cellular Biochemistry* 66, 129–160. https://doi.org/10.1007/978-94-007-5940-4_6

Bardgett, R.D., van der Putten, W.H., 2014. **Belowground biodiversity and ecosystem functioning**. *Nature* **515**, 505–511. https://doi.org/10.1038/nature13855

Bardoel, B.W., Ent, S. van der, Pel, M.J.C., Tommassen, J., Pieterse, C.M.J., Kessel, K.P.M. van, Strijp, J.A.G. van, 2011. ***Pseudomonas* evades immune recognition of flagellin in both mammals and plants**. *PLOS Pathogens* **7**, e1002206. https://doi.org/10.1371/journal.ppat.1002206

Berendsen, R.L., van Verk, M.C., Stringlis, I.A., Zamioudis, C., Tommassen, J., Pieterse, C.M.J., Bakker, P.A.H.M., 2015. **Unearthing the genomes of plant-beneficial *Pseudomonas* model strains WCS358, WCS374 and WCS417**. *BMC Genomics* **16**, 539. https://doi.org/10.1186/s12864-015-1632-z

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**, 1496–1507. https://doi.org/10.1038/s41396-018-0093-1

Bernal, P., Allsopp, L.P., Filloux, A., Llamas, M.A., 2017. **The *Pseudomonas putida* T6SS is a plant warden against phytopathogens**. *The ISME Journal* **11**, 972–987. https://doi.org/10.1038/ismej.2016.169

Beskrovnaya, P., Melnyk, R.A., Liu, Z., Liu, Y., Higgins, M.A., Song, Y., Ryan, K.S., Haney, C.H., 2020. **Comparative genomics identified a genetic locus in plant-associated *Pseudomonas* spp. that is necessary for induced systemic susceptibility**. *mBio* 11. https://doi.org/10.1128/mBio.00575-20

Bickel, S., Or, D., 2023. **Aqueous habitats and carbon inputs shape the microscale geography and interaction ranges of soil bacteria**. *Communications Biology* 6, 322. https://doi.org/10.1038/s42003-023-04703-7

Birt, H.W.G., Tharp, C.L., Custer, G.F., Dini-Andreote, F., 2022. **Root phenotypes as modulators of microbial microhabitats**. *Frontiers in Plant Science* **13**.

Bittner-Eddy, P.D., Crute, I.R., Holub, E.B., Beynon, J.L., 2000. ***RPP13* is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica***. *The Plant Journal: For Cell and Molecular Biology* 21, 177–188. https://doi.org/10.1046/j.1365-313x.2000.00664.x

Blake, C., Nordgaard, M., Maróti, G., Kovács, Á.T., 2021. **Diversification of *Bacillus subtilis* during experimental evolution on *Arabidopsis thaliana* and the complementarity in root colonization of evolved subpopulations**. *Environmental Microbiology* 23, 6122–6136. https://doi.org/10.1111/1462-2920.15680

Bodilis, J., Nsigue-Meilo, S., Besaury, L., Quillet, L., 2012. **Variable copy number, intra-genomic heterogeneities and lateral transfers of the *16S rRNA* gene in *Pseudomonas***. *PLOS ONE* 7, e35647. https://doi.org/10.1371/journal.pone.0035647

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. **Near-optimal probabilistic RNA-seq quantification**. *Nature Biotechnology* **34**, 525–527. https://doi.org/10.1038/nbt.3519

Brewer, T.E., Albertsen, M., Edwards, A., Kirkegaard, R.H., Rocha, E.P.C., Fierer, N., 2020. **Unlinked *rRNA* genes are widespread among bacteria and archaea**. *The ISME Journal* **14**,

597–608.
https://doi.org/10.1038/s41396-019-0552-3

Briatte, F., 2023. ggnet2: network visualization with ggplot2.

Bushnell, B., Rood, J., Singer, E., 2017. **BBMerge – Accurate paired shotgun read merging via overlap**. *PLOS ONE* **12**, e0185056.
https://doi.org/10.1371/journal.pone.0185056

Butcher, B.G., Bronstein, P.A., Myers, C.R., Stodghill, P.V., Bolton, J.J., Markel, E.J., Filiatrault, M.J., Swingle, B., Gaballa, A., Helmann, J.D., Schneider, D.J., Cartinhour, S.W., 2011. **Characterization of the Fur regulon in *Pseudomonas syringae* pv. *tomato* DC3000**. *Journal of Bacteriology* **193**, 4598–4611.
https://doi.org/10.1128/JB.00340-11

Butts, C.T., 2008. **network: a package for managing relational data in R**. *Journal of Statistical Software* **24**, 1–36.
https://doi.org/10.18637/jss.v024.i02

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**, 581–583.
https://doi.org/10.1038/nmeth.3869

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. **eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale**. *Molecular Biology and Evolution* **38**, 5825–5829.
https://doi.org/10.1093/molbev/msab293

Cao, Z., Zuo, W., Wang, L., Chen, J., Qu, Z., Jin, F., Dai, L., 2023. **Spatial profiling of microbial communities by sequential FISH with error-robust encoding**. *Nature Communications* **14**, 1477.
https://doi.org/10.1038/s41467-023-37188-3

Case, T.J., 1990. **Invasion resistance arises in strongly interacting species-rich model competition communities**. *Proceedings of the National Academy of Sciences* 87, 9610–9614.
https://doi.org/10.1073/pnas.87.24.9610

Castrillo, G., Teixeira, P.J.P.L., Paredes, S.H., Law, T.F., Lorenzo, L. de, Feltcher, M.E., Finkel, O.M., Breakfield, N.W., Mieczkowski, P., Jones, C.D., Paz-Ares, J., Dangl, J.L., 2017. **Root microbiota drive direct integration of phosphate stress and immunity**. *Nature* **543**, 513–518.
https://doi.org/10.1038/nature21417

Cezairliyan, B.O., Sauer, R.T., 2007. **Inhibition of regulated proteolysis by RseB**. *Proceedings of the National Academy of Sciences* **104**, 3771–3776.
https://doi.org/10.1073/pnas.0611567104

Chandrangsu, P., Dusi, R., Hamilton, C.J., Helmann, J.D., 2014. **Methylglyoxal resistance in *Bacillus subtilis*: contributions of bacillithiol-dependent and independent pathways**. *Molecular Microbiology* 91, 706–715.
https://doi.org/10.1111/mmi.12489

Chang, W., 2023. **extrafont: tools for using fonts**.

Chaparro, J.M., Badri, D.V., Vivanco, J.M., 2014. **Rhizosphere microbiome assemblage is affected by plant development**. *Isme Journal* **8**, 790–803.
https://doi.org/10.1038/ismej.2013.196

Chávez-Jacobo, V.M., Becerra-Rivera, V.A., Guerrero, G., Dunn, M.F., 2023. **The *Sinorhizobium meliloti* NspS-MbaA system affects biofilm formation, exopolysaccharide production and motility in response to specific polyamines**. *Microbiology* 169, 001293.
https://doi.org/10.1099/mic.0.001293

Chen, C., Wang, M., Zhu, J., Tang, Y., Zhang, H., Zhao, Q., Jing, M., Chen, Y., Xu, X., Jiang, J., Shen, Z., 2022. **Long-term effect of epigenetic modification in plant–microbe interactions: modification of DNA methylation induced by plant growth-promoting bacteria mediates promotion process**. *Microbiome* 10, 36–36.
https://doi.org/10.1186/s40168-022-01236-9

Chen, W., Guillaume-Gentil, O., Rainer, P.Y., Gäbelein, C.G., Saelens, W., Gardeux, V., Klaeger, A., Dainese, R., Zachara, M., Zambelli, T., Vorholt, J.A., Deplancke, B., 2022. **Live-seq enables temporal transcriptomic recording of single cells**. *Nature* **608**, 733–740.
https://doi.org/10.1038/s41586-022-05046-9

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., Town, C.D., 2017. **Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome**.

The Plant Journal **89**, 789–804.
https://doi.org/10.1111/tpj.13415

Coenye, T., Vandamme, P., 2003. **Intragenomic heterogeneity between multiple *16S ribosomal RNA* operons in sequenced bacterial genomes**. *FEMS Microbiology Letters* 228, 45–49.
https://doi.org/10.1016/S0378-1097(03)00717-1

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**, e2002860.
https://doi.org/10.1371/journal.pbio.2002860

Colin, R., Ni, B., Laganenka, L., Sourjik, V., 2021. **Multiple functions of flagellar motility and chemotaxis in bacterial physiology**. *FEMS Microbiology Reviews* **45**, fuab038.
https://doi.org/10.1093/femsre/fuab038

Contreras-Moreno, F.J., Moraleda-Muñoz, A., Marcos-Torres, F.J., Cuéllar, V., Soto, M.J., Pérez, J., Muñoz-Dorado, J., 2024. **Siderophores and competition for iron govern myxobacterial predation dynamics**. *The ISME journal* **18**, wrae077.
https://doi.org/10.1093/ismejo/wrae077

Cordero-Alba, M., Bernal-Bayard, J., Ramos-Morales, F., 2012. **SrfJ, a *Salmonella* type III secretion system effector regulated by PhoP, RcsB, and IolR**. *Journal of Bacteriology* **194**, 4226–4236.
https://doi.org/10.1128/JB.00173-12

Corral, J., Sebastià, P., Coll, N.S., Barbé, J., Aranda, J., Valls, M., 2020. **Twitching and swimming motility play a role in *Ralstonia solanacearum* pathogenicity**. *mSphere* **5**, e00740-19.
https://doi.org/10.1128/mSphere.00740-19

da Costa, P.B., van Elsas, J.D., Mallon, C., dos Anjos Borges, L.G., Pereira Passaglia, L.M., 2020. **Efficiency of probiotic traits in plant inoculation is determined by environmental constrains**. *Soil Biology and Biochemistry* **148**, 107893.
https://doi.org/10.1016/j.soilbio.2020.107893

da Silva Neto, J.F., Lourenço, R.F., Marques, M.V., 2013. **Global transcriptional response of *Caulobacter crescentus* to iron availability**. *BMC Genomics* 14, 549.
https://doi.org/10.1186/1471-2164-14-549

Dal Co, A., van Vliet, S., Kiviet, D.J., Schlegel, S., Ackermann, M., 2020. **Short-range interactions govern the dynamics and functions of microbial communities**. *Nature Ecology & Evolution* **4**, 366–375.
https://doi.org/10.1038/s41559-019-1080-2

Damron, F.H., Qiu, D., Yu, H.D., 2009. **The *Pseudomonas aeruginosa* sensor kinase KinB negatively controls alginate production through AlgW-dependent MucA proteolysis**. *Journal of Bacteriology* 191, 2285–2295.
https://doi.org/10.1128/jb.01490-08

Daniels, R., Vanderleyden, J., Michiels, J., 2004. **Quorum sensing and swarming migration in bacteria**. *FEMS Microbiology Reviews* **28**, 261–289.
https://doi.org/10.1016/j.femsre.2003.09.004

Dar, D., Dar, N., Cai, L., Newman, D.K., 2021. **Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution**. *Science* 373, eabi4882.
https://doi.org/10.1126/science.abi4882

de Boer, M., van der Sluis, I., van Loon, L.C., Bakker, P.A.H.M., 1999. **Combining fluorescent *Pseudomonas* spp. strains to enhance suppression of *Fusarium* wilt of radish**. *European Journal of Plant Pathology* **105**, 201–210.
https://doi.org/10.1023/A:1008761729073

de Weert, S., Vermeiren, H., Mulders, I.H.M., Kuiper, I., Hendrickx, N., Bloemberg, G.V., Vanderleyden, J., De Mot, R., Lugtenberg, B.J.J., 2002. **Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens***. *Molecular Plant-Microbe Interactions®* **15**, 1173–1180.
https://doi.org/10.1094/MPMI.2002.15.11.1173

de Weger, L.A., van Boxtel, R., van der Burg, B., Gruters, R.A., Geels, F.P., Schippers, B., Lugtenberg, B., 1986. **Siderophores and outer membrane proteins of antagonistic, plant-growth-stimulating, root-colonizing *Pseudomonas* spp**. *Journal of Bacteriology* **165**, 585–594.

https://doi.org/10.1128/jb.165.2.585-594.1986
Delgado, C., Florez, L., Lollett, I., Lopez, C., Kangeyan, S., Kumari, H., Stylianou, M., Smiddy, R.J., Schneper, L., Sautter, R.T., Smith, D., Szatmari, G., Mathee, K., 2018. ***Pseudomonas aeruginosa* regulated intramembrane proteolysis: protease MucP can overcome mutations in the AlgO periplasmic protease to restore alginate production in nonmucoid revertants**. *Journal of Bacteriology* **200**, e00215-18. https://doi.org/10.1128/JB.00215-18
Deng, X., Li, M., Pan, X., Zheng, R., Liu, C., Chen, F., Liu, X., Cheng, Z., Jin, S., Wu, W., 2017. **Fis regulates type III secretion system by influencing the transcription of *exsA* in *Pseudomonas aeruginosa* strain PA14**. *Frontiers in Microbiology* **8**, 669. https://doi.org/10.3389/fmicb.2017.00669
Deng, X., Liang, H., Chen, K., He, C., Lan, L., Tang, X., 2014. **Molecular mechanisms of two-component system RhpRS regulating type III secretion system in *Pseudomonas syringae***. *Nucleic Acids Research* 42, 11472–11486. https://doi.org/10.1093/nar/gku865
Deng, X., Zhang, N., Li, Y., Zhu, C., Qu, B., Liu, H., Li, R., Bai, Y., Shen, Q., Falcao Salles, J., 2022. **Bio-organic soil amendment promotes the suppression of *Ralstonia solanacearum* by inducing changes in the functionality and composition of rhizosphere bacterial communities**. *New Phytologist* 235, 1558–1574. https://doi.org/10.1111/nph.18221
Dillon, S.C., Dorman, C.J., 2010. **Bacterial nucleoid-associated proteins, nucleoid structure and gene expression**. *Nature Reviews Microbiology* 8, 185–195. https://doi.org/10.1038/nrmicro2261
Dindas, J., DeFalco, T.A., Yu, G., Zhang, L., David, P., Bjornson, M., Thibaud, M.-C., Custódio, V., Castrillo, G., Nussaume, L., Macho, A.P., Zipfel, C., 2022. **Direct inhibition of phosphate transport by immune signaling in *Arabidopsis***. *Current Biology* **32**, 488-495.e5. https://doi.org/10.1016/j.cub.2021.11.063
Dini-Andreote, F., Raaijmakers, J.M., 2018. **Embracing community ecology in plant microbiome research**. *Trends in Plant Science* **23**, 467–469. https://doi.org/10.1016/j.tplants.2018.03.013
Dong, Y., Li, S., Zhao, D., Liu, J., Ma, S., Geng, J., Lu, C., Liu, Y., 2020. Io**IR, a negative regulator of the *myo*-inositol metabolic pathway, inhibits cell autoaggregation and biofilm formation by downregulating RpmA in *Aeromonas hydrophila***. *npj Biofilms and Microbiomes* 6, 1–12. https://doi.org/10.1038/s41522-020-0132-3
Dongen, S.M. van, 2000. **Graph clustering by flow simulation.** [WWW Document]. URL http://localhost/handle/1874/848 (accessed 11.28.21).
Dressler, M.D., Conde, J., Eldakar, O.T., Smith, R.P., 2019. **Timing between successive introduction events determines establishment success in bacteria with an Allee effect**. *Proceedings of the Royal Society B: Biological Sciences* **286**, 20190598. https://doi.org/10.1098/rspb.2019.0598
Durán, P., Ellis, T.J., Thiergart, T., Ågren, J., Hacquard, S., 2022. **Climate drives rhizosphere microbiome variation and divergent selection between geographically distant *Arabidopsis* populations**. *New Phytologist* 236, 608–621. https://doi.org/10.1111/nph.18357
Dziewit, L., Radlinska, M., 2016. **Two inducible prophages of an Antarctic *Pseudomonas* sp. ANT_H14 use the same capsid for packaging their genomes – characterization of a novel phage helper-satellite system**. *PLoS ONE* 11, e0158889. https://doi.org/10.1371/journal.pone.0158889
Eckshtain-Levi, N., Harris, S.L., Roscios, R.Q., Shank, E.A., 2020. **Bacterial community members increase *Bacillus subtilis* maintenance on the roots of *Arabidopsis thaliana***. *Phytobiomes Journal* 4, 303–313. https://doi.org/10.1094/PBIOMES-02-20-0019-R
Ehrensberger, A.H., Wilson, D.K., 2004. **Structural and catalytic diversity in the two family 11 aldo-keto reductases**. *Journal of Molecular Biology* **337**, 661–673. https://doi.org/10.1016/j.jmb.2004.01.059
Emami, S., Alikhani, H.A., Pourbabaee, A.A., Etesami, H., Motasharezadeh, B., Sarmadian, F., 2020. **Consortium of endophyte and rhizosphere phosphate solubilizing bacteria improves phosphorous use efficiency in** wheat cultivars in phosphorus deficient soils.

