data

MDPI

Article

Revealing the Complete Chloroplast Genome of an Andean Horticultural Crop, Sweet Cucumber *(Solanum muricatum)*, and Its Comparison with Other Solanaceae Species

Carla L. Saldaña [1], Julio C. Chávez-Galarza [1,2], Germán De la Cruz [3], Jorge H. Jhoncon [4,5], Juan C. Guerrero-Abad [6], Héctor V. Vásquez [1,7], Jorge L. Maicelo [1,7] and Carlos I. Arbizu [1,*]

1 Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru
2 Departamento Académico de Ciencias Básicas y Afines, Universidad Nacional de Barranca, Av. Toribio de Luzuriaga 376, Urbanización La Florida Mz J, Lima 15169, Peru
3 Laboratorio de Genética y Biotecnología Vegetal, Facultad de Ciencias Agrarias, Universidad Nacional San Cristóbal de Huamanga, Portal Constitución 57, Ayacucho 05003, Peru
4 Centro de Investigación de Plantas Andinas y Nativas, Facultad de Ciencias, Universidad Nacional de Educación Enrique Guzmán y Valle, Av. Enrique Guzmán y Valle s/n, Lima 15472, Peru
5 Unidad de Investigación, Wanxin Group EIRL. Av. La Marina 800, Lima 15084, Peru
6 Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru
7 Facultad de Zootecnia, Agronegocios y Biotecnología, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Chachapoyas 01001, Peru
* Correspondence: carbizu@inia.gob.pe

Abstract: Sweet cucumber *(Solanum muricatum)* sect. *Basarthrum* is a neglected horticultural crop native to the Andean region. It is naturally distributed very close to other two *Solanum* crops of high importance, potatoes, and tomatoes. To date, molecular tools for this crop remain undetermined. In this study, the complete sweet cucumber chloroplast (cp) genome was obtained and compared with seven Solanaceae species. The cp genome of *S. muricatum* was 155,681 bp in length and included a large single copy (LSC) region of 86,182 bp and a small single-copy (SSC) region of 18,360 bp, separated by a pair of inverted repeats (IR) regions of 25,568 bp. The cp genome possessed 87 protein-coding genes (CDS), 37 transfer RNA (tRNA) genes, eight ribosomal RNA (rRNA) genes, and one pseudogene. Furthermore, 48 perfect microsatellites were identified. These repeats were mainly located in the noncoding regions. Whole cp genome comparative analysis revealed that the SSC and LSC regions showed more divergence than IR regions. Similar to previous studies, our phylogenetic analysis showed that *S. muricatum* is a sister species to members of sections *Petota* + *Lycopersicum* + *Etuberosum*. We expect that this first sweet cucumber chloroplast genome will provide potential molecular markers and genomic resources to shed light on the genetic diversity and population studies of *S. muricatum*, which will allow us to identify varieties and ecotypes. Finally, the features and the structural differentiation will provide us with information about the genes of interest, generating tools for the most precise selection of the best individuals of sweet cucumber, in less time and with fewer resources.

Keywords: chloroplast; genome; sweet cucumber; Solanaceae; next-generation sequencing

1. Introduction

Solanum muricatum (*Solanum* section *Basarthrum*), known also as "sweet cucumber" or "pepino dulce", is an Andean domesticated crop [1,2], which is closely related to *S. tuberosum* (potato) and *S. lycopersicum* (tomato). Indeed, they all share the same center of origin. Likewise, it is related to its wild relatives *S. perlongistylum* and *S. catilliflorum*, which are endemic species of Peru [3]. In Peru, sweet cucumber is commonly cultivated by

peasants both in the valleys that are irrigated by the rivers of the desert coastline and in the inter-Andean valleys [4] in many regions especially in Ayacucho, Chimbote, Chiclayo, Lima, Arequipa and Trujillo [5]. It is also found in many American countries, such as the USA, Mexico, Colombia, and Chile. Sweet cucumber is also cultivated in Spain, Australia, New Zealand [2,6,7], Morocco, Israel, and Kenya, diversifying horticultural productions. Reports of *S. muricatum* on the coast and valleys of the Andean highlands of Peru are documented since 1549, forming part of the diet of ancient Peruvians [8]. Nutritionally, fruits of sweet cucumber have low caloric content and are characterized by their very high levels of vitamin C and potassium [9], conferring antidiabetic [10], antitumor, anti-inflammatory [11], and antihyperlipidemic properties [12]. Studies have shown that fruits of *S. muricatum* contain considerable levels of flavonoids and phenols and have high antioxidant capacity due to its activity for the elimination of free radicals [10,13].

The genus *Solanum* contains around 1500 species worldwide [14] and includes very important food crops, as well as most annual or perennial plants cultivated for their ornamental flowers and fruit [15]. Chloroplasts play vital metabolic roles including photosynthesis, amino acid and lipid synthesis [16]. Numerous important events take place in chloroplast genomes: deletions (InDels), insertions, inversions, substitutions, genome rearrangements, and translocations [16–18]. Consequently, chloroplast genome sequences are useful for phylogenetic analyses [19], evolution [20], population structure [21], and genetic engineering [22] studies. Currently, the cp genome of various Solanaceae species has been sequenced, e.g., *S. tuberosum* by Chung et al. [23]. It consists of 155,312 bp and two inverted repeat regions of 25,595 bp. The IR regions are separated by SSC and LSC regions of 18,373 and 85,749 bp, respectively. This cp genome encodes 30 tRNAs, 4 rRNAs and 79 proteins. Daniell et al. [24] compared the cp genomes of *S. lycopersicum*, *S. tuberosum*, *Nicotiana tabacum*, and *Atropa belladonna*. They revealed deletions and insertions, indicating that some genes (*clpP, cemA, ccsA,* and *matK*) for photosynthesis are the most divergent. These results confirmed that *S. tuberosum* and *S. lycopersicum* sequences presented minor number of C-to-U changes, as previously reported by Schmitz-Linneweber et al. [25]. Moreover, *N. tabacum* and *A. belladonna* cp genome did not present C-to-U conversions, but they were observed in *S. tuberosum* and *S. lycopersicon*. Moreover, Mehmood et al. [26] revealed the cp genome of five species of *Nicotiana* and found the distribution and location of repeated sequences as well as divergences of the IR region sequences, SSC and LSC. Another member of the Solanaceae is *S. dulcamara* whose chloroplast genome length is 155,580 bp with a characteristic structure composed of a large and small single-copy region (85,901 bp and 18,449 bp, respectively) separated by a pair of inverted repeats (25,615 bp) [20]. This genome encodes for 112 unique genes, including 27 tRNA, four rRNA and 81 protein-coding genes. On the other hand, Arbizu et al. [27] reported for the first time the complete chloroplast genome of native chili pepper, *Capsicum chinense*, whose genome possessed a 156,931 bp in length with two IR regions (25,847 bp) separated a LSC region (87,325 bp) and a SSC region (17,912 bp). The content of GC was 37.71%. They indicated that the cp genome of this Peruvian landrace encodes for 133 in total (86 protein-coding genes, eight rRNA, 37 tRNA and two pseudogenes).

