

Article

# Identification of High Molecular Variation Loci in Complete Chloroplast Genomes of *Mammillaria* (Cactaceae, Caryophyllales)

Delil A. Chincoya <sup>1</sup>, Alejandro Sanchez-Flores <sup>2</sup> , Karel Estrada <sup>2</sup>, Clara E. Díaz-Velásquez <sup>3</sup>, Antonio González-Rodríguez <sup>4</sup>, Felipe Vaca-Paniagua <sup>3,5</sup>, Patricia Dávila <sup>6</sup>, Salvador Arias <sup>7</sup>  and Sofía Solórzano <sup>1,\*</sup> 

- <sup>1</sup> Laboratorio de Ecología Molecular y Evolución, Unidad de Biotecnología y Prototipos FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Mexico; dela@comunidad.unam.mx
  - <sup>2</sup> Instituto de Biotecnología, Unidad Universitaria de Secuenciación Masiva y Bioinformática, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Chamilpa, Cuernavaca 62250, Mexico; alexsf@ibt.unam.mx (A.S.-F.); karel@ibt.unam.mx (K.E.)
  - <sup>3</sup> Laboratorio Nacional en Salud, Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, FES Iztacala, Universidad Nacional Autónoma de México, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Mexico; cdiaz@comunidad.unam.mx (C.E.D.-V.); felipe.vaca@iztacala.unam.mx (F.V.-P.)
  - <sup>4</sup> Laboratorio de Genética de la Conservación, Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701, Ex-Hacienda San José La Huerta, Morelia 58190, Mexico; agrodrig@cieco.unam.mx
  - <sup>5</sup> Subdirección de Investigación Básica, Instituto Nacional de Cancerología, Ciudad de México 04510, Mexico
  - <sup>6</sup> Laboratorio de Recursos Naturales, Unidad de Biotecnología y Prototipos, FES Iztacala, Universidad Nacional Autónoma de México, Avenida de los Barrios 1, Los Reyes Iztacala, Tlalnepantla de Baz 54090, Mexico; pdavilaa@unam.mx
  - <sup>7</sup> Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Tercer Circuito Exterior, Ciudad Universitaria, Coyoacán, Ciudad de México 04510, Mexico; sarias@ib.unam.mx
- \* Correspondence: solorzanols@unam.mx; Tel.: +52-1-5556231129

Received: 19 June 2020; Accepted: 15 July 2020; Published: 21 July 2020



**Abstract:** In plants, partial DNA sequences of chloroplasts have been widely used in evolutionary studies. However, the Cactaceae family (1500–1800 species) lacks molecular markers that allow a phylogenetic resolution between species and genera. In order to identify sequences with high variation levels, we compared previously reported complete chloroplast genomes of seven species of *Mammillaria*. We identified repeated sequences (RSs) and two types of DNA variation: short sequence repeats (SSRs) and divergent homologous loci. The species with the highest number of RSs was *M. solisioides* (256), whereas *M. pectinifera* contained the highest amount of SSRs (84). In contrast, *M. zephyranthoides* contained the lowest number (35) of both RSs and SSRs. In addition, five of the SSRs were found in the seven species, but only three of them showed variation. A total of 180 homologous loci were identified among the seven species. Out of these, 20 loci showed a molecular variation of 5% to 31%, and 12 had a length within the range of 150 to 1000 bp. We conclude that the high levels of variation at the reported loci represent valuable knowledge that may help to resolve phylogenetic relationships and that may potentially be convenient as molecular markers for population genetics and phylogeographic studies.

**Keywords:** locus; *Mammillaria*; molecular variation; non-coding regions; repeated sequences; SSRs

## 1. Introduction

The analysis of DNA variation is critical when studying biological patterns in evolutionary biology, conservation and molecular ecology. For example, the comparison of complete sequences of chloroplast DNA (cpDNA) has enabled the interpretation of its evolutionary history [1] by tracking variations in genes related to the contemporary cyanobacteria. Genes are not randomly arranged in the circular molecule of cpDNA contained in each chloroplast. On the contrary, they are often located in a relatively conserved position that can be associated to each taxonomic group. Moreover, the total number of genes and the total size of cpDNA tend to be evolutionarily conserved, as does the absence or presence of two blocks of genes organized as inverted repeats [2]. Therefore, the information used for evolutionary studies varies depending on the type of plant.

In land flowering plants, the cpDNA is divided into four sections: two single copies (SCs), one of them larger (LSC) than the other one (SSC), which are separated by the two inverted repeats (IRs) [2]. The two IRs are the main source of structural variation within lineages and among the members of large taxonomic groups. The evolutionary information reflected in the structure of cpDNAs has been described [3], and it has been used as characters in phylogenetic studies [4,5]. The other source of variation in cpDNA is at the nucleotide level, comprising substitutions (transitions/transversions), gains or losses of a single or a block of nucleotides (insertions or deletions), and changes in the orientation of a pair or a large block of nucleotides (inverted sequences) [6,7].

In the cpDNA of land plants, there are two types of non-coding sequences, the intergenic spacers (IGSs) and the introns, which are currently most commonly used as molecular markers in phylogenetic studies [8]. In angiosperms, the IGSs *trnK-matK*, *psbA-trnH*, *trnL-trnF* [9–11] and intron *rpl16* [12] are recognized by their high molecular variation. However, across different taxonomic groups of angiosperms these DNA sequences exhibit unequal levels of molecular variation, and in some groups these loci frequently do not show high levels of variation. Consequently, under these conditions these markers are not useful in resolving phylogenetic relationships among species of a single genus or those of closely related genera [13–17]. However, such molecular markers resolved ancient deep divergent processes in distant phylogenetic taxa [18,19].

