Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome

Carrión, V. J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., De Hollander, M., Ruiz-Buck, D., ... Raaijmakers, J. M.

Title: Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome

Authors: Víctor J. Carrión[1,6], Juan Perez-Jaramillo[1]†, Viviane Cordovez[1]†, Vittorio Tracanna[4]†, Mattias de Hollander[1], Daniel Ruiz-Buck[1], Lucas W. Mendes[2], Wilfred F.J. van Ijcken[5], Ruth Gomez-Exposito[1], Somayah S. Elsayed[6], Prarthana Mohanraju[7], Adini Arifah[7], John van der Oost[7], Joseph N. Paulson[8], Rodrigo Mendes[3], Gilles P. van Wezel[1,6], Marnix H. Medema[4*] and Jos M. Raaijmakers[1,6*]

Affiliations:

[1]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands.

[2]Cell and Molecular Biology Laboratory, Center for Nuclear Energy in Agriculture (CENA), University of Sao Paulo (USP), Piracicaba, Brazil.

[3]Laboratory of Environmental Microbiology, Brazilian Agricultural Research Corporation, Embrapa Environment, Rodovia SP 340 - km 127.5, 13820-000 Jaguariúna, Brazil.

[4]Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands.

[5]Erasmus MC, University Medical Center Rotterdam, Department of Cell Biology, Center for Biomics, The Netherlands.

[6]Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE, Leiden, The Netherlands.

[7]Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

[8]Department of Biostatistics, Product Development, Genentech Inc., South San Francisco, CA 94080, USA.

*Correspondence to: Jos M. Raaijmakers, Netherlands Institute of Ecology (NIOO-KNAW), Department of Microbial Ecology, Droevendaalsesteeg 10, 6708 PB, Wageningen, The Netherlands. Phone: +31 (0)317 473 497. Mail: j.raaijmakers@nioo.knaw.nl; Marnix H. Medema, Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708PB, Wageningen; Phone: +31 (0)317 484 706. Mail: marnix.medema@wur.nl

† These authors contributed equally to this work.

Abstract: Microorganisms living inside plants can promote plant growth and health, but their genomic and functional diversity remain largely elusive. Here, metagenomics and network inference showed that fungal infection of plant roots enriched for *Chitinophagaceae* and *Flavobacteriaceae* in the root endosphere and boosted expression of chitinase genes and various unknown biosynthetic gene clusters encoding the production of nonribosomal peptides (NRPS) and polyketides (PKS). Following strain-level genome reconstruction, a consortium of *Chitinophaga* and *Flavobacterium* was designed that consistently suppressed fungal root disease. Site-directed mutagenesis then revealed that a novel NRPS-PKS gene cluster from *Flavobacterium* was essential for disease suppression by the endophytic consortium. Our results highlight that endophytic root microbiomes harbor a wealth of yet unknown functional traits that, in concert, can protect the plant inside out.

One Sentence Summary: metagenomics-guided microbiome reconstruction

Main Text:

Past and present plant microbiome studies have generated a large amount of sequence data and a wealth of (mostly) descriptive information on the diversity and relative abundance of different taxonomic groups in the rhizosphere, phyllosphere, spermosphere and endosphere of a multitude of plant species (*1, 2*). To date, however, relatively few studies have demonstrated the functional importance of microbiomes for specific plant phenotypes, i.e., plant growth, development and health (*3-9*). Furthermore, the molecular and chemical basis of the causal relationships between these plant phenotypes and microbiome structure and functions are in most cases still unknown. The aim of this study was to investigate the genomic diversity and functional potential of the endophytic root microbiome in protection of plants against fungal infections. To this end, we integrated multiple approaches including network inference and metagenomics to identify root endophytic bacterial consortia and functional gene clusters associated with a soil that is suppressive to disease caused by *Rhizoctonia solani*, a fungal root pathogen of several plant species, including rice, wheat and sugar beet.

Disease-suppressive soils are exceptional ecosystems in which plants are protected from root pathogens as a result of antagonistic activities of the root-associated microbiome. Suppressive soils have been described for various soil-borne pathogens, including fungi, bacteria, oomycetes and nematodes (*3, 5, 10-15*). Disease suppression can be eliminated by selective heat treatment and can be transplanted to non-suppressive (conducive) soils, analogous to fecal transplants in humans (*5, 16*). Specific suppression of soils to fungal root pathogens, such as *R. solani*, is induced in field soils by a disease outbreak during continuous cultivation of a susceptible host plant (*17*). Once established, the suppression can dissipate if non-host plants are grown but is regained in the presence of the host plant and the specific fungal

pathogen. Therefore, the three-way interactions between the fungal pathogen, the host plant and its root microbiome are key elements of the onset and persistence of specific disease suppression. We previously showed that in a soil suppressive to the fungal root pathogen *R. solani,* several bacterial genera inhabiting the rhizosphere of sugar beet, in particular *Paraburkholderia*, *Pseudomonas* and *Streptomyces* (*5, 18, 19*), act as a first line of defense. To understand what role microorganisms that live within plant root tissues (endophytes) play in disease suppression, we conducted a metagenomic analysis of the endosphere of sugar beet seedlings grown in field soil suppressive to *R. solani* and identified the microorganisms associated with disease suppression, characterized biosynthetic gene clusters (BGCs) that were upregulated during infection, reconstructed synthetic endosphere consortia and finally made site-directed mutations to test the role of specific BGCs in disease suppression.

Taxonomic diversity and network inference of the endophytic microbiome

Sugar beet plants were grown in disease-conducive (C) and disease-suppressive (S) soils inoculated (or not) with the root pathogen *R. solani* (Fig. S1). Disease incidence in the pathogen-inoculated suppressive soil (S+R) was 15-30%, whereas disease incidence in the pathogen-inoculated conducive soil (C+R) exceeded 80% (Fig. S1A), typical of our previous studies (*5, 16*). Given the high disease incidence in C+R, there was not enough root material left for in-depth microbiome analysis of this condition. The taxonomic diversity and functional potential of the root endophytic microbiome of plants grown in the remaining three soil conditions (C, S, S+R) was investigated after 4 weeks of plant growth. Following metagenome sequencing and bioinformatic analyses (Fig. S2, table S1 and S2), taxonomic assignment of the microbial cell fraction from the sugar beet endosphere showed that 76.1%, 10.5% and 0.0065% of the sequence reads corresponded to the domains Bacteria, Eukaryotes and Archaea, respectively (Fig. S3, A and B). For the Eukaryotic reads, Constrained Analysis

of Principal Coordinates (CAP) showed significant differences (PERMANOVA, $P < 0.05$) between the endophytic fungal community composition in C, S and S+R (Fig. S4A). This was largely due to a significant increase in *Rhizoctonia*-related sequence reads in the suppressive soil inoculated with *R. solani* (S+R) (Fig. S4, B and C). Most of the other sequence reads could not be reliably assigned to specific fungal phyla. Collectively, these results indicate that after inoculation into the disease-suppressive soil, *R. solani* colonized and penetrated the plant roots but caused little disease.

