Biologia Futura
https://doi.org/10.1007/s42977-024-00240-4

ORIGINAL PAPER

Characterisation of the complete chloroplast genome of *Solanum tuberosum* cv. White Lady

Krisztián Frank[1] · **Erzsébet Nagy**[2] · **János Taller**[2] · **István Wolf**[1] · **Zsolt Polgár**[1,3]

Received: 5 March 2024 / Accepted: 24 August 2024

Abstract

Potato (*Solanum tuberosum*) is considered worldwide as one of the most important non-cereal food crops. As a result of its adaptability and worldwide production area, potato displays a vast phenotypical variability as well as genomic diversity. Chloroplast genomes have long been a core issue in plant molecular evolution and phylogenetic studies, and have an important role in revealing photosynthetic mechanisms, metabolic regulations and the adaptive evolution of plants. We sequenced the complete chloroplast genome of the Hungarian cultivar White Lady, which is 155 549 base pairs (bp) in length and is characterised by the typical quadripartite structure composed of a large- and small single-copy region (85 991 bp and 18 374 bp, respectively) interspersed by two identical inverted repeats (25 592 bp). The genome consists of 127 genes of which 82 are protein-coding, eight are ribosomal RNAs and 37 are transfer RNAs. The overall gene content and distribution of the genes on the White Lady chloroplast was the same as found in other potato chloroplasts. The alignment of *S. tuberosum* chloroplast genome sequences resulted in a highly resolved tree, with 10 out of the 13 nodes recovered having bootstrap values over 90%. By comparing the White Lady chloroplast genome with available *S. tuberosum* sequences we found that gene content and synteny are highly conserved. The new chloroplast sequence can support further studies of genetic diversity, resource conservation, evolution and applied agricultural research. The new sequence can support further potato genetic diversity and evolutionary studies, resource conservation, and also applied agricultural research.

Keywords cpDNA · NGS · Plastome · Potato · Sequencing

Introduction

The chloroplast is a semi-autonomous intracellular organelle of photosynthesising organisms. Its genome encodes several genes for the chloroplast components and photosynthesis system. Chloroplast genomes vary in size but are highly conserved in gene content and organisation in vascular plants. For most plants, the chloroplast genome is between 72 and 220 kb in size and contains around 130 genes (Palmer and Stein 1986; Ghimiray and Sharma 2014; Huang et al. 2019). The typical angiosperm chloroplast genome is circular and has a quadripartite structure composed of a pair of inverted repeats (IRa and IRb) separated by a large single-copy (LSC) and a small single-copy (SSC) region (Chung et al. 2006; Samson et al. 2007; Liu et al. 2019; Guzmán-Díaz et al. 2022). Although the general structure of the chloroplast genome is relatively uniform and conserved, several mutations, duplications, insertions/deletions, inversions and rearrangements have been observed (Chung et al. 2006; Ghimiray and Sharma 2014; Amiryousefi et al. 2018; Fang et al. 2021).

Chloroplast genomes have long been a core issue in plant molecular evolution and phylogenetic studies (Palmer and Stein 1986; Huang et al. 2019; Guzmán-Díaz et al. 2022). The advent of high-throughput sequencing technologies has enabled the efficient generation of large amounts of sequence data at a relatively low cost, from which complete organellar genome sequences can be obtained (Frank et al. 2016; Nagy et al. 2017; Guzmán-Díaz et al. 2022). Thus, more and

✉ Krisztián Frank
frank.krisztian@uni-mate.hu

[1] Potato Research Centre, MATE Agrárcsoport Kft., Keszthely, Hungary

[2] Festetics Bioinnovation Group, Institute of Genetics and Biotechnology, Hungarian University of Agricultural and Life Sciences, Keszthely, Hungary

[3] Department of Agronomy, Hungarian University of Agricultural and Life Sciences, Georgikon Campus, Keszthely, Hungary

Published online: 09 September 2024

more chloroplast genomes have been sequenced and reported (Nagy et al. 2017; Amiryousefi et al. 2018; Chen et al. 2021; Fang et al. 2021; Li et al. 2021b; Zhang et al. 2021; Yin et al 2022; Yin and Gao 2023). In addition to exploring phylogenetic relationships, the chloroplast genome has an important role in revealing photosynthetic mechanisms, metabolic regulations and the adaptive evolution of plants (Chung et al. 2006; Daniell et al. 2006; Samson et al. 2007; Amiryousefi et al. 2018) and even some implications in applied agricultural sciences (Cho et al. 2016; Nagy et al. 2017).

The genus *Solanum* is one of the largest genera of flowering plants and includes several economically important food, ornamental and medicinal plants, as well as various species used as biological model systems (Chung et al. 2006; Daniell et al. 2006; Amiryousefi et al. 2018). Potato (*Solanum tuberosum*) is considered one of the most important non-cereal food crops (FAO 2021) with high nutritional value, adaptability and large yield (Ábrahám and Sárvári 2006; Flis et al. 2014; Tömösközi-Farkas et al. 2016). Being an important crop, there are large efforts put into the research of potatoes, including the molecular basis of biotic and abiotic stress, like virus resistance, insect deterrence and drought tolerance (Cho et al. 2016; Song et al. 2021; Islam et al. 2024) as well as their nutrient content (Tömösközi-Farkas et al. 2016). Traditional plant breeding has resulted in numerous potato cultivars with increased yield, disease and pest resistance, environmental stress tolerance and other favourable agronomic attributes. As a result of its adaptability and worldwide production area, potato displays a vast phenotypical variability as well as genomic diversity (Tang et al. 2022). Biotic and abiotic conditions in lowlands of Central Europe are different from conditions of the South American centre of domestication, thus a major issue is the breeding of varieties suitable for such agro-ecological conditions. The main focus is on resistance against major pests, and abiotic stress (heat and drought) tolerance (Polgár et al. 2016). The cultivar White Lady is the result of decades-long breeding efforts at the Potato Research Centre at Keszthely. White Lady has short oval, tubers with light beige skin and cream-coloured flesh, a relatively high dry matter and starch content (Ábrahám and Sárvári 2006; Kollaricsné Horváth et al. 2019). One main breeding goal was to acquire resistance against major pathogens. For this reason, compatible resistant wild species like *Solanum acaule*, *S. andigenum*, *S. demissum*, *S. stoloniferum* and *S. vernei* which convey different resistance genes were included in the pedigree of White Lady (Polgár et al. 2016), and at least *S. stoloniferum* was used as female partner in the pedigree. Such an approach is a common strategy in the breeding of *Solanum* species (Cho et al. 2016; Li et al. 2021a; Ortiz et al. 1994; Yin et al. 2022).

