Efficient assembly of plant genomes: A case study with evolutionary implications in *Ranunculus* (Ranunculaceae)

Karbstein, Kevin[1,2*], Choudhary, Nancy[3], Xie, Ting[4], Tomasello, Salvatore[1], Wagner, Natascha D.[1], Barke, Birthe H.[1], Paetzold, Claudia[5], Bradican, John P.[1], Preick, Michaela[6], Himmelbach, Axel[7], Stein, Nils[7,8], Papantonis, Argyris[4], Irisarri, Iker[9], de Vries, Jan[10,11,12], Pucker, Boas[3], & 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] *Max Planck Institute for Biogeochemistry, Department of Biogeochemical Integration, Jena, Germany*

[3] *TU Braunschweig, Institute of Plant Biology, Braunschweig, Germany*

[4] *University Medical Center Göttingen, Institute of Pathology, Göttingen, Germany*

[5] *Senckenberg Naturhistorische Sammlungen, Dresden, Germany*

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

[7] *Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany*

[8] *Center of integrated Breeding Research (CiBreed), Department of Crop Sciences, Göttingen, Germany*

[9] *Museo Nacional de Ciencias Naturales (MNCN-CSIC), Department of Biodiversity and Evolutionary Biology, Madrid, Spain*

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

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

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

ORCID

Kevin Karbstein 0000-0003-1424-6557

Nancy Choudhary 0000-0002-2562-7905

Ting Xie 0000-0002-6648-7697

Salvatore Tomasello 0000-0001-5201-1156

Natascha D. Wagner 0000-0001-6623-7623

Birthe H. Barke 0000-0001-5379-6055

Claudia Paetzold 0000-0002-4128-6645

John P. Bradican 0000-0001-5650-6172

Axel Himmelbach 0000-0001-7338-0946

Nils Stein 0000-0003-3011-8731

Argyris Papantonis 0000-0001-7551-1073

Iker Irisarri 0000-0002-3628-1137

Jan de Vries 0000-0003-3507-5195

Boras Pucker 0000-0002-3321-7471

Elvira Hörandl 0000-0002-7600-1128

* **corresponding authors:** *email: kkarb@bgc-jena.mpg.de or kevin.karbstein@tu-ilmenau.de, elvira.hoerandl@biologie.uni-goettingen.de*

Running Head

Strategies to assemble non-model plant genome sequences

Abstract

Currently, it is still a challenge - in terms of laboratory effort and cost, as well as assembly quality - to unravel the sequence of large and complex genomes from non-model plants. This often hampers the study of evolutionarily intricate species groups. The species-rich genus *Ranunculus* (Ranunculaceae) is an angiosperm model system for the study of polyploidy, apomixis, reticulate evolution, and biogeography. However, neither mitochondrial, nor high-quality nuclear genome sequences are available. This limits phylogenomic, functional, and taxonomic analyses thus far. Here, we tested Illumina short-read, Oxford Nanopore Technology (ONT) or PacBio/HiFi long-read, and hybrid-read assembly strategies. We used the diploid progenitor species *R. cassubicifolius* (*R. auricomus* complex), and selected the best assemblies in terms of completeness, contiguity, and quality scores. We first assembled the plastome (156 kbp, 85 genes) and mitogenome (1.18 Mbp, 40 genes) sequences using Illumina and Illumina-PacBio-hybrid strategies, respectively. We also present an updated plastome and the first mitogenome phylogeny of Ranunculaceae, including studies of gene loss (e.g., *infA*, *ycf15*, or *rps*) with evolutionary implications. For the nuclear genome, we favored a PacBio-based assembly three-times polished with filtered reads and subsequently scaffolded into 8 pseudochromosomes by chromatin conformation data (Hi-C) as the representative sequence. We obtained a haploid genome sequence with 2.69 Gbp, 94.5% complete BUSCO 'embryophyta_odb10' genes found, and 31,322 annotated genes. The genomic information presented here will improve phylogenomic analyses in this species complex, and will enable advanced functional, evolutionary, and biogeographic analyses for the genus and beyond Ranunculaceae in the future.

Keywords

gene evolution, large plant genomes, plastome, mitogenome, nuclear genome, Illumina vs. Nanopore vs. PacBio sequencing, *de novo* assembly strategies, *Ranunculus auricomus,* Ranunculaceae

Significance Statement

The genus *Ranunculus* is a model system in flowering plants for polyploidy, apomixis, evolution, and biogeography research. We present the first nuclear and mitochondrial genome sequence, and the plastid evolution of Ranunculaceae. Using Illumina, ONT, and PacBio data, we developed an efficient assembly strategy that can be applied to other non-model plants.

Results presented here are useful for improving population genomic and phylogenomic analyses, and enable better functional analyses at species, genus and family level.

87 **Resource Article**

88

89 ## Introduction

90 Over the past decade, the scientific community has experienced a massive expansion in
91 accessible sequencing technologies. For constructing genome assemblies, short-reads (e.g.,
92 Weisenfeld et al., 2018; Prjibelski et al., 2020), long-reads (e.g., Koren et al., 2017; Cheng et
93 al., 2021), or a combination of both (hybrid-read; e.g., Di Genova et al., 2021; Gatter et al.,
94 2021; Zimin and Salzberg, 2022) have frequently been used. Nevertheless, biologists have
95 moved beyond short-read (100-250 bp) Illumina sequence data to perform genome assemblies.
96 Short reads often produce incomplete assemblies, and lack the contiguity required to resolve
97 repetitive regions that plague large genomes, leading to highly fragmented and incomplete *de*
98 *novo* genome assemblies (Gatter et al., 2021; Rhie et al., 2021; Zimin and Salzberg, 2022).
99 Today, long-read technologies, such as nanopore sequencing (Oxford Nanopore Technologies,
100 ONT) or single molecule real time sequencing (SMRT; Pacific Biosciences), can bridge such
101 gaps and improve the quality of genome assemblies. High-fidelity (HiFi) reads derived from
102 SMRT sequencing are typically of high per-base quality (>99.5%) and up to 25 kbp in length.
103 In contrast, ONT sequencing just crossed the 99% per-base accuracy threshold, read lengths are
104 limited only by the input DNA fragment size thus reaching >1 Mbp, and can be easily
105 performed in any laboratory (reviewed in Pucker et al., 2022).

106 HiFi or ONT sequencing, combined with proximity information obtained by
107 chromosome conformation capture sequencing (Hi-C), has started to become prevalent for
108 chromosome-scale, high-quality model but also non-model plant genome sequences (e.g.,
109 Mascher et al., 2017; Xie et al., 2020; Hoencamp et al., 2021; Qu et al., 2023). The quality and
110 completeness of the assembly can be improved by polishing strategies using available, quality-
111 filtered Illumina, ONT, or PacBio reads (e.g., Zimin and Salzberg, 2020; Dmitriev et al., 2021;
112 Chang et al., 2023). The combination of low- to medium-sequencing coverage (that is, the
113 proportion of the total assembly that is supported by the raw reads) long-read data with highly
114 accurate, often already available Illumina sequences in hybrid assemblies is particularly
115 attractive for improving genome assemblies when resources are limited. This is usually the case
116 when working with non-model plants.

117 Producing "high-quality" genome sequences, that is genome assemblies characterized
118 by > 90% completeness with respect to expected genome size, > 1-10 Mbp contiguity (N_{50}),
119 and > 90% recovery rate of conserved protein-coding genes (BUSCO; see assembly standards

120 of the Earth Biogenome Project; Table S1 in Lawniczak et al., 2022), represent an important
121 target for research groups. However, some metrics are rather adapted to animals. For example,
122 flowering plants have undergone many ancient or recent polyploidization events (Van De Peer
123 et al., 2017; Leebens-Mack et al., 2019; Hörandl, 2022) and are characterized by quite large,
124 repetitive genomes with duplicated sequence content. Consequently, a BUSCO
125 'embryophyta_odb10' score of > 90% single-copy genes can often not be met because genes
126 are often duplicated in the genome (Dmitriev et al., 2021) or is not informative for quality
127 assessment in non-coding, repetitive DNA regions (e.g., LTR assembly index; Mokhtar et al.,
128 2023). Irrespective of chosen metrics, efficient strategies in terms of laboratory effort, cost, and
129 assembly quality are needed to reliably reconstruct large and complex high-quality genome
130 sequences for non-model organisms. This often hampers the study of evolutionary,
131 biogeographic, and ecological relationships within groups characterized by intricate
132 evolutionary processes such as hybridization, polyploidy, and/or apomixis (Ennos et al., 2005;
133 Karbstein et al., 2024).

134 *Ranunculus* represents a cosmopolitan genus with approximately 600 species, of which
135 about 40% are polyploid (Hörandl et al., 2005) and have rather large genomes (ranging from
136 1.8 to 24.6 Gbp; Leitch et al., 2019). The *Ranunculus auricomus* complex is one of the largest
137 polyploid apomictic plant species groups in Eurasia, with more than 840 described apomictic
138 taxa (Hörandl and Raab-Straube, 2015). The complex is a model system for understanding the
139 evolution and biogeography of young flowering plant groups formed by reticulate evolution
140 and allopolyploidy (Paun et al., 2006; Hörandl et al., 2009; Pellino et al., 2013; Hodač et al.,
141 2018; Karbstein et al., 2020b, 2020a, 2021, 2022; Tomasello et al., 2020; Karbstein, 2021;
142 Bradican et al., 2023, 2024; Hodač & Karbstein et al., 2023). It has also been frequently studied
143 for the molecular, cytogenetic, environmental, and evolutionary mechanisms of apomixis (i.e.,
144 the asexual reproduction via seeds; Hörandl et al., 2024).

145 Apomixis is relevant for agricultural applications (e.g., Schmidt, 2020), and it is often
146 associated with hybridization and/or polyploidy, which are considered major evolutionary
147 forces in plant evolution. At least 30% of flowering plant species are of recent polyploid origin,
148 while > 25% of extant plant species are known to be involved in interspecific hybridization or
149 introgression events (Mallet, 2005; Soltis et al., 2015; Van De Peer et al., 2017; Leebens-Mack
150 et al., 2019; Hörandl, 2022). Consequently, improving (phylo)genomic analyses for di- and
151 polyploids, and observing regulatory mechanisms of apomixis (e.g., Brukhin et al., 2019) with
152 *R. auricomus* reference genome sequences will help develop a better understanding of plant
153 evolution and biodiversity, and is therefore of general interest in biological research. Organellar

genome reconstructions are also valuable targets for studying nuclear-organelle compatibility and the chimerism of cyto-nuclear complexes (Sloan, 2013; Møller et al., 2021). Little is known, for instance, about plant mitochondrial genomes, although mitochondrial-nuclear compatibility is important for speciation, evolution of crossing barriers, and emergence of cytoplasmic male sterility (Postel and Touzet, 2020; Møller et al., 2021). Elucidating cyto-nuclear complexes may be important for studying a variety of evolutionary processes. For instance, plastid information may provide additional insights into phylogenetic relationships (Tomasello et al., 2024), maternal progenitors of hybrids (Karbstein et al., 2022), or the ability of hybridization (Sobanski et al., 2019; Postel and Touzet, 2020), or mitochondrial information into plant sex evolution and determination (e.g., male sterility; (Havird et al., 2015; Park et al., 2021; Hörandl, 2024).

The *R. auricomus* complex contains five newly circumscribed, predominantly diploid sexual species (Karbstein et al., 2020b), which gave rise to the >800 described allopolyploid apomictic taxa within the last ca. 800 kyr (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), recalculated as 5.998-6.174 Gbp (Leitch et al., 2019). Therefore, *R. auricomus* genomes are large even at the diploid level. Increasing proportions of repetitive regions are responsible for genome size inflation (Bennetzen, 2005; Peška et al., 2019; Guo et al., 2021; Jayakodi et al., 2023), which could also be relevant for *R. auricomus* genomes (Hörandl et al., 2009; Hodač et al., 2018; Karbstein et al., 2022). However, so far, high-quality reference genome sequences for the species complex, but also for the entire genus *Ranunculus* are missing. In Ranunculaceae, the only other reference genome sequences are available from diploid *Aquilegia* and *Coptis* species (Xie et al., 2020; Liu et al., 2021; Mokhtar et al., 2023), but these genomes are much smaller (ca. 1 Gbp; Leitch et al., 2019), and no polyploid species complexes are known in these genera (TROPICOS database v3.4.2, http://www.tropicos.org; CCDB database v1.66, https://taux.evolseq.net/CCDB_web; Rice et al., 2015).

The evolutionary relationships among the diploid sexual progenitors of *R. auricomus* have recently been clarified (Karbstein et al., 2020b; Tomasello et al., 2020). A diploid *Ranunculus* reference genome sequence 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 in hybrids; it would be the starting point for more sophisticated approaches such as reference-guided genome assemblies (e.g., Eaton and Overcast, 2020), polyploid phasing used for multilabeled species trees (e.g., Dauphin et al., 2018), or

188 allopolyploid network reconstructions (e.g., Jones, 2017; Šlenker et al., 2021). In addition, the
189 annotation of thousands of nuclear genes would also help enable sophisticated structural and
190 functional analyses in the species complex. New developments allow the incorporation of
191 external hints such as RNA-seq data or the annotation of homologous genes through knowledge
192 transfer from other species (Pucker, 2024). Tools are also needed for automatically extracting
193 and comparing group-specific annotations across genome sequences from databases, which are
194 essential for evolutionary studies. In *Ranunculus*, and in Ranunculaceae, such large-scale
195 studies are missing. Annotated high-quality genome sequences are essential for functional
196 genomics in the context of apomixis, adaptation, and gene expression (e.g., Syngelaki et al.,
197 2021), and for studying loci under selection and mutation-selection dynamics (e.g., Del Amparo
198 et al., 2021) that could so far only be performed on (incomplete) transcriptomes (Pellino et al.,
199 2013; Paetzold et al., 2022).
200 To generate plastid, mitochondrial, and high-quality nuclear genome assemblies of a
201 diploid species of the *R. auricomus* complex, we applied different assembly strategies using
202 short Illumina reads and long ONT or PacBio reads under low to medium sequencing coverage,
203 with the best nuclear-genome assembly being scaffolded by Hi-C reads. This represents a
204 typical scenario when working with non-model plant groups on a limited budget. We thus
205 focused on applying different approaches based on the available sequencing data. To study
206 organellar gene evolution in detail, we also downloaded and analyzed all available sequences
207 from NCBI.
208
209
210 **Methods**
211 **Plant Material**
212 In the present study, analyses were performed on diploid sexual individuals (2n = 16)
213 of *Ranunculus cassubicifolius* (*R. auricomus* complex; taxonomic treatment follows Karbstein
214 et al., 2020b). Chromosome numbers, measurements of ploidy and mode of reproduction are
215 published in Hörandl and Greilhuber (2002) under '*R. carpaticola*', and in Karbstein et al.
216 (2020b, 2021). Fresh leaf material was harvested in 2019 from individual EH8483/10 (NCBI
217 BioSample SAMN27753242), sampled in Revúca, Slovakia (48°41'39"N, 20°07'42"E; Hörandl
218 and Greilhuber, 2002; under '*R. carpaticola*'). High-molecular-weight DNA was extracted
219 immediately after harvest and used for Illumina DNA sequencing. The same population was
220 already used for transcriptome (Pellino et al., 2013) and population genetic analyses (Paun et
221 al., 2006; Hodač et al., 2018). In autumn 2021, we harvested fresh, young leaf material for ONT

sequencing, after 12-hour dark-adaptation at room temperature in the early morning from another conspecific individual collected nearby (LH040/02, SAMN27753242; 48°40'38"N, 20°06'24"E; Figure 1; Karbstein et al., 2020b) because the previously used individual died. In 2023, we used individuals from the same population for PacBio and Hi-C sequencing, LH040/06 and LH040/05+06 (SAMN44350262), respectively, because the previously harvested individual did not sprout again. All plants were cultivated at the Old Botanical Garden of the University of Göttingen.

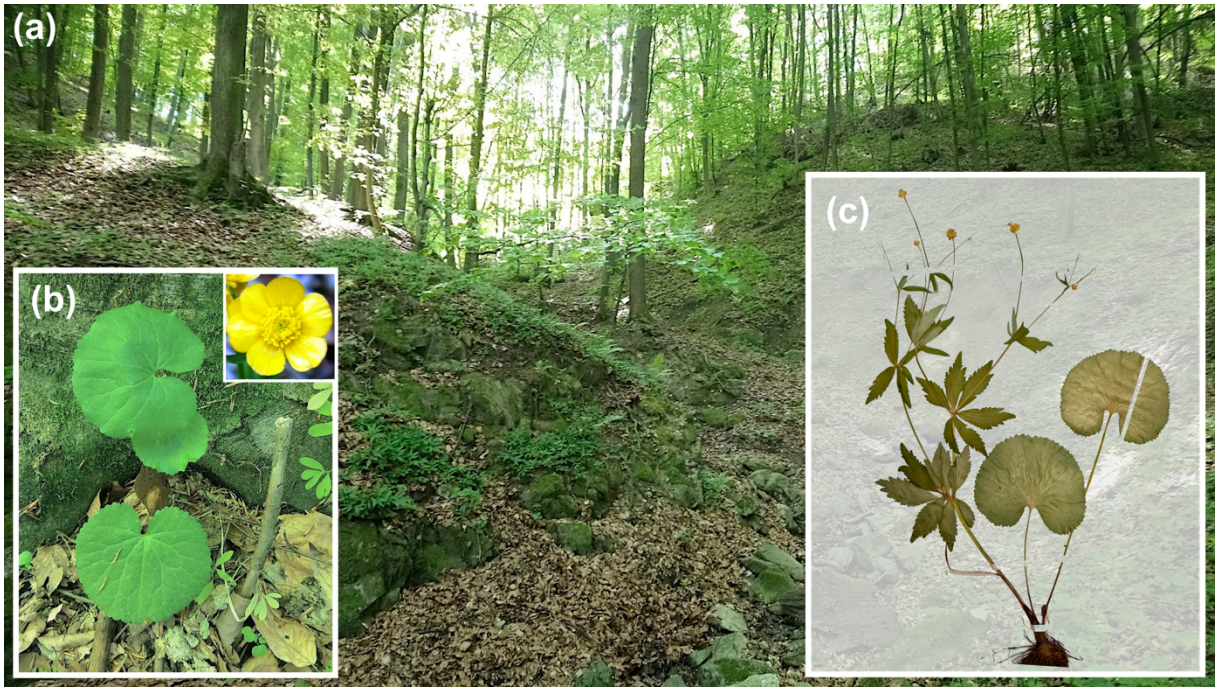


Figure 1. (a) Habitat of *Ranunculus cassubicifolius* s.l. (previously listed as '*R. carpaticola*') population LH040 in a forest near Liptov, Slovakia (4th May 2018, L. Hodač & K. Spitzer; details in Karbstein et al., 2020b). (b) Shown are the typical undissected basal leaves *in situ* and a typical *R. cassubicifolius* flower from a different population (LH009) and (c) the scan of a representative herbarium specimen from LH040. Field images: L. Hodač.

