bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dissecting the co-transcriptome landscape of plants and microbiota members

Tatsuya Nobori[3,4], Yu Cao[3], Frederickson Entila[3], Eik Dahms[3], Yayoi Tsuda[1,2,3], Ruben Garrido-Oter[3,5], Kenichi Tsuda[1,2,3*]

[1]Key Laboratory of Agricultural Microbiology, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China.

[2]The Provincial Key Lab of Plant Pathology of Hubei Province, Huazhong Agricultural University, Wuhan, China

[3]Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, Cologne 50829, Germany.

[4]Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

[5]Cluster of Excellence on Plant Sciences, 40225 Düsseldorf, Germany.

*Corresponding author. Email: tsuda@mail.hzau.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Interactions between plants and each neighboring microbial species are fundamental building blocks that collectively determine the structure and function of the plant microbiota, but the molecular basis of such interactions is poorly characterized. Here, we monocolonized *Arabidopsis* leaves with nine plant-associated bacteria from all major phyla of the plant microbiota and profiled co-transcriptomes of plants and bacteria. These strains elicited quantitatively different plant transcriptional responses including typical pattern-triggered immunity responses. Genes of non-pathogenic bacteria involved in general metabolism and energy production were commonly suppressed *in planta* in contrast to a virulent pathogen. Various nutrient acquisition pathways that are frequently encoded in the genomes of plant-associated bacteria were induced *in planta* in a strain-specific manner, shedding light on bacterial adaptation to the plant environment and identifying a potential driving force of niche separation. Integrative analyses of plant and bacterial transcriptomes suggested that the transcriptional reprogramming of plants is largely uncoupled from that of bacteria at an early stage of interactions. This study provides insights into how plants discriminate among bacterial strains and sets the foundation for in-depth mechanistic dissection of plant-microbiota interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

In nature, plants can reproducibly assemble bacterial communities with well-defined taxonomic structures (the plant microbiota) (Hacquard et al., 2015), which can be harnessed for plant health and survival (Carrión et al., 2019; Durán et al., 2018; Kwak et al., 2018). How plants discriminate among various bacterial strains and establish strain-specific associations in a community context remains an open question in basic plant microbiota research and is key to facilitate the application of microbiota-based strategies to improve plant health in agricultural settings. Answering this question requires a comprehensive and unified understanding of plant and bacterial responses during their interactions.

Plant responses to microorganisms are controlled by the plant innate immune system, which contributes to the assembly and maintenance of healthy bacterial communities (Chen et al., 2020; Lebeis et al., 2015). A crucial part of the plant immune system is the perception of environmental microbes using cell surface receptors that detect conserved microbial epitopes, termed microbe-associated molecular patterns (MAMPs) (Boller and Felix, 2009). Recognition of MAMPs triggers defense responses collectively called pattern-triggered immunity (PTI), which can inhibit pathogen growth (Jones and Dangl, 2006). MAMPs such as the bacterial flagellin peptide flg22 are widely conserved in non-pathogenic bacteria as well as pathogenic bacteria (Hacquard et al., 2017), and some non-pathogenic Proteobacteria strains were shown to trigger defense responses in plant leaves likely via PTI pathways (Vogel et al., 2016). On the other hand, diverse microbiota members can suppress PTI triggered by flg22 in roots (Garrido-Oter et al., 2018; Teixeira et al., 2021; Yu et al., 2019), which can facilitate colonization by the root microbiota (Teixeira et al., 2021; Yu et al., 2019). Thus, PTI activation by divergent MAMPs and subsequent PTI modulation by plant-associated bacteria might steer plant responses in a bacterial strain-specific manner, contributing to microbiota assembly in roots. In leaves,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

however, our current understanding of plant responses to individual microbiota members is largely limited to a specific phylum of bacteria, Proteobacteria, and little is known about how plants respond to bacterial strains belonging to other major phyla such as Actinobacteria, Bacteroidetes, and Firmicutes.

When colonized densely and heterogeneously by various bacterial species, plants might not be able to tailor their responses to individual bacterial strains. Yet, it might be possible that different plant-associated bacterial species respond differently to the same microenvironments created by plants. If so, analyzing plant responses alone does not wholly explain bacterial responses during interactions with hosts. The explanation requires directly interrogating bacterial responses *in planta* at the genome-wide scale. *In planta* bacterial omics approaches, such as transcriptomics, are powerful in understanding bacterial gene functions in the plant microbiome and how plants influence bacterial activities (Levy et al., 2018). To date, however, there is a handful of *in planta* bacterial transcriptome studies, all focusing on pathogenic Proteobacteria strains (Chapelle et al., 2015; Lovelace et al., 2018; Nobori et al., 2018, 2020; Yu et al., 2013, 2014). It is, therefore, unknown whether plant-associated bacteria have any common or phylum-specific signatures in the usage of their genomes and what kind of functions are important for their non-pathogenic and sometimes beneficial traits in plants. Furthermore, to our knowledge, no study has analyzed plant and bacterial transcriptome responses at the same time, limiting our ability to build hypotheses about the molecular dialogue between plants and microbiota members.

Here, in monoassociation conditions, we co-profiled the transcriptomes of the model plant *Arabidopsis thaliana* and various bacterial strains isolated from healthy (asymptomatic) plants in nature (hereafter commensal strains), representing all major phyla of the plant microbiota residing in leaves. Commensal strains commonly induced PTI responses in plants, but these differed in intensity. We found examples of both common and strain-specific

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regulation of commensal gene expression in plants. Bacterial genes enriched in plant-associated strains tended to be induced *in planta*. These included genes involved in sulfur, nitrogen, and carbon transport and metabolism, which were induced *in planta* in a strain-specific manner. This suggests that nutrient status differs for different strains in plants, which may affect bacterial fitness and niche separation. We also observed that plants could elicit different transcriptional responses from different bacterial strains without tailoring their own transcriptional reprogramming. This study provides a framework for dissecting plant-microbiota interactions at the strain level using co-transcriptomics and unravels diverse modes of interactions between plants and commensal bacteria.

Results

Co-transcriptome analysis of plants and plant microbiota members

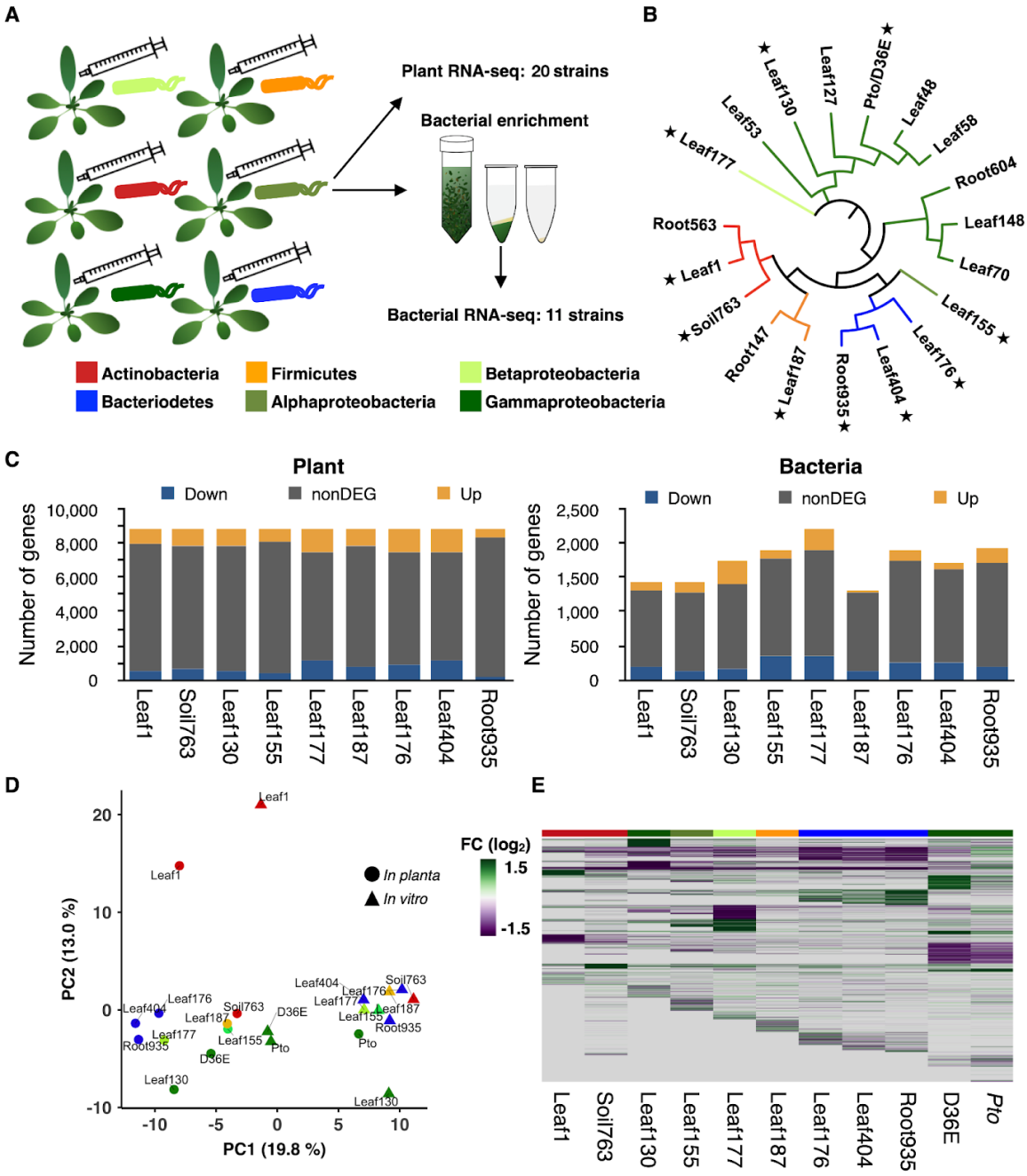
We developed a pipeline to simultaneously investigate the transcriptomes of both plants and bacteria during plant colonization with a single bacterial strain. We monocolonized *A. thaliana* wild-type Col-0 leaves with individual commensal strains by hand-infiltration and profiled transcriptomes of plants and bacteria at 6 hours post-inoculation (hpi) by RNA-seq **(Fig. 1A)**. For *in planta* bacterial RNA-seq, we used a previously developed method with some modifications (see Methods). Briefly, bacterial cells are isolated from plant leaves before extracting RNA, followed by rRNA depletion and RNA-seq (Nobori et al., 2018). For plant and bacterial RNA-seq, respectively, 18 and nine commensal strains were selected covering all major phyla of the plant microbiota **(Fig. 1B and Table 1)**. Three biological replicates from independent experiments were taken for each condition. We used the same strain IDs as in the original study where these bacterial strains were isolated from wild *A. thaliana* plants (leaves and roots) or

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

soil (Bai et al., 2015). A strain ID indicates the original compartment from which the strain was isolated, but many root/soil isolates can also colonize the shoot, indicating extensive niche overlap (Bai et al., 2015).

Nine commensal strains were used for co-transcriptome analysis **(Fig. 1C)**. These strains could colonize to various degrees in the leaf endosphere when inoculated on the leaf surface **(Fig. S1)**. For plants, we compared gene expression changes between bacteria-inoculated plants and water-inoculated plants **(Fig. 1C, left)**. For bacteria, we compared expression changes between *in planta* and *in vitro* (rich media) conditions **(Fig. 1C, right; Fig. S2A)**. We also included previously generated *in planta*/*in vitro* transcriptome data of the virulent pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (*Pto*) and its avirulent mutant D36E (36 type III effectors are depleted) (Nobori et al., 2018) in the analysis. To directly compare bacterial gene expression patterns among phylogenetically diverse bacterial strains, genes of different strains were grouped based on sequence homology, resulting in 6,823 orthologous groups (OGs) **(Fig. S2B)**. Of these OGs, only 454 OGs were shared among all strains **(Data S1)**, indicating that the commensal strains used in this study possess highly diverse gene sets. Principal component analysis using the shared OGs showed that all commensal strains except for Leaf1, showed relatively similar gene expression patterns both *in planta* and under *in vitro* conditions **(Fig. 1D)**. Interestingly, while the gene expression pattern of the avirulent pathogen D36E *in planta* was similar to those of commensals, the virulent *Pto* showed a distinct pattern, i.e., *Pto in planta* resembled commensals *in vitro* **(Fig. 1D)**. Thus, the commensal strains used in this study have highly different genomes, but *in planta* expression of shared genes is similar to each other and highly distinct from that of a virulent pathogen.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



142

143 **Fig. 1: Co-transcriptomics of plants and bacteria (A)** Experimental scheme.
144 Individual bacterial strains were syringe-infiltrated into leaves of *A. thaliana*.
145 Leaves were sampled at 6 h post-inoculation. Total RNA was extracted for plant
146 RNA-seq. For bacterial RNA-seq, bacterial cells were isolated from plant leaves
147 before extracting RNA using a method previously reported (Nobori et al., 2018).
148 **(B)** Bacterial strains used in this study. Stars indicate the strains used for
149 co-transcriptome analysis. Detailed taxonomic information is shown in Table 1.
150 **(C)** The number of genes differentially regulated ($|\log_2 FC| > 1$; FDR < 0.01;
151 two-tailed Student's t test followed by Storey's q-value) in plant or bacterial

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RNA-seq. Plant: bacteria-inoculated vs. water-inoculated. Bacteria: *in planta* 6 h vs. *in vitro* (rich media). Bacterial strains used for co-transcriptomics are shown. **(D)** Principal component analysis of bacterial gene (orthologous group) expression *in planta* and *in vitro*. **(E)** Gene expression fold changes (*in planta* vs. *in vitro*) of bacteria. The data of *Pto* and D36E are from a previous study (Nobori et al., 2018). Genes not detected or missing in the genome are shown in gray. See **Data S2** for gene expression data. **(A, B, D, and E)** The taxonomic affiliation (phylum/class level) of each strain is indicated with different colors.

ID	Phylum	Class	Order	Family	Genus	RNA-seq
Leaf1	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	P, B
Root563	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Janibacter	P
Soil763	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Arthrobacter	P, B
Leaf176	Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	Pedobacter	P, B
Leaf404	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Chryseobacterium	P, B
Root935	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	P, B
Leaf187	Firmicutes	Bacilli	Bacillales	Exiguobacterium	Exiguobacterium	P, B
Root147	Firmicutes	Bacilli	Bacillales	Bacillaceae		P
Leaf155	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Agrobacterium	P, B
Leaf177	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia	P, B
Leaf53	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Erwinia	P
Leaf70	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Xanthomonas	P
Leaf148	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Xanthomonas	P
Root604	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae		P
Leaf48	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	P
Leaf58	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	P
Leaf127	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	P
Leaf130	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	P, B
Pto	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	P, B
D36E	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	P, B

Table 1: List of bacterial strains used in this study RNA-seq data (P, plant; B, bacteria) obtained in this study are indicated.

