THE COMPLETE CHLOROPLAST DNA SEQUENCE OF ELEVEN GRAPE CULTIVARS. SIMULTANEOUS RESEQUENCING METHODOLOGY

Vazha TABIDZE[1,2], Grigol BARAMIDZE[1,2], Ia PIPIA[1,2], Mari GOGNIASHVILI[1,2], Levan UJMAJURIDZE[2], Tengiz BERIDZE[1,2*], Alvaro G. HERNANDEZ[3] and Barbara SCHAAL[4]

1 : Free University of Tbilisi, D. Agmashenebeli Al.13th km, 0159, Tbilisi, Georgia
2 : Institute of Molecular Genetics, Agricultural University of Georgia, D. Agmashenebeli Al.13th km, 0159, Tbilisi, Georgia
3 : Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, IL 61801, USA
4 : Washington University in St Louis, MO 63130-4899, USA

Abstract

Aims : The chloroplast DNA sequence of eight Georgian grape cultivars (Rkatsiteli, Saperavi, Meskhuri Mtsvane, Chkhaveri, Aladasturi, Krakhuna, Tsitska, Tsolikouri) and three French cultivars (Chardonnay, Gouais Blanc, Chasselas), belonging to four different haplogroups (AAA, ATT, ATA, GTA), was determined by Illumina resequencing of genomic DNA. The chloroplast DNA sequence of the Maxxa cultivar was used as reference.

Methods and results : The comparison of sequenced chloroplast DNA gave 100 % identity to Chardonnay and Gouais Blanc, differing from Meskhuri Mtsvane by two insertions/deletions (indels) (all ATA haplogroup). The difference between Chasselas and Saperavi was a single insertion (both ATT haplogroup), while Maxxa, Chkhaveri, Aladasturi, Krakhuna, Tsitska and Tsolikouri were all identical (all members of the GTA haplogroup). Forty-seven identical single nucleotide polymorphisms (SNPs) were detected in the AAA, ATA and ATT haplogroups in comparison to the reference DNA. Additionally, 18 SNPs were detected for the ATT haplogroup, 4 for AAA, 6 for ATA and 11 for both AAA and ATA. The phylogenetic results show that the ATT, AAA and ATA haplogroups are more closely related to each other than to the GTA haplogroup.

Conclusion : In the sequencing data of grape genomic DNA at the coverage (read depth) of chromosomal DNA 30-40, the coverage of chloroplast DNA reaches several thousand reads per bp due to the high number of chloroplast DNA copies in genomic DNA, much higher than necessary for resequencing. Based on these data, a new methodology of simultaneous resequencing of large number of chloroplast DNA was developed without preliminary chloroplast isolation or chloroplast enrichment.

Significance and impact of the study : This method has great potential for expanding both phylogenetic and population genetic information on the evolution of domesticated crops.

Key words : grape DNA, Illumina, resequencing, indels, SNP

Résumé

Objectifs : La séquence d'ADN chloroplastique de huit cultivars Géorgiens (Rkatsiteli, Saperavi, Meskhuri Mtsvane, Chkhaveri, Aladasturi, Krakhuna, Tsitska, Tsolikouri) et trois cultivars Français (Chardonnay, Gouais Blanc, Chasselas), appartenant à quatre haplogroupes différents (AAA, ATT, ATA, GTA), a été obtenue par reséquençage de l'ADN génomique selon la technologie Illumina. La séquence de l'ADN chloroplastique du cultivar Maxxa a été utilisée comme référence.

Méthodes et résultats : La comparaison des ADN chloroplastiques séquencés a révélé une identité de 100 % entre Chardonnay et Gouais Blanc, différente de celle de Meskhuri Mtsvane de deux insertions/délétions (indels) (appartenant tous à l'haplogroupe ATA). Chasselas et Saperavi (de l'haplogroupe ATT) sont différenciés par une seule insertion, tandis que Maxxa, Chkhaveri, Aladasturi, Krakhuna, Tsitska et Tsolikouri (tous membres de l'haplogroupe GTA) sont tous identiques. Quarante-sept polymorphismes mononucléotidiques (SNPs) identiques ont été détectés chez les haplogroupes AAA, ATA et ATT en comparaison de l'ADN de référence. En outre, 18 SNPs ont été identifiés pour l'haplogroupe ATT, 4 pour AAA, 6 pour ATA, et 11 communs pour AAA et ATA. Les analyses phylogénétiques montrent que les haplogroupes ATT, AAA et ATA sont plus étroitement liés les uns aux autres qu'avec l'haplogroupe GTA.

Conclusion : Le séquençage de l'ADN génomique de raisin à partir de fragments d'ADN chromosomiques de 30- à 40 de long permet un recouvrement de l'ADN chloroplastique de plusieurs milliers de lectures par bp en raison du nombre de copies élevé d'ADN chloroplastique, nettement supérieur à ce qui est nécessaire pour le reséquençage. Une nouvelle méthodologie de séquençage d'ADN chloroplastique a été élaborée sans isolement ou enrichissement de chloroplastes.

Signification et impact de l'étude : Cette méthode a un fort potentiel pour accroître nos connaissances en génétique et en phylogénie sur l'évolution des cultures domestiquées.

Mots clés : ADN de raisin, Illumina, reséquençage, indels, SNP

manuscript received 16th January 2014 - revised manuscript received 21st April 2014

*Corresponding author : t.beridze@freeuni.edu.ge

©Vigne et Vin Publications Internationales (Bordeaux, France)

INTRODUCTION

Georgia is home to over 500 grape cultivars (Ketskhoveli *et al.,* 1960). The greater Caucasus region in which Georgia lies is widely believed to be the area where grape domestication began, and the study of genetic diversity in this region is viewed as a key to understanding grape domestication in general (Negrul 1946).

