Journal of
Phylogenetics & Evolutionary Biology

Talat, J Phylogenetics Evol Biol 2018, 6:4
DOI: 10.4172/2329-9002.1000208

Research Article Open Access

Comparative Bioinformatics Analyses of the Chloroplast Genomes of *Vitis vinifera* with Two Caucasica Subspecies of Grape Fruit

F. Talat [1*], S. Shahdparvar[2], and M. Badri Anarjan[3]

[1]*Assistant Professor, Seed and Plant Improvement Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran*
[2]*Master Graduate of Horticultural Biotechnology, Saba Institute of Education, Urmia, Iran*
[3]*Master Graduate of Plant Breeding, Urmia University, Urmia, Iran*

Abstract

Grape (*Vitis vinifera*) is a genus of tree in the family of Vitaceae. *V. vinifera's* species belong to Eurasian grapes. The genome of chloroplast is the most comprehensive genome in plants, and it has many features for evolution analyses due to the unique molecular structure and single-parent inheritance. The goals of this research were study and compare of the complete sequences chloroplast genomes of *Saperavi* and *Meskhuri mtsvane* from Caucasia subspecies with common grape (*Vitis vinifera*), as well as genomes structure analysis, gene content, organization and repetitive sequences, codon usage and comparison among genomes. The chloroplast genome of *Vitis vinifera* is a circular DNA molecule with 160928 base pairs (bp), which is longer than the chloroplast genomes of *Saperavi* and *Meskhuri mtsvane* varieties. The large and small unique regions are separated by two inverted repeat regions a, b. In three genomes, the complete genome contains 131 genes, which include 79 protein coding genes, 4 rRNA genes and 30 tRNA genes. In other words, there are totally 113 single-copy genes and 18 double-copy genes located in inverted repeat region (IR) in the three studied genomes. The SSRs of the chloroplast genomes were identified and the results indicated that the chloroplast genomes of *Vitis vinifera* and *Saperavi* both have 74 SSRs and *Meskhuri mtsvane* has 73 SSRs. The chloroplast SSRs are important and useful for genetic diversity studies. Low GC content is a significant feature of plastidic genomes, which is possibly formed after endosymbiosis by DNA replication and repair.

Keywords: *Vitis vinifera*; *Saperavi*; *Meskhuri mtsvane*; Chloroplast genome; Complete sequence

Introduction

The most plastidic genomes contain a pair of inverted repeat regions (IRa, IRb) with 25000 bp each, which are separated by small and large single copy regions of 20000 bp and 80000 bp, respectively. Some plant like of *Vicia faba* [1] and *Cryptomeria japonica* break this structural conservation by losing of an IR [2,3]. Lack of complete chloroplast genome sequences are still one of the major limitations, and the complete chloroplast genome sequences is useful to expanding chloroplast genetic engineering technology of crops [3]. The using of DNA sequences from all of the shared chloroplast genes provide many characters for phylogeny reconstruction compared to previous studies that they have relied on only one or a few genes to address the same questions [4]. However, the all genomes can limit estimation misleading of relationships because of taxon sampling [5-7] and the using of incorrect models of sequence evolution in concatenated datasets [8,9]. Thus, there is a growing interest in expanding the taxon sampling of complete chloroplast genome sequences and developing new evolutionary models for phylogenetic analysis of chloroplast sequences [10] to overcome these concerns. The completely sequencing of chloroplast genomes provide a rich source of data that it can be used to address phylogenetic questions [2-6]. Chloroplasts have a low mutation rate with a great deal of conservation in their structure, genome size, gene content and organization. In previous studies it has been reported that, chloroplast genes like *16S, 23S, ndhB, psbA, psbB, psbC, psbD, pasA, psaB* and *rbcL* are suitable to study the relationship among higher plants; *ycf1, ycf2, accD, matK, rpoC2* and *ndhF* are more compatible to study the relationship of the close species [11]. The complete sequencing of chloroplast genome of *Vitis* also provides valuable data for using chloroplast genetic engineering for this economically important plant [12]. Transcriptomics have been famed to be a potent tool to improve the plant genetic architecture and high expression of the foreign protein, low risk of the pollen pollution [13] and no gene silencing [14]. So it is necessary to find out the chloroplast genome in order to design our next generation transcriptomics. All the grapes in the world belong to the Ampelidaceae family, which is also called Sarmentaceae or Vitaceae family, that this family has more than fourteen generas and six hundred species. Among them, the most important genus is *vitis*, which has two subgenus with different chromosome numbers. *Muscadinae* subgenus is wild and has fourteen chromosomes (2n=2x=40). In the sub-genus *Ovitis*, which is the second subgenus of this family, there are species that they have thirty-eight chromosomes (2n=2x=38). The only specie that it has been domesticated is *Vinifera*, which has many varieties and is common in many parts of the world due to the high quality and quantity of crop cultivation. This study was managed to study and comparing the complete sequencing of chloroplast of *Vitis vinifera*, analyses of its genome structure, gene content, organization, repeat sequence and codon usage. At the same time, the comparison of three sequencing Grape species were performed.

