DATA NOTE

The genome sequence of strawberry clover, *Trifolium fragiferum* L. (Fabaceae) [version 1; peer review: awaiting peer review]

Sahr Mian[1], Maarten J. M. Christenhusz[1,2], Ilia J. Leitch[1], Andrew R. Leitch[3], Michael F. Fay[1], Royal Botanic Gardens Kew Genome Acquisition Lab, Plant Genome Sizing collective, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team, Wellcome Sanger Institute Scientific Operations: Sequencing Operations, Wellcome Sanger Institute Tree of Life Core Informatics team, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Royal Botanic Gardens Kew, Richmond, England, UK
[2]Curtin University, Perth, Western Australia, Australia
[3]Queen Mary University of London, London, England, UK

v1 **First published:** 20 Mar 2024, **9**:158
https://doi.org/10.12688/wellcomeopenres.21162.1
Latest published: 20 Mar 2024, **9**:158
https://doi.org/10.12688/wellcomeopenres.21162.1

Abstract

We present a genome assembly from an individual *Trifolium fragiferum* (strawberry clover; Tracheophyta; Magnoliopsida; Fabales; Fabaceae). The genome sequence is 512.0 megabases in span. Most of the assembly is scaffolded into 8 chromosomal pseudomolecules. The mitochondrial and plastid genome assemblies have lengths of 298.57 kilobases and 139.15 kilobases in length, respectively.

Keywords

Trifolium fragiferum, strawberry clover, genome sequence, chromosomal, Fabales

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

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: Mian S: Investigation, Resources; **Christenhusz MJM**: Investigation, Resources, Writing – Original Draft Preparation; **Leitch IJ**: Investigation, Resources, Writing – Review & Editing; **Leitch AR**: Investigation, Resources; **Fay MF**: Writing – Original Draft Preparation;

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.

How to cite this article: Mian S, Christenhusz MJM, Leitch IJ *et al.* **The genome sequence of strawberry clover, *Trifolium fragiferum* L. (Fabaceae) [version 1; peer review: awaiting peer review]** Wellcome Open Research 2024, **9**:158 https://doi.org/10.12688/wellcomeopenres.21162.1

First published: 20 Mar 2024, **9**:158 https://doi.org/10.12688/wellcomeopenres.21162.1

Species taxonomy

Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade; NPAAA clade; Hologalegina; IRL clade; Trifolieae; *Trifolium*; *Trifolium fragiferum* L., 1753 (NCBI: txid97023).

Background

Strawberry clover, *Trifolium fragiferum*, is a creeping perennial herb in the bean family Fabaceae. It produces stolons that root at the nodes, forming large mats or clumps. Like most clover species it has leaves composed of three leaflets that have a serrate edge and grow to about 2 cm across. The pale pink flowers are clustered in a head of around a centimetre across (Figure 1a), which enlarges into a 2 cm-diameter globe of inflated sepals that are furry and tinged pink (Figure 1b and c). This gave it its common name, because these globular heads resemble a strawberry.

Trifolium fragiferum is native to Eurasia and Northern Africa, east to India and western China, but it is widely cultivated and often naturalises outside its native range, such as in North and South America, Japan, Australia and New Zealand (POWO, 2023). It is cultivated as a cover crop, as green manure, as a bee plant, and for hay and silage. It is often used in flood-prone areas as it can survive flooding for up to two months. In addition, it can tolerate a saline soil, and short-term drought. Several agricultural cultivars such as 'Fresa', 'Grasslands Onward', 'Grasslands Upward', 'Palestine', 'Prinsep Park', 'O' Connors' and 'Salina' have been developed in the USA and Australia, which have been selected for different attributes (USDA, 2022).

In the UK, *T. fragiferum* is widespread in southern and eastern England. Further west and north, it is mostly coastal, where it can be found on the coast in saltmarshes, on sea-walls and in grazing marshes where this sample was collected. It also occurs inland at low elevations in pastures and along tracks on damp, usually clay soils. It can sometimes occur as a weed in lawns, but it is declining somewhat due to pasture improvements and conversion of pastures to arable land.

Here we present the first chromosomally complete genome sequence for this species, sequenced as part of the Darwin Tree of Life Project. This genome sequence will be useful for the development of new cultivars for high protein fodder grown on flood-prone or saline land.

Genome sequence report

The genome was sequenced from a specimen of *Trifolium fragiferum* (Figure 1) collected from Cuckmere Haven, East Sussex, UK (50.77, –0.15). Using flow cytometry, the genome size (1C-value) was estimated to be 0.62 pg, equivalent to 610 Mb. A total of 33-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C

Figure 1. Photographs of the *Trifolium fragiferum* (drTriFrag1) specimen used for genome sequencing. (**A**) pale pink flowers clustered into the inflorescence, (**B**) and (**C**) globe of inflated sepals.

data. Manual assembly curation corrected 117 missing joins or mis-joins and removed 8 haplotypic duplications, reducing the assembly length by 0.52% and the scaffold number by 79.00%, and also decreasing the scaffold N50 by 6.62%.

The final assembly has a total length of 512.0 Mb in 19 sequence scaffolds with a scaffold N50 of 63.4 Mb (Table 1). The snail plot in Figure 2 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 3. The cumulative assembly plot in Figure 4 shows curves for subsets of scaffolds assigned to different phyla. Most (99.85%) of the assembly sequence was assigned to 8 chromosomal-level scaffolds. This is in agreement with cytological data published for this species which report it to be diploid with 2n = 16 (Lukjanová and Řepková 2021). Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 5; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial and plastid genomes were also assembled and can be found as contigs within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 57.2 with *k*-mer completeness of 99.99%, and the assembly has a BUSCO v5.3.2 completeness of 98.5% (single = 95.3%, duplicated = 3.3%), using the fabales_odb10 reference set (n = 5,366).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/97023.

Table 1. Genome data for *Trifolium fragiferum*, drTriFrag1.1.

Project accession data		
Assembly identifier	drTriFrag1.1	
Species	*Trifolium fragiferum*	
Specimen	drTriFrag1	
NCBI taxonomy ID	97023	
BioProject	PRJEB52577	
BioSample ID	SAMEA10369849	
Isolate information	drTriFrag1: leaf	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	57.2	*≥ 50*
k-mer completeness	99.99%	*≥ 95%*
BUSCO**	C:98.5%[S:95.3%,D:3.3%], F:0.3%,M:1.1%,n:5,366	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.85%	*≥ 95%*
Sex chromosomes	None	*localised homologous pairs*
Organelles	Mitochondrial genome: 298.57 kb Plastid genome: 139.15 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR9793191	
Hi-C Illumina	ERR9682491	
PolyA RNA-Seq Illumina	ERR10378010	
Genome assembly		
Assembly accession	GCA_954870985.1	
Accession of alternate haplotype	GCA_954870555.1	
Span (Mb)	512.0	
Number of contigs	222	
Contig N50 length (Mb)	6.2	
Number of scaffolds	19	
Scaffold N50 length (Mb)	63.4	
Longest scaffold (Mb)	85.4	

* 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 fabales_odb10 BUSCO set using version 5.3.2. 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/drTriFrag1_1/dataset/drTriFrag1_1/busco.

