





DATA NOTE

The genome sequence of the tree of heaven, *Ailanthus altissima* (Mill.) Swingle, 1916 [version 1; peer review: 2 approved]

Rowan J. Schley ¹, Ilia J. Leitch ², Maarten J. M. Christenhusz²,
Royal Botanic Gardens Kew Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective, Plant Genome Sizing collective,
Wellcome Sanger Institute Tree of Life programme,
Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

¹University of Exeter, Exeter, England, UK

²Royal Botanic Gardens Kew, Richmond, England, UK

V1 First published: 25 Jul 2023, 8:321
<https://doi.org/10.12688/wellcomeopenres.19628.1>

Latest published: 25 Jul 2023, 8:321
<https://doi.org/10.12688/wellcomeopenres.19628.1>

Abstract

We present a genome assembly from an individual *Ailanthus altissima* (tree of heaven; Streptophyta; Magnoliopsida; Sapindales; Simaroubaceae). The genome sequence is 939 megabases in span. Most of the assembly is scaffolded into 31 chromosomal pseudomolecules. The mitochondrial and plastid genome assemblies are 661.1 kilobases and 161.1 kilobases long, respectively.

Keywords



Ailanthus altissima, tree of heaven, genome sequence, chromosomal, Simaroubaceae



This article is included in the [Tree of Life gateway](#).

Open Peer Review

Approval Status  

	1	2
version 1 25 Jul 2023	 view	 view

1. **Xin Liu** , BGI (Beijing Genomics Institute)-Shenzhen, Shenzhen, China
2. **Jim Leebens-Mack** , University of Georgia, Athens, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: **Schley RJ:** Writing – Original Draft Preparation, Writing – Review & Editing; **Leitch IJ:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Christenhusz MJM:** Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute (206194, <https://doi.org/10.35802/206194>) and the Darwin Tree of Life Discretionary Award (218328, <https://doi.org/10.35802/218328>). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2023 Schley RJ *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Schley RJ, Leitch IJ, Christenhusz MJM *et al.* **The genome sequence of the tree of heaven, *Ailanthus altissima* (Mill.) Swingle, 1916 [version 1; peer review: 2 approved]** Wellcome Open Research 2023, 8:321 <https://doi.org/10.12688/wellcomeopenres.19628.1>

First published: 25 Jul 2023, 8:321 <https://doi.org/10.12688/wellcomeopenres.19628.1>

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetales; rosids; malvids; Sapindales; Simaroubaceae; *Ailanthus*; *Ailanthus altissima* (Mill.) Swingle, 1916 (NCBItxid:23810).

Background

The tree of heaven (*Ailanthus altissima*, Simaroubaceae) is a deciduous tree native to China and Taiwan, known for its fast growth and attractive foliage. For these reasons, it has been widely planted as an ornamental tree (Figure 1). *Ailanthus altissima* was first introduced to the UK in 1751 as a garden plant (Hu, 1979), during a period when European fascination with Chinese culture ('Chinoiserie') was at its height. *Ailanthus altissima* has long been used in Chinese traditional medicine as well as in silk production, as a foodplant for the *Ailanthus* silk moth (*Samia cynthia*) (Swingle, 1916). *Ailanthus altissima* has since become an invasive species across North America, southern South America, South Africa, New Zealand and Europe due to its rapid growth, resistance to pollution, fecundity, vegetative reproduction and use of allelopathic chemicals to stifle competition (Demeter *et al.*, 2021).

The genus name *Ailanthus* is derived from the Ambonese word 'ailanto', meaning 'heaven-tree', and the specific epithet *altissima* is Latin for 'tallest', referring to the size of the mature tree. Three varieties of *A. altissima* have been described – *A. altissima* var. *altissima*, native to mainland China, *A. altissima* var. *sutchuenensis* (Dode) Rehder and E. H. Wilson, restricted to southern China and *A. altissima* var. *tanakae* (Hayata) Kaneh. and Sasaki, endemic to northern Taiwan. Multiple chromosome counts have been reported for *A. altissima*, ranging from $2n = 80$ for arboretum specimens examined at the Royal Botanic Gardens, Kew (Desai, 1960), down to $2n = 64$ for samples collected from naturalised populations in mainland Europe (Ivanova *et al.*, 2006), suggesting polyploidy with different cytotypes between individuals.

Here we present the first high-quality genome of *A. altissima* var. *altissima*, which we believe will be a useful resource both for those studying the tree in its native range and those aiming to understand what underlies its success as an invasive species. This genome complements existing studies which identified microsatellites for use in population genetic analysis (Dallas *et al.*, 2005; Saina *et al.*, 2021), and those which sequenced the plastome of the species (Saina *et al.*, 2018).

