Mapping phyllosphere microbiota interactions in planta to establish genotype-phenotype relationships

Martin Schäfer, Christine M. Vogel, Miriam Bortfeld-Miller, Maximilian Mittelviefhaus and Julia A. Vorholt

Host-associated microbiomes harbour hundreds of bacterial species that co-occur, creating the opportunity for manifold bacteria–bacteria interactions, which in turn contribute to the overall community structure. The mechanisms that underlie this self-organization among bacteria remain largely elusive. Here, we studied bacterial interactions in the phyllosphere microbiota. We screened for microbial interactions in planta by adding 200 endogenous strains individually to a 15-member synthetic community that was co-cultivated in planta. We identified interactions with a range of different outcomes, including negative, amensalistic, mutualistic, and parasitic interactions. In planta interactions were more diverse than those previously identified in vitro. The identification of specific bacterial interactions in planta allowed us to uncover a strong microbial interaction that occurs despite a spatially heterogeneous environment.

In the environment, bacteria often occur in multispecies communities. These may consist of a few species or comprise hundreds of species as is often found in host-associated microbiomes. Due to the complexity of these systems, it is particularly challenging to identify microbes that interact and thereby shape the community. Interactions among members of the microbiota are often studied in vitro, but the observations are difficult to translate to the host-associated context due to differences in spatial and nutritional environments.

The plant microbiome has been described for many species and the different plant compartments. Members of the microbiota can provide beneficial functions to the plant host, such as growth promotion or increased resistance to biotic and abiotic stresses. While there is variation in the composition of colonizing bacteria at the species level, the community is largely conserved at higher phylogenetic level over time and different locations.

The mechanisms that govern microbial community assembly in the plant host have only started to be unravelled. Apart from host factors, microbe–microbe interactions also contribute to community assembly, but little is known about the overall impact and specific mechanisms that prevail in planta. In general, interactions can be categorized on the basis of the outcome for the interaction partners. Positive interactions include commensalism and mutualism, in which one or both microbes benefit. Negative interactions include competition (negative for both microbes), amensalism (negative for one, neutral for the other microbe) and parasitism (negative for one and positive for the other microbe).

The genomes of bacterial phyllosphere and rhizosphere colonizers harbour a plethora of biosynthetic gene clusters. Therefore, the production of antimicrobial compounds to inhibit competitors is proposed to be one mechanism that shapes microbial communities. Previous studies have also shown that the presence of fungi or oomycetes can alter bacterial community composition. So far, most interactions have been described in the context of pathogen infection in diseased plants. Interactions among bacteria in healthy plants and hence the fundamental mechanisms driving community assembly are less studied.

Investigating bacteria–bacteria interactions directly in the host is challenging and usually relies on co-occurrence networks based on natural variation over time or between individuals. The development of computational approaches applicable to microbiomes is a rapidly growing field and current limitations of the methods in interpreting species interactions in microbiome studies have been highlighted. An inherent drawback of all co-occurrence network approaches is that they infer ecological associations on the basis of abundance correlations and thus fall short in demonstrating direct interactions. Synthetic community experiments are attractive in this regard because: (1) the complexity of the system under study can be reduced and (2) individual bacteria can be removed or replaced to examine the community response to these controlled perturbations. Over the past years, extensive culture collections have been established for several plant species, including Arabidopsis, allowing synthetic communities to be assembled not only in vitro but also in planta.

Here we describe a systematic in planta screen to identify positive and negative interactions of 200 Arabidopsis thaliana leaf bacteria with a synthetic 15-member fotal community in a gnotobiotic model system. Mapping interactions of a broad range of bacteria in planta...
the phyllosphere allowed the identification of emerging patterns for groups of microbes as well as the identification of specific interactions. We then decipher an exemplary specific microbial interaction in planta to expand our mechanistic understanding of how these interactions work at the molecular level.

Results
Selection and characterization of the synthetic community. We aimed to dissect interactions among leaf strains during phyllosphere colonization of the model plant Arabidopsis thaliana. To this end, we selected 15 strains from the At-LSPHERE culture collection to assemble a focal community against which we probed 200 additional strains of the At-LSPHERE collection and a few selected -LSPHERE culture collection and to identify an ideal starting point to identify interactions directly induced by additional strains while still capturing potentially emerging higher-order interactions.

The focal community strains were selected to span the phylogenetic diversity of the At-LSPHERE culture collection and to include genera that are abundant in the Arabidopsis phyllosphere and part of the phyla Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes. Upon mono-association, most isolates colonized the phyllosphere in the range of $10^7$–$10^8$ colony-forming units per gram plant fresh weight (c.f.u. g$^{-1}$ FW) in a calcine clay-based gnotobiotic model system. We also characterized the composition of the combined focal community in the phyllosphere by 16S ribosomal RNA gene sequencing by combining the strains in approximately equal ratio in the inoculum and provided data across two independent experiments are shown.

Extended Data Fig. 1, Supplementary Figs. 3 and 4, and Table 2). We also included conditions in which the focal strains themselves were

Fig. 1 | A 15-strain focal SynCom to map bacteria–bacteria interactions. a. Graphical representation of experimental design for interaction screening. b. Phylogenetic tree representing the selected strains based on full-length 16S rRNA gene sequences. c. Phyllosphere colonization by each focal strain in Col-0 plants. Seedlings were inoculated with bacterial suspensions 7 d after planting and bacteria were enumerated 21 d later by c.f.u. counting on R-2A + M agar. The limit of detection is indicated by a dashed line. Data across two independent experiments are shown.
**Fig. 2 | Interactions of the At-LSPHERE collection with the focal SynCom in planta.**

**a.** Interaction map showing all interactions identified between At-LSPHERE strains (outer) and the focal community (inner) based on log_{2}-transformed abundance changes (DESeq2-normalized counts, Wald test, Benjamini-Hochberg corrected, $P_{adj} \leq 0.01$, $n_{focal\ community} = 16$, $n_{drop-in} = 3$). Positive (red) and negative (blue) interactions are shown as connecting lines between drop-in strains and focal strains. Line thickness corresponds to fold change. Bars in the outermost ring correspond to the overall effect size of the perturbation on the focal community (PCA, PERMANOVA). Strains are ordered by phylogeny on the basis of full-length 16S rRNA gene sequences and dots are coloured by bacterial phyla or Proteobacteria class. Drop-in strains that belong to the same ASV as a focal strain are labelled in grey as for these, no interactions with the focal strain of the same ASV can be determined.

**b.** Pairwise strain inoculations of selected interacting strains identified in **a**. Top: the focal strain log_{2}FC observed in **a** is shown as a reference for each strain pair. Bottom: the log_{2}FCs (pairwise inoculation vs mono-association) for the focal strain and drop-in strain are shown on the basis of absolute abundances obtained by c.f.u. enumeration for two biological replicates. The colour of the boxes reflects the observed log_{2}FC and the black frames around the boxes indicate a significant difference compared with the mono-association condition (two-sided Wilcoxon rank-sum test, $P \leq 0.05$). Combinations that were not tested in a given experiment are marked with a grey box. Seedlings were inoculated at day 14 (replicate 1, R1) or day 7 (replicate 2, R2) after planting. Exact $P$ values and number of replicates are provided in the source data.