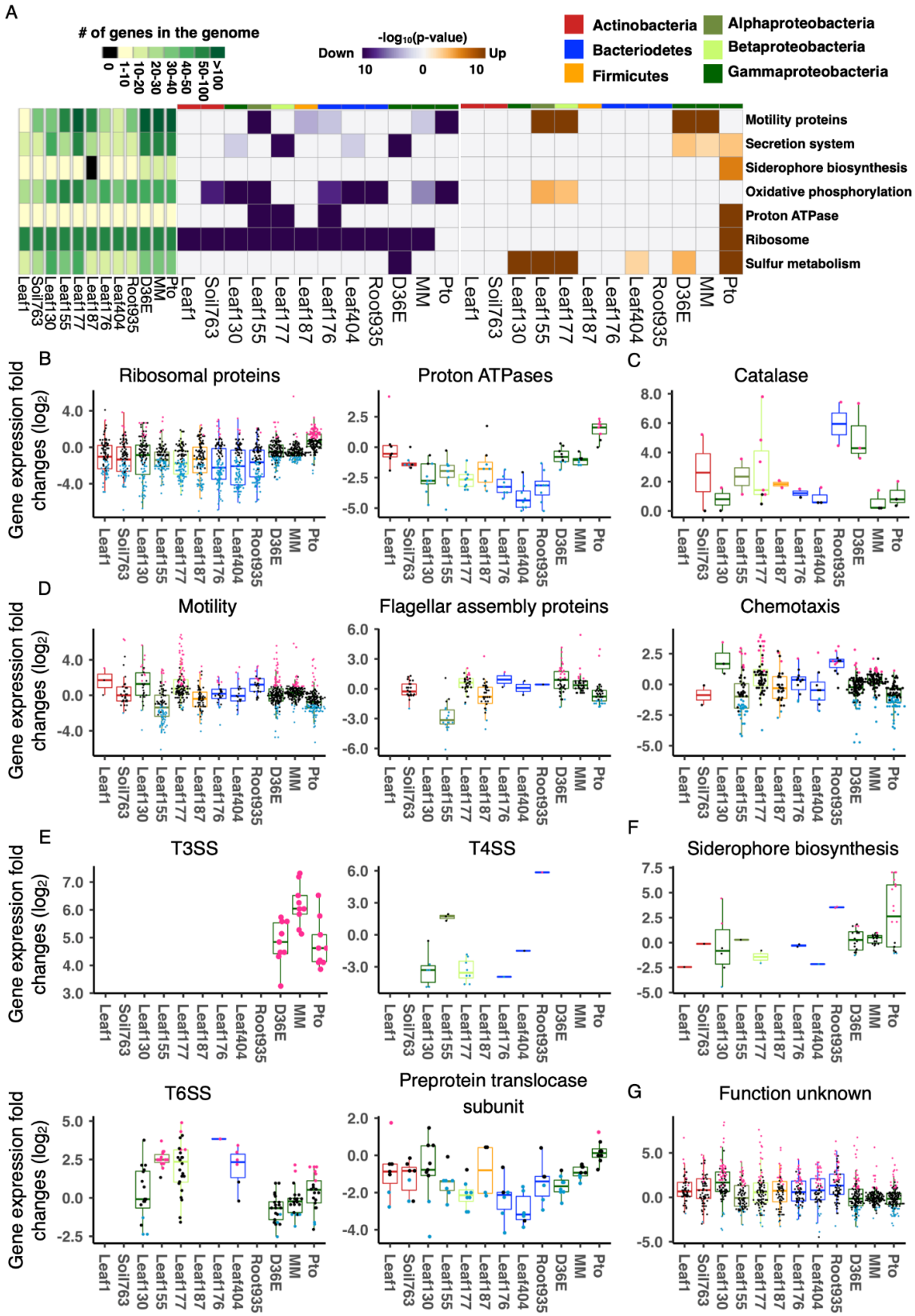
Conserved and strain-specific regulation of commensal functions *in planta*

The high variability in bacterial genomes complicates a genome-wide comparison of bacterial gene expression among phylogenetically diverse strains **(Fig. 1E)**. To overcome this problem, we sought to compare the regulation of bacterial functions rather than of individual genes. For each strain, we performed

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functional enrichment analysis on genes significantly up- or down-regulated *in planta* using KEGG ontologies (KOs) assigned to individual OGs (see Method for details). Then, enrichment scores (*p* values) for individual KOs were summarized for all the strains **(Fig. 2A and Fig. S3A)**. Data from *Pto* grown in a minimal medium were included to determine the effect of nutrient availability on gene expression changes in a pathogen. A clear pattern distinguishing virulent and avirulent strains was seen in the process “ribosome” **(Fig. 2A and 2B)**. Genes encoding ribosomal subunits were significantly suppressed *in planta* in all the commensal strains tested and the avirulent pathogen D36E, while these genes were induced in the virulent pathogen *Pto* **(Fig. 2A and 2B)**. The population density of *Pto* remains unchanged at this time point (Nobori et al., 2018), indicating that the gene expression changes are not the consequence of bacterial growth *in planta*. This process was also suppressed in *Pto* grown in a minimal medium compared with a rich medium **(Fig. 2A and 2B)**. Similarly, genes encoding proton ATPases, which are involved in energy production, were induced in *Pto in planta*, but suppressed or not altered in the commensal strains **(Fig. 2A and 2B)**. Together, these results suggest that commensal strains are metabolically less active *in planta* compared with a virulent pathogen, which is consistent with the notion that commensal strains do not proliferate in plants to the same extent as virulent pathogens. Since D36E is a mutant of *Pto* lacking PTI-suppressing effector molecules, PTI is likely responsible for suppressing bacterial metabolism *in planta*, at least in this strain. Catalase genes were commonly induced in most commensals and D36E at varying degrees but not in *Pto* **(Fig. 2C)**, suggesting that commensals are responding to plant ROS burst, a characteristic PTI response.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 2: Conserved and strain-specific regulation of bacterial functions in plants (A) KEGG ontology (KO) terms enriched in genes that are significantly up- (orange) or down (purple)-regulated *in planta* compared with *in vitro* (rich media). The heatmaps indicate $-\log_{10}$ p-value (FDR corrected by Benjamini–Hochberg method). A KO can be both significantly up and down-regulated in the same strain. The left green panel shows the number of genes involved in each KO term. The top color bars indicate the taxonomic affiliation (phylum/class level) of each strain. See **Fig. S2** for a more comprehensive list of KOs. **(B-G)** Expression fold changes (*in planta* vs. *in vitro*) of genes involved in (B) general metabolism, **(C)** catalase metabolic pathway, **(D)** motility, **(E)** secretion systems, **(F)** siderophore biosynthesis, and **(G)** unknown functions. T3SS, type III secretion system. T4SS, type IV secretion system. T6SS, type VI secretion system. MM, *Pto* grown in a minimal medium. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

Genes involved in bacterial motility were differentially regulated among bacteria in plants. Many of these genes were suppressed in *Pto in planta* but induced in D36E **(Fig. 2A and 2D)**. Leaf177, a *Burkholderia* (Betaproteobacteria) strain, showed a similar pattern to D36E **(Fig. 2A and 2D)**. However, the *Rhizobia* (Alphaproteobacteria) Leaf155 more closely resembled virulent *Pto* – a majority of the genes were suppressed *in planta*, whereas some genes were significantly induced **(Fig. 2D)**. Motility-related genes can be classed into two major functional categories, flagellar assembly, and chemotaxis. Genes encoding flagellar assembly proteins were globally suppressed *in planta* in Leaf155 as in *Pto*, and many Leaf177 chemotaxis-related genes were induced *in planta* in contrast to *Pto* **(Fig. 2D)**. Thus, physiological processes are differentially regulated among different plant-associated commensal bacteria strains, with some species even exhibiting similarity to a virulent pathogen.

The type III secretion system, an essential component of the virulence of bacterial pathogens, including *Pto* (Toruño et al., 2016), was strongly induced in *Pto* and D36E, but these genes were absent in the commensals **(Fig. 2E)**. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

type IV secretion system is involved in multiple processes such as translocating proteins and DNA into other cells and bacterial motility (Costa et al., 2015). This process was globally suppressed in Leaf130 and Leaf177, but not in Leaf155 **(Fig. 2E)**. The type VI secretion system is an injection machine involved in bacteria-host and bacteria-bacteria interactions (Russell et al., 2014). This machinery was globally induced in Leaf155 and Leaf404 and partially induced in Leaf177 and *Pto* **(Fig. 2E)**. Lastly, preprotein translocase subunits, which are involved in the bacterial general secretory pathway (Osborne et al., 2005), tended to be suppressed in all commensals and D36E, but not in *Pto* **(Fig. 2E)**.

It has been shown that genes encoding iron-chelating siderophores are strongly induced in *Pto* upon plant infection, and the induction of these genes is blocked by plant immunity to suppress bacterial growth (Nobori et al., 2018) **(Fig. 2F)**. Commensal strains either harbor only a few or completely lack genes encoding siderophores, and most of these genes were not induced *in planta* **(Fig. 2F)**. We speculate that siderophores are not actively used by commensals in leaves and/or that plant immunity suppresses the production of siderophores to control commensal growth. Notably, many genes annotated as “Function unknown” were significantly induced *in planta* in various commensals **(Fig. 2G)**. These functionally unannotated genes induced *in planta* may have unique roles in plant-bacterial interactions.

Phylum and strain-specific gene expression

To compare expression of individual genes between different strains, we conducted comparative transcriptome analysis focusing on specific phyla **(Fig. S4A and S4C)**. This approach allows more comprehensive comparative transcriptome analysis as more genes are shared among closely related strains. We focused on Bacteroidetes and Proteobacteria, in which 1,422 and 1,122 OGs were shared, respectively (far more than the 454 OGs shared among the nine

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

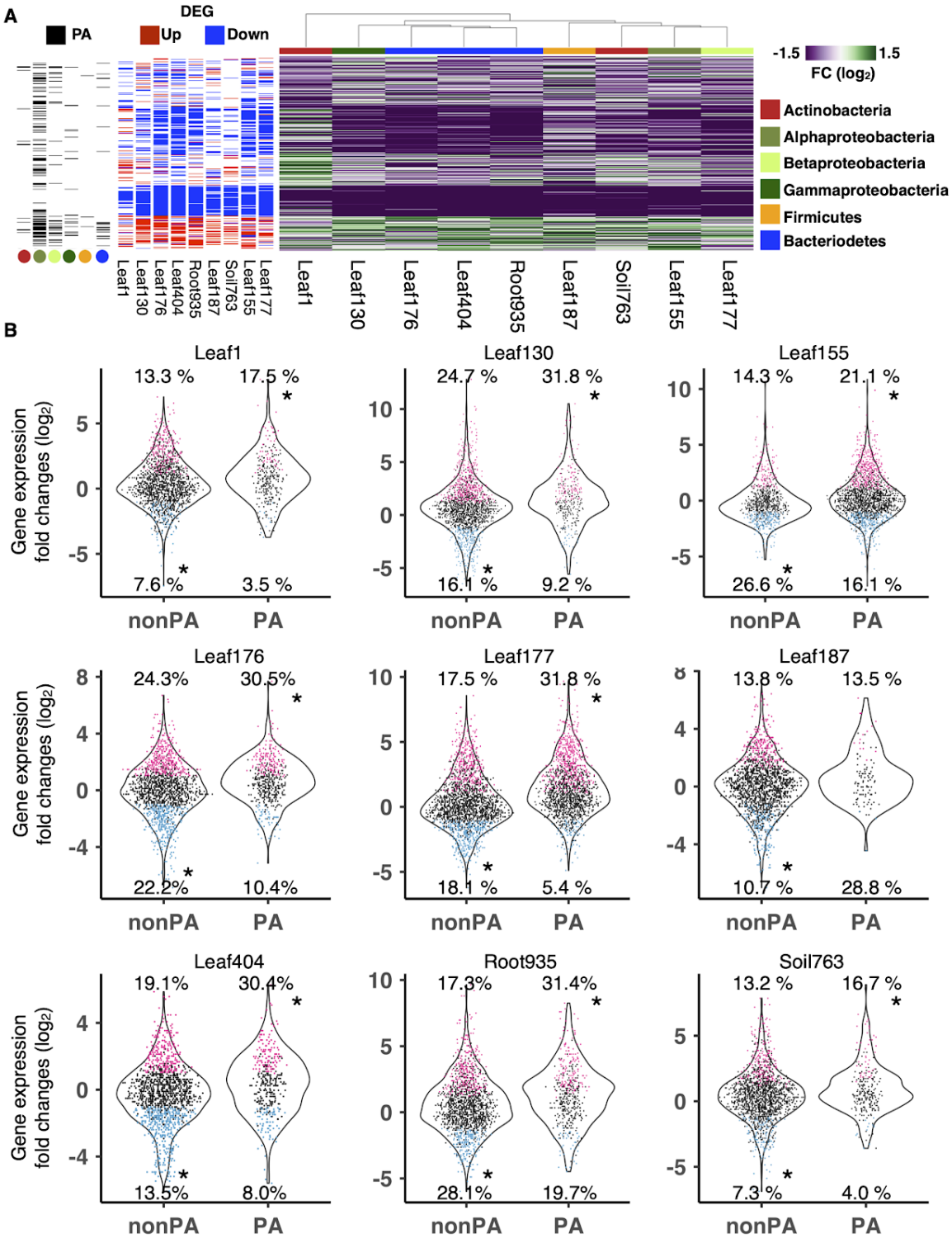
commensals) **(Fig. S2B, S4A, and S4C)**. Overall, many genes were differentially regulated in a single strain **(Fig. S2B, S4A, and S4C)**. In both of these phyla, a larger number of genes were commonly suppressed among the three strains in plants than commonly induced **(Fig. S4B, S4D, and S5)**. Clusters of genes commonly suppressed *in planta* (clusters 7 & 8 in Bacteroidetes and clusters 1 & 4 in Proteobacteria) were enriched with "ribosome"-related genes **(Fig. S4E)**. "Transporters" were enriched in multiple clusters with various expression patterns **(Fig. S4F and S6A)**, suggesting that transporters can be separated into sub-groups based on regulation in plants. Also, genes annotated as part of a "two-component system" showed strain-specific expression patterns **(Fig. S4G and S6B)**. Taken together, our intraphylum analysis reveals that even relatively closely related commensal strains respond differently *in planta* at the transcriptional level.

In planta bacterial transcriptomics illuminates bacterial adaptation to the leaf environment

Various bacterial functions were differentially regulated in plants in a strain-specific manner. Is such functional regulation relevant for bacterial fitness in plants? Comparative genomics is one way to infer bacterial functions associated with adaptation to the plant environment. A previous study compared the genomes of nearly 4,000 plant-associated and non-plant-associated bacterial strains and defined "plant-associated (PA) genes" that are significantly enriched in plant-associated strains (Levy et al., 2017). We analyzed how PA genes are regulated in plants in our transcriptome data. When analyzing the genes shared among nine commensal strains, we observed that genes induced *in planta* tended to be enriched with PA genes, whereas genes suppressed *in planta* tended to be enriched with nonPA genes **(Fig. 3A)**. Remarkably, PA and nonPA genes were significantly enriched with plant-induced and plant-suppressed

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

281 genes, respectively, for all the commensals, except for the *Firmicutes* strain
282 Leaf187 **(Fig. 3B)**. Therefore, our data suggest that bacterial genes associated
283 with adaptation to the plant environment were indeed activated *in planta*.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 3: Genes enriched in plant-associated bacteria are induced *in planta* **(A)** (Right panel) Bacterial gene expression fold changes (FC) in plants compared with *in vitro* (rich media). (Middle panel) Genes differentially expressed *in planta* compared with *in vitro* ($|\log_2 FC| > 1$; FDR < 0.01; two-tailed Student's t test followed by Storey's q-value). (Left panel) Genes previously shown to be "plant-associated" (Levy et al., 2017) are shown as black. The bar and dots indicate the taxonomic affiliation (phylum/class level) of each strain. **(B)** Boxplots showing expression changes of plant-associated (PA) and non-plant-associated (nonPA) genes between *in planta* and *in vitro*. Each dot represents a gene. Genes significantly up- or downregulated *in planta* are colored in red and blue, respectively. Asterisks indicate statistically significant enrichment (FDR < 0.05; Hypergeometric test corrected by Benjamini-Hochberg method) of up or down-regulated genes in the PA or nonPA category. The proportion of genes up- or downregulated are shown. For the full expression data with the orthologous group, KEGG annotation, DEG, and PA information, see **Data S3**.

We then performed KO enrichment analysis for PA genes induced in plants and nonPA genes suppressed in plants. Ribosome-related genes were conserved among all strains (and are thus nonPA genes) and were generally suppressed in plants **(Fig. 4A)**, which may be a strategy by which plants control bacterial growth. Glycan degradation genes were highly plant-associated and induced in Bacteroidetes strains Leaf176 and Root935 **(Fig. 4A and 4B)**. Among such genes were homologs of beta-galactosidase, alpha-L-fucosidase, and glucosylceramidase, which can degrade plant cell wall components. Thus, Leaf176 and Root935 may have evolved the ability to degrade the plant cell wall enabling the establishment of favorable niches during plant colonization.

