nature portfolio

Corresponding author(s): Steven Hallam

Last updated by author(s): Jul 27, 2023

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	No software was used for data collection
Data analysis	Statistical analysis and figure generation code is provided as a supplemental file. Open-source software was used for data analysis. QIIME2 v2021.4; BLASTN v2.5.0, SPAdes v3.11; bbmap v37.78; MetaBAT v2.12.1; CheckM v1.0.12; GTDB-tk v0.2.2; tRNAscan-SE 2.0.6; INFERNAL v1.1.3; HMMER v3.0; bowtie2 v2.2.5; Picard Toolkit; R packages: DESeq2, limma, ggplot2, stats, phyloseq, networkD3, dendextend, vegan, SpiecEasi, indicspecies

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 generated for this study can be found deposited in public repositories. For 16S rRNA amplicon sequences, reference the NCBI GenBank database under

BioProject PRJNA902729 as BioSamples SAMN32228586 - SAMN32228684. For whole-genome shotgun metagenome sequences, reference the JGI GOLD database under IMG study Gs0127566, with sequencing project IDs Gp0293654, Gp0293652, Gp0293653, Gp0293650, Gp0293651, Gp0293648, Gp0293649, Gp0293647, Gp0406093, Gp0406094, Gp0406085, Gp0406096, Gp0406083, Gp0406099, Gp0406100, Gp0406104, Gp0406081 (See Table S3 for details). For mRNA shotgun metatranscript sequences, reference the NCBI SRA database under BioProject PRJNA902729 as BioSamples SAMN34106045 - SAMN34106058. For the code used to conduct statistical analysis and create data visualizations in R v4.1.2, reference the accompanying file "lulu_microbes_viz.Rmd".

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	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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	N/A
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

Behavioural & social sciences study design

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

Study description	N/A
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing	N/A
Data exclusions	N/A
Non-participation	N/A
Randomization	N/A

Ecological, evolutionary & environmental sciences study design

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

Study description	A 2-year time series study was conducted in a full-scale wastewater treatment plant anaerobic digester (AD) to determine microbiome composition in relation to physicochemistry and biogas production across 4 sample groups. Data generated represented 1) the physicochemical state of the AD material and 2) the microbial community taxonomic and function composition at specific timepoints. Samples were delineated into 4 groups spanning 52 samples taken throughout the 2-year study (16 for Standard Operation I, 10 for Chemically Enhanced Pretreated Operation, 11 for Standard Operation II, and 6 for Serial Operation). Statistical analysis to determine whether these 4 Operations selected for distinct microbial communities was conducted in a pairwise manner.
Research sample	Physical samples were taken from activated sludge material within the AD tank in order to understand the metabolic and chemical processes occurring within. These samples were roughly 50mL and meant to represent the whole AD tank (>10,000 L). Sampling size was constrained by the port available to sample from. Datasets generated from this sample were numerical matrices and DNA/RNA sequences describing the samples.
Sampling strategy	Activated sludge was collected from a sampling port on the side of an anaerobic digester by wastewater treatment plant operators. Sample size was precalculated from an estimate of microbial biomass concentration (cells/mL) and based on the amount of material needed to perform physicochemical assays. The sample size chosen is within the range of standard samples taken from environmental matrices for microbial community analysis.
Data collection	Physicochemical data were collected from samples by plant operators using industry-standard assays. DNA/RNA sequence data were generated by short-read Illumina sequencing technologies.
Timing and spatial scale	Samples were taken between Oct 13th, 2016 and August 1, 2018, approximately every 2 weeks. Samples were removed from the AD tank by a sampling port near the bottom of the tank (the same spatial location each time).
Data exclusions	Some data were obtained from a second digester, but these samples did not encompass the entire time series and were excluded
Reproducibility	Each sample had 3 biological replicates which were sampled one after the other (from the same port) on each sampling day. These replicates were treated independently during analysis. Spanning 2 years, the study reproduced each season twice.
Randomization	There was 1 AD tank and 1 sampling port available, so no spatial randomization was done. There are no co-variates to account for in the study because the intent was to sample the same location at constant intervals to determine its properties.
Blinding	Blinding was not relevant for this study.

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	N/A
Location	N/A
Access & import/export	N/A
Disturbance	N/A

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

Antibodies

Antibodies used	N/A
Validation	N/A

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	N/A
Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See ICLAC register)	N/A

Palaeontology and Archaeology

Specimen provenance	N/A
Specimen deposition	N/A
Dating methods	N/A

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	N/A

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	N/A
Wild animals	N/A
Reporting on sex	N/A
Field-collected samples	N/A
Ethics oversight	N/A

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	N/A
Study protocol	N/A
Data collection	N/A
Outcomes	N/A

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	N/A
Novel plant genotypes	N/A
Authentication	N/A

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	N/A

Files in database submission	N/A
Genome browser session (e.g. UCSC)	N/A

Methodology

Replicates	N/A
Sequencing depth	N/A
Antibodies	N/A
Peak calling parameters	N/A
Data quality	N/A
Software	N/A

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	N/A
Instrument	N/A
Software	N/A
Cell population abundance	N/A
Gating strategy	N/A

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	N/A
Design specifications	N/A
Behavioral performance measures	N/A

Acquisition

Imaging type(s)	N/A
Field strength	N/A
Sequence & imaging parameters	N/A
Area of acquisition	N/A

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	N/A
Normalization	N/A
Normalization template	N/A
Noise and artifact removal	N/A
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	N/A
Effect(s) tested	N/A

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	N/A
Correction	N/A

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	N/A
Graph analysis	N/A
Multivariate modeling and predictive analysis	N/A