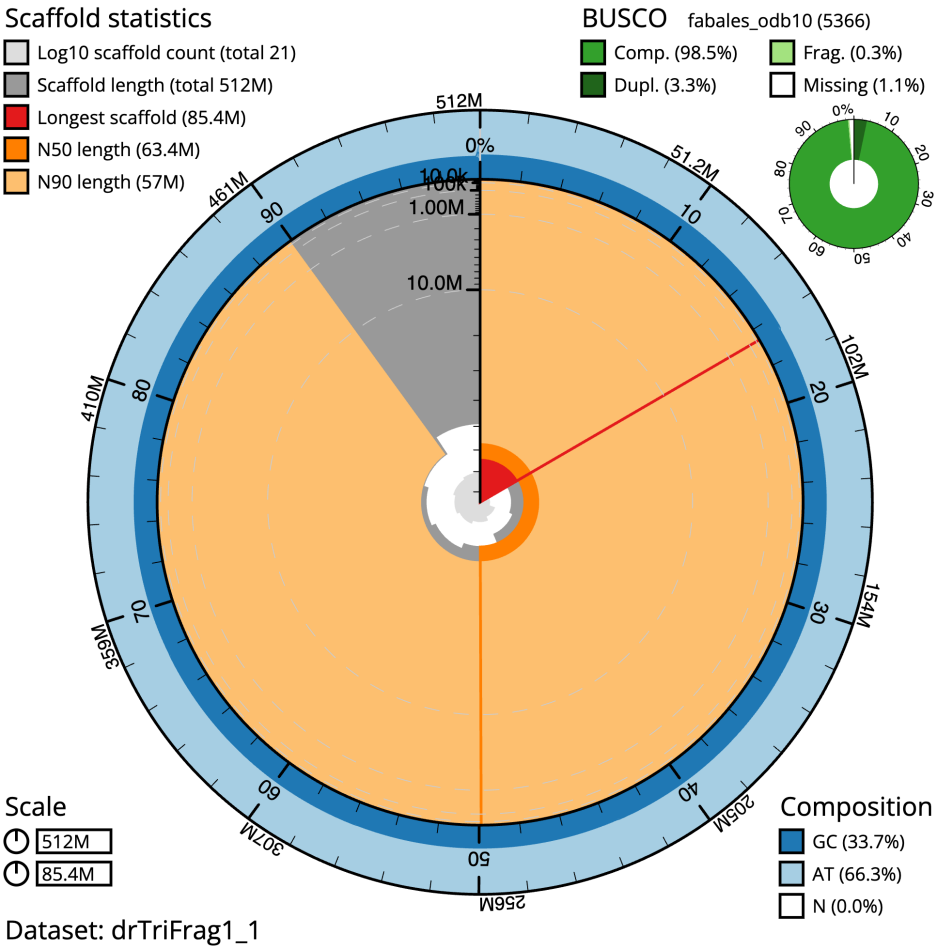

Figure 2. Genome assembly of *Trifolium fragiferum*, drTriFrag1.1: metrics. The BlobToolKit snail plot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 512,414,909 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (85,398,858 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (63,352,505 and 57,030,410 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 fabales_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/drTriFrag1_1/dataset/drTriFrag1_1/snail.

Methods

Sample acquisition, genome size estimation and nucleic acid extraction

A specimen of *Trifolium fragiferum* (specimen ID KDTOL10398, ToLID drTriFrag1) was collected by hand from Cuckmere Haven, East Sussex, UK (latitude 50.77, longitude –0.15) on 2021-09-07 by Sahr Mian, Maarten Christenhusz, Ilia Leitch (Royal Botanic Gardens Kew) and Andrew Leitch (Queen Mary University of London) and identified by Maarten Christenhusz, and 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). 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 *Solanum lycopersicum* 'Stupiké polní rané' with an assumed 1C-value of 968 Mb (Doležel *et al.*, 2007).

The workflow for high molecular weight (HMW) DNA extraction at the Wellcome Sanger Institute (WSI) includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the drTriFrag1 sample was weighed and dissected on dry ice (Jay *et al.*, 2023). For sample homogenisation,

