The plant NADPH oxidase RBOHD is required for microbiota homeostasis in leaves

Sebastian Pfeilmeier, Gabriella C. Petti, Miriam Bortfeld-Miller, Benjamin Daniel, Christopher M. Field, Shinichi Sunagawa and Julia A. Vorholt

The plant microbiota consists of a multitude of microorganisms that can affect plant health and fitness. However, it is currently unclear how the plant shapes its leaf microbiota and what role the plant immune system plays in this process. Here, we evaluated Arabidopsis thaliana mutants with defects in different parts of the immune system for an altered bacterial community assembly using a gnotobiotic system. While higher-order mutants in receptors that recognize microbial features and in defence hormone signalling showed substantial microbial community alterations, the absence of the plant NADPH oxidase RBOHD caused the most pronounced change in the composition of the leaf microbiota. The rbohD knockout resulted in an enrichment of specific bacteria. Among these, we identified Xanthomonas strains as opportunistic pathogens that colonized wild-type plants asymptotically but caused disease in rbohD knockout plants. Strain dropout experiments revealed that the lack of RBOHD unlocks the pathogenicity of individual microbiota members driving dysbiosis in rbohD knockout plants. For full protection, healthy plants require both a functional immune system and a microbial community. Our results show that the NADPH oxidase RBOHD is essential for microbiota homeostasis and emphasizes the importance of the plant immune system in controlling the leaf microbiota.

In nature, plants are colonized by a diverse microbiota. The phyllosphere constitutes a vast microbial habitat in which microorganisms populate plant surfaces as epiphytes and the intercellular space inside leaves as endophytes. The functional repertoire of the microbiome expands the capacity of the plant to adapt to its environment, as beneficial microbes can promote plant growth and increase tolerance to abiotic and biotic stress. Plants must be prepared to recognize pathogens and mount defence responses against them, while being colonized by a complex microbiota that can potentially also trigger immunity and provide indirect plant protection. Therefore, plant immunity must be balanced to allow accommodation of a microbiota. Emerging evidence suggests that the plant immune system has a significant impact on the bacterial microbiota in the rhizosphere and phyllosphere. However, our current understanding of specific host factors involved in the perception and structuring of the microbiota and the molecular mechanisms underlying them is limited.

In recent years, experiments with synthetic communities (SynCom) have been introduced to identify host genetic factors that drive microbiota assembly under controlled environmental conditions. The availability of representative collections of strains isolated from Arabidopsis thaliana and the reproducible reconstitution of the microbiota to similar phylogenetic structures as found in nature provide opportunities to quantify community changes and probe available genetic resources to establish gene–phenotype causal relationships.

Here, we used a reductionist approach to link host immune components to the assembly of the bacterial microbiota. To test the impact of the plant immune system on the microbiota, we used single and higher-order mutants of well-described key immunity factors, and probed changes in the microbiota composition compared to the one that establishes on wild-type plants. To identify specific plant–microbiota interactions we adopted a SynCom approach to achieve strain-level resolution. We found a number of A. thaliana immunity mutants with altered bacterial community composition and demonstrate a critical function of an immunity-related NADPH oxidase in shaping the leaf microbiota. Ultimately, we explore the interplay between the host and individual community members and their roles in plant health and community composition.

Results

Bacterial community in the A. thaliana phyllo- and endosphere.

For microbiota reconstitution experiments, we used an established gnotobiotic growth system consisting of A. thaliana and a bacterial synthetic community based on the At-LSPHERE collection of genome-sequenced strains. The microbiota inoculum communities consisted of either 222 or 223 strains (termed SynCom-222 or SynCom-223), which represent the maximal genomic diversity of the At-LSPHERE and included strains with identical amplicon sequence variant (ASV) but different genomes. In follow-up experiments, we chose a subset of 137 strains (SynCom-137) in which each strain represents an ASV with a unique 16S ribosomal RNA gene (rDNA) amplicon sequence (Supplementary Table and Methods). Germ-free 10-day-old A. thaliana seedlings were inoculated with the SynCom and the aboveground parts of the plants were harvested 5.5 weeks after germination to examine the bacterial communities by 16S rDNA amplicon sequencing (Fig. 1a). The relative abundance of phyllosphere strains (Fig. 1b) was similar to the natural microbiota of A. thaliana at the phyllosphere level and was consistent with SynCom experiments from previous studies.

The community structure in the A. thaliana Col-0 phyllosphere comprised Alphaproteobacteria (36%), Betaproteobacteria (30%), Gammaproteobacteria (2%), Actinobacteria (19%), Bacteroidetes (12%) and Firmicutes (0.0005%) (Fig. 1b). In addition to these phyllosphere samples, we analysed the endosphere-enriched bacterial fraction (Fig. 1a and Methods). Comparing the community composition, we observed a relative enrichment of Gammaproteobacteria in the endosphere compared to the entire phyllosphere (Fig. 1b and...
Extended Data Fig. 1). The quantification of bacterial abundance by quantitative PCR (qPCR) using 16S rDNA copies per plant gene revealed reduced colonization of the endosphere. Similarly, colony-forming units (c.f.u.) per gram of fresh weight recovered from endosphere samples were approximately 1,000-fold lower than those recovered from phyllosphere samples (Fig. 1c), as expected.

The analysis of the relative abundance of strains revealed that a few bacteria were significantly enriched or depleted in the endosphere across three SynCom experiments (Fig. 1d). Xanthomonas Leaf131 in particular but also Pseudomonas Leaf59 were enriched in the endosphere samples, thus explaining the higher relative abundance of Gammaproteobacteria (Fig. 1b). In contrast, Methylphilus Leaf416, Aeromonas Leaf350 and Pseudorhodoferax Leaf265 were depleted, indicating that these strains or ASVs have a greater relative abundance on the leaf surface compared to the community in the apoplast (Fig. 1d). The experiment was repeated with SynCom-223 and SynCom-137 and revealed similar results (Source Data Fig. 1d), which shows that the strains in SynCom-137 are representative for the ASVs.

Because the differences in relative abundance suggest a distinct adaptation to an endophytic or epiphytic lifestyle of some strains, we conducted leaf infiltration experiments to test bacterial growth in the apoplast. Indeed, the endophytic-enriched strains Xanthomonas Leaf131 and Pseudomonas Leaf59 were able to grow to a greater extent (37- and 3.5-fold increase, respectively) after low-dose infiltration into leaves of soil-grown Col-0 plants, in contrast to the depleted strain Pseudorhodoferax Leaf265 (2.5-fold increase) (Extended Data Fig. 2), confirming the potential of the former, especially Xanthomonas Leaf131, to thrive as endophytes. No disease symptoms were observed after leaf infiltration, and bacterial titres were lower than those of the model pathogen Pseudomonas syringae pv. tomato DC3000 (Pst) (Extended Data Fig. 2).

Screening A. thaliana mutants for an altered bacterial community. To identify host genetic factors that affect the bacterial community composition in the phyllosphere, we tested A. thaliana mutants with defects in distinct parts of the immune system. These included mutants for immune signalling or regulators that are involved in pattern-triggered immunity (PTI)4, effector-triggered immunity (ETTI)5 and homeostasis of defence-associated plant hormones (Supplementary Table).

A. thaliana Col-0 wild-type and different mutants were inoculated with SynCom-222, and the community composition was determined by 16S rDNA amplicon sequencing. We examined the impact of the host genotype on the community composition in a principal component analysis (PCA), which quantifies how much variation in the strain (or ASV) abundances can be explained by individual genotypes, and determined the effect size and statistical significance by permutational multivariate analysis of variance (PERMANOVA). Bacteria inhabiting the apoplast are in closer contact with host cells and, as such, may be subject to stronger host control than bacteria on the leaf surface. Therefore, we primarily focused our analysis on the bacterial community in the endosphere compartment.

