The effect of plant domestication on the seed microbiome

Riccardo Soldan

Wolfson College

University of Oxford

Supervised by

Professor Gail Margaret Preston

Thesis submitted for the degree of Doctor of Phylosophy

December 2

Abstract

There is a growing body of literature demonstrating the role of microbial communities in plant growth, health, and resilience. It is becoming increasingly clear that advances in our understanding of plant-microbe interactions could make a significant difference in sustainable agricultural production.

In my thesis, we approach this ever-evolving topic to try to advance our understanding of plant-microbe interactions within the agricultural ecosystem. One key question that has attracted recent research efforts is whether plant domestication has influenced host-microbiome assembly and composition.

In Chapter 2, we develop an evolutionary framework explaining that domesticated plants could showcase a reduction in positive microbe-to-host effects as a result of plant phenotypic traits selected by the domesticator through artificial selection. In Chapter 3, we provide evidence of the effect of domesticated plant traits on seed microbiome composition and assembly, partially validating the framework developed in Chapter 2. In Chapter 4, we approach the plant and its microbiome from a broader perspective and reflect on the consequences of plant domestication from a community evolution point of view. In Chapter 5, we develop a tool that will allow researchers to study bacterial plant colonization and how bacteria could eventually be transmitted to the seed compartment. Overall, we show that domesticated plant traits have consistent effects on seed microbiome composition and assembly, independently of the domestication event. We also demonstrate that monitoring bacterial localization in plants can be effectively achieved by combining bacterial bioluminescence with fluorescence.

Acknowledgments

I would like to thank my supervisor Gail Margaret Preston for the support and trust she gave me; my wife for her understanding, patience, and advice; my family for always being there for me when I needed them; my friends, old and new, for making my life way more enjoyable and finally my funding bodies (BBSRC, DTC, and Zegna), which allowed me to embark into my Ph.D. journey.

Table of Contents

Chapter 1: General introduction and thesis outline

Understanding how plants showcase certain phenotypic traits has been one of the first key questions in genetics. Since the pioneering work of Gregor Johann Mendel, our understanding of the genetic basis of plant phenotypes has certainly expanded (Diamond, 2002). Nonetheless, only in recent years have we come to appreciate the importance of microorganisms in influencing the manifestation of plant phenotypes.

Historically, a plant phenotype was thought to be the result of an interaction between the environment and the plant genotype. We are now aware that the plant-associated microbiome also plays a role in determining the plant phenotype, interacting with the environment, and the plant genotype (Henry *et al.*, 2021). Confounding the influence of the microbiome in the manifestation of a plant phenotype with the host genotype seems increasingly inadequate (Friesen *et al.*, 2011; Stappenbeck and Virgin, 2016). The host-associated microbiome could influence plant phenotype manifestation by increasing host phenotypic variance or shifting the phenotypic mean of the population (Henry *et al.*, 2021). The results of such contributions can lead to changes in the evolutionary trajectories of the host. It is increasingly clear that a better understanding of plant-microbe interactions could unlock the potential of the microbiome to promote plant growth and protect plants from pathogens. This requires i) the conceptualization of evolutionary frameworks that create meaningful hypotheses that can support the development of research projects, ii) a better understanding of factors that drive microbiome assembly, and iii) the development of tools to visualize spatiotemporal relationships between plants and microbes. In my view, these are three important keys needed to unlock the potential of the plant microbiome, which could contribute to a more sustainable agricultural production.

Three important keys to understanding plant-microbe interactions

1. Evolutionary frameworks

Evolutionary frameworks can provide meaningful hypotheses that direct the development of research projects.

For example, understanding why and how the host-associated microbiome would influence the manifestation of certain plant phenotypic traits could help plant breeders to include the microbiome in breeding schemes (Henry *et al.*, 2021). Moreover, we mainly rely on crop plants that have been domesticated and grown in agricultural ecosystems. Therefore, as we will see in the next paragraphs, evolutionary frameworks could help us contextualize plant-microbe interactions in non-natural environments. This is important because we still do not fully understand the consequences of agriculture for the mutualistic interactions between plants and microbes.

2. Factors driving host-microbiome assembly and composition

In recent years, the development of Next Generation Sequencing (NGS) techniques has allowed us to embrace microbial communities' complexity in an unprecedented way. This technology has led to a better understanding of factors driving microbial community assembly and composition. We have come to appreciate that the host-associated microbiome is influenced by multiple factors, namely: i) the environment in which the plant is grown. For example, environment-specific microbial communities can be found on grapes (Gilbert *et al.*, 2014), indicating the presence of a "microbiome *terroir*"; ii) the plant genotype and phenotype. For instance, the rhizosphere microbiome of *Phaseolus*

vulgaris is influenced by the plant genotype but also by the plant trait root length (Pérez-Jaramillo *et al.*, 2017); and iii) the ability of microorganisms to exploit plant-compartment-specific resources. For example, the root endosphere contains a subselection of the microbial communities inhabiting the rhizosphere in *Arabidopsis thaliana* (Bulgarelli *et al.*, 2012) and the seed microbiome contains bacterial species harboring coding sequences only found associated with seed-transmitting bacterial strains (Chesneau *et al.*). Additionally, the environmental effect (*e.g* location where the plant is grown) can have different effects on microbial communities depending on the plant compartment. For example, the environment has a stronger effect on root-associated microbial communities than on phyllosphere microbiota (Wagner *et al.*, 2016), indicating how microbiome composition in different plant compartments can be differentially affected by the same factor. Practical applications of this knowledge are constantly increasing and this is reflected in the growing market of microbial inoculants (Santos *et al.*, 2019).

3. Visualizing plant-microbe interactions

NGS techniques often provide limited insight into the spatio-temporal aspects of host-microbe and microbe-microbe interactions. This is because such methods typically homogenise a limited number of biological samples to isolate macromolecules or to enumerate microbial cells, potentially obscuring variation in microbial physiology at the macro and micro-scale. For these reasons, visualization techniques remain an important and complementary approach to NGS study to better study the temporal and spatial distribution of microorganisms (Soldan, Sanguankiattichai, *et al.*, 2021). This is important because the output of an interaction between a bacterial pathogen or Plant-Growth-Promoting bacteria (PGPB) and its host greatly depends on the type of tissue that

is colonized, the local microenvironment (spatial relationship), and the movement of the bacterium over time within the host (spatiotemporal relationship) (Lamichhane *et al.*, 2015).

Some of these visualization techniques include fluorescent *in situ* Hybridization (FISH) coupled with Confocal Laser Scanning Microscopy (CLSM) (Earle *et al.*, 2015; Soldan *et al.*, 2019) used to study the spatial relationships of microbe-microbe interactions. Another technique is using bacteria tagged with fluorescent proteins (*e.g* YFP, GFP) to visualize the host-microbe interactions (Godfrey *et al.*, 2010; Cerutti *et al.*, 2017; Donati *et al.*, 2018; Planas-Marquès *et al.*, 2020). In general, considerable effort is being invested into developing new tools to allow the visualisation of bacteria in the host environment (Caldwell and Iyer-Pascuzzi, 2019; Noirot-Gros *et al.*, 2020). The development of new tools to study bacterial colonization, coupled with advances in our understanding of the factors that drive microbial community assembly both conceptually and practically, can ultimately unlock the microbiome's potential in the context of agricultural production. In my thesis, I approach these three key aspects of plant-microbe interactions to advance our knowledge of i) how domesticated plant microbiomes are assembled (evolutionary framework), ii) whether plant domestication has ultimately influenced the host-associated microbiome (factors driving domesticated plant microbiomes assembly and composition), iii) how to improve existing visualization techniques in the context of plant-microbe interactions (development of a new tool to monitor bacterial localization).

Plant-microbe interactions in the agricultural ecosystem

One of the biggest achievements of human history has been plant domestication, which contributed to the formation of sedentary agricultural groups (Ross-Ibarra *et al.*, 2007;

Pérez-Jaramillo *et al.*, 2016). Most common domestication traits are involved in flowering time, seed size, seed chemical composition, loss of seed dispersal mechanisms, and determinate growth and apical dominance (Bitocchi *et al.*, 2012). However, domestication has also led to a reduction in genetic diversity as shown for common bean (Bitocchi *et al.*, 2012; Bellucci *et al.*, 2014), lima bean (Chacón-Sánchez and Martínez-Castillo, 2017), and rice (Ram *et al.*, 2007). Only in recent years, have scientists begun to study the effect of domestication on the host-microbiome assembly and composition (Zachow *et al.*, 2014; Pérez-Jaramillo *et al.*, 2016, 2017; Leff *et al.*, 2017). Even though to date the number of studies is still limited, researchers agree that there are differences between the host microbiome of domesticated and wild plants. Specifically, domesticated crops show a greater abundance of *Proteobacteria* and a decrease in abundance of *Bacteroidetes* compared to ancestral wild types (Zachow *et al.*, 2014; Leff *et al.*, 2017; Pérez-Jaramillo *et al.*, 2017). Moreover, Jaramillo (Pérez-Jaramillo *et al.*, 2017) showed that root traits, like root length, could partially determine the differences between the root microbiome of domesticated and wild common bean plants. Nonetheless, evolutionary frameworks aimed at explaining why these differences occur are lacking. Additionally, different plant compartments are colonized by different microbial communities, thus the choice of which plant compartments to investigate in the context of plant domestication is not trivial.

Nonetheless, if we adopt the extended phenotype concept (Whitham *et al.*, 2003), and we realize that host phenotypes can lead to heritable changes in the communities of interacting organisms, the most remarkable phenotypic changes occurred in the seed compartment (Figure 1). Thus, if we are looking for an effect of domesticated plant

phenotypes on microbiome assembly and composition, the seed is probably a good candidate.

Figure 1. Difference in seed sizes between a domesticated seed (left) and wild-type seeds (right).

The seed microbiome

The term rhizosphere was first coined by Hiltner in 1904 to define the soil zone influenced by root exudates. Since 1904, more and more studies have started focusing on the rhizosphere microbiome, its importance for plant growth (Berendsen *et al.*, 2012)**,** and environmental factors shaping its composition (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Tkacz *et al.*, 2015). However, in recent years other plant niches colonized by microorganisms have started to arouse interest in the scientific community, such as pollen (Ambika Manirajan *et al.*, 2016; Obersteiner *et al.*, 2016) and seed (López-López *et al.*, 2010; Johnston-Monje and Raizada, 2011; Berg and Raaijmakers, 2018; Shahzad *et al.*, 2018).

Studies on seed colonization by bacteria started in the last century, although these mainly focused on seed-borne diseases (Singh and Mathur, 2004). For example, Common bean is affected by several seed-borne bacterial pathogens, including *Pseudomonas syringae*

pv *phaseolicola* (Arnold *et al.*, 2011) (Pph), *Xanthomonas axonopodis* pv *phaseoli* (Jacques *et al.*, 2005)**,** and *Xanthomonas citri* pv *fuscans* (Darrasse *et al.*, 2018). However, it is noticeable that most of the knowledge of seed colonization by bacterial pathogens comes from the last century (Singh and Mathur, 2004) when most of the tools that are now available to study plant-microbe interactions were yet to be discovered.

The seed microbiome is often characterized by a high relative abundance of *Proteobacteria* and *Firmicutes* (Simonin *et al.*, 2021) and shows low alpha-diversity compared to the rhizosphere microbiome (Guo *et al.*). It is also characterised by a high number of specialists, that is a high number of microbial members present at low prevalence (Guo *et al.*, 2021).

The assembly of the seed microbiome seems to be strongly driven by a plant species and a plant genotype effect. For example, the common bean seed microbiome has been reported to be characterised by a high abundance of members of the *Pseudomonadaceae* family (Klaedtke *et al.*, 2016; Chesneau *et al.*) in contrast to the barley seed microbiome (Abdullaeva *et al.*, 2020). A genotype effect was observed for the *Brassica napus* seed microbiome (Rybakova *et al.*, 2017), meaning that different plant genotypes differentially affected microbiome assembly.

It is now clear that microorganisms can reach the seeds at different stages of seed development (Chesneau *et al.*, 2020) and from different sources (Figure 2). In general, microorganisms can reach the seed through the flower (horizontal transmission) or the vascular system (vertical transmission). Sources of horizontally transmitted microbes are pollinators and pollen. Vertically transmitted microorganisms can originate from the rhizosphere, phyllosphere, and soil. We discuss these colonization routes in more detail in chapter 2.

Interestingly, plant domestication has often influenced many of these colonization routes, with unknown consequences on the assembly of the seed microbiome (Soldan, Fusi, *et al.*, 2021).

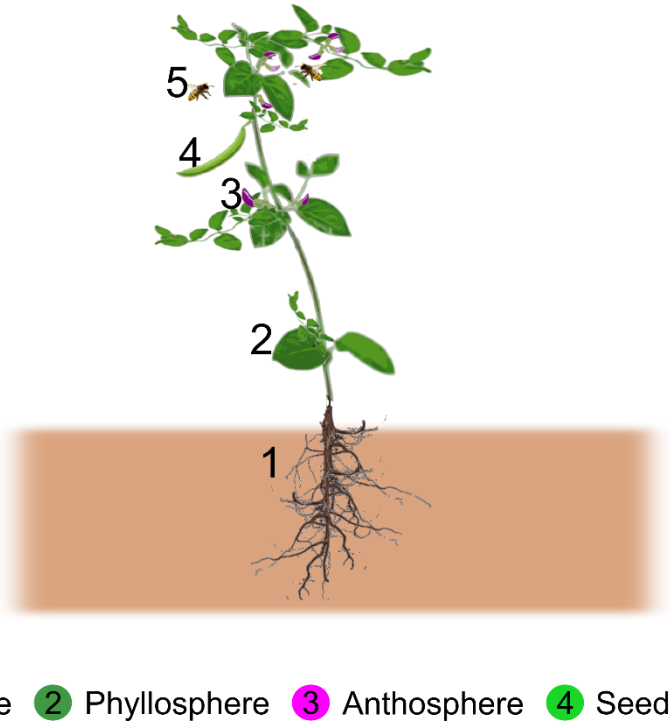


Figure 2. Seed-associated microorganisms can originate from different plant compartments and be both horizontally and vertically transmitted. The seed microbiome could originate from the rhizosphere microbiome through vertical transmission, but also from the phyllosphere microbiome and anthosphere microbiome. Pollinators can also transmit microbes to the seed during flower visits.

Phaseolus spp as a model genus

Phaseolus vulgaris (common bean) is an important crop, being the major source of calories and proteins in developing countries (Myers and Kmiecik, 2017). Moreover, the

genus *Phaseolus* is an important model species in the context of plant domestication and for understanding crop evolution (Bitocchi *et al.*, 2017).

The domestication scenario for common bean (*Phaseolus vulgaris*) shows two independent domestication events that occurred in Mesoamerica and Andes (Rodriguez *et al.*, 2016). Specifically, *P. vulgaris* has a Mesoamerican origin and was introduced later in the Andes giving origin to another gene pool (Bitocchi *et al.*, 2013). In central and South America, wild *P. vulgaris* underwent two independent domestication events, which lead to two new domesticated gene pools, one in South America and one in Mesoamerica. A similar domestication scenario has been shown for lima bean (*Phaseolus lunatus*) (Chacón-Sánchez and Martínez-Castillo, 2017).

This means that four independent domestication events have happened within two closely related plant species. This offers the unique opportunity to have a replication of an event lasting about 13,000 thousand years. Therefore, testing an evolutionary framework developed to explain why and how domesticated plant microbiomes are assembled on *Phaseolus* spp represents a good choice.

We have covered the rationale for why evolutionary frameworks are important in the context of agricultural plant-microbe interactions and how this can guide the development of research projects. We have also reflected on the importance of selecting a model plant species and a plant compartment to study domesticated plant microbiomes and factors that drive their assembly. Moreover, as we will see in the next paragraph, common bean also represents a model plant species for the 3rd key for understanding plant-microbe interactions, which is visualization techniques.

Studying bacterial colonization. *Pseudomonas syringae* and its host *Phaseolus vulgaris* as model organisms

As highlighted above, NGS technologies have allowed researchers to embrace the complexity of microbial ecosystems. However, they cannot be used to deeply investigate spatio-temporal dynamics of plant-microbe interactions. For this reason, various visualization techniques exist. For example, bacteria tagged with fluorescent proteins (e.g YFP, GFP) have been extensively used to study both detrimental (Godfrey *et al.*, 2010; Cerutti *et al.*, 2017; Donati *et al.*, 2018; Planas-Marquès *et al.*, 2020) and beneficial (Compant *et al.*, 2005; Fan *et al.*, 2011; Mitter *et al.*, 2017) host-microbe interactions. However, monitoring bacterial movements ideally requires both macro (whole plant/organ) and micro (tissue) resolution (Soldan, Sanguankiattichai, *et al.*, 2021). A simple and easy-to-use technique allowing researchers to study host-microbe interactions at both macroscopic and microscopic levels has not been developed yet. This is important for several reasons. For pathogenic bacteria, the ability to become systemic and spread in plant tissues has great consequences for the management of the disease (Cerutti *et al.*, 2017). Fluorescent proteins cannot be used to understand bacterial spreading as imaging a whole plant organ is not feasible. Similarly, bacterial spreading is also important for beneficial plant-microbe interactions. Therefore, developing a tool that allows researcher to monitor bacterial localization possibly at different resolution levels is of great importance and have many applications. To achieve this goal, we used *Phaseolus vulgaris* and its bacterial pathogen *Pseudomonas syringae* pv *phaseolicola* (Pph) as model organisms. Pph represents an ideal candidate to develop a monitoring tool for several reasons. Pph is a Gram-negative bacterium and the agent of halo blight of *Phaseolus*

vulgaris (common bean), where it causes water-soaked lesions on both leaves and pods, chlorosis, reduced photosynthesis, and consequently reduced yields (Hernández-Morales *et al.*, 2009). The primary source of inoculum is infected seeds, but the route of colonization remains to be fully elucidated (Soldan, Sanguankiattichai, *et al.*, 2021). Therefore, it represents a good candidate to develop a tool to monitor bacterial movement and colonization habits since Pph has to show some degree of motility within the plant to reach and infect the seed compartment. Moreover, Pph's genome is sequenced and it is a widely studied bacterium (Arnold *et al.*, 2011).

Thesis outline

The overall aim of my thesis is to advance our understanding of plant-microbe interactions in the three complementary aspects mentioned above. In Chapter 2 I introduce why advancing our understanding of assembly processes in domesticated plants is important. Moreover, I propose an evolutionary framework that could explain domesticated plant microbiome assembly to help the development of new research hypotheses. In Chapter 3, I extend the framework developed in Chapter 2 and conceptualise the effect of plant domestication on microbiome heritability. In Chapter 4, I provide experimental evidence for the framework proposed in Chapter 2 by analysing the seed microbiome of hundreds of wild and domesticated plants of *Phaseolus* spp. In Chapter 5, I describe the development and validation of a tool that can be used to monitor and study bacterial colonization. Chapter 6 brings together the research findings of this thesis and discusses their implications.

References

Abdullaeva, Y., Ambika Manirajan, B., Honermeier, B., Schnell, S., and Cardinale, M. (2020) Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *Journal of Advanced Research*.

Ambika Manirajan, B., Ratering, S., Rusch, V., Schwiertz, A., Geissler-Plaum, R., Cardinale, M., and Schnell, S. (2016) Bacterial microbiota associated with flower pollen is influenced by pollination type, and shows a high degree of diversity and species-specificity. *Environ Microbiol* 18: 5161–5174.

Arnold, D.L., Lovell, H.C., Jackson, R.W., and Mansfield, J.W. (2011) *Pseudomonas syringae* pv. *phaseolicola*: from "has bean" to supermodel. *Mol Plant Pathol* 12: 617–627.

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., et al. (2014) Decreased Nucleotide and Expression Diversity and Modified Coexpression Patterns Characterize Domestication in the Common Bean. *The Plant Cell* tpc.114.124040.

Berendsen, R.L., Pieterse, C.M.J., and Bakker, P.A.H.M. (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci* 17: 478–486.

Berg, G. and Raaijmakers, J.M. (2018) Saving seed microbiomes. *The ISME Journal* 12: 1167–1170.

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol* 197: 300–313.

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P.S., et al. (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *PNAS* 109: E788–E796.

Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M.L., Gioia, T., et al. (2017) Beans (*Phaseolus s*sp.) as a Model for Understanding Crop Evolution. *Frontiers in Plant Science* 8: 722.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., et al. (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 488: 91–95.

Caldwell, D. and Iyer-Pascuzzi, A.S. (2019) A Scanning Electron Microscopy Technique for Viewing Plant−Microbe Interactions at Tissue and Cell-Type Resolution. *Phytopathology*.

Cerutti, A., Jauneau, A., Auriac, M.-C., Lauber, E., Martinez, Y., Chiarenza, S., et al. (2017) Immunity at Cauliflower Hydathodes Controls Systemic Infection by *Xanthomonas campestris* pv *campestris*. *Plant Physiology* 174: 700–716.

Chacón-Sánchez, M.I. and Martínez-Castillo, J. (2017) Testing Domestication Scenarios of Lima Bean (*Phaseolus lunatus* L.) in Mesoamerica: Insights from Genome-Wide Genetic Markers. *Front Plant Sci* 8:.

Chesneau, G., Torres-Cortes, G., Briand, M., Darrasse, A., Preveaux, A., Marais, C., et al. Temporal dynamics of bacterial communities during seed development and maturation. *FEMS Microbiol Ecol*.

Chesneau, G., Torres-Cortes, G., Briand, M., Darrasse, A., Preveaux, A., Marais, C., et al. (2020) Temporal dynamics of bacterial communities during seed development and maturation. *FEMS Microbiology Ecology* 96:.

Compant, S., Reiter, B., Sessitsch, A., Nowak, J., Clément, C., and Ait Barka, E. (2005) Endophytic colonization of *Vitis vinifera* L. by plant growth-promoting bacterium *Burkholderia* sp. strain PsJN. *Appl Environ Microbiol* 71: 1685–1693.

Darrasse, A., Barret, M., Cesbron, S., Compant, S., and Jacques, M.-A. (2018) Niches and routes of transmission of *Xanthomonas citri* pv. *fuscans* to bean seeds. *Plant Soil* 422: 115–128.

Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418: 700–707.

Donati, I., Cellini, A., Buriani, G., Mauri, S., Kay, C., Tacconi, G., and Spinelli, F. (2018) Pathways of flower infection and pollen-mediated dispersion of *Pseudomonas syringae* pv. *actinidiae*, the causal agent of kiwifruit bacterial canker. *Hortic Res* 5:.

Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., et al. (2015) Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host & Microbe* 18: 478–488.

Fan, B., Chen, X.H., Budiharjo, A., Bleiss, W., Vater, J., and Borriss, R. (2011) Efficient colonization of plant roots by the plant growth promoting bacterium *Bacillus amyloliquefaciens* FZB42, engineered to express green fluorescent protein. *Journal of Biotechnology* 151: 303–311.

Friesen, M.L., Porter, S.S., Stark, S.C., von Wettberg, E.J., Sachs, J.L., and Martinez-Romero, E. (2011) Microbially Mediated Plant Functional Traits. *Annual Review of Ecology, Evolution, and Systematics* 42: 23–46.

Gilbert, J.A., Lelie, D. van der, and Zarraonaindia, I. (2014) Microbial terroir for wine grapes. *PNAS* 111: 5–6.

Godfrey, S. a. C., Mansfield, J.W., Corry, D.S., Lovell, H.C., Jackson, R.W., and Arnold, D.L. (2010) Confocal imaging of *Pseudomonas syringae* pv. *phaseolicola* colony development in bean reveals reduced multiplication of strains containing the genomic island PPHGI-1. *Mol Plant Microbe Interact* 23: 1294–1302.

Guo, J., Ling, N., Li, Y., Li, K., Ning, H., Shen, Q., et al (2021). Seed-borne, endospheric and rhizospheric core microbiota as predictor for plant functional traits across rice cultivars are dominated by deterministic processes. *New Phytologist* n/a:

Henry, L.P., Bruijning, M., Forsberg, S.K.G., and Ayroles, J.F. (2021) The microbiome extends host evolutionary potential. *Nat Commun* 12: 5141.

Hernández-Morales, A., De la Torre-Zavala, S., Ibarra-Laclette, E., Hernández-Flores, J.L., Jofre-Garfias, A.E., Martínez-Antonio, A., and Álvarez-Morales, A. (2009) Transcriptional profile of *Pseudomonas syringae* pv. *phaseolicola* NPS3121 in response to tissue extracts from a susceptible *Phaseolus vulgaris* L. cultivar. *BMC Microbiology* 9: 257.

Jacques, M.-A., Josi, K., Darrasse, A., and Samson, R. (2005) *Xanthomonas axonopodis* pv. *phaseoli* var. *fuscans* Is Aggregated in Stable Biofilm Population Sizes in the Phyllosphere of Field-Grown Beans. *Appl Environ Microbiol* 71: 2008–2015.

Johnston-Monje, D. and Raizada, M.N. (2011) Conservation and Diversity of Seed Associated Endophytes in Zea across Boundaries of Evolution, Ethnography and Ecology. *PLOS ONE* 6: e20396.

Klaedtke, S., Jacques, M.-A., Raggi, L., Préveaux, A., Bonneau, S., Negri, V., et al. (2016) Terroir is a key driver of seed-associated microbial assemblages. *Environmental Microbiology* 18: 1792–1804.

Lamichhane, J.R., Messéan, A., and Morris, C.E. (2015) Insights into epidemiology and control of diseases of annual plants caused by the *Pseudomonas syringae* species complex. *J Gen Plant Pathol* 81: 331–350.

Leff, J.W., Lynch, R.C., Kane, N.C., and Fierer, N. (2017) Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, *Helianthus annuus*. *New Phytologist* 214: 412–423.

López-López, A., Rogel, M.A., Ormeño-Orrillo, E., Martínez-Romero, J., and Martínez-Romero, E. (2010) Phaseolus vulgaris seed-borne endophytic community with novel bacterial species such as *Rhizobium endophyticum* sp. nov. *Systematic and Applied Microbiology* 33: 322–327.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488: 86–90.

Mitter, B., Pfaffenbichler, N., Flavell, R., Compant, S., Antonielli, L., Petric, A., et al. (2017) A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. *Front Microbiol* 8:.

Myers, J.R. and Kmiecik, K. (2017) Common Bean: Economic Importance and Relevance to Biological Science Research. In *The Common Bean Genome*. Compendium of Plant Genomes. Pérez de la Vega, M., Santalla, M., and Marsolais, F. (eds). Cham: Springer International Publishing, pp. 1–20.

Noirot-Gros, M.-F., Shinde, S.V., Akins, C., Johnson, J.L., Zerbs, S., Wilton, R., et al. (2020) Functional Imaging of Microbial Interactions With Tree Roots Using a Microfluidics Setup. *Frontiers in Plant Science* 11: 408.

Obersteiner, A., Gilles, S., Frank, U., Beck, I., Häring, F., Ernst, D., et al. (2016) Pollen-Associated Microbiome Correlates with Pollution Parameters and the Allergenicity of Pollen. *PLoS ONE* 11: e0149545.

Pérez-Jaramillo, J.E., Carrión, V.J., Bosse, M., Ferrão, L.F.V., de Hollander, M., Garcia, A.A.F., et al. (2017) Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J* 11: 2244–2257.

Pérez-Jaramillo, J.E., Mendes, R., and Raaijmakers, J.M. (2016) Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Molecular Biology* 90: 635–644.

Planas-Marquès, M., Kressin, J.P., Kashyap, A., Panthee, D.R., Louws, F.J., Coll, N.S., and Valls, M. (2020) Four bottlenecks restrict colonization and invasion by the pathogen *Ralstonia solanacearum* in resistant tomato. *J Exp Bot* 71: 2157–2171.

Ram, S.G., Thiruvengadam, V., and Vinod, K.K. (2007) Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. *J Appl Genet* 48: 337–345.

Rodriguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., et al. (2016) Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol* 209: 1781–1794.

Ross-Ibarra, J., Morrell, P.L., and Gaut, B.S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences* 104: 8641–8648.

Rybakova, D., Mancinelli, R., Wikström, M., Birch-Jensen, A.-S., Postma, J., Ehlers, R.-U., et al. (2017) The structure of the *Brassica napus* seed microbiome is

cultivar-dependent and affects the interactions of symbionts and pathogens. *Microbiome* 5: 104.

Santos, M.S., Nogueira, M.A., and Hungria, M. (2019) Microbial inoculants: reviewing the past, discussing the present and previewing an outstanding future for the use of beneficial bacteria in agriculture. *AMB Express* 9: 205.

Shahzad, R., Khan, A.L., Bilal, S., Asaf, S., and Lee, I.-J. (2018) What Is There in Seeds? Vertically Transmitted Endophytic Resources for Sustainable Improvement in Plant Growth. *Front Plant Sci* 9:.

Simonin, M., Briand, M., Chesneau, G., Rochefort, A., Marais, C., Sarniguet, A., and Barret, M. (2021) Seed microbiota revealed by a large-scale meta-analysis including 50 plant species.

Singh, D. and Mathur, S.B. (2004) Histopathology of Seed-Borne Infections, CRC Press.

Soldan, R., Fusi, M., Cardinale, M., Daffonchio, D., and Preston, G.M. (2021) The effect of plant domestication on host control of the microbiota. *Commun Biol* 4: 1–9.

Soldan, R., Mapelli, F., Crotti, E., Schnell, S., Daffonchio, D., Marasco, R., et al. (2019) Bacterial endophytes of mangrove propagules elicit early establishment of the natural host and promote growth of cereal crops under salt stress. *Microbiological Research* 223–225: 33–43.

Soldan, R., Sanguankiattichai, N., Bach-Pages, M., Bervoets, I., Huang, W.E., and Preston, G.M. (2021) From macro to micro: a combined bioluminescence-fluorescence approach to monitor bacterial localization. *Environmental Microbiology* 23: 2070–2085.

Stappenbeck, T.S. and Virgin, H.W. (2016) Accounting for reciprocal host–microbiome interactions in experimental science. *Nature* 534: 191–199.

Tkacz, A., Cheema, J., Chandra, G., Grant, A., and Poole, P.S. (2015) Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. *The ISME Journal* 9: 2349–2359.

Wagner, M.R., Lundberg, D.S., del Rio, T.G., Tringe, S.G., Dangl, J.L., and Mitchell-Olds, T. (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 7: 12151.

Whitham, T.G., Young, W.P., Martinsen, G.D., Gehring, C.A., Schweitzer, J.A., Shuster, S.M., et al. (2003) Community and Ecosystem Genetics: A Consequence of the Extended Phenotype. *Ecology* 84: 559–573.

Zachow, C., Müller, H., Tilcher, R., and Berg, G. (2014) Differences between the rhizosphere microbiome of *Beta vulgaris* ssp. *maritima*—ancestor of all beet crops—and modern sugar beets. *Front Microbiol* 5:.

Chapter 2: The effect of plant domestication on host control of the microbiota

Riccardo Soldan[1*], Marco Fusi[2], Massimiliano Cardinale[3], Daniele Daffonchio[4], Gail M. Preston[1*]

[*]corresponding authors

Communication Biology 4, 936 (2021)

doi: 10.1038/s42003-021-02467-6

Abstract

Macroorganisms are colonized by microbial communities that exert important biological and ecological functions, the composition of which is subject to host control and has therefore been described as “an ecosystem on a leash”. However, domesticated organisms such as crop plants are subject to both artificial selection and natural selection exerted by the agricultural ecosystem. Here, we propose a framework for understanding how host control of the microbiota is influenced by domestication, in which a double leash acts from domesticator to host and host to microbes. We discuss how this framework applies to a plant compartment that has demonstrated remarkable phenotypic changes during domestication: the seed.

Introduction

According to FAO estimates, consumption of seeds comprises from 40 to 70% of both calories and protein in the human diet (FAO, 1970), while most of the remaining 60 to 30% is constituted by derivates of animals, which again are mainly fed on seeds (Mandal and Mandal, 2000). Each year, we produce about 2,5 billion tons of cereal grains (https://data.worldbank.org/indicator/ag.prd.crel.mt). This accomplishment can be considered to be the consequence of plant domestication (Diamond, 1998, 2002). In the last 12,000 years, humans have domesticated and semi-domesticated about 2,500 plant species as crops (Dirzo and Raven, 2003; Meyer and Purugganan, 2013), which together with changes in agricultural practices, have enabled high yielding harvests. Many domesticated crops show common traits that distinguish them from their wild counterparts. This pool of common traits is called the domestication syndrome (Hammer, 1984) and includes increased fruit size and changes to reproductive strategy, branching, secondary metabolites, and seed-shattering (Ross-Ibarra *et al.*, 2007; Meyer *et al.*, 2012). The most remarkable domestication traits frequently involve the seed. For example, grain crops produce seeds that are up to 15.2 times bigger than seeds produced by their wild progenitors (Kluyver *et al.*, 2017).

Since the pioneering work of Darwin (1868) and lately Vavilov (Vavilov *et al.*, 1992) and Engelbrecht (Dr Th. H. Engelbrecht's views on the origin of cultivated plants, 1973), we have furthered our understanding of the origin of domestication (Diamond, 2002; Piperno, 2011; Meyer and Purugganan, 2013) and its genetic basis (Ross-Ibarra *et al.*, 2007; Kantar *et al.*, 2017). However, except for a few notable exceptions, little is still known about how plant domestication influences the community phenotypes of

interacting organisms (Chen *et al.*, 2015). Among the different organisms that live in associations with plants, microorganisms have particular importance because of growth and fitness (Cheplick, 2008; Berendsen *et al.*, 2012; Haney *et al.*, 2015; Vries *et al.*, 2020). These microbial communities are so closely associated with the plant that they can form, together with the plant host, a holobiont (Postler and Ghosh, 2017; Hassani *et al.*, 2018). These intimate relationships are crucial for plant well-being, resistance and resilience, as well as for basic processes of the plant life cycle, such as germination (Puente *et al.*, 2009). Since plant domestication, led by artificial selection, has shaped and changed many aspects of plant growth and development, it is logical to ask how these phenotypic changes have influenced the evolution of plant-microbe interactions. In recent years, we have witnessed an increasing research effort to understand the effect of plant domestication on the plant microbiome (Bulgarelli *et al.*, 2015; Berg and Raaijmakers, 2018; Pérez-Jaramillo *et al.*, 2018; Abdullaeva *et al.*, 2020; Kim *et al.*, 2020) but evolutionary theories that could explain how the assembly of domesticated plant microbiomes differs from that of wild progenitors are still lacking.

In this work, we first adopt and expand the concept of the microbiome as an ecosystem on a "leash" using the framework proposed by Foster *et al.*, (Foster *et al.*, 2017) to incorporate the effect of domestication, thus artificial selection and natural selection exerted by the agricultural ecosystem, on the host control of the microbiota. We then apply our framework for a specific plant compartment, the seed, and discuss the potential effect of plant domestication (phenotypic changes) on the seed microbiome and plant-microbe interactions in the spermosphere. We conclude by highlighting the implications of the double-leash framework from a biotechnological perspective.

Domesticated plant microbiomes as an ecosystem on a "double leash"

To better explain the evolution of hosts and their microbiomes, Foster *et al.,* (Foster *et al.*, 2017) propose that the microbiome can be broken down into three classes of effect (host-to-microbe, microbe-to-host, and microbe-to-microbe), each one with its evolutionary features. Because the host is under strong selective pressure to favor beneficial interactions with its microbiome (Schluter and Foster, 2012), and the host can control the environment in which its microbiome lives, the microbiome becomes "*a dynamic microbial ecosystem held on an ever-evolving leash by the host"* (Foster *et al.*, 2017). In this work, we adopt this framework with a particular focus on the host-to-microbe effects caused by domestication, thereby extending this concept to incorporate artificial selection (or direct selection) and natural selection (or indirect selection) acting through the agricultural ecosystem. Understanding the consequences of domesticated plant phenotypes on microbe-to-host effects requires an understanding of how these phenotypes could change host-to-microbe effects, particularly when domesticated plant phenotypes are often similar between different phylogenetically distant crops.

In a natural environment, plants would be under selective pressure to maximize positive microbe-to-host effects, through control of the host microbiome (the leash) (Foster *et al.*, 2017) (Figure 1a). However, under domestication, host evolution would be also driven by artificial selection and natural (indirect) selection acting through the agricultural ecosystem (Purugganan, 2019). Under these circumstances, the host-associated microbiome would become an ecosystem on a leash held by the host but modified by the domesticator, who influences the evolution of the host through a second leash (Figure 1b). The second leash, which is exerted by the domesticator through direct selection and/or indirect selection, mainly acts on the host at a genotypic level. Artificial selection

would act by increasing the contribution of the genotypic component in the three-way interaction between genotype, environment, and host-microbiome that determines the plant phenotype, to maximize the heritability of artificially selected traits in different environments. Moreover, plant traits directly selected by the domesticator are not necessarily linked to higher fitness, thus we predict a lower contribution of the microbiome to the manifestation of these phenotypes compared to wild progenitors.

Indirect selection would also act at a genotypic level, but traits under indirect selection would confer a fitness advantage to individuals growing in agricultural ecosystems. For this reason, we cannot exclude a contribution of microbe-to-host effects in the manifestation of certain plant traits under indirect selection. Nonetheless, the domesticator could still guide indirect selection through changes in agricultural practices, ultimately limiting the selection of unwanted phenotypes (Zohary, 2004).

Since phenotypic changes caused by artificial selection did not arise as a plant response to control its microbiota for positive microbe-to-host effects, it is logical to predict that changes in the domesticated plant microbiome assembly relative to wild progenitors would be driven by the effect of these artificially selected phenotypic changes, which may not necessarily favour host control of the microbiota to the benefit of the host, with a few notable exceptions such as artificial selection for disease resistance. This does not suggest weaker host-to-microbe effects (Figure 1b), but simply an effect mainly dictated by plant phenotypes directly selected through artificial selection on the microbiome. Therefore, while the plant will still exert control of its microbiota, artificial selection would lower the effect of the plant leash, and host-microbiome assembly will be shaped by the genotypic leash exerted by the domesticator.

When phenotypic changes are induced by indirect selection, host control of the microbiota could still favor positive microbe-to-host effects. Therefore, indirect selection would not necessarily lower the effect of the plant leash, but potentially drive a different microbiome assembly compared to one of the wild progenitor that confers positive microbe-to-host effects. However, several studies have indicated a reduction in positive microbe-to-host effects in domesticated hosts compared to wild progenitors, including studies involving wheat, maize (Hetrick *et al.*, 1992; Sangabriel-Conde *et al.*, 2014), and soybean (Kiers *et al.*, 2007) (reviewed in (Porter and Sachs, 2020)), possibly suggesting a weak effect of indirectly selected plant traits in determining microbiome assembly.

The results of host-to-microbe effects dictated by domesticated plant phenotypes could lead to microbial species loss, gain and species replacement depending on the ecological processes driving host microbiome assembly (Figure 1c). When host-to-microbe effects caused by domesticated plant phenotypes lead to species loss, and the lost microbial species were selected through host-to-microbe effects in wild progenitors, then a decrease in the overall importance of selection in driving community assembly compared to wild plants is expected (Figure 1cd). Our double-leash framework predicts this could happen when domesticated plant phenotypes are the result of artificial selection. Species loss could also be caused by host-to-microbe effects induced by plant traits under indirect selection, where the lost microbial species negatively affect plant fitness under the agricultural ecosystem. Moreover, the agricultural ecosystem could also influence microbe-to-microbe effects, potentially influencing the outcome of host-to-microbe effects.

When domesticated plant traits exert strong host-to-microbe effects, species replacement and/or species gain could dominate the difference between wild and domesticated plant

microbiomes. In the case of species replacement, there could be no difference in overall selection exerted by the wild and domesticated plants (Figure 1ce). However, the double-leash hypothesis would predict that wild and domesticated plants would be associated with different microbial communities that exert different microbe-to-host effects. In the case of species gain, an increase in the overall importance of selection in driving community assembly compared to wild plants is expected (Figure 1cf). Both species replacement and species gain are predicted to be an effect of domesticated plant phenotypes under artificial or indirect selection. However, we might expect that under indirect selection, the resulting microbiome is more likely to provide positive host-to-microbe effects.

To date, research studies seem to agree with the predictions of this framework, particularly in relation to species loss and species replacement, and reduced selection caused by plant domestication. For example, this has been shown for the wheat rhizosphere microbiome as a consequence of plant dwarfing (Kavamura *et al.*, 2020) and also for the wheat seedling microbiome (Özkurt *et al.*, 2020).

The effect of agricultural practices

As host control of the microbiome in domesticated plants acts in the context of agricultural ecosystems, we need to also contextualize their effect on microbe-to-microbe interactions during domestication, which is a multi-staged process in which both host genotypes and environments have changed (Purugganan, 2019). During the early stages of domestication, plants and their microbiome were adapted to their geographically local, predictably unstable environment, not yet fully agricultural. The differences, if any, in the plant microbiomes of wild and early domesticated crops would be the result of variation

in interactions between host, environment, and microorganisms mainly caused by host genetic and environmental differences. Differences in host control of the microbiome in domesticated plants compared to wild plants would arise through artificial and indirect selection but agricultural practices would exert a low influence on microbe-to-microbe effects.

During later stages of the domestication process plants were grown over a wider geographical area and plant genotypes adapted to different agricultural environments (landraces), and new cultivars were bred. During these stages, host-to-microbe effects would be influenced by the double leash but also by differences in environment-to-host and microbe-to-microbe effects caused by different agricultural ecosystems and new environments. For example, the effect of agricultural practices, particularly intensive fertilization and pest control methods could act to lower soil microbial diversity (Hartmann *et al.*, 2015; Hendgen *et al.*, 2018), influencing the outcome of host-to-microbe effects.

Below, we further discuss these effects for the plant compartment that has shown the most remarkable phenotypic changes, the seed.

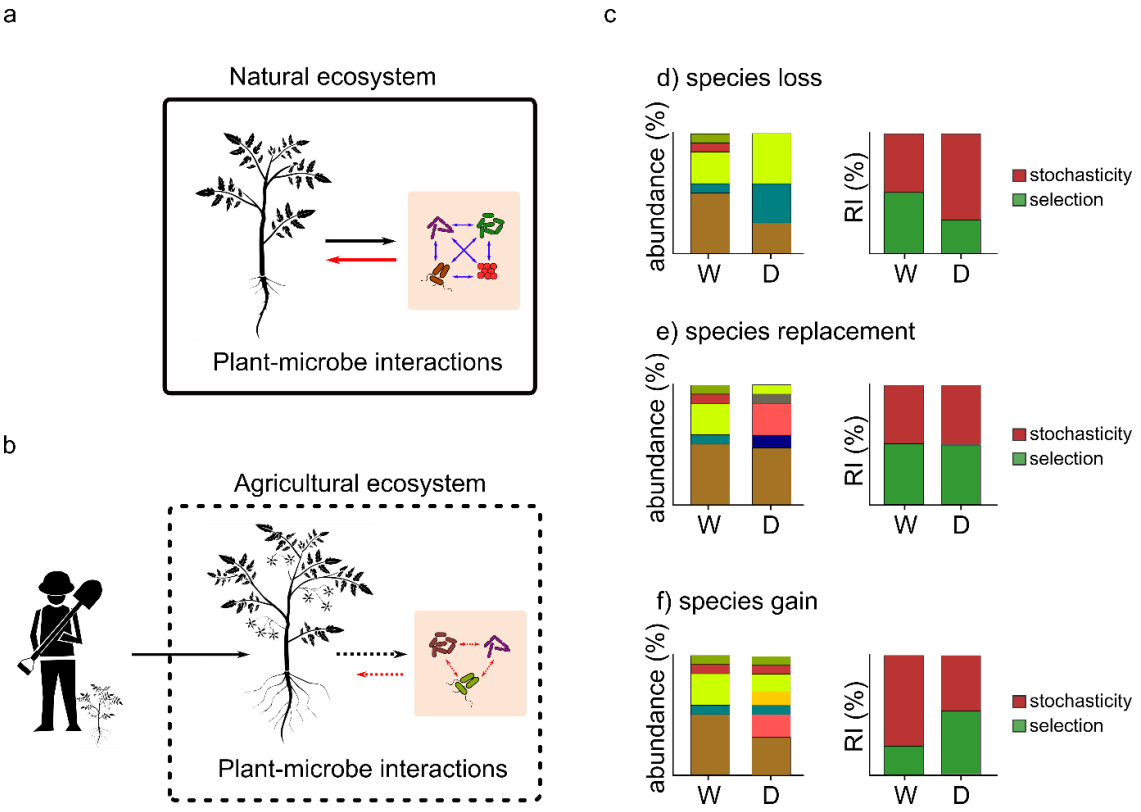


Figure 1. Differences in plant microbiomes between wild and domesticated plants are governed by host-to-microbe and microbe-to-microbe effects dictated by plant domestication. a: Wild plants are under natural selection to control their microbiota for positive microbe-to-host effects. Black arrow indicates host-to-microbe effects (the leash) and red arrow microbe-to-host effects. Microbe-to-microbe effects are indicated by blue arrows. The black line surrounding host and microbiome represents the natural environment. b: Domesticated plants are under artificial and indirect selection. The domesticator strongly influences the plant genotype through artificial selection (first leash, black arrow) and indirect selection exerted by the agricultural ecosystem (black dotted line surrounding host and microbiome), so that the positive microbe-to-host-effects arising from host control of the microbiome (second leash, black dotted arrow) are reduced (red dotted arrow). Under these circumstances, domesticated plant microbiomes are ecosystems on a "double leash" from domesticator to host and host to microbe. Additionally, the agricultural ecosystem can affect microbe-to-microbe and host-to-microbe interactions, with consequences for microbe-to-host effects. cd-f: The double leash hypothesis could predict species loss (cd),

replacement (ce), and gain (cf) caused by different host-to-microbe effects induced by domesticated plant phenotypic traits. Different colours in the relative abundance charts indicate different microbial members. W; wild-type plants. D; domesticated plants. RI; Relative importance of the ecological processes driving community assembly. Plant drawings were downloaded from iStock under standard license agreement.

The seed provides an ideal context in which to study the effect of plant domestication on plant microbiomes

The seed microbiome and plant-microbe interactions in the spermosphere have important ecological functions (for recent reviews see (Shade *et al.*, 2017; Nelson, 2018)). Seed-associated microorganisms can be inherited through vertical transmission (Morse *et al.*, 2007; Johnston-Monje and Raizada, 2011; Truyens *et al.*, 2015; Shahzad *et al.*, 2018; Vannier *et al.*, 2018; Walitang *et al.*, 2019); they provide beneficial effects for seedling growth and establishment (Puente *et al.*, 2009; Pitzschke, 2018; Shahzad *et al.*, 2018; Rochefort *et al.*, 2019; Soldan *et al.*, 2019; Matsumoto *et al.*, 2021) and determine, together with soil, the rhizosphere microbiome (Johnston-Monje *et al.*, 2016). Moreover, artificially introducing plant growth-promoting bacteria into seeds has been shown to increase wheat crop yield (Mitter *et al.*, 2017), opening the possibility of a new era of bacterial inoculants and plant breeding (Wei and Jousset, 2017).

