

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 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

The tools described are at the following versions within KBase at the time of this publication. KBase apps are regularly evaluated for update, or upon request from users. If the app has been contributed by a member of the KBase community, KBase informs the community developer when their app needs updating and supports them in the update process. In the time from the initial development of the example Narratives for this study, we have updated metaSPAdes from v3.13.0 to v3.15.3, dbCAN2 from v6 to v10, GTDB-Tk from v1.1.0 to v1.7.0, and BLAST from v2.7.0 to v2.12.0. Except for metaSPAdes, as metagenome assembly is an expensive operation, we have re-run each analysis with the latest version of the software.

- FastQC v0.11.5 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.5_source.zip
- Trimmomatic v0.36 <http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.36.zip>
- Cutadapt v1.18
<https://github.com/marcelm/cutadapt>
- Kaiju v1.7.3
<https://github.com/bioinformatics-centre/kaiju>
- GOTTCHA2 v2.1.6
<https://github.com/LANL-Bioinformatics/kbase-gottcha2>
- MEGAHIT v1.2.9 https://github.com/voutcn/megahit/releases/download/v1.2.9/MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz
- IDBA-UD v1.1.3 <https://github.com/loneknightpy/idba/releases/download/1.1.3/idba-1.1.3.tar.gz>
- metaSPAdes v3.15.3
<http://cab.spbu.ru/files/release3.13.0/SPAdes-3.13.0-Linux.tar.gz>
- MaxBin2 v2.2.7 <https://sourceforge.net/projects/maxbin2/files/MaxBin-2.2.7.tar.gz/download>
- MetaBAT2 v1.7
<https://gitlab.com/jfroula/kbase-metabat>
- CONCOCT v1.1.0
<https://github.com/BinPro/CONCOCT>

- DAS-Tool v1.1.2
https://github.com/cmks/DAS_Tool/archive/1.1.2.tar.gz
- CheckM v1.0.18
<https://github.com/CheckM/CheckM>
- RASTtk v1.073
https://github.com/kbaseapps/RAST_SDK
- Prokka v1.14.5
<https://github.com/tseemann/prokka>
- DRAM v1.2.0
https://github.com/shafferm/kb_DRAM
- GTDB-Tk v1.7.0
<https://github.com/CheckM/GTDBTk>
- FastTree-2 v2.1.9
https://github.com/kbaseapps/kb_fasttree
- FastANI v1.1
<https://github.com/ParBLISS/FastANI>
- dRep v3.1.0
<https://github.com/MrOlm/drep>
- MUSCLE v3.8.425
http://drive5.com/muscle/muscle3.8.425_binaries.tar.gz
- BLAST+ v2.12.0
<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ncbi-blast-2.12.0+-x64-linux.tar.gz>
- HMMER3 v3.3.2
<http://eddylab.org/software/hmmer/hmmer-3.3.2.tar.gz>
- OrthoMCL v2.0.9
<http://orthomcl.org/common/downloads/software/v2.0/orthomclSoftware-v2.0.9.tar.gz>
- dbCAN2 v10
<http://bcb.unl.edu/dbCAN2/download/dbCAN-HMMdb-V10.txt>

Data analysis

All KBase code is Open Source under the MIT license and available from Github at <https://github.com/kbase> and <https://github.com/kbaseapps>. All externally developed software run in KBase is also Open Source by policy and available from their respective repositories, typically Github, Gitlab, Bitbucket, or Sourceforge (see Code Versions section in Materials above).

Source code for the following KBase Apps is in Github or Gitlab in the following repositories:

