nature portfolio

Corresponding author(s): Bhatt

Last updated by author(s): Jul 23, 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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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	Candidate isolate long-read sequencing datasets were identified on NCBI with the following search criteria: "(Bacteria[Organism] OR Archaea[Organism]) AND ("pacbio smrt"[Platform] OR "oxford nanopore"[Platform]) AND genomic[Source]". Datasets were further filtered by removing datasets with the "amplicon" flag, and removing datasets with less than 50 Mbp of sequencing in total. Individual read datasets were downloaded with fastq-dump, a part of the sratoolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). Nanostat was run on each remaining readset to measure dataset characteristics.
Data analysis	Long-read datasets were then analyzed with PhaVa (v0.1.0) with default parameters. PhaVa is available at https://github.com/patrickwest/PhaVa

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

Short-read adult HCT stool sequencing data was previously published and is available at (NCBI BioProject ID PRJNA707487). Short-read pediatric HCT stool sequencing data was previously published and is available at (NCBI BioProject ID PRJNA787952). Long-read metagenomic sequencing data was previously published and is available at BioProject PRJNA820119 and BioProject PRJNA940499. Assembled metagenomic contigs are available at https://doi.org/10.5281/zenodo.7662825. A list of accession numbers for long-read isolate sequencing data is available in supplementary file Data S5. MS raw files (.d) generated in this study have been deposited to the public repository, ProteomeXchange Consortium, through the PRIDE partner repository (project accession is PXD054577). Long-read sequencing data for the locked thiC intragenic inverton strains and RNA sequencing data are available NCBI BioProject ID PRJNA1118344. Accession codes for long read datasets are listed in the extended data section. The reference genome for B. thetaiotaomicron VPI-5482 is the NCBI Reference Sequence AE015928.1. The reference genome for B. fragilis FDAARGOS_1225 is the NCBI Reference Sequence NZ_CP069563.1.

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	No sample-size calculation was performed. For our in vitro assays with significance testing, we followed standard laboratory procedure to repeat experiments in biological duplicate or triplicate on at least three separate occasions. For in vitro assays, this has historically been sufficient to identify statistical differences between groups using routine statistical tests.
Data exclusions	Candidate isolate long-read sequencing datasets were chosen by a predetermined set of criteria described in detail in the methods section. No in vitro data was excluded.
Replication	For the in vitro competitive growth assays, antibiotic markers were flipped between strains to control for any fitness advantages/ disadvantages these markers would provide. Experiments replicated as expected
Randomization	No randomization was performed for in vitro assays as inclusion in a group was determined by the bacterial genotype. We controlled for confounding variables by treating all samples in exactly the same way and with the same media.
Blinding	Blinding was not done as there was no opportunity for this to occur.

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