Protective role of the *Arabidopsis* leaf microbiota against a bacterial pathogen

Christine M. Vogel, Daniel B. Potthoff, Martin Schäfer, Niculò Barandun and Julia A. Vorholt

The aerial parts of plants are host to taxonomically structured bacterial communities. Members of the core phyllosphere microbiota can protect *Arabidopsis thaliana* against foliar pathogens. However, whether plant protection is widespread and to what extent the modes of protection differ among phyllosphere microorganisms is not clear. Here, we present a systematic analysis of plant protection capabilities of the At-LSPHERE, which is a collection of >200 bacterial isolates from *A. thaliana*, against the bacterial pathogen *Pseudomonas syringae* pv. tomato DC3000. In total, 224 bacterial leaf isolates were individually assessed for plant protection in a gnotobiotic system. Protection against the pathogen varied, with ~10% of leaf microbiota strains providing full protection, ~10% showing intermediate levels of protection and the remaining ~80% not markedly reducing disease phenotypes upon infection. The most protective strains were distributed across different taxonomic groups. Synthetic community experiments revealed additive effects of strains but also that a single strain can confer full protection in a community context. We also identify different mechanisms that contribute to plant protection. Although pattern-triggered immunity coreceptor signalling is involved in protection by a subset of strains, other strains protected in the absence of functional plant immunity receptors BAK1 and BKK1. Using a comparative genomics approach combined with mutagenesis, we reveal that direct bacteria-pathogen interactions contribute to plant protection by *Rhizobium* Leaf202. This shows that a computational approach based on the data provided can be used to identify genes of the microbiota that are important for plant protection.

Plants do not grow as solitary organisms in nature but are host to a plethora of microorganisms, the plant microbiota. In particular, the bacterial microbiota have been intensively studied across plant species, geographical locations and years. They show remarkable convergence at genus level and higher taxonomic rank with apparent differences among plant species at finer taxonomic resolution. The microbiota contribute to the plant phenotype and affect processes such as development and resistance to abiotic and biotic stresses. Plant-associated microorganisms can increase the availability of limiting nutrients, influence flowering time, improve resistance to drought and salinity stress and mitigate disease progression on pathogen challenge, making the plant microbiota a target of interest for agricultural applications. However, successful application of microbiome-mediated phenotypes requires an understanding of the mechanisms underlying the beneficial functions of the microbiota and how these emerge from the properties of its members, host genotypes and environmental conditions. Regarding plant protection specifically, the importance of the microbiota is evident from the addition of individual strains that lead to pathogen attenuation. It is also apparent from suppressive soils, where the resident soil microbiota prevent pathogens from becoming established and/or causing disease. Insights into protective mechanisms have been obtained for individual bacteria. These include direct microbe–microbe interactions such as the production of antimicrobial compounds directed against the pathogen, competition for the niche (for example, nutrient competition) or interference with pathogen virulence (for example, via quorum quenching). Indirect plant-mediated protection can occur via the induction of plant defence strategies as part of the innate immune system. The latter enables the recognition of pathogens, as well as other microorganisms or their products via a multilayered perception system. In addition, there is evidence that the immune system can alter the composition of the microbiota, which in turn may also have consequence for the outcome of pathogen infection.

For the phyllosphere, the aboveground plant parts, strains of the two most common bacterial families, the Methylobacteriaceae and Sphingomonadaceae, show differences in their potential to protect the host plant against infection by the pathogen *Pseudomonas syringae* DC3000 (refs. 27,28). Other often-detected leaf bacteria belong to the Pseudomonadaceae family and members of this family can have positive and negative impact on the plant host. Currently, it is unclear how widespread such protective functions are in endogenous leaf-associated bacterial communities and what the implications are for leaf microbiota assembly.

Here, we systematically tested more than 200 representative bacterial leaf isolates from a publicly available collection of genome-sequenced strains for the ability to protect *Arabidopsis thaliana* from infection with the foliar pathogen *P. syringae* DC3000. Following the screening of individual strains, we examined consequences of building synthetic communities for plant protection and used genome-inferred analysis to test a mechanism of plant protection.

Results

Plant protection potential in tripartite screening experiments. To uncover how the microbiome contributes to host phenotypes, in this case plant protection, the consequences of colonization by a broad representation of individual microbiota members was systematically assessed. We tested the potential of 224 *A. thaliana* leaf isolates, collectively referred to as the At-LSPHERE strain collection, for plant protection. This strain collection contains representatives of ~50% of the diversity observed by 16S ribosomal DNA sequencing of natural *A. thaliana* plants on the basis of 97% operational taxonomic units (OTUs). Furthermore, these strains assemble into communities resembling the natural microbiota at a phylum level in
a gnotobiotic system. We used a tripartite *A. thaliana* model system and screened each of the strains in planta for protection against a luxCDABE-tagged *P. syringae* DC3000 (Pst). Pathogen infection or lack thereof was scored in two ways. Luminescence quantification was used as a proxy for pathogen colonization. Disease scoring based on symptoms was used to calculate the protection score relative to the disease observed in control plants. Furthermore, we used uninfected plants to determine phyllosphere colonization by the commensal strains by counting c.f.u. after a wash protocol.

Correlation of plant protection by *At*-LSPHERE strains to phylogeny and colonisation. Next, we examined plant protection by members of the microbiota in terms of their phylogenetic distribution and plant colonization. The distribution of protective and non-protective strains within the phylogenetic tree of the *At*-LSPHERE revealed the presence of clusters of protective strains (Fig. 2). Indeed, we found a significant phylogenetic signal in the plant protection trait (Pagel's $\lambda = 0.979$, $P = 1 \times 10^{-45}$; Abouheif's $C_{\text{mmu}} = 0.569$, $P = 0.001$) that was robust also to subsampling (Supplementary Fig. 3) indicating that closely related strains are more likely to have a similar protective phenotype than two randomly picked strains.

We also measured phyllosphere colonization levels of the *At*-LSPHERE strains as colony-forming units (c.f.u.). Most strains reached densities $>10^5$ c.f.u. mg$^{-1}$, with the majority colonizing at $10^4$–$10^5$ c.f.u. mg$^{-1}$ (Fig. 3c and Supplementary Tables 1 and 2). The highest colonization densities were found for strains belonging to the families Enterobacteriaceae, Pseudomonadaceae, Nocardioidaceae and Microbacteriaceae (Fig. 3d). A few strains were either not or only sporadically recovered from the phyllosphere (Supplementary Data 1).

The integration of phylosphere colonization capacity and protection revealed that both traits are positively correlated when all strains are considered (Pearson's correlation of log$_{10}$(c.f.u. mg$^{-1}$) and protection score, $r=0.47$, $t=7.97$, d.f. = 223, $P=8 \times 10^{-14}$; Spearman's $\rho = 0.49$, $P=9.8 \times 10^{-15}$; Fig. 3c). Strains with a protection score $>75$ were also good colonizers with colonization densities $>10^5$ c.f.u. mg$^{-1}$. However, the opposite was not necessarily the case. Many strains showed a high colonization capacity but were not protective. Notably, the correlation between mean colonization and mean protection score was not evenly distributed across taxa (Fig. 3d). For example, most *Pseudomonas* spp. tested showed full protection, one strain (Leaf48) showed reduced protection and one strain showed no protection (Leaf83). This difference in protection correlated with differences in colonization densities ($R=0.98$, $t=13.328$, d.f. = 7, $P=3 \times 10^{-4}$), Leaf83 was the only *Pseudomonas* that did not reproducibly colonize the phyllosphere of *A. thaliana* in our experimental system (Supplementary Table 2). On the contrary, for *Methyllobacteriaceae* the majority of strains colonized well ($>10^5$ c.f.u. mg$^{-1}$) but none of the strains showed protection. For *Sphingomonadaceae* and *Rhizobiaceae*, the best colonizing strains showed some of the highest protection potential of all strains; notably, however, other members of the same bacterial family did not protect.