Thus, it is evident that the levels of molecular variation in non-coding and coding sequences of cpDNA that have for a long time been traditionally used to establish phylogenetic relationships are insufficient for all the taxonomic groups of angiosperms. Hence, it is necessary to find other sources of variation in the nuclear, chloroplast and mitochondrial genomes, in order to resolve phylogenies and other evolutionary issues. In this study, we focused on studying complete chloroplast genomes, since the comparative genomics perspective allows for a better identification of deep structural changes and punctual mutation changes that may provide insight into evolutionary processes. Recently, a high structural variation of chloroplast genomes was identified among species of the short-globose cacti of the genus *Mammillaria* [20], which showed strong differences with respect to the columnar cactus *Carnegiea gigantea* [21]. However, the amount of molecular variation in DNA sequences has not been evaluated in the cpDNA of the rich genus *Mammillaria*. This genus has received attention in previously published phylogenetic studies, but the phylogenetic relationship between the existent species has only been partially resolved, as it also occurs with other related genera. These studies obtained large polytomies, and even species of different genera were placed at the same internal nodes (e.g., [22,23]). Such unresolved phylogeny topologies are a common result for many angiosperm families and are not exclusive to Cactaceae. This is a consequence of a lack of molecular markers in many plant groups exhibiting high levels of variation [13,17].

*Mammillaria* Haw. (Cactaceae, Cactoideae, Cactaeae) is the richest cactus genus; it groups 163 species [24]. Besides the diversity relevance, 192 species and subspecies of this genus are a world conservation concern [25]. Previous phylogenetic studies based on *trnL-trnF* [13], *trnK-matK* [22] and *psbA-trnH* and *rpl16* intron [23] did not resolve the phylogenetic relationship within the group. In addition, the study based on the IGS *psbA-trnH* and *rpl16* intron concluded that the members of *Mammillaria* do not have a monophyletic origin [23]. Moreover, these studies did not resolve the

phylogenetic relationships of *Mammillaria* to six other cacti genera (*Coryphantha*, *Cumarina*, *Escobaria*, *Neolloydia*, *Ortegocactus* and *Pelecyphora*). It seems that the low levels of variation found in the molecular markers that were used may be associated to the relatively recent origin of *Mammillaria*, which is estimated to be <8 Ma [26,27]. In addition, the small number of molecular markers used in most of the phylogenetic studies may be an additional factor in explaining the unresolved phylogenies that were obtained. Considering the current scenario, finding molecular markers with the appropriate molecular variation levels in order to resolve the phylogenetic relationship between species and genera remains a challenge. Based on the genomic comparison of the cpDNA of phylogenetically close species of *Mammillaria*, we aimed to identify sources of high molecular variation, assuming that they might fully resolve phylogenetic relationships not only in *Mammillaria* but also in other Cactaceae and Caryophyllales members.

## 2. Materials and Methods

### 2.1. Data Source

We used the complete chloroplast genome of seven species of *Mammillaria* recently assembled: *M. albiflora*, *M. crucigera*, *M. huitzilopochtli*, *M. pectinifera*, *M. solisioides*, *M. supertexta* and *M. zephyranthoides* [20]. The seven chloroplast genomes are uploaded to the NCBI digital database and freely available in genbank (Table S1). By comparing the complete chloroplast genomes of these seven *Mammillaria* species, we identified three types of cpDNA sequences as sources of molecular variation: repeated sequences (RSs), short sequences of the microsatellite type (SSRs) and divergent DNA sequences in homologous loci.

### 2.2. Repeats Characterization and Short Sequence Repeats Identification

We compared the cpDNA of the seven studied species (Table S1) and identified forward, reverse, palindromic and complementary RSs with a minimum size of 30 bp, using a Hamming distance of three with REPUTER [28]. In addition, for each species, SSRs were identified with the MISA perl script [29]. The minimum number of repeat units was ten for homopolymer repeats, six for dinucleotide and five for trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats.

### 2.3. Sequence Divergence in Homologous Loci Among Species

A total of 180 homologous DNA regions were identified; of these, 77 were coding and 103 were non-coding sequences. Each of these 180 loci was aligned with MAFFT v7.310 [30] by paired comparisons between species. For each of the aligned loci, the proportion of nucleotide substitutions was documented (total number of divergent changes/locus length). The changes in the nucleotide position considered transitions and transversions. Each gap  $\geq 1$  of contiguous bases was counted as a single evolutionary change. The percentage of variation by locus among 21 paired comparisons was obtained and averaged. The percentage of variation by locus was calculated with the formula: % variation =  $[(NS + ID)/L] \times 100$ , where NS = nucleotide substitutions, ID = number of indels (indels of any size were quantified as 1) and L = locus length.