Illumina DNA extraction and sequencing

For Illumina short-read sequencing, genomic DNA (gDNA) was directly extracted from ca. 200 mg of fresh leaf material from a single *R. cassubicifolius* s.l. individual (EH8483/10) using liquid nitrogen and the DNeasy PowerPlant Pro Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA size and concentration were measured with the Qubit Fluorometer 2.0 and the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Waltham,

243 USA), and the Genomic DNA ScreenTape device (Agilent Technologies, Santa Clara, USA).
244 The extraction yielded 16 ng/µl of DNA with a median fragment length of 19 kb.
245 Library preparation followed the 10x Chromium Controller Genomics Kit (10x
246 Genomics, Pleasanton, USA) according to the manufacturer's instructions. Library quality was
247 assessed by a qPCR based on the NEBNext Library Quant Kit (New England Biolabs Inc.,
248 Ipswich, USA). Sequencing with standard Illumina primers was performed on a NextSeq
249 500/550 instrument (Illumina, San Diego, USA) with 2 x 150 bp reads.
250 We converted the raw output into FASTQ files using SUPERNOVA v2.1.1 (10x
251 Genomics, San Francisco, USA; Marks et al., 2017; Weisenfeld et al., 2018) and BLC2FASTQ
252 v2.20 (Illumina, San Diego, USA), resulting in 51 Gb raw FASTQ files. We conducted polyX-
253 tail and adapter trimming, and quality filtering of Illumina raw reads using ATRIA v3.1.2
254 (Chuan et al., 2021) with default settings, and BBTools/BBMap v39.01 ('repair.sh'; (Bushnell,
255 2022) to reorder filtered paired-end reads.
256
257 **Nanopore DNA extraction and sequencing**
258 For ONT, gDNA extraction was mainly performed according to the Qiagen protocol for
259 the isolation of gDNA from plants and fungi, and the manufacturer's instructions of the Qiagen
260 Genomic DNA Buffer Set using Genomic Tip 20/G (see details in Text S1). Several protocol
261 modifications were applied to increase DNA yield and to optimize DNA quality (see also
262 Vaillancourt and Buell, 2019; Li et al., 2020). Three gDNA extractions were performed in
263 parallel from a single *R. cassubicifolius* s.l. individual (LH040). Fresh leaf material (ca. 75 mg)
264 was frozen in liquid nitrogen and ground with a mortar and pestle, and transferred into a 2.0 ml
265 Eppendorf tube containing 400 µl API Qiagen extraction buffer (Qiagen DNeasy Plant Mini
266 Kit).
267 The mixture was incubated at 40°C for 1.5 hours with gentle agitation. Then, 4 µl
268 Qiagen RNase A (100 mg/ml) was added, and the mixture was incubated again at 37°C for 30
269 minutes, followed by the addition of 18 µl Qiagen Proteinase K (20 mg/ml), and incubation at
270 50°C for 2 h with gentle agitation. The mixture was centrifuged at 12,000 x g for 20 min, and
271 ca. 400 µl of the lysate was transferred with wide-bore tips to a buffer QBT-equilibrated
272 Genomic Tip. DNA was washed with 5 x 1 ml QC Buffer, and eluted with 0.8 ml 55°C pre-
273 warmed QF Buffer into a new 2.0 ml Eppendorf tube. DNA was precipitated by adding 70%
274 room-temperature isopropanol and inverting the tube 20 times, followed by incubation at room
275 temperature for 15 min. Using a glass rod, the DNA was rolled up and placed into a new 1.5 ml
276 Eppendorf tube containing ca. 0.8 ml ice-cold 80% ethanol. The DNA was air dried,

277 resuspended in 200 µl autoclaved ddH_2O, and all three extractions were pooled into a single
278 tube, stored at 4 °C until library preparation.
279 DNA concentration was assessed using the Qubit Fluorometer 3.0 and the Qubit dsDNA
280 HS Assay Kit, and the DNA purity was checked with the NanoDrop 2000 Spectrophotometer
281 (ThermoFisher Scientific, Waltham, USA; Text S1). DNA fragment length distribution and
282 RNA absence were checked on a 2% agarose gel via electrophoresis (50 V, 90 min) and the
283 Quick-Load 1 kb DNA Ladder (New England Biolabs Inc., Ipswich, USA; Text S1, Fig. S1).
284 After several dilution steps, we obtained a final DNA yield of 61.2 µg (51.0 ng/µl * 1200 µl
285 ddH_2O).
286 Library preparation was conducted using the ONT Ligation Sequencing Kit SQK-
287 LSK110 optimized for high throughput and long reads (ONT, Oxford, UK), and applicable for
288 singleplex gDNA sequencing. We adjusted the DNA concentration to 1000 ng in 47 µl (ca. 21.5
289 µl/ng) and followed the manufacturer's instructions for library preparation (protocol
290 vGDE_9108_v110_revL_10Nov2020, available at community.nanoporetech.com) with a few
291 modifications. Incubation times were increased up to 15 min, the concentration of ethanol wash
292 buffer was increased to 80%, and DNA fragments > 3 kbp were enriched using the buffers
293 provided with the kit. Active pores ranged from ca. 800-1500. Flow cell 1 was loaded with
294 freshly prepared libraries, flow cell 2 with one freshly prepared and one library stored at -80°C
295 for two days, and flow cells 3-6 with two libraries stored at -80°C for 1-2 weeks. Freezing of
296 gDNA had no apparent effect on N_{50} fragment size (see also Results & Discussion and flow
297 cell reports on FigShare). We used a MinION 101B device, the ONT software MinKNOW
298 v21.11.7, 21.11.9, and 22.03.5 on a local Linux system, and six R9.4.1 flow cells following the
299 manufacturer's instructions for priming and loading. After 1-2 days of run time, the old library
300 was exhausted and replaced with a second library using the supplied flow cell wash kit (protocol
301 vWFC_9120_v1_revB_08Dec2020). Sequencing was run for 3-5 days per flow cell until the
302 majority of pores were inactive (> 99%).
303 Reports per flow cell can be accessed via FigShare. Basecalling was performed with the
304 ONT software GUPPY v6.0.1 specifying the configuration file "dna_r9.4.1_450bps_fast.cfg"
305 for fast base-calling (in preruns, the difference of median per-base quality between 'fast' and
306 the substantially slower 'high accuracy' was quite small, i.e., 97 vs. 98%, see FastQC files on
307 Figshare). After that, NanoFilt v2.2.0 (De Coster et al., 2018) was applied to remove the first
308 50 bp of the reads that were usually of low quality.

309 **PacBio DNA extraction and sequencing**

310 High-molecular-weight gDNA was isolated from fresh leaf material of a single *R.*
311 *cassubicifolius* s.l. individual (LH040) with the NucleoBond HMW DNA kit (Macherey Nagel,
312 Germany), following the manufacturer's instructions. The integrity of the DNA was determined
313 by the Femto Pulse System (Agilent Technologies Inc, CA, USA). The DNA concentration was
314 measured with the Qubit dsDNA High Sensitivity assay kit (Thermo Fisher Scientific, MA,
315 USA). For the construction of a HiFi library, 7.5 µg HMW DNA was fragmented (speed 29)
316 using the Megaruptor 3 device (Diagenode SA, Seraing, Belgium).

317 In total, two HiFi libraries were constructed and sequenced into one Revio 25M SMRT
318 cell. Libraries were generated according to the 'Procedure & Checklist - Preparing whole
319 genome and metagenome libraries using SMRTbell® prep kit 3.0' manual (102-166-600,
320 Pacific Biosciences of California Inc., USA). For size fractionation, a SageELF (Sage Science,
321 USA) device was employed. The final sequencing library size (16.2 kb) was measured using
322 the Femto Pulse System (Agilent Technologies Inc, USA). Polymerase-bound SMRTbell
323 complexes were formed according to standard manufacturer's protocols. Sequencing (HiFi
324 CCS) was conducted using the Pacific Biosciences Revio instrument (24 h movie time, loading
325 concentration 230 pM, 2 h pre-extension time, diffusion loading, mean insert length according
326 to SMRT link raw data report: 18.0 kb) following standard manufacturer's protocols. The Revio
327 25M SMRT cell generated 82 Gb HiFi CCS data. All steps were done at IPK Gatersleben.

328 HiFiAdapterFilt v2.0.0 (Sim et al., 2022) was run with default settings to convert the
329 HiFi reads from BAM to FASTQ format, and to remove reads with remnant PacBio adapter
330 sequences. The quality of the sequenced reads was checked by FastQC v0.11.4 (Andrews,
331 2010) (reports on FigShare).

332

333 **Chromosome conformation capture (Hi-C) DNA extraction and sequencing**

334 The Hi-C sequencing library was generated from fresh leaf material of two *R.*
335 *cassubicifolius* s.l. individuals (LH040) with the *DpnII* enzyme, essentially following the
336 method described in Padmarasu et al. (2019). The library was sequenced (paired-end: 2 x 111
337 cycles, two indexing reads: 8 cycles, v1.5 chemistry) using the NovaSeq6000 instrument
338 (Illumina, Inc., San Diego, USA) according to standard manufacturer's protocols. All steps
339 were performed at IPK Gatersleben.

340 **Genomic Data Analysis**

341 **The plastid genome.** The plastome sequence was assembled using Illumina short-read
342 data and the software GetOrganelle v1.7.5.3 (Jin et al., 2020) as recommended by Freudenthal
343 et al. (2020), specifying the 'embplant_pt' database, k-mer size range between 45 and 121, and
344 as seed the *Ranunculus repens* plastome (154,247 bp, GenBank accession number NC_036976;
345 Dann et al., 2017). The average embplant_pt base sequencing depth was 560x. The plastome
346 was annotated using CPGAVAS2 (Shi et al., 2019) with default settings and 2,544 reference
347 plastomes, and visualized with CPGView (Liu et al., 2023). We checked and edited the
348 annotation results with Geneious R11 2023.1.2 (Kearse et al., 2012). Results were also visually
349 inspected using Bandage v0.9.0 (Wick et al., 2015). The final annotated plastome sequence
350 (Fig. 2) was deposited in GenBank (NC_077490). We applied MAFFT v7.490 to align and
351 investigate detailed differences between *R. cassubicifolius* and available *Ranunculus* plastome
352 sequences. The plastid raw data (SRA: PRJNA826743) and the annotated plastome (GenBank:
353 NC_077490) are deposited in NCBI.
354

355 **The mitochondrial genome.** Two alternative assembly strategies were followed. First,
356 we used GetOrganelle with the same settings as described above, except that we specified
357 "embplant_mt" and as seed used the closest available mitogenome (*Hepatica maxima*,
358 GenBank accession number NC_053368.1; Park and Park, 2020). The average embplant_mt
359 base sequencing depth (i.e., the average number of reads per position) of Illumina reads was
360 estimated at 92x. To improve the fragmented mitogenome assembly, we first mapped ONT
361 reads against the contigs derived from GetOrganelle and extracted the mapped reads using
362 SAMtools v1.9 ('faidx' and 'view'; Danecek et al., 2021), MINIMAP2 v2.27 ('-ax map-ont';
363 Li, 2018), and BEDTools v2.29.1 ('bamtofastq'; Quinlan and Hall, 2010). We applied
364 MaSuRCA v4.0.9 (Zimin and Salzberg, 2022) to perform a hybrid assembly based on extracted
365 ONT reads and "embplant_mt" hitting Illumina reads derived from the GetOrganelle assembly
366 process.

367 Then, we retained an ONT-based mitogenome sequence in 18 contigs and a total length
368 of 351,681 bp. We aligned contigs in Geneious to merge overlapping regions, which resulted
369 in 4 contigs, onto which Illumina and ONT reads were mapped using minimap2 ('-ax sr' and '-
370 ax map-ont', respectively). Alignments were inspected with the Integrative Genomics Viewer
371 (IGV) v2.16.1 (Thorvaldsdottir et al., 2013) and Geneious to ensure that previous gaps were
372 spanned by raw reads (see alignments and information concerning contig borders on FigShare).
373 Using Geneious, we inferred a mean Illumina depth of 62x (2.87% sites with no coverage) and

374 ONT depth of 3640x (9.88% sites with no coverage), respectively. Remaining gaps were closed
375 manually based on mapping results of Illumina and ONT reads. The assembly was annotated
376 using MitoHiFi pipeline v3.2.1 (Uliano-Silva et al., 2023) including gene annotation with *H.*
377 *maxima* as reference, and visualized with PMGmap (Zhang et al., 2024;
378 http://www.1kmpg.cn/pmgmap) web browser tool. We ran RepeatModeler2 v2.0.5 to
379 additionally check for dispersed repeats.

380 In a second assembly strategy, we used PacBio HiFi reads, which were subjected to the
381 MitoHiFi all-in-one pipeline to assemble, annotate, and draw the mitogenome based on the
382 PacBio HiFi reads. The pipeline represents a similar strategy as described before by first
383 mapping length-filtered HiFi reads to the reference mitogenome, followed by assembly using
384 Hifiasm and annotation with MitoFinder (Allio et al., 2020). The pipeline resulted in a very
385 small mitogenome of about 125 kb (about 10% of the expected length). Therefore, we
386 performed a hybrid assembly strategy with the HiFi reads as described above. This resulted in
387 40 raw contigs with 1,913,766 bp summarized into 9 contigs that were aimed to be arranged
388 like the ones in the ONT hybrid mitogenome. We inferred a mean Illumina sequencing depth
389 of 269x (1.96% sites with no coverage) and HiFi depth of 3324x (0.12% sites with no coverage).
390 Finally, the mitogenome sequences were retained in a single contig based on ONT- and PacBio-
391 hybrid assemblies. Due to substantial intragenomic rearrangements, we performed
392 progressiveMauve (2015-02-25; Darling et al., 2010) instead of MAFFT in Geneious to
393 investigate sequence differences between both assemblies, and between *R. cassubicifolius* and
394 available Ranunculaceae assemblies.

395 Geneious was used to inspect and edit the annotation results. The resulting GeneBank
396 file was converted into a feature table file for GenBank submission using the GB2SEQUIN tool
397 (Lehwark and Greiner, 2019) within the MPI-MP CHLOROBOX webserver. Finally, we
398 selected the PacBio-based mitogenome because of its (4x) higher sequencing depth of mapped
399 Illumina reads and 1.5x and 82x fewer Illumina and long-read sites without coverage, although
400 the gene content (but not the gene block arrangement, see FigShare) was the same compared to
401 the ONT-based mitogenome. The annotated mitogenome is illustrated in Figure 3. The
402 mitochondrial raw data (SRA: PRJNA831351) and annotated mitogenome (GenBank:
403 PP657143) are deposited in NCBI.

404

405 **Phylogenomics and organellar gene evolution within Ranunculacaeae.** To study
406 gene evolution within Ranunculaceae, we downloaded all available organellar genome
407 sequences from NCBI using a custom Python script 'findGenome.py'

408 (https://github.com/KK260/NCBI-Genome-Tools/). Applying the following specifications
409 "complete genome" AND "Ranunculaceae" AND "chloroplast"/ "mitochondrion", we found
410 1161 plastome and 16 mitogenome sequences on NCBI that were initially filtered by duplicate
411 sequence removal, a maximum of 1 individual per species and 2 species per genus, and selection
412 of the latest RefSeq 'NC' records. Taxon names were checked for synonyms using information
413 from the Catalogue of Life on GBIF. This yielded 308 plastomes including 6 outgroup taxa of
414 Berberidaceae and Papaveraceae (295 taxa), and 10 mitogenomes including 1 outgroup taxon
415 of Papaveraceae (none available for Berberidaceae, 10 taxa) (see Table S1).
416 Plastome sequences were aligned using MAFFT with mode 'FFT-NS-2' (Katoh and
417 Standley, 2013), following the deletion of sites with less than 50%, 70%, and 90% of available
418 samples using Goalign v0.3.7 (Lemoine and Gascuel, 2021). We then ran ModelTest-NG v0.1.7
419 (Darriba et al., 2020), which determined 'TVM+I+G4' (unfiltered), 'TIM1+I+G4' (min50), and
420 'GTR+I+G4' (min70, min90) as best-fit evolutionary models for the datasets. Maximum
421 likelihood trees were inferred with RAxML-NG v1.2.1 (Kozlov et al., 2019) specifying the
422 estimated sequence model and 100 pseudoreplicates of non-parametric bootstrapping, and
423 transfer bootstrap expectation (TBE) metric. To further assess branch support, Quartet
424 Sampling (QS) v1.3.1 (Pease et al., 2018;
425 https://www.github.com/fephyfofum/quartetsampling) scores were calculated with default
426 settings and visualized using the supplied R script
427 (https://github.com/ShuiyinLIU/QS_visualization). Following the approach of Karbstein et al.
428 (2020b), the final phylogenetic tree was selected according to the highest mean TBE support
429 and optimal QS scores (QC=1, QD=0, QI=1) using the R package PHYTOOLS v2.0-3 (Revell,
430 2012, 2024; https://github.com/liamrevell/phytools). Finally, we chose 'min90' (126 kbp) with
431 the lowest number of gaps excluding poorly assembled non-coding DNA regions, and the
432 highest TBE and QS values in the phylogeny as the optimal alignment (see details in Figs. S2-
433 S4 and Text S2, and alignments on FigShare). Branches of all tribes except *Ranunculeae* were
434 collapsed with FigTree v1.4.4 (Rambaut, 2014), gene gain and loss events per genus were
435 mapped onto the tree (Fig. 4a).
436 The mitogenome phylogeny was built based on coding regions due to substantial
437 genomic rearrangements and size differences present among sequences. Using a custom Python
438 script 'assembleGenome.py' (https://github.com/KK260/NCBI-Genome-Tools/), we run Astral
439 v5.7.8 with 100 bootstraps based on collapsed gene trees of low support previously inferred by
440 RAxML_NG with 100 bootstraps, and mapped gene features onto the resulting tree. In
441 Ranunculaceae, the loss of genes per group derived from the feature table of

442 ‘assembleGenome.py’ was manually verified by mapping extracted genes from closely related
443 Ranunculaceae species to the sequence of interest in Geneious, and edited if necessary. In *R.*
444 *cassubicifolius*, we aimed to find lost organellar genes by mapping extracted genes from closely
445 related *Ranunculus* species to the nuclear genome sequence trees (Fig. 4b).

446 **The nuclear genome.** We applied different genome assembly approaches with default
447 settings and a haploid genome size of 3.2 Gbp and RAM of at least 1.5 TB, unless otherwise
448 specified: (i) using Illumina short-reads, we performed analyses with Supernova (incorporation
449 of all reads, acceptance of extreme coverages) and SPAdes v3.13.2 (Prjibelski et al., 2020); (ii)
450 using Illumina short-reads and ONT or PacBio/HiFi long-reads, we run SPAdes (untrusted
451 Supernova contigs for gap filling), MaSuRCA v4.1.0 (Zimin et al., 2017), MuCHSALSA v0.02
452 (Gatter et al., 2021) and WENGAN (M) v0.2 (Di Genova et al., 2021); (iii a) using ONT long-
453 reads, we performed analyses with Canu v2.1.1 (Koren et al., 2017) and Flye v2.9.1
454 (Kolmogorov et al., 2019) option ‘--nano-raw’), followed by the best polishing strategy of
455 Dmitriev et al. (2021) of two-times via filtered ONT reads using Racon v1.4.3 (Vaser et al.,
456 2017) and Medaka v0.7.1 (https://github.com/nanoporetech/medaka), and once using Illumina
457 reads via the POLCA function (Zimin and Salzberg, 2020); (iii b) using HiFi long-reads, we
458 run Canu, Flye, and Hifiasm v0.19.8 (Cheng et al., 2021), followed by the previous polishing
459 strategy except for the use of polishCLR v1.1.0 (Chang et al., 2023) instead of Medaka. To
460 ensure a fair comparison between assemblies based on ONT or PacBio reads within the same
461 tool, we downsampled PacBio reads to 16x coverage using reformat.sh from BBTools and reran
462 all analyses.