The largest impact on the endophytic community was observed in the A. thaliana mutant rbohD, which lacks the plasma membrane-localized NADPH oxidase Respiratory Burst Oxidase Homologue (RBOH) D, with a 25% effect size (Fig. 2a–c). Among the genotypes with defects in PTI signalling, the effect of rbohD was larger than that of other genotypes with defects in pattern recognition receptors (for example, fts2efcrerk1 (fec)4, lym3ym1, ref. 25, pepr1pepr2, ref. 26) or coreceptors (for example, bak1-5bbk1cerk1 (bbc)27). The genotypes compromised in plant hormone biosynthesis or perception had an effect size between 17% by jar1 (JA signalling)48 and 7% by ein2 (ethylene signalling)49. Notably, the higher-order mutants with defects in multiple hormone systems, jar1ein2npr1 (jen)50 and deez2e2pad4sid2 (deps)51, did not have a stronger impact on the community composition than the single mutants tested (Fig. 2a). The salicylic acid-dependent immunoregulatory mutant npr1 (ref. 52) and the stomatal opening regulator ost1 (ref. 53) had a minor impact on the endophytic community composition. The quadruple mutants min7jef (mjef) and min7bcb (mbbc52) also affected community assembly, corroborating a recent study reporting that these mutants are involved in bacterial microbe homeostasis53. The hyperimmune genotypes cpr5 (ref. 54) and cpr6 (ref. 55), which have constitutively high salicylic acid levels, also had an impact on the community.

In fact, all genotypes with large effects on the endosphere community also affected the entire phyllosphere community (Extended Data Fig. 3a). The malectin-like receptor-like kinase Peronia knock-out fer7 had the highest effect size (28%) on the phyllosphere community but was not tested for the endosphere community due to poor germination and impaired development56. In addition to the genotype effect on data variance in a PCA, we also analysed the variability of community composition on colonization of immunity mutants (Fig. 2c). The microbtiota variability in rbohD plants was higher than that in wild-type plants, while it was equally variable in jar1 and wild-type plants in all SynCom experiments.

Analysis of the community composition at higher taxonomic levels, such as phylum (or class for Proteobacteria), between individual genotypes revealed that several taxa were significantly altered in their relative abundance in rbohD, mbbc, bbc, mjef, cpr5, jar1 and fer (fer tested only for phyllosphere) compared to wild-type plants (Fig. 2d and Extended Data Fig. 3b).
Hierarchical clustering of the differential strain abundances between individual genotypes and Col-0 of endosphere-enriched strains grouped rbohD separately from the other genotypes, but no other significant clusters were observed (Extended Data Fig. 4a). Clustering analysis of the phyllosphere community profiles revealed clustering of all genotypes but fer and cpr5 separately (Extended Data Fig. 4b).

Overall, we found striking genotype effects on the leaf bacterial community in both endosphere-enriched and total phyllosphere samples.
Fig. 2 | Effect of plant genotype on the leaf endosphere community.  

**a.** The endosphere bacterial community (SynCom-222) of each plant genotype was compared to Col-0 wild-type plants in PCA (n = 12) followed by PERMANOVA (permutations, 10,000), and the effect size of the genotype was plotted in decreasing order. **b.** PCA of bacterial endosphere communities in Col-0 (blue), bbc (orange), jar1 (pink) and rbohD (green). PC1 and PC2 are principal components PC1 and PC2 with their explained variance (%). **c.** Exemplary PCA of endosphere communities in rbohD, jar1 and bbc. Effect size represents variance explained by genotype, and statistical significance is expressed with P values determined by PERMANOVA (Benjamini–Hochberg adjusted, n = 12). **d.** Relative abundance of phyla (and classes for Proteobacteria) of endosphere bacteria on the indicated plant genotypes. Genotypes are ordered by decreasing abundance of Gammaproteobacteria. Asterisks (or hashtags for Firmicutes) denote significant differences in taxa on a genotype compared to Col-0 in a two-sided t-test (P < 0.05, Benjamini–Hochberg adjusted, n = 12).  

**e.** Community spread of genotypes rbohD, jar1 and bbc relative to Col-0 in PCA. The community spread was calculated as the Euclidean distance of data points to the centroid in PCA. Distances of genotypes were normalized to the median distance of Col-0 as the z score. Community spread was calculated for SynCom-222, SynCom-223 and SynCom-137. Box plots show the median with upper and lower quartiles and whiskers present 1.5× interquartile range. Statistical differences were determined by two-sided t-test (n = 12; NS, not significant; *P < 0.05; **P < 0.01; ***P < 0.001).
Fig. 3 | The plant mutant rbohD shows a dysbiosis phenotype and assembles a microbiota enriched in Gammaproteobacteria. a, Representative pictures of 5-week-old Col-0 and rbohD inoculated with SynCom-137. Scale bar, 1 cm. b, Fresh weight of aboveground plant tissue of Col-0 and rbohD inoculated with SynCom-137 and SynCom-222. Columns show phyllosphere and endosphere samples from plants inoculated with SynCom-222, SynCom-223 or SynCom-137. Black rectangles show significant changes, \( P < 0.05 \) (n = 12, two-tailed Mann–Whitney U-test; \( *** P < 0.001 \)). c, Heatmap shows log2 fold changes of strains in phylogenetic order in rbohD compared to Col-0 wild-type plants. d, Volcano plot shows the relative abundance of rbohD-enriched strains of SynCom-137 in rbohD endosphere (log2 fold changes in rbohD compared to Col-0 with adjusted \( P < 0.05 \)).
**rbohD assembles a dysbiotic microbiota.** Because *rbohD* showed the largest impact on the endosphere community with an effect size of 25% (Fig. 2a), compared to 17% and lower for other immunity mutants, we examined the impact of RBOHD on the microbiota more closely. RBOHD is part of a family of NADPH oxidases that generate extracellular reactive oxygen species (ROS) and function in diverse physiological processes, including biotic and abiotic stress signalling. In particular, RBOHD is regulated through activated immune receptors and is responsible for inductive apoplastic ROS production during PTI and ETI, playing critical roles in microbe-associated molecular pattern (MAMP)-induced stomatal closure, cell wall damage-induced lignification and resistance to pathogens.

In addition to assembling an altered endophytic community, as described above, *rbohD* plants developed disease symptoms after inoculation with a SynCom (Fig. 3a), while axenic *rbohD* plants were indistinguishable from the wild-type. SynCom-induced disease was also reflected in the reduced average plant weight (Fig. 3b) and points towards a dysbiosis phenotype. Although symptoms could be observed in most *rbohD* plants, the disease severity varied (Fig. 3a). We classified SynCom-induced disease into five categories (1, healthy, to 5, dead), and increasing disease severity matched with decreasing plant fresh weight (Extended Data Fig. 5b). Next, we examined the dysbiotic microbiota of *rbohD* at the strain level by using the SynCom-137. A defined set of bacteria was affected in their relative abundance in the endosphere and phyllosphere compared to Col-0 (Fig. 3c). Mostly, Gammaproteobacteria (*Xanthomonas* Leaf131, *Serratia* spp. Leaf50 and Leaf51, *Pseudomonas* spp. Leaf83, Leaf58, Leaf127, Leaf15 and Leaf434), an Alphaproteobacterium (*Sphingobium* Leaf26), and two Actinobacteria (*Sanguibacter* spp. Leaf50, *Serratia* spp. Leaf53 and *Erwinia* Leaf265) showed a significant increase in relative abundance in *rbohD* compared to wild-type plants (Fig. 3c) and were among the top colonizers in the endosphere of *rbohD* (Fig. 3d). Notably, the same strains were consistently enriched in *rbohD* in three SynCom experiments, which underlines the robustness of the observed effect. SynCom-137 recapitulated the results of the full community, thus showing that single strains were representative of the ASV with respect to their behaviour in *rbohD*. Furthermore, most of the affected strains were increased in both the endosphere and phyllosphere of *rbohD* (Fig. 3c).

To investigate whether disease in *rbohD* is accompanied by an increased bacterial load, we assessed the bacterial cell numbers in the phyllosphere and endosphere. Overall bacterial abundance measured by c.f.u. and qPCR normalized to plant weight or plant genomes, respectively, was not significantly increased in *rbohD* compared to Col-0 across multiple experiments (Extended Data Fig. 5c). This finding is in line with previous plant pathology studies that did not observe a correlation between disease severity and pathogen titre, although a correlation had been described in other studies.

In addition to RBOHD, RBOHF also plays a role in ROS production and disease resistance, although to a lesser extent. We tested the double knockout mutant *rbohDrbohF* and found that it had an even stronger impact on SynCom-137 than *rbohD* alone. For example, multiple strains were specifically enriched in the endosphere of *rbohDrbohF* but not in *rbohD* (Extended Data Fig. 6b), such as *Rhizobium* Leaf311 and *Brevundimonas* Leaf168 (Alphaproteobacteria); *Pseudorhodoferax* Leaf265 and *Acidovorax* Leaf160 (Betaproteobacteria); *Serratia* Leaf50, *Erwinia* Leaf53 and *Pseudomonas* Leaf59 (Gammaproteobacteria); *Arthrobacter* strains Leaf145 and Leaf141, and *Plantibacter* Leaf314 (Actinobacteria).

