Efficient hybrid strategies for assembling the plastome, mitochondriome, and large nuclear genome of diploid *Ranunculus cassubicifolius* (Ranunculaceae)

Karbstein, Kevin[1,2,4*], Tomasello, Salvatore[1], Wagner, Natascha[1], Barke, Birthe Hilkka[1], Paetzold, Claudia[3], Bradican, John Paul[1], Preick, Michaela[5], Irisarri, Iker[6,7,9], de Vries, Jan[6,7,8] & Hörandl, Elvira[1]

[1] *University of Göttingen, Albrecht-von-Haller Institute for Plant Sciences, Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), Göttingen, Germany*

[2] *University of Göttingen, Georg-August University School of Science (GAUSS), Göttingen, Germany*

[3] *Senkenberg Naturhistorische Sammlungen, Dresden, Germany*

[4] *Max Planck Institute for Biogeochemistry, Department of Biogeochemical Integration, Jena, Germany*

[5] *Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany*

[6] *University of Göttingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Göttingen, Germany*

[7] *University of Göttingen, Campus Institute Data Science (CIDAS), Göttingen, Germany*

[8] *University of Göttingen, Göttingen Center for Molecular Biosciences (GZMB), Department of Applied Bioinformatics, Göttingen, Germany*

[9] *Leibniz Institute for the Analysis of Biodiversity Change (LIB), Centre for Molecular Biodiversity Research, Museum of Nature, Hamburg, Germany*

ORCID:

Kevin Karbstein 0000-0003-1424-6557

Salvatore Tomasello 0000-0001-5201-1156

Natascha Wagner 0000-0001-6623-7623

Birthe Hilkka Barke 0000-0001-5379-6055

Claudia Paetzold 0000-0002-4128-6645

Iker Irisarri 0000-0002-3628-1137

Jan de Vries 0000-0003-3507-5195

Elvira Hörandl 0000-0002-7600-1128

* **corresponding author:** *email: kkarb@bgc-jena.mpg.de, elvira.hoerandl@biologie.uni-goettingen.de*

Running Head

Efficient hybrid assembly strategies for plant genomes

Abstract

Currently, we lack efficient strategies - in terms of laboratory work and cost, as well as assembly quality - to reliably reconstruct large and complex genomes from non-model plants. This often impairs the study of evolutionary intricate groups characterized by hybridization, polyploidy, and/or apomixis. The large *Ranunculus auricomus* polyploid complex is an angiosperm model system for apomixis, evolution, and biogeography research. However, neither plastid, mitochondrial, nor high-quality nuclear genomes are available, and a genus-wide draft genome has yet to be assembled, limiting phylogenomic, ecological, and taxonomic analyses thus far. We tested different Illumina (short, highly accurate reads), Oxford Nanopore Technology (ONT; long, less accurate reads), and Illumina-ONT-hybrid assembly strategies using the diploid species *R. cassubicifolius*, a sexual diploid progenitor of the *R. auricomus* complex, and selected the best assemblies in terms of completeness, contiguity, and BUSCO quality scores. We first assembled the entire plastome (156 Mbp; 78 protein-coding genes, 30 tRNAs, and 4 rRNAs) and mitochondriome (1,260 Mbp; 39 protein-coding genes, 13 tRNAs, 3 rRNAs), using an Illumina- and hybrid-based strategy, respectively. The best out of 11 strategies used raw Nanopore reads assembled and polished by filtered Nanopore and Illumina reads accompanied by genome annotation yielded a 3.45 Gbp high-quality genome with (>99% completeness), 97% single-copy and duplicated complete BUSCO genes found, median contig length (N_{50}) of 104 Mbp, and 39,544 predicted

genes. The genomic information presented herein helps to improve phylogenomic analyses in this species complex, and will enable improved functional, biogeographic, and population genomic analyses for the *R. auricomus* complex, the genus, and beyond in Ranunculaceae in the future.

Keywords

hybrid assembly, large diploid plant genomes, *de novo* assembly, plastome, mitochondriome, genome, Illumina sequencing, Nanopore sequencing, gDNA, *Ranunculus auricomus*

Significance Statement

The *Ranunculus auricomus* complex is a model system in flowering plant apomixis, evolution, and biogeography research, but thus far no genus-wide complete genome exists. Using Illumina and/or Nanopore data, we assembled the plastome, mitochondriome, and high-quality diploid genome. We created efficient assembly strategies, which are applicable to other non-model systems and useful for improving phylogenomic analyses, enabling better functional (e.g., apomixis), biogeographic, and population genomic analyses at the complex, but also at the genus and family level.

Resource Article

Introduction

Over the past decade, the scientific community has experienced a massive expansion in accessible sequencing technologies. Biologists have moved beyond short-read (100-250 bp) Illumina sequence data, which often produce incomplete genetic assemblies and lack the information required to assemble repetitive regions that plague large genomes. This results in highly fragmented, erroneous *de novo* genome assemblies (Gatter et al., 2021; Rhie et al., 2021; Zimin and Salzberg, 2022). Today, long-read sequencing technologies, such as Nanopore (Oxford Nanopore Technologies, ONT) or PacBio (Pacific Biosciences), can bridge such gaps and improve the quality of genome assemblies. DNA reads derived from PacBio are typically up to 25 kbp in length; in contrast, ONT sequencing is typically less accurate per base (ca. 90%, although new ONT generations can also reach up to 99%) but can provide much longer reads of up to 500 kbp or even 1.5 Mbp, and can be easily performed in any laboratory (reviewed in Pucker et al., 2022). Three strategies are currently available for genome assembly: (i) Illumina short-read assembly (e.g., Lu et al., 2020; Zimin and Salzberg, 2022), (ii) ONT long-read assembly (e.g., Dmitriev et al., 2021), or (iii) Illumina and ONT hybrid assembly (e.g., Prjibelski et al., 2020; Gatter et al., 2021; Zimin and Salzberg, 2022). For any of the three cases, assembly quality and completeness can be improved with gap filling using trusted contigs from prior assemblies, or with different polishing strategies using available, quality-filtered Illumina- and/or ONT-reads.

In the latter two cases, the combination of low- to medium-coverage long-read data with highly accurate Illumina sequences is particularly attractive for improving genome assemblies when resources are limited, as is usually the case when working with non-model plants. Scientists therefore aim to produce “high-quality” genomes, i.e., genome assemblies characterized by, for example, high completeness (>90%), contiguity (N_{50}>1 Mpb), and >90% recovery rate of conserved single-copy genes (e.g., BUSCO; assembly standards of the Earth Biogenome Project; Lawniczak et al., 2022). However, some metrics are rather adapted to animals, for example, flowering plants have undergone many ancient or recent polyploidization events (Van De Peer et al., 2017; Leebens-Mack et al., 2019; Hörandl, 2022) and are characterized by quite large, repetitive genomes, and thus a BUSCO score of >90% single-copy genes can often not be met (Dmitriev et al., 2021) or is not informative for quality assessment of large, repetitive genomes (Mokhtar et al., 2023). Therefore, other thresholds,

metrics, or characteristics have been proposed (Kress et al., 2022; Hotaling et al., 2023; Mokhtar et al., 2023). Irrespective of chosen metrics, efficient strategies in terms of laboratory effort, cost, and assembly quality are needed to reliably reconstruct large and complex high-quality genomes for non-model organisms. This often hampers the study of evolutionary, biogeographic, and ecological relationships within taxonomically complex groups (TCGs), which are characterized by intricate evolutionary processes such as hybridization, polyploidy, and/or apomixis (Karbstein et al., under review; Ennos et al., 2005).

The *Ranunculus auricomus* complex represents one of such TCGs and is one of the largest polyploid apomictic plant species groups in Europe, comprising more than 840 described morphospecies (Hörandl and Raab-Straube, 2015). Individuals are distributed across Eurasia, ranging from Greenland to Europe, Siberia, and Alaska, and from the Arctic through to temperate and Mediterranean climate zones (Jalas and Suominen, 1989; GBIF Secretariat, 2021). The *R. auricomus* complex is a model system for understanding the evolution and biogeography of phylogenetically young, reticulately evolving flowering plant groups characterized by reticulate evolution (Paun et al., 2006; Hörandl et al., 2009; Pellino et al., 2013; Hodač et al., 2018; Karbstein et al., 2020b, 2020a, 2021, 2022; Tomasello et al., 2020; Karbstein, 2021). It is also a model system for unraveling the molecular, cytogenetic, and evolutionary mechanisms of apomixis (i.e., the asexual reproduction via seeds; Hojsgaard et al., 2014; Barke et al., 2018, 2020; Hodač et al., 2019; Pellino et al., 2020; Paetzold et al., 2022). Apomixis is relevant for agricultural applications (e.g., Schmidt, 2020), and is often associated with hybridization and/or polyploidy, which are considered major evolutionary forces in plant evolution. All flowering plants share a polyploid ancestor, and at least 30% of flowering plant species are of recent polyploid origin, while more than 25% of extant plant species are known to be involved in interspecific hybridization or introgression events (Mallet, 2005; Soltis et al., 2015; Van De Peer et al., 2017; Leebens-Mack et al., 2019; Hörandl, 2022).

Consequently, improving (phylo)genomic approaches for di- and polyploids, and observing regulatory mechanisms of apomixis (e.g., Brukhin et al., 2019) with *R. auricomus* reference genomes will help to develop a better understanding of plant evolution and biodiversity, and is therefore of general interest in biological research. Organellar genome reconstruction is valuable for studying nuclear-organelle compatibility and the chimerism of cyto-nuclear complexes (Sloan et al. 2018, Moller et al. 2021). Little is known about plant mitochondrial genomes although mitochondrial-nuclear compatibility is important for

speciation, evolution of crossing barriers, and for emergence of male cytoplasmic sterility (Postel and Touzet, 2020; Møller et al., 2021). Elucidating cyto-nuclear complexes could also be important in the study of the evolution of sex (Havird et al., 2015).

The *R. auricomus* complex contains five newly circumscribed, predominantly diploid sexual species (Karbstein et al., 2020b), which have generated the more than 800 described polyploid apomictic taxa within the last 0.8 million years (Tomasello et al., 2020; Karbstein et al., 2022). The diploid genome size of the *R. auricomus* complex (2n = 16) ranges between 6.12 and 6.30 pg DNA based on DNA image cytometry (Hörandl and Greilhuber, 2002). Therefore, *R. auricomus* genomes are large even at the diploid level (diploid genome sizes are estimated to be between 5.998-6.174 Gbp, Leitch et al., 2019). Increasing proportions of repetitive regions have usually been considered responsible for genome size inflation (Bennetzen, 2005), which could also be relevant for *R. auricomus* genomes (Hörandl et al., 2009; Hodač et al., 2018; Karbstein et al., 2022). However, so far, reference genomes for the species complex, but also for the entire genus *Ranunculus*, are still missing. Among Ranunculaceae, the only other reference genomes are available from diploid *Aquilegia* species (Xie et al., 2020; Mokhtar et al., 2023), but these genomes are much smaller (mean 0.9 Gbp, Leitch et al. 2019), and no polyploid complexes appear in this genus (TROPICOS database vers. 3.4.2, http://www.tropicos.org; CCDB database vers. 1.66, http://ccdb.tau.ac.il/home/, Rice et al., 2015).

