

A plant genetic network for preventing dysbiosis in the phyllosphere

<https://doi.org/10.1038/s41586-020-2185-0>

Received: 1 September 2019

Accepted: 19 February 2020

Published online: 08 April 2020

 Check for updates

Tao Chen^{1,2,3,8}, Kinya Nomura^{1,8}, Xiaolin Wang⁴, Reza Sohrabi^{1,5}, Jin Xu⁶, Lingya Yao⁴, Bradley C. Paasch¹, Li Ma¹, James Kremer¹, Yuti Cheng^{1,3}, Li Zhang^{1,3}, Nian Wang⁶, Ertao Wang⁴, Xiu-Fang Xin^{4,7} & Sheng Yang He^{1,3,5}

The aboveground parts of terrestrial plants, collectively called the phyllosphere, have a key role in the global balance of atmospheric carbon dioxide and oxygen. The phyllosphere represents one of the most abundant habitats for microbiota colonization. Whether and how plants control phyllosphere microbiota to ensure plant health is not well understood. Here we show that the *Arabidopsis* quadruple mutant (*min7 fls2 efr cerk1*; hereafter, *mfec*)¹, simultaneously defective in pattern-triggered immunity and the MIN7 vesicle-trafficking pathway, or a *constitutively activated cell death1* (*cad1*) mutant, carrying a S205F mutation in a membrane-attack-complex/perforin (MACPF)-domain protein, harbour altered endophytic phyllosphere microbiota and display leaf-tissue damage associated with dysbiosis. The Shannon diversity index and the relative abundance of Firmicutes were markedly reduced, whereas Proteobacteria were enriched in the *mfec* and *cad1*^{S205F} mutants, bearing cross-kingdom resemblance to some aspects of the dysbiosis that occurs in human inflammatory bowel disease. Bacterial community transplantation experiments demonstrated a causal role of a properly assembled leaf bacterial community in phyllosphere health. Pattern-triggered immune signalling, MIN7 and CAD1 are found in major land plant lineages and are probably key components of a genetic network through which terrestrial plants control the level and nurture the diversity of endophytic phyllosphere microbiota for survival and health in a microorganism-rich environment.

The phyllosphere is inhabited by a diverse microbiota, with some phyllosphere microorganisms living on the surface of plants as epiphytes and others residing inside leaves as endophytes^{2,3}. In contrast to the intensively studied roles of root-colonizing microbiota in plant health^{4–11}, the collective community-level contribution of phyllosphere microbiota to plant growth, development and health is not well understood. The phyllosphere is functionally distinct from the belowground rhizosphere. For example, compared with roots, leaves have a larger apoplast, which facilitates the gas exchange essential for photosynthesis and provides a largely air-filled internal space for microbiota colonization. The composition of leaf microbiota can be influenced by host genotypes^{12–14}, and a recent ecological study showed a positive correlation between leaf bacterial diversity and terrestrial ecosystem productivity¹⁵. However, whether these variations in phyllosphere microbiota make a causal contribution to (or are merely a consequence of) plant health remains an unresolved fundamental question.

In a previous study aimed at identifying plant pathways that are attacked by the bacterial pathogen *Pseudomonas syringae* pv. tomato (*Pst*) DC3000, we identified *Arabidopsis* quadruple mutants (for example, *mfec* and *min7 bak1 bkk1 cerk1* (hereafter *mbbc*)¹ that allowed

increased proliferation of a nonpathogenic mutant of *Pst* DC3000 and harboured a greater abundance of endophytic leaf microorganisms under high humidity, one of the most common environmental conditions plants encounter in nature. *mfec* and *mbbc* mutants are defective in two pathways—pattern-triggered immune signalling^{16,17} and the MIN7 vesicle-trafficking pathway, which is involved in modulating an aqueous microenvironment in the apoplast¹. The *mfec* and *mbbc* mutants also exhibited spontaneous leaf necrosis and chlorosis in the absence of pathogen inoculation under high humidity¹. However, the lack of well-controlled soil-based gnotobiotic plant growth systems (the *mfec* and *mbbc* mutants showed most obvious phenotypes when grown in soil) prevented us from answering the fundamental question of whether altered leaf endophytic microbiota are the cause or a consequence of poor phyllosphere health of *mfec* and *mbbc* plants.

Genotype-dependent shift of microbiota

Because *mfec* and *mbbc* exhibited similar phenotypes in initial experiments, we conducted further characterization using the *mfec* mutant and observed additional phenotypes. First, chlorosis and/or necrosis

¹Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, MI, USA. ²State Key Laboratory of Agriculture Microbiology, Huazhong Agricultural University, Wuhan, China. ³Howard Hughes Medical Institute, Michigan State University, East Lansing, MI, USA. ⁴National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. ⁵Plant Resilience Institute, Michigan State University, East Lansing, MI, USA. ⁶Citrus Research and Education Center, Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL, USA. ⁷CAS-JIC Center of Excellence for Plant and Microbial Sciences (CEPAMS), Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. ⁸These authors contributed equally: Tao Chen, Kinya Nomura. [✉]e-mail: xinx@ippe.ac.cn; hes@msu.edu

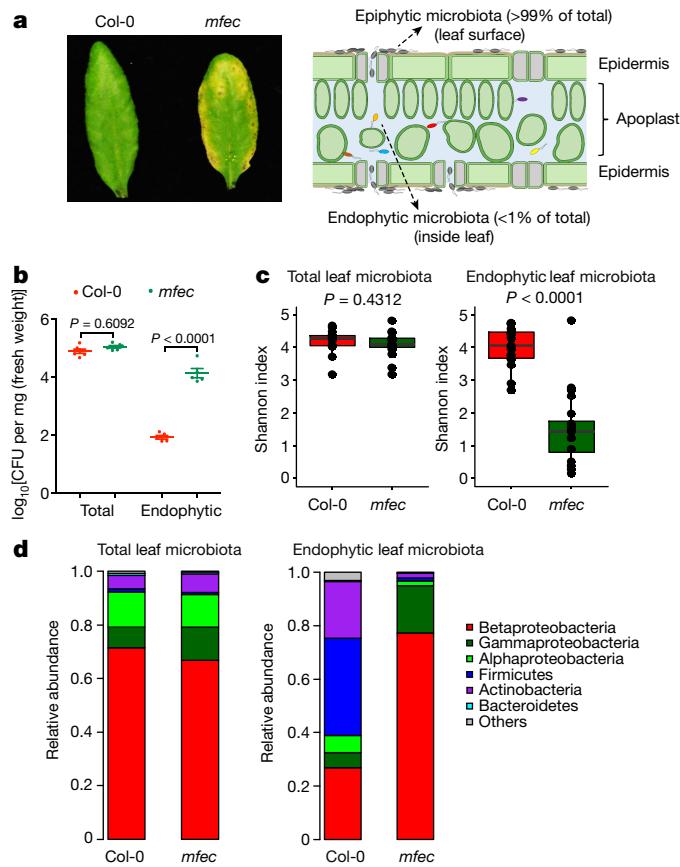


Fig. 1 | Total and endophytic leaf microbiota in Col-0 and *mfec* plants. **a, b,** Leaf appearance (a) and population size of leaf microbiota (b) in five-week-old Col-0 and *mfec* plants grown in *Arabidopsis mix* potting soil. **a,** Left, images at day 5 after plants (grown at approximately 60% humidity) were exposed to approximately 95% humidity. **a,** Right, cartoon showing epiphytic and endophytic microbiota in a leaf cross-section. Experiments in **a, b,** were repeated three times with similar results. One-way ANOVA with Tukey's test. $n = 6$ (total bacteria populations) and $n = 5$ (endophytic bacteria populations) biological replicates. Data are mean \pm s.e.m. **c, d,** Shannon indexes (c) and the relative abundance of bacteria at the phylum level (d), obtained from 16S rRNA gene-sequence profiles of total and endophytic bacteria in Col-0 and *mfec* plants grown in *Arabidopsis mix* potting soil. In box plots, the centre line represents the median, box edges show the 75th and 25th percentiles, and whiskers extend to $1.5 \times$ the interquartile range. Two-tailed Mann–Whitney *U*-test. $n = 15$ (Col-0) and $n = 15$ (*mfec*) biological replicates for analysis of total leaf bacterial microbiota across 3 independent experiments; $n = 18$ (Col-0) and $n = 20$ (*mfec*) biological replicates for analysis of leaf endophytic bacterial microbiota across 4 independent experiments.

phenotypes in the *mfec* mutant were seen in plants grown in different soil types in air-circulating growth chambers, albeit to varying degrees (Fig. 1a, Extended Data Fig. 1a). Second, tissue damage appeared to be restricted mostly to leaves, as roots of Col-0 and the *mfec* mutant appeared similar (Extended Data Fig. 1b). Third, in contrast to the marked difference in the levels of endophytic bacteria (estimated after surface sterilization to remove epiphytic bacteria), little difference was observed between Col-0 and *mfec* plants in total leaf bacteria (without surface sterilization), which include both epiphytic and endophytic bacteria (Fig. 1b). The total bacteria count was usually at least 100-fold higher than that of endophytic bacteria (Fig. 1b; in Col-0). Our results provided evidence for compartment-specific modulation of phyllosphere bacteria and suggest that the bulk epiphytic phyllosphere bacteria may have a less intimate interaction with and, therefore, may be less influenced by host innate immune signalling and the MIN7 vesicle-trafficking pathway.

Next, we conducted 16S rRNA gene-sequencing analysis of leaf bacterial communities in Col-0 and *mfec* plants. Plants were grown in '*Arabidopsis mix*' potting soil in air-circulating growth chambers for colonization of phyllosphere microbiota. We observed that the endophytic leaf community in Col-0 plants was substantially more diverse than that in the *mfec* plants, as judged by Shannon index and observed operational taxonomic units (OTUs) (Fig. 1c, Extended Data Fig. 2a, b). Furthermore, Firmicutes and Actinobacteria were abundantly observed in Col-0 leaves, whereas their relative abundance was greatly reduced in the *mfec* mutant. Conversely, Betaproteobacteria and Gammaproteobacteria were highly enriched in *mfec* leaves (Fig. 1d). By contrast, there was no significant difference in bacterial composition of total leaf bacteria between Col-0 and *mfec* (Fig. 1c, d), providing further evidence for profound compartment-specific modulation of the level and composition of endophytic leaf microbiota by pattern-triggered immunity and the MIN7 vesicle-trafficking pathway.

To determine the individual contributions of pattern-triggered immunity and the MIN7 vesicle-trafficking pathway to the endophytic leaf microbiota shift in the *mfec* mutant, we performed further 16S rRNA gene sequencing. In these experiments, we inoculated the *Arabidopsis mix* potting soil with a 48-member leaf endophytic bacterial community derived from healthy Col-0 plants (SynCom^{Col-0}; Supplementary Table 1) as a consistent microbiota source, in addition to presumably variable soil and/or air-derived communities to which plants were exposed in growth chambers. Again, we observed a significant reduction in the overall diversity and a substantial shift in the composition of the endophytic bacterial community in the *mfec* leaves, but not in *fec* (defective in pattern-triggered immunity alone) or *min7* (defective in MIN7 pathway alone) leaves, compared with that in Col-0 leaves (Fig. 2a, b, Extended Data Fig. 2c). Correspondingly, only *mfec* quadruple-mutant plants displayed necrosis and chlorosis and had a higher level of endophytic bacterial microbiota (Fig. 2c, d). These results show non-redundant and essential roles of pattern-triggered immunity and the MIN7 vesicle-trafficking pathway in controlling the endophytic leaf microbiota in *Arabidopsis*.

Further analysis of endophytic microbiota 16S-profiling data from SynCom^{Col-0}-supplemented experiments revealed changes in specific amplicon sequence variants (ASV) representing distinct bacterial 16S rRNA gene sequences between Col-0 and *mfec* (Supplementary Table 2). ASVs belonging to Comamonadaceae (ASV1, ASV113, ASV141 and ASV280), Xanthomonadaceae (ASV12 and ASV386), Alcaligenaceae (ASV3) and Sphingomonadaceae (ASV245) were enriched in *mfec* plants (together representing 91.97% of reads in *mfec* plants compared with only 39.69% in Col-0 plants). Conversely, 33 Paenibacillaceae ASVs were depleted in *mfec* plants (representing 24.83% and 0.52% of reads in Col-0 and *mfec* plants, respectively). However, the observed modest enrichment of the ASVs—on the basis of relative abundance—may not account for the approximately 100-fold increase of the total endophytic microbiome population in the *mfec* mutant, suggesting that other ASVs could contribute to the increase of the total load of endophytic microbiota in *mfec* leaves without being reflected in their relative abundance. A clear resolution of all ASVs will require a further study using methods that are more appropriate for estimating the absolute abundance of ASVs.

Role of *mfec* microbiota in dysbiosis

The reduction of the overall relative bacterial diversity and conversion of a Firmicutes-rich community to a Proteobacteria-rich community in the *mfec* mutant was intriguing because these changes bear some resemblance to those observed in human microbiome dysbiosis associated with inflammatory bowel disease^{18,19}. This raised the possibility that tissue damage in *mfec* plants may result from a form of dysbiosis in plants. However, true dysbiosis implies a causative role of altered microbiota in inducing symptoms. To test this possibility, we grew plants in sterile 0.5× Murashige–Skoog agar plates and in a peat-based gnotobiotic

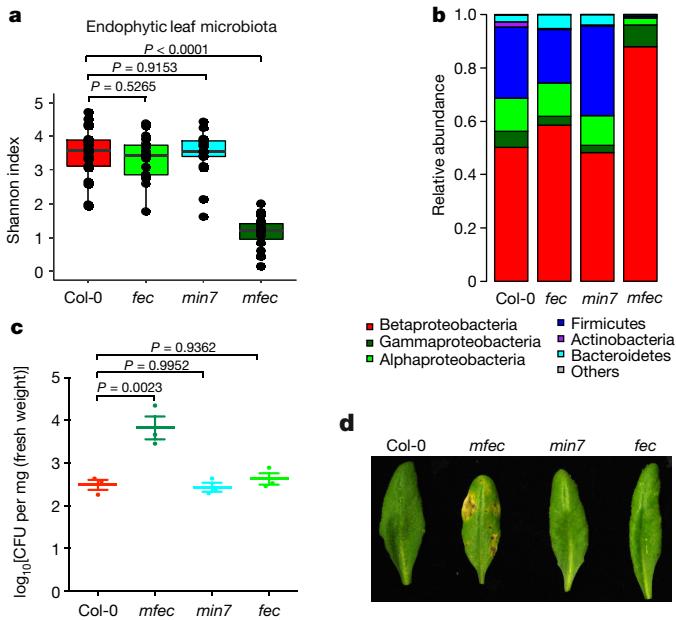


Fig. 2 | Endophytic leaf microbiota in Col-0, fec, min7 and mfec plants.

a, b, Shannon indexes (**a**) and the relative abundance of bacteria at the phylum level (**b**), obtained from 16S rRNA gene-sequence profiles of endophytic leaf bacteria in plants grown in *Arabidopsis* mix soil supplemented with SynCom^{Col-0}. Data presentation and statistical analysis as in Fig. 1c, d. $n = 20$ (Col-0), $n = 19$ (fec), $n = 19$ (min7) and $n = 19$ (mfec) biological replicates passing quality control across 4 independent experiments. **c, d**, Population size of endophytic leaf microbiota (**c**) and leaf appearance (**d**) in 5-week-old plants 6 days after plants were shifted to high humidity (approximately 95%). One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n = 3$ biological replicates; experiments were repeated four times with similar results.

plant growth system (hereafter, GnotoPot; Methods) and found that the mfec plants appeared healthy in the absence of microbiota (Extended Data Fig. 2d, e). By contrast, we observed chlorosis and some necrosis in mfec plants in the presence of leaf endosphere-derived SynCom^{Col-0}, whereas wild-type Col-0 plants remained healthy in the presence of SynCom^{Col-0}. Thus, SynCom^{Col-0} is sufficient to partially recapitulate host genotype-dependent dysbiotic symptoms in the phyllosphere.