Rhizosphere 14, 100196.
https://doi.org/10.1016/j.rhisph.2020.100196

Emms, D.M., Kelly, S., 2019. **OrthoFinder: phylogenetic orthology inference for comparative genomics**. *Genome Biology* 20, 238.
https://doi.org/10.1186/s13059-019-1832-y

Emonet, A., Zhou, F., Vacheron, J., Heiman, C.M., Dénervaud Tendon, V., Ma, K.-W., Schulze-Lefert, P., Keel, C., Geldner, N., 2021. **Spatially restricted immune responses are required for maintaining root meristematic activity upon detection of bacteria**. *Current Biology* **31**, 1012-1028.e7.
https://doi.org/10.1016/j.cub.2020.12.048

Eng, T., Herbert, R.A., Martinez, U., Wang, B., Chen, J.C., Brown, J.B., Deutschbauer, A.M., Bissell, M.J., Mortimer, J.C., Mukhopadhyay, A., 2020. **Iron supplementation eliminates antagonistic interactions between root-associated bacteria**. *Frontiers in Microbiology* **11**, 1742.
https://doi.org/10.3389/fmicb.2020.01742

Engelhardt, I.C., Patko, D., Liu, Y., Mimault, M., de las Heras Martinez, G., George, T.S., MacDonald, M., Ptashnyk, M., Sukhodub, T., Stanley-Wall, N.R., Holden, N., Daniell, T.J., Dupuy, L.X., 2022. **Novel form of collective movement by soil bacteria**. *The ISME Journal* **16**, 2337–2347.
https://doi.org/10.1038/s41396-022-01277-w

Entila, F., Han, X., Mine, A., Schulze-Lefert, P., Tsuda, K., 2024. **Commensal lifestyle regulated by a negative feedback loop between *Arabidopsis* ROS and the bacterial T2SS**. *Nature Communications* 15, 456.
https://doi.org/10.1038/s41467-024-44724-2

Escudero-Martínez, C., Coulter, M., Alegría Terrazas, R., Foito, A., Kapadia, R., Pietrangelo, L., Maver, M., Sharma, R., Aprile, A., Morris, J., Hedley, P.E., Maurer, A., Pillen, K., Naclerio, G., Mimmo, T., Barton, G.J., Waugh, R., Abbott, J., Bulgarelli, D., 2022. **Identifying plant genes shaping microbiota composition in the barley rhizosphere**. *Nature Communications* **13**, 3443–3443.
https://doi.org/10.1038/s41467-022-31022-y

Estrela, S., Sánchez-Gorostiaga, A., Vila, J.C., Sánchez, Á., 2021. **Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments**. *eLife* 10, e65948.
https://doi.org/10.7554/eLife.65948

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics* 32, 3047–3048.
https://doi.org/10.1093/bioinformatics/btw354

Fabian, B.K., Foster, C., Asher, A., Hassan, K.A., Paulsen, I.T., Tetu, S.G., 2024. **Identifying the suite of genes central to swimming in the biocontrol bacterium *Pseudomonas protegens* Pf-5**. *Microbial Genomics* **10**, 001212.
https://doi.org/10.1099/mgen.0.001212

Fan, B., Carvalhais, L.C., Becker, A., Fedoseyenko, D., von Wirén, N., Borriss, R., 2012. **Transcriptomic profiling of *Bacillus amyloliquefaciens* FZB42 in response to maize root exudates**. *BMC Microbiology* **12**, 116.
https://doi.org/10.1186/1471-2180-12-116

Ferrández, A., Hawkins, A.C., Summerfield, D.T., Harwood, C.S., 2002. **Cluster II *che* genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response**. *Journal of Bacteriology* **184**, 4374–4383.
https://doi.org/10.1128/JB.184.16.4374-4383.2002

Ferreira, D.A., da Silva, T.F., Pylro, V.S., Salles, J.F., Andreote, F.D., Dini-Andreote, F., 2021. **Soil microbial diversity affects the plant-root colonization by arbuscular mycorrhizal fungi**. *Microbial Ecology* **82**, 100–103.
https://doi.org/10.1007/s00248-020-01502-z

Figueroa, W., Cazares, A., Ashworth, E.A., Weimann, A., Kadioglu, A., Floto, R.A., Welch, M., 2025. **Mutations in *mexT* bypass the stringent response dependency of virulence in *Pseudomonas aeruginosa***. *Cell Reports* **44**, 115079.
https://doi.org/10.1016/j.celrep.2024.115079

Fleming, D., Niese, B., Redman, W., Vanderpool, E., Gordon, V., Rumbaugh, K.P., 2022. **Contribution of *Pseudomonas aeruginosa* exopolysaccharides Pel and Psl to wound infections**. *Frontiers in Cellular and Infection Microbiology* **12**, 835754.
https://doi.org/10.3389/fcimb.2022.835754

Freeman, B.C., Chen, C., Yu, X., Nielsen, L., Peterson, K., Beattie, G.A., 2013. **Physiological and transcriptional responses to osmotic stress of two *Pseudomonas syringae* strains that differ in epiphytic fitness and osmotolerance**. *Journal of Bacteriology* 195, 4742–4752.
https://doi.org/10.1128/JB.00787-13

Fröschel, C., Komorek, J., Attard, A., Marsell, A., Lopez-Arboleda, W.A., Le Berre, J., Wolf, E., Geldner, N., Waller, F., Korte, A., Dröge-Laser, W., 2021. **Plant roots employ cell-layer-specific programs to respond to pathogenic and beneficial microbes**. *Cell Host & Microbe* **29**, 299-310.e7.
https://doi.org/10.1016/j.chom.2020.11.014

Fry, J., Wood, M., Poole, P.S., 2001. **Investigation of *myo*-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness**. *Molecular plant-microbe interactions: MPMI* 14, 1016–1025.
https://doi.org/10.1094/MPMI.2001.14.8.1016

Furneaux, B., Anslan, S., Somervuo, P., Hultman, J., Abrego, N., Roslin, T., Ovaskainen, O., 2025. **OptimOTU: Taxonomically aware OTU clustering with optimized thresholds and a bioinformatics workflow for metabarcoding data**. *arXiv*.
https://doi.org/10.48550/arXiv.2502.10350

Galbraith, M.P., Feng, S.F., Borneman, J., Triplett, E.W., de Bruijn, F.J., Rossbachl, S., 1998. **A functional *myo*-inositol catabolism pathway is essential for rhizopine utilization by *Sinorhizobium meliloti***. *Microbiology* **144**, 2915–2924.
https://doi.org/10.1099/00221287-144-10-2915

Gao, C.-H., 2022. **ggVennDiagram: a "ggplot2" implement of Venn diagram**.

Gao, Y., Wu, M., 2023. **Accounting for *16S rRNA* copy number prediction uncertainty and its implications in bacterial diversity analyses**. *ISME Communications* 3, 59.
https://doi.org/10.1038/s43705-023-00266-0

García-Cárdenas, E., Ortiz-Castro, R., Ruiz-Herrera, L.F., Valencia-Cantero, E., López-Bucio, J., 2022. ***Bacillus* sp. LC390B from the maize rhizosphere improves plant biomass, root elongation, and branching and requires the phytochromes PHYA and PHYB for phytostimulation**. *Journal of Plant Growth Regulation*.
https://doi.org/10.1007/s00344-022-10770-7

Garde, S., Chodisetti, P.K., Reddy, M., 2021. **Peptidoglycan: structure, synthesis, and regulation**. *EcoSal Plus* **9**.
https://doi.org/10.1128/ecosalplus.ESP-0010-2020

Gavriilidou, A., Kautsar, S.A., Zaburannyi, N., Krug, D., Müller, R., Medema, M.H., Ziemert, N., 2022. **Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes**. *Nature Microbiology* **7**, 726–735.
https://doi.org/10.1038/s41564-022-01110-2

Geels, F.P., Schippers, B., 1983. **Selection of antagonistic fluorescent *Pseudomonas* spp. and their root colonization and persistence following treatment of seed potatoes**. *Journal of Phytopathology* 108, 193–206.
https://doi.org/10.1111/j.1439-0434.1983.tb00579.x

Gilchrist, C.L.M., Booth, T.J., van Wersch, B., van Grieken, L., Medema, M.H., Chooi, Y.-H., 2021. **cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters**. *Bioinformatics Advances* **1**, vbab016.
https://doi.org/10.1093/bioadv/vbab016

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. **Microbiome datasets are compositional: and this is not optional**. *Frontiers in Microbiology* **8**, 1664–302X.
https://doi.org/10.3389/fmicb.2017.02224

Gomes, C.I., Obendorf, R.L., Horbowicz, M., 2005. ***myo*-Inositol, D-*chiro*-inositol, and D-pinitol synthesis, transport, and galactoside formation in soybean explants**. *Crop Science* 45, 1312–1319.
https://doi.org/10.2135/cropsci2004.0247

Gonzalez, M.E., Castagno, L.N., Sannazzaro, A.I., Rios, C., Gudesblat, G.E., Altabe, S.G., Bedoya-Castañeda, D.C., Estrella, M.J., Pieckenstain, F.L., 2025. **Description of *Pseudomonas gorinensis* sp. nov., a plant growth-promoting bacterium that modulates stomatal aperture**. *International Journal of Systematic and Evolutionary Microbiology* **75**, 006739.
https://doi.org/10.1099/ijsem.0.006739

Goo, E., Majerczyk, C.D., An, J.H., Chandler, J.R., Seo, Y.-S., Ham, H., Lim, J.Y., Kim, H., Lee, B., Jang, M.S., Greenberg, E.P., Hwang, I., 2012. **Bacterial quorum sensing, cooperativity, and anticipation of stationary-phase stress**. *Proceedings of the National Academy of Sciences* **109**, 19775–19780. https://doi.org/10.1073/pnas.1218092109

Goossens, P., Spooren, J., Baremans, K.C.M., Andel, A., Lapin, D., Echobardo, N., Pieterse, C.M.J., Van den Ackerveken, G., Berendsen, R.L., 2023. **Obligate biotroph downy mildew consistently induces near-identical protective microbiomes in *Arabidopsis thaliana***. *Nature Microbiology* **8**, 2349–2364. https://doi.org/10.1038/s41564-023-01502-y

Gu, S., Shao, Z., Qu, Z., Zhu, S., Shao, Y., Zhang, D., Allen, R., He, R., Shao, J., Xiong, G., Jousset, A., Friman, V.-P., Wei, Z., Kümmerli, R., Li, Z., 2025. **Siderophore synthetase-receptor gene coevolution reveals habitat- and pathogen-specific bacterial iron interaction networks**. *Science Advances* **11**, eadq5038. https://doi.org/10.1126/sciadv.adq5038

Gu, S., Wei, Z., Shao, Z., Friman, V.-P., Cao, K., Yang, T., Kramer, J., Wang, X., Li, M., Mei, X., Xu, Y., Shen, Q., Kümmerli, R., Jousset, A., 2020. **Competition for iron drives phytopathogen control by natural rhizosphere microbiomes**. *Nature Microbiology* **5**, 1002–1010. https://doi.org/10.1038/s41564-020-0719-8

Gu, Y., Banerjee, S., Dini-Andreote, F., Xu, Y., Shen, Q., Jousset, A., Wei, Z., 2022. **Small changes in rhizosphere microbiome composition predict disease outcomes earlier than pathogen density variations**. *The ISME Journal* 1–9. https://doi.org/10.1038/s41396-022-01290-z

Gupta, R., Elkabetz, D., Leibman-Markus, M., Jami, E., Bar, M., 2022. **Cytokinin-microbiome interactions regulate developmental functions**. *Environmental Microbiome* **17**, 2–2. https://doi.org/10.1186/s40793-022-00397-2

Haas, D., Défago, G., 2005. **Biological control of soil-borne pathogens by fluorescent pseudomonads**. *Nature Reviews Microbiology* 3, 307–319. https://doi.org/10.1038/nrmicro1129

Hahsler, M., Buchta, C., Hornik, K., Barnett, D., Brusco, M., Friendly, M., Koehn, H.-F., Murtagh, F., Stahl, S., 2025. **seriation: infrastructure for ordering objects using seriation**.

Hamilton, C.D., Steidl, O.R., MacIntyre, A.M., Hendrich, C.G., Allen, C., 2021. ***Ralstonia solanacearum* depends on catabolism of *myo*-inositol, sucrose, and trehalose for virulence in an infection stage-dependent manner**. *Molecular Plant-Microbe Interactions*. https://doi.org/10.1094/MPMI-10-20-0298-R

Hansen, M.L., Wibowo, M., Jarmusch, S.A., Larsen, T.O., Jelsbak, L., 2022. **Sequential interspecies interactions affect production of antimicrobial secondary metabolites in *Pseudomonas protegens* DTU9.1**. *The ISME Journal* 1–11. https://doi.org/10.1038/s41396-022-01322-8

Harrell, F., 2022. **Hmisc: Harrell Miscellaneous**.

He, D., Singh, S.K., Peng, L., Kaushal, R., Vílchez, J.I., Shao, C., Wu, X., Zheng, S., Morcillo, R.J.L., Paré, P.W., Zhang, H., 2022. **Flavonoid-attracted *Aeromonas* sp. from the *Arabidopsis* root microbiome enhances plant dehydration resistance**. *The ISME Journal* 1–11. https://doi.org/10.1038/s41396-022-01288-7

He, M., Tao, Y., Mu, K., Feng, H., Fan, Y., Liu, T., Huang, Q., Xiao, Y., Chen, W., 2025. **Coordinated regulation of chemotaxis and resistance to copper by CsoR in *Pseudomonas putida***. *eLife* **13**, RP100914. https://doi.org/10.7554/eLife.100914

Heiman, C.M., Maurhofer, M., Calderon, S., Dupasquier, M., Marquis, J., Keel, C., Vacheron, J., 2022. **Pivotal role of O-antigenic polysaccharide display in the sensitivity against phage tail-like particles in environmental *Pseudomonas* kin competition**. *The ISME Journal* 16, 1683–1693. https://doi.org/10.1038/s41396-022-01217-8

Hellinckx, J., Heermann, R., Felsl, A., Fuchs, T.M., 2017. **High binding affinity of repressor IolR avoids costs of untimely induction of *myo*-inositol utilization by *Salmonella typhimurium***. *Scientific Reports* 7, 44362. https://doi.org/10.1038/srep44362

Helmann, T.C., Deutschbauer, A.M., Lindow, S.E., 2019. **Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the**

leaf surface and apoplast. *Proceedings of the National Academy of Sciences* 116, 18900–18910.
https://doi.org/10.1073/pnas.1908858116

Hennessy, R.C., Glaring, M.A., Olsson, S., Stougaard, P., 2017. **Transcriptomic profiling of microbe–microbe interactions reveals the specific response of the biocontrol strain *P. fluorescens* In5 to the phytopathogen *Rhizoctonia solani***. *BMC Research Notes* 10, 376.
https://doi.org/10.1186/s13104-017-2704-8

Herren, C.M., 2020. **Disruption of cross-feeding interactions by invading taxa can cause invasional meltdown in microbial communities**. *Proceedings of the Royal Society B: Biological Sciences* 287, 20192945.
https://doi.org/10.1098/rspb.2019.2945

Hothorn, T., Bretz, F., Westfall, P., 2008. **Simultaneous inference in general parametric models**. *Biometrical Journal* **50**, 346–363.
https://doi.org/10.1002/bimj.200810425

Hu, J., Wei, Z., Friman, V.-P., Gu, S., Wang, X., Eisenhauer, N., Yang, T., Ma, J., Shen, Q., Xu, Y., Jousset, A., 2016. **Probiotic diversity enhances rhizosphere microbiome function and plant disease suppression**. *mBio* **7**, e01790-16.
https://doi.org/10.1128/mBio.01790-16

Hu, J., Wei, Z., Kowalchuk, G.A., Xu, Y., Shen, Q., Jousset, A., 2020. **Rhizosphere microbiome functional diversity and pathogen invasion resistance build up during plant development**. *Environmental Microbiology* **22**, 5005–5018.
https://doi.org/10.1111/1462-2920.15097

Hu, J., Yang, T., Friman, V.-P., Kowalchuk, G.A., Hautier, Y., Li, M., Wei, Z., Xu, Y., Shen, Q., Jousset, A., 2021. **Introduction of probiotic bacterial consortia promotes plant growth via impacts on the resident rhizosphere microbiome**. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20211396.
https://doi.org/10.1098/rspb.2021.1396

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. **eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses**. *Nucleic Acids Research* **47**, D309–D314.
https://doi.org/10.1093/nar/gky1085

Huertas-Rosales, Ó., Romero, M., Chan, K.-G., Hong, K.-W., Cámara, M., Heeb, S., Barrientos-Moreno, L., Molina-Henares, M.A., Travieso, M.L., Ramos-González, M.I., Espinosa-Urgel, M., 2021. **Genome-wide analysis of targets for post-transcriptional regulation by Rsm proteins in *Pseudomonas putida***. *Frontiers in Molecular Biosciences* **8**, 624061.
https://doi.org/10.3389/fmolb.2021.624061

Humphrey, P.T., Satterlee, T.T., Whiteman, N.K., 2020. **Competitive hierarchies, antibiosis, and the distribution of bacterial life history traits in a microbiome**. *bioRxiv* 2020.06.15.152272.
https://doi.org/10.1101/2020.06.15.152272

Hunter, R.C., Newman, D.K., 2010. **A putative ABC transporter, HatABCDE, is among molecular determinants of pyomelanin production in *Pseudomonas aeruginosa***. *Journal of Bacteriology* 192, 5962–5971.
https://doi.org/10.1128/JB.01021-10

Hynen, A.L., Lazenby, J.J., Savva, G.M., McCaughey, L.C., Turnbull, L., Nolan, L.M., Whitchurch, C.B., 2021. **Multiple holins contribute to extracellular DNA release in *Pseudomonas aeruginosa* biofilms**. *Microbiology* **167**, 000990.
https://doi.org/10.1099/mic.0.000990

Irie, Y., Starkey, M., Edwards, A.N., Wozniak, D.J., Romeo, T., Parsek, M.R., 2010. ***Pseudomonas aeruginosa* biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA**. *Molecular Microbiology* 78, 158–172.
https://doi.org/10.1111/j.1365-2958.2010.07320.x

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**, 5114.
https://doi.org/10.1038/s41467-018-07641-9

Jiang, G., Krishnan, A.H., Kim, Y.-W., Wacek, T.J., Krishnan, H.B., 2001. **A functional *myo*-inositol**

dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] Merr.). *Journal of Bacteriology* 183, 2595–2604. https://doi.org/10.1128/JB.183.8.2595-2604.2001

Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. **Evaluation of *16S rRNA* gene sequencing for species and strain-level microbiome analysis**. *Nature Communications* **10**, 5029. https://doi.org/10.1038/s41467-019-13036-1

Joshi, H., Khan, A., 2024. **Competition-driven phenotypic plasticity in iron acquisition and aromatic utilization confers a fitness advantage to *Pseudomonas putida* in an iron-limited rhizospheric environment**. *World Journal of Microbiology & Biotechnology* 40, 386. https://doi.org/10.1007/s11274-024-04192-8

Kanehisa, M., Goto, S., 2000. **KEGG: Kyoto Encyclopedia of Genes and Genomes**. *Nucleic Acids Research* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies**. *PeerJ* 7, e7359. https://doi.org/10.7717/peerj.7359

Kassambara, A., 2022. **rstatix: pipe-friendly framework for basic statistical tests**.

Kassambara, A., 2020. **ggpubr: "ggplot2" Based Publication Ready Plots**.

Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., Bakker, H.C. den, Deng, X., Carleton, H.A., 2019. **Mashtree: a rapid comparison of whole genome sequence files**. *Journal of Open Source Software* **4**, 1762. https://doi.org/10.21105/joss.01762

Kawa, D., Brady, S.M., 2022. **Root cell types as an interface for biotic interactions**. *Trends in Plant Science* **27**, 1173–1186. https://doi.org/10.1016/j.tplants.2022.06.003

Kehe, J., Ortiz, A., Kulesa, A., Gore, J., Blainey, P.C., Friedman, J., 2021. **Positive interactions are common among culturable bacteria**. *Science Advances* **7**, eabi7159. https://doi.org/10.1126/sciadv.abi7159

Keppler, A., Roulier, M., Pfeilmeier, S., Petti, G.C., Sintsova, A., Maier, B.A., Bortfeld-Miller, M., Sunagawa, S., Zipfel, C., Vorholt, J.A., 2025. **Plant microbiota feedbacks through dose-responsive expression of general non-self response genes**. *Nature Plants* **11**, 74–89. https://doi.org/10.1038/s41477-024-01856-z

Kiesewalter, H.T., Lozano-Andrade, C.N., Strube, M.L., Kovács, Á.T., 2020. **Secondary metabolites of *Bacillus subtilis* impact the assembly of soil-derived semisynthetic bacterial communities**. *Beilstein Journal of Organic Chemistry* **16**, 2983–2998. https://doi.org/10.3762/bjoc.16.248

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nature Biotechnology* **37**, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, J.Y., Yi, M., Kim, M., Lee, S., Moon, H.S., Yong, D., Yong, T.-S., 2021. **Measuring the absolute abundance of the microbiome by adding yeast containing *16S rRNA* gene from a hyperthermophile**. *MicrobiologyOpen* 10, e1220. https://doi.org/10.1002/mbo3.1220

Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M., 2000. ***rRNA* operon copy number reflects ecological strategies of bacteria**. *Applied and Environmental Microbiology* **66**, 1328–1333. https://doi.org/10.1128/AEM.66.4.1328-1333.2000

Klonowska, A., Melkonian, R., Miché, L., Tisseyre, P., Moulin, L., 2018. **Transcriptomic profiling of *Burkholderia phymatum* STM815, *Cupriavidus taiwanensis* LMG19424 and *Rhizobium mesoamericanum* STM3625 in response to *Mimosa pudica* root exudates illuminates the molecular basis of their nodulation competitiveness and symbiotic evolutionary history**. *BMC Genomics* 19, 105. https://doi.org/10.1186/s12864-018-4487-2

Ko, D., Kang, J., Kiba, T., Park, J., Kojima, M., Do, J., Kim, K.Y., Kwon, M., Endler, A., Song, W.-Y.,

Martinoia, E., Sakakibara, H., Lee, Y., 2014. ***Arabidopsis* ABCG14 is essential for the root-to-shoot translocation of cytokinin**. *Proceedings of the National Academy of Sciences* **111**, 7150–7155. https://doi.org/10.1073/pnas.1321519111

Kobayashi, K., 2021. **Diverse LXG toxin and antitoxin systems specifically mediate intraspecies competition in *Bacillus subtilis* biofilms**. *PLOS Genetics* 17, e1009682. https://doi.org/10.1371/journal.pgen.1009682

Kohler, P.R.A., Choong, E.-L., Rossbach, S., 2011. **The RpiR-like repressor IolR regulates inositol catabolism in *Sinorhizobium meliloti***. *Journal of Bacteriology* **193**, 5155–5163. https://doi.org/10.1128/JB.05371-11

Kohler, P.R.A., Zheng, J.Y., Schoffers, E., Rossbach, S., 2010. **Inositol catabolism, a key pathway in *Sinorhizobium meliloti* for competitive host nodulation**. *Applied and Environmental Microbiology* **76**, 7972–7980. https://doi.org/10.1128/AEM.01972-10

Kolde, R., 2019. **pheatmap: Pretty Heatmaps. R package version 1.0.12.** [WWW Document]. URL https://CRAN.R-project.org/package=pheatmap

Konstantinidis, K.T., Tiedje, J.M., 2005. **Genomic insights that advance the species definition for prokaryotes**. *Proceedings of the National Academy of Sciences* **102**, 2567–2572. https://doi.org/10.1073/pnas.0409727102

Koprivova, A., Schuck, S., Jacoby, R.P., Klinkhammer, I., Welter, B., Leson, L., Martyn, A., Nauen, J., Grabenhorst, N., Mandelkow, J.F., Zuccaro, A., Zeier, J., Kopriva, S., 2019. **Root-specific camalexin biosynthesis controls the plant growth-promoting effects of multiple bacterial strains**. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 15735–15744. https://doi.org/10.1073/pnas.1818604116

Kopylova, E., Noé, L., Touzet, H., 2012. **SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data**. *Bioinformatics* 28, 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kraigher, B., Butolen, M., Stefanic, P., Mandic Mulec, I., 2022. **Kin discrimination drives territorial exclusion during *Bacillus subtilis* swarming and restrains exploitation of surfactin**. *The ISME Journal* 16, 833–841. https://doi.org/10.1038/s41396-021-01124-4

Kröger, C., Rothhardt, J.E., Brokatzky, D., Felsl, A., Kary, S.C., Heermann, R., Fuchs, T.M., 2018. **The small RNA RssR regulates *myo*-inositol degradation by *Salmonella enterica***. *Scientific Reports* **8**, 17739. https://doi.org/10.1038/s41598-018-35784-8

Kröger, C., Srikumar, S., Ellwart, J., Fuchs, T.M., 2011. **Bistability in *myo*-inositol utilization by *Salmonella enterica* serovar *Typhimurium***. *Journal of Bacteriology* **193**, 1427–1435. https://doi.org/10.1128/JB.00043-10

Krysciak, D., Grote, J., Rodriguez Orbegoso, M., Utpatel, C., Förstner, K.U., Li, L., Schmeisser, C., Krishnan, H.B., Streit, W.R., 2014. **RNA sequencing analysis of the broad-host-range strain *Sinorhizobium fredii* NGR234 identifies a large set of genes linked to quorum sensing-dependent regulation in the background of a *traI* and *ngrI* deletion mutant**. *Applied and Environmental Microbiology* **80**, 5655–5671. https://doi.org/10.1128/AEM.01835-14

Kuiper, I., Bloemberg, G.V., Noreen, S., Thomas-Oates, J.E., Lugtenberg, B.J.J., 2001. **Increased uptake of putrescine in the rhizosphere inhibits competitive root colonization by *Pseudomonas fluorescens* strain WCS365**. *Molecular Plant-Microbe Interactions* 14, 1096–1104. https://doi.org/10.1094/MPMI.2001.14.9.1096

Kupferschmied, P., Maurhofer, M., Keel, C., 2013. **Promise for plant pest control: root-associated pseudomonads with insecticidal activities**. *Frontiers in Plant Science* 4, 287. https://doi.org/10.3389/fpls.2013.%2000287

Kupferschmied, P., Péchy-Tarr, M., Imperiali, N., Maurhofer, M., Keel, C., 2014. **Domain shuffling in a sensor protein contributed to the evolution of insect pathogenicity in plant-beneficial *Pseudomonas protegens***. *PLOS Pathogens* 10, e1003964. https://doi.org/10.1371/journal.ppat.1003964

Kurkjian, H.M., Akbari, M.J., Momeni, B., 2021. **The impact of interactions on invasion and colonization resistance in microbial communities**. *PLOS Computational Biology* 17,

e1008643.
https://doi.org/10.1371/journal.pcbi.1008643
Langmead, B., Salzberg, S.L., 2012. **Fast gapped-read alignment with Bowtie 2**. *Nature Methods* **9**, 357–359.
https://doi.org/10.1038/nmeth.1923
Lee, S.-M., Kong, H.G., Song, G.C., Ryu, C.-M., 2021. **Disruption of Firmicutes and Actinobacteria abundance in tomato rhizosphere causes the incidence of bacterial wilt disease**. *The ISME Journal* 15, 330–347.
https://doi.org/10.1038/s41396-020-00785-x
Lenth, R.V., 2023. **emmeans: Estimated Marginal Means, aka least-squares means**.
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., Grant, S.R., Doty, S.L., Ley, R.E., Zhao, B., Venturi, V., Pelletier, D.A., Vorholt, J.A., Tringe, S.G., Woyke, T., Dangl, J.L., 2018. **Genomic features of bacterial adaptation to plants**. *Nature Genetics* **50**, 138–150.
https://doi.org/10.1038/s41588-017-0012-9
Lewis, J.D., Wu, R., Guttman, D.S., Desveaux, D., 2010. **Allele-specific virulence attenuation of the *Pseudomonas syringae* HopZ1a type III effector via the *Arabidopsis* ZAR1 resistance protein**. *PLoS genetics* 6, e1000894.
https://doi.org/10.1371/journal.pgen.1000894
Li, E., de Jonge, R., Liu, C., Jiang, H., Friman, V.-P., Pieterse, C.M.J., Bakker, P.A.H.M., Jousset, A., 2021a. **Rapid evolution of bacterial mutualism in the plant rhizosphere**. *Nature Communications* **12**, 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., 2021b. **Experimental-evolution-driven identification of *Arabidopsis* rhizosphere competence genes in *Pseudomonas protegens***. *mBio* **12**, e00927-21.
https://doi.org/10.1128/mBio.00927-21
Li, H., La, S., Zhang, X., Gao, L., Tian, Y., 2021. **Salt-induced recruitment of specific root-associated bacterial consortium capable of enhancing plant adaptability to salt stress**. *The ISME Journal* 15, 2865–2882.
https://doi.org/10.1038/s41396-021-00974-2
Li, M., Pommier, T., Yin, Y., Wang, J., Gu, S., Jousset, A., Keuskamp, J., Wang, H., Wei, Z., Xu, Y., Shen, Q., Kowalchuk, G.A., 2022. **Indirect reduction of *Ralstonia solanacearum* via pathogen helper inhibition**. *The ISME Journal* **16**, 868–875.
https://doi.org/10.1038/s41396-021-01126-2
Li, M., Wei, Z., Wang, J., Jousset, A., Friman, V.-P., Xu, Y., Shen, Q., Pommier, T., 2019. **Facilitation promotes invasions in plant-associated microbial communities**. *Ecology Letters* **22**, 149–158.
https://doi.org/10.1111/ele.13177
Li, Y., Ye, Z., Lu, P., Lu, L., 2021. **Pyomelanin produced by *Streptomyces* sp. ZL-24 and its protective effects against SH-SY5Y cells injury induced by hydrogen peroxide**. *Scientific Reports* 11, 16649.
https://doi.org/10.1038/s41598-021-94598-3
Lin, R., Wu, H., Kong, X., Ren, H., Lu, Z., 2023. ***Ribosomal RNA* gene operon copy number, a functional trait indicating the hydrocarbon degradation level of bacterial communities**. *Journal of Hazardous Materials* 459, 132100.
https://doi.org/10.1016/j.jhazmat.2023.132100
Ling, N., Wang, T., Kuzyakov, Y., 2022. **Rhizosphere bacteriome structure and functions**. *Nature Communications* **13**, 836–836.
https://doi.org/10.1038/s41467-022-28448-9
Liu, F., Hewezi, T., Lebeis, S.L., Pantalone, V., Grewal, P.S., Staton, M.E., 2019. **Soil indigenous microbiome and plant genotypes cooperatively modify soybean rhizosphere microbiome assembly**. *BMC Microbiology* **19**, 1–19.
https://doi.org/10.1186/s12866-019-1572-x
Liu, X., Le Roux, X., Salles, J.F., 2022. **The legacy of microbial inoculants in agroecosystems and potential for tackling climate change challenges**. *iScience* 25, 103821.
https://doi.org/10.1016/j.isci.2022.103821
Liu, Z., Beskrovnaya, P., Melnyk, R.A., Hossain, S.S., Khorasani, S., O'Sullivan, L.R., Wiesmann,

C.L., Bush, J., Richard, J.D., Haney, C.H., 2018. **A genome-wide screen identifies genes in rhizosphere-associated *Pseudomonas* required to evade plant defenses**. *mBio* **9**. https://doi.org/10.1128/mBio.00433-18

Llamas, M.A., Imperi, F., Visca, P., Lamont, I.L., 2014. **Cell-surface signaling in *Pseudomonas*: stress responses, iron transport, and pathogenicity**. *FEMS Microbiology Reviews* **38**, 569–597. https://doi.org/10.1111/1574-6976.12078

Lopes, M.J. dos S., Dias-Filho, M.B., Gurgel, E.S.C., 2021. **Successful plant growth-promoting microbes: inoculation methods and abiotic factors**. *Frontiers in Sustainable Food Systems* **5**. https://doi.org/10.3389/fsufs.2021.606454

López, J.L., Fourie, A., Poppeliers, S.W.M., Pappas, N., Sánchez-Gil, J.J., de Jonge, R., Dutilh, B.E., 2023. **Growth rate is a dominant factor predicting the rhizosphere effect**. *The ISME journal* **17**, 1396–1405. https://doi.org/10.1038/s41396-023-01453-6

Love, M.I., Huber, W., Anders, S., 2014. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biology* 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Luan, T., Muralidharan, H.S., Alshehri, M., Mittra, I., Pop, M., 2023. **SCRAPT: an iterative algorithm for clustering large 16S rRNA gene data sets**. *Nucleic Acids Research* **51**, e46. https://doi.org/10.1093/nar/gkad158

Lukan, T., Kraigher, B., Pogačar, K., Stare, K., Kovačič, T.G., Zagorščak, M., Petek, M., Stefanic, P., Vozelj, A., Levak, V., Povalej, T.M., García, J.M., Pozo, M.J., Mandić-Mulec, I., Gruden, K., 2025. **An ERF transcription factor PTI5, a novel regulator of endophyte community maintenance in potato**. *bioRxiv* 2025.04.24.650297. https://doi.org/10.1101/2025.04.24.650297

Lynch, J.A., Kierstead, M.E., 1985. **Antibiotic resistance associated with inositol metabolism in salmonellae from Ontario cattle**. *The Canadian Veterinary Journal* 26, 251.

Ma, K.-W., Niu, Y., Jia, Y., Ordon, J., Copeland, C., Emonet, A., Geldner, N., Guan, R., Stolze, S.C., Nakagami, H., Garrido-Oter, R., Schulze-Lefert, P., 2021. **Coordination of microbe–host homeostasis by crosstalk with plant innate immunity**. *Nature Plants* **7**, 814–825. https://doi.org/10.1038/s41477-021-00920-2

Ma, L., Conover, M., Lu, H., Parsek, M.R., Bayles, K., Wozniak, D.J., 2009. **Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix**. *PLOS Pathogens* **5**, e1000354. https://doi.org/10.1371/journal.ppat.1000354

Ma, L., Wang, J., Wang, S., Anderson, E.M., Lam, J.S., Parsek, M.R., Wozniak, D.J., 2012. **Synthesis of multiple *Pseudomonas aeruginosa* biofilm matrix exopolysaccharides is post-transcriptionally regulated**. *Environmental microbiology* **14**, 1995–2005. https://doi.org/10.1111/j.1462-2920.2012.02753.x

Mallon, C.A., Elsas, J.D. van, Salles, J.F., 2015. **Microbial invasions: the process, patterns, and mechanisms**. *Trends in Microbiology* **23**, 719–729. https://doi.org/10.1016/j.tim.2015.07.013

Mallon, C.A., Le Roux, X., van Doorn, G.S., Dini-Andreote, F., Poly, F., Salles, J.F., 2018. **The impact of failure: unsuccessful bacterial invasions steer the soil microbial community away from the invader's niche**. *The ISME Journal* **12**, 728–741. https://doi.org/10.1038/s41396-017-0003-y

Markel, E., Stodghill, P., Bao, Z., Myers, C.R., Swingle, B., 2016. **AlgU controls expression of virulence genes in *Pseudomonas syringae* pv. *tomato* DC3000**. *Journal of Bacteriology* **198**, 2330–2344. https://doi.org/10.1128/JB.00276-16

Martin, M., 2011. **Cutadapt removes adapter sequences from high-throughput sequencing reads**. *EMBnet.journal* 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Martínez-García, E., de Lorenzo, V., 2011. **Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440**. *Environmental Microbiology* 13, 2702–2716. https://doi.org/10.1111/j.1462-2920.2011.02538.x

Matilla, M.A., Cano-Muñoz, M., Monteagudo-Cascales, E., Krell, T., 2025. **Phosphorylated CheV interacts with a subset of chemoreceptors**. *bioRxiv* 2025.02.06.636884.

https://doi.org/10.1101/2025.02.06.636884
Matilla, M.A., Espinosa-Urgel, M., Rodríguez-Herva, J.J., Ramos, J.L., Ramos-González, M.I., 2007. **Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere**. *Genome Biology* **8**, R179. https://doi.org/10.1186/gb-2007-8-9-r179
Matilla, M.A., Krell, T., 2024. **Bacterial amino acid chemotaxis: a widespread strategy with multiple physiological and ecological roles**. *Journal of Bacteriology* 206, e00300-24. https://doi.org/10.1128/jb.00300-24
Mavrodi, O.V., McWilliams, J.R., Peter, J.O., Berim, A., Hassan, K.A., Elbourne, L.D.H., LeTourneau, M.K., Gang, D.R., Paulsen, I.T., Weller, D.M., Thomashow, L.S., Flynt, A.S., Mavrodi, D.V., 2021. **Root exudates alter the expression of diverse metabolic, transport, regulatory, and stress response genes in rhizosphere *Pseudomonas***. *Frontiers in Microbiology* **12**.
Mawarda, P.C., Lakke, S.L., Dirk van Elsas, J., Salles, J.F., 2022. **Temporal dynamics of the soil bacterial community following *Bacillus* invasion**. *iScience* **25**, 104185. https://doi.org/10.1016/j.isci.2022.104185
Mawarda, P.C., Le Roux, X., Dirk van Elsas, J., Salles, J.F., 2020. **Deliberate introduction of invisible invaders: a critical appraisal of the impact of microbial inoculants on soil microbial communities**. *Soil Biology and Biochemistry* 148, 107874. https://doi.org/10.1016/j.soilbio.2020.107874
McMillan, H.M., Kuehn, M.J., 2023. **Proteomic profiling reveals distinct bacterial extracellular vesicle subpopulations with possibly unique functionality**. *Applied and Environmental Microbiology* 89, e01686-22. https://doi.org/10.1128/aem.01686-22
Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., Thomas, P.D., 2021. **PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API**. *Nucleic Acids Research* **49**, D394–D403. https://doi.org/10.1093/nar/gkaa1106
Miao, J., Chen, T., Misir, M., Lin, Y., 2024. **Deep learning for predicting *16S rRNA* gene copy number**. *Scientific Reports* **14**, 14282. https://doi.org/10.1038/s41598-024-64658-5
Molina, L., Ramos, C., Duque, E., Ronchel, M.C., García, J.M., Wyke, L., Ramos, J.L., 2000. **Survival of *Pseudomonas putida* KT2440 in soil and in the rhizosphere of plants under greenhouse and environmental conditions**. *Soil Biology and Biochemistry* **32**, 315–321. https://doi.org/10.1016/S0038-0717(99)00156-X
Molina-Santiago, C., Vela-Corcía, D., Petras, D., Díaz-Martínez, L., Pérez-Lorente, A.I., Sopeña-Torres, S., Pearson, J., Caraballo-Rodríguez, A.M., Dorrestein, P.C., de Vicente, A., Romero, D., 2021. **Chemical interplay and complementary adaptative strategies toggle bacterial antagonism and co-existence**. *Cell Reports* **36**, 109449–109449. https://doi.org/10.1016/j.celrep.2021.109449
Mooij, M.J., Drenkard, E., Llamas, M.A., Vandenbroucke-Grauls, C.M.J.E., Savelkoul, P.H.M., Ausubel, F.M., Bitter, W., 2007. **Characterization of the integrated filamentous phage Pf5 and its involvement in small-colony formation**. *Microbiology* **153**, 1790–1798. https://doi.org/10.1099/mic.0.2006/003533-0
Mosimann, C., Oberhänsli, T., Ziegler, D., Nassal, D., Kandeler, E., Boller, T., Mäder, P., Thonar, C., 2017. **Tracing of two *Pseudomonas* strains in the root and rhizoplane of maize, as related to their plant growth-promoting effect in contrasting soils**. *Frontiers in Microbiology* 7. https://doi.org/10.3389/fmicb.2016.02150
Muriel-Millán, L.F., Montelongo-Martínez, L.F., González-Valdez, A., Bedoya-Pérez, L.P., Cocotl-Yañez, M., Soberón-Chávez, G., 2024. **The alternative sigma factor RpoS regulates *Pseudomonas aeruginosa* quorum sensing response by repressing the *pqsABCDE* operon and activating *vfr***. *Molecular Microbiology* **121**, 291–303. https://doi.org/10.1111/mmi.15224
Murphy, P.J., Wexler, W., Grzemski, W., Rao, J.P., Gordon, D., 1995. **Rhizopines—Their role in symbiosis and competition**. *Soil Biology and Biochemistry* **27**, 525–529. https://doi.org/10.1016/0038-0717(95)98627-Z
Nguyen, N.-P., Warnow, T., Pop, M., White, B., 2016. **A perspective on *16S rRNA* operational taxonomic unit clustering using sequence similarity**. *npj Biofilms and Microbiomes* **2**, 1–8. https://doi.org/10.1038/npjbiofilms.2016.4

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**, E2450–E2459. https://doi.org/10.1073/pnas.1616148114

Nobori, T., Velásquez, A.C., Wu, J., Kvitko, B.H., Kremer, J.M., Wang, Y., He, S.Y., Tsuda, K., 2018. **Transcriptome landscape of a bacterial pathogen under plant immunity**. *Proceedings of the National Academy of Sciences* **115**, E3055–E3064. https://doi.org/10.1073/pnas.1800529115

Nolan, L.M., McCaughey, L.C., Merjane, J., Turnbull, L., Whitchurch, C.B., 2020. **ChpC controls twitching motility-mediated expansion of *Pseudomonas aeruginosa* biofilms in response to serum albumin, mucin and oligopeptides**. *Microbiology (Reading, England)* 166, 669–678. https://doi.org/10.1099/mic.0.000911

Nordgaard, M., Blake, C., Maróti, G., Hu, G., Wang, Y., Strube, M.L., Kovács, Á.T., 2022. **Experimental evolution of *Bacillus subtilis* on *Arabidopsis thaliana* roots reveals fast adaptation and improved root colonization**. *iScience* 25, 104406. https://doi.org/10.1016/j.isci.2022.104406

Nuccio, E.E., Starr, E., Karaoz, U., Brodie, E.L., Zhou, J., Tringe, S.G., Malmstrom, R.R., Woyke, T., Banfield, J.F., Firestone, M.K., Pett-Ridge, J., 2020. **Niche differentiation is spatially and temporally regulated in the rhizosphere**. *Isme Journal*. https://doi.org/10.1038/s41396-019-0582-x

O'Banion, B.S., Jones, P., Demetros, A.A., Kelley, B.R., Wagner, A.S., Chen, J.-G., Muchero, W., Reynolds, T.B., Jacobson, D., Lebeis, S.L., 2023. **Plant inositol transport influences bacterial colonization phenotypes**. *bioRxiv*. https://doi.org/10.1101/2023.05.03.538913

O'Callaghan, M., Ballard, R.A., Wright, D., 2022. **Soil microbial inoculants for sustainable agriculture: limitations and opportunities**. *Soil Use and Management* **38**, 1340–1369. https://doi.org/10.1111/sum.12811

Ooms, J., 2021. **magick: advanced graphics and image-processing in R**.