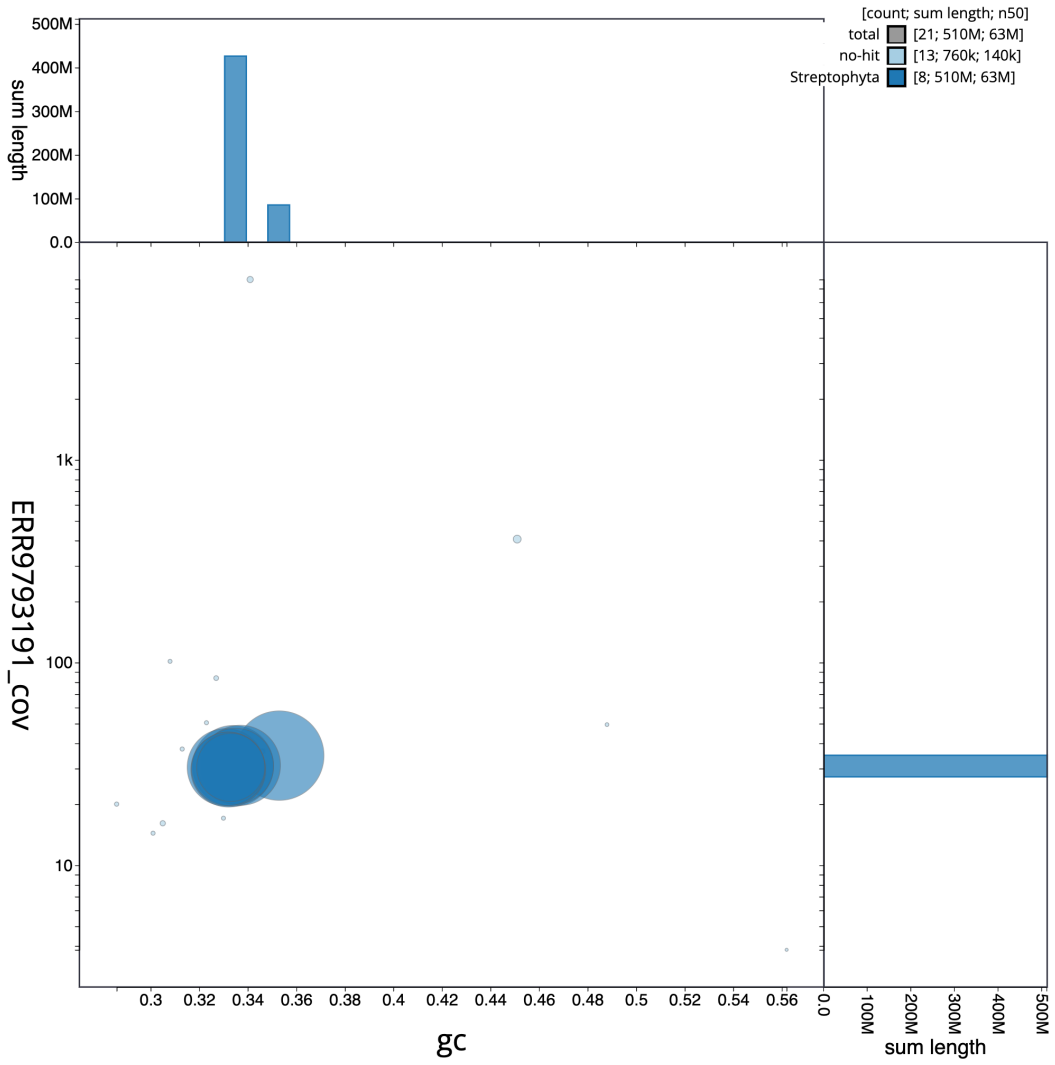

Figure 3. Genome assembly of *Trifolium fragiferum*, drTriFrag1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/drTriFrag1_1/dataset/drTriFrag1_1/blob.

leaf tissue was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023).

HMW DNA was extracted using the Automated Plant MagAttract v2 protocol (Todorovic *et al.*, 2023a). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system with speed setting 30 (Todorovic *et al.*, 2023b). Sheared DNA was purified by solid-phase reversible immobilisation (Strickland *et al.*, 2023): in brief, the method employs a 1.8X ratio of AMPure PB beads to sample to eliminate shorter fragments and concentrate the DNA. 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 drTriFrag1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Protocols developed by the WSI Tree of Life core laboratory are publicly available on protocols.io (Denton *et al.*, 2023).