As with plant protection, a phylogenetic signal was observed for plant colonization (Abouheif's $C_{\text{mmu}} = 0.41$, $P=0.001$, Pagel's Spearman's $\rho = 0.64$, $P=2.9 \times 10^{-9}$). The correlation was robust also to subsampling (Supplementary Fig. 3) indicating that closely related strains are more likely to have a similar colonization capacity than two randomly picked strains.
λ = 0.988, \( P = 5 \times 10^{-26} \); Supplementary Fig. 3). Thus, although both the colonization and protection phenotypes were strain specific, closely related strains tended to show similar in planta characteristics.

### Synthetic communities tested for protection

Next, we tested combinations of strains and investigated their impact on protection. We explored randomly assembled synthetic communities (SynComs) of ten strains (for more details on SynCom experiments...
Fig. 3 | Protective strains of the At-LSPHERE colonize the phyllosphere at high density. a, b. Histograms of mean protection score (n=222 strains) (a) and of mean colonization (n=224 strains) (b) for the At-LSPHERE collection. c. Correlation of mean colonization (log10-transformed) and mean protection score (Pearson’s $R = 0.47$, $t = 7.97$, d.f. = 223, $P = 8 \times 10^{-14}$). The linear regression line is indicated. The dashed box highlights mean colonization $>10^5$ c.f.u. mg$^{-1}$ and mean protection score $>75$. d. Mean colonization plotted against mean protection score for each family separately. Colours throughout the figure correspond to phyla and class. a.u., arbitrary units. Exact numbers of independent experiments and biological replicates within experiment for each strain are provided in Supplementary Tables 1 and 2, respectively.
see Supplementary Note). While random SynComs of strains with mean protection scores <25 did not improve protection (Extended Data Fig. 3 and Supplementary Table 4), the majority of SynComs, in which strains with mean protection scores <65 were included, showed better protection than the best individual strain within the SynCom (Fig. 4a, Extended Data Fig. 4 and Supplementary Table 5). Notably, two SynComs (M10.35 and M10.21) showed a protection score >85, which is a clear improvement relative to the protection conferred by the best individual strains. We thus wondered whether the better protection might be attributed to one or two strains within the SynComs. Some mixes in which we removed individual strains only partially lost protection conferred by the community. On the contrary, in SynCom M10.35, dropout of the two most abundant strains *Rhodococcus* Leaf278 and *Curtobacterium* Leaf261, resulted in a strongly reduced protection with a protection score of 38 compared to a protection score of 92 obtained with the full SynCom (Fig. 4b, Extended Data Fig. 5 and Supplementary Table 6). *Rhodococcus* Leaf278 showed reduction of luminescence but not a great improvement in plant phenotype with infected plants showing a distinct stressed and chlorotic phenotype (Supplementary Fig. 2 and Fig. 4d). As *Rhodococcus* Leaf278 was the most abundant strain in the SynCom, we tested whether removal of Leaf278 by itself would also abolish the protective effect of the M10.35 mix. Indeed, the SynCom without Leaf278 showed higher *Pst* colonization and stronger disease than plants inoculated with the full SynCom (Fig. 4c–e). Plant colonization by both *Rhodococcus* Leaf278 and the SynCom M10.35 were comparable (Fig. 4f and Supplementary Table 7). Thus, the improved protection of SynCom M10.35 was not due to higher overall plant colonization but rather requires at least two different strains within M10.35, one of which needs to be Leaf278. However, not every combination of strains containing Leaf278 shows high protection. We tested another SynCom with and without Leaf278 (M10.48) and found no increased protection, with the full SynCom protecting to the same extent as Leaf278 alone (Fig. 4b), indicating that the effect of Leaf278 is conditional. It will thus be interesting to further deduce the mechanism of interaction leading to the improved protection by Leaf278, also in light of an arsenal of natural product gene clusters present in this strain; see ref. 30.
In addition, we used smaller synthetic communities to evaluate potential synergistic effects of strains. We hypothesized that potential additive or synergistic effects could best be identified in small mixes of strains that, by themselves, show quantifiable but not strong protection. We assembled SynComs of three strains and tested these side-by-side with the individual strains for plant protection. For two of the three mixes (M3.1 and M3.3), the community improved plant protection with regard to protection score and luminescence relative to the individual strains (Extended Data Fig. 6 and Supplementary Table 8; for details see Supplementary Note). The strains already individually provided an intermediate protection score, thus suggesting additive effects. Overall, this indicates that the combination of several strains can improve protection, whereas in mix M3.2, a more complex community did not lead to an improved phenotype.

**Evidence for complementary mechanisms of plant protection.** Protection against infection can be due to fundamentally different mechanisms. One of these mechanisms involves the plant immune system and defence reactions by the host that are triggered by certain microbiota members\(^\text{51}\). Bacteria are perceived by a large arsenal of dedicated receptors (for example LRR-RLKs) that are dependent on the common coreceptors BAK1 and BKK1 (ref. \(^\text{13}\)). We tested all protective strains of the At-LSSHARE collection (mean protection score >75) as well as some strains showing intermediate (mean protection score between 50 and 75) or no protection on bak1/bkk1 plants for loss of protection. Indeed, luminescence and hence pathogen colonization was increased in nine of the 28 protective At-LSSHARE strains in the bak1/bkk1 background (Fig. 5; for other strains see Supplementary Fig. 4).

Notably, Sphingomonas Leaf205 and Pseudomonas Leaf127 were compromised in plant protection in the bak1/bkk1 background, showing higher increase in luminescence by the pathogen compared to the axenic control as well as stronger disease symptoms on bak1/bkk1 plants (Fig. 5, Supplementary Data 2 and Supplementary Tables 9 and 10). This compromised protection cannot be attributed to reduced plant colonization by the At-LSSHARE strains in the bak1/bkk1 background (Supplementary Table 10). For other protective strains that resulted in higher pathogen titre on bak1/bkk1 plants, it is not clear whether the observation is solely due to increased susceptibility of bak1/bkk1 to the pathogen per se, as the increase in luminescence was not greater than in the axenic infected control plants (Supplementary Table 9). However, they showed a clear difference when compared to strains, which did not show higher luminescence on bak1/bkk1 infected plants. Notably, Xanthomonas spp. Leaf131 and Leaf148 behaved as opportunistic pathogens on bak1/bkk1 plants and killed numerous plant even in the absence of Pst, which was not the case in wild-type plants (Supplementary Fig. 5). This is in line with recent observations that Leaf131 and Leaf148 are opportunistic pathogens on immune-compromised rbohD plants\(^\text{26}\). Of note, six of the 18 protective strains with mean protection score >90 showed luminescence at background level of uninfected plants and thus also completely protected bak1/bkk1 plants. At this time, it cannot be excluded that plant-independent protection mechanisms are strong enough in the best protective strains to mask any potential plant-mediated processes. Alternatively, it is also possible that plants recognize the presence of specific strains in a BAK1/BKK1-independent manner. Importantly, in none of the strains with mean protection score >75 was a complete loss of protection observed, suggesting multiple mechanisms contributing to plant protection.
**Comparative genomics to identify protection traits.** Next, we wondered whether the genomes of the At-lSPHERE strains could be used to identify genetic features associated with plant protection. As the strains are phylogenetically highly diverse, we anticipated that differences potentially related to protection would be more readily identified in genera containing both protective and non-protective strains. As proof of concept, we focused on *Rhizobium* spp. that showed a range of protection with protective, intermediate and non-protective strains in a balanced distribution in our dataset (Fig. 6a). We identified 25 clusters of orthologous genes (COGs) present in protective strains that were absent in all non-protective strains (Supplementary Table 11 and Fig. 6a). Interestingly, 14 of these 25 COGs were predicted to be type VI secretion system (T6SS) components or T6SS-associated. T6SS have previously...
been described to mediate bacterial interactions by Agrobacterium and Pseudomonas spp. both in vitro and in planta\textsuperscript{51-59}.

