

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	No software had been used for data collection. Data were sequenced from PacBio and Illumina platform.
Data analysis	Hifiasm (v0.16.1), HiCUP (v0.6.1), ALLHIC (v0.9.8), Juicebox (v.1.11.08), Merqury (v.1.3), BWA mem2 (v.0.7.17), Picard (v. 2.20.7), GATK (v. 4.2.2.0), TBtools (v2.056), IQ-TREE2 (v. 2.1), GetOrganelle (v. 1.7.4.1), FigTree, iTOL, Mesquite (v.3.81), ADMIXTURE (v. 1.3.0), Plink (v. 1.90), PopLDdecay (v. 3.41), GeneHapR, GWASpoly (v. 2.10), VCFtools (v.0.1.16), ArcGIS (v.10.2), TreeMix (v.1.13), OptM R (v.0.1.6), Dsuite, Stairway Plot (v.2.1), R (v4.3.1), JCVI, QuarTeT, Genewise (v. 2.4.1), Augustus (v. 3.3), TopHat (v. 2.1.1), EVIDENCEModeler (v. 1.1.1), RepeatMasker (v.4.0.6), RepeatModeler (v.1.0.11), LTR-FINDER (v1.0.7), NGenomeSyn, Orthofinder (v 2.5.4), BUSCO (v.5.4.5). Specific parameters used during run-time are provided in the methods. All softwares or scripts are available from official websites or GitHub as indicated in the methods and supplementary methods.

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 phased genome assembly of *R. persica* has been deposited in the National Center for Biotechnology Information BioProject database with the BioProject PRJNA1216766 (Haplotype 1) and PRJNA1216767 (Haplotype 2). The previously published reads used in this study are available from NCBI and public GDR database (<https://www.rosaceae.org>). Specific source data are listed in the Supplementary Tables 10 and 13.

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.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

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	215 accessions were used to represent species of the genus <i>Rosa</i> . The logic of this selection was based on <i>Rosa</i> species that are collectible. Our collection encompasses 84% (80/95) of <i>Rosa</i> species documented in the Flora of China. We have also included subspecies, varieties, and cultivars for comprehensive analyses. No statistical methods were used to predetermine sample size. Our samples were all from wild type and did not use processed samples and groups.
Data exclusions	No data were excluded.
Replication	The genome sequence was taken and sequenced with more than 247 fold coverage. Robustness of phylogenetic analysis was tested using bootstrap test. All measurements of the phenotypic traits were taken at least three biological repetition. Multiple tests were used to identify gene flow, including TreeMix and D statistics. We confirm that all attempts at replication were successful.
Randomization	No random sampling is required for genome sequencing, because the genome differences are very small within the wild population, thus any wild plant is allowed for genome sequencing.
Blinding	Blinding is not applicable in our study because it does not involve subjects which receive different treatments.

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

Methods

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

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	The plant accessions used in this study were extensively collected across China and preserved in germplasm gardens in Yunnan, China.
Novel plant genotypes	No novel plant genotypes were applied in this study.
Authentication	We introduced these resources to our germplasm gardens for years of observations, including the morphology of flowers, leaves and hips, as well as the shedding of glandular trichomes and sepals. After comparing with a large number of specimens and consulting with experts and colleagues, the collected plant accessions were accurately identified and used for this study.

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

Fresh leaves of *Rosa persica* were collected and chopped, then put into dissociation solution for 10 min, followed by filtering to collect nucleus cells. Leaf material of *Solanum pimpinellifolium* was used as an internal standard, with known genome size, and co-chopped with the rose callus.

Instrument

BD FACScalibur flow cytometry (Becton, Dickinson and Company)

Software

Modifit 3.0 was used for DNA content analysis.

Cell population abundance

The fluorescence intensity was detected under 488 nm blue light excitation. For every analysis 10,000 nuclei were analyzed. Genome sizes were calculated based on the ratio of the peak position of the *Rosa persica* callus material and the *Solanum pimpinellifolium* with the known genome size (900Mbp).

Gating strategy

No gating was applied. The analysis of the plant genome size results in a non-gated one parameter histogram output on a linear scale.

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