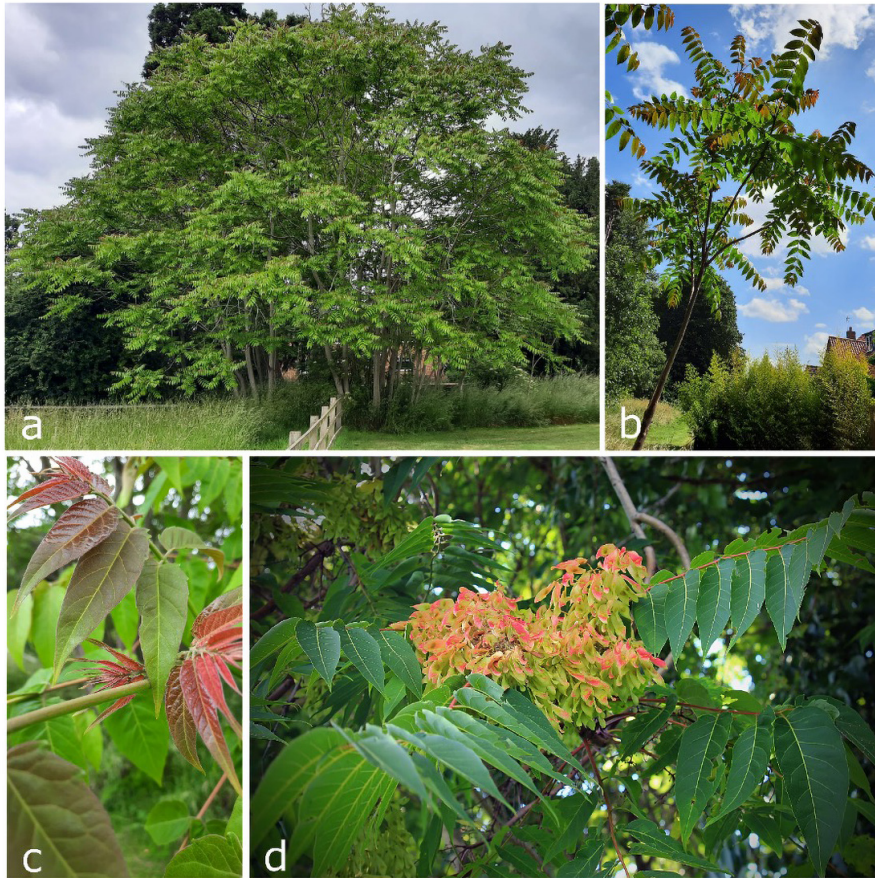


Figure 1. *Ailanthus altissima* (not the sampled specimen). **a)** mature tree; **b)** young sapling, and **c)** new leaves in Nottinghamshire (photographs © Carlton Roberts); **d)** *A. altissima* fruits (photograph © Nicholas A. Tonelli).

Fruitful directions for future work may include investigating the population genetics of invasion, building on previous work showing population bottlenecks in invasive populations of *A. altissima* (Neophytou *et al.*, 2020).

Genome sequence report

The genome was sequenced from the leaves of an *A. altissima* specimen (drAilAlt1) collected from the Royal Botanic Gardens, Kew (latitude 51.482, longitude -0.2879). Using flow cytometry, the genome size (1C-value) was estimated as 1.20 pg, equivalent to 1,170 Mb. A total of 45-fold coverage in Pacific Biosciences single-molecule HiFi long reads and 35-fold coverage in 10X Genomics read clouds were generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 17 missing joins or misjoins and removed one haplotypic duplication, reducing the assembly length by 0.11% and the scaffold number by 33.93%, and decreasing the scaffold N50 by 3.45%.

The final assembly has a total length of 939 Mb in 37 sequence scaffolds with a scaffold N50 of 29.4 Mb (Table 1). Most of the assembly sequence (97.72%) was assigned to 31 chromosomal-level scaffolds (Figure 2–Figure 5; Table 2). Several small repetitive scaffolds localise equally well to both chromosome 28 and chromosome 9. They have been placed with chromosome 9, although their placement is uncertain. While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited.

The estimated Quality Value (QV) of the final assembly is 61.1 with *k*-mer completeness of 100%, and the assembly has a BUSCO v5.3.2 completeness of 98.7% (single 54.7%, duplicated 44.0%), using the eudicots_odb10 reference set ($n = 2,326$).

Metadata for specimens, spectral estimates, sequencing runs, contaminants and pre-curation assembly statistics can be found at <https://links.tol.sanger.ac.uk/species/23810>.

Methods

Sample acquisition, genome size estimation and nucleic acid extraction

Ailanthus altissima grows as a weed along pavements and in neglected plots in the UK. Leaves from an *A. altissima* individual (drAilAlt1) were harvested from a plant growing in the pavement near the Jodrell Gate, Royal Botanic Gardens Kew, Surrey, UK (latitude 51.482, longitude -0.2879). The sample was collected and identified by Maarten Christenhusz (Royal Botanic Gardens, Kew). The leaves were preserved by freezing at -80°C.