Next, we addressed the question of whether the mfec-associated (that is, ‘improperly assembled’) microbiota alone is sufficient to cause dysbiotic symptoms in wild-type Col-0 plants. For this purpose, we assembled a 52-member mfec leaf-derived endophytic bacterial community (SynCom^{mfec}; Supplementary Table 1), which was prepared in parallel with SynCom^{Col-0}. Genome sequencing of individual isolates in SynCom^{Col-0} and SynCom^{mfec} confirmed a more diverse and balanced bacterial composition in SynCom^{Col-0} compared with SynCom^{mfec} (Extended Data Fig. 3a, Supplementary Tables 3–5), partially reflecting the endophytic bacterial composition found in Col-0 and mfec leaves as revealed by 16S rRNA gene sequencing (Fig. 1d). In particular, Firmicutes isolates were relatively abundant (20.8% of isolates) in SynCom^{Col-0}, whereas no culturable Firmicutes were recovered from SynCom^{mfec}. Conversely, 96.2% of isolates were Proteobacteria in SynCom^{mfec}, compared with 62.5% in SynCom^{Col-0} (Supplementary Table 1). There were additional taxonomic differences in the two synthetic communities even though they were derived from Col-0 and mfec plants that were grown in the same soil and growth chamber at the same time (Methods), illustrating the powerful influence of the mfec genotype on the assembly of the leaf endophytic bacterial community.

We conducted three types of functional assays to rigorously test whether mfec-associated (that is, incorrectly assembled) microbiota could cause health-damaging dysbiosis. First, in Linsmaier–Skoo agar plate assays, Col-0 plants inoculated with SynCom^{mfec} had significantly

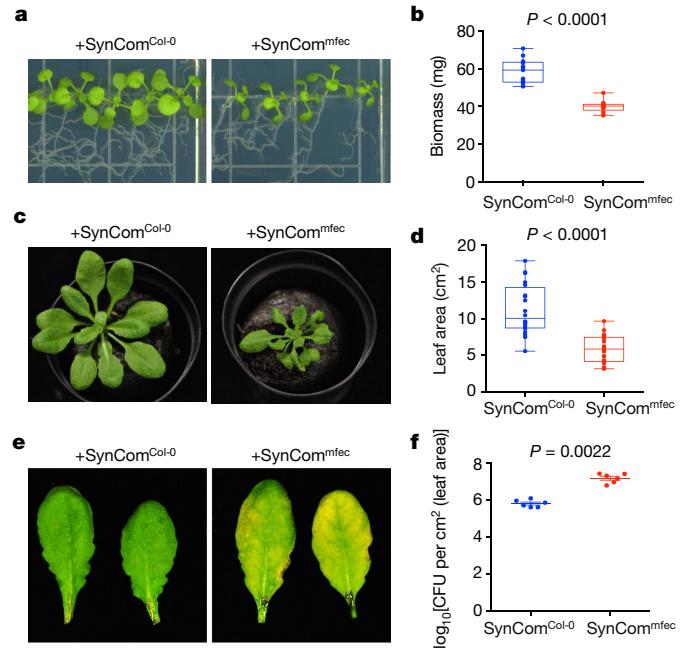


Fig. 3 | Functional effect of SynCom^{Col-0} and SynCom^{mfec} on plant health.

a, b, Phenotype (**a**) and biomass (**b**) of Col-0 seedlings inoculated with SynCom^{Col-0} or SynCom^{mfec}. Twelve 14-day-old seedlings were weighed as one biological replicate (see Methods). $n = 13$ biological replicates. **c, d**, Appearance (**c**) and the total leaf area per plant (as one biological replicate) (**d**) of Col-0 plants grown in GnotoPots in the presence of SynCom^{Col-0} or SynCom^{mfec} for 26 days. $n = 20$ biological replicates. **e, f**, Col-0 leaves were infiltrated with SynCom^{Col-0} or SynCom^{mfec} by syringe injection of $1 \times 10^8 \text{ CFU ml}^{-1}$, and leaf images (**e**) and bacterial populations (**f**) were recorded 5 days after infiltration. Data are mean \pm s.e.m., $n = 6$ biological replicates. In box plots in **b** and **d**, the centre line is the median, box edges show the 75th and 25th percentiles, and whiskers cover the full range of values. In **b**, **d** and **f**, two-tailed Mann–Whitney *U*-test was used for statistical analysis. All experiments were repeated three times with similar results.

reduced biomass relative to those inoculated with SynCom^{Col-0} (Fig. 3a, b). Second, when grown in peat-based GnotoPots, Col-0 plants appeared healthy in the presence of SynCom^{Col-0}, but showed varying degrees of seedling stunting and an overall reduction in rosette size in the presence of SynCom^{mfec} (Fig. 3c, d). Third, when infiltrated at a concentration of 1×10^8 colony-forming units (CFU) per ml into otherwise healthy leaves of fully grown and colonized Col-0 plants, SynCom^{mfec}, but not SynCom^{Col-0}, induced prominent necrosis and chlorosis (Fig. 3e). In addition, SynCom^{mfec} grew to a higher population than SynCom^{Col-0} in Col-0 leaves (Fig. 3f). When infiltrated at a lower concentration of $1 \times 10^7 \text{ CFU ml}^{-1}$ (equivalent to the approximately 10^4 CFU per mg (leaf tissue) of microbiota in *Arabidopsis*; Extended Data Fig. 3b), which simulates the level of endophytic microbiota in mfec leaves (Fig. 1b), SynCom^{mfec} was still capable of causing tissue damage, but to a lesser degree and more sporadically (Extended Data Fig. 3c). These combined results, from three independent assays, demonstrated that a dysbiotic microbiota (SynCom^{mfec}) is sufficient to confer a negative health effect in wild-type plants and provided evidence of the importance of assembling a normal leaf endophytic microbiota to ensure phyllosphere health.

Next, we investigated whether individual strains in SynCom^{Col-0} and SynCom^{mfec} could cause tissue damage when infiltrated into leaves of Col-0 plants grown in *Arabidopsis* mix. With an inoculum containing $1 \times 10^8 \text{ CFU ml}^{-1}$, more SynCom^{mfec} isolates (32 strains) than SynCom^{Col-0} isolates (17 strains) caused tissue damage, supporting the hypothesis that mfec leaves are enriched for tissue-damaging bacteria (Supplementary Table 1). With an inoculum containing $1 \times 10^7 \text{ CFU ml}^{-1}$ (equivalent to

approximately 10^4 CFU per mg (leaf tissue), similar to the total endophytic microbiota in *mfec* leaves), ten SynCom^{mfec} strains, but only four SynCom^{Col-0} strains, induced mild tissue damage (Supplementary Table 1). Of note, none of the ‘symptom-inducer’ strains multiplied at a rate similar to *Pst* DC3000, a virulent pathogen of *Arabidopsis* (Extended Data Fig. 4a), consistent with the hypothesis that these symptom-inducer strains are not canonical pathogens per se, but probably represent potentially harmful members of a normal leaf microbiota that are kept at low, non-damaging levels in a healthy wild-type phyllosphere. We further assembled a five-member synthetic phyllosphere community (SynCom^{mix}⁵), consisting of Proteobacteria strains derived from SynCom^{mfec}, that induced robust leaf tissue damage (Extended Data Fig. 4b). Each of the five strains were sufficient to cause leaf damage on their own, suggesting functional redundancy within SynCom^{mfec} in the induction of phyllosphere dysbiosis. However, simple removal of Firmicutes from SynCom^{Col-0} was not sufficient to produce a dysbiotic bacterial community (Extended Data Fig. 3c). This is consistent with the many fine taxonomical differences between strains in SynCom^{Col-0} and SynCom^{mfec}.

Mechanisms of bacterial community shift

Next, we investigated the underlying mechanism by which *mfec* plants lost the ability to maintain endophytic leaf bacterial diversity. We hypothesized that, in addition to host genetic influences, antagonistic bacterial interactions might be involved. To test this hypothesis, we performed binary inhibition assays (2,116 combinations), on R2A medium, of 46 strains that represent all bacterial species we identified in SynCom^{Col-0} and SynCom^{mfec}. This assay revealed a pattern of almost unidirectional antibiosis: most Firmicutes were strongly inhibited by a subset of Proteobacteria in vitro (Extended Data Fig. 5).

This in vitro observation was unexpected as it cannot explain the coexistence of Firmicutes and Proteobacteria in Col-0 leaves. We therefore considered the possibility that the largely unidirectional antibiosis observed on R2A medium (that is, in vitro) may become biologically relevant only when the two bacteria are in close proximity in vivo (that is, when endophytic bacterial populations become relatively high, as observed in *mfec* leaves; Fig. 1b). To investigate this possibility, we examined in vivo pairwise interactions of bacteria that showed strong binary interactions in vitro (for example, Proteobacteria strain C13 and Firmicutes strain C3, Proteobacteria strain C45 and Firmicutes strain C3, and Proteobacteria strain C13 and Firmicutes strain C41). These pairs were infiltrated into leaves of Col-0 plants grown in *Arabidopsis* mix potting soil at two different bacterial concentrations: 1×10^4 CFU ml⁻¹ (equivalent to 1×10^2 CFU per cm² (leaf area)) and 1×10^6 CFU ml⁻¹ (equivalent to 1×10^4 CFU cm² (leaf area)). For all three binary interactions examined, Firmicutes strains were outcompeted only at the higher inoculation level (Extended Data Fig. 6a, c, d). No such competition was observed between Proteobacterium strain C52 and Firmicutes strain C3, which did not show inhibition in vitro (Extended Data Figs. 5b, 6b). These results suggest a possible mechanism to explain why *mfec* plants lost the ability to maintain endophytic leaf bacterial diversity. In wild-type Col-0 leaves, pattern-triggered immunity and the MIN7 vesicle-trafficking pathway restrain the growth of leaf endophytic bacteria, including specific Proteobacteria that could inhibit other leaf endophytic bacteria, most notably, Firmicutes. In the *mfec* mutant, excess proliferation of leaf endophytic bacteria probably leads to the inhibition of Firmicutes by Proteobacteria strains, contributing to the reduction of overall relative bacterial diversity and conversion of a Firmicutes-rich community in Col-0 leaves to a Proteobacteria-rich community in the *mfec* mutant leaves.

A framework for microbiota homeostasis

Our results thus far provided strong evidence for the importance of pattern-triggered immunity and the MIN7 vesicle-trafficking pathway in

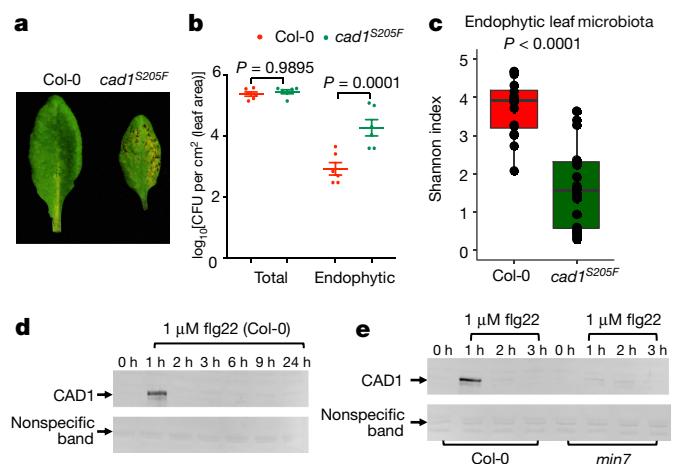


Fig. 4 | Microbiota phenotypes in the *ben3* mutant. **a, b**, Leaf appearance (**a**) and population sizes of total and endophytic leaf microbiota (**b**) in Col-0 and *ben3* (hereafter referred to as *cad1S205F*) plants grown in *Arabidopsis* mix soil supplemented with SynCom^{Col-0} for 4 weeks before plants were shifted to high humidity (approximately 95%) for 2 days (see Methods). One-way ANOVA with Tukey’s test. Data are mean \pm s.e.m., $n = 6$ biological replicates. **c**, Shannon indexes of 16S rRNA gene-sequence profiles of endophytic leaf bacteria in Col-0 and *cad1S205F* plants supplemented with SynCom^{Col-0}. Data presentation and statistical analysis as in Fig. 1c. $n = 20$ (Col-0) and $n = 20$ (*cad1S205F*) biological replicates. **d, e**, Western blot analyses of CAD1 protein in Col-0 (**e**) and *min7* (**f**) plants. Five-week-old Col-0 and *min7* leaves were infiltrated with 1 μ M flg22 and collected at the indicated time points. CAD1 protein was detected with a CAD1 antibody; nonspecific bands show equal loading. The uncropped gel images are shown in Supplementary Fig. 1. All experiments in this figure were repeated three times with similar results.

controlling endophytic phyllosphere microbiota; however, it remained unclear whether these two processes are mechanistically separate or are components of a common molecular framework. Fortunately, during this study we discovered *mfec*-like phenotypes in *ben3*, an *Arabidopsis* mutant that was initially isolated on the basis of a genetic screen for a defect in intracellular vesicle trafficking^{20,21}. The *ben3* mutant carries a mutation in the *BREFELDIN A-INHIBITED GUANINE NUCLEOTIDE-EXCHANGE PROTEIN2* (*BIG2*) gene, which encodes an ADP ribosylation factor (ARF) family of guanine nucleotide exchange factor that is closely related to MIN7 and, like MIN7, is localized in the *trans*-Golgi network and early endosome²¹. The *ben3* mutant phenocopied the *mfec* quadruple mutant in (1) exhibiting spontaneous dysbiosis-like symptoms (Fig. 4a, Extended Data Fig. 7a, c), (2) harbouring a higher level of leaf endophytic microbiota compared with Col-0 plants (Fig. 4b, Extended Data Fig. 7b) and (3) hosting a leaf endophytic microbial community that is reduced in Shannon diversity index, enriched in Proteobacteria and depleted in Firmicutes (Fig. 4c, Extended Data Fig. 9a). However, during further characterization, we found that independent *big2* mutants carrying transfer DNA (T-DNA) insertions (*big2-1* and *big2-2*) did not show dysbiosis phenotypes (Extended Data Fig. 8a, b, g) and, using bulk segregation analysis and next-generation sequencing (Methods), we identified the causal mutation of dysbiosis in *ben3* to be a S205F substitution in a MACPF-domain protein—encoded by the *CONSTITUTIVELY ACTIVATED CELL DEATH1* (*CAD1*) gene (Extended Data Fig. 8c–g, Supplementary Table 6)—localized to the plasma membrane^{22–24}.