463 Statistics for genome assemblies were calculated using QUAST v5.0.1 (Gurevich et al.,
464 2013), and are summarized in Table 1 (reports on FigShare). Assembly completeness in terms
465 the expected single-copy genes in inherent to land plants was estimated by ‘Benchmarking
466 Universal Single-copy Orthologs’ (BUSCO v5.0.0; Waterhouse et al., 2018; Manni et al., 2021)
467 with the ‘embryophyta_odb10’ 2020-09-10 dataset containing 60 species and 1375 core genes.
468 The assembly with the highest completeness, largest contig size, and best BUSCO scores (Table
469 1, Fig. 3, reports on FigShare) was selected for scaffolding with Hi-C data and genome
470 annotation. Filtered Hi-C reads were mapped to the HiFi polished contigs using Juicer v1.6
471 (Durand et al., 2016) with default parameters. Paired reads mapping to different contigs were
472 used for the Hi-C-associated scaffolding. We applied the 3D-DNA pipeline v170123
473 (Dudchenko et al., 2017) to order and orient the clustered contigs with sizes > 15,000 bp to
474 produce chromosome-level scaffolds. Misassembles were identified and fixed manually using
475 assembling mode in the Juicebox software, which was manually corrected based on the

476 neighboring interactions. The resulting Hi-C contact map is shown in Fig. S5. To fulfill NCBI
477 quality standards, we screened for non-plant contigs using FCS-GX v0.5.4 (Astashyn et al.,
478 2024;https://github.com/ncbi/fcs) and for mitochondrial contigs using seqkit v2.8.2 (Shen et al.,
479 2016) removing 552 short contigs in total. The functional annotation was performed using
480 InterProScan5 (Jones et al., 2014) and eggNOG-mapper (Cantalapiedra et al., 2021) using the
481 Funannotate "annotate" function.

482 The genome assembly was then subjected to the identification of tandem repeats (TRs)
483 and transposable elements (TEs). Tandem Repeats Finder v4.10.1 (Benson, 1999) using
484 parameters 'Match=2, Mismatch=5, Delta=7, PM=80, PI=10, Minscore=5, and
485 MaxPeriod=2000', and EDTA v2.2.1. The EDTA pipeline integrates the results of multiple
486 high-performing tools: LTR-FINDER (Xu and Wang, 2007; Ou and Jiang, 2019),
487 LTR_retriever (Ou and Jiang, 2018), Generic Repeat Finder (Shi and Liang, 2019), TIR-
488 Learner (Xiong et al., 2014; Su et al., 2019), HelitronScanner (Xiong et al., 2014; Zhang et al.,
489 2022), and TEsorter, to achieve accurate and efficient TE identification and classification.

490 Given the limited availability of RNA-seq reads from *R. cassubicifolius*, we used a
491 multi-species RNA-seq mapping approach. This strategy included RNA reads from multiple
492 *Ranunculus* species (37 individuals, Table S1), retrieved using prefetch and fasterq-dump from
493 NCBI SRA toolkit v3.1.0 (https://github.com/ncbi/sra-tools). RNA data was mapped onto the
494 genome using HISAT2 v2.2.1 (Kim et al., 2019).

495 Structural annotation of protein-coding genes used a combination of three approaches.
496 (i) First, homology-based gene prediction using GeMoMa v1.9 (Keilwagen et al., 2016, 2018),
497 using annotations from three Ranunculaceae and eleven Ranunculales species (Table S2). The
498 coding genes from these 14 species were aligned to the *R. cassubicifolius* genome sequence
499 using MMseqs2 v15-6f452 (Steinegger and Söding, 2017) following the default GeMoMa
500 setting. GeMoMa is based on the conservation of amino acid sequences and intron position
501 conservation among closely related species. To refine intron boundaries, RNA-seq data from
502 five recently published diploid *R. auricomus* transcriptomes (Paetzold et al., 2022) were
503 incorporated. The resulting fourteen gene annotation sets were merged and filtered using the
504 GeMoMa Annotation Filter (GAF) with parameters: f="start=='M' and stop=='*' and
505 (isNaN(score) or score/aa>=1.50 or (score/aa>0.8 and avgCov>50)) and avgCov>0.0 and
506 (iAA>=0.5 or pAA>=0.5) and (evidence>1 or tpc>=0.8)" and atf="tie==1 and sumWeight>4".

507 (ii) Second, we used the BRAKER3 pipeline (Gabriel et al., 2024), which uses ab initio
508 gene prediction using AUGUSTUS v3.5.0 (Stanke et al., 2006, 2008) and GeneMark-ETP
509 v1.02 (Brůna et al., 2024), which is corrected by extrinsic evidence provided by RNA-seq and

510 protein homology. For protein evidence, we used 'Viridiplantae' library (Kuznetsov et al.,
511 2023) and annotated NCBI genomes of Ranunculaceae species (*Aquilegia caerulea*
512 (GCA_002738505.1), *Thalictrum thalictroides* (GCA_013358455.1), and *Coptis chinensis*
513 (GCA_015680905.1)). The resulting BRAKER3 annotation was filtered using GeMoMa
514 Annotation Filter (GAF) with parameters: f="start=='M' and stop=='*' and avgCov>0.0 and
515 tpc>0.5" and atf="tie==1 or sumWeight>1".

516 (iii) Third, we used Funannotate v1.8.15 (https://github.com/nextgenusfs/funannotate;
517 Palmer and Stajich, 2022), initially developed for fungi but now capable of handling larger
518 genomes. Repetitive contigs were cleaned from the genome using minimap2 v2.26 and simple
519 repeats were masked using TANTAN ('-s arabidopsis'; Frith, 2011). Approximately 8.04%
520 (217 Mbp) of the genome sequence was masked. Funannotate's gene prediction involved three
521 steps: train, predict, and update. In all steps, the parameter '--max_intronlen 1000000' was used.
522 The filtered genome and the transcriptome sequences were used to train the annotation process
523 by a genome-guided transcriptome assembly using Trimmomatic v0.39 (Bolger et al., 2014),
524 Trinity v2.8.5 (Grabherr et al., 2011), HISAT2, kallisto v0.46.1 (Bray et al., 2016), and PASA
525 v2.5.3 (Haas et al., 2008) to filter, normalize, and cluster (genome-guided) transcriptomic data.
526 The sorted alignments, transcripts, UniProt protein library, and transcriptome annotations were
527 the input for the gene prediction process performed by AUGUSTUS v3.5.0, GlimmerHMM
528 (Majoros et al., 2004), and SNAP (Korf, 2004), and results were summarized with
529 EVidenceModeler v1.1.1 (Haas et al., 2008). Additional parameters '--busco_db
530 eudicots_odb10 --organism other --busco_seed_species arabidopsis' were used in this step.

531 The annotations resulting from GeMoMa, BRAKER3, and Funannotate were compared
532 based on BUSCO completeness scores. Based on these comparisons, the GeMoMa and
533 BRAKER3 gene predictions were merged using GeMoMa Annotation Filter (GAF) with the
534 following parameters: f="start=='M' and stop=='*' and avgCov>0 and tpc>0 and aa >=50 and
535 ((isNaN(score) and (tie==1 or isNaN(tie) and tpc>0.5)) or (tie>0.5 or tpc>0.5))" and atf= "
536 tie==1 and tpc>0.5 and (isNaN(score) or sumWeight>2)". Gene models in disagreement with
537 RNA evidence data were fixed manually. The annotation quality was evaluated using BUSCO
538 v5.6.1 with the embryophyta_odb10 database.

539 The nuclear genome sequences and annotated protein-coding genes (Fig. 6) are
540 deposited in NCBI under SAMN44350262 in the Bioproject PRJNA831351. Non-coding RNA
541 features, including rRNAs, tRNAs, and short ncRNAs were identified using Infernal v1.1.5
542 (Nawrocki and Eddy, 2013) cmscan, which scanned covariance models from the Rfam database
543 v15 (Ontiveros-Palacios et al., 2024). Additionally, tRNAscan-SE v2.0.12 (Chan et al., 2021)

544 and barrnap v0.9 (available at https://github.com/tseemann/barrnap) were used to enhance the
545 detection of tRNAs and rRNAs, respectively.
546 All data analyses were performed on the HPC cluster of the GWDG (Göttingen,
547 Germany), except for the genome annotation which was run on the de.NBI cloud.

548 ## Results and Discussion

549 More than 504 million filtered Illumina reads (150 bp) were generated of *R.*
550 *cassubicifolius*, resulting in 70.7 Gbp raw data and genome coverage of 22x (mean base
551 accuracy of 99.97%), assuming a haploid genome size of 3.2 Gbp. In addition, the final ONT
552 long-read dataset contains 51.6 Gbp and 5.83 million reads (N_{50}=22.6 kbp), yielding a coverage
553 of 16x (mean base accuracy of 96.62%). The final HiFi (PacBio) long-read dataset harbors 79.1
554 Gbp and 4.40 million reads (N_{50}=17.9 kbp), resulting in a coverage of 25x (mean base accuracy
555 of 99.98%).

556 **Plastome features and gene evolution within *Ranunculus* and Ranunculaceae.**
557 Based on short Illumina reads, we retained a circular plastid genome sequence with 156,233 bp
558 in two states caused by the differing orientations of the inverted repeat (IR) regions ("flip-flop";
559 Stein et al., 1986; Walker et al., 2015). The plastome shows a typical tetrapartite structure of a
560 small single-copy region (SSC, length: 19,826 bp) and a large single-copy region (LSC, length:
561 85,629 bp) separated by two IR regions (length: 2 x 25,389 bp). The plastome contains 85 (78
562 unique) protein-coding genes, 37 (30 unique) tRNAs, and 8 (4 unique) rRNAs, resulting in a
563 coding content of 53.2%. We also observed 35 SSR, 32 long tandem, and 34 dispersed repeat
564 motifs (2.2% repeat content). For example, in comparison with the closely related *R. sceleratus*
565 (nested within the sister clade to sect. *Auricomus*) and the more distantly related *R. repens*
566 (nested within a different subgenus within Ranunculaceae), *R. cassubicifolius* shows a similar
567 number of unique protein-coding genes (78/77/78), tRNAs (30/28/29), and rRNAs (4/3/4). The
568 plastome size of *R. cassubicifolius* was similar to that of *R. sceleratus* (+96 bp, 156,329 bp),
569 but larger compared to *R. repens* (-1,986 bp; 154,247 bp), mainly due to additional mono-,
570 oligo-, or polynucleotide insertions between different protein-coding genes (e.g., see positions
571 at 15,754 bp or 126,944 bp in alignment file(s) on FigShare).

572 Within Ranunculaceae, the plastome sequence of *R. cassubicifolius* is quite average,
573 considering the longest (167 kbp; *Helleborus atrorubens*) and shortest (150 kbp; *Myosurus*
574 *apetalus*) plastomes. A clear phylogenetic trend of size evolution in Ranunculaceae is not
575 recognized (Fig. 4). With similar gene contents, length variation can be mainly attributed to
576 non-coding regions (Fig. 4a, and feature tables and alignments on FigShare). However, Long
577 et al. (2024) observed in *Myosurus* that gene length and number can also be responsible for
578 smaller plastomes, possibly reflecting adaptation to new environments, or an effect of annual
579 life cycles, as known from nuclear genomes (e.g., Lunkova and Ivanov, 2022). In general,
580 compared to other flowering plant lineages, we detected usual size variation and no substantial

plastome rearrangement among investigated plastome sequences of *Ranunculus* and Ranunculaceae, typical for an obligate autotrophic lifestyle (Petersen et al., 2018; Konupková et al., 2024).

The maximum likelihood plastome phylogeny with 292 taxa revealed a fully resolved and well-supported phylogeny of Ranunculaceae (TBE>85; Figs. 4 and S4, Text S3). Low branch support occurs within species-rich genera such as *Anemone, Hepatica,* or *Aconitum* and is probably due to fast radiation and the slow evolutionary rates of plastomes (Wang et al., 2016; Zhai et al., 2019). Almost all subfamilies or tribes are monophyletic, and branching is widely consistent with previous studies on plastome sequences (35 taxa) or four nuclear and plastid loci (105 taxa) (Wang et al., 2009; Zhai et al., 2019). Within *Ranunculeae*, *Oxygraphis*, *Ficaria*, *Myosurus,* and *Ceratocephala* represent close relatives of *Ranunculus*, in agreement with Emadzade et al. (2010). Compared to Zhai et al. (2019) and Ji et al. (2023), the here presented plastome phylogeny is the most recent and taxon-rich to date of Ranunculaceae. The plastome phylogeny of *Ranunculus* widely fits the one presented in Ji et al. (2023), that is, an Eurasian-American (clade 1), European-Asian clade including *R.* sect. *Auricomus* (clade 2), and the predominant Asian clade including aquatic taxa (clade 3). The phylogeny is also largely congruent with combined plastid-nuclear marker trees in Hörandl and Emadzade (2012), but based on a smaller taxon sampling and therefore not representative for this big genus. Further plastome sequencing efforts would allow a more taxon-rich and -balanced phylogenomic reconstruction of Ranunculaceae.

Regarding gene evolution within Ranunculaceae and particularly *Ranunculus*, we frequently did not detect some specific genes (*infA*, *pbf1*, *rpl32*, *rps16*, and *ycf15*) suggesting a potential loss (Fig. 4a; see feature tables on FigShare). Gene transfers from the chloroplast to the nucleus have been reported for *infA* but also for *accD, rpl22, rpl20, rpl32, rpl23, rps7, rps16, ycf1*, and *ycf2* in seed plants (Park et al., 2015; Zhou et al., 2023). In our study, the plastid gene *rpl32* was lost in the common ancestor of *Caltheae, Asteropyreae*, *Cimifugeae,* and *Nigelleae*, and *Delphinieae* and reappeared in *Callianthemeae* and all remaining subfamilies/tribes. The ribosomal gene *rpl32* is involved in protein biosynthesis and plastid function. Its putative loss and functional transfer to the nuclear genome have been observed in other clades within Ranunculaceae (Park et al., 2015; Zhai et al., 2019; Park and Park, 2020). The gene *rpl32* may be involved in rapid adaptation to novel environments, such as in *Myosurus* adapting from mesophytic to wet habitats during climate change over the last 5 Ma (Long et al., 2024). In *R. cassubicifolius, rpl32* shows a typical length of 162 bp and is not reduced as in *Myosurus* (138 bp), or in alpine *Clematis alpina* (69 bp).

615 The plastid gene *infA* is present in the outgroups and all tribes of Ranunculaceae but
616 appears to have been lost several times (22x in *Ranunculus*), including *R. cassubicifolius*.
617 Independent losses and transfers to the nuclear genome are already documented for various
618 angiosperm lineages, making *infA* one of the most mobile plastome genes in plants (Millen et
619 al., 2001). In the case of *R. cassubicifolius*, we were even able to confirm the loss of *infA* in the
620 plastome sequence and its transfer to the nuclear genome (two presumably functional, size-
621 reduced (ca. 150-160 bp) copies exist in chromosomes 1 and 4), a result not previously reported
622 in any other plant. The gene *infA* encodes the translation initiation factor 1 important for
623 ribosome assembly, translation initiation, and thus cell viability (Cummings and Hershey,
624 1994). Transfer to the nucleus may reduce the metabolic burden of the chloroplast, enhance the
625 functional integration and regulation of plastid- and nucleus-encoded proteins, or even escape
626 from the accumulation of deleterious mutations in asexual organellar genomes (Muller's
627 Ratchet; Millen et al., 2001). Studies have shown that *infA* plays a crucial role in adaptation to
628 cold shocks (*Escherichia coli*) or to different temperatures (slipper orchids; Giangrossi et al.,
629 2007; Hu et al., 2022). The transfers of *infA* into the nuclear genome of *R. cassubicifolius* could
630 thus, even more, promote the adaptation of sexual diploids, and particularly of allopolyploid
631 apomicts distributed from Arctic to Mediterranean climates (Karbstein et al., 2020b, 2021,
632 2022; Hodač & Karbstein et al., 2023).
633 In addition, the gene *ycf15* appears to have been lost early in the evolution of
634 Ranunculaceae (e.g., *Hydrastideae*, *Caltheae*), and regained several times in the plastome also
635 within *Ranunculus* (12x in *Ranunculus)* including *R. cassubicifolius*. The gene *ycf15* probably
636 originated at the roots of the angiosperms and has been pseudogenized or lost many times, and
637 compared to other (pseudo)genes like *infA* or *rpl32*, no transfer to the nuclear genome was
638 observed across many angiosperm lineages (Shi et al., 2013). For the genus *Ranunculus*, the
639 complete loss of this gene had been assumed (Shi et al., 2013) but the recent study by Kim et
640 al. (2023) for *R. austro-oreganus* and *R. occidentalis,* and the results presented here contradict
641 this assumption. Although the gene is only 54 bp long in *R. cassubicifolius* and thus apparently
642 degraded (e.g., compare to 198 bp in *R. austro-oreganus*), and seemingly without essential
643 function to photosynthesis and survival of plants (but see its application for species
644 delimitation; Shi et al., 2013; Li et al., 2021), we found two copies of *ycf15* in the IRA and IRB
645 regions with regular ATG start and TGA stop codons, suggesting that it is functional in *R.*
646 *cassubicifolius*. However, the functions of *ycf15* and the co-transcribed large plastid gene *ycf2*
647 are still unclear, although *ycf2* and *accD* influence plastid structure and competitiveness and
648 thus may play a role in uniparental vs. biparental inheritance of plants that is known to shape

649 the ability of hybridization (Drescher et al., 2000; Shi et al., 2013; Liu et al., 2017; Sobanski et
650 al., 2019; Postel and Touzet, 2020). Its role in the hybridization of sexual progenitors that
651 produced the hundreds of allopolyploid *R. auricomus* nothotaxa (Karbstein et al., 2022) needs
652 further investigation.

