

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. [For final submission](#): please carefully check your responses for accuracy; you will not be able to make changes later.

▶ Experimental design

1. Sample size

Describe how sample size was determined.

We chose 35 eukaryote genomes to train our models based on balanced sampling across available eukaryote diversity.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

The computer code used to build models described in the study will be freely available pending publication of the manuscript. The code can use the models as reported on the same or novel genomes, and can regenerate the models given the same input parameters, or can generate new models given altered inputs. All computational predictions were verified by 1000 times repeated 10-fold cross-validation. That verification is reported in the manuscript.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Organisms were allocated based on known trophic mode. Test genomes were chosen based on availability.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Researchers were not blinded to group allocation as the study represents a form of supervised machine learning.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The <u>exact sample size</u> (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A statement indicating how many times each experiment was replicated |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used and whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Test values indicating whether an effect is present
<i>Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Clearly defined error bars in <u>all</u> relevant figure captions (with explicit mention of central tendency and variation) |

See the web collection on [statistics for biologists](#) for further resources and guidance.

► Software

Policy information about [availability of computer code](#)

7. Software

Describe the software used to analyze the data in this study.

Custom code was written and is part of the focus of the study. That code is available with the manuscript as a computational tool.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* [guidance for providing algorithms and software for publication](#) provides further information on this topic.

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

no unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

no antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

no eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

no eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

no eukaryotic cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

no cell lines were used

► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

no animals were used

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

this study did not involve human research participants