16S rRNA data from the metagenome sequences (Fig. S2) showed that Proteobacteria and Bacteroidetes dominated the endophytic bacterial community with ten OTUs spanning *Pseudomonadaceae* (2), *Xanthomonadaceae* (4), *Chitinophagaceae* (1), *Flavobacteriaceae* (2) and *Veillonellaceae* (1) (Fig. S5), all of which became enriched in the S+R condition compared with the S condition (Fig. 1A). Co-occurrence network analysis revealed increased complexity in the S+R condition (Table S3, Fig. S6, A, B and C) compared with C and S conditions (Table S3). Highly connected networks, like those in the S+R samples, can occur when microbiota face environmental perturbation, such as pathogen invasion (*20*). Interestingly, 80% of the interacting nodes in the S+R network belonged to *Chitinophaga*, *Flavobacterium* and *Pseudomonas* spp. (Table S4). When sequence reads from the Bacteroidetes were removed from the datasets, the endophytic signal from the C and S soils were indistinguishable (Fig. S7, A and B), once again indicating an association of the Bacteroidetes genera *Chitinophaga* and *Flavobacterium* with the disease-suppressive phenotype.

Functional diversity of the endophytic microbiome

Fifty to seventy percent of the genes retrieved from the metagenome data were assigned to a known function (Fig. S3, C, D and E). For the other genes, grouping annotations indicated 56,175 taxa-associated functions of which 402 functions were significantly enriched in the

endophytic bacterial community of plants grown in the S soil compared with plants in the C soil (FDR < 0.1; Fig. S8, B and C). In the S+R condition, this proportion of functional enrichment increased over ten-fold (4,443) (FDR < 0.1; Fig. 1B). These genes belonged mainly to pathways classified as 'carbohydrate transport and metabolism' and 'signal transduction mechanisms'. Several endophytic bacterial families, including *Chitinophagaceae* and *Flavobacteriaceae* (Bacteroidetes), *Pseudomonadaceae* and *Xanthomonadaceae* (Gammaproteobacteria), *Hyphomicrobiaceae* and *Rhizobiaceae* (Alphaproteobacteria), and *Burkholderiaceae* (Betaproteobacteria) were specifically associated with the functional enrichment we observed (Fig. 1B, C and Fig. S9A). The majority of the overrepresented genes in S+R (3,138 genes of 4,443) were associated with *Chitinophagaceae* and *Flavobacteriaceae* (Fig. 1B, Fig. S9A). When we used a more stringent significance level of $P<0.05$, 2,063 of 56,175 taxa-associated functions were overrepresented with 461 functions associated mainly with *Chitinophagaceae* and *Flavobacteriaceae*. Cumulative differential abundance analyses of all Bacteroidetes' genes between samples highlighted that genes from COG category Q (secondary metabolites biosynthesis, transport, and catabolism) were among the most differentially abundant between S+R and S, while genes from category G (carbohydrate transport and metabolism) were among the most differentially abundant between S and C (Fig. 1C).

For more detailed resolution of the specific functions associated with COGs G and Q, we searched for carbohydrate-active enzymes and secondary metabolite biosynthetic gene clusters within the metagenome sequences using dbCAN (*21, 22*) and antiSMASH (*23*), respectively. Using dbCAN we were able to annotate 1,822 genes in the endophytic metagenome with glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide lyase (PL), and carbohydrate esterase (CE) domains as well as non-catalytic carbohydrate-binding modules (CBMs). Because many of these domains are evolutionary related and have related

functionalities, we mapped the domain diversity in a protein family similarity network using the hhsearch algorithm (*24*). Glycoside hydrolases and glycosyltransferases were more abundant in the S+R endophytic microbiome and correlated with disease suppression (Fig. 2A, Fig. S9, B and C). Three endophyte families (*Chitinophagaceae, Burkholderiaceae, Xanthomonadaceae*) showed statistically significant differences in CAzyme composition between S+R and S (FDR < 0.1, Fig. 2A, Fig. S9, A and B). Further, we found that *Chitinophagaceae* harbored several enzymes with domains associated with fungal cell wall degradation, such as chitinases, beta-glucanases, endoglucanases (Fig. 2A), and also possessed de-branching enzymes, including α-1,6-mannanase and α-L-rhamnosidase. *Burkholderiaceae* and *Xanthomonadaceae* families (Fig. S9B, C) also contributed two chitinase domains and three other enzymes involved in chitin degradation, including chitin deacetylase and chitosanase. Only five domains were shared between *Chitinophagaceae, Burkholderiaceae* and *Xanthomonadaceae* (Fig. 2B), indicating limited functional redundancy among these endophytes for this trait. The enrichment of genes encoding chitin-degrading enzymes points to a role in disease suppression for these endophytes (*25*).

Bacterial genomes contain a large diversity of biosynthetic gene clusters (BGCs), the vast majority of which have not yet been linked to specific molecules or functions (*5, 26-28*). Our antiSMASH analysis for secondary metabolites revealed a total of 730 BGCs associated with the biosynthesis of nonribosomal peptides, polyketides, terpenes, aryl polyenes, ribosomally synthesized and post-translationally modified peptides (RiPPs), phosphonates, phenazines, and siderophores (Fig. 3A, figs. S10, 11, 12). Of these 730 BGCs, only 12 BGCs have previously been described and the chemical structure of their products elucidated (Table S5, Fig. S11). Among these were the BGCs for thanamycin and brabantamide, which are two NRPS-derived products previously detected in the rhizosphere microbiome of plants grown in *Rhizoctonia*-suppressive soil (*5, 26, 29*). For the other 718 BGCs, no near or exact matches

were found for their genetic architecture and predicted products in the MIBiG repository (*27*). Of the BGCs detected, several proteobacterial RiPPs and NRPSs were noted (Fig. 3C), as well as NRPS and aryl polyene clusters originating from Bacteroidetes (mainly *Flavobacterium* and *Chitinophaga* [Fig. 3D]) and a larger proportion of NRPS clusters from a group of unclassified phyla (Fig. 3E). Altogether, 117 BGCs were significantly overrepresented (two-tailed Welch's T-test, $P < 0.1$) in the endosphere under the S+R conditions with 34 BGCs belonging to Bacteroidetes (Fig. 3A-F, Fig. S10, S11and S12). Notably, these did not include the thanamycin and brabantamide BGCs identified previously for the disease-suppressive *Pseudomonas* spp. from the rhizosphere (*5, 29*). For the Bacteroidetes species, 10 NRPS gene clusters out of the 117 were overrepresented under S+R conditions and none of these had a match in antiSMASH with gene clusters from MIBiG.