The genome size was estimated by flow cytometry using the fluorochrome propidium iodide and following the ‘one-step’ method as outlined in (Pellicer *et al.*, 2021). Specifically for

this species, the General Purpose Buffer (GPB) supplemented with 3% PVP and 0.08% (v/v) beta-mercaptoethanol was used for isolation of nuclei (Loureiro *et al.*, 2007), and the internal calibration standard was *Petroselinum crispum* ‘Champion Moss Curled’ with an assumed 1C-value of 2,200 Mb (Obermayer *et al.*, 2002).

DNA was extracted at the Tree of Life laboratory, Wellcome Sanger Institute (WSI). The drAilAlt1 sample was weighed and dissected on dry ice with tissue set aside for Hi-C sequencing. Leaf tissue was cryogenically disrupted to a fine powder using Qiagen Plant Magattract, receiving multiple impacts. High molecular weight (HMW) DNA was extracted using the Qiagen MagAttract HMW DNA extraction kit. Low molecular weight DNA was removed from a 20 ng aliquot of extracted DNA using 0.8X AMPure XP purification kit prior to 10X Chromium sequencing; a minimum of 50 ng DNA was submitted for 10X sequencing. HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30. Sheared DNA was purified by solid-phase reversible immobilisation using AMPure PB beads with a 1.8× ratio of beads to sample to remove the shorter fragments and concentrate the DNA sample. The concentration of the sheared and purified DNA was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from leaf tissue of drAilAlt1 in the Tree of Life Laboratory at the WSI using TRIzol, according to the manufacturer’s instructions. RNA was then eluted in 50 µL RNase-free water and its concentration was assessed using a Nanodrop spectrophotometer and Qubit Fluorometer using the Qubit RNA Broad-Range (BR) Assay kit. Analysis of the integrity of the RNA was done using Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus and 10X Genomics read cloud DNA sequencing libraries were constructed according to the manufacturers’ instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing were performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi), Illumina HiSeq 4000 (RNA-Seq) and Illumina NovaSeq 6000 (10X) instruments. Hi-C data were also generated from material from drAilAlt1 using the Arima v2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). One round of polishing

Table 1. Genome data for *Ailanthus altissima*, drAilAlti1.1.

Project accession data		
Assembly identifier	drAilAlti1.1	
Species	<i>Ailanthus altissima</i>	
Specimen	drAilAlti1	
NCBI taxonomy ID	23810	
BioProject	PRJEB47393	
BioSample ID	SAMEA7522284	
Isolate information	Monoecious; leaves	
Assembly metrics*		Benchmark
Consensus quality (QV)	61.1	≥ 50
<i>k</i> -mer completeness	100%	≥ 95%
BUSCO**	C:98.7%[S:54.7%,D:44.0%], F:0.3%,M:0.9%,n:2,326	C ≥ 95%
Percentage of assembly mapped to chromosomes	97.72%	≥ 95%
Sex chromosomes	Not applicable	<i>localised homologous pairs</i>
Organelles	MT single scaffold of 661,054 bp Plastid single scaffold of 161,079 bp	<i>complete single alleles</i>
Raw data accessions		
PacificBiosciences SEQUEL II	ERR6808067, ERR6939282	
10X Genomics Illumina	ERR6745740-ERR6745743	
Hi-C Illumina	ERR6745744	
PolyA RNA-Seq Illumina	ERR9435026, ERR9435027	
Genome assembly		
Assembly accession	GCA_946807835.1	
<i>Accession of alternate haplotype</i>	GCA_946808065.1	
Span (Mb)	939	
Number of contigs	49	
Contig N50 length (Mb)	29.4 Mb	
Number of scaffolds	37	
Scaffold N50 length (Mb)	29.3 Mb	
Longest scaffold (Mb)	38.4	

* Assembly metric benchmarks are adapted from column VGP-2020 of “Table 1: Proposed standards and metrics for defining genome assembly quality” from (Rhie *et al.*, 2021).

** BUSCO scores based on the eudicots_odb10 BUSCO set using v5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at <https://blobtoolkit.genomehubs.org/view/CAMPEX01/dataset/CAMPEX01/busco>.