653 Other genes frequently known to be transferred into the nuclear genome like *accD, yc1*,
654 or *ycf2* remain in the plastomes of Ranunculaceae. Noteworthy, we observed that annotations
655 of (pseudo)genes are sometimes missing (e.g., *clpp*, *psbG*, *rpl22*, or *rpl32*) or renamed (*pbsN*
656 vs. *pbf1*), which complicates plastome feature comparisons and requires additional manual
657 work such as mapping gene sequences from closely related species to the plastome sequence of
658 interest. Therefore, we advocate for a more careful annotation of plastome sequences through
659 cross-species alignments of homologous genes.
660
661
662

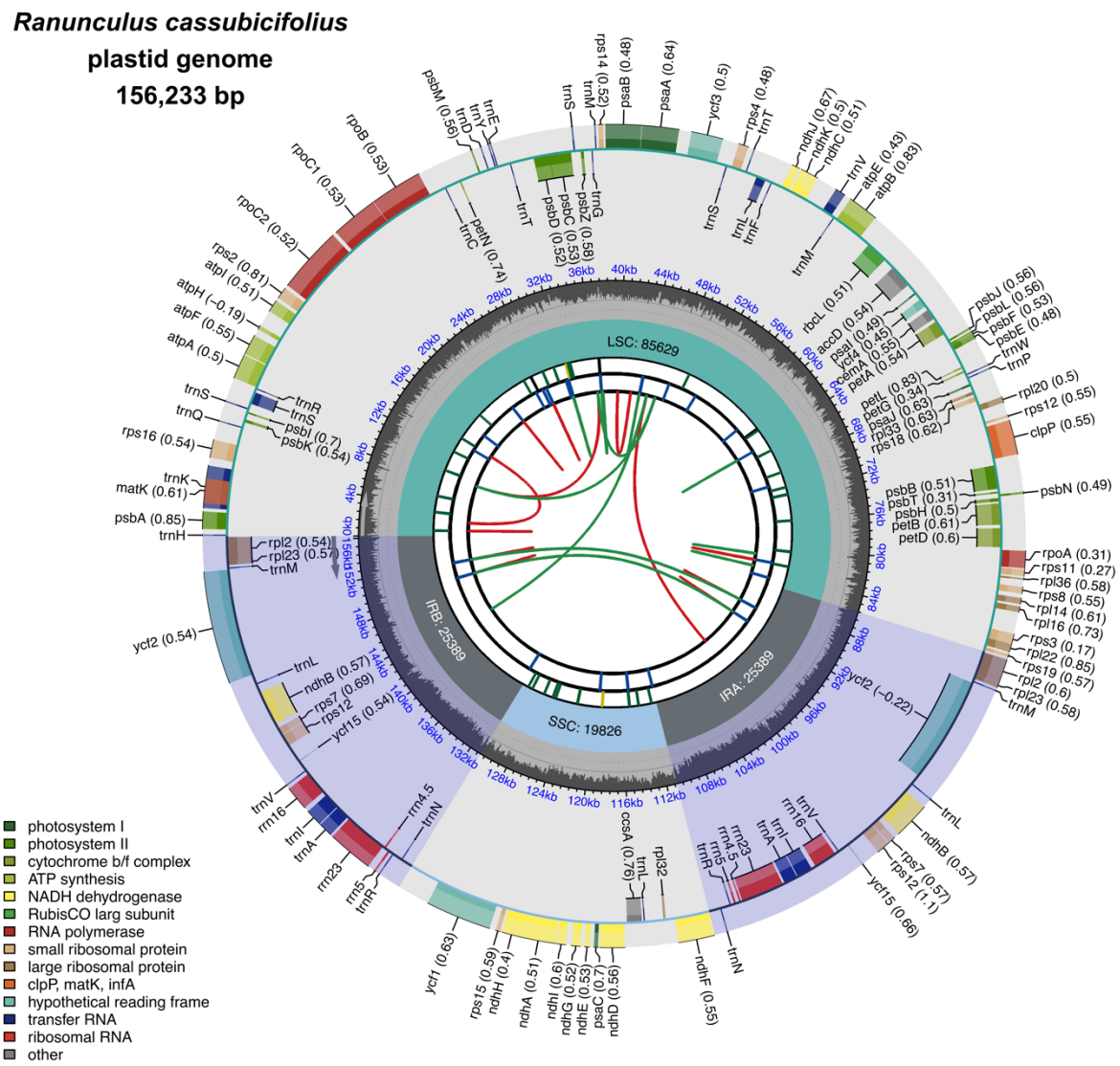


Figure 2. (a) The annotated plastome sequence of *Ranunculus cassubicifolius* (LH040). The plastome assembly is a 156,233 bp-long single circular contig. It contains 85 (78 unique) protein-coding genes, 37 (30 unique) tRNAs, and 8 (4 unique) rRNAs, as shown in the outer circle. Genes and t/rRNAs on the inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively (gene codon usage bias in brackets). The first inner circle represents GC content (%), and 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 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 details.

675 **Mitogenome features and gene evolution within *Ranunculus* and Ranunculaceae.**
676 Based on the Illumina-PacBio hybrid assembly, we assembled a mitochondrial *R.*
677 *cassubicifolius* genome sequence of 1,183,731 bp. It contains 40 (40 unique) protein-coding
678 genes, 77 (18 unique) tRNAs, and 7 (3 unique) rRNAs, resulting in a coding content of 8.0%.
679 This is consistent with observations in angiosperms (Skippington et al., 2015; Dong et al., 2018;
680 Park and Park, 2020; Jiang et al., 2023). In comparison with *Hepatica maxima* (the closest
681 available reference within Ranunculaceae, see Fig. S4a; Zhai et al., 2019), both mitogenomes
682 show similar numbers of unique protein-coding genes (40/39), t-RNAs (18/18), and rRNAs
683 (3/3), but the *R. cassubicifolius* mitogenome is slightly longer (+61,185 bp). In Ranunculaceae,
684 variation in mitogenome size is high, ranging from 0.207 Mbp in *Isopyrum anemonoides* to
685 1.184 Mbp in our study species *R. cassubicifolius*. In angiosperms, mitogenomes can even
686 range from 66 kbp to 11.7 Mbp (Skippington et al., 2015; Putintseva et al., 2020). The
687 mitogenome of *R. cassubicifolius* is therefore relatively small, but one of the largest known for
688 Ranunculaceae.
689 Size differences in the mitogenomes of Ranunculaceae can only be explained to a
690 limited extent by variation in repeat content. In *R. cassubicifolius*, many short (187) and long
691 (168) tandem repeats and a few dispersed repeats were detected, leading to 2.1% repeat content.
692 This is lower compared to *H. maxima* (6.9%) and substantially lower (28x) than the smallest
693 known mitogenome of *I. anemonoides* (0.207 Mbp, 56%). In contrast to plastomes (and animal
694 mitogenomes), plant mitogenomes are recombinationally active, can exist as populations of
695 alternative structures even within individuals, and are considered to be organized in noncircular
696 molecules (Sloan, 2013; Møller et al., 2021). Sequence differences despite similar gene content
697 are thus mainly attributable to sequence recombination and rearrangements in non-coding
698 regions, supported by whole genome alignment results in the present study (Fig. S5a).
699 Significant structural differences were also found between the PacBio and ONT mitogenome
700 assemblies (Fig. S5b), which might stem from the fact that different individuals were sequenced
701 rather than from the use of different sequencing technologies (see Table 1). In addition, despite
702 reads bridging contig boundaries, assembly results may encourage the presence of at least 3
703 chromosomes instead of a single circular molecule in the active cell state (Skippington et al.,
704 2015; Feng et al., 2021).
705 The coalescent-based mitogene phylogeny of Ranunculaceae was well-supported
706 (FBP>85%; Fig. 4b), marking the first comprehensive one for the family and one of the first
707 for plants (e.g., Lin et al., 2022). Like in the plastome tree, low branch support occurs especially
708 within fast radiating, species-rich genera such as *Pulsatilla* (*Anemoneae*). Branching of tribes

709 is also widely consistent with the here presented and previously described plastome phylogenies
710 (Fig. 4a; Wang et al., 2009; Zhai et al., 2019), that is, *Thalictroideae*, *Delphinieae*,
711 *Ranunculeae*, and *Anemoneae*. Regarding mitochondrial gene evolution within Ranunculaceae
712 and particularly *Ranunculus*, we observed that some genes (*rps1*, *rps2*, *rps10*, *rps11*, *rps12*,
713 *rps14*, *rps19*, or *sdh3*) were frequently lost (Fig. 4b; see raw feature table on FigShare). In
714 flowering plants, gene loss and transfer to the nuclear genome have been frequently observed
715 for *rps*, *sdh3,* and *sdh4* (Adams et al., 2000, 2001; Bi et al., 2020).

716 In our study, the mitochondrial genes *rps10* and *rps14* appeared to be lost from
717 *Ranunculeae* to *Anemoneae* and putatively transferred to the nuclear genome (no reference
718 genome sequence available). The *rps* genes encode small ribosomal subunit proteins and are
719 thus essential for mitochondrial ribosome function, as well as vegetative and reproductive plant
720 development (Adams et al., 2000, 2001; Robles and Quesada, 2017). These two genes were
721 shown to be under evolutionary positive selection in certain plant groups (e.g., *rps1*, *rps10,*
722 *rps14* in legumes: Bi et al., 2020; *rps13* in rosids: Liu and Adams, 2008). In addition, *rps1*,
723 *rps2*, *rps11*, and *rps19* were not found in *I. anemonoides*; *rps2* and *rps12* in *Aconitum*
724 *kusnezoffii*; and *rps2*, *rps11*, and *rps14* in *A. carmichaelii*. In plant evolution, the majority of
725 mitochondrial genes have already been transferred to the nuclear genome. This enhances
726 regulatory host control of these genes and integration with nuclear-encoded mitochondrial
727 proteins, and has important consequences for ATP synthesis, metabolism, and abiotic cold or
728 heat and biotic stress responses (Adams et al., 2000; Bonen and Calixte, 2006; Woodson and
729 Chory, 2008; Saha et al., 2017; Fakih et al., 2023).

730 Interestingly, only in *R. cassubicifolius,* all ribosomal genes are present. In contrast to
731 our obligate self-incompatible species of interest (Hörandl, 2008; Karbstein et al., 2020a), all
732 other genera examined, such as *Anemone*, *Pulsatilla*, and partially *Aconitum*, are known to be
733 largely self-compatible (Lindell, 1998; Zhigang et al., 2006; Liao et al., 2009). Selfing (and
734 clonal reproduction), in the absence of deleterious mutation accumulation in the mitogenome,
735 is known to increase the transfer of mitochondrial genes to the nuclear genome and provides an
736 adaptive advantage due to fixation of co-adapted gene combinations (Brandvain and Wade,
737 2009). Therefore, it would be worth investigating whether self-compatible, apomictic
738 polyploids retain as many *rps* genes in their mitogenomes as their sexual diploid *R. auricomus*
739 progenitors, such as *R. cassubicifolius*. However, also the putative transfer of all *rps10* and
740 *rps14* genes to the nuclear genome within the fast radiating, species-rich clade Anemoneae
741 (e.g., only *Anemone* and *Clematis* with >100 and >400 accepted species on gbif.org) needs
742 further investigation. In Ranunculaceae, the lost and putatively transferred *rps* genes are known

743 to be regulated under temperature heat and cold stress (Alafari and Abd-Elgawad, 2021; Fakih
744 et al., 2023), and in the investigated taxa it could play a crucial role in adaptation to a broad
745 range of subtropical to subarctic and lowland to mountainous habitats in central and
746 (south)eastern Asia. Similar to its common forest understory ancestor, *R. cassubicifolius* grows
747 in temperate habitats in humid forests or along stream-sides (Tomasello et al., 2020; Karbstein
748 et al., 2021), facing low selective environmental pressure, which, along with its reproductive
749 traits, may explain the retention of all *rps* genes in its mitogenome sequence. Whether this
750 pattern is consistent across the 600 *Ranunculus* species with worldwide distribution in various
751 environments, needs further investigation.
752 The protein-coding gene *sdh3* was detected neither in *R. cassubicifolius* nor *I.*
753 *anemonoides* (see raw feature table on FigShare). It encodes the subunit 3 of the succinate
754 dehydrogenase and plays a role in electron transfer in complex II of the respiratory chain, and
755 is thus important for energy production (Gebert et al., 2011). Ancient and recent events of *sdh*3
756 loss with and without transfer to the nuclear genome have been observed in several flowering
757 plant lineages (>40x; (Adams et al., 2001; Choi et al., 2019; Huang et al., 2019). Surprisingly,
758 transfer of *sdh3* to the nuclear genome could not be confirmed in *R. cassubicifolius*, and for *I.*
759 *anemonoides* no reference genome sequence is available. Loss of *sdh3* might reflect
760 optimization of cellular efficiency, and that the gene is not essential in the investigated plant
761 lineages.
762
763

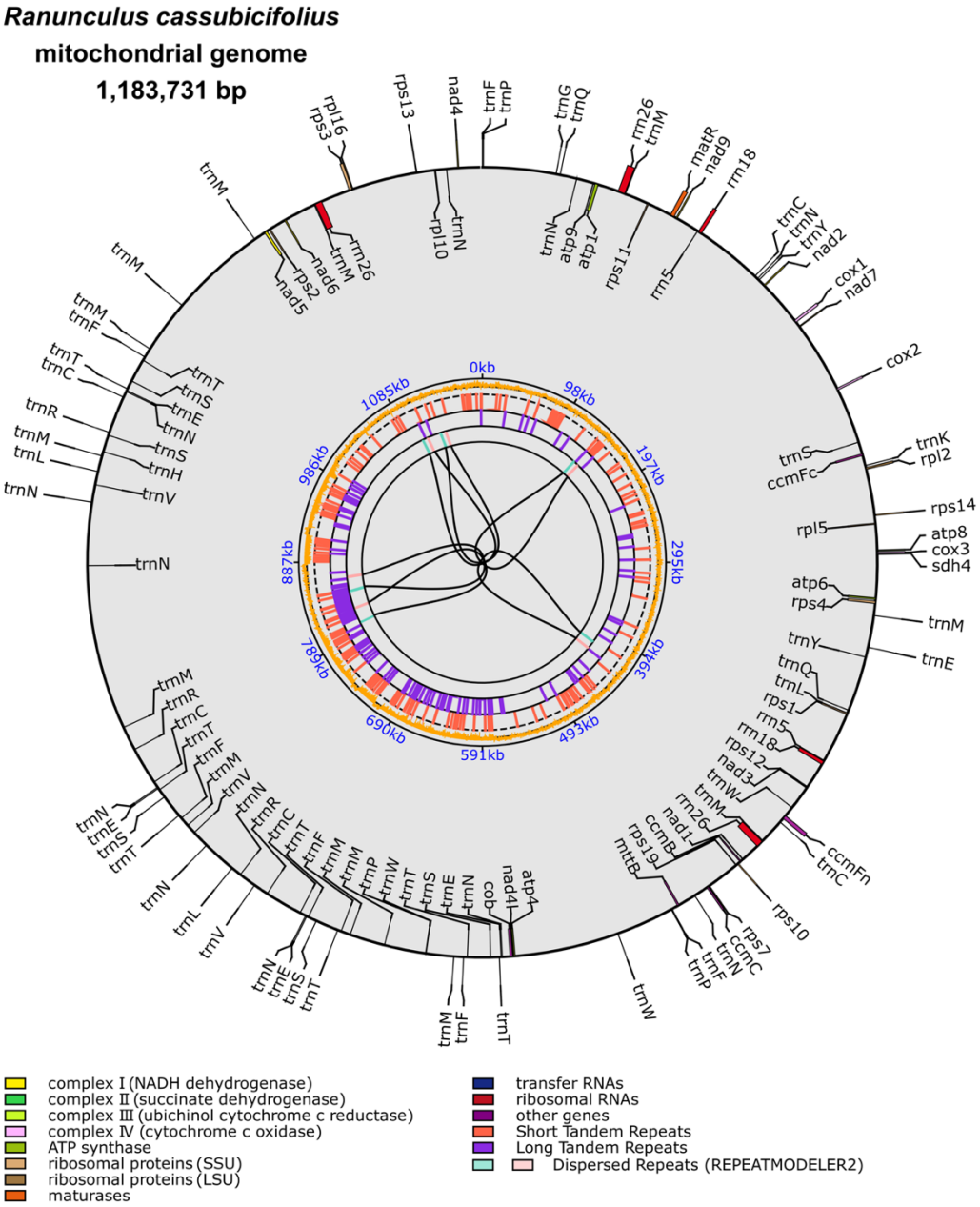
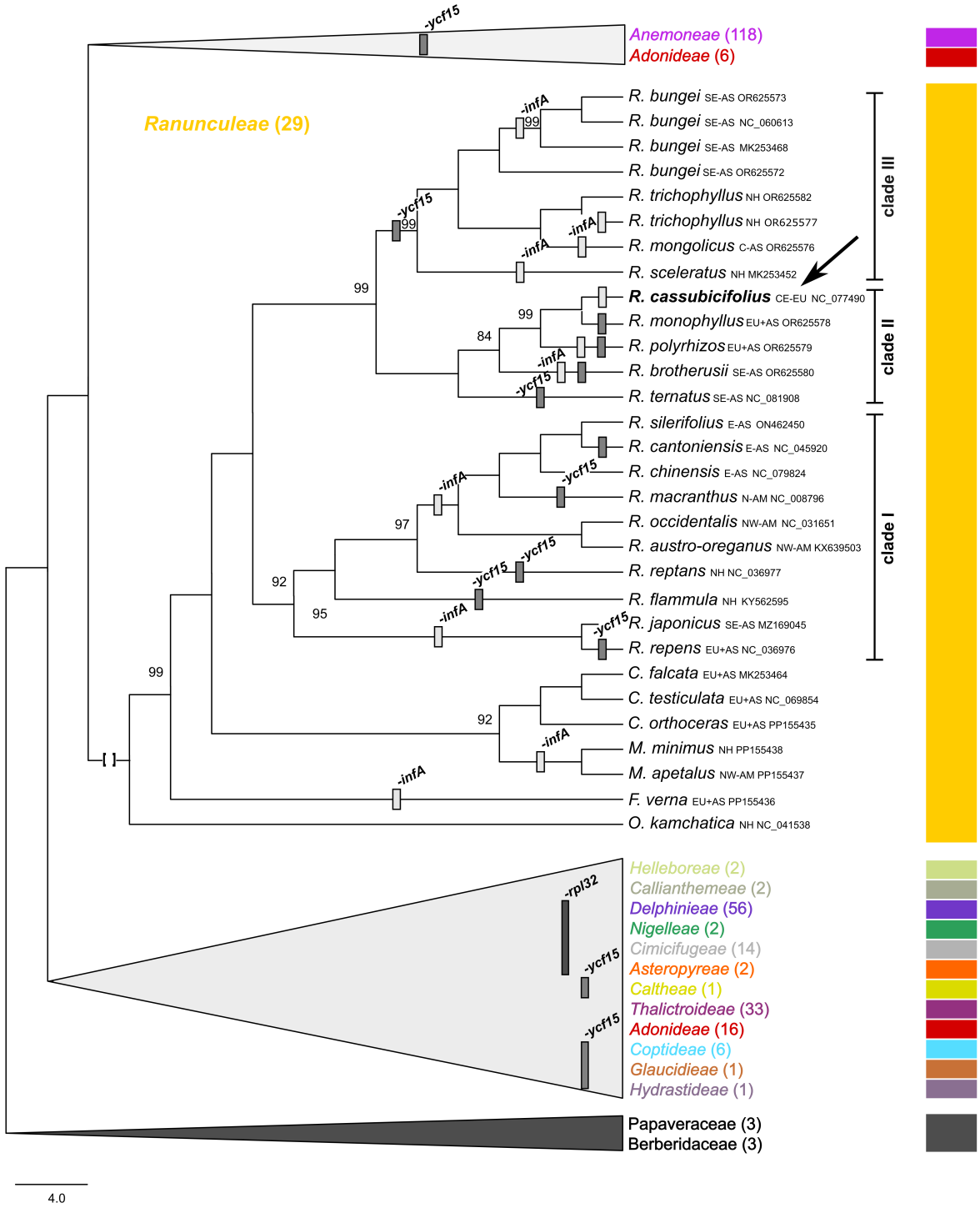


Figure 3. The mitogenome sequence of *Ranunculus cassubicifolius* (LH040). The mitogenome is 1,183,731 bp in size represented in a single, circular contig, and contains 40 (40 unique) protein-coding genes, 77 (unique 18) tRNAs, and 7 (unique 3) rRNAs, as shown in the outer circle. Genes and t/rRNAs on the inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. The first inner circle represents GC content (%). The second inner circle reflects repeat content, i.e., short tandem repeats/microsatellites (1-6 bp per repeat), long tandem repeats (>6 bp per repeat), and dispersed repeats detected by RepeatModeler2 (reports on FigShare). See (Zhang et al., 2024) and http://www.1kmpg.cn/pmgmap for more details.