### 2.4. Phylogenetic Analysis

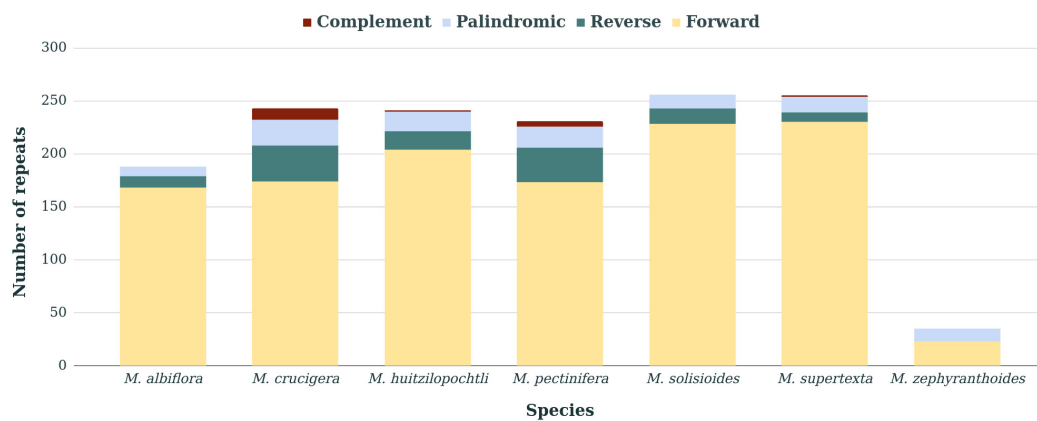
A total of 14 concatenated loci were used in order to reconstruct the phylogenetic relationships of the seven species. The 14 loci were selected from the 20 most variable homologous loci among the *Mammillaria* species that were also found in the outgroup genomes. The giant columnar cactus *Carnegiea gigantea* (Cactaceae) [21] and purslane *Portulaca oleracea* (Portulacaceae) [31] of the order Caryophyllales were used as outgroups. The sequences of the 14 loci were aligned with MAFFT v7.310 [30], adding a total length of 10,698 bp. The best evolutionary model was obtained with JMODELTEST v2.1.10 [32], which used Akaike Information Criterion (AIC). The best model obtained was GTR + G, which was used to construct the phylogenetic tree based on ML in RAXML-HPC v8.2.10 [33] with 1000 replicates.

### 3. Results

After analyzing the complete chloroplast genomes of seven *Mammillaria* species, we estimated the different levels of molecular variation among the RSs, SSRs and divergent DNA sequences in homologous loci.

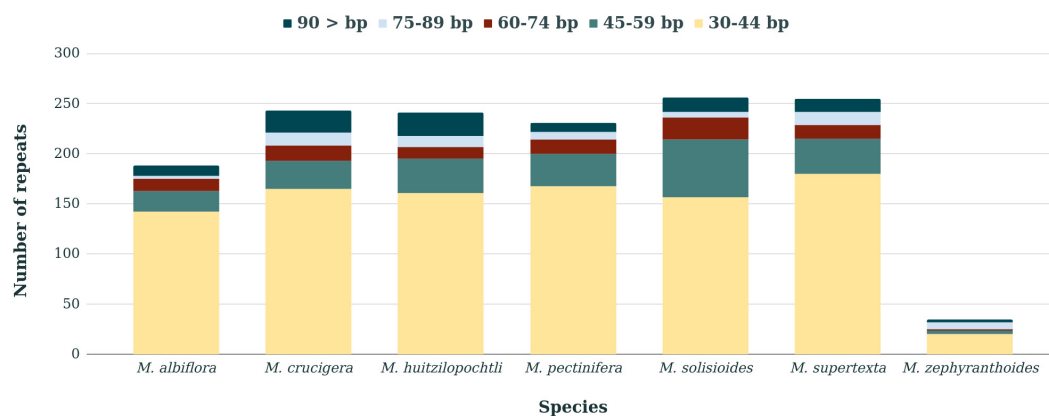
#### 3.1. Repeated Sequences

Repeated sequences longer than 30 bp were found in the chloroplast genomes of the seven species of *Mammillaria*. The number of large RSs varied from 35 in *M. zephyranthoides* to 256 in *M. solisioides* (Figure 1). In *M. zephyranthoides*, we found only forward and palindromic RSs. In contrast, in the six other species, forward, reverse and palindromic RSs were detected. Additionally, *M. crucigera*, *M. pectinifera*, *M. huitzilopochtli* and *M. supertexta* had between one to 11 complement repeats. The forward RSs were the most abundant in the chloroplast genomes of the seven species (Figure 1).



**Figure 1.** Number of repeated sequences (RSs) by type (complementary, palindromic, reverse and forward) identified in the seven species of *Mammillaria*.

The number of RSs showed a tendency to decrease with the length of the repeated sequences in the seven species. The repeats from 30 to 44 bp represented >50% of the total repeated sequences in the seven species, whereas longer repeated sequences from 75 to 89 bp were the scarcest in *M. albiflora*, *M. crucigera*, *M. huitzilopochtli*, *M. pectinifera*, *M. solisioides* and *M. supertexta*. Likewise, in *M. zephyranthoides*, the less frequently repeated sequences were those from 60 to 74 bp (Figure 2).



**Figure 2.** Number of repeated sequences identified in seven *Mammillaria* species. The color of the bars indicates the number of RSs by length in bp.

