nature portfolio

Corresponding author(s):	Shyamal Peddada
Last updated by author(s):	Sep 13, 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 <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

< ∙	トつ	1		Ηı	\sim
.)	ıa	ш	15	u	CS

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	nfirmed
	\boxtimes	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	\boxtimes	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	\boxtimes	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.
	\boxtimes	A description of all covariates tested
	\boxtimes	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
	\boxtimes	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)
	\boxtimes	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>
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
\boxtimes		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 <u>statistics for biologists</u> 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

ANCOM-BC2 has been implemented in the R package ANCOMBC, which is available on Bioconductor at https://www.bioconductor.org/packages/release/bioc/html/ANCOMBC.html. The entirety of the analyses, conducted in RStudio (R version 4.2.2) on aarch64-apple-darwin20 (64-bit) running macOS Monterey 12.6.8, with the sole exception of the soil microbiome richness trend test, is provided in the related GitHub repository, which can be found at https://github.com/FrederickHuangLin/ANCOM-BC2-Code-Archive. For the aforementioned exception, ORIOGEN 4.01 was employed and is available at https://www.niehs.nih.gov/research/resources/software/biostatistics/oriogen/index.cfm.

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

Policy information about studies involving human research participants and Sex and Gender in Research.

The upper respiratory tract (URT) data were sourced from the LOCOM R package (https://github.com/yijuanhu/LOCOM-Archive). The Quantitative Microbiome Project (QMP) data are accessible via the SPRING R package (https://github.com/GraceYoon/SPRING) or the ANCOMBC package (https://www.bioconductor.org/ packages/release/bioc/html/ANCOMBC.html). Data pertaining to soil microbiome for aridity and gut microbiome in IBD patients are hosted on Qiita, with respective links available at https://giita.ucsd.edu/study/description/10360 and https://giita.ucsd.edu/study/description/11546, respectively.

		1		
⊢	liiman	research	narticii	nantc
	iuiiiaii	1 C3Cal CII	partici	pants

Reporting on sex and gender	Reporting on sex and gender No experiment in this study			
Population characteristics	No experiment in this study			
Recruitment	No experiment in this study			
Ethics oversight	No experiment in this study			
Note that full information on the appr	oval of the study protocol must also be provided in the manuscript.			
Field-specific re	eporting			
Please select the one below that i	s the best fit for your research. If you are not sure, read the appropriate sections before making your selection.			

Ecological, evolutionary & environmental sciences

Life sciences study design

X Life sciences

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

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

Sample size	We utilized two publicly available datasets: soil microbiome for aridity and gut microbiome in IBD patients. No sample size calculation was performed.
Data exclusions	We utilized all samples provided by the authors on the public website for the two datasets. No exclusions were made.
Replication	Not relevant. We used publicly available datasets and did not participate in sample generation.
Randomization	Not relevant. For this paper, we utilized two publicly available datasets: soil microbiome for aridity and gut microbiome in IBD patients. We did not conduct any randomization.
Blinding	Not relevant. Two real datasets used for illustrations in this paper, i.e., soil microbiome for aridity and gut microbiome in IBD patients, are publicly available. We were not involved in the blinding or unblinding process.

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.

portfolio reporting summar	ומנטות	Du+1 FD -
rtfolio reporti		5
olio reporti	Ξ	7
reporti		<u>)</u>
_글.	5	2.

Ma	terials & experimental systems	Me	thods
n/a	Involved in the study	n/a	Involved in the study
\boxtimes	Antibodies	\boxtimes	ChIP-seq
\boxtimes	Eukaryotic cell lines	\boxtimes	Flow cytometry
\boxtimes	Palaeontology and archaeology	\boxtimes	MRI-based neuroimaging
\boxtimes	Animals and other organisms		
\boxtimes	Clinical data		
\boxtimes	Dual use research of concern		