The plastid genome is an effective tool for interspecific phylogenetic and intraspecific phylogeographic studies of angiosperms (Aoki *et al.,* 2006; Gutiérrez-Rodríguez *et al.,* 2011). Few papers are devoted to grape chloroplast DNA. Arroyo-Garcia *et al.* (2006) analyzed chloroplast DNA variation at nine polymorphic microsatellite loci of 1201 *V. vinifera* genotypes belonging to both *sativa* and *sylvestris* subspecies. Genotypic analyses for these microsatellite loci identified eight different chlorotypes (A to H) (Arroyo-Garcia *et al.,* 2006). Among them, only four (A, B, C and D) had global frequencies greater than 5 %. The intermediate relationship of chlorotype B to all other chlorotypes suggests that it could be an ancestral *V. vinifera* chlorotype.

Plastid DNA of Caucasian (Georgian) grape varieties was studied by sequencing of some noncoding regions of grape chloroplast DNA (Beridze *et al.,* 2011). During the investigation of 113 samples of a world-wide set of grape cultivars including Georgian cultivars, four plastid DNA haplotypes were evident and were designated by their character-states at each of the three polymorphic positions (Beridze *et al.,* 2011) (Table 1).

The AAA plastid haplotype was found only in the cultivars from Georgia. More specifically, twenty-three (57.5 %) of the 40 included Georgian cultivars exhibited this haplotype, of which the “Rkatsiteli” cultivars originating from Eastern Georgia prevailed. This contrasts with the nine cultivars (22.5 %) of the “Chkhaveri” group (GTA), which are mostly cultivated in Western Georgia near the Black Sea coast. Six other Georgian cultivars exhibited the “Saperavi” (ATT) haplotype. Among these was the well-known Saperavi cultivar, which is now mainly distributed in Eastern Georgia but is believed to have originated in south-west Caucasus. Only two Georgian cultivars exhibited the “Meskhuri” group haplotype (ATA), as this group comprises mainly West European cultivars (Beridze *et al.,* 2011). Both the B and C chlorotypes of Arroyo-Garcia *et al.* (Arroyo-Garcia *et al.,* 2006) are combined in the haplotype ATA according to these data (Beridze *et al.,* 2011).

The genomic DNA of Georgian grape cultivars was also studied by nuclear microsatellite analysis based on 20 nuclear microsatellites and no close relationship between Georgian and Western European cultivars was found (Imazio *et al.,* 2013).

In the present investigation, the whole chloroplast DNA of four Georgian grape cultivars, one from each haplotype group (as defined by (Beridze *et al.,* 2011)), was sequenced. These cultivars were: Rkatsiteli (haplotype AAA), Saperavi (haplotype ATT), Meskhuri Mtsvane (haplotype ATA) and Chkhaveri (haplotype GTA).

In the sequencing data of the four Georgian grape cultivars (coverage of chromosomal DNA 30-40), the coverage of chloroplast DNA reaches several thousand because of the high number of chloroplast DNA copies in genomic DNA, much higher than necessary for resequencing. The idea of the new methodology of simultaneous sequencing of large number of chloroplast DNA was developed by mixing and sequencing the genomic DNA from many cultivars in one Illumina lane upon individually barcoding each library without preliminary chloroplast isolation or chloroplast enrichment. The chloroplast DNA of ten different grape cultivars and species was sequenced simultaneously in one Illumina run. As a result, the number of sequenced chloroplast DNA of grape cultivars increased to eleven: to the four Georgian grape cultivars mentioned above, seven others, three French - Chardonnay, Gouais Blanc (both ATA haplogroup) and Chasselas (ATT haplogroup) - and

Table 1. Haplogroup definition for investigated cultivars.

Haplogroup	Nucleotide position			Investigated cultivars
	205	86715	86721	
AAA (1 haplotype)	A	A	A	Rkatsiteli
ATA (2 haplotypes)	A	T	A	Chardonnay, Gouais blanc, Meskhuri Mtsvane
ATT (2 haplotypes)	A	T	T	Chasselas, Saperavi
GTA (1 haplotype)	G	T	A	Pinot noir, Maxxa, Chkhaveri, Aladasturi, Krakhuna, Tsitska, Tsolikouri

©Vigne et Vin Publications Internationales (Bordeaux, France)

four Georgian cultivars - Aladasturi, Krakhuna, Tsitska and Tsolikouri (all GTA haplogroup) - were added.

MATERIALS AND METHODS

1. Plant material and DNA isolation

The grape cuttings of four grape cultivars (Rkatsiteli, Saperavi, Meskhuri Mtsvane and Chkhaveri) were collected at the Saguramo National Centre for Grapevine and Fruit Tree Planting Material Propagation (Mtskheta, Georgia) and grown in water at room temperature. Total genomic DNA was extracted from young grape leaves. The leaves were ground in liquid nitrogen and DNA was isolated using the CTAB-based method (Lodhi *et al.,* 1994). All four DNA preparations were preliminarily checked by sequencing of three noncoding plastid DNA regions (the *trnH-psbA* intergenic spacer, the *accD-psa*I intergenic spacer and the *rpl*16 intron).

The DNA of other grape cultivars (Muscat Blanc à Petits Grains, Chardonnay, Cabernet Sauvignon, Cabernet Franc, Veltliner Rot, Mourvedre, Alvarelhao, Traminer Rot, Gouais Blanc, Chasselas, Aladasturi, Krakhuna, Tsitska, Tsolikouri) was also analyzed. The origin of these cultivars (total of 59 cuttings received from the Institut National de la Recherche Agronomique (INRA), Montpellier, France) and DNA isolation procedure have been described in our previous publication (Beridze *et al.,* 2011).