**c.** Phyllosphere colonization by *Methylobacterium* Leaf88 in mono-association or in combination with *Xylophilus* Leaf220, *Methylphillus* Leaf408 or both. Shown are the median and individual datapoints of log_{10}-transformed c.f.u. g^{-1} FW across two independent experiments. Exact $P$ values (Kruskal-Wallis and post-hoc Dunn test, Bonferroni adjusted $P$) are indicated above and log_{2}FCs below the graph. For the corresponding colonization levels of Leaf220 and Leaf408, see Extended Data Fig. 3.
added as a drop-in strain in the inoculum. These control conditions had no significant impact on the composition of the focal community, except for the Leaf427 drop-in that affected Leaf 33 (log-fold change (FC): −2.52, $P_{adj}$: 0.0064). Thus, overall this result validated the robustness of the assembled focal community.

In total, we identified 84 interactions between strains (Wald test, Benjamini-Hochberg adjusted $P \leq 0.01$) upon addition of 57 of the 200 strains tested. For the selected focal community, 90% of identified interactions were negative, indicating competitive interactions. We also observed that phylogenetically closely related isolates showed similar effects on the focal community, indicating trait conservation. The majority of negative interactions were observed with phylogenetically closely related strains, except for *Pseudomonas* Leaf 129, which also frequently interacted with distantly related strains (Extended Data Fig. 2). For focal strains that showed at least three negative interactions within the genus, we tested whether within-genus interactions were more frequent than expected by chance (Fisher’s exact test, Bonferroni adjusted $P \leq 0.05$). This was the case for *Rhizobium* Leaf371, *Sphingomonas* Leaf67 and *Methylobacterium* Leaf88, suggesting competition among closely related isolates (Supplementary Table 3). Intriguingly, among the interactions found were some that have previously been identified in a larger 62 strain SynComx (for example Leaf88–Leaf416, Leaf129–Leaf233 and Leaf220–Leaf262), demonstrating that these interactions are also observed in even more complex communities.

We also estimated the impact of each added strain on the focal community by calculating the observed effect size on the basis of principal component analysis (PCA) (Fig. 2a and Supplementary Table 4). Except for *Aeromicrobium* Leaf245, all strains with an effect size larger than 20% interacted with two or more focal strains. *Serratia* Leaf50 interacted with four focal community strains—the maximum observed—and had an effect size of 27%, suggesting strong remodelling of the community. In fact, this strain and the plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 were the only strains that induced disease symptoms in some plants when present in the community (Supplementary Fig. 5). This is in line with the plant pathogenic behaviour of *Serratia* Leaf50 and may explain the strong impact on the focal community composition.

To test whether the identified interactions were binary or higher-order interactions, and to independently validate the screen above using a complementary read-out allowing absolute quantification, we performed pairwise co-inoculations of Arabidopsis and compared colonization relative to mono-colonization using c.f.u. counts. Notably, 7 out of 12 interactions tested could be recapitulated upon co-inoculation, indicating that the observed interactions do not require a community context (Fig. 2b). This complementary approach also allowed us to study the effect of the interaction on the drop-in strain during co-inoculation. For *Rhizobium* Leaf68 and *Acidovorax* Leaf76 and Leaf160, which all had negative effects on the interacting focal strains, no significant abundance changes were observed upon co-colonization compared with mono-association, indicating amensalism. For the interaction between *Xanthomonas* Leaf148 and the focal strain *Aeromicrobium* Leaf272, Leaf272 abundance increased when co-inoculated while Leaf148 abundance was not affected, suggesting commensalism.

Because higher-order interactions would not manifest in pairwise inoculations, we also tested a trio in planta. We speculated that the addition of *Xylophilus* Leaf220 might be necessary to recapitulate the interaction of *Methylobacterium* Leaf88 and *Methylobacterium* Leaf408, or *Methylobacterium* Leaf88 (Extended Data Fig. 3), suggesting a similar mechanism to support the growth of Leaf220, possibly linked to the methylotrophic lifestyle of both strains.

In summary, our screening approach allowed us to map bacterial interactions emerging due to targeted perturbation of a community by strain addition. These interactions are readily reproducible with a complementary read-out in minimal synthetic communities of two or three strains.

**Nocardioides** Leaf374–*Aeromicrobium* Leaf245 interaction. We had previously shown the potential of the leaf isolates to produce novel antimicrobial compounds. Out of 18 inhibitory interactions observed in vitro between the tested strains, 2 were also identified in planta (Supplementary Table 5), one of which was between *Nocardioides* Leaf374 (Leaf374) and *Aeromicrobium* Leaf245 (Leaf245). This interaction was not only conspicuous because of the overlap with previous in vitro data but, with a 50-fold reduction in relative abundance of Leaf374, was also particularly prominent in planta (Fig. 2a). We then further validated this strain pair by sequential phyllosphere inoculation. The order of inoculation did not matter (Extended Data Fig. 4) and Leaf245 also reduced an already established population of Leaf374 (Fig. 3a and Supplementary Fig. 6). The extent of Leaf374 reduction in the binary validation was two orders of magnitude (log,FC: −7.6 to −4.1), which is in a similar range to that in the in planta screen (log,FC: −5.7). To further characterize this conspicuous interaction, we reproduced it in vitro and found that also cell-free supernatant prepared from suspended Leaf245 cells was sufficient for inhibition of Leaf374 (Methods and Fig. 3b). Further, the supernatant not only inhibited growth of Leaf374, but also efficiently lysed the bacterial cells (Supplementary Video). Since the activity was lost upon boiling or addition of organic solvents (Supplementary Fig. 7), we speculated that the active molecule was probably a protein. Next we fractionated the culture supernatant of Leaf245, identified several protein candidates that could cause the observed inhibition (Fig. 3c, d and Extended Data Fig. 5a) and heterologously expressed these in *Escherichia coli* (see Supplementary Note for details). Among these proteins was a putative M23 family endopeptidase (ASF05_00205), which resulted in inhibition and lysis of the target cells (Fig. 3e and Extended Data Fig. 5b).

To further substantiate that the observed in planta reduction of Leaf374 colonization was caused by the proteins found in the Leaf245 supernatant, we aimed to generate loss-of-function mutants. All attempts to genetically manipulate Leaf245 were unsuccessful (Supplementary Note). We therefore used chemical mutagenesis and searched for mutants with loss of in vitro inhibition. We screened more than 10,000 clones from an ethyl methanesulfonate (EMS)-treated population and identified 56 mutants with complete or partial loss of Leaf374 inhibition. We validated these mutants by comparing the activity of cell-free supernatants to the wild-type strain (Supplementary Fig. 8) and selected 33 clones that showed a maximum of 2% residual activity compared with the wild type for genome re-sequencing (Fig. 4a). Using this approach, we identified a gene locus, ASF05_00210, that was mutated in 40% of all clones analysed. The locus encodes a 3 kb annotated transcriptionsal activator and is located directly upstream of the previously identified bacteriolytic protein ASF05_00205 (Fig. 4b). ASF05_00205 was also mutated in one clone (E_60, S70L mutation) and the corresponding purified mutant protein showed attenuated inhibition of Leaf374 (about 10-fold less) compared with wild-type protein. Furthermore, the altered protein failed to lyse Leaf374 cells grown in an agar pad (Extended Data Fig. 5d), thus identifying the S70L mutation as crucial for protein function. In the remaining identified clones, including those described above, no mutations in other predicted effector proteins (Extended Data Fig. 5a) were found. We also identified two other clones (E_12 and E_48) with potential regulatory mutations...
and included these for further analyses. The analysis of supernatant fractions of these and the ASF05_00210 mutants showed that they no longer secreted the ASF05_00205 endopeptidase (Fig. 4c, see Supplementary Note for details).