To test whether the presence of the T6SS was important for plant protection, we generated T6SS mutants in Rhizobium Leaf202 (Fig. 6b). In vitro assays revealed that Rhizobium Leaf202 inhibited Pst and that inhibition was indeed partly dependent on a functional T6SS (Fig. 6c). We then tested Leaf202 wild type as well as the tssL mutant for plant protection. The tssL mutant allowed higher Pst colonization than the wild type, indicating that indeed a functional T6SS contributes to plant protection (Fig. 6d). This was independent of Leaf202 colonization density as both the wild type and tssL mutant colonized the phyllosphere comparably (Fig. 6e). Overall, T6SS are widely distributed in At-LSPHERE strains and present in more protective strains than expected on the basis of the overall distribution (Fisher’s exact test, odds ratio 3.29, \( P = 9 \times 10^{-4} \)) (Extended Data Fig. 7). Thus, T6SS could be one of the mechanisms contributing to plant protection in a subset of strains.

\section*{Discussion}

Plant-associated microbes are important for the host phenotype, including plant protection. Here, we screened the At-LSPHERE collection composed of 224 genome-sequenced strains for protection against P. syringae DC3000 infection in planta, thus generating systematic genotype–phenotype (plant protection and colonization) correlated data. We identified 28 of 222 strains that protected A. thaliana against infection with Pst. The most strongly protecting strains identified belong to the phyla Proteobacteria and Actinobacteria. Only one less protective strain was identified in the phylum Firmicutes and none within Bacteroidetes. Our analysis revealed that protection against a foliar pathogen shows a weak phylogenetic signal within the At-LSPHERE (Fig. 2 and Supplementary Fig. 3). This points to phylogenetic conservatism of the microbial trait of protection by vertical gene inheritance and is in line with the observation that phylogenetic trait conservation is widespread, in particular for genetically complex traits\textsuperscript{38}. Nonetheless, we also observed examples of strain specificity, for example in Pseudomonas and Sphingomonas spp., which is in line with earlier observations that complex in planta phenotypes are often strain specific\textsuperscript{39-42}.

Many of the protective strains are part of the core phyllosphere microbiota taxa that are reproducibly found in association with plant leaves. Albeit at a relatively low percentage of about 10\%, our screening results suggest that it is likely that protective community members are present in an environmental leaf system. Because all protective strains are colonizing at high density (Fig. 3) and these strains might be more competitive compared to others, an enrichment of protective strains might occur. It will therefore be interesting to test whether protective strains are preferentially enriched in a community context and, more specifically, under plant stress. Such observations are also relevant in the context of biocontrol phenotypes that can be dose-dependent\textsuperscript{43-45}. Ultimately, under environmental conditions, higher colonization density might be a requirement for protection; however, a high colonization fitness could also be selected for as the consequence of protection. This poses the more general question whether protective strains are more often associated with and/or selected for by plants. There is evidence that plants can assemble a protective community\textsuperscript{46,47}, recruit beneficial microbes\textsuperscript{48,49} or enrich for groups of bacteria that are mostly beneficial\textsuperscript{50,51}. Interestingly, in our SynCom experiments where we observed community-dependent protection, we note that in some of the dropout experiments where we removed the most abundant strain, we indeed eliminated the most protective strain (Fig. 4). However, whether this observation is robust to more complex communities or whether also removal of less abundant strains affects protection needs to be tested. In addition, it will be of interest to test other pathogens to learn the extent to which strains that are protective against Pst are also protective when encountering another pathogen and a possible pathobiome\textsuperscript{52}.

The broad screening approach conducted here also allowed circumventing likely bias that emerges from in vitro interaction pre-screening before testing strains in planta\textsuperscript{51}. On the contrary, the in planta-generated phenotypes can be integrated with other data sets that have already been assembled for the At-LSPHERE reference collection or will be in the future. The At-LSPHERE strain collection has previously been tested for antagonistic interactions in vitro\textsuperscript{52}. Only two strains (Novosphingobium Leaf2 and Pseudomonas Leaf58) were identified to inhibit Pst under the tested in vitro conditions\textsuperscript{53}. While Pseudomonas Leaf58 is among the identified strains to protect A. thaliana (Supplementary Table 1), Novosphingobium Leaf2 does not provide plant protection in our assay. This result is in line with the notion that antibiotic observed among strains on synthetic media is a rather poor predictor of in planta protection\textsuperscript{54,55}. This could be due to the prevalence of other mechanisms mostly underlying in planta protection phenotypes but also the lack of production of the inhibitory compound at sufficient concentrations under the environmental conditions encountered on the leaf surface, to a lack of sufficient colonization density of the commensal to cover the phylloplane and the spatial distribution of leaf strains that is known to be patchy\textsuperscript{56-58}.

Next to direct microbe–microbe interactions such as antibiotics or competition for nutrients\textsuperscript{59}, plants can also be indirectly protected by enhanced plant resistance\textsuperscript{60}. Of the 28 strains with a mean protection score >75, two showed a strong reduction in plant protection in the pattern-triggered immunity-compromised bku1/bkk1 plant background. Another six strains allowed higher pathogen colonization, indicating that plant immunity is important for the full extent of protection observed by these strains. Notably, none of these showed a complete loss of protection. This finding indicates that multiple mechanisms act by complementary means to confer protection. This is also supported by the observation that additive effects occur among synthetic communities, as evidenced by, for example, the ten and three member communities M10.35 and M3.1 showing superior effects than their respective best protecting strains alone (Fig. 1 and Extended Data Fig. 6). All bacterial families for which we found protective strains (Fig. 3) harbour at least one known member described to be pathogenic\textsuperscript{61,62}, bringing up the question on the perception and potential discrimination of pathogen versus ‘commensal’ strains in the plant microbiota. It will thus be instructive to use the data generated here to test how far protective and non-protective strains can be discriminated via potential plant responses they elicit.

The in planta screen also provided a basis for genome-inferred analyses as a way to identify modes of protection. Although systematically gathered phenotypic data can be used to identify genotype–phenotype associations\textsuperscript{62} they pose the difficulty of distinguishing spurious phylogenetic associations from true genotype–phenotype associations\textsuperscript{63}. Here, we exemplarily focused on one bacterial family, which showed a range of protective and non-protective strains under our experimental conditions. Genome comparisons allowed the identification of an association between the T6SS and plant protection in Rhizobium Leaf202 (Fig. 6). T6SSs are enriched in plant-associated microbes\textsuperscript{64} and the T6SS can also have functions beyond microbe–microbe interactions with T6SS having also been linked to symbiosis, biofilm formation and virulence\textsuperscript{65-67}. Our observation extends on other reports of T6SSs involved in microbe–microbe interactions in the plant environment\textsuperscript{68-70}. The Leaf202 tssL mutant showed attenuated but not complete loss of protection, thus hinting towards several modes of protection acting in concert as highlighted already above. Thus, it will be interesting to test additional genes found exclusively in the protective versus the non-protective Rhizobium spp. for their potential contribution to plant protection. Another promising bacterial family for future
genome-based analyses are *Sphingomonas* spp. that were initially thought to be composed mainly of plant protective strains on the basis of a limited selection of strains but were shown to harbour both protective and non-protective strains in this study with the latter outnumbering the former.

