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*

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* + *Eubersosum*. 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.

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

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.
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—psbf*), 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*.

### 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*, and *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/hZgcwsMMwANRaCSNTBfH7eAFCGvaYULbrkemVOLs6q-A](https://datadryad.org/stash/share/hZgcwsMMwANRaCSNTBfH7eAFCGvaYULbrkemVOLs6q-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 [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 Gesseq 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). 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.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data7090123/s1, Table S1: Genes identified in the cp genome of S. muricatum; Table S2: Codon usage and codon-anticodon recognition pattern in the cp genome of S. muricatum; Table S3: Number and types of microsatellites in S. muricatum and seven Solanaceae species; Figure S1: Identity matrix of LSC, SSC, IRa, IRb regions between eight cp genomes; Figure S2: Genome structure comparison and location of microsatellites of the eight cp genomes, with S. muricatum as reference; Figure S3: Simple sequence repeat (SSR) number and motif in the S. muricatum cp genome. The colored bars indicate the different repeats within SSR.


Funding: This research was funded by project “Creacióndel servicio de agricultura de precisión en los Departamentos de Lambayeque, Huancavelica, Ucayali y San Martín 4 Departamentos” of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government with grant number CUI 2449640. J.C.Ch.-G. and C.L.S. were supported by PP0068 “Reducción de la vulnerabilidad y atención de emergencias por desastres” of MIDAGRI.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Written informed consent was obtained from the owner of the sweet cucumber crop.

Data Availability Statement: All data generated during this study are included in this published article.
Acknowledgments: We thank Rubén Ferro for providing technical assistance in the laboratory. We also thank Ivan Ucharima, Maria Angélica Puyo, Cristina Aybar, and Erick Rodriguez for supporting the logistic activities in the laboratory. Finally, the authors thank the Bioinformatics High-performance Computing server of Universidad Nacional Agraria la Molina for providing resources for data analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Herraz, F.J.; Blanco, 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]
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, 17, 134. [CrossRef]


26. Mehmood, F.; Abdulllah, U.; 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). *Peer*. 2020, 8, e9552. [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]


38. 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]


40. 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 *Sativa miltiorrhiza*. *PLoS ONE* 2013, 8, e5760. [CrossRef]


46. Le Fléche, 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]


50. 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]
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]
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 panicoidaceae species. Agronomy 2018, 8, 159. [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]
65. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 2013, 17, 10–12. [CrossRef]
67. Saldaña, C.L.; Rodríguez-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]
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]