Next we tested whether Leaf374 colonization in planta was no longer reduced when co-inoculated with Leaf245 EMS mutants that do not secrete the ASF05_00205 effector in vitro (E_12, E_13, E_34, E_48 and E_61; Extended Data Fig. 5d and Supplementary Fig. 8). In mono-association studies, Leaf245 mutants reached similar colonization densities as the wild type, indicating that they are not affected in plant colonization per se (Extended Data Fig. 6). Notably, Leaf374 reached 20-fold higher colonization levels in the presence of the Leaf245 ASF05_00210 mutants (E_13, E_34 and E_61) and 10-fold higher levels in combination with clone E_60 carrying the ASF05_00205 S70G mutation, compared with colonization in the presence of wild-type Leaf245 (Fig. 4d). This is in line with the observed supernatant activities in vitro and the partial loss of activity for the modified ASF05_00205 protein. In contrast, the E_12 and E_48 mutant clones did not significantly affect Leaf374 populations compared with the wild type, despite their loss of in vitro inhibition. This might be due to different regulation in planta compared with the tested in vitro conditions. Furthermore, increased Leaf374 levels did not impact colonization by Leaf245, as we did not detect
a significant difference between the wild-type and mutant strains in combination with Leaf374 (Extended Data Fig. 6).

In summary, we show that the reduction of Leaf374 colonization in planta by Leaf245 is caused by the putative M23 family endopeptidase ASF05_00205 that is regulated by ASF05_00210. We thus named the proteins Nocardioides lysis-associated peptidase (NlaP) and regulator (NlaR), respectively.

**NlaP target range and prevalence in other species.** Having established the causative effect of NlaP, we wondered whether the protein was present in other bacterial species. For this purpose, we conducted a search on the Integrated Microbial Genomes (IMG) database\(^{39}\), considering also synteny (gene’s bidirectional best hits). This search yielded three other Aeromicrobium species of the At-LSPHERE collection—Leaf289, Leaf272 and Leaf291—and for one of these (Leaf289), a significant interaction with Leaf374 was also identified in the in planta screen above (Fig. 2a). Additionally, a group of Terrabacter spp. was identified, this group being a part of the At-RSPHERE collection or related soil isolates\(^{34}\). These contained *nlaP* and *nlaR*, which are located next to each other in the genomes, but with opposite direction (Fig. 5a).

All *Terrabacter* spp. we tested showed strong inhibition of Leaf374 in vitro (Fig. 5b). Furthermore, we showed that *Aeromicrobium* spp. that lack *nlaP*, that is Root236, Root344 and Root495, did not inhibit *Nocardioides* Leaf374 in vitro (Fig. 5b), further supporting our finding of the relevance of NlaP.

Next we investigated the target range of *Aeromicrobium* spp. with endopeptidase activity by selecting a panel of Actinobacteria from both the At-L- and At-R-Sphere collections, which cover isolates from leaves and roots, and tested whether they were inhibited by living cells or cell-free supernatant of At-LSPHERE *Aeromicrobium* spp. or At-RSPHERE *Terrabacter* Root85 (Fig. 5c). Living cells consistently inhibited *Nocardioides* spp. from both the At-RSPHERE and At-LSPHERE collections and few other isolates. These inhibitions were density dependent as exemplarily shown for four strain pairs, especially if the target strain densities were high (Extended Data Fig. 7). The activity of cell-free supernatant, on the other hand, was restricted to *Nocardioides* spp. and a closely related *Marmoricola*.
sp. (Leaf446). Notably, the supernatants of Aeromicrobium Leaf245 and Terrabacter Leaf85 were more active than the other supernatants, suggesting that these strains release more protein or have proteins with higher activity. This may also explain the observed differences in interaction strength between Aeromicrobium spp. and Leaf374 in the in planta interaction screen.

Lastly, we also analysed the activity spectrum of the purified proteins NlaP, ASF05_05370 and ASF05_12180 that all inhibit Leaf374 (Fig. 3e). These were also consistently active against Nocardioides sp. and Marmoricola sp. Leaf446, demonstrating the specificity of the endopeptidase-based negative interaction for this group of bacteria and providing a basis for future investigations regarding the target specificity within this so far poorly studied bacterial group.

**Discussion**

Identifying the nature of microbial interactions that manifest in complex microbiomes remains a challenge. Large numbers of colonizing bacteria create the possibility of numerous interactions, making the discovery of genotype–phenotype relationships difficult. In contrast to correlation-based interaction mapping, synthetic communities (SynCom) provide the opportunity to establish causal interactions by adding or removing strains\(^3\), and thus to empirically determine the consequence of inter-strain interactions on the remainder of the community. It also allows for the adjustment of community complexity, which is important because of the inherent complexity of the interaction network in which effects might superpose. There are, however, also drawbacks to the SynCom approach. While complexity reduction is favourable to identify individual...
interactions, fewer species combinations can be probed at once, and specific metabolic functions associated with only few strains may be absent in the community. Therefore, the size of the SynCom has to be carefully considered.

Here we selected a focal community of 15 strains and mapped bacterial interactions observed when 200 leaf isolates were individually added during colonization of the Arabidopsis thaliana phyllosphere. We found that 90% of the observed interactions were negative (Fig. 2a). This suggests that interactions are mainly competitive in nature, which is consistent with previous studies5,37,40 and not surprising given the oligotrophic nature of the phyllosphere41. The constructed interaction map revealed high consistency of observed interactions for closely related isolates, suggesting similar functions of these within the community. This hints towards shared evolutionary history of these strains and phylogenetic trait conservation56. However, we also noted within-genus interactions, which were all negative. It has been a longstanding question in ecology whether closely related species can co-exist in the same niche due to phenotypic similarity and hence intense competition57. This hypothesis, termed competition-relatedness hypothesis58, was much debated in recent years43–46. In this study, we found that within-genus negative interactions were more frequent than expected by chance. However, some strains (Methyllobacterium Leaf88, Xylophilus Leaf220 and Pseudomonas Leaf129) also showed negative interactions with distantly related groups (Extended Data Fig. 2).

To assess whether the presence of the community is required to observe the identified interactions, we tested a subset of interactions in a binary inoculation model (Fig. 2b). Notably, we found that two thirds of the interactions were reproducible without community background, meaning that these strains affect each other independently of other microbiota members. Further studies are necessary for each specific pair to demonstrate whether this is due to direct bacteria–bacteria interaction or to indirect effects through host modulation59, given that plants respond to most of its endogenous leaf microbiota members by inducing a conserved set of defence-related genes43. Interactions between host-associated bacteria in binary and tertiary combinations have been characterized before40,50,60. However, most of the corresponding studies were conducted in vitro under conditions that are substantially different compared with the host-associated context. This is particularly pertinent when studying phyllosphere communities, since the leaf architecture provides a multitude of microenvironments that are spatially segregated and the oligotrophic nature of leaf surfaces is difficult to mimic in vitro41,43.

We also identified one higher-order interaction52. A marked reduction of Methyllobacterium Leaf88 was observed in the presence of at least two other strains that did not occur in the presence of either alone (Fig. 2c). Such higher-order interactions are proposed to be important in stabilizing species co-existence in complex communities61; however, there is still little empirical evidence to assess their importance43. It was found that the presence of multiple strains can dampen the competition previously observed in a binary situation, which was attributed to higher-order interactions62. Here we found the opposite scenario, where a higher-order interaction led to the emergence of a competitive interaction. Additional interactions observed in planta (Fig. 2a) might be the result of higher-order interactions. In planta observed interactions that could not be validated using binary strain inoculations thus provide valuable starting points to test combinatorial permutations to uncover potentially underlying higher-order interactions.