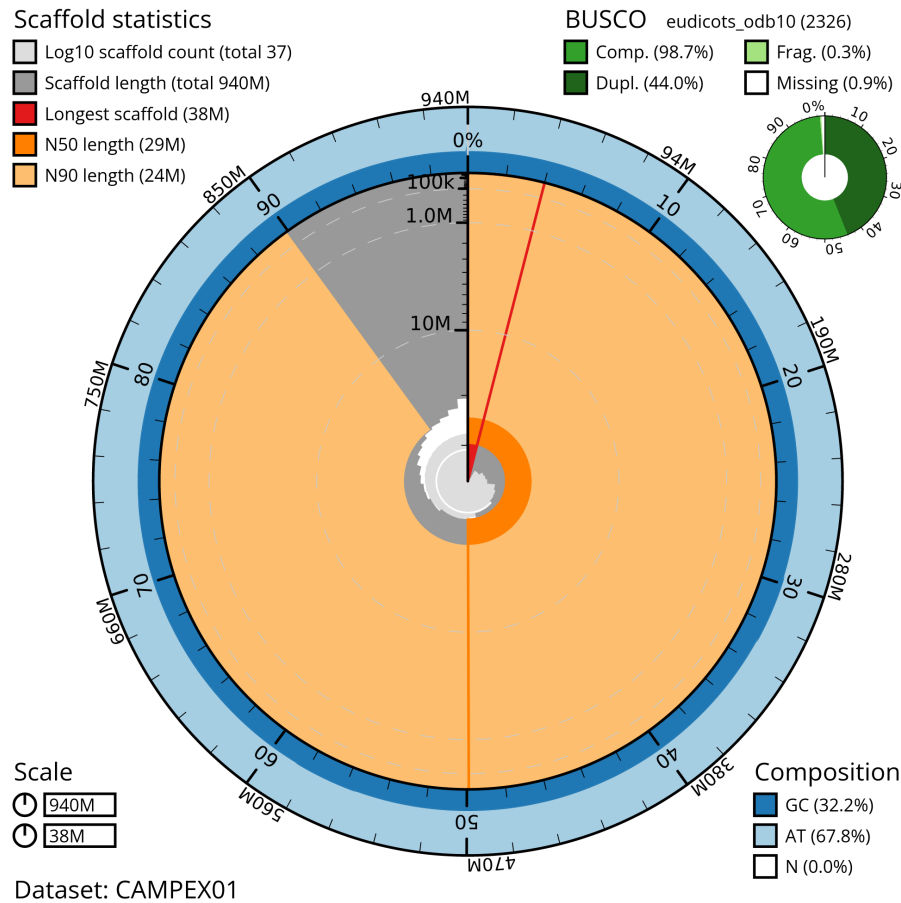


Figure 2. Genome assembly of *Ailanthus altissima*, drAilAlt1.1: metrics. The BlobToolKit Sniplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 939,218,608 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (38,427,504 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (29,421,908 and 24,184,688 bp, respectively). The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the eudicots_odb10 set is shown in the top right. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAMPEX01/dataset/CAMPEX01/sniplot>.

was performed by aligning 10X Genomics read data to the assembly with Long Ranger ALIGN, calling variants with FreeBayes (Garrison & Marth, 2012). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using SALSA2 (Ghurye *et al.*, 2019). The assembly was checked for contamination and corrected using the gEVAL system (Chow *et al.*, 2016) as described previously (Howe *et al.*, 2021). Manual curation was performed using gEVAL, HiGlass (Kerpedjiev *et al.*, 2018) and Pretext (Harry, 2022). The mitochondrial and plastid genomes were assembled using MBG (Rautiainen & Marschall, 2021) from PacBio HiFi reads mapping to related genomes. A representative circular sequence was selected for each from the graph based on read coverage.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format (Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines “sanger-tol/readmapping” (Surana *et al.*, 2023a) and “sanger-tol/genomenote” (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021; Simão *et al.*, 2015) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