The substantial impact of *rbohD* and *rbohDrbohF* compared to other PTI-associated genotypes in our experiments (Fig. 2a and Extended Data Fig. 6) could be explained by the central roles of NADPH oxidases in converging signalling pathways in immunity and beyond. Overall, our results suggest that dysbiosis of *rbohD* is caused by a shift in community structure with an increased abundance of certain taxa dominating the community.

**The bacterial leaf community contains opportunistic pathogens.** Many plant-associated bacteria carry immunomag MAMPs and can potentially trigger immunity. We selected phylogenetically diverse species, including *rbohD*-enriched strains, and tested whether they can elicit an apoplastic ROS burst. Extracts of most *rbohD*-enriched strains and other strains triggered plant ROS, suggesting that ROS-eliciting potential is common, especially among Beta- and Gammaproteobacteria (Extended Data Fig. 5e).

To identify the causal agent(s) of SynCom-induced disease in *rbohD*, we tested the pathogenicity of the individual SynCom-137 strains. We inoculated single strains onto germ-free *rbohD* seedlings and monitored the plant phenotype. On the basis of this screening we identified 28 phenotype-inducing strains that caused plant symptoms, such as necrotic lesions or curled leaves, indicating a strong host–microbe interaction (Extended Data Fig. 5a), with *Xanthomonas* Leaf131 triggering the most severe disease symptoms when inoculated alone (Fig. 4a,b). The At-LSPHERE strain collection contains another *Xanthomonas* strain, Leaf148, which shares the same 16S rDNA amplicon sequence but has only an 87% average nucleotide identity with Leaf131. We tested both *Xanthomonas* strains for their phenotype on Col-0, *rbohD* and *rbohDrbohF* complementation lines. While *rbohD* plants became diseased or died after inoculation with Leaf131 or Leaf148, Col-0 and the complemented *rbohD*/*rbohF* looked healthier and had higher plant weights (Fig. 4a,b). In addition, bacterial extracts of both *Xanthomonas* strains triggered a plant ROS burst, indicating that the plant is able to perceive these bacteria (Fig. 4c). Moreover, infiltration of *Xanthomonas* Leaf131 into leaves of soil-grown Col-0 wild-type and *rbohD* plants resulted in mild disease symptoms only in *rbohD* (Fig. 4d). We concluded that Leaf131 and Leaf148 are opportunistic pathogens that develop their deleterious potential on immunocompromised *rbohD* plants.

RBOHD activation in *A. thaliana* occurs through N-terminal phosphorylation by receptor-like cytoplasmic kinases, such as Botrytis-Induced Kinase 1 (BIK1). Multiple complementary and overlapping RBOHD phosphorylation sites have been reported to be critical for PTI signalling. We proved whether RBOHD phosphorylation sites can be linked to microbiota-induced disease. After validation of impaired MAMP-induced ROS production in RBOHD mutants (Extended Data Fig. 7a), we tested whether RBOHD with mutated phosphorylation sites, that is, *rbohD*/RBOHD-S39A-S393A-S343A and *rbohDrbohF*/RBOHD-S343A-S347A, could complement the *rbohD* disease phenotype after inoculation with SynCom-137 or single strains of *Xanthomonas* Leaf131 and Leaf148. The presence of SynCom-137 caused a significant weight reduction of *rbohD* and *rbohDrbohF* and, to a lesser extent, of the phosphorylation site mutants compared to their respective
controls (Fig. 4b). The intermediate disease phenotype of the RBOHD phosphorylation site mutants compared to the knockout lines rbohD and rbohD/RBOHD could be due to regulatory redundancy of multiple posttranslational modifications resulting in tunable activation of RBOHD and ROS production [46–48]. Inoculation with Xanthomonas Leaf131 and Leaf148 caused disease symptoms and reduced the average plant weight of rbohD and the two phosphorylation site mutants compared to their control lines Col-0, rbohD/RBOHD and rbohDrbohF/RBOHD, respectively (Fig. 4b and Extended Data Fig. 7b). In contrast, inoculation with Pst reduced the plant weight to a similar level on all genotypes irrespective of RBOHD presence, while colonization by the non-virulent Pst hrcC- did not affect the plant weight of the different genotypes (Extended Data Fig. 7b).

In summary, our data highlight the importance of RBOHD and its regulation via immunity-associated phosphorylation sites for plant health during colonization with opportunistic Xanthomonas pathogens and for microbiota homeostasis.

**Dysbiosis of rbohD is driven by opportunistic pathogens.** Due to the enrichment of Xanthomonas Leaf131 in rbohD (Fig. 3c,d) and the phenotype it caused in mono-association (Fig. 4a), we next wondered whether the opportunistic pathogen was sufficient to explain the phenotype in a community context and speculated on the relative contribution of plant immunity and the microbiota to plant health. We designed several synthetic communities by removing either the opportunistic pathogen Xanthomonas Leaf131 or all rbohD-enriched and plant phenotype-inducing (REPI) strains (Supplementary Table). We inoculated germ-free Col-0, rbohD and rbohD/RBOHD with the selected bacterial mixes. The full community control, that is, SynCom-137, reduced plant health and weight compared to axenic control plants only in rbohD, not in Col-0 and rbohD/RBOHD (Fig. 5), confirming our previous observation that SynCom causes disease in rbohD (Fig. 3a).

Xanthomonas Leaf131 single inoculation was detrimental to rbohD and reduced plant weight in Col-0 and rbohD/RBOHD, corroborating its pathogenic potential (Fig. 4). We then mimicked the dysbiotic community observed in rbohD plants by assembling SynCom-REPI, consisting of the 32 strains for which we observed phenotypes or enrichment in rbohD (Fig. 3c and Extended Data Fig. 5a). SynCom-REPI severely reduced the plant weight of rbohD and had a mild effect on Col-0. When Leaf131 or all SynCom-REPI members were removed from SynCom-137, the plant weight difference between Col-0 and rbohD disappeared and both genotypes had phenotypes that were not distinguishable from axenic control plants (Fig. 5). Because Xanthomonas Leaf131 is also part of the
REPI strains, we assumed that it is responsible for the negative effect on \( \textit{rbohD} \). Thus, we repeated the dropout SynCom experiment with an additional condition consisting of Leaf131 dropout from SynCom-REPI. Again, we observed that whenever the SynCom did not contain Leaf131, plant health and weight were increased in Col-0, \( \textit{rbohD} \) and \( \textit{rbohD}/\textit{RBOHD} \) (Extended Data Fig. 8). This shows that \textit{Xanthomonas} Leaf131 is responsible and sufficient for SynCom-induced disease in \( \textit{rbohD} \). Notably, a diverse bacterial community can alleviate the detrimental effects of the opportunistic pathogen \textit{Xanthomonas} Leaf131 on immuno-compromised \( \textit{rbohD} \); however, a healthy plant requires functional RBOHD (Fig. 5).

**Fig. 5 | \textit{Xanthomonas} Leaf131 causes dysbiosis in \( \textit{rbohD} \).** a, Fresh weight of Col-0, \( \textit{rbohD} \) and \( \textit{rbohD}/\textit{RBOHD} \) plants inoculated with 10 mM MgCl\(_2\) (axenic), SynCom-137, single strain Leaf131, SynCom-REPI (containing 32 \( \textit{rbohD} \)-enriched and phenotype-inducing strains), SynCom-137 without (w/o) Leaf131, SynCom-137 w/o REPI. Germ-free 10-day-old seedlings were inoculated with OD of 0.02 of SynCom-137, and other inocula with lower strain numbers were diluted with 10 mM MgCl\(_2\) to obtain equal amounts of cells from each strain in the inoculum. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. Significant differences were calculated with ANOVA and Tukey’s HSD post hoc test (\( n = 20 \), letters indicate significance groups, \( \alpha = 0.05 \)). b, Pictures of plants treated as indicated above. Genotypes Col-0, \( \textit{rbohD} \) and \( \textit{rbohD}/\textit{RBOHD} \) were grown in the same microbox in rows, and different genotypes are highlighted by coloured arrows (Col-0, blue; \( \textit{rbohD}/\textit{RBOHD} \), light blue and \( \textit{rbohD} \), green). Scale bar, 1 cm. The experiment was repeated twice with additional treatments (Extended Data Fig. 8).
Discussion
The assembly and maintenance of a healthy microbiome is crucial for host fitness\textsuperscript{6,48}. Previous knowledge about the function of key plant immunity genes makes it now possible to test how these genes influence not only pathogens, but also the entire plant microbiota and interactions within. By using a reverse genetic screen, we found multiple plant immunity mutants that affect bacterial leaf community assembly. In particular, we uncovered the pivotal role of the NADPH oxidase RBOHD to prevent microbiota dysbiosis in the phyllosphere. NADPH oxidases are evolutionarily conserved in multicellular eukaryotes with functions in defence, development and redox-dependent signalling\textsuperscript{6,24}. NADPH oxidases play a major role in gut microbe homeostasis in mammals\textsuperscript{69,70}, zebrafish\textsuperscript{45} and insects\textsuperscript{69,70}, highlighting the importance of understanding the interplay between ROS production, host immunity and microbiota.