The evolutionary relationships among diploid sexual *R. auricomus* progenitors have recently been clarified (Tomasello et al., 2020; Karbstein et al. 2020b), as noted above, and initial insights into the allopolyploid origins and genome evolution of derivatives have been obtained (Karbstein et al. 2021a, 2022). A *Ranunculus* reference genome would be a major step forward in improving the reconstruction of the species by facilitating read mapping to orthologous loci, allele phasing, and subgenome assignments; it would be the starting point for sophisticated approaches like reference-guided assemblies (e.g., Eaton and Overcast, 2020), polyploid phasing used for multilabeled, subgenome species trees (e.g., Dauphin et al., 2018) or allopolyploid network reconstructions (e.g., Jones, 2017; Šlenker et al., 2021), or functional genomics (e.g., in relation to apomixis, adaptation, and gene expression; Syngelaki et al., 2021). High-quality genomes are essential for studying loci under selection and mutation-selection dynamics (e.g., Del Amparo et al., 2021), which could so far only be performed on (incomplete) transcriptomes (Pellino et al., 2013; Paetzold et al., 2022). In addition, the reconstruction of reference plastomes could help to reveal nuclear-plastid discordances and thus reticulation signals (e.g., hybridizations, McKain et al., 2018; Karbstein

et al., 2022). Mitochondrial information would provide additional insights into discordances with the nuclear genome and plant sex determination (e.g., male sterility; Park et al., 2021). To generate a plastome, mitochondriome, and a high-quality genome of a diploid *R. auricomus*, we used different assembly strategies using short but highly accurate Illumina reads with medium genome coverage and long but less accurate ONT reads with low genome coverage, as it is a typical scenario for working with non-model plant groups.

Methods

Plant Material

In the present study, analyses were performed on two diploid sexual individuals of the species *Ranunculus cassubicifolius* within the *R. auricomus* complex (taxonomic treatment follows Karbstein et al., 2020a). Measurements of ploidy and reproduction mode are published in Hörandl and Greilhuber (2002; under *R. carpaticola*), and Karbstein et al. (2020) and (2021). Fresh leaf material was directly harvested in 2019 from individual EH8483/10, originally sampled in Revúca, Slovakia (48°41'39"N, 20°07'42"E) in Slovakia (Hörandl and Greilhuber, 2002; under *R. carpaticola*). High-molecular-weight DNA was extracted immediately after harvest and used for Illumina DNA sequencing. The same population was already used for transcriptome (Pellino et al., 2013) and microsatellite analyses (Paun et al., 2006; Hodač et al., 2018). We harvested fresh young, 12 hours dark-adapted, leaf material from another conspecific individual originated from nearby (LH040/02; 48°40'38"N, 20°06'24"E, Karbstein et al., 2020a) for ONT gDNA sequencing in 2021 because the previously used individual died. All plants were cultivated at the Old Botanical Garden of the University of Göttingen.

DNA extraction

For Illumina short-read sequencing, gDNA was directly extracted from approximately 200 mg of fresh leaf material from a single *R. cassubicifolius* s.l. individual (EH8483) using liquid nitrogen and the DNeasy PowerPlant Pro Kit (Qiagen, Hilden, Germany) according to the manufacturer´s instructions. The DNA concentration was assessed using the Qubit Fluorometer 2.0 and the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Waltham, USA), and the DNA quality was checked by the Genomic DNA ScreenTape device (Agilent Technologies, Santa Clara, USA). The extraction yielded 16 ng/µl of DNA with a median fragment length of 19 kb.

For ONT long-read sequencing, gDNA extraction was mainly performed according to the Qiagen protocol specifically developed for the isolation of gDNA from plants and fungi and the manufacturer´s instructions of the Qiagen Genomic DNA Buffer Set using Genomic Tip 20/G (details in Text S1). We again used fresh leaf material from a single *R. cassubicifolius* s.l. individual (LH040), and performed three gDNA extractions in parallel. Several protocol modifications were applied to increase DNA yield and to optimize DNA quality (see also Vaillancourt and Buell, 2019; Li et al., 2020). Fresh leaf material (ca. 75 mg) was frozen with liquid nitrogen and immediately finely ground using a mortar and pestle. The still frozen leaf powder was transferred into a 2.0 ml Eppendorf tube containing 400 µl API Qiagen extraction buffer (Qiagen DNeasy Plant Mini Kit). The mixture was incubated at 40°C for 1.5 hours with gentle agitation. Then, 4 µl Qiagen RNase A (100 mg/ml) was added, and the mixture was incubated again for 30 minutes at 37°C, followed by the addition of 18 µl Qiagen Proteinase K (20 mg/ml), and incubation for 2 h at 50°C with gentle agitation. The mixture was centrifuged at 12,000 x g for 20 minutes, and ca. 400 µl of the lysate was transferred (using wide-bore pipettes) to a buffer QBT-equilibrated Genomic Tip. The DNA was washed with 5 x 1 ml Buffer QC, and eluted with 0.8 ml pre-warmed (55°C) Buffer QF into a new 2.0 ml Eppendorf tube. The DNA was precipitated by adding 0.7 volume of room-temperature isopropanol and inverting the tube up to 20 times, followed by 15 minutes of incubation. Using a glass rod, the DNA was rolled up and placed into a new 1.5 ml Eppendorf tube containing ice-cold 80% ethanol. The DNA was air dried, resuspended in 200 µl autoclaved ddH_2O, and samples were pooled in a single tube. The tube was stored at 4 °C until library preparation for ONT sequencing was finished.

The DNA concentration was assessed using the Qubit Fluorometer 3.0 and the Qubit dsDNA HS Assay Kit, and the DNA quality was checked using the NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific, Waltham, USA; Text S1). The DNA length and RNA absence were assessed by 2% agarose gel electrophoresis (50 V, 90 min) including the Quick-Load 1 kb DNA Ladder (New England Biolabs Inc., Ipswich, USA; Text S1). The extracted gDNA was not fragmented and larger than 10 kb (Fig. S1). We obtained a final DNA yield (after some dilution steps) of 61.2 µg (51.0 ng/µl * 1200 µl ddH_2O).

Genome Sequencing on the Illumina Platform

Library preparation was performed using the 10x Chromium Controller Genomics Kit (10x Genomics, Pleasanton, USA) following the manufacturer's instructions. The DNA library quality was assessed by a qPCR based on the NEBNext Library Quant Kit (New

England Biolabs Inc., Ipswich, USA). Sequencing with standard Illumina primers was performed on a NextSeq 500/550 instrument (Illumina, San Diego, USA) with a read length of 150 + 150 bp (paired-end).

Genome Sequencing on the ONT Platform

Library preparation was conducted using an ONT Ligation Sequencing Kit SQK-LSK110 optimized for high throughput and long reads (ONT, Oxford, UK) and applicable for singleplex gDNA sequencing. We adjusted the DNA concentration to 1000 ng in 47 µl (ca. 21.5 µl/ng). We followed the manufacturer's instructions for library preparation with a few modifications (protocol vers. GDE_9108_v110_revL_10Nov2020, accessible via community.nanoporetech.com). Incubation times were increased up to 15 min, ethanol wash buffer concentration was increased to 80%, and we enriched DNA fragments of more than 3 kb. A hardware check was performed for each device, and a flow cell check was run prior to sequencing runs. Active pores ranged from ca. 800-1500. Flow cell no. 1 was loaded with freshly prepared libraries, flow cell no. 2 with one freshly prepared and one at -80°C (two days) stored libraries, and flow cell no. 3-6 with two at -80°C for 1-2 weeks stored libraries. We used a MinION 101B device, the ONT software MINKNOW vers. 21.11.7, 21.11.9, and 22.03.5 were installed on a local Linux system, and six R9.4.1 flow cells following the manufacturer's instructions for priming and loading. After one or two days of runtime, the old library was exhausted and thus replaced with a second library using the supplied flow cell wash kit (protocol vers. WFC_9120_v1_revB_08Dec2020). Sequencing was run for 3-5 days per flow cell until the majority of pores were inactive (> 99%).

Genomic Data Analysis

For Illumina sequence data, we first converted the raw output into *fastq files using SUPERNOVA vers. 2.1.1 (10x Genomics, San Francisco, USA; Marks et al., 2017; Weisenfeld et al., 2018) and BLC2FASTQ vers. 2.20 (Illumina, San Diego, USA), resulting in 51 Gb raw *fastq files. Read quality was assessed by FASTQC vers. 0.11.4 (reports on FigShare; Andrews, 2010). We performed polyX-tail and adapter trimming, and quality filtering of Illumina raw reads, using ATRIA vers. 3.1.2 (Chuan et al., 2021) with default settings, and BBMAP vers. 37.66 ('repair.sh'; sourceforge.net/projects/bbmap/) to reorder filtered paired-end reads.

For long-read ONT sequencing, six flow cells generated 391 Gb *.fast5 (28-104) output files, 5.77 (0.48-1.81) million reads, with a median read fragment length of 21 (19.6-

23.2) kbp with the longest read being 849 kbp, and an estimated output of 51 Gbp (3.3-11.7). Reports per flow cell can be accessed via FigShare. Subsequent analyses were performed on the local HPC cluster (GWDG, Göttingen, Germany). Basecalling was performed with the ONT software GUPPY vers. 6.0.1 specifying the configuration file "dna_r9.4.1_450bps_fast.cfg" for fast base-calling (in preruns, the difference between 'fast' and much slower 'high accuracy' was very small, see FASTQC files on Figshare). After that, NANOFILT vers. 2.2.0 (De Coster et al., 2018) was applied to remove the first 50 bp of the reads, which were usually of low quality. The read quality of all datasets was checked using FASTQC (reports on FigShare).

The plastid genome. The *R. cassubicifolius* plastome was assembled using Illumina short-read data and the software GETORGANELLE vers. 1.7.5.3 (Jin et al., 2020; as recommended by Freudenthal et al., 2020), specifying the "embplant_pt" database, k-mer size range between 45 and 121, use of all Illumina reads, and as seed the *Ranunculus repens* plastome (154,247 bp, GenBank accession number NC_036976; Dann et al., 2017). The average embplant_pt base coverage was 560. The plastome was annotated for protein-coding genes, tRNA, rRNA, and repeat content (microsatellites, tandem repeats, and dispersed repeats), using CPGAVAS2 (Shi et al., 2019) with default settings and 2,544 reference plastomes (Shi et al., 2019), and drawn using CPGVIEW (Liu et al., 2023). Geneious R11 2023.1.2 (Kearse et al., 2012) was used to check and edit the annotation results. The final plastome sequence incl. annotations was deposited into GenBank (OP250948). The plastome with annotations is also shown in Figure 1.

To investigate sequence differences between *R. cassubicifolius* and *R. repens*, we performed progressiveMAUVE (Darling et al., 2004) whole genome alignment analysis in Geneious. In addition, mapping TEG off-target reads to reference plastomes from sampled or herbarium specimens is a common strategy for constructing plastome phylogenies in evolutionary studies (Folk et al., 2015; Tomasello et al., 2020). Therefore, we followed the procedures described in the study by Karbstein et al. (2022) to examine the effect of read mapping and lengths of the resulting alignments using the distantly related plastome of *R. repens*, the *R. cassubicifolius* plastome assembled here as a closely related reference, and 87 di- and polyploid, sexual and apomictic *R. auricomus* TEG samples (SRA BioProject PRJNA628081; Tomasello et al., 2020; Karbstein et al., 2022).

The mitochondrial genome. To assemble the mitogenome, we used GETORGANELLE with the same settings as described above, except that we specified "embplant_mt" and as seed the closest available mitochondriome (*Hepatica maxima*,

GenBank accession number NC_053368.1; Park and Park, 2020). The estimated average embplant_mt base coverage was 92. To improve mitogenome assembly, we mapped ONT reads against the contigs derived from GETORGANELLE and extracted mapped reads (442,704 reads, 953 Mbp, 10% of all ONT reads) with an estimated mitogenome coverage of 756 (see Results section) using SAMTOOLS2 vers. 1.9 ('sort' and 'index'; Danecek et al., 2021) and BEDTOOLS vers. 2.29.1 ('bamtofastq'; Quinlan and Hall, 2010). We applied MASURCA vers. 4.0.9 (Zimin and Salzberg, 2022) to perform hybrid assembly based on extracted ONT reads and "embplant_mt" hitting Illumina reads derived from the GETORGANELLE assembly process. Then, we retained a mitochondrial genome in 18 scaffolds and a total length of 351,681 bp. We aligned contigs in Geneious to reduce overlapping regions and thus contig number, and mapped Illumina and ONT reads against the retained three contigs using MINIMAP2 ('-ax sr' and '-ax map-ont' options, respectively), converted *.sam files to *.bam files using SAMTOOLS2 ('sort' and 'index'), and inspected alignments by IGV Viewer vers. 2.16.1 (Thorvaldsdottir et al., 2013) (reports on FigShare). Using SAMTOOLS2 option 'depth', we inferred a mean Illumina coverage of 1404, and ONT coverage of 618. We completed the mitochondrial genome by inspecting the mapping results and editing the gap regions usually spanned by Illumina- and/or ONT-reads.

The assembly was annotated for protein-coding genes, tRNA, rRNA, and repeat content, and visualized using the webservers of organellar genome annotation tools AGORA (Jung et al., 2018) with *H. maxima* as reference and a maximum number of matched sub gene's count per each contig as 1, CPGAVAS2 (Shi et al., 2019) with the same settings as described above, and MFannot (Lang et al., 2023) and CPGView (Liu et al., 2023) with default settings. For visual identification of repetitive regions, the BLASTN webserver tool (Zhang et al., 2000; Camacho et al., 2009) was used to compare the *R. cassubicifolius* mitogenome against itself (e-value cutoff of 1×10^{-6}, word size of 16; following Park and Park (2020). Geneious was used to inspect and edit the annotation results. The resulting *gb file was converted into a feature table file for GenBank submission using the GB2SEQUIN tool (Lehwark and Greiner, 2019) within the CHLOROBOX webserver. The final annotated mitochondrial genome is illustrated in Figure 2. The mitochondrial raw data and the annotated mitochondrial genome will be made available upon publication (SRA: PRJNA831351, and BankIt: 2722616, respectively). To investigate sequence differences between *R. cassubicifolius* and *H. maxima* as the closest reference mitogenome, we again performed progressiveMAUVE in Geneious.

The nuclear genome. We applied different genome assembly approaches (default settings with haploid genome size of 3.3 Gbp and RAM of at least 1.5 TB, unless otherwise specified): (i) using moderate-coverage Illumina short-reads, we performed analysis with SUPERNOVA (specifying the incorporation of all reads, acceptance of extreme coverages) and SUPERNOVA with subsequent gap filling by filtered ONT reads, and SPADES vers. 3.13.2 (Prjibelski et al., 2020); (ii) using moderate-coverage Illumina short-reads and low-coverage ONT long-reads, we performed analysis with SPADES (incl. the option of untrusted SUPERNOVA contigs for gap filling), MASURCA vers. 4.1.0 (Zimin et al., 2017), MUCHSALSA/LazyB vers. 0.0.2 (Gatter et al., 2021), and WENGAN (M) vers. 0.2 (Di Genova et al., 2021); (iii) using low-coverage ONT long-reads, we performed analyses with CANU vers. 2.1.1 (Koren et al., 2017) and FLYE vers. 2.9.1 (Kolmogorov et al., 2019) (first: CANU corrected reads with the option '--nano-hq' as recommended, and second ONT raw reads with the option '--nano-raw' without polishing) using ONT long-reads followed by the optimized polishing strategy of Dmitriev et al. (2021), i.e., two-times via filtered ONT reads using RACON vers. 1.4.3 (Vaser et al., 2017) and MEDAKA vers. 0.7.1 (https://github.com/nanoporetech/medaka), and once using Illumina reads via the POLCA function (Zimin and Salzberg, 2020) within the MASURCA pipeline.

Statistics for genome assemblies were calculated using QUAST vers. 5.0.1 (Gurevich et al., 2013), and summarized in Table 1 (reports on FigShare). Genomic completeness in terms of expected single-copy genes inherent to land plants was estimated by 'Benchmarking Universal Single-copy Orthologs' (BUSCO vers. 5.0.0; Waterhouse et al., 2018; Manni et al., 2021) with the 'embryophyta_odb10' vers. 2020-09-10 dataset containing 60 species and 1375 core genes, and shown in Figure 3. The assembly with the highest completeness, largest contig size, and best BUSCO scores (Table 1, Fig. 3, reports on FigShare), was selected for genome annotation. We thus annotated the assembled genome from FLYE+Polishing (3.45 Gbp) with FUNANNOTATE pipeline vers. 1.8.15 (https://github.com/nextgenusfs/funannotate; Palmer and Stajich, 2022). First, the assembly was cleaned (only contigs >1 kbp, and duplicated contigs; -14 bp removed) and sorted by size using MINIMAP2 (Li, 2018), and simple repeats were masked using TANTAN ('-s arabidopsis'; Frith, 2011) as they easily align to each other, which causes false homology predictions and which in turn negatively impact gene annotation. Approximately 6.33% (218 Mbp) of the genome was masked. Then, five recently published diploid *R. auricomus* transcriptomes from flower buds (*R. cassubicifolius*, *R. notabilis*, and their F_2 hybrids;

Paetzold et al., 2022) were downloaded from NCBI BioProject PRJNA870679 and quality checked with FASTQC vers. 0.11.4.

To ensure successful genome annotation based on transcriptomic data, RNA reads from different *R. auricomus* species were mapped to the filtered genome, resulting in 81% mapped reads. This represents a good result in the context of multispecies mapping (target: >90% of mapped reads from the same individual as BioGenome Standard; report on FigShare). Then, the filtered genome and the transcriptomes were used to train the annotation process by a genome-guided transcriptome assembly using TRIMMOMATIC (Bolger et al., 2014), TRINITY (Grabherr et al., 2011), HISAT2 (Kim et al., 2015), KALLISTO (Bray et al., 2016), and PASA (Haas et al., 2008) to filter, normalize, and genome-guided cluster transcriptomic data. Sorted alignments, transcripts, and transcriptome annotations are the input for the genome gene prediction process performed by AUGUSTUS (Stanke et al., 2008), GLIMMERHMM (Majoros et al., 2004), and SNAP (Korf, 2004), and results were summarized with EVIDENCEMODELER (Haas et al., 2008). Using the update function, UTRs were added and gene models in disagreement with RNA data were fixed (partly manually, as recommended). Functional genome annotation was then done with INTERPROSCAN (Jones et al., 2014), EGGNOG-MAPPER (Cantalapiedra et al., 2021), and ANTISMASH (Blin et al., 2021). The article will be updated soon with annotation results, and the annotated genome will be deposited in NCBI under Bioproject PRJNA831351 upon publication.

Results and Discussion

More than 610 million, 150 bp raw Illumina reads were generated leading to 91 Gbp and genome coverage of ca. 28x (mean base accuracy > 99.9%) assuming a haploid genome size of 3.3 Gbp (see Introduction). In addition, the final ONT long-read dataset contains 50.9 Gbp and 4.26 million reads, yielding a coverage of 16x (mean base accuracy of 97%) on an estimated haploid genome size of 3.3 Gbp.

The plastid genome. Based on short Illumina reads, we retained a circular plastid genome with 156,233 bp in two different states caused by the differing orientations of the inverted repeat (IR) regions ("flip-flop"; Stein et al., 1986; Walker et al., 2015). Results were visually inspected using Bandage vers. 0.9.0 (Wick et al., 2015). The plastome shows a typical tetrapartide structure of a small single-copy region (SSC, length: 19,826 bp) and a large single-copy region (LSC, length: 85,629 bp) separated by two IR regions (length: 2 x

25,389 bp). The plastome contains 85 (78 unique) protein-coding genes, 37 (30 unique) tRNAs, and 8 (4 unique) rRNAs (duplicated ones in IRs). We also observed 35 SSR, 32 long tandem repeat, and 34 dispersed repeat motifs. In comparison with the closely related, recently published plastome of *R. sceleratus* (sister section to sect. *Auricomus*; Genbank accession number ON755204, Kim et al., 2023) and the more distantly related plastome of *R. repens* (different subgenus within Ranunculaceae), the plastome of *R. cassubicifolius* shows a maximum number of unique protein-coding genes (78/77/78), tRNAs (30/28/29), and rRNAs (4/3/4). The plastome size of *R. cassubicifolius* was slightly smaller compared to *R. sceleratus* (+96 bp, 156,329 bp), but smaller compared to *R. repens* (-1,986 bp; 154,247 bp), mainly due to differences in repeat motives in the SSC (19,826/18,434 bp) and LSC (85,629/84,224 bp) regions (e.g., the *R. cassubicifolius* plastome has additional, long repeats between various genes), but not in the IR regions (25,389/25,794 bp).

Interestingly, the plastome of *R. cassubicifolius* additionally contains the gene *ycf15* (table for plastome comparisons on FigShare), which is controversial in terms of origin and function. The gen *ycf15* probably originated at the roots of the angiosperms and not in eubacteria, and has been disabled or lost in many ancestral angiosperm lineages, and compared to other pseudogenes like *infA* or *rpl22*, no intracellular transfer to the nucleus has been observed across many angiosperm lineages (Shi et al., 2013). For the genus *Ranunculus*, the complete loss of this gene was assumed (Shi et al., 2013) but the recent study of Kim et al. (2023) for *R. austro-oreganus* and *R. occidentalis*, and the results presented here for *R. cassubicifolius* contradict this assumption. Despite the gene is only 54 bp long and apparently degraded (e.g., Shi et al., 2013; Li et al., 2021), we found two copies of *ycf15* in the IRA and IRB regions with regular ATG start and TGA stop codon, suggesting that it is functional in *R. cassubicifolius*. In addition, in contrast to *R. repens*, the plastome of *R. cassubicifolius* (and of *R. sceleratus*) has lost the *rpl32* ribosomal gene (involved in protein biosynthesis and the function of the plastid, but often non-functional in other Ranunculaceae), potentially as the result of complete loss or gene transfer into the nucleus, as also observed in other clades within Ranunculaceae (Park et al., 2015; Zhai et al., 2019; Park and Park, 2020). However, the function of *ycf15* and the co-transcribed, large plastid gene *ycf2* is still unclear, although *ycf2* together with *accD* influence plastid structure and competitiveness, and thus may play a role in uniparental/biparental inheritance of plants that is known to shape the ability of hybridization (Shi et al., 2013; Liu et al., 2017; Sobanski et al., 2019; Postel and Touzet, 2020). In general, we detected no substantial plastome rearrangement and gene transfer in

mapping results (see reports on FigShare), additionally supporting phylogenetically close relationships among these plastome lineages.

Interestingly, mapping off-target plastid reads from a previous target enrichment project (Karbstein et al., 2022) against the coding regions of reference plastomes from *R. repens* and *R. cassubicifolius* yielded slightly better assemblies for the closely-related *R. cassubicifolius* with 24.30% compared to 24.07% plastome completeness i.e., assembly lengths) resulting in 827 bp longer alignments on average (37,968 vs. 37,141 bp; see plastome mapping report on FigShare). The estimated sequence difference in plastomes among these species is little (<2% in coding regions), and thus this approach did not improve mapping substantially. However, read mapping to the closest possible reference ensures reliable locus assembly, particularly in non-coding regions often characterized by insertions and deletions even among individuals of a species (Freeland et al., 2011; Folk et al., 2015; Scheunert et al., 2020). Although mapping to a closer reference was only marginally more efficient (0.23%), off-target read mapping resulted in a significantly larger alignment of 2.2% (+ 827 bp). This might be even more effective when no close reference plastome is available within the genus, with the potential to improve subsequent phylogenomic analysis due to increased sequence length and information content.

