nature portfolio

Corresponding author(s): Alex Harkess

Last updated by author(s): Sep 12, 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

Data analysis: The manuscript main text contains version numbers and citations for all software used in the materials and methods section. A significant number of programs were run, and the main text contains all this information.

We annotated repeats using EDTA v2.0.0.
To identify tandem repeats, we used Tandem Repeats Finder v.4.09.1 (parameters ??2 7 7 80 10 50 500 -f -d -m -h).
We ran StainedGlass v0.593 to visualize the massive tandem repeat arrays for chromosomes in both haplotypes.
To build the repeat landscapes for assessing recent expansion events, we followed the methods outlined in EDTA Github Issue #92: Draw Repeat Landscapes, utilizing a library generated from an independent annotation on the combined haplotypes with EDTA v2.0.1.
To plot comparisons between the two haplotypes, including genes and repeats, we used GENESPACE v.1.3.1 94. We assessed the gene annotations using compleasm v0.2.6 89 with the embryophyta database from v.odb10.
To generate synteny between the two haplotypes, we first performed genome alignments. HAP1 and HAP2 were aligned using AnchorWave v1.0.1 using the 'genoAli' method and '-IV' parameter to allow for inversions. Alignment was performed using only "chromosome" sequence f or each haplotype. The alignment was converted to SAM format using the 'maf-convert' tool provided in 'last' v460 and used for calling variants with SyRI v1.6.3.
The output from SyRI was used to make chromosome-level synteny and SV plots using plotsr v0.5.4.
We used whole-genome sequencing data to identify the sex-determining region (SDR) of the W. All paired-end Illumina data had adapters removed and were quality filtered using TRIMMOMATIC v0.39 99 with leading and trailing values of 3, sliding window of 30, jump of 10, and a minimum remaining read length of 40.
We found all canonical 21-mers in each isolate using Jellyfish v2.3.0 100 and used the bash comm command to find all k-mers shared in all female isolates and not found in any male isolate (W-mers).
We mapped the W-mers to both haplotype assemblies using BWA-MEM v0.7.17 80, with parameters '-k 21' '-T 21' '-a' '-c 10'. W-mer mapping was visualized by first calculating coverage in 100,000-bp sliding windows (10,000 bp jump) using BEDTools v2.28.0 101 and plotted using karyoploteR v1.26.0 102.To identify structural variants between the haplotypes, we mapped PacBio reads using minimap2 v2.24 in HiFi mode, added the MD tag using samtools v1.10 calmd, and called structural variants using Sniffles v2.0.7. We also performed whole-genome

dotplot using pafR v0.0.2 105. To identify one-to-one orthologs on the ZW to examine protein evolution, we ran OrthoFinder v.2.5.2 106,107 using only the Amborella haplotypes. We calculated synonymous (Ks) and nonsynonymous (Ka) changes in codons using Ka/Ks Calculator v2.0. To identify the boundaries of evolutionary strata, we used the R package mcp v0.3.4 109 on dXY and Ks. For Ks, we first ran a test for outliers using PMCMRplus v1.9.10 110 to run Rosner's generalized extreme studentized deviate many-outlier test 111. For mcp, we used the model, 'y ~ 1, ~ 1' to identify the change point between two plateaus, and we used 100,000 iterations, 3 chains, and a burn-in of 100,000 (i.e., 'adapt').BWA v0.7.17 80 was used to map reads, and bcftools v1.9 mpileup and call 112 functions were used to call variants using the Island-wide sampling (nine male and six female plants; Table S11). We filtered the vcf file using 'QUAL>20 & DP>5 & MQ>30', minor allele frequency of 0.05, and dropped sites with > 25% missing data. To calculate Nei's nucleotide diversity between the sexes (dXY), we used pixy v1.2.7.beta1.To call PAV, reads for the samples were aligned to each haplotype using BWA v0.7.17 and orted BAM files were converted to bedgraph format using bedtools v2.30.0. For differential expression, filtered reads were mapped to the HAP1 genome assembly using STAR v2.7.9a and expression estimated for the annotated gene models using StringTie v2.1.7 and we performed differential gene expression analyses using DESeq2 v1.32.0.

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

The genome assemblies and annotations (v.2.1) are available on Phytozome v13 (https://phytozome-next.jgi.doe.gov/) and have been deposited on NCBI under BioProjects PRJNA1100625 and PRJNA1167780. Sequencing libraries for the genome assembly and annotation are publicly available on NCBI under BioProject PRJNA1100625 and the whole-genome sequencing of additional isolates are under PRJNA1161132. Individual accession numbers are provided in Supplementary Tables S10-11.

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	None
Reporting on race, ethnicity, or other socially relevant groupings	None
Population characteristics	None
Recruitment	None
Ethics oversight	None

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	This study was performed to sequence and assemble a chromosome-scale genome for Amborella trichopoda. Here we present a haplotype-resolved genome assembly, including highly contiguous assemblies of the Z and W chromosomes.
Research sample	The research sample used here is Amborella 75-5, an accession at the Atlanta Botanic Garden. Additional population-level Illumina short-read sequencing was performed on all individuals of a segregating F1 population of males and females.
Sampling strategy	No sample size calculation was performed, as only reference genome sequence was produced. For population-level sequencing, as many individuals were collected as possible from the different subpopulations.
Data collection	Tissue for the reference genome line was collected by Dr. Jim Leebens-Mack and sequenced and assembled by the DOE Joint

Data collection	Genome Institute. The JGI also performed all RNA isolation and sequencing and analysis.
Timing and spatial scale	None
Data exclusions	We identified two significant Ks outliers and excluded them from the change point analysis.
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	Randomization is not relevant to this study as it is a genome assembly manuscript with population-level resequencing of known male and female isolates.
Blinding	Blinding was not used in this study as it was necessary to know the sex of individuals before sequencing.

Did the study involve field work? ☐ Yes ☒ No

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	The reference genome individual 75-5 is available as a living individual at the Atlanta Botanic Garden
Novel plant genotypes	None
Authentication	None