Like to dysbiosis in the human intestine, which has been characterized by an increase in the relative abundance of Gammaproteobacteria\textsuperscript{68} and dysbiotic communities being more diverse than communities in healthy patients\textsuperscript{68}, we also found an enrichment of multiple Gammaproteobacteria (Fig. 3) and a higher variability among the microbiota between individual rbohD plants compared to wild-type (Fig. 2e). However, in contrast to observations made in the gut, we identified Xanthomonas spp. as the most conspicuous, acting as opportunistic pathogens when inoculated on plants alone (Fig. 4).

Xanthomonas spp. are common plant pathogens and are ubiquitously found in the phyllosphere microbiota on a variety of plants\textsuperscript{71,72}. The type-3 secretion system (T3SS) represents a chief virulence determinant in pathogenic Xanthomonas\textsuperscript{6}. A preliminary genome analysis of Xanthomonas Leaf131 and Leaf148 revealed the absence of a T3SS (ref. \textsuperscript{17}), which is in accordance with their opportunistic lifestyle. Note also that the opportunistic Xanthomonas pathogens identified here have been isolated from phenotypically asymptomatic A. thaliana plants in their natural environment\textsuperscript{73}. Correspondingly, numerous ecological studies have detected pathogenic species, including Xanthomonas, as members of the plant microbiota that are asymptomatic for the host in a community context but cause disease during mono-association\textsuperscript{74,75}. Our data support the concept of conditional pathogens\textsuperscript{76}, which defines species with pathogenic potential as an integral part of the microbiota that are kept in check by other community members and the host, and underlines the importance of both the genotype and the microbiota to plant protection (Fig. 5).

The mechanisms by which the microbiota reduces the prevalence of opportunistic pathogens might be manifold and remain to be elucidated. Screening of bipartite interactions between At-LSPHERE strains in vitro did not reveal growth inhibition of Xanthomonas Leaf131 and Leaf148 by other community members\textsuperscript{77}; nonetheless, the plant could be protected through a concerted action of multiple commensals or indirectly via stimulation of plant immunity\textsuperscript{78,79}. Our data support the concept of conditional pathogens\textsuperscript{80}, which defines species with pathogenic potential as an integral part of the microbiota that are kept in check by other community members and the host, and underlines the importance of both the genotype and the microbiota to plant protection (Fig. 5).

The mechanisms by which the microbiota reduces the prevalence of opportunistic pathogens might be manifold and remain to be elucidated. Screening of bipartite interactions between At-LSPHERE strains in vitro did not reveal growth inhibition of Xanthomonas Leaf131 and Leaf148 by other community members\textsuperscript{77}; nonetheless, the plant could be protected through a concerted action of multiple commensals or indirectly via stimulation of plant immunity\textsuperscript{78,79}. Our data support the concept of conditional pathogens\textsuperscript{80}, which defines species with pathogenic potential as an integral part of the microbiota that are kept in check by other community members and the host, and underlines the importance of both the genotype and the microbiota to plant protection (Fig. 5).

Apoplastic ROS production by RBOHD is a typical response during plant-pathogen interactions. As the community shift in rbohD could be observed in endosphere and phyllosphere samples, the internal and external bacterial populations might form a continuum\textsuperscript{1}. The absence of RBOHD could facilitate endophytic growth of certain taxa, which then relocate to the leaf surface. RBOHD-produced ROS can act directly as antimicrobials as well as in local and systemic signalling\textsuperscript{81,82}. In our experiments, many bacteria were able to induce a ROS burst, indicating that microbiota members have the capacity to trigger PTI. Future experiments will help to disentangle whether RBOHD selectively affects individual bacteria, for example, through antimicrobial ROS, cell wall modification via cross linking or downstream immune responses\textsuperscript{83,84,85}, and whether pathogenic Xanthomonas act as keystone species by affecting other microbes either directly or indirectly via the plant.

Recently, Chen et al. proposed a genetic network to control the microbiota and prevent dysbiosis\textsuperscript{86}. Our data indicate that RBOHD and RBOHF could be central parts of such a network and integrate signals from converging pathways, as rbohD and rbohDrbohF knockout lines had a more pronounced impact on the endophytic community than other immunity mutants under our experimental conditions. While Gammaproteobacteria generally benefited from gene knockouts in bbc mbbc, mfcf, rbohD and rbohDrbohF, we could see overlap but also differences at the strain level between the genotypes (Fig. 2d and Extended Data Fig. 6). Notably, RBOHD has been identified as a central link between the immune signalling pathways PTI and ETI\textsuperscript{47,48}, which could explain its extraordinary effect on the microbiota. In addition, RBOHD could be a convergence point of biotic and abiotic stress signalling\textsuperscript{87,88} that jointly regulate microbiota homeostasis. It remains to be determined how the different host components interact to control the microbiota. The impact of RBOHD on selected members of the bacterial microbiota potentially has strong implications for plant defence against pathogens as these bacteria may prime plant immunity and convey health-protective properties.

In summary, we identified a pivotal function of the NADPH oxides RBOHD and RBOHF in maintaining microbiota homeostasis in the phyllosphere to prevent dysbiosis. Our findings emphasize the role of the commensal microbiota together with a functional immune system in the control of opportunistic pathogens.

Methods
Plant growth conditions. Gnotobiotic plants were grown in calcined clay (Diamond Pro Calcined Clay Drying Agent) supplemented with 0.5x Murashige and Skoog (MS) medium including vitamins, pH 5.8 (M0222.0050, Duchefa), in round gamma-irradiated microboxes (no. O118/80 + OD118 with green filter lid, Saco2) as described previously\textsuperscript{89}. A. thaliana seeds were surface sterilized using 70% EtOH for 2 min and 7% NaOCl containing 0.2% Triton X-100 for 8 min (ref. \textsuperscript{90}). Seeds were washed with water six times and stratified for 4 d in the dark at 4°C. Several seeds were individually placed in five positions per microbox to account for unsuccessful seed germination. Additional seedlings were removed 1 d before inoculation to keep five plants per microbox. Microboxes only contained plants of one genotype and treatment for the initial screening experiment with SynCom-223. Each plant was watered with 200 µl of 0.5x MS medium, pH 5.8, at 9, 24 and 38 d after germination. A list of plant genotypes and their references can be found in the Supplementary Table.

Growth chambers (CU-41L4, Percival) had air flow from the bottom of each shelf and were equipped with full-spectrum lights (Philips Master TL D 18W/950 Graphica, Philips) and UVA-/UVB-emiting light (Sylvania Reptistar F18 W/6500K, Sylvania) to simulate natural conditions. The combined light intensity was set to 220 µmol m\textsuperscript{-2} s\textsuperscript{-1} for wavelengths of 400–700 nm (measured with a PAR quantum sensor SKP215, Skye Instruments) and 5.4 µmol m\textsuperscript{-2} s\textsuperscript{-1} for wavelengths of 300–400 nm (ultraviolet (UV) light measured with MU-250, Apogee). The growth chambers were set to 22°C and 54% relative humidity and run under an 11-h light cycle. Due to light activity during the day, the temperature inside microboxes increased to 24°C and water condensates appeared on sidewalls.

