nature portfolio

Corresponding author(s): Marcela Mendoza-Suárez
Stig Uggerhøj Andersen

Last updated by author(s): May 21, 2024

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	All the codes are available in GitHub and links are provided in the manuscript
Data analysis	Fully described in the Methods section

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

The data that support the findings of this study are openly available in:
Supplementary Data Table 1. Greenhouse setup. https://github.com/tyakyol/rhizobialStrains2_Data/blob/main/greenhouse.xlsx
Supplementary Data Table 2. Genotype table for the faba bean cultivars. https://github.com/tyakyol/rhizobialStrains2_Data/blob/main/snps.csv.zip
Supplementary Data Table 3. Rhizobial strain read counts in nodule samples (raw).

https://github.com/tyakyol/rhizobialStrains2_Data/blob/main/countsPlantRaw.csv
Supplementary Data Table 4. Rhizobial strain read counts in inoculum samples (raw).
https://github.com/tyakyol/rhizobialStrains2_Data/blob/main/countsInoculum.csv
Supplementary Data Table 3. Rhizobial strain read counts in nodule samples (adjusted).
https://github.com/tyakyol/rhizobialStrains2_Data/blob/main/countsPlantAdjusted.csv

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	603 samples
Data exclusions	We excluded five samples with read counts below 2,000
Replication	3
Randomization	Random position on the greenhouse and for analysis we perform Random forest prediction
Blinding	NA

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	Seed stocks from the ProFaba panel described in (Skovbjerg et al., 2023). C. K. Skovbjerg, D. Angra, T. Robertson-Shersby-Harvie, J. Kreplak, G. Keeble-Gagnère, S. Kaur, W. Ecke, A. Windhorst, L. K. Nielsen, A. Schiemann, J. Knudsen, N. Gutierrez, V. Tagkouli, L. I. Fechete, L. Janss, J. Stougaard, A. Warsame, S. Alves, H. Khazaei, W. Link, A. M. Torres, D. M. O'Sullivan, and S. U. Andersen, “Genetic analysis of global faba bean diversity, agronomic traits and selection signatures,” Theor. Appl. Genet., vol. 136, no. 5, p. 114, 2023, doi: 10.1007/s00122-023-04360-8.
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	NA