De novo assembly of endophytic bacterial genomes

From the 730 BGCs identified in the metagenome by antiSMASH, 157 were found in a set of 25 metagenome-assembled genomes (MAGs) we reconstructed (Table S6, Fig. S13 and S14). The MAGs, housekeeping genes and identified BGCs were subsequently used to generate specific primer sets for transcriptome analyses and to associate the BGCs to isolates in the bacterial endophyte collection.

The initial collection of 935 bacterial endophyte isolates (Fig. S1) were taxonomically characterized by 16S rRNA sequencing (Fig. S15, A and B, Table S7), revealing eight different genera, mostly represented by Bacteroidetes and Gammaproteobacteria. Although no BGCs associated with *Chitinophaga* or *Pseudomonas* spp. (Table S8) were detected, four BCGs (298, 396, 471 and 592) were found in the endophytic *Flavobacterium* isolates obtained from the S+R condition. Three of these encoded an NRPS (BCGs 396, 471, 592) and the fourth a hybrid NRPS-PKS gene cluster (BGC298, Fig. 4A). A similar approach confirmed the presence of glycosyl hydrolase (GH18) genes in three endophytic

Chitinophaga isolates obtained from the S+R condition (Fig. 2A). Subsequent *in vitro* assays with the bacterial isolate collection showed that the three *Chitinophaga* isolates also had extracellular chitinolytic activity.

Subsequent genome sequencing of the three *Chitinophaga* and four *Flavobacterium* isolates showed >99% similarity among the isolates within each genus (table S9, Fig. S15C and S16A). The isolate genomes also clustered with MAGs assigned to each of these genera (Fig. S15B and C), confirming that they correspond to taxa abundant in the microbiome. For the key BGCs, no signs of metagenome mis-assemblies were identified based on comparisons with the complete genome sequences of the Bacteroidetes isolates (Fig. S16B and C, and S17A, B and C).

Reconstruction and functional analysis of disease-suppressive consortia

We selected the seven sequenced Bacteroidetes isolates for root colonization assays and BGC-transcript analysis. All isolates colonized the rhizosphere and the root endosphere of sugar beet seedlings (Fig. S18 and S19). Transcriptional analysis showed that chitinase expression was significantly ($P < 0.05$) higher in the consortium colonizing the rhizosphere and endosphere compartments inoculated with the fungal pathogen (Fig. 4B and C, Fig. S20). Of the four *Flavobacterium* gene clusters, BGC298 was expressed at significantly ($P < 0.05$) higher levels in the endosphere than in the rhizosphere when the plant roots were challenged with the fungal pathogen *R. solani* (Fig. 4C). This BGC was consistently assembled in all four *Flavobacterium* genomes and in a MAG (Fig. S16B) and showed no match with known BGCs in MIBiG (Fig. S17).

The central place of *Flavobacterium* and *Chitinophaga* in the functional network of plants grown in the disease suppressive soil, their ability to colonize the endosphere and the fact that expression of BGC298 and chitinase genes in the synthetic consortium are induced by the fungal pathogen indicate a role in *R. solani*-disease suppression. To test this

hypothesis, three independent bioassays showed that the consortium of *Chitinophaga* and *Flavobacterium* conferred more significant and more consistent protection against fungal root infection than the individual consortium members (Fig. 4D, E and F, and Fig. S21A, B and C). Even when single isolates showed little benefit against disease, consortia always showed a greater degree of protection (Fig. 4D-F, Fig. S21A-C). The apparent 'minimal' consortium to reconstitute the plant phenotype consisted of one *Chitinophaga* isolate and one *Flavobacterium* isolate (syncom-2) as this consortium showed the same level of disease control observed for the seven-member consortium (syncom-7; Fig. 4F).

To confirm the role of the *Flavobacterium* BGC298 in the disease-suppressive activity, we developed a SpyCas9-mediated system for introduction of double-stranded DNA breaks in *Flavobacterium* sp. 98. We obtained two independent BGC298 mutants (table S10, 11 and 12, and Fig. S22A, B, C and D), for which the PKS gene deletion was verified by Sanger sequencing with specific primers (Fig. S22D). The two mutants colonized the rhizosphere and endosphere to the same extent as wild type *Flavobacterium* sp. 98 when introduced alone or with *Chitinophaga* sp. 94 (Table S13). When the two independent BGC298 mutants were tested in the disease bioassay, the mutation reduced disease-suppressive activity of Fl98 alone and when paired with the *Chitinophaga* isolate (Fig. 4F).

Conclusions

In our previous studies on soils suppressive to fungal root diseases, we have shown that rhizosphere bacteria act as first line of defense (*5-7, 10*). If the pathogen breaks through this first line of defense, it will encounter the basal and induced defense mechanisms of the plant (*30*). Here, we show that in this second stage of pathogen invasion of the plant roots, the endophytic microbiome can provide an additional layer of protection. Our experiments showed that on pathogen invasion, members of the *Chitinophagaceae* and *Flavobacteriaceae*

became enriched within the plant endosphere and showed enhanced enzymatic activities associated with fungal cell wall degradation, as well as secondary metabolite biosynthesis encoded by NRPSs and PKSs. Following *de novo* assembly of 25 bacterial genomes from metagenome sequences we were able to reconstruct a synthetic community (syncom) of *Flavobacterium* and *Chitinophaga* that provided disease protection. Site-directed mutagenesis further confirmed the contribution of BGC298 in *Flavobacterium* to this phenotype. Where these two bacterial genera are localized inside the root tissue and how they interact at the molecular level in the endosphere is not yet known. Possibly, chitinase-generated chitooligosaccharides induce expression of the *Flavobacterium* BGC298. Whether BCG298 encodes a metabolite that exerts direct antifungal activity or acts as a regulator of other protective traits is not yet known. Another consideration is that the consortium may have indirect effects via induction of local or systemic resistance in the roots. The results of this study highlight the wealth of yet unknown microbial genera and functional traits in the endophytic root microbiome. Adopting metagenome-guided analyses and network inference was successful in pinpointing taxa and functions for targeted design of microbial consortia to attain a specific microbiome-associated plant phenotype.

