Genomics 113 (2021) 4173–4183

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

The draft chromosome-level genome assembly of tetraploid ground cherry (*Prunus fruticosa* Pall.) from long reads

Thomas W. Wöhner [a,*], Ofere F. Emeriewen [a], Alexander H.J. Wittenberg [b], Harrie Schneiders [b], Ilse Vrijenhoek [b], Júlia Halász [c], Károly Hrotkó [d], Katharina J. Hoff [e,h], Lars Gabriel [e,h], Janne Lempe [a], Jens Keilwagen [f], Thomas Berner [f], Mirko Schuster [a], Andreas Peil [a], Jens Wünsche [g], Stephan Kropop [i], Henryk Flachowsky [a]

[a] *Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Fruit Crops, Pillnitzer Platz 3a, D-01326, Dresden, Germany*
[b] *Keygene N.V., P.O. Box 216, 6700 AE Wageningen, Netherlands*
[c] *Department of Genetics and Plant Breeding, Faculty of Horticultural Science, Szent István University, Ménesi Str. 44, Budapest 1118, Hungary*
[d] *Department of Floriculture and Dendrology, Institute of Landscape Architecture, Urban Planning and Ornamental Horticulture, Hungarian University of Agriculture and Life Science, Villányi Str. 35-43, Budapest 1118, Hungary*
[e] *Institute of Mathematics and Computer Science, University of Greifswald, Walther-Rathenau-Str. 47, 17489 Greifswald, Germany*
[f] *Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27, D-06484 Quedlinburg, Germany*
[g] *University of Hohenheim, Institute of Special Crops and Crop Physiology, 70593 Stuttgart, Germany*
[h] *Center for Functional Genomics of Microbes, University of Greifswald, Felix-Hausdorff-Str. 8, 17489 Greifswald, Germany*
[i] *we.heart.codes, Jagdweg 1-3, 01159 Dresden, Germany*

ARTICLE INFO

Keywords:
Genome assembly
Long read
P. fruticosa
Ground cherry
Tetraploid

ABSTRACT

Cherries are stone fruits and belong to the economically important plant family of *Rosaceae* with worldwide cultivation of different species. The ground cherry, *Prunus fruticosa* Pall., is an ancestor of cultivated sour cherry, an important tetraploid cherry species. Here, we present a long read chromosome-level draft genome assembly and related plastid sequences using the Oxford Nanopore Technology PromethION platform and R10.3 pore type. We generated a final consensus genome sequence of 366 Mb comprising eight chromosomes. The N50 scaffold was ~44 Mb with the longest chromosome being 66.5 Mb. The chloroplast and mitochondrial genomes were 158,217 bp and 383,281 bp long, which is in accordance with previously published plastid sequences. This is the first report of the genome of ground cherry (*P. fruticosa*) sequenced by long read technology only. The datasets obtained from this study provide a foundation for future breeding, molecular and evolutionary analysis in *Prunus* studies.

1. Introduction

Cherries are stone fruits belonging to the important family of *Rosaceae* fruit crops, which are produced for fresh fruit consumption or industrial processing [1]. The worldwide production of cherries was 4 million metric tons on an area of 6.7 million ha [2] in 2019. Nevertheless, cherry production worldwide is threatened by changing climatic conditions, which promote pests, *e.g.*, *Drosophila suzukii* and *Rhagoletis cerasi*, diseases, *e.g.*, *Monilinia laxa* and *Blumeriella jaapii*, as well as unfavourable abiotic conditions, *e.g.*, hail or late frost [1,3]. Breeding of new cultivars that are resistant to biotic stress factors and adapted to local climate conditions could contribute to sustainable cultivation in the long-term and secure future production. Donors for breeding and introgression of new characters and traits can be found in wild/related species of the genus *Prunus* [4–6]. The ground cherry (*Prunus fruticosa* Pall.) is a wild *Prunus* species with a small shrub-like habitus that is native from middle Europe to Western Siberia and Western China [7,8]. The natural habitats vary from open landscapes with steppe characteristics to the edges of open forests [9–11] or hillsides with stony soils [12]. *P. fruticosa* is a self-incompatible [13] tetraploid ($2n = 4\times = 32$)

[*] Corresponding author.
E-mail address: thomas.woehner@julius-kuehn.de (T.W. Wöhner).

https://doi.org/10.1016/j.ygeno.2021.11.002
Received 10 June 2021; Received in revised form 28 October 2021; Accepted 2 November 2021
Available online 11 November 2021

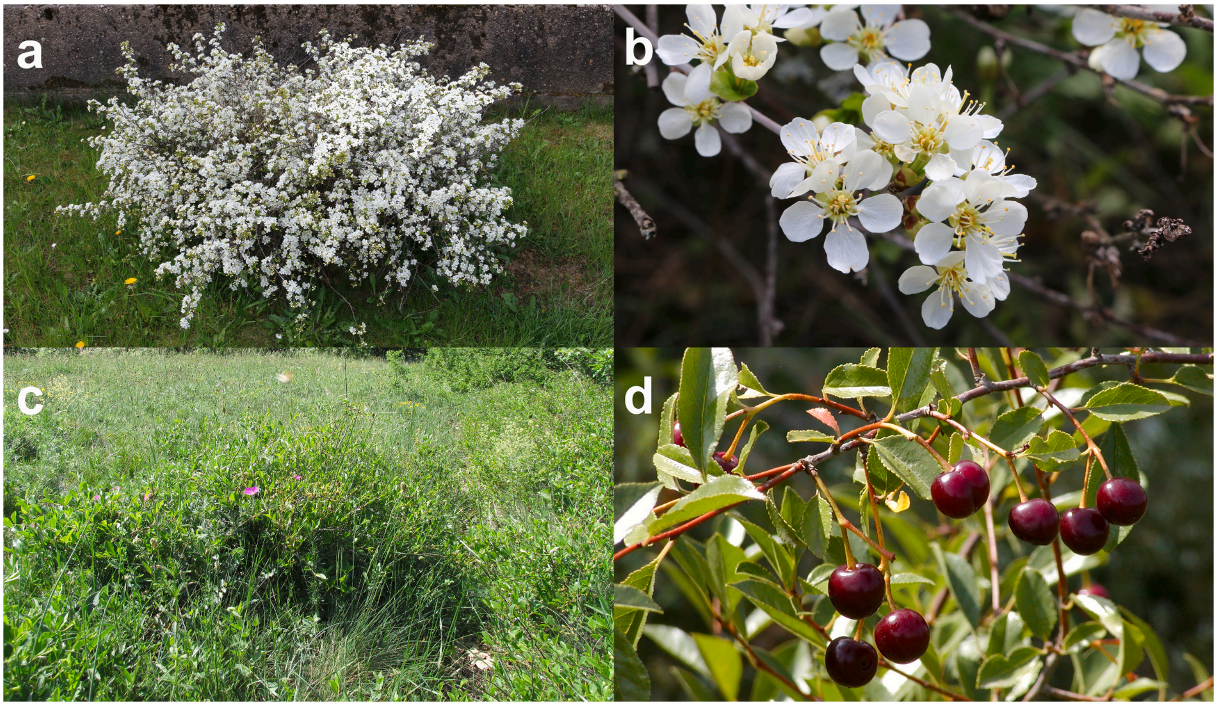

Fig. 1. Morphology of *P. fruticosa* Pall.. (a) flowering habitus, (b) inflorescence, (c) mature shrub in the natural habitat in Hungary and (d) leaves and fruits.

species with an estimated genome mass of 1.31 pg determined by flow cytometry analysis [14]. It is the progenitor of sour cherry (*P. cerasus* L.), which developed by natural hybridization from unreduced pollen of sweet cherry (*P. avium* L.) with *P. fruticosa* [15,16]. The relatedness to allochthonous *P. cerasus* and autochthonous *P. avium* is one reason why the species is endangered by crop-to –wild hybridization events in its endemic habitats in Europe [11]. However, *P. fruticosa* is a valuable genetic resource for breeding of varieties adapted to drought and low temperatures [17,18] because of its growth at cold and semi-arid sites and its edible fruits [7]. Due to its dwarf habitus, the species has been used as a donor for cherry rootstock breeding in several programmes [19–21]. Like other *Rosaceae* fruit species, cherries are perennial crops and breeding of new cultivars is labour intensive and time consuming [22]. Therefore, genome sequencing advances breeding processes enormously by providing insights into evolution and comparative studies with related species, determining the positions of putative genes, which may control different traits, and allowing for the possibility for marker-assisted selection. Hence several genomes of other *Prunus* species [23–30] as well as other members of the *Rosaceae* family [31–33] have been sequenced in recent years. The sizes of *Prunus* genomes so far sequenced range between 250 and 300 Mbp with high synteny of the eight basic chromosomes [3]. However, sequencing and assembling plant genomes is still a challenging task. The commercialization of third-generation sequencing technology has enabled rapid generation of giga-bases of data but most genome sequences are still fragmented or incomplete due to size, composition and structure (repeat content). The availability of long read sequencing technologies can solve these problems and offers many more advantages [34].