Even though cp genomes for multiple *Solanum* species were deciphered employing next-generation sequencing techniques, some other groups of plants have not received the same attention. Hence, molecular tools were not developed for them. To date, *S muricatum* is still considered a neglected Andean crop [9] since genomic tools for sweet cucumber remain still scarce. We sequenced for the first time the complete cp genome of this orphan crop and compared it with other eight important species within the Solanaceae family. This study revealed the genome organization in the sweet cucumber cp genome, as well as information concerning its phylogeny, which will generate advances in genetic and evolutionary studies for this Andean horticultural crop.

2. Results

2.1. S. muricatum cp Genome Assembly

The complete cp sequence was deposited into the database of GenBank with accession number: OK326864 (https://www.ncbi.nlm.nih.gov/nuccore/OK326864.1/ accessed on 1 March 2022). The associated Bioproject, Biosample and SRA numbers are PRJNA742505, SAMN19957695, SRX11310971, respectively. *S. muricatum* cp genome was 155,861 bp in length and exhibited the typical circular tetrad structure and consisted of a pair of IR regions of 25,568 bp separated by a LSC region of 86,182 bp and a SSC region of 18,360 bp (Figure 1). The percentage of GC of the IR region (43.03%) was higher than that of the LSC (35.90%) and SSC regions (32.02%). The GC abundance in the structural regions of the cp was repeated in the other seven species (Table 1). In the *S. muricatum* cp genome, 133 genes were predicted in total; 87 were protein-coding genes, 37 tRNAs, eight rRNAs, and one pseudogene. A total of 115 were unique. (Table S1). In the IR regions, seven tRNAs, four rRNAs, and seven protein coding genes were duplicated. The protein-coding genes included nine genes that encode large ribosomal proteins; *rpl2* and *rpl23* genes had two copies in the IR region. In addition, *rpl2* and *rpl16* genes had one intron. There were 12 small ribosomal protein genes; two of them (*rps7* and *rps12*) showed two copies. *Rps12* gene was separated into two independent transcription units. Furthermore, the cp genome of sweet cucumber possessed five genes that encoded photosystem I components, and 15 genes were related to photosystem II. Adenosine triphosphate (ATP) synthase and electron transport chain component were encoded by six genes (Table S1).
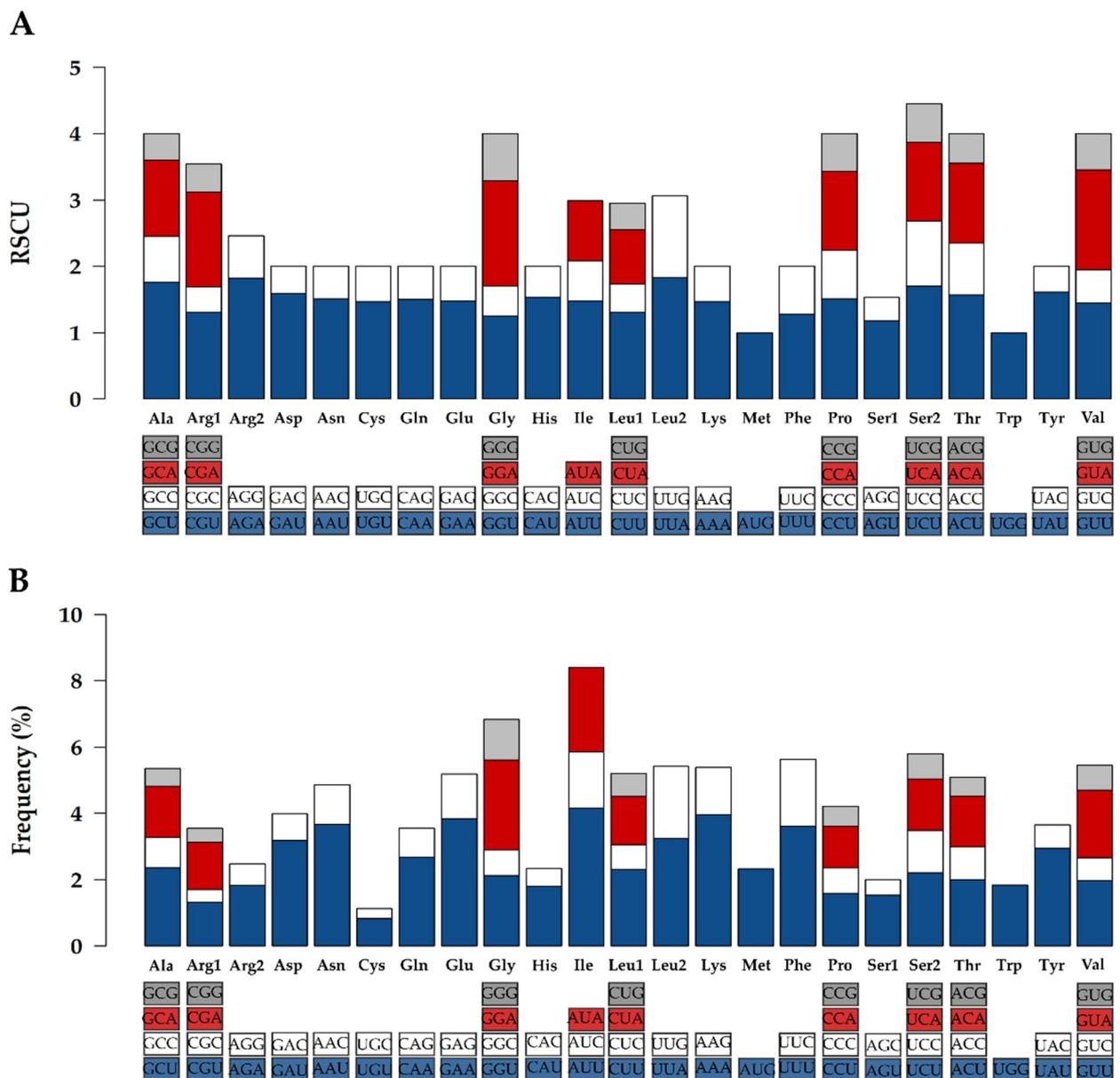
Table 1. Comparison of the cp genomes of *S. muricatum* and seven Solanaceae species.