The patchy colonization observed on the leaf surface45,52,53 could also explain why higher-order interactions in communities in the phyllosphere are indeed difficult to detect at the overall population level, as more than two strains have to co-localize for direct bacteria–bacteria interactions to occur. These events are less probable in spatially segregated environments. Similarly, binary interactions could also be affected as strains do not necessarily have a chance to encounter each other at the microscale, which mitigates the observed effect.

Spatial segregation may also explain why antibiotic interactions previously found in vitro52 were rarely observed in planta (Supplementary Table 5). Other reasons for the lack of antibiotic effects in planta, apart from leaf colonization patterns, include insufficient amounts or lack of production of the compound under the conditions found in situ51. While antibiotic production observed in vitro could not explain most interactions observed in planta, we found one prominent interaction based on antimicrobial production.

Intriguingly, the interaction is mediated by a protein, an endopeptidase termed NlaP. We hypothesize that proteins might be more stable in the phyllosphere compared with small molecules, which might be subject to rapid degradation due to exposure to UV irradiation or reactive oxygen species, and hence explain the strong effect observed in planta. Moreover, proteins may directly impair the target cell. We speculate that NlaP cleaves Noxocardioides peptidoglycan cross-linkages with high specificity, since other members of this protein family, including lysostaphin from Staphylococcus simulans63–65 and LasA from Pseudomonas aeruginosa64, have glycin-glycine endopeptidase activity. As observed for other bacteriolysin enzymes66,67, the target range of NlaP is taxonomically narrow (Fig. 2c) and potentially confined to strains with similar surface structure. Therefore, the production of NlaP probably provides a strategy to defend favourable niches on the leaf surface.

In summary, this study demonstrates the power of synthetic communities to systematically identify bacterial interactions in the phyllosphere. Since the microbes involved are all culturable, the identified interactions can be further characterized to decipher mechanisms that contribute to community assembly. Here we demonstrate a role that a bacteriolysin enzyme can play in community assembly. Further mechanistic understanding of how communities assemble will ultimately be critical to building lasting plant growth-promoting and disease-suppressive microbial communities necessary for sustainable agriculture.

Methods

Strains and cultivation. Unless stated otherwise, At-SPHERE isolates were cultivated at room temperature on R-2A agar (Sigma-Aldrich) at or at 28°C in R-2A broth (0.5 g yeast extract (Oxoid), 0.5 g Proteose peptone No. 3 (Becton, Dickinson and Company), 0.5 g Casamino acids (Becton, Dickinson and Company), 0.5 g d-glucose monohydrate (Sigma-Aldrich), 0.5 g starch (from potato, Fluka), 0.3 g sodium pyruvate (Sigma-Aldrich), 0.3 g K2HPO4 (AppliChem) and 0.05 g magnesium sulphate heptahydrate (Sigma-Aldrich) dissolved in 11 deionized water), both supplemented after sterilization with 0.5% (v/v) methanol (R-2A+M). Minimal medium was prepared as described previously and supplemented with 5 mM malonate (Fluka). For Leaf374 cultivation, minimal medium agar was supplemented with vitamins (500 μg l−1 pantothenic acid hemi calcium salt, 100 μg l−1 biotin, 400 μg l−1 riboflavin, 400 μg l−1 thiamine HCl, 200 μg l−1 pyridoxal HCl, 150 μg l−1 p-amino benzoic acid, 200 μg l−1 cobalamin, 50 μg l−1 ipionic acid, 150 μg l−1 nicotinic acid and 100 μg l−1 folic acid). For selective growth of Leaf272, R-2A+M was supplemented with 50 μg ml−1 rifampicin (Sigma-Aldrich), since this strain has a natural resistance towards this antibiotic.

E. coli BL21 (DE3) gold was cultivated in LB-Lennox at 37°C, supplemented with 50 μg ml−1 kanamycin sulphate (AppliChem).

Plant growth conditions. Plants were cultivated as described previously, with some modifications. In brief, Arabidopsis thaliana Columbia-0 (Col-0) seeds were surface-sterilized and stratified for 4d at 4°C. Plants were cultivated in 6- or 12-well tissue culture plates (TechnoPlasticProducts) filled with 5 or 2 ml washed and heat-sterilized calcined clay, and 2.5 or 1 ml half-strength Murashige & Skoog medium pH 5.8 including vitamins (½ MS, Duchefa), respectively. Individual seeds were placed in the centre of each well. If seeds did not germinate, a seedling was transplanted to the corresponding well 6d after seeding. Starting at day 6, each well was supplemented twice per week with 200 or 100 μl ½ MS, respectively. Plants were incubated in a growth chamber (Percival, CU41-14) set to 22°C and 54% relative humidity, with 11 h light:13 h dark regime fitted with full spectrum lights (Philips Master TL-D 18 W/950 Graphica) and lights emitting a higher fraction of UVA and UVB (Sylvania Reptistar F18W/6500 K). Combined light intensity was...
Inoculation with bacteria. Bacteria were cultivated at 22 °C on R-2A+M agar plates for 4 d, re-streaked onto fresh R-2A+M plates and grown for another 3 d before inoculation. Inoculation suspensions were prepared by suspending individual strains in 10 mM MgCl₂ solution. For the focal community, focal community drop-out and replacement, as well as for binary and tripartite strain inoculations, the optical density at 600 nm (OD₆₀₀) of each strain was adjusted to 0.2 and the strains were mixed at equal ratios. Finally, the strain mixtures were diluted to a final OD₆₀₀ of 0.02 for inoculation. For strains added in the drop-in screen, similar volumes of cell paste were suspended in 10 mM MgCl₂ solution corresponding to an OD₆₀₀ of 0.1–0.3. Strains were diluted 20-fold when added to the premixed focal community (all strains adjusted to an OD₆₀₀ of 0.2) or 100-fold when added to each of the binary or tripartite strains. The phyllosphere was placed into tubes containing a metal bead (5 mm diameter) and Silwet-L77 in the wash solution. For harvest with the revised protocol, the leaf was cleaned twice by bead clean-up (AMPure XP, Beckman Coulter) with a ratio of 0.8:1 and finally with a ratio of 0.6:1 bead to sample ratio. The final length of the ampiclon library was checked on a 2200 TapeStation (Agilent Technologies) using high-sensitivity D1000 screen tape. Sequencing was performed on the Illumina MiSeq platform using a v3 cycle kit (2 x 300 bp, paired-end) at the Genetic Diversity Center (ETH Zurich). In all cases, the combined phyllosphere was diluted and diluted to a final concentration of 12·Pₘ₉ with addition of 20% PHEX. Sequencing was performed with custom sequencing primers as previously described10.

Phylophore harvest for 16S rRNA gene amplicon sequencing or c.f.u. enumeration. The entire phyllosphere of 28- or 29-day-old plants was harvested with sterilized tweezers and scalpels. Plants for community profiling by 16S rRNA gene amplicon sequencing were directly placed into lysis matrix E tubes (FastDNA SPIN Kit for Soil, MP Biomedicals), frozen in liquid nitrogen and stored at −80 °C. For the in planta screen (Fig. 2a), two plants grown on separate plates were pooled for one biological replicate. In all other experiments, individual plants were analysed.