Plant domestication has influenced and changed many seed-related phenotypes, and while root architecture has also changed from wild to domesticated crops (Szoboszlay *et al.*, 2015), the increase in seed size exceeds any other difference noticeable in roots. For example, maize (*Zea mays* ssp. *mays*) kernel weight has increased more than 10-fold compared to teosinte (*Z. mays* ssp. *parviglumis*) (Liu *et al.*, 2016). If the same effect was

to be observed in roots, domesticated maize would have on average 8 meter-long roots. Differences in root length have already been linked to differences in the composition of the rhizosphere bacterial community between wild and domesticated common bean (Pérez-Jaramillo *et al.*, 2017) but the remarkable increase in seed size of domesticated plants and its consequences for plant-microbe interactions has thus far been neglected. Seed chemical composition has also changed through plant domestication (Beleggia *et al.*, 2016), with unknown consequences for the seed microbiome and plant-microbe interactions in the spermosphere (the area surrounding a germinating seed that is influenced by the seed).

We propose that under the double leash framework, seeds represent an ideal model to study the effect of domestication on plant microbiomes. Seeds are often the main target of direct and indirect selection and they often showcase the most remarkable phenotypic changes associated with domestication compared to other plant compartments, such as roots. Seed microbiomes are characterised by low alpha diversity values, potentially favoring host control of microbial communities (Foster *et al.*, 2017). Therefore, the strength of the double leash in shaping microbial communities could be stronger for the seed compartment.

Seeds are also involved in the vertical transmission of microbes. For example, this has been widely reported for the fungal symbiont *Neotyphodium* in grasses (Morse *et al.*, 2007; Wiewióra *et al.*, 2015) and also for members of the bacterial microbiota in maize (Johnston-Monje and Raizada, 2011). Vertical transmission, or inheritance, is important because it can favour a higher degree of co-evolution between host and symbionts (Opstal and Bordenstein, 2015) (although see (Douglas and Werren, 2016)). Vertically transmitted microorganisms could increase the fitness of their host by promoting

competition and seed establishment, thus increasing the chances of seedlings developing into mature plants (Puente *et al.*, 2009; Soldan *et al.*, 2019; Verma *et al.*, 2019). The effect of phenotypic changes (host-to-microbe effects) caused by plant domestication on vertical transmission remains unknown.

The effect of domestication on seed microbiome assembly: the double leash paradigm

In our proposed view, domesticated plant microbiomes are ecosystems on a double leash, which can be exemplified by consideration of the seed microbiome. Changes in the seed-microbe interactions of domesticated plants compared to wild progenitors could arise through host-to-microbe effects acting pre-dispersal (when the seed is developing), such as changes in the plant compartments and routes used by microorganisms to colonize seeds (Figure 2a), as well as through the biochemical and physical properties of the developing and mature seed, which act both pre-seed and post-seed dispersal (Figure 2b). We discuss the mechanisms and potential impact of both of these interactions in more detail below.

a) Host-to-microbe effects of plant domestication on seed microbiome assembly (pre-seed dispersal)

Under our "double-leash" framework, the domesticator has influenced and changed plant traits that have led to host-to-microbe effects that differ from those occurring in the wild plant progenitors. The phenotypic changes caused by domestication and their effect on seed microbiome assembly could be represented as changes in microbial connectivity (sharing of microbial members), between plant compartments and the environment (Figure 2a). Even though the exact routes used by microorganisms to colonize seeds

remain under-investigated, it is now widely recognized that the seed microbiome can originate from the anthosphere, rhizosphere and phyllosphere, each of which could exert host-microbe effects on the composition of the seed microbiome (for recent reviews see (Truyens *et al.*, 2015; Shade *et al.*, 2017; Rodríguez *et al.*, 2018; Shahzad *et al.*, 2018)). The colonization of seeds via the anthosphere is particularly well-documented for plant pathogens. For example, *Xanthomonas euvescicatoria* infects pepper seeds through the flower stigma (Dutta *et al.*, 2014). Similarly, *Acidovorax citrulli* can colonize seeds of watermelon through a floral pathway (Dutta *et al.*, 2016). Pollen can also be a source of microorganisms, potentially transmitted to the seeds through the flower stigma and style. For example, *Pseudomonas syringae* pv. *actinidiae* colonizes the anthers and pollen of infected male plants (Donati *et al.*, 2018). In turn, the microbiome of flowers and pollen is influenced by the visits of pollinators (Aizenberg-Gershtein *et al.*, 2013; Aleklett *et al.*, 2014; Ambika Manirajan *et al.*, 2016; McFrederick *et al.*, 2017), which could also be a source of potentially horizontally transmitted microbes to the seeds. For example, *Lactobacillus*, *Spiroplasma* and *Frischella* genera found on honey bees were found to be more abundant in the seed microbiome of bee-pollinated compared to hand-pollinated oilseed rape (Prado *et al.*, 2020). Multiple phenotypic changes observed in domesticated plants could have direct and indirect consequences for seed colonisation via the anthosphere, including earlier flowering time, flower size, colour and shape, and floral reward chemistry (Egan *et al.*, 2018). Changes in floral chemical composition and flowering time are known to lead to different flower-pollinator interactions, and potentially to differences in the importance of horizontal transmission in contributing to the seed microbiome (Prado *et al.*, 2020).

Transmission of bacteria via the vascular system from the rhizosphere and other plant tissues is also known to be important for seed colonisation. For instance, co-localization of bacteria with plant vascular tissues of seeds has been demonstrated for seeds of *Anadenanthera colubrina* (Alibrandi *et al.*, 2018) and bacteria potentially originating from soil were found to be endophytic on flowers, berries, and seeds of grapevine (Compant *et al.*, 2011; Zarraonaindia *et al.*, 2015). Rodriguez et al. (Rodríguez *et al.*, 2019) suggested that the soil is likely to be an important reservoir of microorganisms that contributes to the seed microbiome assembly of *Setaria viridis.* The authors conducted an experiment growing *Setaria viridis* in a sterile substrate for successive plant generations and detected an altered and simplified seed microbial community, concluding that at least part of the seed microbiome could originate from the soil. Similar findings were observed for the seed microbiome of *Prosopis velutina* (Valdez, 2019). The shorter life cycles typical of domesticated plants could reduce connectivity between the rhizosphere and seed microbiomes, lowering microbial inheritance; also changes in root exudates, root architecture and plant immunity are likely to substantially influence the ability of rhizosphere and phyllosphere-colonising organisms to colonise seeds.

As host-to-microbe effects caused by domestication have directly or indirectly influenced the colonization routes used by microorganisms to colonize seeds, thereby affecting the microbial connectivity between different plant compartments and the environment, the seed microbiome of wild progenitors is likely to harbor different members of the microbiota or to show differences in community composition when compared to that of domesticated plants (Kim *et al.*, 2020). However, the extent to which different flowering times, duration of plant cycles, and even changes in reproductive strategy affect the seed microbiome assembly remains to be elucidated. It is also important to bear in mind that

the ability of microorganisms to reach the seed via the vascular system, or other routes, is not sufficient for these organisms to become established in the seed, as antimicrobial proteins and secondary metabolites present in the developing seed limit microbial colonisation. Therefore, another unknown effect acting on the seed microbiome both pre- and post-dispersal is whether the reduction in secondary metabolites often observed in seeds of domesticated crops (Wink, 1988; Shlichta *et al.*, 2018), along with other changes in seed chemistry influences seed microbiome composition and viability. As many plant pathogenic microorganisms are transmitted via seed, it will be particularly important in future studies to investigate whether changes in seed properties associated with domestication positively or negatively affect the ability of specific microorganisms to survive the dispersal stage and be transmitted to the emerging seedling, or even to transition along the free-living/parasite/mutualist continuum (Drew *et al.*, 2021).

Another aspect of plant domestication which may influence the final pre-dispersal assembly of the seed microbiome is grafting. Grafting has been used for thousands of years to foster the domestication of fruit trees (Lee *et al.*, 2010). Many domesticated plants are grown as grafted plants (Lee *et al.*, 2010), such as several horticultural annual species and fruit tree species, including grapevine. It has been shown that rootstock type influences the rhizosphere and root endosphere microbiome in grapevine (Marasco *et al.*, 2018) and tomatoes (Poudel *et al.*, 2019), possibly leading to changes in microbial connectivity between plant compartments. Grafting can occur in nature, but it is used as a systematic agronomic practice in many domesticated species, thus potentially contributing as a factor steering the seed microbiome assembly by changing the microbial connectivity between plant compartments.

b) Host-to-microbe effects of plant domestication on plant-microbe interaction in the spermosphere (post-seed dispersal)

As mentioned above, modern crops usually have larger seeds compared to their wild ancestors, even in crops in which seeds are not directly consumed by humans or livestock, accompanied by changes in chemical composition. To what extent this change may have affected the seed-associated microbiome and plant-microbe interactions in the spermosphere, remains unknown. It is not within the scope of this article to discuss why domestication has led to plants with bigger seed mass, but this is thought to be the result of both indirect and artificial selection, and is observed in both grain crops and in crops that are not harvested for their grains. Indeed, a seed mass increase has been reported for lettuce, carrots and potatoes, and is thought to have been partially driven by unconscious selection in which indirect selection was exerted by agricultural practices (Kluyver *et al.*, 2017).

As a result, domestication has altered resource allocation and the trade-off relationship between seed mass (SM) and seed output (SO) (Sadras, 2007; Martin, 2018). If a fixed amount of energy has to be devoted towards the production of seeds, then a plant can either produce a large number of small seeds, or a small number of larger seeds (Smith and Fretwell, 1974). Overall, domesticated crops tend to have larger SM and smaller SO compared to wild plants (Martin, 2018) (for further reading on the ecological implications of seed size see (Turnbull *et al.*, 1999, 2004; Rodríguez-Gironés *et al.*, 2003)). We speculate that this change in seed size is likely to play an important role in determining the spatial-temporal dynamics of plant-microbe interactions in the spermosphere. The concept of the spermosphere was fully developed by Onorato Verona in 1958 (O, 1958) and since then, numerous studies have investigated the importance of this dynamic plant

compartment (for comprehensive reviews see (Nelson, 2004, 2018)). Eric Nelson, in his review (Nelson, 2004) wrote: "*The carbon released from seeds during germination represents the major driving force behind plant-microbe and microbe-microbe interactions in the spermosphere*". Therefore, seed size, which may be directly associated with differences in the amount and/or the nature of carbon released, may directly affect the microbial interactions of the spermosphere. Additionally, if seed enlargement, driven by domestication, alters the dynamics of water uptake, in turn, this could influence the deposition of seed exudates. In soybean, small seeds imbibe and germinate faster compared to larger seeds, regardless of the moisture levels (Edwards and Hartwig, 1971). Similar negative correlation trends between seed size, water uptake and germination time were also observed for maize (Muchena and Grogan, 1977; Schneider, 1998) and potatoes (Bhatt *et al.*, 1989).

A more detailed study investigated why smaller seeds have a higher percentage of water uptake in soybeans (Calero *et al.*, 1981). The authors reported that in soybean smaller seeds have large rounded pores whereas medium and large seeds have smaller and elongated pores. Pores in small seeds were open and functional, whereas they were distorted and plugged in large seeds. As microbial interactions in the spermosphere are influenced by seed exudates (Nelson, 2004), the release of which depends on imbibition (Simon and Harun, 1972; Simon and Mathavan, 1986), we speculate that seed and pore size could affect the dynamics of these interactions.

We are still far from understanding what compounds, present in seed exudates, are used by microorganisms to proliferate in the spermosphere. Nonetheless, we know that the spores of *Pythium ultimum*, a seed-infecting pathogenic oomycete, are stimulated to germinate rapidly in the presence of oleic and linoleic acids, which are fatty acids

commonly found in seed exudates (Nelson, 2018). Also, the quantitative aspects of seed exudation are important. For example, *Enterobacter cloacae*, a spermosphere-colonizing bacterium, has higher metabolic activity on pea seed exudates than on cucumber seed exudates. The latter has an order of magnitude fewer nutrients (Roberts *et al.*, 2009). Barret (Barret *et al.*, 2015) reported that germination influences the structure of the microbiota and that the observed shift in composition between seed microbiota and seedling microbiota was associated with a higher relative abundance of bacterial (*Bacillus*, *Massilia*, *Pantoea*, and *Pseudomonas*) and fungal (*Trichodermam*, *Chaetomium*) taxa that exhibit fast growth rates.

Together with seed size, seed color, which is linked to morphological features, has been linked to the dynamics of seed imbibition. For example, in proso millet (*Panicum miliaceum*), darker seeds have heavier seed coats, imbibe, and germinate more slowly and suffer less imbibition damage (Khan *et al.*, 1997). In *Phaseolus vulgaris*, cultivars with white seed coats imbibe faster and have higher solute leakage compare with seeds with a darker color (Powell *et al.*, 1986). As domestication has not only caused seed enlargement but also expanded the color range of seeds (Meyer and Purugganan, 2013), and both seed size and color affect imbibition and germination, they should be considered together when assessing whether there is an effect on the spermosphere microbiome. We propose that one of the consequences of the production of large seeds, which may have slower imbibition and germination rates compared to smaller seeds, is different temporal dynamics of spermosphere-microbe interactions. The consequences of these different temporal dynamics of spermosphere colonization for seedling establishment and plant growth remains unexplored.

Similar to seed size and seed color, seed chemical composition has also been affected by domestication. For example, domesticated lima bean (*Phaseolus lunatus*) seeds have been reported to have up to 20 times lower concentrations of cyanogenic glycosides than seeds of wild populations (Shlichta *et al.*, 2018), and legume seeds of domesticated plants have fewer carotenoids than their wild counterparts (Fernández-Marín *et al.*, 2014). Moreover, mineral composition can also differ between wild and domesticated seeds. For example, calcium concentration was found to be negatively correlated with seed mass in common bean seeds (Moraghan and Grafton, 2001) and minerals have been reported to influence microbiome assembly in the rhizosphere (Whitman *et al.*, 2018).

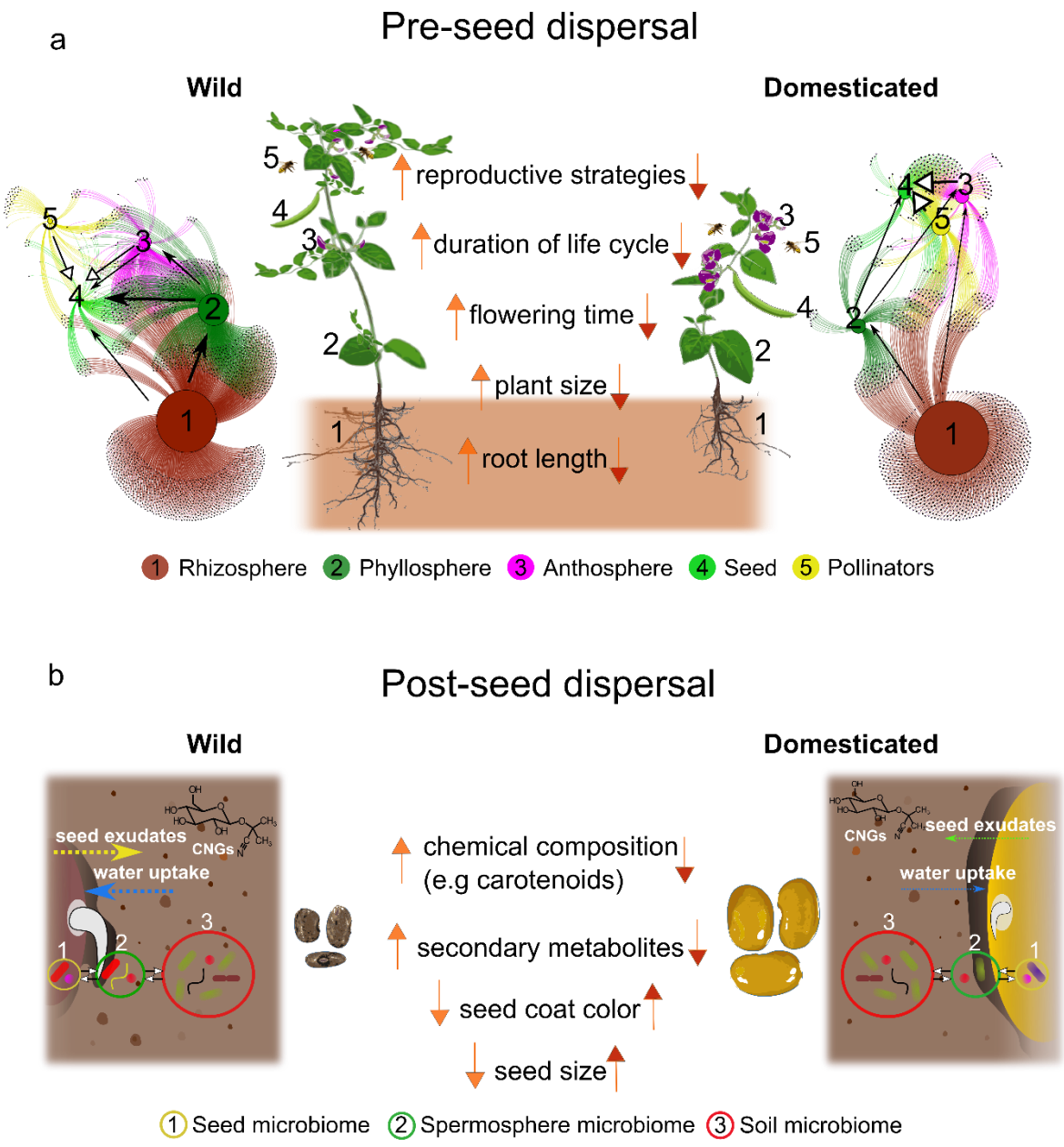


Figure 2. Phenotypic changes introduced with plant domestication have consequences for the microbial connectivity between different plant compartments and the spatial and temporal dynamics of plant-microbe interactions in the spermosphere. a: Pre-seed dispersal: Phenotypic changes caused by plant domestication affect both the below and above-ground plant compartments. In wild plants, we propose that the microbial connectivity (sharing of microbial members) between the environment and plant compartments, visualized as a bipartite microbial network, could reflect a higher degree of vertical transmission of microorganisms to plant seeds

(represented by wider black arrows), under the influence of host microbiome control mechanisms. Vertical transmission can occur from the rhizosphere (1) to the seed compartment (4); from the rhizosphere (1) and phyllosphere (2) to the seed compartment (4); or from the rhizosphere (1), phyllosphere (2) and anthosphere (3) to the seed compartment (4). Horizontal transmission (white arrow) can occur from plant to plant in the rhizosphere and phyllosphere via air-borne, soil-borne and water-borne inocula, vectors, and mechanical transmission (not illustrated), from pollinators (5) to the seed compartment (4) or from the anthosphere (3) to the seed compartment (4). We speculate that domesticated plants exhibit differences in microbial connectivity compared to their wild counterparts. In this example, we predict higher importance of horizontal transmission (represented by wider white arrows) in the seed microbiome assembly. b: Post-seed dispersal: Phenotypic changes caused by plant domestication can affect the spatial and temporal dynamics of plant-microbe interactions in the spermosphere. Wild plants have smaller seed sizes and different colors of the seed testa compared to domesticated plants. Water uptake may be negatively correlated with seed or seed pore size and seed colour is linked to morphological features of the seed testa influencing imbibition rate. The dynamics of seed germination, imbibition and seed exudation differ from domesticated plants. Secondary metabolites such as cyanogenic glycosides (CNGs) and antimicrobial proteins may also be more abundant or present at higher concentrations in seeds of wild plants. We speculate that these phenotypic changes are likely to lead to different spatial and temporal dynamics of interactions between the seed microbiome (1), the spermosphere microbiome (2), and the soil microbiome (3). In the example shown, seeds of domesticated plants germinate slowly, resulting in a different spermosphere microbial community compared to wild plant seeds. This situation could be reversed for seeds of wild plants that manifest dormancy. In this case, seeds of domesticated plants would germinate faster, regardless of seed size.

Testing the double-leash framework

In general, our double-leash hypothesis predicts that domesticated plant traits, either under direct or indirect selection, explain the difference in ecological processes driving community assembly between wild and domesticated plants. The effects of domesticated plant traits resulting from direct and indirect selection on host-microbiome assembly could lead to species loss, species gain, or species replacement (Figure 1c). To test the double-leash hypothesis, researchers could perform experiments in which wild and domesticated plants are grown in a greenhouse with non-native soil (to avoid an effect of plant adaptation to agricultural or natural ecosystems). The selection of wild and domesticated plants should be performed to provide a wide range of phenotypic traits or researchers could create experimental populations (*e.g.* from a cross between a domesticated and wild accession of the same species) that are segregating for a certain domestication trait. For example, to study the effect of seed size on seed microbiome assembly, accessions of both wild and domesticated plants should be selected that showcase a wide range of seed sizes.

The double leash could then be tested by looking at significant correlations between the plant phenotype distance matrix and the distance matrix calculated from the relative importance of ecological processes driving microbiome assembly. This approach has been already used to understand the environmental factors driving microbiome assembly in grasslands as a response to warming (Ning *et al.*, 2020). If the plant trait(s) under selection during domestication (*e.g* flowering time, root length, seed size) is significantly correlated with the observed relative importance of selection, then the double-leash hypothesis is supported.

This approach, in which ecological processes driving microbial community assembly are broken down into their components, namely stochasticity and selection, and explained by domesticated plant phenotypes, provides a quantitative approach to identify the causes of differences in microbiome assembly and thereby test the double-leash hypothesis. However, it should be noted that this approach rests on the assumption that host selection acts on microbial community assembly. The double-leash hypothesis would not predict the outcome of assembly processes mainly governed by stochasticity (weak selection), for example when hosts exert low selection on microbial communities (weak host-to-microbe effects).

Biotechnological implications of the double-leash framework

Understanding how to re-introduce lost microbial species that exert positive microbe-to host effects in domesticated plants is an important and current area of research (Liu *et al.*, 2020; Porter and Sachs, 2020). Porter *et al* (Porter and Sachs, 2020) suggest the possibility to re-introduce into domesticated plant genomes loci responsible for symbiosis, similarly to the introgression of genes responsible for disease resistance (Arora *et al.*, 2019), which have been lost during domestication. Similarly, other studies suggest acting at a plant genotypic level to re-establish positive plant-microbe interactions (Pérez-Jaramillo *et al.*, 2016). Our framework, focusing on the effect of plant phenotypes on driving microbiome assembly, does not focus on microbe-to-host effects but tries to understand whether a certain plant phenotype affects the selection of microbial species. This suggests an alternative approach for enhancing plant productivity, based on the identification of microbial groups that are under positive selection in the context of the double leash, which can be screened or engineered for positive-host-to-microbe effects. While this approach

presents several limitations at the moment, including challenges associated with the release of GMOs in the environment, advances in bacterial genome editing and biocontainment make this an increasingly feasible approach for developing synthetic plant growth-promoting communities.

Conclusion and future directions

Domesticated plants may have lost, or show a reduction in the ability to establish beneficial associations with microbial communities through changes in host-microbe or microbe-microbe interactions (Hetrick *et al.*, 1992; Sangabriel-Conde *et al.*, 2014). However, how host control of the microbiota is influenced by domestication through different host-to-microbe effects remains poorly understood. We advocate that by extending the concept of the host microbiome as an ecosystem on a leash to a double leash framework that includes the domesticator and the consequences of direct and indirect selection, host-to-microbe effects can be framed in a more comprehensive perspective for understanding their role in determining the assembly of domesticated plant microbiomes. Moreover, our double-leash hypothesis could lead to the identification of microbial species selected by host-to-microbe effects driven by domesticated plant phenotypes that could be applied or engineered for positive microbe-to-host effects. To that end, the seed compartment is not only emblematic in terms of the phenotypic changes caused by domestication but also an ideal candidate to investigate the consequences and potential for exploitation of host-to-microbe effects. In the long term, unveiling the mechanism and consequences of domestication of the plant microbiome, and in particular the seed microbiome, can help to explain new adaptative patterns of plants and provide new insights toward more sustainable management of arable land.

References

Abdullaeva, Y., Ambika Manirajan, B., Honermeier, B., Schnell, S., and Cardinale, M. (2020) Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *Journal of Advanced Research*.

Aizenberg-Gershtein, Y., Izhaki, I., and Halpern, M. (2013) Do Honeybees Shape the Bacterial Community Composition in Floral Nectar? *PLoS One* 8:.

Aleklett, K., Hart, M., and Shade, A. (2014) The microbial ecology of flowers: an emerging frontier in phyllosphere research. *Botany* 92: 253–266.

Alibrandi, P., Cardinale, M., Rahman, M.M., Strati, F., Ciná, P., de Viana, M.L., et al. (2018) The seed endosphere of *Anadenanthera colubrina* is inhabited by a complex microbiota, including *Methylobacterium* spp. and *Staphylococcus* spp. with potential plant-growth promoting activities. *Plant Soil* 422: 81–99.

Ambika Manirajan, B., Ratering, S., Rusch, V., Schwiertz, A., Geissler-Plaum, R., Cardinale, M., and Schnell, S. (2016) Bacterial microbiota associated with flower pollen is influenced by pollination type, and shows a high degree of diversity and species-specificity. *Environ Microbiol* 18: 5161–5174.

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., et al. (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology* 37: 139–143.

Barret, M., Briand, M., Bonneau, S., Préveaux, A., Valière, S., Bouchez, O., et al. (2015) Emergence Shapes the Structure of the Seed Microbiota. *Appl Environ Microbiol* 81: 1257–1266.

Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M., et al. (2016) Evolutionary Metabolomics Reveals Domestication-Associated Changes in Tetraploid Wheat Kernels. *Molecular Biology and Evolution* 33: 1740–1753.

Berendsen, R.L., Pieterse, C.M.J., and Bakker, P.A.H.M. (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci* 17: 478–486.

Berg, G. and Raaijmakers, J.M. (2018) Saving seed microbiomes. *The ISME Journal* 12: 1167–1170.

Bhatt, A.K., Bhalla, T.C., Agrawal, H.O., and Upadhya, M.D. (1989) Effect of seed size on protein and lipid contents, germination and imbibition in true potato seeds. *Potato Res* 32: 477–481.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., et al. (2015) Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host & Microbe* 17: 392–403.

Calero, E., West, S.H., and Hinson, K. (1981) Water Absorption of Soybean Seeds ans Associated Causal Factors1. *Crop Science* 21: cropsci1981.0011183X002100060030x.

Chen, Y.H., Gols, R., and Benrey, B. (2015) Crop Domestication and Its Impact on Naturally Selected Trophic Interactions. *Annu Rev Entomol* 60: 35–58.

Cheplick, G.P. (2008) Host genotype overrides fungal endophyte infection in influencing tiller and spike production of *Lolium perenne* (Poaceae) in a common garden experiment. *American Journal of Botany* 95: 1063–1071.

Compant, S., Mitter, B., Colli-Mull, J.G., Gangl, H., and Sessitsch, A. (2011) Endophytes of grapevine flowers, berries, and seeds: identification of cultivable

bacteria, comparison with other plant parts, and visualization of niches of colonization. *Microb Ecol* 62: 188–197.

Darwin, C. R (1868) The variation of animals and plants under domestication. London: John Murray. First edition, first issue. Volume 2.

Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418: 700–707.

Diamond, J.M. (1998) Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years, Random House.

Dirzo, R. and Raven, P.H. (2003) Global State of Biodiversity and Loss. *Annual Review of Environment and Resources* 28: 137–167.

Donati, I., Cellini, A., Buriani, G., Mauri, S., Kay, C., Tacconi, G., and Spinelli, F. (2018) Pathways of flower infection and pollen-mediated dispersion of *Pseudomonas syringae* pv. *actinidiae*, the causal agent of kiwifruit bacterial canker. *Hortic Res* 5:.

Douglas, A.E. and Werren, J.H. (2016) Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *mBio* 7:.

Dr Th. H. Engelbrecht's views on the origin of cultivated plants (1973) *Euphytica* 22: 279–286.

Drew, G.C., Stevens, E.J., and King, K.C. (2021) Microbial evolution and transitions along the parasite–mutualist continuum. *Nat Rev Microbiol* 1–16.

Dutta, B., Gitaitis, R., Sanders, H., Booth, C., Smith, S., and Langston, D.B. (2014) Role of Blossom Colonization in Pepper Seed Infestation by *Xanthomonas euvesicatoria*. *Phytopathology* 104: 232–239.

Dutta, B., Schneider, R.W., Robertson, C.L., and Walcott, R.R. (2016) Embryo Localization Enhances the Survival of *Acidovorax citrulli* in Watermelon Seeds. *Phytopathology* 106: 330–338.

Edwards, C.J. and Hartwig, E.E. (1971) Effect of Seed Size upon Rate of Germination in Soybeans1. *Agronomy Journal* 63: 429–450.

Egan, P.A., Adler, L.S., Irwin, R.E., Farrell, I.W., Palmer-Young, E.C., and Stevenson, P.C. (2018) Crop Domestication Alters Floral Reward Chemistry With Potential Consequences for Pollinator Health. *Front Plant Sci* 9:.

FAO (1970) The state of food and agriculture. 289.

Fernández-Marín, B., Milla, R., Martín-Robles, N., Arc, E., Kranner, I., Becerril, J.M., and García-Plazaola, J.I. (2014) Side-effects of domestication: cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than their wild counterparts. *BMC Plant Biol* 14:.

Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017) The evolution of the host microbiome as an ecosystem on a leash. *Nature* 548: 43–51.

Hammer, K. (1984) Das Domestikationssyndrom. *Die Kulturpflanze* 32: 11–34.

Haney, C.H., Samuel, B.S., Bush, J., and Ausubel, F.M. (2015) Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants* 1: 1–9.

Hartmann, M., Frey, B., Mayer, J., Mäder, P., and Widmer, F. (2015) Distinct soil microbial diversity under long-term organic and conventional farming. *The ISME Journal* 9: 1177–1194.

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

Hendgen, M., Hoppe, B., Döring, J., Friedel, M., Kauer, R., Frisch, M., et al. (2018) Effects of different management regimes on microbial biodiversity in vineyard soils. *Scientific Reports* 8: 9393.

Hetrick, B.A.D., Wilson, G.W.T., and Cox, T.S. (1992) Mycorrhizal dependence of modern wheat varieties, landraces, and ancestors. *Can J Bot* 70: 2032–2040.

Johnston-Monje, D., Lundberg, D.S., Lazarovits, G., Reis, V.M., and Raizada, M.N. (2016) Bacterial populations in juvenile maize rhizospheres originate from both seed and soil. *Plant Soil* 405: 337–355.

Johnston-Monje, D. and Raizada, M.N. (2011) Conservation and Diversity of Seed Associated Endophytes in Zea across Boundaries of Evolution, Ethnography and Ecology. *PLOS ONE* 6: e20396.

Kantar, M.B., Nashoba, A.R., Anderson, J.E., Blackman, B.K., and Rieseberg, L.H. (2017) The Genetics and Genomics of Plant Domestication. *BioScience* 67: 971–982.

Kavamura, V.N., Robinson, R.J., Hughes, D., Clark, I., Rossmann, M., Melo, I.S. de, et al. (2020) Wheat dwarfing influences selection of the rhizosphere microbiome. *Scientific Reports* 10: 1–11.

Khan, M., Cavers, P.B., Kane, M., and Thompson, K. (1997) Role of the pigmented seed coat of proso millet (*Panicum miliaceum* L.) in imbibition, germination and seed persistence. *Seed Science Research* 7: 21–26.

Kiers, E.T., Hutton, M.G., and Denison, R.F. (2007) Human selection and the relaxation of legume defences against ineffective rhizobia. *Proceedings of the Royal Society B: Biological Sciences* 274: 3119–3126.

Kim, H., Lee, K.K., Jeon, J., Harris, W.A., and Lee, Y.-H. (2020) Domestication of Oryza species eco-evolutionarily shapes bacterial and fungal communities in rice seed. *Microbiome* 8: 20.

Kluyver, T.A., Jones, G., Pujol, B., Bennett, C., Mockford, E.J., Charles, M., et al. (2017) Unconscious selection drove seed enlargement in vegetable crops. *Evolution Letters* 1: 64–72.

Lee, J.-M., Kubota, C., Tsao, S.J., Bie, Z., Echevarria, P.H., Morra, L., and Oda, M. (2010) Current status of vegetable grafting: Diffusion, grafting techniques, automation. *Scientia Horticulturae* 127: 93–105.

Liu, J., Yu, X., Qin, Q., Dinkins, R.D., and Zhu, H. (2020) The Impacts of Domestication and Breeding on Nitrogen Fixation Symbiosis in Legumes. *Front Genet* 11:.

Liu, Z., Garcia, A., McMullen, M.D., and Flint-Garcia, S.A. (2016) Genetic Analysis of Kernel Traits in Maize-Teosinte Introgression Populations. *G3: Genes, Genomes, Genetics* 6: 2523–2530.

Mandal, S. and Mandal, R.K. (2000) Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. *Current Science* 79: 576–589.

Marasco, R., Rolli, E., Fusi, M., Michoud, G., and Daffonchio, D. (2018) Grapevine rootstocks shape underground bacterial microbiome and networking but not potential functionality. *Microbiome* 6: 3.

Martin, A.R. (2018) Crops and the seed mass-seed output trade-off in plants, Ecology.

Matsumoto, H., Fan, X., Wang, Yue, Kusstatscher, P., Duan, J., Wu, S., et al. (2021) Bacterial seed endophyte shapes disease resistance in rice. *Nature Plants* 7: 60–72.

McFrederick, Q.S., Thomas, J.M., Neff, J.L., Vuong, H.Q., Russell, K.A., Hale, A.R., and Mueller, U.G. (2017) Flowers and Wild Megachilid Bees Share Microbes. *Microb Ecol* 73: 188–200.

Meyer, R.S., DuVal, A.E., and Jensen, H.R. (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* 196: 29–48.

Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* 14: 840–852.

Mitter, B., Pfaffenbichler, N., Flavell, R., Compant, S., Antonielli, L., Petric, A., et al. (2017) A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. *Front Microbiol* 8:.

Moraghan, J.T. and Grafton, K. (2001) Genetic diversity and mineral composition of common bean seed. *Journal of the Science of Food and Agriculture* 81: 404–408.

Morse, L.J., Faeth, S.H., and Day, T.A. (2007) Neotyphodium interactions with a wild grass are driven mainly by endophyte haplotype. *Functional Ecology* 21: 813–822.

Muchena, S.C. and Grogan, C.O. (1977) Effects of Seed Size on Germination of Corn (zea Mays) Under Simulated Water Stress Conditions. *Can J Plant Sci* 57: 921–923.

Nelson, E.B. (2004) Microbial dynamics and interactions in the spermosphere. *Annu Rev Phytopathol* 42: 271–309.

Nelson, E.B. (2018) The seed microbiome: Origins, interactions, and impacts. *Plant Soil* 422: 7–34.

Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al. (2020) A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nature Communications* 11: 4717.

O, V. (1958) [The spermosphere]. *Ann Inst Pasteur (Paris)* 95: 795–798.

Opstal, E.J. van and Bordenstein, S.R. (2015) Rethinking heritability of the microbiome. *Science* 349: 1172–1173.

Özkurt, E., Hassani, M.A., Sesiz, U., Künzel, S., Dagan, T., Özkan, H., and Stukenbrock, E.H. (2020) Seed-Derived Microbial Colonization of Wild Emmer and Domesticated Bread Wheat (*Triticum dicoccoides* and *T. aestivum*) Seedlings Shows Pronounced Differences in Overall Diversity and Composition. *mBio* 11:.

Pérez-Jaramillo, J.E., Carrión, V.J., Bosse, M., Ferrão, L.F.V., de Hollander, M., Garcia, A.A.F., et al. (2017) Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J* 11: 2244–2257.

Pérez-Jaramillo, J.E., Carrión, V.J., de Hollander, M., and Raaijmakers, J.M. (2018) The wild side of plant microbiomes. *Microbiome* 6: 143.

Pérez-Jaramillo, J.E., Mendes, R., and Raaijmakers, J.M. (2016) Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Molecular Biology* 90: 635–644.

Piperno, D.R. (2011) The Origins of Plant Cultivation and Domestication in the New World Tropics: Patterns, Process, and New Developments. *Current Anthropology* 52: S453–S470.

Pitzschke, A. (2018) Molecular dynamics in germinating, endophyte-colonized quinoa seeds. *Plant Soil* 422: 135–154.

Porter, S.S. and Sachs, J.L. (2020) Agriculture and the Disruption of Plant–Microbial Symbiosis. *Trends in Ecology & Evolution* 35: 426–439.

Postler, T.S. and Ghosh, S. (2017) Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System. *Cell Metab* 26: 110–130.

Poudel, R., Jumpponen, A., Kennelly, M.M., Rivard, C.L., Gomez-Montano, L., and Garrett, K.A. (2019) Rootstocks Shape the Rhizobiome: Rhizosphere and Endosphere Bacterial Communities in the Grafted Tomato System. *Appl Environ Microbiol* 85: e01765-18.

Powell, A.A., Oliveira, M.D.A., and Matthews, S. (1986) Seed vigour in cultivars of dwarf French bean (*Phaseolus vulgaris*) in relation to the colour of the testa. *The Journal of Agricultural Science* 106: 419–425.

Prado, A., Marolleau, B., Vaissière, B.E., Barret, M., and Torres-Cortes, G. (2020) Insect pollination: an ecological process involved in the assembly of the seed microbiota. *Scientific Reports* 10: 1–11.

Puente, M.E., Li, C.Y., and Bashan, Y. (2009) Endophytic bacteria in cacti seeds can improve the development of cactus seedlings. *Environmental and Experimental Botany* 66: 402–408.

Purugganan, M.D. (2019) Evolutionary Insights into the Nature of Plant Domestication. *Current Biology* 29: R705–R714.

Roberts, D.P., Baker, C.J., McKenna, L., Liu, S., Buyer, J.S., and Kobayashi, D.Y. (2009) Influence of host seed on metabolic activity of *Enterobacter cloacae* in the spermosphere. *Soil Biology and Biochemistry* 41: 754–761.

Rochefort, A., Briand, M., Marais, C., Wagner, M.-H., Laperche, A., Vallée, P., et al. (2019) Influence of Environment and Host Plant Genotype on the Structure and Diversity of the *Brassica napus* Seed Microbiota. *Phytobiomes Journal* 3: 326–336.

Rodríguez, C.E., Antonielli, L., Mitter, B., Trognitz, F., and Sessitsch, A. (2019) Heritability and Functional Importance of the *Setaria viridis* Bacterial Seed Microbiome. *Phytobiomes Journal* 4: 40–52.

Rodríguez, C.E., Mitter, B., Barret, M., Sessitsch, A., and Compant, S. (2018) Commentary: seed bacterial inhabitants and their routes of colonization. *Plant Soil* 422: 129–134.

Rodríguez-Gironés, M.A., Sandsten, H., and Santamaría, L. (2003) Asymmetric competition and the evolution of propagule size. *Journal of Ecology* 91: 554–562.

Ross-Ibarra, J., Morrell, P.L., and Gaut, B.S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences* 104: 8641–8648.

Sadras, V.O. (2007) Evolutionary aspects of the trade-off between seed size and number in crops. *Field Crops Research* 100: 125–138.

Sangabriel-Conde, W., Negrete-Yankelevich, S., Maldonado-Mendoza, I.E., and Trejo-Aguilar, D. (2014) Native maize landraces from Los Tuxtlas, Mexico show varying mycorrhizal dependency for P uptake. *Biol Fertil Soils* 50: 405–414.

Schluter, J. and Foster, K.R. (2012) The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. *PLoS Biol* 10:.

Schneider, A. (1998) Variability of maize seed imbibition rates as influenced by seed size distribution and coating application. *Agronomie* 18: 247–260.

Shade, A., Jacques, M.-A., and Barret, M. (2017) Ecological patterns of seed microbiome diversity, transmission, and assembly. *Current Opinion in Microbiology* 37: 15–22.

Shahzad, R., Khan, A.L., Bilal, S., Asaf, S., and Lee, I.-J. (2018) What Is There in Seeds? Vertically Transmitted Endophytic Resources for Sustainable Improvement in Plant Growth. *Front Plant Sci* 9:.

Shlichta, J.G., Cuny, M.A.C., Hernandez-Cumplido, J., Traine, J., and Benrey, B. (2018) Contrasting consequences of plant domestication for the chemical defenses of leaves and seeds in lima bean plants. *Basic and Applied Ecology* 31: 10–20.

Simon, E.W. and Harun, R.M.R. (1972) Leakage during Seed Imbibition. *J Exp Bot* 23: 1076–1085.

Simon, E.W. and Mathavan, S. (1986) The time-course of leakage from imbibing seeds of different species. *Seed sci technol* 14: 9–13.

Smith, C.C. and Fretwell, S.D. (1974) The Optimal Balance between Size and Number of Offspring. *The American Naturalist* 108: 499–506.

Soldan, R., Mapelli, F., Crotti, E., Schnell, S., Daffonchio, D., Marasco, R., et al. (2019) Bacterial endophytes of mangrove propagules elicit early establishment of the natural host and promote growth of cereal crops under salt stress. *Microbiological Research* 223–225: 33–43.

Szoboszlay, M., Lambers, J., Chappell, J., Kupper, J.V., Moe, L.A., and McNear, D.H. (2015) Comparison of root system architecture and rhizosphere microbial communities of Balsas teosinte and domesticated corn cultivars. *Soil Biology and Biochemistry* 80: 34–44.

Truyens, S., Weyens, N., Cuypers, A., and Vangronsveld, J. (2015) Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports* 7: 40–50.

Turnbull, L.A., Coomes, D., Hector, A., and Rees, M. (2004) Seed mass and the competition/colonization trade-off: competitive interactions and spatial patterns in a guild of annual plants. *Journal of Ecology* 92: 97–109.

Turnbull, L.A., Rees, M., and Crawley, M.J. (1999) Seed mass and the competition/colonization trade-off: a sowing experiment. *Journal of Ecology* 87: 899–912.

Valdez, R.A. (2019) Land Use History Influences Recruitment Of Soilborne Microbes To Seeds, And Seed Germination, Of *Prosopis Velutina* (velvet Mesquite).

Vannier, N., Mony, C., Bittebiere, A.-K., Michon-Coudouel, S., Biget, M., and Vandenkoornhuyse, P. (2018) A microorganisms' journey between plant generations. *Microbiome* 6: 79.

Vavilov, N.I., Vavylov, M.I., Vavílov, N.Í., and Dorofeev, V.F. (1992) Origin and Geography of Cultivated Plants, Cambridge University Press.

Verma, S.K., Kharwar, R.N., and White, J.F. (2019) The role of seed-vectored endophytes in seedling development and establishment. *Symbiosis*.

Vries, F.T. de, Griffiths, R.I., Knight, C.G., Nicolitch, O., and Williams, A. (2020) Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* 368: 270–274.

Walitang, D.I., Kim, C.-G., Jeon, S., Kang, Y., and Sa, T. (2019) Conservation and transmission of seed bacterial endophytes across generations following crossbreeding and repeated inbreeding of rice at different geographic locations. *MicrobiologyOpen* 8: e00662.

Wei, Z. and Jousset, A. (2017) Plant Breeding Goes Microbial. *Trends in Plant Science* 22: 555–558.

Whitman, T., Neurath, R., Perera, A., Chu-Jacoby, I., Ning, D., Zhou, J., et al. (2018) Microbial community assembly differs across minerals in a rhizosphere microcosm. *Environmental Microbiology* 20: 4444–4460.

Wiewióra, B., Żurek, G., and Pańka, D. (2015) Is the Vertical Transmission of Neotyphodium lolii in Perennial Ryegrass the Only Possible Way to the Spread of Endophytes? *PLoS ONE* 10:.

Wink, M. (1988) Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theoret Appl Genetics* 75: 225–233.

Zarraonaindia, I., Owens, S.M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., et al. (2015) The Soil Microbiome Influences Grapevine-Associated Microbiota. *mBio* 6:.

Zohary, D. (2004) Unconscious selection and the evolution of domesticated Plants.

Econ Bot 58: 5–10.

Chapter 3: Consistent effects of independent domestication events on seed microbiome assembly and composition

Riccardo Soldan*, Marco Fusi, Massimiliano Cardinale, Felix Homma, Luis Guillermo Santos, Peter Wenzl, Macel Bach-Pages, Elena Bitocchi, Maria Isabel Chacon Sanchez, Daniele Daffonchio, Gail M. Preston*

*Corresponding authors

Manuscript under submission (Proceedings of the National Academy of Sciences)

Abstract

Plant domestication has been decisive in the development of human societies. However, the consequences of plant domestication at the ecosystem level are still poorly understood. Among different organisms that live in association with plants, microbial communities exert important biological and ecological functions. Whether and how plant domestication changed plant-microbe interactions remains difficult to prove but could have important repercussions for sustainable agriculture. In this study, we investigate whether there is a domestication effect on plant-microbe interactions and the processes that drive the assembly of the microbiome in wild and domesticated plants. To achieve this, we focused on *Phaseolus*, which is one of the few plant genera that have undergone four independent domestication events within two species (*Phaseolus vulgaris* and *Phaseolus lunatus*), providing us with a replication of an event lasting thousands of years. We specifically examined the effect of domestication on seed-associated microbial communities, as the seed is a plant compartment that has been subject to large phenotypic changes during domestication, making it an ideal candidate to test a domestication effect on the microbiome. We analysed the seed bacteriome of wild accessions of *P. vulgaris* and *P. lunatus* and accessions derived from all four domestication events and observed evidence of a common domestication effect on the seed microbiome, which can be explained, at least partially, by domesticated plant phenotypes. Our results provide evidence of a convergent evolution of domesticated plant phenotypes and seed microbiome assembly leading to a reduction in ecological selective forces during plant domestication.

Significance statement

Plant domestication has been one of the most important achievements in human history. However, our understanding of its impact on the associations between plants and microorganisms remains limited, partially because it is difficult replicate an event lasting thousands of years. We conceived an experimental design that allowed us to use multiple independent domestication events for two plant species to show that the impact of plant domestication on the seed microbiome is consistent for independent domestication events. We provide evidence that this is partially due to similar plant phenotypic traits that are selected by the domesticator regardless of the domestication event. Our results also suggest that seed traits selected directly or indirectly by domesticators, such as increased seed size, reduced concentrations of antimicrobial compounds and changes in mineral content, may have contributed to a reduction in the extent to which plants are able to control the composition of the seed microbiota.

Introduction

Over the last 13,000 years of human history, one notable human-driven process has allowed the rise of civilization as we know it today, namely the domestication of plant and animal species (Diamond, 1998). Virtually all vegetables, grains and fruits we currently rely upon come from domesticated plants (Diamond, 2002). One remarkable feature of domesticated plants is the set of common phenotypes that often distinguish them from their wild counterparts. This pool of common traits is called domestication syndrome (Hammer, 1984) and includes loss of seed shattering in the case of grain crops, diminished secondary metabolite content, and changes in reproductive strategies and branching (Meyer *et al.*, 2012). However, while we have advanced our knowledge of the origin of domestication (Meyer and Purugganan, 2013) and its genetic basis (Kantar *et al.*, 2017), we still know little about the consequences of plant domestication for interacting organisms (Chen *et al.*, 2015; Fernandez *et al.*, 2021).