Genome Characteristics	Solanum muricatum	S. bulbocastanum	S. dulcamara	S. etuberosum	S. lycopersicum	S. peruvianum	S. phureja	S. tuberosum
Genome size (bp)	155,681	155,371	155,580	155,302	155,461	155,561	155,492	155,296
SSC length (bp)	18,360	18,380	18,448	18,357	18,362	18,376	18,375	18,372
LSC length (bp)	86,182	85,814	85,901	85,758	85,874	85,906	85,930	85,737
IRA length (bp)	25,568	25,587	25,614	25,592	25,611	25,638	25,592	25,592
IRB length (bp)	25,568	25,587	25,614	25,592	25,611	25,638	25,592	25,592
No. of different protein-coding genes	88	88	88	88	88	88	88	88
No. of different rRNA genes	4	4	4	4	4	4	4	4
No. of tRNA genes	37	37	37	37	37	37	37	37
No. of different genes	117	117	117	117	117	117	117	117
%GC content in LSC	35.90	36.01	36.01	36.07	35.99	35.98	35.07	36.01
%GC content in SSC	32.02	32.13	32.07	32.25	32.02	32.02	32.10	32.09
%GC content in IR	43.03	43.06	43.06	43.09	43.09	43.03	43.09	43.10

A total of 26,235 codons were identified (Table S2). The most abundant amino acid was Leucine (Leu) followed by isoleucine (Ile). Cysteine (Cys) was present in less quantity. Moreover, for tryptophan (Trp) and methionine (Met) amino acids, only one codon was recognized. Thirty-one codons revealed a RSCU < 1, and the third position of the biased codons were A/U, and 29 codons were detected to be used more frequently than the expected usage at equilibrium (RSCU > 1). Biased codons Arg (AGA), Ser (UCU), Ala (GCU), Tyr (UAU), and Asp (GAU) presented the highest RSCU values (Figure 2, Table S2).
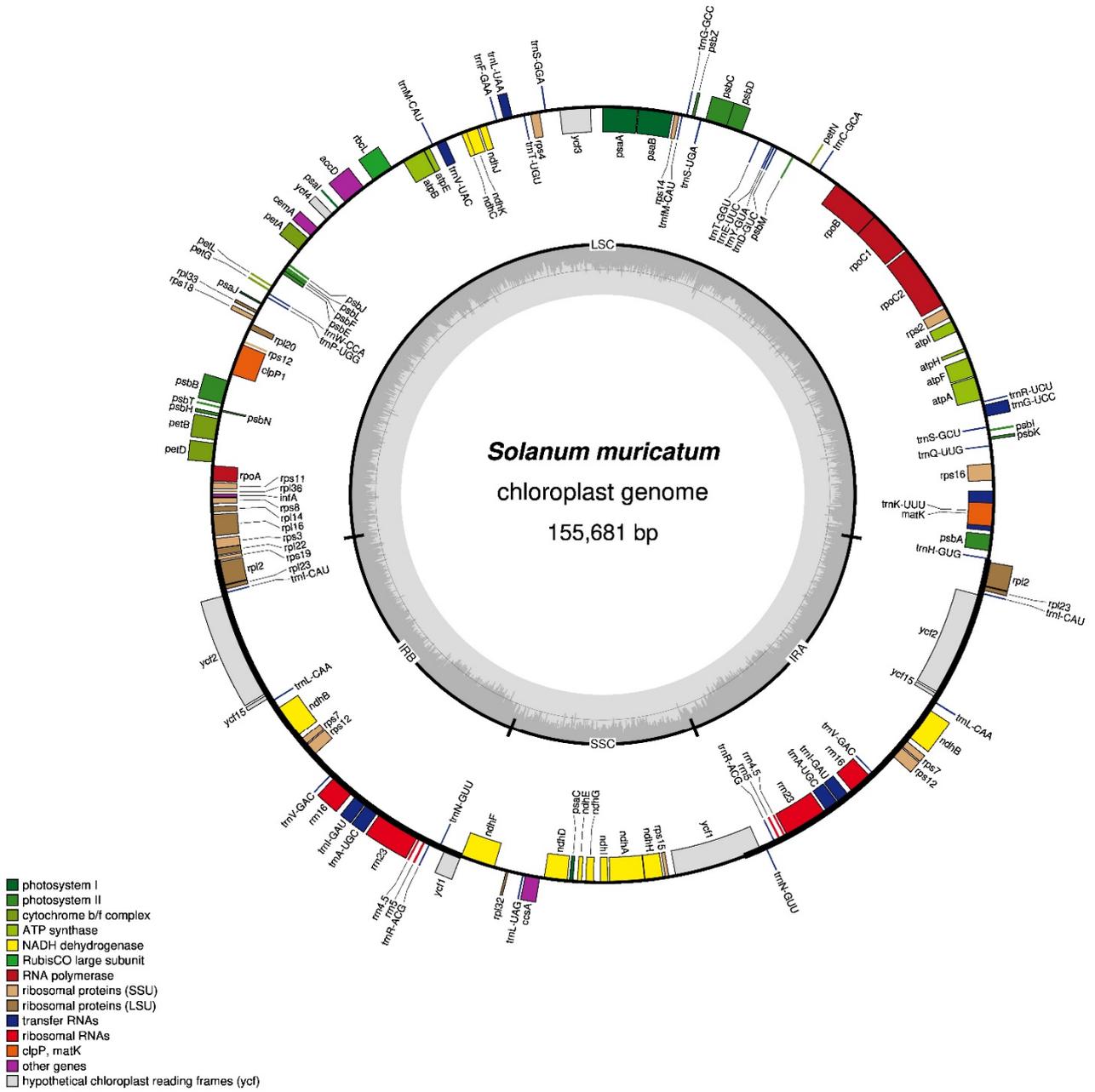
Figure 1. Gene map of *S. muricatum* cp genome. Genes drawn outside circle are transcribed counter clockwise, and genes inside this circle are transcribed in a clockwise direction. Genes belonging to different functional groups are in colors.

A

B

Figure 2. Codon usage analysis and amino acid frequencies in cp genome of sweet cucumber. (**A**). Relative synonymous codon usage (RSCU). (**B**). Amino acid frequencies. *X-axis* show codons that are represented by different colors.

2.2. Comparative Analysis of Genome Structure

We explored the structural characteristics of the cp genome of *S. muricatum,* and compared with other seven Solanaceae species: *S. bulbocastanum, S. dulcamara, S. etuberosum, S. lycopersicum, S. peruvianum, S. phureja, S. tuberosum.* The cp genomes of these species differed in 310 pb, 101 pb, 379 pb, 220 pb, 120 pb, 189 pb, and 385 pb, respectively (Table 1). We were able to identify less divergence between the coding regions than the noncoding regions. In addition, the IR regions showed less divergence than the SSC and LSC regions (Figures 3 and S1).

We identified that five pairs of genes (*trnL-UAG_1—ccsA_1, rps16—trnG-UCC, trnC-GCA—psbC, rps4—atpE, ycf10—psbJ*), that belong to the intergenic region, varied greatly. On the other hand, *ndhF_1, ycf1, ycf2, matK* and *rpoC2* in the coding regions were more conserved (Figure 3). The identity matrix showed that the values in the IR region fluctuated

between 0.99 and 1. The SSC and LSC region showed the values that varied from 0.96 to 1 (Figure S1). We found greater divergences between *S. muricatum* and *S. dulcamara*.

