nature research

Corresponding author(s): Justin L. Sonnenburg

Last updated by author(s): 2021/05/23

Reporting Summary

Nature Research 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 Research 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	Metabolomics data were collected on Agilent QTOF instruments (models 6530 and 6545) using Agilent's LC/MS Data Acquisition software (version 10.1). Metabolomics data were also collected on the ThermoFisher Q Exactive HF using Thermo Scientific Xcalibur Data Acquisition software (version 4.3). Optical density data were collected using the BioTek Gen5 software V3.03.
Data analysis	Data were analyzed using Agilent Qualitative Analysis (version B.07.00), the MS-DIAL software (version 3.83), Python-based custom code, and Prism version 8.0. The bioinformatics pipeline for LC/MS MS/MS library construction, in vitro data processing, and in vivo data processing were done with custom Python-based code available at the Sonnenburg lab Github site (https://github.com/SonnenburgLab/Han_and_Van_Treuren_et_al_2021). The JavaScript code for the interactive, web-based software (Metabolomics Data Explorer) is also available at the Sonnenburg lab Github site (https://github.com/SonnenburgLab/Metabolomics_Data_Explorer). The dependencies for the Python based code can be found in this .yml file (https://github.com/SonnenburgLab/Han_and_Van_Treuren_et_al_2021/blob/master/environment.yml) and in the specific scripts found at the Sonnenburg lab Github. Bioinformatics processing of 16S gene sequences was done with QIIME1 (legacy release available via Conda install: http://qiime.org/install/install.html). MultiGeneBlast version 1.1.13, SPAdes version 3.9.1, and prokka version 1.14.5 were used for comparative genomics in search of polyamine biosynthetic genes. Custom Python code was written to enable the construction of the MS/MS spectra library, the processing and visualization of the in vitro and in vivo LC/MS data, the optical density and growth curve data, the bioinformatic analysis of 16S and whole genomes, and the analysis of the metabolomics data. Full code for each of these steps is provided at the Sonnenburg lab GitHub site (https://github.com/SonnenburgLab/Han_and_Van_Treuren_et_al_2021).

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All metabolomics raw data are publicly accessible on the Metabolomics Workbench under study number ST001683 for all in vivo data and study number ST001688 for in vitro data. The MS/MS spectra library constructed on the qTOF and QE instruments are publicly accessible on the MONA spectra database.

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 for the microbial culture (in vitro data) were originally chosen as n = 5 (5 biological replicates). Initial work identified that biological triplicates were sufficient to detect outlier samples (e.g. machine injection failures) and these were used subsequently. See Extended Data Fig. 2 and 4b, and the in vitro bioinformatics pipeline (https://github.com/SonnenburgLab/Han_and_Van_Treuren_et_al_2021/tree/master/in_vitro_pipeline_and_analysis) for a full description of these outlier detection methods and results. When a highly variable sample was identified, it was discarded and the microbe was regrown and the new triplicates compared with the old triplicates. In some cases, microbes were grown repeatedly to verify inter-experimental variability. All metadata associated with sample sizes and inter-experiment variability repeats can be found in Supplementary Table 5. Sample sizes for gnotobiotic animals (in vivo data) were determined based on animal housing and experimental design matching constraints. In particular, n = 5 was chosen for mouse groups because our mouse facility could supply this number of age, sex, and litter-mate matched mice for most experiments.
Data exclusions	Metabolomics samples went through a rigorous, multi-step process for quality control. Samples were eliminated if internal standards were detected significantly below expected values, if the correlation with biological replicates was two standard deviations below the mean, and if a random forest classifier identified a sample as having been produced by a bacterium found in a different phylum than the actual producing microbe. These steps, and the samples that were excluded because of them, are given completely in the in vitro bioinformatics pipeline (https://github.com/SonnenburgLab/Han_and_Van_Treuren_et_al_2021/tree/master/in_vitro_pipeline_and_analysis). The criteria for the exclusion of samples were a mixture of pre-established (internal standard filter, correlation coefficient filter) and post-hoc (random forest analysis). Metabolite features detected on the mass spectrometry instrument from all experiments also underwent a quality control filtering process in the MS-DIAL software (Extended Data Fig. 1f). Based on the list of feature (or peak) identified for each experiment, each set of aligned peaks was manually checked in MS-DIAL. Select metabolite features were removed from this list based on pre-established criteria for misidentification, poor peak shapes, or background contamination peaks. Annotated features that passed this inspection were reported in the final output file. See details in Materials and Methods.
Replication	Growth curves (Fig. 3) were repeated in at least three independent experiments. In vitro experiments for metabolomic profiling were repeated with at least 3 biological replicates (3 independent cultures) in one or more independent experiments. In vivo mouse experiments for metabolomic profiling were repeated with at least three mice per condition in one or more independent experiments. Please see details on the number of biological replicates and independent experimental repeats described in relevant figure legends.
Randomization	For in vitro studies, randomization of sample injection order was not conducted. Internal standards were monitored for injection-order, sample storage time, and carry-over effects. No significant effects were found. The injection order of all samples reported in this study can be found in Supplementary Table 5. The microbes selected for each growth and LC/MS measurement were selected pseudo-randomly; there was an imbalance of phylogenetic covariates (e.g., some experiments had more Bacteroidetes than others) but inter-experiment replication was carried out to mitigate any statistical effects. For mouse experiments (in vivo studies), mice were assigned to treatment groups randomly taking into account age, sex, and litter-mate matching.
Blinding	Mouse experiments (in vivo studies) were not blinded because no subjective measurement modalities were employed (e.g. no tissue histology scoring). All mouse metabolomic samples were prepared in the same way regardless of group and monitored as described in the "Data Exclusions" section.

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

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse experiments were performed on gnotobiotic Swiss Webster germ-free mice (males, 10-14 weeks of age, n = 3-8 per group for all experiments) or Swiss-Webster Excluded Flora mice ("conventional mice", males, 10-14 weeks of age, n = 3 per group) maintained in aseptic isolators, and originally obtained from Taconic Bioscience. Mice were maintained on a 12-hour light/dark cycle at 69˚F in ambient humidity, fed ad libitum, and maintained in flexible film gnotobiotic isolators for the duration of all experiments (Class Biologically Clean, Madison WI).
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All animal experiments were performed in accordance with the Stanford Institutional Animal Care and Use Committee.

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