"Sulfur metabolism"-related genes were classified as PA genes and were induced *in planta* in the three *Proteobacteria* strains **(Fig. 4A and 4C)**. "ABC transporters" was another PA function that was induced in *Proteobacteria* strains *in planta* **(Fig. 4A and 4C)**. Genes involved in both categories were sulfur uptake transporters **(Fig. 4C)**. A previous proteomics study showed that the expression of proteins involved in sulfur metabolism and uptake was induced on the leaf surface in two commensal Proteobacteria, *Sphingomonas melonis* and

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

316 *Methylobacterium extorquens* (Müller et al., 2016). These results suggest that
317 sulfate acquisition is important for the adaptation of commensal Proteobacteria to
318 the plant environment. On the other hand, sulfur metabolism-related genes were
319 not found to be PA genes and were suppressed *in planta* in the Bacteroidetes
320 strains Leaf176 and Leaf404 **(Fig. 4A and 4C)**. The number of genes predicted
321 to be involved in sulfur metabolism was lower in Bacteroidetes strains than in
322 Proteobacteria strains **(Fig. 4A)**. This may indicate that Bacteroidetes strains are
323 less reliant on sulfur acquisition during plant colonization.

324 Urea transporters are other ABC transporters that are PA genes and were
325 induced specifically in some Proteobacteria strains *in planta* **(Fig. 4D)**. Also,
326 genes encoding ureases, which hydrolyze urea in the bacterial cytoplasm, were
327 induced in some Proteobacteria strains **(Fig. 4D)**. It has been shown that
328 *Yersinia enterocolitica*, a Gammaproteobacteria strain, can use urea as a
329 nitrogen source (Young et al., 1996). These results suggest that Proteobacteria
330 (especially Leaf177 and D36E) might require urea as a nitrogen source in the
331 plant apoplast. Also, genes involved in the nitrate transport system were induced
332 in Leaf130, Leaf155, and D36E *in planta*, but not in *Pto* **(Fig. S7)**, suggesting that
333 some commensal Proteobacteria strains activate nitrogen acquisition systems in
334 plants. Similarly, ribose transporters and glycerol transporters are PA genes and
335 were commonly induced *in planta* in commensal or avirulent Proteobacteria
336 strains (Leaf155, Leaf177, and D36E) but not in the virulent *Pto* **(Fig. 4F)**.
337 Moreover, arabinose and xylose transporters (both monosaccharide transporters)
338 were induced in Leaf177 and D36E *in planta*, but not in *Pto* **(Fig. S7)**. Thus,
339 these Proteobacteria strains may use various types of sugars as carbon sources
340 in plants. The induction of urea and sugar acquisition systems may indicate that
341 commensal bacteria activate nutrient starvation responses in the leaf apoplast.
342 We speculate that plants may sequester nitrogen and carbon sources from the
343 apoplast to limit the growth of commensal and avirulent pathogenic bacteria. In
344 line with this hypothesis, the high-affinity plant urea transporter *AtDUR3* was

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

induced upon inoculation with many commensal strains while suppressed by the virulent *Pto* **(Fig. 4E)**. A previous study showed that plants sequester extracellular sugars by activating the sugar influx transporter *AtSTP13* via the PTI pathway (Yamada et al., 2016). Indeed, our plant transcriptome data showed that *AtSTP13* is induced by the commensals as well as D36E and *Pto* **(Fig. 4G)**. On the other hand, *Pto* can induce plant sugar efflux transporters (Chen et al., 2010), which might increase sugar availability in the apoplast and explain why *Pto* did not activate its sugar transporters in plants. A non-mutually exclusive possibility is that the virulent *Pto* switches its metabolic preference to other substrates during successful infection in plants. In summary, we revealed bacterial phylum/strain-specific gene repertoires and gene regulation, which may be actively controlled by plants and drive bacterial niche separation *in planta*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

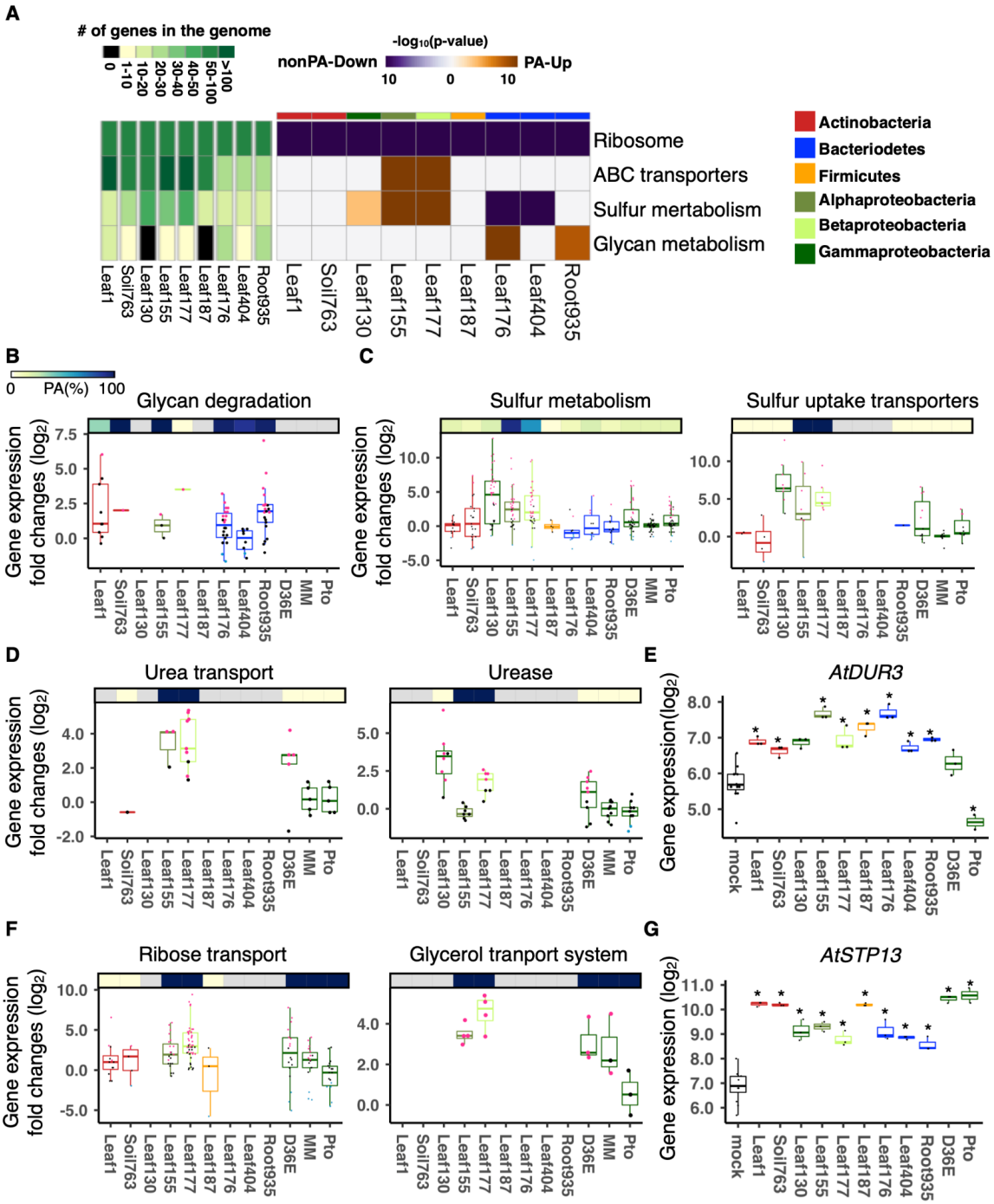


Fig. 4: Nutrient acquisition systems are associated with bacterial adaptation to the plant environment in a strain-specific manner (A) KEGG enrichment analysis of genes that are plant-associated (PA) and significantly induced *in planta* compared with *in vitro* (rich media) (orange) and genes that are nonPA and significantly suppressed *in planta* compared with *in vitro* (purple). The left panel shows the number of genes involved in each KEGG term. **(B-D, F)** Expression fold changes (*in planta* vs. *in vitro*) of genes with different functions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The top bars indicate the ratio of PA genes in each strain. All individual data points (genes) are overlaid to the box plots with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG). **(E and G)** Expression of the plant gene **(E)** *AtDUR3* (urea transporter) and **(G)** *AtSTP13* (sugar transporter) based on the RNA-seq data. Asterisks indicate statistically significant difference ($|\log_2 FC| > 1$; FDR < 0.01; two-tailed Student's t test followed by Storey's q-value) compared with the mock (water-inoculated) condition. In the box plots, boxes display the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. For the full expression data with the orthologous group, KEGG annotation, DEG, and PA information, see **Data S3**.

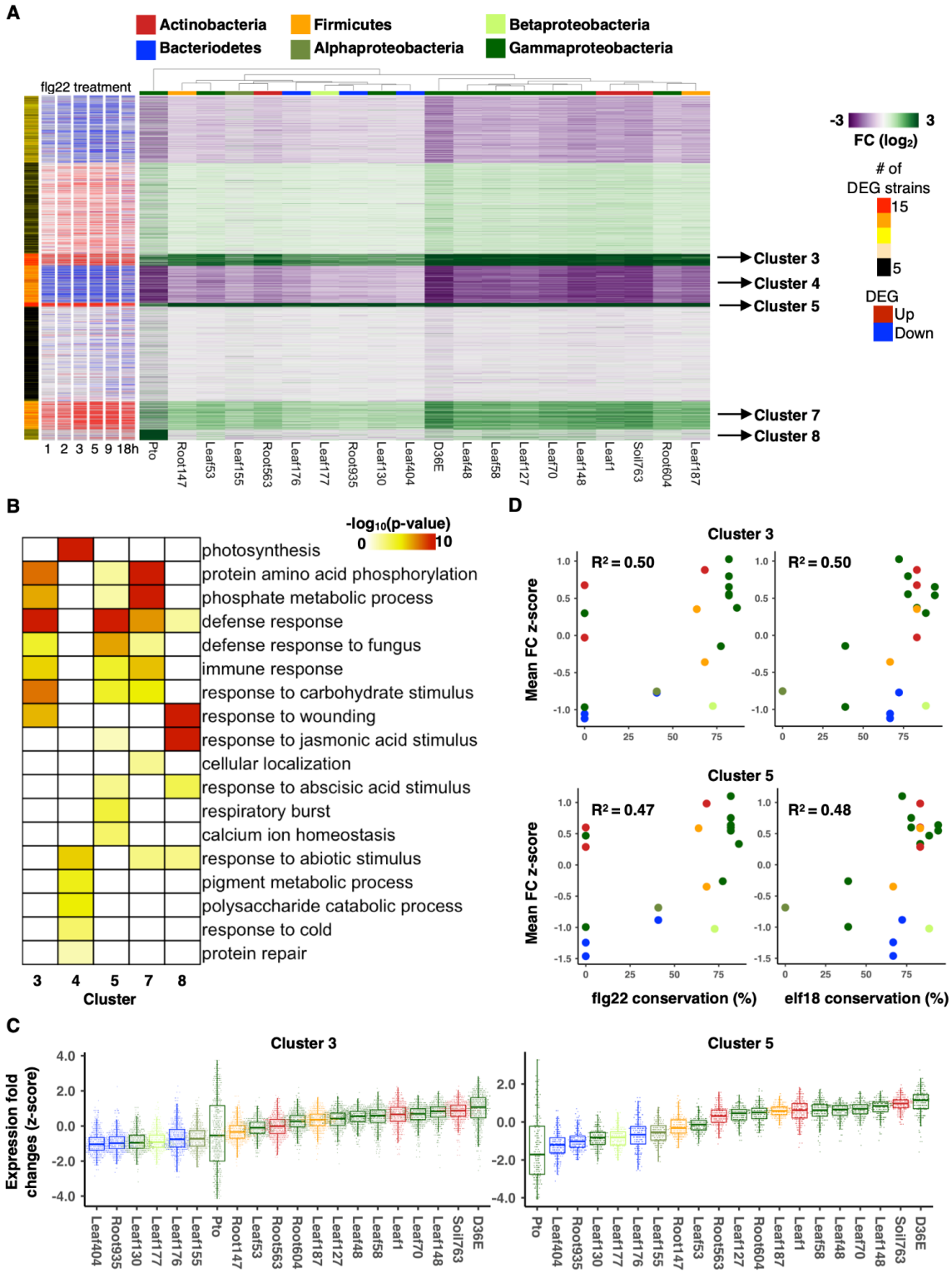
Commensals activate plant PTI in a strain-specific manner

As commensal strains showed differing responses in plants, indicating strain specificity in the interactions of plants with bacteria, we investigated genome-wide plant responses to individual commensals by RNA-seq. In addition to the nine strains used for the bacterial transcriptome analysis, we added nine more commensal strains to enrich our dataset. Global gene expression changes (bacteria-inoculated vs. water-inoculated) were similar among all commensal strains as well as D36E at 6 hpi **(Fig. 5A)**. Interestingly, plant gene expression changes triggered by commensals overlapped markedly with responses to flg22 (Hillmer et al., 2017), a potent PTI inducer. PTI-inducible genes accounted for clusters of genes commonly and strongly induced by most of the commensals (clusters 3 & 5 in **Fig. 5A**). GO enrichment analysis showed that these clusters are enriched with genes related to defense responses (**Fig. 5B**). Thus, commensal strains, when infiltrated into plant leaves, can induce MTI.

The degree of PTI induction varied among strains in a manner that is partly determined by phylogeny: Gammaproteobacteria and Actinobacteria strains induced stronger PTI than Bacteroidetes strains **(Fig. 5C)**. We then investigated the amino acid sequences of the major MAMPs flg22 and elf18 across the different strains. Intriguingly, strains with flg22 and elf18 sequences similar to those known to be particularly potent PTI inducers (Felix et al., 1999; Kunze et al., 2004) tended to elicit strong PTI induction (gene expression fold changes in

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

397 clusters 3 or 5) **(Fig. 5D)**. Thus, sequence variation in these MAMPs may partly
398 determine the degree of PTI induced by some of these commensal strains.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 5: Plant transcriptome responses to phylogenetically diverse commensals (A) (Green/purple heatmap) Gene expression fold changes (FCs) between bacteria-inoculated plants and water-inoculated plants. (Red/blue heatmap) Plant genes significantly induced or suppressed upon flg22 treatment at different time points (Hillmer et al., 2017). The number of strains causing differential gene expression ($|\log_2 FC| > 1$; FDR < 0.01; two-tailed Student's t test followed by Storey's q-value) are indicated in the sidebar (# of DEG strains). DEG, differentially expressed gene. Genes were clustered by k-mean clustering (k = 8). The bars on the heatmaps indicate the taxonomic affiliation (phylum/class level) of each strain. See **Data S4** for gene expression data. **(B)** Gene ontology enrichment analysis for genes in clusters 3, 4, 5, 7, and 8 of **(A)**. $-\log_{10}$ p-values (FDR corrected by Benjamini-Hochberg method) were shown. **(C)** Expression fold changes (FC; z-score) of genes in clusters 3 and 5. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. **(D)** Relationships between amino acid (AA) sequence conservation of flg22 or elf18 and normalized expression FCs of genes in clusters 3 and 5. AA sequence conservation of flg22 and elf18 compared with the canonical sequences known to induce strong defense responses in plants (Elf18: SKEKFERTKPHVNVGTIG. Flg22: QRLSTGSRINSAKDDAAGLQIA). The Pearson correlation coefficients are shown. **(C-D)** The same color code was used for the taxonomic affiliation.