Materials and Methods

Complete chloroplast genome sequence of *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* in FASTA format, with respectively access number (NC_007957.1), (AB856290.1), (AB856291.1) downloaded from NCBI (//http:www.ncbi.nlm.nih.gov/nuccore/term). The studied genomes were annotated with using of DOGMA (Dual Organellar Genome Annotator) [15], after uploading a FASTA-formation file of the complete plastid genome to the program's server. BLASTX and BLASTN searches against a custom database of previously published plastid genomes identified *Vitis* putative protein-coding genes and

***Corresponding author:** Assistant Professor, Seed and Pant Improvement Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran, Tel: 0098 9104055682; E-mail: f.talat@areeo.ac.ir

Received November 20, 2018; **Accepted** December 13, 2018; **Published** December 21, 2018

Citation: Talat F, Shahdparvar S, Anarjan MB (2018) Comparative Bioinformatics Analyses of the Chloroplast Genomes of *Vitis vinifera* with Two Caucasica Subspecies of Grape Fruit. J Phylogenetics Evol Biol 6: 208. doi: 10.4172/2329-9002.1000208

tRNAs or rRNAs. For genes with low sequence identity, the manual annotation were performed, after identifying the position of the start and stop codons, as well as the translated amino acid sequence with using of the plastid genetic code. Gene map and gene distribution of *Vitis vinfera*, *Saperavi* and *Meskhuri mtsvaneh* were performed by OGDRAW V1.1 (Organellar Genome DRAW, http://ogdraw.mpimpgolm.mpg.de/) [16]. REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer/) [17] was used in order to locate and count the forward, reverse, complement and palindromic repeats within the intended genomes. For repeat identification, the following constraints were set to REPuter: (i) minimum repeat size of 30 bp and (ii) 90% or greater sequence identity, based on hamming distance of 3 [3]. Numbers of codons (NCs), the synonymous relative of codon usage (RSCU) and GC composition of codons were calculated for each gene. The analysis was carried out by CODONW 1.4 (http://codonw.sourceforge.net/). Correspondence analysis (COA) has become the method of choice for multivariate statistical analysis of codon usage [18,19]. Since there are 59 synonymous codons (61 senses codons less the unique methionine and tryptophan codons).

Results and Discussion

Overall structure

The complete chloroplast genome of *Vitis vinifera* has 160,928 bp length (Figure 1), which is longer than *Saperavi* (160,927 bp) and *Meskhuri mtsvane* (160,906 bp). Total genome of *Vitis vinifera* includes a pair of inverted repeats with 26,358 bp long, separated by a small and large single copy regions with 19,065 bp and 89,147 bp, respectively. The length of LSC region in *Vitis vinifera* is longer than *Saperavi* and *Meskhuri mtsvane*. But the length of SSC and IR regions are same in *Vitis vinifera* and *Saperavi*, which are different from *Meskhuri mtsvane*. The coding region of *Vitis vinifera's* chloroplast genome is 91,872 bp in length, accounting for 57.08% of the whole plastidic genome, which is same with *Saperavi*, and similar to *Meskhuri mtsvane* by 57.14%, genus *Alsophila* 53.2% [20], *Dendrocalamus latiforus* 53.4% [21], genus *Megaleranthis* 52.4% [11]. The chloroplast genome of *Vitis vinifera* codes for proteins (49.94%), tRNA genes (1.73%) and rRNA (5.61%), similar to coffee [22] and *M. esculenta* [23]. The non-coding region has a length with 69,661 bp (43.28% of the genome). The proportions of intragenic spacers and intron are 31.10% and 12.18% respectively.