Figure 3. Identity plot comparison of the chloroplast genome of *S. muricatum* with other cp genomes. The *x*-axis represents the coordinate in the chloroplast genome. The percent identity ranging from 50–100% is presented on the *y*-axis. Genome regions are color-coded.

2.3. SSR Analysis in Solanaceae

Forty-eight perfect simple sequence repeats (SSR) were identified in the *S. muricatum* cp genome. We also detected that the mononucleotide repeats (32) were the most abundant, followed by tetranucleotide (6) and dinucleotides (5). SSR with trinucleotides (3), pentanucleotide (1), and hexanucleotide (1) repeat motifs were recognized in less quantity (Figure S3). Similarly, 49, 54, 49, 48, 44, and 46 SSRs were found in *S. bulbocastanum*, *S.dulcamara*, *S. etuberosum*, *S. lycopersicum*, *S. peruvianum*, *S. phureja*, *S. tuberosum* genomes, respectively (Table S3). We also identified that all these *Solanum* species possessed not only the highest number of A/T mononucleotides but also AT/TA dinucleotides. On the other hand, only *S. bulbocastarum* and *S. muricatum* presented hexanucleotide repeats; and *S. etuberosum* did not present pentanucleotide repeats. The SSC and LSC regions presented greater quantity of SSR (Figure S2).

2.4. Phylogenetic Inference of S. muricatum

We employed sequence alignments of 33 Solanum species and one outgroup (*Capsicum chinense*, Solanaceae) that were submitted to Dryad (https://datadryad.org/stash/share/hZgcwsMMwANRaCSNTBfH7eAFGVaYULbrkemV0Lsfq-A accessed on

9 April 2022). Our maximum likelihood (ML) phylogenetic tree was highly resolved. All nodes possessed 100% bootstrap support (BS), except for two (98% and 99%). Members of the *Petota* section were clustered into one group with 100% BS. Similarly, the nine species of the *Lycopersicum* section were placed within one cluster with very high BS. *Solanum etuberosum* was sister species to these two sections (*Petota*, *Lycopersicum*), and sister to them was *S. muricatum* (section *Basarthrum*) with 100% BS (Figure 4).
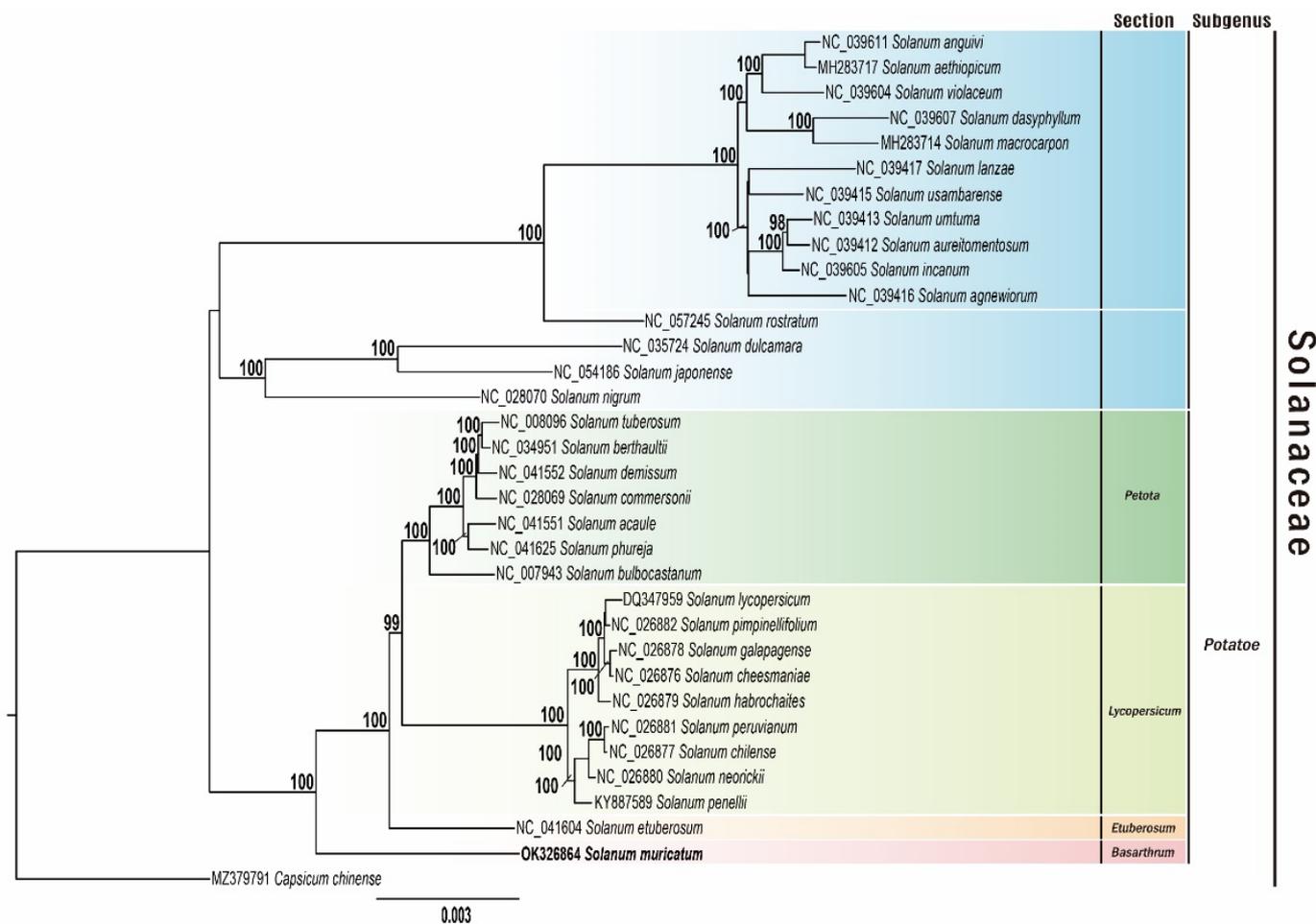
Figure 4. Maximum likelihood tree of the Solanaceae family examined here with complete chloroplast genome data. The numbers above the nodes represent bootstrap values, with only values higher than 70% depicted. The outgroup taxon is *Capsicum chinense*.

3. Discussion

Employing next-generation sequencing (NGS) techniques, here, we sequenced for the first time the *S. muricatum* cp genome with accession number: OK326864, and compared to other cp genomes of seven Solanaceae species, which revealed similar genomic features. The *S. muricatum* cp genome is similar in gene content to most angiosperm species [26,28–30]. The complete cp genome of *S. muricatum* was 155,681 pb in length, which was very similar to other Solanaceae genomes with a typical structure (two IR regions, LSC, and SSC). This structure is a characteristic of higher plants 43 [30]. The annotated sweet cucumber cp genome proposed 88 protein coding genes (CDS), differing from the CDS of other Solanaceae species [20,26]. Similar to other studies [19,20], we also identified two genes, *ycf3* and *clpP1*, that contain two introns in the chloroplast genome of sweet cucumber. These characteristics are important because they give us light into the metabolism of chloroplast of *S. muricatum*. Gene *clpP1* (caseinolytic protease P1) is indispensable for the metabolism of plants [31] and is involved in the expression of plastid genes [32,33]. Furthermore, the

ycf3 gene is important in the conformation of photosystem I (PSI) as it interacts at the post-translational level with the PSI subunits [34].