In recent years researchers have come to appreciate the importance of the microbial community that lives in close association with plants (Trivedi *et al.*, 2020). Its influence on host growth and development is so important that host and host-microbiome have been defined as a meta-organism, called a holobiont (Vandenkoornhuyse *et al.*). As plant domestication has influenced plant phenotypes, and phenotypes, when heritable, can affect the community of interacting organisms (Whitham *et al.*, 2003), a potential effect of domesticated plant phenotypes on host-associated microbial communities may be expected.

In a recent perspective we conceptualised this effect as a “double-leash” acting from domesticator to host and host to microbes (Soldan *et al.*, 2021). Foster *et. al* (Foster *et al.*,

2017) first conceived the idea that host-associated microbial communities are ecosystems "*held on an ever-evolving leash by the host*", as the host is under strong selective pressure to favour beneficial interactions with its microbiome, and the host can control the environment in which its microbiome lives. Consequently, we (Soldan *et al.*, 2021) expanded this framework to acknowledge the importance of artificial selection and natural selection acting in the agricultural ecosystem on host control of the microbiota. This newly extended framework predicts that microbiome assembly in domesticated crops could be highly influenced by domesticated plant phenotypes. The "double-leash" framework predicts that when the host microbiome is influenced by artificially selected plant traits, there could be a reduction in positive microbe-to-host effects.

Among different plant compartments influenced by domestication, the seed often showcases the most remarkable phenotypic changes (Soldan *et al.*, 2021). The seed microbiome has recently acquired importance in the field of plant-microbe interactions as it can influence disease resistance (Matsumoto *et al.*, 2021), plant germination (Puente *et al.*, 2009; Soldan *et al.*, 2019), and overall plant growth and development (Berg and Raaijmakers, 2018). For these reasons, we (Soldan *et al.*, 2021) suggested that the seed microbiome could hold the key to understanding the effect of plant domestication on microbiome assembly.

Evidence of domesticated plant phenotypes partially explaining microbial community assembly is found in the literature. Differences in root length, as well as plant size, have been correlated to microbiome composition in common bean and wheat (Pérez-Jaramillo *et al.*, 2017; Kavamura *et al.*, 2020). Nonetheless, clear evidence that plant domestication has driven changes in host-microbiome assembly through domesticated plant phenotypes remains challenging.

A noteworthy feature of the *Phaseolus* genus is that two of its species, namely *Phaseolus vulgaris* (common bean) (Bitocchi *et al.*, 2013; Bellucci *et al.*, 2014; Rodriguez *et al.*, 2016) and *P. lunatus* (Lima bean) (Chacón-Sánchez and Martínez-Castillo, 2017; Garcia *et al.*, 2021), have been domesticated at least twice independently (Bitocchi *et al.*, 2017; Garcia *et al.*, 2021), over two geographical areas in the Americas. As a result, for each plant species, two gene pools exist (Mesoamerican and Andean gene pools), containing both wild and domesticated accessions. This distinctive feature of four independent domestication events that have occurred within two species, offers the possibility of testing whether domesticated plant phenotypes have independently led to a similar effect on microbiome assembly.

In this paper we initially characterized the seed microbiome of wild and domesticated *Phaseolus vulgaris* plants in a greenhouse experiment for both domestication events, to test if a common selection effect is induced by domesticated plant phenotypes on seed microbiome assembly. We then expanded this analysis to seeds isolated from plant genotypes of *P. vulgaris* and *P. lunatus* obtained from the gene bank of the Alliance Biodiversity International and the International Center for Tropical Agriculture (CIAT), to examine whether the effects observed in the greenhouse experiment extend to seeds grown in diverse, non-controlled environments and transcend environmental influences. While plant accessions grown at CIAT are not propagated in controlled conditions, CIAT follows a standardised procedure for regenerating and conserving *Phaseolus* accessions. We detected consistent changes in members of the seed microbial community for both domestication events in both species, which significantly correlated with domesticated plant phenotypes. Our results provide evidence of convergent evolution of the

domesticated plant phenotypes and the seed microbiome assembly during plant domestication.

Results

Independent domestication events led to a consistent decrease in calcium concentration in *Phaseolus* spp. seeds.

We examined the effect of domestication on *Phaseolus* spp. seeds in three experiments. In experiment 1 (PVG) we analysed seed collected from 16 representative wild and domesticated accessions of *P. vulgaris* (one wild and one domesticated subpopulation per domestication scenario, four accessions per subpopulation. This encompass both domestication events) grown in controlled greenhouse conditions and collected directly from the pod. We subsequently analysed seeds from 70 wild and domesticated accessions of *P. vulgaris* (experiment 2 (PVC)) and 61 accessions of *P. lunatus* (experiment 3 (PLC)) obtained from CIAT, to check whether we could replicate the results obtained in experiment PVG when including more accessions and subpopulations and without controlling for the environment (CIAT grows plant accessions in field trial and different years). We analysed the seed mineral composition of accessions as a proxy for domesticated plant phenotypes (Supplementary Table 1, available from GitHub - see code availability section; Materials and Methods section). Mineral composition is important for human nutrition but it is also correlated with protein content (Silva *et al.*, 2012) and secondary metabolites (Sun *et al.*, 2019) and has been shown to be influenced by plant domestication (Celmeli *et al.*, 2018).

Principal component analysis (PCA) of z-standardised seed cation concentrations revealed clustering of plant samples based on domestication event and domestication status (Andean wild, Andean domesticated, Mesoamerican wild, and Mesoamerican domesticated) for all three experiments, namely PVG, PVC, and PLC (Figure 1, a-c). In the case of *P. vulgaris* genotypes grown in controlled conditions (PVG), a subpopulation effect could also be seen for wild genotypes (Figure 1a). Both domesticated *P. vulgaris* and *P. lunatus* accessions showed higher seed mass and shorter flowering time compared to their wild counterparts (*P. vulgaris*; average weight of 100 seeds 7.26±3.36 and 35.04±16.03 g for wild and domesticated seeds respectively. Average flowering time 48.55±3.36 and 33.80±7.09 days for wild and domesticated seeds respectively. *P. lunatus*; average weight of 100 seeds 11.48±4.23 and 55.53±22.78 g for wild and domesticated seeds respectively. Average flowering time 62.88±21.61 and 96.37±55.11 days for wild and domesticated seeds respectively. Data provided by CIAT) (Supplementary Table 1, available from GitHub, see code availability section).

In the PVG dataset, PCA shows that the two most important cations determining the clustering of wild genotypes were Fe (positive PC1) and P (negative PC2) (Figure 1a. Supplementary Table S1. Note that Fe has a similar score to Ca). Similar patterns were observed for Ca and P when expanding the analysis to multiple subpopulations and plant accessions of *P. vulgaris* propagated at CIAT (experiment PVC), indicating an overall decrease in the concentration of Ca and P in domesticated *P. vulgaris* seeds (Figure 1b, Supplementary Table S1). Wild accessions of *P. lunatus* were also characterized by higher Ca concentration (Figure 1c. Supplementary Table S1), however, Mg was the most important cation explaining the clustering of domesticated accessions for positive values of PC1 (Figure 1c. Supplementary Table S1).

We detected a consistent significant negative correlation between seed weight and Ca concentration (Figure 2, d-f). However, in the case of *P. lunatus*, domesticated seeds also had a higher concentration of Mg, which was significantly negatively correlated with flowering time (Figure 2f).

Overall, a consistent statistically significant decrease in Ca concentration was observed in both independent domestication events for both *Phaseolus* species in all three experiments (Figure 2, g-i. Supplementary Table S3). Mg concentration was also consistently affected by domestication status in *P. lunatus* seeds, with wild seeds showcasing lower Mg concentration (Figure 1i. Supplementary Table S3), opposite to *P. vulgaris* seeds.

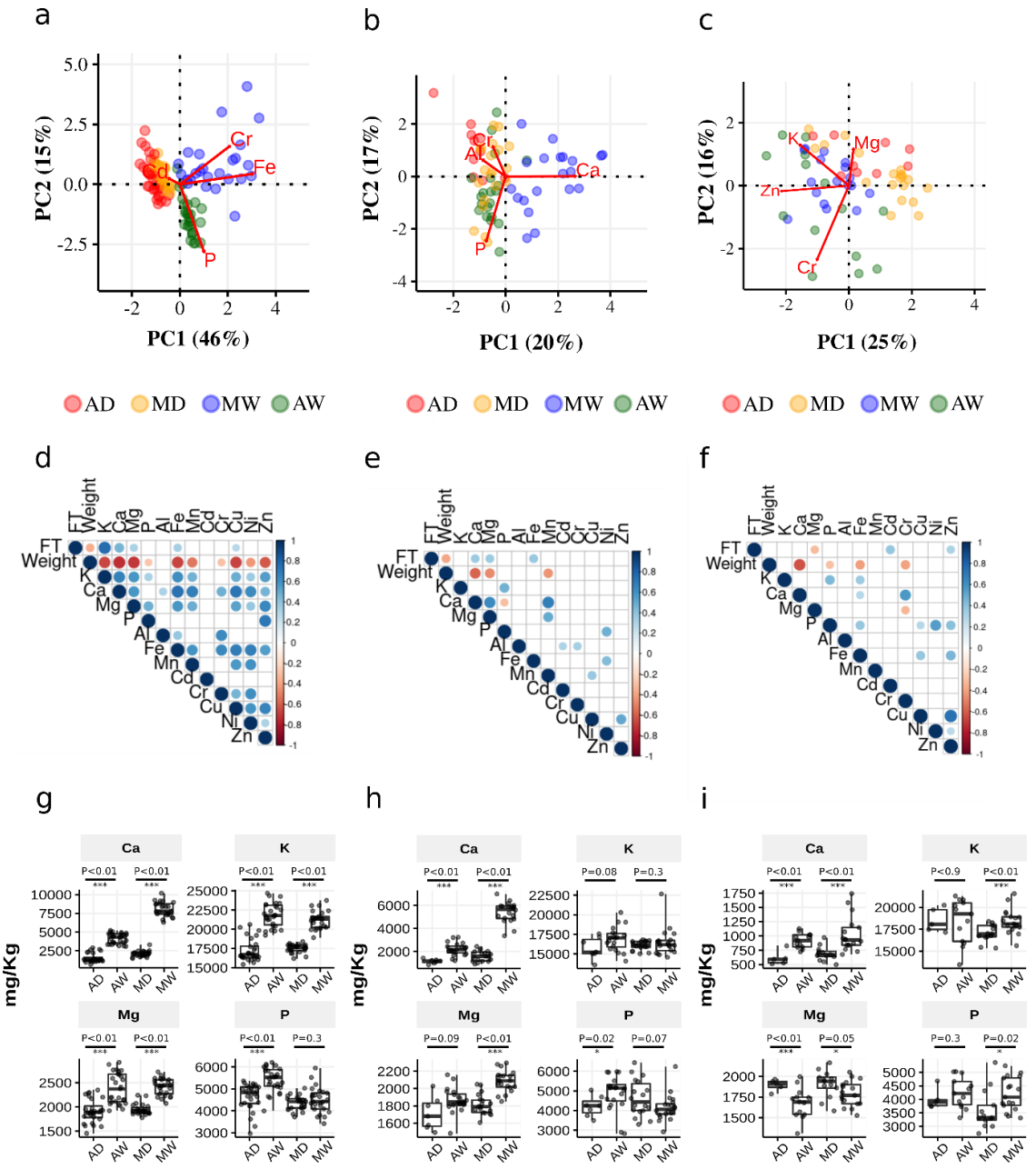


Figure 1. Seed mineral compositions distinguish wild and domesticated *Phaseolus* sp. accessions. (a), (d), and (g) refer to experiment PVG; (b), (e), and (h) to experiment PVC; (c), (f), and (i) to experiment PLC. (a-c); PCA on z-standardised cation concentrations. Andean domesticated (AD), Mesoamerican Domesticated (MD), Mesoamerican Wild (MW) and Andean Wild (AW). Red arrows indicate the most important cations for both PCs. (d-f); Matrix of significant correlations between plant phenotypes. Seed weight (weight), flowering time (FT). (g-i); Concentration of Ca, K, Mg, and P per domestication event and domestication status. Welch's t-test was used to assess

the statistical significance of differences between means. Andean Domesticated (AD), Andean Wild (AW), Mesoamerican domesticated (MD) and Mesoamerican Wild (MW).

Environment, plant species and domestication status significantly shape seed bacterial communities

The composition of the seed microbiome was determined in each experiment by amplicon sequencing of the 16S rRNA gene. A total of 1,344, 1,173, and 875 bacterial sequence variants (SVs) were identified in the seeds of *P. vulgaris* and *P. lunatus* accessions (Supplementary Table S4) of experiments PVG, PVC and PLC respectively. Bipartite network analysis of all three libraries (SVs clustered into OTUs at 99% similarity), constructed based on microbial connectivity (sharing of microbial members), showed a clear separation of plant samples based on environment of cultivation, species, and domestication status (Figure 2a).

In experiment PVG, none of the members of the core microbiome (80% prevalence) of domesticated accessions were shared with members of the core microbiome of domesticated plants grown at CIAT (experiment PVC) (Figure 2a. Supplementary Table S5). However, wild- plants shared 7 OTUs belonging to their core microbiomes (Figure 2a. Supplementary Table S5) potentially indicating stronger inheritance or heritability for these members of the microbial community. Additionally, the core microbiome of wild accessions counted more members than the corresponding domesticated plants in all three experiments (Supplementary Table S5), as well as containing an OTU belonging to the family *Rhizobiaceae* in experiment PLC (Supplementary Table S5).

Principal Coordinate analysis (PCoA) based on Bray Curtis dissimilarity matrix calculated after Cumulative Sum Scaling normalization (Paulson *et al.*, 2013) also

showed clear clustering due to environment, plant species, and domestication status (for experiments PVG and PLC) (Figure 2b). The environment factor significantly affected beta-diversity (PERMANOVA: df = 1,214, F = 67.8, $p = 0.001$), explaining 24% of the observed variation, as well as domestication status (PERMANOVA: df=2, 212, F=15, p=0.001, R^2=9.4%) and plant species (PERMANOVA: df=1, 50.9, F=19, p=0.001, R^2=6.2%).

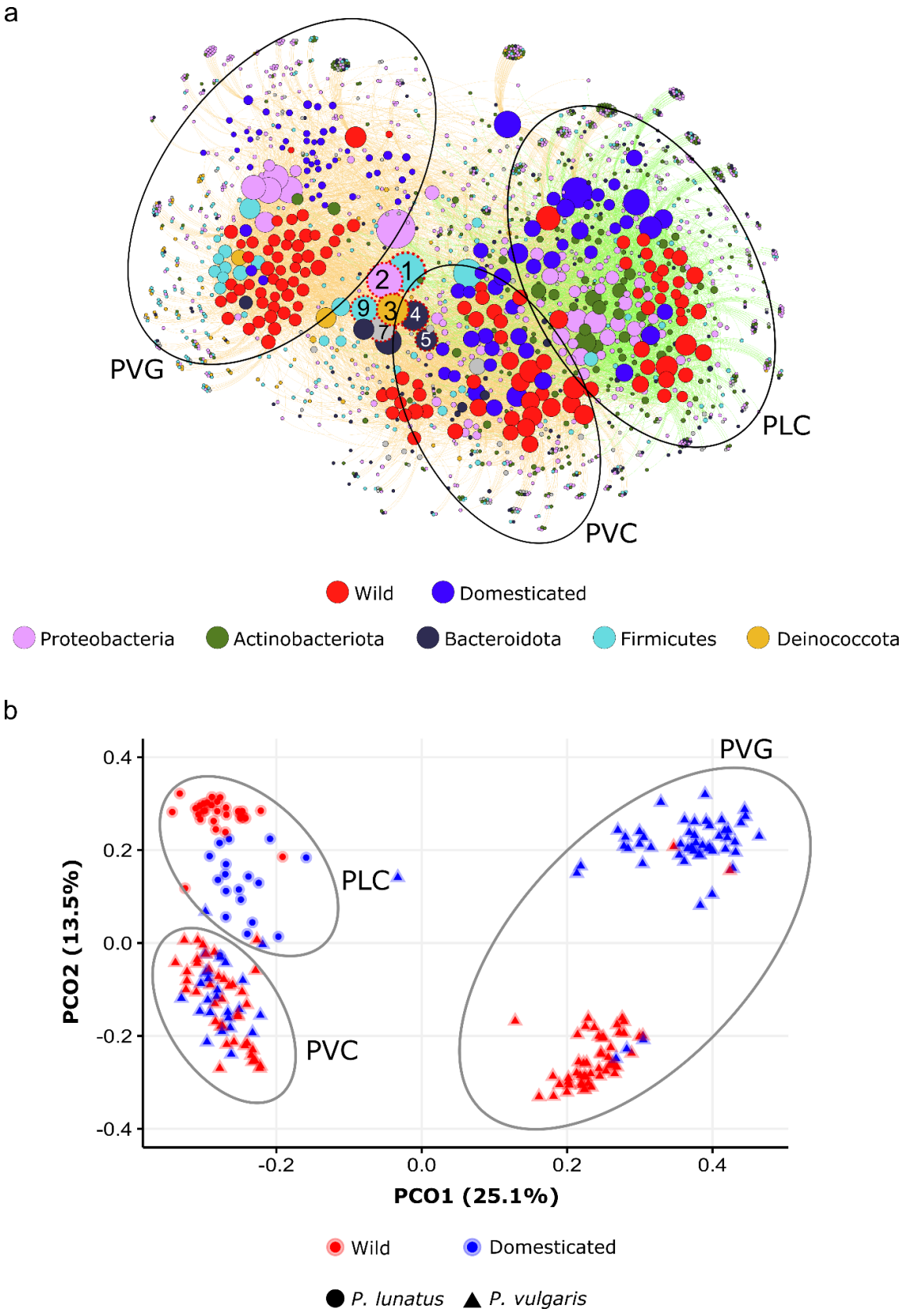

a
PVG
PLC
PVC
1
2
3
4
5
7
9
Wild
Domesticated
Proteobacteria
Actinobacteriota
Bacteroidota
Firmicutes
Deinococcota
b
PVG
PLC
PVC
PCO2 (13.5%)
PCO1 (25.1%)
0.4
0.2
0.0
-0.2
-0.4
-0.2
0.0
0.2
0.4
Wild
Domesticated
P. lunatus
P. vulgaris

Figure 2. Bipartite network and PCoA analysis of seed-associated microbial communities. (a); bipartite network representing sample/OTU interactions. In the network, node size is proportional to the number of degrees (number of connections). Edge color (line connecting different nodes) indicates different plant species (green: *P. lunatus*; orange: *P. vulgaris*). Numbered nodes show shared members of the core microbiomes of wild plants grown in different environments (Supplementary Table S5). Ellipses highlight the experiment. PVG (*Phaseolus vulgaris* greenhouse); PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT). (b); PCoA analysis based on Bray Curtis dissimilarity matrix after Cumulative Sum Scaling normalization (CSS). Ellipses highlight the experiment. PVG (*Phaseolus vulgaris* greenhouse); PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Independent domestication events led to similar footprints on seed microbial communities

According to the taxonomic affiliation of the SVs, the seed microbiome of *Phaseolus* spp. was dominated by *Proteobacteria* (average relative abundance: 20-98%), *Firmicutes* (average relative abundance: 0.7-35%), *Actinobacteriota* (average relative abundance: 0-37%), and *Bacteroidota* (average relative abundance: 0-23%), depending on the experiment, plant species, and domestication status (Figure 3a,c,e; Supplementary Table S6).

At the family level, *Pseudomonadaceae* dominated the seed microbiome of *P. vulgaris* accessions grown in controlled conditions (average relative abundance: 13-85%) followed by *Bacillaceae* (average relative abundance: 0.2-27%) (Supplementary Figure 1; Supplementary Table S6). The seed microbiome of *Phaseolus vulgaris* propagated at CIAT (experiment PVC) was mainly dominated by *Bacillaceae* (average relative abundance: 15-26%) followed by *Bacteroidia* (average relative abundance: 8-15%)

(Supplementary Figure 1c; Supplementary Table S6) while *P. lunatus* seeds were mainly colonized by *Microbacteriaceae* (mean relative abundance:4-14%) and *Sphingomonadaceae* (mean relative abundance:9-12%) (Supplementary Figure 1e; Supplementary Table S6). Interestingly, *Rhizobiaceae* was the fifth most abundant family in the seed microbiome of *P. lunatus* seeds (mean relative abundance:1-6%) (Supplementary Figure 1e; Supplementary Table S6).

We found more similarities between the seed microbiome profile of *P. vulgaris* accessions (experiment PVG and PVC), than between the seed microbiome of the *P. vulgaris* and *P. lunatus* accessions selected in experiments PVG and PLC (Figure 1a,c,e; Supplementary Figure 1a,c,e; Supplementary Table S6) indicating a plant species effect in shaping seed microbiome assembly. For example, members of the families *Thermaceae* and *Hydrogenophilaceae* were found to be abundant (0.2-16%) in both PVG and PVC experiments but were not detected in PLC (Supplementary Figure 1a,c,e; Supplementary Table S6).

We then performed differential abundance analysis for each experiment. For experiment PVG, we detected significantly increased abundance (FDR < 0.01, "BH" method) of several members of the *Proteobacteria* phylum when comparing domesticated *vs* wild accessions within each independent domestication event (Figure 3b; Supplementary Figure 1b). In particular, *Pseudomonadaceae* were significantly more abundant in domesticated accessions, regardless of the domestication event (Supplementary Figure 1b). Most of the differentially abundant SVs detected were the same for both independent domestication events (Figure 3b; Supplementary Figure 1b). For example, seeds from both Mesoamerican and Andean wild plants had a higher abundance of *Bacteroidetes*, *Firmicutes*, and *Deinoccoccota* (Figure 3b; Supplementary Figure 1b).

When we extended the analysis to encompass all subpopulations of *P. vulgaris* accessions grown in a non-controlled environment at CIAT, we could still detect a consistent differential change in abundance for both domestication events for some members of the seed microbiome (Figure 1d). For example, *Pasteurellaceae* and *Hydrogenophilaceae* were found to be significantly more abundant (FDR < 0.01, "BH" method) in seeds of wild-type plants (Supplementary Figure 1d). The higher abundance (FDR < 0.01, "BH" method) of *Hydrogenophilaceae* in wild seed microbiomes was also observed for both domestication events in accessions grown under controlled conditions (Supplementary Figure 1b) and for *P. lunatus* (Supplementary Figure 1f). For *P. lunatus*, in addition to the higher abundance of *Hydrogenophilaceae*, we also detected a consistent higher abundance of SVs belonging to *Dermabacteraceae, Sphingomonadaceae, Rhizobiaceae*, and *Pseudomonadaceae* in wild accessions regardless of domestication scenario (Figure 3f; Supplementary Figure 1f).

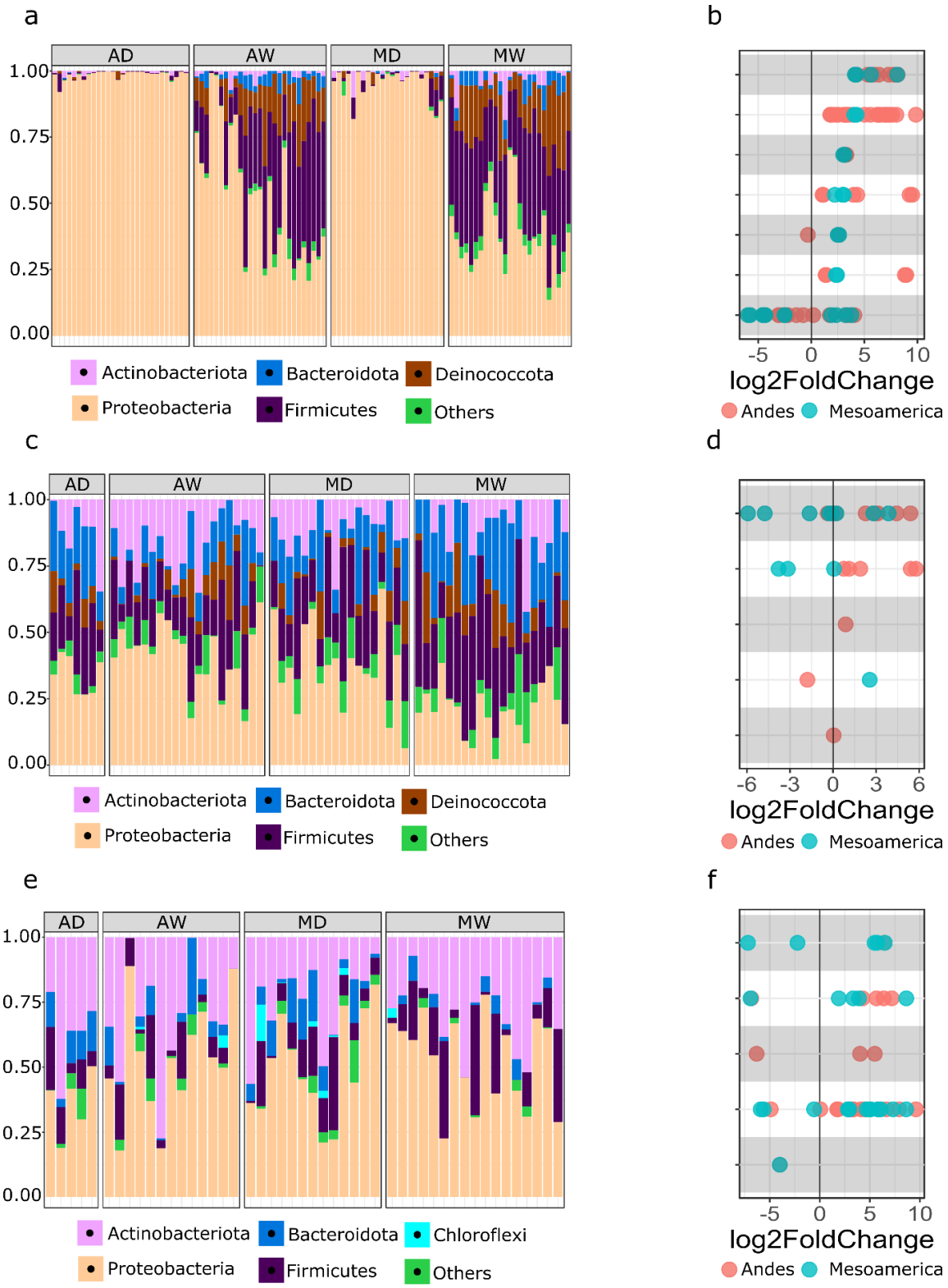


Figure 3. Independent domestication events lead to similar effects on members of the seed microbiome. (a) and (b) refer to experiment PVG; (c), and (d) to experiment PVC; (e) and (f) to experiment PLC. (a,c,e); Relative abundance of the top 5 Phyla. Andean Domesticated (AD),

Andean Wild (AW), Mesoamerican Domesticated (MD) and Mesoamerican Wild (MW). (b,d,f); Differentially abundant members of the seed microbiome in wild and domesticated seeds grouped by phylum and by domestication event calculated with DESeq2 (FDR<0.01, "BH" method). The same data, but grouped by family, is provided in Supplementary Figure 1.

Domesticated plant phenotypes partially explain seed microbiome assembly, independently of the domestication event

To understand whether similar changes in seed microbiome assembly occurred in domesticated plants from both domestication events and whether these changes correlated with changes in seed phenotype, we applied model-based analysis to the three experiments separately. For experiment PVG, in which accessions were grown in the greenhouse and seeds removed from pods under axenic conditions, wild accessions clustered together according to PCoA, regardless of their origin (Figure 4a). For experiments PVC and PLC (Figure 4b and c respectively), we also observed a similar pattern.

AIC information criteria showed that the best minimum adequate model for the plant phenotypes, which explains seed microbiome differences, accounted calcium as a covariate for both PVG and PVC experiments (lowest AIC), and magnesium for the PLC experiment. We did not directly use flowering time as an explanatory variable in the models for PVC and PLC experiments, but it should be noted that in the analyses summarised in Figure 1, flowering time correlated with both calcium and magnesium. The AIC information criteria showed that calcium and magnesium were responsible for similar changes in seed microbiome assembly in each independent domestication event (Figure 4d-f). Interestingly, although the plant accessions used in experiments PVG and

PVC were grown in completely different environments and partially different genotypes were used, we still found Ca as the best plant phenotypic covariate (Figure 4d,e) (PVG: LRT=3191, *p*=0.001***; PVC: LRT=3090, *p*=0.0009***). Moreover, for both libraries PVG and PVC, increased seed calcium concentration led to an overall positive effect on *Acidobacteriota*, *Bacteroidota*, *Deinococcota*, and *Firmicutes* and a decrease in members of the *Proteobacteria* and *Actinobacteriota* (Figure 4g,h). The seed microbiome of *P. lunatus* wild and domesticated accessions was partially explained by Mg concentration (LRT=2134, *p*=0.004***) (Figure 4f) with an overall positive effect on members of the *Deinoccocota* (Figure 4i).

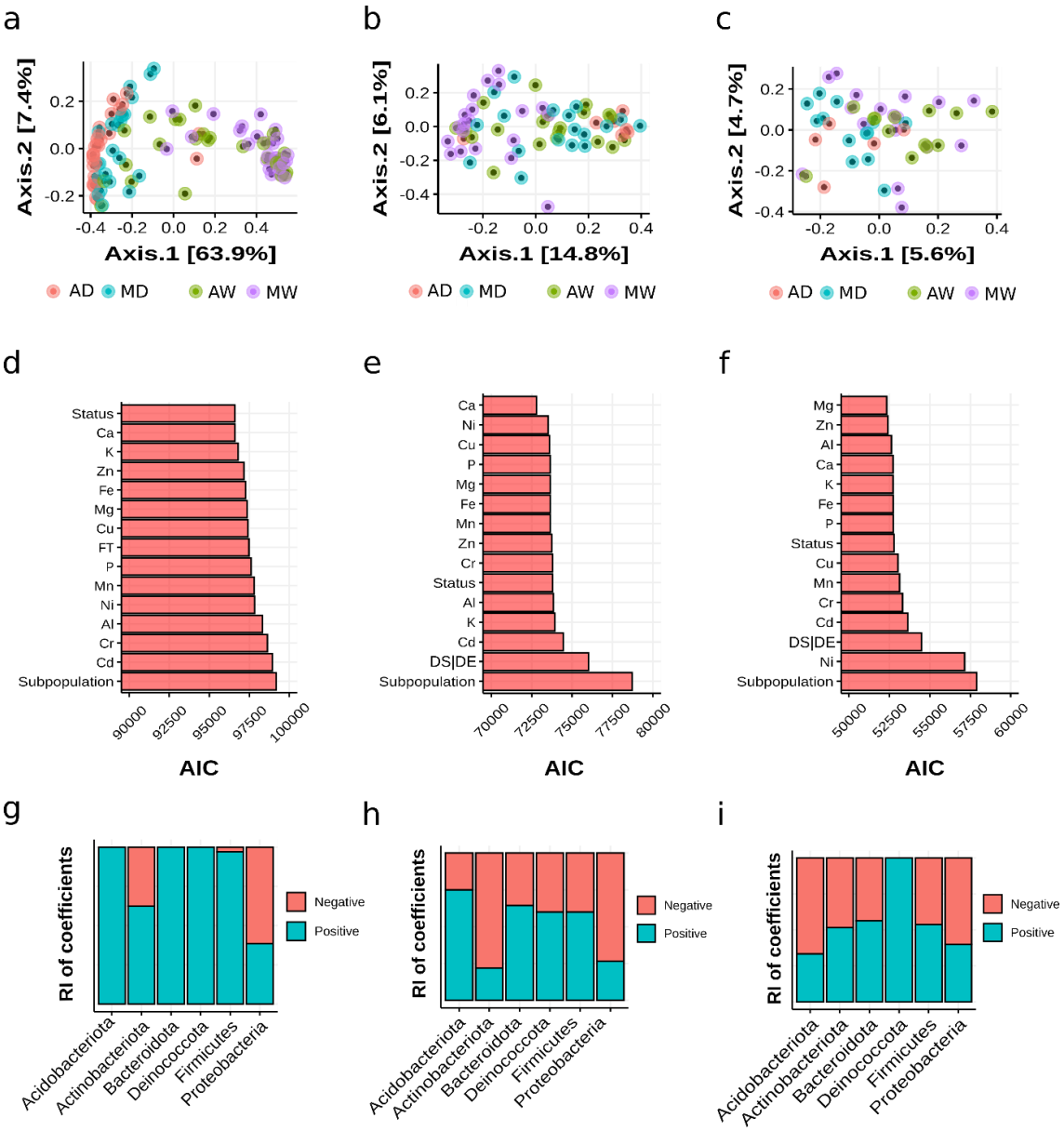


Figure 4. Domesticated plant phenotypes significantly predict seed microbial community composition, independently of domestication events. (a), (d) and (g) refer to experiment PVG; (b), (e), and (h) to experiment PVC; (c), (f), and (i) to experiment PLC. (a-c); PCoA analysis based on Bray Curtis dissimilarity matrix. Andean Domesticated (AD), Andean Wild (AW), Mesoamerican domesticated (MD) and Mesoamerican Wild (MW). (d-f); AIC values of several MGLMs (Multivariate Generalized Linear Models). The explanatory variable used in each model is represented in the y-axis. For all quantitative variables (*e.g* mineral concentration and flowering time (FT)), two parameters are being estimated, namely, intercept and slope. Covariate

domestication status (status) has two levels (Wild and Domesticated) while explanatory variable DS|DE (Domestication Status within Domestication Event) has four levels (AD, AW, MD, MW). Factor subpopulation has 4, 6, and 7 levels in experiment PVG, PVC, and PLC respectively (Supplementary Table 1). Note that for experiment PVC and PLC we did not include FT, as we did not measure this directly, but included as covariates both collection date and regeneration site of the accessions (Supplementary Table 1). However, they were not significant ($P > 0.05$; Supplementary Table S7), thus the minimum adequate model for all three experiments included one covariate only. (g-i); percentage of microbial taxa that were negatively or positively affected by calcium concentration (g,h) and magnesium (i) after filtering SVs based on 5% prevalence.

Seed calcium concentration is correlated with the importance of selection processes in the assembly of domesticated seed microbial communities

We applied iCAMP, a framework to quantitatively infer community assembly mechanisms by phylogenetic bin-based null model analysis (Ning *et al.*, 2020), to disentangle ecological processes governing seed microbiome assembly. Based on iCAMP analysis, for the PVG experiment, 72 % of the seed microbiome assembly in domesticated plants was governed by ecological drift (DR) (Figure 5a; Supplementary Table S8) while homogeneous selection (HoS) dominated processes in the assembly of wild seed microbiomes (average relative importance of 53%) (Figure 5a; Supplementary Tabe 6). The increased importance of ecological drift in the assembly of domesticated seed microbial communities was significant (Cohens'd=7.06, p<0.001) as well as the reduction in HoS (Cohens'd=-6.8, p<0.001), suggesting a relaxation of selective forces during the domestication process (Supplementary Table S8). Since iCAMP can provide

information on the relative importance of different ecological processes in individual lineages (bins), our results revealed that HoS dominated 29 and 39% of the bins in domesticated and wild seeds respectively (Figure 5d). Five bins accounted for 80% of relative abundance, namely 1 PVG (phylum: *Firmicutes*; family: *Bacillaceae*), 14 PVG (phylum: *Deinococcota*; family: *Thermaceae*), 25 PVG (phylum: *Proteobacteria*; family: *Burkholderiaceae*), 27 PVG (phylum: *Proteobacteria*; family: *Hydrogenophilaceae*) and 29 PVG (phylum: *Proteobacteria*; family: *Pseudomonadaceae*) (Figure 5d). Bin 29 alone accounted for 53% of relative abundance, but it was the only one among the top five most abundant bins with low importance of HoS (HoS< 12%, DR>85%) (Figure 6d).

For plants regenerated at CIAT, seed microbiome assembly was mainly driven by stochastic processes (DL+DR+HD) (Figure 5b,c,d; Supplementary Table S8). However, we found an overall increase in the relative importance of stochasticity in the assembly of domesticated seed microbiomes, specifically for members of the *Firmicutes* and *Bacteroidota* phyla in all three experiments and an overall decrease of stochasticity for members of the *Proteobacteria* phylum in experiments PVG and PLC (Supplementary Figure 3).

In the double-leash framework (Soldan *et al.*, 2021), we proposed that where domesticated plant phenotypes exert strong host-to-microbe effects, they should explain differences in selection processes governing the assembly of microbiomes between wild and domesticated plants. Since HoS was the dominant process in seed microbiome assembly of wild seeds (PVG experiment), we tested this hypothesis by performing a Mantel test between pairwise differences in seed calcium concentration and the pairwise differences in homogeneous selection (Soldan *et al.*, 2021). We found a statistically significant correlation between the two distance matrices (Mantel $R^2 = 0.19$, $p = 0.001$;

Supplementary Table S9), suggesting that a decrease in calcium concentration (or other variables co-correlated with calcium), typically observed in domesticated seeds (Figure 1g), led to a decrease in the relative importance of HoS.

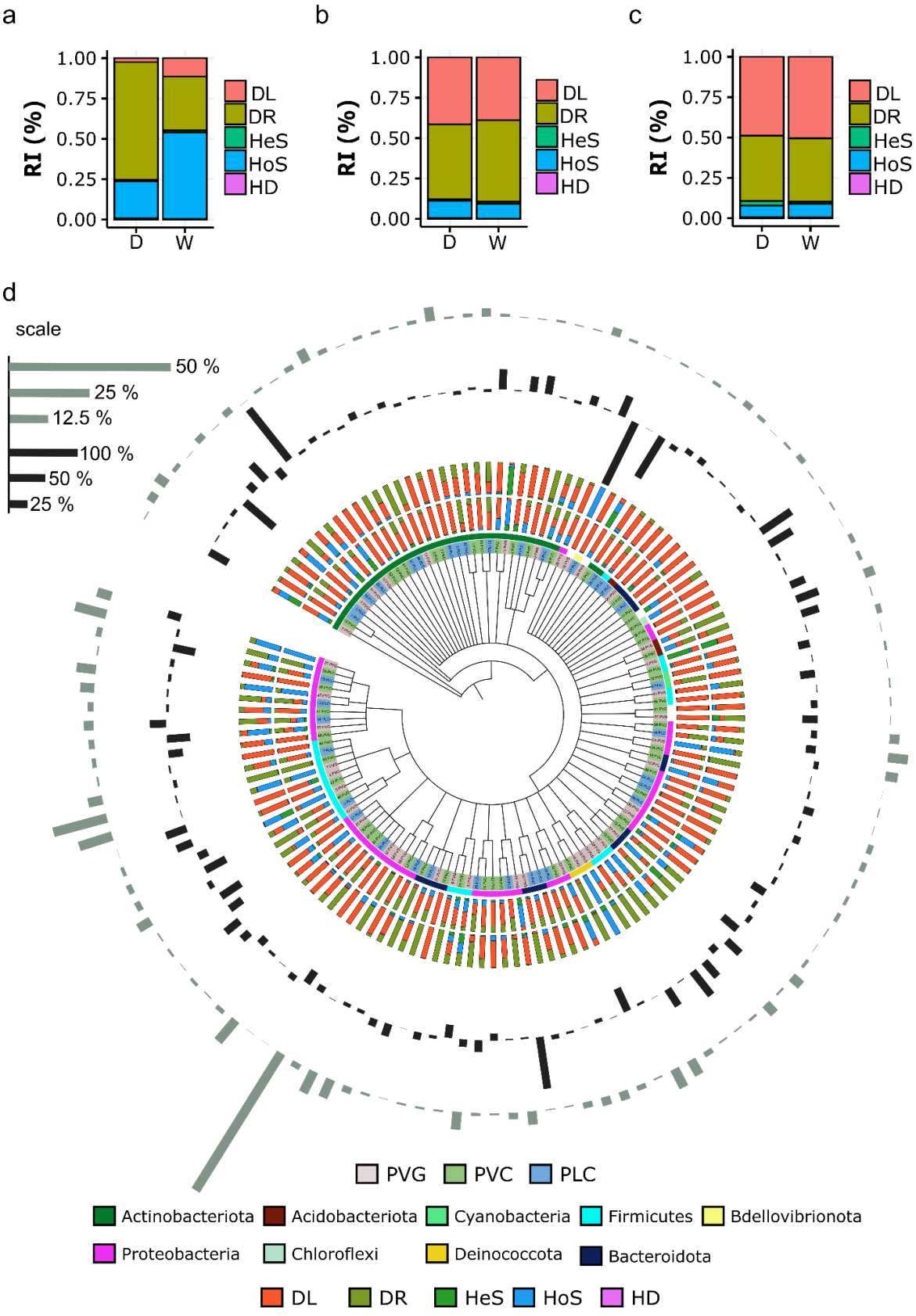

a
b
c
RI (%)
1.00
0.75
0.50
0.25
0.00
D
W
DL
DR
HeS
HoS
HD
d
scale
50 %
25 %
12.5 %
100 %
50 %
25 %
PVG
PVC
PLC
Actinobacteriota
Acidobacteriota
Cyanobacteria
Firmicutes
Bdellovibrionota
Proteobacteria
Chloroflexi
Deinococcota
Bacteroidota
DL
DR
HeS
HoS
HD

Figure 5. Ecological processes driving seed microbiome assembly in wild and domesticated plants. (a), (b) and (c) refer to experiment PVG, PVC and PLC respectively. (a-c) Overall relative importance of dispersal limitation (DL), ecological drift (DR), hetereogeneous selection (HeS), homogeneous selection (HoS) and homogenizing dispersal (HD) in seed microbiome assembly of wild (W) and domesticated (D) accessions. (d) ecological processes at phylogenetic bin level. Maximum likelihood phylogenetic tree showing branches with bootstrap values higher than 75%. Color of the inner circle (1st annulus) refers to experiment. Taxonomic affiliation of each bin (2nd annulus). Relative importance of DL, DR, HeS, HoS and HD in seed microbiome assembly of wild (W) (3rd annulus) and domesticated (D) (4th annulus) accessions. Differences (domesticated – wild; outward bars = positive, inward bars = negative) in the relative importance of stochastic processes (DL+HD+DR) in the assembly of the seed microbiome between domesticated and wild plants (5th annulus). Relative abundance of each bin (6th annulus).

Seed-associated bacterial communities can be used to predict the domestication status of plant accessions

We applied a random forest classifier on members of the seed microbiome to investigate whether community composition could be used to classify plant accessions based on the domestication status and domestication status within each domestication event, namely AD (Andean Domesticated), AW (Andean Wild), MD (Mesoamerica Domesticated), MW (Mesoamerica Wild). For all experiments, the accuracy of the RF model was significantly higher when classifying domestication status (acc > 90 %) compared to classifying domestication status within domestication events (acc < 80%) (Figure 6a,b,c). The highest accuracy (acc =97.5%) was reached for experiment PLC (Figure 6c).

Confusion matrices, showing where misclassification occurred, revealed that most of the misclassifications happened within domesticated and wild accessions of different

domestication events. For example, for experiment PVG, 96.6% of the classification errors for MW accessions were with AW plant genotypes (Figure 6a). Similarly, 89% of the misclassification for MD accessions was with AD plant samples (Figure 1a). For experiments PVC and PLC, since we had fewer observations for AD genotypes (Figure 6b,c), we report the classification errors for the wild accessions only. Again, most of the errors (66% and 67%) happened between AW and MW plant genotypes for experiments PVC and PLC respectively (Figure 6b,c).

The microbial members that allowed high accuracy in the classification of wild vs domesticated accessions mainly belonged to three phyla, namely *Proteobacteria*, *Firmicutes*, and *Bacteroidota* (Figure 6d), and were pre-selected by the Boruta algorithm (Kursa and Rudnicki, 2010). In particular, for experiment PVG, two SVs among the pre-selected features by Boruta alone allowed a classification accuracy of 96.6% (Figure 6a,d). The two SVs, namely SV_13 and SV_37 belonged to the *Proteobacteria* and *Bacteroidota* phyla respectively and were exclusively found in the seed microbiome of wild accessions at low relative abundance (Supplementary Figure 2a), like most of the indicator taxa in all three experiments (Supplementary Figure 2a,b,c). Moreover, some of the indicator taxa are recurring features in different experiments (Figure 6d). For example, the genus *Pseudomonas* and *Anoxibacillus* were important features in all three experiments, potentially indicating recurrent differential recruitment of certain microbial members in the seed microbiome of wild vs domesticated plants.

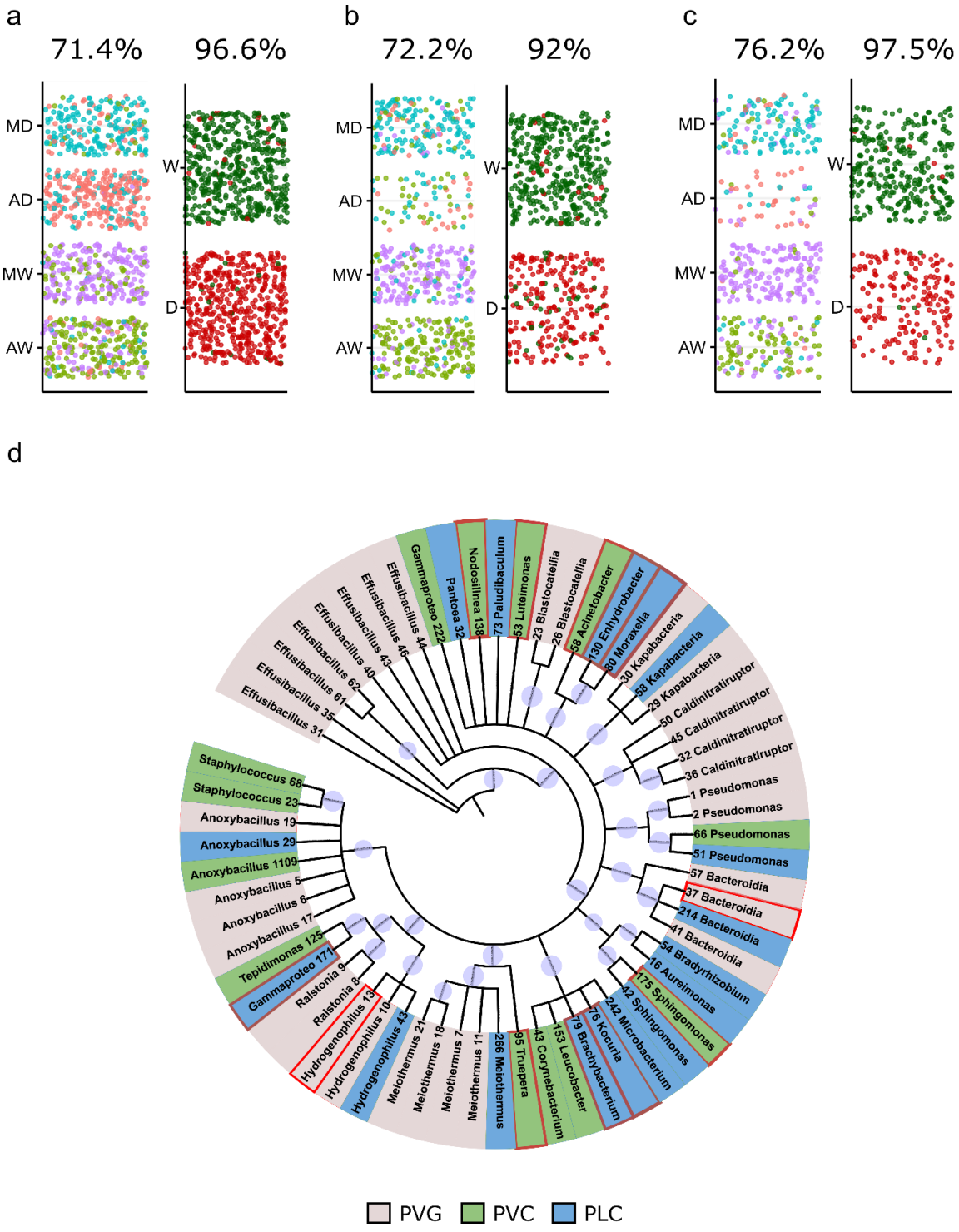


Figure 6. Seed bacterial communities can be used to distinguish wild vs. domesticated plant accessions. (a-c); accuracy and confusion matrix of the random forest model for experiment PVG (a), PVC (b), and PLC (c) for classification task domestication status (2 levels; W=wild, D=domesticated) and domestication status within domestication event (4 levels; AD=Andean

Domesticated, AW=Andean Wild, MD=Mesoamerican domesticated and MW=Mesoamerican Wild). (d); Maximum likelihood phylogenetic tree showing the indicator taxa selected by the Boruta algorithm for all 3 experiments. Red lines indicate the sub-selection of the indicator taxa used for the random forest-based classifier. Branches with bootstrap values lower than 75% are not shown. Internal node sizes are proportional to bootstrapping values (> 75%).

