

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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ 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
<i>Give P values as exact values whenever suitable.</i> |
| ☐ | ☒ 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 specific software was used in data collection.
Data analysis	The BEAST v2.4.8 software can be downloaded from https://www.beast2.org/. The Babel package is a BEAST package for performing linguistic analysis (URL: https://github.com/rbouckaert/Babel). The BayesTrait program can be downloaded from http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html. The SPLITSTREE v4.16.6 was used which can be downloaded from http://www.splittree.org/ The Matlab codes used in this study are available in the supplementary files ('Matlab codes for estimation of ST evolutionary tempo.zip' and 'Matlab codes for Four-Point analysis.zip').

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

The data supporting the findings of this study are available in the supplementary information files.

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	We used 109 Sino-Tibetan language samples with 949 binary-coded lexical Root-meanings in this study. To the extent that it is possible to increase this sample size, we do not expect this will improve the accuracy or precision of our date and geographic homeland estimates, because our sample already represents all the available languages in the database.
Data exclusions	The languages and lexical root-meanings used in this study came from the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) project (http://stedt.berkeley.edu/). 1. We manually checked the word entries with a STEDT-labelled etymon and a meaning within the range of Swadesh 100 words, and excluded the incomplete etymon annotation and the inexhaustive investigation of a language. 2. We filtered only those languages with at least 90 lexical meanings of Swadesh 100 word-list recorded and 30 – 120 Root-Meanings (RMs). The reason for this filtering process was that if too few meanings or RMs existed for a language, we considered the language was either not well investigated in the literatures or not well labelled for etyma in the database. And too many RMs might indicate that the language was excessively recorded. 3. Furthermore, we removed those language that were possibly duplicated, or commonly believed to have suffered from strong lateral transfers (or horizontal language influence) among the Sino-Tibetan language family, or could not be located to a certain time point. 4. Finally, we retained 109 ST language samples with 949 binary-coded lexical RMs from STEDT database.
Replication	Findings were reproduced under a range of different model assumptions as reported in the manuscript.
Randomization	This is not relevant to our study because it does not include an experimental treatment.
Blinding	This is not relevant to our study because it does not include an experimental treatment.

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