In this study, we present a draft assembly of the *P. fruticosa* Pall. genome generated with long read Oxford Nanopore Technology (ONT) on the PromethION platform and the latest R10.3 pore type. Compared to the R9.4.1 pore type, the R10.3 pore has a longer barrel and dual reader head, enabling improved resolution of homopolymeric regions and improving the consensus accuracy of nanopore sequencing data. The improved consensus accuracy (https://nanoporetech.com/accuracy) allows for *de novo* assemblies that do not rely anymore on short-read data for polishing. An additional risk with polishing using short reads is that reads get mapped incorrectly on the assembly and therefore introduce mistakes during the polishing process [35,36]. Using the final assembly for reference based scaffolding; eight chromosome scale pseudomolecules were constructed and subsequently used for gene annotation. This data provides additional information, which may be useful for breeding and genetic diversity studies in cherry and the genus *Prunus* in general.

2. Material and methods

2.1. Plant material, DNA extraction and ONT sequencing

Prunus fruticosa Pall. young leaf material (tetraploid, short type, size *ca.* 30–50 cm) was collected in its natural habitat [8] from a single tree (*in situ*) in Budapest, Hármashatárhegy (Fig. 1, coordinates 47°33′15.322"N, 18°59′49.623″E). Snap frozen plant material was sent to the sequencing service provider KeyGene N.V. (Wageningen, The Netherlands) for high molecular weight DNA extraction, purification and nanopore sequencing analysis. High molecular weight DNA was extracted by KeyGene N.V. using the nuclei isolated from frozen leaves ground under liquid nitrogen, as described elsewhere [37,38]. Genomic DNA was quality controlled with a Qubit device (Thermo Fisher Scientific, Waltham, MA, USA) and length was determined using the Femto Pulse instrument (Agilent, California). Short DNA fragments were removed using the Circulomics SRE XL kit (Circulomics, Baltimore, MD, USA) following the manufacturer's instruction. Finally, 2 μg AMPure purified genomic DNA per flow cell (AMPure PB, Pacific Biosciences, California) was used as input for library construction using the 1D Genomic DNA ligation SQK_LSK110 library prep kit (Oxford Nanopore Technologies, Oxford, UK). Subsequently, the library was loaded on three PromethION FLO PRO003 (R10.3 pore, early access pore) flow cells and run on PromethION P24 platform according to the manufacturer's recommendations. Base calling was performed in real-time on the compute module (PromethION version: 20.06.9/Guppy4.0.11). Only passed reads with a Q-value threshold of seven were used for further data analysis.

2.2. De novo assembly and scaffolding

Raw data assembly was performed using a combination of the aligner Minimap2 (2.16-r922) and the assembler Miniasm (0.2-r137-dirty) using a 20×, 30× and 50× coverage/length cut-offs at default settings. Three runs of Racon (v1.4.10) subsequently improved base accuracy of the interim contig assembly using a 10 Kb length cut-off (https://github.com/lbcb-sci/racon) and one run of Medaka (1.01) using all raw reads for consensus calling (https://denbi-nanopore-training-course.readthedocs.io/en/latest/polishing/medaka/racon.html). The sequences of the obtained contig assembly were collapsed with two runs of Purge Dups (V1.0.1) using default settings. The BUSCO (Benchmark Universal Single-Copy Orthologs - Galaxy Version 4.1.4) software was used for quantitative and quality assessment of the genome assemblies based on near-universal single-copy orthologs. The genome sequence of *P. avium* 'Tieton' ([39], GenBank assembly accession: GCA_014155035.1) was used as a matrix for reference guided scaffolding of the final assembly (purged2) using RAGOO (v1.11) with the standard settings [40]. Final sequence statistics were calculated with CLC Mainworkbench (v20.0.4). The generated *P. fruticosa* genome (Pf_1.0) was hard masked with NCBI WindowMasker [41] implementation on the CoGe platform [42]. Synteny comparisons between *P. avium* 'Tieton' and *P. persica* 'Lovell' ([24], GenBank assembly accession: GCA_000346465.2) with Pf_1.0 were performed with SynMap2 [43] using the standard program settings. The LTR assembly Index (LAI) [44] was calculated with LTR_retriever 2.9.0 (https://github.com/oushujun/LTR_retriever) to evaluate the assembly continuity between the final genome sequence Pf_1.0 and *P. avium* 'Tieton' and *P. persica* 'Lovell'. LTR_harvest (genometools 1.6.1 implementation) was used to obtain LTR-RT candidates.

2.3. Annotation

A species-specific repeat library for Pf_1.0, *P. avium* 'Tieton' and *P. persica* 'Lovell' was first generated with RepeatModeler 1.0.11 [45]. The obtained dataset was then used for repetitive sequence identifcation and masking in Pf_1.0 with ReapeatMasker 4.0.7 [46]. As no RNA-seq data for *P. fruticosa* was available, publicly available RNA-seq data [47] from the close relative *P. cerasus* 'Schattenmorelle' (SRR2290965) was downloaded from NCBI and mapped to Pf_1.0 using HISAT2 2.1.0 [48].

The structural gene annotation of genomic features is result of a combination of *ab initio* and homology-based gene annotation. *Ab initio* gene prediction was performed with both BRAKER1 [49,50] and BRAKER2 [51]. The BRAKER pipeline in general leverages extrinsic data, such as spliced alignments from short read RNA-Seq or large-scale protein to genome alignments for executing self-training GeneMark-ET/EP [52] [53,54] with help of SAMtools [55], and BamTools [56], or GeneMark-EP+ [57], with DIAMOND [58], GeneMark-ES [59], and Spaln2 [60,61] for generating an evidence-supported training gene set for the gene finder AUGUSTUS. AUGUSTUS then predicts genes with evidence where available [62] and in *ab initio* mode in local absence of evidence [63]. OrthoDB v.10 *Plantae* partition [64] and related species proteins [*P. armeniaca* (GCA_903112645.1), *P. persica* (GCF_000346465.2), *Prunus mume* (GCF_000346735.1), *P. dulcis* (GCF_902201215.1) and *P. avium* (GCF_002207925.1)] obtained from GenBank were used as reference protein dataset for BRAKER2. Gene predictions from BRAKER1 and BRAKER2 were combined into one transcript set by filtering the union of transcripts from both predictions in context with their support by the evidence generated with PrEvCo v. 0.1.0 (https://github.com/LarsGab/PrEvCo). The obtained *ab initio* annotation was augmented with additional GFF attributes using the GeMoMa module AnnotationEvidence.

Homology-based gene annotation was performed with GeMoMa version 1.7.2beta [65] using the mapped RNA-seq data from 'Schattenmorelle' [47] and the genome and gene annotation from the following reference organisms that are available at NCBI: *A. thaliana* (TAIR10.1, RefSeq GCF_000001735.4), *M. domestica* (Golden Delicious Doubled haploid version 1, GCF_002114115.1), *F. vesca* (FraVesHawai_1.0, GCF_000184155.1), *P. avium* (PAV_r1.0, GCF_002207925.1), *P. persica* (Prunus_persica_NCBIv2, GCF_000346465.2), *P. mume* (P.mume_V1.0, GCF_000346735.1), *P. dulcis* (ALMONDv2, GCF_902201215.1) and *P. armeniaca* (pruArmRojPasHapCUR, GCA_903112645.1).

The augmented *ab initio* gene annotation from BRAKER and the eight homology-based gene predictions from GeMoMa were combined using the GeMoMa module GAF yielding a final gene annotation. BUSCO with set embryophyta_odb10 (Galaxy Version 4.1.4) was used for the assessment of protein completeness. For handling alternative transcripts correctly and not as duplicates, a custom script was ran on the BUSCO full table, assigning gene ID instead of transcript ID. The functional annotation was performed with the obtained protein files using InterproScan at Galaxy Europe using default parameters [66–68] and [69].

Noncoding RNA prediction was performed with tRNAscan (Galaxy version 0.4), Aragorn (Galaxy version 0.6), barrnap (Galaxy version 1.2.1) and INFERNAL (cmsearch with rFAM 11.0, Galaxy Version 1.1.2.0).

The chloroplast and mitochondria sequences were annotated with GeSEq. [70] using the references for chloroplast from *P. fruticosa* (GenBank accession MT916286) published by [71] and mitochondria from *P. avium* (GenBank accession MK816392) published by [72]. GeSeq pipeline analysis was performed using the annotation packages ARAGORN, blatN, blatX, Chloe and HMMER.

2.4. Haplotype mining of the self-incompatibility locus (S-locus)

Coding sequences from the S-locus of *P. avium S-locus F box-like1* (*SLFL1* AB360342.1), *S-RNAse* (AY259115.1), *S haplotype-specific F-box gene* (*SFB*, AY805052.1), *SLFL2* (AB280954.1) and *P. persica* (*SLFL1* ppa021716m, *SLFL2* evm.model.contig77.461_ppc_v1.0) were obtained from NCBI database (www.ncbi.nlm.nih.gov) and genome database of Rosaceae (GDR, www.rosaceae.org). BLASTs (BLASTN 2.9.0+) and alignments of cds sequences of the S-locus containing genes were conducted against *P. avium* 'Tieton', *P. persica* and *P. fruticosa* genome sequences with the CLC Main Workbench software (21.0.1, QIAGEN Aarhus A/S). The following parameters were used for BLASTN: match/mismatch and gap costs = Match 2 Mismatch 3 Existence 5 Extension 2; Expectation value = 10.0; Word size 11; Mask lower case = No; Mask low complexity regions = Yes; Maximum number of hits = 250; Number of threads = 4; Filter out redundant results = No.