The GC content of *Solanum* chloroplast genomes were in agreement with previous studies in other members of the Solanaceae family [24]. Concordant with previous work in other Solanaceae species [26,35,36] and in other angiosperms cp genomes [30,37,38], the GC percentage was high in the IR, probably due to the structure of the ribosomal RNA [19,39]. It is important to determine the guanine-cytosine (GC) content as it is used to characterize the behavior of genomes in general terms [40,41]. That is, variation of GC content across genomes is a key feature of genomic organization and varies strongly between species [40–42]. Moreover, GC content also sheds light on gene density, compactness, and recombination rates [43]. Our results revealed similarities between the cp genomes, indicating that *Solanum* cp DNA sequences were highly conserved. Previous reports depicted that the sequences were more conserved in IR regions compared to LSC and SSC regions. A plausible explanation for this are the copy corrections between inverted repeat sequences by gene conversion [44]. Further research is needed to confirm this possible explanation. *Solanum dulcamara* cp genome showed higher divergence values in comparison with the other seven *Solanum* species (Figure S2). Genomic rearrangements or some other biological events (deletions, inversions, or insertions) may be producing higher levels of divergence between *S. dulcamara* and other *Solanum* cp genomes. To determine the exact reason for divergence, further research should be conducted. In the present study, we identified five regions that were highly divergent. We consider that these variable regions may be employed for phylogenetic studies and the development of molecular markers in Solanaceae species.

In addition, SSR could be employed for phylogenetic studies [45,46]. We identified 48 perfect microsatellites in *S. muricatum*, which included dispersed, palindromic, and tandem repeats. In agreement with other studies [19,47], these SSRs are located in the IR, LSC, and SSC regions with high A/T bias. The majority of these repeats were found in non-coding regions, which were the regions with the greatest divergence in the cp genomes, as reported by Asaf et al. [48], Raman and Park [30], and Zhang et al. [49]. Moreover, the LSC region contained most of the microsatellites compared to IR and SSC regions, which is concordant with other angiosperm plastid genomes [48,50,51]. The identification of SSR in the chloroplast genome of *S. muricatum* will allow us to examine the genetic diversity and population structure between and within the populations of *S. muricatum*. In addition, these markers could also be applied for the characterization and selection of sweet cucumber individuals for the development of a modern program of conservation and genetic improvement [27].

The translation of mRNA into proteins is based on the ordering of the genetic code. This is redundant as two or more codons can be encoding the majority of amino acids (known as synonymous), except tryptophan and methionine. Several studies regarding the usage frequency of synonymous codons revealed that synonymous codons varied within the same species and between different species, despite encoding the same amino acid [52,53], phenomenon named as codon usage bias [54]. Hence, highly expressed genes are using more frequent codons than lowly expressed genes [55]. However, codon usage bias of the chloroplast genome is highly conserved, but its encoding genes do not have the same codon usage bias [56]. Our results revealed the preference of 29 codons of *S. muricatum* more frequently than expected based on RCSU values > 1. A similar preference of these codons was reported in other Solanaceae and higher plant genomes [24,27,56–58]. In addition, nucleotide constitution at the third codon site of *S. muricatum* evidenced codons ending in T or A more often than codons ending with G or C, as reported by other studies [24,27,56] where plant chloroplast genomes dis-played low GC content and tended to employ A/T ending codons. Taking together this information, we concluded that encoding genes present in the chloroplast genome of sweet cucumber tended to employ highly expressed codons with A/T termination.

The phylogenomics of Solanaceae using complete chloroplast genomes was well resolved. Weese and Bohs [59] employed three DNA sequence regions (chloroplast *ndhF* and *trnTF*, and nuclear *waxy*) and identified major Solanum clades. Similarly, Särkinen et al. [60] identified clades of Solanaceae, particularly *Solanum* by using five plastid (*ndhF*, *matK*, *trnS-G psbA-trnH*, *trnL-F*), and two nuclear regions (*ITS* and *waxy*). The phylogenetic position of sweet cucumber was previously reported by Anderson et al. [1], Herraiz et al. [2]. Särkinen et al. [60], Prohens et al. [61], and Blanca et al. [62]. They demonstrated that *S. muricatum* (i) is closely related to tomatoes and potatoes, and (ii) is member of section *Basarthrum* in the Potatoe clade, as reported in the present work. Similar to Weese and Bohs [59], sweet cucumber was placed, with 100% BS, as sister species to a clade formed by members of sections *Petota, Lycopersicum,* and *Etuberosum*, confirming that chloroplast genomes can successfully determine relationships of Solanum members, as demonstrated by Huang et al. [63] who analyzed the plastid sequence of 202 cultivated and wild potatoes (section *Petota*). However, further molecular studies, including additional collections of sweet cucumber from a wider geographic area, are needed to confirm its origin and relationships.

4. Materials and Methods

4.1. Plant Material and DNA Isolation

One plant of sweet cucumber was selected from Cañete province, in Lima department to be sequenced. It was collected and deposited at the INIA Germplasm Bank (https://www.inia.gob.pe/sdrg/ accessed on 20 March 2022), under the accession code PER1002426. Fresh leaves were used to purify genomic DNA using the CTAB method [64] adapted for this species. Finally, we evaluated the DNA quantity and quality using the Qubit™ 4 Fluorometer (Invitrogen, Waltham, MA, USA) and agarose gel (1%), respectively.

4.2. DNA Sequence and Genome Assembly

Whole genome sequencing was performed by the Illumina HiSeq 2500 and paired-end (PE) 150 library (Illumina, San Diego, CA, USA). Low-quality reads were removed with Trim Galore software [65] with the original configuration. To assemble the cp genome of *S. muricatum*, GetOrganelle v1.7.2 [66] was used with the same parameters as followed by Saldaña et al. [67], using *S. tuberosum* (NC_008096) as reference. SPAdes v3.11.1 [68], bowtie2 v2.4.2 [69], and BLAST+ v2.11 [70] were also run with default options.

4.3. Annotation of S. muricatum Chloroplast

Sweet cucumber gene annotation was performed in Geseq online tool [71] with original options. We included plastid genomes of related species within Solanaceae available in NCBI associated with this server. The chloroplast genome of *S. muricatum* was manually curated. We used the MEGA X software for codon usage analysis [72]. Finally we visualized the architecture of *S. muricatum* cp genome employing OGDRAW 1.3.1 [73].