Opota, O., Vallet-Gély, I., Vincentelli, R., Kellenberger, C., Iacovache, I., Gonzalez, M.R., Roussel, A., Goot, F.-G. van der, Lemaitre, B., 2011. **Monalysin, a novel ß-pore-forming toxin from the *Drosophila* pathogen *Pseudomonas entomophila*, contributes to host intestinal damage and lethality**. *PLOS Pathogens* **7**, e1002259. https://doi.org/10.1371/journal.ppat.1002259

Oyserman, B.O., Flores, S.S., Griffioen, T., Pan, X., van der Wijk, E., Pronk, L., Lokhorst, W., Nurfikari, A., Paulson, J.N., Movassagh, M., Stopnisek, N., Kupczok, A., Cordovez, V., Carrión, V.J., Ligterink, W., Snoek, B.L., Medema, M.H., Raaijmakers, J.M., 2022. **Disentangling the genetic basis of rhizosphere microbiome assembly in tomato**. *Nature Communications* **13**, 3228–3228. https://doi.org/10.1038/s41467-022-30849-9

Pacheco-Moreno, A., Stefanato, F.L., Ford, J.J., Trippel, C., Uszkoreit, S., Ferrafiat, L., Grenga, L., Dickens, R., Kelly, N., Kingdon, A.D., Ambrosetti, L., Nepogodiev, S.A., Findlay, K.C., Cheema, J., Trick, M., Chandra, G., Tomalin, G., Malone, J.G., Truman, A.W., 2021. **Pan-genome analysis identifies intersecting roles for *Pseudomonas* specialized metabolites in potato pathogen inhibition**. *eLife* **10**, e71900. https://doi.org/10.7554/eLife.71900

Pagès, H., Aboyoun, P., Gentleman, R., DebRoy, S., 2023. **Biostrings: efficient manipulation of biological strings**. https://doi.org/10.18129/B9.bioc.Biostrings

Pan, P., Gu, Y., Sun, D.-L., Wu, Q.L., Zhou, N.-Y., 2023. **Microbial diversity biased estimation caused by intragenomic heterogeneity and interspecific conservation of *16S rRNA* genes**. *Applied and Environmental Microbiology* 89, e02108-22. https://doi.org/10.1128/aem.02108-22

Pan, S., Zhu, C., Zhao, X.-M., Coelho, L.P., 2022. **A deep siamese neural network improves metagenome-assembled genomes in microbiome datasets across different environments**. *Nature Communications* 13, 2326–2326. https://doi.org/10.1038/s41467-022-29843-y

Parnell, J.J., Berka, R., Young, H.A., Sturino, J.M., Kang, Y., Barnhart, D.M., DiLeo, M.V., 2016. **From the lab to the farm: an industrial perspective of plant beneficial microorganisms.** *Frontiers in plant science* **7**, 1110–1110.

https://doi.org/10.3389/fpls.2016.01110
Passarelli-Araujo, H., Franco, G.R., Venancio, T.M., 2022. **Network analysis of ten thousand genomes shed light on *Pseudomonas* diversity and classification**. *Microbiological Research* 254, 126919.
https://doi.org/10.1016/j.micres.2021.126919
Pedersen, T.L., 2022. **patchwork: the composer of plots**.
Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L., Poles, M., Brown, S.M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K., Pei, Z., 2010. **Diversity of *16S rRNA* genes within individual prokaryotic genomes**. *Applied and Environmental Microbiology* **76**, 3886–3897.
https://doi.org/10.1128/AEM.02953-09
Pel, M.J.C., van Dijken, A.J.H., Bardoel, B.W., Seidl, M.F., van der Ent, S., van Strijp, J.A.G., Pieterse, C.M.J., 2014. ***Pseudomonas syringae* evades host immunity by degrading flagellin monomers with alkaline protease AprA**. *Molecular Plant-Microbe Interactions®* **27**, 603–610.
https://doi.org/10.1094/MPMI-02-14-0032-R
Peterson, B., Carl, P., 2020. **PerformanceAnalytics: econometric tools for performance and risk analysis**.
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**, 852–864.
https://doi.org/10.1038/s41564-021-00929-5
Philippot, L., Raaijmakers, J.M., Lemanceau, P., van der Putten, W.H., 2013. **Going back to the roots: the microbial ecology of the rhizosphere**. *Nature Reviews Microbiology* **11**, 789–799.
https://doi.org/10.1038/nrmicro3109
Piccardi, P., Alberti, G., Alexander, J.M., Mitri, S., 2022. **Microbial invasion of a toxic medium is facilitated by a resident community but inhibited as the community co-evolve**s. *The ISME Journal* 1–9.
https://doi.org/10.1038/s41396-022-01314-8
Pieterse, C.M.J., Berendsen, R.L., de Jonge, R., Stringlis, I.A., Van Dijken, A.J.H., Van Pelt, J.A., Van Wees, S.C.M., Yu, K., Zamioudis, C., Bakker, P.A.H.M., 2021. ***Pseudomonas simiae* WCS417: star track of a model beneficial rhizobacterium**. *Plant and Soil* **461**, 245–263.
https://doi.org/10.1007/s11104-020-04786-9
Pini, F., East, A.K., Appia-Ayme, C., Tomek, J., Karunakaran, R., Mendoza-Suárez, M., Edwards, A., Terpolilli, J.J., Roworth, J., Downie, J.A., Poole, P.S., 2017. **Bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion**. *Plant Physiology* **174**, 1289–1306.
https://doi.org/10.1104/pp.16.01302
Poppeliers, S.W., Sánchez-Gil, J.J., de Jonge, R., 2023. **Microbes to support plant health: understanding bioinoculant success in complex conditions**. *Current Opinion in Microbiology* **73**, 102286.
https://doi.org/10.1016/j.mib.2023.102286
Poppeliers, S.W.M., Sánchez-Gil, J.J., López, J.L., Dutilh, B.E., Pieterse, C.M.J., Jonge, R. de, 2024. **High-resolution quantification of the rhizosphere effect along a soil-to-root gradient shows selection-driven convergence of rhizosphere microbiomes**. *bioRxiv* 2024–06, 2024.06.21.600027.
https://doi.org/10.1101/2024.06.21.600027
Price, M.N., Deutschbauer, A.M., Arkin, A.P., 2022. **Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics**. *PLOS Genetics* 18, e1010156–e1010156.
https://doi.org/10.1371/journal.pgen.1010156
Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y., Carlson, H.K., Esquivel, Z., Sadeeshkumar, H., Chakraborty, R., Zane, G.M., Rubin, B.E., Wall, J.D., Visel, A., Bristow, J., Blow, M.J., Arkin, A.P., Deutschbauer, A.M., 2018. **Mutant phenotypes for thousands of bacterial genes of unknown function**. *Nature* **557**, 503–509.
https://doi.org/10.1038/s41586-018-0124-0
Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A.H., Nieuwdorp, M., Levin, E., 2020. **Comparing bioinformatic pipelines for microbial *16S rRNA* amplicon sequencing**. *PLOS ONE* **15**, e0227434.

https://doi.org/10.1371/journal.pone.0227434
Purtschert-Montenegro, G., Cárcamo-Oyarce, G., Pinto-Carbó, M., Agnoli, K., Bailly, A., Eberl, L., 2022. ***Pseudomonas putida*** **mediates bacterial killing, biofilm invasion and biocontrol with a type IVB secretion system**. *Nature Microbiology* 1–11. https://doi.org/10.1038/s41564-022-01209-6
Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., Zanini, F., 2022. **Analysing high-throughput sequencing data in Python with HTSeq 2.0**. *Bioinformatics* **38**, 2943–2945. https://doi.org/10.1093/bioinformatics/btac166
Qi, S.S., Bogdanov, A., Cnockaert, M., Acar, T., Ranty-Roby, S., Coenye, T., Vandamme, P., König, G.M., Crüsemann, M., Carlier, A., 2021. **Induction of antibiotic specialized metabolism by co-culturing in a collection of phyllosphere bacteria**. *Environmental Microbiology* **23**, 2132–2151. https://doi.org/10.1111/1462-2920.15382
Qian, J.J., Akçay, E., 2020. **The balance of interaction types determines the assembly and stability of ecological communities**. *Nature Ecology & Evolution* 4, 356–365. https://doi.org/10.1038/s41559-020-1121-x
R Core Team, 2022. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria.
Rasmussen, T.B., Bjarnsholt, T., Skindersoe, M.E., Hentzer, M., Kristoffersen, P., Köte, M., Nielsen, J., Eberl, L., Givskov, M., 2005a. **Screening for quorum-sensing inhibitors (QSI) by use of a novel genetic system, the QSI selector**. *Journal of Bacteriology* **187**, 1799–1814. https://doi.org/10.1128/JB.187.5.1799-1814.2005
Rasmussen, T.B., Skindersoe, M.E., Bjarnsholt, T., Phipps, R.K., Christensen, K.B., Jensen, P.O., Andersen, J.B., Koch, B., Larsen, T.O., Hentzer, M., Eberl, L., Hoiby, N., Givskov, M., 2005b. **Identity and effects of quorum-sensing inhibitors produced by** ***Penicillium species***. *Microbiology* **151**, 1325–1340. https://doi.org/10.1099/mic.0.27715-0
Reiser, L., Bakker, E., Subramaniam, S., Chen, X., Sawant, S., Khosa, K., Prithvi, T., Berardini, T.Z., 2024. **The** ***Arabidopsis*** **Information Resource in 2024**. *Genetics* **227**, iyae027. https://doi.org/10.1093/genetics/iyae027
Rensburg, H.J. van, Schandry, N., Waelchli, J., Stengele, K., Cadot, S., Jandrasits, K., Becker, C., Schlaeppi, K., 2025. **A TNL receptor mediates microbiome feedbacks in** ***Arabidopsis***. *bioRxiv* 2025.02.25.640125. https://doi.org/10.1101/2025.02.25.640125
Rich-Griffin, C., Eichmann, R., Reitz, M.U., Hermann, S., Woolley-Allen, K., Brown, P.E., Wiwatdirekkul, K., Esteban, E., Pasha, A., Kogel, K.-H., Provart, N.J., Ott, S., Schäfer, P., 2020. **Regulation of cell type-specific immunity networks in** ***Arabidopsis*** **roots**. *The Plant Cell* **32**, 2742–2762. https://doi.org/10.1105/tpc.20.00154
Rico, A., Preston, G.M., 2008. ***Pseudomonas syringae*** **pv.** ***tomato*** **DC3000 uses constitutive and apoplast-induced nutrient assimilation pathways to catabolize nutrients that are abundant in the tomato apoplast**. *Molecular Plant-Microbe Interactions* 21, 269–282. https://doi.org/10.1094/MPMI-21-2-0269
Rieusset, L., Rey, M., Muller, D., Vacheron, J., Gerin, F., Dubost, A., Comte, G., Prigent-Combaret, C., 2020. **Secondary metabolites from plant-associated** ***Pseudomonas*** **are overproduced in biofilm**. *Microbial Biotechnology* **13**, 1562–1580. https://doi.org/10.1111/1751-7915.13598
Rillig, M.C., Muller, L.A.H., Lehmann, A., 2017. **Soil aggregates as massively concurrent evolutionary incubators**. *The ISME Journal* **11**, 1943–1948. https://doi.org/10.1038/ismej.2017.56
Rincon-Enriquez, G., Crété, P., Barras, F., Py, B., 2008. **Biogenesis of Fe/S proteins and pathogenicity: IscR plays a key role in allowing** ***Erwinia chrysanthemi*** **to adapt to hostile conditions**. *Molecular Microbiology* 67, 1257–1273. https://doi.org/10.1111/j.1365-2958.2008.06118.x
Roca, A., Rodríguez-Herva, J., Duque, E., Ramos, J.L., 2008. **Physiological responses of** ***Pseudomonas putida*** **to formaldehyde during detoxification**. *Microbial Biotechnology* **1**, 158–169. https://doi.org/10.1111/j.1751-7915.2007.00014.x

Roller, B.R.K., Stoddard, S.F., Schmidt, T.M., 2016. **Exploiting *rRNA* operon copy number to investigate bacterial reproductive strategies**. *Nature Microbiology* 1, 16160. https://doi.org/10.1038/nmicrobiol.2016.160

Romsang, A., Duang-Nkern, J., Saninjuk, K., Vattanaviboon, P., Mongkolsuk, S., 2018. ***Pseudomonas aeruginosa nfuA*: gene regulation and its physiological roles in sustaining growth under stress and anaerobic conditions and maintaining bacterial virulence**. *PloS One* 13, e0202151. https://doi.org/10.1371/journal.pone.0202151

Saarenpää, S., Shalev, O., Ashkenazy, H., de Oliveira-Carlos, V., Lundberg, D.S., Weigel, D., Giacomello, S., 2022. **Spatially resolved host-bacteria-fungi interactomes via spatial metatranscriptomics**. *bioRxiv* 2022.07.18.496977-2022.07.18.496977. https://doi.org/10.1101/2022.07.18.496977

Salas-González, I., Reyt, G., Flis, P., Custódio, V., Gopaulchan, D., Bakhoum, N., Dew, T.P., Suresh, K., Franke, R.B., Dangl, J.L., Salt, D.E., Castrillo, G., 2021. **Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis**. *Science* **371**, eabd0695. https://doi.org/10.1126/science.abd0695

Schindelin, J., Rueden, C.T., Hiner, M.C., Eliceiri, K.W., 2015. **The ImageJ ecosystem: an open platform for biomedical image analysis**. *Molecular Reproduction and Development* **82**, 518–529. https://doi.org/10.1002/mrd.22489

Schloss, P.D., 2021. **Amplicon sequence variants artificially split bacterial genomes into separate clusters**. *mSphere* e0019121. https://doi.org/10.1128/mSphere.00191-21

Schwengers, O., Jelonek, L., Dieckmann, M.A., Beyvers, S., Blom, J., Goesmann, A., 2021. **Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification**. *Microbial Genomics* **7**, 000685. https://doi.org/10.1099/mgen.0.000685

Seemann, T., 2014. **Prokka: rapid prokaryotic genome annotation**. *Bioinformatics (Oxford, England)* **30**, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Selten, G., Gómez-Repollés, A., Lamouche, F., Radutoiu, S., de Jonge, R., 2025. **SyFi: generating and using sequence fingerprints to distinguish SynCom isolates**. *Microbial Genomics* **11**, 001461. https://doi.org/10.1099/mgen.0.001461

Selten, G., Lamouche, F., Gómez-Repollés, A., Blahovska, Z., Kelly, S., Jonge, R. de, Radutoiu, S., 2024a. **Functional capacities drive recruitment of bacteria into plant root microbiota**. *bioRxiv* **2024–08**, 2024.08.22.609090. https://doi.org/10.1101/2024.08.22.609090

Selten, G., Stassen, M.J.J., De Rooij, P., Berendsen, R.L., Stringlis, I., de Jonge, R., 2024b. **Draft genome sequences of *Arabidopsis thaliana*-associated micro-organisms from Reijerscamp soil, the Netherlands**. https://doi.org/10.5281/zenodo.10992416

Seto, D., Koulena, N., Lo, T., Menna, A., Guttman, D.S., Desveaux, D., 2017. **Expanded type III effector recognition by the ZAR1 NLR protein using ZED1-related kinases**. *Nature Plants* **3**, 17027. https://doi.org/10.1038/nplants.2017.27

Seyedsayamdost, M.R., 2014. **High-throughput platform for the discovery of elicitors of silent bacterial gene clusters**. *Proceedings of the National Academy of Sciences* 111, 7266–7271. https://doi.org/10.1073/pnas.1400019111

Shalev, O., Karasov, T.L., Lundberg, D.S., Ashkenazy, H., Pramoj Na Ayutthaya, P., Weigel, D., 2022. **Commensal *Pseudomonas* strains facilitate protective response against pathogens in the host plant**. *Nature Ecology & Evolution* **6**, 383–396. https://doi.org/10.1038/s41559-022-01673-7

Shao, Z., Gu, S., Zhang, X., Xue, J., Yan, T., Guo, S., Pommier, T., Jousset, A., Yang, T., Xu, Y., Shen, Q., Wei, Z., 2024. **Siderophore interactions drive the ability of *Pseudomonas* spp. consortia to protect tomato against *Ralstonia solanacearum***. *Horticulture Research* **11**, uhae186.