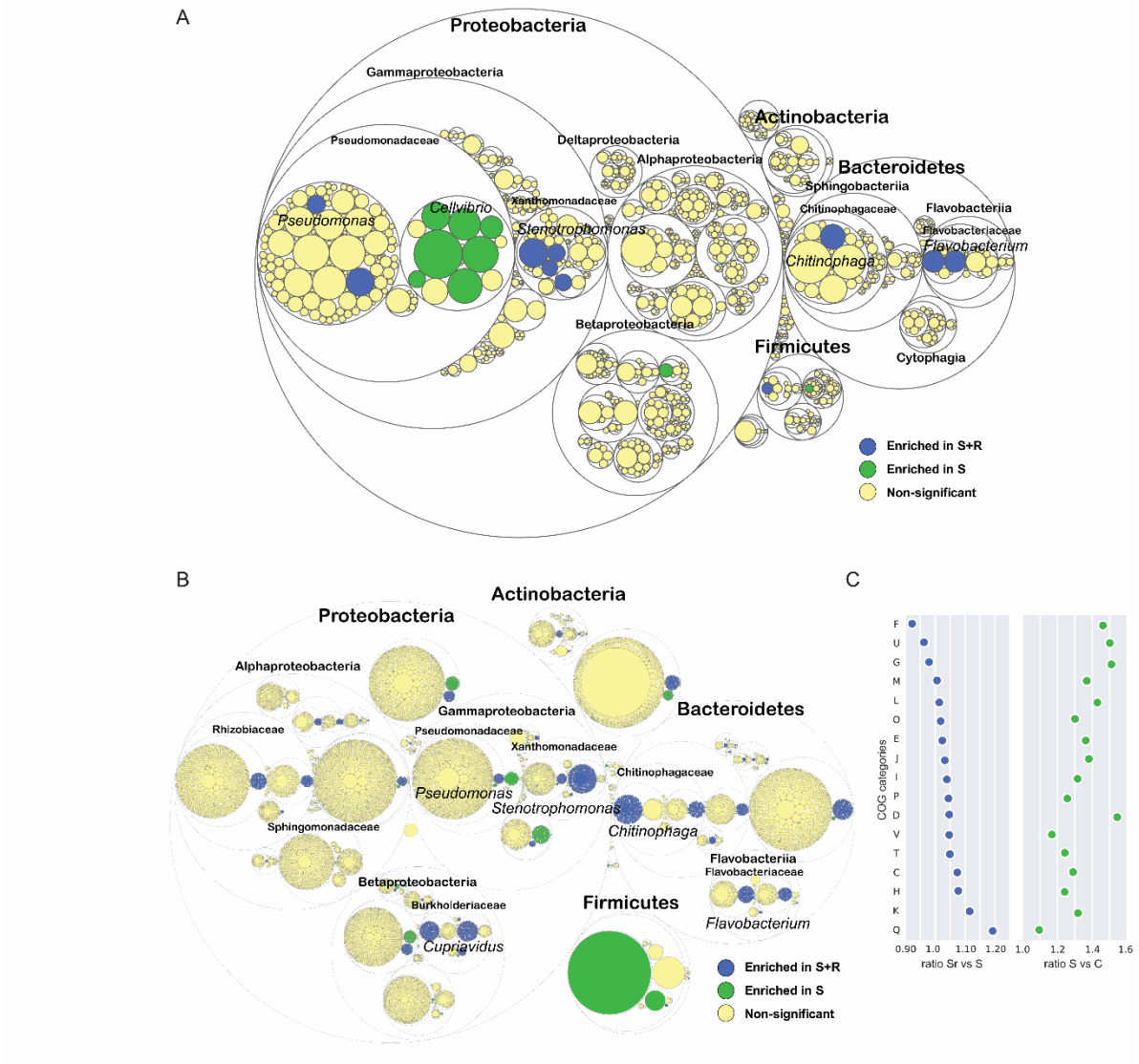

Fig. 1. Pathogen-induced changes in endophytic microbiome diversity and functions. Differential abundance of endophytic bacterial communities from plants grown in suppressive soil (S) or in suppressive soil inoculated with the fungal root pathogen *R. solani* (S+R). (**A**) Taxonomic differences are based on 16S rRNA sequences extracted from the metagenome. The largest circles represent Phylum level, the inner circles represent Class, Family and Genus. (**B**) Functional differences are based on the metagenome sequence data and assigned to taxonomic groups. The smallest circles represent the COG categories groups. The circle sizes represent the mean read relative abundance of the differentially abundant taxa and functions. Bacterial taxa or functions that are significantly enriched (FDR<0.1) in

the comparison between S and S+R are indicated in green for S and in blue for S+R; non-significant taxa and functions are indicated in yellow. (**C**) Strip plot depicting the average abundance ratios of all genes from Bacteroidetes belonging to core COG functional categories that contain significantly enriched genes in S+R compared with S, and in S compared with C. Categories are sorted by S+R/S and S/C ratios. Each COG type has been abbreviated as follows: C: energy production and conversion, D: cell cycle control, cell division, and chromosome partitioning, E: amino acid transport and metabolism, F: nucleotide transport and metabolism, G: carbohydrate transport and metabolism, H: coenzyme transport and metabolism, I: lipid transport and metabolism, J: translation, ribosomal structure, and biogenesis, K: Transcription, L: replication, recombination, and repair, M: cell wall/membrane/envelope biogenesis, O: post-translational modification, protein turnover, and chaperones, P: inorganic ion transport and metabolism, Q: secondary metabolites biosynthesis, transport, and catabolism, T: signal transduction mechanisms, U: intracellular trafficking, secretion, and vesicular transport, V: defense mechanisms, W: extracellular structures.