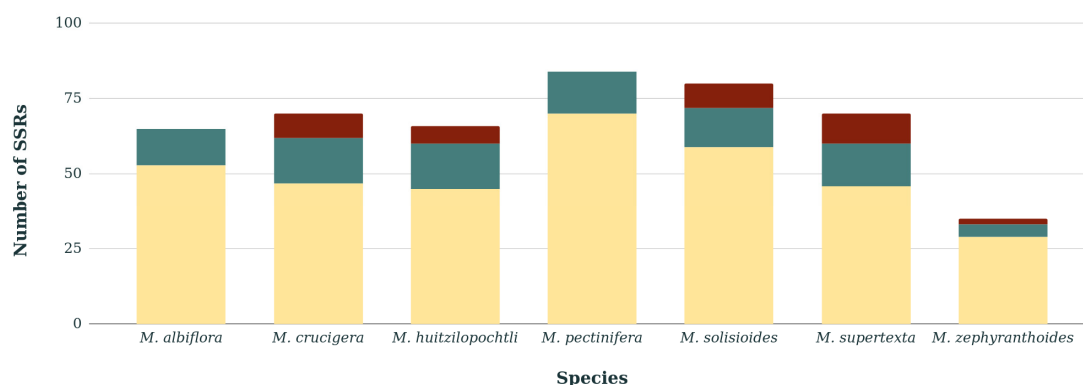
### 3.2. Variation and Localization of the Short Sequence Repeats

We found a number of short sequence repeats ranging from 35 (*M. zephyranthoides*) to 84 (*M. pectinifera*) in the chloroplast genomes of the seven *Mammillaria* species, with different lengths and nucleotide compositions (Table 1 and Table S2). The SSRs that were composed of a single nucleotide (i.e., homopolymer repeats) or two nucleotides (i.e., dinucleotide repeats) were the most abundant, and they were detected in the seven species. On the other hand, the SSRs composed of three nucleotides were scarce; however, five species (*M. albiflora*, *M. crucigera*, *M. pectinifera*, *M. solisioides* and *M. supertexta*) contained one to four of any of the three types of trinucleotide repeat sequences (Table 1). Lastly, SSRs composed of four nucleotides (i.e., tetranucleotide) were identified only in *M. crucigera* and *M. supertexta*, whereas repeated sequences composed of five nucleotides were only found in *M. solisioides*, and those SSRs with six nucleotides were only present in *M. zephyranthoides*.

**Table 1.** Type and number of Short Sequence Repeats (SSRs) in the complete chloroplast genomes of the seven analyzed species.

Type of SSRs	<i>M. albiflora</i>	<i>M. crucigera</i>	<i>M. huitzilopochtli</i>	<i>M. pectinifera</i>	<i>M. solisioides</i>	<i>M. supertexta</i>	<i>M. zephyranthoides</i>
1. Homopolymer							
A/T	56	60	57	69	65	60	31
C/G	1	1	1	2	2	1	0
2. Dinucleotide							
AT/AT	5	6	7	9	9	6	3
3. Trinucleotide							
AAG/CTT	2	1	0	3	2	1	0
AGG/CCT	1	0	0	0	0	0	0
AAT/ATT	0	0	0	1	1	0	0
4. Tetranucleotide							
AAAT/ATTT	0	2	0	0	0	2	0
5. Pentanucleotide							
AATAT/ATATT	0	0	0	0	1	0	0
6. Hexanucleotide							
ACGAGG/CCTCGT	0	0	0	0	0	0	1
Total number of SSRs							
	65	70	65	84	80	70	35

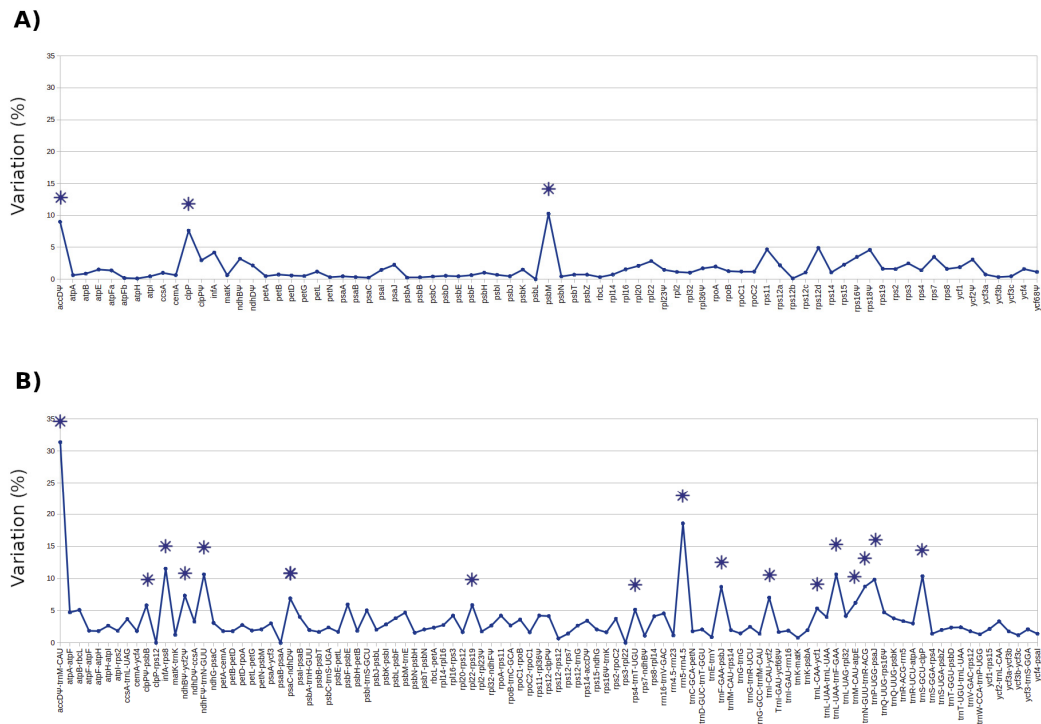
Most of the SSRs were identified in the LSC region of the chloroplast genome of the seven species, whereas SSRs were scarce in the IRs, and particularly in *M. albiflora* and *M. pectinifera* they were absent in the IRs (Figure 3). Within the homopolymer repeats, we found five homologous SSRs that were shared in the seven species, and of these only three had length variations between species (Table S3).



