nature portfolio

Corresponding author(s): Min Yue

Last updated by author(s): 2026-2-13

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	This study comprises two datasets: one containing 11,361 genomes of invasive non-typhoidal Salmonella and another containing 3,115 ST313 genomes. Data were collected from GenBank, EnteroBase (data cutoff: March 1, 2024), and relevant literature sources. After deduplication, the species of each strain were confirmed using KmerFinder v3.2. Genomic quality control was performed using Python 3. The serovar type of each Salmonella genome was determined using SISTR v1.1.2 and SeqSero v2. Additionally, the sequence types (ST) of each Salmonella genome were identified using MLST v2.22 with the senterica_achtman_2 scheme.
Data analysis	The data analysis for this study was conducted using the following software tools: Snippy v4.4.5 was utilized to identify core-genome single nucleotide polymorphisms (SNPs) and pseudogene-associated SNPs. Gubbins v2.3.4 was employed to calculate recombination regions. IQ-TREE v1.6.2 was used to construct the basic phylogenetic tree of ST313 using the best-fit model: TVM+F+ASC+R2. iTOL v6.0 was applied to visualize the phylogenetic tree generated by IQ-TREE. RhierBAPS v1.1.4 was used to determine the lineage types of different ST313 isolates. Population structure was calculated using R v4.3.1, supplemented by the R packages "phytools" v2.0.3, "ape" v5.7.1, and "rhierbaps" v1.1.4. TreeTime v0.11 assessed the temporal structure by performing regression analysis of root-to-tip branch distances on the maximum likelihood tree. BactDating v1.1 was employed to further detect temporal signals in ST313 through randomization tests. BEAST v2.5 was used to analyze the divergence times of different ST313 lineages, perform phylogeographical reconstruction. NSLogAnalyser determined the optimal molecular clock type and model by comparing marginal likelihood values across different models. Tracer v1.7.2 evaluated the convergence of BEAST runs and infer effective population sizes. TreeAnnotator v2.6.7 generated maximum clade credibility trees.

Key divergence time points (with 95% highest posterior density [HPD]) were estimated, and the phylogenetic tree was visualized using FigTree v1.4.3.
SNP-dists v0.8.2 calculated SNP distances between pairs of isolates.
Clustering of isolates based on SNP distances was performed using complete-linkage hierarchical clustering in Python 3 with SciPy v1.14.1.
MOB-suite v3.1.9 identified plasmids present in ST313 genomes and determined their types.
BacAnt v3.4.0 detected integrons and transposons carried by Salmonella.
The Phastest pipeline identified prophages carried by Salmonella.
Abricate v1.0.1 with ResFinder detected antimicrobial resistance genes.
Staramr v0.10.0 identified antimicrobial resistance-associated point mutations.
Prokka v1.14.6 was used for genome annotation.
Pan-genome construction was carried out using Panaroo v1.5.1.
Scoary v1.6.16 identified lineage-specific differentially present genes.
Functional annotation of key differentially present genes was conducted using eggNOG-mapper with the eggNOG v5 database.
Statistical analyses were performed using SciPy v1.14.1.
The code used for calculating the invasiveness index is publicly available at: https://github.com/Gardner-BinfLab/invasive_salmonella.

Additionally, this study includes three pipelines developed in our laboratory, with the corresponding code publicly available on GitHub:
The pipeline for calculating the horizontal gene transfer frequencies of antimicrobial resistance genes is available at: https://github.com/tjiaa/Cal_HGT_Frequency. This method has been previously published in eLife (https://elifesciences.org/articles/101241).
Scripts for calculating the co-localization between ARGs and mobile genetic elements are available at: https://github.com/tjiaa/INTS_Code.
The script used to calculate potential transmission events based on SNP distance and strain isolation dates can be found at: https://github.com/tjiaa/SNP-Distance-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

This study analyzed genomic data from 11,361 iNTS isolates. The accession numbers and their corresponding database sources are provided in Supplementary Data 1. Additionally, a dataset comprising 3,115 ST313 genomes was utilized, with accession numbers listed in Supplementary Data 2. The per capita GDP data for different countries in 2024 are sourced from the World Bank (https://data.worldbank.org). 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	In this study, we examined the prevalence of Salmonella ST313 among different sexes, considering sex as a biological factor. Classification into male and female groups was based on data reported by public databases. The collected sex information is provided in the Supplementary Table 2, with 226 male infections and 214 female infections recorded. We conducted sex- and gender-based analyses wherever possible. In cases where such analyses were not performed, we have provided justifications for their absence.
Reporting on race, ethnicity, or other socially relevant groupings	We used per capita GDP data from different countries to assess sampling bias. The 2024 per capita GDP data were sourced from the World Bank (https://data.worldbank.org).
Population characteristics	In this study, we did not recruit human participants.
Recruitment	Our study did not involve human participants. Therefore, there were no recruitment processes or potential self-selection biases to consider. As a result, there are no biases related to participant recruitment that could impact the results of this study.
Ethics oversight	Not necessary in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Life sciences study design

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

Sample size	In this genomic epidemiology study, invasive non-typhoidal Salmonella (iNTS) was defined as non-typhoidal and non-paratyphoidal Salmonella isolates recovered from human specimens of non-fecal origin, including blood, urine, abscesses, and other normally sterile sites. We subsequently retrieved all publicly available Salmonella genomes from EnteroBase and GenBank (data cutoff: 1 March 2024; total, 488,232 genomes). Strains with ambiguous isolation sources—for example, those labeled only as "clinical" or "human"—were excluded. Only isolates with clearly defined sources that met the above criteria for iNTS were retained. Information on country and date of isolation was also collected where available. This filtering procedure yielded a final dataset of 11,361 iNTS genomes (Supplementary Data 1). On this basis, we further constructed a global dataset comprising 3,115 ST313 genomes (Supplementary Data 2), representing all publicly available ST313 genomes retrieved as of 1 March 2024. This dataset included not only human-derived ST313 isolates but also those obtained from animals, plants, and isolates annotated with an "unknown" source. Genome quality was assessed in accordance with the European Reference Laboratory for Salmonella genomic quality control standards.
Data exclusions	Genome data from low-quality strains that did not meet the aforementioned genomic quality control standards were excluded.
Replication	This study relies on bioinformatic analysis of genomic data. No biological replicates were generated as no new wet-lab experiments were conducted. However, to ensure computational reproducibility and statistical robustness: Phylogenetic trees were inferred using 1,000 bootstrap replicates to assess branch support.
Randomization	Randomization was not applicable to this study as the research as isolates were included based on a sampling strategy with no predetermined groups defined before the study commenced.
Blinding	Blinding was not relevant because sources of bias that could be avoided by blinding did not play a role in our investigation, i.e. all results were objectively derived from genomic analysis.

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	No plants used in this study.
Novel plant genotypes	No plants used in this study.
Authentication	No plants used in this study.