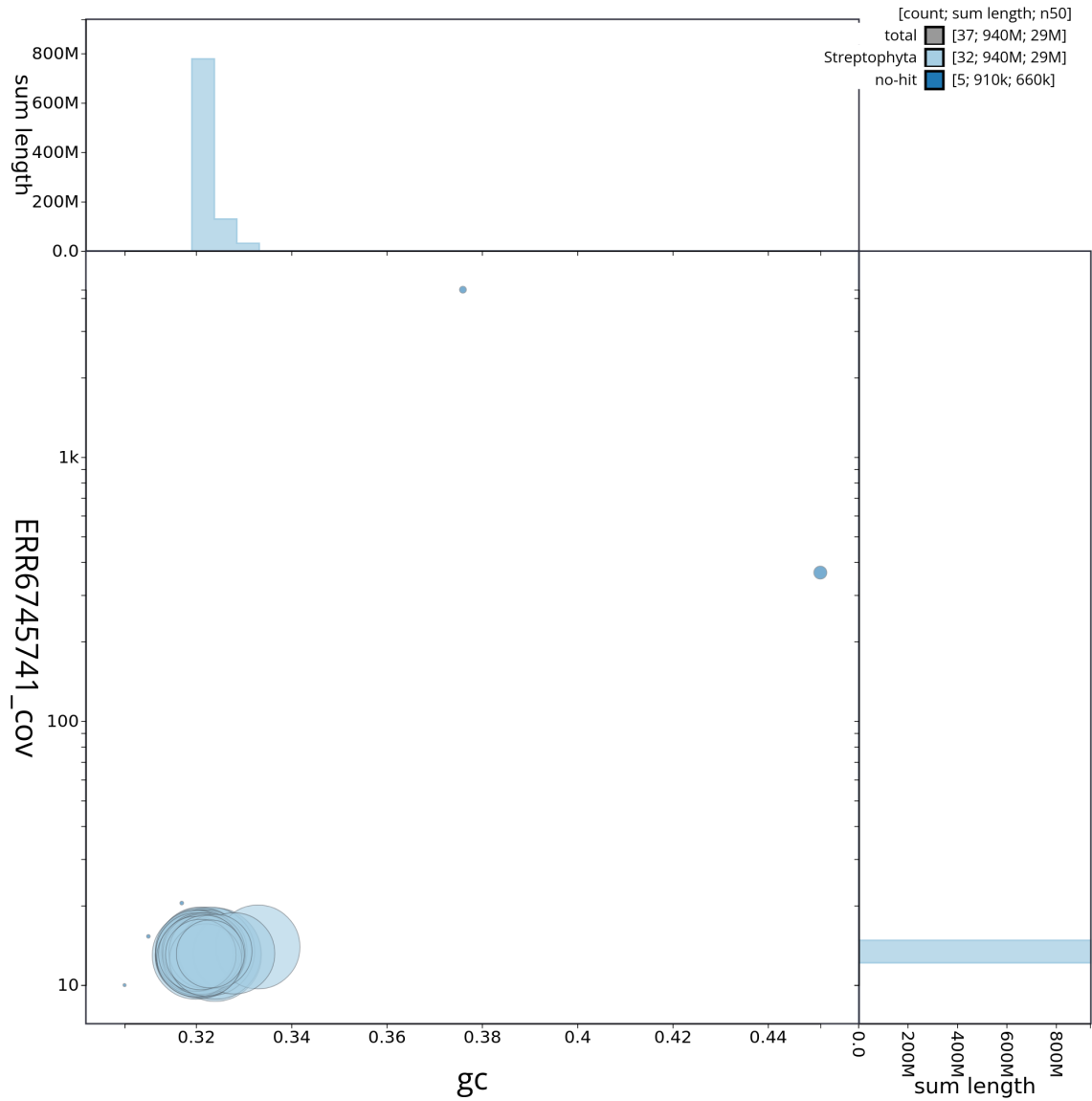


Figure 3. Genome assembly of *Ailanthus altissima*, drAilAlti1.1: GC coverage. BlobToolKit GC-coverage plot. Chromosomes are coloured by phylum. Circles are sized in proportion to chromosome length. Histograms show the distribution of chromosome length sum along each axis. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAMPEX01/dataset/CAMPEX01/blob>.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the ‘**Darwin Tree of Life Project Sampling Code of Practice**’, which can be found in full on the Darwin Tree of Life website [here](#). By agreeing with and signing up to the Sampling Code

