

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                                                                                                                                                                                                                                                                                      |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                                          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                                   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                                      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), 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.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Data analysis   | <p>Dataset quality control and curation was performed with Aquamis v1.3.9. Software involved in WGS data analysis include chewieSnake v3.1.1, chewBBACA v2.8.5, MentaliST v1.0.0, SeqSphere+ v8.3.4 (Ridom), BioNumerics v8.1 (Biomérieux), snippySnake v1.1.0, CSI Phylogeny v1.4, SnapperDB v1.0.6, WGSBAC v2.2, cultevo package v1.0.2, stats package v 0.1.0.</p> <p>Output harmonization and clustering at all possible resolution levels was performed with ReporTree v1.0.1.</p> <p>In silico traditional typing was performed with mlst v2.22.0 and SeqSero2 v1.1.1 or patho_typing v1.0. PopPUNK v2.6.5 was used to determine the main genomic lineages of <i>E. coli</i> and <i>C. jejuni</i> datasets.</p> <p>The script <code>comparing_partitions_v2.py</code> (<a href="https://github.com/insapathogenomics/ComparingPartitions">https://github.com/insapathogenomics/ComparingPartitions</a>) implemented in ReporTree v1.0.1 was used for identification of stability regions and assessment of cluster congruence at all possible threshold levels. The script <code>get_best_part_correspondence.py</code> was used for identification of inter-pipeline corresponding points. The scripts <code>comparison_outbreak_level.py</code>, <code>stats_outbreak_analysis.py</code>, <code>stats_outbreak_analysis_snp_dists.py</code> and <code>wgmlst_exercise.py</code> were used for pipeline comparison at outbreak level. These scripts are available at: <a href="https://github.com/insapathogenomics/WGS_cluster_congruence">https://github.com/insapathogenomics/WGS_cluster_congruence</a>.</p> <p>INNUca v4.2.2 and chewBBACA v3.3.5 were used as alternative software for assessing the impact of modifying pipeline components.</p> |

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](#)

Anonymized sequencing reads of the “BeONE dataset” are deposited in the European Nucleotide Archive (ENA) database under the BioProjects PRJEB57166, PRJEB57179, PRJEB57098 and PRJEB57119. Genome assemblies are deposited in the Zenodo repository (L. monocytogenes: 10.5281/ZENODO.7267486; S. enterica: 10.5281/ZENODO.7267785; E. coli: 10.5281/ZENODO.7267844; C. jejuni: 10.5281/ZENODO.7267879). The public dataset data was retrieved from Zenodo (L. monocytogenes: 10.5281/ZENODO.7116878; S. enterica: 10.5281/ZENODO.7119735; E. coli: 10.5281/ZENODO.7120057; C. jejuni: 10.5281/ZENODO.7120166). The collection of scripts used to conduct these analyses are available at the github repository [https://github.com/insapathogenomics/WGS\\_cluster\\_congruence](https://github.com/insapathogenomics/WGS_cluster_congruence). Supplementary data are available in the Zenodo repository 10.5281/zenodo.12805750.

## 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 Not applicable. This study is not based on the recruitment of human participants.

Reporting on race, ethnicity, or other socially relevant groupings Not applicable. This study is not based on the recruitment of human participants.

Population characteristics Not applicable. This study is not based on the recruitment of human participants.

Recruitment Not applicable. This study is not based on the recruitment of human participants.

Ethics oversight Not applicable. This study is not based on the recruitment of human participants.

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](https://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 order to accomplish the objectives of this study, we aimed to compile a diverse dataset that captures the genomic diversity within the populations of L. monocytogenes, S. enterica, E. coli and C. jejuni. To this end, the Institutes involved in the study shared anonymized sequencing reads under a Material Transfer Agreement. In order to ensure the dataset diversity, we complemented this shared dataset with publicly available WGS data for the four species of interest, and for which the QC and assembly followed the exact same methodology. The method for compiling this public dataset was described in a previous publication (10.1186/s13073-023-01196-1). Briefly, we collected information about the genetic diversity (in terms of STs or serotype) of each of the four pathogens of interest in public databases (as of November 2021), and then selected public WGS data of previous works aiming at obtaining a similar diversity in the final dataset.

Data exclusions WGS data that did not pass the predefined QC criteria (e.g. low coverage or contamination) were excluded from the datasets.

Replication The findings of this study rely on the independent analysis of four large datasets (>2000 isolates each) by each institute in order to perform a comprehensive inter-laboratory cluster congruence analysis. The datasets are publicly available, and the software parameters and versions used in this study are indicated throughout the manuscript. Therefore, these results are reproducible.

Randomization Not applicable. As described in the manuscript, depending on the species or analysis, the pipelines under comparison used the same dataset as input to assess inter-pipeline comparability.

Blinding Blinding is not applicable. All samples were analysed strictly for genomic purposes.

## 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                                  |
| <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                        |

## Methods

|                                     |                                                 |
|-------------------------------------|-------------------------------------------------|
| n/a                                 | Involved 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.