For c.f.u. enumeration were harvested by two different protocols. The first protocol was performed as described previously12. The phyllosphere was placed in tubes containing 1.3 ml 100 mM phosphate buffer pH 7 containing 0.2% Silwet-L77 (Leu-Gyga). Plant fresh weight was measured on an analytical balance (Mettler-Toledo) with an accuracy of 0.1 mg. Bacteria were subsequently washed off by shaking tubes for 15 min at 25 Hz with a Tissuelyser II (Qiagen), followed by sonication (ultrasonic bath, Branson) for 5 min. Tubes containing wash-off solution were vortexed and 100 µl of the solution was transferred to a 96-well plate to prepare a 10-fold serial dilution in 100 mM phosphate buffer pH 7.

The protocol was changed for experiments including Nocardioides Leaf374, since we observed that the strain was reduced by the presence of Silwet-L77 in the wash solution. For harvest with the revised protocol, the phyllosphere was placed into tubes containing a metal bead (5 mm diameter) and 100 mM phosphate buffer pH 7 (200 µl). Plants were crushed with a Tissuelyser II for 45 s at 30 Hz. Phosphate buffer (400 µl) was added to the crushed plant material and mixed by vortexing. Samples (100 µl) were transferred to a 96-well plate to prepare a 10-fold serial dilution in 100 mM phosphate buffer pH 7.

For dilution series prepared with either of the harvest protocols, 4 µl of each dilution were spotted on R-2A+M agar square plates (Greiner) and dried under laminar flow. For samples inoculated with strain combinations or synthetic communities, 50 µl of the appropriate dilutions were plated on round (9 cm) R-2A+M agar or appropriate selective agar plates to count individual strains on the basis of morphology. Plates were incubated at room temperature and c.f.u. were counted after 1–3 d on square plates and after 4–7 d on round plates. Strains in the combinations had different colony morphology to allow identification and quantification of each strain in the mix.

Phylogenetic tree construction. The phylogeny of the At-LSPhERE strains was based on full-length 16S rRNA gene sequences as described previously11. The phylogeny of At-L- and At-R-SPHERE Actinobacteria was based on AMPHORA genes as described in Bai et al.11. The phylogenetic tree was divided with the ‘keep’ tip command of the R package ape 5.1-1. Leaf374 was placed manually in the tree with the command ‘bind tip’ in the R package phytools on the basis of full-length 16S rRNA gene tree (New England BioLabs and Exonuclease I (New England BioLabs). Barcoding-PCR (10 cycles) was performed in triplicate with DFS Tnp polymerase and plate-specific forward and well-specific reverse primers with the same cycling parameters as the first PCR. Triplicates were pooled and the identity of each sample was verified by loading equal amounts (5 µl) on a 1.5% (w/v) agarose gel. Samples were pooled on the basis of intensity of the obtained agarose gel band. The library was loaded on a 1.5% agarose gel, and the band at a size of approximately 500 bp was cut out and DNA was cleaned up with the QiAquick gel extraction kit (Qiagen) according to the manufacturer's protocol. The library was cleaned twice by bead clean-up (AMPure XP, Beckman Coulter) with a ratio of 0.8:1 and finally with a ratio of 0.6:1 bead to sample ratio. The final length of the ampiclon library was checked on a 2200 TapeStation (Agilent Technologies) using high-sensitivity D1000 screen tape. Sequencing was performed on the Illumina MiSeq platform using a v3 cycle kit (2 x 300 bp, paired-end) at the Genetic Diversity Center (ETH Zurich). In all cases, the combined phyllosphere was diluted and diluted to a final concentration of 12·Pₘ₉ with addition of 20% PHEX. Sequencing was performed with custom sequencing primers as previously described10.

Leaf245 supernatant preparation, fractionation and protein identification. Aeromicrobium Leaf245 was grown on 3% R-2A+M agar, re-streaked as a thin layer onto two square R-2A+M agar plates and grown for 24 h. The complete cell material was collected with a sterile inoculation loop and suspended in 800 µl sterile 100 mM ammonium bicarbonate adjusted to pH 7.5 with HCl. Cells were vortex-mixed for 1 min and the OD₆₀₀ of the cell suspension was measured for normalization purposes when different clones were compared. Cells were pelleted by centrifugation for 2 min at 8,000 x g and the supernatant was passed through a 0.22 µm PES filter to remove residual cells. Finally, the supernatant was concentrated to a final volume of approximately 250 µl using an Amicon Ultra-4 spin filter (3 KDa, Merck KGaA) and lyophilised.

A Superdex 75 column (GE Healthcare) fitted to an AKTA Purifier 10 fast protein liquid chromatography (FPLC) system (GE Healthcare) was equilibrated
with 100 mM ammonium bicarbonate buffer pH 7.5 before injecting 150–200 μl of concentrated supernatant. For fractionation, the flow rate was set to 0.2 ml min⁻¹ for the first 3 ml and then increased to 0.5 ml min⁻¹. Fractions (1 ml) were collected while the run was monitored by absorption measurement at 280 and 215 nm. Single or pooled active fractions were concentrated with 3 KDa cut-off Amicon Ultra-4 spin filter tubes. Proteins were separated by SDS–PAGE on precast 10% acrylamide gels (Merck) and bands of interest were cut out or whole fractions were sent directly for protein identification by shotgun proteomics at the Functional Genomics Center Zurich.

Heterologous expression of effectors in E. coli. Plasmids for expression of selected candidate proteins were ordered from Twist Bioscience. First, secretion signals of isopropyl-β-D-thiogalactopyranosid (IPTG, Biosynth) at a final concentration of 0.1 mM was used to induce expression in smaller zones of the PetE28a (+) (Novagen) vector in E. coli BL21 (DE3) Gold by electroporation and transformants were selected on LB agar supplemented with kanamycin. After overnight growth in 20 ml LB supplemented with kanamycin, cells were inoculated with cell material from several transformant colonies and grown to mid-exponential phase. Bacteria were added to 100 or 200 ml main cultures at a starting OD₅₆₂ of 0.01 and grown at 37°C. When cell cultures reached an OD₅₆₂ of 0.5, they were shifted to 16°C and expression was induced 30 min later by addition of isopropyl-β-D-thiogalactopyranosid (IPTG, Biosynth) at a final concentration of 1 mM. Bacterial cell cultures were harvested by centrifugation (2 min, 8,000 × g) and supernatant was completely removed. Cells were suspended in 1 ml 80 mM sodium phosphate buffer containing 32 mM sodium phosphate buffer containing 1 mM sodium thiosulfate and 50 μl of dilutions 10⁻¹ and 10⁻² were plated on R-2A+M agar plates to estimate survival of cells. The sample with a survival rate of ~5% (reached after 1 h EMS treatment) was chosen for further experiments.

Mutagenized cells were thawed, diluted 10⁻¹-fold into 100 mM sodium phosphate buffer pH7, and 50 μl aliquots plated on R-2A+M and incubated for 5 d. Individual clones were picked with sterile toothpicks and streaked on top of overlay plates and containing Leaf374. Clones that showed complete loss of in-frame glyA and kanamycin resistance. Aliquots of suspensions (200 μl) were mixed with the same amount of sterile 50% (v/v) glycerol and stored at –80°C. For each sample, a 10-fold dilution series was prepared in 80 mM sodium phosphate buffer containing 32 mM sodium thiosulfate and 50 μl of dilutions 10⁻¹ and 10⁻² were plated on R-2A+M agar plates to estimate survival of cells. The cell sample with a survival rate of ~5% (reached after 1 h EMS treatment) was chosen for further experiments.