4.4. Comparison of Solanaceae cp Genomes

We used mVISTA program with Shuffle-LAGAN model (http://genome.lbl.gov/vista/mvista/ accessed on 1 March 2022) [74] to compare the cp genome of sweet cucumber with seven species of Solanaceae family (*S. bulbocastanum*—NC_007943, *S.dulcamara*—NC_035724, *S. etuberosum*—NC_041604.1, *S. lycopersicum*—DQ347959, *S. peruvianum*—NC_02688, *S. phureja*—NC_041625.1, *S. tuberosum*—NC_008096.2). We used the annotated *S. muricatum* cp genome as reference. Sequence alignment was conducted by using MAFFT v7.475 [75], following the same parameters of Saldaña et al. [67]. Manual alignment corrections were performed using MacClade v4.08a [76]. Finally an identity plot was created using the *ggplot2* [77] package in the R software [78]. We also used extension packages including *ggtext* (https://github.com/wilkelab/ggtext/issues accessed on 5 March 2022) and *ggpubr* [79].

MISA software [80] was employed to identify microsatellites of seven Solanaceae cp genomes. The repeats for mononucleotides, dinucleotides, trinucleotides, tetranucleotides,

pentanucleotides, and hexanucleotides were 10, 5, 4, 3, 3, and 3, respectively [49]. A plot with the structure and location of the SSRs in the seven cp genomes was generated using *gggenomes* (https://github.com/thackl/gggenomes accessed on 6 March 2022) and *genoPlotR* [81] packages in the R software [78]. Finally, we employed MEGA X software [72] to analyze frequency, the codon usage, and relative synonymous codon usage (RCSU) of the *S. muricatum* cp genome. We used the original parameters.

4.5. Phylogenetic Analyses

We constructed a maximum likelihood (ML) phylogenetic tree using RAxML v8.2.11 software [82] with 1000 nonparametric bootstrap replicates under the GTR + nucleotide substitution model of evolution. The 33 cp genomes selected from the Organelle Genome Resources of the NCBI (accessed on 09 March 2022) were compared and aligned by MAFFT [75] with *S. muricatum* cp genome. *Capsicum chinense* (accession number: MK379791) was included as an outgroup. The resulting tree was viewed with FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/ accessed on 25 March 2022).

5. Conclusions

Here, we report for the first time the complete chloroplast genome sequence of a neglected native crop from the Andean region. The complete cp sequence of *S. muricatum* is similar to other Solanaceae species in structure and gene content, providing further evidence of common chloroplast evolutionary lineage within this family. We identified 48 microsatellites to be used as molecular markers to study the population structure and genetic diversity of sweet cucumber landraces. The gene content, order, and genome structure were much conserved for all *Solanum* species. However, *S. dulcamara* showed high values of dissimilarity compared with the other seven species. Similar to previous studies employing different genes, we showed that sweet cucumber is placed within section *Petota* and is closely related to tomatoes and potatoes. We will continue to develop molecular tools for this neglected Andean horticultural crop as they may throw light on deciphering its evolution and establishing conservation practices and genomics-assisted breeding.

References

1. Anderson, G.J.; Jansen, R.K.; Kim, Y. The origin and relationships of the pepino, Solanum muricatum (Solanaceae): DNA restriction fragment evidence. *Econ. Bot.* **1996**, *50*, 369–380. [CrossRef]
2. Herraiz, F.J.; Blanca, J.; Ziarsolo, P.; Gramazio, P.; Plazas, M.; Anderson, G.J.; Prohens, J.; Vilanova, S. The first de novo transcriptome of pepino (*Solanum muricatum*): Assembly, comprehensive analysis and comparison with the closely related species *S. caripense*, potato and tomato. *BMC Genom.* **2016**, *17*, 321. [CrossRef]
3. Anderson, G.J.; Martine, C.T.; Prohens, J.; Nues, F. *Solanum* Perlongistylum and *S. Catilliflorum*, New Endemic Peruvian Species of *Solanum*, Section Basarthrum, Are Close Relatives of the Domesticated Pepino, *S. Muricatum. Novon A J. Bot. Nomencl.* **2006**, *16*, 161–167. [CrossRef]
4. Jones, R.A.C.; Koenig, R.; Lesemann, D.E. Pepino mosaic virus, a new potexvirus from pepino (*Solanum muricatum*). *Ann. Appl. Biol.* **1980**, *94*, 61–68. [CrossRef]
5. Pickersgill, B. Domestication of plants in the Americas: Insights from Mendelian and molecular genetics. *Ann. Bot.* **2007**, *100*, 925–940. [CrossRef]
6. Gorbe, M.; Bhat, R.; Aznar, E.; Sancenón, F.; Marcos, M.D.; Herraiz, F.J.; Prohens, J.; Venkataraman, A.; Martínez-Máñez, R. Rapid biosynthesis of silver nanoparticles using pepino (*Solanum muricatum*) leaf extract and their cytotoxicity on HeLa cells. *Materials* **2016**, *9*, 325. [CrossRef]
7. Gülçin, I. Antioxidant activity of food constituents: An overview. *Arch. Toxicol.* **2012**, *86*, 345–391. [CrossRef]
8. Prohens, J.; Ruiz, J.J.; Nuez, F. The pepino (*Solanum muricatum*, Solanaceae): A "new" crop with a history. *Econ. Bot.* **1996**, *50*, 355–368. [CrossRef]
9. Rodríguez-Burruezo, A.; Prohens, J.; Fita, A.M. Breeding strategies for improving the performance and fruit quality of the pepino (*Solanum muricatum*): A model for the enhancement of underutilized exotic fruits. *Food Res. Int.* **2011**, *44*, 1927–1935. [CrossRef]
10. Hsu, C.C.; Guo, Y.R.; Wang, Z.H.; Yin, M. chin Protective effects of an aqueous extract from pepino (*Solanum muricatum* Ait.) in diabetic mice. *J. Sci. Food Agric.* **2011**, *91*, 1517–1522. [CrossRef]
11. Ma, C.T.; Chyau, C.C.; Hsu, C.C.; Kuo, S.M.; Chuang, C.W.; Lin, H.H.; Chen, J.H. Pepino polyphenolic extract improved oxidative, inflammatory and glycative stress in the sciatic nerves of diabetic mice. *Food Funct.* **2016**, *7*, 1111–1121. [CrossRef] [PubMed]
12. Wang, Z.; Hsu, C.; Yin, M. Aqueous Extract from Pepino (*Solanum muricatum* Ait.) Attenuated Hyperlipidemia and Cardiac Oxidative Stress in Diabetic Mice. *ISRN Obes.* **2012**, *2012*, 1–6. [CrossRef] [PubMed]
13. Sudha, G.; Sangeetha Priya, M.; Indhu Shree, R.B.; Vadivukkarasi, S. Antioxidant Activity of Ripe and Unripe Pepino Fruit (*Solanum muricatum* Aiton). *J. Food Sci.* **2012**, *77*, 1131–1135. [CrossRef] [PubMed]
14. David, G. Frodin History and concepts of the big plant genera. *Taxon* **1394**, *53*, 753–776.
15. Wang, L.; Xing, H.; Yuan, Y.; Wang, X.; Saeed, M.; Tao, J.; Feng, W.; Zhang, G.; Song, X.; Sun, X. Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS ONE* **2018**, *13*, e0194372. [CrossRef]
16. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [CrossRef]
17. Ahmed, I.; Biggs, P.J.; Matthews, P.J.; Collins, L.J.; Hendy, M.D.; Lockhart, P.J. Mutational dynamics of aroid chloroplast genomes. *Genome Biol. Evol.* **2012**, *4*, 1316–1323. [CrossRef]
18. Sloan, D.B.; Triant, D.A.; Forrester, N.J.; Bergner, L.M.; Wu, M.; Taylor, D.R. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). *Mol. Phylogenet. Evol.* **2014**, *72*, 82–89. [CrossRef]
19. Zhao, Z.; Wang, X.; Yu, Y.; Yuan, S.; Jiang, D.; Zhang, Y.; Zhang, T.; Zhong, W.; Yuan, Q.; Huang, L. Complete chloroplast genome sequences of *Dioscorea*: Characterization, genomic resources, and phylogenetic analyses. *PeerJ* **2018**, *2018*, e6032. [CrossRef]
20. Amiryousefi, A.; Hyvönen, J.; Poczai, P. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. *PLoS ONE* **2018**, *13*, e0196069. [CrossRef]
21. Powell, W.; Morgante, M.; Mcdevitt, R.; Vendramin, G.G.; Rafalski, J.A. Polymorphic Simple Sequence Repeat Regions in Chloroplast Genomes: Applications to the Population Genetics of Pines. *Proc. Natl. Acad. Sci. USA* **2016**, *92*, 7759–7763. [CrossRef] [PubMed]
22. Bock, R.; Khan, M.S. Taming plastids for a green future. *Trends Biotechnol.* **2004**, *22*, 311–318. [CrossRef] [PubMed]
23. Chung, H.J.; Jung, J.D.; Park, H.W.; Kim, J.H.; Cha, H.W.; Min, S.R.; Jeong, W.J.; Liu, J.R. 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.* **2006**, *25*, 1369–1379. [CrossRef] [PubMed]

