nature portfolio

Corresponding author(s): Alessandra Luchini

Last updated by author(s): 1/17/2025

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	The mass spectrometry data were collected using Proteome Discoverer 3.1 (Thermo Fisher).
Data analysis	Data were analyzed using R software, a published algorithm called Bacphlip (PMID: 33996289) available on Github (https://github.com/adamhockenberry/bacphlip)

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

Proteomics data are available on the PRIDE archive at https://www.ebi.ac.uk/pride/archive. Materials used in this manuscript will be made available upon inquiry to aluchini@gmu.edu.

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	Urine proteomic analysis was performed on 30 volunteers. Sex (biological attribute) is reported in Table 1 of the manuscript. 30 participants were enrolled in the study, 17 self-reported being female and 13 males. A disaggregate analysis of sex and gender was not performed. Sex based analysis was not performed because the main goal of the study was to derive information about the interaction between bacteria and phages in the urinary tract system from an overall perspective.
Reporting on race, ethnicity, or other socially relevant groupings	Socially constructed and social relevant categorization variables were not used in this study. The demographic variables reported include age and sex. Race, ethnicity were not reported.
Population characteristics	The population characteristic relevant for this study is a clinical diagnosis of urinary tract infection, which was provided by the treating urologist. Participants are either positive or negative for urinary tract infection. Age and sex are reported but not included in the analysis.
Recruitment	Patients were recruited by their treating physician. Proper randomization with random sampling was implemented to minimize selection bias.
Ethics oversight	The George Mason University institutional review board reviewed 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	The study includes a limma analysis to identify differentially abundant proteins (Figs. 1 and 2). The Benjamini and Hochberg method to adjust the p value was used. There are three classes (urinary tract infection N=10, in country non urinary tract infection N=10, and non-urinary tract infection in the US N=10). The number of observations was dictated by sample availability. We calculated the average power for the given sample size. We used the check.power function in the ssizeRNA package in R and published in http://dx.doi.org/10.1186/s12859-016-0994-9. Our data parameters were as follows: proportion of non-differentially abundant proteins = 0.93, number of proteins = 798, mean counts in control group for simulation = 10, dispersion parameter = 0.1, fold change = 2, FDR = 0.05. The average power for the given sample size was 0.95.
Data exclusions	No data were excluded from the analysis.
Replication	Steps to ensure reproducibility of experimental findings included the spike in of an exogenous control in the protein mixture for mass spectrometry analysis (chicken lysozyme), and the verification of the lysogenic phenotype of phages in culture using the agar overlay technique.
Randomization	Allocation of participants to classes was based on a clinical diagnosis of urinary tract infection. No treatment was administered to participants. Participants donated urine under informed consent. The urine samples were anonymized and sent to the laboratory in dry ice.
Blinding	Laboratory scientists were blinded to the group allocation when the experimental analysis was run. The scientist responsible for data analysis used the class labels to conduct data analysis without revealing the group allocation to laboratory scientists.

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.