In conclusion, we show that ~10% of the *At-LSPHERE* strain collection protect *Arabidopsis* from *P. syringae* infection in tripartite interactions, that the potential for protection is elevated in *SynComs* and that different mechanisms contribute to plant protection. Some of the latter will require spatial proximity such as bacterial warfare, while others might be indirect via the plant and thus might act systemically. The data presented here highlight how emergent properties of microbial communities can be identified on the basis of the abilities of their individual community members, and help establish causal genotype–phenotype relationships.

**Methods**

**Plant growth conditions.** *A. thaliana* Columbia (Col-0) and *bak1-5/bkk1-1* (ref. 31) were grown on R2A plates (Sigma-Aldrich) supplemented with ~0.1% w/v methanol (R2A + M) and incubated at room temperature (~22 °C). Cell material was resuspended in 10 mM MgCl₂, adjusted to an optical density (OD) of 0.2 and then diluted 1:10 to an OD of 0.02, corresponding to around 5 × 10⁷ to 5 × 10⁵ c.f.u. ml⁻¹. Plants were inoculated at 10–11 d by distributing four to five small droplets of suspension in water at 4 °C for 3–4 d before being placed in 24-well plates containing 1.5 ml of plant medium.

**Inoculation with *At-LSPHERE* strains.** *At-LSPHERE* strains were grown on R2A plates (Sigma-Aldrich) supplemented with 0.5% v/v methanol (R2A + M) and incubated at room temperature (~22 °C). Cell material was resuspended in 10 mM MgCl₂, adjusted to an optical density (OD₆0₀) of 0.2 and then diluted 1:10 to an OD₆0₀ of 0.02, corresponding to around 5 × 10⁷ to 5 × 10⁵ c.f.u. ml⁻¹. Plants were inoculated at 10–11 d by distributing four to five small droplets of suspension (10 µl in total) to the leaves and the centre of the plants. This titre was chosen on the basis of earlier experiments and corresponds roughly to the carrying capacity of *A. thaliana* for the positive control strain *S. melonis* Fr1 at this plant size.

Briefly, one loop (corresponding to ~1 µl) of cell material was resuspended in 1 ml of 10 mM MgCl₂, and spotted on R2A + M for c.f.u. determination. When mixtures of strains were tested, *SynComs* were established by mixing roughly equal ratios of the different strains (*At-LSPHERE phyllosphere colonization is robust towards an imbalance in the inoculum*). Briefly, one loop (corresponding to ~1 µl) of cell material was resuspended in 1 ml of 10 mM MgCl₂, and spotted on R2A + M for c.f.u. determination. When mixtures of strains were tested, *SynComs* were established by mixing roughly equal ratios of the different strains (*At-LSPHERE phyllosphere colonization is robust towards an imbalance in the inoculum*). Briefly, one loop (corresponding to ~1 µl) of cell material was resuspended in 1 ml of 10 mM MgCl₂, and spotted on R2A + M for c.f.u. determination. When mixtures of strains were tested, *SynComs* were established by mixing roughly equal ratios of the different strains (*At-LSPHERE phyllosphere colonization is robust towards an imbalance in the inoculum*). Briefly, one loop (corresponding to ~1 µl) of cell material was resuspended in 1 ml of 10 mM MgCl₂, and spotted on R2A + M for c.f.u. determination. When mixtures of strains were tested, *SynComs* were established by mixing roughly equal ratios of the different strains (*At-LSPHERE phyllosphere colonization is robust towards an imbalance in the inoculum*). Briefly, one loop (corresponding to ~1 µl) of cell material was resuspended in 1 ml of 10 mM MgCl₂, and spotted on R2A + M for c.f.u. determination.

**Infection with *P. syringae* DC3000 lux.** Fifteen-day-old plants were infected with a luxCDABE-tagged *P. syringae* pv. tomato DC3000 (ref. 31) (designated Pst) similarly as described. A dense suspension of Pst was plated on King’s B medium and incubated overnight at 28 °C. The lawn of Pst was scraped off after incubation with 10 ml of 10 mM MgCl₂ for 10 min. The OD₆0₀ of the suspension was adjusted to 0.3 and the suspension diluted to a final OD₆0₀ of 0.00003. Plants were infected with 15 µl of suspension by distributing small droplets on the leaves (~13 µl) and the centre (~2 µl) of the plants. This corresponded to ~250–300 pathogen c.f.u. per plant. Plates were kept open to dry for about 2–3 min. Wells, in which the seeds had not germinated or in which plants were growing that could not be scored for disease development (either because growing inside the agar or upside down, other odd phenotypes) were treated with 10 mM MgCl₂ instead of Pst suspension.

**Plant protection assessment.** Infected *A. thaliana* plants were scored for disease development and plant protection in two different ways. As a proxy for Pst colonization, luminescence images of 24-well plates were taken at 6 d post infection (dpi) using the IVIS Spectrum Imaging System (Xenogen). To detect the Pst luminescence signal but block plant photosynthesis a 500 nm emission filter was used. The total photon flux per well was deduced by integration of the signal over regions of interest drawn on the different wells in the imaging software. Wound elicitation and luminescent images were acquired at 4.2. At time point of infection or mock-treatment, the plants were retested and uninfected control plants harvested to record overall colonization densities. For a subset of *SynComs*, 50–100µl of selected dilutions were additionally plated on round Petri dishes to identify the most abundant strain in the *SynCom*.

**Computational methods.** A phylogenetic tree of all *At-LSPHERE* isolates as well as *S. melonis* Fr1 was prepared using ezTree. RefSeq assemblies were downloaded from NCBI for all published genomes (Supplementary Table 12). For the remaining *At-LSPHERE* strains, the cultured strains on R2A + M
agar or in R2A + M liquid medium and isolated genomic DNA using the EpiCentre MasterPure DNA purification kit according to the manufacturer’s recommendations. DNA libraries were prepared using the Illumina TruSeq DNA Nano kit and sequenced on the Illumina HiSeq4000 platform (2 × 150 base pairs, bp) or on the Illumina HiSeq2000 platform (2 × 250 bp). Sequencing reads were quality filtered and trimmed using the BBTools suite (v.37.36; ref. 41) and quality of reads was assessed using FastQC (v.0.11.5; ref. 42). Draft genomes were assembled with SPAdes 3.11 (ref. 43) and annotated with prokka (v.1.12)44.

Average nucleotide identities were calculated using FastANI45 for members of the same family/order (Supplementary Table 13). The presence of phylogenetic signal in the plant protection against P. syringae and in plant colonization was tested by calculating Pageli’s J using the function phyllosig of the package phylows46 and by calculating Abouheif’s θ using the function abouheif.moran of the adephylog package47. To test whether the unknown taxonomic contribution affects the final calculation and conclusion of this analysis, we additionally calculated the metrics on randomly subsampled strains (Supplementary Fig. 3). In one set of analyses, we randomly subsampled all families with more than nine isolates. In another set, we first randomly picked one strain from clusters with >99.9% average nucleotide identity and then randomly subsampled all families with more than nine isolates. We performed 1,000 iterations for each.

To identify genes potentially associated with protection in *Rhizobium* spp. a list of all assigned COGs for the publicly available genomes of *Rhizobium* spp. of the Art-LSPIERER collection was downloaded from IMG/MER48. Subsequently, the list was filtered by COGs present in protective strains but not in non-protective strains. To the list of 39 genes associated with 6 secreted proteins (SPIH) in other *Rhizobium* genomes, we looked for the presence of orthologues of the T6SS after annotating the proteins using the egg-megmapper and eggNog database v.4.5 (ref. 49). We called the presence of a T6SS if the presence of a T6SS system was either predicted by TxSScan80 or if there were more than five of the core T6SS-associated orthologues present in the genome (COG5542, COG3137, COG3455, COG5501 and COG5516-3523).