(a)


Anemoneae (118)
Adonideae (6)
Ranunculeae (29)
R. bungei SE-AS OR625573
R. bungei SE-AS NC_060613
R. bungei SE-AS MK253468
R. bungei SE-AS OR625572
R. trichophyllus NH OR625582
R. trichophyllus NH OR625577
R. mongolicus C-AS OR625576
R. sceleratus NH MK253452
R. cassubicifolius CE-EU NC_077490
R. monophyllus EU+AS OR625578
R. polyrhizos EU+AS OR625579
R. brotherusii SE-AS OR625580
R. ternatus SE-AS NC_081908
R. silerifolius E-AS ON462450
R. cantoniensis E-AS NC_045920
R. chinensis E-AS NC_079824
R. macranthus N-AM NC_008796
R. occidentalis NW-AM NC_031651
R. austro-oreganus NW-AM KX639503
R. reptans NH NC_036977
R. flammula NH KY562595
R. japonicus SE-AS MZ169045
R. repens EU+AS NC_036976
C. falcata EU+AS MK253464
C. testiculata EU+AS NC_069854
C. orthoceras EU+AS PP155435
M. minimus NH PP155438
M. apetalus NW-AM PP155437
F. verna EU+AS PP155436
O. kamchatica NH NC_041538
clade III
clade II
clade I
-ycf15
-infA
-rpl32
99
84
97
92
95
Helleboreae (2)
Callianthemeae (2)
Delphinieae (56)
Nigelleae (2)
Cimicifugeae (14)
Asteropyreae (2)
Caltheae (1)
Thalictroideae (33)
Adonideae (16)
Coptideae (6)
Glaucidieae (1)
Hydrastideae (1)
Papaveraceae (3)
Berberidaceae (3)
4.0

774

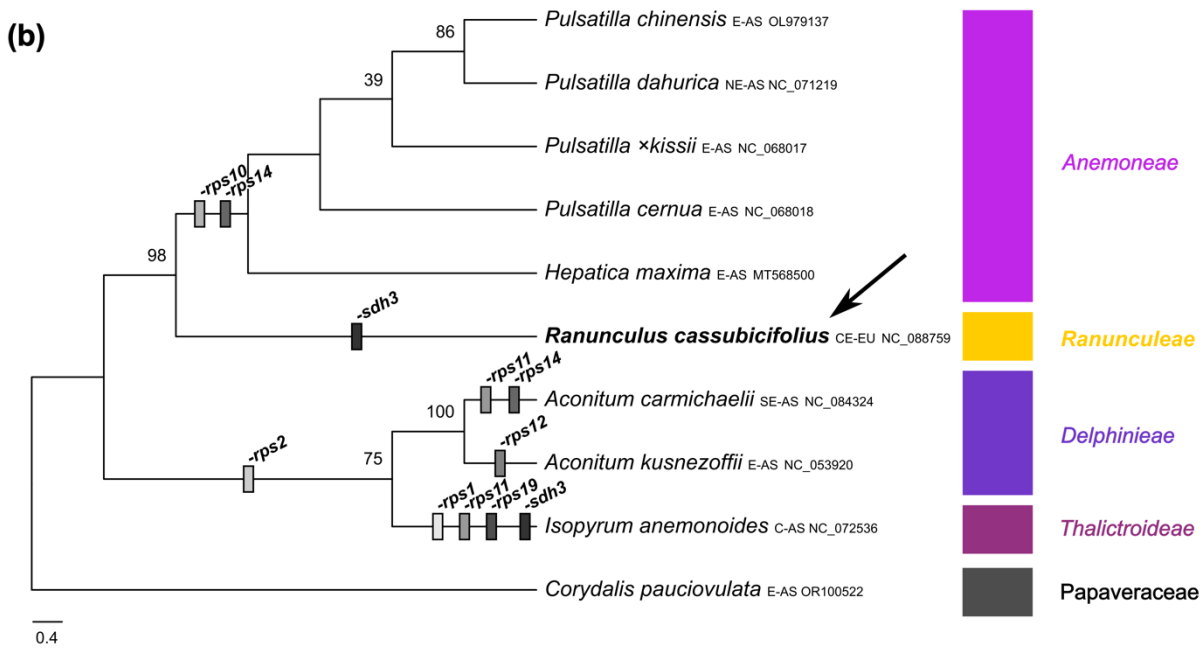


Figure 4. Phylogenetic relationships and gene evolution within Ranunculaceae. (a) Maximum-likelihood phylogeny based on 308 plastome sequences (295 taxa) and the min90 alignment (126 kbp) of the plant family Ranunculaceae. Branches of tribes except for *Ranunculeae* were collapsed, and gained or lost genes are shown in different colored gray bars along the tree. All branches received full (1) transfer bootstrap expectation (TBE) values, unless otherwise shown. See also Quartet Sampling (QS) metrics in Fig. S2, the complete phylogeny in Fig. S4a,b, and more technical details in Text S3. (b) Coalescent-based phylogeny of 10 mitogenome sequences and 42 genes of Ranunculaceae. All branches received full (100%) Felsenstein Bootstrap Proportions (FBP) values, unless otherwise shown. All NCBI species names were checked by gbif.org, and synonyms were replaced by accepted names (synonyms in brackets). The color of clades corresponds to tribe/subfamily names (tribal classification follows Wang et al. (2016) and Zhai et al. (2019). *Ranunculus cassubicifolius* is highlighted in bold with an arrow. Abbreviations for native distribution ranges of taxa: EU = Europe, AS = Asia, AM = America, NH = Northern Hemisphere, C = Central, E = Eastern, N= Northern, S = Southern, and W = Western.

792 **Comparison of nuclear genome assembly strategies.**

793 This study presents a systematic and comprehensive comparison of assembly strategies
794 for large non-model plant genomes based on short- and long-read sequencing data. Related
795 studies have mainly focused on model plants with high coverage and few tested tools (e.g.,
796 Mascher et al., 2021), or non-model plants with small-sized genomes (e.g., Murigneux et al.,
797 2020). Here, the 18 genome assembly strategies yielded substantial differences concerning
798 completeness, continuity, and quality (Table 1, Fig. 5). SUPERNOVA and SPAdes assemblies
799 based on Illumina reads delivered mixed results in terms of completeness (86/17%), with too
800 many (6.92/1.23 million contigs), short (N_{50}: 2/3 kbp) contigs. BUSCO genes were also highly
801 fragmented (F: 36/28%), with low levels of complete single-copy (S: 60/20%) and duplicated
802 (D: 11/2%) genes, and many missing genes (M: 34/10%). This is in line with studies showing
803 highly fragmented, gappy, low-quality assemblies of large genomes when only short Illumina
804 reads are available (Gatter et al., 2021; Rhie et al., 2021; Zimin and Salzberg, 2022; Hotaling
805 et al., 2023). Especially in *R. cassubicifolius* with 86% repetitive regions, short reads with
806 moderate coverage cannot resolve these long and complex patterns.

807 Hybrid assemblies of SPAdes, MaSuRCA, and WENGAN based on Illumina and ONT
808 or PacBio reads performed slightly better in terms of contiguity, with fewer but larger contigs
809 (368-1.18/0.530-6.57 million contigs, with N_{50}: 5-21/5-6 kbp), but not in terms of genome
810 completeness (0.16-13/14-38%) and BUSCO completeness as many genes were often missing
811 (M: 7.7-99.6 / 7.5-90%, F: 0.2-24 / 9-23 %, D: 0-2.4 / 0-2.4%, S: 0.2-66 / 1-67%). SPAdes
812 produced the best assemblies in terms of completeness, contig size, and BUSCO scores, despite
813 being developed and tested for very small microbial genomes (<0.1 Gbp; Prjibelski et al., 2020).
814 MaSuRCA and WENGAN (MuCHSALSA, results not shown) performed worse. Many hybrid
815 assemblers are considered by developers to be particularly useful under low long-read coverage
816 (<10-15x) or large/complex genomes (e.g, Zimin et al., 2017; Di Genova et al., 2021; Gatter et
817 al., 2021). However, this cannot be confirmed in our study. These tools have been tested on
818 model species like *Arabidopsis thaliana* with small genomes (0.135 Gbp) and very low levels
819 of TE (ca. 21%, reviewed in Quesneville, 2020) or in human data with similar genome size (3.2
820 Gbp) to *R. cassubicifolius* but also lower levels of repeat content (ca. 50%, reviewed in Liao et
821 al., 2023), or they need Illumina data at higher coverage (>30-40x) than we provide to build
822 initial robust reads for subsequent long read mapping, which may explain their poorer
823 performance here.

824 Compared to already discussed strategies, the ONT or PacBio-based assemblies
825 conducted with Canu, Flye, and Hifiasm were generally superior in terms of completeness (25-

826 158%), contig number (1,453- 65,232), median contig size (N_{50}: 0.037-20.0 Mbp), and quality
827 as the majority of BUSCO genes were found (M: 0.2-29%, F: 0.3-7.1%, D: 7.9-88%, S: 11-76
828 %). This is also consistent with the recent review by Hotaling et al. (2023), which concludes
829 that long reads allow for more complete and contiguous genome assemblies of higher quality,
830 particularly for large repeat-rich genomes. However, we also observed that PacBio delivered
831 better assemblies than ONT using Canu and Flye. PacBio produced the longer (N_{50}: 617/853
832 vs. 37-103 kbp) and fewer contigs (16,262/12,506 vs. 30,612/64,618). The lower base accuracy
833 of ONT compared to PacBio sequence data (97 vs. 98%), despite a slightly larger median
834 sequence length (23 vs. 18 kbp), can explain this observation (Hotaling et al., 2023). It is
835 important to note that the ONT read N_{50} length of the runs in this study was ca. 20 kbp, while
836 other studies have achieved runs with more than twice this value when using fresh libraries
837 based on different DNA extraction protocols (Hakim et al., 2024; Horz et al., 2024; Nowak et
838 al., 2024). Using the latest ONT flow cell generation and fresh DNA extracts from different kits
839 would probably deliver much longer, higher quality reads and thus improved assemblies (see
840 Text S5 for further discussion). However, the PacBio-based assemblies generated by Canu and
841 Flye were far too large considering the expected haploid genome size (158/153 vs. 25/112%)
842 and characterized by too many duplicated BUSCO genes (70/88 vs. 8/14%). This suggests
843 uncollapsed heterozygous regions in the assembly.
844 However, Hifiasm – both using the full PacBio 25x coverage or the 16x downsampled
845 set – outperformed Canu and Flye by being closest to expected genome size (101/113 vs.
846 158/153%), median contig length (2,890/20,000 vs. 617/853 kbp) and contig number
847 (3449/1453 vs. 16,262/12,506), and runtime (2-3 vs. 7-14 days), but all delivered similar
848 BUSCO gene completeness (99/99% vs. 99/99%). This agrees with recent benchmark analyses
849 on large and complex eukaryotic and synthetic genomes (Cheng et al., 2021; Yu et al., 2024).
850 Interestingly, our polishing strategy slightly improved assembly in terms of missing and
851 fragmented BUSCO scores (mean M: 8.41 vs. 3.67%, and F: 2.51 vs. 1.16%), and contig
852 number of >1 Mbp (mean 21.9 vs. 19.8 Mbp). Percentage of single-copy BUSCO genes
853 improved significantly only for ONT-Canu assemblies (56 vs. 66%). After polishing, genome
854 completeness (mean 118 vs. 117%) and N_{50} contig length (mean 4.13 vs. 4.13 Mbp) remained
855 constant. We generally observed that polishing improved BUSCO statistics more for lower-
856 quality assemblies rather than for good assemblies. Thus, sequencing errors, misassemblies,
857 and duplications can be improved by polishing, but existing contigs are not significantly
858 enlarged. The need for polishing thus depends on the obtained assembly quality, and is
859 particularly attractive when working under low coverage scenarios (<30x). In addition, 25x

860 compared to downsampled 16x coverage PacBio data delivered assemblies of substantially
861 fewer and larger contigs (mean 26,873 vs. 21,268, and 0.832 vs. 1.83 Mbp), more single-copy
862 BUSCO genes (mean 37.6 vs. 31.7%) and less duplicated (mean 40 vs. 46%), fragmented (mean
863 5.23 vs. 5.63%) and missing ones (mean 8.4 vs. 16%), but had similar genome completeneness
864 (mean 105 vs 106%) across all tested hybrid and long-read strategies (Tables 1, S4). Coverage
865 much below 25x is therefore not recommended for either strategy.
866 The PacBio-based Hifiasm assembly polished twice by PacBio and once more by
867 Illumina reads delivered the best results among all tested strategies (100% completeness, 1,319
868 contigs, 20 Mbp N_{50}, and 99% complete BUSCO genes). This represents the best assembly
869 strategy for the large and complex *R. cassubicifolius* nuclear genome with low to moderate
870 sequencing depths, and agrees with findings in smaller non-model plant assemblies (e.g.,
871 Dmitriev et al., 2021).
872
873 **The nuclear genome features.**
874 Based on the best PacBio + Hi-C assembly, we obtained a nuclear *R. cassubicifolius*
875 genome sequence of 2,691,150,139 bp. Approximately 98.4% (2.65 Gbp) of the scaffolds were
876 anchored to form 8 pseudochromosomes based on Hi-C data. This high-quality, chromosome-
877 level assembly has an N_{50} of 355 Mbp, with the longest pseudochromosome spanning 447 Mbp
878 (Fig. 6, Table 3). BUSCO analysis revealed that 94.1% (78.8% single copy, 15.3% duplicated)
879 of the BUSCO embyophyta_odb10 genes were present, indicating the completeness of the
880 assembly. The BUSCO score is slightly lower than before Hi-C scaffolding (99%). The Hi-C
881 scaffolding procedure automatically removes contigs below 1,500 bp, but all contigs prior to
882 scaffolding were far above this threshold, so contig removal does not explain the reduced
883 BUSCO gene score. More likely are misassemblies or incorrect gene annotations based on the
884 original assembly. The retained genome sequences excel the recommended minimum standard
885 of the Earth Biogenome Project in terms of continuity (scaffolds >10 Mbp), functional
886 completeness (>90% BUSCO genes), and chromosome status (>80% contigs assigned to
887 chromosomes; https://www.earthbiogenome.org/report-on-assembly-standards, Lawniczak et
888 al., 2022).
889 The genome sequence contains 35,482 (31,322 unique) protein-coding genes with an
890 average gene length of 3,548 bp and an average coding sequence (CDS) length of 1,169 bp.
891 The gene annotation showed a high level of BUSCO completeness (94.5% using
892 embryophyta_odb10). We predicted 984 tRNA, 99 spliceosomal RNA (snRNA), and 6,448
893 rRNA (3,013 5S rRNA, 1,077 5.8S rRNA, 1,154 18S rRNA, and 1,204 28S rRNAs) gene

models. Further annotation details are available on Github. In comparison with *Aquilegia kansuensis* (293 Mbp; Xie et al., 2020), *Coptis chinensis* (959 Mbp; Chen et al., 2021), and *Coptis teeta* (932 Mbp; Wang et al., 2024) as the only available references within Ranunculaceae, the *R. cassubicifolius* nuclear genome is up to 10-fold larger in size (+2.42 Gbp / +1.75 Gbp /+1.78 Gbp) but has similar numbers of chromosomes (8 vs. 7/9/9). *Ranunculus*, however, has mostly the long, strongly winding chromosome type (R-type) of the family, while *Aquilegia* and *Coptis* have small, simply curved chromosome types (T-type like *Thalictrum*) (Tamura, 1993). Within *Ranunculus*, our target species shows an ancestral karyotype with large chromosomes (4 metacentric and 4 submetacentric, Baltisberger and Hörandl, 2016). Comparing all available genome sequences, *Coptis teeta* shows the most genes (43,979 in Hap1 /46,311 in Hap2), followed by *C. chinensis* (34,109), *R. cassubicifolius* (31,322), and *A. kansuensis* (25,571), suggesting that despite different gene annotation pipelines, gene content and genome size are probably decoupled (Michael, 2014; Li et al., 2017). In angiosperms, haploid genome size varies ca. 2400-fold from 0.063 to 149 Gbp (median = 0.6 / mean = 5.7 Gbp, Dodsworth et al., 2015). The nuclear genome of *R. cassubicifolius* is larger than most flowering plants, and one of the largest assembled yet for Ranunculaceae. Within diploid *Ranunculus* species, haploid genome sizes measured with karyological methods were estimated to range from 1.86 to 8.43 Gbp (Leitch et al., 2019). *R. cassubicifolius* thus represents a typical, middle-sized genome reference for the genus.

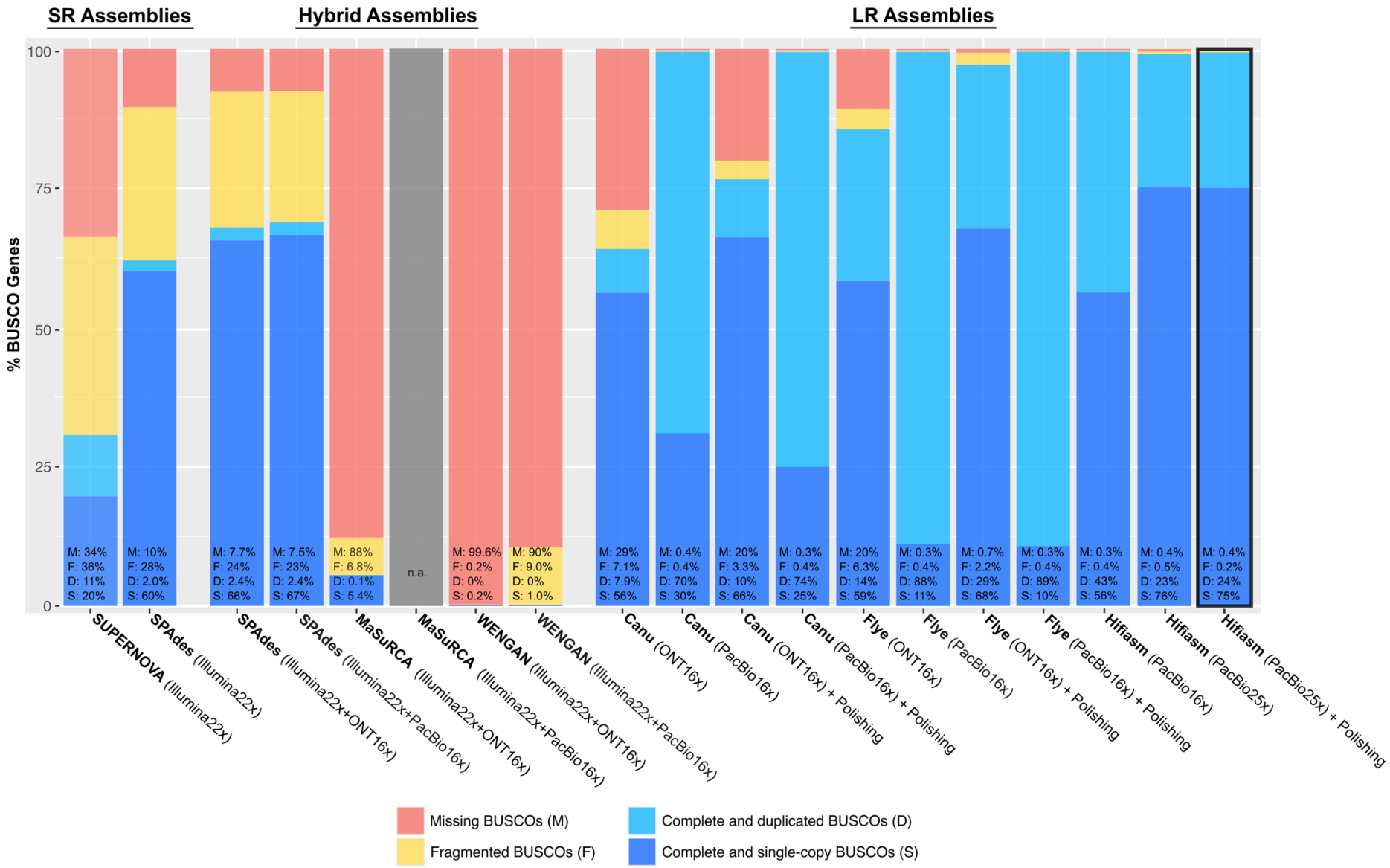
The low proportion of duplicated BUSCO genes (15%) hints at no recent polyploidization event. In ancestral angiosperms, a baseline of 12,000 – 14,000 protein-coding genes is postulated (Sterck et al., 2007). In Liu et al. (2021) and He et al. (2022), ancient whole genome duplication (WGD) events were detected in Beberidaceae and Papaveraceae as close outgroups to Ranunculaceae, and also shared and lineage-specific ones (e.g., *Coptis chinensis* with ca. 39,500 genes, Coptideae) at the base of Ranunculaceae. *Ranunculus* is nested within a far later branching clade (Ranunculeae) within Ranunculaceae, but *R. cassubicifolius* shows only slightly higher gene number. Either no further shared WGD event occurred and gene number remained quite conserved in the several million years of Ranunculaceae evolution, or re-diploidization masked another, more recent WGD event. Interestingly, the *A. kansuensis* (Thalictroideae) genome with only ca. 25,000 genes is placed shortly after *Coptis* and before *Ranunculus* in the Ranunculaceae phylogeny. Whether this re-diploidisation represents a lineage-specific or general trend followed by subsequent WGD events needs to be tested using higher quality genome assemblies and additional analysis tools (e.g., Sun et al., 2022b).