Synthetic community selection. SynCom composition was selected on the basis of the At-LSPHERE strain collection\textsuperscript{17}. SynCom-222 contained 223 strains (Supplementary Table) and covered maximal phylogenetic diversity to investigate changes in the community assembly on A. thaliana genotypes. SynCom-223 was used in a follow-up experiment and contained 223 strains (SynCom-223 additionally included Serratia Leaf50, which was left out in SynCom-222 due to concerns of being a strong pathogen, which turned out not to be the case). The VS-V7 region of the 16S rDNA gene amplified with the primers 799F\textsuperscript{92} and 1518R\textsuperscript{92} “allows us to distinguish 137 of these strains with 100% sequence identity representing ASVs. To be able to attribute changes in strain abundance to individual strains, we selected SynCom-137 containing one strain as a representative for each ASV. To distinguish whether the data shown in the figures belong to individual strains or an ASV containing multiple strains (as indicated in Supplementary Table), the ASVs with more than one strain are marked by a superscript circle after the strain name (that is, Xanthomonas Leaf131\textsuperscript{3} in the axis labels. Experiments with SynCom-137 were always done with a single strain for each ASV.)
Synthetic community mixing and plant inoculation. Bacteria were individually streaked on R2Aagar (Sigma-Aldrich) plates containing 0.5% v/v methanol (R2A-MeOH), grown for 4 d at 22°C, restreaked on fresh R2A-MeOH plates and grown for a further 3 d. To prepare Single-AMPhurl (SAM) inoculum, a ‘loopful’ of biomass from each strain was scraped off with a sterile 1-µl plastic loop and resuspended individually in 1 ml of 10 mM MgCl₂. Tubes containing the resuspended cells were vortexed for 5 min to disperse cell aggregates. Strains that were not fully resuspended after vortexing were treated with TissueLyzer II (Qiagen) at 30 Hz for 2×45 s. The SynCom inoculum was obtained by mixing each strain in an equal volume ratio. The SynCom inoculum was adjusted to an optical density (OD₆₀₀) of 0.02, and each plant was inoculated with 200 µl of bacterial solution or 10 mM MgCl₂.

Aliquots of 1 ml from the SynCom mixtures were taken for DNA extraction to determine the community composition of the inoculum. All strains could be detected in the SynCom inoculum mixes, and undetected strains were most probably due to insufficient sequencing depth (Extended Data Fig. 1 and Supplementary Table). In addition, a dilution series was made from all resuspended strains, and 4 µl was spotted on R2A-MeOH plates to count c.f.u. and verify the viability of each strain in the inoculum.

Dropout SynCom inocula was prepared as described previously. The resuspension of strains in a 2×10⁷ c.f.u.-µl-¹ volume containing 300 nM of each primer and 10 µl volume containing 50 µM each primer was used as described previously17. The 16S rDNA amplicon band was excised from the gel and DNA was extracted using a QIAquick Gel Extraction Kit (Qiagen). The DNA concentration of each subpool was measured and combined in volumes according to respective sample numbers.

Sequencing was performed on a MiSeq desktop sequencer (Illumina) at the Genetic Diversity Centre Zurich using the MiSeq reagent kit v3 (paired end, 2 × 300bp). Denaturation, dilution and addition of 15% PhiX to the DNA library were performed according to the manufacturer’s instructions. Custom sequencing primers were used as described previously17.

DNA extraction and 16S rDNA amplicon sequencing. Samples were freeze-dried at −40°C and 0.12 mbar for 16 h (Alpha 2-4 LD Plus, Christ) and pulverized at 30 Hz for 2×45 s (TissueLyzer II, Qiagen). DNA was extracted using the FastDNA SPIN Kit for Soil (MP Biomedicals) following the manufacturer’s instructions. The samples were transferred to DNA low-binding 96-well plates (Frame Star 96, SPIN Kit for Soil (MP Biomedicals) following the manufacturer’s instructions. The data from each SynCom experiment were analysed separately.

The 16S rDNA reference sequences of SAM strains were extracted as described previously17 and each strain is represented by one ASV reference. Paired-end sequencing reads were merged using the USEARCH v11.0.667-i86 linux64 command fastq_mergepairs, with a minimum overlap of 16 bp and a minimum identity of 90%. Merged reads were then filtered using fastq_filter with a maximum expected error of one and a minimum length of 200 bp. The filtered dataset was classified using the USEARCH v11.0.667-i86 linux64 command classifyﷺ using 100% identity to 16S rDNA reference sequences to generate an initial ASV table with a count for each reference per sample. Unidentified sequences of each sample (potentially originating from additional 16S rRNA gene copies of SynCom strains, from sample or plant contamination, or from DNA artifactuals of PCR amplification or sequencing) were dereplicated (fastx_uniques) and clustered using USEARCH v11.0.667-i86 linux64 (with a minimum cluster size of 1 and a fixed identity threshold of 97%). De novo operational taxonomic unit (OTU) clusters were annotated using the SILVA SSU Ref NR database (release 132) and the initial ASV table. The final ASV table was log-normalized and variance-stabilized by DESeq2 v.1.14.1 (ref. 94). The data from each SynCom experiment were analysed separately.

The effect on individual strains between the test and control conditions was visualized as heatmaps or volcano plots.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100. The strain phylogeny was bootstrapped using the explained by the compared factor and was calculated on Euclidean distances followed by a PERMANOVA to test for statistical significance using the adonis command of the package vegan v.2.5-4.

To summarize the relative abundance of each strain, the relative abundance values were calculated by proportional normalization of each sample by its sequencing depth.

To examine the effect on each variable between the test and control conditions, the output of DESeq2 log, fold-change values and P values (Wald test, Benjamin–Hochberg adjusted). The differential strain abundances between the test and control conditions were visualized as heatmaps or volcano plots.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.

Phylogenetic analysis. The strain phylogeny was constructed based on full-length 16S rRNA gene sequences (see Supplementary Table for accession numbers). Sequences were aligned using SINA v.1.3.3 (ref. 99) and SILVA SSU Ref NR database release 132. To build a maximum likelihood phylogeny from the alignment, PhyML v.3.3.2018214 was used with default parameters100.
subsequently washed in water three times. Leaf discs were taken with a cork borer filled with 200
g/ml of bacterial biomass in 1 ml of 10 mM MgCl2 (final inocula had an OD600 between 0.02
and 0.04). Each inoculum was prepared from a 1:10 dilution of the resuspension of a ‘loopful’
from the primers for background amplification by titrating increasing amounts of
to the plate. Cells were lysed by boiling at 100 °C for 10 min with intermediate
to the plant. Bacterial infections by leaf infiltration. A. thaliana seeds were sterilized using
grown in potting soil (Substrat 1, Klasmann-Deilmann) for 4–5 weeks in growth
for another 3 d at 22 °C. Cells were resuspended in 10 mM MgCl2 and the OD600
was adjusted to 10 in a 1-ml volume. Cells were homogenized with TissueLyzer (Qiagen) at 30 Hz for 2
samples were taken from infiltrated leaves at the indicated time points.
leaves were surface sterilized by submerging them in 70% ethanol for 30 s and
subsequently washed in water three times. Leaf discs were taken with a cork borer
(7.5 mm), and two leaf discs from the same plant were combined in tubes filled
with 200 µl of 10 mM MgCl2 and two glass beads (3 mm diameter, Merck). Leaf discs were homogenized with TissueLyzer (QIagen) at 30 Hz for 2× 45 s. The bacterial solution was plated as a dilution series on R2A-MeOH agar plates to
count cfu.
Gnotobiotic plant inoculation with single strains. Plants were grown in microboxes as described for SynCom experiments. Bacteria were grown on R2A-MeOH plates and bacterial suspension for inoculum was prepared as described above for the SynCom experiments. Germ-free plants were inoculated with 200 µl of bacterial solution of an OD600 of 0.02 in 10 mM MgCl2. For screening of bacteria that induce a plant phenotype during mono-association on rhod1, each inoculum was prepared from a 1:10 dilution of a loopful of a ‘loopful’
bacterial biomass in 1 ml of 10 mM MgCl2 (final inocula had an OD600 between 0.02
and 0.08). The plant phenotype was examined at 25 d after inoculation.