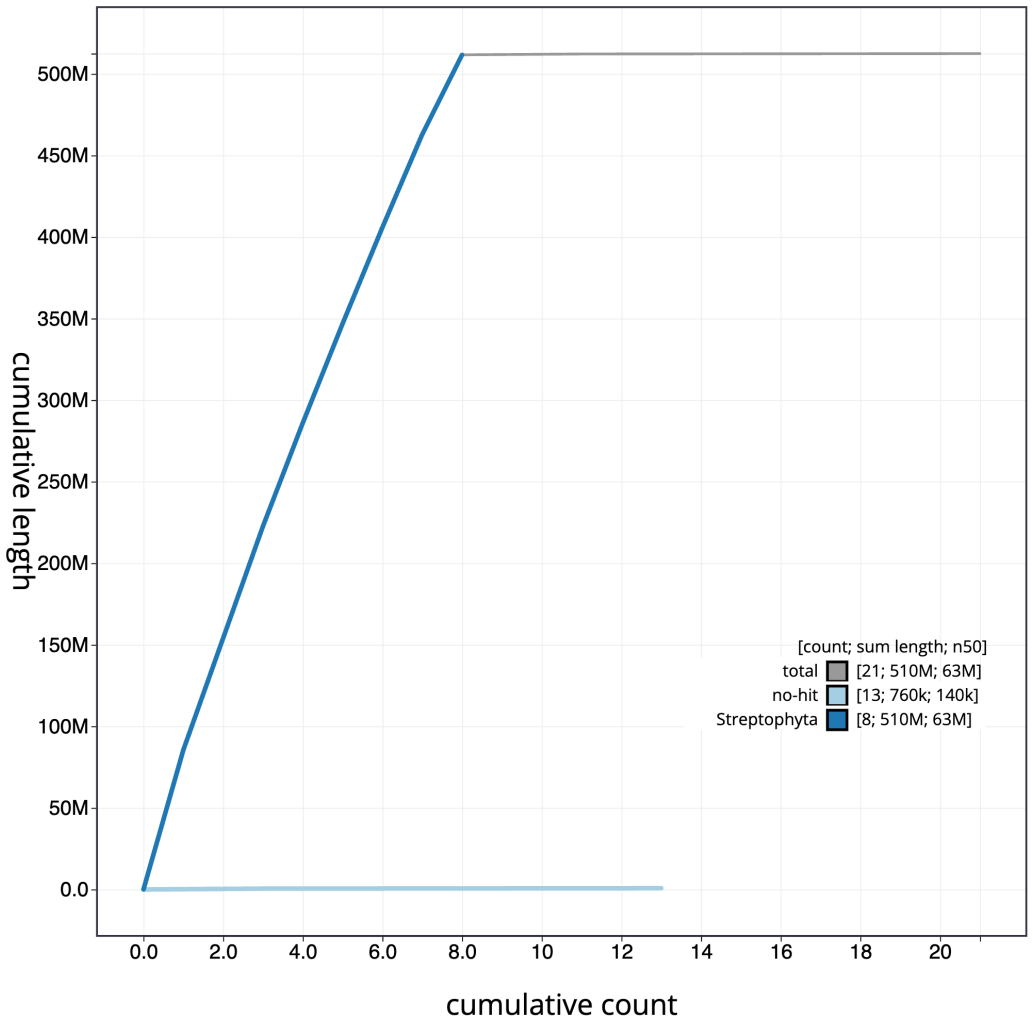

Figure 4. Genome assembly of *Trifolium fragiferum*, drTriFrag1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/drTriFrag1_1/dataset/drTriFrag1_1/cumulative.

Sequencing

Pacific Biosciences HiFi circular consensus 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 was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from leaf tissue of drTriFrag1 using the Arima2 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) The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (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