3. Results & discussion

We report the use of Oxford Nanopore technology to assemble a high-quality reference genome of *P. fruticosa* – the first report in a tetraploid *Prunus* species. Previously described genomes in *Prunus* applied Illumina, PacBio or shotgun sequencing techniques [25,26,29]. However, Wang et al. [28] reported a combination of Oxford Nanopore and Illumina technologies for sweet cherry. Table S1 and S2 summarize the raw read statistics of our study. We generated 4.5 million raw reads (124.7 Gb), which is considerably lower compared to the read output of *P. avium* cultivars [25,28]. After cleaning, approximately 4.0 million reads comprised 117.3 Gb in total (mean q = 9.96), which were generated by the R10.3 PromethION flow cells representing ~97× coverage of the estimated tetraploid genome size of 1.2 Gb. Compared to Wang et al. [28], the R10.3 flow cells produced longer reads with higher quality (Table S2). A mean of 1,347,740 (SD = 135,304) reads with a N50 length of 41,236 (SD = 275) bp and 39.1 (SD = 4.2) Gb per flow cell were obtained (Table S1). Based on the results of the raw data assemblies (Table S3), it was decided to continue with the obtained 30× coverage Miniasm assembly with a length cut-off at 62.3 kb. After three runs of Racon and one run of Medaka consensus calling, the final assembly covered approximately four times the estimated haploid genome size of ~0.3 Gb, indicating we were able to separate the parental haplotypes (4n), to a large extent. Consensus calling resulted in a total assembly size of 1161.5 Mb, represented by 4426 contigs with an N50

Table 1
Statistics of different datasets and assemblies from *P. fruticosa.*

Data set/ assembly	Ploidy	Number of contigs	Contig N50 (bp)	Longest contig (bp)	Total contig length (Mb)
All reads	4n	4,525,811	40,963	1,257,508	1247,375
Passed reads	4n	4,043,192	41,244	732,658	1172,679
Miniasm/ Minimap	4n	4399	324,889	5,840,253	1147,459
Racon	4n	4381	326,739	5,954,545	1161,2
Medaka	4n	4426	325,453	5,956,772	1161,5
Purge dups_1x	1n	1516	501,505	5,956,772	480,6
Purge dups_2x	1n	1275	533,462	5,956,772	376,7

contig size of 325 Kb and the longest contig of 5.9 Mb (Table 1). Two runs of Purge Dups were performed to collapse the haplotype-separated assembly in order to reduce the duplicated content to a haplotype consensus sequence (1n). The purged_2x assembly data set has a size of 376.7 Mb and consists of 1275 contigs with an N50 contig size of 533,426 bp. Compared to Wang et al. [28], the final purged_2x assembly did not reach the contiguity metrics that had been obtained by Illumina Hi-C and ONT 9.4.1.

Although it is recommended to use Hi-C or bio-nano optical mapping to resolve structural variants within the chromosomes [73], the final assembly was used as input for reference-guided scaffolding using RaGoo and the genome sequence of *P. avium* 'Tieton' [28] to generate a chromosome scale genome sequence. The obtained sequence file consists of nine scaffolds representing eight chromosomes and one sequence with concatenated unmapped data (unassigned). The final *Prunus fruticosa* 1.0 genome sequence (Fig. 2) consists of 366.5 Mb with a N50 size of scaffolds about 43,818,497 bp and G + C of 37.74%, A + T of 62.22% and only 0.03% gaps (N). The longest scaffold is 66,497,422 bp (Table 2). Compared to the genome sequences available so far in *Prunus* [24,25,28], the genome of *P. fruticosa* is the most complete obtained from long read sequencing only.

Although contiguity was comparably low, the obtained BUSCO analysis resulted in 98.6% - 98.7% completeness for the representing 4n Racon and Medaka generated data sets. This is in agreement with the report by [74] that measures of contiguity are not strongly correlated to biological completion or assembly correctness. The comparison of BUSCO results (Fig. 3) on assembly completeness between the Racon only and the Racon and Medaka data sets (Table 1) indicates that consensus generation by Medaka increases the number of duplicated genes (from 89.7% to 92.4%) and improves the consensus accuracy. The obtained assembly sequences (1n) after haplotig removal showed a decrease of duplicated BUSCOS (from 92.4% to 12.5%) and an increase

Table 2
Pseudomolecule statistics for Pf_1.0.

Pseudomolecule	Total size (bp)	%
Pf_1.0_chr1	66,497,422	18.1
Pf_1.0_chr2	59,585,028	16.3
Pf_1.0_chr3	39,930,086	10.9
Pf_1.0_chr4	42,034,286	11.5
Pf_1.0_chr5	31,043,513	8.5
Pf_1.0_chr6	46,922,205	12.8
Pf_1.0_chr7	36,673,485	10.0
Pf_1.0_chr8	43,818,497	12.0
	366,504,522	100

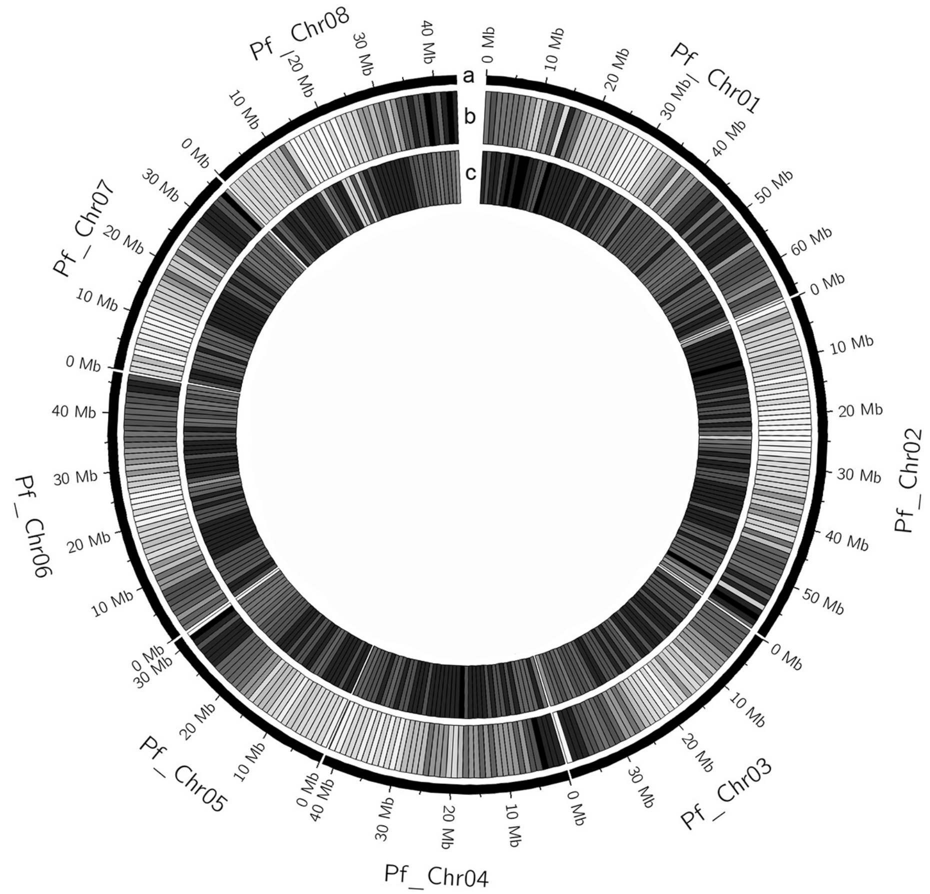

Fig. 2. The genome of *P. fruticosa.* Circos plot of the 8 pseudomolecules. (a) Chromosome length (Mb); (b) gene density in blocks of 1 MB; (c) repeat density in blocks of 1 Mb.

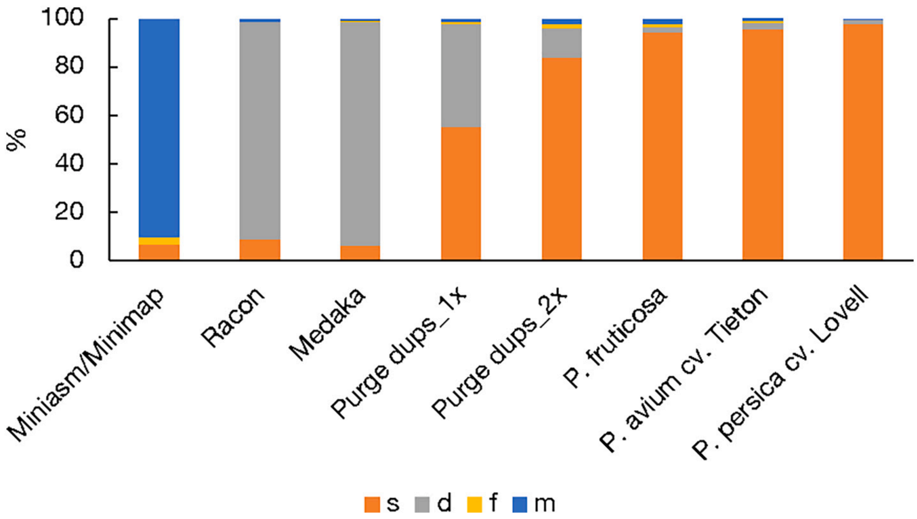

Fig. 3. Analysis of completeness of different *P. fruticosa* datasets compared to *P. avium* cv. 'Tieton' and *P. persica* cv. Lovell by mapping of a set of universal single-copy orthologs using BUSCO. The bar charts indicate complete **s**ingle copy (orange), complete **d**uplicated (gray), **f**ragmented (yellow) and **m**issing (blue) genes. For evaluation the embryophyta_odb10 BUSCO dataset (n = 1614) was used. *P. fruticosa* 1.0 show a 96.4% completeness (S: 94.1%, D: 2.3%, F: 1.3%, M: 2.3%, n: 1614) which almost reaches the completeness of *P. avium* cv. 'Tieton' (C: 98.3%, S: 95.6%, D: 2.7%, F: 0.5%, M:1.5%, n:1614) and *P. persica* 'Lovell' (C: 99.3%, S: 97.5%, D: 1.8%, F: 0.1%, M: 0.6%, n:1614). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

