nature portfolio

Corresponding author(s): Lei Qin, Bo Lv, Haiyang Jia and Chun Li

Last updated by author(s): Dec 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<br>*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<br>*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 | Retrobiosynthesis data were obtained using Python 3.11 with RDKit (Fingerprint Similarity) and NetworkX (shortest_simple_paths) packages, and RetroRules reaction rules. Genome sequencing was performed on a PacBio Sequel platform (Pacific Biosciences) and Hi-C data processed by Juicer v1.6 and 3D-DNA v201008. Transcriptome sequencing libraries were constructed on BGISEQ500RS (BGI Genomics), and RNA-seq reads were mapped using Hisat2 v2.2.1. Phylogenetic trees were generated using IQ-TREE v2.1.4-beta and visualized by iTOL. Protein structures were predicted by local AlphaFold 3; ligand SMILES were retrieved from PubChem. Metabolite profiling was performed on a Thermo Ultimate 3000 UHPLC coupled to Q Exactive Plus Orbitrap MS, and images were acquired on a Nikon AXR super-resolution confocal microscope with NIS-Elements software. |
| Data analysis | Genome assembly and polishing were conducted with Canu v2.2, Smartdenovo v1.0.0, Racon v1.4.3, Pilon v1.24, and Purge Haplotigs v1.1.2, with BUSCO v5.8.3 for completeness evaluation. Transcript abundance was quantified using Stringtie, co-expression heatmaps plotted with R package pheatmap v1.0.12, and Pearson correlation coefficients calculated in R. Protein-ligand complexes were analyzed in PyMOL v2.5.7 with APBS Electrostatics. Ligand geometry optimization was performed using ORCA 6.0, and molecular dynamics simulations were carried out in GROMACS 2023 with Amber ff14SB force field. LC-MS raw data were processed in Compound Discoverer 3.3 with mzCloud/mzVault libraries. Fluorescence images were analyzed using ImageJ v1.54f, and ordinary differential equations were solved with Python (solve_ivp, scipy v1.10.1). |

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

Source data are provided with this paper. Data supporting the findings of this study are publicly available at Figshare through the identifier https://10.6084/m9.figshare.29221010. Genome and transcriptome raw sequencing data generated in this study are available at https://kw.ibc.tsinghua.edu.cn/database/genome and https://kw.ibc.tsinghua.edu.cn/database/transcriptome, respectively. Metabolomics data generated in this study have been deposited in the MetaboLights database under accession code MTBLS13519 (https://www.ebi.ac.uk/metabolights/MTBLS13519). Gene accession numbers for all biosynthetic genes characterized in this study are provided in Supplementary Data 2.

## 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 | n/a |
| Reporting on race, ethnicity, or other socially relevant groupings | n/a |
| Population characteristics | n/a |
| Recruitment | n/a |
| Ethics oversight | n/a |

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 | At least three independent biological replicates were analyzed for each condition unless otherwise indicated, as described in the figure legends. |
| Data exclusions | No data was excluded in the analysis. |
| Replication | All key experiments were independently repeated at least three times with similar results. The data are reproducible. |
| Randomization | Randomization was not applicable, as the study did not involve subject allocation to experimental groups. |
| Blinding | Blinding was not applicable, as no subjective assessment was involved. All data were analyzed using predefined computational pipelines. |

# 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 |

# Dual use research of concern

Policy information about dual use research of concern

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No | Yes |
|---|---|
| ☒ | ☐ Public health |
| ☒ | ☐ National security |
| ☒ | ☐ Crops and/or livestock |
| ☒ | ☐ Ecosystems |
| ☒ | ☐ Any other significant area |

### Experiments of concern

Does the work involve any of these experiments of concern:

| No | Yes |
|---|---|
| ☒ | ☐ Demonstrate how to render a vaccine ineffective |
| ☒ | ☐ Confer resistance to therapeutically useful antibiotics or antiviral agents |
| ☒ | ☐ Enhance the virulence of a pathogen or render a nonpathogen virulent |
| ☒ | ☐ Increase transmissibility of a pathogen |
| ☒ | ☐ Alter the host range of a pathogen |
| ☒ | ☐ Enable evasion of diagnostic/detection modalities |
| ☒ | ☐ Enable the weaponization of a biological agent or toxin |
| ☒ | ☐ Any other potentially harmful combination of experiments and agents |

# Plants

| | |
|---|---|
| Seed stocks | Glycyrrhiza glabra specimens were collected from the Alar region of Xinjiang, China, and other wild populations located in Turpan, Urumqi, Emin, Tumxuk, Aksu (Xinjiang), and Longdong (Gansu). Wild plants were sampled during April, May, June, July, and September in 2018. No seed stocks from commercial sources or germplasm banks were used. |
| Novel plant genotypes | No novel plant genotypes were generated in this study; only wild-type Glycyrrhiza glabra accessions were analyzed. |
| Authentication | The authenticity of Glycyrrhiza glabra specimens was confirmed based on morphological characteristics examined by experienced botanists. No transgenic or gene-edited materials were used. |