**Generation of T6SS mutants.** We targeted the structural gene *tssL* as an essential gene for T6SS in *Agrobacterium tumefaciens*50. Mutants of the *tssL* gene were generated in *Rhizobium* Leaflol2 by gene replacement with a kanamycin resistance cassette according to Ledermann et al.51. DNA regions upstream and downstream of the target gene were amplified from genomic DNA and cloned into the SpeI restriction site into pREDlSIX1 to obtain plasmid pREDlSIX HR1/2. The kanamycin resistance cassette (Kmr) was cut out of pRGD_KmR with SpeI and gel-purified. The Kmr fragment was then ligated into the linearized pREDlSIX HR1/2 and transformed into E. coli DH5α. Transformants were selected on kanamycin and the origin of the kanamycin cassette was identified by colony PCR. The confirmed mutagenesis constructs were transformed into *Rhizobium* Leaflol2 by electroporation. A single colony of *Leaflol2* was inoculated into half-strength LB-Lennox liquid medium and incubated at 28 °C overnight. The culture was plated on ice for 15 min and then washed three times with ice-cold water, followed by one wash step with ice-cold 10% glycerol and concentrated 100-fold. To 10% glycerol (50% of) electrolytically competent cells were added with ~500 ng of purified plasmid DNA and electroporated at 2.2 kV.

Immediately after the pulse, 1 ml of half-strength LB-Lennox was added and cells were regenerated at 28 °C with agitation for 4–5 h. Transformants were selected on half-strength LB-Lennox supplemented with 50 µg ml−1 kanamycin and tested for double homologous recombination by PCR. Confirmed mutants (isolated from two independent plates) were restreaked at least three times before storing at −80 °C. A list of primers used is available in Supplementary Table 14.

**T6SS assay in vitro.** T6SS in vitro sensitivity was tested using a protocol adapted from previous experiences15,17,28. Attacker (triplicate) and target strains were grown in liquid culture at 28 °C in half-strength LB-Lennox (attacker) and LB-Lennox (target) with appropriate antibiotics, respectively. Cells were pelleted, washed once and resuspended in LB-Lennox. The OD600 were adjusted to 0.3 (attacker) and 0.01 (target). Suspensions were mixed at 1:1 ratio and 5 µl spotted onto LB-Lennox or R2A + M agar. After 17–24 h of incubation at 28 °C agar plugs containing the spots were placed in buffer (10 mM MgCl2, or 100 mM phosphate buffer pH 7.0), washed for 5 min and diluted suspension spotted onto R2A + M and selective R2A + M for determination of attacker and target c.f.u. (R2A + M supplemented with 25 µg ml−1 rifampicin). The experiment was performed three times with similar outcomes.

**Testing of T6SS dependency for plant protection.** Leaflol2 and its T6SS mutant (Leaflol2 tsL::Kmr) were tested in the 24-well plate system described above for plant protection with minor modifications. The Leaflol2 inocula were prepared from liquid cultures grown in half-strength LB-Lennox and supplemented with kanamycin when appropriate. Well-grown overnight cultures (OD > 1) were pelleted at room temperature, washed once with 10 mM MgCl2, and resuspended in 100 mM MgCl2, 0.1 M NaCl, was adjusted to OD600 = 1. Further diluted 1:100 strains were inoculated and infected as described above, with one half of the plate inoculated with the wild type and the other with the T6SS mutant. Four independent plates were prepared per experiment. Two axenic control plates were inoculated with 10 mM MgCl2, before infection. At 7 dpi, 18 plants per condition were harvested individually from three plates using the protocol described above but without pooling. For *Pst*, c.f.u. were determined on R2A supplemented with 20 µg ml−1 rifampicin. The experiment was performed twice with similar results.

**Statistics.** Data were analysed in RStudio54 with R 3.6.3 (ref. 55). No statistical methods were used to predetermine sample size. Sample size for protection assays was chosen on the basis of previous experiences and ensured each condition was within each experiment present on at least three independent 24-well plates. The disease phenotype scores were used to calculate the disease severity index (DSI) and the protection scores. The DSI reflects the occurrence of disease relative to the maximal possible disease outcome with all plants scored as dead (DSI of 100%) or all plants scored as completely healthy (DSI of 0%). The protection score of each strain was then deduced by subtracting the DSI from the DSI of the axenic control and dividing by the latter. Thus a protection score of 100 corresponds to all plants completely healthy and a protection score of 0 corresponds to no improvement relative to axenic controls. When several subsets of experiments were performed in parallel, values for axenic controls were combined to decrease the effect of single plants on the scaling. Luminescence data and data based on bacterial c.f.u. counts were log10-transformed before analysis. Data distribution was then assumed to be normal but this was not formally tested. For luminescence, log10-transformed values were tested for differences relative to the axenic control by two-sided Welch’s t-tests and corrected over all experiments for multiple testing using Holm’s method. To compare samples of log10-transformed luminescence values were scaled relative to the values for axenic and Fr1-inoculated controls of the same experiment. The arbitrary scale was adjusted so that 100 reflected a 1.3X stronger luminescence reduction than the positive control Fr1 and 0 no luminescence reduction relative to the axenic control. To identify the effect of plant genotype on protection by individual strains, we performed first three one-sided Welch’s t-tests per inoculation treatment, testing whether log10-transformed luminescence values were significantly higher in infected plants relative to non-infected control plants in each plant genotype (that is, *Pst* Col-0 versus CTL_Col-0 and *Pst bak1/bkk1* versus CTL_bak1/bkk1) and testing whether they were higher in infected bak1/bkk1 plants relative to Col-0 plants (that is, *Pst bak1/bkk1 versus Pst Col-0*). P values were corrected using Benjamini–Hochberg’s method. Furthermore, a linear model of infected log10-transformed luminescence values was fitted for each strain in combination with the respective axenic control treatments. We used the full interaction model of strain treatment and genotype using experiment as a blocking factor when appropriate. When results were heteroscedastic experiment 1Levene’s test in the R package varPower56, a generalized least squares model was fitted using the function gls within the R package nlme with weights = varPower. P values of the obtained Strain:Genotype:bak1:bkk1 interaction estimates were corrected for multiple testing using Benjamini–Hochberg’s method. For mixes of three strains, a one-way analysis of variance (ANOVA) followed by Tukey’s post-hoc test implemented in the R package emmeans57 was performed for each mix and experiment individually, including plate information as a blocking factor. Overall results were obtained by averaging the results from all individual experiments per treatment.

**Data availability**

Sequencing data for this study have been deposited in the European Nucleotide Archive under accession PRJEB47672. Other genome data are available from NCBI under accession PRJNA297956, PRJNA471493 and PRJNA84361. The eggNog database is available from http://eggno.g45.embl.de/#/app/home. *Rhizobium* genomes used for comparative genomics are available from https://img.jgi.doe.gov/sites/IMG_146857740/db/146857740-2634221832/2634221860, 2634221889, 2634221891, 2634221905, 2634221915, 2634221931 and 2634221933-36. Source data are provided with this paper.

**Code availability**

The scripts used for sequencing data processing, genome assembly and data analysis are available at https://gitlab.ethz.ch/chvogel1/vogel_natmicro_2021/.

Received: 27 January 2021; Accepted: 15 October 2021; Published online: 24 November 2021.
2. Udvardi, M. & Poole, P. S. Transport and metabolism in legume–rhizobia symbioses. Annu. Rev. Plant Biol. 64, 781–805 (2013).

3. Parniske, M. Arabidopsis mycorrhiza: the mother of plant root endosymbioses. Nat. Rev. Microbiol. 6, 763–775 (2008).

4. Bulgarelli, D., Schläpfer, K., Spaepen, S., Ver Loren van Themaat, E. & Schulze-Lefert, P. Structure and functions of the bacterial microbiota of plants. Annu. Rev. Plant Biol. 64, 807–838 (2013).