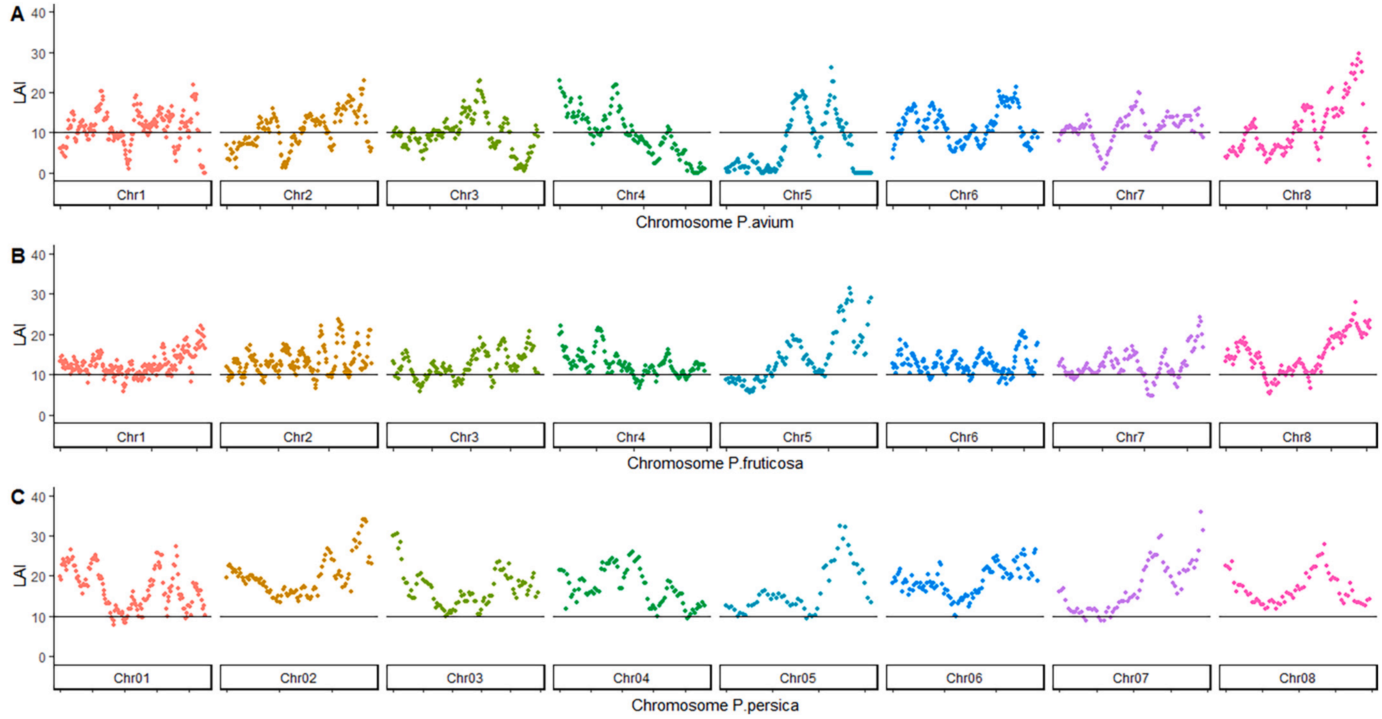

Fig. 4. Assessing the assembly quality of repetitive sequences between the chromosome sequences of *P. avium* 'Tieton' (A), *P. fruticosa* (B) and *P. perisca* (C) using the LAI index. The genome of *P. avium* was sequenced with ONT 9.4.1 and Illumina [28], *P. fruticosa* with ONT 10.3 and *P. persica* with Illumina and Sanger sequencing of fosmid and BAC clones [23,24].

of single BUSCOS (from 6.3% to 83.6%). *P. fruticosa* 1.0 results outlined in Fig. 3 show a 96.4% completeness. Compared to the genome sequence of *P. persica* (99.3%) and *P. avium* (98.3%) which represent the highest genome completeness of published *Prunus* datasets, the obtained long read only assemblies (98.7%) and consensus genome sequence (96.4%) from this study shows a comparably high genome completeness. To assess the continuity of the final genome, we calculated the LAI index for *P. fruticosa, P.avium* and *P. persica* using the chromosome sequences only (unscaffolded sequences were not integrated into the calculation). The LAI index is independent of genome size, LTR-RT content and gene space and reflects the correctness of the assembly [44]. The whole genome LAI obtained for *P. fruticosa* was 12.5 compared to 9.6 for *P. avium* 'Tieton' and 16.4 for *P. persica.* This is in accordance with other long-read assemblies [44], but Wang et al. [28] presented a much higher LAI for *P. avium* 'Tieton' (19.7) and *P. persica* (18.8). Fig. 4 contains the LAI index over the single chromosomes from all three compared genome sequences and reveals the robust continuity of the *P. fruticosa* genome assembly. Due to the proposed genome classification system from [44], the genome of *P. fruticosa* can be classified as reference.

Our approach detected 189.7 Mb of repetitive sequences (51.75% of the genome) and 42.1 Mb (11.5%) unknown elements. Repetitive sequences observed in other *Prunus* species [25–27,29,33] ranged from 37.1% in *P. persica* [53] to 59.4% in *P. avium* [28]. Due to sequencing technologies and repeat prediction software used for repeat modeling, the number of transposable elements can vary [75]. We additionally modeled the repeats of *P. avium* and *P. persica* (Table 3). A total of 30.5% and 42.6% of the genomes were repetitive. However, similar to *P. avium* [25] and *P. persica*, the repeated sequences observed in our study

Table 3
Characterization of repetitive sequences of *P. fruticosa* 1.0 compared to *P. avium* and *P. persica.Repetitive elements which represent the highest proportion or uniqueness only for P. fruticosa are indicated in bold.*