927 Almost 86% of the *R. cassubicifolius* assembly consists of repetitive elements (Table
928 4), among the highest in flowering plants (85% in *Zea mays*; Schnable et al., 2009; Wang et al.,
929 2021). Long terminal repeat (LTR) retrotransposons are most abundant, making up to ca. 77%
930 of the genome. Additionally, DNA transposons, and long interspersed nuclear elements
931 (LINEs) and short interspersed nuclear elements (SINEs) retrotransposons account for 5.7%,
932 0.81%, and 0.2% of the genome, respectively. In angiosperms, genome size correlates well with
933 repetitive element content and particularly with LTR-retrotransposon content but not with
934 polyploidization (Li et al., 2017; Wang et al., 2021). Size differences between Ranunculaceae
935 genomes can be largely explained by variation in LTR-retrotransposon contents. In *R.*
936 *cassubicifolius*, LTR-RTs account for 77.6% of the genome whereas they constitute ca. 36% of
937 the genome of *Aquilegia kansuensis* (Xie et al., 2020). In *Coptis teeta,* LTR-RTs contribute to
938 ca. 36% of genome (Wang et al., 2024), and in *Coptis chinensis*, it is 50% (Chen et al., 2021).
939 Interestingly, LTR retrotransposon bursts have recently been shown to have occurred during
940 the last three million years (Pleistocene) characterized by severe climatic variability, potentially
941 promoting adaptation through novel genomic functions and altered expression patterns (Wang
942 et al., 2021). The progenitor species of the *R. auricomus* complex also speciated by vicariance
943 processes due to range contraction and expansion during cold and warm phases 830-580 kya in
944 Europe (Tomasello et al., 2020). Therefore, climatic variations and the need to adapt to them
945 may be related to the high LTR content in *R. cassubicifolius*.
946 Given the overall high completeness of our assembly (94.1% of BUSCO genes), we
947 think it could represent a very important resource for exploring the genetic background of
948 reproductive mode shifts from sex to apomixis in *R. cassubicifolius* and beyond. The apomixis
949 candidate genes under selection in apomictic hybrids (*ASY1, XRI1*, and *APC1*) recognized by
950 Paetzold et al., 2022) based on RNA-seq data were also detected in the genome of *R.*
951 *cassubicifolius* at pseudochromosomes 2 and 5. However, the apomixis candidate gene *MSP1*
952 was not found and needs further investigation. The annotated genes will be also an important
953 resource for comprehensive analysis of loci under selection in the genus (e.g., via dN/dS ratios;
954 (Pellino et al., 2013), which could give insights into general genome evolution, adaptation, and
955 speciation.
956

957 **Table 1.** QUAST statistics for different genome assembly strategies of *R. cassubicifolius* (LH040) (see also reports on Figshare). SUPERNOVA,
958 and partially SPAdes, MaSuRCA, and WENGAN used Illumina short-reads ('Illumina'). Canu, Flye, and Hifiasm, and partially SPAdes,
959 MaSuRCA, and WENGAN used Oxford Nanopore (ONT) or PacBio long-reads. Coverage is given per read type. Assembly polishing included
960 the following steps: polishing via filtered ONT reads using Racon and Medaka / via filtered PacBio reads using Racon and polishCLR, respectively,
961 and via filtered Illumina reads using POLCA (MaSuRCA pipeline). The best assembly in terms of completeness and contig length are highlighted
962 in bold. N_{50} describes the shortest contig length required to summarize at least half (50%) of the bases of the entire assembly; L_{50} is defined as the
963 minimum number of contigs that encompass half (50%) of the bases of the entire assembly. See Table S3 for comparisons with not downsampled
964 PacBio reads. - means no results produced.

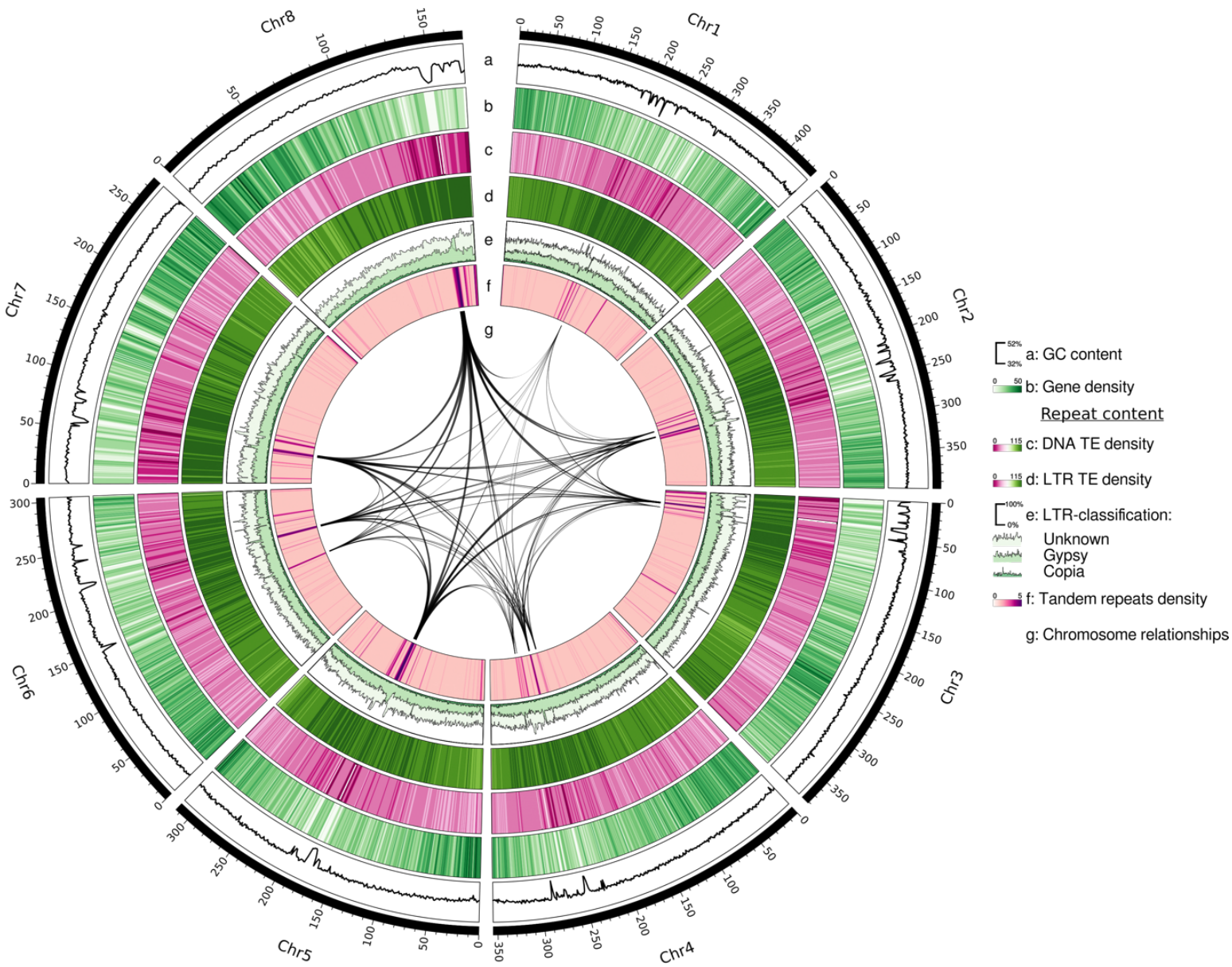
Metric	**SUPERNOVA** (Illumina22x)	**SPAdes** (Illumina22x)	**SPAdes** (Illumina22x + ONT16x/ PacBio16x)	**MaSuRCA** (Illumina22x + ONT16x/ PacBio16x)	**WENGAN** (Illumina22x + ONT16x/ PacBio16x)	**Canu** (ONT16x/ PacBio16x)	**Canu** (ONT16x/ PacBio16x) +3x Polish	**Flye** (ONT16x/ PacBio16x)	**Flye** (ONT16x/ PacBio16x) +3x Polish	**HIFIASM** (PacBio25x/ PacBio16x)	**HIFIASM** (PacBio25x) +3x Polish
Total assembly length, Mbp	2,76 (86%)	0,557 (17%)	0.408/ (13%)/ 0.464 (14%)	0.076/ (2.4%)/ -	0.005 (0.16%)/ 1.21 (38%)	0.80 (25%)/ 5.05 (158%)	0.81 (25%)/ 5.04 (157%)	3.58 (112%)/ 4.90 (153%)	3.45 (108%)/ 4.90 (153%)	3.22 (101%)/ 3.62 (113%)	**3.21 (100%)**
No. contigs	6,920,808	1,231,317	1,184,036/ 529,574	38,650/ -	368/ 6,565,544	30,612/ 16,262	32,959/ 15,711	65,232/ 12,514	54,310/ 12,356	1,453/ 3,449	**1,319**
No. contigs > 1 Mbp	562,716	122,407	110,594/ 108,266	12,799/ -	368/ 5,046	30,612/ 16,262	31,072/ 15,711	64,618/ 12,506	53,572/ 12,356	1,453/ 3,449	**1,319**
Largest contig, Mbp	0.039	0.069	0.096/ 0.112	0.145/ -	0.095/ 0.009	0.540/ 14.3	0.542/ 14.3	0.976/ 6.57	0.974/ 6.57	84/ 32	**84**
GC, %	40	39	39/ 39	40/ -	40 / 39	40/ 42	40/ 42	43/ 42	43/ 42	42/ 42	**42**
N_{50}, Mbp	0.002	0.003	0.005/ 0.005	0.010/ -	0.021/ 0.006	0.037/ 0.617	0.038/ 0.618	0.103/ 0.853	0.104/ 0.852	20.0/ 2.89	**20.0**
L_{50}	259,069	29,900	23,910/ 23,533	1,507/ -	76/ 28,344	6,935/ 2,374	6,937/ 2,366	10,305/ 1728	9,889/ 1727	45/ 350	**45**

965


SR Assemblies
Hybrid Assemblies
LR Assemblies
% BUSCO Genes
100
75
50
25
0
M: 34% F: 36% D: 11% S: 20%
M: 10% F: 28% D: 2.0% S: 60%
M: 7.7% F: 24% D: 2.4% S: 66%
M: 7.5% F: 23% D: 2.4% S: 67%
M: 88% F: 6.8% D: 0.1% S: 5.4%
n.a.
M: 99.6% F: 0.2% D: 0% S: 0.2%
M: 90% F: 9.0% D: 0% S: 1.0%
M: 29% F: 7.1% D: 7.9% S: 56%
M: 0.4% F: 0.4% D: 70% S: 30%
M: 20% F: 3.3% D: 10% S: 66%
M: 0.3% F: 0.4% D: 74% S: 25%
M: 20% F: 6.3% D: 14% S: 59%
M: 0.3% F: 0.4% D: 88% S: 11%
M: 0.7% F: 2.2% D: 29% S: 68%
M: 0.3% F: 0.4% D: 89% S: 10%
M: 0.3% F: 0.4% D: 43% S: 56%
M: 0.4% F: 0.5% D: 23% S: 76%
M: 0.4% F: 0.2% D: 24% S: 75%
SUPERNOVA (Illumina22x)
SPAdes (Illumina22x)
SPAdes (Illumina22x+ONT16x)
SPAdes (Illumina22x+PacBio16x)
MaSuRCA (Illumina22x+ONT16x)
MaSuRCA (Illumina22x+PacBio16x)
WENGAN (Illumina22x+ONT16x)
WENGAN (Illumina22x+PacBio16x)
Canu (ONT16x)
Canu (PacBio16x)
Canu (ONT16x) + Polishing
Canu (PacBio16x) + Polishing
Flye (ONT16x)
Flye (PacBio16x)
Flye (ONT16x) + Polishing
Flye (PacBio16x) + Polishing
Hifiasm (PacBio16x)
Hifiasm (PacBio25x)
Hifiasm (PacBio25x) + Polishing
Missing BUSCOs (M)
Fragmented BUSCOs (F)
Complete and duplicated BUSCOs (D)
Complete and single-copy BUSCOs (S)

966

967 **Figure 5.** BUSCO assessments for different genome assembly strategies of *Ranunculus cassubicifolius* (LH040). SUPERNOVA, and partially
968 SPAdes, MaSuRCA, and WENGAN used Illumina short-reads (SR). Canu, Flye, and HIFIASM, and partially SPAdes, MaSuRCA, and WENGAN
969 used ONT/PacBio long-reads alone (LR) or in combination with short reads (Hybrid). The best assembly in terms of complete BUSCO genes
970 (C+D) is highlighted with a black square (i.e., 'HIFIASM (PacBio) + Polishing'). Analyses are based on 1,614 BUSCO groups searched. See the
971 Materials & Methods section, and Table 1 and Table S4 (Excel Sheet) for more details.
972
973


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
a
b
c
d
e
f
g
a: GC content
b: Gene density
Repeat content
c: DNA TE density
d: LTR TE density
e: LTR-classification:
Unknown
Gypsy
Copia
f: Tandem repeats density
g: Chromosome relationships

974

975 **Figure 6.** Circos plot of the chromosome-scaffolded nuclear genome of *Ranunculus cassubicifolius* (LH040). The genome is 2.69 Gbp in size, has
976 8 pseudochromosomes and 31,322 annotated genes. Illustrated are (from the outer to the inner circle): Chromosome number, chromosome size in
977 Mbp, GC content (a), protein-coding gene density (b), DNA transposon element density (c), Long terminal repeat (LTR) retrotransposon density
978 (d), percentage of unknown, Gypsy, and copia LTR-type Tes (e), tandem repeat density (f), and major interchromosomal synteny (g).
979

980 **Table 3.** Assembly statistics of the annotated plastome, mitogenome, and chromosome-scaffolded nuclear genome of *Ranunculus cassubicifolius*
981 (LH040). Statistics were calculated with Geneious and R. See Table 4 for a detailed repeat content of the nuclear genome, Materials and Methods,
982 and Results section for more details.

Metric	Plastome	Mitogenome	High-Quality Nuclear Genome
Assembly Strategy	Illumina SR (GetOrganelle)	Best strategy: Illumina (filtered by GetOrganelle) assembled with previously mapped HiFi (MaSuRCA)	Best strategy: PacBio (HIFIASM) + 3x Polishing [HiFi (Racon), HiFi (polishCLR), Illumina (POLCA)] + Hi-C chromosome-level scaffolding
Assembled Genome Size	156,233 bp	1,183,731 bp	2,691,150,139 bp
Number of Contigs	1 (circular)	1 (circular)	8 (pseudochromosomes)
N_{50} of Contigs	-	-	355 Mbp
Largest Contig	156,233 bp	1,183,731 bp	447 Mbp
BUSCO Genes (S+D/S)	-	-	94.5% / 61%
Protein Coding Genes	85 (78 unique)	40 (40 unique)	35,482 (31,322 unique)
tRNA Genes	37 (28 unique)	77 (unique 18)	984
rRNA Genes	8 (4 unique)	7 (unique 3)	6,448
GC-Content (%)	37.9%	43.5%	42.1%
Repeat Content (%)	2.15%	2.13%	86.0%
Gene Content (%)	53.2%	7.99%	4.1%

983

984

985 **Table 4.** Repeat content of the chromosome-scaffolded nuclear genome of *Ranunculus*
986 *cassubicifolius* (LH040). Long Interspersed Nuclear Elements (LINEs), Short Interspersed
987 Nuclear Elements (SINEs), Long Terminal Repeats (LTRs), and Terminal Inverted Repeats
988 (TIR). A detailed explanation of repeat classes can be found in Wicker et al. (2007).
989 Annotation reports are deposited in Github.

	Total Sequences	**Total Length**	
	320	2,691,150,139	
Class	**Count**	**bpMasked**	**%masked**
Class I (retrotransposons)			
LINE			
L1	31,842	21,678,526	0.81%
LTR			
Copia	147,394	80,810,425	3.00%
Gypsy	1,105,626	654,391,560	24.32%
unknown	2,075,881	1,353,935,235	50.32%
SINE			
tRNA	309	114,573	0.00%
unknown	347	444,295	0.02%%
Class II (DNA transposons)			
TIR			
CACTA	85,851	32,242,719	1.20%
Mutator	134,007	38,209,632	1.42%
PIF_Harbinger	86,910	30,275,684	1.13%
Tc1_Mariner	32,184	12,038,336	0.45%
hAT	133,119	40,436,252	1.50%
nonTIR			
helitron	123,174	38,735,549	1.44%
rDNA			
45S	191	42,204	0.00%
Repeat Fragment	44,541	11,822,897	0.44%
Satellite DNA	186	83,887	0.00%
total interspersed	4,001,562	2,315,261,774	86.04%
Total	**4,001,562**	**2,315,261,774**	**86.04%**

990

991 **Conclusions**

992 In this study, we assembled the complete plastome, mitogenome, and large nuclear
993 genome sequences of *R. cassubicifolius,* a diploid sexual progenitor species within the
994 *Ranunculus auricomus* complex, using different combinations of Illumina short-reads and ONT
995 and PacBio long-reads under low to medium sequencing depth scenarios. We generated a
996 plastid and the first intricate mitogenome sequences for *Ranunculus* based on Illumina and
997 Illumina + PacBio data, respectively. An updated plastome phylogeny and the first mitogenome
998 phylogeny of Ranunculaceae indicated frequent gene losses (e.g., *infA*, *ycf15*, and *rps* genes),
999 potentially in the context of environmental adaptation and reproductive features. To assemble
1000 the large and complex nuclear genome, we performed 18 assemblies using short-, long-, and
1001 hybrid-read strategies, followed by gap filling and/or polishing with various state-of-the-art
1002 bioinformatic tools. The best result in terms of contiguity and completeness was achieved by
1003 Hifiasm using PacBio data, polished twice with filtered PacBio reads (Racon, polishCLR) and
1004 once with Illumina reads (POLCA), representing the first high-quality, chromosome-scaffolded
1005 genome sequence within *Ranunculus* and the fourth within Ranunculaceae (although the two
1006 currently available genome sequences on NCBI are below 1 Gbp genome size; Xie et al., 2020b;
1007 Chen et al., 2021). The large *R. cassubicifolius* genome with 2.69 Gb size, containing 31,322
1008 annotated genes and 86% repetitive elements, represents a typical medium-sized plant genome
1009 and is probably representative for *Ranunculus*.