Frozen cell pellets were thawed on ice. Cells were suspended in 50 μl freshly prepared 80 mM sodium phosphate buffer pH7 containing 3 mM sodium thiosulfate and 150 μl of diluted 1:200000 1 μl freshly prepared 80 mM sodium phosphate buffer pH7 containing 3 mM sodium thiosulfate and 150 μl of diluted 1:200000

Protein purification. Affinity purification was carried out on an ÄKTa Purifier 10 FPLC system (GE Healthcare). Cleared cell lysate (4–6 ml) was passed through a 0.22 μm PES filter (ThermoTechnoProducts) and loaded on a HiTrap HP column (1 ml, Cytiva) equilibrated with 20 mM TRIS-HCl pH 7.5 containing 250 mM NaCl and 40 mM imidazole. The column was washed with 10 column volumes of the same buffer and bound proteins were eluted over a gradient of 10 ml to a final concentration of 500 mM imidazole. Elution of proteins was monitored by absorption at 215 and 280 nm and 1 ml fractions were collected. The basis of the SDS–PAGE gel band pattern, fractions containing purified protein were pooled and imidazole concentration was reduced by 2-fold dilution in buffer without imidazole, followed by volume reduction with an Amicon Ultra-4 spin filter tube (3 KDa, Merck).

Protein quantification. The purified proteins were quantified by absorption measurement at 280 nm with a NanoDrop spectrophotometer (Thermo Fisher) calibrated with extinction coefficients calculated for each purified protein with the Exasy ProtParam tool.

Inhibition assay with living cells, supernatant and purified proteins. Novacordiaves Leaf374 and other target strains were grown on R-2A+M agar plates for 2–4 d. Cells were suspended in 10 mM MgCl₂ solution. To prepare overlay plates, R-2A+M agar was resuspended in PBS pH7.4 (10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl), spotted on an LB agar pad and incubated at 37 °C for 30 min. After the lysozyme pretreatment, plates were spotted (2 μl) on top of overlay plates and dried for 5 min before elution in 10 mM Tris-HCl pH 8 containing 50 mM NaCl and 40 mM imidazole. The size distribution of the libraries was checked with a 2100 Bioanalyzer (Agilent Technologies) and quantified by real-time PCR. The combined library was sequenced on a NovaSeq 6000 sequencing system (2 x 150 bp, paired-end). Reads containing adaptors, unassigned bases (N > 10%) or low quality bases (Qscore ≤ 5 for >50% of bases) were removed from the raw data by the sequencing centre. Illumina short reads were adapter-trimmed (ktrim=r, k=23, min_k=11, h=1, t=1, f=1, r=5), contaminant-filtered (k=31, hdist=1) and quality-filtered (trimq=14, maq=20, minlength=36, maxn=0, qtrim=r) using BBduk of BBtools (v.38.87, sourceforge.net/projects/bbmap). Mutations in EMS clones of Leaf245 were identified with Breset (v.0.35.4). Variations also observed in the wild-type-resequenced strain were removed from the list.

Sequencing of Leaf245 with Nanopore long reads and Illumina read polishing. Leaf245 wild-type DNA was also sequenced using ONT long-read sequencing technology to close the genome before EMS mutant identification. DNA was fragmented and rapid-barcode (RB03) with the SKT-RBB004 kit (Nanopore) and pooled with other tagged DNA samples (not included here) according to the manufacturer’s instructions. The tagged DNA pool was concentrated with SPEED beads and half of the mixture was size-selected with a 4 kb high-pass cut-off on a BluePippin (Sage Science) with a BLF-7510 cassette. The size-selected DNA was cleaned and concentrated using an equal volume of SPEED beads before elution in 100 mM Tris-HCl pH 8 containing 50 mM NaCl. Nanopore sequencing adapters were added according to the SKT-RBB004 kit instructions before sequencing on a MinION with a FLO-MIN160 flow cell. Base calling and demultiplexing were done post-sequencing using Guppy (v.3.2.4.4+dPd22F, Nanopore) and reads from two sequencing runs/flow cells were combined. Adapters were trimmed and possible chimeras removed using porechop (v.2.0.4, https://github.com/porechop/Porechop). The resulting long reads were assembled with Flye (v.2.8+b1674) into a single circular chromosome, followed by five rounds of illumina short-read polishing with Pilon (v.1.23). One circular chromosome of 3,463,989 bp was obtained and annotated using Proka (v.1.13.3).

