

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

Data analysis

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](#)

The newly sequenced genomic data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) GenBank database under BioProject PRJNA1177432 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1177432>) and the China National Microbiology Data Center (NMDC, <https://nmdc.cn/en>) under BioProject NMDC10018536 (<https://nmdc.cn/resource/genomics/project/detail/NMDC10018536>). Background information of newly sequenced strains is listed in Supplementary Table 1. The publicly available data used in this study are available in the NCBI GenBank and Sequence Read Archive (SRA) databases, with accession numbers provided in Supplementary Data 2. Source data are provided with this paper.

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	Sex and gender were not considered in the study design
Reporting on race, ethnicity, or other socially relevant groupings	Race, ethnicity, or other socially relevant groupings were not used in our study
Population characteristics	In our study, we focus solely on the isolation sources of <i>Y. pestis</i> strains, differentiating between rodents, vectors, humans, or environmental samples (like soil, etc.). Population characteristics were not considered in our analysis.
Recruitment	Given that plague is classified as a Class A infectious disease in China, annual surveillance of natural plague foci is conducted. The 2,336 newly sequenced <i>Y. pestis</i> strains were collected from these surveillance activities. There was no active recruitment involved.
Ethics oversight	Our study did not involve ethical oversight.

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	In this study, we analyzed 3,318 <i>Y. pestis</i> genomes, including 2,336 newly sequenced strains and 982 genomes publicly available from the NCBI SRA database. The newly sequenced strains were obtained from natural plague foci across 14 provinces in China between 1948 and 2011, with the exception of two strains from Myanmar and 13 of unknown origin. Additionally, by adding 164 GenBank assemblies without SRA data and 93 ancient genomes from the SRA, we analyzed a total of 3,575 strains to construct the most comprehensive phylogenetic tree for hierarchical nomenclature analysis.
Data exclusions	For publicly available genomes in NCBI databases, we excluded data of lab-passaged strains, genomes with less than 20X sequencing depth, and strains with duplicate BioSample IDs.
Replication	The sequencing data and software used for data analysis are publicly available, and the custom scripts along with the processed data can be found on GitHub. Details of the analysis are described in the Methods section.
Randomization	Strains were assigned to specific phylogroups based on genomic similarity, not randomly. The Methods section provides detailed information on the random sampling process used for identifying and statistically characterizing hot regions.
Blinding	Blinding was not relevant to the study design, as the strains were collected during annual surveillance of all sylvatic plague foci in China.

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

Methods

- | n/a | Involvement in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

- | n/a | Involvement in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.