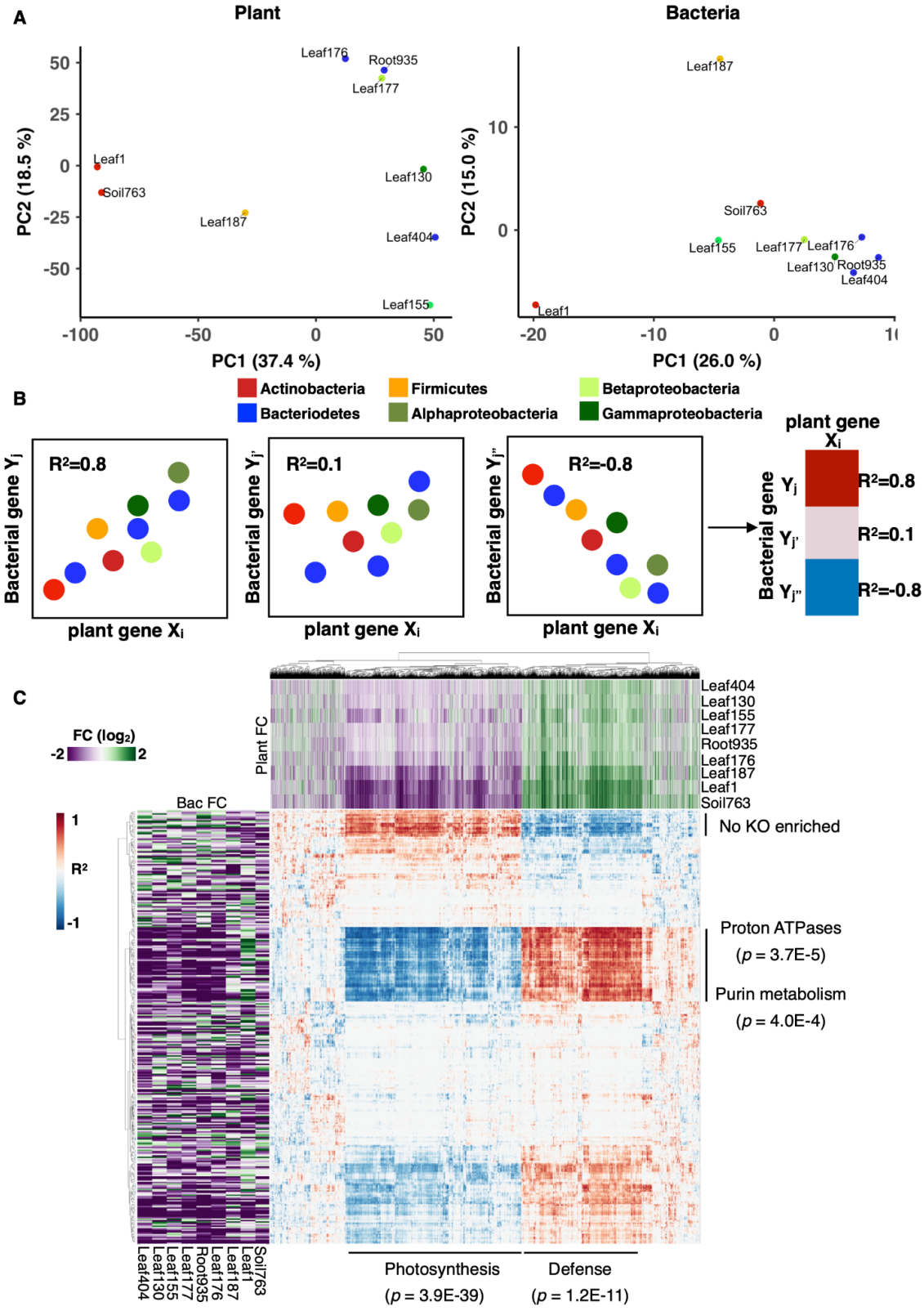
Plant responses do not correlate with bacterial responses in plants

Principal component analysis showed that plant and bacterial transcriptome patterns were incongruent with each other **(Fig. 6A)**. For instance, we observed similarity between the plant transcriptome changes elicited by different Actinobacteria strains (Leaf1 and Soil763), but the strains themselves respond highly differently *in planta* **(Fig. 6A)**. Also, *Bacteroidetes* strains (Leaf176, Leaf404, and Root935) showed similar transcriptional changes in plants, but plant transcriptome changes triggered by these strains differed more **(Fig. 6A)**. To get deeper insights into the relationships between plant and bacterial gene expression, we measured the correlation between gene expression changes of individual plant genes and shared bacterial OGs using co-transcriptome data of nine commensal strains. To prevent a single outlier strain from impacting correlation scores, we took a bootstrapping approach in which correlations were calculated using all the combinations of eight strains as well as all the nine

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strains and then combined (see Methods in detail) **(Fig. 6B and Fig. S8)**. This analysis revealed that the expression of a majority of plant and bacterial genes is not correlated, further indicating that the plant and bacterial responses are largely uncoupled in our dataset **(Fig. 6C)**. For instance, in many cases, commensal strains that triggered similar plant transcriptional responses (e.g., Soil763 and Leaf1; **Fig. 6C**) showed distinct gene expression in plants **(Fig. 6C)**. However, a subset of plant and bacterial genes showed a stronger correlation **(Fig. 6C)**. KO enrichment analysis showed that expression of genes annotated as belonging to the bacterial processes “proton ATPases” and “purine metabolism” positively correlates with plant defense-related genes **(Fig. 6C)**. More specifically, the expression of such bacterial genes was higher when plants showed stronger PTI activation. The biological relevance of this observation remains to be tested. Together, our data indicate that plant and bacterial gene expression can be largely uncoupled at an early stage of interaction.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 6: Plant and bacterial transcriptomes are largely uncoupled (A) Principal component analysis of gene expression fold changes (FCs) of plants (left: bacteria-inoculated vs. water-inoculated) and bacteria (right: *in planta* vs. *in vitro*). Orthologous groups (OGs) of bacterial genes shared among all strains are used for the analysis. **(B)** Schematic diagram of the integration of plant and bacterial RNA-seq data. For each interaction condition, the correlation coefficients between individual plant genes and bacterial OGs were calculated. The correlation coefficient data was corrected by bootstrapping (see Methods and **Fig. S8**) **(C)** A map of correlation coefficients between plant genes and bacterial OGs calculated as described in **(B)**. Rows and columns are bacterial OGs and plant genes, respectively. The top and left heatmaps indicate gene expression FCs of plants and bacteria, respectively. See **Data S5** for the full correlation data. KEGG enrichment analysis was performed for the clusters of plant and bacterial genes with strong correlation.

Discussion

Previous studies of the plant microbiota have suggested that plants assemble bacterial communities and regulate their functions by interacting with commensals in a strain-specific manner. However, there are a limited number of studies that have interrogated the responses of plants and commensal bacteria at a genome-wide scale, and we thus have no comprehensive understanding of the two-way molecular dialogue between plants and microbiota members. Here, we profiled, for the first time, co-transcriptomes of plants and commensal bacteria under monoassociation conditions using diverse strains covering all major phyla of the plant microbiota. Our dataset demonstrated that different commensal strains 1) trigger qualitatively similar yet quantitatively different immune responses in plants and 2) show both common and highly strain-specific responses in plants.

We found that suppression of genes related to general metabolic activity and energy production *in planta* is a common trait among phylogenetically diverse commensals, in marked contrast to a virulent pathogen, which elicited the opposite response **(Fig. 2)**. PTI was commonly induced by the commensal strains **(Fig. 5A)**, suggesting that plant immunity might act to keep commensal

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metabolic activity in check to avoid overgrowth. This notion is in line with a previous finding that commensals can proliferate in an unrestrained manner in the leaf apoplast of plant mutants lacking key immune components under high humidity, which further compromises plant immunity (Chen et al., 2020; Xin et al., 2016). Further transcriptome analysis of commensals in immunocompromised plants and environmental conditions will unravel how different immune pathways tailor their responses to effectively control commensal growth and function.

We provide evidence that bacterial genes enriched in the genomes of plant-adapted strains are induced *in planta* **(Fig. 3A and 3B)**, suggesting that those genes which enable bacteria to thrive in the plant environment are indeed activated in plants. This finding is somewhat in contrast to a previous study, which showed that gene expression of a bacterial pathogen *in planta* does not correlate with fitness scores determined by transposon insertion mutagenesis (Helmann et al., 2019). Importantly, loss-of-function screening with single mutants has limitations in assigning gene function owing to functional redundancy. In this case, a gain-of-function assay is a complementary, albeit also limited, approach. For instance, we previously showed that *in planta* bacterial transcriptome data can predict bacterial genes that contribute to bacterial growth *in planta* when overexpressed (Nobori et al., 2018, 2019). It will be necessary, in the future, to rigorously validate the biological relevance of bacterial transcriptional regulation in plants at the mechanistic level.

We found that processes involved in the uptake of nutrients, such as sulfur, urea, and sugars, were enriched in plant-associated Alpha- and Betaproteobacteria (L155 and L177) and induced *in planta* **(Fig. 4)**. We observed that plants induce transporters that could sequester urea and sugars from the apoplast during interactions with commensals **(Fig. 4E and 4G)**, which potentially affects bacterial nutrient acquisition processes and eventually bacterial fitness in plants. Notably, we also found many other bacterial nutrient transporters to be regulated *in planta* in a strain-specific manner **(Fig. S7)**. The results imply that

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different commensals experience distinct nutrient status in the plant apoplast, which might affect bacterial fitness *in planta*. Our co-transcriptome data sets the stage for investigating whether plants control nutrient availability for particular strains to drive bacterial niche separation in plants and shape the plant microbiota.

We did not observe a strong association between gene expression changes in plants and commensals **(Fig. 6A and 6C)**, implying that different commensals could respond divergently to similar programs of plant gene expression. This seems plausible, given that plants have to deal with complex bacterial communities residing in an area smaller than the plant cell. However, our single time point experiment does not allow us to draw a firm conclusion regarding this idea, as there may be a time lag between plant responses and bacterial responses. In the future, time-course analysis of a larger number of strains combined with plant and bacterial genetics will facilitate the prediction of biologically meaningful links between plant and bacterial responses.

In this study, bacterial cells were syringe-infiltrated into leaves to bypass stomatal entry as different commensals might have different abilities to access the apoplast. Transcriptomes were profiled at 6 hpi, where the population density of even the virulent pathogen *Pto* has not yet increased (Nobori et al., 2018), and thus we assumed that the population density of commensals remained the same at this time point. Therefore, our experimental setup allowed us to characterize strain-specific co-transcriptomes under controlled conditions without the influence of stomatal immunity and differences in the sizes of bacterial populations. It is, however, important to note that transcriptome analysis under more natural conditions would reveal additional layers of plant-microbiota interactions. This requires technological innovations that enable *in planta* transcriptome analysis of bacteria with much smaller populations.

Since plants used in this study were not grown in a strictly sterile condition, we do not exclude the possibility that the pre-existing plant microbiota influenced

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on plant and bacterial responses. However, influence of the pre-existing microbiota on data interpretation was minimized by including the mock control (for plant transcriptomics) and randomizing sampling and taking three independent replicates (for plant and bacterial transcriptomics). We demonstrated that an *in planta* bacterial transcriptome approach can be applied to all major phyla of the plant microbiota, opening a new avenue for *in planta* transcriptome analysis of synthetic communities that are generated by mixing bacterial strains in a desired manner. This *in planta* bacterial metatranscriptome approach can capture more complex traits such as microbe-microbe interactions, which are important to understand the functions of the microbiota as a community.

This study has provided a wealth of information regarding gene regulation of both plants and commensals during monoassociations. In **Fig. S9-12**, we provided additional insights on the regulation of genes related to diverse functions, including biosynthesis/metabolism of various compounds, transporters, and nucleic acid regulation. Notably, we found many genes with unknown functions to be dynamically regulated in commensals during interactions with plants **(Fig. 2G)**. To explore these commensal functions, it will be critical in the future to link bacterial transcriptome responses to bacterial niche preference and reproductive fitness in plants. Our co-transcriptome dataset will provide a robust platform for hypothesis-driven functional investigation of plant and bacterial genes that play critical roles in plant-microbiota interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments:

We thank Paul Schulze-Lefert and Julia Vorholt for providing the commensal strains, Alan Collmer for providing the Pto DC3000 D36E strain, the Max Planck Genome Centre for sequencing support, Dieter Becker, Ryosuke Kanaoka, Shinpei Shimokawa, and Ying Tang for research assistance, and Neysan Donnelly, You Lu, and Natsuki Ohmae for critical comments on the manuscript. This work was supported by the Fundamental Research Funds for the Central Universities (Program No. 2662020ZKPY009) (to K.T.), the Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (to K.T.), the Max Planck Society (to R.G.-O. and K.T.), a German Research Foundation grant (SPP2125) (to R.G.-O. and K.T.), a predoctoral fellowship from the Nakajima Foundation (to T.N.), and a Chinese Scholarship Council PhD stipend (CSC Student ID 201808440401) (to Y.C.).

Author contributions:

T.N. and K.T. designed the research. T.N., Y.C., F.E., and Y.T. performed experiments. T.N., E.D., and R.G.O. performed analysis. T.N. and K.T. wrote the paper with input from all authors.

Data availability:

The RNA sequencing data used in this study are deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession no. GSE150422). Key data and scripts are available at https://github.com/tnobori/co-transcriptomics.

Conflict of interest:

The authors declare no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Plant materials and growth conditions

The *Arabidopsis thaliana* accession Col-0 plants were grown in a chamber at 22°C with a 10-h light period and 60% relative humidity for 24 days and then in another chamber at 22°C with a 12-h light period and 60% relative humidity. For all experiments, 31- to 33-day-old plants were used.

Bacterial strains

Commensal strains were previously isolated from wild *A. thaliana* plants (Bai et al., 2015) (http://www.at-sphere.com/) **(Table 1)**. The *Pto* mutant D36E was previously described (Wei et al., 2015). Bacterial strains were cultured at 20ºC (commensal strains) or 28ºC (*Pto* and D36E) at 200 rpm in liquid 50% TSB medium (Sigma-Aldrich, USA).

Sampling of bacteria *in vitro*

Commensal strains were grown in liquid 50% TSB medium and harvested at the late log phase. 0.1 volume of the stop buffer (95% EtOH, 5% Phenol) was added to bacterial cultures before centrifuging to collect bacterial cells. Target OD_{600} for each strain (the late log phase): Leaf1 = 0.7, Leaf130 = 1.2, Leaf155 = 0.5, Leaf176 = 0.9, Leaf177 = 0.6, Leaf187 = 0.8, Leaf404 = 0.6, Root935 = 0.8, Soil763 = 1.8.

Bacterial inoculation to plant leaves and sampling

Commensal strains were grown in the liquid 50% TSB medium. For each strain, multiple cultures were prepared with different bacterial densities to ensure that unsaturated cultures were used for experiments. Bacterial cells were harvested by centrifugation, washed twice with sterile water, and resuspended in sterile

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

609 water to OD_{600} of 0.5. Plants grown in pots were randomized before bacterial
610 inoculation. For bacterial RNA-seq, 80–100 *A. thaliana* leaves (four fully
611 expanded leaves per plant) were syringe-inoculated with bacterial suspensions
612 using a needleless syringe. For plant RNA-seq, approximately six leaves (two
613 fully expanded leaves per plant) were treated. Mock control (water infiltration)
614 was included in every plant RNA-seq experiment. Leaves were harvested at 6
615 hours after inoculation. Sampling was conducted for one plant genotype at a
616 time, and approximately 5 min were needed per genotype. Leaves were
617 immediately frozen in liquid nitrogen and stored at -80°C. Three biological
618 replicates from independent experiments were taken for each condition of plant
619 and bacterial RNA-seq.

620 Sequencing library preparation and RNA sequencing

621 *In planta* bacterial transcriptome analysis was conducted as described previously
622 (Nobori et al., 2018) with slight modifications. Briefly, bacteria-infected leaves
623 were coarsely pulverized and released into bacterial isolation buffer (9.5%
624 ethanol, 0.5% phenol, 25 mM TCEP (tris(2-carboxyethyl)phosphine) pH 4.5
625 adjusted with NaOH) at 4ºC, filtered, and centrifuged to isolate bacterial cells
626 from plant cells. The original RNA extraction method based on chemical lysis of
627 bacterial cells by TriFast (Nobori et al., 2018) did not work for some bacterial
628 strains, thus we used FastRNA PRO™ BLUE KIT (MP Biomedicals), which
629 involves mechanical cell lysis. rRNA was depleted to enrich mRNA, and the
630 cDNA libraries were prepared using Ovation Complete Prokaryotic RNA-seq kit
631 1-8 (NuGEN).