2. PCR conditions and Sanger sequencing

The PCR conditions included initial denaturing at 94 °C for 1 min, 30 cycles of 94 °C denaturing (1 min), 55 °C annealing (2 min) and 72 °C extension (2 min), followed by a final extension step at 72 °C (5 min). Sigma-Aldrich chemicals were used for the PCR reactions. PCR products were purified with GenElute PCR Clean-Up Kit (Sigma-Aldrich), dye-labeled using a Big Dye Terminator Kit (Applied Biosystems) and analyzed on either an Applied Biosystems 3100 or 3700 genetic analyzer (Laboratory Services Division of the University of Guelph, ON, Canada). Sequences were manually aligned in Se-Al (Rambaut 2002) and haplotype networks were generated using TCS1.18 (Clement *et al.,* 2000).

3. Construction of shotgun genomic DNA libraries

Shotgun genomic DNA libraries were constructed using the TruSeq DNA Sample prep kit (Illumina, San Diego, CA). Briefly, 1 µg of genomic DNA was sonicated on a Bioruptor (Diagenode, NJ) for 20 cycles of 30 secondes ON and 90 sec OFF. After sonication, DNA was blunt-ended, 3'-end A-tailed and ligated to indexed adaptors. The adaptor-ligated DNA was size selected with AMPure-beads using the gel-free protocol described in the TruSeq DNA Sample Prep manual. Size-selected DNA was amplified by PCR to selectively enrich for those fragments that have adapters on both ends. Amplification was carried out for 6 cycles with the Kapa HiFi polymerase (Kapa Biosystems, Woburn, MA) to reduce the likeliness of multiple identical reads due to preferential amplification. The final libraries were quantitated by quantitative qPCR on an ABI 7900 (Life Technologies, Grand Island, NY). Final amplified libraries were also run on Agilent bioanalyzer DNA 7500 LabChips (Agilent, Santa Clara, CA) to determine the average fragment size and to confirm the presence of DNA of the expected size range.

4. Sequencing on an Illumina HiSeq 2000

The libraries were pooled in equimolar concentration and loaded onto one lane of an 8-lane flow cell for cluster formation and sequenced on an Illumina HiSeq 2000. The libraries were sequenced from both ends of the molecules to a total read length of 100 nt from each end. The raw bcl files were converted into demultiplexed compressed fastq files using Casava 1.8.2 (Illumina).

Table 2. Regions sequenced by capillary electrophoresis

Primers	Forward	Reverse
Primer1: 54 bp deletion (position 30,133 – 30,186) (haplotypes - AAA, ATA, ATT)	CTACTGGCCCGGTCTTGAAT	TGGAATCGATGGTGCAGAGT
Primer 2: 54 bp deletion (position 30,133 – 30,186) (haplotypes - AAA, ATA, ATT)	GGGTGTCGCCTGATCAACAA	TTAAAGCAGCCCAAGCGAGA
Primer 3: 33 bp duplication (position 6,658 - 6,691) (haplotypes AAA, ATA, ATT)	TCAAGTCGCACGTTGCTTTC	GATGTAATGATGTCGAAGTAGTCAA
Primer 4: 18 bp insertion (position 10,001 - 10,019) (haplotype ATT)	TTATTCCCACGGCCCGGATA	TTGTGCAAGAATCCATAGTTCCC
Primer 5: 18 bp duplication (position 19,527-19,544) (haplotype AAA)	GGCTGTTCCTAAAGGACCCAA	TCGAAAAGACCCATGCTTCC

©Vigne et Vin Publications Internationales (Bordeaux, France)

The sequencing of Saperavi, Meskhuri Mtsvane and Chkhaveri DNA was carried out at the facilities of Roy J. Carver Biotechnology Center, University of Illinois in Urbana-Champaign (USA); Rkatsiteli DNA was sequenced at the Laboratory Services Division of Guelph University (Canada). For the assembly of Georgian grape cultivar genome, the Maxxa plastid DNA sequence was used (http://www.ncbi.nlm.nih.gov/nuccore/91983971) (Jansen *et al.* 2006). Some disputed regions of chloroplast DNAs were additionally sequenced by capillary sequencing (Table 2).

5. Computational analysis

Raw FASTAQ data files were used to align paired-end reads to the Maxxa chloroplast reference genome using Bowtie 2 (version 2.1.0) (Langmead *et al.,* 2012) using default alignment criteria, in particular:

Reads with more than 15 ambiguous characters (Ns and .s) were filtered out.

Reads with gaps within 4 positions of read ends were disallowed.

Traditionally, variant calling programs are geared towards nuclear DNA of diploid organisms and relatively low read coverage. In case of chloroplasts, however, there is a multitude of DNA molecules per chloroplast as well as a multitude of chloroplasts per cell. This results, on one hand, in a very high coverage of chloroplast DNA (taken as one) during HiSeq sequencing and, on the other hand, in the possibility of heteroplasmy (presence of multiple chloroplast genomes within a cell). Because of this, we used a somewhat different approach in identifying single nucleotide polymorphisms (SNPs) and insertions/deletions (indels).

All of aligned reads allowed by Bowtie 2 were considered and weighted equally. This information was extracted into mpileup files using SAMtools (Li *et al.,* 2009a). The very high coverage (ranging from 8,000 to 20,000 reads per bp) gave us the flexibility of using even lower quality reads, relying on the law of large numbers to make variant calls for high-level mutations while also detecting potential lower-level mutations. To verify this approach computationally, we also applied a more traditional computational pipeline to variant calling (filtering on read qualities and applying various variant calling programs with consequent realignments) to obtain near-identical results. In addition, looking at total read counts allowed us to easily identify a large 54-bp deletion in three out of four cultivars (see Figure 1). The filtered chloroplast reads were assembled in parallel by SOAPdenovo computer program and coinciding results were received (Li *et al.,* 2009b).

The chloroplast DNA sequences of the Georgian grape cultivars have been deposited in the DNA Data Bank of Japan. Accession numbers: AB856289 Rkatsiteli; AB856290 Saperavi; and AB856291 Meskhuri Mtsvane.