In this study, we present the complete chloroplast genome sequence of the potato cultivar White Lady using high-throughput sequencing. Specifically, we used Illumina sequencing technology to assemble the chloroplast sequence. The assembled sequence is used for annotation, and to explore the gene content and structure. We also compare the gene order and examine the variation of structural changes across the species. We presume that the analysis of complete chloroplast genome sequences is useful for evolutionary and phylogenetic studies, and for exploring structural differences between and among species.

Materials and methods

DNA was extracted from freshly collected leaf tissue with NucleoSpin Plant II Mini Kit (Macherey–Nagel, Germany). As part of an ongoing whole-genome sequencing study, four paired-end libraries with an insert size of ~ 400 bp were prepared and sequenced on a NextSeq 500 (Illumina, USA) platform using a High-output 300 v2.5 sequencing kit.

Chloroplast genome assembly was carried out on a Unix platform running Ubuntu v18.04.01. Quality check of the 2×150 bp paired-end reads was performed using FastQC v0.11.9 and Trimmomatic v0.32 (Bolger et al. 2014) was used to filter low-quality reads and perform the adapter trimming, applying a sliding window of 4:15 and removing reads shorter than 60 bases.

NOVOPlasty v4.3.1 (Dierckxsens et al. 2017) assembler was used to perform the de novo assembly using the trimmed reads. The assembly pipeline was initiated by a seed sequence, to retrieve one read from the sequencing data and then iteratively extended bidirectionally by recruiting new reads. The seed used for the assembly was the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL) gene sequence included in the potato chloroplast reference genome under accession number NC_008096.2 available at GenBank. The start and end of the seed sequence are scanned for overlapping reads, but only one sequence read is retrieved at a time by the programme to form the assembly. NOVOPlasty does not try to assemble every read, but will extend the given seed until a circular genome is formed.

The newly assembled sequence was annotated with a two-step procedure. First, we used the web-based AGORA (Annotator for Genes of Organelle from the Reference Sequence Analysis) application (Jung et al. 2018), CPGAVAS2 (Liu et al. 2012) and GeSeq (Tillich et al. 2017) to obtain annotations based on different approaches. The tRNA genes were identified additionally by tRNAscan-SE v2.4.4 (Chan et al. 2021). Then, the annotations were manually compared and confirmed; the genes present in the result of multiple annotators were included in the final annotation. The resulting genome map was generated with the help of CPGView (Liu et al. 2023). Short tandem repeats (STRs) were identified with MISA v2.1 (Beier et al. 2017) using a threshold level of five for mononucleotide repeats and three

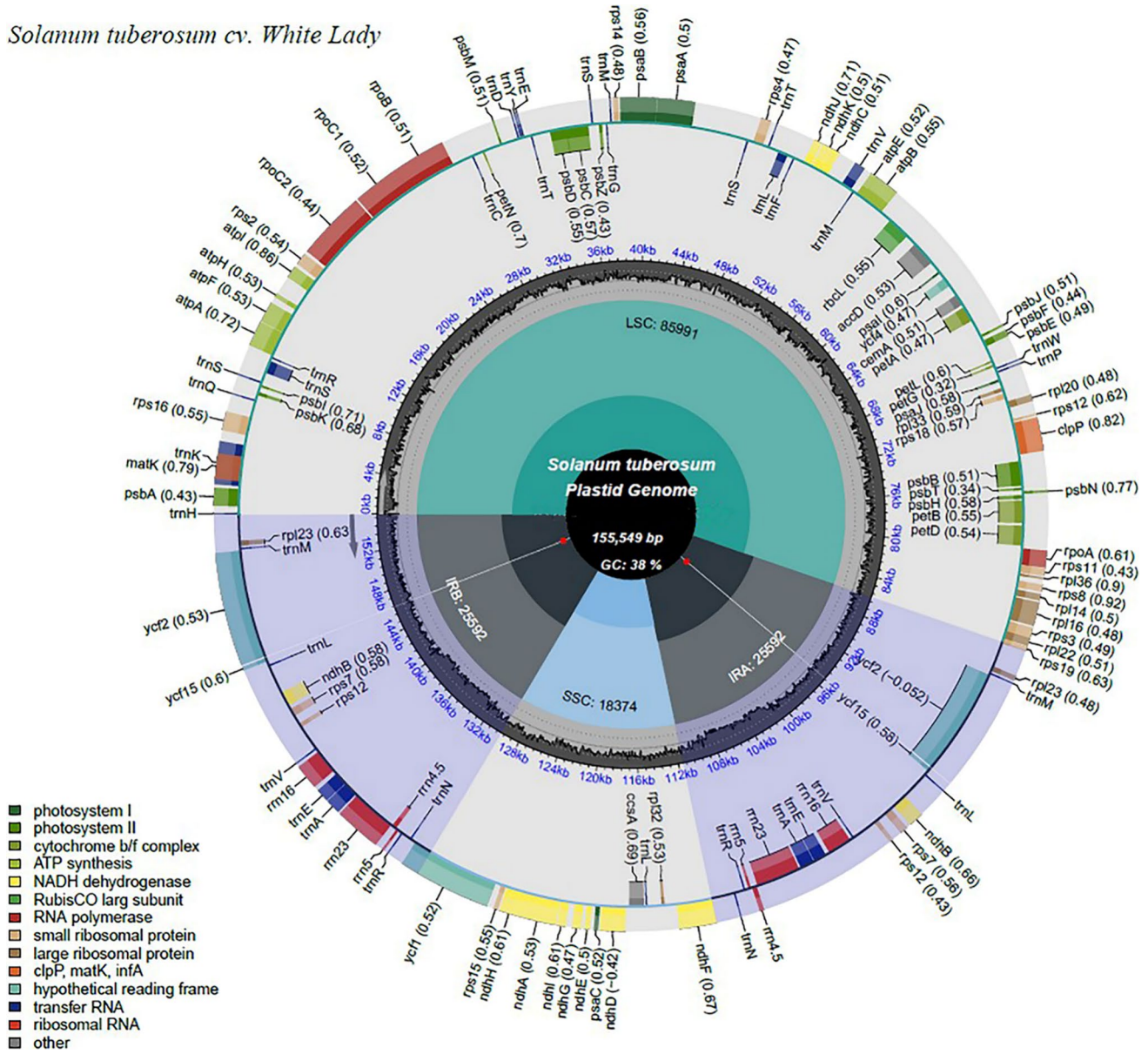

Fig. 1 The circular map of the chloroplast genome of *Solanum tuberosum* cv. White Lady produced by CPGView. Genes are plotted in the outer circle, with the functional classification of the genes shown at the left bottom. The grey inner circle indicates the GC content. The coloured inner circle indicates the quadripartite structure, which consists of the large single-copy (LSC), the small single-copy (SSC) regions and two inverted repeat regions (IRA and IRB)

for di-, tri-, tetra-, penta- and hexanucleotide repeats. The complete chloroplast genome sequence has been deposited into the GenBank (under accession PP432013.1) with assembly metadata and associated feature descriptors. Codon frequency and relative synonymous codon usage (RSCU) were calculated based on protein-coding genes using MEGA11 v11.0.13 (Tamura et al. 2021).

For a phylogenetic analysis, we downloaded all available 53 *Solanum tuberosum* chloroplast genome sequences from the NCBI GenBank (data present on 8. January 2024). We aligned the complete chloroplast genomes with MAFFT v7.310 (Katoh and Standley 2013). A maximum likelihood (ML) tree was constructed using IQ-TREE v1.6.1 (Nguyen et al. 2015) with 1000 bootstraps, with the chloroplast genome sequence of *Capsicum annuum* L. (NC_018552) as an outgroup, because *C. annuum* was the closest sister taxon of the genus *Solanum* with an annotated chloroplast genome (Särkinen et al. 2013).

Table 1 The insertion/deletion (indel) differences found between the *Solanum tuberosum* cv. White Lady and the potato chloroplast reference sequence NC_008096.2. Indels marked with an asterisk are identified as repeats

Indel	Position	Length	In White Lady	Region
Indel01*	6395	17 bp	Insertion	LSC
Indel02	14,618	8 bp	Insertion	LSC
Indel03	52,579	241 bp	Insertion	LSC
Indel04	60,667	8 bp	Insertion	LSC
Indel05	62,636	18 bp	Deletion	LSC
Indel06	78,310	8 bp	Deletion	LSC
Indel07	114,802	8 bp	Deletion	SSC
Indel08*	115,476	9 bp	Insertion	SSC

Asterisks mark indels identified as repeats

Results

Genome sequencing provided more than 1×10^9 reads that aligned to the chloroplast, resulting in an average 65 262 × coverage. The complete chloroplast sequence of White Lady was 155 549 base pairs (bp) in length (Fig. 1) and displayed the typical quadripartite structure of angiosperm plastid genomes, consisting of an LSC (85 991 bp) and SSC (18 374 bp) region, separated by a pair of identical IRs (25 592 bp in length). The overall GC content of the White Lady chloroplast genome was 37.9%, resembling other potato chloroplasts. There were 173 sequence differences found between the White Lady sequence and the potato chloroplast reference sequence NC_008096.2 (Table S1). Most of these, 148 (85.6%), are single nucleotide polymorphisms (SNPs). Eight insertions/deletions (indels) were also found, ranging from 8 to 241 bp in length (Table 1), with all indels and repeat sequences found in intergenic regions. There were also 1864 short tandem repeats (STRs) in the White Lady chloroplast genome (Table 2). The far most abundant STR motifs were poly-A/T stretches (1246, 66.85%), which is characteristic of angiosperm plastids.

The annotation of the new sequence resulted in a total of 127 genes (Table 3), of which 82 corresponded to protein-coding genes, eight to ribosomal RNAs (rRNAs) and 37 to transfer RNAs (tRNAs). Of the genes, 90 were unique, 15 appeared to be duplicated, one triplicated and one

Table 2 Count of simple sequence repeat (SSR) types (mono-, di-, tri-, tetra- and pentanucleotides) found in the *Solanum tuberosum* cv. White Lady chloroplast genome

Repeat	3	4	5	6	7	8	9	10	11	12	13	14	Total
Mono	–	–	852	326	153	39	25	19	10	6	2	2	1434
Di	314	31	4	2									351
Tri	69	2											71
Tetra	8												8
Penta	1												1

Table 3 The gene composition of the *Solanum tuberosum* cv. White Lady chloroplast genome with functional classification. Genes marked with an asterisk were present in more than one copy

Category	Group of genes	Genes
Genes for photosynthesis	ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
	Cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*
	NADH dehydrogenase	*ndhA, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbM, psbN, psbT, psbZ*
	Subunit of RuBisCo	*rbcL*
Self-replication	DNA-dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
	Large subunit of ribosome	*rpl14, rpl16, rpl20, rpl22, rpl23*, rpl32, rpl33, rpl36*
	Small subunit of ribosome	*rps2, rps3, rps4, rps7*, rps8, rps11, rps12*, rps14, rps15, rps16, rps18, rps19*
Other	Other genes	*accD, ccsA, cemA, clpP, matK*
	Hypothetical open reading frames	*ycf1, ycf2*, ycf4, ycf15**
RNA genes	Transfer RNAs	*trnA-UGC*, trnC-GCA, trnD-GUC, trnE-UUC*, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU, trnL-CAA*, trnL-UAA, trnL-UAG, trnM-CAU*, trnN-GUU*, trnP-UGG, trnQ-UUG, trnR-ACG*, trnR-UCU, trnS-CGA, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC*, trnV-UAC, trnW-CCA, trnY-GUA*
	Ribosomal RNAs	*rrn4.5*, rrn5*, rrn16*, rrn23**

Asterisks mark genes with multiple copies

Table 4 The lengths of introns and exons for the intron-containing genes of the *Solanum tuberosum* cv. White Lady chloroplast genome

Gene	Strand	Start	End	Length (bp)				
				Exon I	Intron I	Exon II	Intron II	Exon III
trnK-UUU	–	1795	4380	37	2513	36		
rps16	–	5043	6164	40	855	227		
trnS-CGA	+	9173	9938	31	675	60		
atpF	–	11,924	13,171	152	690	406		
rpoC1	–	21,221	24,024	453	737	1614		
trnL-UAA	+	48,458	49,039	35	497	50		
trnV-UAC	–	52,937	53,580	36	552	56		
rps12	–	71,717	141,615	114	–	243		
clpP	–	71,963	73,968	71	789	294	620	232
petB	+	76,915	78,309	6	747	642		
petD	+	78,504	79,725	6	739	477		
rpl16	–	83,028	84,446	9	1014	396		
trnE-UUC	+	103,820	104,613	32	722	40		
trnA-UGC	+	104,678	105,560	37	810	36		
ndhA	–	121,177	123,426	553	1158	539		
trnA-UGC	–	135,921	136,803	37	810	36		
trnE-UUC	–	136,868	137,661	32	722	40		

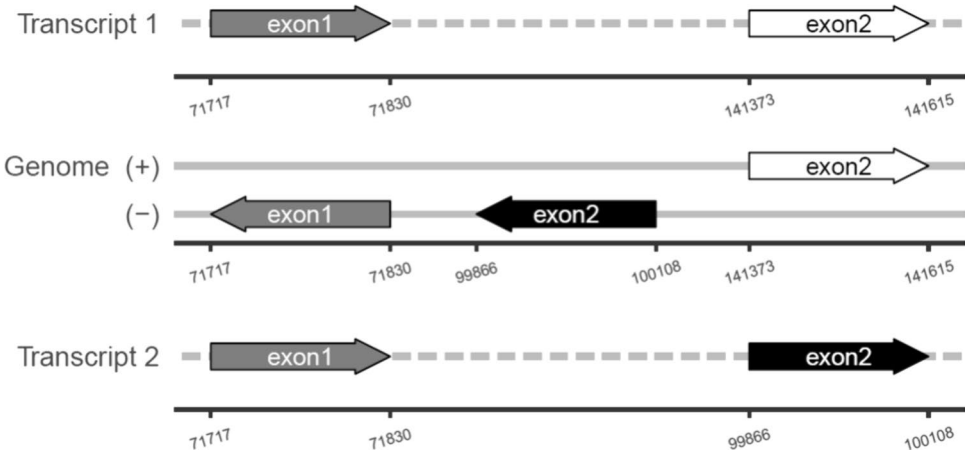

Fig. 2 Schematic representation of the transcription of the *rps12* gene found in the chloroplast genome of *Solanum tuberosum* cv. White Lady

quadrupled. The duplicated genes were the NADH dehydrogenase subunit *ndhB*, two open reading frames (*ycf2* and *ycf15*), three ribosomal proteins (*rps7*, *rps12* and *rpl23*), all four rRNA subunits (*rrn4.5S*, *rrn5S*, *rrn16S* and *rrn23S*) and five tRNAs (*trnA*-UGC, *trnL*-CAA, *trnN*-GUU, *trnR*-ACG, *trnV*-GAC). The tRNA *trnE*-UUC was triplicated and *trnM*-CAU had four copies. Almost all protein-coding genes had a standard ATG start codon, except *ndhD* starting with AAT and *rps19* starting with GTG. None of the genes had internal or non-standard stop codons.

Eighty percentage of the total length of the genome is related to genetic regions. The ATT codon, coding the amino acid isoleucine, was the most frequent codon with an RSCU rate of 41.304 (Table S2). The distribution of the genes on different regions of the plastid exhibited similarity with other *Solanum* species with 13 genes in the SSC region and 17 genes in the IR while the rest were in the LSC. Eight tRNAs and nine protein-coding genes of the White Lady chloroplast genome contained introns (Table 4). The gene *rps12* is trans-spliced with Exon I located in the LSC region and the duplicated Exon II located in the IR regions (Fig. 2). The other genes are cis-spliced, with *clpP* having three exons and two introns and the other genes having two exons separated by a single intron (Fig. 3).

Our phylogenomic analysis of the whole chloroplast genome sequences resulted in a highly resolved tree, with 10 out of the 14 nodes recovered having bootstrap values over 90 (Fig. 4). *S. tuberosum* cv. White Lady belonged to the same node as the cultivars Castle Russet, Meteor, Gusar, Grand, Sudarinya and an unspecified *S. tuberosum* (OR632697), with a bootstrap support of 97.

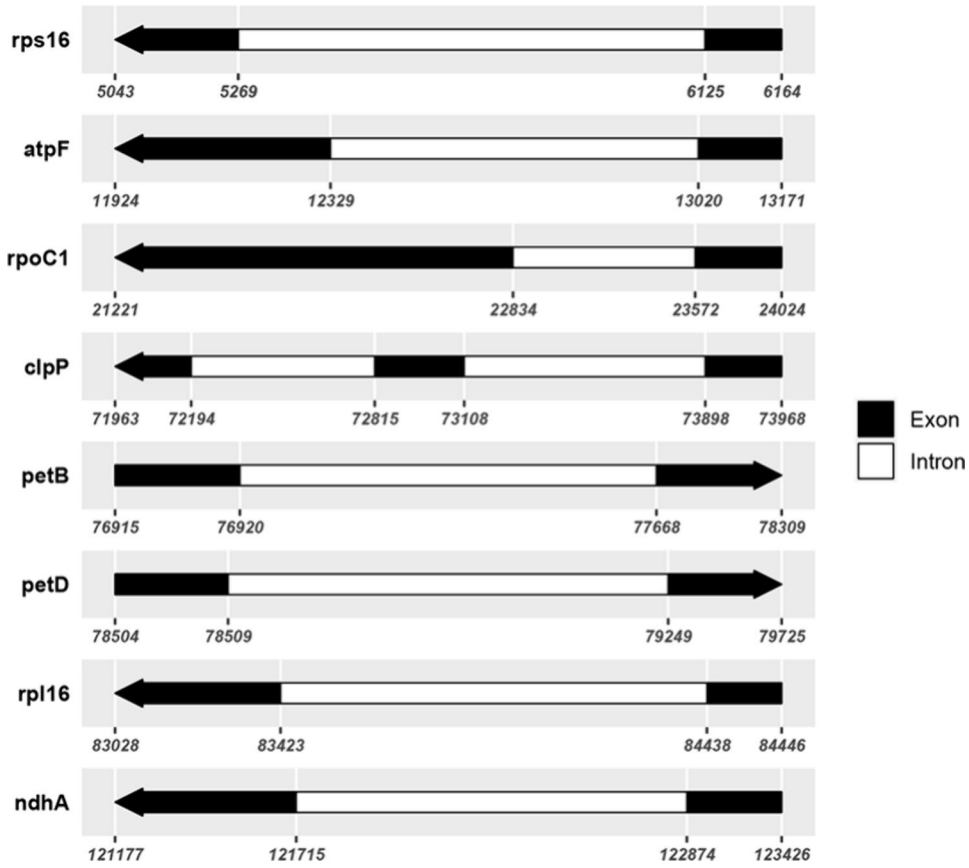

Fig. 3 Schematic representation of the protein-coding genes containing introns found in the chloroplast genome of *Solanum tuberosum* cv. White Lady

Discussion

High-throughput sequencing technologies have been constantly changing biological research with the generation of large amounts of sequence data (Frank et al. 2016, 2017; Nagy et al. 2017; Guzmán-Díaz et al. 2022). These sequences may be used to reconstruct complete chloroplast genomes, which may provide us with the information basis for studying chloroplast evolution and the mechanisms of genomic rearrangements (Chung et al. 2006; Cho et al. 2016; Amiryousefi et al. 2018; Huang et al. 2019). Analysis of the phylogenetic relationships of chloroplast genomes can help understand plant phylogeny and some aspects of their population genetics, or even underline taxonomic status at the molecular level (Daniell et al. 2006; Huang et al. 2019; Liu et al. 2019; Guzmán-Díaz et al. 2022). Furthermore, some herbicides like PSI and PSII inhibitors have their target genes in the chloroplast genome, thus investigation of plastid genome organisation and function could trigger further breakthroughs in applied sciences (Cho et al. 2016; Nagy et al. 2017). The chloroplast genomes of angiosperm plants are generally conservative in sequence and gene content, however, structural variations like gene duplications and sequence rearrangements have been observed across angiosperms (Chung et al. 2006; Samson et al. 2007; Martin et al. 2014; Fang et al. 2021; Guzmán-Díaz et al. 2022). Here, we report the complete chloroplast genome sequence of *Solanum tuberosum* cv. White Lady as a genomic tool for comparative studies.

Unsurprisingly, the new chloroplast genome sequence showed very high similarity to the available potato sequences. The overall gene content and distribution of the genes on the White Lady chloroplast was the same as found in other *Solanum tuberosum* chloroplasts (Chung et al. 2006; Chen et al. 2021; Li et al. 2021b). The White Lady chloroplast genome contains 107 unique genes, 76 protein-coding genes, four rRNAs and 27 tRNAs, which is in line with other potato chloroplasts (Chung et al. 2006; Chen et al. 2021; Li et al. 2021b). The trans-splicing variants of the *rps12* genes, the location of the duplicated exons and the exon–intron boundaries are known in other potato cultivars, and all the protein-coding genes are the same as described in other potato cultivars (Chen et al. 2021; Huang et al. 2019; Li et al. 2021b). Reports on splicing variants in wild *Solanum* species were scarce, but previous studies did not report any differences (Cho et al. 2016; Huang et al. 2019). The low sequence diversity of chloroplast genes in *Solanum* is attributed to the conservation of the functions of the photosynthetic system (Amiryousefi et al. 2018) and contrasted with other economically important plant families

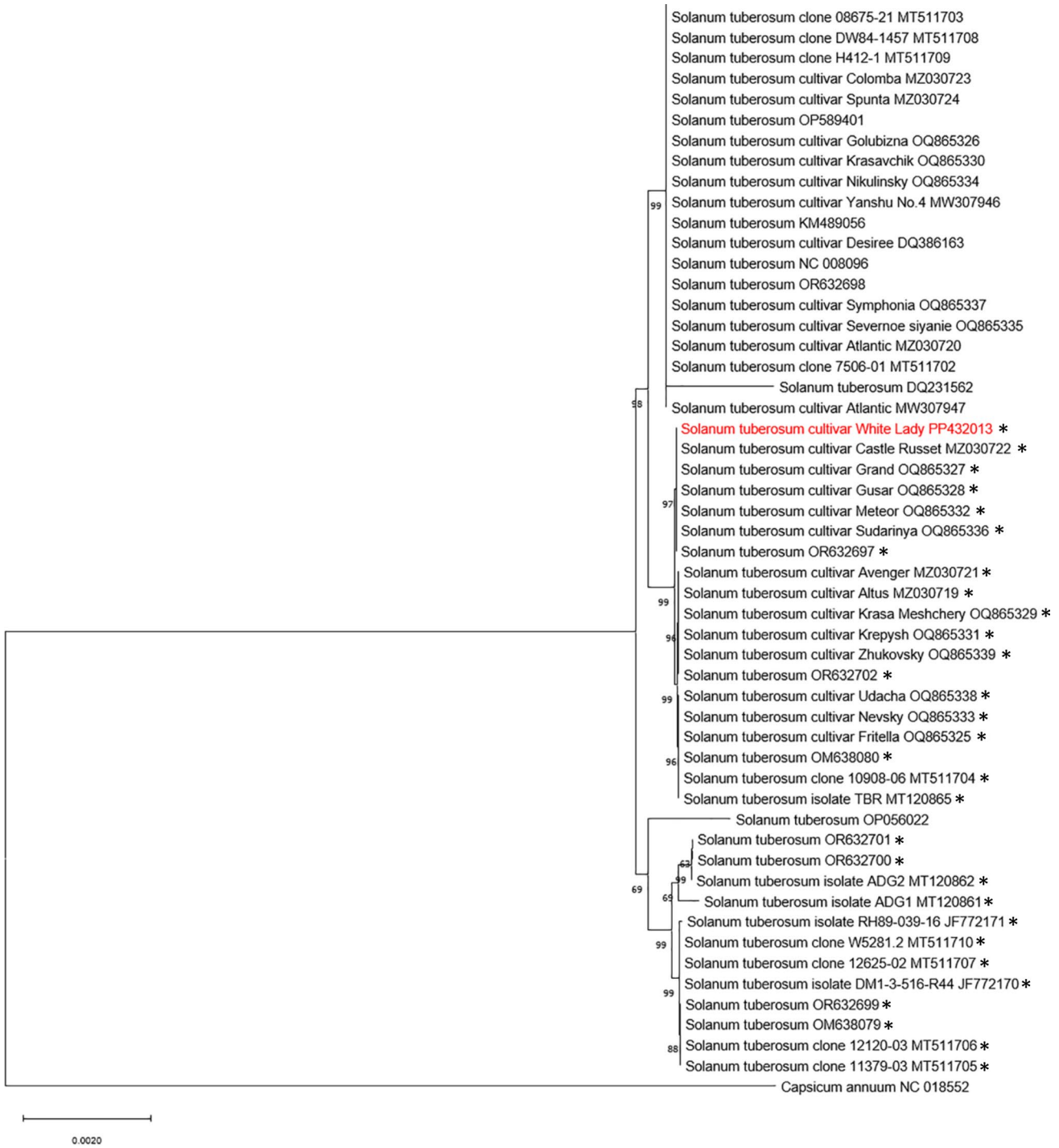

Fig. 4 Maximum-likelihood phylogenetic tree of *Solanum tuberosum* based on chloroplast genomes. The scale bar represents the number of substitutions at each locus; the new sequence of cv. White Lady is denoted with red, and sequences with the 241 bp insertion are marked with an asterisk

where plastid genomes can harbour many structural differences and rearrangements (Samson et al. 2007; Skuza et al. 2023; Yin and Gao 2023).

We found an abundant reservoir of structural differences (SNPs, STRs and InDels) in the chloroplast genome of the potato. Such loci can be used to develop genetic markers for genetic diversity assessment and molecular fingerprinting

(Cho et al. 2016; Daniell et al. 2016; Yan et al. 2022). The cultivated potato is tetraploid, hence multiple copies and heterozygous loci can cause difficulties when using nuclear genetic markers. Compared with nuclear markers, chloroplast DNA markers could make genetic analyses easier and more feasible. The largest structural difference in White Lady was the absence of a 241 bp long deletion in the intergenic region between *ndhC* and *trnV*, which was previously associated with the cultivated potato (Chung et al. 2006). This deletion was thought to represent a genetic difference between cultivated and wild potatoes; however, some potato cultivars carry the deletion, while others, like White Lady, lack this variation. Only 23 of the available 53 *S. tuberosum* chloroplast sequences carried this deletion, while 30 of them and the new White Lady chloroplast harboured the longer sequence variation (marked with an asterisk in Fig. 4). Several wild tuber-bearing *Solanum* species and their hybrids have been used as parents in potato breeding (Ortiz et al. 1994; Cho et al. 2016; Polgár et al. 2016; Li et al. 2021a; Tang et al. 2022) with different degree, this might have had an effect of the presence or absence of this deletion. If the insertion originated in the cultivated potato, using these wild species as maternal partners could have introduced the insertion variant back to cultivated lines, which could happen several times in different locations (Ortiz et al. 1994; Spooner et al. 2005). The presence or absence of the deletion seemingly does not affect the agricultural traits or breeding habits; however, this has not been studied in detail before.