5. Wagner, M. R. et al. Natural soil microbes alter flowering phenology and the intensity of selection on flowering time in a wild Arabidopsis relative. Ecol. Lett. 17, 717–726 (2014).

6. Glick, B. R. Bacteria with ACC deaminase can promote plant growth and help to feed the world. Microbiol. Res. 169, 30–39 (2014).

7. Ritpítatphong, U. et al. The microbiome of the leaf surface of Arabidopsis thaliana supports against a fungal pathogen. New Phytol. 210, 1033–1043 (2016).

8. Comment, S., D. Tassaneon, V., Cospic, I. & Sterflinger, K. A. Plant–soil microbe interactions promoting plant growth-benefiting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. Appl. Environ. Microbiol. 71, 4951–4959 (2005).

9. Santhanam, R. et al. Native root-associated bacteria rescue a plant from a soil-borne wilt disease that emerged during continuous cropping. Proc. Natl Acad. Sci. USA 112, E5013–E5020 (2015).

10. Carrion, V. J. et al. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science 366, 606–612 (2019).

11. Toju, H. et al. Core microbiomes for sustainable agroecosystems. Nat. Plants 4, 247–257 (2018).

12. Vorholt, J. A., Vogel, C., Carlström, C. J. & Müller, D. Establishing causality: opportunities of synthetic communities for plant microbiome research. Cell Host Microbe 22, 142–155 (2017).

13. Busby, P. E. et al. Research priorities for harnessing plant microbiomes in sustainable agriculture. PLoS Biol. 15, e2001793 (2017).

14. Berg, C. G. Plant–microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. Appl. Microbiol. Biotechnol. 84, 11–18 (2009).

15. Cook, R. J. in Encyclopedia of Agriculture and Food Systems Vol. 4 (ed. Van Allen, N. K.) 441–455 (Academic Press, 2014).

16. Woller, D. M., Raaijmakers, J. M., Gardner, B. B. & Thomashow, L. S. Microbial populations responsible for specific soil suppressiveness to plant pathogens. Ann. Rev. Phytopathol. 40, 309–348 (2002).

17. Letz, E., Eisenhauer, N., Rall, B. C., Scheu, S. & Jousset, A. Unravelling linkages between plant community composition and the pathogen-suppressive potential of soils. Sci. Rep. 6, 23584 (2016).

18. Finkel, O. M., Castrillo, G., Herrera Paredes, S., Salas Gonzàlez, I. & Dangl, J. L. Understanding and exploiting plant beneficial microbes. Curr. Opin. Plant Biol. 38, 155–163 (2017).

19. Jacobsen, B. J. in Microbial Ecology of Aerial Plant Surfaces (eds Bailey, M. J. et al.) 133–147 (CAB International, 2006).

20. Barber, C. et al. In planta biocentral of Pectobacterium atrosepticum by Rhodococcus erythropolis involves silencing of pathogen communication by the rhodococcal gamma-lactone catabolic pathway. PLoS ONE 8, e66642 (2013).

21. Pieterse, C. M. et al. Induced systemic resistance by beneficial microbes. Annu. Rev. Phytopathol. 52, 347–375 (2014).

22. Dann, J. L., Harthoorn, R. M. & Staskawicz, B. J. PVX: a model for the study of plant immunity. Plant Mol. Biol. 54, 347–357 (2004).

23. Decoin, V. et al. A type VI secretion system is involved in Pseudomonas type 6 secretion system is related to mucoidy, motility and bacterial competition. BMC Microbiol. 15, 72 (2015).

24. Gobena, M. & Verdu, M. Predicting microbial traits with phylogenies. ISME J. 10, 959–967 (2016).

25. Melnyk, R. A., Hossain, S. S. & Haney, C. H. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. ISME J. 13, 1575–1588 (2019).

26. Haney, C. H., Samuel, B. S., Bush, J. & Ausubel, F. M. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. Nat. Plants 1, 15051 (2015).

27. Savory, E. A. et al. Evolutionary transitions between beneficial and phytopathogenic rhodococcus challenge disease management. eLife 6, e39025 (2017).

28. Haas, D. & Défago, G. Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat. Rev. Microbiol. 3, 307–319 (2005).

29. Raaijmakers, J. M. et al. Dose–response relationships in biological-control of Fusarium wilts by Pseudomonas spp. Phytopathology 85, 1075–1081 (1995).

30. Kwak, M. J. et al. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat. Biotechnol. 36, 1100–1109 (2018).

31. Berg, M. & Koskelka, B. Nutrient- and disease-dependent microbe-mediated protection against a plant pathogen.Curr. Biol. 28, 2487–2492 (2018).

32. Berendsen, R. L. et al. Disease-induced assemblage of a plant-beneficial bacterial consortium. ISME J. 12, 1496–1507 (2018).

33. Rudrappa, T., Czymmek, K. J., Pare, P. W. & Bais, H. P. Root-secreted malic acid recruits beneficial soil bacteria. Plant Physiol. 148, 1547–1558 (2008).

34. Liu, H., Brettell, L. E., Qiu, Z. & Singh, B. K. Microbiome-mediated stress resistance in plants. Trends Plant Sci. 25, 733–743 (2020).

35. Song, Y. et al. FERONIA restricts Pseudomonas in the rhizosphere microbiome via regulation of reactive oxygen species. Nature Plants 7, 644–654 (2021).

36. Bartoli, C. et al. In situ relationships between microbiota and potential pathogens in Arabidopsis thaliana. ISME J. 12, 2024–2038 (2018).

37. Andrews, J. H. Biological control in the phyllosphere. Annu. Rev. Phytopathol. 30, 603–635 (1992).

38. Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. Annu. Rev. Phytopathol. 52, 101–121 (2014).

39. Lindow, S. E. & Brandl, M. T. Microbiology of the phyllosphere. Appl. Environ. Microbiol. 69, 1875–1883 (2003).

40. Monier, J. M. & Lindow, S. E. Frequency, size, and localization of bacterial aggregates on bean leaf surfaces. Appl. Environ. Microbiol. 70, 346–355 (2004).

41. Remus-Emsermann, M. N. P. et al. Spatial distribution analyses of natural phyllosphere-colonizing bacteria on Arabidopsis thaliana revealed by fluorescence in situ hybridization. Environ. Microbiol. 16, 2329–2340 (2014).

42. Wei, Z. et al. Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health. Nat. Commun. 9, 1100–1113 (2018).

43. Decoin, V. et al. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen Xanthomonas. FEMS Microbiol. Rev. 44, 1–32 (2020).

44. Baston, J. M., Newell, P. D. & Douglas, A. E. Metagenome-wide association of microbial determinants of host phenotype in Drosophila melanogaster. Mol. Ecol. 5, e01631–e01614 (2016).

45. MacDonald, N. J. & Beiko, R. G. Efficient learning of microbial genotype–phenotype association rules. Bioinformatics 31, 2329–2340 (2015).

46. Levy, A. et al. Genomic features of bacterial adaptation to plants. Nat. Genet. 50, 138–150 (2018).

47. Ryu, C. M. Against friend and foe: type 6 effectors in plant-associated bacteria. J. Microbiol. 53, 201–205 (2018).

48. Bernal, P., Llamas, M. A. & Filoux, A. Type VI secretion systems in plant-associated bacteria. Environ. Microbiol. 20, 1–15 (2018).

49. Decoin, V. et al. Type VI secretion system is involved in Pseudomonas fluorescens bacterial competition. PLoS ONE 9, e98411 (2014).
64. Mosquito, S. et al. In planta colonization and role of T6SS in two rice Kossakonia endophytes. Mol. Plant Microbe Interact. 33, 349–363 (2020).
65. Vlot, A. C. et al. Systemic propagation of immunity in plants. New Phytol. 229, 1234–1250 (2021).
66. Roux, M. et al. The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SEK3 and BKK1/SEK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. Plant Cell 23, 2440–2455 (2011).
67. Fan, J., Crooks, C. & Lamb, C. High-throughput quantitative luminescence assay of the growth in planta of Pseudomonas syringae chromosomally tagged with Photorhabdus luminescens luxCDABE. Plant J. 53, 393–399 (2008).
68. Wu, Y. W. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. BMC Genomics 19, 921 (2018).
69. Bushnell, B. BBMap: Short read aligner, and other bioinformatic tools (SourceForge, version 37.56); https://sourceforge.net/projects/bbmap/.
70. Andrews, S. FastQC: A quality control tool for high throughput sequence data (Babraham Institute, 2010); https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
71. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. I. Comput. Biol. 19, 455–477 (2012).
72. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).
73. Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114 (2018).
74. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223 (2012).
75. Jombart, T., Balloux, F. & Dray, S. adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics 35, D649–D659 (2019).
76. Chen, I. A. et al. IMG/M v5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucl. Acids Res 47, D666–D677 (2019).
77. Mukherjee, S. et al. Genomes OnLine database (GOLD) v7: updates and new features. Nucl. Acids Res. 47, D649–D659 (2019).
78. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122 (2017).
79. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucl. Acids Res. 44, D286–D293 (2016).
80. Abbo, S. S. et al. Identification of protein secretion systems in bacterial genomes. Sci. Rep. 6, 23080 (2016).
81. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? BMC Genomics 10, 104 (2009).
82. Lin, J. S., Ma, L. S. & Lai, E. M. Systematic dissection of the agrobacterium type VI secretion system reveals machinery and secreted components for subcomplex formation. PLoS ONE 8, e67647 (2013).
83. Ledermann, R., Strelbl, S., Kampik, C. & Fischer, H. M. Versatile vectors for efficient mutagenesis of Bradyrhizobium diaeoeicis and other Alphaproteobacteria. Appl. Environ. Microbiol. 82, 2791–2799 (2016).
84. Wu, C. F. et al. Plant-pathogenic Agrobacterium tumefaciens strains have diverse type VI effector-immunity pairs and vary in in-planta competitiveness. Mol. Plant Microbe Interact. 32, 961–971 (2019).
85. Kassambura, A. rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.6.0 https://CRAN.R-project.org/package=rstatix (2020).
86. Pinheiro J., DebRoy S., Sarkar D. & Team, R. C. nlme: Linear and nonlinear mixed effects models. R package version 3.1-144 https://CRAN.R-project.org/package=nlme (2020).
87. Lenth, R. emmeans: Estimated marginal means, aka least-squares means. R package version 1.6.1 https://CRAN.R-project.org/package=emmeans (2021).
88. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528 (2019).
89. Gu, Z. G., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circize implements and enhances circular visualization in R. Bioinformatics 30, 2811–2812 (2014).
90. Wickham, H. et al. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686 (2019).

Acknowledgements
We thank M. Bertfeld-Miller for technical assistance and A. Levy for helpful discussions. A. Italiana baki/bkk1 was kindly provided by C. Zipfel (University of Zurich, Switzerland). The study was supported by an ERC Advanced Research Grant (PhyMo - 668991 to J.A.V.), ETH Zurich, a grant from the German Research Foundation (DECPRF, no. SPP2125 to J.A.V.) and the NCCR Microbiomes (to J.A.V.), funded by the Swiss National Science Foundation.

Author contributions
C.V., D.B.P. and J.A.V. conceived and designed the study. C.V., D.B.P. and M.S. performed testing SynComs. C.V. tested in vitro and in planta effect of Leaf202 and its T6SS mutant with help from M.S. C.V. and M.S. extracted DNA for genome sequencing. C.V. analysed sequencing data and assembled genomes. C.V. and D.B.P. analysed screening and validation data. C.V. analysed SynCom data with help from M.S. and analysed T6SS data. C.V. and J.A.V. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41564-021-00997-7.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-00997-7.
Correspondence and requests for materials should be addressed to Julia A. Vorholt.
Peer review information Nature Microbiology thanks the anonymous reviewers 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, corrected publication 2022
Extended Data Fig. 1 | Protection potential by protective At-LSPHERE strains scales with infection titre. A. thaliana were inoculated with fully protective strains Leaf15, Leaf154 or Leaf21 (protection score >90), or with protective strains Leaf205 or Leaf233 (protection score >75) and infected with Pst at the regular infection titre (OD 0.00003), a 100x higher (OD 0.003) or a 10'000x higher (OD 0.3) infection titre. a) Luminescence indicative of pathogen colonization was measured at 6 d post infection (dpi). Shown are boxplots and individual data points. Letters indicate significant differences for each infection titre based on ANOVA followed by Tukey’s post-hoc test ($P < 0.05$, $n = 16–24$). Exact $P$ values and number of biological replicates are provided in Supplementary Table 3. Boxplots depict the median and interquartile range with whiskers extending to maximum 1.5x the interquartile range. b) Plants were scored for disease at 21 dpi on a scale of 1 (healthy) to 5 (dead). c) Exemplary images of plants at 21 dpi showing protection of plants by fully protective strains at high infection titre and reduced protection by Leaf205 and Leaf233 at increasing infection titre.
Extended Data Fig. 2 | Protection and luminescence reduction correlate. Mean protection score and mean luminescence reduction (that is pathogen colonization reduction) correlate well for most strains (Pearson’s $R = 0.927, t = 36.8, \text{df} = 221, P < 2.2 \times 10^{-16}$). Colours refer to phylum/class. a.u., arbitrary units.
Extended Data Fig. 3 | Random SynComs of 10 non-protective strains or all Methylobacterium do not protect Arabidopsis against Pst. Plants were inoculated with random SynComs of 10 strains (M10.2-M10.6) containing only non-protective strains, all 32 non-protective Methylobacterium spp. (M10.1), Fr1 or the control SynCom M10.7, which contains one protective strain before infection with lux-tagged *P. syringae* DC3000 (Pst). **a** Distribution of disease scores on a scale of 1 (healthy) to 5 (dead) at 21 d post infection (dpi). **b** Luminescence of the pathogen at 6 dpi. Boxplots depict the median and interquartile range with whiskers extending to maximum 1.5x the interquartile range. *P* values for the comparison to axenic infected controls are indicated (two-sided Welch's t-test, corrected for multiple testing using Holm's method, *n* = 15–23). Exact number of replicates are provided in Supplementary Table 4.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Random SynComs of a mixture of 10 non-protective and intermediate protective strains improve plant phenotype. Plants were inoculated with random SynComs of 10 strains or Fr1 before infection with lux-tagged *P. syringae* DC3000 (Pst). **a** Distribution of disease scores at 13 days post infection (dpi) on a scale of 1 (healthy) to 5 (dead). **b** Luminescence as proxy of pathogen colonization at 6 dpi. Boxplots depict the median and interquartile range with whiskers extending to maximum 1.5x the interquartile range. *P* values for the comparison to axenic infected controls are indicated (two-sided Welch’s *t*-test, corrected for multiple testing using Holm’s method, *n* = 14-24). Exact number of replicates are provided in Supplementary Table 5. **c** Colonization by individual SynComs on non-infected plants at 12 d post inoculation. Shown are the mean and individual data points of 3 replicates consisting of two plants each. **d** Correlation of mean luminescence and mean disease score (*R* = 0.92, *t* = 14.76, d.f. = 39, *P* < 2.2 × 10⁻¹⁶).
Extended Data Fig. 5 | Dropout of one or two strains from a random SynCom of 10 strains can affect plant protection by SynComs. Plants were inoculated with random SynComs of 10 strains, dropout communities thereof or individual strains and infected with lux-tagged *P. syringae* DC3000 (Pst).