Class	Order	Family	No. of elements			Length (bp)			Percentage of the genome (%)		
			Pf	Pa	Pp	Pf	Pa	Pp	Pf	Pa	Pp
I (retrotransposons)	**LTR**	–	2142	1723	2607	472,290	264,621	446,395	0.13	0.08	0.20
		Cassandra	1852	1179	753	910,040	323,669	329,299	0.25	0.09	0.15
		Caulimovirus	793	1586	697	627,333	861,208	933,515	0.17	0.25	0.41
		Copia	**41,192**	**32,612**	**23,578**	**27,822,528**	**12,487,829**	**14,606,294**	**7.59**	**3.64**	**6.47**
		Gypsy	**68,445**	**45,868**	**25,860**	**76,652,400**	**20,554,372**	**19,004,947**	**20.91**	**5.99**	**8.42**
		ERV1	–	94	–	–	17,404	–	–	0.01	–
		ERVK	–	–	195	–	–	41,513	–	–	0.02
		Pao	344	–	200	96,802	–	156,072	0.03	–	0.07
	LINE	I-Jockey	413	464	107	140,619	110,916	38,834	0.04	0.03	0.02
		L1	8844	6349	4286	4,167,515	2,449,897	1,392,308	1.14	0.71	0.62
		L2	434	–	110	64,430	–	21,600	0.02	–	0.01
		Penelope	176	–	218	25,448	–	34,866	0.01	–	0.02
		RTE-BovB	516	214	–	87,801	91,608	–	0.02	0.03	–
		L1-Tx1	–	–	211	–	–	46,145	–	–	0.02
		R2-NeSL	–	–	184	–	–	36,760	–	–	0.02
		Rex-Babar	–	–	89	–	–	30,378	–	–	0.01
		CR1	–	–	712	–	–	609,177	–	–	0.27
		TAD1	–	37	–	–	6991	–	–	0.00	–
	SINE	–	457	620	1886	62,956	49,823	192,330	0.02	0.01	0.09
		ID	–	–	1450	–	–	121,213	–	–	0.05
		tRNA-Core-L2	–	–	1109	–	–	122,977	–	–	0.05
		B2	1517	123	–	122,973	9505	–	0.03	0.00	–
		tRNA	4593	5378	2229	509,966	517,003	203,010	0.14	0.15	0.09
II (DNA transposons)											
Subclass I	TIR	TcMar-Fot1	276	123	–	210,022	48,369	–	0.06	0.01	–
		hAT-Charlie	–	–	1984	–	–	468,323	–	–	0.21
		IS3EU	–	103	–	–	19,992	–	–	0.01	–
		P	–	141	–	–	23,742	–	–	0.01	–
		Sola-3	–	111	–	–	42,070	–	–	0.01	–
		TcMAr	–	–	51	–	–	2513	–	–	0.00
		TcMAr-Tigger	–	–	152	–	–	121,325	–	–	0.05
		Zisupton	–	–	141	–	–	18,330	–	–	0.01
		TcMar-ISRm11	81	55	–	24,496	21,760	–	0.01	0.01	–
		hAT-Ac	11,533	14,886	6511	3,430,880	2,101,683	2,127,280	0.94	0.61	0.01
		hAT-Tag1	5353	5258	3219	1,263,452	971,199	896,657	0.34	0.28	0.40
		hAT-Tip100	9680	8622	6079	2,348,735	1,677,234	1,336,903	0.64	0.49	0.59
		PIF-Harbinger	14,230	14,435	9390	4,268,364	3,498,800	4,123,535	1.16	1.02	1.83
Subclass II	Crypton	**Crypton-H**	**237**	**–**	**–**	**195,974**	**–**	**–**	**0.05**	**–**	**–**
	Maverick	Maverick	576	357	–	155,067	76,204	–	0.04	0.02	–
	Helitron	Helitron	5378	4241	3584	2,220,498	1,496,918	1,255,959	0.61	0.44	0.56
		unknown/Helitron	228	78	92	155,920	17,774	44,752	0.04	0.01	0.02
Other		–	13,120	9946	9302	2,310,744	1,774,664	1,865,367	0.63	0.52	0.83
		Academ	**42**	**–**	**–**	**20,252**	**–**	**–**	**0.01**	**–**	**–**
		CMC-EnSpm	16,958	14,886	10,222	8,879,643	4,747,818	12,856,725	2.42	1.38	5.70
		Ginger	325	–	99	77,794	–	12,430	0.02	–	0.01
		MULE-MuDR	17,464	16,744	9606	4,459,943	3,751,469	3,049,949	1.22	1.09	1.35
rRNA			326	30	223	231,622	10,508	334,205	0.06	0.00	0.15
snRNA			–	55	127	–	4318	21,457	–	0.00	0.01
Satellite			870	397	342	220,737	88,777	137,022	0.06	0.03	0.06
Simple repeat			106,232	81,995	77,870	4,353,840	3,831,673	6,797,949	1.19	1.12	3.01
Low complexity			19,611	15,501	13,876	984,829	756,687	653,107	0.27	0.22	0.29
Unknown			**168,094**	**171,338**	**99,691**	**42,110,587**	**41,991,523**	**21,700,006**	**11.49**	**12.25**	**9.61**
SUM			*522,228*	*450,112*	*319,042*	*189,663,955*	*104,698,028*	*96,191,427*	*51.75*	*30.53*	*42.62*

comprised mainly of the class (I) LTR Gypsy retrotransposons and *Copia*. LTR was the most abundant element in our findings with 20.88% followed by Copia with 7.59% (Table 3). Whereas the LTR element ERV1, LINE element TAD1 and TIR elements IS3EU, P, Sola-3 were unique for *P. avium*, the LTR element ERVK, LINE (L1-Tx1, R2-NeSL, Rex-Babar, Cr1), SINE (tRNA-Core-L2, ID) and TIR (hat-Charlie, TcMAr, TC-Mar-Trigger, Zisuption) elements were unique to *P. persica.* Further, the Academ and Subclass II DNA retrotransposon Crypton-H were uniquely detected in *P. fruticosa.* We employed similar strategy as reported elsewhere namely homology-based, *de novo* and transcriptome supported approaches [28,39] to call repeats, predict protein-coding genes and perform functional annotation. Using RNA-Seq data from *P. cerasus* 'Schattenmorelle' [47] and the augmented gene predictions from BRAKER with eight homology-based gene predictions from GeMoMa, we predicted 58,880 protein-coding transcripts representing 84,524 orthologs within Pf_1.0 with a mean length of 3580 bp and a mean protein length of 355 aa (Table 4). The number of protein-coding transcripts was considerably larger in this study than 38,275 predicted for *P. avium* 'Tieton' [28] and 43,349 transcripts predicted in *P. avium* 'Satonishiki' [25]. A total of 86.7% (75,113) proteins was functionally annotated by InterproScan resulting in 852,470 annotated protein domains and sites from 15 protein databases (Table 4). A total of 2301 (Aragorn) and 2559 (tRNA scan) tRNA and 576 rRNA sequences were detected. Infernal search reveals 36,757 consensus RNA secondary structure profiles. BUSCO analysis for transcriptome completeness (embryophyta_odb10 dataset) reveals 1552 (96.2%) complete (81.8% single and complete, 14.4% duplicated and complete) and 62 (3.8%) fragmented (1.7%) or missing (2.1%) BUSCOs (Fig. 5).

Table 4
Functional annotation results generated by interproscan using BRAKER & GeMoMa combination of *ab-initio* and homology-based structural gene annotation and statistics.

Interproscan annotations	No.
Coils	14,627
Gene3D	82,428
Hamap	1336
PANTHER	150,554
Pfam	95,569
Phobius	197,895
PIRSF	5075
PRINTS	48,332
ProSitePatterns	19,050
ProSiteProfiles	52,557
SignalP_EUK	7914
SMART	42,825
SUPERFAMILY	64,033
TIGRFAM	10,603
TMHMM	59,672
Sum	**852,470**

Transcripts	No.	%
Total	58,880	
Orthologs	84,524	100
Annotated	73,315	86.7
Annotated GO	45,196	53.5
Annotated pathways	5247	6.2
Domains	62,431	73.9
Mean length (bp)	3580	
Mean length of predicted proteins	355	

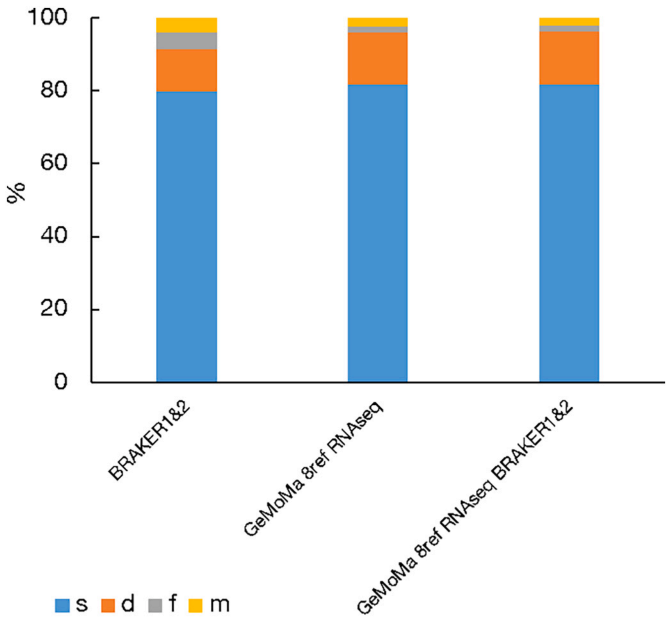

Fig. 5. Analysis of completeness of different protein sets obtained with different structural annotation strategies. The bar charts indicate complete **s**ingle copy (orange), complete **d**uplicated (gray), **f**ragmented (yellow) and **m**issing (blue) genes. For evaluation the embryophyta_odb10 BUSCO dataset (n = 1614) was used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The obtained chloroplast genome sequence (Fig. 6a) was 158,130 bp long (GC 36.6%) with a typical quadripartite structure consisting a large (86,242 bp) and a small (19,143) single-copy region and two inverted repeats (IRA 26,372 bp, IRB 26,373 bp). The GC contents of each region were 34.1% (LSC), 30.1% (SSC) and 42.5% for IRA and IRB each. The size, structure and GC content values are similar to those reported previously for the chloroplast genome of *P. fruticosa* [71]. Forty-five tRNA (ARAGORN), eight rRNA (each with HMMER and blatN) and 116 protein-coding genes (HMMER) were annotated.

We present for the first time a mitochondrial genome for *P. fruticosa* (Fig. 6b) with a length of 383,281 bp and a GC content of 45.7%. The results of the mitochondria genome is similar to the mitochondria genome of *P. avium* 'Summit' [72] where a total of 68 protein coding genes, including 27 tRNA (ARAGORN) and two rRNA (blatN) were annotated.

We compared sequence synteny between *P. fruticosa* and *P. persica* and *P. fruticosa* and *P. avium* (Fig. 7). The synteny analysis involved at least two transcripts of annotated genes in each representative genome (Fig. 7a). As indicated in Table S4, a higher percentage of transcripts (77.5% to 87.3%) were mapped between the homologous chromosomes from *P. persica* and Pf_1.0 compared to the transcripts from *P. avium* (72.1% to 56.3%). In general, the assembled genome of *P. fruticosa* shows good synteny with the genomes of *P. persica* [24] and *P. avium* [28]. Fig. 6b shows the synteny analysis using masked sequences (*i.e.* without repetitive sequences). The results obtained confirm strong synteny between the compared genomes and strongly suggest the high quality of the obtained genome sequence. Since long-read sequencing and referenced based scaffolding will not resolve inter-structural variations within chromosomes, it is therefore recommended to use Hi-C or bionano-optical mapping strategies to model a reference genome.