- Data Import Apps
https://github.com/kbaseapps/kb_uploadmethods
- Assess Read Quality with FastQC App
https://github.com/kbaseapps/kb_fastqc
- Trim Reads with Trimmomatic App
https://github.com/kbaseapps/kb_trimmomatic
- Cutadapt App
https://github.com/kbaseapps/kb_cutadapt
- Run the JGI RQCFilter pipeline App
<https://github.com/kbaseapps/BBTools>
- Classify Taxonomy of Metagenomic Reads with Kaiju App
https://github.com/kbaseapps/kb_kaiju
- Classify Taxonomy of Metagenomic Reads with GOTTCHA2 App
<https://github.com/LANL-Bioinformatics/kbase-gottcha2>
- Merge Reads Libraries App
https://github.com/kbaseapps/kb_ReadsUtilities
- Assemble Reads with metaSPAdes App
https://github.com/kbaseapps/kb_SPAdes
- Assemble Reads with MEGAHIT App
https://github.com/kbaseapps/kb_megahit
- Assemble Reads with IDBA-UD App
https://github.com/kbaseapps/kb_IDBA
- Filter Contigs by Length App
https://github.com/kbaseapps/kb_AssemblyUtilities
- Assess Quality of Assemblies with QUAST App
https://github.com/kbaseapps/kb_quast
- Compare Assembled Contig Distributions App
https://github.com/kbaseapps/kb_assembly_compare
- Bin Contigs using MaxBin2 App
https://github.com/kbaseapps/kb_maxbin
- MetaBAT2 Contig Binning App
<https://gitlab.com/jfroula/kbase-metabat>
- Bin Contigs using CONCOCT App
https://github.com/kbaseapps/kb_concoct
- Optimize Bacterial or Archaeal Binned Contigs using DAS Tool App
https://github.com/kbaseapps/kb_das_tool
- Assess Genome Quality with CheckM, Filter Bins by Quality with CheckM Apps
https://github.com/kbaseapps/kb_Msuite
- Extract Bins as Assemblies from BinnedContigs App
<https://github.com/kbaseapps/MetagenomeUtils>
- Annotate Genome/Assembly with RASTtk App
https://github.com/kbaseapps/RAST_SDK

- Annotate Assembly and Re-annotate Genomes with Prokka App
<https://github.com/kbaseapps/ProkkaAnnotation>
- Classify Microbes with GTDB-Tk App
https://github.com/kbaseapps/kb_gtdbtk
- Build Phylogenetic Tree from MSA using FastTree2 App
https://github.com/kbaseapps/kb_fasttree
- Insert Set of Genomes Into SpeciesTree App
<https://github.com/kbaseapps/SpeciesTreeBuilder>
- Build Microbial SpeciesTree and Build Gene Tree Apps
https://github.com/kbaseapps/kb_phylogenomics
- Compute ANI with FastANI
<https://github.com/kbaseapps/FastANI>
- Dereplicate genomes with dRep
https://github.com/kbaseapps/kb_dRep
- Annotate and Distill Assemblies with DRAM, Annotate and Distill Genomes with DRAM Apps
https://github.com/shafferm/kb_DRAM
- Search with dbCAN2 HMMs of CAZy families and Search with HMMs of MicroTrait Bioelement families Apps
https://github.com/kbaseapps/kb_hmmer
- Build AssemblySet, Batch Create AssemblySet, and Summarize GenomeSet Apps
https://github.com/kbaseapps/kb_SetUtilities

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 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 analyses and data discussed are available via the “dynamic” KBase Narratives <https://narrative.kbase.us/narrative/33233> (Compost) and <https://narrative.kbase.us/narrative/62384> (Moab Desert Crust). Additionally, “static” HTML narratives have been published on KBase [<https://docs.kbase.us/getting-started/narrative/share#publishing-a-static-narrative>] from each of these dynamic Narratives. They are available at <https://kbase.us/n/33233/628/> (Compost, DOI: 10.25982/33233.606/1831502) and <https://kbase.us/n/62384/334/> (Moab Desert Crust, DOI: 10.25982/62384.253/1831503). All input and derived data objects can be exported using standard formats from the Narratives by clicking on the given object, and then on the download arrow in the data panel in the upper left of the dynamic Narrative, as described at <https://docs.kbase.us/data/upload-download-guide/downloads>.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	Protocol for Metagenome-Assembled Genome extraction from microbiomes using KBase
Research sample	User uploads Illumina read libraries
Sampling strategy	Up to user
Data collection	Up to user
Timing and spatial scale	Up to user
Data exclusions	no data were excluded from the analyses
Reproducibility	Narrative notebooks are preserved and can be re-run to reproduce perfectly
Randomization	Random processes are provided with input seed for perfect reproducibility
Blinding	Blinding not relevant to study

Did the study involve field work? ☐ Yes ☒ No

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging