

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](#).

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

No software was used for data collection

Data analysis

Custom python code (using Python version 3.6.6) was used for data analysis and is provided as Supplementary Computer Code (sole documented version therefore unnumbered).

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. 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

All raw data supporting the results are publicly available (Paleobiology Database), secondary data and statistics can be reproduced using the provided supplementary computer code.

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)

Ecological, evolutionary & environmental sciences study design

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

Study description	Machine learning analysis of raw fossil occurrence data for 171,231 independent species downloaded from the public Palaeobiology Database
Research sample	1,273,254 fossil occurrence records of 71,231 independent species downloaded from the public Palaeobiology Database
Sampling strategy	Exhaustive sampling of the databased fossil occurrences from 1000 to 0 million years ago
Data collection	Data were downloaded on 2020-03-19 from http://paleobiodb.org/data1.2/occs/list.csv?datainfo&rowcount&taxon_reso=species&max_ma=1000&min_ma=0&show=class,aconly,coords,paleoloc,env,aconly . J.F.H.C downloaded data from a public repository. No instruments were used to personally record data for this study.
Timing and spatial scale	1000 to 0 million years ago, all global occurrences
Data exclusions	No data were excluded from the main analyses. One repeated statistical analysis excluded recent data to aid identification of an effect as documented in the captions to the relevant table of statistical results (Extended Data Table 1). These data were excluded in order to isolate hypothesised effect after initial data analysis.
Reproducibility	All data analyses were automatically scripted and can be reproduced by re-running the provided computer code. Statistical reproducibility of the machine learning analyses was formally and successfully tested using a bootstrap analysis as described in the paper.
Randomization	This is not relevant to this study which analysed all available data without splitting these into groups
Blinding	This is not relevant to this study which does not involve group allocation or statistical comparison
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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Palaeontology

Specimen provenance	This is not relevant to this study on aggregated fossil occurrence records
Specimen deposition	This is not relevant to this study as no fossil specimens were deposited
Dating methods	This is not relevant to this study as no new dates were obtained
☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	