Discussion

The "double-leash" framework proposes that domesticated plant traits influence the assembly of the host-microbiome (Soldan *et al.*, 2021). Here we show that there is an effect of domesticated plant phenotypes on the seed microbiome. This effect was observed independently of individual domestication events within two *Phaseolus* spp., indicating that the domestication process has led to similar microbiome signatures. However, the domestication effect was consistent only within the plant species. For example, in *P. lunatus*, both Andean and Mesoamerican domesticated seeds harbored similar microbial communities, characterized by an overall higher abundance of *Deinoccoccota*, which were positively correlated with magnesium concentration (higher Mg concentration in domesticated seeds). In contrast, in *P. vulgaris*, *Deinoccocota* were more abundant in wild seeds and positively correlated with calcium (higher Ca concentration in wild seeds). This suggests complex host-to-microbe effects dictated by an interaction between plant host and domesticated phenotypes.

The "double-leash" framework also predicts that when a plant phenotype is associated with host control of the microbiota, changes in this phenotype induced by domestication would decrease the importance of selection processes. We found that the reduction in calcium concentration observed in domesticated seeds of *P. vulgaris* grown in the

greenhouse is significantly correlated with the reduction of homogeneous selection. However, as calcium concentration is only one of many traits that are altered in domesticated seeds, only some of which were included in our analyses, we cannot conclude that calcium is directly responsible for the differences in seed microbial communities between wild and domesticated plants. Nonetheless, calcium has been reported to be important for bacterial spore formation and germination [30], and we found the abundance of spore-forming *Firmicutes*, such as genus *Anoxybacillus,* to be positively affected by calcium concentration. Moreover, it has been recently reported that seed dormancy in *P. vulgaris* is likely caused by a decrease in the activity of pectin acetylesterase-8 (Soltani *et al.*, 2021). This leads to lower calcium-mediated crosslinking among pectin polymers, thus to a looser cell wall that is more prone to micro-fissures (Soltani *et al.*, 2021). The significant reduction in Ca concentration that we found in domesticated seeds supports this finding. Moreover, the varying concentration of cations could affect bacterial regulation of gene expression. Overall, our results support the hypothesis that the domesticator influences the leash used by the plant to control its microbiota by modifying plant phenotypes that might have arisen through co-evolution between plants and microbiomes. This effect is consistent across domestication events because similar plant phenotypes have been selected in plants domesticated multiple times independently.

The seed microbial communities detected were largely composed of *Proteobacteria, Firmicutes,* and *Bacteroidetes,* in agreement with previous studies (Simonin *et al.*, 2021). Post-harvest practices and the environment in which *P. vulgaris* plants are grown have been found to strongly determine bacterial taxa distribution in *P. vulgaris* seeds, as previously reported (Klaedtke *et al.*, 2016). However, in this previous study the authors

found that "*Pseudomonas* was by far the most represented bacterial genus with 48 and 77% of all reads", which largely agrees with our findings for experiment PVG but not for experiment PVC. This difference could be explained by the fact that seeds harvested at CIAT are dried before storing. It has recently been reported that the drying process significantly affect the presence of *Pseudomonas* (reduction from 50% to 0.9% relative abundance) (Chandel *et al.*, 2021) which largely agrees with our findings.

When we classified samples based on microbiome signature, we detected indicator species that allowed us to distinguish the domestication status of the plant accession with high accuracy, indicating the recurring presence or absence of microbial taxa in the seed microbiome of domesticated plants. High accuracy in the classification task was achieved solely when domestication events were not considered, providing additional evidence of a microbial signature caused by domestication.

Interestingly, our results obtained from experiments PVG and PVC showed that the observation that domesticated *P. vulgaris* plants have a lower abundance of *Bacteroidota* in the rhizosphere (Pérez-Jaramillo *et al.*, 2017) holds for the seed compartment, possibly suggesting a rhizospheric origin of this phylum in *P. vulgaris*.

For experiment PVG, we noticed a correlation between the known genetic diversity of *P. vulgaris* subpopulations (Bitocchi *et al.*, 2013) and microbiome alpha diversity indexes (Supplementary Table S4). In practice, AW subpopulations have lower genetic diversity than MW subpopulations because of a known genetic bottleneck that happened before domestication (migration from Mesoamerica to the Andes) (Bitocchi *et al.*, 2013; Rodriguez *et al.*, 2016). Therefore, subpopulation AD has lower genetic diversity than subpopulation MD. When looking at microbiome alpha diversity indexes (Supplementary Table S4), we could see a clear correlation between microbiome diversity and plant

genetic diversity. This could indicate a potential phylosymbiosis between plants and their microbiomes, as already indicated for cereals and their seed microbiomes (Abdullaeva *et al.*, 2020).

Conclusion

Overall, our results suggest that the seed microbiome is highly influenced by plant phenotypes. Since wild plants of a certain plant species are more phenotypically similar to each other than to their corresponding domesticated counterparts, the seed microbiome reflects these similarities at the plant trait level.

Moreover, we observed that this effect is translated in domesticated seeds into a relaxation of the selective forces responsible for host-microbiome assembly in the seeds of wild plants. This could potentially suggest that there is a reduction in positive microbe-to-host effects, with unknown consequences for plant growth and health. For example, one of the most recurrent bacterial diseases affecting *P. vulgaris* yields is halo blight, which is caused by the seed-borne pathogen *Pseudomonas syringae* pv *phaseolicola* (Arnold *et al.*, 2011). We observed a consistent and significant reduction in the abundance of *Pseudomonadaceae* in wild seeds in the PVG experiment. It remains to be elucidated whether the lower abundance of *Pseudomonadaceae* is due to competition with other microbial members found in wild seeds, or to host traits such as secondary metabolites. Nonetheless, this suggests that understanding the differences between wild and domesticated plant microbiomes is important for achieving sustainable agriculture.

Methods

Plant material

Selection of plant genotypes was based on previous studies on population genetic structure of *P. vulgaris* (Rodriguez *et al.*, 2016) and *P. lunatus* accessions (Chacón-Sánchez and Martínez-Castillo, 2017; Garcia *et al.*, 2021). Only accessions that belonged to a specific subpopulation were selected for this study, thus, admixed accessions were excluded. For the greenhouse experiment (PVG), we selected two subpopulations per domestication scenario to have a balanced design (one subpopulation containing all wild genotypes and one containing all domesticated genotypes). Within each subpopulation, we selected 4 accessions with 7 replicates each. Thus, a total of 112 plants were grown in the greenhouse until maturity, but not all replicates cast seeds (Supplementary Table 1). Subsequently, we expanded the analysis to consider 70 different plant genotypes of *P. vulgaris* (experiment PVC) and 61 of *P. lunatus* (experiment PLC) encompassing all identified subpopulations, using the same selection criteria, that is that accessions had to belong to a subpopulation (Rodriguez *et al.*, 2016; Chacón-Sánchez and Martínez-Castillo, 2017). In the case of *P. vulgaris*, all accessions selected for the greenhouse experiment (PVG) were also included in the PVC experiment. The selected accessions encompass the wide geographical distribution of the genus *Phaseolus* in the Americas. Details of the selected accessions and subpopulation genetic cluster can be found in Supplementary Table 1.

Plant growth conditions and processing

Seeds of *P. vulgaris* were washed in 70% ethanol for 1 minute and rinsed 3 times in sterile water before performing scarification to ease the germination process. After 3 days, germinating seeds were transferred to 3-liter pots containing 40% Norfolk topsoil (https://www.norfolktopsoil.co.uk/product-category/topsoil-and-compost/), 50% vermiculite and 10% sand. Pots were arranged according to complete randomization. Accessions were grown at 20–24°C, with artificial light maintained for 12 h periods within the 24-h cycle until maturity. After 3 weeks, 3 g per pot of fertiliser floranid permanent 16-7-15 (Compost Expert, Germany) was added. Drip irrigation was applied to maintain the substrate at field capacity.

At maturity, dry pods were collected and stored at room temperature. Shortly after collection pods were opened under axenic conditions. Two seeds per pod for a total of 5 pods per plant were used for total DNA extraction, performed after 3 months from seed collection. Seeds were crushed in a sterile mortar with liquid nitrogen, under axenic conditions, and 180 mg was used for total DNA extraction.

For PVC and PLC experiments, seeds originated from CIAT were washed in 70% ethanol for 1 minute. Ten seeds per sample were processed as described above.

Seed chemistry

The chemical properties of the seeds were characterized at Forest Research (UK). Approximately 30 seeds were pooled and crushed in a sterile mortar with liquid nitrogen. Each sample (100 mg) was analyzed for the following elements (calcium, magnesium, potassium, phosphate, zinc, molybdenum, cadmium, aluminum, chrome, copper, nickel, manganese) using a dual view ICP-OES (Thermo ICap 6500).

Total DNA extraction

Total DNA was extracted with the *Quick*-DNA Fecal/Soil Microbe Miniprep Kit (Zymo Research, USA) according to the manufacturer's instructions.

Sequencing and Bioinformatics

Investigation of microbial communities was based on paired-end amplicon sequencing of the 16S rRNA gene. The 16S rRNA gene was amplified using the primers 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACNVGGGTWTCTAAT-3′) (Apprill *et al.*, 2015; Parada *et al.*, 2016). Protein nucleic acid PCR clamps (5 μM) targeting plastid (pPNA, 5′-GGCTCAACCCTGGACAG-3′) and mitochondrial (mPNA, 5′-GGCAAGTGTTCTTCGGA-3′) DNA (PNA Bio, Newbury Park, CA, USA) were added to samples, as published previously (Lundberg *et al.*, 2013).

All 3 libraries were constructed with the 96 Nextera XT Index Kit (Illumina) following the manufacturer's instructions with minor modifications. Briefly, the first PCR mixture contents were as follows: Platinum Host-Start PCR Master Mix (2X), 12.5 μl; primers, 1 μl of 10 μM for each; template DNA, 5 μl of 2.5 ng/μl; pPNA and mPNA mix, 5 μl of 25 μM mix; and H_2O to 25 μl. The PCR conditions were 98°C for 2 min, 33 cycles of 98°C for 15 s, 55°C for 15 s, and 72°C for 17 s, and a final elongation step of 72°C for 2 min. The second amplification was performed according to 96 Nextera XT Index Kit instructions. Library sequencing was performed using the Illumina MiSeq platform with 2x300 pair-end sequencing at the Genomics and Bioinformatics Core Facility, Center for Biomedical Research of la Rioja (Spain). Filter tips were used throughout the library preparation steps alongside controls (water and kit reagents) to detect possible contamination.

Primers were removed from raw sequencing data using cutadapt (Martin, 2011). All further read processing namely filtering, trimming, merging, and chimeras removal was done in the dada2 package (Benjamin J. Callahan *et al.*, 2016, 2). Bacterial and archaeal taxonomy was assigned with the naive Bayesian classifier method (Wang *et al.*, 2007) implemented in dada2 to the genus and species level using the SILVA reference database v.138. Both NCs (water and kit reagents) contained the same contaminants, 3 SVs belonging to the genera *Escherichia, Paucibacter,* and *Microbacterium,* accounting for > 99% of the reads in the negative controls from the PVC and PLC experiments. All SVs found in negative controls (kit reagents) with a relative abundance higher than 0.1% in NC samples were removed from all plant samples (libraries PVC and PLC). Subsequently, reads belonging to chloroplast and mitochondria were removed (20-70% of the reads depending on the library). For library PVG we obtained a mean sequencing depth of 83,860 high-quality bacterial reads per sample while for libraries PLC and PVC, we obtained 42,310 and 67,357 bacterial reads per sample respectively. Samples with fewer than 10,000 reads were removed as well as rare SVs (1% prevalence, abundance threshold 20 reads). The SVs table along with metadata information was handled with the R package phyloseq (McMurdie and Holmes, 2013). Phylogenetic trees were made with the phangorn R package by first constructing a neighbor-joining tree and then fitting a GTR+G+I (Generalized time-reversible with Gamma rate variation) maximum likelihood tree using the neighbor-joining tree as a starting point (Ben J. Callahan *et al.*, 2016). For phylogenetic trees in Figures 5 and 6, SV sequences were aligned with MAFFT (Katoh *et al.*, 2002) and the phylogenetic trees constructed with RAxML (Stamatakis, 2014). Phylogenetic trees were visualized in iTOL (Letunic and Bork, 2019).

Statistical analysis and experimental design

We tested whether independent domestication events have led to common changes in the seed microbiome community induced by domesticated plant phenotypes. To answer this question we used two approaches, namely model-based and machine learning applied to 3 independent experiments. Experiment PVG constituted of accessions grown in the greenhouse (complete randomization) until maturity. We grew representative accessions of domesticated and wild subpopulations for both domestication events (Supplementary Table 1) to have a balanced design. Experiment PVG; 112 plants: 2 domestication events * 2 subpopulations (1 wild and 1 domesticated) * 4 accessions * 7 biological replicates. Experiment PVC; 70 accessions: 2 domestication events * 3 subpopulations (two wild and one domesticated) * ~ 12 accessions. Experiment PLC; 61 accessions: 2 domestication events * ~ 4 subpopulations (two wild and two domesticated for the Mesoamerican domestication event and two wild and one domesticated for the Andean domestication event) * ~ 8 accessions.

A model-based approach for multivariate data was developed in the R package mvabund (Wang *et al.*, 2012). This approach fits a separate generalized linear model to each SV member of the microbial community, using a common n-dimensional set of explanatory variables and resampling-based hypothesis testing. Since the number of bacterial reads per SV does not reliably reflect bacterial abundance, we introduce an offset for sequencing depth in our model. We found the best Multivariate Generalized Linear Model (MGLM model, family: negative binomial) based on the sum of the AIC over all variables (seed chemistry and plant phenotypes when appropriate). Briefly, we fitted a MGLM model for each explanatory variable (seed mineral concentrations, plant phenotypes, and plant information). Thus, for experiment PVG we fitted 15 models, 13 models accounted

for quantitative variables as covariates (12 related to seed chemistry and 1 to flowering time), and 2 for categorical variables. One categorical variable, named “status” (two levels) indicates whether a plant genotype is domesticated or wild. The categorical variable subpopulation has 4 levels (1 wild subpopulation per domestication event and 1 domesticated subpopulation per domestication event). We repeated a similar analysis for experiments PVC and PLC. In these cases, we included an additional covariate called “DS|DE” which has 4 levels (AD, AW, MD, MW). This is because, for experiment PVG, subpopulation corresponded to “DS|DE”. Covariate subpopulations has 6 and 7 levels for experiment PVC and PLC respectively.

Following the identification of the minimum adequate model based on AIC values, we included as covariate the regeneration site of the accessions and the collection year for experiments PVC and PLC to account for an environmental effect. The statistical significance of the fitted models was assessed with ANOVA (likelihood ratio tests) using bootstrap iterations via PIT-trap residual resampling, a method that shows low rates of type I errors. For models containing 3 covariates (seed phenotype, regeneration site of the accession, collection year) we used 120 bootstrap iterations (120 cores, 1 bootstrap iteration per core). Depending on the number of response variables, the computation time of this kind of model could be 4-5 days on a high-performance computing cluster. After removing non-significant explanatory variables, we repeated the bootstrapping procedures performing 1,080 iterations (120 cores, 9 bootstrapping per core).

A random forest-based classifier with 10 times 5-fold cross-validation was further used to assess whether higher overall accuracy (acc) was reached when classifying wild vs domesticated genotypes based on the seed microbiome community (indicating similar effects of independent domestication events on seed microbiome assembly) than the

classification of plant genotypes for each domestication status within each domestication event (factor with 4 levels; AW, AD, MW, MD). Microbiome data was converted into relative abundances before model fitting, to avoid biases introduced by different sequencing depths per sample.

The random forest-based classifier was built using mlr R package (Bischl *et al.*, 2016; Lang *et al.*, 2019). Good performances of random forest algorithms in microbiome studies have been already reported (Roguet *et al.*, 2018; Zhou and Gallins, 2019). Feature selection was performed with the Boruta algorithm with the R package Boruta (Kursa and Rudnicki, 2010) with default parameters. Following the identification of the most important features (the most important SVs for the classification task), we randomly subsampled them in combination (Dang and Kishino, 2021) and applied random forest with 10 times 5-fold cross-validation resulting in 25,0000 trees. After selecting the model with the highest accuracy, we tuned the hyperparameters (ntree, mtry and nodesize) of the model with the function tuneParams.

Differential abundance analyses were performed with DEseq2 (Love *et al.*, 2014) with default parameters. Shrunken log2 fold changes (LFC) were calculated with the function lfcshrink, type=”normal”.

Bipartite network

We clustered SVs of the 3 libraries into OTUs based on 99% similarities with the R package DECIPHER (Wright, 2016). The bipartite network was computed with the script make_bipartite_network.py in qiime (Bolyen *et al.*, 2019) and visualized in Gephi (Bastian *et al.*).

Ecological processes

The importance of ecological processes governing seed microbiome assembly was calculated with the iCAMP R package (Ning *et al.*, 2020).

Data availability

Phyloseq objects, containing OUT table, phylogenetic tree and metadata for each library are available from GitHub (git@github.com:RSO9192/Consistent-effects-of-independent-domestication-events-on-seed-microbiome.git).

Code availability

Custom scripts of the main analysis performed in this study are available from GitHub (git@github.com:RSO9192/Consistent-effects-of-independent-domestication-events-on-seed-microbiome.git).

References

Abdullaeva, Y., Ambika Manirajan, B., Honermeier, B., Schnell, S., and Cardinale, M. (2020) Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *Journal of Advanced Research*.

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* 75: 129–137.

Arnold, D.L., Lovell, H.C., Jackson, R.W., and Mansfield, J.W. (2011) Pseudomonas syringae pv. phaseolicola: from “has bean” to supermodel. *Mol Plant Pathol* 12: 617–627.

Bastian, M., Heymann, S., and Jacomy, M. Gephi : An Open Source Software for Exploring and Manipulating Networks. 2.

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., et al. (2014) Decreased Nucleotide and Expression Diversity and Modified Coexpression Patterns Characterize Domestication in the Common Bean. *The Plant Cell* tpc.114.124040.

Berg, G. and Raaijmakers, J.M. (2018) Saving seed microbiomes. *The ISME Journal* 12: 1167–1170.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., et al. (2016) mlr: Machine Learning in R. *Journal of Machine Learning Research* 17: 1–5.

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol* 197: 300–313.

Bitocchi, E., Rau, D., Bellucci, E., Rodriguez, M., Murgia, M.L., Gioia, T., et al. (2017) Beans (*Phaseolus* ssp.) as a Model for Understanding Crop Evolution. *Frontiers in Plant Science* 8: 722.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37: 852–857.

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

Callahan, Ben J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., and Holmes, S.P. (2016) Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses.

Celmeli, T., Sari, H., Canci, H., Sari, D., Adak, A., Eker, T., and Toker, C. (2018) The Nutritional Content of Common Bean (Phaseolus vulgaris L.) Landraces in Comparison to Modern Varieties. *Agronomy* 8: 166.

Chacón-Sánchez, M.I. and Martínez-Castillo, J. (2017) Testing Domestication Scenarios of Lima Bean (Phaseolus lunatus L.) in Mesoamerica: Insights from Genome-Wide Genetic Markers. *Front Plant Sci* 8:.

Chandel, A., Mann, R., Kaur, J., Norton, S., Edwards, J., Spangenberg, G., and Sawbridge, T. (2021) Implications of Seed Vault Storage Strategies for Conservation of Seed Bacterial Microbiomes. *Frontiers in Microbiology* 12: 3512.

Chen, Y.H., Gols, R., and Benrey, B. (2015) Crop Domestication and Its Impact on Naturally Selected Trophic Interactions. *Annu Rev Entomol* 60: 35–58.

Dang, T. and Kishino, H. (2021) Forward variable selection improves the power of random forest for high- dimensional microbiome data.

Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418: 700–707.

Diamond, J.M. (1998) Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years, Random House.

Fernandez, A.R., Sáez, A., Quintero, C., Gleiser, G., and Aizen, M.A. (2021) Intentional and unintentional selection during plant domestication: herbivore damage, plant defensive traits and nutritional quality of fruit and seed crops. *New Phytologist* 231: 1586–1598.

Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017) The evolution of the host microbiome as an ecosystem on a leash. *Nature* 548: 43–51.

Garcia, T., Duitama, J., Zullo, S.S., Gil, J., Ariani, A., Dohle, S., et al. (2021) Comprehensive genomic resources related to domestication and crop improvement traits in Lima bean. *Nat Commun* 12: 702.

Hammer, K. (1984) Das Domestikationssyndrom. *Die Kulturpflanze* 32: 11–34.

Kantar, M.B., Nashoba, A.R., Anderson, J.E., Blackman, B.K., and Rieseberg, L.H. (2017) The Genetics and Genomics of Plant Domestication. *BioScience* 67: 971–982.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

Kavamura, V.N., Robinson, R.J., Hughes, D., Clark, I., Rossmann, M., Melo, I.S. de, et al. (2020) Wheat dwarfing influences selection of the rhizosphere microbiome. *Scientific Reports* 10: 1–11.

Klaedtke, S., Jacques, M.-A., Raggi, L., Préveaux, A., Bonneau, S., Negri, V., et al. (2016) Terroir is a key driver of seed-associated microbial assemblages. *Environmental Microbiology* 18: 1792–1804.

Kursa, M.B. and Rudnicki, W.R. (2010) Feature Selection with the Boruta Package. *Journal of Statistical Software* 36: 1–13.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., et al. (2019) mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software* 4: 1903.

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

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.

Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D., and Dangl, J.L. (2013) Practical innovations for high-throughput amplicon sequencing. *Nature Methods* 10: 999–1002.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10–12.

Matsumoto, H., Fan, X., Wang, Yue, Kusstatscher, P., Duan, J., Wu, S., et al. (2021) Bacterial seed endophyte shapes disease resistance in rice. *Nature Plants* 7: 60–72.

McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8: e61217.

Meyer, R.S., DuVal, A.E., and Jensen, H.R. (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* 196: 29–48.

Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* 14: 840–852.

Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al. (2020) A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nature Communications* 11: 4717.

Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 18: 1403–1414.

Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013) Robust methods for differential abundance analysis in marker gene surveys. *Nat Methods* 10: 1200–1202.

Pérez-Jaramillo, J.E., Carrión, V.J., Bosse, M., Ferrão, L.F.V., de Hollander, M., Garcia, A.A.F., et al. (2017) Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J* 11: 2244–2257.

Puente, M.E., Li, C.Y., and Bashan, Y. (2009) Endophytic bacteria in cacti seeds can improve the development of cactus seedlings. *Environmental and Experimental Botany* 66: 402–408.

Rodriguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., et al. (2016) Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol* 209: 1781–1794.

Roguet, A., Eren, A.M., Newton, R.J., and McLellan, S.L. (2018) Fecal source identification using random forest. *Microbiome* 6: 185.

Silva, C.A., Abreu, Â. de F.B., Ramalho, M.A.P., and Maia, L.G.S. (2012) Chemical composition as related to seed color of common bean. *Crop Breed Appl Biotechnol* 12: 132–137.

Simonin, M., Briand, M., Chesneau, G., Rochefort, A., Marais, C., Sarniguet, A., and Barret, M. (2021) Seed microbiota revealed by a large-scale meta-analysis including 50 plant species.

Soldan, R., Fusi, M., Cardinale, M., Daffonchio, D., and Preston, G.M. (2021) The effect of plant domestication on host control of the microbiota. *Commun Biol* 4: 1–9.

Soldan, R., Mapelli, F., Crotti, E., Schnell, S., Daffonchio, D., Marasco, R., et al. (2019) Bacterial endophytes of mangrove propagules elicit early establishment of the natural host and promote growth of cereal crops under salt stress. *Microbiological Research* 223–225: 33–43.

Soltani, A., Walter, K.A., Wiersma, A.T., Santiago, J.P., Quiqley, M., Chitwood, D., et al. (2021) The genetics and physiology of seed dormancy, a crucial trait in common bean domestication. *BMC Plant Biology* 21: 58.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Sun, L., Liu, Y., Wu, L., and Liao, H. (2019) Comprehensive Analysis Revealed the Close Relationship between N/P/K Status and Secondary Metabolites in Tea Leaves. *ACS Omega* 4: 176–184.

Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., and Singh, B.K. (2020) Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 18: 607–621.

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Van, A.L., and Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytologist* 206: 1196–1206.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73: 5261–5267.

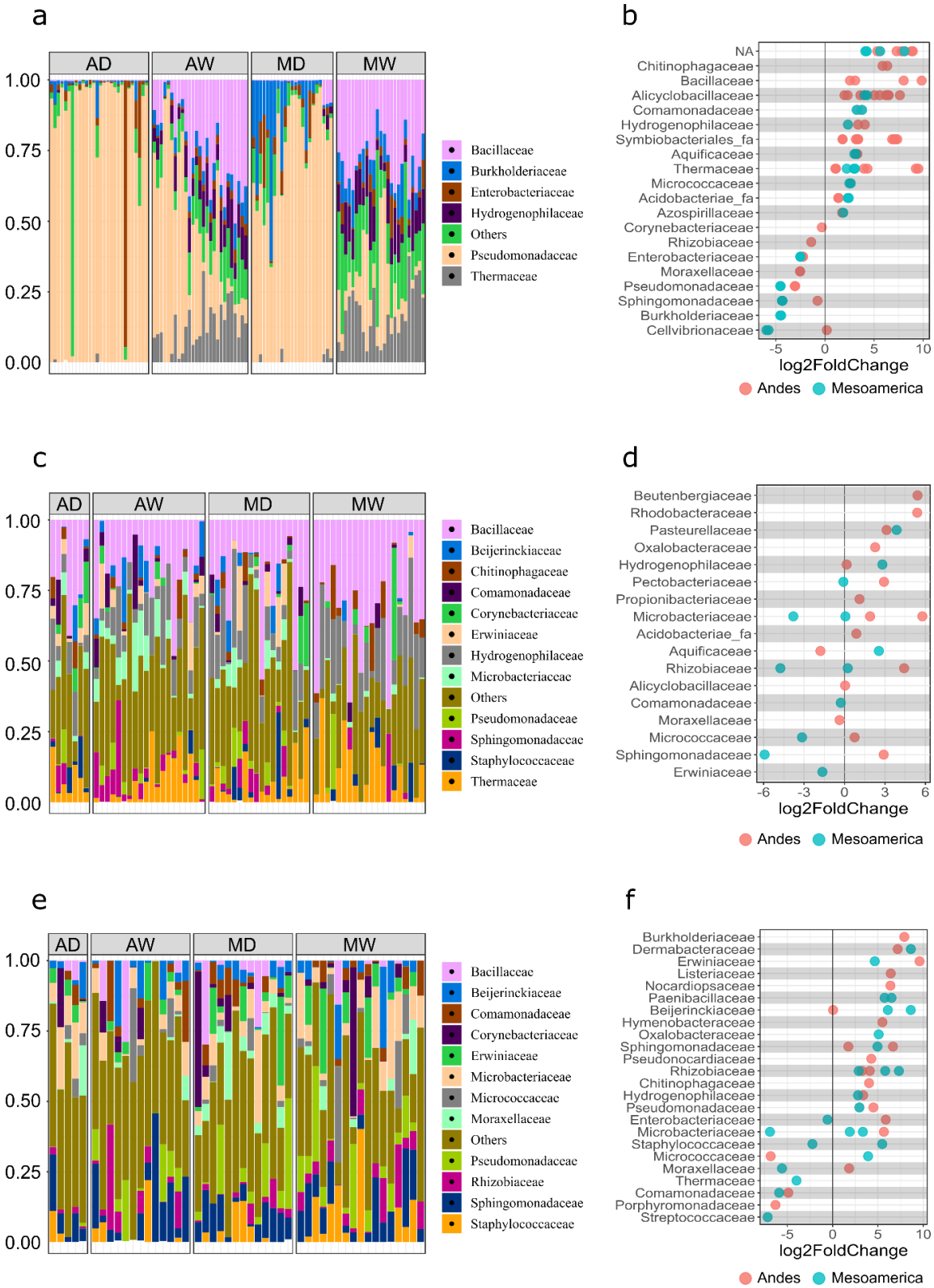
Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012) mvabund– an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3: 471–474.

Whitham, T.G., Young, W.P., Martinsen, G.D., Gehring, C.A., Schweitzer, J.A., Shuster, S.M., et al. (2003) Community and Ecosystem Genetics: A Consequence of the Extended Phenotype. *Ecology* 84: 559–573.

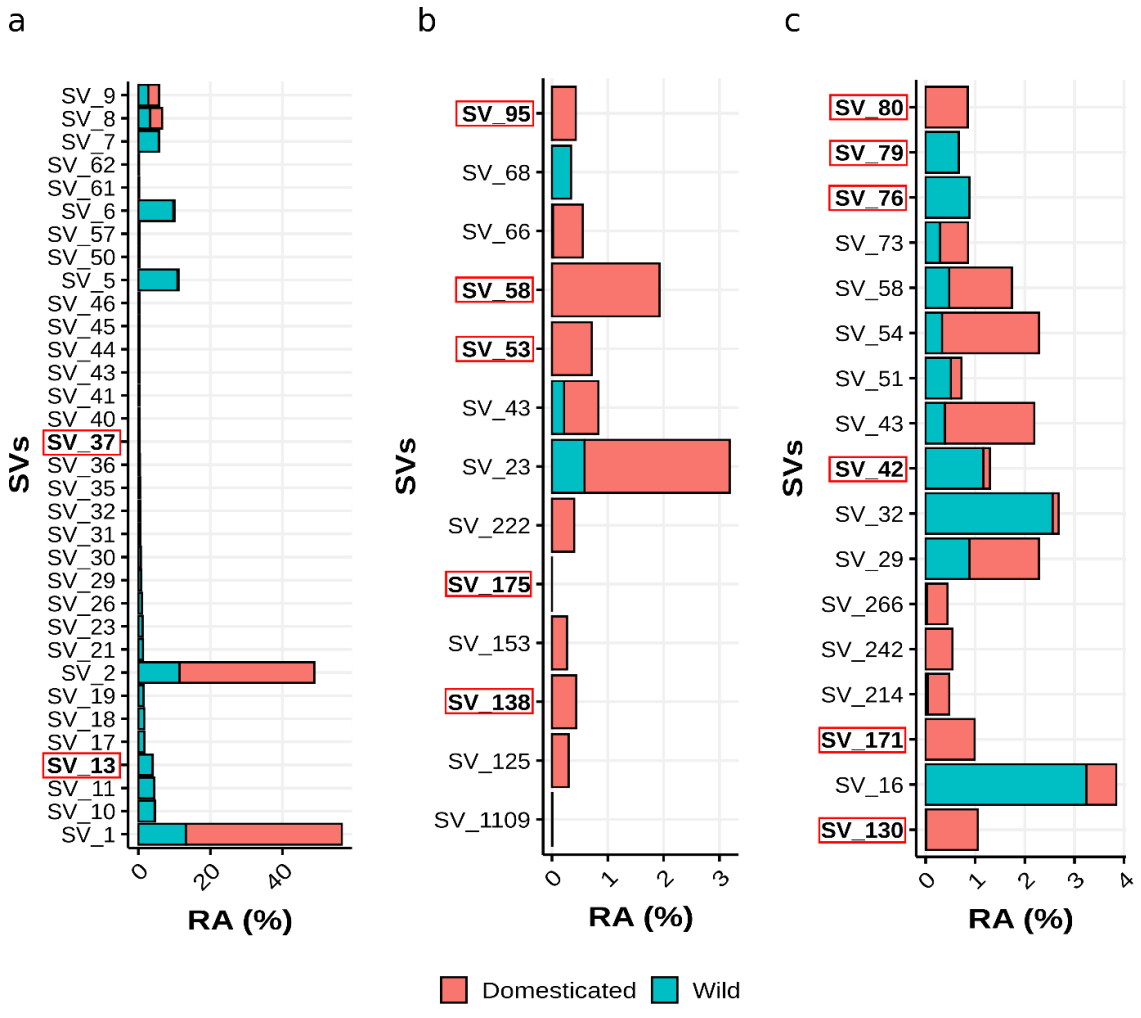
Wright, E.S. (2016) Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal* 8: 352–359.

Zhou, Y.-H. and Gallins, P. (2019) A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction. *Frontiers in Genetics* 10: 579.

Supplementary Figures

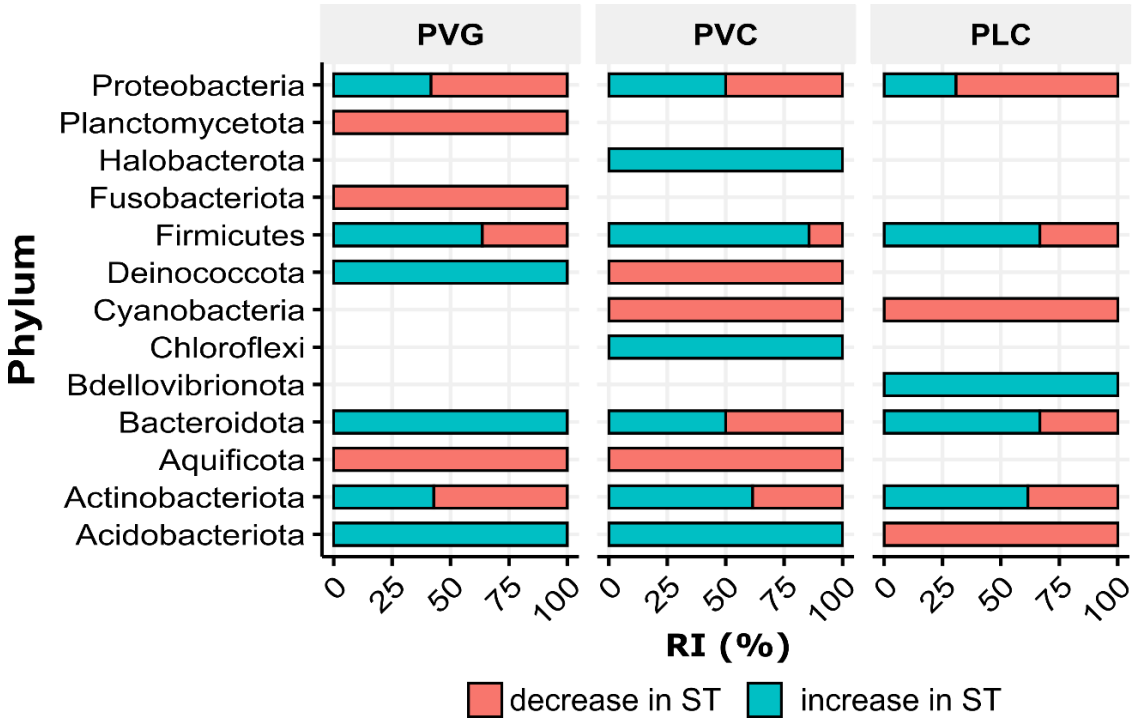


Supplementary Figure 1. Independent domestication events lead to similar effects on members of the seed microbiome. (a) and (b) refer to experiment PVG; (c) and (d) to experiment PVC; (e) and (f) to experiment PLC. (a,c,e); relative abundance of the most abundant families. (b,d,f); differentially abundant members of the seed microbiome in wild and domesticated seeds grouped by family and by domestication event calculated with DESeq2 (FDR<0.01, "BH" method).



Supplementary Figure 2. Relative abundance of the indicator species discriminating wild vs. domesticated accessions. (a) experiment PVG; (b) experiment PVC; (c) experiment PLC. (a,b,c); relative abundance of the indicator species. SV names in bold

and in a red box highlight the subselection of indicator species allowing high accuracy in discriminating wild vs domesticated accessions (see materials and methods).



Supplementary Figure 3. Overall increase in stochasticity processes in governing *Firmicutes* and *Bacteroidota* assembly in domesticated accessions. Differences in the relative importance of stochasticity processes (ST) at Phylum level between domesticated and wild accessions.

Supplementary tables

	PVG		PVC		PLC	
Cations	PC1	PC2	PC1	PC2	PC1	PC2
Fe	1.52	0.21	-0.22	-0.84	-1.15	0.05
Ca	1.47	0.32	1.33	0.00	-0.70	0.80
Cu	1.43	-0.07	0.69	-0.29	-0.87	0.09
Mg	1.27	-0.64	1.09	-0.54	0.07	-0.58
Ni	1.25	0.58	-0.32	-1.07	-0.81	-0.56
K	1.25	-0.39	-0.00	-0.91	-0.78	-0.70
Mn	1.19	0.38	1.31	0.01	-0.46	0.55
Zn	1.18	-1.03	0.32	0.19	-1.16	0.09
Cr	1.04	0.79	-0.24	0.57	-0.51	1.11
Al	0.70	0.76	-0.44	0.31	-0.24	-0.57
P	0.50	-1.41	-0.37	-1.24	-1.05	-0.62
Cd	-0.23	0.12	-0.04	0.16	-0.11	-0.28

Supplementary Table S2. First two principal components scores constructed as linear combinations or mixtures of the initial variables and the contribution of each single response variable(cation) to the new compnent.

Cations	PVG	
	AD-AW	MD-MW
Ca	t(51)=14.8, p<0.01	t(32)=28, p<0.01
Mg	t(48)=6.2, p<0.01	t(42)=11, p<0.01
P	t(52)=5, p<0.01	t(38)=1.0, p=0.3
K	t(52)=11.3, p<0.01	t(35)=14.4, p<0.01

Cations	PVC	
	AD-AW	MD-MW
Ca	t(25)=7.4, p<0.01	t(29)=16, p<0.01
Mg	t(8)=1.9, p=0.09	t(35)=6.8, p<0.01
P	t(18)=2.4, p=0.02	t(30)=-1.8, p=0.07
K	t(11)=1.8, p=0.08	t(23)=0.97, p=0.3

Cations	PLC	
	AD-AW	MD-MW
Ca	t(7)=4.5, p<0.01	t(22)=4.1, p<0.01
Mg	t(15)=-3.6, p<0.01	t(25)=-1.9, p=0.05
P	t(10)=1.0, p=0.3	t(23)=2.3, p=0.02
K	t(13)=-0.0, p=0.9	t(27)=3.1, p<0.01

Supplementary Table S3. Welch's t-test results of the comparisons between major cations concentration's means. PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Index	PVG			
	AD	AW	MD	MW
OTUs	21 ± 15 (a)	56 ± 14 (b)	25 ± 12 (a)	62 ± 19 (b)
Sequence	66584 ± 15174 (a)	87958 ± 24197 (b)	83388 ± 16398 (ab)	87958 ± 20331 (b)
Simpson	0.56 ± 0.08 (a)	0.85 ± 0.11 (b)	0.68 ± 0.10 (c)	0.92 ± 0.03 (b)
Shannon	1.055 ± 0.34 (a)	2.71 ± 0.65 (b)	1.51 ± 0.42 (a)	3.18 ± 0.40 (c)

Index	PVC			
	AD	AW	MD	MW
OTUs	49 ± 10 (ab)	62 ± 15 (a)	56 ± 19 (ab)	45 ± 7 (b)
Sequence	28849 ± 31043 (a)	75579 ± 38470 (a)	66251 ± 42283 (a)	85445 ± 25840 (a)
Simpson	0.91 ± 0.03 (a)	0.91 ± 0.03 (ab)	0.90 ± 0.03 (ab)	0.84 ± 0.07 (c)
Shannon	2.86 ± 0.36 (a)	2.84 ± 0.40 (ab)	2.77 ± 0.43 (ab)	2.22 ± 0.38 (c)

Index	PLC			
	AD	AW	MD	MW
OTUs	49 ± 29 (a)	52 ± 16 (a)	56 ± 18 (a)	41 ± 11 (a)
Sequence	30095 ± 31027 (ab)	81021 ± 60373 (a)	31237 ± 18509 (b)	46719 ± 27452 (ab)
Simpson	0.93 ± 0.02 (a)	0.88 ± 0.07 (a)	0.92 ± 0.02 (a)	0.88 ± 0.06 (a)
Shannon	3.05 ± 0.40 (a)	2.66 ± 0.66 (a)	3.18 ± 0.32 (a)	2.61 ± 0.48 (a)

Supplementary Table S4. Number of sequences, total OTUs and diversity indices expressed as average and standard deviation after pooling individual plant/genotypes into clusters based on domestication event (Andean Domesticated; AD. Andean Wild; AW. Mesoamerican Domesticated; MD. Mesoamerican Wild; MW). Letter in parenthesis indicated the one-way analysis of variance (ANOVA) performed for each index analyzed (Tukey's HSD Multiple Comparison Test, $p<0.01$). PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

	PVG			
OTUs ID	Core microbiome domesticated accessions			
	Phylum	Class	Order	Family
12	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
349	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
710	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
739	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae
772	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae
	Core microbiome wild accessions			
1	Firmicutes	Bacilli	Bacillales	Bacillaceae
2	Proteobacteria	Gammaproteobacteria	Burkholderiales	Hydrogenophilaceae
3	Deinococcota	Deinococci	Thermales	Thermaceae
4	Bacteroidota	Kapabacteria	Kapabacteriales	
5	Bacteroidota	Bacteroidia		
7	Acidobacteriota	Blastocatellia		
9	Firmicutes	Bacilli	Bacillales	Bacillaceae
12	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
20	Deinococcota	Deinococci	Thermales	Thermaceae
190	Proteobacteria	Gammaproteobacteria	Burkholderiales	Hydrogenophilaceae
349	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
409	Bacteroidota	Kapabacteria	Kapabacteriales	
456	Deinococcota	Deinococci	Thermales	Thermaceae
686	Firmicutes	Bacilli	Alicyclobacillales	Alicyclobacillaceae
710	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
721	Firmicutes	Bacilli	Bacillales	Bacillaceae
739	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae
746	Firmicutes	Bacilli	Bacillales	Bacillaceae

772	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae
896	Firmicutes	Bacilli	Alicyclobacillales	Alicyclobacillaceae

PVC				
OTUs ID	Core microbiome domesticated accessions			
	Phylum	Class	Order	Family
1	Firmicutes	Bacilli	Bacillales	Bacillaceae
2	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
	Core microbiome wild accessions			
1	Firmicutes	Bacilli	Bacillales	Bacillaceae
2	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
3	Deinococcota	Deinococci	Thermales	Thermaceae
4	Bacteroidota	Kapabacteria	Kapabacteriales	
5	Bacteroidota	Bacteroidia		
7	Acidobacteriota	Blastocatellia		
9	Firmicutes	Bacilli	Bacillales	Bacillaceae

PLC				
OTUs ID	Core microbiome domesticated accessions			
	Phylum	Class	Order	Family
	Core microbiome wild accessions			
10	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae
12	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae
18	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
94	Actinobacteriota	Actinobacteria	Kineosporiales	Kineosporiaceae
109	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae

Supplementary Table S5. Members of the core microbiome (80% prevalence) per experiment and biological status. PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Phylum	PVG			
	AD	AW	MD	MW
Proteobacteria	98 ± 2	51 ± 24	95 ± 6	37 ± 14
Firmicutes	0.7 ± 1	29 ± 16	2 ± 3	35 ± 7
Deinococcota	0.2 ± 0.6	11 ± 7.5	0.7 ± 0.1	0.7 ± 1
Bacteroidota	0.0 ± 0.1	3 ± 2	0.1 ± 0.4	4 ± 3

Acidobacteriota	0.0 ± 0.0	2 ± 2	0.1 ± 0.3	3 ± 2
Family				
Pseudomonadaceae	85 ± 26	35 ± 30	74 ± 18	13 ± 15
Bacillaceae	0.2 ± 0.1	23 ± 13	1 ± 3	27 ± 7
Thermaceae	0.2 ± 0.6	12 ± 7	0.7 ± 1	16 ± 8
Burkholderiaceae	1 ± 3	5 ± 2	12 ± 16	8 ± 4
Hydrogenophilaceae	0.1 ± 0.1	8 ± 5	0.5 ± 1	10 ± 5

Phylum	PVC			
	AD	AW	MD	MW
Proteobacteria	34 ± 7	41 ± 12	37 ± 16	20 ± 9
Firmicutes	22 ± 11	19 ± 10	23± 12	30 ± 12
Bacteroidota	21 ± 5	11 ± 8	15 ± 11	23 ± 9
Actinobacteriota	13 ± 11	17 ± 11	12 ± 8	10 ± 11
Deinococcota	7 ± 5	6 ± 6	5 ± 5	8 ± 8
Family				
Bacillaceae	17 ± 10	15 ± 10	17 ± 11	26 ± 13
Bacteroidia	15 ± 8	8 ± 8	11 ± 9	14 ±9
Hydrogenophilaceae	8 ± 6	9 ± 5	10 ± 9	13 ± 9
Thermaceae	7 ± 5	8 ± 8	5 ± 5	8 ± 8
Microbacteriaceae	3 ± 3	6 ± 6	3 ± 4	0.4 ± 0.8

Phylum	PLC			
	AD	AW	MD	MW
Proteobacteria	36 ± 12	53 ± 22	50 ± 20	53 ± 17
Actinobacteriota	37 ± 15	30 ± 21	24± 16	29 ± 13
Firmicutes	12 ± 8	8 ± 9	12 ± 10	14 ± 14
Bacteroidota	11 ± 5	6 ± 8	9 ± 7	3 ± 3
Chloroflexi	0 ± 0	0.4 ± 1	2 ± 4	0.2 ± 1
Family				
Microbacteriaceae	14 ± 7	4 ± 4	7 ± 8	9 ± 8
Sphingomonadaceae	9 ± 7	12 ± 12	7 ± 4	10 ± 8

Pseudomonadaceae	0 ± 0	6 ± 7	7 ± 12	8 ± 10
Staphylococcaceae	3 ± 4	3 ± 6	3 ± 5	8 ± 11
Rhizobiaceae	1 ± 1	6 ± 10	2 ± 3	6 ± 6

Supplementary Table S6. Mean relative abundances of the top 5 most abundant Phylum and Family taxonomic classifications of the members of the seed microbiome per experiment and domestication event (Andean Domesticated; AD. Andean Wild; AW. Mesoamerican Domesticated; MD. Mesoamerican Wild; MW). PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Covariate	PVG	PVC	PLC
Site	-	LRT=4606, P=0.2	LRT=3179, P=0.06
Date	-	LRT=2076, P=0.13	LRT=1861, P=0.07

Supplementary Table S7. Deviance and p-value based on Likelihood-Ratio Test (LRT) for the covariates Site and Date. Site represents the regeneration site of the accessions and Date the collection year (Supplementary File 1). PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Process	PVG			
	Domesticated	Wild	Cohens'*d*	p
HoS	0.23	0.53	-6.8	< 0.001
HeS	0.008	0.01	-1.6	0.11
DL	0.024	0.11	-7.6	<0.001
DR	0.72	0.33	8.5	<0.001
HD	0.006	0.00	0.9	0.3

Process	PVC			
	Domesticated	Wild	Cohens'*d*	p
HoS	0.11	0.09	1.0	0.22
HeS	0.00	0.01	-0.8	0.28
DL	0.41	0.38	0.7	0.29
DR	0.46	0.50	-1.0	0.22
HD	0.00	0.00	1.2	0.22

Process	PLC			
	Domesticated	Wild	Cohens'*d*	p
HoS	0.07	0.08	-0.9	0.26
HeS	0.02	0.01	2.1	0.04
DL	0.48	0.50	-0.5	0.33
DR	0.40	0.39	0.46	0.36
HD	0.00	0.00	0.9	0.26

Supplementary Table S8. Average relative importance of each ecological processes in seed microbiome assembly in wild and domesticated plants. Homogeneous selection (HoS), Heterogeneous Selection (HeS), Dispersal Limitation (DL), Drift (DR), Homogenizing dispersal (HD). PVG (*Phaseolus vulgaris* greenhouse), PVC (*Phaseolus vulgaris* CIAT), PLC (*Phaseolus lunatus* CIAT).