https://doi.org/10.1093/hr/uhae186
Sharma-Kuinkel, B.K., Mann, E.E., Ahn, J.-S., Kuechenmeister, L.J., Dunman, P.M., Bayles, K.W., 2009. **The *Staphylococcus aureus* LytSR two-component regulatory system affects biofilm formation**. *Journal of Bacteriology* **191**, 4767–4775. https://doi.org/10.1128/JB.00348-09
Sharp, C., Foster, K.R., 2022. **Host control and the evolution of cooperation in host microbiomes**. *Nature Communications* **13**, 3567–3567. https://doi.org/10.1038/s41467-022-30971-8
Siracusa, L., Napoli, E., Ruberto, G., 2022. **Novel chemical and biological insights of inositol derivatives in Mediterranean plants**. *Molecules* 27, 1525. https://doi.org/10.3390/molecules27051525
Skindersoe, M.E., Alhede, M., Phipps, R., Yang, L., Jensen, P.O., Rasmussen, T.B., Bjarnsholt, T., Tolker-Nielsen, T., Høiby, N., Givskov, M., 2008. **Effects of antibiotics on quorum sensing in *Pseudomonas aeruginosa***. *Antimicrobial Agents and Chemotherapy* **52**, 3648–3663. https://doi.org/10.1128/AAC.01230-07
Slowikowski, K., 2022. **ggrepel: automatically position non-overlapping text labels with "ggplot2"**.
Song, C., van der Voort, M., van de Mortel, J., Hassan, K.A., Elbourne, L.D.H., Paulsen, I.T., Loper, J.E., Raaijmakers, J.M., 2015. **The Rsm regulon of plant growth-promoting *Pseudomonas fluorescens* SS101: role of small RNAs in regulation of lipopeptide biosynthesis**. *Microbial Biotechnology* **8**, 296–310. https://doi.org/10.1111/1751-7915.12190
Song, Y., Wilson, A.J., Zhang, X.-C., Thoms, D., Sohrabi, R., Song, S., Geissmann, Q., Liu, Y., Walgren, L., He, S.Y., Haney, C.H., 2021. **FERONIA restricts *Pseudomonas* in the rhizosphere microbiome via regulation of reactive oxygen species**. *Nature Plants* **7**, 644–654. https://doi.org/10.1038/s41477-021-00914-0
Sperandeo, P., Cescutti, R., Villa, R., Di Benedetto, C., Candia, D., Dehò, G., Polissi, A., 2007. **Characterization of *lptA* and *lptB*, two essential genes implicated in lipopolysaccharide transport to the outer membrane of *Escherichia coli***. *Journal of Bacteriology* 189, 244–253. https://doi.org/10.1128/JB.01126-06
Sprouffske, K., Wagner, A., 2016. **Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves**. *BMC Bioinformatics* 17, 172. https://doi.org/10.1186/s12859-016-1016-7
Stork, I., Gartemann, K.-H., Burger, A., Eichenlaub, R., 2008. **A family of serine proteases of *Clavibacter michiganensis* subsp. *michiganensis*: *chpC* plays a role in colonization of the host plant tomato**. *Molecular Plant Pathology* 9, 599–608. https://doi.org/10.1111/j.1364-3703.2008.00484.x
Stringlis, I.A., de Jonge, R., Pieterse, C.M.J., 2019. **The age of coumarins in plant–microbe interactions**. *Plant and Cell Physiology* **60**, 1405–1419. https://doi.org/10.1093/pcp/pcz076
Stringlis, I.A., Yu, K., Feussner, K., Jonge, R. de, Bentum, S.V., Verk, M.C.V., Berendsen, R.L., Bakker, P.A.H.M., Feussner, I., Pieterse, C.M.J., 2018. **MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health**. *Proceedings of the National Academy of Sciences* **115**, E5213–E5222. https://doi.org/10.1073/pnas.1722335115
Stutz, Emil W., Défago, Geneviève, Kern, Heinz, 1986. **Naturally occurring fluorescent pseudomonads involved in suppression of black root-rot of tobacco**. *American Phytopathological Society* 76(2), 181–185. https://doi.org/10.1094/Phyto-76-181
Sun, X., Xu, Z., Xie, J., Hesselberg-Thomsen, V., Tan, T., Zheng, D., Strube, M.L., Dragoš, A., Shen, Q., Zhang, R., Kovács, Á.T., 2022. ***Bacillus velezensis* stimulates resident rhizosphere *Pseudomonas stutzeri* for plant health through metabolic interactions**. *The ISME Journal* **16**, 774–787. https://doi.org/10.1038/s41396-021-01125-3
Taberman, H., Parkkinen, T., Rouvinen, J., 2016. **Structural and functional features of the NAD(P) dependent Gfo/Idh/MocA protein family oxidoreductases**. *Protein science* 25, 778–786. https://doi.org/10.1002/pro.2877

Tan, X., Calderon-Villalobos, L.I.A., Sharon, M., Zheng, C., Robinson, C.V., Estelle, M., Zheng, N., 2007. **Mechanism of auxin perception by the TIR1 ubiquitin ligase**. *Nature* **446**, 640–645. https://doi.org/10.1038/nature05731

Tecon, R., Or, D., 2016. **Bacterial flagellar motility on hydrated rough surfaces controlled by aqueous film thickness and connectedness**. *Scientific Reports* 6, 19409. https://doi.org/10.1038/srep19409

Teixeira, P.J.P.L., Colaianni, N.R., Law, T.F., Conway, J.M., Gilbert, S., Li, H., Salas-González, I., Panda, D., Del Risco, N.M., Finkel, O.M., Castrillo, G., Mieczkowski, P., Jones, C.D., Dangl, J.L., 2021. **Specific modulation of the root immune system by a community of commensal bacteria**. *Proceedings of the National Academy of Sciences of the United States of America* **118**, e2100678118. https://doi.org/10.1073/pnas.2100678118

Tenenbaum, D., Bioconductor Package Maintainer, 2022. **KEGGREST: client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG)**.

Tian, T., Sun, B., Shi, H., Gao, T., He, Y., Li, Y., Liu, Y., Li, X., Zhang, L., Li, S., Wang, Q., Chai, Y., 2021. **Sucrose triggers a novel signaling cascade promoting *Bacillus subtilis* rhizosphere colonization**. *The ISME Journal* **15**, 2723–2737. https://doi.org/10.1038/s41396-021-00966-2

Tilman, D., 2004. **Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly**. *Proceedings of the National Academy of Sciences* 101, 10854–10861. https://doi.org/10.1073/pnas.0403458101

Tkacz, A., Hortala, M., Poole, P.S., 2018. **Absolute quantitation of microbiota abundance in environmental samples**. *Microbiome* 6, 110. https://doi.org/10.1186/s40168-018-0491-7

Tognon, M., Köhler, T., Luscher, A., van Delden, C., 2019. **Transcriptional profiling of *Pseudomonas aeruginosa* and *Staphylococcus aureus* during *in vitro* co-culture**. *BMC Genomics* 20, 30. https://doi.org/10.1186/s12864-018-5398-y

Tsai, H.-H., Tang, Y., Jiang, L., Xu, X., Dénervaud Tendon, V., Pang, J., Jia, Y., Wippel, K., Vacheron, J., Keel, C., Andersen, T.G., Geldner, N., Zhou, F., 2025. **Localized glutamine leakage drives the spatial structure of root microbial colonization**. *Science* **390**, eadu4235. https://doi.org/10.1126/science.adu4235

Turner, B.L., Papházy, M.J., Haygarth, P.M., Mckelvie, I.D., 2002. **Inositol phosphates in the environment**. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**, 449–469. https://doi.org/10.1098/rstb.2001.0837

Tyc, O., van den Berg, M., Gerards, S., van Veen, J.A., Raaijmakers, J.M., de Boer, W., Garbeva, P., 2014. **Impact of interspecific interactions on antimicrobial activity among soil bacteria**. *Frontiers in Microbiology* **5**.

Tzipilevich, E., Russ, D., Dangl, J.L., Benfey, P.N., 2021. **Plant immune system activation is necessary for efficient root colonization by auxin-secreting beneficial bacteria**. *Cell Host & Microbe* 29, 1507-1520.e4. https://doi.org/10.1016/j.chom.2021.09.005

Vacheron, J., Heiman, C.M., Keel, C., 2021. **Live cell dynamics of production, explosive release and killing activity of phage tail-like weapons for *Pseudomonas* kin exclusion**. *Communications Biology* 4, 87–87. https://doi.org/10.1038/s42003-020-01581-1

Van Wees, S.C.M., Van Pelt, J.A., Bakker, P.A.H.M., Pieterse, C.M.J., 2013. **Bioassays for assessing jasmonate-dependent defenses triggered by pathogens, herbivorous insects, or beneficial rhizobacteria**, *in*: Goossens, A., Pauwels, L. (Eds.), Jasmonate Signaling: Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 35–49. https://doi.org/10.1007/978-1-62703-414-2_4

Vázquez, D., Civantos, C., Durán-Wendt, D., Ruiz, A., Rivilla, R., Martín, M., Bernal, P., 2025. **The *Pseudomonas putida* type VI secretion systems shape the tomato rhizosphere microbiota**. https://doi.org/10.1101/2025.08.14.670259

Venables, W.N., Ripley, B.D., 2002. **Modern Applied Statistics with S**, Statistics and Computing.

Springer, New York, NY.
https://doi.org/10.1007/978-0-387-21706-2

Verbon, E.H., Liberman, L.M., Zhou, J., Yin, J., Pieterse, C.M.J., Benfey, P.N., Stringlis, I.A., de Jonge, R., 2023. **Cell type-specific transcriptomics reveals that root hairs and endodermal barriers play important roles in beneficial plant-rhizobacterium-interactions**. *Molecular plant* 16, 1160–1177.
https://doi.org/10.1016/j.molp.2023.06.001

Veseli, I., Chen, Y.T., Schechter, M.S., Vanni, C., Fogarty, E.C., Watson, A.R., Jabri, B., Blekhman, R., Willis, A.D., Yu, M.K., Fernàndez-Guerra, A., Füssel, J., Eren, A.M., 2025. **Microbes with higher metabolic independence are enriched in human gut microbiomes under stress**. *eLife* **12**, RP89862.
https://doi.org/10.7554/eLife.89862

Vesga, P., Flury, P., Vacheron, J., Keel, C., Croll, D., Maurhofer, M., 2020. **Transcriptome plasticity underlying plant root colonization and insect invasion by *Pseudomonas protegens***. *The ISME Journal* **14**, 2766–2782.
https://doi.org/10.1038/s41396-020-0729-9

Větrovský, T., Baldrian, P., 2013. **The variability of the *16S rRNA* gene in bacterial genomes and its consequences for bacterial community analyses**. *PLOS ONE* 8, e57923.
https://doi.org/10.1371/journal.pone.0057923

Vílchez, J.I., Yang, Y., He, D., Zi, H., Peng, L., Lv, S., Kaushal, R., Wang, W., Huang, W., Liu, R., Lang, Z., Miki, D., Tang, K., Paré, P.W., Song, C.-P., Zhu, J.-K., Zhang, H., 2020. **DNA demethylases are required for *myo*-inositol-mediated mutualism between plants and beneficial rhizobacteria**. *Nature Plants* **6**, 983–995.
https://doi.org/10.1038/s41477-020-0707-2

Visaggio, D., Pasqua, M., Bonchi, C., Kaever, V., Visca, P., Imperi, F., 2015. **Cell aggregation promotes pyoverdine-dependent iron uptake and virulence in *Pseudomonas aeruginosa***. *Frontiers in Microbiology* **6**.

Vliet, S. van, Hauert, C., Fridberg, K., Ackermann, M., Co, A.D., 2022. **Global dynamics of microbial communities emerge from local interaction rules**. *PLOS Computational Biology* **18**, e1009877.
https://doi.org/10.1371/journal.pcbi.1009877

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G.A., 2013. **Micro-scale determinants of bacterial diversity in soil**. *FEMS Microbiology Reviews* **37**, 936–954.
https://doi.org/10.1111/1574-6976.12023

Waite, R.D., Papakonstantinopoulou, A., Littler, E., Curtis, M.A., 2005. **Transcriptome analysis of *Pseudomonas aeruginosa* growth: comparison of gene expression in planktonic cultures and developing and mature biofilms**. *Journal of Bacteriology* 187, 6571–6576.
https://doi.org/10.1128/JB.187.18.6571-6576.2005

Wang, D., Lee, S.-H., Seeve, C., Yu, J.M., Pierson, L.S., Pierson, E.A., 2013. **Roles of the Gac-Rsm pathway in the regulation of phenazine biosynthesis in *Pseudomonas chlororaphis* 30-84**. *MicrobiologyOpen* **2**, 505–524.
https://doi.org/10.1002/mbo3.90

Wang, D., Yu, J.M., Dorosky, R.J., Pierson, L.S., Pierson, E.A., 2016. **The phenazine 2-hydroxy-phenazine-1-carboxylic acid promotes extracellular DNA release and has broad transcriptomic consequences in *Pseudomonas chlororaphis* 30–84**. *PLoS ONE* 11, e0148003.
https://doi.org/10.1371/journal.pone.0148003

Wang, G., Roux, B., Feng, F., Guy, E., Li, L., Li, N., Zhang, X., Lautier, M., Jardinaud, M.-F., Chabannes, M., Arlat, M., Chen, S., He, C., Noël, L.D., Zhou, J.-M., 2015. **The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants**. *Cell Host & Microbe* **18**, 285–295.
https://doi.org/10.1016/j.chom.2015.08.004

Wang, H., Smith, A., Lovelace, A., Kvitko, B.H., 2022. **In planta transcriptomics reveals conflicts between pattern-triggered immunity and the AlgU sigma factor regulon**. *PLOS ONE* 17, e0274009.
https://doi.org/10.1371/journal.pone.0274009

Wang, H., Yang, Z., Swingle, B., Kvitko, B.H., 2021. **AlgU, a conserved sigma factor regulating abiotic stress tolerance and promoting virulence in *Pseudomonas syringae***. *Molecular Plant-Microbe Interactions*® 34, 326–336.

https://doi.org/10.1094/MPMI-09-20-0254-CR
Wang, J., Wang, Y., Lu, S., Lou, H., Wang, X., Wang, W., 2024. **AlgU mediates hyperosmotic tolerance in *Pseudomonas protegens* SN15-2 by regulating membrane stability, ROS scavenging, and osmolyte synthesis**. *Applied and Environmental Microbiology* 90, e00596-24.
https://doi.org/10.1128/aem.00596-24
Wang, L., Zhang, X., Tang, C., Li, P., Zhu, R., Sun, J., Zhang, Y., Cui, H., Ma, J., Song, X., Zhang, W., Gao, X., Luo, X., You, L., Chen, Y., Dai, Z., 2022. **Engineering consortia by polymeric microbial swarmbots**. *Nature Communications* **13**, 3879–3879.
https://doi.org/10.1038/s41467-022-31467-1
Wang, N.R., Wiesmann, C.L., Melnyk, R.A., Hossain, S.S., Chi, M.-H., Martens, K., Craven, K., Haney, C.H., 2022. **Commensal *Pseudomonas fluorescens* strains protect *Arabidopsis* from closely related *Pseudomonas* pathogens in a colonization-dependent manner**. *mBio* 13, e02892-21.
https://doi.org/10.1128/mbio.02892-21
Wang, Z., Chen, Z., Kowalchuk, G.A., Xu, Z., Fu, X., Kuramae, E.E., 2021. **Succession of the resident soil microbial community in response to periodic inoculations**. *Applied and Environmental Microbiology* **87**, e00046-21.
https://doi.org/10.1128/AEM.00046-21
Weber, M., Fuchs, T.M., 2022. **Metabolism in the niche: a large-scale genome-based survey reveals inositol utilization to be widespread among soil, commensal, and pathogenic bacteria**. *Microbiology Spectrum* 10, e02013-22.
https://doi.org/10.1128/spectrum.02013-22
Wei, Z., Gu, Y., Friman, V.-P., Kowalchuk, G.A., Xu, Y., Shen, Q., Jousset, A., 2019. **Initial soil microbiome composition and functioning predetermine future plant health**. *Science Advances* **5**, eaaw0759.
https://doi.org/10.1126/sciadv.aaw0759
Wickham, H., 2019. **assertthat: easy pre and post assertions**.
Wickham, H., 2016. **ggplot2: elegant graphics for data analysis**. Springer-Verlag New York.
Wickham, H., 2007. **Reshaping data with the reshape package**. *Journal of Statistical Software* **21**, 1–20.
https://doi.org/10.18637/jss.v021.i12
Wickham, H., Bryan, J., 2022. **readxl: read Excel files**.
Wickham, H., François, R., Henry, L., Müller, K., 2022. **dplyr: a grammar of data manipulation**.
Wickham, H., RStudio, 2022. **stringr: simple, consistent wrappers for common string operations**.
Wickham, H., Seidel, D., 2022. **scales: scale functions for visualization**.
Wickham, H., Vaughan, D., Girlich, M., 2023. **tidyr: tidy messy data**.
Wickramasinghe, N.N., Ravensdale, J., Coorey, R., Dykes, G.A., Chandry, P.S., 2021. **Transcriptional profiling of biofilms formed on chilled beef by psychrotrophic meat spoilage bacterium, *Pseudomonas fragi* 1793**. *Biofilm* 3, 100045.
https://doi.org/10.1016/j.bioflm.2021.100045
Wiesmann, C.L., Zhang, Y., Alford, M., Hamilton, C.D., Dosanjh, M., Thoms, D., Dostert, M., Wilson, A., Pletzer, D., Hancock, R.E.W., Haney, C.H., 2022. **The ColR/S two-component system is a conserved determinant of host association across *Pseudomonas* species**. *The ISME Journal* 1–11.
https://doi.org/10.1038/s41396-022-01343-3
Wilke, C., Wiernik, B., 2022. **ggtext: improved text rendering support for "ggplot2"**.
Wilkinson, S., 2018. **kmer: an R package for fast alignment-free clustering of biological sequences**.
Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A.V., Elias, D.A., 2019. **Soil aggregate microbial communities: towards understanding microbiome interactions at biologically relevant scales**. *Applied and Environmental Microbiology* **85**, e00324-19.
https://doi.org/10.1128/AEM.00324-19
Winker, S., Woese, C.R., 1991. **A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics**. *Systematic and Applied Microbiology* 14, 305–310.
https://doi.org/10.1016/S0723-2020(11)80303-6

Wippel, K., Tao, K., Niu, Y., Zgadzaj, R., Kiel, N., Guan, R., Dahms, E., Zhang, P., Jensen, D.B., Logemann, E., Radutoiu, S., Schulze-Lefert, P., Garrido-Oter, R., 2021. **Host preference and invasiveness of commensal bacteria in the *Lotus* and *Arabidopsis* root microbiota**. *Nature Microbiology* **6**, 1150–1162. https://doi.org/10.1038/s41564-021-00941-9

Wu, W., Du, K., Kang, X., Wei, H., 2021. **The diverse roles of cytokinins in regulating leaf development**. *Horticulture Research* **8**, 118. https://doi.org/10.1038/s41438-021-00558-3

Xia, T., Wang, Q., 2006. **D-*chiro*-Inositol found in *Cucurbita ficifolia* (Cucurbitaceae) fruit extracts plays the hypoglycaemic role in streptozocin-diabetic rats**. *Journal of Pharmacy and Pharmacology* 58, 1527–1532. https://doi.org/10.1211/jpp.58.10.0014

Xiao, N., 2018. **ggsci: scientific journal and sci-fi themed color palettes for "ggplot2"**.