The mitochondrial genome. Based on short Illumina read assembly bridged with ONT reads, we retained a mitochondrial genome with 1,260,409 bp, which is comparable in size to *H. maxima* with 1,122,546 bp, corresponding to the phylogenetically relatively close relationship between *Hepatica* and *Ranunculus* within the Ranunculaceae family (Zhai et al., 2019). The annotated mitochondriome of *R. cassubicifolius* contains 78 (39 unique) protein-coding genes, 24 (unique 13) tRNAs, and 6 (unique 3) rRNAs. In comparison with *H. maxima*, the mitochondriome of *R. cassubicifolius* shows an equal number of unique protein-coding genes (39/39), tRNAs (13/13), and rRNAs (3/3), but is larger (+137,863 bp; 1,122,546 bp) mainly due to sequence recombination and rearrangements indicated by the presence of additional protein-coding genes (e.g., many genes, tRNAs, and rRNAs could be annotated twice; 78/40, 24/18, 6/3) under similar repeat content (ca. 7%). This is supported by a substantial number of observed mitogenome rearrangements between both species despite their relatively close phylogenetic relationship (mapping results on FigShare). In contrast to plastomes, plant mitochondrial genomes are recombinationally active, can exist as populations of alternative structures even within individuals, and are considered to be organized in noncircular structures (Sloan, 2013; Møller et al., 2021); algae and plant mitochondrial genomes can range among the largest chondromes known from 66 kbp up to 11

Mbp in flowering plants (Smith and Keeling, 2015; Dong et al., 2018). The annotation results presented here are consistent with the general observation that angiosperm mitochondriomes contain 19 to 64 known genes (excluding duplicates and ORFs) and highly variable intergenic regions (Dong et al., 2018).

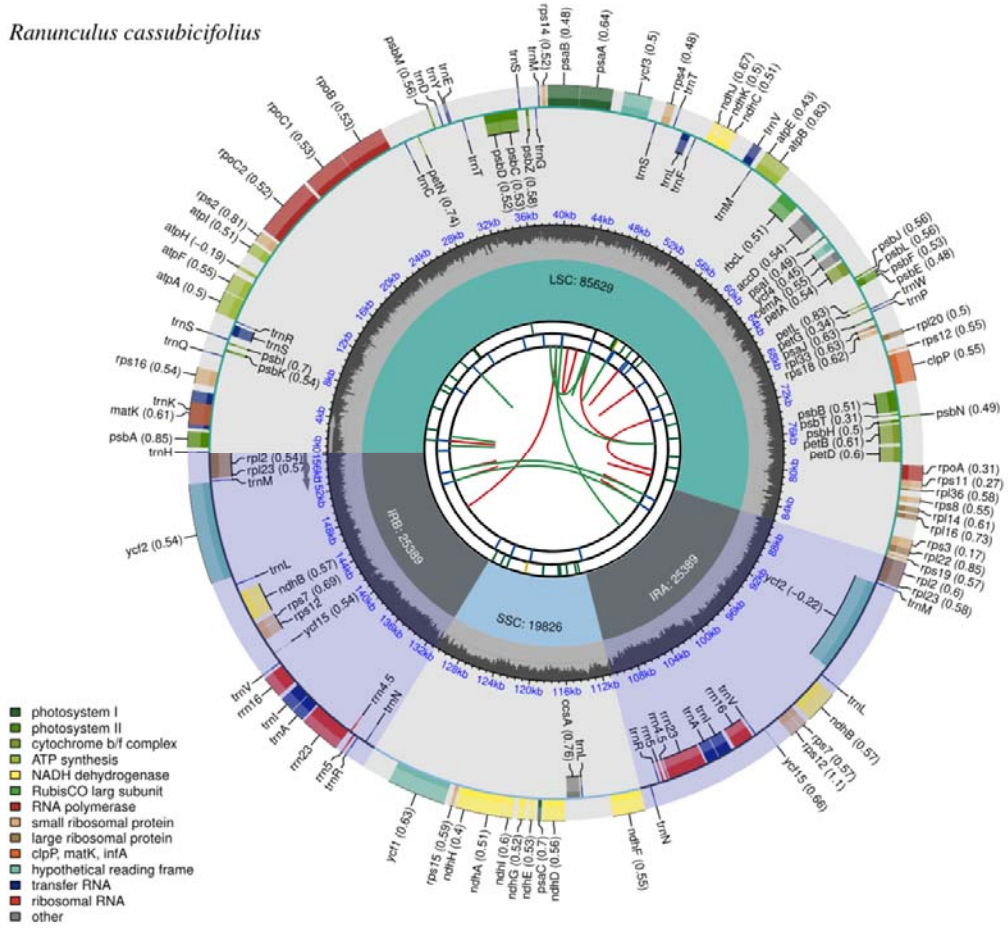


Fig. 1. (a) Annotated plastome of *Ranunculus cassubicifolius* (LH040). Results of plastome annotation using the software CPGAVAS2, talking the default settings and the 2,544 reference plastomes (Shi et al., 2019). The plastome assembly is 156,233 bp large, and represented in a single circular contig). It contains 85 (78 unique) protein-coding genes, 37 (30 unique) tRNAs, and 8 (4 unique) rRNAs, as shown in the outer circle. Genes and t/rRNAs on the inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. The first inner circle represents the GC content (%), the second inner circle the plastome structure, i.e., a small single-copy region (SSC) and a large single-copy region (LSC) separated by two inverted repeat (IRA/B) regions. The third inner circle reflects the repeat content, i.e., microsatellites (1-6 bp per repeat), tandem repeats (>6 bp per repeat), and dispersed repeats (transposable elements, or satellite DNA; reports on FigShare). See Shi et al. (2019) and http://www.1kmpg.cn/cpgview/ for more information.

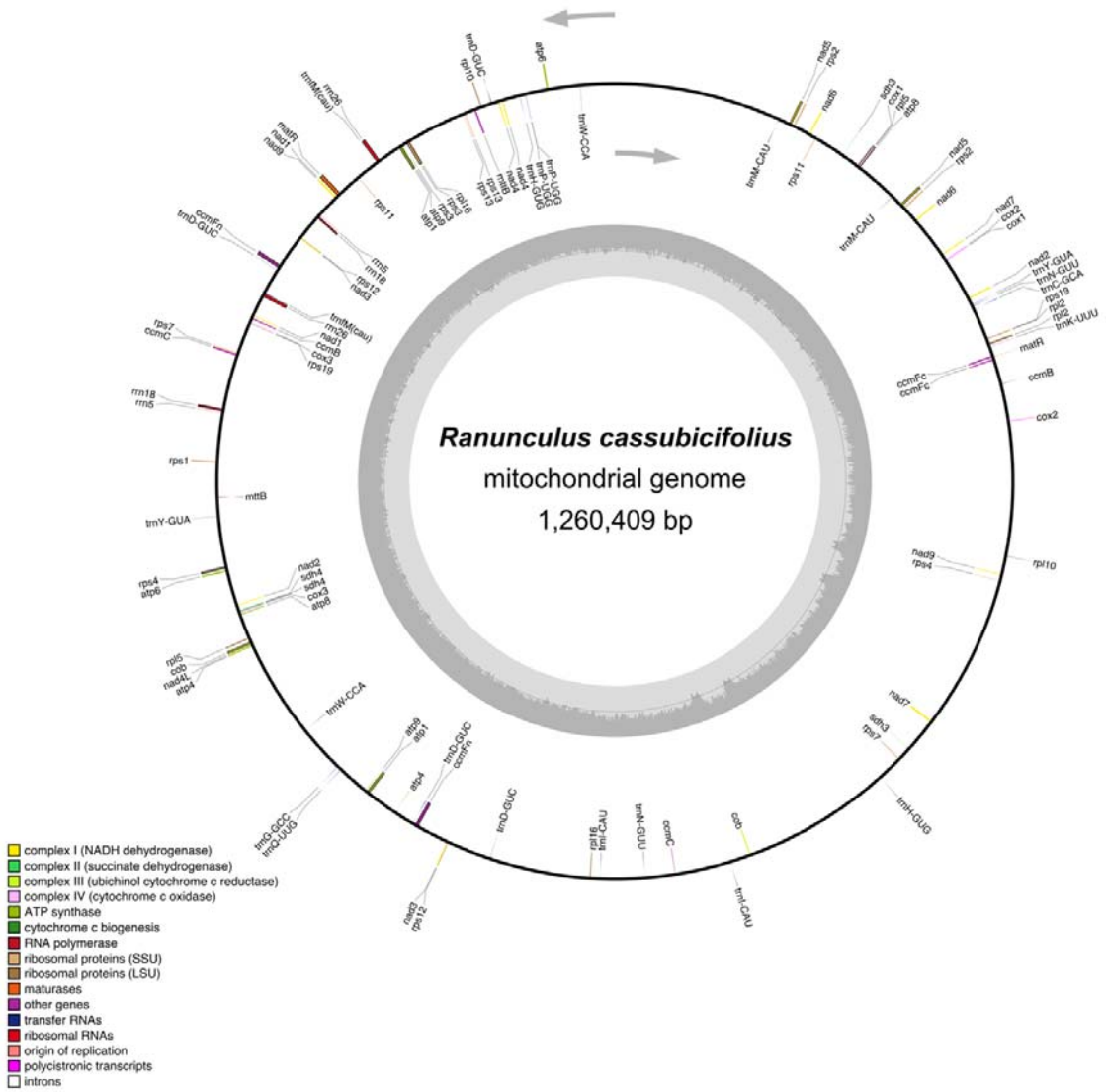


Fig. 2. (b) The assembled mitochondrial genome of *Ranunculus cassubicifolius* (LH040). Results of mitochondriome annotation were generated using AGORA Webserver (Jung et al., 2018). The mitochondriome is 1,260,409 bp in size represented in a single, circular contig, and contains 78 (39 unique) protein-coding genes, 24 (unique 13) tRNAs, and 6 (unique 3) rRNAs, as shown in the outer circle. Genes and t/rRNAs on the inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. The first inner circle represents the GC content (%). The second inner circle reflects the repeat content, i.e., microsatellites (1-6 bp per repeat), tandem repeats (>6 bp per repeat), and dispersed repeats (transposable elements, or satellite DNA; reports on FigShare). See Shi et al. (2019) and http://www.1kmpg.cn/cpgview/ for more information.

The nuclear genome. Genome assembly strategies yielded different results concerning genome completeness and continuity (% from expected genome size, number of contigs, mean contig lengths, N_{50}), and quality (BUSCO scores; Table 1, Fig. 3). Although some genome assemblies based on short Illumina reads performed well in terms of completeness (SUPERNOVA 86-87%, except SAMBA 17%), they produced the largest number of contigs (1.23-1.33 million contigs) with the shortest N_{50} length (1.75-3.8 Mbp). Assemblies were highly fragmented (F; 28-36%) with low to moderate numbers of complete single-copy (S, 20-60%) and duplicated (D, 2-11%), and missing (M, 10-34%) BUSCO genes. This is in line with literature showing highly fragmented, gappy, low-quality assemblies of large and complex genomes when short-read data are used (Gatter et al., 2021; Rhie et al., 2021; Zimin and Salzberg, 2022; Hotaling et al., 2023).

Hybrid assemblies of SPADES, MASURCA, MUCHSALSA, and WENGAN based on short Illumina and long ONT reads performed better in terms of continuity due to fewer but larger contigs (368 up to 1.18 million contigs with 4.7-55 N_{50} length), but not in terms of completeness (0.15-13%) as large portions of BUSCO genes were often missing in assemblies (M: 8-99.6%, F: 0.2-24%, D: 0-2%, S: 0.2-66%). This is interesting because many hybrid assemblers are considered by developers to be particularly useful under low-coverage and/or large-and-complex-genome scenarios (e.g, Zimin et al., 2017; Di Genova et al., 2021; Gatter et al., 2021). However, hybrid assemblers (but also assemblers in general) have rarely been implemented for large and complex plant genomes under both Illumina and ONT low-coverage scenarios, which may explain their lower performance here. Compared to already discussed strategies, the ONT long-read based assemblies conducted with CANU and FLYE were generally superior in terms of completeness (25-99%), continuity (30,612 to 65,232 contigs with 37-104 N_{50} length), and quality as the majority of BUSCO genes were found (M: 1-29%, F: 2-7%, D: 8-29%, S: 56-68%). This is also consistent with results of the recently published review by Hotaling et al. (2023), which indicates that assemblies based on long reads, compared to short reads, provide more complete and continuous genomes of higher quality.