Because the *cad1S205F* mutation phenocopied the *mfec* quadruple mutants that are defective in both pattern-triggered immunity and the MIN7 vesicle-trafficking pathway, we hypothesized that CAD1 could be one of the convergent components downstream of pattern-triggered immunity and the MIN7 vesicle-trafficking pathway. Consistent with

this possibility, we found that the *CAD1* gene and protein were induced in response to flg22 (Fig. 4d, Extended Data Fig. 8h), an inducer of pattern-triggered immunity^{25,26}, and that flg22-induced accumulation of the CAD1 protein was affected by the *min7* mutation (Fig. 4e, Extended Data Fig. 8i). These results suggest that pattern-triggered immunity, the MIN7 vesicle-trafficking pathway and CAD1 are components of a large molecular framework that controls endophytic microbial abundance and diversity in the phyllosphere (Extended Data Fig. 9b, c).

Discussion

Our results begin to highlight conceptual parallels between plants and mammals in the mechanisms that prevent dysbiosis, a condition with severe consequences for the health of the host. In particular, defects in innate immunity pathways seem to be a common determinant underlying dysbiosis in mammals^{27–29} and plants. In addition, CAD1 is a member of the MACPF protein family; members of this family, such as complement protein C9 and perforin, have been shown to be involved in innate and adaptive immunity against pathogens in mammals^{30,31}. Pattern-triggered immunity, MIN7 and CAD1 are broadly conserved across plant taxa (Extended Data Fig. 10, Supplementary Table 7), suggesting that host control of the endophytic phyllosphere population and diversity may be a conserved feature in the plant kingdom. Manipulation of host genetic pathways regulating microbiota homeostasis could lead to a more beneficial and climate-resilient phyllosphere microbiota, which could in turn improve the performance of natural ecosystems and agricultural crops, as discussed further in the Supplementary Discussion.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2185-0>.

1. Xin, X. F. et al. Bacteria establish an aqueous living space in plants crucial for virulence. *Nature* **539**, 524–529 (2016).
2. Beattie, G. A. & Lindow, S. E. Bacterial colonization of leaves: a spectrum of strategies. *Phytopathology* **89**, 353–359 (1999).
3. Lindow, S. E. & Brandl, M. T. Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* **69**, 1875–1883 (2003).
4. Berendsen, R. L., Pieterse, C. M. & Bakker, P. A. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478–486 (2012).
5. Xu, L. et al. Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proc. Natl Acad. Sci. USA* **115**, E4284–E4293 (2018).
6. Edwards, J. A. et al. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol.* **16**, e2003862 (2018).

7. Finkel, O. M., Castrillo, G., Herrera Paredes, S., Salas González, I. & Dangl, J. L. Understanding and exploiting plant beneficial microbes. *Curr. Opin. Plant Biol.* **38**, 155–163 (2017).
8. Pieterse, C. M. J., de Jonge, R. & Berendsen, R. L. The soil-borne supremacy. *Trends Plant Sci.* **21**, 171–173 (2016).
9. Duran, P. et al. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell* **175**, 973–983 (2018).
10. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The plant microbiota: systems-level insights and perspectives. *Annu. Rev. Genet.* **50**, 211–234 (2016).
11. Zhang, J. et al. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* **37**, 676–684 (2019).
12. Horton, M. W. et al. Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat. Commun.* **5**, 5320 (2014).
13. Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M. & Vorholt, J. A. A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet.* **10**, e1004283 (2014).
14. Wagner, M. R. et al. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat. Commun.* **7**, 12151 (2016).
15. Lafosse-Lapointe, I., Paquette, A., Messier, C. & Kembel, S. W. Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* **546**, 145–147 (2017).
16. Tang, D., Wang, G. & Zhou, J. M. Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell* **29**, 618–637 (2017).
17. Boutrot, F. & Zipfel, C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* **55**, 257–286 (2017).
18. Turpin, W., Goethel, A., Bedrani, L. & Croitoru, K. Determinants of IBD heritability: genes, bugs, and more. *Inflamm. Bowel Dis.* **24**, 1133–1148 (2018).
19. Sokol, H. & Seksik, P. The intestinal microbiota in inflammatory bowel diseases: time to connect with the host. *Curr. Opin. Gastroenterol.* **26**, 327–331 (2010).
20. Tanaka, H., Kitakura, S., De Rycke, R., De Groot, R. & Friml, J. Fluorescence imaging-based screen identifies ARF GEF component of early endosomal trafficking. *Curr. Biol.* **19**, 391–397 (2009).
21. Kitakura, S. et al. BEN3/BIG2 ARF GEF is involved in brefeldin A-sensitive trafficking at the *trans*-Golgi network/early endosome in *Arabidopsis thaliana*. *Plant Cell Physiol.* **58**, 1801–1811 (2017).
22. Morita-Yamamoto, C. et al. The *Arabidopsis* gene *CAD1* controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol.* **46**, 902–912 (2005).
23. de Michele, R. et al. Free-flow electrophoresis of plasma membrane vesicles enriched by two-phase partitioning enhances the quality of the proteome from *Arabidopsis* seedlings. *J. Proteome Res.* **15**, 900–913 (2016).
24. Alexandersson, E., Saalbach, G., Larsson, C. & Kjellbom, P. *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol.* **45**, 1543–1556 (2004).
25. Felix, G., Duran, J. D., Volko, S. & Boller, T. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J. Cell Mol. Biol.* **18**, 265–276 (1999).
26. Zipfel, C. et al. Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**, 764–767 (2004).
27. Levy, M., Kolodziejczyk, A. A., Thaiss, C. A. & Elinav, E. Dysbiosis and the immune system. *Nat. Rev. Immunol.* **17**, 219–232 (2017).
28. Hall, A. B., Tolonen, A. C. & Xavier, R. J. Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* **18**, 690–699 (2017).
29. Sun, L., Nava, G. M. & Stappenbeck, T. S. Host genetic susceptibility, dysbiosis, and viral triggers in inflammatory bowel disease. *Curr. Opin. Gastroenterol.* **27**, 321–327 (2011).
30. McCormack, R., de Armas, L., Shiratsuchi, M. & Podack, E. R. Killing machines: three pore-forming proteins of the immune system. *Immunol. Res.* **57**, 268–278 (2013).
31. Spicer, B. A., Conroy, P. J., Law, R. H. P., Voskoboinik, I. & Whisstock, J. C. Perforin—a key (shaped) weapon in the immunological arsenal. *Semin. Cell Dev. Biol.* **72**, 117–123 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Plant materials and growth conditions

For most experiments, *Arabidopsis thaliana* plants were grown in *Arabidopsis* mix greenhouse potting soil (equal parts of Suremix (Michigan Grower Products), medium vermiculite and perlite; autoclaved once) or Redi-Earth soil (Sun Gro Horticulture) in air-circulating growth chambers for colonization of phyllosphere microbiota. Plants were grown under relative humidity set at 60%, temperature at 22 °C, light intensity at 100 $\mu\text{E m}^{-2} \text{s}^{-1}$ and photoperiod at a 12:12 h light:dark cycle. Five-week-old plants were used for bacterial inoculation and microbiota assays.

The wild-type accession Col-0, *min7 fls2 efr cerk1 (mfec)*, *ben3* (that is, *cad1^{S20SF}* used in this study) and *dde2-2/ein2-1/pad4-1/sid2-2 (deps)* mutant derivatives were described previously^{1,21,32} or in this study. The *big2-1* (SALK_033446) and *big2-2* (SALK_016558) T-DNA insertion mutants were obtained from the *Arabidopsis* Biological Resource Centre (ABRC) at The Ohio State University.

Bacterial quantification

To quantify culturable endophytic bacterial communities, leaves were surface-sterilized in 75% ethanol for 1 min and rinsed in sterile water twice. After air-drying to evaporate surface water, leaves were weighed and ground in 1 ml sterile 10 mM MgCl₂ buffer. A serial dilution in 100 μl (to up to 10⁻³ for Col-0 and up to 10⁻⁵ for *mfec* and *cad1^{S20SF}* mutants) was made and plated on R2A plates (100 mm × 100 mm). Under experimental conditions in this study, almost all leaf bacteria grew to large enough microcolonies in R2A plates in 2–3 days to be efficiently counted under a microscope (20× magnification). Pilot trials with longer incubation times, up to 10 days (see below), did not yield significantly more bacterial colonies that would have affected plant genotype differences. Therefore, we counted bacterial colony-forming units (CFU) 3 days after R2A plates were incubated at room temperature. CFUs were normalized to fresh tissue weight or leaf disk area. Total leaf bacterial communities were determined following the same protocol except without surface sterilization. A serial dilution was made in 100 μl volume (up to 10⁻⁵ for Col-0, *mfec* and *cad1^{S20SF}*) and plated on R2A plates.

In pilot trials, two leaves or eight leaf disks were ground in 1 ml sterile 10 mM MgCl₂ buffer. After serial dilutions (see above), 100 μl of leaf homogenate were spread onto square Petri dishes (100 mm × 100 mm) containing 30 ml R2A agar. Degrees of dilution were determined by empirical experiments to ensure that there were 20 to 200 CFUs per plate for accurate counting. As an example, when the following two incubation methods were used: (i) 3 days of incubation at room temperature and (ii) 10 days of incubation (2 days at room temperature, 6 days at 4 °C and 2 days at room temperature), bacterial CFUs (mean ± s.e.m., $n=4$) were 105 ± 20 (2-day incubation) and 101 ± 23 (10-day incubation), respectively.

Gnotobiotic plant growth systems

Three gnotobiotic plant growth systems were used in this study. Murashige–Skoog (MS) or Linsmaier–Skoog (LS) agar plate system: Col-0, *mfec* and *cad1^{S20SF}* seeds were surface-sterilized, cold-stratified and germinated on 0.5× MS or LS agar medium plates. Seedlings were grown under 50 $\mu\text{E m}^{-2} \text{s}^{-1}$ with a photoperiod of 12:12 h light:dark cycle. FlowPot gnotobiotic system: a peat-based gnotobiotic system described previously³³. GnotoPot: a compressed peat pellet-based gnotobiotic system developed in this study as a simpler, alternative potting soil-based gnotobiotic system. In brief, compressed peat pellets (Jiffy Products) were transferred to 2-inch polypropylene pots and

hydrated to saturation with LS medium buffered with 2-(N-morpholino) ethanesulfonic acid (MES) to pH 5.7 (Caisson Labs). GnotoPots were then placed in plant tissue culture microbox (SacO2) that had a no. 40 green filter mounted in the lid. Assembled microboxes with lids loosely placed were placed inside an autoclave bag (Sun Bag, Sigma) and autoclaved twice for 45 min each, with 24 h storage at room temperature in between. After GnotoPots were cooled down, microboxes were sealed and stored until time of use. Surface-sterilized *Arabidopsis* seeds were stratified at 4 °C for 24 h before being sown into GnotoPots under germ-free conditions. GnotoPots were then placed in a tissue culture growth chamber set at 22 °C, 12:12 h light:dark cycle photoperiod and light intensity of 100 $\mu\text{E m}^{-2} \text{s}^{-1}$. Sterility of FlowPots and GnotoPots was routinely monitored by plating samples of plants and peat substrate in R2A plates.

For induction of dysbiosis symptoms shown in Extended Data Fig. 2e, 5.5- to 6.5-week-old GnotoPots-grown plants were sprayed with sterile water, placed under a clear plastic dome to achieve high humidity (~95%) under 40 $\mu\text{E m}^{-2} \text{s}^{-1}$ light intensity and temperature at 23 °C for 10 days. For results shown in Fig. 3c, plants were grown in GnotoPots under a photoperiod of 16:8 h light:dark (100 $\mu\text{E m}^{-2} \text{s}^{-1}$ and temperature at 22 °C) for 26 days and total leaf areas were measured with the Easy Leaf Area software³⁴.

Synthetic communities of leaf endophytic bacteria

To generate SynComs, Col-0 and *mfec* plants were grown in *Arabidopsis* mix potting soil to 5 weeks old and were sprayed with water and kept under high humidity (~95%) for 5 days. Representative leaves were harvested (8 leaves were picked from 4 plants of each genotype) and surface-sterilized in 75% ethanol for 1 min and rinsed in sterile water twice. Leaves were ground in sterile water, and bacterial suspensions were diluted (to 10⁻³ for Col-0 and 10⁻⁵ for *mfec*) and plated on R2A plates, which were kept at 22 °C for 4 days. About 50 colonies from each genotype were randomly picked, constituting SynCom^{Col-0} and SynCom^{*mfec*}, respectively.

For addition of SynCom bacteria as internal control to *Arabidopsis* potting soil for 16S rRNA gene-sequence profiles (Figs. 2, 4 and Extended Data Figs. 2, 9), individual bacterial strains were scraped from R2A plate and suspended in 10 mM MgCl₂ buffer, bacterial suspensions were adjusted to the same OD₆₀₀. Equal volumes of each strain were pooled and diluted to a final OD₆₀₀ of 0.01 (~0.5 × 10⁷ CFU ml⁻¹). Five millilitres of prepared SynCom was added to each *Arabidopsis* mix soil pot. To reduce background microbiota that is naturally present in *Arabidopsis* mix soil, *Arabidopsis* mix soil and meshes were autoclaved twice before addition of SynCom bacteria; pots, flats and plastic domes used in growing plants were sprayed with 75% ethanol. Surface-sterilized Col-0, *mfec*, *min7, fec* or *cad1^{S20SF}* seeds were then added to the assembled soil pots. Plants were watered with autoclaved nutrient water 1–2 times each week.

For experiments with 0.5× LS agar plates (Fig. 3a, b), sterile Col-0 seeds were germinated on 0.5× LS plates (without sucrose) in the presence of 2 μl 10⁷ CFU ml⁻¹ SynCom^{Col-0} or SynCom^{*mfec*} for 14 days. The 2 μl SynCom was applied directly to each seed. For preparation of SynCom^{Col-0} and SynCom^{*mfec*} for inoculation into FlowPots or GnotoPots (Fig. 3c, d and Extended Data Figs. 2, 7), SynCom^{Col-0} and SynCom^{*mfec*} mixtures were prepared as above and the final OD₆₀₀ was adjusted to 0.04 (~2 × 10⁷ CFU ml⁻¹). A single seed was sown to each pot and 1 ml of the SynCom suspension was added. For experiments to compare the effects of different SynComs on leaf health grown in *Arabidopsis* potting soil (Fig. 3e, f and Extended Data Fig. 3b, c), SynCom^{Col-0} and SynCom^{*mfec*} mixtures were prepared as above, and the final OD₆₀₀ was adjusted to 0.2 (~1 × 10⁸ CFU ml⁻¹ for total SynCom mixtures; ~2 × 10⁶ CFU ml⁻¹ for each strain, Fig. 3e, f) or 0.02 (~1 × 10⁷ CFU ml⁻¹ for total SynCom mixtures; ~2 × 10⁵ CFU ml⁻¹ for each strain, Extended Data Fig. 3b, c) before infiltration into 4-week-old leaves.

