nature research

Corresponding author(s): Ami S. Bhatt

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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☒	☐	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used in the process of data collection.
Data analysis	Read cloud data were processed with 10X Genomics Longranger v2.1.3. Read cloud, synthetic long read and short read data were subjected to quality control using cutadapt v1.8.1. Data were assembled with MetaSPAdes v3.11.1, TruSPAdes 3.11.1, CANU v1.5, MEGAHIT v1.1.2, and Athena, the focus of the present manuscript. The Athena read cloud assembler is available at www.github.com/abishara/athena_meta. Assemblies were further analyzed with BWA v0.7.10, Metabat v2.12.1, Metaquast v4.6.0, CheckM v1.0.7, Prokka v1.12, Aragorn v1.2.36, Barrnap v0.7, and Kraken v0.10.6.

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/reviewers upon request. 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

The datasets generated during the current study will be available in the NCBI Sequence Read Archive under Bioproject accession PRJNA380276 upon publication.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

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

Sample size	No sample size calculation was performed. Four independent samples were chosen to evaluate the technical contributions presented in the manuscript. We selected four samples for the technical development and evaluation of this method to meet or exceed the precedence of prior publications that have developed methods for improved genome assembly (Kuleshov et al, NBT, 2015). The samples selected represent a spectrum of species richness that spans nearly two orders of magnitude and which reflects what is observed in most human stool samples and some environmental samples.
Data exclusions	No data were excluded from analysis.
Replication	Both human samples were sequenced in the presence and absence of size selection, as well as with DNA extraction by two alternative approaches. All samples yielded compositionally concordant results.
Randomization	Randomization was not relevant to our study as there were no groups to which participants could be assigned.
Blinding	Blinding was not relevant to our study as there were no differing selection or exclusion criteria between sampled participants.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Unique materials

Obtaining unique materials	The mock mixture DNA is available from ATCC (product MSA-1003). The ocean sediment sample and human stool samples are unique materials and can be made available through a material transfer agreement with the corresponding author.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Samples were obtained from human volunteers that have been consented under a Stanford IRB approved protocol.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging