nature portfolio

Corresponding author(s): William Murphy

Last updated by author(s): Aug 9, 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 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	Genome sequence data was generated on Pacific Biosciences Sequel IIe and Illumina NovaSeq 6000 platforms using standard commercial software.
Data analysis	We used publicly available software/code for data analysis. We used the following programs and versions: FastQC v0.11.8, Trim Galore! v0.6.4, Canu v1.8, NextDenovo v2.2-beta.0, NextPolish v1.3.0, QUAST v5.0.2, BUSCO v4.0.6, mummer3.23, Merqury v1.3, Inspector v1.0.2, bwa mem v0.7.17, Samtools v1.9, BEDTools suite v2.30.0, R v.3.5.1, BLAST v2.9.0, RepeatMasker v4.0.9, seqTK subseq v1.3, SALSA v2.2, RagTag v1.0.1, Juicer v1.5.7, Juicebox v1.11.08, 3d-dna v.180922, Liftoff v1.4.2, Geneious Prime v2021.0.3, FlexiDot v1.06, Mafft Multiple Aligner v1.4.0, RAxML v8.2.11, Mesquite v3.61, FigTree v1.4.4, Mega-X v10.0.5, Tree House Explorer v1.0.2, RepeatMasker-4.1.2-p1, RepeatModeler 2 v2.0.2a, GenomeTools 1.6.2, LTR_Retriever v2.9.0, Ninja 1.10.2 , MAFFT 7.453, CD-HIT 4.8.1, hmmer3.3.2; cdhit4.8.1, ULTRA version 0.99.17, GRM, SEDEF, IQTREE v1.6.12, ReLERNNv1.0.0, PAV, minimap2 v2.24, Long-Read Aligner v1.3.2, PBSV v2.8.0, pbmm2 v1.9.0, Sniffles v2.0.7, SDA, SafFire, Pangenie v1.0.1, pybedtools v0.9.0, SciPy v1.7.3, pandas v1.4.0, biopython v1.79

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

All genome assemblies are available in NCBI, under the following accession numbers: GCA_016509475.2, GCA_016509815.2, GCA_018350155.1, GCA_018350175.1, GCA_018350195.2, GCA_018350215.1. OR gene sequences are found at: https://figshare.com/s/68266360874d5078bdf5

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	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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 statistical method was used to predetermine sample size
Data exclusions	No data was excluded in our analyses.
Replication	All attempts to replicate results were successful
Randomization	The experiments were not randomized
Blinding	The investigators were not blinded to allocation during experiments and outcome assessment.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The single haplotype genome assemblies were derived from DNA sequencing of primary skin fibroblast cell lines.
Authentication	All cell lines were karyotyped using standard methodology from early passages (p1-p3) and verified by comparison to known karyotypes and analyzed with PacBio long read technology to confirm species id.
Mycoplasma contamination	The cells were not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	N/A