16S rRNA gene-sequence profiling

Col-0, *mfec*, *min7*, *fec* or *cad1^{S20F}* plants were grown on experimental *Arabidopsis* mix soil with or without inoculated SynCom^{Col-0} added as described above. Five-week-old healthy plants were sprayed with water and kept under high humidity (~95%) for several days until dysbiosis symptoms appeared. Middle-age leaves from Col-0 and mutants were collected. To analyse endophytic bacterial community, leaves were first surface-sterilized with 5% (v/v) bleach for 1 min, followed with rinse with sterile ddH₂O twice. After blot-drying to remove surface water, two leaves from the same plant were collected in one tube as one biological replicate, which was then snap-frozen in liquid N₂ and stored at -80 °C. For analysis of total bacterial community, two leaves from one plant was collected in a tube as one sample, which was then snap-frozen in liquid N₂ and stored at -80 °C.

To prepare DNA for bacterial 16S rRNA gene-based community analysis, Total DNA from leaf samples was extracted using MoBio Power Soil DNA Isolation kit (Qiagen). PCR was performed using AccuPrime high-fidelity Taq DNA polymerase (Invitrogen) using barcoded primers 799F/1193R³⁵. 799F: **ACACTGACGACATGGTTCTACAAACMGGATTAGATA**CCCKG and 1193R: **TACGGTAGCAGAGACTTGGTCTACGT**CATC-CCCACCTTCC (bold sequences are the Illumina common sequence adapters). PCR was performed in triplicate in 25 µl reaction volumes containing 0.15 µl AccuPrime high-fidelity Taq DNA polymerase, 1 µl DMSO, 2.5 µl Buffer II, 0.5 µl of each primer (10 µM), 2 µl template DNA and 18.35 µl ddH₂O. The PCR program included a hot start at 94 °C for 60 s, 35 cycles of denaturation at 94 °C for 20 s, primer annealing at 53 °C for 30 s and extension at 68 °C for 45 s, followed by a final extension at 68 °C for 2 min and a cool down to 8 °C. PCR products were separated on 1% agarose gel. The 450-bp band of amplified bacterial 16S rRNA gene was extracted using Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instructions. DNA concentration was measured with PicoGreen dsDNA assay kit (Life Technologies) and adjusted to 1 to 10 ng µl⁻¹ for all samples. Samples were submitted to Research Technology Service Facility (RTSF) at Michigan State University for library preparation and 16S rRNA gene sequencing (see below).

RTSF Genomics Core at Michigan State University completed library preparation by PCR with dual-indexed Illumina-compatible adaptors targeting the Fluidigm barcoding oligos. Final PCR products were bulk-normalized using Invitrogen SequalPrep DNA Normalization plates and recovered libraries were pooled. The library pool was cleaned up using AmpureXP magnetic beads and then quantified using a combination of Qubit dsDNA HS (Invitrogen). Agilent 4200 TapeStation DNA 1000 and Kapa Illumina Library Quantification qPCR assays: the library pool was loaded onto an Illumina MiSeq Standard v2 flow cell and sequencing was performed in a 2 × 250 bp paired-end format using a MiSeq v2 500 cycle reagent cartridge. Common sequencing and index primers complementary to the Fluidigm CS1/CS2 oligomers were added to appropriate wells of the reagent cartridge. Base calling was done by Illumina Real Time Analysis (RTA) v.1.18.54 and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v.2.19.1.

Raw Illumina fastq files were quality-filtered and taxonomically analysed using QIIME 2 Core 2018.11 distribution³⁶. In brief, primers of imported sequences were removed via Cutadapt³⁷. DADA2³⁸ was used to filter and denoise sequences, remove chimaeras, identify representative sequences of OTUs and create an OTU table. Representative sequences of OTUs were taxonomically annotated using a pre-trained naive Bayes classifier³⁹ on the basis of the bacterial 16S rRNA Greengenes reference database (13_8 release). From this taxonomic annotation, all unassigned sequences and sequences annotated as mitochondria and chloroplast were removed. The filtered sequences were clustered at 97% similarity and the resulting OTU table was then used to determine taxonomic distributions and alpha (observed OTUs

and Shannon's species diversity index). For alpha-diversity calculations, samples were rarefied to the same number of reads.

The total bacterial community in Col-0 leaves at phylum level are dominated by Proteobacteria (92.2% relative abundance), Actinobacteria (5%), Firmicutes (1.1%) and Bacteroidetes (0.7%) in this study, which are similar (Pearson's correlation = 0.96, *P* = 0.037) to the *Arabidopsis* leaf natural communities found by Bai and colleagues³⁵. At the order level, we found that 9 of the top-10 orders (Burkholderiales, Actinomycetales, Methylophilales, Sphingomonadales, Rhizobiales, Pseudomonadales, Enterobacterales, Flavobacterales and Caulobacterales) in our total Col-0 leaf bacterial community are also identified as the top-10 orders of *Arabidopsis* leaf natural communities found by Bai and colleagues³⁵.

Bacterial genome assembly and taxonomic classification

To isolate high molecular weight genomic DNA of bacterial isolates in SynCom^{Col-0} and SynCom^{*mfec*}, we used a modified EZNA Bacterial DNA Kit Protocol (OMEGA Bio-Tek, UAS). Quality of genomic DNA was analysed by agarose gel (1% (w/v)) and quantified by Qubit. Approximately 1 µg of genomic DNA was used for Oxford Nanopore bacterial sequencing and 30 µg of genomic DNA for Illumina sequencing.

Using the software Canu v.1.7 with default parameters⁴⁰, raw nanopore reads were corrected, trimmed and then assembled into long contigs. A majority of the assembled genomes of SynCom^{Col-0} and SynCom^{*mfec*} are complete or near complete, with genome size ranging from 3.54 to 9.42 Mb (Supplementary Table 3). To further improve the quality of the assembled genomes, we resequenced full genomes for 27 relevant strains using the Illumina sequencing platform. After removing adaptor sequences, trimming and removing low-quality reads using the software Sickle³⁶ v.1.33, more than 24 Gb high-quality short reads were generated (Supplementary Table 3). We applied high-quality short reads from Illumina sequencing platform to correct the assembled genomes using program Pilon⁴¹ v.1.22. To generate accurate taxonomic information for the sequenced genomes, the average nucleotide identity of the whole genome with references and phylogenetic relationship with references on the basis of 120 marker genes were inferred using program gtdbtk⁴² v.0.1.3. We also used Ribosomal Database Project tools to infer the taxonomy of sequenced isolates based on full-length 16S rRNA genes using software Mothur⁴³ v.1.34.2. The maximum-likelihood phylogenetic tree for sequenced isolates were constructed based on the full-length 16S rRNA gene using MEGA7⁴⁴. A total of 100 bootstrap replicates were made.

Binary interaction

Bacterial strains were individually cultured on R2A plates at 28 °C for 1–2 days. One full inoculation loop of bacteria was suspended in 3 ml R2A medium (OD₆₀₀ of 2.1 to 4.0). For making a bacterial lawn, 2.6 ml of a 'target' bacterial suspension was added to 40 ml of molten R2A agar pre-cooled to 42 °C, gently mixed and then poured into two square Petri dishes (100 × 100 mm). Two microlitres of each 'attacker' strain was spotted onto the plate and incubated at room temperature for 3 days at which photographs were taken to observe inhibition zones. Two technical repeats were performed for each strain, strain had two technical repeats, and experiment was conducted three times. In total, 2,116 binary interactions were examined.

Characterization of the *ben3* mutant

The *ben3(cad1^{S20F})* mutant was backcrossed to wild type Col-0 to generate a mapping-by-sequencing population. Of 24 F1 plants grown in soil, all showed wild type Col-0 phenotype (that is, no spontaneous dysbiosis symptoms), indicating mutant phenotype is a recessive trait. F1 plants were allowed to self and produce a segregating F2 population. Of 376 F2 plants, 88 showed mutant phenotype (that is, spontaneous dysbiosis symptoms), whereas the rest (288 plants) were similar to the wild type. The observed 1:3 (mutant:wild type) phenotype segregation

Article

ratio suggests the mutant phenotype is caused by a single mutation on a nuclear gene. To identify the causative mutation, genomic DNA was extracted from 50 F2 plants exhibiting mutant phenotype and 50 F2 plants exhibiting wild-type phenotype and pooled into mutant-like and wild type-like pools. Pooled DNA samples were submitted to the Michigan State University RTSF Genomic Core facility for library preparation (Illumina TruSeq Nano DNA Library Preparation Kit) and sequenced on Illumina HiSeq 4000 platform in a 2×150 bp paired-end sequencing format. 70.3 and 94.3 million reads were obtained for mutant-like and wild-type-like pools, respectively. Whole-genome resequencing data were analysed following methods developed by Austin and colleagues⁴⁵, with minor package and version changes. Adaptor sequences and poor-quality reads were trimmed off using Trimmomatic (v.0.33). Reads were aligned to *Arabidopsis* TAIR10 genome using bowtie2 (v.2.3.1). Alignments were coordinate-sorted using SAMtools (v.1.5); PCR duplicate reads were removed using picardTools (v.1.89). Variances were called using bcftools (v.1.2). Candidate causative mutations were analysed using SHOREmap v.3.0⁴⁶ with algorithms developed for recessive mutation within a backcrossing population.

Production of 35S::*CAD1* transgenic *Arabidopsis*

The *CAD1* coding sequence was amplified by PCR using the following primers. Sense primer: **CACCATGGAGAACCGTAAAGGAGGAACT** (start codon in bold); antisense primer: **TCAATAATTAGCAACGAACATTC** (stop codon in bold). The amplified *CAD1* fragment was cloned into pENTR/D cloning vector (Invitrogen), and transferred by LR recombination into the binary expression vector pB7-35S::His-Flag-GW⁴⁷ to generate a 35S::His-Flag-*CAD1* (named 35S::*CAD1* for short). The binary vectors containing 35S::*CAD1* gene was introduced into *Agrobacterium tumefaciens* C58C1 by electroporation. *Arabidopsis* plants were transformed using the floral-dip method⁴⁸. Glufosinate ammonium (Basta) was used to select for transgenic T1 plants, which were further screened by western blot using a CAD1-specific antibody. Homozygous T3 plants expressing fusion proteins were used for analyses. CAD1-specific antibody was prepared against the C-terminal 240 amino acids. The NdeI-Xhol fragment of *CAD1*_{322–561} was cloned into the pET28a vector (Novagen) to overexpress a His₆-CAD1 fusion protein. Sense primer: **CATATGTGGGCTCCCGAACAGAGTAACCTCC** (NdeI site bold, start codon underlined); antisense primer: **CTCGAGTCAATAATTAGCAACGAAATCTTC** (XhoI site bold, stop codon underlined). Guinea pigs were injected with purified His₆-CAD1 protein to raise CAD1-specific antibody (Cocalico Biologicals).

Data analysis, statistics and experimental repeats

Plants of different genotypes (Col-0, *mfec* and *cad1*^{S20SF}) were grown side by side to minimize unexpected environmental variations during growth and experimentation. Leaf samples of similar ages were collected and assessed randomly for each genotype. Researchers were not blinded to allocation during experiments and outcome assessment. This is in part because different plant genotypes under study (Col-0, *mfec* and *cad1*^{S20SF}) exhibit very visually distinct phenotypes, making blinding not possible. Routine practices included more than one author observing and assessing phenotypes whenever possible. The specific statistical method used, the sample size, the number of experimental repeats and the results of statistical analyses are described in the relevant figure legends. Sample size was determined on the basis of experimental trials and with consideration of previous publications on similar experiments to allow for confident statistical analyses. Two-tailed Mann–Whitney *U*-test or one-way or two-way ANOVA with Tukey's test was used for multiple comparisons within a dataset, with significance at $P < 0.05$. ANOVA was performed with GraphPad Prism software. Statistical significance of alpha-diversity between plant genotypes were determined via Mann–Whitney *U*-test. The Benjamini–Hochberg method⁴⁹ was applied to correct the *P* values after performing multiple comparisons. Differential ASVs representing unique bacterial

16S rRNA sequences of endophytic phyllosphere microbiota between Col-0 and *mfec* mutants grown in potting soil supplemented with SynCom strains were identified with a negative binomial generalized linear model (GLM) in the edgeR package⁵⁰. The Benjamini–Hochberg method false discovery rate was applied to correct the *P* values after performing multiple comparisons. ASVs with false discovery rate below or equal to 0.05 were considered differentially colonized (that is, enriched or depleted in *mfec* compared to Col-0).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw source 16S rRNA gene sequences from this project are available in the Sequence Read Archive database under BioProject PRJNA554246, accession numbers SAMN12259846 to SAMN12260169. Bacterial genome source data are available in the Sequence Read Archive database under the BioProject PRJNA55902. Source Data for Figs. 1–4 and Extended Data Figs. 3, 4, 6–8 are provided with the paper.

Code availability

Scripts used in the microbiota analyses are available at <https://github.com/godlovesiaolin/A-genetic-network-for-host-control-of-phyllosphere-microbiota-for-plant-health>. All other software used in this study are cited in the text.

32. Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J. & Katagiri, F. Network properties of robust immunity in plants. *PLoS Genet.* **5**, e1000772 (2009).
33. Kremer, J. M. P. et al. FlowPot axenic plant growth system for microbiota research. Preprint at *bioRxiv* <https://doi.org/10.1101/254953> (2018).
34. Easlon, H. M. & Bloom, A. J. Easy Leaf Area: automated digital image analysis for rapid and accurate measurement of leaf area. *Appl. Plant Sci.* **2**, 1400033 (2014).
35. Bai, Y. et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**, 364–369 (2015).
36. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
37. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, ej.171.200 (2011).
38. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
39. Werner, J. J. et al. Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J.* **6**, 94–103 (2012).
40. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
41. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
42. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
43. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
44. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
45. Austin, R. S., Chatfield, S. P., Desveaux, D. & Guttmann, D. S. Next-generation mapping of genetic mutations using bulk population sequencing. *Methods Mol. Biol.* **1062**, 301–315 (2014).
46. Sun, H. & Schneeberger, K. SHOREmap v3.0: fast and accurate identification of causal mutations from forward genetic screens. *Methods Mol. Biol.* **1284**, 381–395 (2015).
47. Lee, C. M., Adamchek, C., Feke, A., Nusinow, D. A. & Gendron, J. M. Mapping protein–protein interactions using affinity purification and mass spectrometry. *Methods Mol. Biol.* **1610**, 231–249 (2017).
48. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J. Mol. Cell Biol.* **16**, 735–743 (1998).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
50. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
51. Nomura, K. et al. A bacterial virulence protein suppresses host innate immunity to cause plant disease. *Science* **313**, 220–223 (2006).
52. Nomura, K. et al. Effector-triggered immunity blocks pathogen degradation of an immunity-associated vesicle traffic regulator in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **108**, 10774–10779 (2011).

Acknowledgements We thank members, including B. Hsueh, of the He laboratory for technical help and insightful discussions; T. Gong for help with testing leaf necrosis phenotypes of plants grown on MS agar plates; and H. Tanaka and J. Friml for *ben3/cad1^{S205F}* seeds. This project was supported by funding from National Institutes of Health (GM109928), the Department of Energy (the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science; DE-FG02-91ER20021 for *ben3/cad1^{S205F}* mutant characterization) and Plant Resilience Institute at Michigan State University for support of optimization of the GnotoPot system (to S.Y.H.) and by funding from the CAS Center for Excellence in Molecular Plant Sciences and the Institute of Plant Physiology and Ecology, Chinese Academy of Sciences (to X.-F.X.).

Author contributions X.-F.X. and S.Y.H. conceptualized, designed the experiments and co-supervised the project. T.C. and K.N. performed most of experiments; X.-F.X. performed initial 16S sequencing set up and sample collection while at Michigan State University; R.S. performed GnotoPot experiments; X.W. performed 16S bioinformatics analysis; J.X. performed bacterial genome analysis; L.Y. performed the MS plate assay for Col-0 and the *mfec* mutant;

B.C.P. performed 16S bioinformatics analysis. L.M. was involved in *cad1*-related experiments; J.K. was involved in initial 16S RNA gene sequencing design; Y.C. was involved in mapping the *cad1* mutation; L.Z. performed phylogenetic analysis of *CAD1* and *MIN7* genes and advised on statistical analyses; N.W. and E.W. advised on bioinformatics and statistical analyses. T.C., X.-F.X. and S.Y.H. wrote the manuscript with input from all co-authors. X.W. and R.S. contributed equally as co-second authors.

Competing interests The authors declare no competing interests.

Additional information

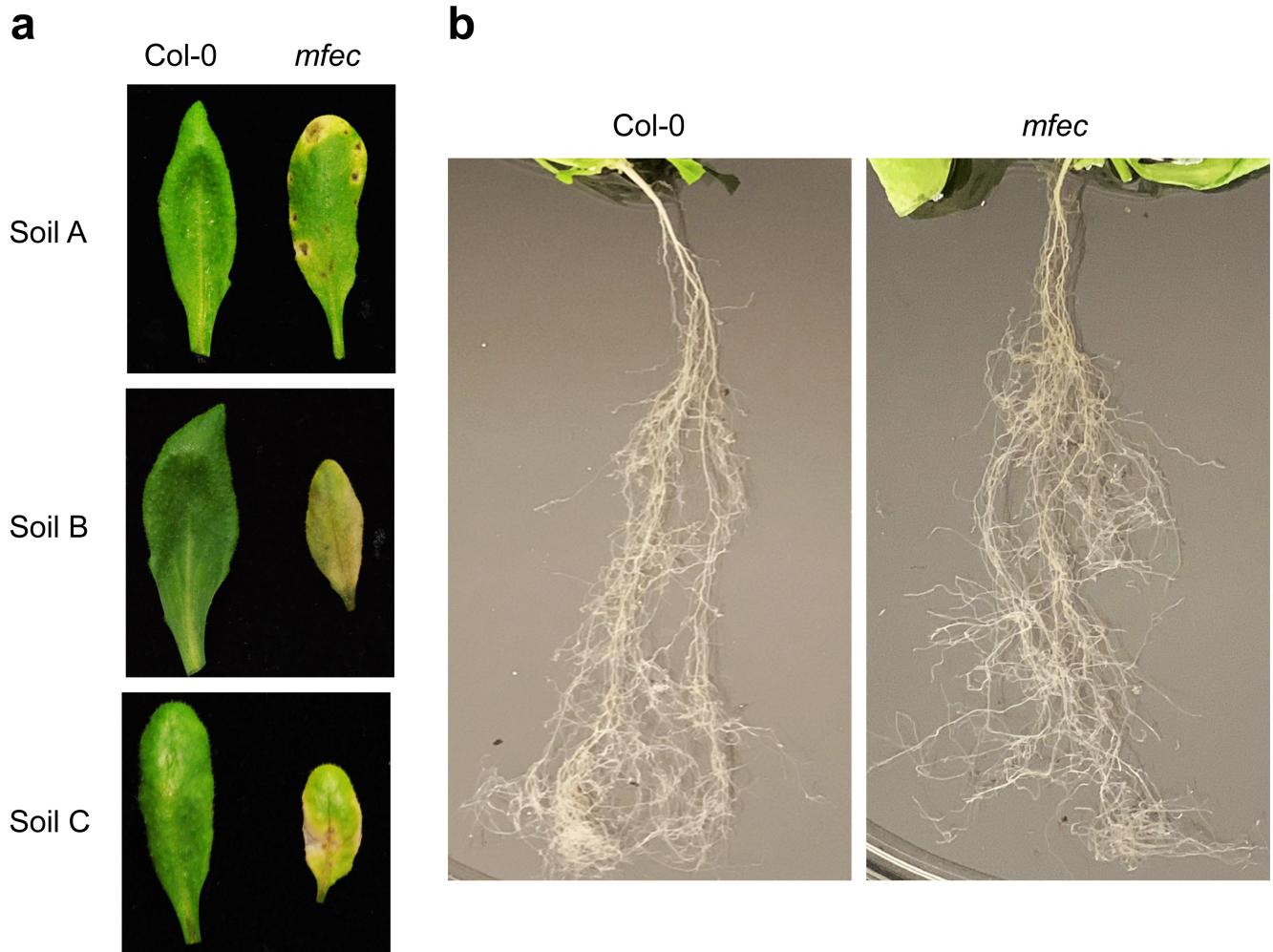
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-020-2185-0>.

Correspondence and requests for materials should be addressed to X.-F.X. or S.Y.H.

Peer review information *Nature* thanks Steven Lindow and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

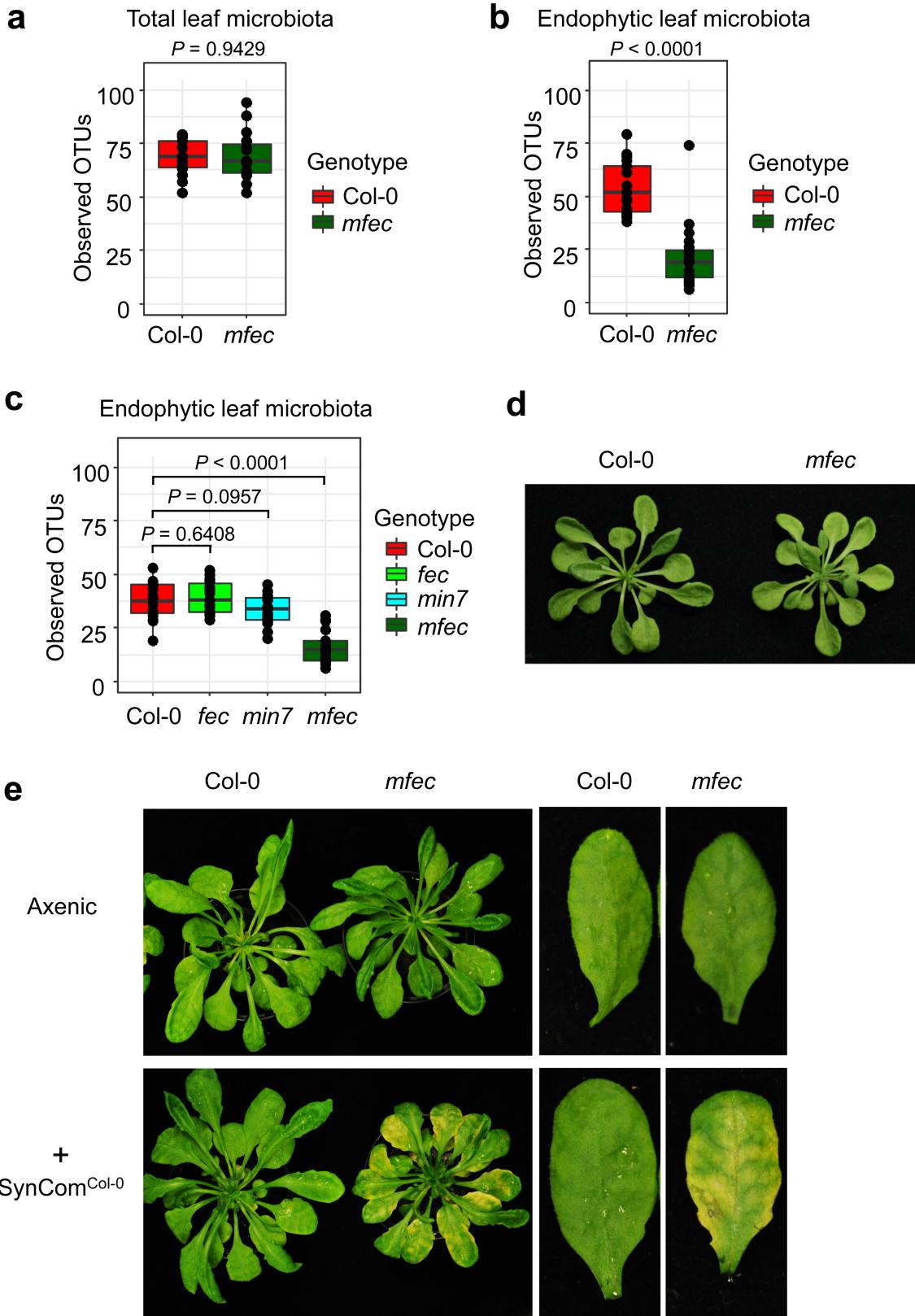
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



Extended Data Fig. 1 | Leaf and root appearance of soil-grown Col-0 and *mfec* plants. **a**, Leaf appearance of Col-0 and *mfec* plants grown in *Arabidopsis* mix soil (soil A) and Michigan State University (MSU) community agricultural soil (soil B; equal parts MSU community soil, medium vermiculite and perlite) or organic seed starter premium potting mix (Espoma) (soil C) for 6.5 weeks. Images were taken 5 days (soil A) or 11 days (soil B and soil C) after plants were

shifted to high humidity (~95%). Representative leaf images are shown. **b**, Root appearance of Col-0 and *mfec* plants grown in *Arabidopsis* mix soil for five weeks and shifted to high humidity (~95%) for 5 days. Representative root images are shown. Experiments in **a**, **b**, were repeated three times with similar results.

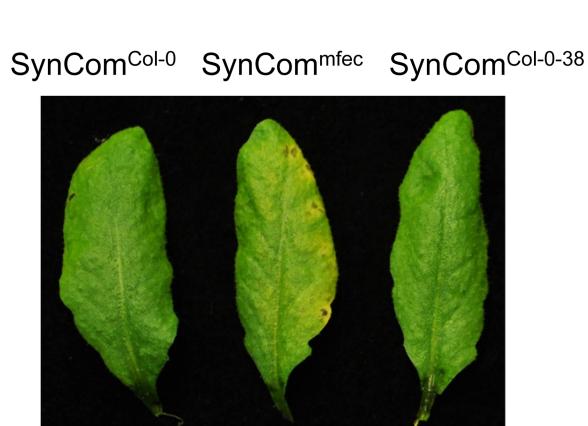
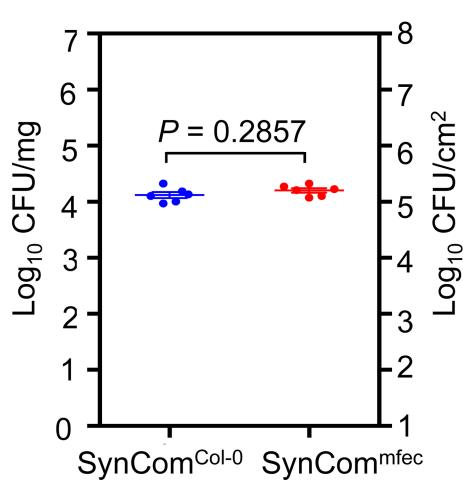
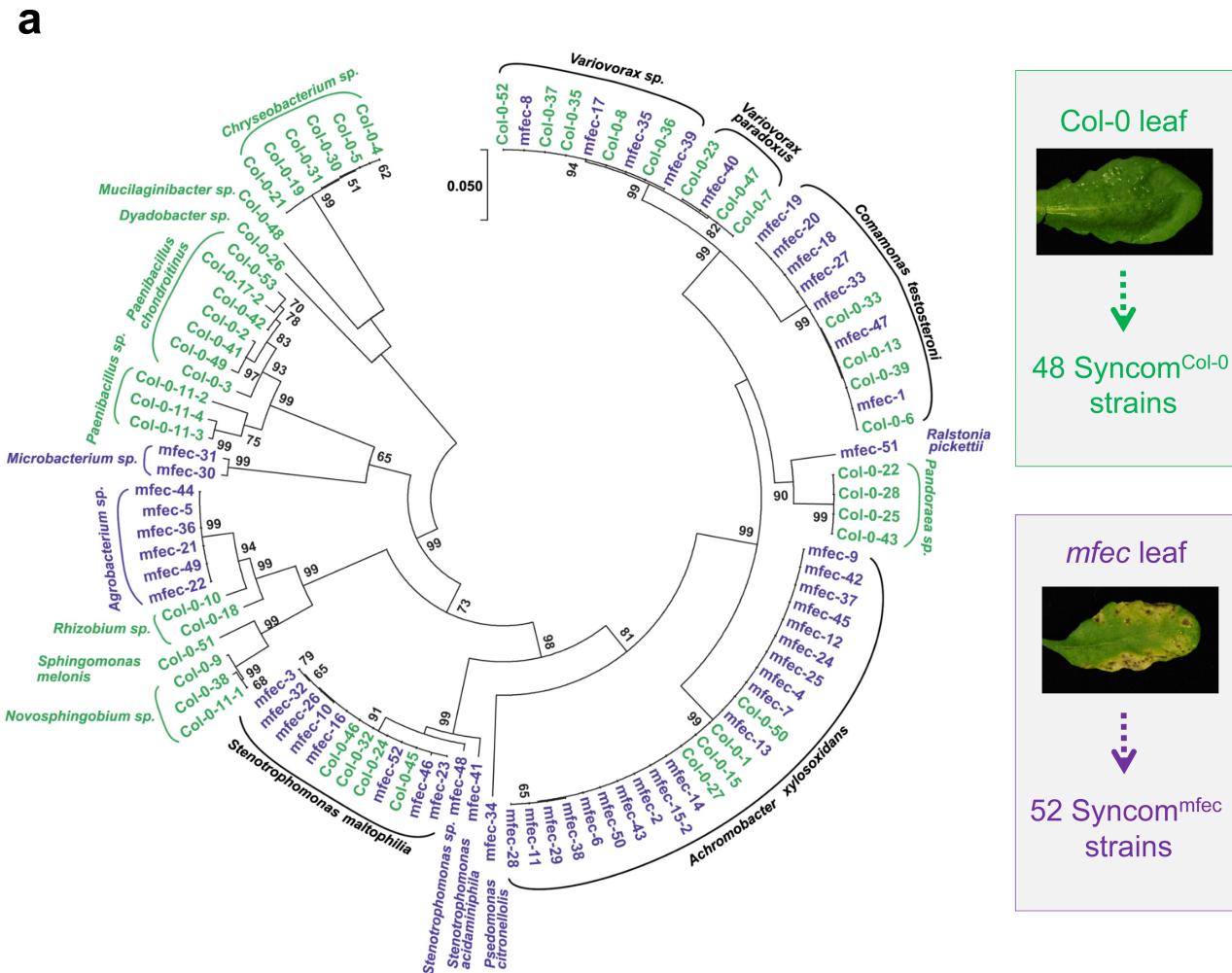


Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Observed OTUs of total and endophytic leaf bacteria in different plant genotypes and requirement of microbiota for appearance of dysbiosis symptoms in *mfec* leaves. **a, b**, Observed OTUs of total (**a**) and endophytic leaf bacteria (**b**) in Col-0 and *mfec* plants, which were grown in *Arabidopsis* mix soil and shifted to high humidity for 5 days. **c**, Observed OTUs of endophytic leaf microbiota in Col-0, *fec*, *min7* and *mfec* plants supplemented with SynCom^{Col-0}. In box plots, the centre line represents the median, box edges show the 75th and 25th percentiles, and whiskers extend to 1.5× the interquartile range. Two-tailed Mann–Whitney *U*-test. *n*=15 (Col-0) and *n*=15 (*mfec*) biological replicates passing quality control for analysis of total leaf bacterial microbiota across 3 independent experiments; *n*=18 (Col-0)

and *n*=20 (*mfec*) biological replicates passing quality control for analysis of leaf endophytic bacterial microbiota across 4 independent experiments. *n*=20 (Col-0), *n*=19 (*fec*), *n*=19 (*min7*) and *n*=19 (*mfec*) biological replicates passing quality control for analysis of leaf endophytic bacterial microbiota with SynCom^{Col-0} across 4 independent experiments. **d**, Leaf appearance of Col-0 and *mfec* plants grown in sterile MS agar plates. Pictures were taken 5 days after shifting plates to high humidity (~95%). **e**, Leaf appearance of Col-0 and *mfec* plants grown in GnotoPots in the absence (axenic) or presence of SynCom^{Col-0} for 6.5 weeks. Plants were then shifted to high humidity (~95%) for 10 days, before images were taken. Rosette leaf images are representative of at least four replicated experiments.