1010 The obtained results and in particular the new nuclear genome assembly are useful for
1011 (i) improving read mapping (see Text S4), orthologous gene/locus assembly, and allele phasing
1012 of subgenomic reduced-representation sequencing (RRS) (e.g., RAD-Seq, target enrichment),
1013 transcriptome (RNA-seq), and whole genome (re)sequencing data (e.g., WGS, WGR), (ii) as a
1014 basis for understanding cyto-nuclear compatibility for both plastid-nuclear and mitochondrial-
1015 nuclear interactions; (iii) as diploid reference for the analysis of the more complex allo- and
1016 autopolyploid nuclear genomes within the *Ranunculus,* (iii) as reference for further studies of
1017 genome evolution, loci under selection, adaptation, and speciation, as well as (iv) for improving
1018 functional analyses (e.g., gene expression studies) related to apomixis, adaptation, and many
1019 other traits.

Data Availability. The basic data supporting the findings are available within the manuscript and Supporting Information. Raw sequencing data and the assembled sequences are deposited at 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 ONT, PacBio, and Hi-C reads, filtered ONT and PacBio mitochondrial reads, and filtered mitochondrial Illumina reads (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA831351/); SAMN44350262 the assembled and annotated genome (https://www.ncbi.nlm.nih.gov/datasets/genome/SAMN44350262/); NC_077490 for the assembled and annotated plastome (https://www.ncbi.nlm.nih.gov/search/all/?term=NC_077490), PP657143 for the assembled and annotated mitogenome (https://www.ncbi.nlm.nih.gov/search/all/?term=PP657143). The non-coding genes and transposable elements annotation are available on Github at https://github.com/NancyChoudhary28/Ranunculus-genomics. We deposited software assembly reports, statistical results, and genome alignments on FigShare (https://doi.org/10.6084/m9.figshare.20488908; private link: https://figshare.com/s/b7fba31a6f2444123906), and the used Python scripts on GitHub at https://github.com/KK260/NCBI-Genome-Tools.

Author Contributions

KK and EH designed research; EH collected plant materials; KK, AH, MP, ST, JPB, and BHB performed lab work; KK, NC, XT, BP, and ST analyzed data; KK and EH wrote the paper, with contributions from NC, II, BP, JdV, XT, AH, NS, ST, AP, NW, BHB, JPB, CP, and MP.

Funding

This work was financially supported by the Lindemann-foundation of the University of Göttingen to E.H., and the 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 program (Grant Agreement No. 852725; ERC-StG “TerreStriAL”) and the DFG via SPP 2237 “MAdLand” (422691801, 528076711). AP, JdV, and EH are part of the GRK 2984/1 “EvoReSt” financed by DFG. I.I. was supported by the Spanish Ministry of Science and

Innovation (MCIN/AEI) and the European Social Fund Plus (ESF+) (grant RYC2022-038245-I).

Acknowledgments

We acknowledge the excellent technical assistance of Susanne König, Sylvia Swetik, and Manuela Knauft during PacBio and Hi-C library preparation and DNA sequencing, Jennifer Krüger for laboratory help during ONT DNA sequencing, Silvia Friedrichs and Gabriele Ließmann for garden work, Ladislav Hodač for material sampling and field image supply, and Magnus Nordborg and Robin Burns for bioinformatic advice. This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A).

Supplementary Material

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

References

Adams, K. L., Daley, D. O., Qiu, Y.-L., Whelan, J., and Palmer, J. D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408, 354–357. doi: 10.1038/35042567

Adams, K. L., Rosenblueth, M., Qiu, Y.-L., and Palmer, J. D. (2001). Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics* 158, 1289–1300. doi: 10.1093/genetics/158.3.1289

Alafari, H. A., and Abd-Elgawad, M. E. (2021). Differential expression gene/protein contribute to heat stress-responsive in Tetraena propinqua in Saudi Arabia. *Saudi J Biol Sci* 28, 5017–5027. doi: 10.1016/j.sjbs.2021.05.016

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., and Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour* 20, 892–905. doi: 10.1111/1755-0998.13160

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Astashyn, A., Tvedte, E. S., Sweeney, D., Sapojnikov, V., Bouk, N., Joukov, V., et al. (2024). Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol* 25, 60. doi: 10.1186/s13059-024-03198-7

Baltisberger, M., and Hörandl, E. (2016). Karyotype evolution supports the molecular phylogeny in the genus Ranunculus (Ranunculaceae). *Perspect Plant Ecol Evol Syst* 18, 1–14. doi: 10.1016/j.ppees.2015.11.001

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

Bi, C., Lu, N., Xu, Y., He, C., and Lu, Z. (2020). Characterization and analysis of the mitochondrial genome of common bean (Phaseolus vulgaris) by comparative genomic approaches. *Int J Mol Sci* 21, 3778. doi: 10.3390/ijms21113778

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

Bonen, L., and Calixte, S. (2006). Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. *Mol Biol Evol* 23, 701–712. doi: 10.1093/molbev/msj080

Bradican, J. P., Tomasello, S., Boscutti, F., Karbstein, K., and Hörandl, E. (2023). Phylogenomics of Southern European Taxa in the Ranunculus auricomus Species Complex: The Apple Doesn't Fall Far from the Tree. *Plants* 12, 3664. doi: 10.3390/plants12213664

Bradican, J. P., Tomasello, S., Vollmer, J., and Hörandl, E. (2024). Converging forms: an examination of sub-Arctic, circumarctic, and Central Asian Ranunculus auricomus agg. populations. *Front Plant Sci* 15, 1–10. doi: 10.3389/fpls.2024.1415059

Brandvain, Y., and Wade, M. J. (2009). The functional transfer of genes from the mitochondria to the nucleus: The effects of selection, mutation, population size and rate of self-fertilization. *Genetics* 182, 1129–1139. doi: 10.1534/genetics.108.100024

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

Brůna, T., Lomsadze, A., and Borodovsky, M. (2024). GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res* 34, 757–768. doi: 10.1101/gr.278373.123

1119 Bushnell, B. (2022). BBTools software package. Available at: http://sourceforgenet/
1120 projects/bbmap
1121 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021).
1122 eggNOG-mapper v2: functional annotation, orthology assignments, and domain
1123 prediction at the metagenomic scale. *Mol Biol Evol* 38, 5825–5829. doi:
1124 10.1093/molbev/msab293
1125 Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved
1126 detection and functional classification of transfer RNA genes. *Nucleic Acids Res* 49,
1127 9077–9096. doi: 10.1093/nar/gkab688
1128 Chang, J., Stahlke, A. R., Chudalayandi, S., Rosen, B. D., Childers, A. K., and Severin, A. J.
1129 (2023). polishCLR: A nextflow workflow for polishing PacBio CLR genome assemblies.
1130 *Genome Biol Evol* 15, 1–5. doi: 10.1093/gbe/evad020
1131 Chen, D., Pan, Y., Wang, Y., Cui, Y.-Z., Zhang, Y.-J., Mo, R., et al. (2021). The
1132 chromosome-level reference genome of Coptis chinensis provides insights into genomic
1133 evolution and berberine biosynthesis. *Hortic Res* 8, 121. doi: 10.1038/s41438-021-
1134 00559-2
1135 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de
1136 novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18, 170–175.
1137 doi: 10.1038/s41592-020-01056-5
1138 Choi, I.-S., Schwarz, E. N., Ruhlman, T. A., Khiyami, M. A., Sabir, J. S. M., Hajarah, N. H.,
1139 et al. (2019). Fluctuations in Fabaceae mitochondrial genome size and content are both
1140 ancient and recent. *BMC Plant Biol* 19, 448. doi: 10.1186/s12870-019-2064-8
1141 Chuan, J., Zhou, A., Hale, L. R., He, M., and Li, X. (2021). Atria: an ultra-fast and accurate
1142 trimmer for adapter and quality trimming. *GigaByte* 2021, 1–18. doi:
1143 10.46471/gigabyte.31
1144 Cummings, H. S., and Hershey, J. W. B. (1994). Translation initiation factor IF1 is essential
1145 for cell viability in Escherichia coli. *J Bacteriol* 176, 198–205. doi:
1146 10.1128/jb.176.1.198-205.1994
1147 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021).
1148 Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi:
1149 10.1093/gigascience/giab008
1150 Dann, M., Bellot, S., Schepella, S., Schaefer, H., and Tellier, A. (2017). Mutation rates in
1151 seeds and seed-banking influence substitution rates across the angiosperm phylogeny.
1152 *bioRxiv*, 156398. doi: 10.1101/156398
1153 Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome
1154 alignment with gene gain, loss and rearrangement. *PLoS One* 5, e11147. doi:
1155 10.1371/journal.pone.0011147
1156 Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020).
1157 ModelTest-NG: a new and scalable tool for the selection of DNA and protein
1158 evolutionary models. *Mol Biol Evol* 37, 291–294. doi: 10.1093/molbev/msz189
1159 Dauphin, B., Grant, J. R., Farrar, D. R., and Rothfels, C. J. (2018). Rapid allopolyploid
1160 radiation of moonwort ferns (*Botrychium*; Ophioglossaceae) revealed by PacBio
1161 sequencing of homologous and homeologous nuclear regions. *Mol Phylogenet Evol* 120,
1162 342–353. doi: 10.1016/j.ympev.2017.11.025
1163 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018).
1164 NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34,
1165 2666–2669. doi: 10.1093/bioinformatics/bty149
1166 Del Amparo, R., Branco, C., Arenas, J., Vicens, A., and Arenas, M. (2021). Analysis of
1167 selection in protein-coding sequences accounting for common biases. *Brief Bioinform*
1168 22, 1–10. doi: 10.1093/bib/bbaa431

1169 Di Genova, A., Buena-Atienza, E., Ossowski, S., and Sagot, M.-F. (2021). Efficient hybrid de
1170 novo assembly of human genomes with WENGAN. *Nat Biotechnol* 39, 422–430. doi:
1171 10.1038/s41587-020-00747-w
1172 Dmitriev, A. A., Pushkova, E. N., Novakovskiy, R. O., Beniaminov, A. D., Rozhmina, T. A.,
1173 Zhuchenko, A. A., et al. (2021). Genome sequencing of fiber flax cultivar Atlant using
1174 Oxford Nanopore and Illumina platforms. *Front Genet* 11, 1–7. doi:
1175 10.3389/fgene.2020.590282
1176 Dodsworth, S., Leitch, A. R., and Leitch, I. J. (2015). Genome size diversity in angiosperms
1177 and its influence on gene space. *Curr Opin Genet Dev* 35, 73–78. doi:
1178 10.1016/j.gde.2015.10.006
1179 Dong, S., Zhao, C., Chen, F., Liu, Y., Zhang, S., Wu, H., et al. (2018). The complete
1180 mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly
1181 repetitive with low recombination. *BMC Genomics* 19, 1–12. doi: 10.1186/s12864-018-
1182 4991-4
1183 Drescher, A., Ruf, S., Calsa, T., Carrer, H., and Bock, R. (2000). The two largest chloroplast
1184 genome-encoded open reading frames of higher plants are essential genes. *The Plant*
1185 *Journal* 22, 97–104. doi: 10.1046/j.1365-313x.2000.00722.x
1186 Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al.
1187 (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-
1188 length scaffolds. *Science (1979)* 356, 92–95. doi: 10.1126/science.aal3327
1189 Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., et al.
1190 (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C
1191 experiments. *Cell Syst* 3, 95–98. doi: 10.1016/j.cels.2016.07.002
1192 Eaton, D. A. R., and Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq
1193 datasets. *Bioinformatics* 36, 2592–2594. doi: 10.1093/bioinformatics/btz966
1194 Emadzade, K., Lehnebach, C., Lockhart, P., and Hörandl, E. (2010). A molecular phylogeny,
1195 morphology and classification of genera of Ranunculeae (Ranunculaceae). *Taxon* 59,
1196 809–828. doi: 10.1002/tax.593011
1197 Ennos, R., French, G., and Hollingsworth, P. (2005). Conserving taxonomic complexity.
1198 *Trends Ecol Evol* 20, 164–168. doi: 10.1016/j.tree.2005.01.012
1199 Fakih, Z., Plourde, M. B., and Germain, H. (2023). Differential participation of plant
1200 ribosomal proteins from the small ribosomal subunit in protein translation under stress.
1201 *Biomolecules* 13, 1160. doi: 10.3390/biom13071160
1202 Feng, Y., Xiang, X., Akhter, D., Pan, R., Fu, Z., and Jin, X. (2021). Mitochondrial
1203 phylogenomics of Fagales provides insights intoplant mitogenomemosaic evolution.
1204 *Front Plant Sci* 12, 1–13. doi: 10.3389/fpls.2021.762195
1205 Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., and Förster, F.
1206 (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biol*
1207 21, 254. doi: 10.1186/s13059-020-02153-6
1208 Frith, M. C. (2011). A new repeat-masking method enables specific detection of homologous
1209 sequences. *Nucleic Acids Res* 39, e23–e23. doi: 10.1093/nar/gkq1212
1210 Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., et al. (2024).
1211 BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence
1212 with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res* 34, 769–777. doi:
1213 10.1101/gr.278090.123
1214 Gatter, T., von Löhneysen, S., Fallmann, J., Drozdova, P., Hartmann, T., and Stadler, P. F.
1215 (2021). LazyB: fast and cheap genome assembly. *Algorithms for Molecular Biology* 16,
1216 8. doi: 10.1186/s13015-021-00186-5
1217 Gebert, N., Gebert, M., Oeljeklaus, S., von der Malsburg, K., Stroud, D. A., Kulawiak, B., et
1218 al. (2011). Dual function of Sdh3 in the respiratory chain and TIM22 protein Translocase

1219 of the mitochondrial inner membrane. *Mol Cell* 44, 811–818. doi:
1220 10.1016/j.molcel.2011.09.025
1221 Giangrossi, M., Brandi, A., Giuliodori, A. M., Gualerzi, C. O., and Pon, C. L. (2007). Cold-
1222 shock-induced de novo transcription and translation of infA and role of IF1 during cold
1223 adaptation. *Mol Microbiol* 64, 807–821. doi: 10.1111/j.1365-2958.2007.05699.x
1224 Grabherr, M. G. ;, Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al.
1225 (2011). Full-length transcriptome assembly from RNA-Seq data without a reference
1226 genome. *Nat Biotechnol* 29, 644–652. doi: 10.1038/nbt.1883
1227 Guo, X., Fang, D., Sahu, S. K., Yang, S., Guang, X., Folk, R., et al. (2021). Chloranthus
1228 genome provides insights into the early diversification of angiosperms. *Nat Commun* 12,
1229 6930. doi: 10.1038/s41467-021-26922-4
1230 Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment
1231 tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi:
1232 10.1093/bioinformatics/btt086
1233 Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008).
1234 Automated eukaryotic gene structure annotation using EVidenceModeler and the
1235 program to assemble spliced alignments. *Genome Biol* 9, R7. doi: 10.1186/gb-2008-9-1-
1236 r7
1237 Hakim, S. E., Choudhary, N., Malhotra, K., Peng, J., Arafa, A., Bültemeier, A., et al. (2024).
1238 Phylogenomics and metabolic engineering reveal a conserved gene cluster in Solanaceae
1239 plants for withanolide biosynthesis. *bioRxiv*, 1–44. doi: 10.1101/2024.09.27.614867
1240 Havird, J. C., Hall, M. D., and Dowling, D. K. (2015). The evolution of sex: A new
1241 hypothesis based on mitochondrial mutational erosion. *BioEssays* 37, 951–958. doi:
1242 10.1002/bies.201500057
1243 He, J., Lyu, R., Luo, Y., Xiao, J., Xie, L., Wen, J., et al. (2022). A phylotranscriptome study
1244 using silica gel-dried leaf tissues produces an updated robust phylogeny of
1245 Ranunculaceae. *Mol Phylogenet Evol* 174, 107545. doi: 10.1016/j.ympev.2022.107545
1246 Hodač, L. & Karbstein, K., Tomasello, S., Wäldchen, J., Bradican, J. P., and Hörandl, E.
1247 (2023). Geometric morphometric versus genomic patterns in a large polyploid plant
1248 species complex. *Biology (Basel)* 12, 418. doi: 10.3390/biology12030418
1249 Hodač, L., Barke, B. H., and Hörandl, E. (2018). Mendelian segregation of leaf phenotypes in
1250 experimental F2 hybrids elucidates origin of morphological diversity of the apomictic
1251 *Ranunculus auricomus* complex. *Taxon* 67, 1082–1092. doi: 10.12705/676.6
1252 Hoencamp, C., Dudchenko, O., Elbatsh, A. M. O., Brahmachari, S., Raaijmakers, J. A., van
1253 Schaik, T., et al. (2021). 3D genomics across the tree of life reveals condensin II as a
1254 determinant of architecture type. *Science (1979)* 372, 984–989. doi:
1255 10.1126/science.abe2218
1256 Hörandl, E. (2008). Evolutionary Implications of self-Compatibility and reproductive fitness
1257 in the apomictic Ranunculus auricomus polyploid complex (Ranunculaceae). *Int J Plant*
1258 *Sci* 169, 1219–1228. doi: 10.1086/591980
1259 Hörandl, E. (2022). Novel approaches for species concepts and delimitation in polyploids and
1260 hybrids. *Plants* 11, 204. doi: 10.3390/plants11020204
1261 Hörandl, E. (2024). Apomixis and the paradox of sex in plants. *Ann Bot*, 1–18. doi:
1262 10.1093/aob/mcae044
1263 Hörandl, E., and Emadzade, K. (2012). Evolutionary classification : A case study on the
1264 diverse plant genus Ranunculus L. (Ranunculaceae). *Perspect Plant Ecol Evol Syst* 14,
1265 310–324. doi: 10.1016/j.ppees. 2012.04.001
1266 Hörandl, E., and Greilhuber, J. (2002). Diploid and autotetraploid sexuals and their
1267 relationships to apomicts in the *Ranunculus cassubicus* group: insights from DNA

content and isozyme variation. *Plant Systematics and Evolution* 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., Hojsgaard, D., Caperta, A., Conceição, S. I. R., Róis, A. S., Dickinson, T., et al. (2024). Apomixis in systematics, evolution and phylogenetics of angiosperms: Current developments and prospects. *CRC Crit Rev Plant Sci* 0, 1–43. doi: 10.1080/07352689.2024.2396259

Hörandl, E., Paun, O., Johansson, J. T., Lehnebach, C., Armstrong, T., Chen, L., et al. (2005). Phylogenetic relationships and evolutionary traits in Ranunculus s.l. (Ranunculaceae) inferred from ITS sequence analysis. *Mol Phylogenet Evol* 36, 305–327. doi: 10.1016/j.ympev.2005.02.009

Hörandl, E., and Raab-Straube, E. (2015). Ranunculaceae. *Euro+Med Plantbase - the information resource for Euro-Mediterranean plant diversity.* Available at: http://ww2.bgbm.org/EuroPlusMed

Horz, J. M., Wolff, K., Friedhoff, R., and Pucker, B. (2024). Genome sequence of the medicinal and ornamental plant Digitalis purpurea reveals the molecular basis of flower color variation. *bioRxiv*, 1–27.