RESULTS

We identified 86 SNPs in three Georgian grape cultivars in comparison to the reference Maxxa chloroplast DNA (Table 3). Forty-seven identical SNPs were characteristic for all three cultivars (Rkatsiteli + Saperavi + Meskhuri Mtsvane), the additional SNPs were: 18 for Saperavi, 4 for Rkatsiteli, 6 for Meskhuri

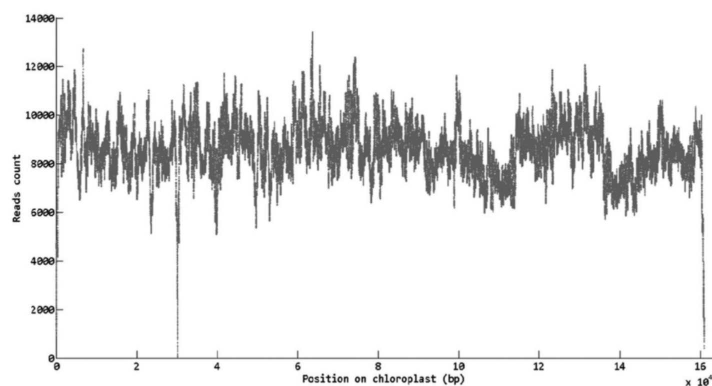

Figure 1. Read counts (depth coverage) for the alignment of Meskhuri Mtsvane cultivar chloroplast DNA. The reads of Meskhuri Mtsvane were aligned against the reference genome of Maxxa chloroplast. The graph shows the depth coverage for each position of the reference sequences after initial alignment performed by Bowtie 2. The dip at around 30,000 bp corresponds to a 54-bp deletion in Meskhuri Mtsvane chloroplast genome.

©Vigne et Vin Publications Internationales (Bordeaux, France)

Mtsvane and 11 for both Rkatsiteli and Meskhuri Mtsvane. The number of noncoding substitutions was 67 and coding substitutions 19. In 8 cases, nonsynonymous substitutions were observed, which altered the amino acid sequence (Table 3). In 13 cases, synonymous substitutions were detected. In gene ycf1, 5 SNPs (3 nonsynonymous and 2 synonymous) were observed.

Short indels were also detected : in Saperavi - 11 insertions and 15 deletions ; in Rkatsiteli - 15 insertions and 9 deletions ; and in Meskhuri Mtsvane - 13 insertions and 9 deletions. In Chkhaveri, no insertions or deletions were detected in comparison to reference DNA. As a result, the length of sequenced chloroplast DNA varied, with Rkatsiteli at 160,927, Saperavi at 160,928 and Meskhuri Mtsvane at 160,906 bp in length.

In the region surrounding position 30,100 - 30,200 bp, a drastic fall of read numbers was observed in all cultivars except Chkhaveri (Figure 1). This region was subsequently sequenced by capillary electrophoresis. A 54-bp deletion at position 30,133 - 30,186 (trnC-GCA – petN intergenic spacer) was detected in Saperavi, Rkatsiteli and Meskhuri Mtsvane, but not in Chkhaveri. The existence of this deletion was checked in other grape cultivars including West Europeans. It was established that this 54-bp deletion is typical for all checked AAA, ATA and ATT haplogroups of the world-wide set of grape cultivars, but not for GTA haplogroup (Figure 2).

Additionally, three other long indels were detected by capillary sequencing in chloroplast DNA (Table 4). Two were in intergenic spacers (Rps16 – trnQ-UUG and trnS-GCU – trnG-GCC) ; the other was an 18-bp duplication in the rpoC2 gene of Rkatsiteli (AAA haplotype ; position 19,527 - 19,544). As a result, a 6-aa-long peptide duplication occurred (TLLNRN ; position 934 - 939 in the rpoC2 protein).

To illustrate the evolutionary relationship between the studied cultivars, a phylogenetic tree based on the multiple alignments was constructed using NCBI BLAST's TreeView Display option that constructs a phylogenetic distance tree based on fast minimum evolution approach (Desper *et al.,* 2002). Figure 3 represents the resulting phylogenetic tree.

Chloroplast isolation or chloroplast enrichment from leaves of grape cultivars or *Vitis* species is a great problem due to the large amount of phenolic compounds present in leaves, which negatively influence DNA isolation. Therefore, a method for determining chloroplast DNA sequence directly from genomic DNA was developed.

In sequencing data, the number of reads for chloroplast DNA can, in some cases, reach 20,000 per base due to a high number of chloroplast DNA copies in genomic DNA, whereas for resequencing of DNA, 40-50 reads are sufficient. On the basis of these facts, the new methodology of chloroplast DNA resequencing was developed. The idea was mixing the genomic DNA from many cultivars and sequencing them in one Illumina lane, barcoding each library so that reads coming from each cultivar can be identified. The number of chloroplast DNA reads is quite high, sufficient to compute chloroplast DNA sequence, but reads from chromosomal are low and are ignored.

During manuscript preparation, other papers appeared where next-generation sequencing techniques also have allowed for simultaneous genomic studies of mitochondrial and chloroplast DNA (Hancock-Hanser *et al.,* 2013 ; Middleton *et al.,* 2014). Here, ten chloroplast DNAs of different grape cultivars

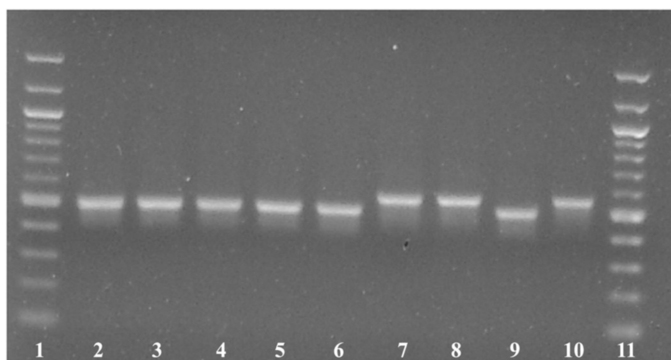

Figure 2. 1% agarose gel electrophoresis of PCR-amplified grape chloroplast DNA fragment 29,945 - 30,506. Lanes: 2. Muscat Blanc à Petits Grains (ATT); 3. Chardonnay (ATA); 4. Cabernet Sauvignon (ATT); 5. Cabernet Franc (ATT); 6. Veltliner Rot (ATT); 7. Mourvedre (GTA); 8. Alvarelhao (GTA); 9. Traminer Rot (ATT); and 10. Chkhaveri (GTA). Lanes 1 and 11, 100-bp DNA marker.

