

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure 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 RNA-seq and small RNA data reported in the manuscript were sequenced at the New York Genome Center. Public RNA-seq and small RNA datasets utilized in the manuscript were downloaded from Gene Expression Omnibus database.

Data analysis RNA-seq data reported in this study was mapped to PacBio genome assemblies from simulans clade (Chakraborty et al. 2021) using HiSat2 aligner. bam2wig.py script from RSeQC package was used to convert resulting bam files from mapping to wig format for visualization. Similarly, small RNA data was mapped to PacBio assemblies from simulans clade using bowtie (with $-v0$ $-best$ $-strata$ options.) The resulting small RNA alignments in SAM format were converted to BED for downstream processing using the BEDOps software and visualized on IGV. Custom shell scripts to run RNA-seq and small RNA mapping pipeline are provided at <https://github.com/Lai-Lab-Sloan-Kettering>

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](#)

RNA-seq and small RNA data generated for this study are now available on the GEO database with the following accession number: GSE185361. Public RNA-seq

datasets used in this study: SRR7961899, SRR8194694, and SRR8194697. Public small RNA datasets used in this study: SRR7961896, SRR7961895, SRR2046463, SRR2549640, SRR2549641, SRR2549643, and SRR989483

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](https://www.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	For RNA-seq and small RNA data in simulans clade, we collected testis sample from 7 days old adult males. We collected 2 replicates per species for RNA-seq and small RNA library preparation.
Data exclusions	We did not exclude any data.
Replication	The study design did not involve mutant samples or differential gene expression analysis. However, we collected two biological replicates for RNA-seq and small RNA data.
Randomization	Not applicable
Blinding	Not applicable

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