Mantel R^2	p
0.19	0.001

Supplementary Table S9. Mantel test result (Spearman's rank correlation method, 999 permutations) between calcium and homogeneous selection distance matrices.

Chapter 4: Approaching the domesticated plant holobiont from a community evolution prospect

Riccardo Soldan*, Marco Fusi, Gail M. Preston*

*Corresponding authors

Manuscript under submission (Microbiology)

Abstract

Plants establish a pivotal relationship with their microbiome and are often conceptualised as holobionts. Nonetheless, holobiont theories have attracted much criticism, especially concerning the fact that the holobiont is rarely a unit of selection. In previous work, we discussed how the plant microbiome can be considered to be an “ecosystem on a leash”, which is subject to the influence of natural selection acting on plant traits. We proposed that in domesticated plants the assembly of the plant microbiome can usefully be conceptualised as being subject to a “double leash”, which encompasses both the effect of artificial selection imposed by the domesticator on plant traits and the leash from plant to the microbiome.

Here we approach the domesticated plant holobiont, simply defined as a community of organisms, from a community evolution point of view, and show how community heritability (a measure of community selection) complements the “double-leash” framework in providing a community-level view of plant domestication and its impact on plant-microbe interactions. We also propose simple experiments that could be performed to investigate whether plant domestication has altered the potential for community selection at the holobiont level.

Introduction

In recent years, we have come to appreciate that as single individuals can have heritable phenotypes, the effect of these phenotypes on a community or at ecosystem level can also lead to heritable community and ecosystem phenotypes [1]. Because microorganisms are ubiquitous in virtually all environments on earth, plant and animal species do not evolve in isolation but are colonized by a rich and diverse microbial community [2]. Therefore, the effect of a heritable host phenotype can lead to a heritable component of the host microbiome. Potentially, the host and its microbiome, a holobiont, can be subject to community selection which may lead to community evolution [3]. Community evolution is the "outcome of selection operating at multiple levels that results in the differential survival and proliferation of communities" [4]. Through community evolution, specific microbial community phenotypes could arise, which may provide positive microbe-to-host effects.

One advantage of approaching plant-microbe interactions from a community selection perspective is that community selection does not make any assumptions about the level at which selection is occurring or whether selection acts on the holobiont as a unit [4]. In this approach, we simply define the holobiont as a community of organisms and microorganisms in which community selection can occur [4]. Central to theories of community evolution is the concept of heritability. This aspect is often confused with inheritance, but they refer to two different processes. Inheritance is the vertical transmission of members of the microbiome from one generation to the other. Heritability measures how much of the variation in the microbiome can be attributed to host genetic variation [5] and it can be calculated for single microbial members or at a community

level. As we will see later, we argue that heritability, when expanded to a community level, can provide useful insight into the evolution of the holobiont in agricultural settings.

Broad-sense community heritability

Broad-sense community heritability (H^2_C) expands the concept of broad-sense heritability calculated for single microbial members to a community level. In practice, it measures how much of the variation at a community level is explained by the host genotype. This can be achieved by condensing information on microbial composition into univariate scores using an ordination method, such as non-metric multidimensional scaling (NMDS), and estimating how much of the variation in the NMDS score can be attributed to the genetic variance of the host through an ANOVA (Analysis of Variance) [6, 7].

H^2_C is important because finding a statistically significant effect of the host genotype in influencing microbial community composition provides evidence of community selection [6]. In other words, it provides evidence of genetic covariance between host and microbiome, documenting a heritable effect on the host-associated microbiome due to plant phenotypes.

When community selection occurs, generating holobionts with differential survival, community evolution could lead to specific host-microbiome interactions spanning from mutualism to parasitism. Practically, broad-sense community heritability can be calculated through common garden experiments [6], in which for several plant genotypes and replicates of plant genotypes (clones) community data is measured (*e.g* abundance of lichens, insects, microorganisms) [8–10] and the ANOVA on the NMDS score is calculated, with the host genotype being the explanatory variable. In this approach, the

value of community heritability is the R^2 of the ANOVA and the significance is given by the p-value of the factor plant genotype.

Conceptually, broad-sense community heritability (H^2_C) is proportional to the product of the broad-sense heritability of the host phenotypic traits ($\boldsymbol{H_\theta^2}$) (for example, the cumulative effect of root length, root architecture, and so forth) and the intensity of the host phenotypic effect at the microbial community level (also referred to as Interspecific Indirect Genetic effects, IIGEs) ($\boldsymbol{\gamma}$), relative to the total selection ($\boldsymbol{\gamma + E_n}$), that is the sum of the community-level selection intensity (the effect of IIGEs) and other sources of variation in the microbiome community ($\boldsymbol{E_n}$) (for example soil chemical and physical properties) [6].

$$\boldsymbol{H_C^2 \propto H_\theta^2 \frac{\gamma}{\gamma + E_n}}$$

IIGEs could be seen as equivalent to the host-to-microbe effects used in the framework of Foster et al. [9], however IIGEs can be more generally applied to any community interactions. When host phenotypic effects on microbial communities (IIGEs) are weak (low γ), and the proportion of total selection due to other sources of variation is relatively high (high $\boldsymbol{E_n}$), broad-sense community heritability (H^2_C) approaches zero [6]. For example, low microbial H^2_C could be the result of transplanting plants into different soils, with the consequence that the soil microbiome is the main determinant of the rhizosphere microbiome [11]. Under these conditions, any IIGEs resulting from certain root phenotypic traits could be overcome by stronger drivers of microbiome assembly, in this case, the different soils (high E_n), and, the rhizosphere microbiome would have low H^2_C.

The effect of plant domestication on broad-sense community heritability

In our recent work [12] we conceptualized how the domesticated plant microbiome could differ from that of wild progenitors due to an effect of domesticated plant phenotypes by expanding the "ecosystem on a leash" framework of Foster et al [13]. We proposed that in domesticated plants the plant microbiome can be conceptualised as being subject to a "double leash" that includes the effect of artificial selection imposed by the domesticator on plant traits and the effect of plant traits on the microbiome. We predicted a reduction in positive-microbe-to-host effects when domesticated plant phenotypes arose due to artificial selection. However, we did not discuss how our double-leash framework could be investigated at a community level. In this sense, broad-sense community heritability could help us to understand how community evolution may have been affected in domesticated vs wild plants. This is important because community evolution could lead to new community phenotypes that could have important agricultural repercussions. For example, domestication may have led to plant holobionts that are less resilient to environmental stresses than their wild progenitors.

We propose that plant domestication could lower H^2_C of the host-microbiome by 1) reducing community-level selection on the microbiome ($\boldsymbol{\gamma}$), and 2) increasing other potential sources of variation in the host-microbiome community ($\boldsymbol{E_n}$).

1) The effect of plant domestication on community-level selection ($\boldsymbol{\gamma}$)

One of the main consequences of plant domestication has been a reduction in plant genetic diversity (domestication bottleneck) [14]. Phenotypic traits in wild ancestors with potentially high IIGEs on the microbial community could have been lost due to the

domestication bottleneck [15], particularly if these traits did not positively contribute to the plant phenotype from the domesticators' point of view [12]. The probability of losing plant phenotypic traits with high IIGEs would increase with the strength of the domestication bottleneck (Figure 1a). Additionally, artificial selection could have directly selected against certain plant phenotypes if those traits were unsuitable for an agricultural ecosystem. For example, the reduction of plant secondary metabolites to make plants more palatable or less toxic is a common feature of domesticated crops [16]. Secondary metabolites are known to play an important role in host-microbe interactions [17, 18], possibly exerting high IIGEs.

Thus, we might generally predict to observe a reduction in community-level selection and in broad-sense community heritability as a consequence of domestication, which is consistent with the double-leash hypothesis. However, an essential condition for community selection to occur is the existence of heritable plant phenotypic variations exerting IIGEs. If domesticated plant phenotypes exert strong host-to-microbe effects , but there is little or no heritable plant phenotypic variation in IIGEs across individuals, community selection at holobiont level cannot occur. However, providing there is heritable phenotypic variation, H^2_C of the domesticated plant microbiome would not necessarily decrease with the strength of the domestication bottleneck as the loss of plant phenotypes could be balanced out by strong IIGEs induced by domesticated plant phenotypes (Figure 1b).

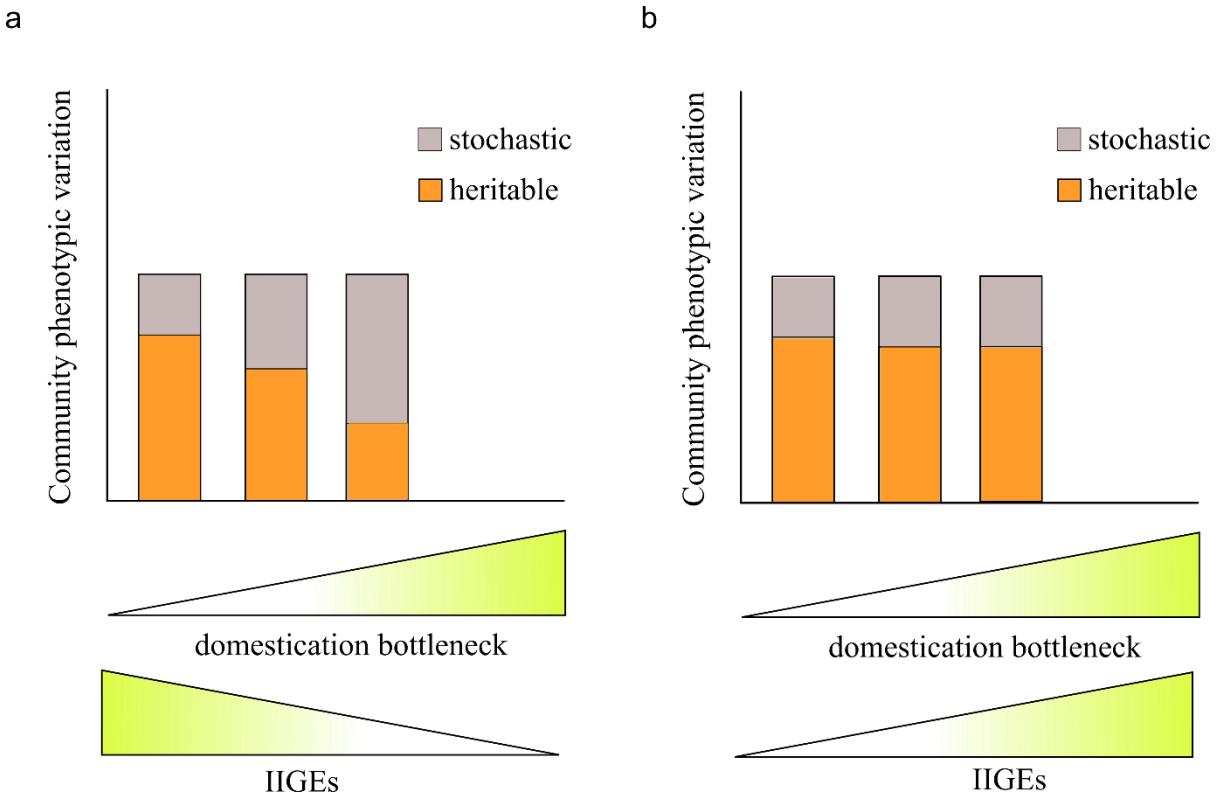


Figure 1. The effect of plant domestication on community heritability depends on the combined effect of the domestication bottleneck and the Interspecific Indirect Genetic Effects (IIGEs) of the host phenotypes. (a) A reduction in plant genetic diversity due to domestication can lead to loss of host phenotypes with strong IIGEs (*e.g* concentration of secondary metabolites), which causes a reduction in community heritability of the microbiome. This could be exacerbated when traits subject to artificial selection in domesticated plant phenotypes exert low IIGEs on the microbial communities (*e.g* traits selected for aesthetic purposes). (b) When domesticated plant phenotypes exert high IIGEs, loss of plant phenotypes due to the domestication bottleneck could be compensated, assuming there are sufficient heritable phenotypic variations to observe the effect of these plant phenotypes.

2) The effect of plant domestication on other sources of variation in the microbiome community ($\boldsymbol{E_n}$)

The effects of domestication on plant phenotypes and the plant microbiome also need to be interpreted in the context of agricultural environments, in which agricultural inputs such as fertilization or soil management practices can have a larger effect on the host-associated microbiome than IIGEs of host traits, thus increasing $\boldsymbol{E_n}$ [19]. For example, inorganic nitrogen applications have been shown to affect both taxonomical and functional profiles of rhizosphere microbial communities of wheat [20]. At the same time, intensive agriculture reduces soil microbiome alpha-diversity, creating biotic homogenization [21]. This could ultimately limit the number of community members affected by IIGEs.

Microbiome broad-sense community heritability of different plant compartments

While in some contexts it is useful to consider broad-sense community heritability at the level of the whole plant, it is also important to consider that different host compartments can be colonized by different, but interconnected microbial communities [11, 22]. Therefore, a host phenotypic trait (*e.g* leaf tannin concentration) could have different IIGEs on microbial communities inhabiting different host compartments (*e.g* phyllosphere microbiome compared to rhizosphere microbiome). In this example, while the phyllosphere microbiome could be affected directly by tannin concentrations in leaves, the rhizosphere microbiome could be also affected as leaves fall on the ground, where they contribute to and are decomposed by the soil microbiota. In turn, the seed microbiome could be affected by changes in the assembly of the phyllosphere and rhizosphere microbiome, which both contribute microorganisms to the seed microbiome.

At the same time, even within the same plant compartment, microbial communities change dynamically based on the plant developmental stage [23]. For this reason, microbial communities inhabiting different plant compartments at different developmental stages will have different H^2_C. The domestication process will, therefore, differentially affect the microbial H^2_C depending on the host compartment, developmental stage, and their interaction. The effect of IIGEs is also likely to be stronger where community diversity is lower (*e.g* root endosphere vs rhizosphere) [13]. For this reason, we would expect H^2_C to be higher for endophytic microbial communities, which can then lead to stronger community-level selection.

Practical implications

Approaching the holobiont from a community evolution perspective could enable us to better understand the evolution of the holobiont and identify plant traits leading to the assembly of specific microbial communities. This view is different but complementary to what we have described in our double-leash framework, which focuses on the host individual level. The double-leash framework posits that as when domesticated plant phenotypes are selected through artificial selection by the domesticator and they have no consequences for plant fitness, these traits are unlikely to have evolved as an attempt of the host to control the microbiota and receive positive microbe-to-host effects. The community evolution perspective also supports the conceptual development of scenarios in which domestication could lead to reduced community selection through a reduction in IIGEs, irrespective of whether interactions have positive implications for the host, and requires us to consider the possibility that a loss of genetic diversity could lead to scenarios where heritable variation in plant phenotypic traits exerting IIGEs is insufficient

for community selection to occur. In such cases introgression of wild germplasm into domesticated populations could be needed to reintroduce variation in IIGEs in order to be able to select effectively for specific microbe-host interactions in plant breeding programmes. Importantly, viewing plant domestication through the lens of community evolution facilitates the design of experiments to detect and quantify IIGEs, which will not only help us to understand whether domestication has led to a loss in plant phenotypic variation that was exerting strong IIGEs, but could also enable us to identify plant traits responsible for certain aspects of microbiome assembly.

A simple initial experiment could be to grow wild and domesticated plants in a small geographical area in their correspondent ecosystems, that is natural ecosystem for wild plants (or as close as possible to natural) and agricultural for domesticated plants. Plant genotypes should be selected to encompass the population genetic diversity of the species being studied. Each genotype would be replicated paying attention that replicates of the same genotype are identical (clones or highly homozygous individuals). The microbiome of different plant compartments would then be sampled and analysed. This simple experiment could help us understand i) how wild and domesticated plants differ in terms of community selection, ii) whether domesticated plants have lower broad-sense community heritability compared to wild progenitors. This would help demonstrate whether domestication has led to i) a reduction in IIGEs, ii) insufficient heritable plant phenotypic variation to detect IIGEs, or iii) a combination of both. Subsequent experiments involving progeny from crosses and backcrosses between wild and domesticated plants would help to resolve these possibilities.

Moreover, we could use a similar approach to understand whether broad-sense community heritability of the microbiome is indeed plant-compartment dependent. It

could also be possible to predict which plant traits are influencing microbiome assembly through multivariate statistics, as reported for root length and the rhizosphere microbiome [24]; and to investigate the genetic basis of IIGEs through forward genetics, quantitative trait locus (QTL) mapping and genome-wide association studies [25].

Concluding remarks

Approaching the holobiont, defined as host and host-associated microbiome, from a community level perspective avoids many of the debated aspects of holobiont theories, such as co-evolution and inheritance. In the context of plant domestication, approaching the holobiont from a community level prospect could give us a more holistic view of plant-microbe interactions in the agricultural ecosystem, pave the way to the identification of plant traits exerting strong IIGEs and cast light on the assembly of domesticated plant microbiomes.

References

1. Bailey JK, Schweitzer JA, Úbeda F, Koricheva J, LeRoy CJ, Madritch MD, et al. From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. *Philos Trans R Soc Lond B Biol Sci* 2009; 364: 1607–1616.
2. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *PNAS* 2013; 110: 3229–3236.
3. Simon J-C, Marchesi JR, Mougel C, Selosse M-A. Host-microbiota interactions: from holobiont theory to analysis. *Microbiome* 2019; 7: 5.
4. Whitham TG, Allan GJ, Cooper HF, Shuster SM. Intraspecific Genetic Variation and Species Interactions Contribute to Community Evolution. *Annual Review of Ecology, Evolution, and Systematics* 2020; 51: 587–612.
5. Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. *Nat Commun* 2021; 12: 5141.
6. Shuster SM, Lonsdorf EV, Wimp GM, Bailey JK, Whitham TG. Community Heritability Measures the Evolutionary Consequences of Indirect Genetic Effects on Community Structure. *Evolution* 2006; 60: 991–1003.
7. Whitham TG, Bailey JK, Schweitzer JA, Shuster SM, Bangert RK, LeRoy CJ, et al. A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* 2006; 7: 510–523.
8. Lamit LJ, Lau MK, Næsborg RR, Wojtowicz T, Whitham TG, Gehring CA. Genotype variation in bark texture drives lichen community assembly across multiple environments. *Ecology* 2015; 96: 960–971.

9. Compson ZG, Hungate BA, Whitham TG, Meneses N, Busby PE, Wojtowicz T, et al. Plant genotype influences aquatic-terrestrial ecosystem linkages through timing and composition of insect emergence. *Ecosphere* 2016; 7: e01331.
10. Ferrier SM, Bangert RK, Hersch-Green EI, Bailey JK, Allan GJ, Whitham TG. Unique arthropod communities on different host-plant genotypes results in greater arthropod diversity. *Arthropod-Plant Interactions* 2012; 6: 187–195.
11. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 2012; 488: 91–95.
12. Soldan R, Fusi M, Cardinale M, Daffonchio D, Preston GM. The effect of plant domestication on host control of the microbiota. *Commun Biol* 2021; 4: 1–9.
13. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. *Nature* 2017; 548: 43–51.
14. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature* 2002; 418: 700–707.
15. Brown TA. Is the domestication bottleneck a myth? *Nature Plants* 2019; 5: 337–338.
16. Moreira X, Abdala-Roberts L, Gols R, Francisco M. Plant domestication decreases both constitutive and induced chemical defences by direct selection against defensive traits. *Scientific Reports* 2018; 8: 1–11.
17. Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, et al. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *PNAS* 2020; 117: 3874–3883.

18. Kudjordjie EN, Sapkota R, Steffensen SK, Fomsgaard IS, Nicolaisen M. Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome* 2019; 7: 59.

19. Hartman K, van der Heijden MGA, Wittwer RA, Banerjee S, Walser J-C, Schlaeppi K. Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. *Microbiome* 2018; 6: 14.

20. Kavamura VN, Hayat R, Clark IM, Rossmann M, Mendes R, Hirsch PR, et al. Inorganic Nitrogen Application Affects Both Taxonomical and Predicted Functional Structure of Wheat Rhizosphere Bacterial Communities. *Front Microbiol* 2018; 9.

21. Olden JD, LeRoy Poff N, Douglas MR, Douglas ME, Fausch KD. Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution* 2004; 19: 18–24.

22. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. *Science* 2009; 326: 1694–1697.

23. Moroenyane I, Mendes L, Tremblay J, Tripathi B, Yergeau É. Plant Compartments and Developmental Stages Modulate the Balance between Niche-Based and Neutral Processes in Soybean Microbiome. *Microb Ecol* 2021.

24. Pérez-Jaramillo JE, Carrión VJ, Bosse M, Ferrão LFV, de Hollander M, Garcia AAF, et al. Linking rhizosphere microbiome composition of wild and domesticated Phaseolus vulgaris to genotypic and root phenotypic traits. *ISME J* 2017; 11: 2244–2257.

25. Doganlar S, Frary A, Tanksley SD. The genetic basis of seed-weight variation: tomato as a model system. *Theor Appl Genet* 2000; 100: 1267–1273.

Chapter 5: From Macro to Micro: a combined bioluminescence-fluorescence approach to monitor bacterial localization

Riccardo Soldan, Nattapong Sanguankiattichai, Marcel Bach-Pages, Indra Bervoets, Wei E. Huang, Gail M. Preston*

*Corresponding authors

Environmental microbiology 23(4): 2070-2085 (2021)

https://doi.org/10.1111/1462-2920.15296

Originality-Significance Statement

The detection of bacterial bioluminescence is a valuable approach to macroscopically monitor bacterial spatial localization. However, imaging at the cellular level cannot be achieved with standard bioluminescence detection techniques, and is therefore typically undertaken using fluorescent proteins (FP) as biomarkers. Identifying specific samples to be imaged with fluorescence microscopy in a host or environmental context can be challenging and time-consuming when there is no *a priori* information on bacterial localization. Here we demonstrate the application of dual *lux*-FP constructs to bridge the gap between macro- and micro-imaging.

We describe the development of a new mini-Tn7 vector library containing *lux* and *lux*-eYFP operons that can be used to chromosomally tag Proteobacteria. We demonstrate the effectiveness of the *lux*-eYFP constructs to investigate systemic infection by plant pathogenic bacteria, and show that the halo blight pathogen *Pseudomonas syringae* pv. *phaseolicola* can colonise the vascular system of common bean (*Phaseolus vulgaris*). We also demonstrate the effectiveness of bioluminescence for monitoring the impact of plant immune responses on bacterial populations. The constructs and approach described here can be used to facilitate the study of a wide range of host-microbe, microbe-microbe and microbe-environment interactions *in situ.*

Abstract

Bacterial bioluminescence is widely used to study the spatiotemporal dynamics of bacterial populations and gene expression *in vivo* at a population level, but cannot easily be used to study bacterial activity at the level of individual cells. In this study we describe the development of a new library of mini-Tn7-*lux* and *lux::eyfp* reporter constructs that provide a wide range of *lux* expression levels, and which combine the advantages of both bacterial bioluminescence and fluorescent proteins to bridge the gap between macro- and micro-scale imaging techniques. We demonstrate that a dual bioluminescence-fluorescence approach using the *lux* operon and eYFP can be used to monitor bacterial movement in plants both macro- and microscopically, and demonstrate that *Pseudomonas syringae* pv *phaseolicola* can colonise the leaf vascular system and systemically infect leaves of common bean (*Phaseolus vulgaris*). We also show that bacterial bioluminescence can be used to study the impact of plant immune responses on bacterial multiplication, viability and spread within plant tissues. The constructs and approach described in this study can be used to study the spatiotemporal dynamics of bacterial colonisation and to link population dynamics and cellular interactions in a wide range of biological contexts.

Introduction

In recent years, advances in DNA sequencing technology and molecular biology techniques, have provided important new insights into the composition and functions of host-associated bacteria for both plants (Bulgarelli *et al.*, 2012; Rybakova *et al.*, 2017) and animals (Turnbaugh *et al.*, 2008; Le Chatelier *et al.*, 2013). However, although these methods are of pivotal importance in furthering our understanding of the biological processes governing microbial communities and their functions, they often provide limited insight into the spatio-temporal aspects of host-microbe and microbe-microbe interactions. Such methods typically aggregate information across spatially-distributed populations, with information being collected for a limited number of samples and timepoints (Prosser, 2010; Fierer and Ladau, 2012). However, the output of an interaction between a bacterial pathogen and its host (animal or plant) greatly depends on the type of tissue that is colonized, the local microenvironment (spatial relationship) and the movement of the pathogen over time within the host (spatiotemporal relationship) (Lamichhane *et al.*, 2015). The homogenisation of a limited number of biological samples to isolate macromolecules or to enumerate microbial cells can obscure variation in microbial physiology at the macro- and micro-scale. For this reason, visualizing and studying bacteria in their environment continues to be important (Propheter and Hooper, 2015) and considerable effort has been invested into developing tools to allow visualisation of bacteria in the host environment (Seleem *et al.*, 2008; Monteiro *et al.*, 2012).

For example, the spatial relationships of microbe-microbe interactions can be studied using Fluorescent *in situ* Hybridization (FISH) coupled with Confocal Laser Scanning

Microscopy (CLSM) (Earle *et al.*, 2015; Soldan *et al.*, 2019). The main advantage of FISH is the observation of spatial relationships between multiple members of a microbial community without necessarily knowing *a priori* the composition of the community itself. However, FISH requires fixed samples, thus preventing the study of real-time interactions. For this reason, bacteria tagged with fluorescent proteins (*e.g* YFP, GFP) have been extensively used to visualize the spatial and temporal relationships of host-microbe interactions for both detrimental (Godfrey *et al.*, 2010; Cerutti *et al.*, 2017; Donati *et al.*, 2018; Planas-Marquès *et al.*, 2020) and beneficial (Compant *et al.*, 2005; Fan *et al.*, 2011; Mitter *et al.*, 2017) relationships.

However, while fluorescent proteins allow researchers to investigate bacterial colonisation at micro-scale resolution, they are not suitable for macro-scale resolution studies, such as analyses of systemic infection. This is partially due to the relatively weak fluorescence of labelled bacteria and the autofluorescence of the host (Wang *et al.*, 2007). This can result in a labour-intensive process of collecting a series of host samples that can be visualised with fluorescence microscopy in an attempt to localize labeled bacteria without knowing *a priori* their location. Light-sheet microscopy, such as the ultramicroscope from La Vision Bio Tec, allows imaging of tissue samples up to centimeters in size (Reynaud *et al.*, 2015), but even with this technology, imaging a whole plant organ, such as a leaf, would not be feasible for many plant species. Moreover, this specialised equipment is not commonly accessible to many laboratories. Wang *et al.* (2007) proposed that a brighter green-fluorescent protein variant, called GFPuv, encoded in a broad-host-range high-copy number plasmid could allow monitoring of bacterial disease at whole-plant level using UV light. However, with a few notable exceptions (Rodríguez-Moreno *et al.*, 2009; Fujie *et al.*, 2010), this approach has not been widely

adopted for macro-scale imaging. In practice, this is commonly due to limitations in sensitivity and difficulties in imaging acquisition and analysis.

An alternative approach is to use bacteria labeled with the *lux* operon. Bacterial bioluminescence (Huang *et al.*, 2006; Rico *et al.*, 2010; Brodl *et al.*, 2018; Fleiss and Sarkisyan, 2019) offers several advantages over fluorescent proteins for macro-scale imaging. Bioluminescence has generally little or no background, excitation light is not needed and emission of light depends on bacterial metabolism, so only active cells are visible (Gregor *et al.*, 2018; Planas-Marquès *et al.*, 2020). Consequently, bacterial bioluminescence has become a valuable tool to monitor bacterial localization in plants (Fukui *et al.*, 1996; Bogs *et al.*, 1998; Seleem *et al.*, 2008; Xu *et al.*, 2010) and animals (Yu *et al.*, 2004; Burkatovskaya *et al.*, 2006; Mortin *et al.*, 2007; Seleem *et al.*, 2008) at a macroscopic scale. Advanced imaging technologies, such as Electron Multiplying CCD (EMCDD) cameras, can be used to detect bioluminescence at a single cell level, as primarily demonstrated in mammalian cell systems to date (Asai *et al.*, 2008; Iwano *et al.*, 2018). However, this equipment is not widely available within the scientific community. Therefore at present bioluminescence is most often used for temporal analyses of cell growth, viability and gene expression in bacterial populations using luminometers and macroscopic imaging using standard CCD cameras.

In an attempt to solve the low signal intensity limitation, Gregor *at al.*, (2018) recently developed an enhanced *lux* operon, called *ilux,* which could allow single-cell imaging of *Escherichia coli*. The *ilux* operon consists of a series of random mutations within the *lux* operon and an *frp* gene from *Vibrio campbellii* inserted at the end of *luxE* to maximize brightness (Gregor et al., 2018). The *frp* gene codes for an FMN reductase responsible for generating $FMNH_2$ which is then oxidized in the process of light production (Gregor

et al., 2018). In *E. coli*, the additional FMN reductase caused a 2.3-fold increase in brightness, suggesting that $FMNH_2$ generated by the endogenous FMN reductase in *E. coli* was limiting for *lux* activity under the conditions tested (Gregor et al., 2018). However, this has not yet been tested in a wide range of bacteria and under a range of environmental conditions.

When using either fluorescent proteins or the *lux* operon as a bioreporter, the system used to tag bacteria has to be chosen carefully. Ideally, the marker genes should integrate into the bacterial chromosome in a single copy and in a neutral position to avoid disrupting endogenous gene functions (Lambertsen *et al.*, 2004). These criteria are satisfied by the bacterial transposon Tn7 (Craig, 1991) which can be used to insert cloned DNA at high efficiency into a specific intergenic site *att*Tn7, present in the chromosome of many Proteobacteria (Koch *et al.*, 2001).

This study aimed to develop a novel library of mini-Tn7 delivery plasmids combining the benefits of both bacterial luminescence and fluorescence for macro- and micro-scale resolution imaging, which can be used to chromosomally tag Proteobacteria. We assessed both *ilux* (Gregor *et al.*, 2018) and modified *lux* operons (*lux CDABE* and *luxCDABE-frp*, this study) in different bacterial strains to select the brightest constructs before developing the mini-Tn7 vector library. We demonstrate that these constructs can be used to visualize both local and systemic host-microbe interactions between *Phaseolus vulgaris* and the bacterial plant pathogen *Pseudomonas syringae* pv. *phaseolicola* (*Pph*), and show that *Pph* can move systemically through bean leaves by colonising the vascular system. While we have demonstrated the applicability of these constructs to study spatio-temporal relationships in plant-microbe interactions, this toolbox will be instrumental to study a

wide range of host-microbe, microbe-microbe and microbe-environment interactions *in situ.*

Results and discussion

Comparison between *lux, ilux* and *lux frp* operons

To select the brightest bioluminescence operon and achieve a low detection threshold when visualizing bacteria in the environment, we tested whether *ilux* (Gregor *et al.*, 2018) was brighter than *lux* in bacteria other than *E. coli.* We developed a low copy-number plasmid, called pRSJ-p_{nptII}::ilux (Supplementary Table 1) derived from the reporter plasmid pIJ11282 (Frederix *et al.*, 2014), in which *lux* was replaced with *ilux.* Additionally, to test the effect of *frp* on *lux*-mediated bioluminescence, we included *frp* from pGEX(-) (Gregor *et al.*, 2018) after *luxE* in pIJ11282, generating pRSJ-p_{nptII}::lux-frp (Supplementary Table 1). The three plasmids pIJ11282, pRSJ-p_{nptII}::ilux and pRSJ-p_{nptII}:lux-frp have the same backbone and promoter.

pRSJ-p_{nptII}::ilux did not yield brighter bioluminescence than pIJ11282 in *Acinetobacter baylyi* strain ADP1, *Pseudomonas fluorescens* NZ011 and *Pseudomonas syringae* pv. *phaseolicola* (*Pph*) strain 1302A (Supplementary Figure 1A). On the contrary, in *P. fluorescens* NZ011, pRSJ-p_{nptII}::ilux was approximately 20-fold less bright than pIJ11282 in both M9 minimal medium and King's B (KB) medium (Supplementary Figure 1A). We speculate that the mutations introduced into the *ilux* operon were optimized for light production under the conditions tested in *E. coli* and similar results are not guaranteed for other bacterial strains. However, pRSJ-p_{nptII}:lux-frp was two-fold brighter than pIJ11282 in *P. fluorescens* NZ011 without affecting bacterial growth (Supplementary Figure 1B),

indicating that, depending on the bacterial strain, an additional FMN reductase can increase the amount of light produced. FMN reductases oxidise NADPH to NADP+ in order to reduce FMN to $FMNH_2$. To test for the activity of the FMN reductase encoded by the *frp* gene, we measured the NADP+ concentration of cell extracts of *P. fluorescens* NZ011 harbouring pIJ11282 or pRSJ-p_{nptII}:lux-frp. We observed a higher concentration of NADP+ in bacterial cells expressing the *frp* gene, which supports the conclusion that the increased bioluminescence of *P. fluorescens* NZ011 (pRSJ-p_{nptII}:lux-frp) relative to *P. fluorescens* NZ011 (pIJ11282) is due to increased FMN reductase activity (Supplementary Figure 2).

Generation of an improved mini-Tn7 vector library for bioluminescence and combined bioluminescence-fluorescence expression

To combine the benefit of using bacterial bioluminescence for macro-scale localization with the advantages of fluorescent proteins for micro-scale localization, we developed a library of mini-Tn7 plasmids containing both a bioluminescent operon and eYFP (Supplementary Figure 3). eYFP was chosen over other fluorescent proteins for its higher brightness over eGFP (Shaner *et al.*, 2005) and for its peak emission spectra (527 nm), which would easily allow visualization in an environment with red autofluorescence (*e.g.* leaves). To provide a wide range of expression levels of the bioluminescent operon, which would allow researchers to modularly choose the best expression level according to their experimental needs, we designed the library to have five different constitutive promoters driving *lux* transcription. Promoters were selected to be highly conserved, constitutive and to provide a range of expression levels. For these reasons, we chose the *recA* promoter (Giliberti *et al.*, 2006), named p_{OXB20} and its derivatives p_{OXB16}, p_{OXB13} and p_{OXB11} (Oxford

Genetics Limited, https://www.oxgene.com/Products, Lynn *et al.*, 2015; Presnell *et al.*, 2019), and the promoter p_{nptII} (Wright and Beattie, 2004). We also included two different constitutive promoters driving the expression of eYFP, namely p_{lac} and $p_{A1/04/03}$ (Koch *et al.*, 2001). p_{lac} is a common and conserved constitutive promoter, while $p_{A1/04/03}$ has been widely adopted for gene expression in *Pseudomonas* (Godfrey *et al.*, 2010)..

To generate the mini-Tn7 vector library, we used the two cassettes with the highest light production in the tested bacterial strains, which were the *lux* and the modified *lux* and *frp* operons (Supplementary Figure 1). Initially, we used pUC18-mini-Tn7T-Gm-lux (Choi *et al.*, 2005) as a backbone. Plasmid pUC18-mini-Tn7T-Gm-lux was developed to easily insert a promoter of interest into a multiple cloning site (MCS) upstream of *luxC*. However, this backbone has a different ribosome binding site (RBS) upstream of *luxC* than the RBS found upstream of *luxC* in pIJ11282 (Frederix *et al.*, 2014). The latter RBS has been predicted to increase the translation initiation rate more than 4-fold (14243 arbitrary units compared to 3371 calculated with the RBS calculator; Farasat *et al.*, 2014; Ng *et al.*, 2015). Additionally, the upstream sequence of *luxC*, encompassing the MCS in pUC18-mini-Tn7T-Gm-lux was found to negatively affect transcription (Glassing and Lewis, 2015) (Supplementary Figure 4A). We therefore constructed a second plasmid (pRS-p_{OXB20}::lux-$p_{A1/04/03}$::eYFP) containing the RBS from pIJ11282 without the long MCS of pUC18-mini-Tn7T-Gm-lux (Supplementary Figure 4B).

We tested both constructs in *Pph* 1302A and found that the promoter with the alternate RBS and without the MCS of pUC18-mini-Tn7T-Gm-lux (p_{OXB20}) resulted in 4-7 times more luminescence compared to the original promoter ($p_{OXB20(1)}$) in both KB and M9 media, with no consequences for bacterial growth (Supplementary Figure 5A,B).

As we found these backbones to be an improvement over constructs based on pUC18-mini-Tn7T-Gm-lux, we also included constructs with the *lux* operon by itself in our library, without eYFP (Table 1; Supplementary Table 1).

The five different constitutive promoters used encompass a 15 fold range between the weakest promoter p_{OXB11} and the strongest promoter p_{nptII} in *Pph* 1302A (Figure 1. Supplementary Figure 6A). We observed a small but significant decrease in bacterial growth both in *vitro* and *in planta* when comparing *Pph* 1302A with *Pph* 1302A p_{nptII}::lux-$p_{A1/04/03}$::eYFP inoculated into the susceptible host plant *P. vulgaris* cultivar Canadian Wonder (Supplementary Figure 6B, Supplementary Figure 7A,B). However, we did not detect any difference between the growth of *Pph* 1302A and *Pph* 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP (intermediate expression) or *Pph* 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP (low expression) *in planta* (Supplementary Figure 7B).

Access to a range of promoters will enable researchers to choose an expression level of the *lux* cassette that is suitable for their experimental conditions, balancing both signal strength and any fitness cost observed (*e.g.* purpose of the experiment, equipment used to detect luminescence). While we would expect the constructs to give different expression levels in different bacterial strains, the constructs should maintain a wide expression range as the OXB promoters (Oxford Genetics Limited, UK) were all derived from the constitutive and conserved *recA* promoter (Weisemann and Weinstock, 1991).

Table 1. Mini-Tn7 vector library constructed in this study[1].

Lux	**Lux and eYFP**	**Lux and eYFP**
pRS-p_{nptII}::lux	**pRS-p_{nptII}::lux-$p_{A1/04/03}$::eYFP**	**pRS-p_{nptII}::lux-p_{lac}::eYFP**
pRS-p_{OXB20}::lux	**pRS-p_{OXB20}::lux-$p_{A1/04/03}$::eYFP**	**pRS-p_{OXB20}::lux-p_{lac}::eYFP**

pRS-p_{OXB16}::lux	**pRS-p_{OXB16}::lux-$p_{A1/04/03}$::eYFP**	**pRS-p_{OXB16}::lux-p_{lac}::eYFP**
pRS-p_{OXB13}::lux	**pRS-p_{OXB13}::lux-$p_{A1/04/03}$::eYFP**	**pRS-p_{OXB13}::lux-p_{lac}::eYFP**
pRS-p_{OXB11}::lux	**pRS-p_{OXB11}::lux-$p_{A1/04/03}$::eYFP**	**pRS-p_{OXB11}::lux-p_{lac}::eYFP**
pRS-p_{nptII}::lux-frp	**pRS-p_{nptII}::lux-frp-$p_{A1/04/03}$::eYFP**	**pRS-p_{nptII}::lux-frp-p_{lac}::eYFP**
pRS-p_{OXB20}::lux-frp	**pRS-p_{OXB20}::lux-frp-$p_{A1/04/03}$::eYFP**	**pRS-p_{OXB20}::lux-frp-p_{lac}::eYFP**

[1]For more details consult Supplementary Table 1.

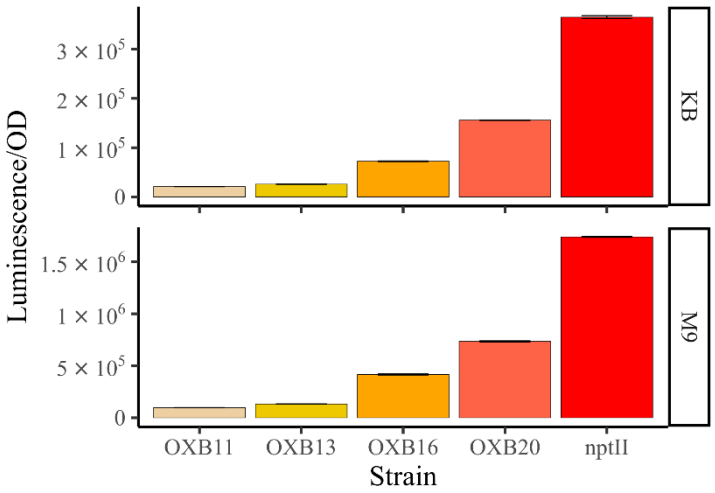


Figure 1. A library of mini-Tn7 constructs with different promoters upstream of the *lux* operon provides a wide dynamic range of luminescence when introduced into *Pseudomonas syringae* pv. *phaseolicola* 1302A (*Pph* 1302A). Luminescence values were detected at 4 hours. Time series data is presented in Supplementary Figure 6. OXB11: *Pph* 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP;

OXB13: *Pph* 1302A p_{OXB13}::lux-$p_{A1/04/03}$::eYFP; OXB16: *Pph* 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP; OXB20: *Pph* 1302A p_{OXB20}::lux-$p_{A1/04/03}$::eYFP; nptII: *Pph* 1302A p_{nptII}::lux-$p_{A1/04/03}$:::eYFP. KB (King's B medium), M9 (M9 minimal medium). OD = optical density at 600 nm. Error bar +/- SE. n=3.

Pph colonizes the vascular system of *Phaseolus vulgaris* leaves

We applied a combined luminescence-fluorescence approach to investigate whether *Pph* systemically infects *P. vulgaris* leaves. *Pph*, the causal agent of halo blight of bean, is known to be a seed-borne pathogen and to be able to cause systemic symptoms and spread within infected plants (Taylor *et al.*, 1979), but to date, there is limited knowledge of the routes and mechanisms used by *Pph* to spread systemically within host plants.

We tagged *Pph* RJ3 (Jackson *et al.*, 2000) using pRS p_{nptII}::lux-$p_{A1/04/03}$::eYFP (Table 1, Supplementary Table 1) as p_{nptII} was the strongest promoter (Figure 1) and $p_{A1/04/03}$ has previously been successfully used in *Pph* (Godfrey *et al.*, 2010). *Pph* RJ3 differs from *Pph* 1302A as it lacks a >40 Kb genomic island that contains the gene *avrPphB*, which encodes a secreted effector protein that elicits effector triggered immunity (ETI) in plants carrying the resistance gene *R3* (Jackson *et al.*, 2000). Due to the loss of *avrPphB*, *Pph* RJ3 can successfully colonize *P. vulgaris* cultivar Tendergreen (TG) without triggering a plant immune response known as the hypersensitive response (HR), which acts to restrict bacterial growth. In experiments with bacteria incubated in agar plates, we calculated that the detection limit for tagged cells, in terms of number of cells detectable per unit area, was approximately 100 CFU/mm^2 (Supplementary Figure 8) and observed a high correlation (0.99) between luminescence and CFU, as also reported by Fan *et al.*, (2008). We used bacterial bioluminescence to macroscopically observe *Pph* RJ3 in *P. vulgaris* cultivar TG leaves. We detected a clear co-localization of *Pph* RJ3 with the leaf

vasculature within 3 days after syringe infiltration (Figure 2A). We then used this information to select and collect leaf samples (vasculature cross-sections) from leaves infected with *Pph* RJ3 to be visualized with confocal microscopy. This confirmed that *Pph* RJ3 was localized inside the leaf vasculature (Figure 2B, Figure 2C). Both data types (bioluminescence and fluorescence) showed the ability of *Pph* RJ3 to move from the infection site.

We were able to detect *Pph* RJ3 cells within the xylem vessels (Figure 2B, Figure 2C), and in the collenchyma (Figure 2B, Figure 2D, Figure 2E). *Pph* RJ3 appeared to accumulate to higher densities in the collenchymatic cells of the leaf vascular system than in the xylem (Figure 2D, Figure 2E,Supplementary Video 1). *Pph* RJ3 movement in the xylem was more rapid than movement observed in collenchymatic cells (Supplementary Video 2), suggesting that *Pph* RJ3 moves within, but does not produce biofilms inside the xylem. We confirmed that *Pph* RJ3 can colonise the vascular system of intact leaves following epiphytic colonisation by spray-inoculating *Pph* RJ3 onto *P. vulgaris* leaves (Supplementary Figure 9). Images of negative controls show that there is little or no background signal in the absence of tagged bacteria (Supplementary Figure 10).

Pph strains have been reported to show substantial variation in their ability to cause systemic symptoms in host plants, while host cultivars have been reported to vary in their susceptibility to systemic symptoms. This has been attributed in some studies to systemic movement of the toxin, phaseolotoxin, within host plants (Mitchell and Bieleski, 1977). Our study confirms that the pathogen itself can move systemically within host tissues, although under our experimental conditions *Pph* RJ3 was not an aggressive systemic pathogen, consistent with the absence of detectable bacterial biofilms in the xylem and wilting symptoms, which are often the manifestation of systemic bacterial infections (Bae

et al., 2015). The constructs and methods established in this study will provide valuable tools to investigate the underlying causes of variation in systemic infection, and will enable researchers to better understand the biology and epidemiology of this disease.

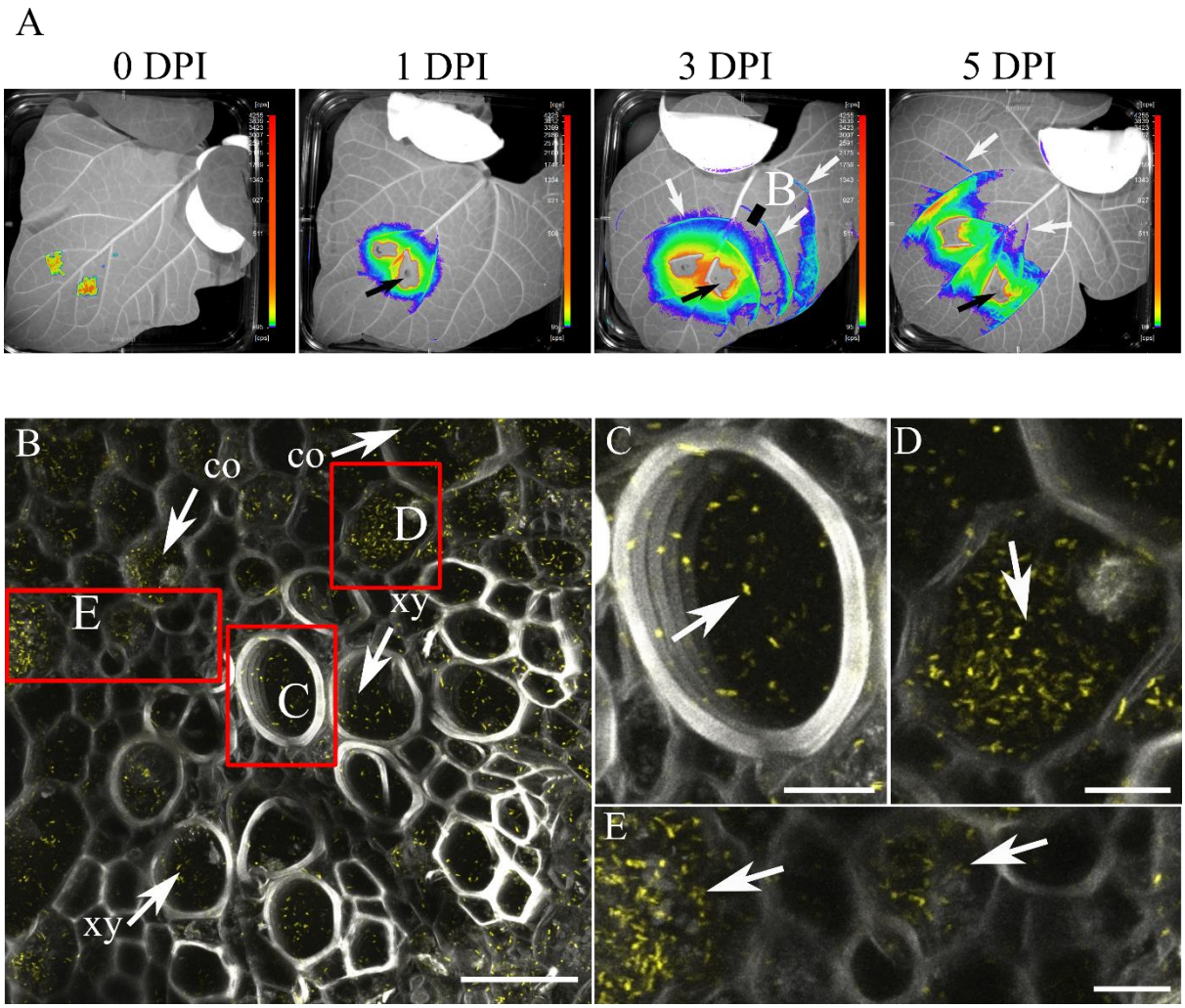


Figure 2. *Pseudomonas syringae* pv. *phaseolicola* RJ3 (*Pph* RJ3) colonizes the leaf vasculature of *P. vulgaris* cultivar TG. A: *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP was syringe-infiltrated in localised areas on the abaxial surface of *Phaseolus vulgaris* cv. Tendergreen leaves. Images were taken after 0, 1, 3, 5 DPI with the nightOWL LB 983 at 10-minute exposure. The black line

indicates cross-sections used for confocal microscopy imaging (B). The white arrows indicate co-localization of *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP with the leaf vasculature. The black arrows indicate regions of the leaves that have cts values higher than the maximum detection limit. B: Maximum projections of the leaf vasculature cross-section. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow). Xylem autofluorescence (bright grey). Collenchyma autofluorescence (dark gray). co; collenchyma. xy; xylem. Red squares indicate areas shown at higher magnification in panels C, D and E. Scale bar: 40 μm. C: *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). Xylem autofluorescence (grey). D: *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). Collenchyma autofluorescence (dark grey). E: Zoom of panel B. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). Collenchyma autofluorescence (dark grey). Scale bar: 10 μm.

Bacterial bioluminescence can be used to monitor the effect of plant immune responses on bacterial growth and viability *in planta*

Bacterial bioluminescence has been proposed as a high-throughput method to detect *in planta* growth of bacterial pathogens (Fan *et al.*, 2008). Fan *et al.*, (2008) showed that *Pseudomonas syringae* tagged with *luxCDABE* can be used as a quantitative assay to monitor bacterial growth and that CFU is correlated with bacterial bioluminescence. Moreover, luminescence allows the detection of bacterial growth without performing tissue extraction, serial dilutions, plating, and scoring, thus reducing the time needed for the experiment. We therefore examined whether the luminescent constructs developed in this study could also be used to rapidly and sensitively monitor the effect of different plant immune responses on bacterial growth and viability *in planta*. As expression of high levels of *lux* activity could have a detrimental effect on bacterial fitness in the plant

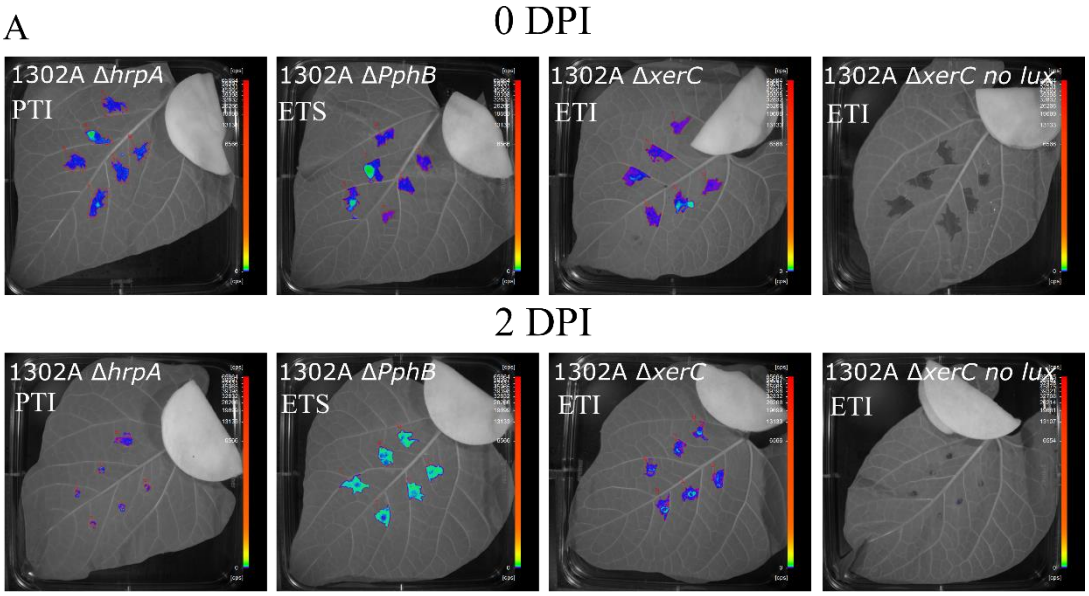
environment, we examined whether a relatively weak promoter driving *lux* expression could be used for this purpose.

We tagged *Pph* 1302A Δ*hrpA* (Supplementary Table 2), *Pph* 1302A Δ*xerC* (Lovell *et al.*, 2009) and *Pph* 1302A Δ*avrPphB* (Supplementary Table 2) using pRS-p_{OXB13}::lux (Table 1; Supplementary Table 1). *Pph* 1302A Δ*xerC* (Lovell *et al.*, 2009) lacks a recombinase (XerC) that is able to excise the genomic island containing *avrPphB,* thus it stably triggers ETI in *P. vulgaris* cultivar TG. On the contrary, *Pph* 1302A is able to excise the island and overcome resistance. *Pph* 1302A Δ*hrpA* (this study; Supplementary Table 2) cannot assemble a functioning type III secretion system (T3SS), thus it cannot secrete effectors to counteract plant defenses. Therefore, its growth *in planta* is suppressed by PAMP-Triggered Immunity (PTI) (Jones and Dangl, 2006). *Pph* 1302A Δ*PphB* (this study. Supplementary Table 2) has a functional T3SS, but lacks the *avr* gene *avrPphB,* thus it can suppress PTI and colonize TG leaves. This compatible interaction is named Effector-Triggered Susceptibility (ETS) (Jones and Dangl, 2006).

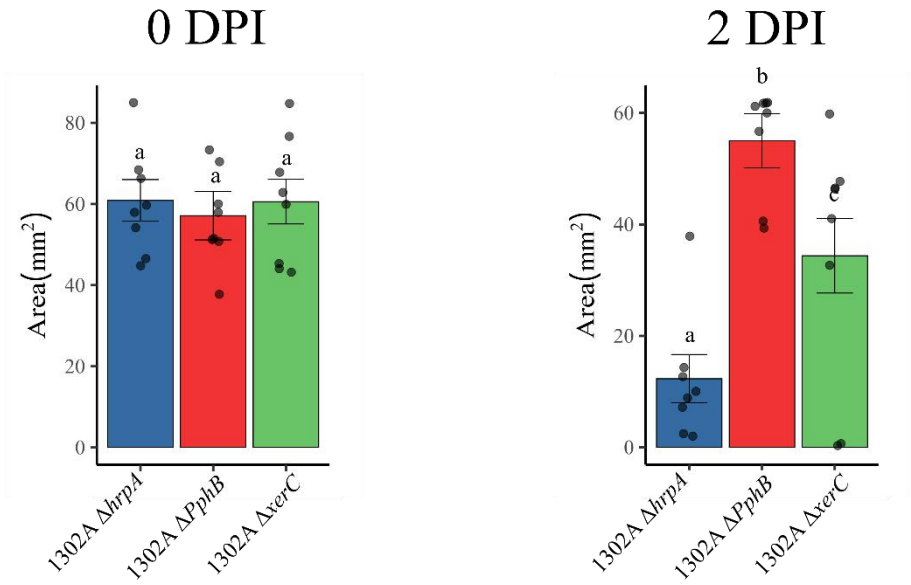
A 0 days post inoculation (DPI) we could not detect any difference between strains for both luminescence value (cts) [ANOVA, $F(2,14)=0.0616$, $p=0.94$] and infected area (mm^2) [ANOVA, $F(2,14)=0.18$, $p=0.83$] (Figure 3A-C). At 2 DPI, both luminescence signal (cts) [ANOVA, $F(2,14)=35.01$, $p<0.0001$] and infected area (mm^2) [ANOVA, $F(2,14)=19.21$, $p<0.0001$] for the different strains displayed significant differences using a two way ANOVA, as well as pairwise comparisons between strains (Figure 3A-C), indicating that both bacterial bioluminescence and infected area can be used to detect the effect of different plant immune responses on bacterial viability and growth, with significantly lower bioluminescence in the context of PTI and ETI.

Colony count analyses performed at 2 DPI agreed with luminescence data [ANOVA, $F(2,14)=56.73$, $p<0.0001$] identifying a statistically significant effect of strain. However, pairwise comparisons failed to distinguish between plants infected with *Pph* 1302A Δ*xerC* and *Pph* 1302A Δ*PphB* [$t(14)=1.6$, $p=0.06$] (Supplementary Figure 11A,B). The higher variability of colony count data compared to bioluminescence data (Figure 3,A-C Supplementary Figure 11B) may be associated with variability introduced during tissue extraction, serial dilution and plating that are avoided in bioluminescence detection. Therefore the bioluminescent reporter constructs developed in this study could be used to effectively monitor the effect of plant immune responses on bacterial growth, spread and viability in interactions between *Pph* and *P. vulgaris.*

A


0 DPI
1302A ΔhrpA
PTI
1302A ΔPphB
ETS
1302A ΔxerC
ETI
1302A ΔxerC no lux
ETI
2 DPI
1302A ΔhrpA
PTI
1302A ΔPphB
ETS
1302A ΔxerC
ETI
1302A ΔxerC no lux
ETI

B


0 DPI
2 DPI
Area(mm2)
a
a
a
b
a
c
1302A ΔhrpA
1302A ΔPphB
1302A ΔxerC

C

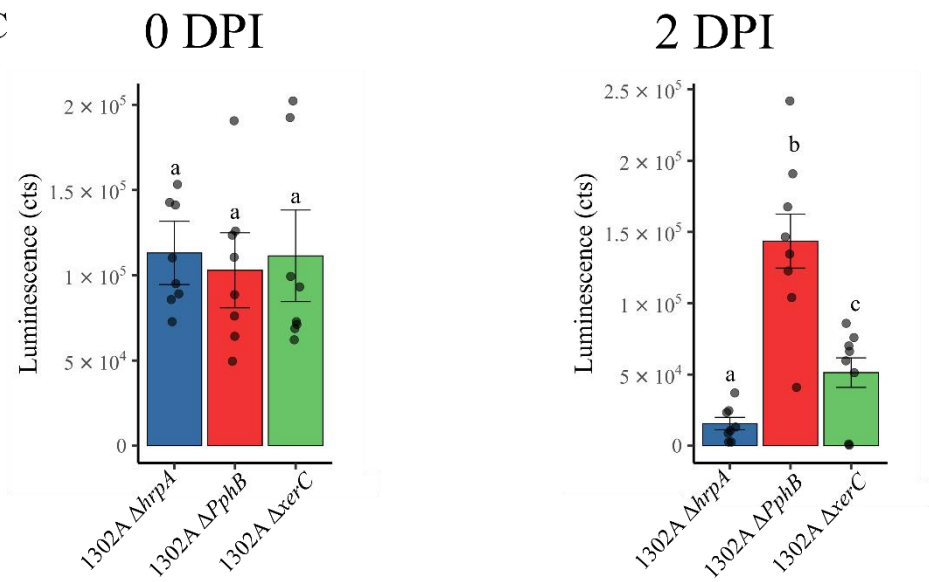

0 DPI
2 DPI
Luminescence (cts)
a
a
a
b
a
c
1302A ΔhrpA
1302A ΔPphB
1302A ΔxerC

Figure 3. Bacterial bioluminescence can be used to study the effect of plant immune responses on bacterial growth and viability. *Phaseolus vulgaris* cultivar Tendergreen leaves were syringe infiltrated with 10^6 CFU ml^{-1} of *Pseudomonas syringae* pv. *phaseolicola* 1302A (*Pph* 1302A) Δ*hrpA*, Δ*PphB* and Δ*xerC* tagged with p_{OXB13}::lux. A: After infiltration, leaves were kept in the dark for 10 minutes and then imaged with the nightOWL LB 983 at 2 minutes exposure. B: After two days post inoculation, leaves were detached from the plants, kept in the dark for 10 minutes and imaged with the nightOWL LB 983 at 5 minute exposure. 1302A Δ*PphB*: Pph 1302A Δ*PphB* p_{OXB13}::lux; 1302A Δ*xerC*: Pph1302A Δ*xerC* p_{OXB13}::lux; 1302A Δ*hrpA*: Pph 1302A Δ*hrpA* p_{OXB13}::lux; 1302A Δ*xerC* no lux: Pph 1302A Δ*xerC*. Luminescence and area values are reported for the eight biological replicates. Technical replicates (spots within the same leaf) were averaged. Error bar, +/- SE. n=8. Significant differences (Student's T-test, $p<0.05$) are indicated by letters.

Conclusion

Dual bioluminescence and fluorescence reporter constructs have previously been described for both gram-negative and gram-positive bacteria and used to study transcription, metabolic activity and host colonisation (Unge *et al.*, 1999; Unge and Jansson, 2001; Perehinec *et al.*, 2007; Benedetti *et al.*, 2012; Kim *et al.*, 2018). Here we describe the development of a flexible library of mini-Tn7 constructs that can be used to tag Gram-negative bacteria with a combination of *lux*, *lux-frp* and *lux*-eYFP operons with a wide selection of promoter strengths to drive *lux* expression. The strongest promoter used (p_{nptII}), together with a modified RBS and removal of the MCS, gives a 4 to 7 fold increase in luminescence over a previously described mini-Tn7-*lux* construct. We also show that introducing an additional FMN reductase can positively increase bioluminescence in some bacterial strains and that the relative performance of *lux* and *ilux* depends on the bacterial strain used.

We have demonstrated the application of these reporter constructs by showing that *Pph* is able to systemically colonise the leaves of *P. vulgaris.* By combining bioluminescence and fluorescence, the process of collecting samples to be imaged with confocal microscopy could be greatly accelerated. Interestingly, *Pph* did not form biofilms within the xylem vessels, where they were observed to be highly mobile. Instead *Pph* preferentially colonised collenchymatic cells. While previous studies have shown that *Pph* causes systemic symptoms in *P. vulgaris* (Zaiter and Coyne, 1984), the question of whether, and how *Pph* is able to move systemically through host plants has not been fully addressed. Having shown that *Pph* can colonise the leaf vascular system it will be of interest to elucidate the molecular mechanisms underpinning vascular colonisation by *Pph*, and to investigate the contribution of vascular colonisation to the development of systemic symptoms and to seed-borne transmission of *Pph*.

Our results also show that bacterial bioluminescence can be used to rapidly detect the effect of different plant immune responses on bacterial viability and growth, and to monitor the effectiveness of plant immune responses in restricting systemic colonisation of plant tissues. Importantly, bioluminescence is a sensitive indicator of bacterial viability and metabolic activity as well as localisation, while the relatively high stability of fluorescent proteins means that they can be used to detect both living and metabolically inactive or dead cells. Thus, it is of interest to note that while ETI and PTI could be observed to restrict the movement and population levels of *Pph* in infected tissue, we were still able to detect a substantial number of bioluminescent cells 2 days, and even 20 days after infection (Figure 3A; data not shown), showing that neither PTI nor ETI eliminated invading bacteria. Bioluminescent imaging may be of particular value in studying diseases such as bacterial leaf streak of wheat caused by *Xanthomonas*

translucens pv. *undulosa*, or bacterial blight of rice caused by *Xanthomonas oryzae* pv. *oryzae*, in which resistance is typically assessed by monitoring lesion length or area, allowing rapid comparative imaging of bacterial spread within host plants even before lesions develop (Kandel *et al.*, 2012; Oliva *et al.*, 2019; Sapkota *et al.*, 2020). We have already confirmed that the constructs described here can be used to transform and produce bioluminescence in *Xanthomonas* (data not shown).

Mini-Tn7 constructs have previously been used to transform a wide range of Proteobacteria by exploiting the presence of a conserved *att*Tn7 site downstream of the *glmS* gene, in which the effects of transposon insertion have generally been found to be relatively neutral (Choi *et al.* 2005, 2006). More recently, researchers have shown that the range of bacteria that can be tagged using this system can be further expanded through the introduction of an artificial *att*Tn7 site, which can also be used to extend the use of this system to bacteria in which insertion into the pre-existing *att*Tn7 site has been found to not be neutral (Figueroa-Cuilan *et al.*, 2016). We therefore believe that the constructs and approaches described in this study will be a valuable resource for researchers who are interested in tracking both the spatio-temporal and physiological dynamics of bacterial populations in a wide range of biological contexts.

Experimental Procedures

Bacterial strains, media and growth conditions

E. coli DH5α strains carrying the plasmids listed in Supplementary Table 1 were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 37 ºC. Single colonies were used to inoculate

10 ml LB cultures, which were incubated at 37 ºC with shaking at 200 r.p.m. *Pseudomonas* and *Acinetobacter* strains were routinely grown in the same manner, but incubated at 28°C. Antibiotics were included where relevant at the following concentrations: gentamicin (10 µg ml^{-1}), tetracycline (10 µg ml^{-1}), kanamycin (20 µg ml^{-1}), carbenicillin (50 µg ml^{-1}), nitrofurantoin (25 µg ml^{-1}).

Plant growth conditions

Phaseolus vulgaris cultivars Tendergreen (TG) and Canadian Wonder were grown at 20-22ºC, 70% humidity and with artificial light maintained for 8 h periods within the 24-h cycle. Two-week old plants (first true leaves fully developed) were used for all experiments.

Cloning, restriction enzyme digestion and polymerase chain reaction (PCR)

Cloning was performed using Gibson assembly (Gibson *et al.*, 2009) and constructs heat-shock transformed into *E. coli* strain DH5α. Transformants harboring the *lux* operon were screened for luminescence using a nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany). For constructs combining luminescence and fluorescence, transformants were screened for both luminescence and fluorescence using a nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) and a FastGene blue light LED illuminator (Geneflow Ltd, UK) respectively. PCR was conducted according to the manufacturer's instructions using Q5 high-fidelity DNA polymerase (Thermo Fisher Scientific) unless otherwise specified. Colony-PCR was performed according to the manufacturer's instructions using GoTaq Green Master Mix (Promega Corporation,

USA). DNA Sanger sequencing was performed by Source Bioscience (UK). Restriction enzyme digestions were performed according to the manufacturer's instructions (Thermo Fisher Scientific).

Generation of a low copy-number plasmid harboring *ilux* and *lux* constructs

pRSJ-p_{nptII}::ilux was derived from pIJ11282 (Frederix *et al.*, 2014) and pGEX(-) (Gregor *et al.*, 2018). Briefly, pIJ11282 was digested with PstI and SnaBI to remove the *lux* operon. *ilux* was amplified with primers OXRS_F1 and OXRS_R1 (Supplementary Table 3) from pGEX(-). The PCR product was ligated into PstI/SnaBI-digested pIJ11282 and transformed into *E. coli.* Potential clones were selected in LB agar supplemented with 10 µg ml^{-1} tetracycline. DNA sequencing was performed with primers listed in Supplementary Table 3.

pRSJ-p_{nptII}::lux-frp was generated by ligating *frp* from pGEX(-) into pIJ11282. Briefly, *frp* was amplified with primer OXRS_F2 and OXRS_R2 (Supplementary Table 3) from pGEX(-) and ligated into PstI-digested pIJ11282. Transformants were selected on LB agar supplemented with 10 µg ml^{-1} tetracycline. DNA sequencing was performed using primer OXRS_S1 (Supplementary Table 5).

Generation of a mini-Tn7 vector library for *lux* expression

pRS-$p_{OXB20(1)}$::lux was generated by ligating the OXB20 promoter (Oxford Genetics Limited, UK) synthesised by Eurofins Genomics into SmaI-digested pUC18-mini-Tn7T-Gm-lux (Choi *et al.*, 2005). To make pUC18-mini-Tn7T-Gm-lux compatible for conjugation, oriT was PCR amplified with primers OXRS_FT and OXRS_RT

(Supplementary Table 3) from pK18mobsacB (Schäfer *et al.*, 1994) and ligated into EheI-digested pUC18-mini-Tn7T-Gm-lux. Potential clones were selected on LB agar supplemented with 10 µg ml^{-1} gentamicin. DNA sequencing was performed using primer OX26 (Supplementary Table 4).

pRS-p_{nptII}::lux was generated by ligating *lux* amplified with primers OXRS_F3 and OXRS_R3 (Supplementary Table 3) from pIJ11282 into the pUC18T backbone amplified with primers OXRS_F4 and OXRS_R4 (Supplementary Table 3) from pUC18T-mini-Tn7T-Gm-dsRedExpress (Choi *et al.*, 2005). Transformants were selected on LB agar supplemented with 10 µg ml^{-1} gentamicin. DNA sequencing was performed using primers listed in Supplementary Table 6.

pRS-p_{OXB20}::lux, pRS-p_{OXB16}::lux, pRS-p_{OXB13}::lux and pRS-p_{OXB11}::lux plasmids were generated by ligating OXB20, OXB16, OXB13 and OXB11 promoters (Oxford Genetics Limited, UK) into PdiI (NaeI)/SnaBI-digested pRS-p_{nptII}::lux. Transformants were selected on LB agar supplemented with 10 µg ml^{-1} gentamicin. DNA sequencing was performed using primer OXB26 (Supplementary Table 4).

For constructs pRS-p_{OXB20}::lux and pRS-p_{nptII}::lux we additionally included *frp* downstream of the *lux* operon. Briefly, *frp* was amplified from pGEX(-) with primers OXRS_F5 and OXRS_R5 (Supplementary Table 3) and ligated into MscI-digested pRS-p_{nptII}::lux and MscI-digested pRS-p_{OXB20}::lux. DNA sequencing was performed using primer OXRS_S1 (Supplementary Table 5).

Generation of a mini-Tn7 vector library for combined bioluminescence and fluorescence expression

$p_{A1/04/03}$::eYFP amplified with OXRS_F6 and OXRS_R6 (Supplementary Table 3) from miniTn7(Gm)PA1/04/03-eyfp-a (Klausen *et al.*, 2003), was ligated into MscI-digested pRS-p_{nptII}::lux, pRS-p_{OXB20}::lux, pRS-p_{OXB16}::lux, pRS-p_{OXB13}::lux and pRS-p_{OXB11}::lux plasmids. $p_{A1/04/03}$::eYFP amplified with OXRS_F7 and OXRS_R7 (Supplementary Table 3) from miniTn7(Gm)PA1/04/03-eyfp-a (Klausen *et al.*, 2003), was ligated into MscI-digested pRS-p_{nptII}::lux-frp and pRS-p_{OXB20}::lux-frp. Transformants were selected on LB agar supplemented with 10 µg ml^{-1} gentamicin. For plasmids pRS-p_{nptII}::lux-$p_{A1/04/03}$::eYFP, pRS-p_{OXB20}::lux-$p_{A1/04/03}$::eYFP, pRS-p_{OXB16}::lux-$p_{A1/04/03}$::eYFP, pRS-p_{OXB13}::lux-$p_{A1/04/03}$::eYFP and pRS-p_{OXB11}::lux-$p_{A1/04/03}$::eYFP, DNA sequencing was performed using primer OXRS_S1 (Supplementary Table 5). For plasmids pRS-p_{nptII}::lux-frp-$p_{A1/04/03}$::eYFP and pRS-p_{OXB20}::lux-frp-$p_{A1/04/03}$::eYFP DNA sequencing was performed using primer OXRS_S2 (Supplementary Table 5).

The same cloning procedure, primers and PCR conditions used to ligate $p_{A1/04/03}$::eYFP into pRS-p_{nptII}::lux, pRS-p_{OXB20}::lux, pRS-p_{OXB16}::lux, pRS-p_{OXB13}::lux, pRS-p_{OXB11}::lux, pRS-p_{nptII}::lux-frp and pRS-p_{OXB20}::lux-frp plasmids were used to amplify p_{lac}::eYFP from pRS p_{lac}::eYFP (this study) and ligate it into MscI-digested pRS-p_{nptII}::lux, pRS-p_{OXB20}::lux, pRS-p_{OXB16}::lux, pRS-p_{OXB13}::lux, pRS-p_{OXB11}::lux, pRS-p_{nptII}::lux-frp and pRS-p_{OXB20}::lux-frp.

pRS-$p_{OXB20(1)}$::lux-$p_{A1/04/03}$::eYFP was generated by ligating $p_{A1/04/03}$::eYFP amplified with primers OXRS_F8 and OXRS_R8 from miniTn7(Gm)PA1/04/03-eyfp-a (Klausen *et al.*, 2003) into StuI-digested pRS-$p_{OXB20(1)}$::lux. Potential clones were selected on LB agar supplemented with 10 µg ml^{-1} gentamicin.

Construction of *Pph* 1302A *PphB* and *hrpA* knock-out mutants

To generate the *Pph* 1302A *avrPphB* knock-out mutant (*Pph* 1302A *PphB*-), 1 kb genomic regions flanking *PphB* were PCR amplified with primers OXRS_F9 and OXRS_R9 (left end) and primers OXRS_F10 and OXRS_R10 (right end) (Supplementary Table 3). Primer OXRS_R9 was designed to include the starting codon of *avrPphB* while primer OXRS_R10 included the stop codon of *avrPphB*. PCR products were ligated into pK18mobsacB (Schäfer *et al.*, 1994) amplified with primers OXRS_F11 and OXRS_R11 (Supplementary Table 3). Transformants were selected in LB agar supplemented with 20 µg ml^{-1} kanamycin. Colony-PCR was performed with primers OXRS_F12 and OXRS_R12 (Supplementary Table 3) to validate the assembly. The plasmid was conjugated into *Pph* 1302A, as described below, and transformants were plated onto LB plates supplemented with kanamycin and nitrofurantoin. Transformants were the result of plasmid integration (a single recombination event). To allow allelic exchange, four colonies were cultured overnight in LB and 20 µl plated onto LB (1.5% agar) supplemented with 10% sucrose. Colony PCR with primers OXRS_F9 and OXRS_R10 (Supplementary Table 3) was used to confirm the deletion of *avrPphB* from the *Pph* 1302A genome.

A similar procedure was used to generate *Pph* 1302A *ΔhrpA*. Briefly, *hrpA* upstream and downstream regions were amplified from the *Pph* 1302A genome with primers OXRS_F13, OXRS_R13 and OXRS_F14, OXRS_R14 (Supplementary Table 3) and ligated into pK18mobsacB amplified with primers OXRS_F11 and OXRS_R11 (Supplementary Table 3). Conjugation and selections of knock-out mutants were carried out as described for *Pph* 1302A *ΔPphB*.

Mini-Tn7 delivery by four-parental mating conjugation

Pph 1302A strains were tagged by four-parental mating conjugation (Choi and Schweizer, 2006) with minor modifications. Details of the protocol used can be found in Supplementary File 1. *Pph* 1302A transformants were screened for luminescence (if tagged with *lux* constructs) and luminescence and fluorescence (if tagged with *lux* and eYFP constructs) as reported for *E. coli.* The correct insertion of the mini-Tn7 delivery construct was verified by colony-PCR using primers OXRS_F15 and OXRS_R15 (Supplementary Table 3). Primer OXRS_F15 anneals in the *glmS* gene of *Pph* and primer OXRS_R15 binds in the Tn7R (right end of the Tn7 transposon).

Three-parental mating conjugation

Pseudomonas and *Acinetobacter* strains (Supplementary Table 2) were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures, which were incubated at 28 ºC with shaking at 200 rpm. Bacterial cultures were then used for three-parental mating conjugation as described for four-parental mating conjugation (Supplementary File 1), using the helper plasmid pRK2013 (Knauf and Nester, 1982).

Plate reader assay

Bacterial cultures were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures, which were incubated at 28 ºC with shaking at 200 rpm. After 24 hours, 500 µl pre-culture was used to set up an overnight 10 ml LB culture. Cultures were centrifuged at 4000 g for 6 minutes and resuspended in 10 ml 10

mM $MgCl_2$. 15 µl of resuspended bacterial cultures were added to 135 µl King's B medium (KB) (King *et al.*, 1954) and M9 medium (20% glucose) (CSH protocols) placed into a 96-well assay black plate (Corning incorporated, USA). Luminescence and OD_{600} were measured using a Infinite M200 plate reader (Tecan Group Ltd, Switzerland) over 20 hours. The temperature was set at 28 ºC and measurements were taken every 30 minutes.

NADP+ assay

Bacterial cultures were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) supplemented with 10 µg ml^{-1} tetracycline and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures supplemented with 10 µg ml^{-1} tetracycline, which were incubated at 28 ºC with shaking at 200 rpm. After 24 hours, 200 µl pre-culture was used to set up 200 ml LB cultures supplemented with 10 µg ml^{-1} tetracycline which were incubated at 28 ºC with shaking at 200 rpm for 6 hours. Cultures were processed according to the manufacturer's instructions (NADP/NADPH assay ab176724, abcam, UK) at a final OD_{600} of 3,3.

Determining the effect of bacterial bioluminescence expression on bacterial growth *in planta*

Pph strains were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures, which were incubated at 28 ºC with shaking at 200 rpm. After 24 hours, 500 µl pre-culture was used to set up an overnight 10 ml LB culture. Cultures were centrifuged at 4000 g for 6 minutes and resuspended in 10 ml 10

mM $MgCl_2$ ($5x10^6$ CFU ml^{-1}). Four plants per bacterial culture were used for the experiment. Briefly, the first two fully developed leaves of 12 two-weeks old plants (*P. vulgaris* cultivar Canadian Wonder) were syringe-infiltrated with *Pph 1302A* p_{nptII}::lux $p_{A1/04/03}$, *Pph 1302A* p_{OXB16}::lux $p_{A1/04/03}$, *Pph 1302A* p_{OXB11}::lux $p_{A1/04/03}$ and *Pph 1302A* ($5x10^6$ CFU ml^{-1}) (Supplementary Table 2). Each leaf was infiltrated four times in four small areas (50-80 mm^2). The four spots within the same leaf were considered technical replicates and were averaged for the calculation of the CFU mm^{-2}. At 2 DPI (Days Post Inoculation) four leaf disks per leaf (four leaves total per bacterial culture) were detached from the four inoculated spots with a 1 cm^2 cork and pooled together. Samples were ground with a TissueLyser (Qiagen) in 1 ml 10 mM $MgCl_2$, and serial dilutions (20 µl drops) were spotted onto LB agar supplemented with nitrofurantoin 25 µg ml^{-1}. Plates were incubated at 28 ºC for three days. At 5 DPI, the same process described above was repeated.

Determining the detection limit of *Pph* RJ3 p_{nptII}::lux $p_{A1/04/03}$::eYFP

Bacterial cultures were prepared as described above. Serial dilutions (100/1000 µl) were performed in 10 mM $MgCl_2$ and 100 µl aliquots were spotted onto LB (1.5% agar) square plates. Plates were incubated at 28 ºC for one hour before imaging with the nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) at 20 minutes exposure. Luminescence values and area (mm^2) of each spot were measured for the 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4} dilutions. Plates were then incubated at 28 ºC for three days to allow bacterial growth and colony count to be performed.

Combined luminescence-fluorescence approach to monitor *Pph* dispersal *in planta*

Syringe-infiltration

Pph RJ3 p_{nptII}::lux $p_{A1/04/03}$::eYFP (Supplementary Table 1) was inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures, which were incubated at 28 ºC with shaking at 200 rpm. After 24 hours, 500 µl pre-culture was used to set up an overnight 10 ml LB culture. Cultures were centrifuged at 4000 g for 6 minutes and resuspended in 20 ml 10 mM $MgCl_2$. The bacterial suspension ($5x10^6$ CFU ml^{-1}) was syringe-infiltrated into abaxial areas of the first true leaves of TG plants without causing any damage to the leaf mesophyll. To minimise damage to the leaf while infiltrating, we modified a 10 ml syringe by attaching to it a P20 pipette tip. We removed the last 1 cm of the P20 pipette tip and attached to it a rubber O-ring to ensure a soft contact between the tip and the leaf mesophyll. Plants were incubated for 0,1,3 and 5 days at 22 ºC and artificial light was maintained for 16 h within the 24-h cycle.

At designated times (0,1,3,5 DPI), leaves were detached and placed on square Petri dishes. Leaf petioles were wrapped in water-soaked cotton disks to allow plant transpiration. Leaves were incubated in the dark for 10 minutes before imaging with the nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) at 10 minutes exposure. Bacterial bioluminescence was used to visualize *Pph* localisation and to identify portions of leaves to be imaged with confocal laser scanning microscopy. Leaf vasculatures showing *Pph* colonization were cross-sectioned by hand-sectioning (50-80 µm) without embedding. Samples were imaged with a confocal laser scanning

microscope Zeiss LSM 880 (Carl Zeiss, Germany). eYFP was excited at 514 nm (5.50%) and detected in the range 520-610 nm; Phenolic compounds were excited at 405 nm (1.10%) and detected in the range 200-450 nm to visualize xylem structure and collenchyma. Confocal stacks were acquired with a Z-step of 50–80 µm using the objectives C-Apochromat 40x/1.2 W Korr FCS M27.

Spray-inoculation

Pph RJ3 p_{nptII}::lux $p_{A1/04/03}$::eYFP (Supplementary Table 1) was incubated prior to inoculation as described above. The bacterial suspension ($5x10^7$ CFU ml^{-1}) was deposited on restricted abaxial areas of the first true leaves of TG plants by spraying the resuspended bacterial culture through a round aperture (about 1 cm^2) cut into a folded aluminum foil. When spraying, close contact between leaves and aluminum foil ensured that there was no contamination of adjacent areas of the leaf. Four inoculated plants were placed inside a closed transparent box (30x20x30 cm) with two Oasis flower foams (Oasis Floral Products, UK) previously soaked in water. Plants were incubated for 3 days at 22 ºC and artificial light was maintained for 16 h within the 24-h cycle. Negative controls were represented by leaves inoculated solely with 10 mM $MgCl_2$.

After three days, leaves were detached and placed on square Petri dishes. Leaf petioles were wrapped in water-soaked cotton disks to allow plant transpiration. Leaves were incubated in the dark for 10 minutes before imaging with the nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) at 20 minutes exposure. Bacterial bioluminescence was used to visualize *Pph* localisation and to identify portions of leaves to be imaged with confocal laser scanning microscopy. Leaf vasculatures showing *Pph* colonization were cross-sectioned by hand-sectioning (50-80 µm) or imaged as part of

leaf disks (50 mm^2). Samples were imaged with a confocal laser scanning microscope Zeiss LSM 880 (Carl Zeiss, Germany).

eYFP was excited at 514 nm (5.50%) and detected in the range 520-610 nm; Phenolic compounds were excited at 405 nm (1.10%) and detected in the range 200-450 nm to visualize xylem structure; Chlorophyll was excited at 514 nm (5.50%) and detected in the range 650-720 nm.

Confocal stacks were acquired with a Z-step of 50–80 µm (for cross-sections) or 10-30 µm (for leaf disks) using the objectives LD LCI Plan-Apochromat 25x/0.8 Imm Korr DIC M27. Leaf volume-rendering and three-dimensional models of the confocal stacks were created with the software Imaris 8 (Bitplane AG, Zürich, Switzerland).

Bacterial bioluminescence assays to detect the effect of plant immune responses on bacterial growth and viability

Pph 1302A Δ*hrpA* (this study), *Pph* 1302A Δ*PphB* (this study) and *Pph* 1302A Δ*xerC* (Lovell *et al.*, 2009) were tagged using pRS-p_{OXB13}::lux (Table 1; Supplementary Table 1) as described above. Bacterial cultures were inoculated from frozen glycerol stock onto Luria Bertani (LB) (Sambrook *et al.*, 1989) agar (1.5% agar) and grown for 24 hours at 28 ºC. Single colonies were used to inoculate 10 ml LB cultures, which were incubated at 28 ºC with shaking at 200 rpm. After 24 hours, 500 µl pre-culture was used to set up an overnight 10 ml LB culture. Cultures were centrifuged at 4000 g for 6 minutes and resuspended in 10 ml 10 mM $MgCl_2$ (10^6 CFU ml^{-1}). Eight plants per bacterial culture were used for the experiment.

The optimal number of replicates were calculated with a jacknife resampling (Hanna, 1989) derived approach before performing the experiment. Briefly, the first two fully

developed leaves of 12 two-weeks old plants (*P. vulgaris* cultivar TG) were syringe-infiltrated with *Pph* 1302A Δ*PphB* p_{OXB13}::lux (10^6 CFU ml^{-1}) (Supplementary Table 1). Each leaf was infiltrated six times in six small areas (50-80 mm^2). The six spots within the same leaf were considered technical replicates and were averaged for the calculation of the optimal sample size. Leaves were detached and incubated in the dark for 10 minutes before imaging with the nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) at 2-minute exposure. Luminescence values were recorded and used for the calculation of the optimal sample size. A custom R script (Supplementary File 2) was used to randomly sample 1000 thousand times the generated data (12 values, in this case) with a jackknife approach. Standard deviation (SD) for each 1000 resampling per each sample size (from 12 to 1) was used to assess the difference in SD between sample sizes (12-n), where n is the sample size. The process was repeated 300 times and statistically significance differences of differences in SD between sample sizes were assessed with Benjamin-Hochberg (BH) post-hoc test. Eight was assessed to be the optimal sample size as the reduction in SD from sample size 8 to sample size 9 was not statistically significant.

Subsequently, two-week old TG plants were syringe-infiltrated with p_{OXB13}::*lux* expressing *Pph* 1302A Δ*hrpA*, *Pph* 1302A Δ*PphB* and *Pph* 1302A Δ*xerC* bacterial cultures (Supplementary Table 2). The experiment was conducted according to a Randomized Complete Block Design (RCBD) (Liu and Berger, 2014) in which the block corresponded to the inoculation time. This was done to account for the passage of time from the first plant inoculations to the last plant inoculations (about 4 hours). At time 1, three plants (two leaves per plant) were syringe-infiltrated with the corresponding bacterial cultures. One leaf per plant was detached and incubated in the dark for 10

minutes before imaging with the nightOWL LB 983 *in vivo* imaging system (Berthold Technologies, Germany) at 2-minute exposure. Luminescence and area of each spot within leaves were recorded. Plants were then incubated for 2 days at 22 ºC and artificial light was maintained for 16 h within the 24-h cycle. The same procedure was repeated for subsequent time points. Within each block (time point and tray) plants were randomized. The same bacterial cultures were used for all time points. Negative controls were represented by plants infiltrated with 10 mM $MgCl_2$ and infiltrated with the non-tagged *Pph* 1302A Δ*xerC* bacterial culture to detect any bioluminescence associated with induction of plant immune responses (Bennett *et al.*, 2005).

After 2 DPI, luminescence and the area where luminescence is detected were measured as described previously, for the remaining leaves. Exposure time was set at 5 minutes. Immediately after measuring bioluminescence signal, leaves were used to perform colony counts. Briefly, six leaf disks per leaf were detached from the six inoculated spots with a 1-cm^2 cork and pooled together. Samples were ground with a TissueLyser (Qiagen) in 1 ml 10 mM $MgCl_2$, and serial dilutions (20 µl drops) were spotted onto LB agar supplemented with gentamicin 10 µg ml^{-1} and nitrofurantoin 25 µg ml^{-1}. Plates were incubated at 28 ºC for three days.

Data analysis and statistics

Processing of data was performed using R (R Core Team, 2017) version 3.5.3 and ggplot2 (Wickham, 2016). Requirements of linear models, namely normal distribution of the residuals and homogeneity of variance were assessed using diagnostic plots (Harrison *et al.*, 2018). When diagnostic plots were not clearly interpretable, we used Leven's test to

assess the homogeneity of variance. Independence of the observations was guaranteed by the experimental design.

References

Asai, S., Takamura, K., Suzuki, H., and Setou, M. (2008) Single-cell imaging of c-fos expression in rat primary hippocampal cells using a luminescence microscope. *Neuroscience Letters* 434: 289–292.

Bae, C., Han, S.W., Song, Y.-R., Kim, B.-Y., Lee, H.-J., Lee, J.-M., et al. (2015) Infection processes of xylem-colonizing pathogenic bacteria: possible explanations for the scarcity of qualitative disease resistance genes against them in crops. *Theor Appl Genet* 128: 1219–1229.

Benedetti, I.M., Lorenzo, V. de, and Silva-Rocha, R. (2012) Quantitative, Non-Disruptive Monitoring of Transcription in Single Cells with a Broad-Host Range GFP-luxCDABE Dual Reporter System. *PLOS ONE* 7: e52000.

Bennett, M., Mehta, M., and Grant, M. (2005) Biophoton Imaging: A Nondestructive Method for Assaying R Gene Responses. *MPMI* 18: 95–102.

Bogs, J., Bruchmüller, I., Erbar, C., and Geider, K. (1998) Colonization of Host Plants by the Fire Blight Pathogen *Erwinia amylovora* Marked with Genes for Bioluminescence and Fluorescence. *Phytopathology*^{TM} 88: 416–421.

Brodl, E., Winkler, A., and Macheroux, P. (2018) Molecular Mechanisms of Bacterial Bioluminescence. *Computational and Structural Biotechnology Journal* 16: 551–564.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F., et al. (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 488: 91–95.

Burkatovskaya, M., Tegos, G.P., Swietlik, E., Demidova, T.N., P Castano, A., and Hamblin, M.R. (2006) Use of chitosan bandage to prevent fatal infections developing from highly contaminated wounds in mice. *Biomaterials* 27: 4157–4164.

Cerutti, A., Jauneau, A., Auriac, M.-C., Lauber, E., Martinez, Y., Chiarenza, S., et al. (2017) Immunity at Cauliflower Hydathodes Controls Systemic Infection by *Xanthomonas campestris* pv *campestris*. *Plant Physiology* 174: 700–716.

Choi, K.-H., Gaynor, J.B., White, K.G., Lopez, C., Bosio, C.M., Karkhoff-Schweizer, R.R., and Schweizer, H.P. (2005) A Tn 7 -based broad-range bacterial cloning and expression system. *Nature Methods* 2: 443.

Choi, K.-H. and Schweizer, H.P. (2006) mini-Tn 7 insertion in bacteria with single att Tn 7 sites: example *Pseudomonas aeruginosa*. *Nature Protocols* 1: 153–161.

Compant, S., Reiter, B., Sessitsch, A., Nowak, J., Clément, C., and Ait Barka, E. (2005) Endophytic colonization of Vitis vinifera L. by plant growth-promoting bacterium *Burkholderia* sp. strain PsJN. *Appl Environ Microbiol* 71: 1685–1693.

Craig, N.L. (1991) Tn7: a target site-specific transposon. *Molecular Microbiology* 5: 2569–2573.

Donati, I., Cellini, A., Buriani, G., Mauri, S., Kay, C., Tacconi, G., and Spinelli, F. (2018) Pathways of flower infection and pollen-mediated dispersion of *Pseudomonas syringae* pv. *actinidiae*, the causal agent of kiwifruit bacterial canker. *Hortic Res* 5:.

Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., et al. (2015) Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host & Microbe* 18: 478–488.

Fan, B., Chen, X.H., Budiharjo, A., Bleiss, W., Vater, J., and Borriss, R. (2011) Efficient colonization of plant roots by the plant growth promoting bacterium *Bacillus amyloliquefaciens* FZB42, engineered to express green fluorescent protein. *Journal of Biotechnology* 151: 303–311.