24. Daniell, H.; Lee, S.B.; Grevich, J.; Saski, C.; Quesada-Vargas, T.; Guda, C.; Tomkins, J.; Jansen, R.K. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* **2006**, *112*, 1503–1518. [CrossRef] [PubMed]

25. Schmitz-Linneweber, C.; Regel, R.; Du, T.G.; Hupfer, H.; Herrmann, R.G.; Maier, R.M. The plastid chromosome of Atropa belladonna and its comparison with that of *Nicotiana tabacum*: The role of RNA editing in generating divergence in the process of plant speciation. *Mol. Biol. Evol.* **2002**, *19*, 1602–1612. [CrossRef]

26. Mehmood, F.; Abdullah; Ubaid, Z.; Shahzadi, I.; Ahmed, I.; Waheed, M.T.; Poczai, P.; Mirza, B. Plastid genomics of *Nicotiana* (Solanaceae): Insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). *PeerJ* **2020**, *8*, e9552. [CrossRef]

27. Arbizu, C.I.; Saldaña, C.L.; Ferro-Mauricio, R.D.; Chávez-Galarza, J.C.; Herrera, J.; Contreras-Liza, S.; Guerrero-Abad, J.C.; Maicelo, J.L. Characterization of the complete chloroplast genome of a Peruvian landrace of *Capsicum chinense* Jacq. (Solanaceae), arnaucho chili pepper. *Mitochondrial DNA Part B Resour.* **2022**, *7*, 156–158. [CrossRef]

28. Li, S.M.; Zheng, X.H.; Duan, H.C.; Dong, Q. The complete chloroplast genome of *Solanum betacea* (Solanaceae, *Solaneae*). *Mitochondrial DNA Part B Resour.* **2021**, *6*, 1642–1644. [CrossRef]

29. Li, S.; Wang, Y.-P.; Zhao, Y.-F.; Zhang, J.-Y.; Zhang, J.-Y.; Ma, H.-R.; Yue, Y.; Du, C.-Y.; Zhao, C.-B.; Han, Y.-Z. Characterization of the complete chloroplast genome of the *Solanum tuberosum* L. cv. Favorita (Solanaceae). *Mitochondrial DNA Part B* **2021**, *6*, 909–911. [CrossRef]

30. Raman, G.; Park, S.J. The complete chloroplast genome sequence of the *Speirantha gardenii*: Comparative and adaptive evolutionary analysis. *Agronomy* **2020**, *10*, 1405. [CrossRef]

31. Kuroda, H.M.P. The plastid clpP1 protease gene is essential for plant development. *Nat. Publ.* **2003**, *425*, 30–33. [CrossRef] [PubMed]

32. Clarke, A.K.; Schelin, J.; Porankiewicz, J. Inactivation of the clpP1 gene for the proteolytic subunit of the ATP-dependent Clp protease in the cyanobacterium Synechococcus limits growth and light acclimation. *Plant Mol. Biol.* **1998**, *37*, 791–801. [CrossRef]

33. Cahoon, A.B.; Cunningham, K.A.; Stern, D.B. The plastid clpP gene may not be essential for plant cell viability. *Plant Cell Physiol.* **2003**, *44*, 93–95. [CrossRef] [PubMed]

34. Naver, H.; Boudreau, E.; Rochaix, J.D. Functional studies of Ycf3: Its role in assembly of photosystem I and interactions with some of its subunits. *Plant Cell* **2001**, *13*, 2731–2745. [CrossRef] [PubMed]

35. Yukawa, M.; Tsudzuki, T.; Sugiura, M. The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: Complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol. Genet. Genom.* **2006**, *275*, 367–373. [CrossRef]

36. Sugiyama, Y.; Watase, Y.; Nagase, M.; Makita, N.; Yagura, S.; Hirai, A.; Sugiura, M. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: Comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genom.* **2005**, *272*, 603–615. [CrossRef]

37. Yang, J.B.; Yang, S.X.; Li, H.T.; Yang, J.; Li, D.Z. Comparative Chloroplast Genomes of *Camellia* Species. *PLoS ONE* **2013**, *8*, e73053. [CrossRef]

38. Raman, G.; Park, V.; Kwak, M.; Lee, B.; Park, S.J. Characterization of the complete chloroplast genome of *Arabis stellari* and comparisons with related species. *PLoS ONE* **2017**, *12*, e0183197. [CrossRef]

39. Qian, J.; Song, J.; Gao, H.; Zhu, Y.; Xu, J.; Pang, X.; Yao, H.; Sun, C.; Li, X.; Li, C.; et al. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *PLoS ONE* **2013**, *8*, e5760. [CrossRef]