Data availability. The 16S rRNA gene amplicon sequencing reads and the reads obtained from Leaf245 EMS mutant genome sequencing are available online at the European nucleotide archive (https://www.ebi.ac.uk/ena/) under the accession numbers...
References
1. Flemming, H. C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. Nat. Rev. Microbiol. 17, 247–260 (2019).
2. Bulgarelli, D., Schuett, K., Spiesecke, S., Ver Loren van Themaat, E. & Schatz-Lefert, P. Structure and functions of the bacterial microbiota of plants. Annu. Rev. Plant Biol. 64, 807–838 (2013).
3. Turnbaugh, P. J. et al. The human microbiome project. Nature 449, 804–810 (2007).
4. Venturelli, O. S. et al. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol. Syst. Biol. 14, e8157 (2018).
5. Foster, K. R. & Bell, T. Competition, not cooperation, dominates interactions among culturable microbial species. Curr. Biol. 22, 1845–1850 (2012).
6. Helfrich, J. W. et al. Bipartite interactions, antibiotic production and biosynthetic potential of the Arabidopsis leaf microbiome. Nat. Microbiol. 3, 909–919 (2018).
7. Coyle, K. Z. & Rakoff-Nahoum, S. Understanding competition and cooperation within the mammalian gut microbiome. Curr. Biol. 29, 538–544 (2019).
8. Turner, T. R. et al. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. ISME J. 7, 2248–2258 (2013).
9. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome interactions: from networks to models. Trends Plant Sci. 18, 607–621 (2020).
10. Müller, D. B., Vogel, C., Bai, Y. & Vorholt, J. A. The plant microbiota: new tools and advanced capabilities. Mol. Syst. Biol. 60, 440–452 (2010).
11. Lunkenberg, B. & Kamilova, F. Plant-growth-promoting Rhizobacteria. Annu. Rev. Phytopathol. 45, 440–452 (2007).
12. Billick, I. & Case, T. J. Higher-order interactions in ecological communities. Annu. Rev. Ecol. Syst. 26, 1–23 (1995).
13. Maier, B. A. et al. A general non-self response as part of plant immunity. EMBO J. 30, 3847–3858 (2011).
14. Kniskern, J. M., Traw, M. B. & Bergelson, J. Salicylic acid and jasmonic acid factors of epiphytic bacterial communities: a review. Annu. Rev. Phytopathol. 52, 475–477 (1989).
15. Schleicher, R. O., Miebach, M. & Remus-Emsermann, M. N. P. Driving factors of epiphytic bacterial communities: a review. J. Adv. Res. 19, 57–65 (2019).
16. Faust, K. & Raes, J. Microbial interactions: from networks to models. Nat. Rev. Microbiol. 10, 538–550 (2012).
17. Grosskopf, T. & Söyeren, O. S. Synthetic microbial communities. Curr. Opin. Microbiol. 18, 72–77 (2014).
18. Blair, P. M. et al. Exploration of the biosynthetic potential of the Populus microbiome. mSystems 3, e00045–00018 (2018).
19. Suda, W., Nagasaki, A. & Shishido, M. Powdery mildew-infection changes bacterial community composition in the phyllosphere. Microbes Environ. 24, 217–223 (2009).
20. Blanching, H. C., Balant-Kurti, P. J. & Stapleton, A. E. Southern leaf blight disease severity is correlated with decreased maize leaf epiphytic bacterial species richness and the phyllosphere bacterial diversity decline is enhanced by nitrogen fertilization. Front. Plant Sci. 5, 403 (2014).
21. Agler, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLoS Biol. 14, 100235 (2016).
22. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 25, 217–228 (2017).
23. Faust, K. et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8, e1002606 (2012).
24. Carr, G. A., Diner, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. ISME J. 13, 2647–2659 (2019).
25. 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).
26. Bai, Y. et al. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528, 364–369 (2015).
27. Knief, C., Frances, L. & Vorholt, J. A. Competitiveness of diverse Methylotractibacter strains in a controlled model system. Microbes Environ. 316, 1746–1748 (2017).
28. Maier, B. A. et al. Plant microbe interactions: from networks to models. Trends Plant Sci. 16, 224–233 (2011).
29. Billick, I. & Case, T. J. Higher-order interactions in ecological communities. Annu. Rev. Ecol. Syst. 26, 1–23 (1995).
30. Agler, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLoS Biol. 14, 100235 (2016).
31. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 25, 217–228 (2017).
32. Faust, K. et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8, e1002606 (2012).
33. Carr, G. A., Diner, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. ISME J. 13, 2647–2659 (2019).
34. 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).
35. Bai, Y. et al. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528, 364–369 (2015).
36. Knief, C., Frances, L. & Vorholt, J. A. Competitiveness of diverse Methylotractibacter strains in a controlled model system. Microbes Environ. 316, 1746–1748 (2017).
37. Billick, I. & Case, T. J. Higher-order interactions in ecological communities. Annu. Rev. Ecol. Syst. 26, 1–23 (1995).
38. Agler, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLoS Biol. 14, 100235 (2016).
39. Layeghifard, M., Hwang, D. M. & Guttman, D. S. Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 25, 217–228 (2017).
40. Faust, K. et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8, e1002606 (2012).
41. Carr, G. A., Diner, C., Baliga, N. S. & Gibbons, S. M. Use and abuse of correlation analyses in microbial ecology. ISME J. 13, 2647–2659 (2019).
42. 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).
43. Bai, Y. et al. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528, 364–369 (2015).
57. Sundarraman, D. et al. Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome. *mBio* **11**, e01667-20 (2020).
58. Morris, C. in *Encyclopedia for Life Sciences* (National Publishing Group, 2002).
59. Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **50**, 403–424 (2012).
60. Iversen, O. J. & Grov, A. Studies on lysostaphin — separation and characterization of 3 enzymes. *Eur. J. Biochem.* **38**, 293–300 (1973).
61. Recsei, P. A., Gruss, A. D. & Novick, R. P. Cloning, sequence, and expression of the lysostaphin gene from *Staphyloccocus aureus*. *Proc. Natl Acad. Sci. USA* **84**, 1127–1131 (1987).
62. Kessler, E., Safrin, M., Abrams, W. R., Rosenbloom, J. & Ohman, D. E. Inhibitors and specificity of *Pseudomonas aeruginosa* LasA. *J. Biol. Chem.* **272**, 9884–9889 (1997).
63. Trayer, H. R. & Buckley, C. E. Molecular properties of lysostaphin, a bacteriocylstic agent specific for *Staphylococcus aureus*. *J. Biol. Chem.* **245**, 4842–4846 (1970).
64. Heymer, B. & Schmidt, W. C. Purification and characterization of a Streptomyces albus endo-N-acetylmuramidase lytic for group A and other beta hemolytic streptococci. *Microbios* **12**, 51–66 (1975).
65. Vollmer, W., Ioris, B., Charlier, P. & Foster, S. Bacterial peptidoglycan (murein) hydrolases. *FEBS Microbiol. Rev.* **32**, 259–286 (2008).
66. Peyraud, R. et al. Demonstration of the ethylmalonyl-CoA pathway by using C-13 metabolomics. *Proc. Natl Acad. Sci. USA* **106**, 4846–4851 (2009).
67. Schlesier, B., Breton, F. & Mock, H. P. A hydroponic culture system for *Streptomyces albus* bacteriolytic agent specific for *Staphylococcus aureus*. *J. Biol. Chem.* **245**, 4842–4846 (1970).
68. Edgarp, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
69. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
70. Integrated Development Environment for R (R Studio, 2020).
71. Pruesse, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
72. Integrated Development Environment for R (R Studio, 2020).
73. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2021).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
75. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
76. Armenteros, J. J. A. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
77. Gasteiger, E. et al. in *The Proteomics Protocols Handbook* 571–607 (ed Walker, J. M.) (Humana Press, 2005).
78. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
79. Bushnell, B. BBMap short read aligner, and other bioinformatic tools (SourceForge, version 38.87); https://sourceforge.net/projects/bbmap
80. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using bresq. *Methods Mol. Biol.* **1151**, 165–188 (2014).
81. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
82. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS ONE* **9**, e112963 (2014).
83. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

**Acknowledgements**

We thank A. Minder and S. Kobel at the Genetic Diversity Center Zurich (GDC, ETH Zurich) for MiSeq amplicon sequencing service. C. Field (Institute of Microbiology, ETH Zurich) for amplicon raw read processing, the Functional Genomics Center Zurich (FGCZ) for protein identification service and P. Schudde-Lefert (Max Planck Institute for Plant Breeding Research, Cologne, Germany) for sharing Ar-BSPIER isolates. This work was supported by the Swiss National Science Foundation through NRP72 (no. 407240_167051) and as part of NCCR Microbiomes, a National Centre of Competence in Research (no. 51NF40_180575) and a European Research Council Advanced Grant (PhyMo; no. 668991).

**Author contributions**

M.S. and J.A.V. designed the research. M.S., M.B.-M. and C.M.V. conducted the in planta interaction screen and binary competition experiments. M.S. prepared the amplicon sequencing libraries and analysed amplicon sequencing data, performed biochemical analyses of Leaf245 supernatants and heterologous expression of proteins. M.M. conducted microscopy time course experiment to monitor Leaf245 lysis. M.S. prepared and M.S. and C.M.V. screened the Leaf245 EMS library. C.M.V. closed the Leaf245 genome and mapped point mutations of genome re-sequenced EMS clones. M.S. screened the inhibition spectrum of NlaP producer strains, performed statistical analyses and visualized the data. M.S. and J.A.V. wrote the manuscript with contributions from all authors.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01132-w.

**Supplementary information**

The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01132-w.

**Correspondence and requests for materials**

should be addressed to Julia A. Vorholt.