Fig. 2

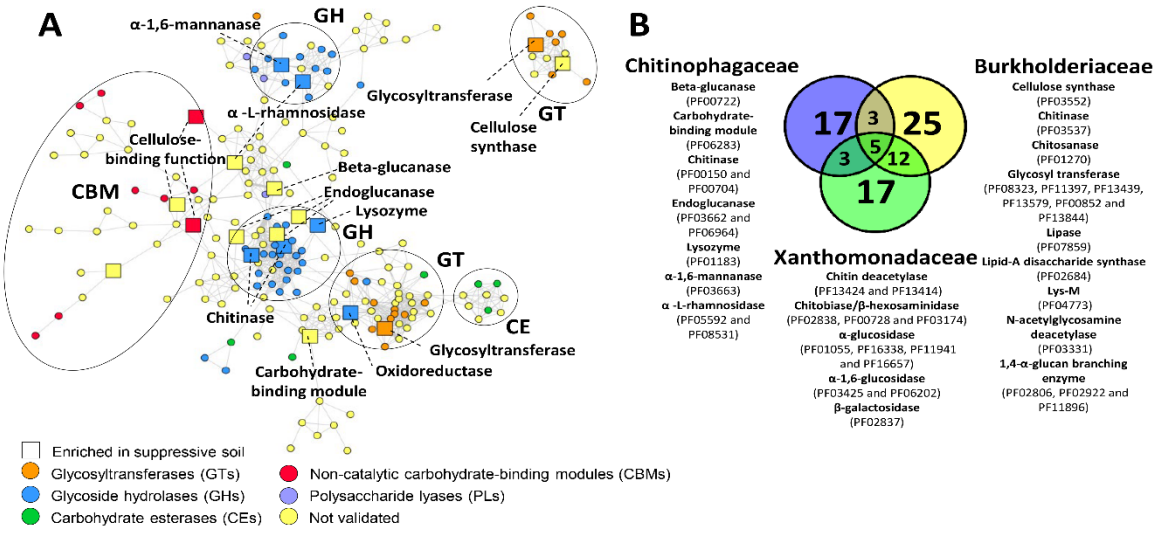

Fig. 2. Diversity and distribution of carbohydrate-active enzymes in the endophytic microbiome (**A**) similarity network of known and putative HMM domains of enzymes involved in carbohydrate metabolism (CAzymes). From the endophytic metagenome of plants grown in suppressive soil (S) or in suppressive soil inoculated with the fungal root pathogen *R. solani* (S+R), a total of 1,822 genes were annotated as CAzymes. Domain-domain distances and their relatedness are shown in the network. Nodes were grouped into five functional classes: glycoside hydrolases (GH, blue), glycosyltransferases (GTs, orange), polysaccharide lyases (PL, purple), carbohydrate esterases (CE, green) and the non-catalytic carbohydrate-binding modules (CBM, red). Unknown domains or domains for which the function has not been experimentally validated are shown in yellow. Squared nodes represent enzymes that are significantly overrepresented (FDR < 0.1) in S+R compared with S, and taxonomically assigned to the *Chitinophagaceae*. Enzymes significantly overrepresented in S+R and taxonomically classified as *Burkholderiaceae* and *Xanthomonadaceae* are shown in Fig. S9B and C, respectively. (**B**) Venn diagram with different CAzymes annotated for three endophytic bacterial families enriched in S+R, i.e. *Burkholderiaceae*, *Chitinophagaceae* and *Xanthomonadaceae*. For each of the CAzymes, the Pfam number is shown between brackets.

The Venn diagram shows the number of domains detected exclusively for each bacterial family and the domains shared by these endophytic bacterial families.

Fig. 3

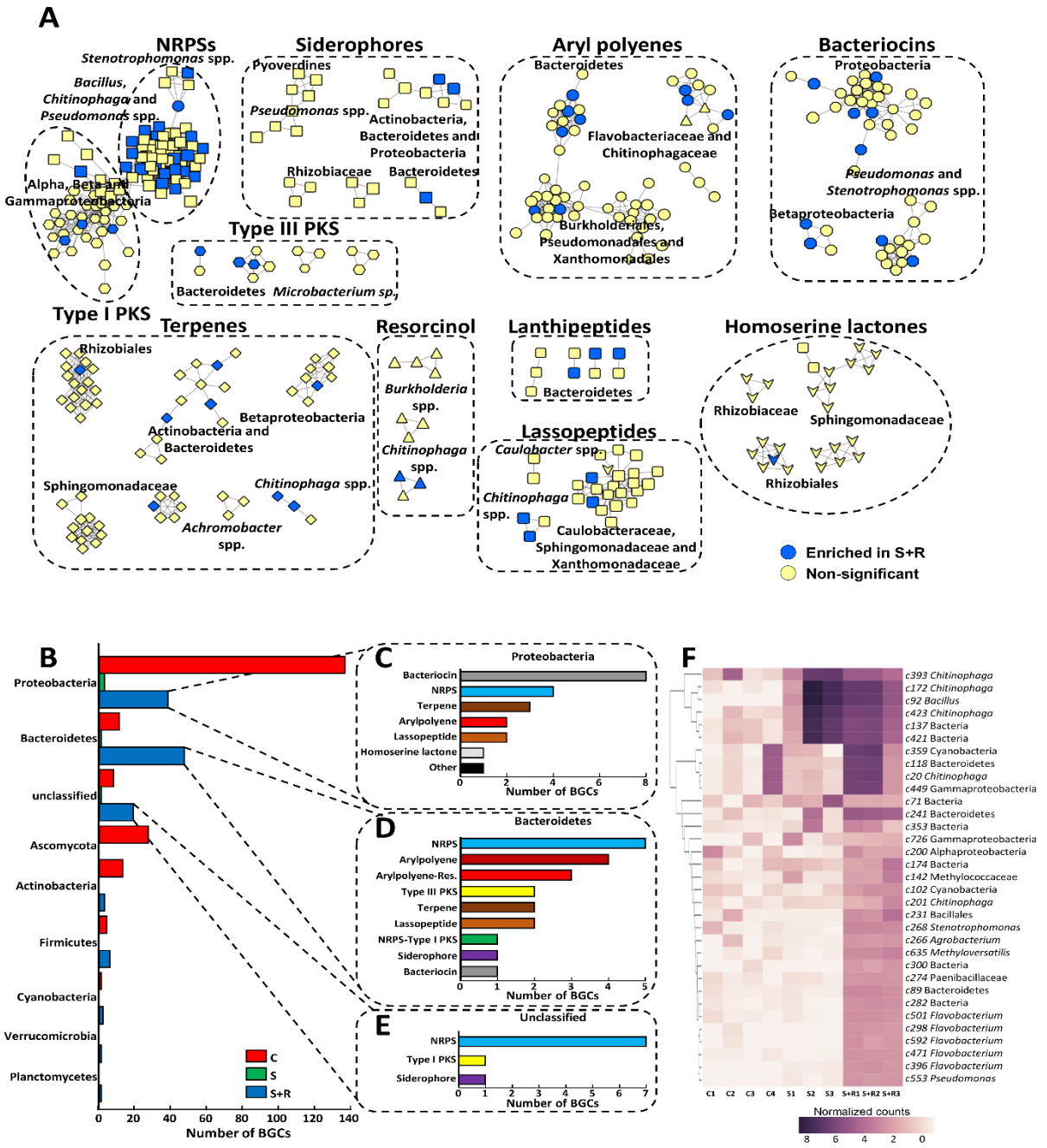

Fig. 3. Diversity and distribution of biosynthetic gene clusters (BGCs) in the endophytic microbiome. (**A**) Sequence similarity network (constructed with BiG-SCAPE [69], threshold: 0.8) of the different classes of BGCs detected in the endophytic microbiome. Taxonomic assignment and BGC class annotation of the nodes are shown. Nodes with less