Our phylogenomic analysis of the complete chloroplast genome resulted in a highly resolved tree, with almost all branches having high support values. The cultivar White Lady belonged to the same node as the cultivars Castle Russet, Meteor, Gusar, Grand, Sudarinya and an unspecified *S. tuberosum*. The branch lengths of the phylogenetic tree were extremely short, indicating also very high sequence similarity among the chloroplast sequences of cultivated potato (Huang et al. 2019). Although our phylogenetic tree was highly resolved, phylogenetic relationships of many species in the genus *Solanum* based on chloroplast genomes are often ambiguous or have low support. This reduced structure is most evident in the cultivated potato that forms a nearly complete polytomy (Chung et al. 2006; Huang et al. 2019; Yan et al. 2022). We failed to find any association between the phylogenetic position and cultivar designation of the sequences. Although many *S. tuberosum* chloroplast genomes did not have any associated cultivar or origin metadata in GenBank, the known cultivars did not cluster according to their origin. These discordant results of individual accessions failing to cluster, or having reduced support may reflect real phylogenetic status but may be also supported by the history of numerous hybridisations and introgression in cultivated potatoes. It is important to note that interspecific hybridisation as anthropogenically introduced into *S. tuberosum* may make chloroplast phylogenetic analyses difficult (Amiryousefi et al. 2018; Huang et al. 2019; Yan et al. 2022).

Conclusions for future biology

High-throughput sequencing has become a standard method and enabled the generation of large amounts of sequences. The chloroplast genomic data reported here can be utilised to study chloroplast evolution and phylogenetics on the species, genus, family or other taxonomic levels. The sequence differences, SNPs, STRs and InDels described here can be used to develop molecular markers for the identification of potato cultivars or the study of population-genetic processes in wild *Solanum* species. The comparison of complete chloroplast genome sequences may also give insights into the mechanisms of genomic rearrangements, especially to address phylogenetic and molecular evolutionary questions regarding the breeding of cultivated potatoes. Furthermore, some herbicides like PSI and PSII inhibitors have their target genes in the chloroplast genome, thus investigation of plastid genome organisation and function may lead to applicable results in agriculture.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42977-024-00240-4.

Funding Open access funding provided by Hungarian University of Agriculture and Life Sciences. This study was financed by the National Research, Development and Innovation Office (grant number INN_139994).

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical approval No ethical approval was required for the study.

References

Ábrahám ÉB, Sárvári M (2006) Effect of year and irrigation on the yield and quantity of different potato varieties. Cereal Res Comm 34:369–372. https://doi.org/10.1556/CRC.34.2006.1.92

Amiryousefi A, Hyvönen J, Poczai P (2018) The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS ONE 13:e0196069. https://doi.org/10.1371/journal.pone.0196069

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

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

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res 49:9077–9096. https://doi.org/10.1093/nar/gkab688

Chen S, Zhao Y, Zhang JY, Zhang JY, Wang YP, Mou B, Ma H, Han Z, Lu Y, Li S, Zhao CB, Han YZ (2021) Characterization of the complete chloroplast genome of the *Solanum tuberosum* L. cv. Shepody (Solanaceae). Mitochondrial DNA Part B 6:2342–2344. https://doi.org/10.1080/23802359.2021.1934135

Cho KS, Cheon KS, Hong SY, Cho JH, Im JS, Mekapogu M, Yu YS, Park TH (2016) Complete chloroplast genome sequences of *Solanum commersonii* and its application to chloroplast genotype in somatic hybrids with *Solanum tuberosum*. Plant Cell Rep 35:2113–2123. https://doi.org/10.1007/s00299-016-2022-y

Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu JR (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep 25:1369–1379. https://doi.org/10.1007/s00299-006-0196-4

Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. Theor Appl Genet 112:1503–1518. https://doi.org/10.1007/s00122-006-0254-x

Daniell H, Lin CS, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45:e18. https://doi.org/10.1093/nar/gkw955

Fang S, Zhang L, Qi J, Zhang L (2021) De novo assembly of chloroplast genomes of *Corchorus capsularis* and *C. olitorius* yields species-specific InDel markers. Crop J 9:216–226. https://doi.org/10.1016/j.cj.2020.05.010

FAO (2021) World Food and Agriculture-Statistical Yearbook (2021) Food and Agriculture Organization of the United Nations. Rome Available: https://doi.org/10.4060/cb4477en

Flis B, Domański L, Zimnoch-Guzowska E, Zs P, Pousa SÁ, Pawlak A (2014) Stability analysis of agronomic traits in potato cultivars of different origin. Am J Potato Res 91:404–413. https://doi.org/10.1007/s12230-013-9364-6

Frank K, Barta E, Bana ÁN, Nagy J, Horn P, Orosz L, Stéger V (2016) Complete mitochondrial genome sequence of a Hungarian red deer (*Cervus elaphus hippelaphus*) from high-throughput sequencing data and its phylogenetic position within the family Cervidae. Acta Biol Hung 67:133–147. https://doi.org/10.1556/018.67.2016.2.2

Frank K, Molnár J, Barta E, Marincs F (2017) The full mitochondrial genomes of Mangalica pig breeds and their possible origin. Mitochondrial DNA Part B 2:730–734. https://doi.org/10.1080/23802359.2017.1390415

Ghimiray D, Sharma BC (2014) Comparative and bioinformatics analyses of the Solanaceae chloroplast genomes: plastome organization is more or less conserved at family level. J Appl Biol Biotechnol 2:021–026. https://doi.org/10.7324/JABB.2014.2305

Guzmán-Díaz S, Núñez FAA, Veltjen E, Asselman P, Larridon I, Samain MS (2022) Comparison of Magnoliaceae plastomes: adding Neotropical *Magnolia* to the discussion. Plants 11:448. https://doi.org/10.3390/plants11030448

Huang B, Ruess H, Liang Q, Colleoni C, Spooner DM (2019) Analyses of 202 plastid genomes elucidate the phylogeny of *Solanum* section Petota. Sci Rep 9:4454. https://doi.org/10.1038/s41598-019-40790-5