©Vigne et Vin Publications Internationales (Bordeaux, France)

Table 3. The intravarietal SNPs and amino acid substitutions in Georgian grape chloroplast DNA.

Nucleotide Position	Locus	Saperavi	Rkatsiteli	Meskhuri Mtsvane	Amino Acid Substitutions
205	Intergenic trnH-psbA	G-A	G-A	G-A	
1552	Intergenic psbA - trnK-UUU		G-A		
4527	Intergenic trnK-UUU - rps16			T-G	
4547	Intergenic trnK-UUU - rps16	A-T			
5591	Intron rps16	A-C			
5978	Intron rps16		T-G	T-G	
8114	Intergenic rps16 - trnQ-UUG	A-G	A-G	A-G	
8175	Intergenic rps16 - trnQ-UUG	T-C	T-C	T-C	
9996	Intergenic trnS-GCU - trnG-GCC	A-G	A-G	A-G	
10266	Intergenic trnS-GCU - trnG-GCC	A-T			
11120	Intron trnG-GCC		A-G	A-G	
11121	Intron trnG-GCC	A-T			
11625	Intergenic trnR-UCU - atpA	T-A	T-A	T-A	
14416	Intron gene atpF	A-G	A-G	A-G	
14794	Intergenic atpF - atpH	T-A	T-A	T-A	
20840	Gene rpoC2	C-T	C-T	C-T	S - N
21046	Gene rpoC2	G-T	G-T	G-T	F - L
22321	Gene rpoC2	A-G	A-G	A-G	syn
24437	Intron rpoC1	C-A	C-A	C-A	
28707	Intergenic rpoB - trnC-GCA			T-G	
29232	Intergenic rpoB - trnC-GCA		T-C		
29507	Intergenic rpoB - trnC-GCA	C-T	C-T	C-T	
29571	Intergenic rpoB - trnC-GCA	G-T	G-T	G-T	
30783	Intergenic petN - psbM	T-C	T-C	T-C	
32670	Intergenic psbM - trnD-GUC	C-G	C-G	C-G	
34037	Intergenic trnE-UUC - trnT-GGU	C-A	C-A	C-A	
36031	Intergenic trnT-GGU - gene psbD	C-A			
36397	Gene psbD	A-C	A-C	A-C	syn
39357	Gene psbZ	T-A	T-A	T-A	syn
39584	Intergenic psbZ - trnG-GCC	T-G	T-G	T-G	
39585	Intergenic psbZ - trnG-GCC	C-G	C-G	C-G	
39586	Intergenic psbZ - trnG-GCC	C-A	C-A	C-A	
39883	Intergenic psbZ - trnG-GCC	T-A	T-A	T-A	
39885	Intergenic psbZ - trnG-GCC	C-T	C-T	C-T	
39961	Intergenic psbZ - trnG-GCC	T-G	T-G	T-G	
42474	Gene psaB	C-T	C-T	C-T	G - S
43685	Gene psaA	T-G			syn
43910	Gene psaA	C-G			syn
50339	Intergenic trnT-UGU - trnL-UAA	T-G			
50627	Intergenic trnT-UGU - trnLUAA	T-A	T-A	T-A	
51700	Gene trnL-UAA			A-T	
54999	Intergenic ndh3 - trnV-UAC		T-G	T-G	
55094	Intergenic ndh3 - trnV-UAC	T-C	T-C	T-C	
59143	Intergenic atpB - rbcL	A-T			

©Vigne et Vin Publications Internationales (Bordeaux, France)

Table 3. The intravarietal SNPs and amino acid substitutions in Georgian grape chloroplast DNA.

Nucleotide Position	Locus	Saperavi	Rkatsiteli	Meskhuri Mtsvane	Amino Acid Substitutions
61186	Intergenic rbcL - accD	T-G	T-G	T-G	
63478	Intergenic accD - psaI	T-A	T-A	T-A	
63819	Gene psaI	A-G	A-G	A-G	syn
67651	Intergenic petA - psbJ	T-C	T-C	T-C	
70248	Intergenic psbE - petL		A-C	A-C	
70595	Intergenic psbE - petL	C-T	C-T	C-T	
70921	Intergenic petL - petG		G-A	G-A	
71588	Intergenic trnP-UGG - psaJ	C-T	C-T	C-T	
73579	Gene rpl20		C-T		syn
73765	Gene rpl20	G-A			syn
75398	Intron clpP			C-T	
77099	Intergenic clpP - psbB	C-T	C-T	C-T	
80022	Intron petB	A-C	A-C	A-C	
80194	Intron petB	C-T			
80276	Intron petB	C-T	C-T	C-T	
86715	Intron rpl16		T-A		
86721	Intron rpl16	A-T			
89112	Gene rps19	C-T			syn
99217	Intergenic ycf2 - trnL-CAA		A-T	A-T	
99218	Intergenic ycf2 - trnL-CAA		T-G	T-G	
99219	Intergenic ycf2 - trnL-CAA		C-A	C-A	
99220	Intergenic ycf2 - trnL-CAA		A-T	A-T	
115531	Intergenic ycf1 - ndhF	A-T			
117310	Gene ndhF	T-G	T-G	T-G	syn
117791	Intergenic ndhF - rpl32	T-A	T-A	T-A	
119481	Intergenic rpl32 - trnL-UAG	G-T	G-T	G-T	
119571	Intergenic rpl32 - trnL-UAG	T-G	T-G	T-G	
121732	Intergenic ycf5 - ndhD	A-C	A-C	A-C	
123321	Intergenic ndh4 - psaC	A-T	A-T	A-T	
123367	Intergenic ndhD - psaC			T-A	
123664	Intergenic psaC - ndhE		T-G	T-G	
123690	Intergenic psaC - ndhE	C-A			
124258	Intergenic ndhE - ndhG	G-A	G-A	G-A	
125681	Gene ndhI	G-A	G-A	G-A	syn
126020	Gene ndhA	G-T			L - M
128420	Gene ndhH	C-T			E - K
131820	Gene ycf1	T-C	T-C	T-C	Q - R
133211	Gene ycf1		T-G	T-G	syn
133359	Gene ycf1	A-C			L - W
133934	Gene ycf1	G-T	G-T	G-T	syn
133982	Gene ycf1			A-C	I - M