Xiao, Y., Chen, H., Nie, L., He, M., Peng, Q., Zhu, W., Nie, H., Chen, W., Huang, Q., 2021. **Identification of c-di-GMP/FleQ-regulated new target genes, including *cyaA*, encoding adenylate cyclase, in *Pseudomonas putida***. *mSystems* 6, e00295-21. https://doi.org/10.1128/mSystems.00295-21

Xu, L., Dong, Z., Chiniquy, D., Pierroz, G., Deng, S., Gao, C., Diamond, S., Simmons, T., Wipf, H.M.-L., Caddell, D., Varoquaux, N., Madera, M.A., Hutmacher, R., Deutschbauer, A., Dahlberg, J.A., Guerinot, M.L., Purdom, E., Banfield, J.F., Taylor, J.W., Lemaux, P.G., Coleman-Derr, D., 2021. **Genome-resolved metagenomics reveals role of iron metabolism in drought-induced rhizosphere microbiome dynamics**. *Nature Communications* **12**, 3209. https://doi.org/10.1038/s41467-021-23553-7

Xu, S., Liu, Y.-X., Cernava, T., Wang, H., Zhou, Y., Xia, T., Cao, S., Berg, G., Shen, X.-X., Wen, Z., Li, C., Qu, B., Ruan, H., Chai, Y., Zhou, X., Ma, Z., Shi, Y., Yu, Y., Bai, Y., Chen, Y., 2022. ***Fusarium* fruiting body microbiome member *Pantoea agglomerans* inhibits fungal pathogenesis by targeting lipid rafts**. *Nature Microbiology* **7**, 831–843. https://doi.org/10.1038/s41564-022-01131-x

Xu, X., Li, H., Qi, X., Chen, Y., Qin, Y., Zheng, J., Jiang, X., 2021. ***cheA*, *cheB*, *cheR*, *cheV*, and *cheY* are involved in regulating the adhesion of *Vibrio harveyi***. *Frontiers in Cellular and Infection Microbiology* **10**. https://doi.org/10.3389/fcimb.2020.591751

Xu, Z., Escamilla-Treviño, L., Zeng, L., Lalgondar, M., Bevan, D., Winkel, B., Mohamed, A., Cheng, C.-L., Shih, M.-C., Poulton, J., Esen, A., 2004. **Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1**. *Plant Molecular Biology* **55**, 343–367. https://doi.org/10.1007/s11103-004-0790-1

Yang, B., Minne, M., Brunoni, F., Plačková, L., Petřík, I., Sun, Y., Nolf, J., Smet, W., Verstaen, K., Wendrich, J.R., Eekhout, T., Hoyerová, K., Van Isterdael, G., Haustraete, J., Bishopp, A., Farcot, E., Novák, O., Saeys, Y., De Rybel, B., 2021. **Non-cell autonomous and spatiotemporal signaling from a tissue organizer orchestrates root vascular development**. *Nature plants* **7**, 1485–1494. https://doi.org/10.1038/s41477-021-01017-6

Yang, C., Yue, H., Ma, Z., Feng, Z., Feng, H., Zhao, L., Zhang, Y., Deakin, G., Xu, X., Zhu, H., Wei, F., 2022. **Influence of plant genotype and soil on the cotton rhizosphere microbiome**. *Frontiers in Microbiology* **13**. https://doi.org/10.3389/fmicb.2022.1021064

Yang, Y., Chen, S., Wu, X., Peng, L., Vílchez, J.I., Kaushal, R., Liu, X., Singh, S.K., He, D., Yuan, F., Lv, S., Morcillo, R.J.L., Wang, W., Huang, W., Lei, M., Zhu, J.-K., Paré, P.W., Zhang, H., 2022. **Plant latent defense response to microbial non-pathogenic factors antagonizes compatibility**. *National Science Review* **9**, nwac109. https://doi.org/10.1093/nsr/nwac109

Yoneyama, Kaori, Xie, X., Nomura, T., Yoneyama, Koichi, 2020. **Do phosphate and cytokinin interact to regulate strigolactone biosynthesis or act independently?** *Frontiers in Plant Science* 11, 438. https://doi.org/10.3389/fpls.2020.00438

Yu, K., Liu, Y., Tichelaar, R., Savant, N., Lagendijk, E., van Kuijk, S.J.L., Stringlis, I.A., van Dijken, A.J.H., Pieterse, C.M.J., Bakker, P.A.H.M., Haney, C.H., Berendsen, R.L., 2019. **Rhizosphere-associated *Pseudomonas* suppress local root immune responses by gluconic acid-mediated lowering of environmental pH**. *Current Biology* **29**, 3913-3920.e4.

https://doi.org/10.1016/j.cub.2019.09.015
Yu, X., Lund, S.P., Scott, R.A., Greenwald, J.W., Records, A.H., Nettleton, D., Lindow, S.E., Gross, D.C., Beattie, G.A., 2013. **Transcriptional responses of *Pseudomonas syringae* to growth in epiphytic versus apoplastic leaf sites**. *Proceedings of the National Academy of Sciences* **110**, E425–E434.
https://doi.org/10.1073/pnas.1221892110
Yuan, B., Li, C., Wang, Q., Yao, Q., Guo, X., Zhang, Y., Wang, Z., 2025. **Identification and functional characterization of the *RPP13* gene family in potato (*Solanum tuberosum* L.) for disease resistance**. *Frontiers in Plant Science* 15.
https://doi.org/10.3389/fpls.2024.1515060
Zboralski, A., Filion, M., 2020. **Genetic factors involved in rhizosphere colonization by phytobeneficial *Pseudomonas* spp.** *Computational and Structural Biotechnology Journal* **18**, 3539–3554.
https://doi.org/10.1016/j.csbj.2020.11.025
Zhang, H., 2023. **Plant latent defense response against compatibility**. *The ISME Journal* 17, 787–791.
https://doi.org/10.1038/s41396-023-01399-9
Zhang, N., Zhu, X., Tao, X., Li, J., Tang, Q., Liu, X., Luo, L.-M., Zhang, P., Zhang, L.-Q., He, Y.-X., Ge, H., 2025. **Interspecies signaling modulates the biosynthesis of antimicrobial secondary metabolites related to biological control activities of *Pseudomonas fluorescens* 2P24**. *Microbiology Spectrum* 13, e01886-24.
https://doi.org/10.1128/spectrum.01886-24
Zhang, X., Dippold, M.A., Kuzyakov, Y., Razavi, B.S., 2019. **Spatial pattern of enzyme activities depends on root exudate composition**. *Soil Biology and Biochemistry* **133**, 83–93.
https://doi.org/10.1016/j.soilbio.2019.02.010
Zhang, Y., Fu, M., Wang, Q., Zhang, Lei, Chang, X., Zhang, Lixin, 2024. **Role of the sigma factor AlgU in regulating growth, virulence, motility, exopolysaccharide production, and environmental stress adaptation of *Pseudomonas syringae* pv. *actinidiae* QSY6**. *Phytopathology Research* 6, 26.
https://doi.org/10.1186/s42483-024-00245-w
Zhao, T., Vink, S.N., Jia, X., Erban, A., Schaarschmidt, S., Kopka, J., Zuther, E., Treder, K., Michałowska, D., Guyoneaud, R., Elzenga, J.T.M., Attard, E., Salles, J.F., 2025. **Unveiling potato cultivars with microbiome interactive traits for sustainable agricultural production**. *Plant, Cell & Environment*.
https://doi.org/10.1111/pce.70019
Zheng, H., Chatfield, C.H., Liles, M.R., Cianciotto, N.P., 2013. **Secreted pyomelanin of *Legionella pneumophila* promotes bacterial iron uptake and growth under iron-limiting conditions**. *Infection and Immunity* 81, 4182–4191.
https://doi.org/10.1128/iai.00858-13
Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D.M., Evans, E.D., Dai, C.L., Lu, P.J., Alm, E.J., Weitz, D.A., 2022. **High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome**. *Science* **376**, eabm1483.
https://doi.org/10.1126/science.abm1483
Zheng, Y., Cao, X., Zhou, Y., Ma, S., Wang, Y., Li, Z., Zhao, D., Yang, Y., Zhang, H., Meng, C., Xie, Z., Sui, X., Xu, K., Li, Y., Zhang, C.-S., 2024. **Purines enrich root-associated *Pseudomonas* and improve wild soybean growth under salt stress**. *Nature Communications* **15**, 3520.
https://doi.org/10.1038/s41467-024-47773-9
Zhou, F., Emonet, A., Dénervaud Tendon, V., Marhavy, P., Wu, D., Lahaye, T., Geldner, N., 2020. **Co-incidence of damage and microbial patterns controls localized immune responses in roots**. *Cell* **180**, 440-453.e18.
https://doi.org/10.1016/j.cell.2020.01.013
Zhou, J., Uribe Acosta, M., Stassen, M.J.J., Qi, R., de Jonge, R., White, F., Kramer, G., Dong, L., Pieterse, C.M.J., Stringlis, I.A., 2025. ***Arabidopsis* root defense barriers support beneficial interactions with rhizobacterium *Pseudomonas simiae* WCS417**. *The New Phytologist* **248**, 2021–2039.
https://doi.org/10.1111/nph.70549
Zielezinski, A., Vinga, S., Almeida, J., Karlowski, W.M., 2017. **Alignment-free sequence comparison: benefits, applications, and tools**. *Genome Biology* **18**, 186.
https://doi.org/10.1186/s13059-017-1319-7

Zuniga-Soto, E., Fitzpatrick, D.A., Doohan, F.M., Mullins, E., 2019. **Insights into the transcriptomic response of the plant engineering bacterium *Ensifer adhaerens* OV14 during transformation**. *Scientific Reports* 9, 10344. https://doi.org/10.1038/s41598-019-44648-8

SUMMARY
ENGLISH

Bioinoculant success as an ecological problem

Modern agriculture increasingly requires biological solutions to enhance crop productivity and resilience while reducing reliance on synthetic inputs. The widespread use of mineral fertilizers and pesticides has contributed to soil degradation, biodiversity loss, and environmental pollution, while also offering limited robustness against increasingly variable climatic conditions. In this context, microbial bioinoculants – beneficial microorganisms applied to plants or soils – have emerged as a promising strategy, supported by extensive experimental evidence demonstrating their capacity to promote plant growth, improve nutrient acquisition, and enhance resistance to biotic and abiotic stress. However, despite this apparent potential, bioinoculant performance remains highly inconsistent in natural soils and field conditions, and even well-characterised strains with reproducible effects in controlled environments often fail to deliver reliable benefits under realistic agricultural settings.

This inconsistency stems from the fact that bioinoculation is fundamentally an ecological process. Introduced microbes must establish, persist, and function within soils that are already densely populated by resident microbial communities. The rhizosphere, in particular, is a highly competitive and dynamic environment shaped by fluctuating resources, microbial interactions, and active host regulation. As a result, bioinoculant success depends not only on intrinsic genetic traits, but also on the capacity of the microbe to sense, respond to, and remain compatible with local biotic and abiotic conditions.

Crucially, plants do not interact with individual microbes in isolation, but rather shape and are shaped by entire microbial communities. Through immune surveillance, metabolic filtering, and spatial structuring of the root environment, plant hosts exert strong selective pressures on microbial colonisation and activity. In this sense, the resident microbiome itself acts as a key regulator of microbial success, constraining which organisms can establish and which microbial functions are expressed. These interactions generate feedback loops in which microbial behaviour both responds to and influences host physiological state, leading to outcomes that can vary substantially among individual plants, even under otherwise uniform conditions.

Framing bioinoculant performance as an ecological invasion problem therefore provides a conceptual frame for understanding why beneficial effects are often conditional and context-dependent. Rather than reflecting inconsistent potential, variable outcomes emerge from differences in host physiology, microbial community structure, and local environmental constraints. As reviewed in Chapter 1, this ecological perspective sets the stage for the experimental chapters of this thesis, which investigate how genetic determinants, physiological plasticity, and ecological context jointly shape the success or failure of root-associated bacteria.

Transcriptional heterogeneity across microbial colonisation events

Chapter 2 of this thesis addresses bioinoculant variability at the level of individual plant–microbe interactions. Using *Arabidopsis thaliana* grown in natural soil, inoculation with the well-characterised plant growth-promoting bacterium *Pseudomonas simiae* WCS417 (hereafter WCS417) resulted in a reproducible yet heterogeneous plant population. While the overall population benefited, individual plants diverged into distinct phenotypic classes ranging from strong growth promotion to no benefit or even reduced growth.

By combining strain-resolved metagenomics and metatranscriptomics, this chapter shows that WCS417 adopts distinct transcriptional programs that correlate with host phenotype. In bulk soil, WCS417 predominantly expresses stress- and competition-related genes, consistent with a survival-oriented state. In the rhizosphere of small, poorly performing plants, the bacterium exhibits signatures of envelope stress and limited metabolic engagement. In contrast, in the rhizosphere of big plants, WCS417 transcriptional activity suggests a growth-active and exploratory state characterised by high expression of ribosomal genes, respiratory pathways, and motility functions.

Importantly, plant transcription mirrors these microbial states. Small plants show markers of nutrient limitation and oxidative stress, whereas big plants exhibit transcriptional profiles consistent with a more permissive and resource-rich rhizosphere. Together, these results reveal that bioinoculant efficacy is conditional, emerging only in specific host–microbe configurations, and that microbial function is best understood as a dynamic physiological response rather than a fixed trait.

Genetic determinants of rhizosphere competence

In Chapter 3, the focus shifts from variability among individual colonisation events to variability observed among different bioinoculant strains. Specifically, this chapter asks which microbial traits are associated with enhanced rhizosphere competence, defined here as the ability to reach and maintain higher population densities on plant roots in natural soil. To address this question, we compared a group of *Pseudomonas* isolates for their ability to colonise *Arabidopsis* roots and performed genomic comparisons across strains exhibiting different levels of colonisation.

This analysis identifies the conserved *iol* gene cluster as a genetic feature consistently enriched in the most competent root colonisers. Experimental assays show that the *iol* locus contributes to bacterial performance during rhizosphere colonisation. Loss of the locus reduces competitive fitness in soil, indicating that *iol*$^{+}$ strains are better able to establish and maintain populations in the rhizosphere. In addition, *iol*$^{+}$ strains exhibit phenotypes

relevant to colonisation dynamics, including enhanced swimming motility and higher production of fluorescent siderophores in response to inositol, whereas Δ*iol* mutant cells display reduced motility and altered physiological responses under the same conditions. Together, these results link the presence of the *iol* locus to behavioural traits associated with movement, exploration, and population expansion in heterogeneous environments.

Beyond targeted assays, analyses of genome-wide fitness and expression datasets reveal that *iol* genes contribute to bacterial performance across a range of host-associated and stress-related conditions. Fitness contributions associated with the locus are observed across multiple environments and are not restricted to growth on inositol alone. Consistent with this, *iol* gene expression varies across environments and is detected during colonisation of plant and insect hosts, but not in non-colonisation contexts, supporting a role for the locus in host-associated growth.

Inferring ecological units from amplicon data

Understanding microbial ecology ultimately depends on how biological units are defined and represented in community data. Ecological interactions, population dynamics, and host–microbe associations are all inferred at the level of discrete microbial entities, yet most high-throughput surveys rely on marker-gene amplicon sequencing that does not directly observe organisms. Chapter 4 addresses this foundational issue by examining how meaningful microbial units can be inferred from *16S rRNA* amplicon sequencing data.

Amplicon datasets capture sequence variants rather than organisms, and intragenomic variation, shared variants across related taxa, and technical artefacts often decouple sequence similarity from biological identity. As a result, commonly used OTU-based representations can obscure the ecological structure of microbial communities and complicate downstream ecological interpretation.

To address this problem, this chapter introduces ASVNet, a theoretical framework that integrates sequence similarity with co-abundance patterns across samples to infer empirical operational taxonomic units (eOTUs). Rather than treating sequence variants as independent entities or clustering them solely by similarity, ASVNet uses sample-to-sample covariation as ecological information to identify groups of variants that behave as coherent biological units.

Benchmarking using synthetic communities and complex soil datasets shows that ASVNet reconstructs microbial clusters that more closely correspond to underlying biological entities than conventional similarity-based clustering. By leveraging ecological structure embedded in abundance variation, ASVNet captures organism-level signals that are otherwise fragmented or lost, enabling more biologically grounded ecological inference from amplicon sequencing data.

General synthesis and implications

This thesis advances a unifying view of bioinoculant performance as an emergent property of genetic potential, physiological plasticity, and ecological context. It shows that beneficial microbes do not act uniformly across hosts or environments, but instead occupy distinct physiological states shaped by local conditions and host feedbacks. Traits such as growth promotion, colonisation, and persistence are therefore conditional outcomes rather than intrinsic guarantees.

By integrating molecular biology, microbial ecology, and computational biology, this work argues that improving bioinoculant reliability will require moving beyond static trait screening toward context-aware diagnostics that capture microbial activity *in situ*. More broadly, the thesis highlights the importance of studying microbes as dynamic ecological agents embedded in complex systems, providing conceptual and practical foundations for the development of more predictable and resilient microbiome-based solutions in agriculture.

SAMENVATTING
NEDERLANDS

Succes van bioinoculanten als een ecologisch probleem

De moderne landbouw heeft in toenemende mate behoefte aan biologische oplossingen om de gewasproductiviteit en -weerbaarheid te vergroten, terwijl de afhankelijkheid van synthetische inputs wordt verminderd. Het wijdverbreide gebruik van minerale meststoffen en bestrijdingsmiddelen heeft bijgedragen aan bodemdegradatie, verlies aan biodiversiteit en milieubelasting, terwijl het tegelijkertijd slechts beperkte robuustheid biedt onder steeds variabelere klimatologische omstandigheden. In deze context zijn microbiële bioinoculanten – gunstige micro-organismen die worden toegepast op planten of bodems – naar voren gekomen als een veelbelovende strategie, ondersteund door uitgebreid experimenteel bewijs dat hun vermogen aantoont om plantengroei te stimuleren, nutriëntenopname te verbeteren en weerstand tegen biotische en abiotische stress te verhogen. Ondanks dit ogenschijnlijke potentieel blijft de prestatie van bioinoculanten echter sterk inconsistent in natuurlijke bodems en veldomstandigheden, en zelfs goed gekarakteriseerde stammen met reproduceerbare effecten in gecontroleerde omgevingen slagen er vaak niet in om betrouwbare voordelen te leveren onder realistische landbouwomstandigheden.

Deze inconsistentie vloeit voort uit het feit dat bio-inoculatie in wezen een ecologisch proces is. Geïntroduceerde micro-organismen moeten zich vestigen, handhaven en functioneren binnen bodems die reeds dichtbevolkt zijn door residente microbiële gemeenschappen. Met name de rhizosfeer vormt een sterk competitieve en dynamische omgeving, gekenmerkt door fluctuerende hulpbronnen, microbiële interacties en actieve regulatie door de gastheer. Het succes van een bioinoculant hangt daardoor niet alleen af van intrinsieke genetische eigenschappen, maar ook van het vermogen van het micro-organisme om lokale biotische en abiotische omstandigheden waar te nemen, erop te reageren en ermee compatibel te blijven.

Cruciaal is dat planten niet in isolatie interageren met afzonderlijke microorganismen, maar dat zij volledige microbiële gemeenschappen vormen en erdoor worden gevormd. Via immuunsurveillance, metabole filtering en ruimtelijke structurering van de wortelomgeving oefenen planten sterke selectieve druk uit op microbiële kolonisatie en activiteit. In die zin fungeert het residente microbioom zelf als een belangrijke regulator van microbiëel succes, doordat het bepaalt welke organismen zich kunnen vestigen en welke microbiële functies tot expressie komen. Deze interacties genereren terugkoppelingslussen waarin microbiëel gedrag zowel reageert op als invloed uitoefent op de fysiologische toestand van de gastheer, wat leidt tot uitkomsten die sterk kunnen variëren tussen individuele planten, zelfs onder verder uniforme omstandigheden.