Received: 10 January 2021; Accepted: 25 May 2021; Published online: 30 June 2021;

References
1. Beattie, G. A. & Lindow, S. E. Bacterial colonization of leaves: a spectrum of strategies. Phytopathol. 89, 353–359 (1999).
2. Vorholt, J. A. Microbial life in the phyllosphere. Nat. Rev. Microbiol. 10, 828–840 (2012).
3. Pieterse, C. M. et al. Induced systemic resistance by beneficial microbes. Annu. Rev. Phytopathol. 52, 347–375 (2014).
4. 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).
5. Lugtenberg, B. & Kamoun, F. Plant growth-promoting rhizobacteria. Annu. Rev. Microbiol. 65, 541–556 (2009).
6. Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T. & Tsuda, K. Evolution of hormone signaling networks in plant defense. Annu. Rev. Phytopathol. 55, 401–425 (2017).
7. Cook, D. E., Mesarch, C. H. & Thomma, B. P. H. J. Understanding plant immunity as a surveillance system to detect invasion. Annu. Rev. Phytopathol. 53, 541–563 (2015).
8. Teixeira, P. J. P. L., Colaianni, N. R., Fitzpatrick, C. R. & Dangl, J. L. Beyond pathogen recognition: microbiota interactions with the plant immune system. Curr. Opin. Microbiol. 49, 7–17 (2019).
9. Xu, Y., Pieterse, C. M. J., Bakker, P. A. H. M. & Berendsen, R. L. Beneficial microbes driving the root microbiome. Plant Cell Environ. 42, 2860–2870 (2019).
10. Haqcuard, S., Spaepen, S., Garrido-Oter, R. & Schulze-Lefert, P. Interplay between innate immunity and the plant microbiota. Annu. Rev. Phytopathol. 55, 565–589 (2017).
11. Lebeis, S. L. et al. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349, 860–864 (2015).
12. Castrillo, G. et al. Root microbiota drive direct integration of phosphate stress and immunity. Nature 543, 513–518 (2017).
13. Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M. & Vorholt, J. A. Synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. PLoS Genet. 10, e1004283 (2014).
14. Chen, T. et al. A plant genetic network for preventing dysbiosis in the phyllosphere. Nature 580, 653–657 (2020).
15. Horton, M. W. et al. Genome-wide association study of Arabidopsis thaliana leaf microbial community. Nat. Commun. 5, 5320 (2014).
16. Vorholt, J. A., Vogel, C., Carlstrom, C. I. & Müller, D. B. Establishing causality: opportunities of synthetic communities for plant microbiome research. Cell Host Microbe 22, 142–155 (2017).
17. Bai, Y. et al. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528, 364–369 (2015).
18. Durán, P. et al. Microbial interkingdom interactions in roots promote Arabidopsis survival. Cell 175, 973–983 (2018).
19. Carlström, C. I. et al. Synthetic microbiota reveal priority effects and key strains in the Arabidopsis phyllosphere. Nature Ecol. Evol. 3, 1445–1454 (2019).
20. Karasov, T. L. et al. The relationship between microbial population size and disease in the Arabidopsis thaliana phyllosphere. Preprint at bioRxiv https://doi.org/10.1101/882814 (2020).
21. Couto, D. & Zipfel, C. Regulation of pattern recognition receptor signalling in plants. Nat. Rev. Immunol. 16, 537–552 (2016).
22. Peng, Y., van Weersch, R. & Zhang, Y. Convergent and divergent signaling in PAMP-triggered immunity and effector-triggered immunity. Mol. Plant Microbe Interact. 31, 403–409 (2018).
23. Torres, M. A., Dangl, J. L. & Jones, J. D. Arabidopsis gp91(phox) and homologues Atb6 and Atb6b are required for accumulation of reactive oxygen intermediates in the plant defense response. Proc. Natl Acad. Sci. USA 99, 517–522 (2002).
24. Gimenez-Ibanez, S., Noutoukakis, V. & Rathjen, J. P. The LysM receptor kinase CERK1 mediates bacterial perception in Arabidopsis. Plant Signal. Behav. 4, 539–541 (2009).
25. Willmann, R. et al. Arabidopsis lyrishin motif proteins LYM1-LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. Proc. Natl Acad. Sci. USA 108, 19824–19829 (2011).
26. Yamaguchi, Y., Hufacker, A., Bryan, A. C., Tax, F. E. & Ryan, C. A. PEP2R is a second receptor for the Pep1 and Pep2 peptides and contributes to defense responses in Arabidopsis. Plant Cell 22, 508–522 (2010).
27. Xin, X. et al. Bacteria establish an aqueous living space in plants crucial for virulence. Nature 539, 524–529 (2016).
28. Stawski, P. E., Tiryaki, I. & Rowe, M. L. Jasmonate response locus (JAR1) and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. Plant Cell 14, 1465–1415 (2002).
29. Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S. & Ecker, J. R. EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. Science 284, 2148–2152 (1999).

30. Clarke, J. D., Volk, S. M., Leiford, H., Ausubel, F. M. & Dong, X. Roles of salicylic acid, jasmonic acid, and ethylene in cpr-induced resistance in Arabidopsis. Plant Cell 12, 2175–2190 (2000).

31. Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J. & Katagiri, F. Network properties of robust immunity in plants. PLoS Genet. 6, e1000772 (2009).

32. Cao, H., Glazebrook, J., Clarke, J. D., Volk, S. & Dong, X. The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. Cell 88, 57–63 (1997).

33. Mustilli, A.-C., Merlot, S., Vavaussert, A., Feni, F. & Giraudat, J. Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell 14, 3089–3099 (2002).

34. Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F. & Dong, X. The cpr5 mutant of Arabidopsis expresses both NPR1-dependent and NPR1-independent resistance. Plant Cell 9, 1573–1584 (1997).

35. Clarke, J. D., Liu, Y., Klessig, D. F. & Dong, X. Uncoupling PR gene expression from ROS production and cell death. Cell 95, 361–371 (1998).

36. Stegmann, M. et al. The receptor kinase FERONIA is a key modulator of environmental-conditions on epiphytic population sizes of Pseudomonas phaseolicola pathogenic pseudomonads on soybean leaves. Nat. Rev. Immunol. 14, 1849–1859 (2015).

37. Deslauriers, S. D. & Larsen, P. B. FERONIA is a key modulator of cell-wall stress as a hexose-dependent and osmosensitive regulator of plant responses. Curr. Opin. Plant Biol. 8, 387–403 (2005).

38. Miller, G. et al. The plant NADPH oxidase RBOHD mediates rapid systemic signaling in response to diverse stimuli. Sci. Signal. 2, ra45–ra45 (2009).

39. Zhang, J. et al. A Pseudomonas syringae effector inactivates MAPKs to suppress PAMP-induced immunity in plants. Cell Host Microbe 1, 175–185 (2007).

40. Qi, J., Wang, J., Gong, Z. & Zhou, J.-M. Apoplastic ROS signaling in plant immunity. Curr. Opin. Plant Biol. 38, 92–100 (2017).

41. Mersmann, S., Bourdais, G., Rietz, S. & Robatzek, S. Ethylene signaling regulates accumulation of the FLS2 receptor and is required for the oxidative burst contributing to plant immunity. Plant Physiol. 154, 391–400 (2010).

42. Denness, L. et al. Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in Arabidopsis. Plant Physiol. 156, 1364–1374 (2011).

43. Hammans, T., Bennett, M., Mansfield, J. & Somerville, C. Identification of cell-wall stress as a hexose-dependent and osmosensitive regulator of plant responses. The Plant J. 57, 1015–1026 (2009).

44. Pogány, M. et al. Dual roles of reactive oxygen species and NADPH oxidase RBOHD in an Arabidopsis–Alternaria pathosystem. Plant Physiol. 151, 1849–1859 (2010).

45. Fagard, M. et al. Arabidopsis thaliana expresses both NPR1-dependent and NPR1-independent resistance. Plant Cell 9, 557–569 (1997).

46. Stegmann, M. et al. The receptor kinase FERONIA is a key modulator of environmental-conditions on epiphytic population sizes of Pseudomonas phaseolicola pathogenic pseudomonads on soybean leaves. Nat. Rev. Immunol. 14, 1849–1859 (2015).

47. Lee, D. et al. Regulation of reactive oxygen species during plant immunity through phosphorylation and ubiquitination of RBOHD. Nat. Commun. 11, 1563–1578 (2020).

48. Kimura, S. et al. CRK2 and C-terminal phosphorylation of NADPH oxidase RBOHD regulate reactive oxygen species production in Arabidopsis. Plant Cell 32, 1063–1080 (2020).

49. Sumimoto, H. Structure, regulation and evolution of Nox-family NADPH oxidases that produce reactive oxygen species. FEBS J. 275, 3249–3277 (2008).

50. Bedard, K. & Krause, K.-H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. Physiol. Rev. 87, 245–313 (2007).

51. Grain, F. & Wu, P. PAR protein kinases: multiple functions for a single molecule. Nat. Rev. Mol. Cell Biol. 12, 1195–1203 (2011).

52. Liu, L. et al. The FLS2-associated kinase BIK1 directly phosphorylates the NADPH oxidase RboHD to control plant immunity. Cell Host Microbe 15, 328–338 (2014).