Repeat structure

Four types of repeats were detected; forward (direct) match, reverse match, complements match and palindromic (inverted) match. Forty nine repeats have a length more than 18 bp that they are shown in Table 1. The complete chloroplast genome of *Vitis vinifera* includes

Repeat sequence detected in chloroplast genome of *Vitis vinifera*			
Number	**Size (bp)**	**Location**	**Match Direction**
1	20	IGS	C
2	18	IGS	F
3	18	IGS	F
4	20	IGS	F
5	20	IGS	F
6	20	IGS,ycf2	F
7	21	Trnfm-CAU, trnP-UGG	F
8	21	Trns-GCU, trns-UGA	F
9	22	IGS	F
10	22	IGS	F
11	22	trnG-GCC	F
12	24	ACCD	F
13	26	IGS,INTRON	F
14	30	ycf2	F
15	30	ycf2	F
16	31	IGS	F
17	48	ycf2	F
18	48	ycf2	F
19	18	Intron, IGS	P
20	18	IGS	P
21	18	IGS	P
22	18	IGS	P
23	19	IGS	P
24	20	IGS	P
25	20	Trns-UGA, trns-GGA	P
26	20	IGS,ycf2	P
27	20	IGS	P
28	21	Trns-UGA, trns-GCA	P
29	22	IGS	P
30	22	IGS	P
31	22	IGS	P
32	26	Intron, IGS	P
33	30	Ycf2	P

34	30	Ycf2	P
35	30	IGS	P
36	30	IGS, trns-GGA	P
37	31	IGS,CCSA	P
38	48	Ycf2	P
39	48	Ycf2	P
40	54	IGS	P
41	18	IGS, intron	R
42	19	IGS	R
43	19	IGS	R
44	20	IGS	R
45	20	atpB	R
46	20	IGS	R
47	21	Intron	R
48	21	IGS	R
49	29	IGS	R

Table 1: Repeat sequences detected in chloroplast genome of *Vitis vinifera*. Note: IGS represents intergenic spacer sequence. F represents forward (direct) match, R represents reverse match, C represents complement match, P represents palindromic (invert) match.

one complementary repeat, seventeen forward repeats, twenty one inverted repeats and nine reverse repeats. The majority of the repeats were located within ycf2 and intergenic spacer regions and few located at ACCD, trnG, trnP, trnS, and intron sequences. The largest repeat in *Vitis vinifera* and *Saperavi* is fifty four bp, which is located in IGS. But the largest repeat in *Meskhuri mtsvane* is seventy five bp, which is located in IGS and rps19, while the most of repeats in all intended genomes of this study are 18-30 bp. Furthermore, 74 simple sequence repeats in *Vitis vinifera*, *Saperavi* and also 73 chloroplast SSRs in *Meskhuri mtsvane* were obtained. The longest repeat is the repeat of "AT", which is 18 bp, but the most of the repeats are A and T in all three genomes of this study (Table 2).

Repeat	Repeat Sequence	Number	Max (bp)
Mononucleotide	A	24	16
	C	1	11
	T	31	15
Dinucleotide	AT	5	16
	TA	4	12
Trinucleotide	AAT	1	12
	ATA	4	15
	CAG	1	12
	GAA	1	12
	TAT	1	15
	TTA	1	12

Table 2: Simple sequence repeat (SSR) in *Vitis vinifera*.

Gene content

All genes coded by the chloroplast genome of *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* are detected and listed in Table 3. The results showed that the genome analyzed in this study contains 113 unique genes, 18 of which are duplicated in the IR, for a total of 131 genes (Figure 1). There are four ribosomal and 30 distinct tRNA genes, seven tRNA genes and all rRNA genes are duplicated within the IR. The identified genes can be classified according to the gene function, the functional genetic system gene, the photosynthetic system, the biosynthesis and some with unknown function. There are five genes with unknown function in *Vitis vinifera*'s chloroplast genome (*ycf* gene) that they were detected and were highly conserved between species [24]. *Rps12* is the name of two genes, which has an intron [11], this gene was separated (by an intron) into two parts with including two exons, locating at LSC (5′- end) and IR (3′-end). The second important gene of chloroplast genome is *matK*, which has 1.5 kbp lengths, and it was detected in the intron of *trnK-UUU,* and also it is the only gene located in an intron and encodes maturase k [14]. This gene has both conserved and variable fragments [25] Thus, it is frequently used in phylogenetic studies [26-29].

Codon usage analysis

The codon usage was analyzed and is in Table 4, ATG and TGG codes for Methionine and Tryptophan have RSCU value=1 [14]. RSCU of the three terminal codons TAA, TGA and TAG checked out for three intended chloroplast genomes. According to RSCU values, *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* prefer TAA as their stop codons. The analysis of the composition for the codons showed that A+T content at the third position and in all three intended genomes, the A+T were 71.3%, which is similar to what was reported for *Aalophile* [20] and *Panax schinseng nees* [1]. In NC-plot distribution, ENC (effective number of codons) and GC3s values were calculated (Figure 2). The heterogeneity of codon usage was further confirmed from the GC3s values ranging from 17% to 58% with a mean of 41% and standard deviation of 10%. If the codon usage bias is completely dictated by GC3s the values of NC should fall on the expected curve between GC3s and NC-plot of the *Vitis vinifera* chloroplast genome shown in Figure 2. The NC values which lie below the expected curve, indicating that these genes have additional codon usage bias apart from compositional bias (Figure 2).