Het benaderen van bioinoculantprestaties als een probleem van ecologische invasie biedt daarom een conceptueel kader om te begrijpen waarom

gunstige effecten vaak conditioneel en contextafhankelijk zijn. In plaats van een weerspiegeling te zijn van inconsistente microbiële potentie, ontstaan variabele uitkomsten uit verschillen in gastheerfysiologie, microbiële gemeenschapsstructuur en lokale omgevingscondities. Zoals besproken in Hoofdstuk 1 vormt dit ecologische perspectief het vertrekpunt voor de experimentele hoofdstukken van dit proefschrift, waarin wordt onderzocht hoe genetische determinanten, fysiologische plasticiteit en ecologische context gezamenlijk het succes of falen van wortelgeassocieerde bacteriën bepalen.

Transcriptionele heterogeniteit tijdens microbiële kolonisatie-gebeurtenissen

Hoofdstuk 2 van dit proefschrift behandelt variatie in bioinoculantprestaties op het niveau van individuele plant–micro-organisme-interacties. Door *Arabidopsis thaliana* te kweken in natuurlijke bodem en deze te inoculeren met de goed gekarakteriseerde plantengroeibevorderende bacterie *Pseudomonas simiae* WCS417 (hierna WCS417), werd een reproduceerbare maar heterogene plantenpopulatie waargenomen. Hoewel de populatie als geheel profiteerde van inoculatie, divergeerden individuele planten naar onderscheiden fenotypische klassen, variërend van sterke groeibevordering tot geen effect of zelfs verminderde groei.

Door het combineren van stam-specifieke metagenomica en metatranscriptomica laat dit hoofdstuk zien dat WCS417 verschillende transcriptionele programma's aanneemt die correleren met het plantfenotype. In bulkbodem brengt WCS417 voornamelijk genen tot expressie die gerelateerd zijn aan stress en competitie, wat consistent is met een overlevingsgerichte toestand. In de rhizosfeer van kleine, slecht presterende planten vertoont de bacterie kenmerken van envelopstress en beperkte metabole activiteit. Daarentegen suggereert de transcriptionele activiteit van WCS417 in de rhizosfeer van grote planten een groei-actieve en exploratieve toestand, gekenmerkt door hoge expressie van ribosomale genen, respiratoire routes en motiliteitsfuncties.

Belangrijk is dat de planttranscriptie deze microbiële toestanden weerspiegelt. Kleine planten vertonen markers van nutriëntenbeperking en oxidatieve stress, terwijl grote planten transcriptionele profielen laten zien die wijzen op een meer permissieve en hulpbronrijke rhizosfeer. Gezamenlijk tonen deze resultaten aan dat de effectiviteit van bioinoculanten conditioneel is en slechts tot stand komt binnen specifieke configuraties van gastheer en micro-organisme, en dat microbiële functie het best begrepen kan worden als een dynamische fysiologische respons in plaats van als een vaste eigenschap.

Genetische determinanten van rhizosfeercompetentie

In Hoofdstuk 3 verschuift de focus van variatie tussen individuele kolonisatiegebeurtenissen naar variatie tussen verschillende bioinoculantstammen. Dit hoofdstuk onderzoekt welke microbiële eigenschappen geassocieerd zijn met verhoogde rhizosfeercompetentie, hier gedefinieerd als het vermogen om hogere populatiedichtheden te bereiken en te behouden op plantenwortels in natuurlijke bodem. Om deze vraag te beantwoorden vergeleken we een groep *Pseudomonas*-isolaten op hun vermogen om *Arabidopsis*-wortels te koloniseren, en voerden we genomische vergelijkingen uit tussen stammen met uiteenlopende kolonisatieniveaus.

Deze analyse identificeert het geconserveerde *iol*-gencluster als een genetische eigenschap die consistent verrijkt is in de meest competente wortelkolonisatoren. Experimentele assays tonen aan dat het *iol*-locus bijdraagt aan bacteriële prestaties tijdens rhizosfeerkolonisatie. Verlies van het locus verlaagt de competitieve fitness in bodem, wat erop wijst dat *iol*$^+$-stammen beter in staat zijn om populaties in de rhizosfeer te vestigen en te handhaven. Daarnaast vertonen *iol*$^+$-stammen fenotypen die relevant zijn voor kolonisatie, waaronder verhoogde zwemmotiliteit en een hogere productie van fluorescerende sideroforen in reactie op inositol, terwijl Δ*iol*-mutanten een verminderde motiliteit en veranderde fysiologische responsen vertonen onder dezelfde omstandigheden. Samen koppelen deze resultaten de aanwezigheid van het iol-locus aan gedragskenmerken die samenhangen met beweging, exploratie en populatie-expansie in heterogene omgevingen.

Naast gerichte experimenten laten analyses van genoom-brede fitness- en expressiedatasets zien dat *iol*-genen bijdragen aan bacteriële prestaties onder uiteenlopende gastheer-geassocieerde en stressgerelateerde omstandigheden. Fitnessbijdragen van het locus worden waargenomen in meerdere omgevingen en zijn niet beperkt tot groei op inositol alleen. In overeenstemming hiermee varieert de expressie van *iol*-genen tussen omgevingen en wordt deze waargenomen tijdens kolonisatie van planten- en insectengastheren, maar niet in niet-koloniserende contexten, wat wijst op een preferentiële inzet tijdens gastheer-geassocieerde groei.

Het afleiden van ecologische eenheden uit amplicondata

Het begrijpen van microbiële ecologie hangt uiteindelijk af van hoe biologische eenheden worden gedefinieerd en gerepresenteerd in gemeenschapsdata. Ecologische interacties, populatiedynamiek en gastheer–micro-organisme-associaties worden allemaal geïnterpreteerd op het niveau van discrete microbiële entiteiten, terwijl de meeste grootschalige

surveys gebaseerd zijn op marker-gen-ampliconsequentieanalyse die organismen niet direct observeert. Hoofdstuk 4 behandelt deze fundamentele kwestie door te onderzoeken hoe betekenisvolle microbiële eenheden kunnen worden afgeleid uit *16S rRNA*-ampliconsequentiedata.

Amplicondatasets leggen sequentievarianten vast in plaats van organismen, en intragenomische variatie, gedeelde varianten tussen verwante taxa en technische artefacten kunnen de relatie tussen sequentiegelijkenis en biologische identiteit verbreken. Als gevolg daarvan kunnen veelgebruikte OTU-gebaseerde representaties de ecologische structuur van microbiële gemeenschappen verhullen en de ecologische interpretatie bemoeilijken.

Om dit probleem aan te pakken introduceert dit hoofdstuk ASVNet, een theoretisch raamwerk dat sequentiegelijkenis combineert met co-abundantiepatronen over monsters heen om empirische operationele taxonomische eenheden (eOTU's) af te leiden. In plaats van sequentievarianten als onafhankelijke entiteiten te behandelen of ze uitsluitend op basis van gelijkenis te clusteren, gebruikt ASVNet covariatie tussen monsters als ecologische informatie om groepen varianten te identificeren die zich gedragen als coherente biologische eenheden.

Benchmarking met synthetische gemeenschappen en complexe bodemdatasets laat zien dat ASVNet microbiële clusters reconstrueert die beter overeenkomen met onderliggende biologische entiteiten dan conventionele gelijkenisgebaseerde clustering. Door gebruik te maken van ecologische structuur die besloten ligt in abundantievariatie, herwint ASVNet signalen op organismaal niveau die anders gefragmenteerd of verloren zouden gaan, wat leidt tot een biologisch beter gefundeerde ecologische interpretatie van ampliconsequentiedata.

Algemene synthese en implicaties

Dit proefschrift presenteert een samenhangende visie op bioinoculantprestaties als een emergente eigenschap van genetisch potentieel, fysiologische plasticiteit en ecologische context. Het laat zien dat gunstige micro-organismen niet uniform functioneren over gastheren of omgevingen heen, maar uiteenlopende fysiologische toestanden aannemen die worden gevormd door lokale omstandigheden en gastheer-gerelateerde terugkoppelingen. Eigenschappen zoals groeibevordering, kolonisatie en persistentie zijn daarom conditionele uitkomsten en geen intrinsieke garanties.

Door moleculaire biologie, microbiële ecologie en computationele biologie te integreren, betoogt dit werk dat het verbeteren van de betrouwbaarheid van bioinoculanten vraagt om een verschuiving van statische traitsscreening naar contextbewuste benaderingen die microbiële activiteit *in situ* vastleggen. In bredere zin onderstreept dit proefschrift het belang van het bestuderen van micro-organismen als dynamische ecologische actoren die ingebed zijn in complexe systemen, en biedt het zowel conceptuele als

praktische grondslagen voor de ontwikkeling van meer voorspelbare en robuuste microbiome-gebaseerde oplossingen in de landbouw.

RESUMEN

CASTELLANO

El éxito de los bioinoculantes como un problema ecológico

La agricultura moderna requiere cada vez más soluciones biológicas que permitan mejorar la productividad y la resiliencia de los cultivos, reduciendo al mismo tiempo la dependencia de compuestos sintéticos. El uso generalizado de fertilizantes minerales y pesticidas ha contribuido a la degradación del suelo, la pérdida de biodiversidad y la contaminación ambiental, además de ofrecer una robustez limitada frente a condiciones climáticas cada vez más variables. En este contexto, los bioinoculantes microbianos – microorganismos beneficiosos aplicados a plantas o suelos – han surgido como una estrategia prometedora, respaldada por una amplia evidencia experimental que demuestra su capacidad para promover el crecimiento vegetal, mejorar la adquisición de nutrientes y aumentar la resistencia frente a estreses bióticos y abióticos. Sin embargo, a pesar de este potencial aparente, el rendimiento de los bioinoculantes sigue siendo altamente inconsistente en suelos naturales y en condiciones de campo, y aun cepas bien caracterizadas con efectos reproducibles en entornos controlados suelen fracasar a la hora de proporcionar beneficios fiables en contextos agrícolas realistas.

Esta inconsistencia se debe a que la bioinoculación es, en esencia, un proceso ecológico altamente complejo. Los microorganismos introducidos deben establecerse, persistir y funcionar en suelos que ya están densamente poblados por comunidades microbianas residentes. La rizosfera, en particular, constituye un entorno altamente competitivo y dinámico, moldeado por la disponibilidad fluctuante de recursos, las interacciones microbianas y una regulación activa por parte del organismo huésped. Como consecuencia, el éxito de un bioinoculante depende no solo de rasgos genéticos intrínsecos, sino también de la capacidad del microorganismo para percibir, responder y mantenerse compatible con las condiciones bióticas y abióticas locales.

De manera crucial, las plantas no interactúan con microorganismos individuales de forma aislada, sino que moldean y son moldeadas por comunidades microbianas completas. A través de mecanismos de vigilancia inmunitaria, filtrado metabólico y estructuración espacial del entorno radicular, las plantas ejercen fuertes presiones selectivas sobre la colonización y la actividad microbiana. En este sentido, el microbioma residente actúa como un regulador clave del éxito microbiano, limitando qué organismos pueden establecerse y qué funciones microbianas llegan a expresarse. Estas interacciones generan bucles de retroalimentación en los que el comportamiento microbiano responde al estado fisiológico del huésped y, a su vez, lo influye, dando lugar a resultados que pueden variar sustancialmente entre plantas individuales incluso bajo condiciones aparentemente uniformes.

Plantear el rendimiento de los bioinoculantes como un problema de invasión ecológica proporciona, por tanto, una base conceptual para comprender

por qué los efectos beneficiosos suelen ser condicionales y dependientes del contexto. En lugar de reflejar un potencial microbiano inconsistente, los resultados variables emergen de diferencias en la fisiología del hospedador, la estructura de la comunidad microbiana y las restricciones ambientales locales. Tal como se revisa en el Capítulo 1, esta perspectiva ecológica sienta las bases de los capítulos experimentales de esta tesis, que investigan cómo los determinantes genéticos, la plasticidad fisiológica y el contexto ecológico configuran conjuntamente el éxito o el fracaso de bacterias asociadas a la raíz.

Heterogeneidad transcripcional entre diferentes eventos de colonización microbiana

El Capítulo 2 de esta tesis aborda la variabilidad de los bioinoculantes a escala de interacciones individuales planta–microorganismo. Utilizando *Arabidopsis thaliana* cultivada en suelo natural, la inoculación con la bien caracterizada bacteria promotora del crecimiento vegetal *Pseudomonas simiae* WCS417 (en adelante, WCS417) dio lugar a una población vegetal reproducible pero heterogénea. Aunque la población en su conjunto se benefició de la inoculación, las plantas individuales divergieron en clases fenotípicas distintas, que abarcaron desde una fuerte promoción del crecimiento hasta la ausencia de beneficio o incluso una reducción del mismo.

Mediante la combinación de metagenómica y metatranscriptómica con resolución a nivel de cepa, este capítulo muestra que WCS417 adopta programas transcripcionales diferenciados que se correlacionan con el fenotipo del hospedador. En el suelo, WCS417 expresa predominantemente genes asociados al estrés y a la competencia, coherentes con un estado orientado a la supervivencia. En la rizosfera de plantas pequeñas y con bajo rendimiento, la bacteria presenta firmas de estrés de la envoltura celular y una implicación metabólica limitada. En contraste, en la rizosfera de plantas grandes, la actividad transcripcional de WCS417 sugiere un estado activo y exploratorio, caracterizado por una elevada expresión de genes ribosomales, vías respiratorias y funciones relacionadas con la motilidad.

De forma importante, la transcripción vegetal refleja estos estados microbianos. Las plantas pequeñas muestran marcadores de limitación nutricional y estrés oxidativo, mientras que las plantas grandes presentan perfiles transcripcionales compatibles con una rizosfera más permisiva y rica en recursos. En conjunto, estos resultados revelan que la eficacia de los bioinoculantes es condicional, emergiendo únicamente en configuraciones específicas huésped–microorganismo, y que la función microbiana se comprende mejor como una respuesta fisiológica dinámica, más que como un rasgo fijo.

Determinantes genéticos de la competencia rizosférica

En el Capítulo 3, el foco se desplaza desde la variabilidad entre eventos individuales de colonización hacia la variabilidad observada entre distintas cepas bioinoculantes. En concreto, este capítulo se pregunta qué rasgos microbianos se asocian con una mayor competencia rizosférica, definida aquí como la capacidad de alcanzar y mantener densidades poblacionales más elevadas en raíces vegetales en suelo natural. Para abordar esta cuestión, comparamos un conjunto de aislados de *Pseudomonas* en su capacidad para colonizar raíces de *Arabidopsis* y realizamos comparaciones genómicas entre cepas que presentaban distintos niveles de colonización.

Este análisis identifica el clúster génico *iol* como una característica genética consistentemente enriquecida en los colonizadores radiculares más competentes. Ensayos experimentales muestran que el locus *iol* contribuye al rendimiento bacteriano durante la colonización de la rizosfera. La pérdida del locus reduce la aptitud competitiva en suelo, lo que indica que las cepas *iol*$^{+}$ son más capaces de establecer y mantener poblaciones en la rizosfera. Además, las cepas *iol*$^{+}$ presentan fenotipos relevantes para la dinámica de colonización, incluyendo una mayor motilidad por natación y una mayor producción de sideróforos fluorescentes en respuesta al inositol, mientras que las células mutantes Δ*iol* muestran una motilidad reducida y respuestas fisiológicas alteradas bajo las mismas condiciones. En conjunto, estos resultados vinculan la presencia del locus *iol* con rasgos comportamentales asociados al movimiento, la exploración y la expansión poblacional en entornos heterogéneos.

Más allá de los ensayos dirigidos, el análisis de conjuntos de datos genómicos de aptitud y expresión revela que los genes *iol* contribuyen al rendimiento bacteriano en una amplia variedad de condiciones asociadas al hospedador y al estrés. Las contribuciones de aptitud asociadas al locus se observan en múltiples entornos y no se restringen al crecimiento sobre inositol. De forma coherente con ello, la expresión de los genes *iol* varía entre entornos y se detecta durante la colonización de hospedadores vegetales e insectos, pero no en contextos de no colonización, lo que indica que el locus se activa preferentemente durante el crecimiento asociado al hospedador.

Inferencia de unidades ecológicas a partir de datos de secuenciación de amplicones

La comprensión de la ecología microbiana depende en última instancia de cómo se definen y representan las unidades biológicas en los datos de comunidades. Las interacciones ecológicas, la dinámica poblacional y las asociaciones huésped–microorganismo se infieren a nivel de entidades

microbianas discretas. Sin embargo, la mayoría de los estudios de alto rendimiento se basan en la secuenciación de amplicones de genes marcadores, que no observa directamente los organismos. El Capítulo 4 aborda esta cuestión fundamental examinando cómo pueden inferirse unidades microbianas significativas a partir de datos de secuenciación de amplicones del gen de ARNr *16S*.

Los conjuntos de datos de amplicones capturan variantes de secuencia en lugar de organismos, y la variación intragenómica, las variantes compartidas entre taxones relacionados y los artefactos técnicos suelen desacoplar la similitud de secuencia de la identidad biológica. Como resultado, las representaciones basadas en OTUs de uso común pueden ocultar la estructura ecológica de las comunidades microbianas y complicar la interpretación biológica posterior.

Para abordar este problema, este capítulo introduce ASVNet, un marco teórico que integra la similitud de secuencia con patrones de coabundancia entre muestras para inferir unidades taxonómicas operativas empíricas (eOTUs). En lugar de tratar las variantes de secuencia como entidades independientes o agruparlas únicamente por similitud, ASVNet utiliza la covariación entre muestras como información ecológica para identificar grupos de variantes que se comportan como unidades biológicas coherentes.

La evaluación mediante comunidades sintéticas y conjuntos de datos complejos de suelo muestra que ASVNet reconstruye agrupaciones microbianas que se corresponden de forma más estrecha con entidades biológicas subyacentes que los enfoques convencionales basados en similitud. Al aprovechar la estructura ecológica incorporada en la variación de abundancias, ASVNet recupera señales a nivel de organismo que, de otro modo, quedarían fragmentadas o perdidas, permitiendo una inferencia ecológica más sólidamente fundamentada a partir de datos de secuenciación de amplicones.

Síntesis general e implicaciones

En conjunto, esta tesis avanza una visión unificadora del rendimiento de los bioinoculantes como una propiedad emergente del potencial genético, la plasticidad fisiológica y el contexto ecológico. Muestra que los microorganismos beneficiosos no actúan de forma uniforme entre hospedadores o entornos, sino que ocupan estados fisiológicos diferenciados moldeados por las condiciones locales y las retroalimentaciones asociadas al hospedador. Rasgos como la promoción del crecimiento, la colonización y la persistencia emergen, por tanto, como resultados condicionales más que como garantías intrínsecas.

Mediante la integración de biología molecular, ecología microbiana y biología computacional, este trabajo sostiene que mejorar la fiabilidad de los bioinoculantes requerirá ir más allá de la selección estática de rasgos,

avanzando hacia enfoques sensibles al contexto que capturen la actividad microbiana *in situ*. De manera más amplia, la tesis subraya la importancia de estudiar a los microorganismos como agentes ecológicos dinámicos integrados en sistemas complejos, y proporciona bases conceptuales y prácticas para el desarrollo de estrategias microbioma-basadas más predecibles y resilientes en agricultura.

