nature portfolio

Corresponding author(s): Leonard C Harrison

Last updated by author(s): Mar 5, 2025

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	Excel
Data analysis	KneadData (v0.7.7 https://github.com/biobakery/kneaddata; Shannon information index (https://github.com/scripts/blob/master/shannons-filter.py; at https://github.com/theo-allnutt-bioinformatics/Indigenous_gut_mircobiome_2023; USEARCH v10.0.240; BLAST (nt database, 28/8/2022; R package;https://github.com/pmartinezarbizu/pairwiseAdonis;

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 original data presented in this study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding author. Data on individual living humans cannot be publicly available due to its sensitive nature, as regulated by privacy legislation.

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	The study aimed to recruit an equal number of each sex, based on assignment at birth. Recruitment of Indigenous infants was based on access to all available subjects under 2 years of age, whose parents or guardians consented after a thorough (written) consent process in the native language. Disaggregated sex data and analysis are provided.
Reporting on race, ethnicity, or other socially relevant groupings	The gut microbiomes of Indigenous Australian infants living in a closed remove community were compared to those of non-Indigenous Caucasian infants living in Western urban environments. Indigenous and non-Indigenous nfants were matched for sex and age. The non-Indigenous infants were participants in a pregnancy-birth cohort study (ENDIA) in which the primary entry criterion was the presence of type 1 diabetes in a first-degree relative of the infant. Diets were not controlled for but access to Western and traditional foods was documented in the Indigenous group.
Population characteristics	Characteristics of the subjects, e.g., breast feeding history etc, are documented in Table 1 of the manuscript.
Recruitment	Recruitment of Indigenous infants was based on access to all available subjects under 2 years of age, whose parents or guardians consented after a thorough (written) consent process in the native language. I is estimated that the number of Indigenous infants recruited (50) was from a total of about 70. Infants not recruited included those currently away from the community or for whom parental or guardian consent was not provided (for reasons that could not be ascertained). Non-Indigenous infants were then selected from the ENDIA pregnancy-birth cohort based on sex and the nearest age match. prior to commencing the study and after community engagement, the local research team participated in a week-long codesign and training program which involved telling the research story in the local language, with discussion of the study protocol, means of recruitment, data collection and consent. In this program, the concept of the microbiome was discussed using metaphors and microscopy to develop an understanding of the role of microscopic organisms in human health. Prior to enrolment, parents or guardians gave written informed consent on behalf of infant participants.
Ethics oversight	The study protocol was approved by the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research (Ref. 2017-2814), Melbourne Health Human Research Ethics Committee (Ref. 2017.064), Miwatj Health Indigenous Corporation Board and the Local Shire Authority.

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 power calculation was performed as samples were obtained based on the maximum number of Indigenous infants who were accessible in the remote community. Previous studies of our own and others indicated that for a comparative study (Indigenous vs non-Indigenous) 50 samples in each group would provide meaningful measures of the gut microbiome for comparison.
Data exclusions	There were no data exclusions.
Replication	The reproducibility of metagenomic sequences was not determined.
Randomization	Samples were allocated into the two comparison groups, Indigenous vs non-Indigenous infant.
Blinding	Samples were coded and investigators blinded to their origin and identify.

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

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

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.