than three connections were removed - original network with all nodes can be found in Fig. S10. Node colors represent statistical significance based on a Welch test (FDR < 0.1): yellow nodes are non-significant and blue nodes are significantly overrepresented in S+R condition. (**B**) Number of overrepresented BGCs (two-tailed Welch's T-test, P <0.1) detected by the antiSMASH and Clusterfinder algorithms for the different bacterial phyla in the endophytic root microbiome of plants grown in conducive (C), suppressive (S) and suppressive soils challenged with the fungal pathogen *R. solani* (S+R). (**C**, **D** and **E**) Number and type of BGCs assigned to Proteobacteria, Bacteroidetes and unclassified bacterial phyla, respectively, that were significantly (two-tailed Welch's T-test, P <0.1) more enriched in S+R. (**F**) Clustered heat map of relative abundances (CSS-normalized RPKM values) of the 33 NRPS gene clusters that were significantly overrepresented in the different replicate samples of S or S+R versus C. The NRPS cluster number and the corresponding taxonomic assignment are shown on the right side of the panel.

Fig. 4

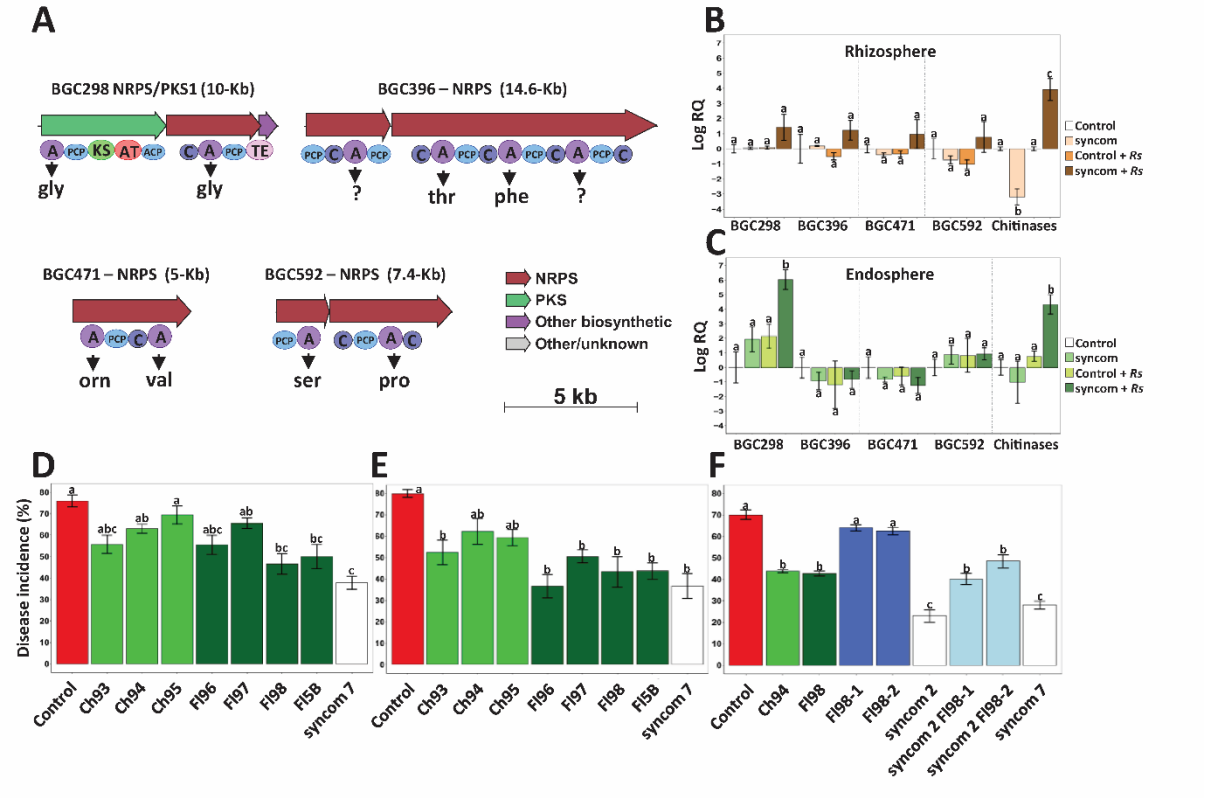

Fig. 4. Transcriptional and functional analyses of disease-suppressive consortia. (**A**) Genetic organization of BGCs 298, 396, 471 and 592 identified in both the *Flavobacterium* MAG nbed44b64, and in the genome sequences of the four endophytic *Flavobacterium* isolates. Shown below the nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) genes are the module and domain organizations of the encoded proteins. The domains are labeled as: C, condensation; A, adenylation; KS, ketosynthase; AT, acyltransferase; and TE, thioesterase. Predicted substrates of the NRPS and PKS modules in BGC298 are glycine, malonyl-CoA and again glycine. (**B** and **C**) qPCR-based analysis of the expression of BGC298, BGC396, BGC471, BGC592, and chitinase genes (GH18) in the rhizosphere and endosphere of sugar beet seedlings treated with the synthetic endophytic consortium of *Chitinophaga* and *Flavobacterium* isolates (syncom). LogRQ represents the gene expression levels by relative quantification scores: values below 0 indicate lower expression of the BGC

relative to that of the housekeeping gene (*glyA*) used for data normalization. Bars represent the average of 3-5 biological replicates per treatment and error bars indicate the standard error of the mean. Different letters indicate statistically significant differences between treatments as determined by one-way ANOVA with post-hoc Tukey HSD ($P < 0.05$). **(D)** and **(E)** *Rhizoctonia* damping-off disease incidence of sugar beet seedlings treated with single *Chitinophaga* (Ch93, 94, 95) and *Flavobacterium* (Fl96, 97,98 and 5B) isolates and with a consortium of all seven endophytic isolates (synthetic community, syncom 7), and **(F)** single *Chitinophaga* (Ch94), *Flavobacterium* (Fl98) isolates, two independent Fl98-mutants (Fl98-1 and Fl98-2) with a deletion in BGC298, syncom 7 and the consortium of *Chitinophaga* sp. 94 and *Flavobacterium* sp. 98 (syncom 2). Single isolates and the two syncoms were applied at an initial density of 10^7 CFU g^{-1} of *Rhizoctonia*-conducive field soil. Bars represent the average of 4-8 biological replicates per treatment and error bars represent the standard error of the mean. Disease incidence was scored 21- 28 days after *R. solani* inoculation. Different letters indicate statistically significant differences between treatments as determined by one-way ANOVA with post-hoc Tukey HSD ($P < 0.05$). Note: for Fig. 4B-F, box plots with the individual data of each replicate are provided in Figures S20 and S21.