53. Julin, J., Mankin, A. & Filipe, S. Heat induced pathogen germination: a potential defense mechanism against fungal pathogen infection. New Phytol. 211, 2160–2175 (2019).
100. Guindon, S. et al. New algorithms and methods to estimate plant immunity. Nature 592, 105–109 (2021).
101. Pfeilmeier, S., Saur, I. M.-L., Rathjen, J. P., Zipfel, C. & Malone, J. G. High-resolution imaging of bacterial communities reveals spatial heterogeneity in Arabidopsis roots. Nature 592, 110–115 (2021).
102. Schlesier, B., Bréton, F. & Mock, H.-P. A hydroponic culture system for growing Arabidopsis thaliana plantlets under sterile conditions. Plant Mol. Biol. Rep. 21, 449–456 (2003).
103. Chebus, M. K. & Tripplett, E. W. The diversity of Xanthomonas bacteria in association with the roots of Zea mays L. Microb. Ecol. 41, 252–263 (2001).
104. Bulgarelli, D. et al. Revealing structure and assembly cues for the plant microbiome. Nature 551, 252–263 (2017).
105. Chelius, M. K. & Triplett, E. W. The diversity of Archaea and Bacteria in association with the roots of Zea mays L. Microb. Ecol. 41, 252–263 (2001).
106. Ngou, B. P. M., Ahn, H.-K., Ding, P. & Jones, J. D. G. Mutual potentiation of plant defense responses: a role for NLR-mediated immunity. Nature 573, 57–61 (2019).
107. Yuan, M. et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. Nature 592, 105–109 (2021).
108. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, 2009).
109. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, 2009).
110. Parslads, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. Bioinform. 20, 289–290 (2004).
111. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
112. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998 (2013).
113. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Gen. Biol. 15, 550 (2014).
114. de Mendiburu, F. & Yaseen, M. agricolae: statistical procedures for agricultural research. R package v.1.4.0 (R Foundation for Statistical Computing, 2020).
115. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. Bioinform. 20, 289–290 (2004).
116. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, 2009).
117. Oksanen, F. et al. vegan: community ecology package. R package v.2.5-6 (R Foundation for Statistical Computing, 2019).
118. Pruesse, E., Peplies, J. & Glockner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinform. 28, 1823–1829 (2012).
119. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321 (2010).
120. Pfeilmeier, S., Saur, I. M.-L., Rathjen, J. P., Zipfel, C. & Malone, J. G. High levels of cyclic-di-GMP in plant-associated Pseudomonas correlate with evasion of plant immunity. Mol. Plant Pathol. 17, 521–531 (2016).

Acknowledgements
We thank R. Maier, A. Imboden, B. Emenegger and P. Kirner for technical assistance, and C. Vogel, M. Schäfer and L. Hemmerle for discussion and critical reading of the manuscript. DNA sequencing was done at the Genetic Diversity Centre Zurich with assistance from S. Kobel. We are grateful to the following laboratories for generous sharing of seeds: C. Zipfel (University of Zurich, Switzerland) for fls2 efr cerk1-1 (flc), bak1-5 bkk1-1 cerk1-2 (bkkc), rbohD, rbohDrbohF, rbohD/prbohD:3xFLAG–RBOHD, rbohD/prbohD:3xFLAG–RBOHD–S343A/S343F; J.-M. Zhou (Chinese Academy of Sciences, Beijing, China) for rbohD/prbohD:3xFLAG–RBOHD–S343A/S343F; K. Shirasu and Y. Kadota (Riken Center for Sustainable Resource Science, Yokohama, Japan) for rbohD/rbohF/pRBOHD:3xFLAG–RBOHD and rbohD/rbohF/pRBOHD:3xFLAG–RBOHD–S343A/S343F; S.Y. He (Michigan State University, USA) for min7 fls2 efr cerk1-2 (mefc) and min7 bak1-5 bkk1-1 cerk1-2 (mefbc); T. Hamann (Norwegian University of Science and Technology, Norway) for pep1–1 pep2–1; T. Nürnberger (University of Tübingen, Germany) for lym3lym1–5–28; X. Dong (Duke University, USA) for cpr5–1 and cpr6–1. We acknowledge funding from ETH Zurich, a grant from the German Research Foundation (DECRyPT, no. SPP2125), the NCCR Microbiomes, funded by the Swiss National Science Foundation, and a European Research Council Advanced grant (no. PhyMo—668991).

Author contributions
The conceptualization came from S.P. and J.A.V. The investigation was carried out by S.P., G.C.P., M.B.-M., B.D. and C.M.F. Software and data curation were done by C.M.F. Visualization was done by S.P., G.C.P., M.B.-M., B.D. and C.M.F. The original draft was written by S.P. and J.A.V. Review and editing of the paper was carried out by S.P., G.C.P., B.D., S.S. and J.A.V. Funding was acquired by S.P. and J.A.V. Supervision of the work was the responsibility of S.P. and J.A.V. All authors read and approved the submitted version.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41564-021-00929-5. Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-00929-5. Correspondence and requests for materials should be addressed to J.A.V.

Peer review information Nature Microbiology thanks Owyn Beattie, Joshua Herr and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints. 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 2021.
Extended Data Fig. 1 | SynCom-222 composition of inoculum, phyllosphere and endosphere. Relative abundance of strains (or ASVs indicated by superscript circle) determined by 16S rDNA amplicon sequencing of samples from SynCom-222 inoculum mix, phyllosphere and endosphere samples of *A. thaliana* Col-0. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. ‘Undetected’ and dot size indicate the number of plant replicates where a given strain was not detected (*n* = 12). Colors represent strain phylogeny.
Extended Data Fig. 2 | Arabidopsis thaliana Col-0 leaves after infiltration with bacterial strains. a, Bacterial load measured as colony forming units (CFU)/cm² at 0 and 5 days post infiltration (dpi). Xanthomonas Leaf131, Pseudomonas Leaf59, Pseudorhodoferax Leaf265, Pseudomonas syringae pv. tomato DC3000 (Pst) hrcC- and wild-type Pst were infiltrated at OD = 0.002 (~10⁶ CFU/ml) into leaves of soil-grown, four-week-old Col-0 plants. Infiltrated leaves were harvested, surface sterilized in 70% ethanol and homogenized before serial dilution plating. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. Statistical differences were calculated with two-tailed Mann–Whitney U-test (0 dpi, n = 4; 5 dpi, n = 8; ns, not significant; * p < 0.05, ** p < 0.01). b, Photographs of leaves from Col-0 five days after treatments as described above. White bar indicates one cm.
Extended Data Fig. 3 | Genotype effect on bacterial community in phyllosphere. a. Effect of plant genotype on phyllosphere community. The bacterial community of each genotype was compared to Col-0 wild-type in principal component analysis (PCA, n = 12) followed by PERMANOVA (permutations = 10,000), and the effect size of the genotype was plotted in decreasing order. Effect size represents variance explained by genotype (p-value<0.05, Benjamini–Hochberg adjusted; n = 12). b. Relative abundance of phyllosphere bacteria on the indicated plant genotypes. Asterisks (or hash tags for Firmicutes) denote significant differences in taxa on genotypes compared to Col-0 in a two-sided t-test (p < 0.05, Benjamini-Hochberg adjusted, n = 12).
Extended Data Fig. 4 | Overview and clustering of community profiles on genotypes versus Col-0. 

**a**, Strain changes in endosphere communities are displayed in a heatmap as log2 fold-change of strains in the endosphere of the individual genotypes versus Col-0 (columns). Strains or ASVs (indicated by superscript circle) of SynCom-222 are ordered and coloured by phylogeny. Hierarchical clustering (R command hclust, method ‘single’) of genotypes was performed on an Euclidean distance matrix of log2 fold changes between test conditions and controls.