Islam S, Li J, Rahman MA, Xie F, Song B, Nie B (2024) Resistance to biotic and abiotic stress in potato: the origin of the genes and corresponding molecular markers. Phytopatol Res 6:4. https://doi.org/10.1186/s42483-023-00222-9

Jung J, Kim JI, Jeong YS, Yi G (2018) AGORA: organellar genome annotation from the amino acid and nucleotide references. Bioinformatics 34:2661–2663. https://doi.org/10.1093/bioinformatics/bty196

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

Kollaricsné Horváth M, Hoffmann B, Cernák I, Sz B, Zs P, Taller J (2019) Nitrogen utilization of potato genotypes and expression analysis of genes controlling nitrogen assimilation. Biol Fut 70:25–37. https://doi.org/10.1556/019.70.2019.04

Li D, Gan G, Li W, Li W, Jiang Y, Liang X, Yu N, Chen R, Wang Y (2021a) Inheritance of *Solanum* chloroplast genomic DNA in interspecific hybrids. Mitochondrial DNA Part B 6:351–357. https://doi.org/10.1080/23802359.2020.1866450

Li S, Wang PY, Zhao YF, Zhang JY, Zhang JY, Ma HR, Yue Y, Du CY, Zhao CB, Han YZ (2021b) Characterization of the complete chloroplast genome of the *Solanum tuberosum* L. cv. Favorita (Solanaceae). Mitochondrial DNA Part B 6:909–911. https://doi.org/10.1080/23802359.2021.1886885

Liu C, Shi L, Chen H, Zhang J, Lin X, Guan X (2012) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genom 13:715. https://doi.org/10.1186/1471-2164-13-715

Liu E, Yang C, Liu J, Jin S, Harijati N, Hu Z, Diao Y, Zhao L (2019) Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. Sci Rep 9:809. https://doi.org/10.1038/s41598-018-37456-z

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

Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, Ferreira de Carvalho J, Aïnouche M, Salmon A, Aïnouche A (2014) The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. Ann Bot 113:1197–1210. https://doi.org/10.1093/aob/mcu050

Nagy E, Hegedűs G, Taller J, Kutasy B, Virág E (2017) Illumina sequencing of the chloroplast genome of common ragweed (*Ambrosia artemisiifolia* L.). Data Brief 15:606–611. https://doi.org/10.1016/j.dib.2017.10.009

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Ortiz R, Iwanaga M, Peloquin SJ (1994) Breeding potatoes for developing countries using wild tuber bearing *Solanum* spp. and ploidy manipulations. J Genet Breed 48:89–98

Palmer JD, Stein DB (1986) Conservation of chloroplast genome structure among vascular plants. Curr Genet 10:823–833

Polgár Zs, Cernák I, Vaszily Zs (2016) Potato breeding, meeting the challenges of climate change. Lucrări Ştiinţifice, Seria Agronomie 59:223–226

Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H (2007) The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. Plant Biotechnol J 5:339–353. https://doi.org/10.1111/j.1467-7652.2007.00245.x

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13:214. https://doi.org/10.1186/1471-2148-13-214

Skuza L, Androsiuk P, Gastineau R, Paukszto Ł, Jastrzębski JP, Cembrowska-Lech D (2023) Molecular structure, comparative and phylogenetic analysis of the complete chloroplast genome sequences of weedy rye *Secale cereale* ssp. segetale. Sci Rep 13:5412. https://doi.org/10.1038/s41598-023-32587-4

Song Y, Feng L, Alyafei MAM, Jaleel A, Ren M (2021) Function of chloroplasts in plant stress responses. Int J Mol Sci 22:13464. https://doi.org/10.3390/ijms222413464

Spooner DM, Nunez J, Rodriguez F, Naik PS, Ghislain M (2005) Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. Theor Appl Genet 110:1020–1026. https://doi.org/10.1007/s00122-004-1917-0

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D, Zhu G, Wang H, Zhou Y, Bryan GJ, Buell CR, Zhang C, Huang S (2022) Genome evolution and diversity of wild and cultivated potatoes. Nature 606:535–541. https://doi.org/10.1038/s41586-022-04822-x

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391

Tömösközi-Farkas R, Adányi N, Gasztonyi-Nagy M, Berki M, Horváth V, Renkecz T, Simon K, Fabulya Z, Zs P (2016) Changes of metabolites and macro-and micro-elements in Hungarian potatoes under organic and conventional farming. J Agricult Sci Technol B. 6:83–92. https://doi.org/10.17265/2161-6264/2016.02.004

Yan LJ, Zhu ZG, Wang P, Fu CN, Guan XJ, Kear P, Zhang CZ, Zhu GT (2022) Comparative analysis of 343 plastid genomes of *Solanum* section *Petota*: Insights into potato diversity, phylogeny, and species discrimination. J System Evol 61:599–612. https://doi.org/10.1111/jse.12898

Yin S, Gao Y (2023) The complete chloroplast genome assembly of *Amorphophallus krausei* Engler, Pflanzenr 1911 (Araceae) from southwestern China. Mitochondrial DNA Part B 8:1339–1342. https://doi.org/10.1080/23802359.2023.2288889

Yin M, Yu Y, Gong Y, Gui M, Li Z, Bao R, Cheng J, Du G, Wu L (2022) The complete chloroplast genome of *Solanum sisymbriifolium* (Solanaceae), the wild eggplant. Mitochondrial DNA Part B 7:886–888. https://doi.org/10.1080/23802359.2022.2077667

Zhang JY, Zhang JY, Zhao YF, Li S, Chen SS, Wang YP, Mou B, Ma HR, Han ZJ, Lu Y, Han YZ (2021) Characterization of the complete chloroplast genome of the *Solanum tuberosum* L. cv. Atlantic (Solanaceae). Mitochondrial DNA Part B 6:73–75. https://doi.org/10.1080/23802359.2020.1845998