References and Notes:

1. J. A. Vorholt, C. Vogel, C. I. Carlström, D. B. Mueller, Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* **22**, 142-155 (2017).
2. V. Cordovez, F. Dini-Andreote, V. J. Carrión, J. M. Raaijmakers, Ecology and evolution of plant microbiomes. *Annu. Rev. Microbiol.* **73**, 69-88 (2019).
3. M.-J. Kwak *et al.*, Rhizosphere microbiome structure alters to enable wilt resistance in tomato. *Nat. Biotech.* **36**, 1100-1109 (2018).
4. S. Hacquard *et al.*, Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **17**, 603-616 (2015).
5. R. Mendes *et al.*, Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**, 1097-1100 (2011).

6. K. Panke-Buisse, A. C. Poole, J. K. Goodrich, R. E. Ley, J. Kao-Kniffin, Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J* **9**, 980 (2015).
7. N. Vannier, M. Agler, S. Hacquard, Microbiota-mediated disease resistance in plants. *PLoS Path.* **15**, e1007740 (2019).
8. B. O. Oyserman, M. H. Medema, J. M. Raaijmakers, Road MAPs to engineer host microbiomes. *Curr. Opin. in Microbiol.* **43**, 46-54 (2018).
9. P. Durán *et al.*, Microbial interkingdom interactions in roots promote Arabidopsis survival. *Cell* **175**, 973-983. e914 (2018).
10. D. M. Weller, J. M. Raaijmakers, B. B. M. Gardener, L. S. Thomashow, Microbial populations responsible for specific soil souppressiveness to plant pathogens. *Annu. Rev. Phytopathol.* **40**, 309-348 (2002).
11. E. Chapelle, R. Mendes, P. A. H. M. Bakker, J. M. Raaijmakers, Fungal invasion of the rhizosphere microbiome. *ISME J* **10**, 265-268 (2016).
12. R. L. Berendsen, C. M. J. Pieterse, P. A. H. M. Bakker, The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478-486.
13. M. Mazzola, Manipulation of rhizosphere bacterial communities to induce suppressive soils. *J. Nematol.* **39**, 213-220 (2007).
14. R. Mendes, P. Garbeva, J. M. Raaijmakers, The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* **37**, 634-663 (2013).
15. J.-Y. Cha *et al.*, Microbial and biochemical basis of a Fusarium wilt-suppressive soil. *ISME J* **10**, 119-129 (2016).
16. M. Voort, M. Kempenaar, M. Driel, J. M. Raaijmakers, R. Mendes, Impact of soil heat on reassembly of bacterial communities in the rhizosphere microbiome and plant disease suppression. *Ecol. Lett.* **19**, 375-382 (2016).
17. J. M. Raaijmakers, M. Mazzola, Soil immune responses. *Science* **352**, 1392-1393 (2016).
18. V. J. Carrión *et al.*, Involvement of *Burkholderiaceae* and sulfurous volatiles in disease-suppressive soils. *ISME J* **12**, 2307-2321 (2018).
19. V. Cordovez *et al.*, Diversity and functions of volatile organic compounds produced by *Streptomyces* from a disease-suppressive soil. *Front. Microbiol.* **6**, 1081 (2015).
20. K. Faust, J. Raes, Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538-550 (2012).
21. Y. Yin *et al.*, dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445-W451 (2012).
22. V. Lombard, H. G. Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-D495 (2014).
23. T. Weber *et al.*, antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237-243 (2015).
24. M. Remmert, A. Biegert, A. Hauser, J. Soding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Meth.* **9**, 173-175 (2012).
25. S. M. Bowman, S. J. Free, The structure and synthesis of the fungal cell wall. *BioEssays* **28**, 799-808 (2006).
26. J. Watrous *et al.*, Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1743–E1752 (2012).
27. M. H. Medema *et al.*, Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11**, 625-631 (2015).
28. P. Cimermancic *et al.*, Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

29. Y. Schmidt *et al.*, Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*. *ChemBioChem* **15**, 259-266 (2014).
30. J. D. G. Jones, J. L. Dangl, The plant immune system. *Nature* **444**, 323-329 (2006).
31. R. Gómez Expósito, J. Postma, J. M. Raaijmakers, I. de Bruijn, Diversity and activity of *Lysobacter* species from disease suppressive soils. *Front. Microbiol.* **6**, 1243 (2015).
32. E. Chapelle *et al.*, A straightforward and reliable method for bacterial in planta transcriptomics: application to the *Dickeya dadantii/Arabidopsis thaliana* pathosystem. *The Plant J.* **82**, 352-362 (2015).
33. S. Courtois *et al.*, Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ. Microbiol.* **3**, 431-439 (2001).
34. P. N. Holmsgaard *et al.*, Bias in bacterial diversity as a result of Nycodenz extraction from bulk soil. *Soil Biol. Biochem.* **43**, 2152-2159 (2011).
35. A. Hevia, S. Delgado, A. Margolles, B. Sánchez, Application of density gradient for the isolation of the fecal microbial stool component and the potential use thereof. *Sci. Rep.* **5**, 16807 (2015).
36. S. Ikeda *et al.*, Development of a bacterial cell enrichment method and its application to the community analysis in soybean stems. *Microb. Ecol.* **58**, 703-714 (2009).
37. N. A. Joshi, J. N. Fass, *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. (Version 1.33 [Software]. Available at https://github.com/najoshi/sickle., 2011).
38. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **9**, 357-359 (2012).
39. A. Bankevich *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
40. D. Hyatt *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
41. C. Trapnell *et al.*, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578 (2012).
42. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Meth.* **12**, 59-60 (2015).
43. D. H. Huson, A. F. Auch, J. Qi, S. C. Schuster, MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377-386 (2007).
44. P. Meinicke, UProC: tools for ultra-fast protein domain classification. *Bioinformatics* **31**, 1382-1388 (2015).
45. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
46. M. Y. Galperin, K. S. Makarova, Y. I. Wolf, E. V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261-269 (2015).
47. R. D. Finn *et al.*, The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-D285 (2016).
48. M. Imelfort, B. Woodcroft, D. Parks, *Ecogenomics/BamM [WWWDocument]* (Available at: https://github.com/Ecogenomics/BamM, 2015).
49. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
50. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]*, (2013).
51. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