of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature

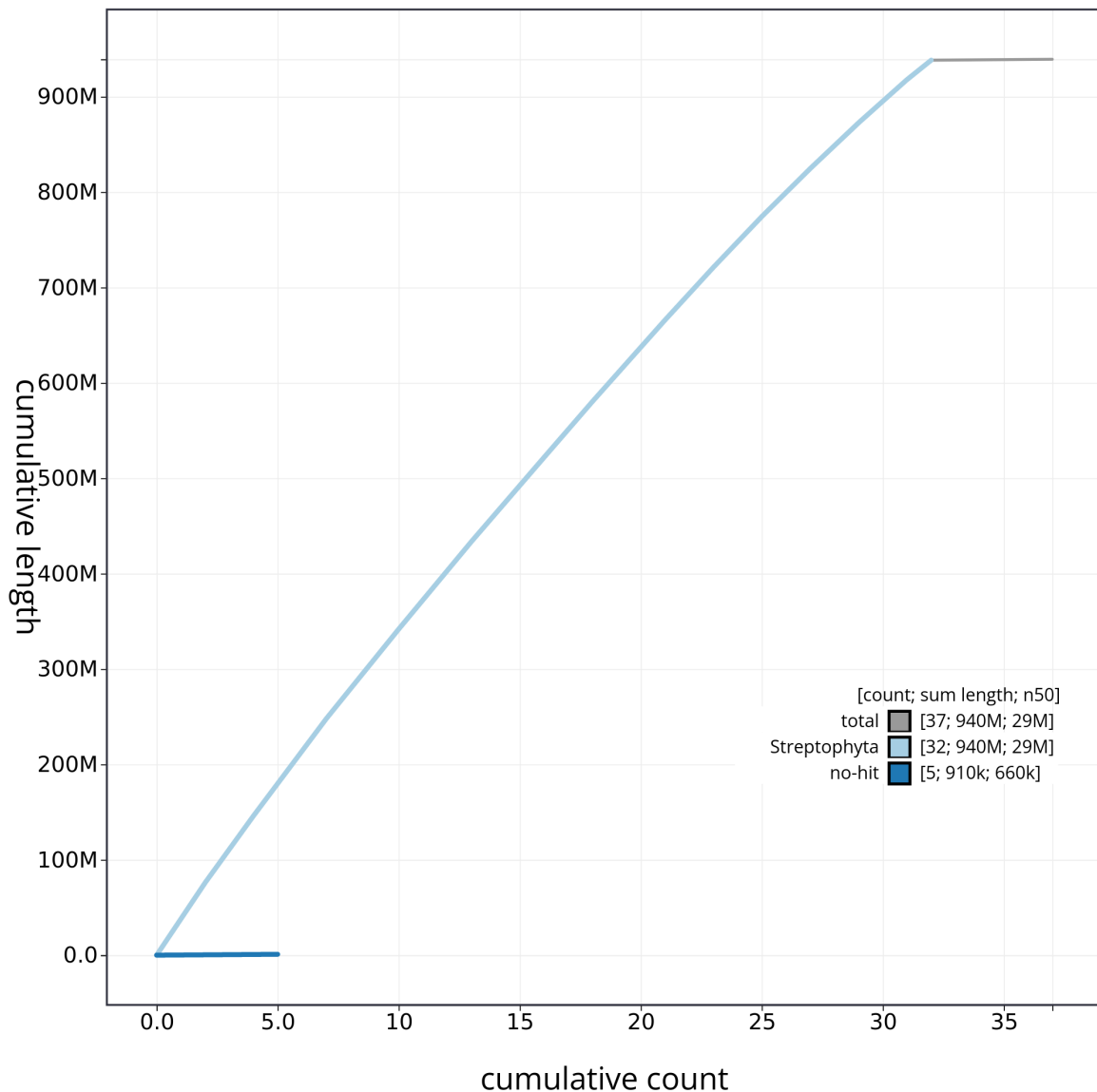


Figure 4. Genome assembly of *Ailanthus altissima*, drAiAlti1.1: cumulative sequence. BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all chromosomes. Coloured lines show cumulative lengths of chromosomes assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at <https://blobtoolkit.genomehubs.org/view/CAMPEX01/dataset/CAMPEX01/cumulative>.

of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material

- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger Institute), and in some circumstances other Darwin Tree of Life collaborators.

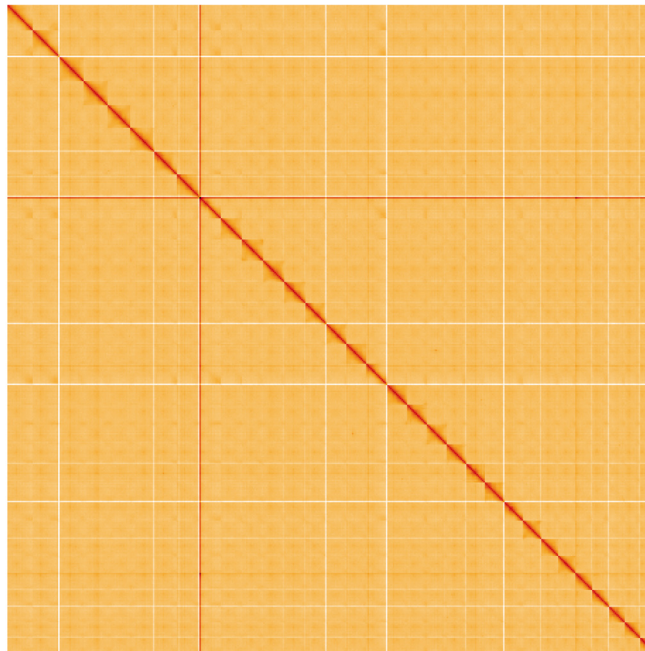


Figure 5. Genome assembly of *Ailanthus altissima*, drAilAlt1.1: Hi-C contact map. Hi-C contact map of the drAilAlt1.1 assembly, visualised using HiGlass. Chromosomes are given in order of size from left to right and top to bottom. An interactive version of this figure is available at <https://genome-note-higlass.tol.sanger.ac.uk/?d=XyRKDNDOTHy19SzkZkKRWw>.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Ailanthus altissima*, drAilAlt1.

INSDC accession	Chromosome	Size (Mb)	GC%
OX327683.1	1	38.43	32.4
OX327684.1	2	37.36	32.4
OX327685.1	3	35.13	32.1
OX327686.1	4	35.07	32.3
OX327687.1	5	34.17	32
OX327688.1	6	33.92	32.1
OX327689.1	7	33.54	32.1
OX327690.1	8	31.7	32.1
OX327691.1	9	31.31	33.3
OX327692.1	10	31.18	32.1
OX327693.1	11	30.58	32.1
OX327694.1	12	30.54	32.2
OX327695.1	13	30.47	32.1
OX327696.1	14	29.5	32
OX327697.1	15	29.42	32
OX327698.1	16	29.33	32