Nevertheless, the assembly of haplotigs will provide the possibility to mount haplotypes of regions of interest. The S-locus of *Prunus* species is a well-studied genomic region [26] and suitable for analyzing the presence of possible haplotypes within this region. It is flanked by the *SLFL1* and *SLFL2* genes. Using a local alignment strategy, we identified the S-locus on chromosome 6 in *P. fruticosa* (43,997,811 - 44,041,281 bp), *P. avium* (34,992,053–35,108,882 bp) and *P. persica* (28,263,530–28,315,441 bp). After alignment based identification of putative homologous genes of *SLFL1, SLFL2, SRNAse* and *SFB* in *P. fruticosa, P. persica* and *P. avium* within the S-locus region, we determined the position of each gene (Fig. S1, Table S5). Using basic local alignments, we identified six contigs (utg003453 = h1, utg003652 = h2, utg003404 = h3, utg003454 = h4, utg001458 = h5, utg001832 = h6) in the racon polished assembly dataset, which represent putative haplotigs of *P. fruticosa.* After assigning the position and gene content analysis, only four (h1-h4) of the six haplotigs contained the four S-locus genes in the same order. Percent identity ranged from 77.3% to 92.2% for *SRNAse* and 81.7% to 100% for *SLFL1*, *SLFL2* and *SFB* (Table S5). There was an absence of *SFB* and *SRNAse* in haplotig 5. Haplotig 6 contained all four genes, but in a different order and a BLAST analysis of the end of haplotig 6 resulted in the localization of the sequence on chromosome 5. It is unclear whether haplotig 6 is another haplotype or a result of incorrect assembly.

4. Conclusion

For the first time, we report a draft genome scale-assembly of tetraploid *Prunus* species. This was achieved using Nanopore sequencing technology, confirming that this technology alone can sufficiently produce a high-quality complete genome draft [76]. This genome will be valuable in exploiting genetic information for breeding programs; will enhance our understanding of genetics of this species relative to breeding as well as molecular and evolutionary analysis in the genus *Prunus*.

Data availability

Data supporting the findings of this study are deposited into the Open Agrar repository [77] and could also be made available on personal request to the corresponding author. The ground cherry genome has been deposited at DDBJ/ENA/GenBank under the accession JAHHUK000000000. The version described in this paper is version JAHHUK010000000. The annotation is submitted at NCBI (SUB10399888).

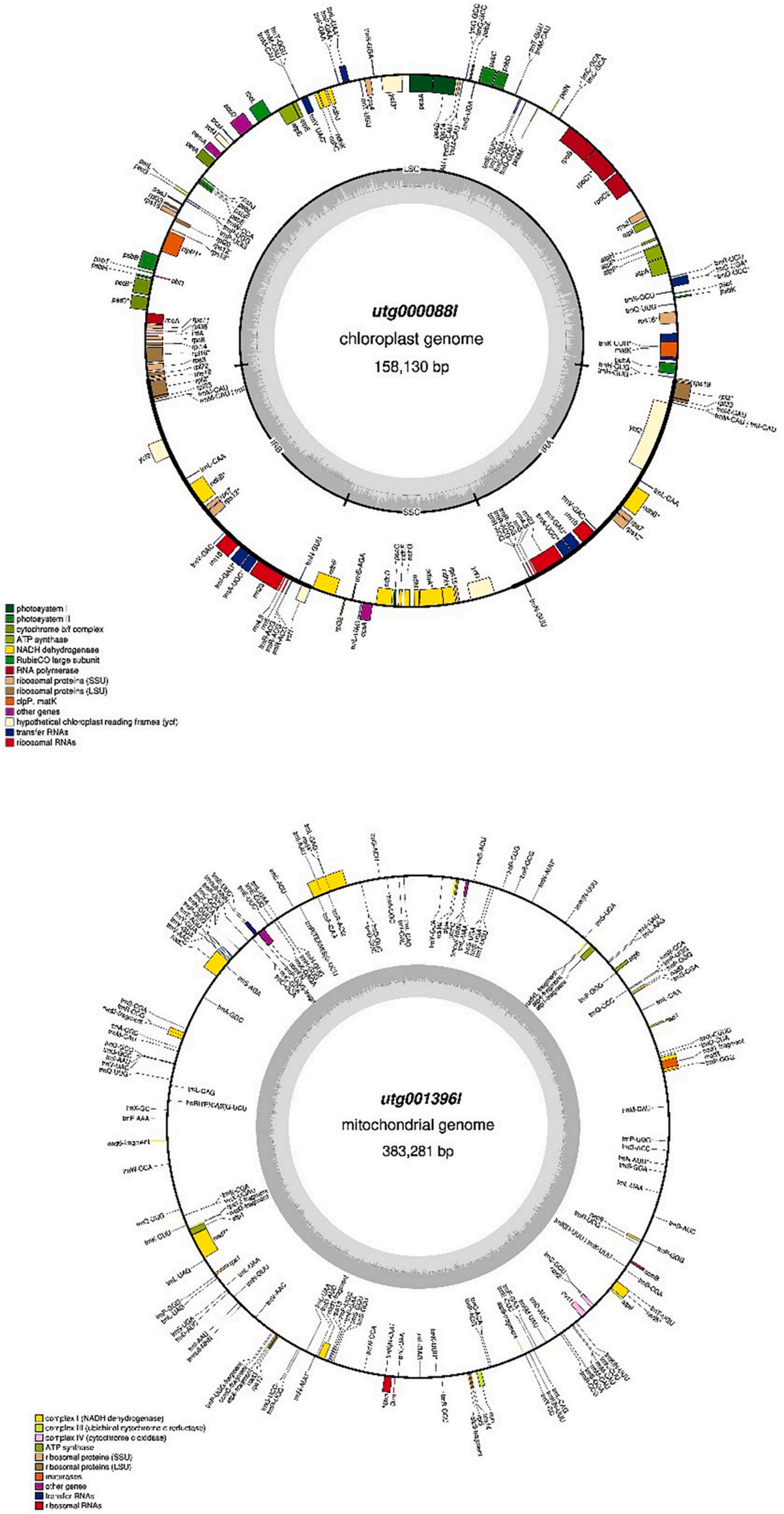

Fig. 6. The chloroplast (a) and mitochondrial (b) genome sequence of *P. fruticosa* 1.0 obtained from the contigs utg000088l and utg001396I in the medaka assembly sequence. Annotation was performed using GeSEq. [70].

