nature portfolio

Corresponding author(s): Lei Zhang, Guoping Zhao

Last updated by author(s): Aug 5, 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	No software was used for data collection.
Data analysis	The AMP generation and prediction codes can be found at https://github.com/W1V1995/AMP_Project. In this study we used CD-HIT (version 4.8.1), scikit-learn (version 1.2.2), FlowJo (version 10.8.1), Biorender, Origin2024Sr1No_H, QIIME2 HTSeq v0.6.1, R (version 4.3.2), DESeq2 (version 1.42.1), ggplot2 (version 3.5.0), vegan (version 2.6.4), cluster (version 2.1.4), pairwiseAdonis (version 0.4.1), tidyr (version 1.3.0), tidyverse (version 2.0.0), ggsci (version 3.0.0), pheatmap (version 1.0.12), stringr (version 1.5.1), ggpmisc (version 0.6.1), ape (version 5.7-1), phyloseq (version 1.40.1), GUniFrac (version 1.8), ade4 (version 1.7-22), clusterProfiler (version 4.10.0), topGO (version 2.54.0) and Rgraphviz (version 2.46.0).

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

Our study contains publicly available proteome, AMP, non-AMP, Toxin and non-Toxin, generated-sequences, 16S rRNA gene sequencing and RNA-seq data. Protein sequences were retrieved from UniProt (http://www.uniprot.org/) (downloaded as of October 2022). AMP data were mainly collected from six public AMP datasets — APD3: https://aps.unmc.edu/AP/, DBAASP: https://www.dbaasp.org/home, DRAMP: http://dramp.cpu-bioinfor.org/, CAMP: http://www.camp3.bicnirrh.res.in/, LAMP: http://biotechlab.fudan.edu.cn/database/lamp, and BaAMPs: http://www.baamps.it, which cover most of AMP sequences from various origins (downloaded as of November 2022). The non-AMP dataset was downloaded from UniProt (https://www.uniprot.org) by setting the 'subcellular location' filter to cytoplasm and removing any entry that matches the following keywords: antimicrobial, antibiotic, antiviral, antifungal, effector or excreted in their functional annotations (downloaded as of November 2022). Toxin and non-Toxin data was collected from https://server.wei-group.net/ToxIBTL/ (downloaded as of April 2023). The relevant data used for model construction, the constructed and generated short peptide datasets, as well as the prediction results had been uploaded to https://zenodo.org/records/16633186. Sequence data generated by five unconstrained generation models were collected from https://zenodo.org/records/7420189#.ZBCo4JHMKUk (downloaded as of October 2024). And the 16S rRNA gene sequencing and RNA-seq data are available under The National Omics Data Encyclopedia (NODE) accession number OEP00006445 and OEP00005095. 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	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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	Sample sizes were estimated on the basis of previous literature (https://doi.org/10.1016/j.cell.2024.07.027; https://doi.org/10.1016/j.cell.2024.05.013; https://doi.org/10.1038/s41551-022-00991-2; https://doi.org/10.1038/s41587-022-01226-0) on similar experimental settings (6-10 mice each group to avoid random error).

Data exclusions	No data were excluded from the analysis.
Replication	MIC and other determinations were replicated at least three times; Resistance to proteolytic degradation assays were replicated two times;all attempts at replications were successful.
Randomization	Mice were randomly allocated in each group receiving treatments.
Blinding	Histopathological diagnosis of tissues from treated mice was performed in a blinded fashion. The remainder of the data acquired did not lend itself to blinding or did not include group-allocation data. Blinding was not relevant for experiments that do not involve subjective assessments.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293 cells were purchased from FUNHENG BIOLOGY (FH0242, Shanghai, China).
Authentication	Cells were authenticated using STR profiling by the vendor and no further authentication was performed in the laboratory
Mycoplasma contamination	Cell line were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No such lines were used in this study.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	5-week-old male BALB/c mice were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. (Beijing, China) and kept under a 12h light/12h dark cycle, humidity of 50% and temperature of 22? in standard specific-pathogen-free (SPF) individually vented cages.
Wild animals	NA
Reporting on sex	Male
Field-collected samples	NA
Ethics oversight	All animal experiments were performed according to the 'Principles of Laboratory Animal Care' (NIH publication No. 86–23, revised 1985) and approved by the Animal Care and Use Committee of Shandong University (LL20240622, Jinan, China).

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

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.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Single colonies of strains were inoculated and cultured to the exponential phase, followed by three washes with 10 mM PBS (pH 7.0), and then adjusted their OD625 values to 0.08~0.13 with PBS. Subsequently, 150 µL of the bacterial suspension was incubated at 37? with 50 µL AMP (dissolved in PBS) for 2h. PI was added with the final concentration was 50 µg mL-1, and the mixture was incubated in the dark at 37? for 30min. Thereafter, bacterial suspensions were centrifuged at 4? 1,825×g for 10min and washed with PBS twice.

Instrument

Data were collected on ThermoFisher AttuneTM NxT flow cytometry.

Software

Flow-cytometry data were analysed on FlowJo (v10.8.1).

Cell population abundance

Not applicable.

Gating strategy

In general, each sample was gated through FSC-A/SSC-A, BL2-A-PI/SSC-A, and BL2-A-PI/Count.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.