Fan, J., Crooks, C., and Lamb, C. (2008) High-throughput quantitative luminescence assay of the growth in planta of *Pseudomonas syringae* chromosomally tagged with *Photorhabdus luminescens* luxCDABE. *The Plant Journal* 53: 393–399.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H.M. (2014) Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular Systems Biology* 10: 731.

Fierer, N. and Ladau, J. (2012) Predicting microbial distributions in space and time. *Nature Methods* 9: 549–551.

Figueroa-Cuilan, W., Daniel, J.J., Howell, M., Sulaiman, A., and Brown, P.J.B. (2016) Mini-Tn7 Insertion in an Artificial attTn7 Site Enables Depletion of the Essential Master Regulator CtrA in the Phytopathogen *Agrobacterium tumefaciens*. *Applied and Environmental Microbiology* 82: 5015.

Fleiss, A. and Sarkisyan, K.S. (2019) A brief review of bioluminescent systems (2019). *Current Genetics*.

Frederix, M., Edwards, A., Swiderska, A., Stanger, A., Karunakaran, R., Williams, A., et al. (2014) Mutation of praR in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Molecular Microbiology* 93: 464–478.

Fujie, M., Takamoto, H., Kawasaki, T., Fujiwara, A., and Yamada, T. (2010) Monitoring growth and movement of *Ralstonia solanacearum* cells harboring

plasmid pRSS12 derived from bacteriophage ϕRSS1. *Journal of Bioscience and Bioengineering* 109: 153–158.

Fukui, R., Fukui, H., McElhaney, R., Nelson, S.C., and Alvarez, A.M. (1996) Relationship between Symptom Development and Actual Sites of Infection in Leaves of Anthurium Inoculated with a Bioluminescent Strain of *Xanthomonas campestris* pv. *dieffenbachiae*. *Applied and environmental microbiology* 62: 1021–1028.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6: 343–345.

Giliberti, G., Baccigalupi, L., Cordone, A., Ricca, E., and De Felice, M. (2006) Transcriptional analysis of the recA gene of *Streptococcus thermophilus*. *Microbial Cell Factories* 5: 29.

Glassing, A. and Lewis, T.A. (2015) An improved Tn7-lux reporter for broad host range, chromosomally-integrated promoter fusions in Gram-negative bacteria. *Journal of Microbiological Methods* 118: 75–77.

Godfrey, S. a. C., Mansfield, J.W., Corry, D.S., Lovell, H.C., Jackson, R.W., and Arnold, D.L. (2010) Confocal imaging of *Pseudomonas syringae* pv. *phaseolicola* colony development in bean reveals reduced multiplication of strains containing the genomic island PPHGI-1. *Mol Plant Microbe Interact* 23: 1294–1302.

Gregor, C., Gwosch, K.C., Sahl, S.J., and Hell, S.W. (2018) Strongly enhanced bacterial bioluminescence with the ilux operon for single-cell imaging. *Proc Natl Acad Sci USA* 115: 962–967.

Hanna, S.R. (1989) Confidence limits for air quality model evaluations, as estimated by bootstrap and jackknife resampling methods. *Atmospheric Environment (1967)* 23: 1385–1398.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., et al. (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:.

Huang, W.E., Huang, L., Preston, G.M., Naylor, M., Carr, J.P., Li, Y., et al. (2006) Quantitative in situ assay of salicylic acid in tobacco leaves using a genetically modified biosensor strain of *Acinetobacter* sp. ADP1. *The Plant Journal* 46: 1073–1083.

Iwano, S., Sugiyama, M., Hama, H., Watakabe, A., Hasegawa, N., Kuchimaru, T., et al. (2018) Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* 359: 935–939.

Jackson, R.W., Mansfield, J.W., Arnold, D.L., Sesma, A., Paynter, C.D., Murillo, J., et al. (2000) Excision from tRNA genes of a large chromosomal region, carrying avrPphB, associated with race change in the bean pathogen, *Pseudomonas syringae* pv. *phaseolicola*. *Mol Microbiol* 38: 186–197.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature* 444: 323.

Kandel, Y.R., Glover, K.D., Tande, C.A., and Osborne, L.E. (2012) Evaluation of Spring Wheat Germplasm for Resistance to Bacterial Leaf Streak Caused by *Xanthomonas campestris* pv. *translucens*. *Plant Disease* 96: 1743–1748.

Kim, Y.-H., Park, P.-G., Seo, S.-H., Hong, K.-J., and Youn, H. (2018) Development of dual reporter imaging system for *Francisella tularensis* to monitor the spatio-temporal pathogenesis and vaccine efficacy. *Clin Exp Vaccine Res* 7: 129–138.

King, E.O., Ward, M.K., and Raney, D.E. (1954) Two simple media for the demonstration of pyocyanin and fluorescin. *J Lab Clin Med* 44: 301–307.

Klausen, M., Heydorn, A., Ragas, P., Lambertsen, L., Aaes-Jørgensen, A., Molin, S., and Tolker-Nielsen, T. (2003) Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Molecular Microbiology* 48: 1511–1524.

Knauf, V.C. and Nester, E.W. (1982) Wide host range cloning vectors: A cosmid clone bank of an *Agrobacterium* Ti plasmid. *Plasmid* 8: 45–54.

Koch, B., Jensen, L.E., and Nybroe, O. (2001) A panel of Tn7-based vectors for insertion of the gfp marker gene or for delivery of cloned DNA into Gram-negative bacteria at a neutral chromosomal site. *J Microbiol Methods* 45: 187–195.

Lambertsen, L., Sternberg, C., and Molin, S. (2004) Mini-Tn7 transposons for site-specific tagging of bacteria with fluorescent proteins. *Environmental Microbiology* 6: 726–732.

Lamichhane, J.R., Messéan, A., and Morris, C.E. (2015) Insights into epidemiology and control of diseases of annual plants caused by the *Pseudomonas syringae* species complex. *J Gen Plant Pathol* 81: 331–350.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* 500: 541–546.

Liu, L. and Berger, V.W. (2014) Randomized Block Design: Nonparametric Analyses. In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society.

Lovell, H.C., Mansfield, J.W., Godfrey, S.A.C., Jackson, R.W., Hancock, J.T., and Arnold, D.L. (2009) Bacterial Evolution by Genomic Island Transfer Occurs via DNA Transformation In Planta. *Current Biology* 19: 1586–1590.

Lynn, G.M., Laga, R., Darrah, P.A., Ishizuka, A.S., Balaci, A.J., Dulcey, A.E., et al. (2015) In vivo characterization of the physicochemical properties of polymer-linked TLR agonists that enhance vaccine immunogenicity. *Nature Biotechnology* 33: 1201–1210.

Mitchell, R.E. and Bieleski, R.L. (1977) Involvement of Phaseolotoxin in Halo Blight of Beans: Transport and Conversion to Functional Toxin. *Plant Physiology* 60: 723–729.

Mitter, B., Pfaffenbichler, N., Flavell, R., Compant, S., Antonielli, L., Petric, A., et al. (2017) A New Approach to Modify Plant Microbiomes and Traits by Introducing Beneficial Bacteria at Flowering into Progeny Seeds. *Front Microbiol* 8:.

Monteiro, F., Genin, S., van Dijk, I., and Valls, M. (2012) A luminescent reporter evidences active expression of *Ralstonia solanacearum* type III secretion system genes throughout plant infection. *Microbiology (Reading, Engl)* 158: 2107–2116.

Mortin, L.I., Li, T., Praagh, A.D.G.V., Zhang, S., Zhang, X.-X., and Alder, J.D. (2007) Rapid Bactericidal Activity of Daptomycin against Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* Peritonitis in Mice as Measured with Bioluminescent Bacteria. *Antimicrobial Agents and Chemotherapy* 51: 1787–1794.

Ng, C.Y., Farasat, I., Maranas, C.D., and Salis, H.M. (2015) Rational design of a synthetic Entner–Doudoroff pathway for improved and controllable NADPH regeneration. *Metabolic Engineering* 29: 86–96.

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J.C., Perez-Quintero, A., Li, T., et al. (2019) Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nature Biotechnology* 37: 1344–1350.

Perehinec, T.M., Qazi, S.N., Gaddipati, S.R., Salisbury, V., Rees, C.E., and Hill, P.J. (2007) Construction and evaluation of multisite recombinatorial (Gateway) cloning vectors for Gram-positive bacteria. *BMC Molecular Biology* 8: 80.

Planas-Marquès, M., Kressin, J.P., Kashyap, A., Panthee, D.R., Louws, F.J., Coll, N.S., and Valls, M. (2020) Four bottlenecks restrict colonization and invasion by the pathogen *Ralstonia solanacearum* in resistant tomato. *J Exp Bot* 71: 2157–2171.

Presnell, K.V., Flexer-Harrison, M., and Alper, H.S. (2019) Design and synthesis of synthetic UP elements for modulation of gene expression in *Escherichia coli*. *Synth Syst Biotechnol* 4: 99–106.

Propheter, D.C. and Hooper, L.V. (2015) Bacteria Come into Focus: New Tools for Visualizing the Microbiota. *Cell Host & Microbe* 18: 392–394.

Prosser, J.I. (2010) Replicate or lie. *Environmental Microbiology* 12: 1806–1810.

R Core Team (2017) R: A language and environment for statistical computing.

Reynaud, E.G., Peychl, J., Huisken, J., and Tomancak, P. (2015) Guide to light-sheet microscopy for adventurous biologists. *Nature Methods* 12: 30–34.

Rico, A., Bennett, M.H., Forcat, S., Huang, W.E., and Preston, G.M. (2010) Agroinfiltration reduces ABA levels and suppresses *Pseudomonas syringae*-elicited salicylic acid production in *Nicotiana tabacum*. *PLoS ONE* 5: e8977.

Rodríguez-Moreno, L., Jiménez, A.J., and Ramos, C. (2009) Endopathogenic lifestyle of *Pseudomonas savastanoi* pv. *savastanoi* in olive knots. *Microbial Biotechnology* 2: 476–488.

Rybakova, D., Mancinelli, R., Wikström, M., Birch-Jensen, A.-S., Postma, J., Ehlers, R.-U., et al. (2017) The structure of the *Brassica napus* seed microbiome is cultivar-dependent and affects the interactions of symbionts and pathogens. *Microbiome* 5: 104.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. *Molecular cloning: a laboratory manual*.

Sapkota, S., Mergoum, M., and Liu, Z. (2020) The translucens group of *Xanthomonas translucens*: Complicated and important pathogens causing bacterial leaf streak on cereals. *Molecular Plant Pathology* 21: 291.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* 145: 69–73.

Seleem, M.N., Ali, M., Boyle, S.M., and Sriranganathan, N. (2008) Reporter genes for real-time in vivo monitoring of *Ochrobactrum anthropi* infection. *FEMS Microbiol Lett* 286: 124–129.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005) A guide to choosing fluorescent proteins. *Nature Methods* 2: 905–909.

Soldan, R., Mapelli, F., Crotti, E., Schnell, S., Daffonchio, D., Marasco, R., et al. (2019) Bacterial endophytes of mangrove propagules elicit early establishment of the

natural host and promote growth of cereal crops under salt stress. *Microbiological Research* 223–225: 33–43.

Taylor, J.D., Dudley, C.L., and Gray), L.P. (née (1979) Studies of halo-blight seed infection and disease transmission in dwarf beans. *Annals of Applied Biology* 93: 267–277.

Turnbaugh, P.J., Bäckhed, F., Fulton, L., and Gordon, J.I. (2008) Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe* 3: 213–223.

Unge, A. and Jansson, J. (2001) Monitoring population size, activity, and distribution of gfp-luxAB-tagged *Pseudomonas fluorescens* SBW25 during colonization of wheat. *Microb Ecol* 41: 290–300.

Unge, A., Tombolini, R., Mølbak, L., and Jansson, J.K. (1999) Simultaneous Monitoring of Cell Number and Metabolic Activity of Specific Bacterial Populations with a Dualgfp-luxAB Marker System. *Appl Environ Microbiol* 65: 813–821.

Wang, K., Kang, L., Anand, A., Lazarovits, G., and Mysore, K.S. (2007) Monitoring in planta bacterial infection at both cellular and whole-plant levels using the green fluorescent protein variant GFPuv. *New Phytologist* 174: 212–223.

Weisemann, J.M. and Weinstock, G.M. (1991) The promoter of the recA gene of *Escherichia coli*. *Biochimie* 73: 457–470.

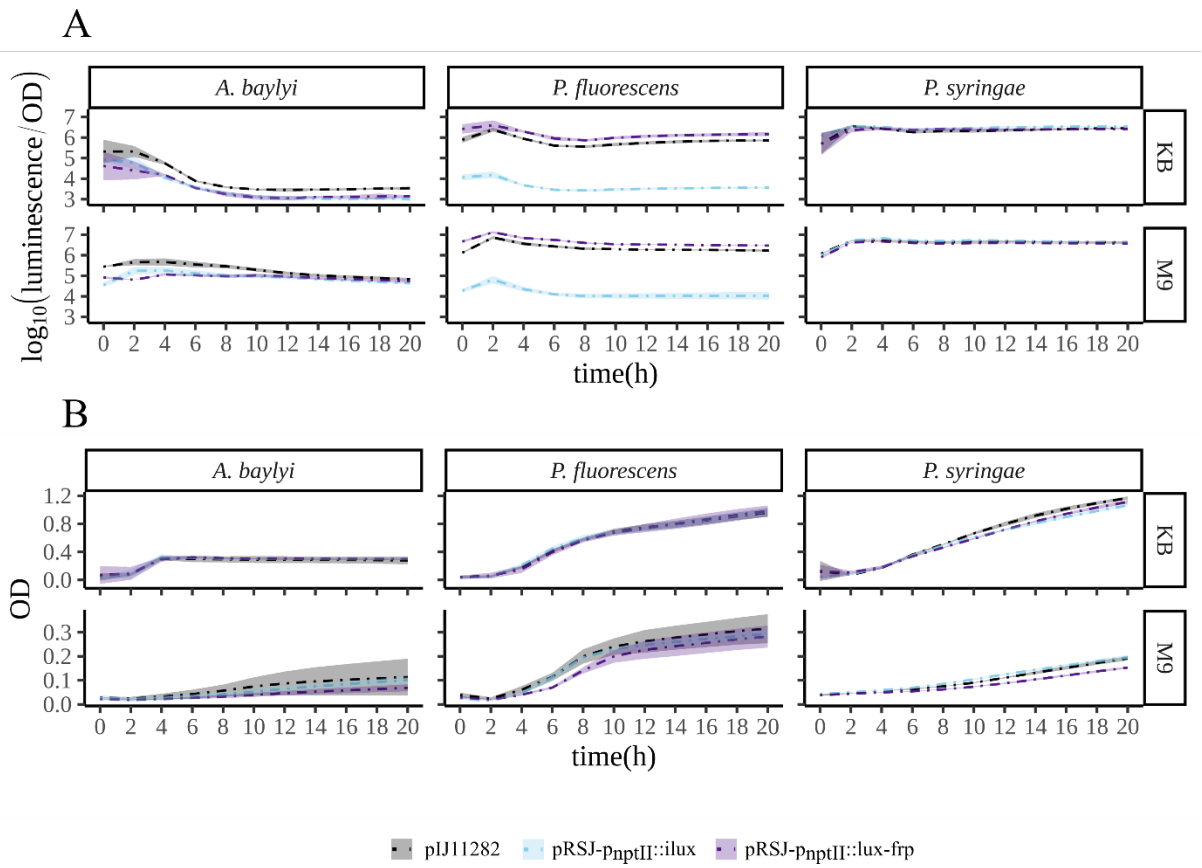
Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis.

Wright, C.A. and Beattie, G.A. (2004) Bacterial Species Specificity in proU Osmoinducibility and nptII and lacZ Expression. *MMB* 8: 201–208.

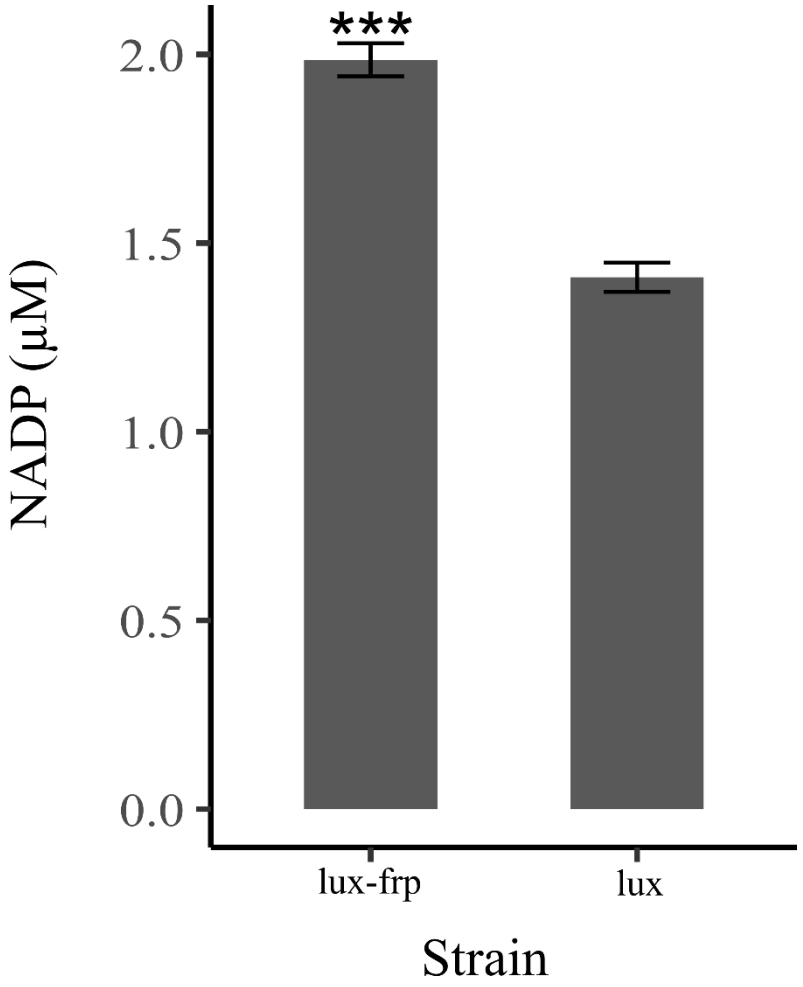
Xu, X., Miller, S.A., Baysal-Gurel, F., Gartemann, K.-H., Eichenlaub, R., and Rajashekara, G. (2010) Bioluminescence Imaging of *Clavibacter michiganensis* subsp. *michiganensis* Infection of Tomato Seeds and Plants. *Applied and Environmental Microbiology* 76: 3978–3988.

Yu, Y.A., Shabahang, S., Timiryasova, T.M., Zhang, Q., Beltz, R., Gentschev, I., et al. (2004) Visualization of tumors and metastases in live animals with bacteria and vaccinia virus encoding light-emitting proteins. *Nat Biotechnol* 22: 313–320.

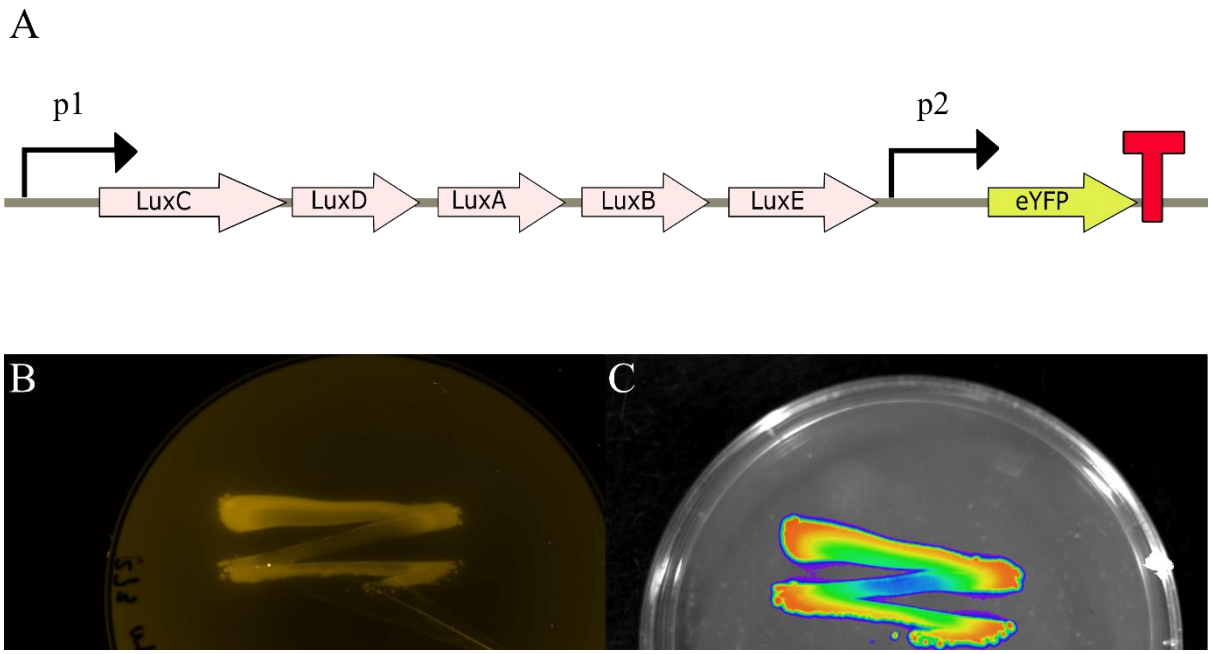
Supplementary Figures



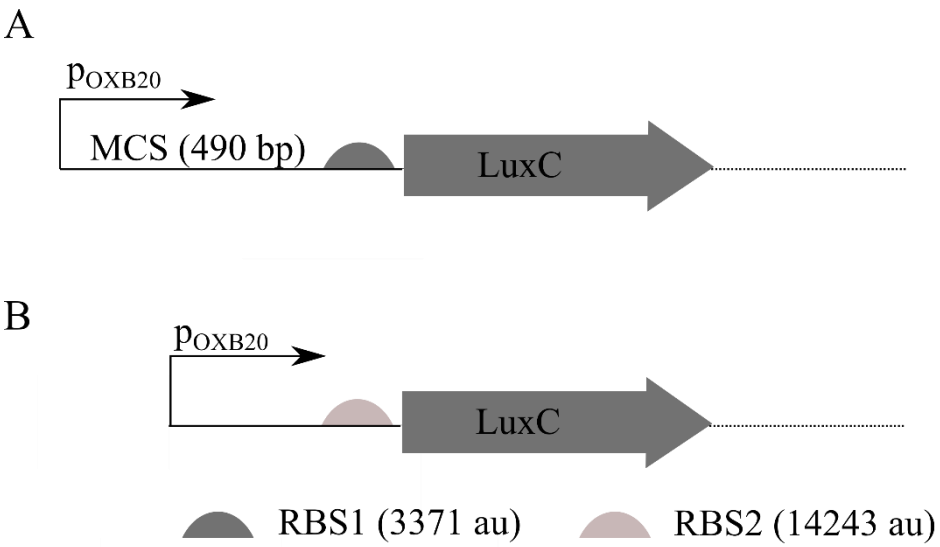
Supplementary Figure 1. Expression of FMN reductase (*frp*) increases luminescence in *P. fluorescens* NZ011 when combined with the *lux* operon. *Acinetobacter* strains: *A. baylyi* ADP1 (pIJ11282), *A. baylyi* ADP1 (pRSJ-p_{nptII}::ilux), *A. baylyi* ADP1 (pRSJ-p_{nptII}::lux-frp). *Pseudomonas fluorescens* strains: *P. fluorescens* NZ011 (pIJ11282), *P. fluorescens* NZ011 (pRSJ-p_{nptII}::ilux), *P. fluorescens* NZ011 (pRSJ-p_{nptII}::lux-frp). *Pseudomonas syringae* pv. *phaseolicola* (*Pph*) strains: *Pph* 1302A (pIJ11282), *Pph* 1302A (pRSJ-p_{nptII}::ilux), *Pph* 1302A (pRSJ-p_{nptII}::lux-frp). Plasmids pIJ11282, pRSJ-p_{nptII}::ilux and pRSJ-p_{nptII}::lux-frp have the same backbone (pIJ11282) and promoter (p_{nptII}). A: Normalised luminescence of reporter strains in KB (King's B medium) and M9 (M9 minimal medium). B: Growth of reporter strains in KB and M9. OD = optical density at 600 nm. Error bar, +/- SE. n=3.



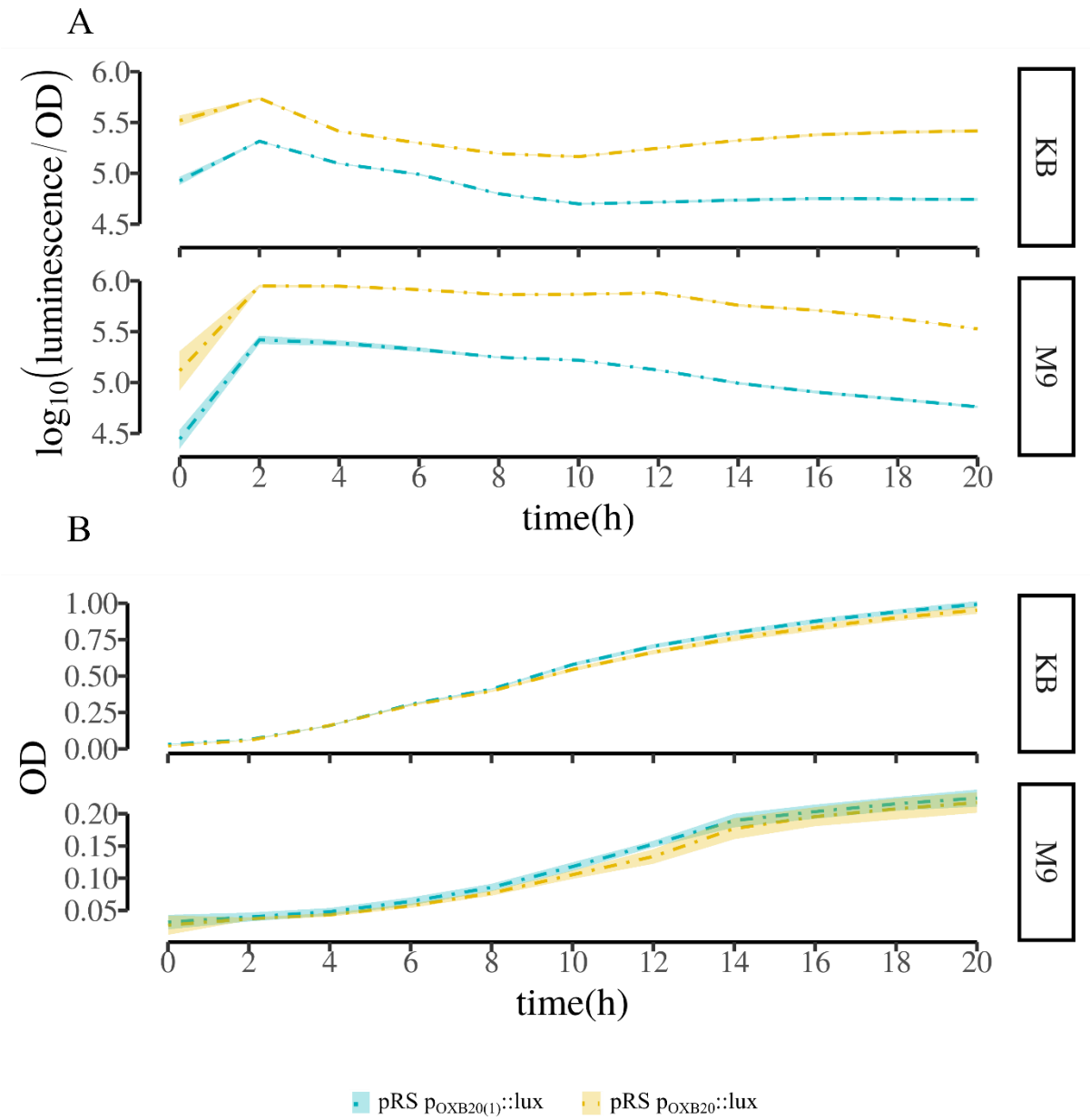
Supplementary Figure 2. Expression of *frp* in *P. fluorescens* NZ011 (pRSJ-p_{nptII}:lux-frp) results in an increase in NADP+ concentration. Lux: *P. fluorescens* NZ011 (pIJ11282). Lux-frp: *P. fluorescens* NZ011 (pRSJ-p_{nptII}:lux-frp). NADP+ concentration refers to 25 ul of bacterial culture, for which a 10-fold dilution was measured as having an optical density (OD600) of 0.33. Error bar +/- SE. Significant differences (t-test, $p<0.05$) are indicated by asterisks. n=6.



Supplementary Figure 3. The *lux*-eYFP operon constructed in this study. A: Schematic representation of the lux-eYFP construct made in this study. p1: p_{nptII}, p_{OXB20}, p_{OXB16}, p_{OXB13}, p_{OXB11}. p2: $p_{A1/04/03}$, p_{lac}. The red "T" represents the T0 terminator. B: *Pseudomonas* syringae pv. *phaseolicola* RJ3 (*Pph* RJ3) p_{nptII}::lux-$p_{A1/04/03}$::eYFP imaged with the Typhoon scanner (Amersham/GE Healthcare, UK) using Cy3 settings. eYFP expression is detected. C: *Pph* RJ3 p_{nptII}::lux $p_{A1/04/03}$::eYFP imaged with the nightOWL LB 983 at 10 s exposure. Bioluminescence is detected.

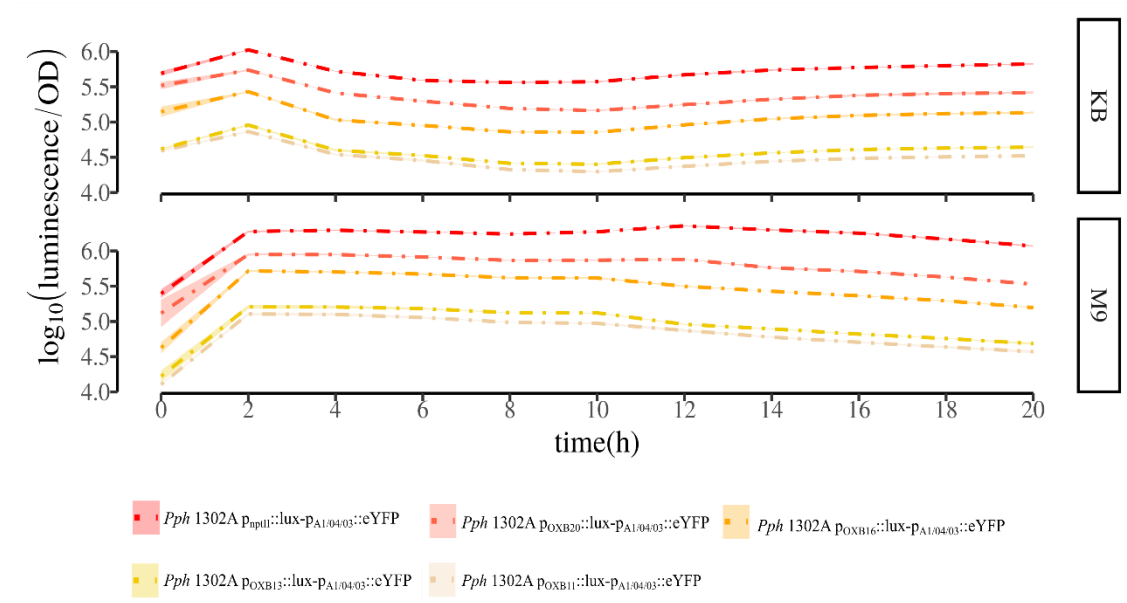


Supplementary Figure 4. A: Schematic representation of the pRS-$p_{OXB20(1)}$::lux plasmid. pRS-$p_{OXB20(1)}$::lux is derived from pUC18-mini-Tn7T-Gm-lux. MCS: Long multiple cloning site (490 bp) of pUC18-mini-Tn7T-Gm-lux plasmid. RBS1: Ribosome Binding site of plasmid pUC18-mini-Tn7T-Gm-lux. B: Schematic representation of the pRS-$_{pOXB20}$::lux plasmid. RBS2: Ribosome-binding site of plasmid pIJ11282. The RBS strength was calculated with the RBS calculator (Farasat *et al.*, 2014; Ng *et al.*, 2015) considering a total DNA length of 45 bp from the starting codon of luxC.

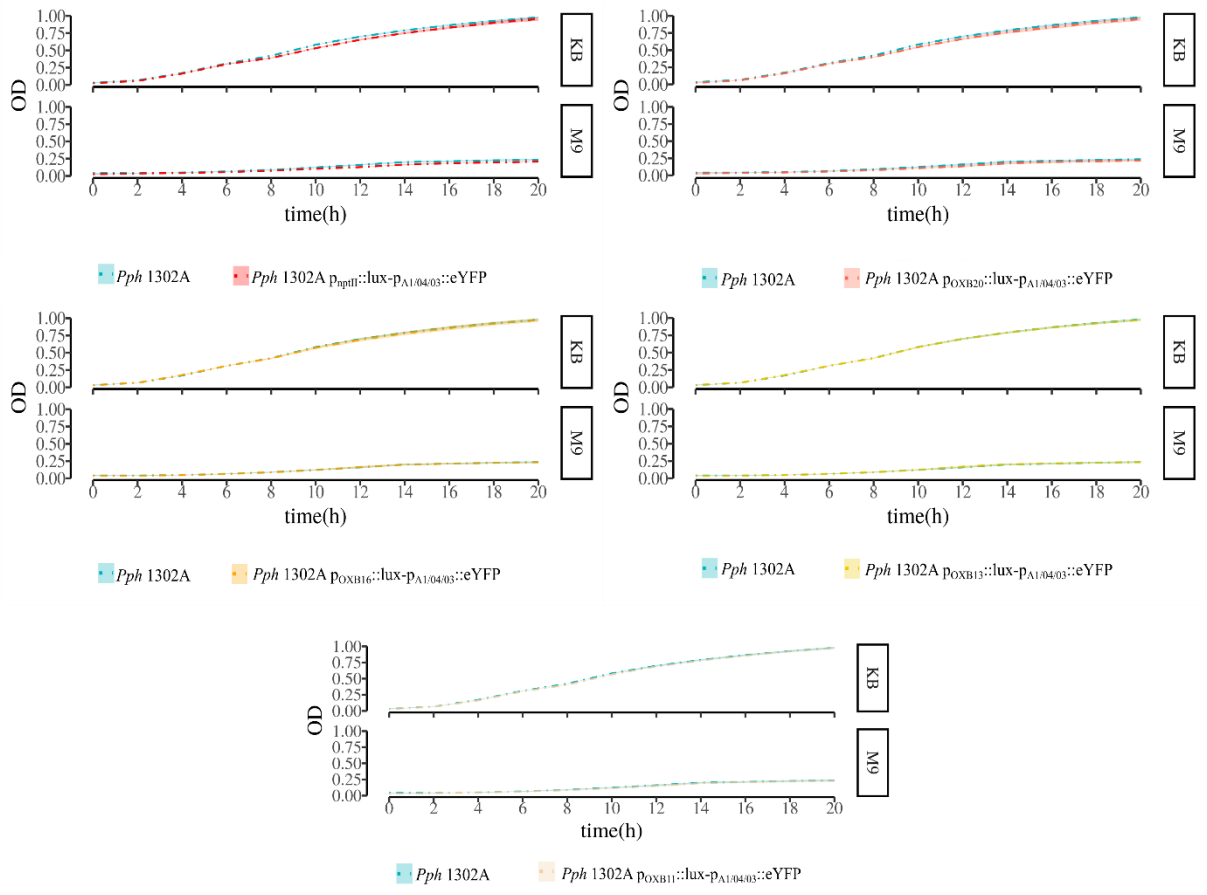


Supplementary Figure 5. pRS-p_{OXB20}::lux generates increased luminescence compared to pRS-$p_{OXB20(1)::}$lux in *Pseudomonas syringae* pv. *phaseolicola* 1302A. Plasmid pRS-$p_{OXB20(1)}$ ::lux was derived from the existing mini-Tn7 plasmid pUC18-mini-Tn7T-Gm-lux (Choi et al., 2005) by inserting oriT and ligating the OXB20 promoter upstream of *luxC*. Plasmid pRS-p_{OXB20}::lux has a stronger RBS (derived from pIJ11282) upstream of *luxC* compared to pUC18-mini-Tn7T-Gm-lux. A: Normalised luminescence of *Pph* reporter strains in KB (King's B medium) and M9 (M9 minimal medium). B: Growth of reporter strains in KB and M9. OD = optical density at 600 nm. Error bar +/- SE. n=3

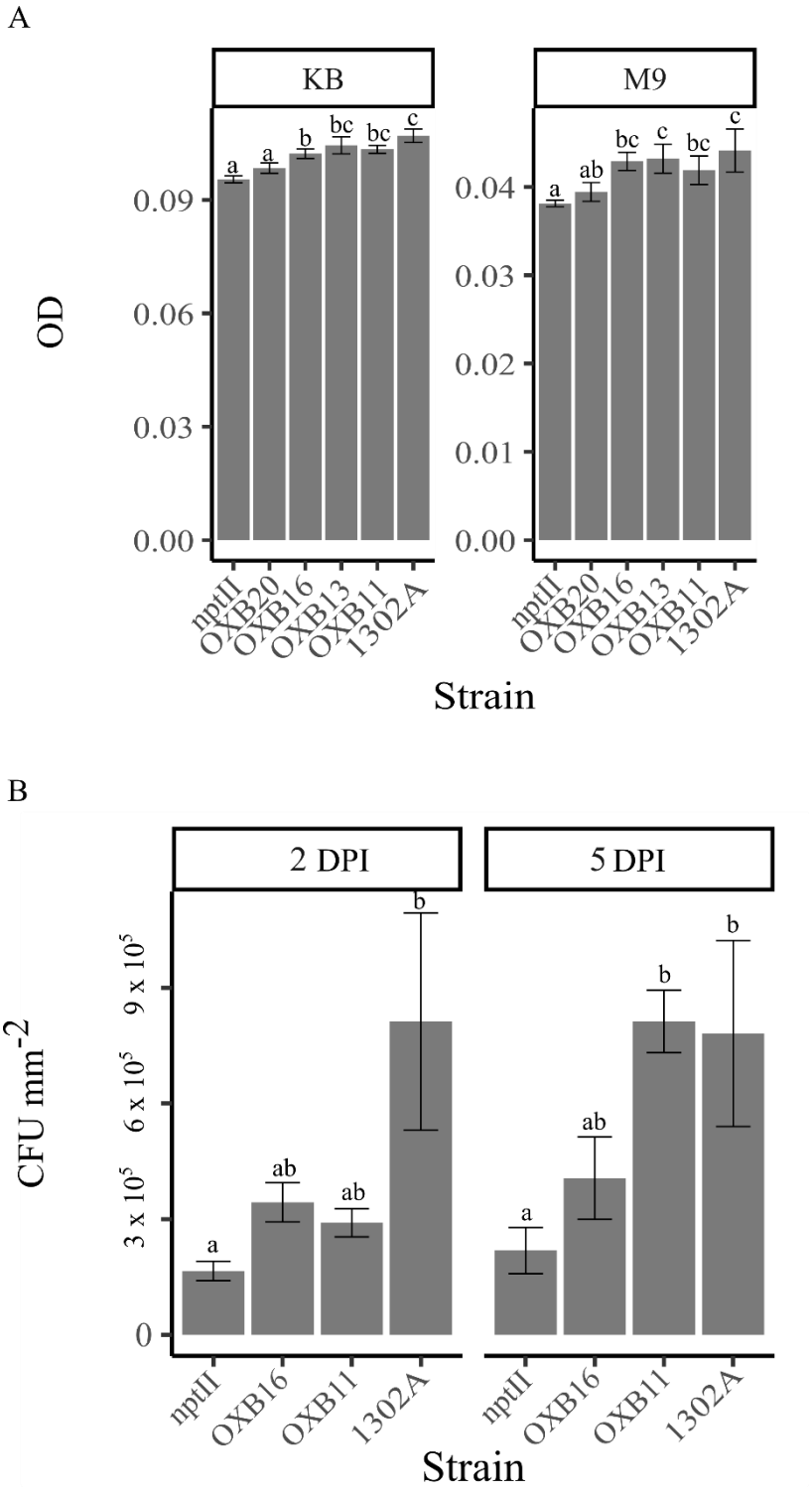
A


log10(luminescence/OD)
time(h)
KB
M9
Pph 1302A pnptII::lux-pA1/04/03::eYFP
Pph 1302A pOXB20::lux-pA1/04/03::eYFP
Pph 1302A pOXB16::lux-pA1/04/03::eYFP
Pph 1302A pOXB13::lux-pA1/04/03::eYFP
Pph 1302A pOXB11::lux-pA1/04/03::eYFP

B

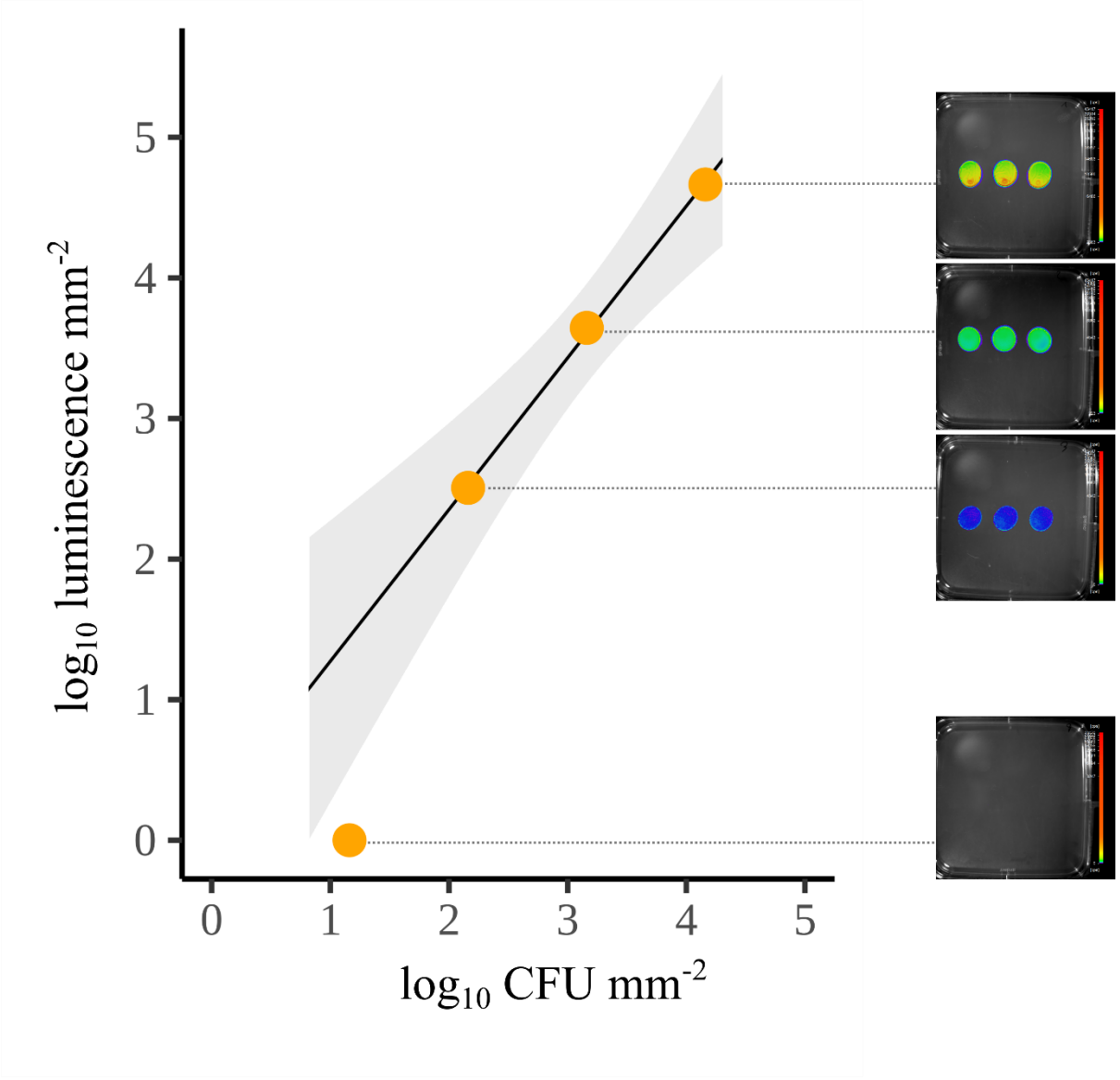

OD
time(h)
KB
M9
Pph 1302A
Pph 1302A pnptII::lux-pA1/04/03::eYFP
Pph 1302A pOXB20::lux-pA1/04/03::eYFP
Pph 1302A pOXB16::lux-pA1/04/03::eYFP
Pph 1302A pOXB13::lux-pA1/04/03::eYFP
Pph 1302A pOXB11::lux-pA1/04/03::eYFP

Supplementary Figure 6. *Pseudomonas syringae* pv. *phaseolicola* 1302A (*Pph* 1302A) reporter strains exhibit growth curves similar to *Pph* 1302A in both rich and minimal medium while providing a wide range of luminescence. Strains: *Pph* 1302A, *Pph* 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP, *Pph* 1302A p_{OXB13}::lux-$p_{A1/04/03}$::eYFP, *Pph* 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP, *Pph* 1302A p_{OXB20}::lux-$p_{A1/04/03}$::eYFP and *Pph* 1302A p_{nptII}::lux-$p_{A1/04/03}$::eYFP. A: Normalised luminescence of *Pph* reporter strains in KB (King's B medium) and M9 (M9 minimal medium) B: Growth of *Pph* 1302A reporter strains compared to *Pph* 1302A in KB and M9. OD = optical density at 600 nm.. Error bar +/- SE. n=3.