COA analysis

Since there are 59 synonymous codons (61 sense codons less the unique methionine and tryptophan codons), according to *Vitis vinifera, Saperavi* and *Meskhuri mtsvane* chloroplast gene functions, 54 sequences can be classified six categories, the number of the first classified gene is 13, which encoding ribosomal protein. The *rpl* and *rps* genes encode large and small subunit ribosomal protein, the number of the second classified genes is 15, including of *psa* gene, *psb* gene, *atp* gene, *pet* gene and *rbcL* gene, also the third category is a conservative gene, *ycf*. The fourth category is translation apparatus genes, including the *rpo* gene of the RNA polymerase gene family. The fifth type is the miscellaneous proteins gene, for example *accD*, the sixth category is unknown function and hypothetical protein gene. Figure 2 shows the diversity among genes in terms of RSCU. In the leftmost of first axis, genes with the greatest codon bias are located, and those with lowest

	Group	Gene Name
Protein genes	Subunit of Acetyl-CoA-carboxylase	*AccD*
	Small subunit of rubisco	*rbcL*
	Subunit of NADH-dehydrogenase	*ndhA**, *ndhB**§, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	Subunits of cytochrome b/f complex	*petN*, *petA*, *petL*, *petG*, *petB**, *petD**
	Subunits of photosystem I AND II	*psbA*, *psbK*, *psbI*, *psbM*, *psbD*, *psbC*, *psaB*, *psaA*, *psaI*, *psbJ*, *psbL*, *psbF*, *psbE*, *psaJ*, *psbT*, *psbN*, *psbH*, *psaC*, *psbZ*, *psbB*
	DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC2*, *rpoC1**
	Large subunit of ribosome	*rpl2**§, *rpl14*, *rpl16**, *rpl20*, *rpl22*, *rpl23*§, *rpl32*, *rpl33*, *rpl36*,
	Small subunit of ribosome	*rps2*, *rps3*, *rps4*, *rps7*§, *rps8*, *rps11*, *rps12**§, *rps14*, *rps15*, *rps16**, *rps18*, *rps19*
	Others	*matK*, *cemA*, *clpP***, *infA*, *ccsA*
	Function unknown	*ycf3***, *ycf4*, *ycf1*§, *ycf2*§
RNA genes	Ribosomal RNA genes	*rrn16*§, *rrn23*§, *rrn4.5*§, *rrn5*§
	Transfer RNA genes	trnH-GUG, trnK-UUU*, trnQ-UUG, trnS-GCU, trnG-GCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA*, trnF-GAA, trnV-UAC*, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU§, trnL-CAA§, trnV-GAC§, trnI-GAU*§, trnA-UGC*§, trnR-ACG§, trnN-GUU§, trnL-UAG,

Note: § reflects gene located in IR; * reflects gene which has one intron; ** reflects gene which has two introns

Table 3: Genes coded by *Vitis vinifera* chloroplast genome.

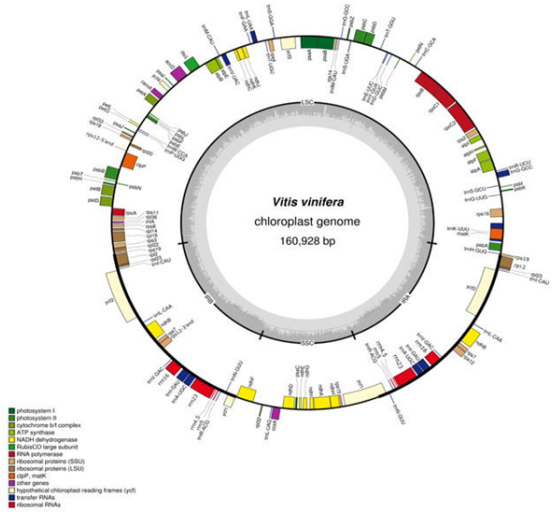

Figure 1: *Vitis vinifera's* chloroplast genome structure and gene organization. Note: Genes shown outside of the circle transcribed anticlockwise and shown inside of the circle transcribed clockwise. tRNA genes are shown by one letter of the coded amino acid followed by anticodon (genome map was created by using OGDRAW V.1.1. [16].