632 For plant RNA-seq, RNA was extracted with FastRNA PRO™ KIT with Lysing
633 Matrix E (MP Biomedicals), and DNA was digested with TURBO DNase
634 (Ambion). RNA quality was determined using a 2100 Bioanalyzer (Agilent
635 Technologies, USA). Initially, 500 ng total RNA was used for polyA enrichment
636 with the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

637 Biolabs). Subsequent library preparation was performed with NEBNext® Ultra™
638 II Directional RNA Library Prep Kit for Illumina® (New England Biolabs)
639 according to the manufacturer's instructions.
640 Libraries were immobilized and processed onto a flow cell with cBot (Illumina)
641 and subsequently sequenced on the HiSeq3000 system (Illumina) with 1 x 150
642 bp single reads. Primary data analysis (incl. image analysis, cluster identification,
643 base calling, assignment of quality scores) has been performed with RTA
644 (real-time analysis software; Illumina) installed on the sequencing platform.
645 For bacterial and plant samples, approximately 10 and 30 million reads,
646 respectively, were obtained. Bacterial reads were mapped onto the
647 corresponding bacterial genomes (Bai et al., 2015) using Bowtie2 (Langmead
648 and Salzberg, 2012). Plant reads were mapped onto the *Arabidopsis* genome
649 (TAIR10) using HISAT2 (Kim et al., 2015). Mapped reads were counted with the
650 Python package HTSeq (Anders et al., 2015). The RNA-seq data used in this
651 study are deposited in NCBI Gene Expression Omnibus database (accession no.
652 GSE150422).

653 Raw data
654 Raw RNA-seq count and bacterial gene annotation files are available at
655 https://github.com/tnobori/co-transcriptomics.

656 Data analysis – plant RNA-seq
657 The statistical analysis of the RNA-seq data was performed in the R
658 environment. Genes with average counts < 5 were excluded from the analysis.
659 The count data were TMM-normalized and log-transformed using the function
660 calcNormFactors in the package edgeR (Robinson et al., 2010) and the function
661 voomWithQualityWeights in the package limma (Ritchie et al., 2015),
662 respectively. To each gene, a linear model was fitted by using the function lmFit
663 in the limma package (Ritchie et al., 2015). The eBayes function in the limma

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

664 package was used for variance shrinkage during the calculation of the p-values.
665 The false discovery rate (FDR; the Storey's q-values) was calculated using the
666 qvalue function in the qvalue package (Storey and Tibshirani, 2003). Genes with
667 q-value <0.01 and $|\log_2$ fold change| > 1 were defined as differentially expressed
668 genes. The prcomp function was used for principal component analysis.
669 Heatmaps were created with the pheatmap function in the R environment.
670 Enriched GO terms were identified using the BiNGO plugin for Cytoscape (Maere
671 et al., 2005). Scatter plots and box plots were generated using the R-package
672 ggplot2.

673 Data analysis – bacterial RNA-seq

674 **Bacterial phylogenetic analysis**

675 The bacterial genomes were searched for the bacterial small ribosomal subunit
676 16S rRNA gene using RNAmmer (Lagesen et al., 2007). Next, a multiple
677 sequence alignment was performed using Clustal Omega (Sievers et al., 2011)
678 with default parameters. Finally, we employed FastTree (Price et al., 2010) to
679 build a maximum-likelihood phylogeny based on the gamma time reversible
680 substitution model (GTR). This tree was visualized (Fig. 1B) using the interactive
681 Tree of Life (Letunic and Bork, 2019).

682 **Orthologous gene prediction and KEGG annotation**

683 De novo orthology prediction was performed by using OrthoFinder (Emms and
684 Kelly, 2015) with default parameters on the predicted protein coding sequences
685 from the bacterial genomes. Individual genes were annotated with the KEGG
686 database as a reference (Kanehisa et al., 2014) using the blastkoala webserver
687 (Prokaryotes group) (Kanehisa et al., 2016). Subsequently, orthologous genes
688 were assigned a single KO annotation by majority vote of individually annotated
689 sequences in each group. The genomes of the commensal strains were

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

690 previously reported (Bai et al., 2015) and are available at our GitHub repository
691 (https://github.com/tnobori/co-transcriptomics).

692 **Data normalization and visualization**

693 RNA-seq data were normalized for each strain. After omitting genes with average
694 count < 5, count data was TMM-normalized and log-transformed as described
695 above. Genes with FDR <0.01 (corrected by Benjamini-Hochberg method) and
696 $|\log_2$ fold change$|$ > 1 were defined as differentially expressed genes.
697 Commensal genes were annotated with OGs to integrate gene expression data
698 of different strains. When multiple genes are annotated with the same OG, the
699 mean expression value was taken. Data visualization was performed as
700 described above. UpSet plots were generated in the R environment using the
701 package UpSetR (Conway et al., 2017).

702 KO enrichment analysis (related to **Fig. 2A and Fig. S3A**)

703 A custom KEGG ontology (KO) database was created by taking only functional
704 terms encoded in at least one bacterial genome (downloaded in January 2019).
705 For each strain, a list of KOs was generated by subsetting the corresponding
706 KEGG IDs from the custom KO database **(Data S6)**. KO enrichment test was
707 performed using a hypergeometric test (FDR corrected by Benjamini-Hochberg
708 method). KOs with FDR < 0.01 and containing more than three genes were
709 defined as significantly enriched KOs. An R script and KO databases are
710 available at https://github.com/tnobori/co-transcriptomics.

711 Generating plots of genes with various functions (related to **Fig. 2B-G, Fig. 4B-D**
712 **and F, Fig. S3B and C, Fig. S6, Fig. S7, Fig. S9-12**)

713 Bacterial genes were selected by KEGG pathway annotations or keyword
714 searches from KEGG BRITE annotations. R scripts for this analysis are available
715 at https://github.com/tnobori/co-transcriptomics.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

716 Intersecting plant-associated bacterial genes and differentially regulated genes *in*
717 *planta* (related to **Fig. 3**)
718 In a previous study (Levy et al., 2017), comparative genomics analyses defined
719 “plant-associated (PA) genes” for each phylum/class using multiple statistical
720 tests. The study defined two groups of Actinobacteria (Actinobacteria1 and
721 Actinobacteria2). The Actinobacteria strains used in the present study are all
722 Actinobacteria1). We defined genes that passed at least one statistical test as
723 "PA genes" and the others were defined as nonPA genes. An R script and
724 PA-gene datasets for this analysis are available at
725 https://github.com/tnobori/co-transcriptomics.

726 MAMP conservation analysis (related to **Fig. 5D**)
727 Canonical flg22 and elf18 sequences were blasted against the bacterial
728 genomes using blastp (Camacho et al., 2009) with standard settings. The results
729 of these homology searches were filtered by retaining hits covering at least 90%
730 of the length of the MAMP sequence in the alignment and subsequently
731 retrieving the alignment with the highest percentage identity.

732 Integration of plant and bacterial RNA-seq data (related to **Fig. 6**)
733 Co-transcriptome fold change data (bacteria: *in planta* vs. *in vitro*; plants: bacteria
734 vs. mock) of nine strains were used for this analysis. Plant genes whose
735 expression was significantly changed by at least one strain were used. Pearson’s
736 correlation coefficients between individual plant genes and bacterial OGs were
737 calculated. The same analysis was performed for all the combinations of eight
738 strains (bootstrapping). Among these 8-strain and 9-strain datasets, the weakest
739 correlation coefficient value was used for each combination of a bacterial OG and
740 a plant gene (Fig. S8). An R script and plant/bacterial gene expression datasets
741 for this analysis are available at https://github.com/tnobori/co-transcriptomics.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Determination of bacterial colony forming units (related to Fig. S1)

Bacterial colonization of the leaves was determined following a previous study (Chen et al., 2020) with slight modifications. The Cl_2-gas-sterilized seeds were stratified for 2 days at 4°C, sown on half Murashige & Skoogs (MS, Duchefa-Biochemie, MO255.0050) agar medium with 1% sucrose, and allowed to germinate for 5 days. Seedlings of the same physiological state were transplanted on half MS agar medium and were grown for another 9 days (a total of 2 weeks) prior to inoculation with bacteria. One day before inoculation, bacterial cultures were grown on half TSB for 24 hours at 22°C with 200 rpm shaking. On the day of inoculation, bacterial cells were harvested by centrifugation at 3000 rpm for 15 min, washed twice with sterile water, and then finally suspended in 10 mM $MgCl_2$. The resulting bacterial suspensions were diluted to a final OD_{600} of 0.5 with sterile water and with this, each plate of 2-week-old seedlings was flood-inoculated for 1 min, drained, and allowed to dry for 15 min. Plants were then grown for 3 days and 2-3 leaves of the same physiological state were harvested aseptically and weighed. To quantify bacteria in the endophytic compartment, leaves were surface-sterilized with 75% ethanol for 30 seconds and washed twice with sterile water, and the leaves were homogenized on 10 mM $MgCl^2$ buffer using TissueLyserII (Qiagen) with the frequency of 30 s^{-1} for 5 min. The samples were then serially diluted (10^0 to 10^5) and spread-plated on 0.5x TSB agar medium. Plates were incubated at ambient temperature, colonies were observed and counted for 1-3 d and colony forming units were expressed per mg FW. The total compartment was assayed similarly but without surface sterilization.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

766 Data S1 OG distribution

767 Data S2 Expression of all OGs

768 Data S3 List of PA and nonPA KOs

769 Data S4 Plant RNA-seq data

770 Data S5 Correlation matrix of plant and bacterial transcriptomes

771 Data S6 KO database for each strain

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

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

Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu. Rev. Plant Biol. *60*, 379–406.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Carrión, V.J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., de Hollander, M., Ruiz-Buck, D., Mendes, L.W., van Ijcken, W.F.J., Gomez-Exposito, R., Elsayed, S.S., et al. (2019). Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science *366*, 606–612.

Chapelle, E., Alunni, B., Malfatti, P., Solier, L., Pédron, J., Kraepiel, Y., and Van Gijsegem, F. (2015). A straightforward and reliable method for bacterial in planta transcriptomics: application to the Dickeya dadantii/Arabidopsis thaliana pathosystem. Plant J. *82*, 352–362.

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.-Q., Guo, W.-J., Kim, J.-G., Underwood, W., Chaudhuri, B., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. Nature *468*, 527–532.

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

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat. Rev. Microbiol. *13*, 343–359.

Durán, P., Thiergart, T., Garrido-Oter, R., Agler, M., Kemen, E., Schulze-Lefert, P., and Hacquard, S. (2018). Microbial Interkingdom Interactions in Roots Promote Arabidopsis Survival. Cell *175*, 973–983.e14.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biol. *16*, 1–14.

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J. *18*, 265–276.

Garrido-Oter, R., Nakano, R.T., Dombrowski, N., Ma, K.-W., AgBiome Team, McHardy, A.C., and Schulze-Lefert, P. (2018). Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. Cell Host Microbe *24*, 155–167.e5.

Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S., McHardy, A.C., Dangl, J.L., Knight, R., Ley, R., et al. (2015). Microbiota and Host Nutrition across Plant and Animal Kingdoms. Cell Host & Microbe *17*, 603–616.

Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P. (2017). Interplay Between Innate Immunity and the Plant Microbiota. Annu. Rev. Phytopathol. *55*, 565–589.

Helmann, T.C., Deutschbauer, A.M., and Lindow, S.E. (2019). Genome-wide identification of Pseudomonas syringae genes required for fitness during colonization of the leaf surface and apoplast. Proc. Natl. Acad. Sci. U. S. A. *116*, 18900–18910.

Hillmer, R.A., Tsuda, K., Rallapalli, G., Asai, S., Truman, W., Papke, M.D., Sakakibara, H., Jones, J.D.G., Myers, C.L., and Katagiri, F. (2017). The highly buffered Arabidopsis immune signaling network conceals the functions of its components. PLOS Genetics *13*, e1006639.

Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature *444*, 323–329.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199–D205.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol. *428*, 726–731.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell *16*, 3496–3507.

Kwak, M.-J., Kong, H.G., Choi, K., Kwon, S.-K., Song, J.Y., Lee, J., Lee, P.A., Choi, S.Y., Seo, M., Lee, H.J., et al. (2018). Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat. Biotechnol.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acids Res. *35*, 3100–3108.

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

Lebeis, S.L., Paredes, S.H., Lundberg, D.S., Breakfield, N., Gehring, J., McDonald, M., Malfatti, S., del Rio, T.G., Jones, C.D., Tringe, S.G., et al. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science *349*, 860–864.

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

Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., et al. (2017). Genomic features of bacterial adaptation to plants. Nat. Genet. *50*, 138–150.

Levy, A., Conway, J.M., Dangl, J.L., and Woyke, T. (2018). Elucidating Bacterial Gene Functions in the Plant Microbiome. Cell Host Microbe *24*, 475–485.

Lovelace, A.H., Smith, A., and Kvitko, B.H. (2018). Pattern-Triggered Immunity Alters the Transcriptional Regulation of Virulence-Associated Genes and Induces the Sulfur Starvation Response in Pseudomonas syringae pv. tomato DC3000. Mol. Plant. Microbe. Interact. *31*, 750–765.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448–3449.

Müller, D.B., Schubert, O.T., Röst, H., Aebersold, R., and Vorholt, J.A. (2016). Systems-level Proteomics of Two Ubiquitous Leaf Commensals Reveals Complementary Adaptive Traits for Phyllosphere Colonization. Mol. Cell. Proteomics *15*, 3256–3269.

Nobori, T., Velásquez, A.C., Wu, J., Kvitko, B.H., Kremer, J.M., Wang, Y., He, S.Y., and Tsuda, K. (2018). Transcriptome landscape of a bacterial pathogen under plant immunity. Proc. Natl. Acad. Sci. U. S. A. *115*, E3055–E3064.

Nobori, T., Wang, Y., Wu, J., Stolze, S.C., and Tsuda, Y. (2019). In planta bacterial multi-omics analysis illuminates regulatory principles underlying plant-pathogen interactions. BioRxiv.

Nobori, T., Wang, Y., Wu, J., Stolze, S.C., Tsuda, Y., Finkemeier, I., Nakagami, H., and Tsuda, K. (2020). Multidimensional gene regulatory landscape of a bacterial pathogen in plants. Nat Plants *6*, 883–896.

Osborne, A.R., Rapoport, T.A., and van den Berg, B. (2005). Protein translocation by the Sec61/SecY channel. Annu. Rev. Cell Dev. Biol. *21*, 529–550.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

maximum-likelihood trees for large alignments. PLoS One *5*.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014). Type VI secretion system effectors: poisons with a purpose. Nat. Rev. Microbiol. *12*, 137–148.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. *100*, 9440–9445.

Teixeira, P.J.P.L., Teixeira, P.J.P., Colaianni, N.R., Law, T.F., Conway, J.M., Gilbert, S., Li, H., Salas-González, I., Panda, D., Del Risco, N.M., et al. (2021). Specific modulation of the root immune system by a community of commensal bacteria. Proceedings of the National Academy of Sciences *118*, e2100678118.

Toruño, T.Y., Stergiopoulos, I., and Coaker, G. (2016). Plant-Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. Annu. Rev. Phytopathol. *54*, 419–441.

Vogel, C., Bodenhausen, N., Gruissem, W., and Vorholt, J.A. (2016). The Arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. New Phytologist *212*, 192–207.

Wei, H.-L., Chakravarthy, S., Mathieu, J., Helmann, T.C., Stodghill, P., Swingle, B., Martin, G.B., and Collmer, A. (2015). Pseudomonas syringae pv. tomato DC3000 Type III Secretion Effector Polymutants Reveal an Interplay between HopAD1 and AvrPtoB. Cell Host & Microbe *17*, 752–762.

Xin, X.-F., Nomura, K., Aung, K., Velásquez, A.C., Yao, J., Boutrot, F., Chang, J.H., Zipfel, C., and He, S.Y. (2016). Bacteria establish an aqueous living space in plants crucial for virulence. Nature *539*, 524–529.

Yamada, K., Saijo, Y., Nakagami, H., and Takano, Y. (2016). Regulation of sugar transporter activity for antibacterial defense in Arabidopsis. Science *354*, 1427–1430.

Young, G.M., Amid, D., and Miller, V.L. (1996). A bifunctional urease enhances survival of pathogenic Yersinia enterocolitica and Morganella morganii at low pH. J. Bacteriol. *178*, 6487–6495.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yu, K., Liu, Y., Tichelaar, R., Savant, N., Lagendijk, E., van Kuijk, S.J.L., Stringlis, I.A., van Dijken, A.J.H., Pieterse, C.M.J., Bakker, P.A.H.M., et al. (2019). Rhizosphere-Associated Pseudomonas Suppress Local Root Immune Responses by Gluconic Acid-Mediated Lowering of Environmental pH. Curr. Biol. *29*, 3913–3920.e4.