**a)** Distribution of disease scores at 13 d post infection (dpi) on a scale of 1 (healthy) to 5 (dead). **b)** Luminescence as proxy of pathogen colonization at 6 dpi. Boxplots depict the median and interquartile range with whiskers extending to maximum 1.5x the interquartile range. *P*-values for indicated comparisons are shown (two-sided Welch’s t-test, n = 21-24 plants per condition). Exact number of replicates are provided in Supplementary Table 6. **c)** Colonization by the individual SynComs on non-infected plants at 12 d post inoculation. Shown are the mean and individual data points of 3 replicates consisting of 2 plants each.
Extended Data Fig. 6 | SynComs of three strains can improve protection phenotypes relative to individual strains. Plants were inoculated with SynComs of three strains (M3.1, M3.2 and M3.3; comprised of non-protective and intermediate protective strains) or the strains individually and infected with lux-tagged P. syringae DC3000 (Pst). a) Distribution of disease scores on a scale of 1 (healthy) to 5 (dead) at 13 d post infection (dpi). b) Pathogen luminescence at 6 dpi. Boxplots depict the median and interquartile range with whiskers extending to maximum 1.5× the interquartile range. Letters indicate statistical significance within each SynCom (one-way ANOVA with Tukey’s post-hoc test, n = 16–24). Exact P values and number of replicates are provided in Supplementary Table 8.
Extended Data Fig. 7 | T6SS gene cluster presence in At-LSPHERE strains and Sphingomonas melonis Fr1. The outer rings reflect mean protection scores against lux-tagged P. syringae DC3000 on Col-0 plants and the presence of predicted T6SS gene clusters, respectively. a.u., arbitrary units.
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
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Data was collected as described in the manuscript. Luminescence data was collected using the IVIS Spectrum Imaging System (Xenogen) with Living Image Software 4.2. Genome data was collected using the Illumina platform either sequencing on a HiSeq4000 or HiSeq2500 instrument.

Data analysis

Data was analyzed and plotted in RStudio 1 with R environment v3.6.3 as described in the manuscript. Apart from functions available in base R, we used the tidyverse (1.3.0) toolbox and rlang (v0.4.6) to manipulate data. For statistical analysis we additionally used the following packages: emmeans v1.6.1, rstanix v0.6.0, nlme (v3.1-144), phytools v0.7-47, adephylo (v1.1-11).

For sequencing data, reads were quality filtered and trimmed using the BBtools suite (v37.56) and quality of reads was assessed using FastQC (v0.11.5). Draft genomes were assembled using SPAdes v 3.11.0. Proteins were identified using prokka v1.12. A phylogenetic tree was prepared using ezTree (v0.1). TxSScan was used to screen the genomes for the presence of TSSS systems.

For plotting and preparing of tables the following R packages were used in addition to the previously mentioned ones: ape (5.4-1), circize (0.4.10), ggpubr (0.3.0), RColorBrewer (1.1-2), gridExtra (2.3), ggforce (0.3.2), scales (1.1.1), broom (v0.7.3), readxl (v. 1.3.1)

Figures were finalized in Adobe Illustrator 2020.

Code is available at https://gitlab.ethz.ch/chvogel1/vogel_natmicro_2021/.

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

Source data are provided with this paper. Sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession PRJEB47672. Previously published genome data are available from NCBI under accessions PRJNA297956, PRJNA471493 and PRJNA84361. The EggNOG-Database is available from http://eggnog45.embl.de/#/app/home. Rhizobium genomes used for comparative genomics are available from https://img.jgi.doe.gov/m under IMG Genome IDs 2643221743, 2643221780, 2643221832, 2643221860, 2643221889, 2643221891, 2643221896, 2643221905, 2643221915, 2643221931, 2643221933-36.

## 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.

- [x] 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 | No statistical methods were used to predetermine sample size. Sample size for protection assays was chosen based on previous experience (Vogel et al 2012, Appl. Environ. Microbiol.; Vogel et al. 2016, New Phytol), the restraints of the growth chamber space and what is feasible to handle in the laboratory at once. As we were primarily focused on protection, we maximised the plants used for protection assays rather than colonization determination. To reduce sampling efforts but still increase the number of plants used for colonization, we generally pooled two plants per sample. For testing the effect of the T6SS mutation on Pst colonization, we harvested individual plants and increased the number of replicates accordingly as we knew from previous experience that Pst colonization tends to be more variable in our screening system than commensal colonization (Vogel et al. 2016, New Phytol). For the T6SS in vitro assay, sample size was chosen based on previous studies (e.g. Levy et al 2017, Nature Genetics). |
| Data exclusions | We excluded data as described in the manuscript. Two treatments (Leaf50 and Leaf75) were completely excluded from assessment for plant protection because of strong plant phenotype in absence of infection (see Supplementary Figure 1). As such a phenotype was not anticipated, this was not based on pre-established exclusion criteria. Furthermore, we did not score disease phenotypes on bak1/bkk1 plants inoculated with Xanthomonads because of Pseudomonas infection-independent disease. We also excluded individual plants when plant development was not according to experience (e.g. no or late germination, growth upside down, strong chlorosis, seed pushed into agar) or when there was a problem with the treatment (e.g. plant wrongly treated, contaminated). For luminescence measurements, we are missing information for a small number of plants due to technical problems during data acquisition. For colonization determination, we seeded almost twice the number of plants needed for harvesting possibly leading to loss of plants as described above. |
| Replication | Due to the amount of work, we screened most strains as well as strain combinations only once. For strains showing intermediate protection, a selection of non-protective strains, as well as for interesting strain combinations, however, we repeated the experiment more than once and found a good reproducibility (see e.g. Supplementary Figure 1). The SynCom10 drop-out was repeated once for the most interesting mix with minor modifications, resulting in similar results. The SynCom3 M3.1 was replicated three times (two replicates shown in the manuscript) with similar results. Scoring of plant disease symptoms was whenever possible done by two researchers. The in vitro phenotype of Leaf202 and its tssL mutants was tested in three independent experiments with similar results. The plant protection by Leaf202 and its tssL mutant was replicated twice (one replicate shown) with similar results. |
| Randomization | For plant experiments, we used 24-well plates that were divided into quarters (for protection assay) or columns (for colonization assay) of the same treatment to reduce errors during plant treatment. The different treatments were randomly distributed to the plates (using a random number generator on www.randomization.com) with minor modifications upon identification of plates with same treatment combinations. For identifying the effect of plant genotype on plant protection, only one inoculation treatment per plate was used for assessing plant protection, however, the position of infected/non-infected plants and of the different plant genotypes was random. Furthermore, plates were randomly placed in the growth chamber and shuffled at least twice a week (both within and between shelves). |
| Blinding | For all plant experiments, the treatments were assigned a number prior to inoculation and plates were labelled with number and letter codes to prevent potential biases due to knowledge of treatment information during data collection. Metadata was then added before analysis. |

### 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). |

April 2020
### Ecological, evolutionary & environmental sciences study design

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

| Field work, collection and transport |
|--------------------------------------|
| **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. |

---

### 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

| Related to study | Yes | No |
|------------------|-----|----|
| Antibodies | ☑ | ☐ |
| Eukaryotic cell lines | ☑ | ☐ |
| Palaeontology and archaeology | ☑ | ☐ |
| Animals and other organisms | ☑ | ☐ |
| Human research participants | ☑ | ☐ |
| Clinical data | ☑ | ☐ |
| Dual use research of concern | ☑ | ☐ |

### Methods

| Related to study | Yes | No |
|------------------|-----|----|
| 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**
  - **Cell line source(s)**: 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**: Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

### 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**: Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

**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.
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

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

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.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

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, MNI305, 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

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.