**Figure 3.** Total number of Short Sequence Repeats (SSRs) identified in the cpDNA of seven species of *Mammillaria*. The color of the bars corresponds to Inverted Repeats (red), Large Single Copy (yellow) and Small Single Copy (green).

### 3.3. Divergent DNA Sequences in Homologous Loci

In the seven species that were analyzed, a total of 180 homologous DNA loci were identified (Figure 4). Among them, 77 were coding and 103 were non-coding loci. The paired comparisons showed that coding loci had a molecular variation that oscillated from 0 to 10.26%, and non-coding loci showed levels from 0 to 31.37% (Figure 4A,B and Table S4).



**Figure 4.** Distribution of the percentage of molecular variation across 180 homologous loci. Average percentage variation of the (A) coding and (B) non-coding homologous loci in the seven chloroplast genomes of *Mammillaria*. The 20 loci with the highest molecular variation are indicated with asterisks (\*).

Of the 180 homologous loci, ~77% were located in the LSC region of the chloroplast genome, ~16% in the SSC region and ~7% in the IRs. Since the seven *Mammillaria* species had three distinct cpDNA arrangements, the number of loci in each structural region differed between species (Table 2).

**Table 2.** Distribution of the 180 homologous loci identified in the seven *Mammillaria* species. The number of loci in each structural region of the chloroplast genome is shown.

Species	Number of Loci		
	LSC Region	SSC Region	IRs Region
Arrangement 1 ( <i>M. albiflora</i> and <i>M. pectinifera</i> )	144	132	142
Arrangement 2 ( <i>M. crucigera</i> , <i>M. huitzilopochtli</i> , <i>M. solisoides</i> and <i>M. supertexta</i> )	34	34	17
Arrangement 3 ( <i>M. zephyranthoides</i> )	2	14	21

Among the 180 loci, we identified the 20 loci with the highest average levels of molecular variation (Figure 4), ranging from 5.14 to 31.37% (Table 3). Among these loci, 17 were IGS, and three were genes or potential pseudogenes.

**Table 3.** Percentage of variation and length of the 20 most variable chloroplast loci recorded in the seven *Mammillaria* species. The type of locus (gene, pseudogene or IGS) and the structural region of the genome where it is located are shown. SSC/IRs indicates that the locus is in the SSC region in all the species, except *M. zephyranthoides*, which has it in IRs. \* indicates the 14 loci that were used to obtain the phylogenetic relationships of the seven studied species. The last column represents the average length (bp) in the seven species.

Locus Name	Locus Type	Location	Variation Percentage (%)	Average Length (bp)
1. <i>accD-trnM</i> *	IGS	LSC	31.37	2176
2. <i>rrn5-rrn4.5</i> *	IGS	SSC	18.56	721
3. <i>infA-rps8</i> *	IGS	LSC	11.52	103
4. <i>trnL-trnF</i> *	IGS	LSC	10.64	325
5. <i>ndhF-trnN</i>	IGS	SSC	10.63	455
6. <i>trnS-clpP</i>	IGS	LSC	10.34	89
7. <i>psbM</i> *	Gene	LSC	10.26	102
8. <i>trnP-psaJ</i> *	IGS	LSC	9.82	399
9. <i>accDψ</i>	Pseudogene	LSC	8.99	1574
10. <i>trnN-trnR</i> *	IGS	SSC/IRs	8.74	360
11. <i>trnF-psbJ</i>	IGS	LSC	8.68	619
12. <i>clpP</i> *	Gene	LSC	7.61	616
13. <i>ndhB-ycf2</i>	IGS	SSC/IRs	7.34	354
14. <i>trnI-ycf2</i> *	IGS	SSC/IRs	7.02	109
15. <i>psaC-ndhD</i> *	IGS	SSC	6.90	146
16. <i>trnM-atpE</i> *	IGS	LSC	6.21	227
17. <i>rpl22-rps19</i> *	IGS	LSC	5.87	67
18. <i>clpP-psbB</i> *	IGS	LSC	5.82	753
19. <i>trnL-ycf1</i>	IGS	SSC	5.34	256
20. <i>rps4-trnT</i> *	IGS	LSC	5.14	277

#### 4. Discussion

In this work, we evaluated the chloroplast genomes of seven *Mammillaria* species (*M. albiflora*, *M. crucigera*, *M. huitzilopochtli*, *M. pectinifera*, *M. solisioides*, *M. supertexta* and *M. zephyranthoides*), in order to characterize their genetic variation and identify potential molecular markers for evolutionary biology studies. Our results showed that the high levels of quantified variation might enable the resolution of phylogenetic relationships between the species and genus levels.