Yu, X., Lund, S.P., Scott, R. a., Greenwald, J.W., Records, A.H., Nettleton, D., Lindow, S.E., Gross, D.C., and Beattie, G. a. (2013). Transcriptional responses of Pseudomonas syringae to growth in epiphytic versus apoplastic leaf sites. Proc. Natl. Acad. Sci. U. S. A. *110*, E425–E434.

Yu, X., Lund, S.P., Greenwald, J.W., Records, A.H., Scott, R.A., Nettleton, D., Lindow, S.E., Gross, D.C., and Beattie, A. (2014). Transcriptional Analysis of the Global Regulatory Networks Active in Pseudomonas syringae during Leaf Colonization. MBio *5*, 01614–e01683.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary figures

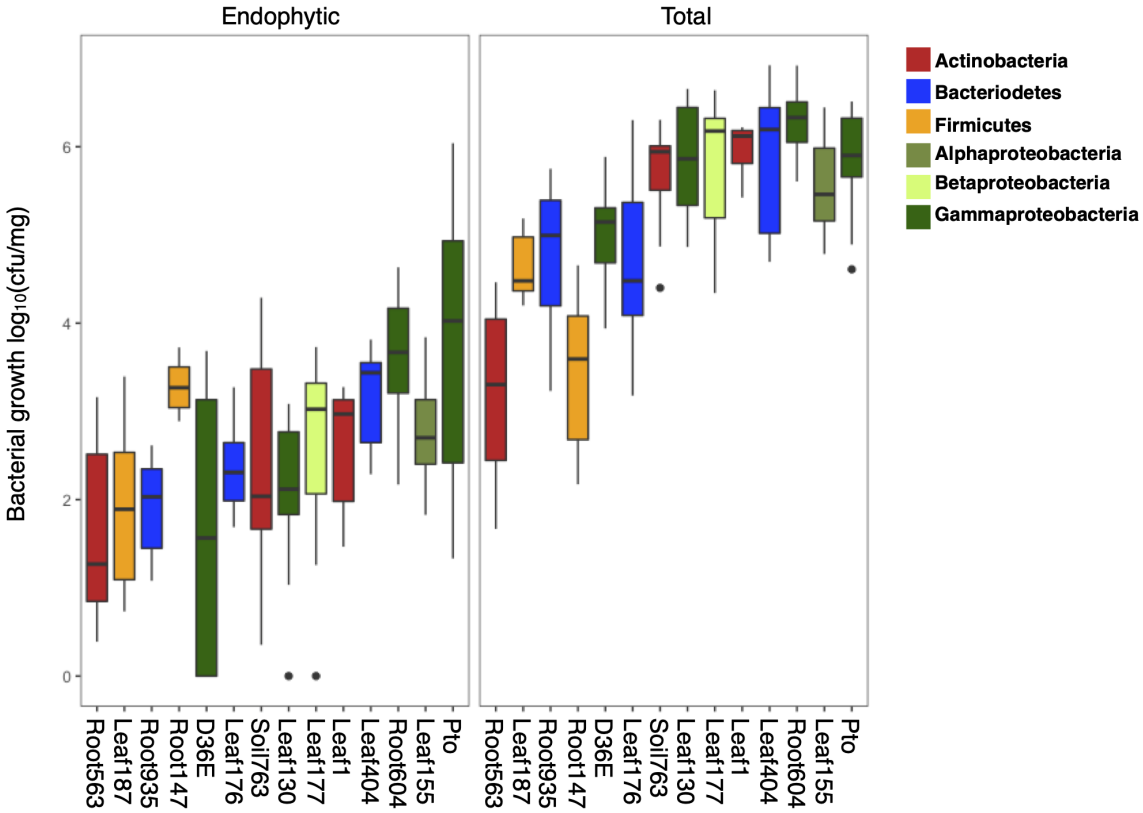


Fig. S1 Bacterial growth in plant leaves Bacteria were flood-inoculated to three-week-old *A. thaliana* Col-0 at OD_{600} = 0.5. Bacterial growth was measured three days after inoculation. Endophytic bacteria were counted after washing and sterilizing the leaf surface, whereas total bacteria were counted without any surface washing and sterilization (see Methods).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

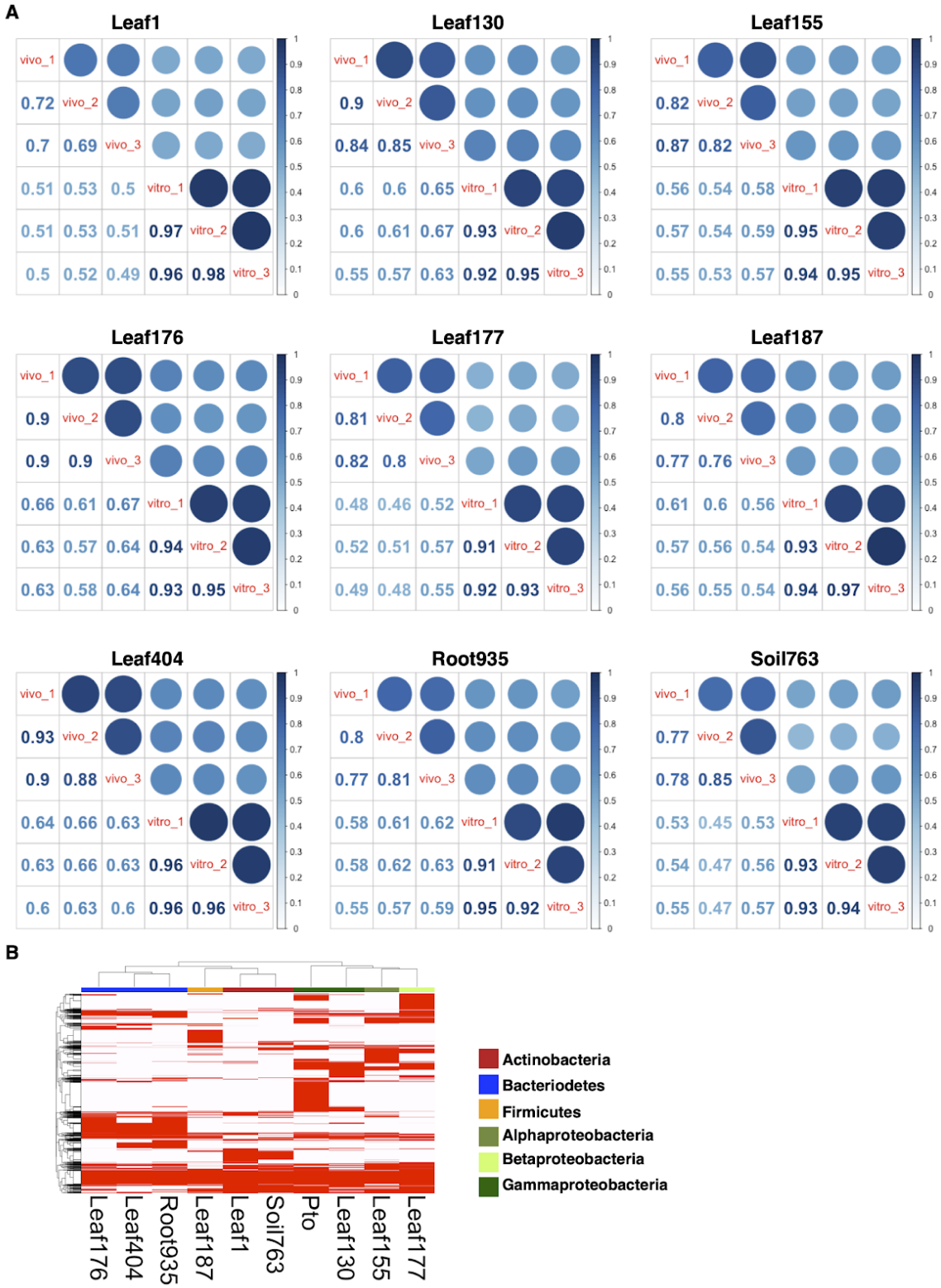


Fig. S2 Transcriptome analysis of bacteria (A) The correlation plot of each replicate of bacterial RNA-seq data for individual strains. vivo: bacteria in plants. vitro: bacteria in rich media. **(B)** Genes that exist are colored in red. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

941 taxonomic affiliation (phylum/class level) of each strain is indicated with different
942 colors. See **Data S1** for the gene presence-absence table.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

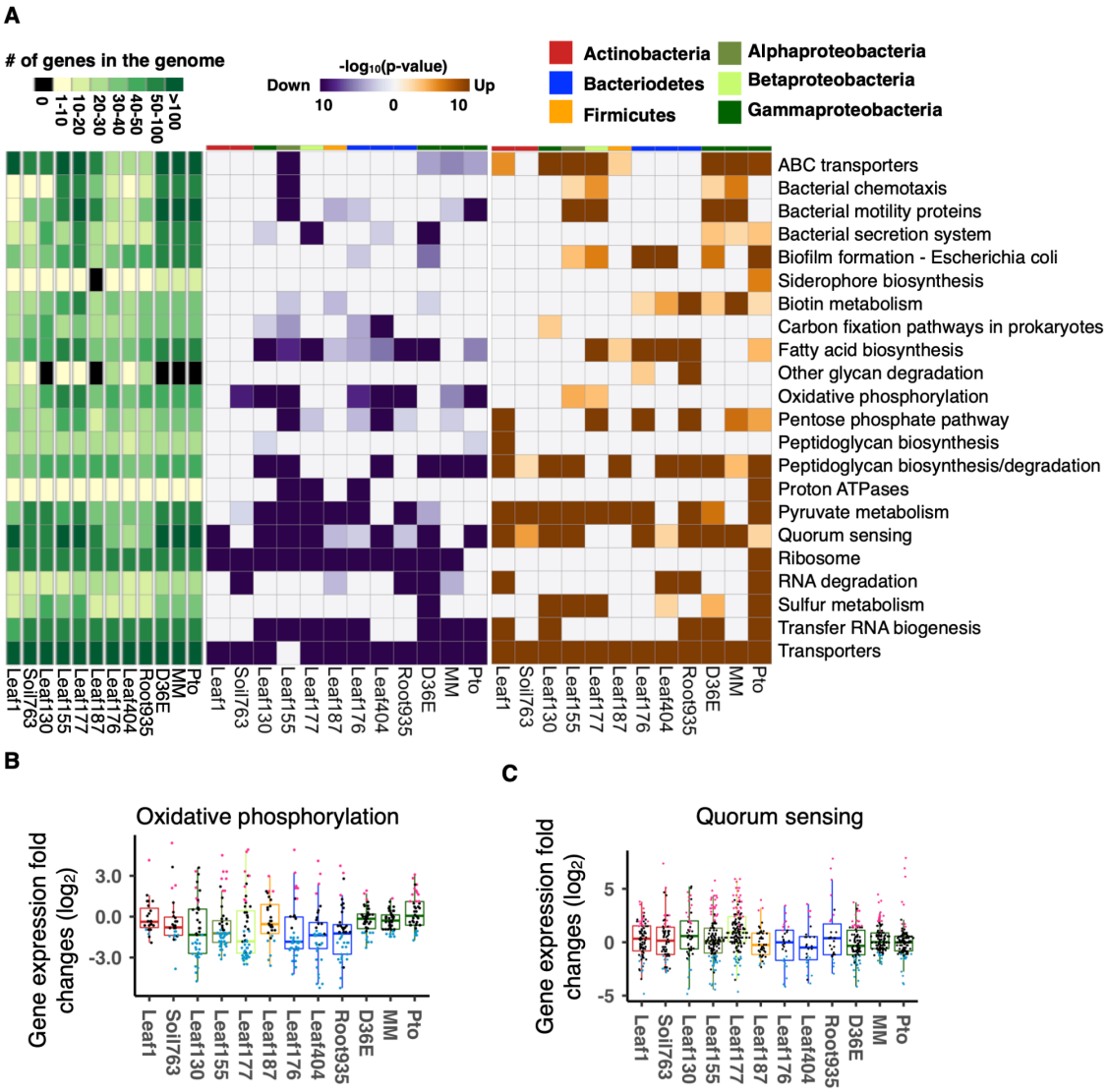
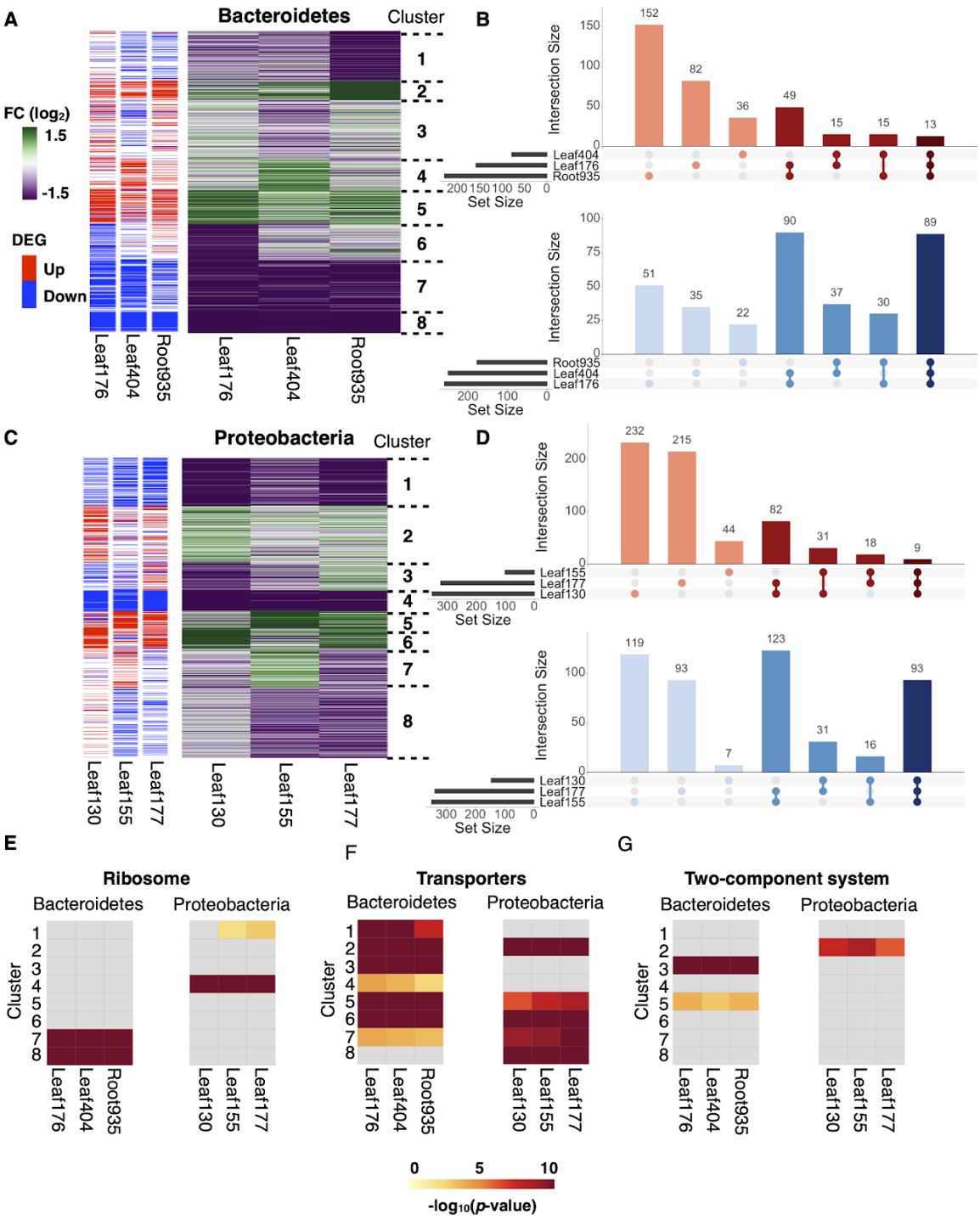