expansion and contraction based on genome size is considerably important. The extent of IR often affects the large size of the genome and also the pseudo genes of IR, LSC or SSC junctions. In this study, the differences of the junctions between *Vitis vinifera*, *Saperavi, Meskhuri mtsvane* and other 3 species (*Solanum lycopersicum*, *Arabidopsis thaliana*, *Spinacia oleracea*) were investigated (Figure 4). The *rps19* gene was identified in the IRb/LSC binding region in each of the six genomes compared, indicating that part of the gene was repeated at the IRA and LSC binding site. Comparisons showed that for this gene, *Saperavi* (38 bp) and *Spinacia* (144 bp) have the shortest and the longest, respectively. On the border between IRb and SSC, *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* were similar, and the *ycf1* fragment was 1112, 1111, and 1109 bp on the IRb border, respectively. In *Solanum*,

AA	Codon	Number	RSCU	AA	Codon	Number	RSCU
Phe	**TTT**	2170	1.19	Ser	**TCT**	1210	1.42
	TTC	1487	0.81		**TCC**	1020	1.20
Leu	**TTA**	1115	1.24		**TCA**	931	1.10
	TTG	1143	1.27		TCG	618	0.73
	CTT	1102	1.22	Pro	**CCT**	638	1.05
	CTC	716	0.80		CCC	600	0.99
	CTA	797	0.89		**CCA**	772	1.27
	CTG	527	0.59		CCG	420	0.69
Ile	**ATT**	1845	1.22	Thr	**ACT**	721	1.21
	ATC	1120	0.74		**ACC**	614	1.03
	ATA	1574	1.04		**ACA**	682	1.14
Met	ATG	965	1		**ACG**	372	1.62
Val	**GTT**	798	1.31	Ala	**GCT**	459	1.25
	GTC	443	0.73		GCC	347	0.95
	GTA	747	1.23		**GCA**	441	1.20
	GTG	446	0.73		GCG	217	0.59
Tyr	**TAT**	1583	1.36	Cys	**TGT**	701	1.19
	TAC	749	0.64		TGC	480	0.81
Ter	**TAA**	1180	1.21	Ter	TGA	945	0.97
	TAG	790	0.81	Trp	TGG	696	1
His	**CAT**	1051	1.42	Arg	CGT	384	0.70
	CAC	433	0.58		CGC	255	0.46
Gln	**CAA**	1124	1.41		**CGA**	574	1.04
	CAG	471	0.59		CGG	401	0.73
Asn	**AAT**	1892	1.42	Ser	AGT	796	0.94
	AAC	772	0.58		AGC	525	0.62
Lys	**AAA**	2076	1.34	Arg	**AGA**	1085	1.97
	AAG	1028	0.66		**AGG**	600	1.09
Asp	**GAT**	1183	1.44	Gly	**GGT**	604	1.04
	GAC	456	0.56		GGC	337	0.58
Glu	**GAA**	1413	1.39		**GGA**	842	1.45
	GAG	624	0.69		GGG	533	0.92

Note: Codons shown in bold represents have RSCU value >1

Table 4: Codon analysis of *Vitis vinifera* chloroplast genes that code for proteins.

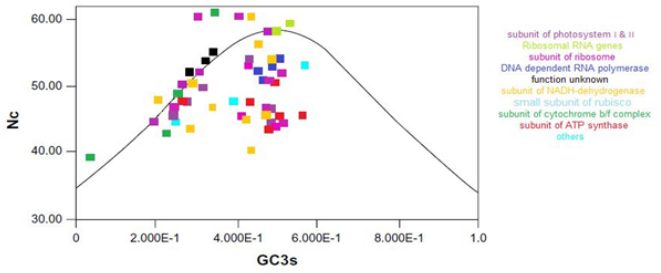

Figure 2: Effective number of codons (NC) used in each gene plotted against GC content at synonymous variable third positions of codons (GC3). The continuous curve plots the relationship between NC and GC3s in the absence of selection.