©Vigne et Vin Publications Internationales (Bordeaux, France)

(Chardonnay, Gouais Blanc, Chasselas, Aladasturi, Krakhuna, Tsolikouri, Tsitska and Rkatsiteli (as a control)) and species (*V. champinii* and *V. rupestris* (not presented in this article)) were sequenced simultaneously in one Illumina lane. Preliminary results show that, together with chloroplast DNA, grape mitochondrial DNA sequences of 10 samples can be computed. In average, the number of reads per base for each chloroplast DNA sample varied between 1000 - 2000. Therefore, it is possible to increase the amount of sequenced chloroplast DNA by 2-3 times based on the fact that for determination of chloroplast DNA sequence, 50-100 reads are sufficient.

The comparison of sequenced chloroplast DNA gave 100 % identity to Chardonnay and Gouais Blanc, differing from Meskhuri Mtsvane by 1-bp insertion (position 5415) and 1-bp deletion (position 121,616) (all ATA) ; the difference between Chasselas and Saperavi was a 1-bp insertion (position 39554) (both ATT) ; Maxxa, Chkhaveri, Aladasturi, Krakhuna, Tsitska and Tsolikouri (all GTA) gave 100 % identity.

DISCUSSION

To date, a sequence of chloroplast DNA is available for only one grape cultivar, *V. vinifera* L. cv Maxxa (Jansen *et al.,* 2006). To that end, a bacterial artificial chromosome (BAC) library of Maxxa genomic DNA was constructed, BAC clones containing the chloroplast genome inserts were isolated, and the nucleotide sequences of the BAC clones were determined by the shotgun method (Jansen *et al.,* 2006).

During the determination of a high quality draft genome sequence of a cultivated clone of *V. vinifera* L. cv Pinot noir, Velasco *et al.* showed that the chloroplast DNA sequence of Pinot noir was identical (without a single mismatch) to that of Maxxa (Velasco *et al.,* 2007).

The most interesting result of the present study is that the chloroplast DNA sequence of Georgian grape cultivars is practically identical to that of West European cultivars : the chloroplast DNA of Chkhaveri and four other Georgian cultivars (Aladasturi, Krakhuna, Tsitska, Tsolikouri) is identical (also without a single mismatch) to the reference Maxxa and consequently to Pinot Noir chloroplast DNA. Saperavi and Chasselas chloroplast DNA differ by one insertion. Meskhuri Mtsvane chloroplast DNA differs by two indels from Chardonnay and Gouais Blanc chloroplast DNA. In general, it can be proposed that all GTA, ATA and ATT haplotypes contain practically identical chloroplast DNA. The single haplotype AAA, to which more than half of the analyzed Georgian cultivars belong, had no analogs in the world-wide set of grape cultivars until now (Beridze *et al.,* 2011).

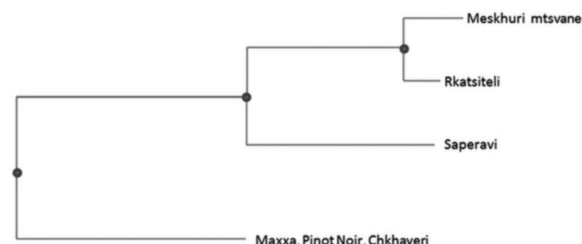

Figure 3. Complete chloroplast genome phylogeny of grape cultivars. The GenBank accessions used for the analyses are NC_007957.1 (Maxxa), Chkhaveri, Rkatsiteli, Saperavi and Meskhuri Mtsvane.

One of the differences between the four haplotypes of grape cultivars, besides SNPs, is a 54-bp deletion in the trnC-GCA – petN intergenic spacer (position 30,133 - 30,186), which is observed in AAA, ATA and ATT haplotypes and absent in GTA haplotype. Additionally, AAA, ATA and ATT haplotypes show a 33-bp duplication in the Rps16 – trnQ-UUG intergenic spacer (position 6,658 - 6,691). In ATT haplotype, an 18-bp duplication in the trnS-GCU – trnG-GCC intergenic spacer (position 10,001 - 10,019) was detected.

The second substantial difference between the studied grape cultivars is the presence of an 18-bp duplication (TATTTCTATTTAATAACG ; position 19,527 - 19,544 ; rpoC2 gene) in Rkatsiteli (AAA haplotype) chloroplast DNA. As a result, a 6-aa-long peptide (TLLNRN ; position 934 - 939 in the rpoC2 protein) duplication is observed.

Table 4. Long indels in grape chloroplast DNA.

Nucleotide position	Locus	Haplogroup	Indellength	Indeltype
6,658-6,691	Rps16 – trnQ-UUG intergenic spacer	AAA, ATA, ATT	33 bp	Duplication
10,001-10,019	trnS-GCU – trnG-GCC intergenic spacer	ATT	18 bp	Duplication
30,133 - 30,186	trnC-GCA –petN intergenic spacer	AAA, ATA, ATT	54 bp	Deletion
19,527-19,544	rpoC2 gene	AAA	18 bp (6 aminoacid)	Duplication

©Vigne et Vin Publications Internationales (Bordeaux, France)

The phylogenetic scheme based on indel priority of grape chloroplast DNA haplotypes is presented (Figure 4). Comparative analysis shows that the separation of Chkhaveri (GTA haplotype) and three other cultivars (AAA, ATA, ATT haplotypes) occurred by a 33-bp duplication and a 54-bp deletion in AAA, ATA, ATT haplotypes (no other *Vitis* species show this 54-bp deletion). Further separation occurred between ATT and the two haplotypes ATA and AAA by an 18-bp duplication and between ATA and AAA again by an 18-bp duplication.