Fig. S3: Conserved and strain-specific regulation of bacterial functions in plants (A) KEGG ontology (KO) terms enriched in genes that are significantly up- (orange) or down (purple)-regulated *in planta* compared with *in vitro* (rich media). The heatmaps indicate -$\log_{10}$ p-values (FDR corrected by Benjamini-Hochberg method). The top color bars indicate the taxonomic affiliation (phylum/class level) of each strain. **(B and C)** Expression fold changes (*in planta* vs. *in vitro*) of genes involved in **(B)** Oxidative phosphorylation and **(C)** Quorum sensing. MM, *Pto* grown in a minimal medium. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



957

Fig. S4: Intraphylum comparative transcriptomics of commensals (A and C) Gene expression fold changes between *in planta* and *in vitro* (rich media) of **(A)** Bacteroidetes and **(C)** Proteobacteria strains. Orthologous groups shared among three strains were used for the analysis. Differentially expressed genes (DEGs) (*in planta* vs. *in vitro*; $|\log_2 FC| > 1$; FDR < 0.01; two-tailed Student's t test followed by Storey's q-value) are indicated in the sidebars. Gene clusters defined by k-mean clustering are shown (k = 8). **(B and D)** UpSet intersection plots of

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

965 DEGs either up- (red) or down (blue)-regulated *in planta* in the **(B)** Bacteroidetes
966 and **(D)** Proteobacteria strains. Intersection size and set size indicate the number
967 of shared DEGs and the number of DEGs in each strain, respectively. **(E-G)**
968 Enrichment analysis of genes with the KEGG ontology terms **(E)** “Ribosome”, **(F)**
969 “Transporters”, and **(G)** “Two-component system”.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

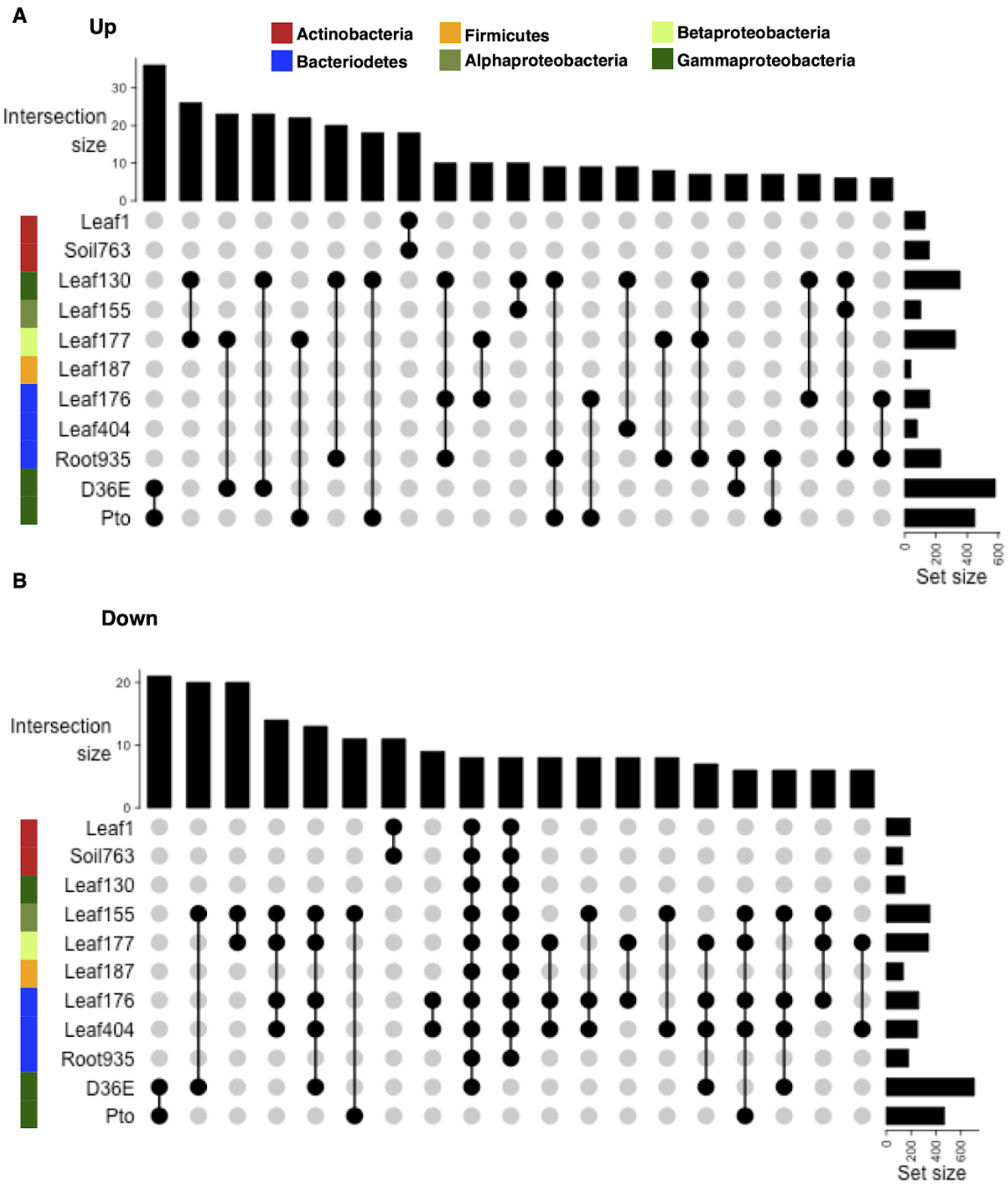
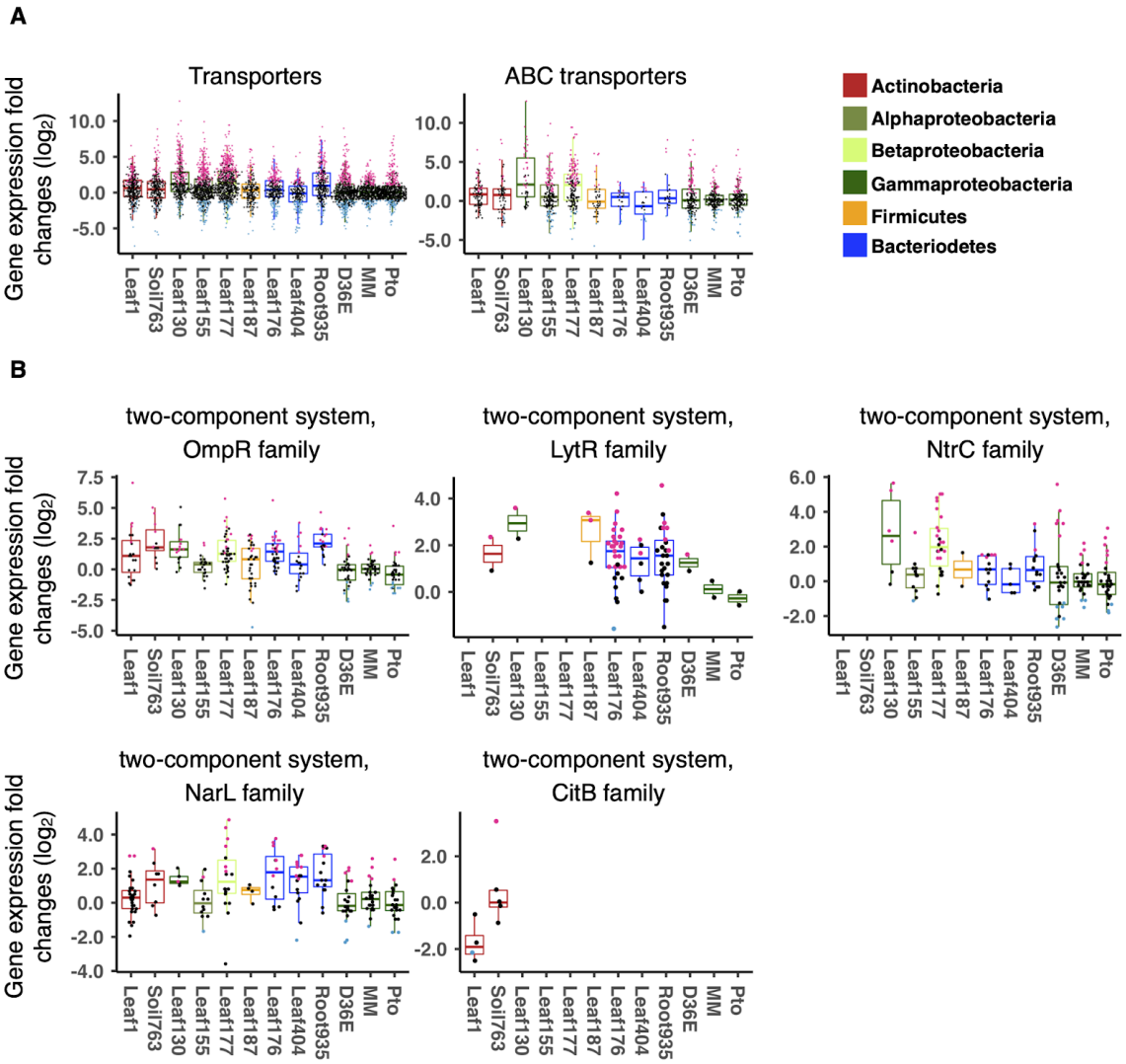


Fig. S5: Bacterial genes differentially regulated in plants UpSet intersection plots of differentially expressed genes (DEGs; $|\log_2 FC| > 1$; FDR < 0.01; two-tailed Student's t test followed by Storey's q-value) either **(A)** up- or **(B)** downregulated *in planta*. Intersection size and set size indicate the number of shared DEGs and the number of DEGs in each strain, respectively. Combinations of more than one strain with intersection size > 5 are shown. The color sidebars indicate the taxonomic affiliation.

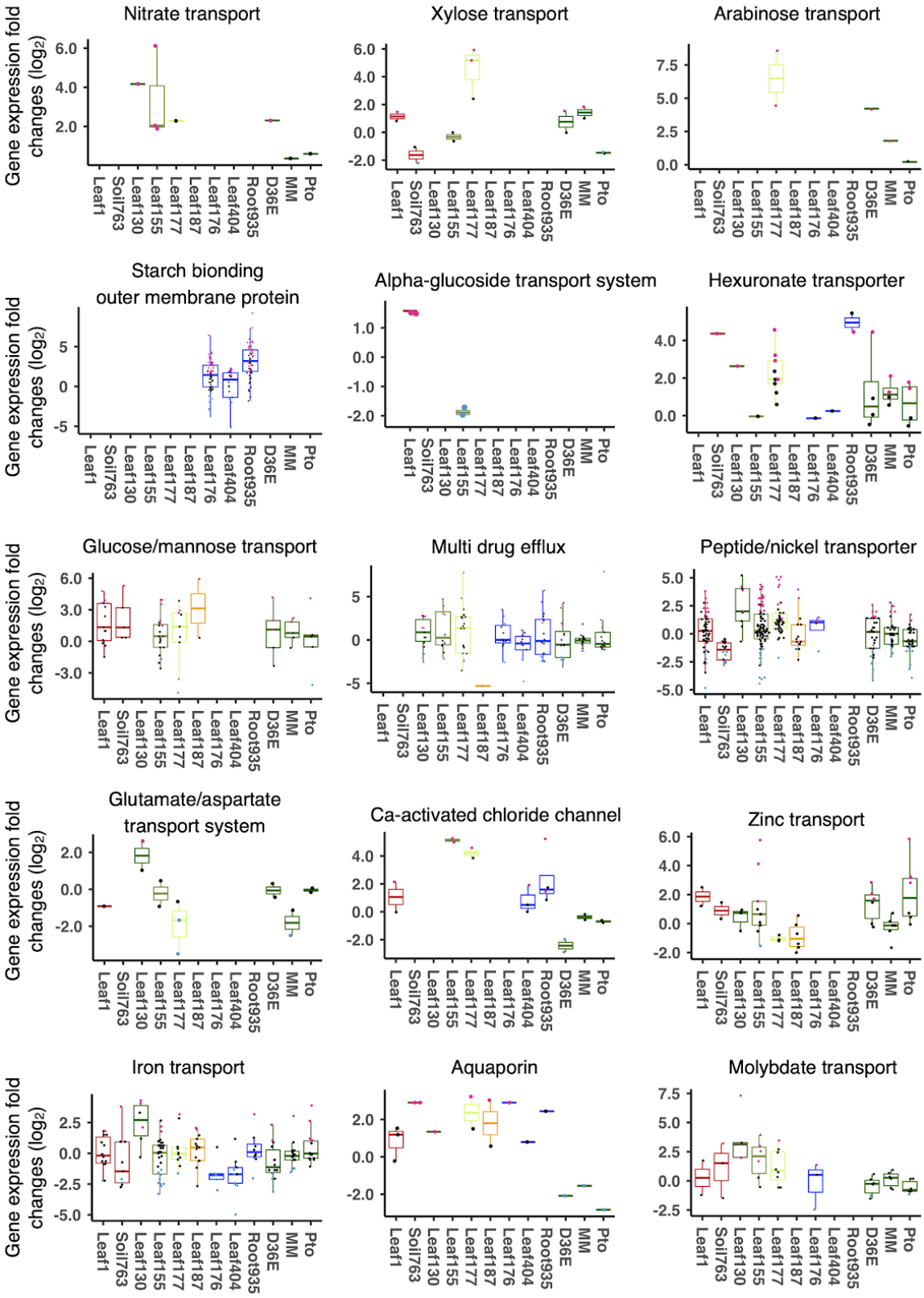
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



977

978 **Fig. S6: Expression of various processes of commensals *in planta* (A and**
979 **B)** Expression fold changes (*in planta* vs. *in vitro*) of genes related to **(A)**
980 transporters and **(B)** two-component system. Results are shown as box plots with
981 boxes displaying the 25th–75th percentiles, the centerline indicating the median,
982 whiskers extending to the minimum, and maximum values no further than 1.5
983 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class
984 level) of each strain. All individual data points (genes) are overlaid with colors for
985 DEGs (red: upregulated, blue: downregulated, black: non-DEG).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



986

987 **Fig. S7: Expression of genes related to nutrient acquisition processes in**
988 **commensals** Expression fold changes (*in planta* vs. *in vitro*) of genes related to

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nutrient transporters. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain (see Fig. S6 for the color code). All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