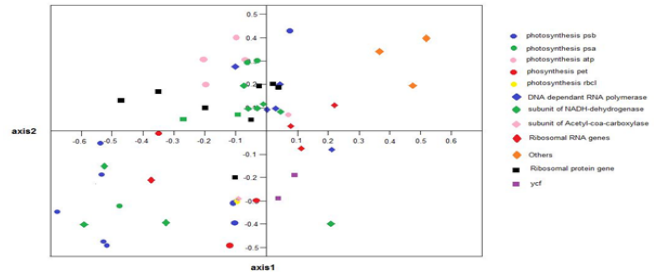

Figure 3: Correspondence analysis of the relative synonymous codon usage in 54 genes from chloroplast genome of *Vitis vinifera*.

codon bias are found in the rightmost of the axis. Figure 3 shows the plots of ENC values towards GC3 values, given that none of these points have been located on the curve, consequently codon selection of them are not limited to mutation bias of GC3. On the other hand, the codon usage of the points below the curve is dependent from compositional constraint.

Chloroplast SSRs

Chloroplast SSRs are useful in analysis of genetic diversity, because of their greater efficiency as opposed to genomic SSRs. In addition and due to the greater level of conservation, the information of the other species can be used to design specific primers for a species with unknown sequence data.

Intron

In all three intended genomes of this study, 18 genes were found containing one or two introns that is the same as in *Panax schinseng* [1]. The number and location of the intron in chloroplast seems to be conserved. The comparison of introns (Table 5) between *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* shows that 18 genes have one or two introns in the chloroplast genome, 6 of which are tRNA coding genes and the rest are protein-coding genes. The longest intron in all three genomes of this study is located in *trnK-UUU* with 2508 bp, which is the only intron in them with another gene, *matK* inside. The smallest intron with 540 bp is placed in *rps12-3 end*, which is situated in IR region. *Ycf3* and *clpP*, which are located in LSC are divided by two introns. Intron lengths for *rps16*, *atpF*, *ycf3-1*, *rps12-3 end*, *petB*, *petD*, *rpl16*, *ndhB*, *trnI-GAU* and *trnA-UGC* are conserved, while the others have small variations. Mentioned introns in Table 5 have a high identity, especially *clpP1* and *ycf3-2*, which has 100% sequence identity among the studied genomes.

Gene loss in chloroplast genome

-	Intron	*Vitis vinifera*	*Vitis vinifera* subsp. *Caucasica* Cultivar: *Saperavi*	*Vitis vinifera* subsp. *caucasica* Cultivar: *Meskhuri mtsvane*	Sequence Identity (%)
1	trnK-UUU	2508	2501	2501	98.57
2	rps16	900	900	900	99.38
3	trnG-GCC	710	682	681	97.86
4	atpF	747	747	747	99.50
5	rpoC1	763	759	763	99.61
6	ycf3-1	743	743	743	99.74
	ycf3-2	728	727	727	100
7	trnL-UAA	516	518	518	99.49
8	trnV-UAC	573	554	554	99.58
9	rps12-3end	540	540	540	99.88
10	clpP1	631	633	632	100
	clpP2	810	810	811	99.79
11	petB	688	688	688	99.95
12	petD	734	734	734	99.63
13	rpl16	1064	1064	1064	99.49
14	rpl2	664	664	664	99.78
15	ndhB	679	679	679	99.94
16	trnI-GAU	944	944	944	99.89
17	trnA-UGC	803	803	803	99.70
18	ndhA	1130	1129	1129	99.95

Table 5: The comparison of introns among *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane*.

The chloroplast genome is a protected genome, but research has shown that a number of genes have been removed or transmitted during the course of evolution. For example *ycf15* is also present in all three genomes of this study. *InfA*, the most mobile gene between chloroplast and nuclear genome, that codes for a translation initial factor 1 is detected in studied genomes of this study. However, some others had the *infA* as a pseudo gene [24], while in some others *infA* identified as an intact gene [22]. The results showed that *trnP-GGG* was eliminated in all three genomes of this study. According to studies, also this gene is absent in angiosperms, but it has been identified in *Cryptomeria japonica* [2]. Therefore, it seems that this gene may have been eliminated before the divergence of angiosperms [14]. *rpl22*, which encodes for large subunit of ribosomal protein 22, was identified in all three chloroplast genomes of this study. In addition, this gene has been deleted in three *legumes*, *Glycine*, *Lotus* and *Medicago* [3].