8 of 11 grape chloroplast DNAs were sequenced by the new methodology of simultaneous resequencing. This methodology can be used also for determination of mitochondrial DNA sequences. It has great potential for expanding both phylogenetic and population genetic information on the evolution of domesticated crops.

Myles *et al.* find support for a Near East origin of *vinifera* and present evidence of introgression from local *sylvestris* as grapes moved into Europe (Myles *et al.,* 2011). But exact DNA data of grape domestication is still absent. The sequencing of wild samples can shed light on grape domestication. Arroyo-Garcia *et al.* suggested the existence of at least two origins for grape cultivars, one in the Near East characterized by chlorotypes C (ATA haplotype in the present publication) and D (ATT haplotype) and a western one related to Iberian Peninsula cultivars and characterized by chlorotype A (GTA haplotype) (Arroyo-Garcia *et al.,* 2006). The geographic distribution observed for some chlorotypes in *sylvestris* groups is still observed in cultivated groups. Cultivars with chlorotype A are highly abundant in Western Europe while they are not observed in the Near and Middle East samples. Similarly, chlorotypes C and D, which are very common among the Near East and Middle East cultivars, are less frequent among Iberian Peninsula cultivars.

Unfortunately the cultivars from South Caucasus were not analyzed in this investigation, though the role of South Caucasus in grape domestication was underlined. Later DNA sequence diversity was investigated in a group of 40 wild grape (*Vitis vinifera* subsp. *sylvestris*) samples from the South Caucasus at four plastid regions (*accD-psa*I intergenic spacer was added ; 63,186 - T, C). This group included 22 samples from Georgia, 9 samples from Azerbaijan, 2 samples from Armenia and 7 samples from Turkey (the base in the parenthesis corresponds to *accD-psa*I intergenic spacer) (Pipia *et al.,* 2012).

As in the case of the world-wide set of grape cultivars, four plastid haplotypes were evident in the 40 wild samples : AAA(T) – 22 samples, ATT(T) – 6 samples, GTA(C) – 1 sample, and ATA(T) – 11 samples. The AAA(T) haplotype was restricted to Georgia and Azerbaijan, the ATA(T) haplotype was distributed across the entire study area, the ATT(T) haplotype was distributed in the southern part of the study area from the Black Sea to the Caspian Sea, and the GTA(C) haplotype was only found in southwestern Georgia. The AAA(T) haplotype is restricted to both wild (*V. vinifera* subsp. *sylvestris*) and cultivated (*V. vinifera* subsp. *vinifera*) grape samples from the Caucasus. The observation that haplotype AAA(T) is found only among Georgian grape cultivars and is restricted to wild samples of Georgia and Azerbaijan suggests that this haplotype was domesticated in South Caucasus. It was proposed that the initial grape dispersal to Greece and Egypt may have occurred not only by land but also by sea, because the AAA(T) haplotype was formed in the Alazani hearth (far from the Black Sea) and this haplotype is absent in the world-wide set of cultivars (Pipia *et al.,* 2012).

As mentioned by Arroyo-Garcia *et al.* many of the current grape varieties can be traced back hundred and even thousand years based on historical records ; they are probably separated from their wild relatives by a

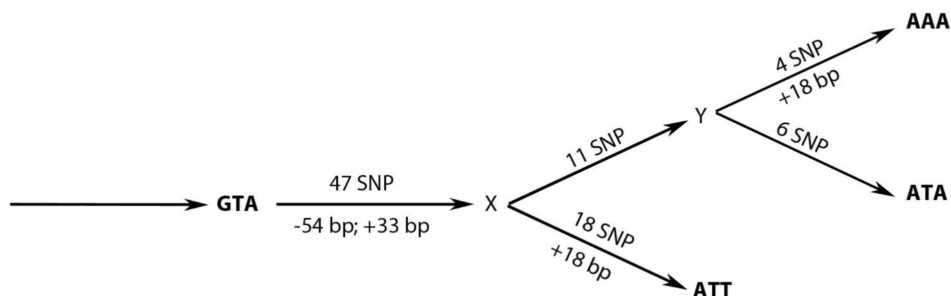

Figure 4. The phylogenetic scheme based on single nucleotide polymorphisms (SNP) and insertions/deletions (indel) priority of grape chloroplast DNA haplogroups.

©Vigne et Vin Publications Internationales (Bordeaux, France)

low number of sexual generations (Arroyo-Garcia L. *et al.,* 2006). Regarding grape domestication, currently only one evidence seems credible : AAA haplotype is restricted to both wild and cultivated grape samples in South Caucasus (Pipia *et al.,* 2012). This haplotype seems to have been domesticated in South Caucasus.

CONCLUSIONS

To date, a complete sequence of chloroplast DNA has been published for the grape cultivar Maxxa. In the present investigation, the chloroplast DNA sequence of eight Georgian grape cultivars (Rkatsiteli, Saperavi, Meskhuri Mtsvane, Chkhaveri, Aladasturi, Krakhuna, Tsitska, Tsolikouri) and three French cultivars (Chardonnay, Gouais Blanc, Chasselas), belonging to four different haplogroups (AAA, ATT, ATA, GTA), was determined by Illumina resequencing of genomic DNA. A new methodology of simultaneous resequencing of large number of chloroplast DNA was developed without preliminary chloroplast isolation or chloroplast enrichment. This method has great potential for expanding both phylogenetic and population genetic information on the evolution of domesticated crops.