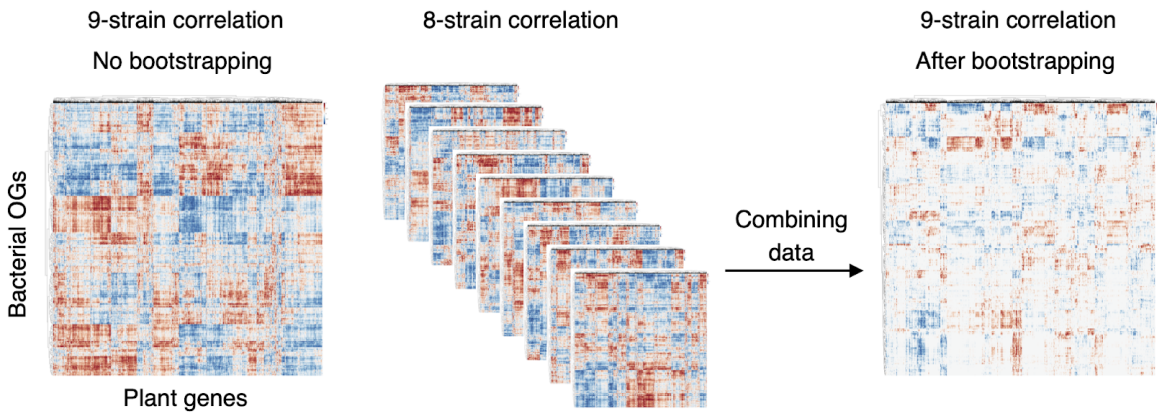
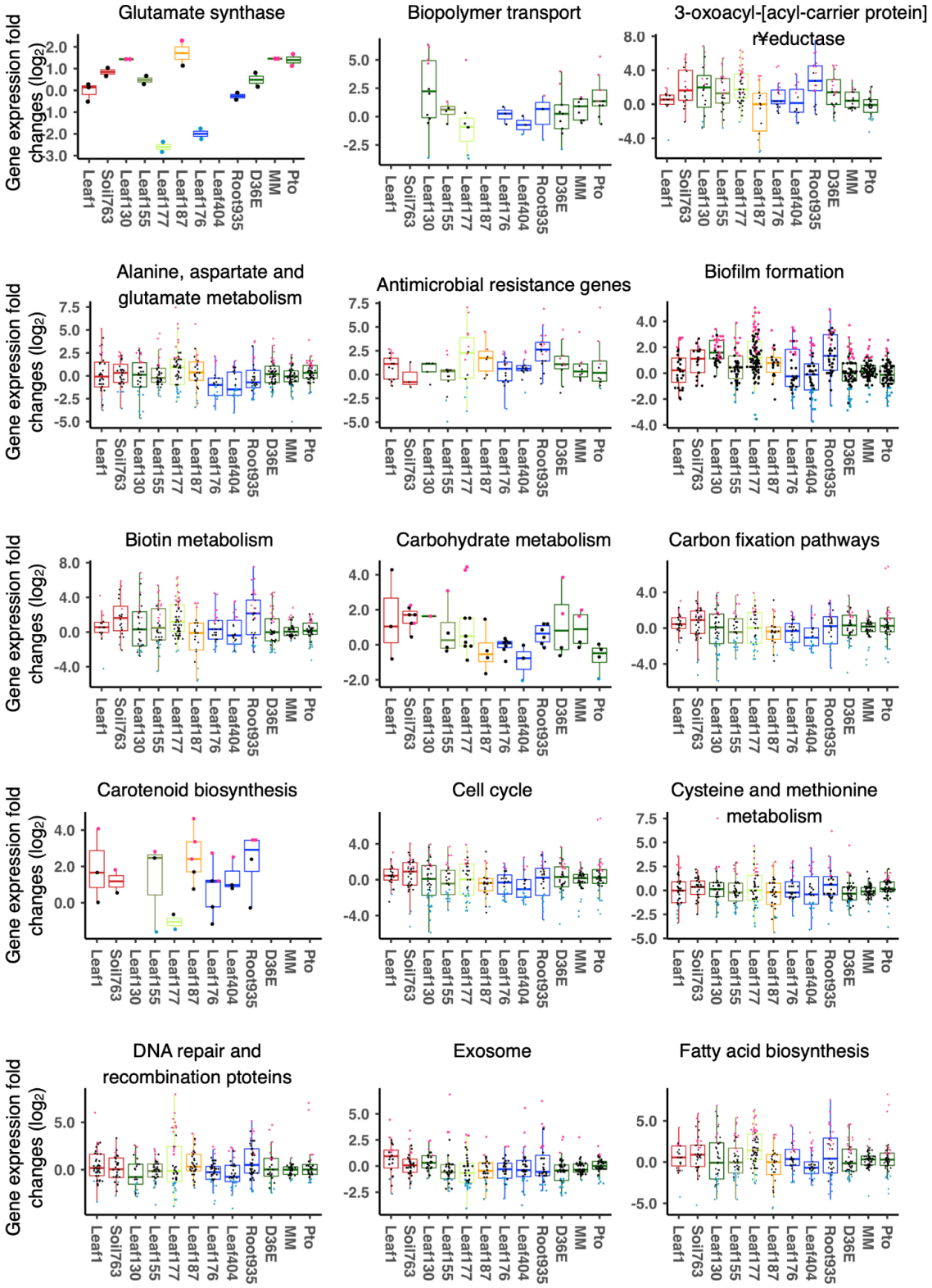


Fig. S8: Integration of plant and bacterial transcriptomes Schematic diagram showing a bootstrapping approach to evaluate correlations between individual plant and bacterial genes. To obtain robust correlation scores, Pearson's correlation coefficients were calculated using all the combinations of eight strains as well as using all the nine strains. Among these 8-strain and 9-strain datasets, the weakest correlation coefficient value was used for each combination of a bacterial OG and a plant gene ("Combining data").

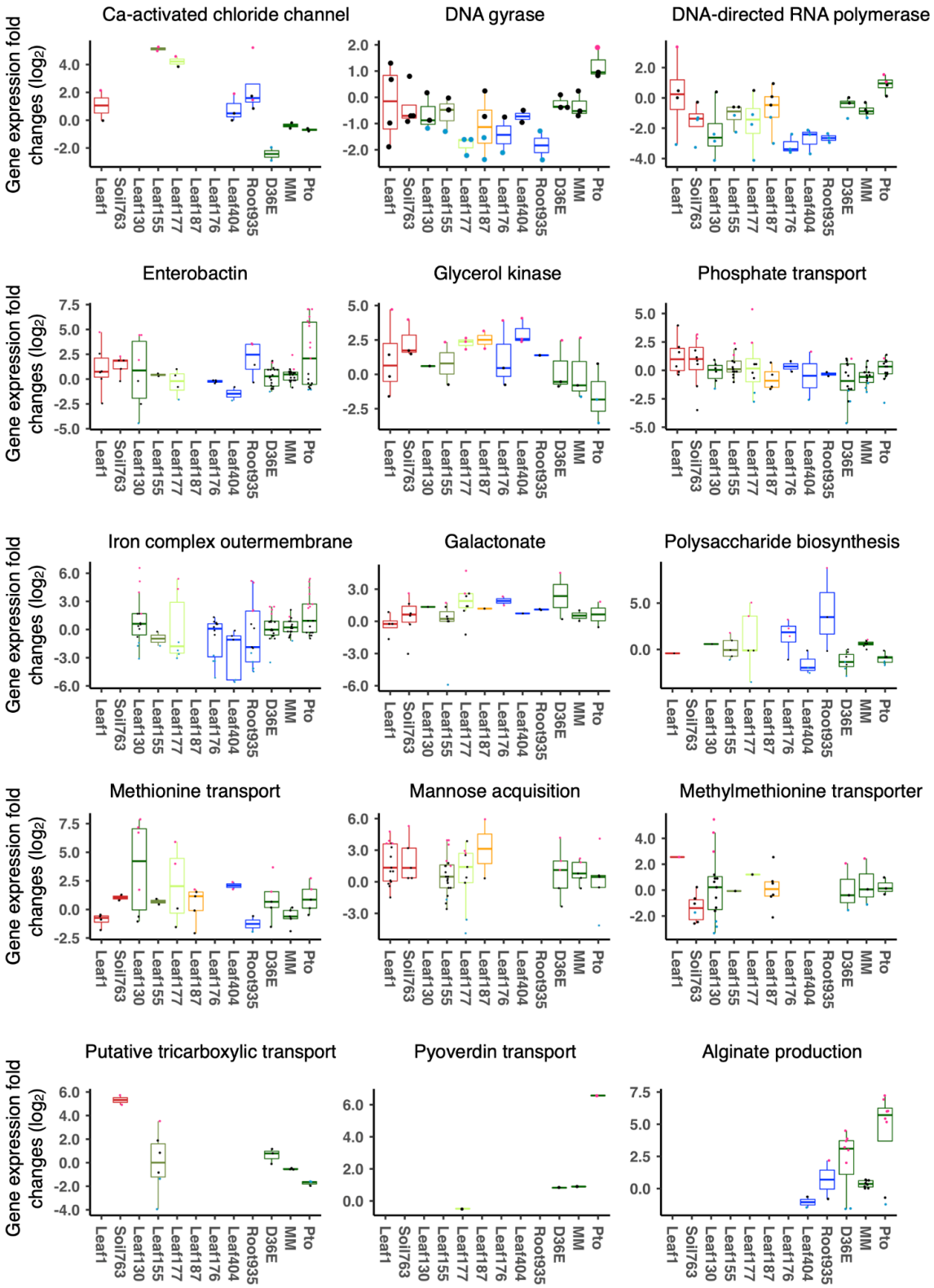
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S9: Expression of commensals genes related to various physiological processes *in planta* Expression fold changes (*in planta* vs. *in vitro*) of genes related to various functions. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



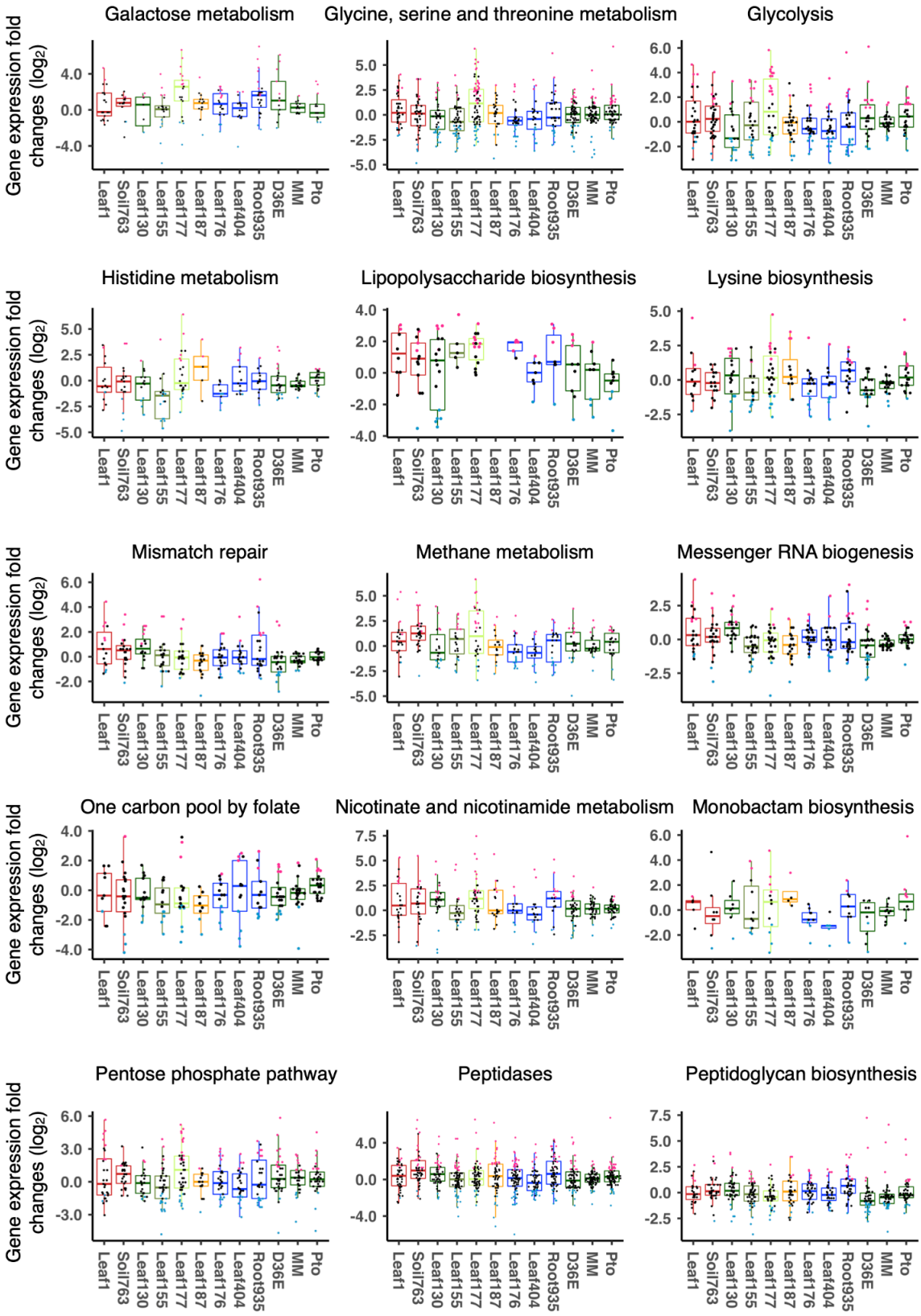
1010

1011 **Fig. S10: Expression of commensal genes related to various physiological**
1012 **processes *in planta*** Expression fold changes (*in planta* vs. *in vitro*) of genes
1013 related to various functions. Results are shown as box plots with boxes

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

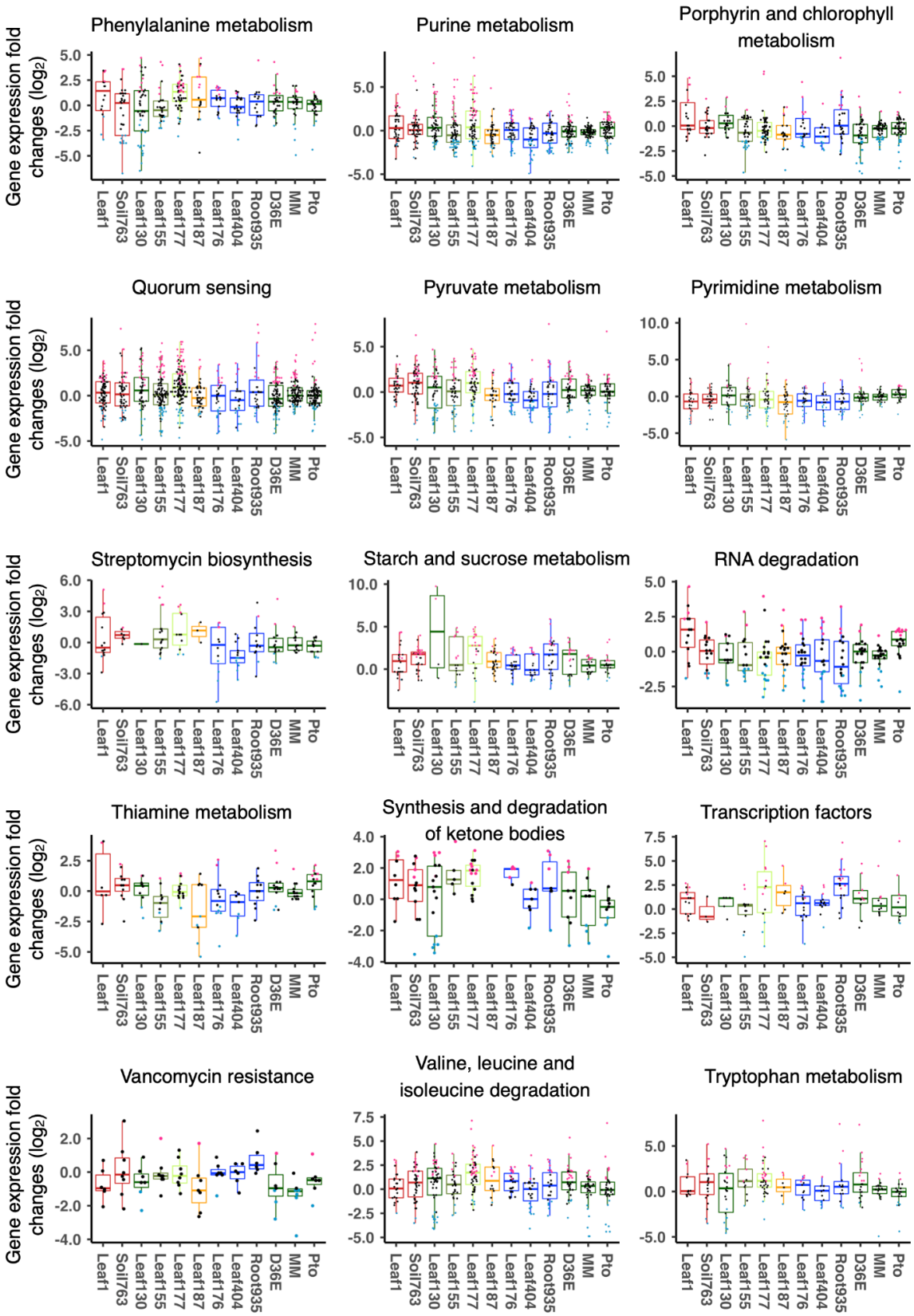
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S11: Expression of commensal genes related to various physiological processes *in planta* Expression fold changes (*in planta* vs. *in vitro*) of genes related to various functions. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.25.440543; this version posted April 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. S12: Expression of commensal genes related to various physiological processes *in planta* Expression fold changes (*in planta* vs. *in vitro*) of genes related to various functions. Results are shown as box plots with boxes displaying the 25th–75th percentiles, the centerline indicating the median, whiskers extending to the minimum, and maximum values no further than 1.5 inter-quartile range. Box color indicates the taxonomic affiliation (phylum/class level) of each strain. All individual data points (genes) are overlaid with colors for DEGs (red: upregulated, blue: downregulated, black: non-DEG).