A previous study found that there are structural variations in the *Mammillaria* chloroplast genomes, identifying three different arrangements in the seven analyzed species [20]. Repeated DNA sequences have been proposed to produce chloroplast genome instability [34]. This idea is based on empirical evidence that plant species with strongly rearranged chloroplast genomes have a larger proportion of repeated sequences than those without a structural rearrangement [35,36]. However, the relatively low proportions of repeated sequences detected in *M. zephyranthoides* do not support such a notion, since a previous study documented that *M. zephyranthoides* shows deep and strong structural changes [20]. Our results do not allow for the conclusion that, in *Mammillaria*, the type of IRs is correlated to the number of RS contained in them.

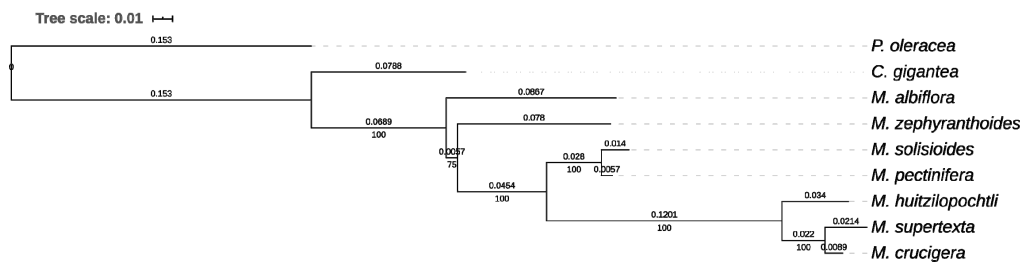
With respect to SSRs, these have been used for a long time as a source of molecular variation. In particular, they show high mutation rates at the intraspecific level [37], mainly due to variations in length [38]. In the cpDNA of *Mammillaria* here analyzed, homopolymer SSRs were the most abundant, and SSRs with >3 nucleotides were scarce or absent, as has been reported in other chloroplast genomes of angiosperms [39,40]. In addition, we concluded that the length and the number of genes involved in the formation of IRs are not the factors that determine the abundance of SSRs, since these were absent in IRs of *M. albiflora* and *M. pectinifera* that have only three genes and a length of <1000 bp [20]. However, *M. zephyranthoides* has divergent IRs of 14 kb and 21 or 22 genes [20], but only contains two SSRs (Figure 3).

In addition, we searched for the variation in SSRs among the seven species, in order to identify potential molecular markers. However, only five SSRs were identified in the seven species, and only three of them showed variation between species (Table S3). Thus, we do not recommend the use of SSRs of the chloroplast genome as unique molecular markers to resolve phylogenetic, phylogeographic or population issues, but these may be used together with other markers. These SSRs may be used not only for *Mammillaria* species, most of them being included in the IUCN Red List [25], but also extensively for the remaining Cactaceae and Caryophyllales members.

The homologous loci are potential sources of variation that should be examined and selected according to the problem that is to be resolved (i.e., population or phylogenetic levels), since the levels of variation differ among them. In particular, there are 20 loci that showed a high molecular variation. We encourage their use for undertaking phylogenetic studies, as well as phylogeographic and population genetics studies. Notably, seven of these loci showed variation levels >10%, which represents an unusually higher level of molecular variation, and, consequently, they might be further tested in population level studies. Indeed, we consider that these loci may be tested in those angiosperms that show rapid radiations or a recent origin, since these are conditions that reduce molecular variation [41,42]. In the particular case of *Mammillaria*, its recent origin (<8 Ma) [26,27] could probably be causing the low levels of variation found in the few markers conventionally used in the preliminary phylogenetic studies carried out on both Cactaceae and *Mammillaria*. However, our results showed that, despite the recent origin of *Mammillaria* [26,27], its chloroplast genomes have DNA sequences with high levels of molecular variation, which were evident only by using the perspective of comparative genomics. Out of the 20 most variable loci here proposed as a source of molecular variations, only the *trnL-trnF* IGS, with an average 10.64% of divergence, was previously used, in a phylogenetic study that analyzed 21 species of *Mammillaria* [15]; however, this marker did not resolve relationships. Based on these results, we recommend a multiloci sampling to improve the resolution of the analysis and expand the taxonomic sampling. Moreover, 12 of the 20 most variable loci analyzed in this paper have amplicons (150–1000 bp) that can be sequenced via the Sanger method. Accordingly, these loci could be easily obtained at relatively low costs. In addition, out of convention, two IGS (*trnK-matK*, *psbA-trnH*) have been recommended and widely used for phylogenetic studies because they were considered as having a high molecular variation in angiosperms [43,44]. They have also been used in *Mammillaria* [23], with poor results. Consequently, we do not recommend the use of these two IGS as molecular markers, since they showed low percentages of variation (0.8 and 2%, respectively).

In order to test the feasibility of using the homologous loci values as molecular markers because of their high variation, we used 14 of them to construct a phylogenetic tree (Figure 5). The tree we obtained showed a topology similar to the one that had been previously obtained using 42 coding regions [20]. This result suggests that the small number of molecular markers here proposed could reduce the costs in phylogenetic studies with satisfactory results. Lastly, we encourage the testing of these 20 loci as molecular markers, in order to resolve the unclear limits of genera such as *Coryphantha*, *Escobaria*, *Neolloydia*, *Ortegocactus* and *Pelecyphora*. In addition, the use of these markers may also be considered for undertaking phylogenetic studies in other land flowering plants that require different options for molecular markers with high levels of variation.





**Figure 5.** Phylogenetic ML tree for the seven studied *Mammillaria* species based on 14 loci of high molecular variation (marked with an asterisk in Table 3), with *Portulaca oleracea* and *Carnegiea gigantea* used as outgroups. The branch length is shown above the branches, whereas bootstrap values are displayed below the branches.