Acknowledgements : The authors would like to thank M. K. Bendukidze for permanent interest and support. This research was funded by Knowledge Fund. Knowledge Fund is the funding organization of the Free University of Tbilisi and Agricultural University of Georgia.

REFERENCES

Aoki K., Matsumura T., Hattori T. and Murakami N., 2006. Chloroplast DNA phylogeography of *Photinia glabra* (Rosaceae) in Japan. *Am. J. Botany* **93** (12): 1852-1858.

Arroyo-Garcia R., Ruiz-Garcia L., Bolling L., Ocete R., Lopez M.A., Arnold C., Ergul A., Söylemezoğlu G., Uzun H.I., Cabello F., Ibanez J., Aradhya M.K., Atanassov A., Atanassov I., Balint S., Cenis J.L., Costantini L., Gorislavets S., Grando M.S., Klein B.Y., McGovern P.E., Merdinoglu D., Pejic I., Pelsy F., Primikirios N., Risovannaya V., Roubelakis-Angelakis K.A., Snoussi H., Sotiri P., Tamhankar S., This P., Troshin L., Malpica J.M., Lefort F. and Martinez-Zapater J.M., 2006. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology* **15** (12) : 3707-3714.

Beridze T., Pipia I., Beck J., Hsu S.-C., Gamkrelidze M., Gogniashvili M., Tabidze V., This P., Bacilieri R., Gotsiridze V., Glonti M. and Schaal B., 2011. Plastid DNA sequence diversity in a worldwide set of grapevine cultivars (*Vitis vinifera* L. subsp. *vinifera*). Bulletin of the Georgian National Academy of Sciences **5** (N1) : 98-103.

Clement M., Posada D. and Crandall K.A., 2000. TCS : a computer program to estimate gene genealogies. *Molecular Ecology* **9** (10) : 1657-1659.

Desper R. and Gascuel O., 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Computational Biology* **9** (5) : 687-705.

Gutiérrez-Rodríguez C., Ornelas J.F. and Rodríguez-Gómez F., 2011. Chloroplast DNA phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past fragmentation and demographic expansion in Mexican cloud forests. *Molecular Phylogenetics and Evolution* **61** (3) : 603-615.

Hancock-Hanser B.L., Frey A., Leslie M.S., Dutton P.H., Archer F.I. and Morin P.A., 2013. Targeted multiplex next-generation sequencing : advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources* **13** (2) : 254-268.

Imazio S., Maghradze D., de Lorenzis G., Bacilieri R., Laucou V., This P., Scienza A. and Failla O., 2013. From the cradle of grapevine domestication : molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. *Tree Genetics & Genomes* **9** (3) : 641-658.

Jansen R.K., Kaittanis C., Saski C., Lee S.-B., Tomkins J., Alverson A. and Daniell H., 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 Evolutionary *Biology* **6** (1) : 32.

Ketskhoveli N., Ramishvili M. and Tabidze D., 1960. *Ampelography of Georgia*. Metsniereba Publishing House, Tbilisi, Georgia.

Langmead B. and Salzberg S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9** (4) : 357-359.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and G.P.D.P. Subgroup, 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Li R., Yu C., Li Y., Lam T.-W., Yiu S.-M., Kristiansen K. and Wang J., 2009b. SOAP2 : an improved ultrafast tool for short read alignment. *Bioinformatics* **25** (15) : 1966-1967.

Lodhi M.A., Ye G.-N., Weeden N.F. and Reisch B.I., 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* **12** (1) : 6-13.

Middleton C.P., Senerchia N., Stein N., Akhunov E.D., Keller B., Wicker T. and Kilian B., 2014. Sequencing of chloroplast genomes from wheat, barley, rye and

©Vigne et Vin Publications Internationales (Bordeaux, France)

their relatives provides a detailed insight into the evolution of the *Triticeae* tribe. *PLoS One* **9** (3) : e85761.

Myles S., Boyko A.R., Owens C.L., Brown P.J., Grassi F., Aradhya M.K., Prins B., Reynolds A., Chia J.-M., Ware D., Bustamante C.D. and Buckler E.S., 2011. Genetic structure and domestication history of the grape. *Proc. the National Academy of Sciences of the United States of America* **108** (9) : 3530-3535.

Negrul A.M., 1946. *Ampelography of USSR*. Pishchepromizdat.

Pipia I., Gogniashvili M., Tabidze V., Beridze T., Gamkrelidze M., Gotsiridze V., Melyan G., Musayev M., Salimov V., Beck J.B. and Schaal B., 2012. Plastid DNA sequence diversity in wild grapevine samples (*Vitis vinifera* subsp. *sylvestris*) from the Caucasus regio*n. Vitis* **51**(3): 119-124.

Rambaut A., 2002. SE-AL Sequence Alignment Program. Department of Zoology, University of Oxford, UK.

Velasco R., Zharkikh A., Troggio M., Cartwright D.A., Cestaro A., Pruss D., Pindo M., FitzGerald L.M., Vezzulli S., Reid J., Malacarne G., Iliev D., Coppola G., Wardell B., Micheletti D., Macalma T., Facci M., Mitchell J.T., Perazzolli M., Eldredge G., Gatto P., Oyzerski R., Moretto M., Gutin N., Stefanini M., Chen Y., Segala C., Davenport C., Demattè L., Mraz A., Battilana J., Stormo K., Costa F., Tao Q., Si-Ammour A., Harkins T., Lackey A., Perbost C., Taillon B., Stella A., Solovyev V., Fawcett J.A., Sterck L., Vandepoele K., Grando S.M., Toppo S., Moser C., Lanchbury J., Bogden R., Skolnick M., Sgaramella V., Bhatnagar S.K., Fontana P., Gutin A., Van de Peer Y., Salamini F. and Viola R. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2** (12) : e1326.

©Vigne et Vin Publications Internationales (Bordeaux, France)