**b**, Strain changes in phyllosphere communities shown in the heatmap with genotype clustering as described above. Differential strain abundance was calculated using DESeq2, and statistical significance was expressed with p-values (two-sided Wald test, Benjamini-Hochberg adjusted): the black cell rectangle highlights significant changes \( p < 0.05 \).
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Microbiota-induced disease in rbohD is linked to the enrichment of Xanthomonadaceae. a, Screening of plant phenotypes and disease symptoms in rbohD after colonization with individual bacterial strains. Germ-free, 10-day-old rbohD seedlings were inoculated with a bacterial suspension (OD$_{600}$ ranging from 0.02 to 0.08), and the plant phenotype was examined after 3.5 weeks (n=8). Inoculated plants showed a variety of phenotypes, for example, stunted plants, necrotic lesions, curled leaves or dead plants. Phenotype-inducing strains were selected based on symptoms and are highlighted by red rectangles in the phylogeny. b, Disease index was assessed in rbohD plants: 1, healthy; 2, mild symptoms on individual leaves; 3, stronger symptoms on multiple leaves; 4, strong symptoms on whole plant; 5, severe symptoms or dead plant. The graph displays the plant fresh weight (mg) of SynCom-137-inoculated rbohD plants (n=20) per disease category. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. c, Bacterial load in SynCom-137-inoculated Col-0 and rbohD plants measured as colony forming units (CFU) per gram of plant fresh weight isolated from the endosphere and phyllosphere (n=16; two-tailed Mann-Whitney U-test; ns, not significant; **, p<0.01). Bacterial load in Col-0 and rbohD plants inoculated with the indicated SynCom represented by qPCR of the bacterial 16S rDNA gene relative to plant gene. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. Statistical significance was calculated by two-tailed Mann-Whitney U-test (n=3, each pool of four DNA samples; ns, not significant). qPCR data for SynCom-222 in Col-0 are the same as in Fig. 1c. d, Correlation analysis between the relative abundance of Xanthomonadaceae in phyllosphere samples of rbohD inoculated with the indicated SynCom and plant fresh weight. The Spearman coefficient ρ (p-value<0.01) was calculated using ggscatter command (ggpubr, R), grey area shows 95% confidence interval of regression line in green. Correlation data on endosphere samples was not possible due to bulk surface sterilization of plants. e, ROS accumulation was measured in leaf discs with a luminol-based assay after treatment with extracts from heat-killed bacteria. ROS production was recorded for 45 min, and luminescence counts were integrated over time. ROS triggered by individual treatments were normalized to ROS production by 10 nM flg22. Normalized ROS accumulation is shown for each bacterial strain. Barplots show mean and error bars show standard deviation (n=16; combined data from two independent experiments). Red dots indicate rbohD-enriched strains.
See next page for caption.
Extended Data Fig. 6 | The community assembly of rbohD and rbohDrbohF substantially differs from that of other PTI genotypes. The heatmap shows log2 fold-changes in strain abundance on different genotypes compared to Col-0 wild-type. Columns show endosphere and phyllosphere samples from plants inoculated with either a, SynCom-222 or b, SynCom-223 or SynCom-137. Strains (or ASVs indicated by superscript circles) are ordered and colored by phylogeny. Statistical significance of differential strain abundances was calculated with the two-sided Wald test (DESeq2 package, R) and highlighted with a black cell rectangle for $p < 0.05$ ($n = 12$; Benjamini-Hochberg adjusted).
Extended Data Fig. 7 | PTI-associated phosphorylation sites of RBOHD are involved in resistance to opportunistic pathogens. a, ROS production of A. thaliana Col-0, rbohD, rbohF, rbohD/RBOHD, rbohD/RBOHD-S39AS339AS343A, rbohD/RBOHD-S343A-S347A, rbohD/RBOHD-S343A-S347A and rbohD/rbohF, after treatment with 100 nM flg22. ROS production was measured in leaf discs from soil-grown plants with a luminol-based assay and expressed as integrated luminescence over 45 min (AU, arbitrary units). Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. b, Fresh weight of Col-0, rbohD, rbohD/RBOHD, and rbohD/RBOHD-S343A-S347A inoculated with individual strains of Xanthomonas spp. Leaf131, Leaf148, Pst hrcC, and Pst wild-type or mock-inoculated with 10 mM MgCl2. Germ-free 10-day-old seedlings were inoculated with OD = 0.02 of the respective strains. Box plots show the median with upper and lower quartiles and whiskers present 1.5x interquartile range. Significant differences were calculated between the mutants and their respective controls (n = 20; two-tailed Mann-Whitney U-test). Brackets above bar plots indicate comparison groups with p-values displayed above and fold-change below. Data from two independent experiments are shown in separate graphs.
Extended Data Fig. 8 | Leaf131 is required and sufficient for disease in rbohD. Two independent replicate experiments with dropout synthetic communities as presented in Fig. 5. In addition, SynCom-REPI without (w/o) Leaf131 was tested and compared to SynCom-REPI. Significant differences were calculated with ANOVA and Tukey’s HSD post-hoc test (n = 20, letters indicate significance groups, α = 0.05).
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 our Editorial Policies 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.

- 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
- 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

Data collection as described in the manuscript and source data made available for each figure.

Data analysis

Description of published software used for data analysis is provided and cited in the Methods section and figure legends. Analysis of microbiota data with customized scripts is described in the manuscript and code is available on GitHub repository: https://github.com/MicrobiologyETHZ/phylloR/releases/tag/v1.1

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 and 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

16S rRNA gene sequences were generated in three Illumina MiSeq runs. Demultiplexed reads are available in the ENA (European Nucleotide Archive) database under the Study accession PRJEB44158 (ERP128175). There is no restriction on data availability.
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 statistical analysis are described in the relevant figure legends. The sample size was chosen according to experimental trials or previously published studies to allow for confident statistical analysis. No statistical methods were used to predetermine sample sizes. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data was excluded in the analysis that fulfilled the set quality standards. During plant experiments, individual plants were excluded before treatment if plant development was not according to standards and previous experiences. |
| Replication | Each experiment was performed at least twice independently (but mostly three times) as shown in individual figures and indicated in respective figure legends and Methods. Synthetic community experiments were repeated with minor changes in the synthetic community (as indicated in the manuscript). |
| Randomization | Samples from different treatments of 16S amplicon sequencing experiment were harvested and processed in random order to avoid batch effects. |
| Blinding | Experimenters were blinded to treatments during harvest of gnotobiotic experiments. |

**Behavioural & social sciences study design**

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

| Study description | Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study). |
| Research sample | State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source. |
| Sampling strategy | Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed. |
| Data collection | Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection. |
| Timing | Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort. |
| Data exclusions | If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established. |
| Non-participation | State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation. |
| Randomization | If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled. |

**Ecological, evolutionary & environmental sciences study design**

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

| Study description | Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates. |
| Research sample | Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National |
Research sample

Yes

Involved in the study

State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Field work, collection and transport

Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access & import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance

Describe any disturbance caused by the study and how it was minimized.

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

Antibodies

Eukaryotic cell lines

Palaeontology and archaeology

Animals and other organisms

Human research participants

Clinical data

Dual use research of concern

Methods

n/a

Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging

Antibodies

Antibodies used

Describe all antibodies used in the study, as applicable, provide supplier name, catalog number, clone name, and lot number.

Validation

Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer’s website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about cell lines

State the source of each cell line used.
Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines

None any commonly misidentified cell lines used in the study and provide a rationale for their use.

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Policy information about studies involving human research participants

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g., collection, storage, sample pretreatment and measurement), where they were obtained (i.e., lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals, ARRIVE guidelines, recommended for reporting animal research

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g., age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.

Study protocol

Note where the full trial protocol can be accessed OR if not available, explain why.

Data collection

Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.

Outcomes

Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.
Dual use research of concern

Policy information about dual use research of concern

Hazards
Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No  Yes
☐  Public health
☐  National security
☐  Crops and/or livestock
☐  Ecosystems
☐  Any other significant area

Experiments of concern
Does the work involve any of these experiments of concern:

No  Yes
☐  Demonstrate how to render a vaccine ineffective
☐  Confer resistance to therapeutically useful antibiotics or antiviral agents
☐  Enhance the virulence of a pathogen or render a nonpathogen virulent
☐  Increase transmissibility of a pathogen
☐  Alter the host range of a pathogen
☐  Enable evasion of diagnostic/detection modalities
☐  Enable the weaponization of a biological agent or toxin
☐  Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition
☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
Confirm that both raw and final processed data have been deposited in a public database such as GEO.

Files in database submission
Provide a list of all files available in the database submission.

Genome browser session
Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates
Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth
Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies
Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters
Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality
Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software
Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.
Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

Magnetic resonance imaging

Experimental design

Design type

Indicate task or resting state, event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI

☐ Used

☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template

Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI152, ICBM152) OR indicate that the data were not normalized.

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).
**Volume censoring**  
Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

### Statistical modeling & inference

| Model type and settings | Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation). |
|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Effect(s) tested        | Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used. |

**Specify type of analysis:**  
- [ ] Whole brain  
- [ ] ROI-based  
- [ ] Both

**Statistic type for inference**  
(See Eklund et al. 2016)  
Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

**Correction**  
Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

### Models & analysis

- **n/a**  
- [ ] Involved in the study  
- [ ] Functional and/or effective connectivity  
- [ ] Graph analysis  
- [ ] Multivariate modeling or predictive analysis

#### Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

#### Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

#### Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.