## 5. Conclusions

In this study, the comparison of the complete chloroplast genome revealed new loci of high molecular variation in *Mammillaria* that were overlooked without a comparative genomics perspective. The high levels of molecular variation of these loci may open a new era in phylogenetic, population and taxonomic studies, as well as, in conservation issues with *Mammillaria* genus, the richest taxa of Cactaceae.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/11/7/830/s1>. Table S1: The species of *Mammillaria* analyzed in this study with their respective GeneBank accession number. Table S2: Nucleotide composition of SSRs (in parenthesis) and number of repeats (out of parenthesis) identified in the complete chloroplast genome of seven *Mammillaria* species. The location (start-end) of these SSRs is referred to the DNA sequences with the accession number shown in Table S1. Table S3: Composition and location of the five homologous SSRs identified in the seven *Mammillaria* species. Table S4: Percentage of molecular variation estimated between the seven *Mammillaria* species in paired comparisons of homologous loci (77 coding regions, followed by 103 non-coding regions).

**Author Contributions:** S.S. is the researcher leading of this study; she designed research and obtained the financial support. She and D.A.C. prepared the draft of the complete manuscript. D.A.C. and K.E. processed genomic data, and they were supervised by A.S.-F.; P.D. provided financial support. A.G.-R., C.E.D.-V., F.V.-P., P.D. and S.A., reviewed draft versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is supported by grants from the Dirección General de Personal Académico of UNAM for research projects (DGAPA-PAPIIT IN222216 and IN228619).

**Acknowledgments:** This article is part of Master Studies of Delil A. Chincoya (310003715), who carried out her studies at Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México she was distinguished with a fellowship of CONACyT (855061). We are grateful to Unidad de Secuenciación Masiva y Bioinformática of the Laboratorio Nacional de Apoyo Tecnológico a las Ciencias Genómicas, at the Instituto de Biotecnología, UNAM for advice in bioinformatics analysis. The comments of three anonymous reviewers, and editor M. Wartenberg improved the quality of this Article.

**Conflicts of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Lemieux, C.; Otis, C.; Turmel, M. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **2000**, *403*, 649–652. [[CrossRef](#)]
- Wicke, S.; Schneeweiss, G.M.; dePamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)]
- Lavin, M.; Doyle, J.J.; Palmer, J.D. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* **1990**, *44*, 390–402. [[CrossRef](#)] [[PubMed](#)]
- Downie, S.R.; Palmer, J.D. A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeat restriction site variation. *Syst. Bot.* **1994**, *19*, 236–252. [[CrossRef](#)]