INSDC accession	Chromosome	Size (Mb)	GC%
OX327699.1	17	29.31	32.8
OX327700.1	18	29.28	32
OX327701.1	19	28.68	32
OX327702.1	20	28.67	32.1
OX327703.1	21	28.14	32.2
OX327704.1	22	27.98	32.2
OX327705.1	23	27.39	32
OX327706.1	24	27.08	32.1
OX327707.1	25	26.09	32.1
OX327708.1	26	25.42	32.1
OX327709.1	27	24.94	32
OX327710.1	28	24.18	32.4
OX327711.1	29	23.78	32.2
OX327712.1	30	23.17	32.2
OX327713.1	31	22.13	32.1
OX327714.1	MT	0.66	45.1
OX327715.1	Pltd	0.16	

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	3.3.8	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
FreeBayes	1.3.1-17-gaa2ace8	https://github.com/freebayes/freebayes
gEVAL	N/A	https://geval.org.uk/
Hifiasm	0.15.3	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
Long Ranger ALIGN	2.2.2	https://support.10xgenomics.com/genome-exome/software/pipelines/latest/advanced/other-pipelines
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MBG	-	https://github.com/maickrau/MBG
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
SALSA	2.2	https://github.com/salsa-rs/salsa

Data availability

European Nucleotide Archive: *Ailanthus altissima* (tree of heaven). Accession number [PRJEB47393](https://identifiers.org/ena.embl/PRJEB47393); <https://identifiers.org/ena.embl/PRJEB47393> (Wellcome Sanger Institute, 2022)

The genome sequence is released openly for reuse. The *A. altissima* genome sequencing initiative is part of the Darwin Tree of Life (DTOL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the [Ensembl](#) pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in [Table 1](#).

Author information

Members of the Royal Botanic Gardens Kew Genome Acquisition Lab are listed here: <https://doi.org/10.5281/zenodo.4786680>.

Members of the Plant Genome Sizing collective are listed here: <https://doi.org/10.5281/zenodo.7994306>.

Members of the Darwin Tree of Life Barcoding collective are listed here: <https://doi.org/10.5281/zenodo.4893703>.

Members of the Wellcome Sanger Institute Tree of Life programme are listed here: <https://doi.org/10.5281/zenodo.4783585>.

Members of Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective are listed here: <https://doi.org/10.5281/zenodo.4790455>.

Members of the Tree of Life Core Informatics collective are listed here: <https://doi.org/10.5281/zenodo.5013541>.

References

- Abdennur N, Mirny LA: **Cooler: Scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chow W, Brugger K, Caccamo M, *et al.*: **gEVAL – a web-based browser for evaluating genome assemblies.** *Bioinformatics.* 2016; **32**(16): 2508–2510. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dallas JF, Leitch M, Hulme PE: **Microsatellites for tree of heaven (*Ailanthus altissima*).** *Mol Ecol Notes.* 2005; **5**(2): 340–342. [Publisher Full Text](#)
- Demeter A, Saláta D, Tormáné Kovács E, *et al.*: **Effects of the invasive tree species *Ailanthus altissima* on the floral diversity and soil properties in the Pannonian region.** *Land.* 2021; **10**(11): 1155. [Publisher Full Text](#)

Desai S: **Cytology of Rutaceae and Simarubaceae.** *Cytologia*. 1960; **25**(1): 28–35.

[Publisher Full Text](#)

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol*. 2017; **35**(4): 316–319.

[PubMed Abstract](#) | [Publisher Full Text](#)

Garrison E, Marth G: **Haplotype-based variant detection from short-read sequencing.** 2012.

[Publisher Full Text](#)

Ghurye J, Rhie A, Walenz BP, *et al.*: **Integrating Hi-C links with assembly graphs for chromosome-scale assembly.** *PLoS Comput Biol*. 2019; **15**(8): e1007273.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics*. 2020; **36**(9): 2896–2898.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Harry E: **PretextView (Paired REad TEXTure Viewer): A desktop application for viewing pretext contact maps.** 2022; [Accessed 19 October 2022].

[Reference Source](#)

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience*. Oxford University Press, 2021; **10**(1): g1aa153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hu SY: **Ailanthus.** *Arnoldia*. 1979; **39**(2): 29–50.

[Reference Source](#)

Ivanova D, Dimitrova D, Vladimirov V: **Chromosome numbers of selected woody species from the Bulgarian flora.** *Phytologia Balcanica*. 2006; **12**(1): 79–84.

[Reference Source](#)

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol*. 2018; **19**(1): 125.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Loureiro J, Rodriguez E, Dolezel J, *et al.*: **Two new nuclear isolation buffers for plant DNA flow cytometry: A test with 37 species.** *Ann Bot*. 2007; **100**(4): 875–888.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Manni M, Berkeley MR, Seppely M, *et al.*: **BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol*. 2021; **38**(10): 4647–4654.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Neophytou C, Pötzelberger E, Curto M, *et al.*: **Population bottlenecks have shaped the genetic variation of *Ailanthus altissima* (Mill.) Swingle in an area of early introduction.** *Forestry: An International Journal of Forest Research*. 2020; **93**(4): 495–504.