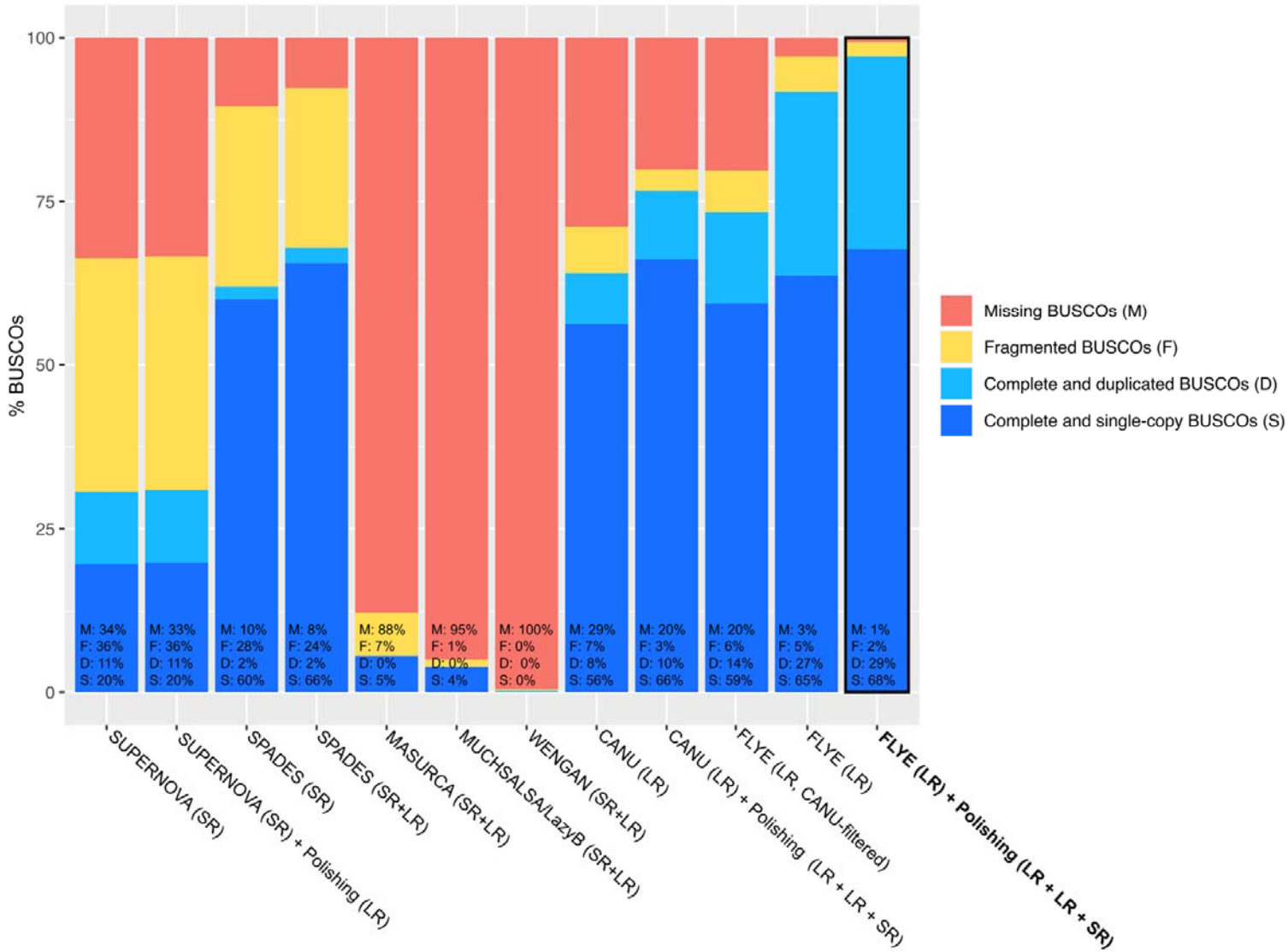
In detail, FLYE compared to CANU assemblies were slightly better in quality (M: 20 vs. 29%, F: 6 vs. 7%, D: 14 vs. 8%, S: 59 vs. 56%), and much better in completeness (49 vs. 25%) and continuity (80 vs. 37 Mbp N_{50} length). Interestingly, concerning CANU and FLYE assemblies, polishing improved quality (M: 29/20% and 3/1%, F: 7/3% and 5/2%, D: 8/10% and 27/29%, S: 66/56% and 65/68%) but not completeness (25/25% and 100/100%) and

continuity (37/38 Mbp and 103/104 Mbp). Consequently, long-read FLYE assembly followed by various polishing strategies with both filtered short- and long-reads represents the best assembly strategy for the large and complex *R. cassubicifolius* plant nuclear genome under both Illumina and ONT low- to moderate-coverage scenarios, and is supported by similar findings for smaller plant genome assemblies (e.g., Dmitriev et al., 2021).

The genome fulfills some standards of the high-quality genome requirements proposed by the Earth Biogenome Project (https://www.earthbiogenome.org/assembly-standards, Lawniczak et al., 2022), and by other authors (Hotaling et al., 2023; Mokhtar et al., 2023) (discussion part will be updated soon). The assembly was specifically successful regarding gene completeness (BUSCO 97%), which will be important for i.a. understanding the genetic background of shifts in mode of reproduction from sex to apomixis (Paetzold et al., 2023). The quite high-quality assembly of BUSCO genes will be further an important resource for high-quality multi-assemblies in the genus for a comprehensive analysis of loci under selection (e.g., via dN/dS ratios; Pellino et al., 2013), which gives insights into general genome evolution, adaptation, and speciation. We assembled the complete organelles but did not yet aim at a complete chromosome-level assembly of the nuclear genome as set as standard by Lawniczak et al. (2022). This requirement, however, is most important for organisms with sex chromosomes. In *Ranunculus* all species are hermaphroditic and no sex chromosomes are known (these are in general very rare in plants, e.g., Charlesworth, 2015, 2016). A chromosome-level assembly using the Hi-C technology is under work.

Table 1. Genome assembly results of different strategies for the diploid sexual species *R. cassubicifolius* (LH040). Assembly statistics are derived from QUAST (see also reports on Figshare). Using Illumina short- or ONT long-reads (SR, LR), we carried out different SR, hybrid (SR+LR), and LR assembly strategies (see Methods section for details). The polishing of the CANU and FLYE assemblies includes the following steps: polishing via filtered ONT reads using RACON and MEDAKA, respectively, and via Illumina reads using POLCA (MASURCA pipeline). Key parameters, and the best assembly in terms of completeness and contig length are highlighted in bold. N_{50}/N_{75} describes the shortest contig length required to summarize at least half (50%) or three quarters (75%) of the bases of the entire assembly; L_{50}/L_{75} is defined as the minimum number of contigs that encompass half (50%) or three quarters (75%) of the bases of the entire assembly.

Metric	**SUPERNOVA** (SR)	**SUPERNOVA** (SR) + **Polishing** (LR)	**SPADES** (SR)	**SPADES** (SR+LR)	**MASURCA** (SR+LR)	**MUCHSALSA** (SR+LR)	**WENGAN** (SR+LR)	**CANU** (LR)	**CANU** (LR) +**Polishing** (SR+LR)	**FLYE** (LR, CANU-filtered)	**FLYE** (LR)	**FLYE** (LR) +**Polishing** (LR+LR+SR)
Total assembly length, Mbp	2,761 (86%)	2,788 (87%)	0,557 (17%)	0,408 (13%)	0,076 (2.4%)	0,043 (1.3%)	0,005 (0.15%)	0,80 (25%)	0,81 (25%)	1,56 (49%)	3,58 (99%)	**3,45 (99%)**
Number of contigs	6,920,808	6,917,617	1,231,317	1,184,036	38,650	915	368	30,612	32,959	42,987	65,232	**54,310**
Number of contigs > 1 Mbp	562,716	560,357	122,407	110,594	12,799	915	368	30,612	31,072	41,985	64,618	**53,572**
Largest contig, Mbp	39	183	69	96	145	131	95	540	542	735	976	**974**
GC, %	40	40	39	39	40	42	40	40	40	42	43	**43**
N_{50}, Mbp	1.75	1.80	3.8	4.7	9.7	52.9	20.6	37.2	37.7	80.5	103	**104**
N_{75}, Mbp	0.91	0.92	1.91	2.31	1.11	40.3	10.4	21.9	22.2	35.9	52.9	**56.2**
L_{50}	259,069	248,906	29,900	23,910	1,507	307	76	6,935	6,937	5,520	10,305	**9,889**
L_{75}	639,776	628,182	67,267	54,888	10,467	537	170	13,891	13,885	12,873	22,076	**20,928**


% BUSCOs
100
75
50
25
0
M: 34%
F: 36%
D: 11%
S: 20%
M: 33%
F: 36%
D: 11%
S: 20%
M: 10%
F: 28%
D: 2%
S: 60%
M: 8%
F: 24%
D: 2%
S: 66%
M: 88%
F: 7%
D: 0%
S: 5%
M: 95%
F: 1%
D: 0%
S: 4%
M: 100%
F: 0%
D: 0%
S: 0%
M: 29%
F: 7%
D: 8%
S: 56%
M: 20%
F: 3%
D: 10%
S: 66%
M: 20%
F: 6%
D: 14%
S: 59%
M: 3%
F: 5%
D: 27%
S: 65%
M: 1%
F: 2%
D: 29%
S: 68%
SUPERNOVA (SR)
SUPERNOVA (SR) + Polishing (LR)
SPADES (SR)
SPADES (SR+LR)
MASURCA (SR+LR)
MUCHSALSA/LazyB (SR+LR)
WENGAN (SR+LR)
CANU (LR)
CANU (LR) + Polishing (LR + LR + SR)
FLYE (LR, CANU-filtered)
FLYE (LR)
FLYE (LR) + Polishing (LR + LR + SR)
Missing BUSCOs (M)
Fragmented BUSCOs (F)
Complete and duplicated BUSCOs (D)
Complete and single-copy BUSCOs (S)

575

Fig. 3. BUSCO assessment results for different draft genome assemblies of the diploid sexual species *Ranunculus cassubicifolius*. Using Illumina short- or ONT long-reads (SR, LR), we carried out different SR, hybrid (SR+LR), and LR assembly strategies (see Methods section for details). The polishing of the CANU and FLYE assemblies includes the following steps: polishing via filtered ONT reads using RACON and MEDAKA, respectively, and via Illumina reads using POLCA (MASURCA pipeline). The best assembly in terms of complete BUSCO genes (S+D) is highlighted with a black square (i.e., "FLYE (LR) + Polishing (LR+ LR + SR)"). Analyses based on 1,614 BUSCO groups searched. See table on FigShare for more details.

Conclusions

In this study, we assembled the complete plastome, mitochondriome, and large nuclear genome of a diploid sexual progenitor species within the *Ranunculus auricomus* complex, using different combinations of highly accurate Illumina short-reads and less accurate ONT long-reads under low to medium-coverage scenarios. To enable successful ONT sequencing, we optimized extraction and library preparation protocols for pure high-molecular-weight DNA (gDNA) from leaves of two single, genetically highly similar *R. cassubicifolius* individuals of the same geographic region. Using Illumina NextSeq sequencing, we obtained filtered reads with 28x genome coverage (610 million of 150 + 150 paired-end reads yielding 91 Gbp). DNA sequencing on ONT MinION 101B devices and subsequent quality filtering resulted in 16x genome coverage (51 Gbp, N_{50}=21 kbp).

We generated, for the first, the complete plastid and intricate mitochondrial genome for the *R. auricomus* complex based on Illumina and Illumina+ONT read assemblies, respectively. We then performed various assembly short-, long-, and hybrid-read strategies for large, complex plant genomes followed by gap filling and/or polishing with various state-of-the-art bioinformatic tools. The best result in terms of contiguity and completeness was achieved by ONT read FLYE-assembly polished several times with filtered ONT (RACON, MEDAKA) and Illumina reads (POLCA), representing the first genome within the genus *Ranunculus* and the third (2 *Aquilegia* genomes, Filiault et al., 2018; Xie et al., 2020) high-quality genome within the Ranunculaceae (compare with Xie et al., 2020).

The obtained results and in particular genomes are useful for (i) improving read mapping, locus assembly, and allele phasing of subgenomic reduced-representation sequencing (RRS) (e.g., RAD-Seq, target enrichment) as shown here (Fig. X), transcriptome (RNA-Seq), and whole genome (re)sequencing data (e.g., WGS, WGR), (ii) as a basis for understanding cyto-nuclear compatibility for both plastid-nuclear and mitochondria-nuclear interactions; (iii) as diploid reference for the analysis of the more complex allo- and autopolyploid nuclear genomes within the genus *Ranunculus* (iii) as reference for further studies of genome evolution, loci under selection, adaptation and speciation (iv) For improving functional analyses (e.g., gene expression studies) related to apomixis, adaptation, and many other traits.

Data Availability. The authors declare that basic data supporting the findings are available within the manuscript and Supporting Information. Raw sequencing data and the assembled genomes are deposited on the National Center for Biotechnology Information (NCBI) database under the BioProject accession numbers: PRJNA826743 for Illumina raw reads (https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample_bioproject&from_uid=27582367), PRJNA831351 for filtered Nanopore reads, filtered mitochondrial Illumina and Nanopore reads, and the assembled and annotated genome (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA831351/), OP250948 for the assembled and annotated plastome (https://www.ncbi.nlm.nih.gov/search/all/?term=OP250948), OP2722616 for the assembled and annotated mitochondrial genome. We deposited software reports on FigShare (https://figshare.com/s/b7fba31a6f2444123906, upon publication: doi: 10.6084/m9.figshare.20488908).

Author Contributions

KK, II, JdV, and EH designed research; EH collected plant materials; KK, ST, JPB, and BBH performed lab work; KK, II, JdV, ST, and NW analyzed data; KK and EH wrote the paper with contributions of all authors.

Funding

This work (gDNA extraction, genome sequencing, assembly, and annotation) was financially supported by German Research Foundation (DFG) grant Ho4395/10-2 to E.H. within the priority program SPP 1991 “Taxon-Omics: New Approaches for Discovering and Naming Biodiversity”. Work in the lab of JdV is supported by the European Research Council through funding under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 852725; ERC-StG “TerreStriAL”).

Acknowledgments

We acknowledge Jennifer Krüger for technical help, Silvia Friedrichs and Gabriele Ließmann for garden work, Ladislav Hodač for material sampling, and Magnus Nordborg and Robin Burns for bioinformatic advice.

Supplementary Material

The Supplementary Material for this article can be found attached to this article.

References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barke, B. H., Daubert, M., and Hörandl, E. (2018). Establishment of apomixis in diploid F2 hybrids and inheritance of apospory from F1 to F2 hybrids of the *Ranunculus auricomus* complex. *Front. Plant Sci.* 9, 1–12. doi: 10.3389/fpls.2018.01111.

Barke, B. H., Karbstein, K., Daubert, M., and Hörandl, E. (2020). The relation of meiotic behaviour to hybridity, polyploidy and apomixis in the *Ranunculus auricomus* complex (Ranunculaceae). *BMC Plant Biol.* 20, 523. doi: 10.1186/s12870-020-02654-3.

Bennetzen, J. L. (2005). Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 95, 127–132. doi: 10.1093/aob/mci008.

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi: 10.1093/nar/gkab335.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519.

Brukhin, V., Osadtchiy, J. V., Florez-Rueda, A. M., Smetanin, D., Bakin, E., Nobre, M. S., et al. (2019). The *Boechera* genus as a resource for apomixis research. *Front. Plant Sci.* 10, 1–19. doi: 10.3389/fpls.2019.00392.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi: 10.1186/1471-2105-10-421.

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293.

Charlesworth, D. (2015). Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* 208, 52–65. doi: 10.1111/nph.13497.

Charlesworth, D. (2016). Plant sex chromosomes. *Annu. Rev. Plant Biol.* 67, 397–420. doi: 10.1146/annurev-arplant-043015-111911.

Chuan, J., Zhou, A., Hale, L. R., He, M., and Li, X. (2021). Atria: an ultra-fast and accurate trimmer for adapter and quality trimming. *Gigabyte* 2021, 1–18. doi: 10.46471/gigabyte.31.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi: 10.1093/gigascience/giab008.

Dann, M., Bellot, S., Schepella, S., Schaefer, H., and Tellier, A. (2017). Mutation rates in seeds and seed-banking influence substitution rates across the angiosperm phylogeny. *bioRxiv*, 156398. doi: 10.1101/156398.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704.

Dauphin, B., Grant, J. R., Farrar, D. R., and Rothfels, C. J. (2018). Rapid allopolyploid radiation of moonwort ferns (*Botrychium*; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions. *Mol. Phylogenet. Evol.* 120, 342–353. doi: 10.1016/j.ympev.2017.11.025.

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149.

Del Amparo, R., Branco, C., Arenas, J., Vicens, A., and Arenas, M. (2021). Analysis of selection in protein-coding sequences accounting for common biases. *Brief. Bioinform.* 22, 1–10. doi: 10.1093/bib/bbaa431.

Di Genova, A., Buena-Atienza, E., Ossowski, S., and Sagot, M.-F. (2021). Efficient hybrid de novo assembly of human genomes with WENGAN. *Nat. Biotechnol.* 39, 422–430. doi: 10.1038/s41587-020-00747-w.

Dmitriev, A. A., Pushkova, E. N., Novakovskiy, R. O., Beniaminov, A. D., Rozhmina, T. A., Zhuchenko, A. A., et al. (2021). Genome sequencing of fiber flax cultivar Atlant using Oxford Nanopore and Illumina platforms. *Front. Genet.* 11, 1–7. doi: 10.3389/fgene.2020.590282.

Dong, S., Zhao, C., Chen, F., Liu, Y., Zhang, S., Wu, H., et al. (2018). The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics* 19, 1–12. doi: 10.1186/s12864-018-4991-4.

Eaton, D. A. R., and Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq

datasets. *Bioinformatics* 36, 2592–2594. doi: 10.1093/bioinformatics/btz966.

Ennos, R., French, G., and Hollingsworth, P. (2005). Conserving taxonomic complexity. *Trends Ecol. Evol.* 20, 164–168. doi: 10.1016/j.tree.2005.01.012.

Filiault, D. L., Ballerini, E. S., Mandáková, T., Aköz, G., Derieg, N. J., Schmutz, J., et al. (2018). The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *Elife* 7, 1–31. doi: 10.7554/eLife.36426.

Folk, R. A., Mandel, J. R., and Freudenstein, J. V. (2015). A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: A phylogenomic example from *Heuchera* (Saxifragaceae). *Appl. Plant Sci.* 3, 1500039. doi: 10.3732/apps.1500039.

Freeland, J. R., Kirk, H., and Petersen, S. D. (2011). *Molecular Ecology*. sec. ed. Chichester, England: Wiley doi: 10.1002/9780470979365.

Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., and Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biol.* 21, 254. doi: 10.1186/s13059-020-02153-6.

Frith, M. C. (2011). A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res.* 39, e23–e23. doi: 10.1093/nar/gkq1212.

Gatter, T., von Löhneysen, S., Fallmann, J., Drozdova, P., Hartmann, T., and Stadler, P. F. (2021). LazyB: fast and cheap genome assembly. *Algorithms Mol. Biol.* 16, 8. doi: 10.1186/s13015-021-00186-5.

GBIF Secretariat (2021). *Ranunculus auricomus* L. *GBIF Backbone Taxon.* doi: https://doi.org/10.15468/39omei.

Grabherr, M. G. ., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7.

Havird, J. C., Hall, M. D., and Dowling, D. K. (2015). The evolution of sex: A new

hypothesis based on mitochondrial mutational erosion. *BioEssays* 37, 951–958. doi: 10.1002/bies.201500057.

Hodač, L., Barke, B. H., and Hörandl, E. (2018). Mendelian segregation of leaf phenotypes in experimental F2 hybrids elucidates origin of morphological diversity of the apomictic *Ranunculus auricomus* complex. *Taxon* 67, 1082–1092. doi: 10.12705/676.6.

Hodač, L., Klatt, S., Hojsgaard, D., Sharbel, T. F., and Hörandl, E. (2019). A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. *BMC Evol. Biol.* 19, 170. doi: 10.1186/s12862-019-1495-z.

Hojsgaard, D., Greilhuber, J., Pellino, M., Paun, O., Sharbel, T. F., and Hörandl, E. (2014). Emergence of apospory and bypass of meiosis via apomixis after sexual hybridisation and polyploidisation. *New Phytol.* 204, 1000–1012. doi: 10.1111/nph.12954.

Hörandl, E. (2022). Novel approaches for species concepts and delimitation in polyploids and hybrids. *Plants* 11, 204. doi: 10.3390/plants11020204.

Hörandl, E., and Greilhuber, J. (2002). Diploid and autotetraploid sexuals and their relationships to apomicts in the *Ranunculus cassubicus* group: insights from DNA content and isozyme variation. *Plant Syst. Evol.* 234, 85–100. doi: 10.1007/s00606-002-0209-x.

Hörandl, E., Greilhuber, J., Klímová, K., Paun, O., Temsch, E., Emadzade, K., et al. (2009). Reticulate evolution and taxonomic concepts in the *Ranunculus auricomus* complex (Ranunculaceae): insights from analysis of morphological, karyological and molecular data. *Taxon* 58, 1194–1215. doi: 10.1002/tax.584012.

Hörandl, E., and Raab-Straube, E. (2015). Ranunculaceae. *Euro+Med Plantbase - Inf. Resour. Euro-Mediterranean plant Divers.* Available at: http://ww2.bgbm.org/EuroPlusMed.

Hotaling, S., Wilcox, E. R., Heckenhauer, J., Stewart, R. J., and Frandsen, P. B. (2023). Highly accurate long reads are crucial for realizing the potential of biodiversity genomics. *BMC Genomics* 24, 117. doi: 10.1186/s12864-023-09193-9.

Jalas, J., and Suominen, J. (1989). *Atlas florae europaeae, vol. 8, Nymphaeaceae to Ranunculaceae*. Helsinki, Finland: The committee for mapping the flora of Europe and Societas Biologica Fennica Vanamo.

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., DePamphilis, C. W., Yi, T.-S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21, 241. doi: 10.1186/s13059-020-02154-5.

Jones, G. (2017). Bayesian phylogenetic analysis for diploid and allotetraploid species

networks. *bioRxiv*, 1–15. doi: 10.1101/129361.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031.

Jung, J., Kim, J. I., Jeong, Y.-S., and Yi, G. (2018). AGORA: organellar genome annotation from the amino acid and nucleotide references. *Bioinformatics* 34, 2661–2663. doi: 10.1093/bioinformatics/bty196.

Karbstein, K. (2021). Untying Gordian knots: The evolution and biogeography of the large European apomictic polyploid *Ranunculus auricomus* plant complex. *eDiss*. doi: 21.11130/00-1735-0000-0008-5946-6.

Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., et al. Species Delimitation 4.0: Integrative Taxon-omics and Artificial Intelligence. *Submitt. to Trends Ecol. Evol.*

Karbstein, K., Rahmsdorf, E., Tomasello, S., Hodač, L., and Hörandl, E. (2020a). Breeding system of diploid sexuals within the *Ranunculus auricomus* complex and its role in a geographical parthenogenesis scenario. *Ecol. Evol.* 10, 14435–14450. doi: 10.1002/ece3.7073.

Karbstein, K., Tomasello, S., Hodač, L., Dunkel, F. G., Daubert, M., and Hörandl, E. (2020b). Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae). *Taxon* 69, 1191–1220. doi: 10.1002/tax.12365.

Karbstein, K., Tomasello, S., Hodač, L., Lorberg, E., Daubert, M., and Hörandl, E. (2021). Moving beyond assumptions: Polyploidy and environmental effects explain a geographical parthenogenesis scenario in European plants. *Mol. Ecol.* 30, 2659–2675. doi: 10.1111/mec.15919.

Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B. H., et al. (2022). Untying Gordian knots: Unraveling reticulate polyploid plant evolution by genomic data using the large *Ranunculus auricomus* species complex. *New Phytol.* 235, 2081–2098. doi: 10.1111/nph.18284.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low

memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317.

Kim, K.-R., Park, S. Y., Kim, H., Hong, J. M., Kim, S.-Y., and Yu, J.-N. (2023). Complete chloroplast genome determination of *Ranunculus sceleratus* from Republic of Korea (Ranunculaceae) and comparative chloroplast genomes of the members of the *Ranunculus* genus. *Genes (Basel).* 14, 1149. doi: 10.3390/genes14061149.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5, 1–9. doi: 10.1186/1471-2105-5-59.

Kress, W. J., Soltis, D. E., Kersey, P. J., Wegrzyn, J. L., Leebens-Mack, J. H., Gostel, M. R., et al. (2022). Green plant genomes: What we know in an era of rapidly expanding opportunities. *Proc. Natl. Acad. Sci.* 119, 1–9. doi: 10.1073/pnas.2115640118.

Lang, B. F., Beck, N., Prince, S., Sarrasin, M., Rioux, P., and Burger, G. (2023). Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. *Front. Plant Sci.* 14, 1–11. doi: 10.3389/fpls.2023.1222186.

Lawniczak, M. K. N., Durbin, R., Flicek, P., Lindblad-Toh, K., Wei, X., Archibald, J. M., et al. (2022). Standards recommendations for the Earth BioGenome Project. *Proc. Natl. Acad. Sci.* 119, 1–8. doi: 10.1073/pnas.2115639118.

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2.

Lehwark, P., and Greiner, S. (2019). GB2sequin - A file converter preparing custom GenBank files for database submission. *Genomics* 111, 759–761. doi: 10.1016/j.ygeno.2018.05.003.

Leitch, I., Johnston, E., Pellicer, J., Hidalgo, O., and Bennett, M. (2019). Angiosperm DNA C-values database (release 9.0, Apr 2019). Available at: https://cvalues.science.kew.org/.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191.

Li, X., Li, Y., Sylvester, S. P., Zang, M., El□Kassaby, Y. A., and Fang, Y. (2021).

Evolutionary patterns of nucleotide substitution rates in plastid genomes of Quercus. *Ecol. Evol.* 11, 13401–13414. doi: 10.1002/ece3.8063.

Li, Z., Parris, S., and Saski, C. A. (2020). A simple plant high-molecular-weight DNA extraction method suitable for single-molecule technologies. *Plant Methods* 16, 38. doi: 10.1186/s13007-020-00579-4.

Liu, L.-X., Li, R., Worth, J. R. P., Li, X., Li, P., Cameron, K. M., et al. (2017). The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): Implications for understanding the evolution of Fagales. *Front. Plant Sci.* 8, 1–15. doi: 10.3389/fpls.2017.00968.

Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H., et al. (2023). CPGView: A package for visualizing detailed chloroplast genome structures. *Mol. Ecol. Resour.* 23, 694–704. doi: 10.1111/1755-0998.13729.

Lu, L., Zhao, J., and Li, C. (2020). High-quality genome assembly and annotation of the big-eye mandarin fish (Siniperca knerii). *G3 Genes, Genomes, Genet.* 10, 877–880. doi: 10.1534/g3.119.400930.

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi: 10.1016/j.tree.2005.02.010.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199.

Marks, P., Garcia, S., Martinez, A., Belhocine, K., Bernate, J., Bharadwaj, R., et al. (2017). Resolving the full spectrum of human genome variation using linked-reads. *bioRxiv*, 1–43. doi: 10.1101/230946.

McKain, M. R., Johnson, M. G., Uribe-Convers, S., Eaton, D., and Yang, Y. (2018). Practical considerations for plant phylogenomics. *Appl. Plant Sci.* 6, e1038. doi: 10.1002/aps3.1038.

Mokhtar, M. M., Abd-Elhalim, H. M., and El Allali, A. (2023). A large-scale assessment of the quality of plant genome assemblies using the LTR assembly index. *AoB Plants* 15, 1–8. doi: 10.1093/aobpla/plad015.

Møller, I. M., Rasmusson, A. G., and Van Aken, O. (2021). Plant mitochondria – past, present

and future. *Plant J.* 108, 912–959. doi: 10.1111/tpj.15495.

Paetzold, C., Barke, B. H., and Hörandl, E. (2022). Evolution of transcriptomes in early-generation hybrids of the apomictic *Ranunculus auricomus complex* (Ranunculaceae). *Int. J. Mol. Sci.* 23, 13881. doi: 10.3390/ijms232213881.

Palmer, J. M., and Stajich, J. E. (2022). Funannotate (Version 1.8.13) [Computer software].

Park, H.-S., Lee, W. K., Lee, S.-C., Lee, H. O., Joh, H. J., Park, J. Y., et al. (2021). Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. *Sci. Rep.* 11, 2506. doi: 10.1038/s41598-021-81988-w.

Park, S., Jansen, R. K., and Park, S. (2015). Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the rpl32 gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biol.* 15, 40. doi: 10.1186/s12870-015-0432-6.

Park, S., and Park, S. (2020). Large-scale phylogenomics reveals ancient introgression in Asian Hepatica and new insights into the origin of the insular endemic *Hepatica maxima*. *Sci. Rep.* 10, 16288. doi: 10.1038/s41598-020-73397-2.

Paun, O., Stuessy, T. F., and Hörandl, E. (2006). The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex. *New Phytol.* 171, 223–236. doi: 10.1111/j.1469-8137.2006.01738.x.

Pellino, M., Hojsgaard, D., Hörandl, E., and Sharbel, T. (2020). Chasing the apomictic factors in the *Ranunculus auricomus* complex: exploring gene expression patterns in microdissected sexual and apomictic ovules. *Genes (Basel).* 11, 1–22. doi: 10.3390/genes11070728.

Pellino, M., Hojsgaard, D., Schmutzer, T., Scholz, U., Hörandl, E., Vogel, H., et al. (2013). Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. *Mol. Ecol.* 22, 5908–5921. doi: 10.1111/mec.12533.

Postel, Z., and Touzet, P. (2020). Cytonuclear genetic incompatibilities in plant speciation. *Plants* 9, 487. doi: 10.3390/plants9040487.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Curr. Protoc. Bioinforma.* 70, 1–29. doi: 10.1002/cpbi.102.

Pucker, B., Irisarri, I., de Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the era of long reads: Progress, challenges and future directions. *Quant. Plant Biol.* 3, e5. doi: 10.1017/qpb.2021.18.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033.

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592, 737–746. doi: 10.1038/s41586-021-03451-0.

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191.

Scheunert, A., Dorfner, M., Lingl, T., and Oberprieler, C. (2020). Can we use it? On the utility of de novo and reference-based assembly of Nanopore data for plant plastome sequencing. *PLoS One* 15, e0226234. doi: 10.1371/journal.pone.0226234.

Schmidt, A. (2020). Controlling apomixis: Shared features and distinct characteristics of gene regulation. *Genes (Basel).* 11, 329. doi: 10.3390/genes11030329.

Shi, C., Liu, Y., Huang, H., Xia, E.-H., Zhang, H.-B., and Gao, L.-Z. (2013). Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One* 8, e59620. doi: 10.1371/journal.pone.0059620.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345.

Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandáková, T., Lysak, M. A., et al. (2021). Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan Mountain endemic *Cardamine barbaraeoides*. *Front. Plant Sci.* 12, 1–22. doi: 10.3389/fpls.2021.659275.

Sloan, D. B. (2013). One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol.* 200, 978–985. doi: 10.1111/nph.12395.

Smith, D. R., and Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc. Natl. Acad. Sci.* 112, 10177–10184. doi: 10.1073/pnas.1422049112.

Sobanski, J., Giavalisco, P., Fischer, A., Kreiner, J. M., Walther, D., Schöttler, M. A., et al. (2019). Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *Proc. Natl. Acad. Sci. U. S. A.* 116, 5665–5674. doi: 10.1073/pnas.1811661116.

Soltis, P. S., Marchant, D. B., Van de Peer, Y., and Soltis, D. E. (2015). Polyploidy and

genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013.

Stein, D. B., Palmer, J. D., and Thompson, W. F. (1986). Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. *Curr. Genet.* 10, 835–841. doi: 10.1007/BF00418530.

Syngelaki, E., Paetzold, C., and Hörandl, E. (2021). Gene expression profiles suggest a better cold acclimation of polyploids in the Alpine species *Ranunculus kuepferi* (Ranunculaceae). *Genes (Basel).* 12, 1818. doi: 10.3390/genes12111818.

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017.

Tomasello, S., Karbstein, K., Hodač, L., Pätzold, C., and Hörandl, E. (2020). Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate montane plant species: a case study on the *Ranunculus auricomus* species complex. *Mol. Ecol.* 29, 2031–2049. doi: 10.1111/mec.15458.

Vaillancourt, B., and Buell, C. R. (2019). High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv*, 1–13. doi: 10.1101/783159.

Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116.

Walker, J. F., Jansen, R. K., Zanis, M. J., and Emery, N. C. (2015). Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. *Am. J. Bot.* 102, 1751–1752. doi: 10.3732/ajb.1500299.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., and Jaffe, D. B. (2018). Corrigendum: Direct determination of diploid genome sequences. *Genome Res.* 28, 606.1-606.1. doi:

10.1101/gr.235812.118.

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383.

Xie, J., Zhao, H., Li, K., Zhang, R., Jiang, Y., Wang, M., et al. (2020). A chromosome-scale reference genome of *Aquilegia oxysepala var. kansuensis*. *Hortic. Res.* 7, 113. doi: 10.1038/s41438-020-0328-y.

Zhai, W., Duan, X., Zhang, R., Guo, C., Li, L., Xu, G., et al. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenet. Evol.* 135, 12–21. doi: 10.1016/j.ympev.2019.02.024.

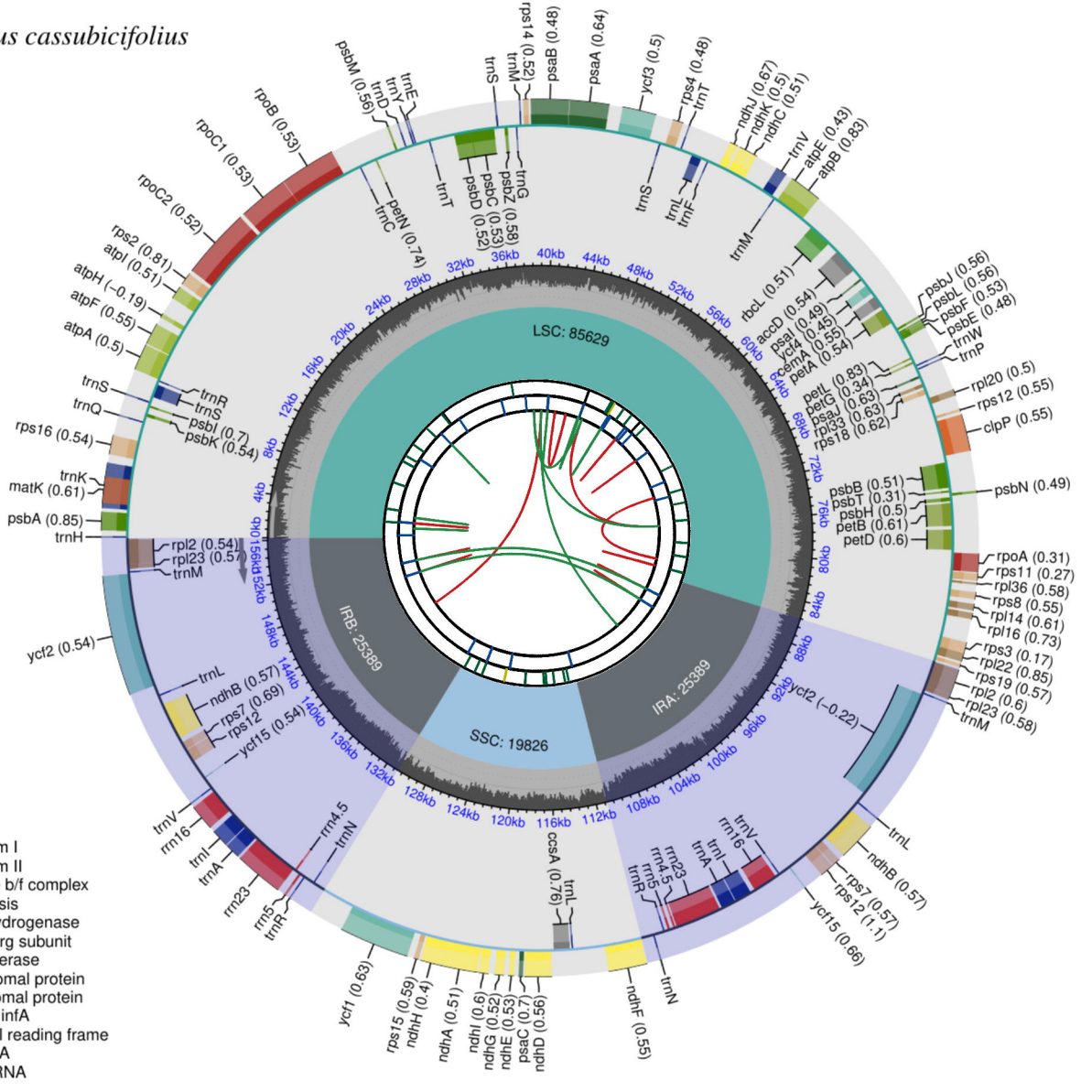
Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* 7, 203–214. doi: 10.1089/10665270050081478.

