

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 used for data collection.

Data analysis

Raw sequencing reads were pre-processed using QIIME. Overlapping sequences were compiled and low-quality reads were removed. Chimeric sequences were removed, and zero-radius Operational Taxonomic Units (zOTUs) constructed using the UNOISE 3 algorithm from Vsearch. The analysis was based on zero-radius OTUs (zOTUs) using RDP-based taxonomic assignment at the lowest unambiguous taxonomic rank. Additionally, the taxonomy of all the taxa that belonging to genus *Lactobacillus* has been updated for the new taxonomic description of the genus. Samples with less than 5000 reads (8 samples in total) were removed and zOTUs found in $\leq 1\%$ of the samples were purged from the zOTU table.

Data analysis was done using R version 4.1.2 (2021-11-01) using packages Vegan v2.5-7, Phyloseq v1.38.0 and DESeq2 v1.34.0. Logistic regression analysis adjusted for enrolment season and enrolment site in order to identify factors associated with diarrhea and between subcategories of diarrhea cases. Putative explanatory variables with a p value less than 0.05 were considered statistically significant.

Alpha diversity measures were based on the number of observed zOTUs and Shannon diversity metrics. Bray-Curtis dissimilarity metrics were determined after cumulative sum scaling (CSS) normalization. Unconstrained principal coordinate analysis (PCoA) and constrained PCoA plots conditioned for age, enrolment site, enrolment season, and sex were subsequently generated. Pairwise permutation multivariate analysis of variance (PERMANOVA) was used for pairwise comparisons based on Bray-Curtis dissimilarity metrics between groups (clinical /demographic data). In addition, DESeq2 was used for differential abundance testing between diarrhea cases and non-diarrheal controls, as well as between the subcategories diarrhea and non-diarrheal controls (illustrated in heatmaps). Subsequently, Wilcoxon rank sum test was carried out on the bacterial taxa that were found to differ significantly in abundance by DESeq2 ($q < 0.05$, with the q value corrected by the Benjamini-Hochberg method). All q values corrected by the Benjamini-Hochberg method reported in the differential abundance testing were from Wilcoxon rank sum test and relative abundances were presented as means. The DESeq2 results were augmented by sparse partial least square discriminatory analysis (sPLS-DA) with 10-fold cross validation used to determine the taxa that discriminates best between diarrhea case and

non-diarrheal control status as well as between the subcategories of diarrhea cases and non-diarrheal controls. To further characterize the bacterial community clusters, we used de novo clustering based on community composition. Overall relationships between clinical and demographic parameters and GM composition in children were illustrated using distance-based Redundancy Analysis (db-RDA) conditioned for enrolment site and enrolment season based on Bray-Curtis dissimilarity metrics.

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 sequence data reported in this study published at Sequence Read Archive (SRA) <https://www.ncbi.nlm.nih.gov/sra/PRJNA1036533> and clinical data used in this study cannot be made freely available to protect the privacy of the participants, in accordance with the Danish Data Protection Act and European Regulation 2016/679 of the European Parliament and of the Council (GDPR). However, data can be made available upon request after agreement reached with University of Copenhagen and other participating institutions via the corresponding authors Getnet Tesfaw (gettesfaw2@gmail.com) and Dennis Sandris Nielsen (dn@food.ku.dk), who aim at responding to any request within two weeks. Once an agreement has been reached between the parties the data will be available to any type of research covered by the ethical permission under which the original study was carried out. Access will be granted for as long as needed to reasonably carry out the envisioned analysis.

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 included both males and females (see Table 1 and Table 2). In all analysis we considered sex as confounding variable and adjusted for sex.

Reporting on race, ethnicity, or other socially relevant groupings

We did not collect on ethnic background. All enrolled children were Ethiopian.

Population characteristics

The study included 1313 Ethiopian children under five years of age, of these, 650 children were presenting with diarrhea and 663 were frequency matched, non-diarrheal controls.

Recruitment

Children 0-59 months old with diarrhea were recruited at Jimma University Specialized Hospital (JUSH) and Serbo Health Centre (SHC). Non-diarrheal controls recruited from communities surrounding Jimma University Specialized Hospital, Jimma, Ethiopia and Serbo Health Centre, Jimma, Ethiopia were children without diarrhea in the previous 48-hours who were frequency matched in strata by residential district, age, and enrolment week. Details can be found: Johansen, Ø.H., et al. Performance and operational feasibility of two diagnostic tests for cryptosporidiosis in children (CRYPTO-POC): a clinical, prospective, diagnostic accuracy study. *Lancet Infect Dis* 21, 722-730 (2021).

Ethics oversight

The study received ethical approval from the Ethiopian National Ethics Review Committee, Ethiopia (Reference MoSHE/RD/14.1/7188/19), the Regional Committee for Medical and Health Research Ethics of Western Norway (2016/1096/REK Vest) and the Minimal Risk IRB (Health Sciences) University of Wisconsin, USA (ID: 2019-0842).

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

The present study is a substudy of a prospective case-control study (CRYPTO-POC). We used available stool samples from the parent study described in details here: Johansen, Ø.H., et al. Performance and operational feasibility of two diagnostic tests for cryptosporidiosis in children (CRYPTO-POC): a clinical, prospective, diagnostic accuracy study. *Lancet Infect Dis* 21, 722-730 (2021).

Data exclusions	Samples with less than 5000 sequencing reads (8 samples in total) were removed from the analysis.
Replication	The study is based on one (large = 1300+ children) cohort. Where possible bioinformatic analysis were carried out using several complementary methods as lined out in Materials and Methods.
Randomization	Not relevant as the study is based on a case-control design.
Blinding	Not relevant as the study is based on a case-control design.

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	Not relevant in our study.
Novel plant genotypes	Not relevant in our study.
Authentication	Not relevant in our study.