

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Metagenomic sequencing data were generated on an Illumina NovaSeq 6000 platform (PE150, 2x150 bp) using the instrument's default control software.
Data analysis	Natural-community comparisons used Kaiju v1.8.2 (genus-level classification) for AtSC metagenomes and QIIME2 2022.8 (DADA2 + VSEARCH at 97% identity, V5–V7 in-silico extraction) for HvSC 16S data. Experimental metagenomic reads were then quality-filtered, with host sequences removed using Bowtie2 v2.2.5. Cleaned reads were pseudoaligned with Salmon v1.9.0 (--minScoreFraction 0.95) against inoculum-specific or combined SSC reference indices, and outputs were assembled into isolate-level count tables normalized by genome length. Functional predictions were inferred with PICRUSt2 v2.5.1 to obtain KO abundance profiles, which were converted to relative abundances. Phylogenetic relationships among isolates were reconstructed with OrthoFinder (STAG/STRIDE). Alpha and beta diversity analyses were conducted using phyloseq v1.44.0 and vegan v2.6-4 (Bray–Curtis distances, PCoA, PERMANOVA), including centroid-based dispersion analyses. Functional contribution of KOs was assessed with sPLS-DA (optimized in mixOmics v6.24.0). KO differential abundance testing employed DESeq2 v1.40.0 (apeglm shrinkage, BH correction), complemented by Wilcoxon tests with Bonferroni adjustment. Additional simulation and enrichment analyses (including Barnard's exact test) were performed in R and Python 3.12.5 with SciPy v1.13.1, and figures generated in ggplot2 v3.5.1.

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](#)

Data deposition: The data reported in this article have been deposited in the National Center for Biotechnology Information Short Read Archive BioProject database. The relevant accession numbers can be found in the separate, specific sections below. Arabidopsis culture collection Genomic reads from the bacterial cultures are deposited under accession number PRJNA1131834, while the genomic sequences are deposited under accession numbers PRJNA1138681 (isolated at three days post transplantation) and PRJNA1139421 (isolated at seven days post transplantation). Genomes, PROKKA gene predictions, and EggNOG annotations are also available at <https://zenodo.org/records/10992416> (Selten et al., 2024). Barley culture collection Genomic reads from the bacterial cultures are deposited under accession number PRJNA1131819, while the genomic sequences are deposited under accession number PRJNA1139693. Lotus culture collection Genomic reads as well as genome assemblies from the bacterial cultures have been deposited previously from Wippel et al. (2021) under accession number PRJEB37696. Genomic sequences can also be found at <https://www.at-sphere.com/>. Bacterial isolates LjRoot204, LjRoot204_2, LjRoot205, LjRoot206, and LjRoot208 were not included under this accession number and can be found under accession numbers PRJNA1142758 (genomic reads) and PRJNA1142862 (genomes). Metagenome sequencing data Shotgun metagenome-sequenced reads from the root microbiomes of Arabidopsis, Barley, and Lotus inoculated with the AtSC, HvSC, LjSC, and SSC generated in this study have been deposited under accession number PRJNA1131994. The shotgun metagenome-sequenced reads from the Arabidopsis multi-SynCom root colonization experiment are deposited under accession number PRJNA1449013, while the shotgun metagenome-sequenced reads from the Lotus bacterial family drop out SynCom root colonization experiment are deposited under accession number PRJNA1444504. Shiny app To assist the community in investigating specific KOs in our data, we have built a shiny app under the following link: https://pm-bacterial-genetics-au.shinyapps.io/SSC_community_app/. This app will allow users to investigate KO abundances across hosts and inocula, abundances of isolates with a specific KO/pathway across hosts and inocula, and the functional enrichment of KOs for each host that is similarly calculated in this manuscript. The user can specify the original dataset as well as the dataset in which the dominator isolates, Lotus symbionts (LjNodule 214, P1_H10, P2_A12, and P2_D6), and the HvSC-dominant P2_G4, are computationally excluded. The data to generate the figures in the shiny app are included in Data S33 (original dataset) and S34 (exclusion of nodulators and P2_G4 dataset).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	na
Reporting on race, ethnicity, or other socially relevant groupings	na
Population characteristics	na
Recruitment	na
Ethics oversight	na

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample sizes were determined based on feasibility of plant growth and harvesting within experimental cycles, and were sufficient to allow robust detection of treatment effects. No formal statistical power calculation was performed.
Data exclusions	Data with the Rhizobiaceae drop out were excluded as a Rhizobiaceae isolate was found in these root samples. No data were excluded from the other studies.
Replication	All key experiments were performed twice, independently, each representing a full biological replicate.
Randomization	Randomization was not applied when assigning plants to groups, as experimental design prioritized minimizing cross-contamination between

Randomization	inocula and hosts. At the time of the experiments, no specific outcomes were anticipated, as microbiome community composition is inherently variable.
Blinding	Blinding was not performed due to the practical requirements of inoculation and sample handling.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

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

Plants

Seed stocks

Lotus japonicus accession Gifu seeds were obtained from the Aarhus University MBG greenhouse. Hordeum vulgare cultivar Golden Promise seeds were sourced from Norwich, UK. Arabidopsis thaliana accession Col-0 seeds were used from standard laboratory stocks.

Novel plant genotypes

No novel plant genotypes were generated in this study. Experiments were conducted using Lotus japonicus accession Gifu, Hordeum vulgare cultivar Golden Promise, and Arabidopsis thaliana accession Col-0.

Authentication

NA