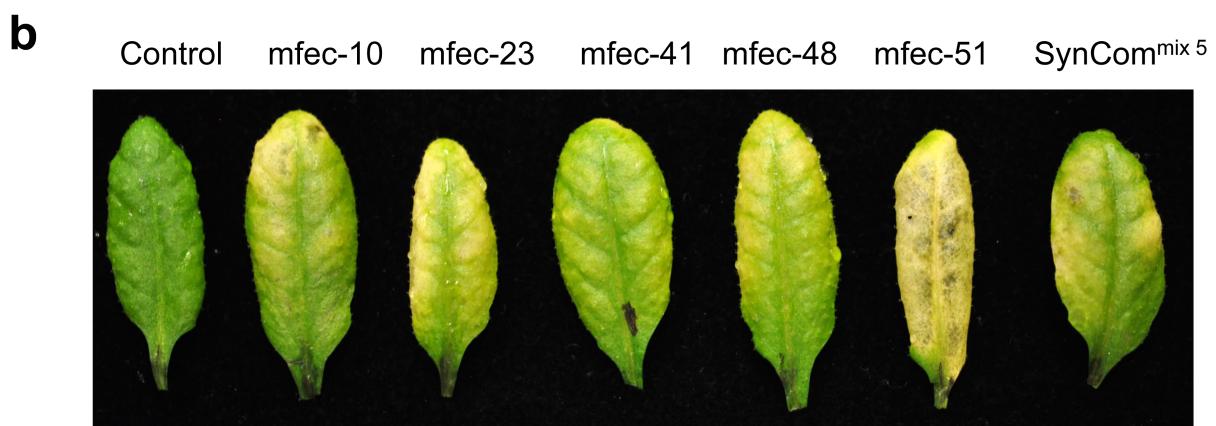
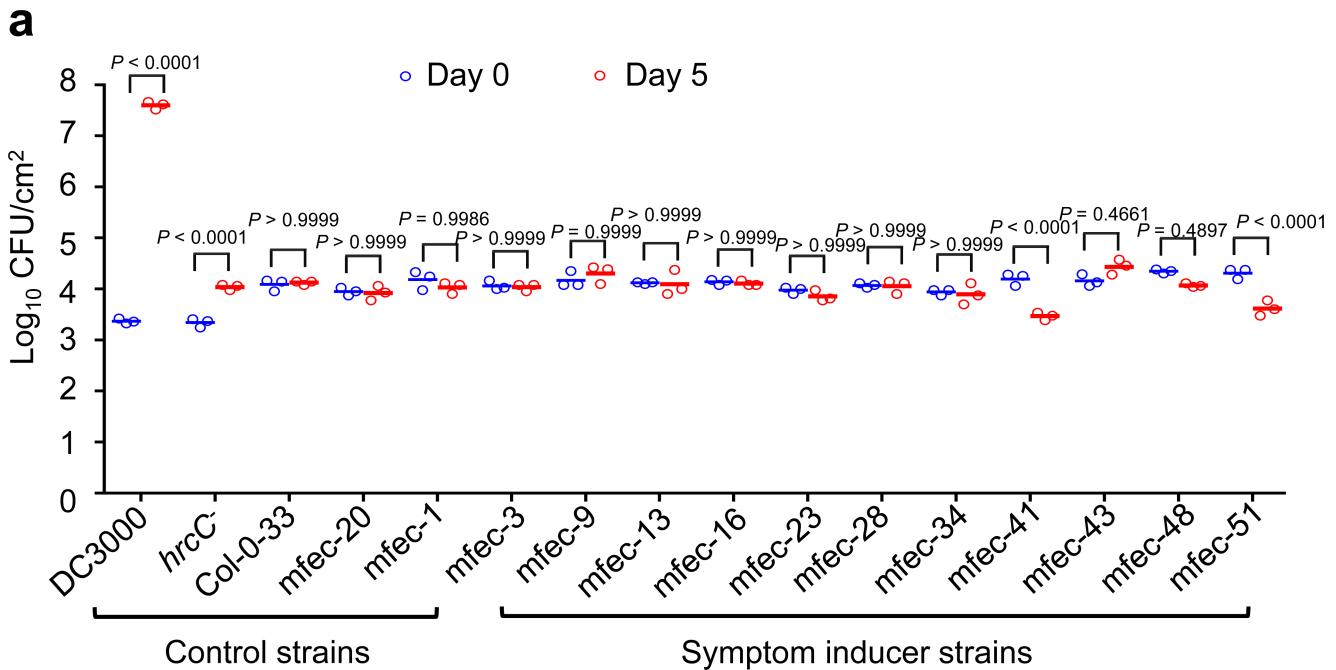


Extended Data Fig. 3 | See next page for caption.

Article

Extended Data Fig. 3 | A Maximum-likelihood phylogenetic tree for genome-sequenced bacterial isolates in *SynCom*^{Col-0} and *SynCom*^{mfec}. **a.** Tree was constructed on the basis of the full-length 16S rRNA gene using MEGA7. A total of 100 bootstrap replicates were made, and bootstrap values are indicated at the branch points. Colours represent bacterial isolates from different plant genotypes: *mfec* mutant (purple); Col-0 (green). In total, 48 strains were derived from healthy Col-0 endophytic leaves and 52 strains were derived from *mfec* endophytic leaves displaying dysbiosis symptoms. **b.** Col-0 leaves were syringe-infiltrated with *SynCom*^{Col-0} and *SynCom*^{mfec} at 1×10^7 CFU ml⁻¹; infiltrated plants were kept under ambient humidity for 1 h for water to evaporate. Bacterial populations were then determined after plant

leaves returned to pre-infiltration appearance. Colony-forming units were normalized to tissue fresh weight (left) and leaf disk areas (right). Statistical significance was determined by two-tailed Mann–Whitney *U*-test. $n=6$ biological replicates, data are mean \pm s.e.m. Experiments were repeated three times with similar results. **c.** Col-0 plants were syringe-infiltrated with *SynCom*^{Col-0}, *SynCom*^{mfec} or *SynCom*^{Col-0-38} (with 10 Firmicutes removed from *SynCom*^{Col-0}) at 1×10^7 CFU ml⁻¹. Inoculated plants were kept under high humidity (~95%), and leaf images were taken 7 days after infiltration. Experiments were repeated three times with similar results. Images are representative of leaves from four plants.



Extended Data Fig. 4 | Multiplication- and dysbiosis-symptom phenotypes of bacterial strains in Col-0 leaves. **a**, Population sizes (\log_{10} CFU/cm²leaf area) of bacterial strains in Col-0 leaves on day 0 (1h after leaf infiltration) and day 5 after leaf infiltration with each strain at 1×10^6 CFU ml⁻¹. The experiment was carried out at ~95% humidity. DC3000, *Pst* DC3000 (pathogenic on Col-0 plants); *hrcC*⁻, a nonpathogenic mutant of DC3000 defective in type III secretion; Col-0-33, mfec-20 and mfec-1, control strains that do not induce dysbiosis symptoms (Supplementary Table 1); other mfec strains, induce

dysbiosis symptoms (Supplementary Table 1). Statistical analysis was performed by two-way ANOVA with Tukey's test. $n = 3$ biological replicates, data are mean \pm s.e.m. Experiments were repeated twice with similar results. **b**, Leaf dysbiosis symptoms 7 days after infiltration of leaves of 4.5-week-old Col-0 plants with indicated mfec strains or SynCom^{mix5} at 1×10^7 CFU ml⁻¹. The experiment was carried out at ~95% humidity. SynCom^{mix5} is a mix of mfec-10, mfec-23, mfec-41, mfec-48 or mfec-51 with equal OD₆₀₀ values. Experiments were repeated three times with similar results.

