nature portfolio

Corresponding author(s): George F. Gao, Yanan Wang

Last updated by author(s): Apr 13, 2025

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	wget 1.21.4, SISTR v1.1.2, SeqSero2 v1.3.1, MLST v2.0, ResFinder 4.1, PlasmidFinder v2.1, PointFinder, BLAST, etc. Full details are in Methods section.
Data analysis	GraphPad Prism v9.0, excel 2021, MapChart, Parsnp v2.0.5, iTOL v6.9, ImageGP platform, office 2021, etc. The software parameters used in this study refer to our previous studies, analysis is available in the following gitee repository: https://gitee.com/wangyanan999/salmonella_enterica, https://github.com/NANYW123/Salmonella-genome-database-CLSGDB-v2. For AMR levels in different sources, we mainly focused on the following categories: chicken, pigs, cattle, turkey, aquatic animals, food, wild animals, humans, and the environment. The linear regression models were used to measure the association between indicators and AMR genes per genome over time, adjusting for 12 representative country-level data available across six continents ranging from 2000 to 2023.

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

A total of 471,617 genomes used in this study are available from the NCBI under the BioProject number PRJNA766315 and public database from the NCBI assembly (https://www.ncbi.nlm.nih.gov/assembly, as of June 17, 2023). The Illumina sequence data generated by us (PRJNA766315) also have been deposited in the National Microbiology Data Center (NMDC) under BioBroject number: NMDC10017893 and NMDC10018145. The Chinese local Salmonella genome database version 2 (CLSGDB v2) we built in previous work including 7997 genomes (freely available at https://nmdc.cn/clsgdbv2) were also included. The reference sequence (a point mutation (−44G>T) in the csgD gene promoter, upregulation of the csgDEFG operons) for MRDRST was obtained from the previous study.

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	In this study, gender was determined based on self-reporting. Participants of both sexes were include.
Reporting on race, ethnicity, or other socially relevant groupings	All genomes from the six continents used in this study were downloaded from the NCBI, and all metadata information was based on their original information.
Population characteristics	This work did not include any participants. Not applicable for this study.
Recruitment	This work did not include any participants, all the genomes used in this study were downloaded from the NCBI.
Ethics oversight	This work did not include any participants. Not applicable for this study.

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	A total of 471,617 publicly available Salmonella genomes were downloaded (as of June 17, 2023), after screening by both bioinformatic software (including STSTR, SeqSero2, and MLST) and clear metadata, a total of 208,233 genomes were obtained. A total of 708 serovars and 3,006 sequence types (STs) were detected.
Data exclusions	First, we screened and summarized the metadata of publicly available Salmonella genomes from the NCBI BioSample (as of June 17, 2023), and isolates with information on the collection time, isolation source, and sampling sites are used for next step analysis. A total of 253,872 genomes were discarded because no background information on the collection time, isolation source, and sampling sites was simultaneously recorded. Second, in silico serotyping was carried out by both the Salmonella in silico typing resource (SISTR v1.1.2) and Salmonella Serotyping by whole genome sequencing (SeqSero2 v1.3.1). Only the prediction result matching between the SeqSero2 and SISTR is used for downstream analysis. Third, multilocus sequence typing (MLST v2.0) was used to assign sequence types (STs) to these genomes based on seven housekeeping genes for examining bacterial relatedness. A total of 9,512 genomes were discarded because no serotype or sequence typing information was predicted.
Replication	In this work, we used the established analysis pipeline which has been used to analyze partial data of the genome (https://gitee.com/wangyanan999/salmonella_enterica, https://github.com/NANYW123/Salmonella-genome-database-CLSGDB-v2).
Randomization	This work did not include any interventions and thus the randomization (as used in clinical trials or intervention studies) was not appropriate.
Blinding	This work did not include any interventions and thus the conventional blinding (as used in clinical trials or intervention studies was not appropriate. To obtain a high-quality Salmonella genome dataset, background information for each isolate, sampling time, isolation source (host), geographic location (country and continent), serovar, and ST were screened.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ 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, mosiacism, off-target gene editing) were examined.