52. B. Bushnell, *BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7065E. Retrieved from https://escholarship.org/uc/item/1h3515gn*. (2014).
53. C. Quast *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590-D596 (2013).
54. J. G. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* **7**, 335-336 (2010).
55. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: Versatile open-source tool for metagenomics. *PeerJ 4:e2584 https://doi.org/10.7717/peerj.2584*, (2016).
56. D. McDonald *et al.*, The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *GigaScience* **1**, 7 (2012).
57. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520-2522 (2012).
58. J. Kuczynski *et al.*, Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr. Protoc. Microbiol.* **27**, 1E. 5.1-1E. 5.20 (2012).
59. J. Oksanen *et al.*, The vegan package. *Community ecology package* **10**, 631-637 (2007).
60. J. N. Paulson, O. C. Stine, H. C. Bravo, M. Pop, Differential abundance analysis for microbial marker-gene surveys. *Nat. Meth.* **10**, 1200-1202 (2013).
61. P. J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* **8**, e61217 (2013).
62. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47-e47 (2015).
63. J. Friedman, E. J. Alm, Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
64. M. E. Newman, The structure and function of complex networks. *SIAM review* **45**, 167-256 (2003).
65. M. E. Newman, Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577-8582 (2006).
66. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504 (2003).
67. M. Bastian, S. Heymann, M. Jacomy, in *Third international AAAI conference on weblogs and social media*. (2009).
68. T. Weber *et al.*, antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237-W243 (2015).
69. J. Navarro-Muñoz *et al.*, A computational framework for systematic exploration of biosynthetic diversity from large-scale genomic data. *BioRxiv*, 445270 (2018).
70. M. H. Medema, E. Takano, R. Breitling, Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* **30**, 1218-1223 (2013).
71. J. Alneberg *et al.*, Binning metagenomic contigs by coverage and composition. *Nat. Meth.* **11**, 1144 (2014).
72. Y.-W. Wu, Y.-H. Tang, S. G. Tringe, B. A. Simmons, S. W. Singer, MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 1-18 (2014).
73. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
74. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055 (2015).

75. D. E. Wood, S. L. Salzberg, Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biol.* **15**, R46 (2014).
76. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. J. B. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).
77. S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, S. R. J. Eddy, Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439-441 (2003).
78. K. D. Pruitt, T. Tatusova, D. R. J. N. a. r. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-D65 (2006).
79. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
80. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
81. I. Letunic, P. Bork, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475-W478 (2011).
82. A. Untergasser *et al.*, Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71-W74 (2007).
83. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotech.* **37**, 540-546 (2019).
84. B. D. Ondov *et al.*, Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
85. I. Lee, Y. O. Kim, S.-C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Micr.* **66**, 1100-1103 (2016).
86. N. Segata, D. Börnigen, X. C. Morgan, C. Huttenhower, PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, 2304 (2013).
87. Y. Bai *et al.*, Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**, 364-369 (2015).
88. M. McBride, M. Kempf, Development of techniques for the genetic manipulation of the gliding bacterium *Cytophaga johnsonae*. *J. Bacteriol.* **178**, 583-590 (1996).
89. S. Chen, M. Bagdasarian, M. Kaufman, E. Walker, Characterization of strong promoters from an environmental Flavobacterium hibernum strain by using a green fluorescent protein-based reporter system. *Appl. Environ. Microbiol.* **73**, 1089-1100 (2007).
90. M. J. McBride, S. A. Baker, Development of techniques to genetically manipulate members of the genera *Cytophaga*, *Flavobacterium*, *Flexibacter*, and *Sporocytophaga*. *Appl. Environ. Microbiol.* **62**, 3017-3022 (1996).
91. J. A. Gagnon *et al.*, Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS one* **9**, e98186 (2014).
92. S. Chen, M. Bagdasarian, M. G. Kaufman, A. K. Bates, E. D. Walker, Mutational analysis of the *ompA* promoter from *Flavobacterium johnsoniae*. *J. Bacteriol.* **189**, 5108-5118 (2007).
93. H. Xie, P. Nie, B. Sun, Characterization of two membrane-associated protease genes obtained from screening out-membrane protein genes of *Flavobacterium columnare* G4. *J. Fish Dis.* **27**, 719-729 (2004).
94. L.-E. Jao, S. R. Wente, W. Chen, Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13904-13909 (2013).

95. S. Agarwal, D. W. Hunnicutt, M. J. McBride, Cloning and characterization of the *Flavobacterium johnsoniae* (*Cytophaga johnsonae*) gliding motility gene, *gldA*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12139-12144 (1997).
96. Data, scripts and code used for statistical and bioinformatic analyses are available at: http://doi.org/10.5281/zenodo.3405564.

Acknowledgments: We thank Irene de Bruijn, Silvia P. Vega-Hernández, Victor de Jager and Roos Keijzer for their valuable advice for genomic and metagenomic DNA extractions. We also would like to thank the BSc and MSc students Cristina Rotoni, Ryan Hijkoop, Hannah McDermott, Azkia Nurfikari, Jelle Spooren and Rik Peters for their valuable help in the *Flavobacterium* and *Chitinophaga* isolation and phenotyping. We thank Mark J McBride at the University of Wisconsin for providing the plasmids for *Flavobacterium* transformation. This is publication number --- of the NIOO-KNAW. **Funding:** J.M.R. and V.J.C. were supported by grants from the Netherlands BE-Basic Foundation and the Dutch TTW-Perspective program 'Back to the Roots'. V.J.C. was also supported by a fellowship from the Junta de Andalucía (Proyecto de Excelencia, P07-AGR-02471) and Plan Propio of the University of Málaga. V.C. was supported by the Dutch STW-program 'Back to the Roots'. JEP-J was financially supported by the Department of Science, Technology and Innovation of Colombia—COLCIENCIAS through the doctoral grant 568-2012-15517825. **Author contributions:** The experiments were designed and conducted by V.J.C., V.C., D.R.B., S.S.E. and J.E.P.J.. R.G.E. contributed to the soil sample collection. Metagenome data were generated by V.J.C and W.F.v.J.. P.M., A.A. and J.v.d.O. made the *Flavobacterium* mutants. Data were analyzed by V.J.C., J.E.P.J., V.T., L.W.M., M.H.M., J.N.P and M.d.H. Funding acquisition and project management was provided by J.M.R. The manuscript was written by V.J.C., R.M., G.P.v.W., M.H.M. and J.M.R.. All authors read and approved the final manuscript. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The sequencing data are available under EBI submission number PRJEB8920. Data, scripts and code used for statistical and bioinformatic analyses are available at: http://doi.org/10.5281/zenodo.3405564 (*96*).