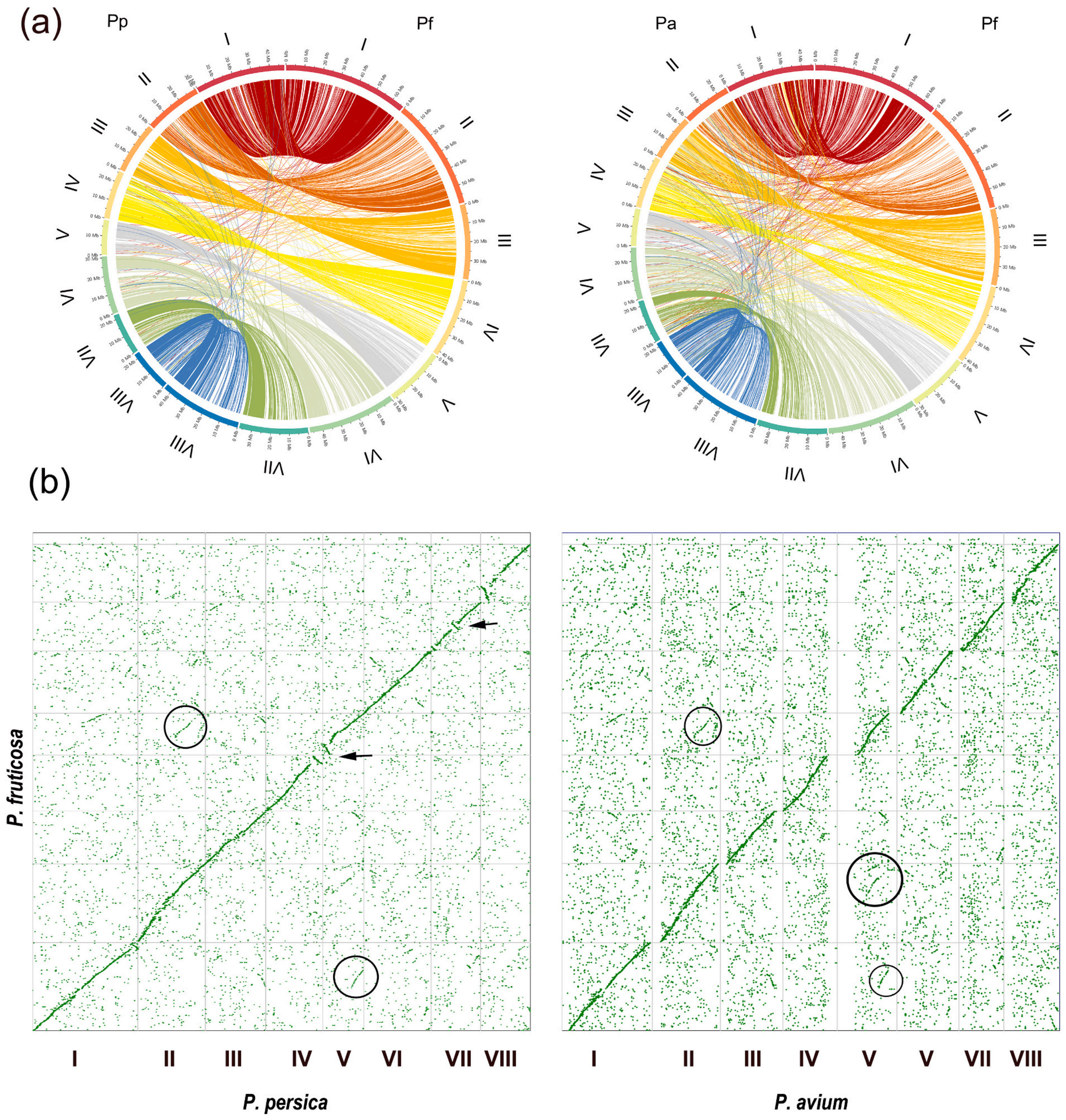

Fig. 7. Synteny between *P. fruticosa*, *P. persica* 'Lovell' and *P. avium* 'Tieton'. (a) Circos plots showing transcripts of *P. persica* (Pp, left) and *P. avium* (Pa, right) annotated in *P. fruticosa* (Pf). Each string represents at least two transcripts in a 50 k bp cluster. (b) Syntenic dot plot of the nucleotide sequences between *P. fruticosa*, *P. persica* and *P. avium*. Before plotting, the sequences were hard masked by the NCBI window maker implication on the CoGe webpage. Several inversions (arrows) and out-paralogs (circles) were identified between the sequences.

Research involving plants

Experimental research on plants and plant materials in this study comply with institutional, national, or international guidelines.

Authors' contribution

TW, OE wrote the manuscript. AW, HS and IV performed DNA isolation, sequencing and genome assembly. JL calculated the LAI index, JH and KH provided the plant material. KH, JK, LG and TB performed annotation of the dataset. SK performed the scaffolding and TW the did the interproscan and synteny analysis. HF, JW, MS and AP conceived the study and made substantial contributions to its design, acquisition, analysis and interpretation of data. All authors contributed equally to the finalization of the manuscript.

Declaration of Competing Interest

The R10.3 flow cells were provided by Keygene for the project. Keygene wanted to gain experience with this new flow cells on a biologically difficult object. Keygene had no influence on the interpretation of the results and the writing of the manuscript.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.11.002.

References

[1] J. Quero-Garcia, A. Iezzoni, G. Lopez-Ortega, C. Peace, M. Fouche, E. Dirlewanger, M. Schuster, Advances and Challenges in Cherry Breeding, Burleigh Dodds Science Publishing, 2019.
[2] FAO, Crops. http://www.fao.org/faostat/en/#data/QC/, 2021 (accessed 17 February 2021).
[3] M.J. Aranzana, V. Decroocq, E. Dirlewanger, I. Eduardo, Z.S. Gao, K. Gasic, A. Iezzoni, S. Jung, C. Peace, H. Prieto, Prunus genetics and applications after de novo genome sequencing: achievements and prospects, Horticul. Res. 6 (2019) 1–25.
[4] M. Schuster, C. Grafe, E. Hoberg, W. Schütze, Interspecific hybridization in sweet and sour cherry breeding, Acta Hortic. 976 (2013) 79–86.
[5] J. Wen, S.T. Berggren, C.-H. Lee, S. Ickert-Bond, T.-S. Yi, K.-O. Yoo, L. Xie, J. Shaw, D. Potter, Phylogenetic inferences in Prunus (Rosaceae) using chloroplast ndhF and nuclear ribosomal ITS sequences, J. Syst. Evol. 46 (2008) 322–332.
[6] T. Stegmeir, M. Schuster, A. Sebolt, U. Rosyara, G.W. Sundin, A. Iezzoni, Cherry leaf spot resistance in cherry (Prunus) is associated with a quantitative trait locus on linkage group 4 inherited from P. canescens, Mol. Breed. 34 (2014) 927–935.
[7] M. Faust, D. Surányi, Origin and dissemination of cherry, Hortic. Rev. 19 (1997) 263–317.
[8] K. Hrotkó, Y. Feng, J. Halász, Spontaneous hybrids of Prunus fruticosa pall. In Hungary, Genet. Resour. Crop. Evol. 67 (2020) 489–502.
[9] E.J. Jäger, D. Seidel, Unterfamilie Prunoideae, Hegi Illustrierte Flora von Mitteleuropa 4, 1995.
[10] H. Meusel, E.J. Jäger, E. Weinert, Vergleichende Chorologie der Zentraleuropaischen Flora, 1965.
[11] L. Macková, P. Vít, T. Urfus, Crop-to-wild hybridization in cherries—empirical evidence from Prunus fruticosa, Evol. Appl. 11 (2018) 1748–1759.
[12] E. Mratinić, M. Kojić, Wild Fruit Species of Serbia, Agricultural Research Institute of Serbia, Belgrade, 1998.
[13] K. Pruski, Tissue culture propagation of Mongolian cherry (*Prunus fruticosa* L.) and Nanking cherry (*Prunus tomentosa* L.), in: Protocols for Micropropagation of Woody Trees and Fruits, Springer, 2007, pp. 391–407.
[14] L. Macková, P. Vít, Ľ. Ďurišová, P. Eliáš, T. Urfus, Hybridization success is largely limited to homoploid Prunus hybrids: a multidisciplinary approach, Plant Syst. Evol. 303 (2017) 481–495.
[15] E.J. Olden, N. Nybom, On the origin of *Prunus cerasus* L, Hereditas 59 (1968) 327–345.
[16] T.S. Brettin, R. Karle, E.L. Crowe, A.F. Iezzoni, Chloroplast inheritance and DNA variation in sweet, sour, and ground cherry, J. Hered. 91 (2000) 75–79.
[17] I.W. Mitschurin, Ausgewählte Schriften, in: Verlag Kultur, Fortschritt (Eds.), Ausgewählte Schriften 1951, 1951.
[18] R.H. Bors, Dwarf sour cherry breeding at the University of Saskatchewan, Acta Hortic. 667 (2005) 135–140.
[19] J.N. Cummins, Vegetatively propagated selections of Prunus fruticosa as dwarfing stocks for cherry, Fruit Var. Hort. Dig. 26 (1972) 76–79.
[20] H. Plock, Bedeutung der Prunus fruticosa Pall. als Zwergunterlage fur Suss-und Sauerkirschen, Mitt. Rebe. Wein. Obstbau. Fruchteverwert (1973) 137–140.
[21] K. Hein, Zwischenbericht über eine Prüfung der Steppenkirsche (*P. fruticosa*) und anderen Süsskirchenunterlagen und Unterlagenkombinationen, Erwerbsobstbau 21 (1979) 219.
[22] C. Peace, J. Norelli, Genomics approaches to crop improvement in the Rosaceae, in: Genetics and Genomics of Rosaceae, Springer, 2009, pp. 19–53.
[23] R. Ahmad, D.E. Parfitt, J. Fass, E. Ogundiwin, A. Dhingra, T.M. Gradziel, D. Lin, N. A. Joshi, P.J. Martinez-Garcia, C.H. Crisosto, Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection, BMC Genomics 12 (2011) 1–7.
[24] I. Verde, J. Jenkins, L. Dondini, S. Micali, G. Pagliarani, E. Vendramin, R. Paris, V. Aramini, L. Gazza, L. Rossini, The peach v2. 0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity, BMC Genomics 18 (2017) 1–18.
[25] K. Shirasawa, K. Isuzugawa, M. Ikenaga, Y. Saito, T. Yamamoto, H. Hirakawa, S. Isobe, The genome sequence of sweet cherry (Prunus avium) for use in genomics-assisted breeding, DNA Res. 24 (2017) 499–508.
[26] S. Baek, K. Choi, G.-B. Kim, H.-J. Yu, A. Cho, H. Jang, C. Kim, H.-J. Kim, K. S. Chang, J.-H. Kim, Draft genome sequence of wild Prunus yedoensis reveals massive inter-specific hybridization between sympatric flowering cherries, Genome Biol. 19 (2018) 1–17.
[27] K. Shirasawa, T. Esumi, H. Hirakawa, H. Tanaka, A. Itai, A. Ghelfi, H. Nagasaki, S. Isobe, Phased genome sequence of an interspecific hybrid flowering cherry,'Somei-Yoshino'(Cerasus× yedoensis), DNA Res. 26 (2019) 379–389.
[28] J. Wang, W. Liu, D. Zhu, P. Hong, S. Zhang, S. Xiao, Y. Tan, X. Chen, L. Xu, X. Zong, Chromosome-scale genome assembly of sweet cherry (Prunus avium L.) cv. Tieton obtained using long-read and Hi-C sequencing, Horticult. Res. 7 (2020) 1–11.
[29] Q. Zhang, W. Chen, L. Sun, F. Zhao, B. Huang, W. Yang, Y. Tao, J. Wang, Z. Yuan, G. Fan, The genome of Prunus mume, Nat. Commun. 3 (2012) 1–8.
[30] A.M. Callahan, T.N. Zhebentyayeva, J.L. Humann, C.A. Saski, K.D. Galimba, L. L. Georgi, R. Scorza, D. Main, C.D. Dardick, Defining the 'HoneySweet' insertion event utilizing NextGen sequencing and a de novo genome assembly of plum (*Prunus domestica*), Horticult. Res. (2021) 1–13.
[31] R. Velasco, A. Zharkikh, J. Affourtit, A. Dhingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S.K. Bhatnagar, M. Troggio, D. Pruss, The genome of the domesticated apple (Malus× domestica Borkh.), Nat. Genet. 42 (2010) 833–839.
[32] N. Daccord, J.-M. Celton, G. Linsmith, C. Becker, N. Choisne, E. Schijlen, H. van de Geest, L. Bianco, D. Micheletti, R. Velasco, High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development, Nat. Genet. 49 (2017) 1099–1106.
[33] F. Jiang, J. Zhang, S. Wang, L. Yang, Y. Luo, S. Gao, M. Zhang, S. Wu, S. Hu, H. Sun, The apricot (Prunus armeniaca L.) genome elucidates Rosaceae evolution and beta-carotenoid synthesis, Horticult. Res. 6 (2019) 1–12.
[34] M. Jain, H.E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community, Genome Biol. 17 (2016) 1–11.
[35] C. Liu, X. Yang, B.F. Duffy, J. Hoisington-Lopez, M. Crosby, R. Porche-Sorbet, K. Saito, R. Berry, V. Swamidass, R.D. Mitra, High-resolution HLA typing by long reads from the R10. 3 Oxford nanopore flow cells, Hum. Immunol. 82 (2021) 288–295.
[36] S.M. Karst, R.M. Ziels, R.H. Kirkegaard, E.A. Sørensen, D. McDonald, Q. Zhu, R. Knight, M. Albertsen, High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing, Nat. Methods 18 (2021) 165–169.
[37] M. Zhang, Y. Zhang, C.F. Scheuring, C.-C. Wu, J.J. Dong, H.-B. Zhang, Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research, Nat. Protoc. 7 (2012) 467–478.
[38] E. Datema, R.J.M. Hulzink, L. Blommers, J.E. Valle-Inclan, N. van Orsouw, A.H. J. Wittenberg, M. de Vos, The megabase-sized fungal genome of Rhizoctonia solani assembled from nanopore reads only, BioRxiv (2016) 1–15.
[39] C. Liu, C. Feng, W. Peng, J. Hao, J. Wang, J. Pan, Y. He, Chromosome-level draft genome of a diploid plum (*Prunus salicina*), GigaScience 9 (2020), giaa130.
[40] M. Alonge, S. Soyk, S. Ramakrishnan, X. Wang, S. Goodwin, F.J. Sedlazeck, Z. B. Lippman, M.C. Schatz, RaGOO: fast and accurate reference-guided scaffolding of draft genomes, Genome Biol. 20 (2019) 1–17.
[41] A. Morgulis, E.M. Gertz, A.A. Schäffer, R. Agarwala, WindowMasker: window-based masker for sequenced genomes, Bioinformatics 22 (2006) 134–141.
[42] E.H. Lyons, CoGe, a New Kind of Comparative Genomics Platform: Insights into the Evolution of Plant Genomes, University of California, Berkeley, 2008.
[43] A. Haug-Baltzell, S.A. Stephens, S. Davey, C.E. Scheidegger, E. Lyons, SynMap2 and SynMap3D: web-based whole-genome synteny browsers, Bioinformatics 33 (2017) 2197–2198.
[44] S. Ou, J. Chen, N. Jiang, Assessing genome assembly quality using the LTR assembly index (LAI), Nucleic Acids Res. 46 (2018) 1–11.
[45] A.F. Smit, R. Hubley, P. Green, RepeatModeler Open-1.0. 2008–2015, Seattle, USA: Institute for Systems Biology. Available from: http//www.repeatmasker.org, Last Accessed May 1 (2015) 2018.
[46] A.F. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013–2015, 2015.
[47] Y. Jo, H. Chu, J.K. Cho, H. Choi, S. Lian, W.K. Cho, De novo transcriptome assembly of a sour cherry cultivar, Schattenmorelle, Genomics Data 6 (2015) 271–272.
[48] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915.
[49] K.J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS, Bioinformatics 32 (2016) 767–769.
[50] K.J. Hoff, A. Lomsadze, M. Borodovsky, M. Stanke, Whole-genome annotation with BRAKER, in: Gene Prediction, Springer, 2019, pp. 65–95.
[51] T. Brůna, K.J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database, NAR Genomics Bioinformat. 3 (2021) 1–11.
[52] A. Lomsadze, P.D. Burns, M. Borodovsky, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm, Nucleic Acids Res. 42 (2014) 1–8.
[53] I. Verde, A.G. Abbott, S. Scalabrin, S. Jung, S. Shu, F. Marroni, T. Zhebentyayeva, M.T. Dettori, J. Grimwood, F. Cattonaro, The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution, Nat. Genet. 45 (2013) 487–494.
[54] I. Verde, A.G. Abbott, S. Scalabrin, S. Jung, S. Shu, F. Marroni, T. Zhebentyayeva, M.T. Dettori, J. Grimwood, F. Cattonaro, The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution, Nat. Genet. 45 (2013) 487–494.
[55] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[56] D.W. Barnett, E.K. Garrison, A.R. Quinlan, M.P. Strömberg, G.T. Marth, BamTools: a C++ API and toolkit for analyzing and managing BAM files, Bioinformatics 27 (2011) 1691–1692.
[57] T. Brůna, A. Lomsadze, M. Borodovsky, GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins, NAR Genomics Bioinformat. 2 (2020) 1–14.
[58] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[59] A. Lomsadze, V. Ter-Hovhannisyan, Y.O. Chernoff, M. Borodovsky, Gene identification in novel eukaryotic genomes by self-training algorithm, Nucleic Acids Res. 33 (2005) 6494–6506.
[60] H. Iwata, O. Gotoh, Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features, Nucleic Acids Res. 40 (2012) 1–9.
[61] O. Gotoh, M. Morita, D.R. Nelson, Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment, Bmc Bioinform. 15 (2014) 1–13.