**Peer review information**

*Nature Microbiology* thanks Gwyn Beattie, Steven Lindsey, Linda Thomasnow 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 2022.
Extended Data Fig. 1 | Changes in focal community strain abundances upon strain drop-in. Heatmap showing log2-transformed fold changes (treatment vs control) for each focal community strain (top) upon strain drop-in (left). Significant changes (DESeq2-normalized counts, Wald test, Benjamini-Hochberg corrected, $p_{\text{adj}} \leq 0.01$, $n_{\text{focal community}} = 16$, $n_{\text{drop-in}} = 3$) are indicated with a black frame. Drop-in strains are colored by phylum or Proteobacteria class. Exact $p$-values are provided in Supplementary Table 2.
Extended Data Fig. 2 | Phylogenetic distance of interacting strains. Focal strains that showed ≥ 1 interaction are shown. The phylogenetic distance of the focal strain to all added strains is indicated on the x-axis. Log2-fold changes (log2FC) are shown on the y-axis. Significant fold-changes are highlighted with blue (negative interactions) or red (positive interactions) fill color. Strains that belong to the same ASV as a focal strain have a light grey fill color.
Extended Data Fig. 3 | Phyllosphere colonization of Leaf220 and Leaf408. Phyllosphere colonization of a) Leaf220 or b) Leaf408 in mono-association or in combination with other strains (indicated below). Shown are the median and individual data points of log10-transformed CFU per gram fresh weight across two independent experiments. Colors/shapes refer to experiment. Exact p-values (Kruskal-Wallis test with post-hoc Dunn test, Bonferroni adjusted p) and log2 fold-changes (log2FC) for comparisons to the mono-association condition are indicated above or below the graph, respectively. For corresponding Leaf88 colonization levels see Fig. 2c.
Extended Data Fig. 4 | Validation of Leaf374-Leaf245 interaction by sequential inoculation of *A. thaliana*. Phyllosphere colonization of a) *Nocardioides* Leaf374 and b) *Aeromicrobium* Leaf245 in mono-association or in combination with the other strain. Shown are the median and individual data points of log10-transformed CFU per gram fresh weight recovered after plant colonization (n = 12). CFU were enumerated on MM maltose agar plates and R2A + M for *Nocardioides* Leaf374 and *Aeromicrobium* Leaf245, respectively. The timepoint of inoculation with Leaf245 or Leaf374 is indicated below the graph as days (d) after planting. For mono-association controls half of the replicates were inoculated on day7 (black) or day8 (grey) and mock-treated with 10 mM MgCl₂ on the other day. Exact p-values (two-sided Wilcoxon rank sum test) and log2 fold changes (log2FC) compared to the mono-association control are indicated above or below the graph, respectively.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Characterization of selected proteins identified in the supernatant of Leaf245. a) List of proteins identified in the supernatant of Leaf245 and selected for heterologous expression in *E. coli*. For each gel band in Fig. 3d as well as the most active fractions from two independent experiments the rank abundance of the proteins based on total ion count as identified by mass spectrometry are shown. b, c, d) Leaf374 inhibition and lysis by heterologously produced candidate proteins. b) Activity of cleared cell lysate after expression of proteins (top label) in *E. coli* BL21 DE3 gold. Lysate was applied on Leaf374 overlay plates directly after preparation (left panel) or on Leaf374 overlay plates that were pre-incubated for 24 h to assess lysis (right panel). Lysate concentration was normalized to the final OD$_{600}$ that each expression culture reached. Up to 50-fold dilutions in the buffer used for lysis (see methods) were prepared as indicated on the left. c) Activity of purified proteins against 24 h pre-grown Leaf374. d) Activity of ASF05_00205 native and mutant (S70L) protein against Leaf374. Concentrations of both purified proteins were normalized (0.1 mg mL$^{-1}$) and a 10- and 100-fold dilution was prepared and assayed on a Leaf374 overlay plate directly after preparation (left panel) or after the overlay was pre-incubated for 24 h (right panel) to test for growth inhibition and cell lysis, respectively. Data shown for ASF05_00205 native protein is the same as shown in Fig. 3e and panel c) of this figure.
Extended Data Fig. 6 | Aeromicrobium Leaf245 wild type and EMS mutant colonization level. Phyllosphere colonization of Leaf245 in mono-association or in combination with Leaf374. Shown are the median and individual data points of log10-transformed CFU per g fresh weight across 2–3 independent experiments. Log2 fold changes and exact p-values (two-sided Wilcoxon rank sum test) between Leaf245 mono-association and co-colonization treatments with Leaf374 are shown above the graph. For the corresponding Leaf374 colonization see Fig. 4d.
Extended Data Fig. 7 | *In vitro* inhibition assay with different producer and target cell densities. The concentrations of the NlaP producer strains (Leaf245 and Root85) spotted on top (2 µL) and target strains (Leaf374 and Root122) within the agar are indicated on top and on the left of the graph, respectively. Pictures were taken 36 hours after overlay preparation and spotting. The standard concentrations used for all other assays performed in this study are highlighted in red.
Reporting Summary

Nature Portfolio 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 Portfolio 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.

n/a  Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

*Give P values as exact values whenever suitable.*

☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

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

Software and code

Policy information about availability of computer code

Data collection

Data was collected as described in the manuscript. Amplicon sequencing data was collected using the Illumina MiSeq platform. Genome sequencing short read data was collected on a Illumina NovaSeq instrument and long read data with a Oxford Nanopore MinION.

Data analysis

Data was analysed and plotted in RStudio 1 with R environment v.4.0.4 or GraphPad Prism 9. The following R packages were used for data analysis: phyloR 1.0.1 (https://github.com/MicrobiologyETHZ/phyloR/), DESeq2 v.1.14.1, vegan 2.5-7, ape 5.4-1, phytools 0.7-70. For amplicon sequencing raw read analysis the functions from USEARCH v.11.0.667–86 linux64 pipeline were used. Illumina short reads and Nanopore long reads were assembled using the following data analysis packages: BBtools v.38.87, porechop v.0.2.4, Flye v.2.8.1b1674, Guppy v.3.2.4, Pilon v.1.23 and PROKKA v.1.13.3. Mutations were mapped on the reference genome of Leaf245 using Baseseq v.0.35.4. Microscopy pictures were assembled to timelapse video with Fiji 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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Demultiplexed 16S rRNA gene amplicon reads and reads obtained from Leaf245 EMS mutant re-sequencing were deposited in the European nucleotide archive under
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-list.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 determine sample size. Sample size was chosen based on previous experience (Carlström et al., 2019, Pfeilmeier et al., 2021) and in case of the screen to ensure feasibility of the experiment. |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions   | We excluded data as described in the manuscript. One focal community replicate was excluded because the community composition was completely changed compared to all other focal community samples i.e. several focal strains were absent and community was dominated by two strains. We also excluded two treatments. Leaf72 was excluded because it was detected neither in planta nor in the inoculum. Leaf69 was excluded because two replicates were contaminated with another At-LSPHERE strain. Exclusion criteria were not pre-established. |
| Replication       | Due to the the large number of treatments the screen was only performed once. The binary in planta experiments were performed twice and data for both replicates is shown. The activity of candidate proteins expressed in E. coli was tested twice in independent experiments with similar results. Identification of proteins in the supernatant of Leaf245 wt and EMS mutant strains was performed twice (protein identification by mass spectrometry only once) including two replicates of the wild type and two independent EMS clones with similar results (one replicate is shown). The inhibition spectrum of Aeromonas strains (cells and supernatant) was performed twice (one replicate shown) with similar results. In general all attempts at replication made were successful. |
| Randomization     | For the in planta screen, three plants of the same four treatments were present on two 12-well plates and the position of the treatments was changed between plates. For all other plant experiments the twelve replicates were distributed randomly over six 6-well plates with the restriction that each treatment was present at each position of the plate and the same treatment was never found on the same plate. Furthermore for all plant experiments the plates were distributed and shuffeled randomly in the plant growth chamber. For DNA extraction and amplicon sequencing library preparation the samples were randomized. |
| Blinding          | Each treatment was assigned a number prior to inoculation and used throughout the experiment until data was collected. For CFU counting treatments containing strain combinations were known to ensure plating of the appropriate dilutions or the correct selective media. |

**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 |
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 Steroceros tharberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. 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 cells

#### 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 (See ICLAC register) | 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). Permits should encompass collection and, where applicable, export. |
|---------------------|-------------------------------------------------------------------------------------------------|
| 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. |

- Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.
- 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
May remain private before publication.
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
Sequencing depth

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 each 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 (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. |