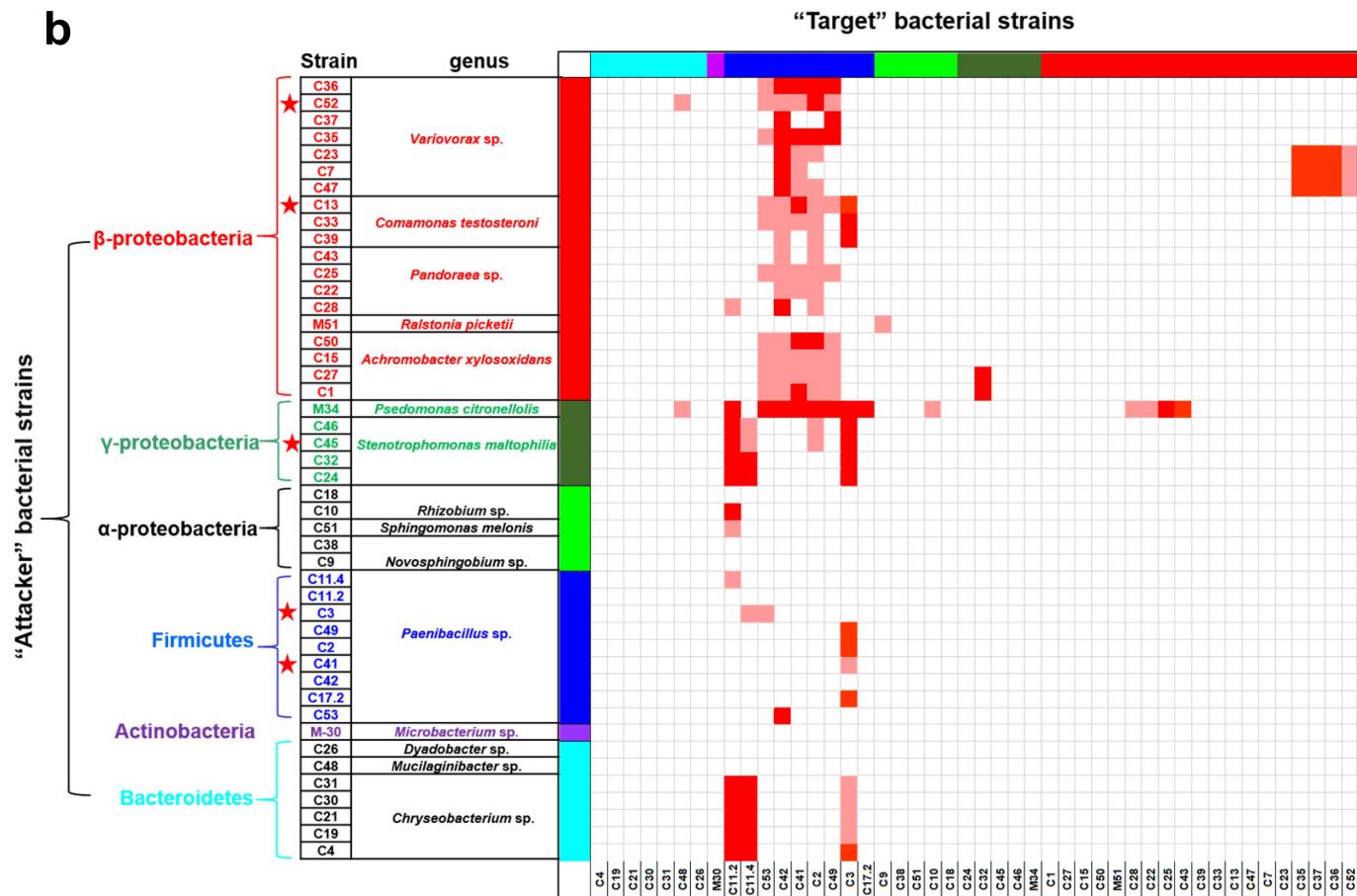
Article

a

Strong halo weak halo no halo

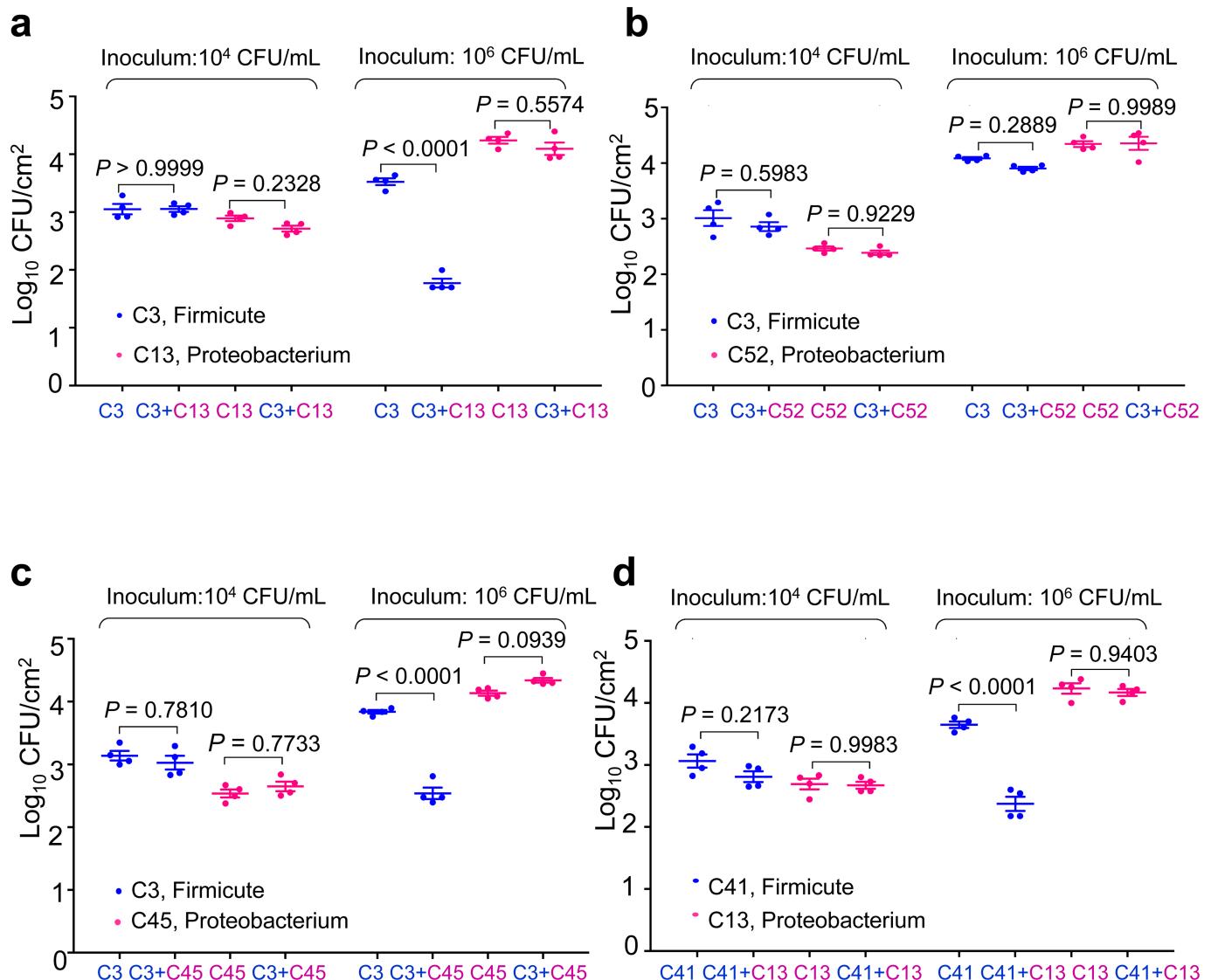


b



Extended Data Fig. 5 | Binary inter-bacterial inhibition. a, Examples of inhibitory halos are labelled with strong, weak or no inhibition. b, Binary inhibition assays (2,116 combinations) on a R2A plate of 46 strains that represent all bacterial species identified in SynCom^{Col⁰} and SynCom^{mfec}. Target bacterial strains are presented along the horizontal axis, whereas attacker bacterial strains are listed vertically. A large or clear halo, indicative of strong

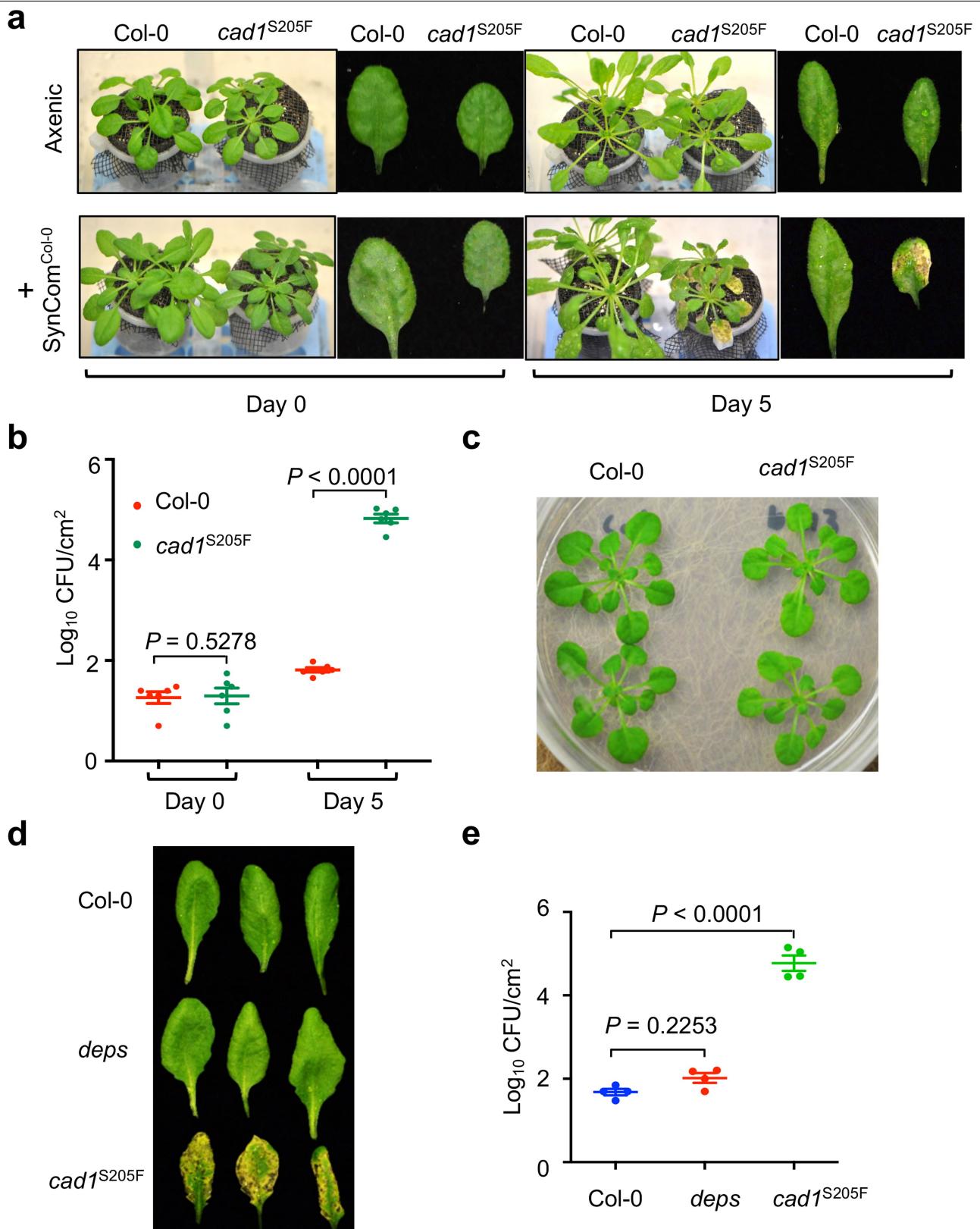
binary inhibition, is represented by a red-filled cell; a small or less transparent halo, indicative of weaker binary inhibition, is represented by a pink-filled cell; the absence of halo is represented in white. Strains labelled with a star were used for the in planta binary inhibition assay in Extended Data Fig. 6. Experiments were repeated three times with similar results.



Extended Data Fig. 6 | In planta binary inhibition. **a**, In planta inhibition of Firmicutes by Proteobacteria strains that displayed a strong inhibitory effect in R2A agar plate assay. Leaves of Col-0 plants were syringe-infiltrated with *Paenibacillus chondroitinus* (C3; a Firmicutes) alone, *Comamonas testosteroni* (C13, a Proteobacterium) alone or C3 and C13 together at 1×10^4 CFU ml⁻¹, corresponding to approximately 1×10^2 CFU cm⁻² leaf area; or 1×10^6 CFU ml⁻¹, corresponding to approximately 1×10^4 CFU cm⁻² leaf area. After infiltration plants were maintained under high humidity (~95%) for 5 days before bacterial populations (\log_{10} CFU/cm² leaf area) were determined. **b**, Similar to **a**, but with a non-inhibitory binary interaction between strains C3 and *Variovorax* sp. C52 (a Proteobacterium). **c**, Leaves of Col-0 plants were syringe-infiltrated with *P. chondroitinus* (C3; a Firmicutes) alone, *Stenotrophomonas maltophilia* (C45, a Proteobacterium) alone or C3 and C45 together at 1×10^4 CFU ml⁻¹ or 1×10^6 CFU ml⁻¹. **d**, Leaves of Col-0 plants were syringe-infiltrated with *P. chondroitinus* (C41; a Firmicutes) alone, *C. testosteroni* (C13, a Proteobacterium) alone or C41 and C13 together at 1×10^4 CFU ml⁻¹ or 1×10^6 CFU ml⁻¹. After infiltration, plants were maintained under high humidity (~95%) for 5 days before bacterial populations were determined. One-way ANOVA with Tukey's test. $n = 4$ biological replicates, data are mean \pm s.e.m. Experiments were repeated three times with similar results.

(a) *P. chondroitinus* (C3; a Firmicutes) alone, *Stenotrophomonas maltophilia* (C45, a Proteobacterium) alone or C3 and C45 together at 1×10^4 CFU ml⁻¹ or 1×10^6 CFU ml⁻¹. (b) *P. chondroitinus* (C41; a Firmicutes) alone, *C. testosteroni* (C13, a Proteobacterium) alone or C41 and C13 together at 1×10^4 CFU ml⁻¹ or 1×10^6 CFU ml⁻¹. After infiltration, plants were maintained under high humidity (~95%) for 5 days before bacterial populations were determined. One-way ANOVA with Tukey's test. $n = 4$ biological replicates, data are mean \pm s.e.m. Experiments were repeated three times with similar results.

Article

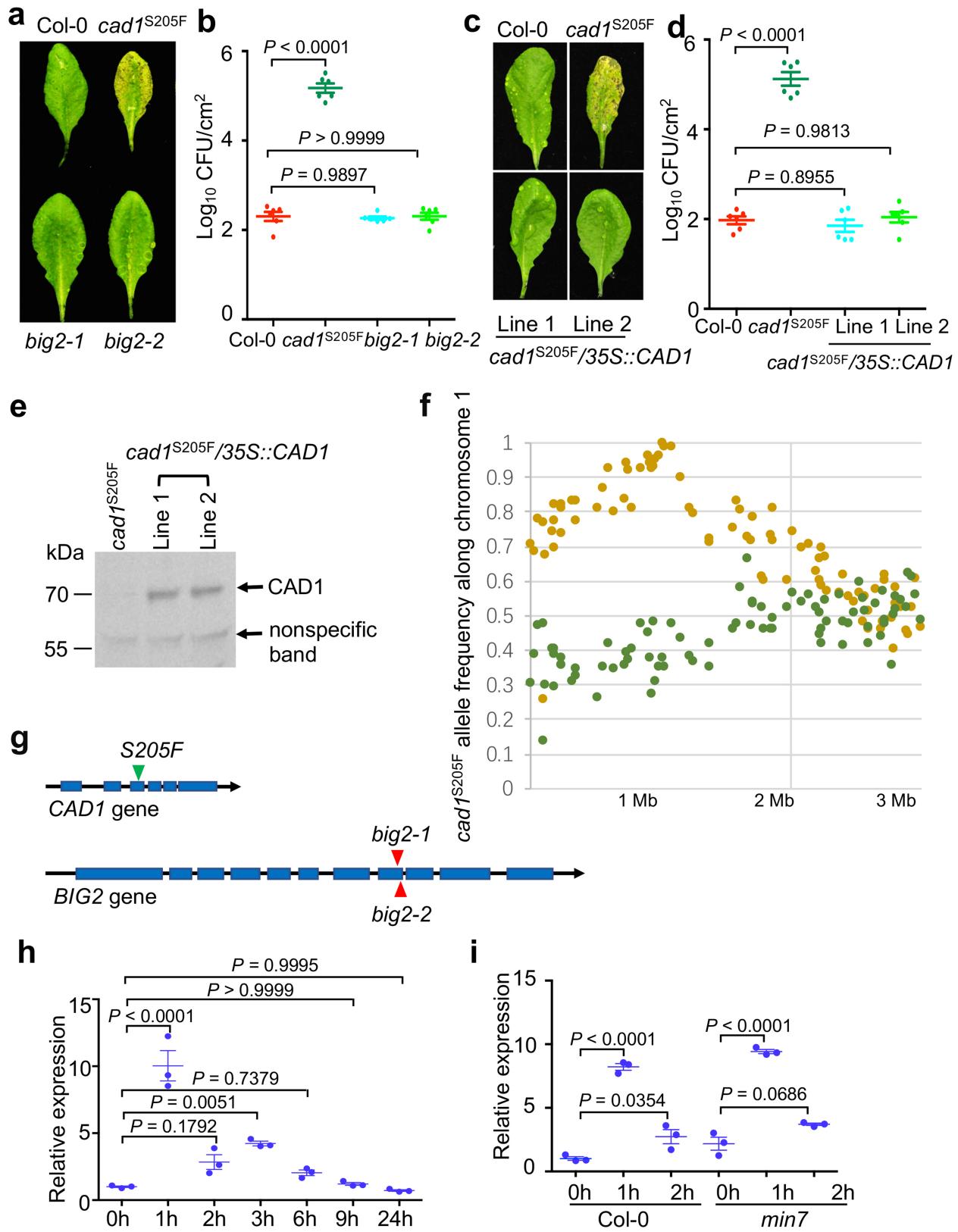


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Appearance and bacterial populations in Col-0 and *cad1*^{S20SF} plants before and after shift to 95% humidity. **a**, Leaf appearance of 5-week-old Col-0 and *cad1*^{S20SF} plants grown in the absence (axenic) or presence of SynCom^{Col-0} in the FlowPot gnotobiotic system (see Methods). Images were taken before (day 0) and 5 days after plants were shifted to high humidity (~95%). **b**, Levels of endophytic bacterial community (\log_{10} CFU/cm²leaf area) in the presence of SynCom^{Col-0} in the FlowPot gnotobiotic system. One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n=6$ biological replicates. Experiments were repeated three times with similar results. **c**, Leaf appearance

of Col-0 and *cad1*^{S20SF} plants grown in enclosed sterile LS agar plates for 4 weeks. **d**, Leaf appearance of 5-week-old Col-0, *deps* and *cad1*^{S20SF} plants grown in *Arabidopsis* mix 5 days after plants were shifted to ~95% relative humidity. **e**, Levels of endophytic leaf microbiota (\log_{10} CFU/cm²leaf area) in 5-week-old Col-0, *deps* and *cad1*^{S20SF} plants 5 days after plants were exposed to high humidity (~95%). One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n=4$ biological replicates. Experiments were repeated three times with similar results.

Article



Extended Data Fig. 8 | See next page for caption.