[62] M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding, Bioinformatics 24 (2008) 637–644.

[63] M. Stanke, O. Schöffmann, B. Morgenstern, S. Waack, Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources, Bmc Bioinform. 7 (2006) 1–11.

[64] E.V. Kriventseva, D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F.A. Simão, E. M. Zdobnov, OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs, Nucleic Acids Res. 47 (2019) 807–811.

[65] J. Keilwagen, F. Hartung, J. Grau, GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data, in: Gene Prediction, Springer, 2019, pp. 161–177.

[66] P.J.A. Cock, B.A. Grüning, K. Paszkiewicz, L. Pritchard, Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology, PeerJ 1 (2013) 1–22.

[67] E.M. Zdobnov, R. Apweiler, InterProScan: protein domains identifier, Bioinformatics (Oxford, Engl.) 17 (2001) 847–848.

[68] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier, Nucleic Acids Res. 33 (2005) 116–120.

[69] S. Hunter, R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, InterPro: the integrative protein signature database, Nucleic Acids Res. 37 (2009) 211–215.

[70] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq–versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (2017) 6–11.

[71] Y.-X. Yang, M.-H. Tian, X.-H. Liu, Y. Li, Z.-S. Sun, Complete chloroplast genome of Prunus fruticosa and its implications for the phylogenetic position within Prunus sensulato (Rosaceae), Mitochondrial DNA Part B 5 (2020) 3624–3626.

[72] M. Yan, X. Zhang, X. Zhao, Z. Yuan, The complete mitochondrial genome sequence of sweet cherry (*Prunus avium* cv.'summit'), Mitochondrial DNA Part B 4 (2019) 1996–1997.

[73] M. Imakaev, G. Fudenberg, R.P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, L.A. Mirny, Iterative correction of hi-C data reveals hallmarks of chromosome organization, Nat. Methods 9 (2012) 999–1003.

[74] A. Thrash, F. Hoffmann, A. Perkins, Toward a more holistic method of genome assembly assessment, Bmc Bioinform. 21 (2020) 1–8.

[75] O.K. Tørresen, B. Star, P. Mier, M.A. Andrade-Navarro, A. Bateman, P. Jarnot, A. Gruca, M. Grynberg, A.V. Kajava, V.J. Promponas, Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases, Nucleic Acids Res. 47 (2019) 10994–11006.

[76] Y.-T. Huang, P.-Y. Liu, P.-W. Shih, Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing, Genome Biol. 22 (95) (2021) 1–17, https://doi.org/10.1186/s13059-021-02282-6.

[77] T.W. Wöhner, O.F. Emeriewen, A.H.J. Wittenberg, H. Schneiders, I. Vrijenhoek, J. Halász, K. Hrotkó, K.J. Hoff, L. Gabriel, J. Keilwagen, T. Berner, M. Schuster, A. Peil, J. Wünsche, S. Kropop, H. Flachowsky, Supporting Materials for - The Draft Chromosome-level Genome Assembly of Tetraploid Ground Cherry (*Prunus fruticosa* Pall.) from Long Reads. https://www.openagrar.de/receive/openagrar_mods_00070329, 2021 (accessed 1 June 2021).