40. Gibson, G.; Muse, S.V. *A Primer of Genome Science*; Sinauer Associates: Sunderland, MA, USA, 2009.

41. Li, W.; Graur, D. *Fundamentals of Molecular Evolution*; Sinauer Associates: Sunderland, MA, USA, 1991; 284p.

42. Glémin, S.; Clément, Y.; David, J.; Ressayre, A. GC content evolution in coding regions of angiosperm genomes: A unifying hypothesis. *Trends Genet.* **2014**, *30*, 263–270. [CrossRef]

43. Duret, L.; Arndt, P.F. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* **2008**, *4*, e1000071. [CrossRef] [PubMed]

44. Zhang, Y.; Iaffaldano, B.J.; Zhuang, X.; Cardina, J.; Cornish, K. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biol.* **2017**, *17*, 34. [CrossRef] [PubMed]

45. Le Flèche, P.; Hauck, Y.; Onteniente, L.; Prieur, A.; Denoeud, F.; Ramisse, V.; Sylvestre, P.; Benson, G.; Ramisse, F.; Vergnaud, G. A tandem repeats database for bacterial genomes: Application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiol.* **2001**, *1*, 2. [CrossRef] [PubMed]

46. Rokas, A.; Holland, P.W.H. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **2000**, *15*, 454–459. [CrossRef]

47. Kuang, D.Y.; Wu, H.; Wang, Y.L.; Gao, L.M.; Zhang, S.Z.; Lu, L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): Implication for DNA barcoding and population genetics. *Genome* **2011**, *54*, 663–673. [CrossRef]

48. Asaf, S.; Khan, A.L.; Khan, M.A.; Waqas, M.; Kang, S.M.; Yun, B.W.; Lee, I.J. Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata ssp. petraea*: Structures and comparative analysis. *Sci. Rep.* **2017**, *7*, 7556. [CrossRef]

49. Zhang, Y.; Zhang, J.W.; Yang, Y.; Li, X.N. Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* Gaertn. f. and Related Rubiaceae Species. *Forests* **2019**, *10*, 1000. [CrossRef]

50. Shahzadi, I.; Abdullah; Mehmood, F.; Ali, Z.; Ahmed, I.; Mirza, B. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: Comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. *Genomics* **2020**, *112*, 1454–1463. [CrossRef]

51. Mehmood, F.; Abdullah; Shahzadi, I.; Ahmed, I.; Waheed, M.T.; Mirza, B. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. *Genomics* **2020**, *112*, 1522–1530. [CrossRef]

52. Grantham, R.; Gautier, C.; Gouy, M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genotype. *Nucleic Acids. Res.* **1980**, *8*, 1883–1912. [CrossRef]

53. Plotkin, J.B.; Kudla, G. Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **2011**, *12*, 32–42. [CrossRef] [PubMed]

54. Liu, Y. A code within the genetic code: Codon usage regulates co-translational protein folding. *Cell Commun. Signal.* **2020**, *18*, 145. [CrossRef] [PubMed]

55. Nie, X.; Zhao, X.; Wang, S.; Zhang, T.; Li, C.; Liu, H.; Tong, W.; Guo, Y. Complete chloroplast genome sequence of broomcorn millet (*Panicum miliaceum* L.) and comparative analysis with other panicoideae species. *Agronomy* **2018**, *8*, 159. [CrossRef]

56. Zhang, R.; Zhang, L.; Wang, W.; Zhang, Z.; Du, H.; Qu, Z.; Li, X.Q.; Xiang, H. Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *Solanum* species. *Int. J. Mol. Sci.* **2018**, *19*, 3142. [CrossRef] [PubMed]

57. Yu, X.; Zuo, L.; Lu, D.; Lu, B.; Yang, M.; Wang, J. Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis. *Gene* **2019**, *689*, 141–151. [CrossRef]

58. Dong, F.; Lin, Z.; Lin, J.; Ming, R.; Zhang, W. Chloroplast genome of rambutan and comparative analyses in sapindaceae. *Plants* **2021**, *10*, 283. [CrossRef]

59. Weese, T.L.; Bohs, L. A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Syst. Bot.* **2007**, *32*, 445–463. [CrossRef]

60. Särkinen, T.; Bohs, L.; Olmstead, R.G.; Knapp, S. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evol. Biol.* **2013**, *13*, 214. [CrossRef]

61. Prohens, J.; Anderson, G.J.; Blanca, J.M.; Cañizares, J.; Zuriaga, E.; Nuez, F. The implications of AFLP data for the systematics of the wild species of Solanum section Basarthrum. *Syst. Bot.* **2006**, *31*, 208–216. [CrossRef]

62. Blanca, J.M.; Prohens, J.; Anderson, G.J.; Zuriaga, E.; Nuez, F.; Prohens, J.; Anderson, J. AFLP and DNA sequence variation in an Andean domesticate, pepino (*Solanum muricatum*, Solanaceae): Implications for evolution and domestication. *Am. J. Bot.* **2007**, *94*, 1219–1229. [CrossRef]

63. Huang, Y.; Yang, Z.; Huang, S.; An, W.; Li, J.; Zheng, X. Comprehensive analysis of rhodomyrtus tomentosa chloroplast genome. *Plants* **2019**, *8*, 89. [CrossRef] [PubMed]

64. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

65. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2013**, *17*, 10–12. [CrossRef]

66. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; DePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef]

67. Saldaña, C.L.; Rodriguez-Grados, P.; Chávez-Galarza, J.C.; Feijoo, S.; Guerrero-Abad, J.C.; Vásquez, H.V.; Maicelo, J.L.; Jhoncon, J.H.; Arbizu, C.I. Unlocking the Complete Chloroplast Genome of a Native Tree Species from the Amazon Basin, Capirona (*Calycophyllum spruceanum*, Rubiaceae), and Its Comparative Analysis with Other Ixoroideae Species. *Genes* **2022**, *13*, 113. [CrossRef]

68. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]

69. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]

70. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]

71. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq-Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [CrossRef]

72. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

73. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef] [PubMed]

74. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [CrossRef] [PubMed]

75. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]

76. Maddison, D.R.; Maddison, W.P. *MacClade 4.08a: Analysis of Phylogeny and Character Evolution*; Sinauer: Sunderland, MA, USA, 2005. [CrossRef]

77. Wickham, H. *ggplot2-Elegant Graphics for Data Analysis*; Springer International Publishing: Cham, Switzerland, 2016.

78. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; p. 326864.

79. Kassambara, A. ggpubr: "ggplot2" Based Publication Ready Plots. 2020. Available online: https://CRAN.R-project.org/package=ggpubr (accessed on 15 April 2022).

80. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]

81. Guy, L.; Kultima, J.R.; Andersson, S.G.E.; Quackenbush, J. GenoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* **2011**, *27*, 2334–2335. [CrossRef]

82. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]