ACKNOWLEDGEMENTS

THANK YOU

One.

Two.

Three.

Four.

Five.

Six.

Seven.

Eight. Eight f*cking years, ladies and gentlemen. What. A. Roller. Coaster. Qué hartura, niño.

Although it has become a cliché by now, nobody expected to have to reinvent a PhD halfway because of a pandemic. Being abroad, isolated and already deep into the PhD was anything but easy. We had to abandon plans, put ongoing ideas on hold, focus on the data we already had and invent projects that could be developed from home, all without knowing when, or whether, normality would return. The idea behind ASVNet actually emerged in an online meeting somewhere in the middle of that chaos, and only 200 years later I finally have a version ready to publish.

PMI has marked a before and after in my life. There have been highs, lows and more lessons than I ever asked for. A lot of things have happened in my time in this team. It is hard to explain exactly what has changed in me since I started. All I can say is that the person who arrived in the Netherlands felt about 15 years old, and that person is now approximately 85. More than professionally, this journey has hit me more on the personal level. The collision of similar-enough-but-different cultures, expectation crashes, the PhD itself, the pandemic again, and the simple fact (but not a minor one) of growing older in the middle of that storm has left me like an old lady whose wig has been blown off by the wind. Or, at this point, like the wig itself. During this time, I forgot who I was and rebuilt myself again, gained confidence, lost it again, win it back, manage situations and complex emotions, faced my own issues and, ultimately, grow as a person. I am very thankful for this journey.

But if you are in, say, your second year and suspect it might take you eight, quit immediately. Pretty please, don't take this long, there's no need to do that to yourself. Start a business near the beach. Open a restaurant. Launch a foot OnlyFans. Do whatever you need to do, but run away and never look back.

And now, after a bit of whining, let's proceed with the actual acknowledgements, shall we? I hope I have not forgotten anyone, but if I have accidentally left anyone behind, I apologise. Over the years, many people have left a part of themselves in this journey and in me, and I really want to thank all of you for having taken part in my story.

First of all. Thank you, Ronnie, Corné and Bas, for trusting me and giving me this opportunity. I still remember receiving inviting me for the interview and immediately thinking: “wait, are these people any sane? Why the hell would they want me to go there?”. Ronnie, I couldn’t have hoped for a better supervisor than you. I truly think you are a great scientist, and I am honoured to have learned next to you. No matter how desperate I was about the results, you always found what was worth seeing in them and helped me understand what they were actually saying. Thank you for being so supportive and an example to follow. But I don’t think there is any need to sleep together in a conference again. Don’t get me wrong, you’re great, but I am sure we can both agree that separate rooms are better for everyone involved, right?

Sanne, doesn’t it seem like yesterday when we first saw each other in the general meeting room? How many things have happened since then? Insane. It’s been a pleasure to share this journey together with you. We have both been through so much, professionally and personally, and we have always been there to listen to and understand each other. And do you remember that strange tension at the beginning? Hahaha so silly of us. I am so happy that I could share this process with you. I loved having your critical thinking and ecological perspective beside me throughout this process. You challenged the way I approached things and made me grow enormously, and I am deeply grateful for that.

Giannis, getting to know you was an inflection point in this journey. You helped me so much, directly and indirectly. You gave me confidence and created a space in which I felt more at home, welcomed and confident, something I needed so much in that moment. I really thank you for who you are, for being inspiring and for your life philosophy. You are a great scientist, but most importantly a great person. Also, thanks to you I started talking to Alberto, my longest-running bromance so far. Alberto, ’o mio Alberto... e che t’aggia dicere, duttò d’ ’o core mio, ca nun t’aggio già ditto? You are one of the people who made me considerably less likely to jump in front of a train. That’s actually quite (b)romantic, I guess. We clicked from the very beginning, and I always knew I had someone I could trust and be completely myself with. You are an ameising person, but please get a salary ASAP. I am emotionally unprepared to visit you under a bridge while you sell your body and crack.

Max, my partner in crime in microbial genomics. I felt a special connection with you from the beginning, and I enjoy every conversation we have. You’re sweet, critical and frighteningly smart. I hope I get to watch you continue growing in academia and fully become the great scientist you already are. The world needs people like you around. Shu, I know you don’t get along with the pipetting robot, but give it a chance one day. It might eventually become like having a second Shu pipetting next to you. I love how you always make things work and thrive in any circumstance, always calm and confident. Guys, I wish your mini-Shu (thankfully a mini-Shu for Shu’s peace of mind) brings you buckets of happiness in your life. I promised I would sow nice baby clothes, but since I still have pants from Melissa and shirts from Alberto, both

likely from 4-5 years ago, you should probably not trust me much...

Brandon, after so many marriage proposals and rejections, I am beginning to suspect that you may not actually want to marry me. If that is truly how you feel, please have the courage to say it and stop playing these cruel games with my heart. In any case, I want you to know how much I enjoy talking to you. You are such a cultured gentleman, so cute and so genuinely good that you make everybody around you feel good. Thanks for being such a good person. And please keep considering my proposal.

María, thanks to our (extremely) toxic relationship I have recovered some faith in my badly-atrophied southern social skills. I know that, at some point in the future, our emotional codependency on each other will inevitably end with one of us annihilating the other. But while we wait to discover who passionately kills whom and who ultimately perishes from a lethal love, I intend to fun as much fun as possible with you. Or, who knows, perhaps this telenovela will end with both of us either (1) confined to a sanatorium, throwing our faeces at the staff in the middle of a sensual and "histernic" twerk, or (2) drinking poison together in a deeply Shakespearean/Lorquian manner. In any case, big fat drama will be served. Of that I am 1000% certain. Diagonal drama. Diagonally. Σ' αγαπώ, πανέμορφη!

Roeland, thank you very much for trusting me for the potato team. I know it has been a bit bumpy sometimes, but we managed to have something very good to show to the world. Yang, it was a pleasure to share office with you at the beginning and later work together on your potato project. I admire your steadiness and resilient temperament, always pushing forward with a nice word and a smile. You are great. David, I love the motivation and dedication you bring to everything you do. Your journey has not been an easy one, but you are strong, hardworking and constantly growing. I am very happy that your work is finally bearing fruit and taking you closer to the thesis you deserve. Afke, it is a pleasure to work with someone like you. Afke, it is a pleasure to work with someone like you. You are a deeply interesting person with such a rich inner world, and I feel I could trust you with anything.

There are so many more people I want to thank that this section could go on forever. Unfortunately, both my time and my remaining will to live are limited, and all I want now is to get rid of this damn thing. I love my current office with Sietske, Sophie, Stijn and Darío. Stijn, it's very fun to talk to you, I like your chill humour and I enjoy our conversations a lot. Sophie, I love your cheerful attitude, you're so creative and a wonderful person. It's so fun to be around you. Sietske, what can I say? We have known each other almost from the moment our paths here began. I love how calm and positive you are, and of course your nerdy humour with just the right pinch of darkness. I'm very happy to have shared this experience with you! Fortunately, not everyone I like is trapped in the same office. Jochem, I always appreciated your cheerful, positive attitude and the determination you showed from the very beginning. Jelle, I enjoyed a lot sharing the office with you and the journeys to Ecuador and Granada. I love how smart you are, your politically-

incorrect humour and the confidence you always show. Thank you for all the moments. Hao, you were the very first person I met after my interview. You were so nice that you made me feel, unfortunately, like this was a good place with nice people. Thank you for all the laughters and conversations!

Melissa, muchísimas gracias por todos los momentos divertidos y por la ayuda que me diste cuando lo necesitaba. Tienes una inteligencia emocional y una capacidad para dar cariño que está fuera de lo normal. Y paciencia, mucha paciencia para aguantarme jajaja. Te dije en su día que te admiraba y lo vuelvo a repetir.

Darío, eres posiblemente una de las personas más inteligentes que conozco. Siento que siempre que hablo contigo aprendo algo nuevo, y cualquier conversación siempre se siente interesante, viva y divertida. Creo que todos sabemos que vas a llegar muy alto, y espero estar ahí para ver cómo te conviertes en el mejor PI del mundo.

Noelia... Mira, con que sobrevivas un par de añitos más, yo ya me daré por satisfecho. Me encanta que seamos tan parecidos en tantas cosas. Somos ese tipo de persona que fuma en ayunas, con la mirada perdida en la telaraña del techo, preguntándose si el Señor tendrá la decencia de llevársela pronto. O a lo mejor no, igual la dramática soy yo. Disfruto mucho estando contigo, creo que eres una persona muy interesante con mucho mundo interior y, aunque vayas a sufrir un poquito, me alegro de que te quedes aquí más tiempo.

Julia, bichaca mía de mi corasssón! Apareciste estando yo medio moñeco, hecho un trapo de mocos, y me diste muchísimo apoyo y cariño. Eres súper fuerte e inteligente y tienes un corazón tan grande que preocupa. Deberías hacértelo mirar. Muchísimas gracias por ser quien eres y por todo lo que me has dado en las poquitas veces que hemos encontrado tiempo para hablar. Eres mu grande, coño! Y espero que se te quite un día eso de no querer ser PI, que hace muuucha faltita gente como tú. Y después me contratas a mí, que no quiero complicaciones.

Natalia, necesitas a Jesús. En serio, no estás bien de lo tuyo. Y mira que te lo dice el que mascaría mierda de gato seca colgado bocabajo de los pies. No sé de dónde sacas tanta bondad y ganas de reír, pero es maravilloso estar cerca de ti. Me alegras muchísimos días grises y sé que algún día podremos vernos fuera del maldito Kruyt, te lo prometo. ¡Te quiero, guapi!

Y por supuesto que no me he olvidado de mis queridísimas hadas madrinas, María José y Coral. Nada de esto habría sido posible sin vosotras. Muchísimas gracias por creer en mí y por vuestra confianza, aunque yo haya sido un maldito descastado en todos estos años. Sois muy grandes, y os estaré siempre agradecido por haberme dado el empujón para que este libro esté aquí.

Mamá, papá, gracias por haberme apoyado siempre en todas mis decisiones. Siempre habéis creído y confiado en mí, me habéis aceptado como soy y nunca me habéis hecho sentir que tuviera que elegir un camino u otro. Me

dejasteis crecer y elegir a mi manera, pero permanecisteis siempre a mi lado, acompañándome y compartiendo conmigo todos los momentos importantes. Sois unos padres maravillosos. Angy, si no hubieras estado tú ahí muchas cosas habrían sido muy diferentes. Desde siempre hemos estado el uno para el otro y nos hemos apoyado mutuamente. En los últimos tiempos han pasado muchas cosas, pero eso solo demuestra otra vez que eres una de las personas más fuertes y buenas que conozco, porque ser fuerte o valiente no significa no tener miedo, sino enfrentarte a él y seguir viviendo. Al final, como ya hablamos, aunque nuestros caminos sean de cada uno y diferentes, siento que en el fondo somos la misma persona que intenta atravesar el mundo de dos maneras distintas. Te apoyaré siempre y, aunque tengamos que estar separados, siempre estaré contigo. Leo y Karolina, espero que, aunque nunca os leáis este libro (yo igual tampoco lo haría), podáis tenerlo como ejemplo de que con esfuerzo y tiempo podéis conseguir lo que os propongáis. Sed siempre quienes queráis ser sin importar lo que os digan, intentad ser buenas personas, quereos mucho y haced lo que os haga felices. Pero portaos bien, que sois unos petardos. A los cinco, sé que es difícil que estemos lejos. Hace 15 años que me fui de Algeciras para buscar mi propio camino. El viaje no ha sido fácil y ha tenido muchos baches y valles, pero al mismo tiempo me alegra que podáis estar leyendo estas palabras en los agradecimientos de mi tesis doctoral. Siempre os he tenido y os tengo muy presentes y espero que, en un futuro no muy lejano, podamos estar todos más cerca y pertenecer algo más a vuestro día a día. Este libro de aquí también es, por supuesto, un logro vuestro. Gracias por estar siempre ahí apoyándome. Os quiero muchísimo.

Julia, Luis y Vanesa, vosotros también tenéis mucho que ver con esto. Desde la primera vez que fui a Carcabuey y os conocí me incluisteis directamente, y desde entonces me habéis querido como a uno más de la familia. Sois unas personas maravillosas y me encanta estar con vosotros. Os quiero.

Y, por supuesto, millones de gracias a ti, gordito. Todavía recuerdo esa primera conversación en la que hablamos de la posibilidad de venirnos los dos a vivir a Holanda. Tenía tantísimo miedo de que eso nos fuera a separar... Pero aquí estamos, ocho años después, juntos los dos contra viento y marea. Han pasado miles de cosas en este tiempo. Hemos crecido juntos, hemos atravesado altibajos, nos hemos sostenido cuando las cosas se complicaban y hemos construido algo muy fuerte entre los dos. Gracias por quererme tanto y por apoyarme siempre de forma incondicional. Eres el mejor novio que se puede desear. Y encima ahora eres jefazo de PostNL, solo te falta que te entallen la camisa una poquita. Te quiero mucho, gordito. Pero quiero un perro, no te olvides ♥.

CURRICULUM VITAE

My name is Juan José Sánchez Gil, and I was born in 1993 in Algeciras, on the southern coast of Spain. I have always been deeply curious and eager to learn, with interests ranging from literature, languages, and philosophy to dance, music, art, mathematics, physics, and biology. After completing secondary school, I moved to Granada in 2011 to pursue a Bachelor's degree in Biology. I soon developed a strong interest in computational biology, a field that combined my enjoyment of solving complex problems with creativity.

During the final year of my degree, I joined the Genetics Department under the supervision of Professor Miguel Burgos Poyatos. There, I worked on the design of a synthetic interaction between *Saccharomyces boulardii* and *Escherichia coli* aimed at detecting and neutralising the TcdB toxin produced by *Clostridioides difficile*, using a BioBrick-based system.

In 2015 I began the Master's degree in Biotechnology at the University of Granada. Under the supervision of Professor Antonio Osuna, I designed and constructed a CRISPRi system targeting the expression of the MASP52 protein in *Trypanosoma cruzi*, the causative agent of Chagas disease.

After completing my studies, my professional path briefly took me beyond academic research. I worked in a bar, at a real estate company, and later as a technician at the Animal Breeding and Experimentation Service of the University of Granada.

In 2018, I moved to the Netherlands to begin my PhD in the Plant–Microbe Interactions group at Utrecht University, under the supervision of Dr Ronnie de Jonge. My doctoral research focused on the genetic and physiological determinants of bacterial ecological competence in the rhizosphere, as well as on computational approaches for studying root-associated microbial communities.

Since 2023, I have worked as a postdoctoral researcher in Dr Roeland Berendsen's team at Utrecht University, where I investigate plant-associated microbiomes as a computational biologist.



LIST OF PUBLICATIONS

Sánchez-Gil, J.J., Jonge, R. de, 2025. **ASVNet: Inferring microbes from *16S rRNA* amplicon sequencing data**. *bioRxiv* 2025.11.21.689717. https://doi.org/10.1101/2025.11.21.689717

Song, Y., Atza, E., **Sánchez-Gil, J.J.**, Akkermans, D., de Jonge, R., de Rooij, P.G.H., Kakembo, D., Bakker, P.A.H.M., Pieterse, C.M.J., Budko, N.V., Berendsen, R.L., 2025. **Seed tuber microbiome can predict growth potential of potato varieties**. *Nature Microbiology* 10, 28–40. https://doi.org/10.1038/s41564-024-01872-x

Poppeliers, S.W.M., **Sánchez-Gil, J.J.**, López, J.L., Dutilh, B.E., Pieterse, C.M.J., Jonge, R. de, 2024. **High-resolution quantification of the rhizosphere effect along a soil-to-root gradient shows selection-driven convergence of rhizosphere microbiomes**. *bioRxiv* 2024.06.21.600027. https://doi.org/10.1101/2024.06.21.600027

Fourie, A., Lopez, J.L., **Sánchez-Gil, J.J.**, Poppeliers, S.W.M., Jonge, R. de, Dutilh, B.E., 2024. **Bacterial family-specific enrichment and functions of secretion systems in the rhizosphere**. *bioRxiv* 2024.05.07.592589. https://doi.org/10.1101/2024.05.07.592589

Sánchez-Gil, J.J., Poppeliers, S.W.M., Vacheron, J., Zhang, H., Odijk, B., Keel, C., Jonge, R. de, 2023. **The conserved *iol* gene cluster in *Pseudomonas* is involved in rhizosphere competence**. *Current Biology* 33, 3097-3110.e6. https://doi.org/10.1016/j.cub.2023.05.057

Sánchez-Gil, J.J., Poppeliers, S.W., de Jonge, R., 2023. **Microbes to support plant health: understanding bioinoculant success in complex conditions**. *Current Opinion in Microbiology* 73, 102286. https://doi.org/10.1016/j.mib.2023.102286

López, J.L., Fourie, A., Poppeliers, S.W.M., Pappas, N., **Sánchez-Gil, J.J.**, de Jonge, R., Dutilh, B.E., 2023. **Growth rate is a dominant factor predicting the rhizosphere effect**. *The ISME Journal* 17, 1396–1405. https://doi.org/10.1038/s41396-023-01453-6

Bakker, P.A.H.M., Berendsen, R.L., Pelt, J.A.V., Vismans, G., Yu, K., Li, E., Bentum, S.V., Poppeliers, S.W.M., **Sánchez-Gil, J.J.**, Zhang, H., Goossens, P., Stringlis, I.A., Song, Y., Jonge, R. de, Pieterse, C.M.J., 2020. **The soil-borne identity and microbiome-assisted agriculture: looking back to the future**. *Molecular Plant* 13, 1394–1401. https://doi.org/10.1016/j.molp.2020.09.01

"Si os dan papel pautado,
escribid por el otro lado"

Juan Ramón Jiménez?

Modern agriculture increasingly requires biological solutions to enhance crop productivity and resilience while reducing reliance on synthetic inputs. In this context, microbial bioinoculants – beneficial microorganisms that are applied to plants or soils – have emerged as a promising strategy, supported by extensive experimental evidence demonstrating their capacity to promote plant growth, improve nutrient acquisition, and enhance resistance to biotic and abiotic stress. However, despite this apparent potential, bioinoculant performance remains highly inconsistent in natural soils and field conditions, and even well-characterised strains with reproducible effects in controlled environments often fail to deliver reliable benefits under realistic agricultural settings.

This inconsistency stems from the fact that bioinoculation is fundamentally an ecological process. Introduced microbes must establish, persist, and function within the rhizosphere – the zone surrounding and including plant roots. The rhizosphere is a densely populated, highly competitive and dynamic environment shaped by fluctuating resources, microbial interactions, and active host regulation. As a result, bioinoculant success depends not only on intrinsic genetic traits, but also on the capacity of the microbe to sense, respond to, and remain compatible with local biotic and abiotic conditions.

Framing bioinoculant performance as an ecological invasion problem provides a conceptual foundation for understanding why beneficial effects are often conditional and context-dependent. Rather than reflecting inconsistent microbial potential, variable outcomes emerge from differences in host physiology, microbial community structure, and local environmental constraints. This ecological perspective sets the stage for the experimental chapters of this thesis, which investigate how genetic determinants, physiological plasticity, and ecological context jointly shape the success or failure of root-associated bacteria.