Zimin, A. V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., et al. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* 27, 787–792. doi: 10.1101/gr.213405.116.

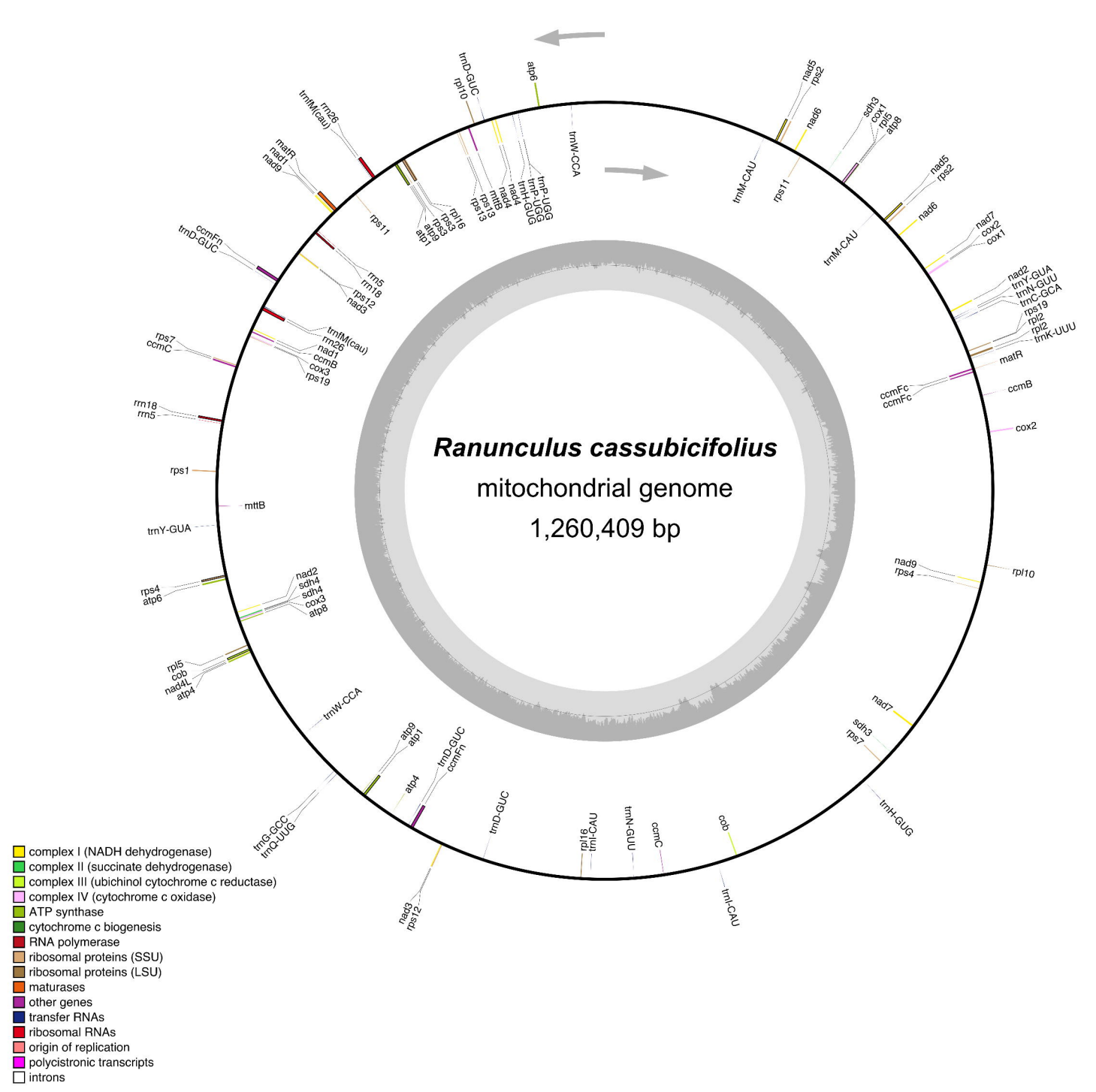
Zimin, A. V., and Salzberg, S. L. (2020). The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLOS Comput. Biol.* 16, e1007981. doi: 10.1371/journal.pcbi.1007981.

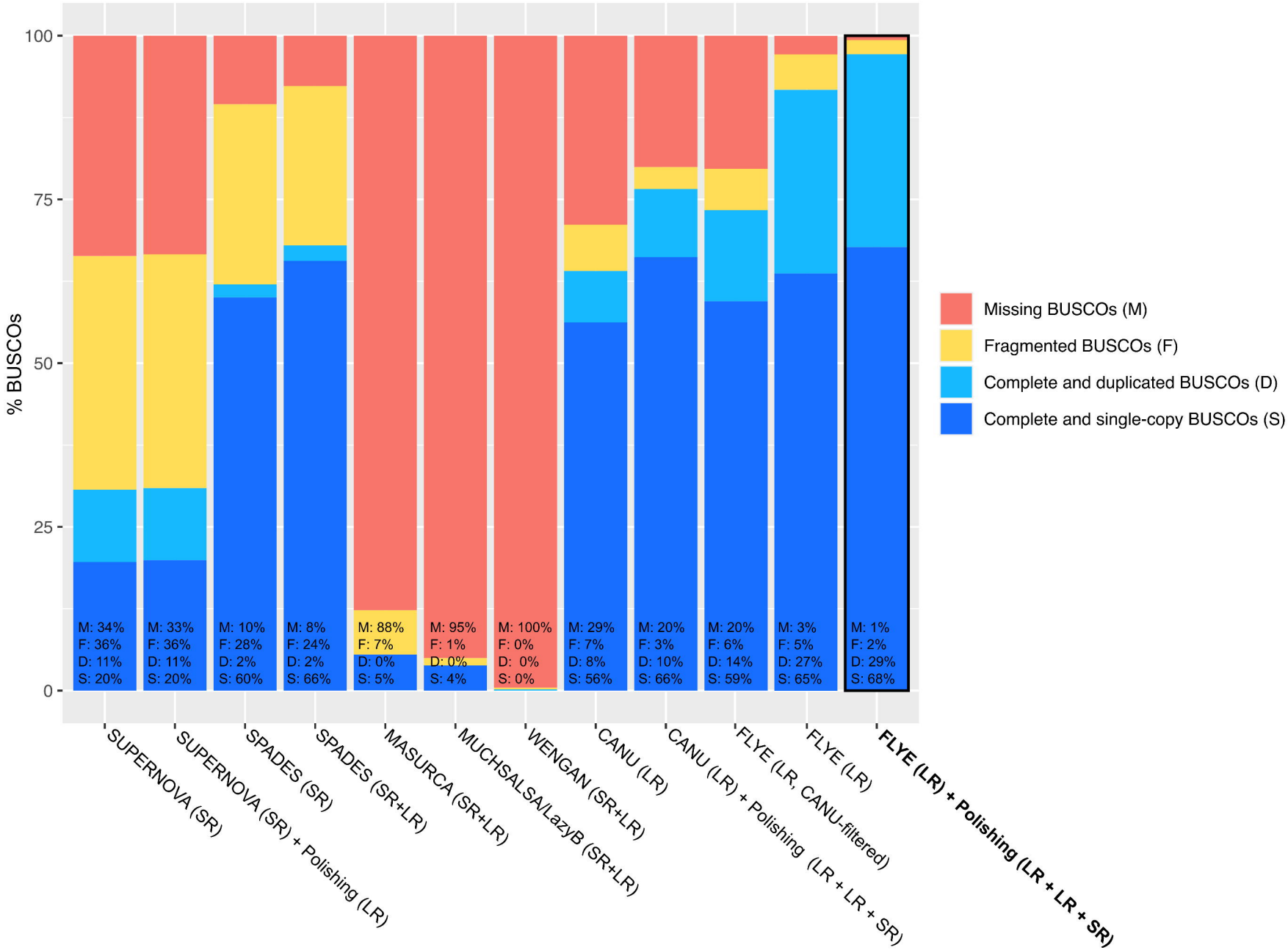
Zimin, A. V., and Salzberg, S. L. (2022). The SAMBA tool uses long reads to improve the contiguity of genome assemblies. *PLoS Comput. Biol.* 18, 1–11. doi: 10.1371/journal.pcbi.1009860.

Ranunculus cassubicifolius



- photosystem I
- photosystem II
- cytochrome b/f complex
- ATP synthesis
- NADH dehydrogenase
- RubisCO larg subunit
- RNA polymerase
- small ribosomal protein
- large ribosomal protein
- clpP, matK, infA
- hypothetical reading frame
- transfer RNA
- ribosomal RNA
- other


Ranunculus cassubicifolius
mitochondrial genome
1,260,409 bp
complex I (NADH dehydrogenase)
complex II (succinate dehydrogenase)
complex III (ubichinol cytochrome c reductase)
complex IV (cytochrome c oxidase)
ATP synthase
cytochrome c biogenesis
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
maturases
other genes
transfer RNAs
ribosomal RNAs
origin of replication
polycistronic transcripts
introns


% BUSCOs
100
75
50
25
0
M: 34%
F: 36%
D: 11%
S: 20%
M: 33%
F: 36%
D: 11%
S: 20%
M: 10%
F: 28%
D: 2%
S: 60%
M: 8%
F: 24%
D: 2%
S: 66%
M: 88%
F: 7%
D: 0%
S: 5%
M: 95%
F: 1%
D: 0%
S: 4%
M: 100%
F: 0%
D: 0%
S: 0%
M: 29%
F: 7%
D: 8%
S: 56%
M: 20%
F: 3%
D: 10%
S: 66%
M: 20%
F: 6%
D: 14%
S: 59%
M: 3%
F: 5%
D: 27%
S: 65%
M: 1%
F: 2%
D: 29%
S: 68%
SUPERNOVA (SR)
SUPERNOVA (SR) + Polishing (LR)
SPADES (SR)
SPADES (SR+LR)
MASURCA (SR+LR)
MUCHSALSA/LazyB (SR+LR)
WENGAN (SR+LR)
CANU (LR)
CANU (LR) + Polishing (LR + LR + SR)
FLYE (LR, CANU-filtered)
FLYE (LR)
FLYE (LR) + Polishing (LR + LR + SR)
Missing BUSCOs (M)
Fragmented BUSCOs (F)
Complete and duplicated BUSCOs (D)
Complete and single-copy BUSCOs (S)