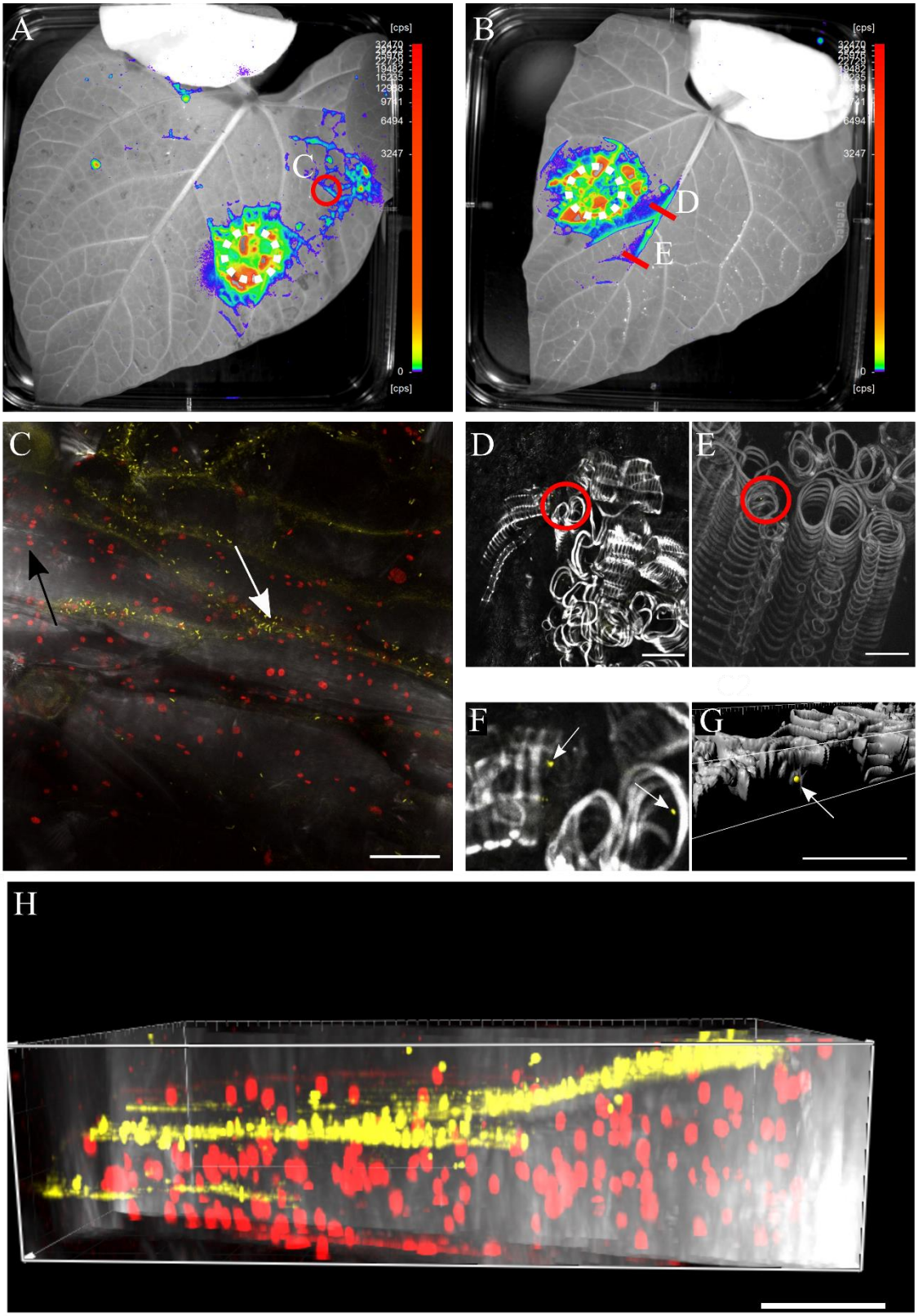


Supplementary Figure 7. Strong expression of the *lux* cassette has a small but significant effect on *Pph* 1302A bacterial growth both *in vitro* and *in planta*. A: OD values at 3 hours of data presented in Supplementary Figure 6. nptII: *Pph* 1302A p_{nptII}::lux-$p_{A1/04/03}$:::eYFP; OXB20: *Pph* 1302A p_{OXB20}::lux-$p_{A1/04/03}$::eYFP; OXB16: *Pph* 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP; OXB13: *Pph*

1302A p_{OXB13}::lux-$p_{A1/04/03}$::eYFP; OXB11: *Pph* 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP; 1302A: *Pph* 1302A. Error bar +/- SE. Significant differences (Tukey HSD, $p<0.05$) are indicated by letters. n=3. B: *In planta* growth of *Pph* 1302A reporter strains. TG leaves were syringe infiltrated with *Pph* 1302A reporter strains at $5x10^6$ CFU/ml in 10 mM $MgCl_2$. nptII: *Pph* 1302A p_{nptII}::lux-$p_{A1/04/03}$:::eYFP; OXB16: *Pph* 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP; OXB11: *Pph* 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP; 1302A: *Pph* 1302A. Error bar +/- SE. Significant differences (Tukey HSD, $p<0.05$) are indicated by letters. n=4.

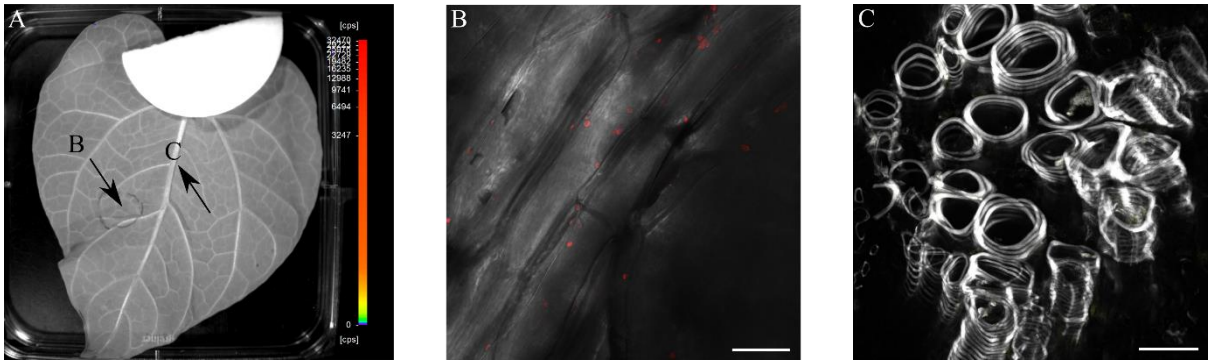


Supplementary Figure 8. Detection limit for bioluminescent *Pseudomonas syringae* pv. *phaseolicola* RJ3 (*Pph* RJ3) in agar plates. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP was spotted in serial dilution on an LB agar plate. Images were taken after 1 hour using the nightOWL LB 983 at 20 minutes exposure. Approximately 100 cells/mm^2 can be detected. Colour scales have different minimum values to avoid oversaturation. CFU/mm^2 was estimated by colony counting. SE for CFU and SE for luminescence constitute on average 27.7% and 2.66% of the means respectively. n=3.

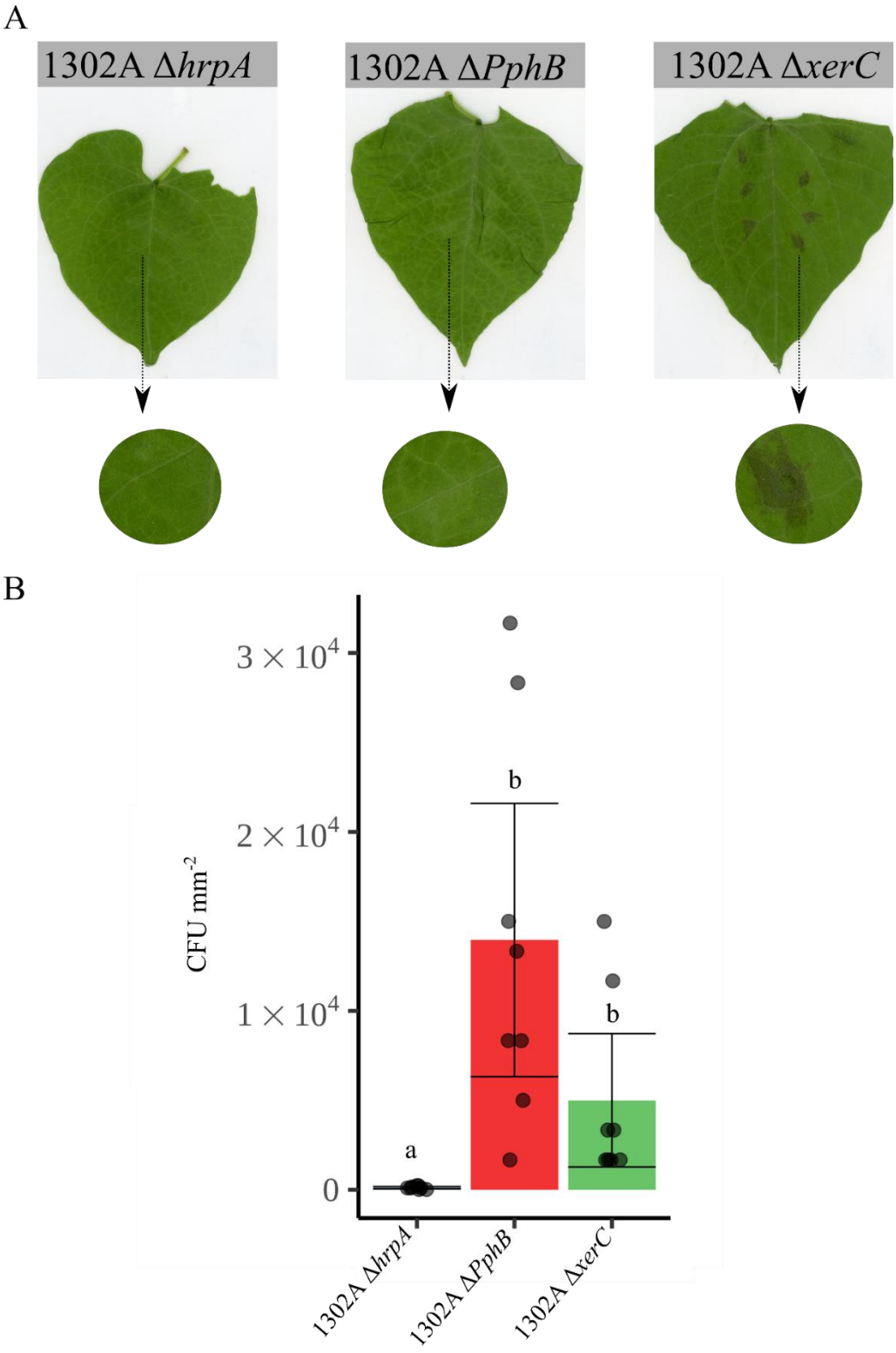


Supplementary Figure 9. *Pseudomonas syringae* pv. *phaseolicola* RJ3 (*Pph* RJ3) colonizes the leaf vasculature of *P. vulgaris* cultivar TG following spray inoculation onto the leaf surface. A,

B: *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP was sprayed in a localised area on the abaxial surface of *Phaseolus vulgaris* cv. Tendergreen leaves through an aperture in an aluminium foil (white dotted line). Images were taken after 2 DPI with the nightOWL LB 983 at 20 minute exposure. Red circle indicates leaf disks used for confocal microscopy imaging (C). Red lines indicates cross sections used for confocal microscopy imaging (D,E). C: Maximum projections of the leaf vasculature. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). Chlorophyll (red; black arrow). D: image of vasculature cross section. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow). Xylem autofluorescence (grey). Red circular line indicates area zoomed in panel F. E: Maximum projection of vasculature cross section. Xylem autofluorescence (grey). Red circular line indicates area modelled in panel G. F: zoom of panel D. *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). Xylem autofluorescence (grey). G: Three-dimensional model of panel E (Imaris 9 Bitplane AG, Zürich, Switzerland). *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP colonies (yellow; white arrow). H: Three-dimensional model of panel C (Imaris 9 Bitplane AG, Zürich, Switzerland). Co-localization of chlorophyll signal (red) and eYFP (yellow), indicates endophytic localization of *Pph* RJ3 p_{nptII}::lux-$p_{A1/04/03}$::eYFP in the leaf vasculature. Scale bars: 40 μm.



Supplementary Figure 10. *Phaseolus vulgaris* cultivar Tendergreen leaves do not display background bioluminescence or fluorescence under conditions used to image luminescence or eYFP. A: nightOWL LB 983 image. The image is automatically generated by the nightOWL LB 983 by overlaying the image acquired with CCD camera with a photo of the sample. Exposure time 20 minutes. The leaf was incubated for 10 minutes in darkness before imaging. Black arrows indicate tissue used for confocal microscopy imaging. B: details of the *Phaseolus vulgaris* cultivar TG vascular system. In non-inoculated plants, eYFP signal is not detectable. Chlorophyll (red), eYFP (yellow). C: details of xylem vessels of non-inoculated plants. eYFP signal is not detectable. Xylem (grey), eYFP (yellow). Settings used were the same as for inoculated plants. Scale bars: 40 µm.



Supplementary Figure 11. The colony count approach is able to detect the effect of PTI but not ETI on the growth of *Pseudomonas syringae* pv. *phaseolicola* in the early stages of infection (2 DPI). *P. vulgaris* cultivar Tendergreen (TG) leaves were syringe infiltrated with *Pph* 1302A

Δ*xerC*, Δ*hrpA* and Δ*PphB* at 10^6 CFU/ml in 10 mM $MgCl_2$. A: Details of the symptoms caused by *Pph* 1302A knock-out mutants in TG leaves at 2 days post inoculation (DPI). The hypersensitive response (HR) is evident on leaves inoculated with *Pph* 1302A Δ*xerC*. B: Colony count data for inoculated leaves at 2 DPI. Error bar +/- SE. Significant differences (Student's T-test, $p<0.05$) are indicated by letters. n=8.

Supplementary Video 1. *Pph RJ3*-eYFP colonizes the collenchyma of the *Phaseolus vulgaris* cultivar Tendergreen (TG) leaf vasculature. *Pph RJ3*-eYFP (Godfrey *et al.* 2010) was syringe-infiltrated at 10^6 CFU/ml into the first true leaves of TG plants. Leaf disks were detached at 3 days post inoculation and imaged with a TCS Leica SP5 confocal microscope in a xyz series (objective used: HCX PL APO CS 20.0x0.70 IMM UV). eYFP was excited with the 488 nm wave length laser. Cyan (*Pph* RJ3 cells); Red (Chlorophyll autofluorescence); Grey (bright field). Available in the onlint version.

Supplementary Video 2. *Pph RJ3*-eYFP (Godfrey *at al.*, 2010) moves systemically inside the vasculature of *Phaseolus vulgaris* cultivar Tendergreen (TG). *Pph RJ3*-eYFP was syringe-infiltrated at 10^6 CFU/ml into the first true leaves of TG plants. Leaf disks were detached at 1 day post inoculation and imaged with the a TCS Leica SP5 confocal microscope in a xyz series (objective used: HC PL APO CS2 63.0x1.20 WATER UV). eYFP was excited with the 488 nm wave length laser. Cyan (*Pph* RJ3 cells); Red (Chlorophyll autofluorescence); Grey (bright field). Available in the online version.

Supplementary Tables

Supplementary Table 1. Plasmids used in this study.

Plasmid	**Relevant characteristics**[1]
pRS-p_{nptII}::lux	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE*. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after *luxE*. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p_{OXB20}::lux	Suicide vector for mini-Tn7 delivery. OXB20 (*recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE*. Backbone pUC18T, suitable for conjugation. MscI opens the plasmid after *luxE*. Gm^r. This study**.**
pRS-p_{OXB16}::lux	Suicide vector for mini-Tn7 delivery. OXB16 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE*. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after *luxE*. Gm^r. This study.
pRS-p_{OXB13}::lux	Suicide vector for mini-Tn7 delivery. OXB13 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE*. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after *luxE*. Gm^r. This study.
pRS-p_{OXB11}::lux	Suicide vector for mini-Tn7 delivery. OXB11 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE*. Backbone pUC18T, suitable for

	conjugation. MscI digestion opens the plasmid after *luxE*. Gm^r. This study.
pRS-p$_{nptII}$::lux-frp	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE*. FMN reductase (*frp*) downstream of *luxE*. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after *frp*. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p$_{OXB20}$::lux-frp	Suicide vector for mini-Tn7 delivery. OXB20 (*recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE*. FMN reductase (*frp*) downstream of *luxE*. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after *frp*. Gm^r. This study.
pRS-p$_{nptII}$::lux-p$_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p$_{OXB20}$::lux-p$_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB20 promoter (*recA* promoter, Oxford Genetics Limited) driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.

pRS-p_{OXB16}::lux-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB16 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{OXB13}::lux-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB13 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{OXB11}::lux-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB11 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{nptII}::lux-frp-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE-frp* with A1/04/03 promoter driving expression of eYFP after *frp*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0

	terminator. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p_{OXB20}::lux-frp-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB20 promoter (*recA* promoter, Oxford Genetics Limited) driving expression of *luxCDABE frp* with A1/04/03 promoter driving expression of eYFP after *frp*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{nptII}::lux-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE* with lac promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p_{OXB20}::lux-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. OXB20 (*recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with lac promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{OXB16}::lux-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. OXB16 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with lac promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream eYFP. Backbone pUC18T, suitable for conjugation. MscI

	digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{OXB13}::lux-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. OXB13 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with lac promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{OXB11}::lux-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. OXB11 (modified from *recA* promoter, Oxford Genetics Limited) promoter driving expression of *luxCDABE* with lac promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-p_{nptII}::lux-frp p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. nptII promoter driving expression of *luxCDABE-frp* with lac promoter driving expression of eYFP after *frp*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. PdiI (NaeI)/SnaBI digestion removes nptII promoter. Gm^r. This study.
pRS-p_{OXB20}::lux-frp-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. OXB20 promoter (*recA* promoter, Oxford Genetics Limited) driving expression of *luxCDABE-frp* with lac promoter driving expression of eYFP

	after *frp*. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. MscI digestion opens the plasmid after lambda T0 terminator. Gm^r. This study.
pRS-$p_{OXB20(1)}$::lux	Suicide vector for mini-Tn7 delivery. OXB20 promoter (*recA* promoter, Oxford Genetics Limited) driving expression of *luxCDABE*. Backbone pUC18-mini-Tn7T-Gm-lux with oriT, suitable for conjugation. Gm^r. This study.
pRS-$p_{OXB20(1)::}$lux-$p_{A1/04/03}$::eYFP	Suicide vector for mini-Tn7 delivery. OXB20 promoter (*recA*, Oxford Genetics Limited) driving expression of *luxCDABE* with A1/04/03 promoter driving expression of eYFP after *luxE*. Lambda T0 terminator added downstream eYFP. Backbone pUC18-mini-Tn7T-Gm-lux with oriT, suitable for conjugation. Gm^r. This study.
pRSJ-p_{nptII}::ilux	Broad host range, low-copy number plasmid. nptII promoter driving expression of *ilux*. Backbone of pIJ11282 (Frederix *et al.*, 2014). Tet^r
pRSJ-p_{nptII}::lux-frp	Broad host range, low-copy number plasmid. nptII promoter driving expression of *luxCDABE frp*. Backbone of pIJ11282 (Frederix *et al.*, 2014). Tet^r
pRS-p_{lac}::eYFP	Suicide vector for mini-Tn7 delivery. lac promoter driving expression of eYFP. Lambda T0 terminator added downstream of eYFP. Backbone pUC18T, suitable for conjugation. Gm^r. This study.
pIJ11282	Broad host range, low-copy number plasmid. nptII promoter driving expression of *luxCDBAE*. (Frederix *et al.*, 2014). Tet^r

pGEX(-)	High-copy number plasmid harbouring *ilux*. (Gregor *et al.*, 2018). Amp^r
pK18mobsacB	Schäfer *et al.*, (1994). Kan^r
pUC18T-mini_Tn7T-Gm-dsRedExpress	Choi *et al.*, (2005). Gm^r
miniTn7(Gm)PA1/04/03-eyfp-a	Klausen *et al.*, (2003). Gm^r
pUC18-mini-Tn7T-Gm-lux	Choi et al., (2005). Gm^r.
pUX-BF13	Miller and Mekalanos, (1988). Carb^r
pRK2013	Yoshimoto *et al.*, (1991). Kan^r

[1] Abbreviations: Gm, gentamicin. Tet, tetracycline. Carb, carbenicillin. Kan, kanamycin. Amp, ampicillin.

Supplementary Table 2. List of strains used

Strain	Features
P. syringae pv. phaseolicola 1302A (*Pph* 1302°)	Taylor *et al.*, (1996). Carries the genomic island PPHGI-1. Causes disease in *P. vulgaris* cv. Canadian Wonder and induces AvrPphB-mediated effector-triggered immunity (ETI) in *P. vulgaris* cv. Tendergreen.
Pph RJ3	Jackson *et al.*, (2000). Derivative of *Pph* that lacks the genomic island PPHGI-1. Causes disease in *P. vulgaris* cv. Canadian Wonder and *P. vulgaris* cv. Tendergreen.

Pph 1302A Δ*PphB*	*avrPphB* deletion mutant. This study. Causes disease in *P. vulgaris* cv. Canadian Wonder and *P. vulgaris* cv. Tendergreen.
Pph 1302A Δ*hrpA*	*hrpA* deletion mutant. This study. Lacks a functional type III secretion system and elicits PAMP-triggered immunity (PTI) in *P. vulgaris*.
Pph 1302A Δ*xerC*	*xerC* mutant. Lovell *et al.*, (2009). Lacks a recombinase (XerC) that acts to excise PPHGI-1. Causes disease in *P. vulgaris* cv. Canadian Wonder and stably elicits ETI in *P. vulgaris* cv. Tendergreen.
Pph 1302A Δ*PphB* p_{OXB13}::lux	*Pph* 1302A Δ*PphB* expressing *lux* under weak constitutive promoter OXB13 (plasmid used: pRS-p_{OXB13}::lux). This study.
Pph 1302A ΔhrpA p_{OXB13}::lux	*Pph* 1302A Δ*hrpA* expressing *lux* under weak constitutive promoter OXB13 (plasmid used: pRS-p_{OXB13}::lux). This study.
Pph 1302A ΔxerC p_{OXB13}::lux	*Pph* 1302A Δ*xerC* expressing *lux* under weak constitutive promoter OXB13 (plasmid used: pRS-p_{OXB13}::lux). This study.
Pph 1302A p_{nptII}::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive nptII promoter and eYFP under constitutive A1/04/03 promoter (plasmid used: pRS-pnptII::lux-$p_{A1/04/03}$::eYFP). This study.
Pph 1302A p_{OXB20}::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive OXB20 promoter and eYFP under constitutive A1/04/03 promoter (plasmid used: pRS-pOXB20::lux-$p_{A1/04/03}$::eYFP). This study.
Pph 1302A p_{OXB16}::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive OXB16 promoter and eYFP under constitutive A1/04/03 promoter. (plasmid used: pRS-pOXB16::lux-$p_{A1/04/03}$::eYFP). This study.

Pph 1302A p_{OXB13}::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive OXB13 promoter and eYFP under constitutive A1/04/03 promoter. (plasmid used: pRS-pOXB13::lux-$p_{A1/04/03}$::eYFP). This study.
Pph 1302A p_{OXB11}::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive OXB11 promoter and eYFP under constitutive A1/04/03 promoter. (plasmid used: pRS-pOXB11::lux-$p_{A1/04/03}$::eYFP). This study.
Pph 1302A $p_{OXB20(1)}$::lux-$p_{A1/04/03}$::eYFP	*Pph* 1302A expressing *lux* under constitutive OXB20 promoter and eYFP under constitutive A1/04/03 promoter (plasmid used: pRS-pOXB20(1)::lux-$p_{A1/04/03}$::eYFP). This study.
Pph 1302A (pIJ11282)	*Pph* 1302A expressing *lux* under constitutive nptII promoter. This study.
Pph 1302A (pRSJ-p_{nptII}::ilux)	*Pph* 1302A expressing ilux under constitutive nptII promoter. This study. This study.
Pph 1302A (pRSJ-p_{nptII}::lux-frp)	*Pph* 1302A expressing lux frp under constitutive nptII promoter. (plasmid used: pRSJ-p_{nptII}::lux-frp). This study.
Acinetobacter baylyi ADP1	WT. Metzgar *et al.*, (2004)
Pseudomonas fluorescens NZ011	WT. Godfrey *et al.*, (2001)
Acinetobacter baylyi ADP1 (pIJ11282)	*A. baylyi* ADP1 expressing *lux* under constitutive nptIIpromoter. This study.
Acinetobacter baylyi ADP1 (pRSJ-p_{nptII}::ilux)	*A. baylyi* ADP1 expressing *ilux* under constitutive nptII promoter. This study.

Acinetobacter baylyi ADP1 (pRSJ-p_{nptII}::lux-frp)	*A. baylyi* ADP1 expressing *lux-frp* under constitutive nptII promoter. This study.
Pseudomonas fluorescens NZ011 (pIJ11282)	*P. fluorescens* NZ011 expressing *lux* under constitutive nptII promoter. This study.
Pseudomonas fluorescens NZ011 (pRSJ-p_{nptII}::ilux)	*P. fluorescens* NZ011 expressing *ilux* under constitutive nptII promoter. This study.
Pseudomonas fluorescens NZ011 (pRSJ-p_{nptII}::lux-frp)	*P. fluorescens* NZ011 expressing *lux-frp* under constitutive nptII promoter. This study.
Pseudomonas fluorescens NZ011 p_{nptII}::lux-frp	*P. fluorescens* NZ011 expressing *lux-frp* under constitutive nptII promoter (plasmid used: pRS-p_{nptII}::lux-frp). This study.
Pseudomonas fluorescens NZ011 p_{nptII}::lux	*P. fluorescens* NZ011 expressing *lux* under constitutive nptII promoter (plasmid used: pRS-p_{nptII}::lux). This study.
Xanthomonas campestris 8004 p_{OXB20}::lux-$p_{A1/04/03}$::eYFP	*X. campestris* 8004 expressing *lux* under constitutive OXB20 promoter and eYFP under constitutive A1/04/03 promoter (plasmid used: pRS-p_{OXB20}::lux-$p_{A1/04/03}$::eYFP). This study.
*E. coli DH5*α (pIJ11282)	Frederix *et al.*, (2014)
E. coli SM10/λ pir (pUX-BF13)	Miller and Mekalanos, (1988)
E. coli HB101 (pRK2013)	Yoshimoto *et al.*, (1991)

Supplementary Table 3. List of primers used for cloning

Primer name	Sequence 5'-3'
OXRS_FT	GTAAGGAGAAAATACCGCATCAGGCCGGCCAGCCTCGCAGA

OXRS_RT	CAGCCTGAATGGCGAATGGCTTTCCGCTGCATAACCCTG
OXRS_F1	GACGGATCGATCCGGGGAATTCAGGCTTGGAGGATACGTATGACTAAAAAAATTTCATTCATTATTACCGGCCAGGTTG
OXRS_R1	GGGTCAGTTCCGGCTGGGGGTTCAGCAGCCACCTGCAGTTACCTTCTGGCAAGGCCCTTGG
OXRS_F2	CCGAAGCGTTTGATAGTTGATGACCTGCAGTCGACCTAAGGAGAAAGAAATGGTGAAGATACAGCCCATCCCCAC
OXRS_R2	CCTTGTGGGGTCAGTTCCGGCTGGGGGTTCAGCAGCCACCTGCAGCACGTGTTACCTTCTGGCAAGGCCCTTGG
OXRS_F3	TTTTGAAGCTAATTCGATCATGCATGAGGCATTTCACGGACTTTCATGGG
OXRS_R3	ACTAGATTTCACTTATCTGGTTGGCCCCTTGTGGGGTCAGTTCCGG
OXRS_F4	GGCCAACCAGATAAGTGAAATCTAGT
OXRS_R4	TCATGCATGATCGAATTAGCTTCAAAA
OXRS_F5	CCCCCAGCCGGAACTGACCCCACAAGGTGGGTCGACCTAAGGAGAAAGAAATGGTGAAGATACAGCCCATCCCCAC
OXRS_R5	ACTAGATTTCACTTATCTGGTTGGCCATTACCTTCTGGCAAGGCCCTTGG
OXRS_F6	CCGGAACTGACCCCACAAGGTGGCAATACGCAAACCGCCTCTC
OXRS_R6	ACTAGATTTCACTTATCTGGTTGGCCAGAGAGTCATTACCCCAGGCGTT
OXRS_F7	TGAACTCCAAGGGCCTTGCCAGAAGGTAATGGCAATACGCAAACCGCCTCTC
OXRS_R7	TTTGGAACTAGATTTCACTTATCTGGTTGGCCAGAGAGTCATTACCCCAGGCGTT
OXRS_F8	GGAATTGGGTACCTCGCGAAGGCAATACGCAAACCGCCTCTC
OXRS_R8	CACTTATCTGGTTGGCCTGCAAGGCCTGAGAGTCATTACCCCAGGCGTT
OXRS_F9	TACCGCCTTTGAGTGAGCTGATACCGCACTGCGTCAACTGCTCGAA

OXRS_R9	TTACGAAACTCTAAACTCGTTTACGCAGGTGGCCTGCGTACCTATTTTCAT
OXRS_F10	ATGAAAATAGGTACGCAGGCCACCTGCGTAAACGAGTTTAGAGTTTCGTAA
OXRS_R10	TTCCCTTGTCCAGATAGCCCAGTAGCCACAAGGTGTAGCCTGAGCG
OXRS_F11	GCTACTGGGCTATCTGGACAAGGGAA
OXRS_R11	CGGTATCAGCTCACTCAAAGGCGGTA
OXRS_F12	TTGCTCACATGTTCTTTCCTGCG
OXRS_R12	TTCGCTTGCTGTCCATAAAACCG
OXRS_F13	TACCGCCTTTGAGTGAGCTGATACCGATGGATCTTGATGAGGGGTTTGATGA
OXRS_R13	AGAAATCAGAATCGGACGACCGAAGAACTCATGATGTTCATAACGTGATACC
OXRS_F14	GGTATCACGTTATGAACATCATGAGTTCTTCGGTCGTCCAGTTCTGATTTCT
OXRS_R14	TTCCCTTGTCCAGATAGCCCAGTAGCCAACAAGCCACCGAGTAACTGAC
OXRS_F15	AATCCGACCTGACCCTGCT
OXRS_R15	CACGCCCCTCTTTAATACGACGG

Supplementary Table 4. List of primers used to sequence the *ilux* operon

Primer name	Sequence 5’-3’
OX26	AATGTCATGCAACCGTAATTCG
OX27	TTCAACACAAATAACGTGGTCG
OX28	TCGACCACGTTATTTGTGTTG
OX29	CAAGCAATTTCTCTGTCTTAAAATCTATTGAG
OX30	ACCAGAAACAGTCAGAGGAG

OX31	CGCTCCTCTGACTGTTTCTG
OX32	CAGGTATGACTTCATATGTTGATAAAC
OX33	ATGCGTCAGCAACCAGTTATC

Supplementary Table 5. List of primers used to sequence eYFP and *frp*.

Primer name	Sequence 5'-3'
OXRS_S1	ATGCGTCAGCAACCAGTTATC
OXRS_S2	AACCAGCAACCAGAAGCAGAGTG

Supplementary Table 6. List of primers used to sequence the *lux* operon

Primer name	Sequence 5'-3'
OX26	AATGTCATGCAACCGTAATTCG
OX36	CCCACCAAGGTGATACATCACTC
OX37_	TTCTCATGAAAGGCCATCTAAC
OX38	AACGATAATTGGGTCAAGCAAG
OX39	AAAGGACGATTTCGGTTTGG
OX40	ATGCAGGAAATAACGGAGTATG
OX41	TGCTGAAAGATATAAAGCCGTTGC
OX42	CAAGAAATTACAGCAAGCTCAG
OX43	AGATCTTTGTCTTATTGGTTCGCC

References for Supplementary Tables

Choi, K.-H., Gaynor, J.B., White, K.G., Lopez, C., Bosio, C.M., Karkhoff-Schweizer, R.R., and Schweizer, H.P. (2005) A Tn 7 -based broad-range bacterial cloning and expression system. *Nature Methods* 2: 443.

Frederix, M., Edwards, A., Swiderska, A., Stanger, A., Karunakaran, R., Williams, A., et al. (2014) Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Molecular Microbiology* 93: 464–478.

Godfrey, S. a. C., Harrow, S.A., Marshall, J.W., and Klena, J.D. (2001) Characterization by 16S rRNA Sequence Analysis of Pseudomonads Causing Blotch Disease of Cultivated *Agaricus bisporus*. *Appl Environ Microbiol* 67: 4316–4323.

Jackson, R.W., Mansfield, J.W., Arnold, D.L., Sesma, A., Paynter, C.D., Murillo, J., et al. (2000) Excision from tRNA genes of a large chromosomal region, carrying avrPphB, associated with race change in the bean pathogen, *Pseudomonas syringae* pv. *phaseolicola*. *Mol Microbiol* 38: 186–197.

Klausen, M., Heydorn, A., Ragas, P., Lambertsen, L., Aaes-Jørgensen, A., Molin, S., and Tolker-Nielsen, T. (2003) Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Molecular Microbiology* 48: 1511–1524.

Lovell, H.C., Mansfield, J.W., Godfrey, S.A.C., Jackson, R.W., Hancock, J.T., and Arnold, D.L. (2009) Bacterial Evolution by Genomic Island Transfer Occurs via DNA Transformation In Planta. *Current Biology* 19: 1586–1590.

Metzgar, D., Bacher, J.M., Pezo, V., Reader, J., Döring, V., Schimmel, P., et al. (2004) *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res* 32: 5780–5790.

Miller, V.L. and Mekalanos, J.J. (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J Bacteriol* 170: 2575–2583.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium* glutamicum. *Gene* 145: 69–73.

Taylor, J.D., Teverson, D.M., Allen, D.J., and Pastor-Corrales, M.A. (1996) Identification and origin of races of *Pseudomonas syringae* pv. *phaseolicola* from Africa and other bean growing areas. *Plant Pathology* 45: 469–478.

Yoshimoto, T., Higashi, H., Kanatani, A., Lin, X.S., Nagai, H., Oyama, H., et al. (1991) Cloning and sequencing of the 7 alpha-hydroxysteroid dehydrogenase gene from *Escherichia coli* HB101 and characterization of the expressed enzyme. *Journal of Bacteriology* 173: 2173–2179.

Chapter 6: General conclusion

Plant-microbe interactions are of pivotal importance for plant growth and development and they have been recently recognized as an important player for achieving zero-hunger, one of the sustainable development goals (SDG) of the United Nations (FAO, 2019). Attempts to bridge the gap between science and policymakers in the context of plant microbiomes at the international level are evident by recent funding devoted to this aspect (FAO, 2021. Internal communication). Nonetheless, a better understanding of plant-microbe interactions in the agricultural ecosystem would foster applications of plant microbiomes. Formulating research hypothesis clearly aimed at understanding plant-microbe interactions in wild-type plants and domesticated plants and their environment seems increasingly appropriate as well as the development of new evolutionary frameworks.

Starting with the simple idea that whenever a plant phenotype is heritable it can have heritable consequences at the ecosystem level (Whitham *et al.*, 2003), and that domesticated plants showcase a remarkable plethora of specific heritable phenotypes, it seemed logical to apply this concept to plant domestication and its consequences for plant-microbe interactions. I hypothesized that domesticated plant microbiomes could show differences in assembly from those of wild progenitors because of an effect of domesticated phenotypes, and that when domesticated phenotypes are a consequence of artificial selection, a reduction in positive microbe-to-host effects is expected (chapter 2) (Soldan, Fusi, *et al.*, 2021). Indeed, in chapter 3, I demonstrated that there is a link between domesticated plant microbiomes and microbiome signature, and this link is independent of the domestication event because different domestication scenarios lead to similar plant phenotypes.

Importantly, by using two different plant species that were both domesticated twice independently, I observed that the domestication effect is consistent within plant species but not between. From a statistical model perspective, this suggests an interaction between domestication and plant species. A similar experimental design could allow us to understand whether this plant phenotype effect on microbiome assembly is also observed in other plant compartments and with which magnitude. According to my double-leash framework, we would expect stronger differences in microbiome assembly between wild and domesticated plants for plant compartments that have showcased higher phenotypic changes. Moreover, to consider that for plant compartments where microbiome diversity is low, such as the endopshere, host-control of the microbiota through host-to-microbe effects is expected to be higher (Brown, 1999). Therefore, we could predict that domesticated plant phenotypes would have a stronger effect when microbial diversity is low.

For example, in cereal crops, the seed microbiome of wild and domesticated plants would be predicted to differ more than the correspondent rhizosphere microbiome because the seed microbiome has lower alpha-diversity (Guo *et al.*, 2021) and because the seed compartment has showcased the most phenotypic changes. Our results in chapter 4 and the results of Jaramillo and co-authors (Pérez-Jaramillo *et al.*, 2017) support this hypothesis. Changes in the rhizosphere microbiome due to domestication are less pronounced than changes in the seed microbiome.

It could be interesting to perform an experiment in which several domesticated crops and their wild-type progenitors are grown and the microbiomes of different compartments compared. The selection of the crops would depend on the plant compartment mainly targeted during domestication. For example, the seed compartment was targeted in cereals

and legumes while the tuber compartment was targeted in plants that provide tuber crops. Therefore, if my hypothesis is correct, we would expect stronger differences in microbiome assembly between wild and domesticated plants for the tuber compartment (tuber-associated microbiome) in tuberous plants and the seed compartment for cereals and legumes.

At later stages, investigating the consequences at the microorganism-to-host effects level will also be important, meaning investigating the functions performed by the microbiome. This task will surely require whole genome sequencing and prediction of bacterial phenotypes, but could ultimately help us to understand the consequences of plant domestication on plant-microbe interactions. For example, it could help us understand whether the differences between wild and domesticated plant microbiomes translate to different microbiome functions being performed or whether taxonomically different individuals are performing similar functions. This is important to better understand the role of plant microbiomes in plant evolution.

However, whole-genome sequencing could impose some technical challenges. For example, host reads would need to be mapped against a reference genome and subsequently removed. This requires a published and available high-quality genome of the crop being investigated. To date, to reduce the number of plant reads, whole genome sequencing has been perform after homogenization and size fractioning, for example, to capture leaf endophytes (Nobori *et al.*, 2018). Nonetheless, new sequencing technologies could yield enough sequencing depth to sufficiently cover bacterial genomes even in the presence of a high ratio between plant and microbial reads.

Another important aspect of plant-microbe interactions that we discuss in chapter 3, is the holobiont concept and how it requires further thinking when referring to plants that have

been domesticated and that are growing in an agricultural ecosystem. We acknowledge that the holobiont concept has received much criticism (Douglas and Werren, 2016), especially regarding the inheritance of the microbiome and what new insights the holobiont theory brings to biological sciences (Simon *et al.*, 2019). We argue that the holobiont concept is important because it makes clear the linkage between plants and the microbiome and how the microbiome can ultimately influence the evolution of its host and vice-versa. We propose that broad-sense community heritability could be a useful matric providing researchers with indications about potential phylosymbiosis or genetic covariance between host and microbes. Indeed, if community heritability is significant, it implies that variation in plant genotypes is affecting host microbial communities and that some microbial communities are specific to some plant genotypes, suggesting community selection is occuring. (Whitham *et al.*, 2006).

Nonetheless, we also reflect that H^2_C is not a metric without cons. For example, we acknowledge that condensing multivariate data into univariate scores through NMDS on a community matrix is not optimal. Better approaches exist, such as model-based approaches implemented in the R package mvabund (Wang *et al.*, 2012). This is currently being investigated and will lead to another future manuscript. Moreover, another aspect that H^2_C does not consider is phylogeny. In other words, broad-sense community heritability could achieve low scores even when phylosymbiosis is occurring. This is the case of high within plant genotype variation in community data due to phylogenetically related individuals. It is also true that very often in multivariate statistics, phylogenetics is not considered. This is another aspect I am currently working on and we hope will lead to another future publication.

In chapter 4 I show that for plants grown under controlled conditions, differences in seed Calcium concentration explains selection processes acting at a microbial community level. Whether these selected communities provide specific microbe-to-host effects would require further investigation. For now, I could conclude that the differences in microbiome assembly between wild and domesticated plants are caused by differences in host phenotypic traits that have been influenced by domestication. This also likely means that community selection is occurring, as these plant phenotypic traits are heritable.

In addition, it is noteworthy that our results showing a high and low abundance of *Pseudomonadales* in fresh and dried seeds respectively agree with a recent research article showing that the abundance of *Pseudomonadaceae* is mainly affected by the seed drying process and not by storing conditions or time (Chandel *et al.*, 2021). Indeed, we have observed a high abundance of *Pseudomonadales* for experiment PVG, but not for experiment PVC, in which seeds were dried before storage.

Another aspect of experiment PVG that deserves consideration is that *Pseudomonadaceae* abundance was significantly lower in wild-type seeds for both domestication events. Understanding why is complex as it could be the results of host-to-microbe or microbe-to-microbe effects. Nonetheless, if we consider that two of the major bacterial pathogens of common bean is *Pseudomonas syringae* pv *phaseolicola* (Pph), a seed-borne pathogen (Arnold *et al.*, 2011), it seems reasonable to hypothesize that this pathogen could show increased virulence towards domesticated plants as a consequence of *Pseudomonadaceae* being selected in domesticated seed plants. It could be interesting to perform an experiment in which the transmission rate of Pph to common bean seeds and transmission from seed to shoot is measured in wild and domesticated plants and correlated with several domesticated plant traits to try to identify the cause of different

transmission rates. Some of these traits could explain differences in transmission rate at a mechanistic level, for example, the lower concentration of secondary metabolites in domesticated seeds (Shlichta *et al.*, 2018)

In chapter 5 I use *Pph* to develop a new tool that can be used to monitor bacterial localization within plant tissues and transmission to and from seeds. The technique developed would allow researchers to quickly monitor a bacterium of interest from both macro and micro resolution using common laboratory equipment. Thanks to this technique, which combines fluorescence and bioluminescence, I showed that *Pph* is systemic and colonizes the leaf vasculature. This is an important finding as it could suggest that *Pph* can reach the seeds through the vasculature system of infected leaves. Moreover, I showed the versatility of using bioluminescence and how it can be useful to study the impact of plant immune responses on bacteria viability. Applications of this simple idea have been successfully used for the development of a bioluminescent strain of *Agrobacterium* by Jutras, Sodan, and co-authors (Jutras *et al.*) that can be used to study the effect of plant immune responses on bacteria.

In the close future, developing bioluminescent constructs with different wave-length emissions coupled with the development of more sensitive photon-detecting cameras would allow researchers to dive dip into plant-microbe interactions and microbe-microbe interactions. In this process, ensuring that the developed biosensors do not have a strong effect on bacterial fitness is also important, as we show in chapter 5. These further developments of my technique would allow researchers to better understand when and how bacteria are transferred to seeds and in general how microbiome connectivity between plant compartments is achieved because multiple microorganisms can be studied simultaneously. For example, it could be interesting to tag two or three members of the

seed microbiome community and *monitor* their colonization habit in the developing seedling through recolonization experiments. As a starting point, *Proteobacteria* are the easiest to isolate and tag (Soldan, Sanguankiattichai, *et al.*, 2021), but further efforts to develop molecular toolboxes for phyla such as *Bacteroidetes* are also important.

In the last twenty years, plant-microbe interaction research has seen incredible developments. It has become apparent that this field has become important not only within academia but also at the industry and policy level and could ultimately allow us to achieve a more sustainable agricultural production. However, this requires a multidisciplinary approach to tackle the many problems associated with studying complex systems such as the interaction between plants and their microbiome. I believe that the chapters of my thesis bring together advancements in different fields of plant-microbe interaction research, showing how they are all highly interconnected. I conceptualized how domesticated plant microbiomes could assemble and the potential consequences for beneficial microbe-to-host effects. I have also validated this framework by statistically showing that independent domestication events led to similar microbiome signatures on the seed microbiome and provided a tool that can be used to further investigate the colonization habit of seed-associated bacteria.

In summary, the chapters of my thesis present advancements on the 3 important aspects of plant-microbe interactions, namely i) evolutionary frameworks, ii) factors driving community assembly, and iii) visualization of plant-microbe interactions. I think that the studies I performed could serve as a base to further develop new research hypotheses and tools that would allow us to better understand and conceptualize the role of the plant microbiome in agricultural ecosystems.

References

Arnold, D.L., Lovell, H.C., Jackson, R.W., and Mansfield, J.W. (2011) Pseudomonas syringae pv. phaseolicola: from “has bean” to supermodel. *Mol Plant Pathol* 12: 617–627.

Brown, S.P. (1999) Cooperation and conflict in host–manipulating parasites. *Proceedings of the Royal Society of London Series B: Biological Sciences* 266: 1899–1904.

Chandel, A., Mann, R., Kaur, J., Norton, S., Edwards, J., Spangenberg, G., and Sawbridge, T. (2021) Implications of Seed Vault Storage Strategies for Conservation of Seed Bacterial Microbiomes. *Frontiers in Microbiology* 12: 3512.

Douglas, A.E. and Werren, J.H. (2016) Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *mBio* 7:.

FAO (2019) Microbiome: The missing link?: Science and innovation for health, climate and sustainable food systems, Rome, Italy: FAO.

Guo, J., Ling, N., Li, Y., Li, K., Ning, H., Shen, Q., et al. (2021) Seed-borne, endospheric and rhizospheric core microbiota as predictors of plant functional traits across rice cultivars are dominated by deterministic processes. *New Phytologist* 230: 2047–2060.

Jutras, P.V., Soldan, R., Dodds, I., Schuster, M., Preston, G.M., and Hoorn, R.A.L. van der AgroLux: bioluminescent Agrobacterium to improve molecular pharming and study plant immunity. *The Plant Journal* n/a:

Nobori, T., Velásquez, A.C., Wu, J., Kvitko, B.H., Kremer, J.M., Wang, Y., et al. (2018) Transcriptome landscape of a bacterial pathogen under plant immunity. *PNAS* 115: E3055–E3064.

Pérez-Jaramillo, J.E., Carrión, V.J., Bosse, M., Ferrão, L.F.V., de Hollander, M., Garcia, A.A.F., et al. (2017) Linking rhizosphere microbiome composition of wild and domesticated *Phaseolus vulgaris* to genotypic and root phenotypic traits. *ISME J* 11: 2244–2257.

Shlichta, J.G., Cuny, M.A.C., Hernandez-Cumplido, J., Traine, J., and Benrey, B. (2018) Contrasting consequences of plant domestication for the chemical defenses of leaves and seeds in lima bean plants. *Basic and Applied Ecology* 31: 10–20.

Simon, J.-C., Marchesi, J.R., Mougel, C., and Selosse, M.-A. (2019) Host-microbiota interactions: from holobiont theory to analysis. *Microbiome* 7: 5.

Soldan, R., Fusi, M., Cardinale, M., Daffonchio, D., and Preston, G.M. (2021) The effect of plant domestication on host control of the microbiota. *Commun Biol* 4: 1–9.

Soldan, R., Sanguankiattichai, N., Bach-Pages, M., Bervoets, I., Huang, W.E., and Preston, G.M. (2021) From macro to micro: a combined bioluminescence-fluorescence approach to monitor bacterial localization. *Environmental Microbiology* 23: 2070–2085.

Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012) mvabund– an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3: 471–474.

Whitham, T.G., Bailey, J.K., Schweitzer, J.A., Shuster, S.M., Bangert, R.K., LeRoy, C.J., et al. (2006) A framework for community and ecosystem genetics: from genes to ecosystems. *Nature Reviews Genetics* 7: 510–523.

Whitham, T.G., Young, W.P., Martinsen, G.D., Gehring, C.A., Schweitzer, J.A., Shuster, S.M., et al. (2003) Community and Ecosystem Genetics: A Consequence of the Extended Phenotype. *Ecology* 84: 559–573.