5. Cosner, M.E.; Raubeson, L.A.; Jansen, R.K. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. *BMC Evol. Biol.* **2004**, *4*, 27. [[CrossRef](#)]
6. Graham, S.W.; Reeves, P.A.; Burns, A.C.E.; Olmstead, R.G. Microstructural Changes in Noncoding Chloroplast DNA: Interpretation, Evolution, and Utility of Indels and Inversions in Basal Angiosperm Phylogenetic Inference. *Int. J. Plant Sci.* **2000**, *161*, S83–S96. [[CrossRef](#)]
7. Kelchner, S.A. The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics. *Ann. Mo. Bot. Gard.* **2000**, *87*, 482. [[CrossRef](#)]
8. Borsch, T.; Quandt, D.; Koch, M. Molecular evolution and phylogenetic utility of non-coding DNA: Applications from species to deep level questions. *Plant Syst. Evol.* **2009**, *282*, 107–108. [[CrossRef](#)]
9. Taberlet, P.; Gielly, L.; Pautou, G.; Bouvet, J. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **1991**, *17*, 1105–1109. [[CrossRef](#)] [[PubMed](#)]
10. Hamilton, M.B. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molec. Ecol.* **1999**, *8*, 521–523.
11. Steele, K.P.; Wojciechowski, M.F. Phylogenetic analyses of tribes Trifolieae and Viciae, based on sequences of the plastid gene *matK* (Papilionoideae: Leguminosae). In *Advances in Legume Systematics, Part 10, Higher Level Systematics*; Klitgaard, B.B., Bruneau, A., Eds.; Royal Botanic Garden: Kew, UK, 2003; pp. 355–370.
12. Kelchner, S.A.; Clark, L.G. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Mol. Phylogenet. Evol.* **1997**, *8*, 385–397. [[CrossRef](#)] [[PubMed](#)]
13. Prince, L.M.; Parks, C.R. Phylogenetic relationships of Theaceae inferred from chloroplast DNA sequence data. *Am. J. Bot.* **2001**, *88*, 2309–2320. [[CrossRef](#)] [[PubMed](#)]
14. Miller, J.T.; Bayer, R.J. Molecular phylogenetics of *Acacia* subgenera *Acacia* and *Aculeiferum* (Fabaceae: Mimosoideae), based on the chloroplast *matK* coding sequence and flanking *trnK* intron spacer regions. *Aust. Syst. Bot.* **2003**, *16*, 27–33. [[CrossRef](#)]
15. Harpke, D.; Peterson, A.; Hoffmann, M.H.; Röser, M. Phylogenetic evaluation of chloroplast *trnL-trnF* DNA sequence variation in the genus *Mammillaria* (Cactaceae). *Schlechtendalia* **2006**, *14*, 7–16.
16. Nyffeler, R. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/matK* and *trnL-trnF* sequences. *Am. J. Bot.* **2002**, *89*, 312–326. [[CrossRef](#)]
17. Mast, A.R.; Kelso, S.; Richards, A.J.; Lang, D.J.; Feller, D.M.S.; Conti, E. Phylogenetic Relationships in *Primula* L. and Related Genera (Primulaceae) Based on Noncoding Chloroplast DNA. *Int. J. Plant Sci.* **2001**, *162*, 1381–1400.
18. Cuenoud, P.; Savolainen, V.; Chatrou, L.W.; Powell, M.; Grayer, R.J.; Chase, M.W. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Am. J. Bot.* **2002**, *89*, 132–144. [[CrossRef](#)] [[PubMed](#)]
19. Borsch, T.; Hilu, K.W.; Quandt, D.; Wilde, V.; Neinhuis, C.; Barthlott, W. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *J. Evol. Biol.* **2003**, *16*, 558–576. [[CrossRef](#)]
20. Solórzano, S.; Chincoya, D.A.; Sanchez-Flores, A.; Estrada, K.; Díaz-Velásquez, C.E.; González-Rodríguez, A.; Vaca-Paniagua, F.; Dávila, P.; Arias, S. De novo assembly discovered novel structures in genome of plastids and revealed divergent inverted repeats in *Mammillaria* (Cactaceae, Caryophyllales). *Plants* **2019**, *8*, 392. [[CrossRef](#)]
21. Sanderson, M.J.; Copetti, D.; Búrquez, A.; Bustamante, E.; Charboneau, J.L.M.; Eguiarte, L.; Kumar, S.; Lee, H.O.; McMahon, M.; Steele, K.; et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*). *Am. J. Bot.* **2015**, *102*, 1115–1127. [[CrossRef](#)]
22. Bárcenas, R.T.; Yesson, C.; Hawkins, J.A. Molecular systematics of the Cactaceae. *Cladistics* **2011**, *27*, 470–489. [[CrossRef](#)]
23. Butterworth, C.A.; Wallace, R.S. Phylogenetic studies of *Mammillaria* (Cactaceae)—Insights from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *Am. J. Bot.* **2004**, *91*, 1086–1098. [[CrossRef](#)] [[PubMed](#)]
24. Hunt, D.; Taylor, N.; Charles, G. *The New Cactus Lexicon*; DH Books: Milborne Port, UK, 2006.
25. The IUCN Red List of Threatened Species. Available online: <https://www.iucnredlist.org/search/grid?query=mammillaria&searchType=species> (accessed on 15 January 2020).
26. Arakaki, M.; Christin, P.-A.; Nyffeler, R.; Lendel, A.; Eggli, U.; Ogburn, R.M.; Spriggs, E.; Moore, M.J.; Edwards, E.J. Contemporaneous and recent radiations of the world’s major succulent plant lineages. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8379–8384. [[CrossRef](#)]

27. Hernández-Hernández, T.; Brown, J.W.; Schlumpberger, B.O.; Eguiarte, L.E.; Magallón, S. Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol.* **2014**, *202*, 1382–1397. [[CrossRef](#)] [[PubMed](#)]
28. Kurtz, S. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)]
29. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)]
30. Katoh, K. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
31. Liu, X.; Yang, H.; Zhao, J.; Zhou, B.; Li, T.; Xiang, B. The complete chloroplast genome sequence of the folk medicinal and vegetable plant purslane (*Portulaca oleracea* L.). *J. Hortic. Sci. Biotech.* **2018**, *93*, 356–365. [[CrossRef](#)]
32. Posada, D. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* **2008**, *25*, 1253–1256. [[CrossRef](#)]
33. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
34. Weng, M.-L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [[CrossRef](#)] [[PubMed](#)]
35. Haberle, R.C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **2008**, *66*, 350–361.
36. Blazier, J.C.; Jansen, R.K.; Mower, J.P.; Govindu, M.; Zhang, J.; Weng, M.-L.; Ruhlman, T.A. Variable presence of the inverted repeat and plastome stability in *Erodium*. *Ann. Bot.* **2016**, *117*, 1209–1220. [[CrossRef](#)] [[PubMed](#)]
37. Sunnucks, P. Efficient genetic markers for population biology. *Trends Ecol. Evol.* **2000**, *15*, 199–203. [[CrossRef](#)]
38. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445.
39. Huang, H.; Shi, C.; Liu, Y.; Mao, S.-Y.; Gao, L.-Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **2014**, *14*, 151. [[CrossRef](#)]
40. Dong, W.-L.; Wang, R.-N.; Zhang, N.-Y.; Fan, W.-B.; Fang, M.-F.; Li, Z.-H. Molecular Evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* **2018**, *19*, 716. [[CrossRef](#)]
41. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [[CrossRef](#)]
42. Whitfield, J.B.; Lockhart, P.J. Deciphering ancient rapid radiations. *Trends Ecol. Evol.* **2007**, *22*, 258–265. [[CrossRef](#)]
43. Kress, W.J.; Wurdack, K.J.; Zimmer, E.A.; Weigt, L.A.; Janzen, D.H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8369–8374. [[CrossRef](#)]
44. Wicke, S.; Quandt, D. Universal primers for the amplification of the plastid *trnK/matK* region in land plants. *Anales Jard. Bot. Madrid* **2009**, *66*, 285–288.