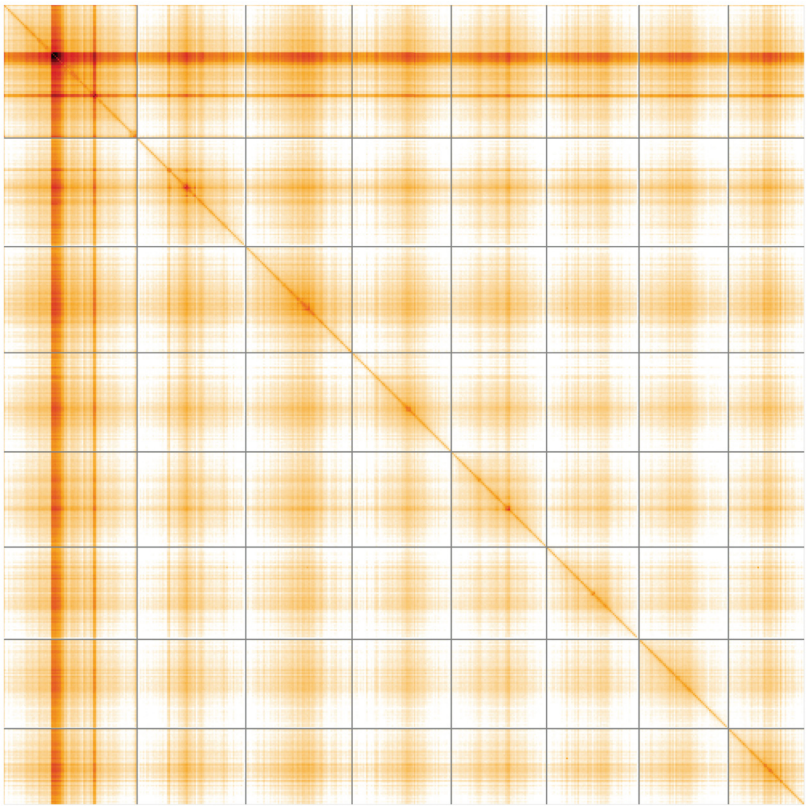

Figure 5. Genome assembly of *Trifolium fragiferum*, drTriFrag1.1: Hi-C contact map of the drTriFrag1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=B-bHEkujSUu9DCM1xffcJw.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Trifolium fragiferum*, drTriFrag1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX940789.1	1	85.4	35.5
OX940790.1	2	69.39	33.5
OX940791.1	3	67.84	33.5
OX940792.1	4	63.35	33.0
OX940793.1	5	60.9	33.5
OX940794.1	6	59.02	33.0
OX940795.1	7	57.03	33.0
OX940796.1	8	48.73	33.5
OX940797.1	MT	0.3	45.0
OX940798.1	Pltd	0.14	34.0

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.

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.

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://github.com/chhylp123/hifiasm
HiGlass	1.11.6	https://github.com/higlass/higlass
MBG	-	https://github.com/maickrau/MBG
Merqury	MerquryFK	https://github.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://github.com/marcelauliano/MitoHiFi
OATK	0.2	https://github.com/c-zhou/oatk
PretextView	0.2	https://github.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://github.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://github.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://github.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://github.com/c-zhou/yahs

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

Data availability

European Nucleotide Archive: *Trifolium fragiferum*. Accession number PRJEB52577; https://identifiers.org/ena.embl/PRJEB52577 (Wellcome Sanger Institute, 2023). The genome sequence is released openly for reuse. The *Trifolium fragiferum* 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 Garden Edinburgh Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4786682.

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 Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

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

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

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

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life Wet Laboratory Protocol Collection V.1.** *protocols.io.* 2023.
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

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA Extraction: Automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Doležel J, Greilhuber J, Suda J, *et al.*: **Estimation of nuclear DNA content in plants using flow cytometry.** *Nat Protoc.* 2007; **2**(9): 2233–2244.
PubMed Abstract | Publisher 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): giaa153.
PubMed Abstract | Publisher Full Text | Free Full Text

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life Sample Preparation: Triage and Dissection.** *protocols.io.* 2023.
Publisher Full Text

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

Lukjanová E, Řepková J: **Chromosome and genome diversity in the genus *Trifolium* (Fabaceae).** *Plants (Basel).* 2021; **10**(11): 2518.
PubMed Abstract | Publisher Full Text | Free Full Text

Manni M, Berkeley MR, Seppey 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

Narváez-Gómez JP, Mbye H, Oatley G, *et al.*: **Sanger Tree of Life Sample Homogenisation: Covaris cryoPREP® Automated Dry Pulverizer V.1.** *protocols.io.* 2023.
Publisher 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. (Clifton, N.J.)*. New York, NY: Humana, 2021; **2222**: 325–361.
PubMed Abstract | Publisher Full Text

POWO: **Plants of the World Online**. Royal Botanic Gardens, Kew, 2023.
Reference Source

Rao SSP, Huntley MH, Durd 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: **MBG: 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

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

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life Fragmented DNA clean up: Manual SPRI.** *protocols.io.* 2023.
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).** *Zenodo.* 2023b.
Publisher Full Text

Todorovic M, Oatley G, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated Plant MagAttract v.2.** *Protocols.Io.* 2023a.
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA Fragmentation: Diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023b.
Publisher Full Text

USDA: **Strawberry clover.** United Stated Department of Agriculture, Natural Resources Conservation Service - Plant Guide, 2022; Accessed 2 August 2022.
Reference Source

Vasimuddin M, 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 strawberry clover, *Trifolium fragiferum* L. (Fabaceae).** European Nucleotide Archive, [dataset], accession number PRJEB52577, 2023.

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | Publisher Full Text | Free Full Text