[Publisher Full Text](#)

Obermayer R, Leitch IJ, Hanson L, *et al.*: **Nuclear DNA C-values in 30 species double the familial representation in pteridophytes.** *Ann Bot*. 2002; **90**(2):

209–217.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Pellicer J, Powell RF, Leitch IJ: **The application of flow cytometry for estimating genome size, ploidy level endopolyploidy, and reproductive modes in plants.** In: Besse, P. (ed.) *Methods Mol Biol*. New York, NY: Humana, 2021; **2222**: 325–361.

[PubMed Abstract](#) | [Publisher Full Text](#)

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell*. 2014; **159**(7): 1665–1680.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rautiainen M, Marschall T: **Minimizer-based sparse de Bruijn Graph construction.** *Bioinformatics*. 2021; **37**(16): 2476–2478.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature*. 2021; **592**(7856): 737–746.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol*. 2020; **21**(1): 245.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Saina JK, Li ZZ, Gichira AW, *et al.*: **The complete chloroplast genome sequence of Tree of Heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree.** *Int J Mol Sci*. 2018; **19**(4): 929.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Saina JK, Li ZZ, Mekbib Y, *et al.*: **Transcriptome sequencing and microsatellite marker discovery in *Ailanthus altissima* (Mill.) Swingle (Simaroubaceae).** *Mol Biol Rep*. 2021; **48**(3): 2007–2023.

[PubMed Abstract](#) | [Publisher Full Text](#)

Simão FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics*. 2015; **31**(19): 3210–3212.

[PubMed Abstract](#) | [Publisher Full Text](#)

Surana P, Muffato M, Qi G: **sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo*. 2023a.

[Publisher Full Text](#)

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote v1.0.dev (v1.0.dev).** *Zenodo*. 2023b; [Accessed 17 April 2023].

[Publisher Full Text](#)

Swingle WT: **The early European history and the botanical name of the Tree of Heaven, *Ailanthus altissima*.** *J Wash Acad Sci*. 1916; **6**: 490–498.

[Reference Source](#)

Vasimuddin Md, Misra S, Li H, *et al.*: **Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2019; 314–324.

[Publisher Full Text](#)

Wellcome Sanger Institute: **The genome sequence of the tree of heaven *Ailanthus altissima* (Mill.) Swingle, 1916.** European Nucleotide Archive, [dataset], accession number PRJEB47666, 2022.

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 31 August 2023

<https://doi.org/10.21956/wellcomeopenres.21744.r63916>

© 2023 Leebens-Mack J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jim Leebens-Mack 

Department of Plant Biology, 4505 Miller Plant Sciences, University of Georgia, Athens, Georgia, USA

The authors present a reference quality, chromosomal genome assembly for *Ailanthus altissima*. As far as I know, this is the first chromosomal genome assembly for any member of the family Simaroubaceae (Sapindales) and it will be quite useful for comparative genome analyses across rosids, the eudicots and all flowering plants.

A very minor point for this report and all Darwin Tree of Life genome reports - scaffold N50 stats are meaningless for chromosomal assemblies made using "3rd generation" sequencing technologies. The N50 stats reflect the chromosomal size distributions rather than the quality of the assembly. Numbers of gaps in the scaffold assemblies are more informative, but again the gap number may say more about the distribution of repeats rather than the quality of the assembly. If the authors chose to keep the N50 statistics in this report, it would be helpful to add a sentence or two on how they should be interpreted.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Phylogenomics and comparative plant genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 16 August 2023

<https://doi.org/10.21956/wellcomeopenres.21744.r63917>

© 2023 Liu X. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Xin Liu 

State Key Laboratory of Agricultural Genomics, BGI (Beijing Genomics Institute)-Shenzhen, Shenzhen, China

The manuscript by Rowan *et al.* described the genome assembly of *Ailanthus altissima*. They applied PacBio, 10X and HiC sequencing to obtain the sequencing data, and they assembled a chromosome-level genome, with good quality. The method used should be appropriate thus the generated dataset should be valuable for future studies. This type of manuscript is quite straightforward, and I think it is OK for publish as data note. I only have several suggestions for the authors to consider.

1. With the cytometric evaluation, the genome size was estimated to be 1,170 Mb, which is larger than the assembled genome. Although there would be some acceptable reasons for this, it would be good to have some more information. Have you estimated the genome size using Kmer analysis? If so, it would be good to know the estimation using the sequencing data.
2. The authors mentioned about the manual curation of the genome assembly to remove some assembly errors. Especially for the haplotypic duplication, it would be better to mention the definition of this type of assembly error. Also, for the contigs corresponding to the second haplotype, it would be more informative to mention the size of the contigs.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