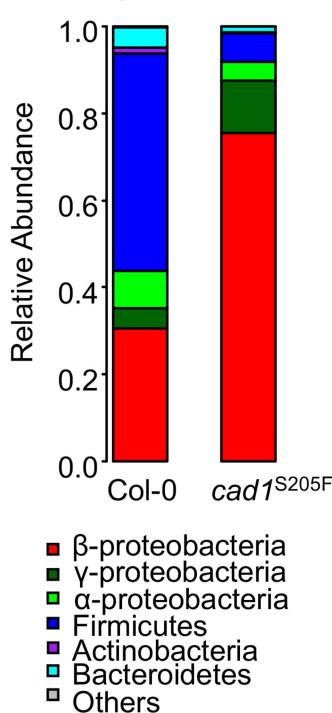
Extended Data Fig. 8 | Identification of a *cad1* mutation responsible for dysbiosis in the *cad1* mutant. **a**, Leaf appearance of 4.5-week-old Col-0, *cad1*^{S20SF} and *big2* plants grown in redi-earth potting soil. Images were taken at day 5 after plants were shifted to 95% humidity. **b**, Bacterial populations (\log_{10} CFU/cm² leaf area) of the endophytic bacterial community. One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n = 6$ biological replicates. Experiments were repeated three times with similar results. Two independent T-DNA insertion lines of *BIG2* were analysed with similar results (*big2*-1, SALK_033446 and *big2*-2, SALK_016558). **c–e**, Appearance (**c**) and endophytic bacterial populations (**d**; \log_{10} CFU/cm² leaf area) in Col-0, *cad1*^{S20SF} and *cad1*^{S20SF}/35S::*CAD1* plants at day 5 after plants were shifted to high humidity. Plants were grown in redi-earth potting soil for 4.5 weeks before they were shifted to high humidity. One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n = 6$ biological replicates. Experiments were repeated three

times with similar results. **e**, Two independent different complementation lines (*cad1*^{S20SF}/35S::*CAD1* line 1 and *cad1*^{S20SF}/35S::*CAD1* line 2) were analysed with similar results and protein levels were confirmed by western blot with the CAD1 antibody. Uncropped gel image is shown in Supplementary Fig. 2. **f**, *cad1*^{S20SF} genomic mapping. Green and brown dots indicate wild type-like and *cad1*^{S20SF}-like allele frequencies, respectively (detailed information in Supplementary Table 6). **g**, Schematic of mutations in *big2* and *cad1*^{S20SF} mutants. **h, i**, Quantitative PCR analyses of *CAD1* transcript in Col-0 (**h**) and *min7* (**i**) plants grown in *Arabidopsis* mix soil. Five-week-old Col-0 and *min7* leaves were infiltrated with 1 μ M flg22 and collected at the indicated time points. Transcript levels were normalized to that of the *PP2AA3* gene. One-way ANOVA with Tukey's test. Data are mean \pm s.e.m., $n = 3$ biological replicates. Experiments were repeated three times with similar results.

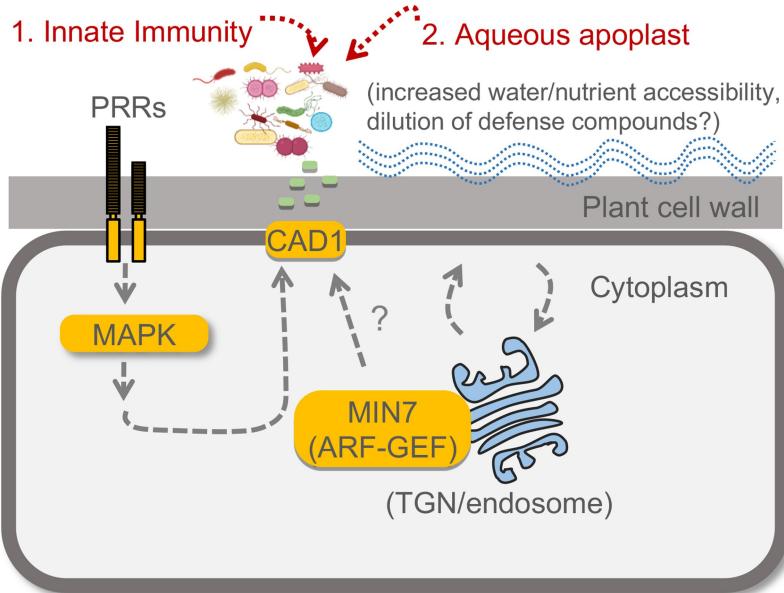
Article

a

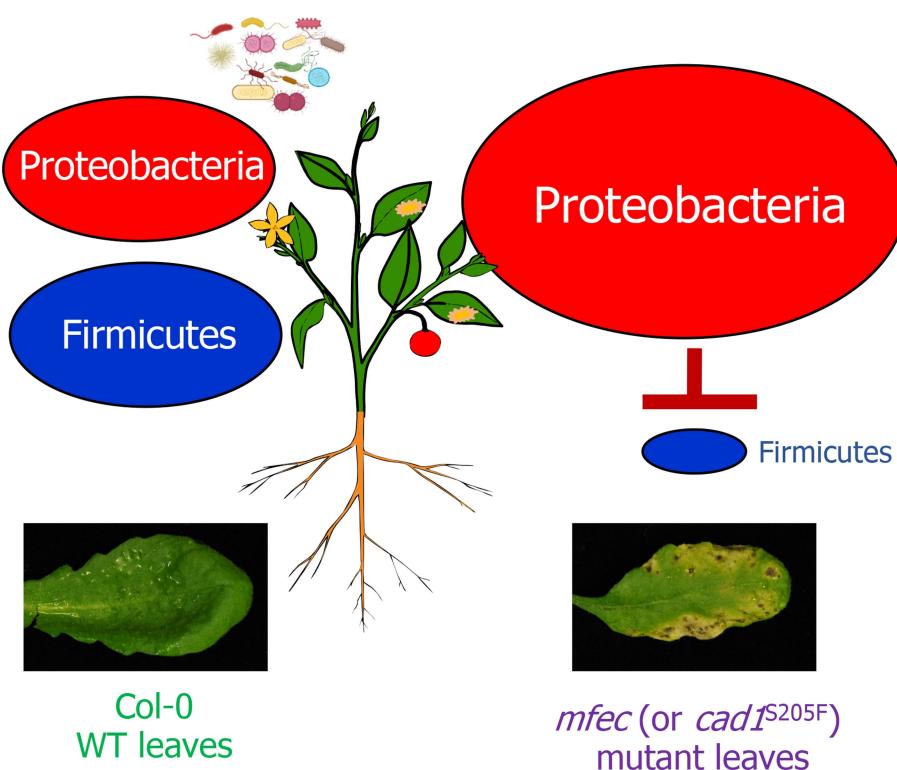
Endophytic leaf microbiota



b

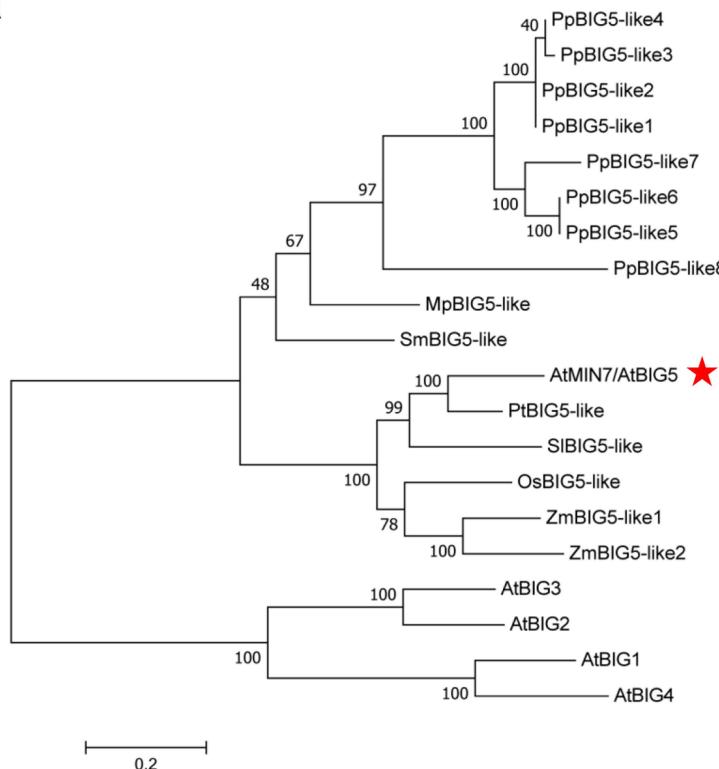
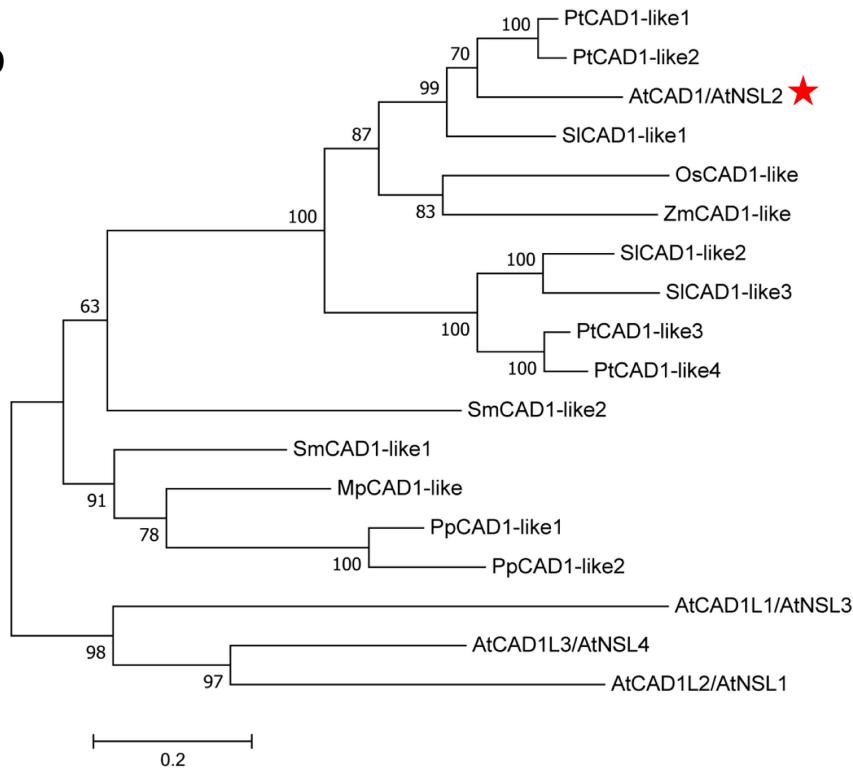


c



Extended Data Fig. 9 | A model for plant control of endophytic phyllosphere microbiota. **a**, The 16S rRNA gene-sequence profiles of endophytic leaf bacteria in Col-0 and *cad1*^{S205F} plants supplemented with SynCom^{Col-0}. Data presentation and statistical analysis as in Fig. 1d. $n=20$ (Col-0) and $n=20$ (*cad1*^{S205F}) biological replicates. **b**, A simplified diagram depicting pattern-triggered immune signalling, MIN7 and CAD1 as three components of a putative genetic framework for controlling endophytic bacterial microbiota,

which live outside a plant cell. MIN7 has previously been shown to be involved in regulating callose deposition^{51,52} and aqueous microenvironment in the leaf apoplast (that is, extracellular space)¹. **c**, Large shifts in the level and composition of endophytic leaf microbiota in wild-type Col-0 versus *mfec* (or *cad1*^{S205F}) leaves in part via competition between Proteobacteria and Firmicutes. Some components in **b** and **c** were drawn using tools in biorender.com.

a**b**

Extended Data Fig. 10 | See next page for caption.

Article

Extended Data Fig. 10 | Phylogenetic trees of protein sequences of MIN7 and CAD1 homologues from different plant species. **a, b,** Protein sequences of *A. thaliana* AtMIN7 (also known as AtBIG5) (AT3G43300.1) (**a**) and AtCAD1 (also known as AtNSL2) (AT1G29690.1) (**b**) were used for comparisons by Blast search against the proteome of *Arabidopsis* and seven other plant species (<https://phytozome.jgi.doe.gov/>). Homologues with *E* values lower than E^{100} were selected to generate phylogenetic trees across taxa, and only homologues specific to the AtMIN7 or AtCAD1 clade were presented with selected proteins from *Arabidopsis* as outgroups. Bootstrap values were

obtained from 1,000 replicates using the maximum-likelihood algorithm using MEGA7. The scale bar represents 0.2 substitutions per amino acid site. The genes are listed in Supplementary Table 7. AtMIN7 and AtCAD1 are highlighted with red stars. Abbreviations: BIG, BREFELDIN A-INHIBITED GUANINE NUCLEOTIDE-EXCHANGE PROTEIN; NSL, NECROTIC SPOTTED LESIONS; At, *A. thaliana*; Mp, *Marchantia polymorpha*; Os, *Oryza sativa*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Sm, *Selaginella moellendorffii*; Sl, *Solanum lycopersicum*; Zm, *Zostera marina*.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Not applicable.

Data analysis

Description of published software used for data analysis is provided and cited in the Methods section (pages 9–14) and figure legends. These include:

- Canu version 1.7: <https://canu.readthedocs.io/en/latest/>
- Sickle version 1.33: <https://github.com/najoshi/sickle>
- Pilon version 1.22: <https://github.com/broadinstitute/pilon>
- Gtdbtk version 0.1.3: <https://github.com/Ecogenomics/GTDBTk>
- Mothur version v.1.34.2: <https://www.mothur.org/>
- MEGA7: <https://www.megasoftware.net/>
- Easy Leaf Area: <https://github.com/heaslon/Easy-Leaf-Area>
- QIIME 2 Core 2018.11 distribution: <https://qiime2.org/>
- Cutadapt: <https://github.com/qiime2/q2-cutadapt>
- DADA2: <http://benjneb.github.io/dada2/>
- edgeR package: <https://bioconductor.org/packages/release/bioc/html/edgeR.html>
- Trimmomatic (v0.33): <http://www.usadellab.org/cms/?page=trimmmatic>
- bowtie2 (v2.3.1): <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>
- SAMTools (v1.5): <http://www.htslib.org/>
- picardTools (v1.89): <https://broadinstitute.github.io/picard/>
- bcfTools (v1.2): <http://www.htslib.org/>
- SHOREmap v3.0: <http://bioinfo.mpiipz.mpg.de/shoremap/>

Scripts Scripts used in the microbiota analyses are available at <https://github.com/godlovesxiaolin/A-genetic-network-for-host-control-of-phyllosphere-microbiota-for-plant-health>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw 16S rRNA gene sequences from this project are available in the SRA (Sequence Read Archive) database under the BioProject PRJNA554246, accession no. SAMN12259846- SAMN12260169. Bacterial genome data are available in the SRA database under the BioProject PRJNA555902.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size and the results of statistical analyses are described in the relevant figure legends. Sample size was determined based on experimental trials and with consideration of previous publications on similar experiments to allow for confident statistical analyses. No statistical methods were used to predetermine sample sizes.
Data exclusions	No data that pass quality control were excluded from analysis.
Replication	The number of replication for each experiment (at least two repeats, but mostly three times) is described in the relevant figure legends. Results were reproducible in all repeats with the same trend.
Randomization	Plants of different genotypes (Col-0, mfec, and cad1) were grown side by side to minimize unexpected environmental variations during growth and experimentation. Leaf samples of similar ages were collected and assessed randomly for each genotype.
Blinding	Researchers were not blinded to allocation during experiments and outcome assessment. This is in part because different plant genotypes under study (Col-0, mfec, and cad1) exhibit very distinct phenotypes visually; blinding was not possible. Routine practices included more than one author observing/assessing phenotypes, whenever possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

A Guinea Pig antibody to the *Arabidopsis* CAD1 protein; custom-made at Cocalico Biologicals, Inc (animal number: Guinea pig MSU-GP-19 and 20). Dilution at 1:5,000.

Validation

CAD1 antibody was validated by immunoblot of the CAD1 protein present in wild-type and absence in the cad1 mutant *Arabidopsis*.