

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 (<i>n</i>) 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> 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 <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	All sequencing data were collected from NCBI database (https://www.ncbi.nlm.nih.gov/)
Data analysis	Data were processed using diverse bioinformatic tools, as indicated both in the manuscript and in the GitHub repository (https://github.com/PaulaRamiro/NpAUREO/tree/main). Data was analyzed using Entrez Direct (v20.6), SRA-toolkit (v2.11.3), CoverM (v0.6.1), COPLA (v1.0), mob_suite (v3.1.8). Datasets were analyzed using R software version 4.1.2.

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 manuscript includes a complete data availability statement

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 NA

Reporting on race, ethnicity, or other socially relevant groupings NA

Population characteristics NA

Recruitment NA

Ethics oversight NA

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://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

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

Study description	We developed a custom bioinformatic pipeline that leverages extensive DNA sequencing data extracted from the NCBI database (https://www.ncbi.nlm.nih.gov/) to calculate the plasmid copy number for 6,327 phylogenetically diverse plasmids.
Research sample	In this study, we focused on genome assemblies and their associated plasmids from nine different bacterial genera from the phyla Pseudomonadota and Bacillota (i.e., Gram-negative and -positive bacteria). Specifically, we analyzed assemblies from the following genera: Acinetobacter, Bacillus, Enterobacter, Enterococcus, Escherichia, Klebsiella, Pseudomonas, Salmonella and Staphylococcus.
Sampling strategy	We selected complete and high-quality genome assemblies from publicly available databases, focusing on bacterial genera with clinical relevance. Our selection criteria included genome completeness, assembly quality, and the presence of plasmid sequences, ensuring a representative dataset for the studied taxa.
Data collection	Genomic data were retrieved from publicly available repositories, such as NCBI RefSeq and GenBank. Metadata associated with each assembly, including taxonomic classification and sequencing quality metrics, were used to ensure consistency and accuracy in data selection.
Timing and spatial scale	All data used in this study were obtained from publicly available genomic databases. Therefore, the timing and spatial scale of data collection are determined by the original sources from which the assemblies were retrieved.
Data exclusions	Assemblies that did not meet the criteria of completeness and high-quality or lacked associated plasmid data were excluded from the analysis.
Reproducibility	All analyses were conducted using publicly available genomic data and standardized bioinformatic pipelines, ensuring full reproducibility. The datasets and scripts used in this study are available to facilitate replication of our results.
Randomization	Since this study is based on publicly available genomic data, no experimental groups were assigned, and randomization was not applicable.

Blinding

Blinding was not applicable, as the study relied entirely on computational analyses of genomic data from public repositories.

Did the study involve field work?

☐ Yes

☒ No

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	NA
Novel plant genotypes	NA
Authentication	NA