Vastness of IR

The IR range between species is usually different, and the IR

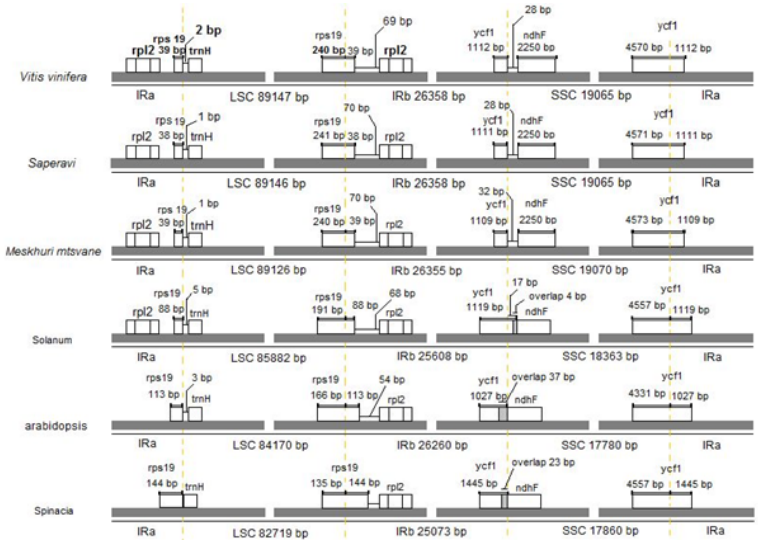

Figure 4: Comparison among LSC, IR and SSC border regions of three common reference species with studied genomes. (Fourth, fifth and sixth rows are derived from [14]).

GC content of *Vitis vinifera* chloroplast genome											
Coding Region						Non-Coding Region					
	Protein	tRNA		Total	IGS	Intron	Total	Complete Genome	LSC	SSC	IR
Length	80374	2794	9036	91872	50057	19604	69661	160928	89147	19065	26358
Proportion	49.94	1.73	5.61	57.08	31.1	12.18	43.28	100	55.39	11.84	16.37
T%	31.6	22.92	22.23	30.46	33.85	31.65	33.37	31.74	33.06	34.13	28.6
A%	30.53	24.05	22.35	29.5	33.93	30.72	33.04	30.91	31.77	34.35	28.43
C%	19.58	26.96	27.88	20.54	16.24	19.69	17.11	19.12	18.24	16.65	22.27
G%	18.52	26.33	27.82	19.61	16.28	18.24	16.63	18.37	17.22	15.01	20.79
A+T%	62.17	46.9	44.53	59.97	67.79	62.27	66.32	62.63	64.73	68.32	57.04
C+G%	37.95	53.21	55.64	40.13	32.44	37.8	33.79	37.45	35.46	31.71	43.09

Table 6: GC content of *Vitis vinifera*.

Arabidopsis and *Spinacia*, the discovery of the *ycf1* and *ndhf* genes at the IRb / SSC binding site was identified, with the largest overlapping region in *Arabidopsis* with 37 bp and the shortest in *Solanum* (17 bp). The *ycf1* is located between SSC and IRb regions, and it was duplicated in IRb at the border of IRb and SSC. In *Spinacia*, the *ycf1* has the longest repeat (1445 bp). While at *Meskhuri mtsvane*, the shortest duplication was detected (1109 bp).

GC content

The GC content of the total chloroplast genome of *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane* chloroplasts is similar and equals 37.4%, the same number for other plants such as *Solanum lycopersicum* (37.86%), *Nicotiana tabacum* (37.85%), *Atropa belladona* (37.56%) and Glycine max (34%) have also been reported [14]. The coding and non-coding regions in each of the three genomes have a low GC content, which is reported as 40.13% and 33.79% respectively. The differences in GC content for four regions of all genomes of this study were analyzed and the *Vitis vinifera's* is presented in Table 6, and the results showed that IR region was the richest in all three genomes of this study. It seems that the ribosomal genes (*rrna 4.5, rrna 5, rrna 16, rrna 23*) and the coding regions responsible for the rich content of GC in IR [11,20,23]. GC content in the IR region in *Vitis vinifera* was 43.09%, while GC content in SSC and LSC was 31.71% and 35.46%, respectively. The distribution of GC content in each region similar with other species [11]. Studies on *Alsophila spinulosa* in 2009 have showed that GC content in the chloroplast genomes are not the same between genes in different functional groups, rRAN (55.18%) > tRNA (54.55%) > photosynthetic (43.85%) > genetic system (40.80%) > NADH (39.54%) [8]. Similar results were reported for *Vitis vinifera*, *Saperavi* and *Meskhuri mtsvane*. In the coding region of *Vitis vinifera*, rRNA genes have the highest GC content (55.64%) and the protein coding genes have the lowest (37.95%). In non-coding region, GC content of IGS and introns is 32.44% and 37.80%, respectively. GC content is a significant property of a genome, which is correlated to the gene expression regulation, number of microRNA binding sites, gene distribution and recombination rate and physical location of functional elements.

References

1. Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11: 247-261.

