nature portfolio

Corresponding author(s): Li He

Last updated by author(s): Jul 23, 2024

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	Summit v.5.2, Jellyfish (v2.3.0), Genomescope (v2.0), Hifiasm (v0.16.1-r375), Juicer (v1.5.6), 3d-dna (v180922), Juicebox(v1.11.08), LR_Gapcloser (v1.1), Nextpolish (v1.3.1), GetOrganelle (v1.7.5), Redundans (v0.13c), Trinity (v2.0.6), HISAT2 (v2.1.0), StingTie (v1.3.5), CD-HIT (v4.6), PASA (v2.4.1), AUGUSTUS (v3.4.0), MAKER (v2.31.9), Exonerate (2.2.0), RepeatMasker (http://www.repeatmasker.org/RepeatMasker/), rmblast(2.11.0), blast+ (v2.13.0+), EvidenceModeler (v1.1.1), TEsorter (v1.4.1), tRNAScan-SE (v1.3.1), RfamScan (v1.0), Barrnap (https://github.com/tseemann/barrnap), eggNOG-mapper (v2.0.0), DIAMOND (v0.9.24), InterProScan (v5.27-66.0), EDTA (v1.9.9), OrthoFinder (v2.5.5), MAFFT (v7.515), IQ-TREE (v2.2.0.3), ASTRAL (v5.7.8), PAML (4.10.0), HomBlocks.pl (v1.0), Nucmer (v4.0.0beta2), RepeatModeler (http://www.repeatmasker.org/), ggplot2 R package (v3.4.4), fastp (v0.23.2), KMC (v3.1.1), Bowtie (v0.12.9), cq-calculate.pl software (https://sourceforge.net/projects/cqcalculate/files/CQ-calculate.pl/download), bwa (v0.7.12), Samtools (v0.1.19), sambamba (v0.7.1), Genome Analysis Toolkit (v4.1.8.1), VCFtools (v0.1.16), Changepoint R package (v2.2.4), Python version of MCScan (v1.2.7), BLASTP (v2.5.0+), BLASTN (v2.5.0+), HISAT2 (v2.2.1), featureCounts (v2.0.1), R, Python, sRNAminer (v1.1.2), IGV (v2.17.1)

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

The genome assembly sequences generated in this study have been deposited in the National Genomics Data Center (NGDC) database under BioProject accession code PRJCA016000 [https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA016000] and in NCBI database under BioProject accession code: PRJNA1130083 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1130083], PRJNA1130084 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1130084], PRJNA1130085 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1130085], and PRJNA1130087 [https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1130087]. The sequencing datasets have been deposited in NCBI under the BioProject accession numbers PRJNA882493, PRJNA1020583, & PRJNA1020619.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	Allopolyploidization from two dioecious ancestors leads to recurrent evolution of sex chromosomes
Research sample	The study collected young leaves from a male Salix dunnii (FNU-M-1) and female weeping willow S. babylonica (saba01F) plant for genome sequencing. Young leaf, catkin, stem and root samples for transcriptome sequencing were collected from FNU-M-1 and saba01F, as well as catkins from two other female and three male plants of weeping willow, one other female plant of S. dunnii. In October 2023 (stage 1), December 2023 (stage 2), and February 2024 (stage 3), flower buds from the three male and three female individuals of S. babylonica for small RNA sequencing and mRNA sequencing. The leaves of 40 individuals of S. babylonica were sampled and dried in silica-gel for resequencing.
Sampling strategy	The individual number of S. babylonica was according to previously works (He et al., 2021,10.1111/1755-0998.13362; He et al., 2023, 10.1111/mec.16902).
Data collection	The study downloaded genome sequence data of 38 individuals of S. dunnii published by He et al.(2021, 10.1111/1755-0998.13362), the genome of S. arbutifolia (Wang et al., 2024, https://doi.org/10.1111/nph.19744, with gap-free 15X and 15Y sex chromosomes), and other available willows genomes and Populus trichocarpa for relevant analysis.
Timing and spatial scale	Not applicable.
Data exclusions	No data was excluded in this analysis.
Reproducibility	Not applicable.
Randomization	Randomization was not relevant to this study.

Blinding	Analyses were unblinded.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The samples were collected during flowering time of Salix babylonica and S. dunnii to confirm the sex of each individual.
Location	Salix babylonica, 31.077222 °N, 121.1741667 °E; Salix dunnii, 26.02675 °N, 119.209329 °E, and 26.02675 °N, 119.209329 °E.
Access & import/export	No permit is required to collect these plants.
Disturbance	No disturbance.

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	Not applicable.
Novel plant genotypes	Not applicable.
Authentication	Not applicable.

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	The samples' leaves were dried in silica-gel. The leaf tissue was incubated for 80 min in 1 mL LB01 buffer and chopped with a razor blade.

Instrument	MoFlo-XDP flow cytometer.
Software	Summit v.5.2.
Cell population abundance	Flow cytometry was used for ploidy level estimation purpose only, and no post-sorting fraction was collected.
Gating strategy	The FI3-Height gate method was used to eliminate debris, cell fragments, and dead cells.

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