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

Hu, C., Jiao, Z., Deng, X., Tu, X., Lu, A., Xie, C., et al. (2022). The ecological adaptation of the unparalleled plastome character evolution in slipper orchids. *Front Plant Sci* 13, 1–13. doi: 10.3389/fpls.2022.1075098

Huang, S., Braun, H., Gawryluk, R. M. R., and Millar, A. H. (2019). Mitochondrial complex <scp>II</scp> of plants: subunit composition, assembly, and function in respiration and signaling. *The Plant Journal* 98, 405–417. doi: 10.1111/tpj.14227

Jayakodi, M., Golicz, A. A., Kreplak, J., Fechete, L. I., Angra, D., Bednář, P., et al. (2023). The giant diploid faba genome unlocks variation in a global protein crop. *Nature* 615, 652–659. doi: 10.1038/s41586-023-05791-5

Ji, J., Luo, Y., Pei, L., Li, M., Xiao, J., Li, W., et al. (2023). Complete plastid genomes of nine species of Ranunculeae (Ranunculaceae) and their phylogenetic inferences. *Genes (Basel)* 14, 2140. doi: 10.3390/genes14122140

Jiang, M., Ni, Y., Li, J., and Liu, C. (2023). Characterisation of the complete mitochondrial genome of Taraxacum mongolicum revealed five repeat-mediated recombinations. *Plant Cell Rep* 42, 775–789. doi: 10.1007/s00299-023-02994-y

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

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

Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., et al. (2024). Species delimitation 4.0: integrative taxonomy meets artificial intelligence. *Trends Ecol Evol* 39, 771–784. doi: 10.1016/j.tree.2023.11.002

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 Phytologist* 235, 2081–2098. doi: 10.1111/nph.18284

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30, 772–780. doi: 10.1093/molbev/mst010

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

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* 19, 189. doi: 10.1186/s12859-018-2203-5

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res* 44, e89–e89. doi: 10.1093/nar/gkw092

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907–915. doi: 10.1038/s41587-019-0201-4

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

Konupková, A., Tomečková, L., Záhonová, K., Oborník, M., and Füssy, Z. (2024). "Easier lost than found? What we know about plastid genome reduction," in *Endosymbiotic Organelle Acquisition*, (Cham: Springer International Publishing), 147–181. doi: 10.1007/978-3-031-57446-7_5

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

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V, et al. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res* 51, D445–D451. doi: 10.1093/nar/gkac998

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. *Proceedings of the National Academy of Sciences* 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 Bennet, M. (2019). Plant DNA C-values database (release 7.1, Apr 2019). Available at: http://data.kew.org/cvalues/

Lemoine, F., and Gascuel, O. (2021). Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genom Bioinform* 3, 1–5. doi: 10.1093/nargab/lqab075

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

Li, S.-F., Su, T., Cheng, G.-Q., Wang, B.-X., Li, X., Deng, C.-L., et al. (2017). Chromosome evolution in connection with repetitive sequences and epigenetics in plants. *Genes (Basel)* 8, 290. doi: 10.3390/genes8100290

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

Liao, W.-J., Hu, Y., Zhu, B.-R., Zhao, X.-Q., Zeng, Y.-F., and Zhang, D.-Y. (2009). Female reproductive success decreases with display size in monkshood, Aconitum kusnezoffii (Ranunculaceae). *Ann Bot* 104, 1405–1412. doi: 10.1093/aob/mcp237

Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., et al. (2023). Repetitive DNA sequence detection and its role in the human genome. *Commun Biol* 6, 954. doi: 10.1038/s42003-023-05322-y

Lin, Q., Braukmann, T. W. A., Soto Gomez, M., Mayer, J. L. S., Pinheiro, F., Merckx, V. S. F. T., et al. (2022). Mitochondrial genomic data are effective at placing mycoheterotrophic lineages in plant phylogeny. *New Phytologist* 236, 1908–1921. doi: 10.1111/nph.18335

Lindell, T. (1998). Breeding systems and crossing experiments in Anemone patens and in the Anemome pulsatilla group (Ranunculaceae). *Nord J Bot* 18, 549–561. doi: 10.1111/j.1756-1051.1998.tb01535.x

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

1414 Liu, S.-L., and Adams, K. (2008). Molecular adaptation and expression evolution following
1415 duplication of genes for organellar ribosomal protein S13 in rosids. *BMC Evol Biol* 8, 25.
1416 doi: 10.1186/1471-2148-8-25
1417 Liu, Y., Wang, B., Shu, S., Li, Z., Song, C., Liu, D., et al. (2021). Analysis of the Coptis
1418 chinensis genome reveals the diversification of protoberberine-type alkaloids. *Nat*
1419 *Commun* 12, 3276. doi: 10.1038/s41467-021-23611-0
1420 Long, J., He, W.-C., Peng, H.-W., Erst, A. S., Wang, W., and Xiang, K.-L. (2024).
1421 Comparative plastome analysis of the sister genera Ceratocephala and Myosurus
1422 (Ranunculaceae) reveals signals of adaptive evolution to arid and aquatic environments.
1423 *BMC Plant Biol* 24, 202. doi: 10.1186/s12870-024-04891-2
1424 Lunkova, N. F., and Ivanov, V. B. (2022). Features of the genome correlated with the
1425 annuality of plant. *Russian Journal of Plant Physiology* 69, 87. doi:
1426 10.1134/S1021443722050132
1427 Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two
1428 open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi:
1429 10.1093/bioinformatics/bth315
1430 Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol Evol* 20, 229–237.
1431 doi: 10.1016/j.tree.2005.02.010
1432 Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO
1433 update: Novel and streamlined workflows along with broader and deeper phylogenetic
1434 coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38,
1435 4647–4654. doi: 10.1093/molbev/msab199
1436 Marks, P., Garcia, S., Martinez, A., Belhocine, K., Bernate, J., Bharadwaj, R., et al. (2017).
1437 Resolving the full spectrum of human genome variation using linked-reads. *bioRxiv*, 1–
1438 43. doi: 10.1101/230946
1439 Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al.
1440 (2017). A chromosome conformation capture ordered sequence of the barley genome.
1441 *Nature* 544, 427–433. doi: 10.1038/nature22043
1442 Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C. S., et al. (2021). Long-read
1443 sequence assembly: a technical evaluation in barley. *Plant Cell* 33, 1888–1906. doi:
1444 10.1093/plcell/koab077
1445 Michael, T. P. (2014). Plant genome size variation: bloating and purging DNA. *Brief Funct*
1446 *Genomics* 13, 308–317. doi: 10.1093/bfgp/elu005
1447 Millen, R. S., Olmstead, R. G., Adams, K. L., Palmer, J. D., Lao, N. T., Heggie, L., et al.
1448 (2001). Many parallel losses of infA from chloroplast DNA during angiosperm evolution
1449 with multiple independent transfers to the nucleus. *Plant Cell* 13, 645–658. doi:
1450 10.1105/tpc.13.3.645
1451 Mokhtar, M. M., Abd-Elhalim, H. M., and El Allali, A. (2023). A large-scale assessment of
1452 the quality of plant genome assemblies using the LTR assembly index. *AoB Plants* 15,
1453 1–8. doi: 10.1093/aobpla/plad015
1454 Møller, I. M., Rasmusson, A. G., and Van Aken, O. (2021). Plant mitochondria – past, present
1455 and future. *The Plant Journal* 108, 912–959. doi: 10.1111/tpj.15495
1456 Murigneux, V., Rai, S. K., Furtado, A., Bruxner, T. J. C., Tian, W., Harliwong, I., et al.
1457 (2020). Comparison of long-read methods for sequencing and assembly of a plant
1458 genome. *Gigascience* 9, 1–11. doi: 10.1093/gigascience/giaa146
1459 Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology
1460 searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
1461 Nowak, M. S., Harder, B., Meckoni, S. N., Friedhoff, R., Wolff, K., and Pucker, B. (2024).
1462 Genome sequence and RNA-seq analysis reveal genetic basis of flower coloration in the
1463 giant water lily Victoria cruziana. *bioRxiv*, 1–23. doi: 10.1101/2024.06.15.599162

Ontiveros-Palacios, N., Cooke, E., Nawrocki, E. P., Triebel, S., Marz, M., Rivas, E., et al. (2024). Rfam 15: RNA families database in 2025. *Nucleic Acids Res*, 1–10. doi: 10.1093/nar/gkae1023

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* 176, 1410–1422. doi: 10.1104/pp.17.01310

Ou, S., and Jiang, N. (2019). LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA* 10, 48. doi: 10.1186/s13100-019-0193-0

Padmarasu, S., Himmelbach, A., Mascher, M., and Stein, N. (2019). In situ Hi-C for plants: An improved method to detect long-range chromatin interactions. *Methods in Molecular Biology* 1933, 441–472. doi: 10.1007/978-1-4939-9045-0_28

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 1.8.15) [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 Phytologist* 171, 223–236. doi: 10.1111/j.1469-8137.2006.01738.x

Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am J Bot* 105, 385–403. doi: 10.1002/ajb2.1016

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

Peška, V., Mandáková, T., Ihradská, V., and Fajkus, J. (2019). Comparative dissection of three giant genomes: Allium cepa, Allium sativum, and Allium ursinum. *Int J Mol Sci* 20, 733. doi: 10.3390/ijms20030733

Petersen, G., Zervas, A., Pedersen, H. Æ., and Seberg, O. (2018). Genome reports: Contractedgenes and dwarfed plastome in mycoheterotrophic Sciaphila thaidanica (Triuridaceae, Pandanales). *Genome Biol Evol* 10, 976–981. doi: 10.1093/gbe/evy064

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 Bioinformatics* 70, 1–29. doi: 10.1002/cpbi.102

Pucker, B. (2024). Functional annotation - How to tackle the bottleneck in plant genomics. *Preprints.org*, 1–15. doi: 10.20944/preprints202402.0645.v1

1512 Pucker, B., Irisarri, I., de Vries, J., and Xu, B. (2022). Plant genome sequence assembly in the
1513 era of long reads: Progress, challenges and future directions. *Quantitative Plant Biology*
1514 3, e5. doi: 10.1017/qpb.2021.18
1515 Putintseva, Y. A., Bondar, E. I., Simonov, E. P., Sharov, V. V., Oreshkova, N. V., Kuzmin, D.
1516 A., et al. (2020). Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled
1517 using both short and long nucleotide sequence reads is currently the largest known
1518 mitogenome. *BMC Genomics* 21, 654. doi: 10.1186/s12864-020-07061-4
1519 Qu, Z., Wang, W., and Adelson, D. L. (2023). Chromosomal level genome assembly of
1520 medicinal plant Sophora flavescens. *Sci Data* 10, 572. doi: 10.1038/s41597-023-02490-8
1521 Quesneville, H. (2020). Twenty years of transposable element analysis in the Arabidopsis
1522 thaliana genome. *Mob DNA* 11, 28. doi: 10.1186/s13100-020-00223-x
1523 Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing
1524 genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
1525 Rambaut, A. (2014). Figtree, a graphical viewer of phylogenetic trees. Available at:
1526 http://tree.bio.ed.ac.uk/software/figtree
1527 Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other
1528 things). *Methods Ecol Evol* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x
1529 Revell, L. J. (2024). phytools 2.0: an updated R ecosystem for phylogenetic comparative
1530 methods (and other things). *PeerJ* 12, e16505. doi: 10.7717/peerj.16505
1531 Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021).
1532 Towards complete and error-free genome assemblies of all vertebrate species. *Nature*
1533 592, 737–746. doi: 10.1038/s41586-021-03451-0
1534 Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al.
1535 (2015). The Chromosome Counts Database (CCDB) - a community resource of plant
1536 chromosome numbers. *New Phytologist* 206, 19–26. doi: 10.1111/nph.13191
1537 Robles, P., and Quesada, V. (2017). Emerging roles of mitochondrial ribosomal proteins in
1538 plant development. *Int J Mol Sci* 18, 2595. doi: 10.3390/ijms18122595
1539 Saha, A., Das, S., Moin, M., Dutta, M., Bakshi, A., Madhav, M. S., et al. (2017). Genome-
1540 wide identification and comprehensive expression profiling of ribosomal protein small
1541 subunit (RPS) genes and their comparative analysis with the large subunit (RPL) genes
1542 in rice. *Front Plant Sci* 8, 1–21. doi: 10.3389/fpls.2017.01553
1543 Schmidt, A. (2020). Controlling apomixis: Shared features and distinct characteristics of gene
1544 regulation. *Genes (Basel)* 11, 329. doi: 10.3390/genes11030329
1545 Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The
1546 B73 maize genome: complexity, diversity, and dynamics. *Science (1979)* 326, 1112–
1547 1115. doi: 10.1126/science.1178534
1548 Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for
1549 fasta/q file manipulation. *PLoS One* 11, e0163962. doi: 10.1371/journal.pone.0163962
1550 Shi, C., Liu, Y., Huang, H., Xia, E.-H., Zhang, H.-B., and Gao, L.-Z. (2013). Contradiction
1551 between plastid gene transcription and function due to complex posttranscriptional
1552 splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One*
1553 8, e59620. doi: 10.1371/journal.pone.0059620
1554 Shi, J., and Liang, C. (2019). Generic Repeat Finder: A high-sensitivity tool for genome-wide
1555 de novo repeat detection. *Plant Physiol* 180, 1803–1815. doi: 10.1104/pp.19.00386
1556 Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an
1557 integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* 47, W65–W73.
1558 doi: 10.1093/nar/gkz345
1559 Sim, S. B., Corpuz, R. L., Simmonds, T. J., and Geib, S. M. (2022). HiFiAdapterFilt, a
1560 memory efficient read processing pipeline, prevents occurrence of adapter sequence in

PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics* 23, 157. doi: 10.1186/s12864-022-08375-1

Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (2015). Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. *Proceedings of the National Academy of Sciences* 112, E3515–E3524. doi: 10.1073/pnas.1504491112

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 Phytologist* 200, 978–985. doi: 10.1111/nph.12395

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

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62. doi: 10.1186/1471-2105-7-62

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

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 35, 1026–1028. doi: 10.1038/nbt.3988

Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P., and Vandepeer, Y. (2007). How many genes are there in plants (… and why are they there)? *Curr Opin Plant Biol* 10, 199–203. doi: 10.1016/j.pbi.2007.01.004

Su, W., Gu, X., and Peterson, T. (2019). TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Mol Plant* 12, 447–460. doi: 10.1016/j.molp.2019.02.008

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol Plant* 15, 1841–1851. doi: 10.1016/j.molp.2022.10.018

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

Tamura, M. (1993). "Ranunculaceae," in *The Families and Genera of Vascular Plants. 2. Flowering Plants. Dicotyledons, Magnoliid, Hamamelid, and Caryophyllid Families*, eds. K. Kubitzki, J. G. Rohwer, and V. Bittrich, 563–583.

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

1610 Tomasello, S., Karbstein, K., Hodač, L., Pätzold, C., and Hörandl, E. (2020). Phylogenomics
1611 unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane
1612 plant species: a case study on the *Ranunculus auricomus* species complex. *Mol Ecol* 29,
1613 2031–2049. doi: 10.1111/mec.15458
1614 Tomasello, S., Manzo, E., and Karbstein, K. (2024). Comparative plastome assembly of the
1615 yellow ironweed (Verbesina alternifolia) using Nanopore and Illumina reads. *Front*
1616 *Plant Sci* 15, 1–10. doi: 10.3389/fpls.2024.1429494
1617 Uliano-Silva, M., Ferreira, J. G. R. N., Krasheninnikova, K., Blaxter, M., Mieszkowska, N.,
1618 Hall, N., et al. (2023). MitoHiFi: a python pipeline for mitochondrial genome assembly
1619 from PacBio high fidelity reads. *BMC Bioinformatics* 24, 288. doi: 10.1186/s12859-023-
1620 05385-y
1621 Vaillancourt, B., and Buell, C. R. (2019). High molecular weight DNA isolation method from
1622 diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv*, 1–13. doi:
1623 10.1101/783159
1624 Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of
1625 polyploidy. *Nat Rev Genet* 18, 411–424. doi: 10.1038/nrg.2017.26
1626 Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome
1627 assembly from long uncorrected reads. *Genome Res* 27, 737–746. doi:
1628 10.1101/gr.214270.116
1629 Walker, J. F., Jansen, R. K., Zanis, M. J., and Emery, N. C. (2015). Sources of inversion
1630 variation in the small single copy (SSC) region of chloroplast genomes. *Am J Bot* 102,
1631 1751–1752. doi: 10.3732/ajb.1500299
1632 Wang, D., Zheng, Z., Li, Y., Hu, H., Wang, Z., Du, X., et al. (2021). Which factors contribute
1633 most to genome size variation within angiosperms? *Ecol Evol* 11, 2660–2668. doi:
1634 10.1002/ece3.7222
1635 Wang, W., Lin, L., Xiang, X.-G., Ortiz, R. D. C., Liu, Y., Xiang, K.-L., et al. (2016). The rise
1636 of angiosperm-dominated herbaceous floras: Insights from Ranunculaceae. *Sci Rep* 6,
1637 27259. doi: 10.1038/srep27259
1638 Wang, W., Lu, A.-M., Ren, Y., Endress, M. E., and Chen, Z.-D. (2009). Phylogeny and
1639 classification of Ranunculales: Evidence from four molecular loci and morphological
1640 data. *Perspect Plant Ecol Evol Syst* 11, 81–110. doi: 10.1016/j.ppees.2009.01.001
1641 Wang, Y., Liu, Y., Miao, K., Hou, L., Guo, X., and Ji, Y. (2024). A haplotype-resolved
1642 genome assembly of Coptis teeta, an endangered plant of significant medicinal value. *Sci*
1643 *Data* 11, 1012. doi: 10.1038/s41597-024-03861-5
1644 Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et
1645 al. (2018). BUSCO applications from quality assessments to gene prediction and
1646 phylogenomics. *Mol Biol Evol* 35, 543–548. doi: 10.1093/molbev/msx319
1647 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., and Jaffe, D. B. (2018). Corrigendum:
1648 Direct determination of diploid genome sequences. *Genome Res* 28, 606.1-606.1. doi:
1649 10.1101/gr.235812.118
1650 Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive
1651 visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi:
1652 10.1093/bioinformatics/btv383
1653 Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A
1654 unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8,
1655 973–982. doi: 10.1038/nrg2165
1656 Woodson, J. D., and Chory, J. (2008). Coordination of gene expression between organellar
1657 and nuclear genomes. *Nat Rev Genet* 9, 383–395. doi: 10.1038/nrg2348

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

Xiong, W., He, L., Lai, J., Dooner, H. K., and Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences* 111, 10263–10268. doi: 10.1073/pnas.1410068111

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35, W265–W268. doi: 10.1093/nar/gkm286

Yu, W., Luo, H., Yang, J., Zhang, S., Jiang, H., Zhao, X., et al. (2024). Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes. *Genome Res* 34, 326–340. doi: 10.1101/gr.278232.123

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, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S., et al. (2022). TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hortic Res* 9, 0–3. doi: 10.1093/hr/uhac017

Zhang, X., Chen, H., Ni, Y., Wu, B., Li, J., Burzyński, A., et al. (2024). Plant mitochondrial genome map (PMGmap): A software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. *Mol Ecol Resour*, 1–15. doi: 10.1111/1755-0998.13952

Zhigang, Z., Guozhen, D., Xianhui, Z., Mantang, W., and Qingji, R. (2006). Variations with altitude in reproductive traits and resource allocation of three Tibetan species of Ranunculaceae. *Aust J Bot* 54, 691. doi: 10.1071/BT05015

Zhou, W., Armijos, C. E., Lee, C., Lu, R., Wang, J., Ruhlman, T. A., et al. (2023). Plastid genome assembly using long-read data. *Mol Ecol Resour* 23, 1442–1457. doi: 10.1111/1755-0998.13787

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