2. Hirao T, Watanabe A, Kurita M, Kondo T, Takata K (2008) Complete nucleotide sequence of the Cryptomeria japonica D. chloroplast genome and comparative chloroplast genomics: Diversified genomic structure of coniferous species. BMC Plant Biol 23: 8-70.

3. Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, et al. (2005) Complete chloroplast genome sequence of Gycine max and comparative analyses with

other legume genomes. Plant Mol Biol 59: 309-322.

4. Robert K, Jansen CK, Seung-Bum L (2006) Phylogenetic analyses of Vitis (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol. 32: 1471-2148.
5. Leebens-Mack J, Raubeson LA, Cui L, Kuehl J, Fourcade MH, et al. (2005) Identifying the basal angiosperms in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol 22: 1948-1963.
6. Soltis DE, Soltis PS (2004) Amborella not a "basal angiosperm"? Not so fast. Amer J Bot 91: 997-1001.
7. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, et al. (2004) Genome-scale data, angiosperm relationships, and 'ending incongruence': A cautionary tale in phylogenetics. Trends Plant Sci 9: 477-483.
8. Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of Acorus calamus chloroplast genome and its phylogenetic implications. Mol Biol Evol 22: 1813-1822.
9. Lockhart PJ, Penny D (2005) The place of Amborella within the radiation of angiosperms. Trends Plant Sci 10: 201-203.
10. Ané C, Burleigh JG, McMahon MM, Sanderson MJ (2005) Covarion structure in plastid genome evolution: a new statistical test. Mol Biol Evol 22: 914-924.
11. Kim YK, Park CW, Kim KJ (2009) Complete chloroplast DNA sequence from a Korean endemic genus, megaleranthis saniculifolia, and its evolutionary implications. Mol Cells 27: 365-381.
12. Daniell H, Kumar S, Duformantel N (2005) Breakthrough in chloroplast genetic engineering of agronomical important crops. Trends Biotechnol 23: 238-245.
13. Ruf S, Karcher D, Bock R (2007) Determining the transgene containment level provided by chloroplast transformation. Proc Natl Acad Sci USA. 104: 6998-7002.
14. Talat F, Wang K (2015) Comparative bioinformatics analysis of the chloroplast genome of a wild diploid Gossypium and two cultivated allotetraploid species. Iran J Biotech 13: 47-56.
15. Wyman SK, Boore JL, Jansen RK (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252-3255.
16. Lohse M, Drechsel O, Bock R (2007) Organellar Genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet 52: 267-274.
17. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29: 4633-4642.
18. Green-acre MJ (1984) Theory and applications of correspondence analysis. Academic Press.
19. Maria DE (2001) Synonymous codon usage in bacteria. Curr Issu Mol Biol 3: 91-97.
20. Gao L, Yi X, Yang YX, Su YJ, Wang T (2009) Complete chloroplast genome sequence of a tree fern Alsophila spinulosa: insights into evolutionary changes in fern chloroplast genomes. BMC Evol Biol 9: 130.
21. Wu FH, Kan DP, Lee SB, Daniell H, Lee YW (2009) Complete nucleotide sequence of Dendrocalamus latiflorus and Bambusa oldhamii chloroplast genomes. Tree Physiol 29: 847-856.
22. Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H (2007) The complete nucleotide sequence of the coffee (Coffea arabica L.) chloroplast genome: Organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. Plant Biotechnol J 5: 339-353.
23. Daniell H, Wurdack KG, Kanagaraj A, Lee SB, Saski C (2008) The complete nucleotide sequence of the cassava (Manihot esculenta) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. Theor Appl Genet 116: 723-737.
24. Steane DA (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, Eucalyptus globulus (Myrtaceae). DNA Res 12: 215-220.
25. Wilson CA (2004) Phylogeny of Iris based on chloroplast matK gene and trnK intron sequence data. Mol Phylogenet Evol 33: 402-412.
26. Millen RS, Olmstead RG, Adams KL (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13: 645-658.
27. Ohsako T, Ohnishi O (2001) Nucleotide sequence variation of the chloroplast trnK/matK region in two wild Fagopyrum (Polygonaceae) species, F.leptopodum and F.statice, Genes Genet Syst 76: 39-46.
28. Wilson CA (2004) Phylogeny of Iris based on chloroplast matK gene and trnK intron sequence data. Mol Phylogenet Evol 33: 402-412.
29. Yang M, Zhang X, Liu G (2010) The complete chloroplast genome sequence of date palm (Phoenix dactylifera L.) PLoS One 5: 143-152.