Ia Pipia[1*], Thierry Lacombe[2], Adam Kotorashvili[3], Nato Kotaria[3], Vazha Tabidze[1]

Complete chloroplast genome sequencing of *Vitis vinifera* subsp. *sylvestris* – wild ancestors of cultivated grapevines

Affiliations
[1] Institute of Molecular Genetics, Agricultural University of Georgia, Tbilisi, Georgia
[2] AGAP, Université de Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
[3] National Center for Disease Control and Public Health (NCDC&PH), Tbilisi, Georgia

Correspondence
Ia Pipia*: i.pipia@agruni.edu.ge, Thierry Lacombe: Thierry.lacombe@supagro.fr, Adam Kotorashvili: adam.kotorashvili@gmail.com, Nato Kotaria: natokotaria@gmail.com, Vazha Tabidze: v.tabidze@agruni.edu.ge

Summary

Wild grapevine – *Vitis vinifera* subsp. *sylvestris* – ancestors of cultivated grapevines are the main players in understanding the molecular bases of the grapevine domestication process. The goal of the presented research was to assess the genetic diversity of wild grapevine samples from several regions encompassing Europe (Spain, France, Germany, Hungary, Greece), the Mediterranean basin (Algeria, Tunisia, Morocco), and South Caucasus (Georgia), using a complete chloroplast DNA sequencing. The results suggest the existence of three different chloroplast DNA haplotypes reflecting the geographical distribution of the analyzed samples. This study represents the first report focused on analysis of a wide range of wild grapevine samples (*Vitis vinifera* subsp. *sylvestris*), applying next-generation technologies, and tracing the grapevine ancestry.

Keywords

***Vitis vinifera* subsp. *sylvestris*, Plastid DNA, Next-generation sequencing.**

Introduction

From ancient times grapevine has been one of the most important agricultural crops in the world. The cultivated forms of Eurasian species *Vitis vinifera* L. represent the vast majority of grapevines producing high quality wines and table grapevines. Considering the basis of the morphological, biological, ecological, and genetic differences, two subspecies of *Vitis vinifera* have been described, *V. vinifera* subsp. *vinifera* (or subsp. *sativa*), which includes thousands of cultivars (up to 7,000), and *V. vinifera* subsp. *sylvestris*, which corresponds to the wild type of *Vitis vinifera*. Considering that wild grapevines represent the progenitors of cultivated varieties, the analysis of their genetic profiles is of key interest for understanding the grapevine domestication process, which molecular bases are still not clearly known (Myles *et al.*, 2011; Grassi and Arroyo-García, 2020). Thus, the comparison between wild and cultivated population haplotypes from a species makes it possible to elucidate the geographic area where ancestral wild populations were domesticated. In this sense, the geographical analysis of genetic variations, commonly referred to as Phyogeography (Avise *et al.*, 1987; Avise, 2000), has long been used to investigate plant domestication of important crop species (Olsen and Schaal, 1999; Olsen and Purugganan, 2002; Singh, 2001; Matsuoka *et al.*, 2002).

Wild grapevines *V. vinifera* subsp. *sylvestris* represent the only taxon of the *Vitis* genus aboriginal to Eurasia. It first appeared approx. 65 million years ago, they are predominantly forest climbers and distributed in disjunct populations from the Atlantic coast to Tadzhikistan and the western Himalayas. Subspecies *sylvestris* is dioecious, exhibiting female and male flowers on different individuals, and a similar proportion in the populations; berries are small and acidic compared to cultivated grapevines (Zohary and Hopf, 2000).

Although *sylvestris* grapevines were spread over Southern Europe and Western and Central Asia during the Neolithic period, archeological and historical evidence suggest that primary domestication events would have occurred in the Near-East (McGovern and Rudolph, 1996). It seems that specific climatic conditions in this area were favorable for the diversification of the wild varieties, and from which cultivated grapevines could be chosen for domestication. Also, in this region the natural spread of *V. vinifera* most closely approaches the probable origin of Western agriculture (Jackson, 2014). Furthermore, several studies supported the existence of secondary domestication events along the Mediterranean basin (Aradhya *et al.*, 2003; Grassi *et al.*, 2003; Arroyo-García *et al.*, 2006; Lopes *et al.*, 2009; Andrés *et al.*, 2012). According to many researchers, cultivated grapevine was domesticated around 6,000 BC. Chemical analyses of organic compound deposits found in ancient pottery fabrics in Georgia (South Caucasus) since the early Neolithic period, represent the earliest biomolecular evidence to date for grape wine and viniculture from the Near East, corresponding to ca. 6,000–5,800 BC.

Submitted/accepted for publication: August 19, 2022/March 15, 2023

This discovery is crucial to trace the later history of wine in Europe and the rest of the world (McGovern *et al.*, 2017). In addition, based on different archeological and paleobotanical findings, several studies have pointed Georgia as the center of grapevine domestication (Zohary and Hopf, 2000; McGovern, 2003; Laucou *et al.*, 2011; Imazio *et al.*, 2013; Ekhvaia *et al.*, 2014; De Lorenzis *et al.*, 2015; Laucou *et al.*, 2018; Sargolzaei *et al.*, 2021).

Plastid DNA analysis enables insights into many aspects of evolutionary process, such as plant origin and population history. Considering the relatively small size of plastid DNA, its low rate of nucleotide substitutions, and conserved molecular structure constitute some of the reasons for its extensive use in molecular genetic studies. Using nine chloroplast microsatellite markers, Arroyo-Garcia *et al.*, (2006) analyzed chlorotype variation and distribution in samples of *sylvestris* and *sativa* genotypes from the eight large population groups according to physical geographic barriers: Iberian Peninsula, Central Europe, Northern Africa, Italian Peninsula, Balkan Peninsula, Eastern Europe, Near East and Middle East, and areas where *sylvestris* is distributed, determining that the studied genomes were divided into eight different chlorotypes. In the case of the wild grapevines, the four most frequent chlorotypes were A, B, and C, while in *sativa* samples was the chlorotype D. Although, the chlorotype A is prevalent in European *sylvestris* populations, it was not found in the Near East. In the case of chlorotypes C and D, they are frequent in populations of the Near East, but were not found farther west (Arrayo-Garcia *et al.*, 2006). Concordantly, our previous works analyzing the plastid DNA sequence on three non-coding regions (*trn*H-*psb*A intergenic spacer, *rpl*16 intron, *ac*D-*psa*I intergenic spacer) showed that the worldwide set of cultivated and South Caucasian wild grapevines share four DNA haplotypes of a so called 'maternal line' (Beridze *et al.*, 2011; Pipia *et al.*, 2012). Each haplotype contains the identifier of the number of each grapevine sample, designated with a three-letter acronym which corresponds to the substitution type in plastid genomes and the informal group name (Fig. 1). Three of the four haplotypes (Chkhaveri-Pinot noir haplotype – GTA, Meskhuri Mtsvane-Chardonnay haplotype – ATA and Saperavi-Cabernet Sauvignon haplotype – ATT), were found in both Georgian and worldwide grape varieties sets, while the Rkatsiteli haplotype (AAA) was identified in most (>50%) of the analyzed Georgian cultivars, representing a genetically unique group which has no analogue in the world to date. All the four haplotypes were found in South Caucasian wild grapevines, where the global restriction of Rkatsiteli haplotype (AAA) was also detected (Pipia *et al.*, 2012). The sharing of plastid haplotypes between cultivated and wild grapevines indicates genetic relationships of worldwide grapevine cultivars and South Caucasian wild populations, supporting earlier evidence about this geographic region as a possible center of grapevine domestication (Myles *et al.*, 2011).

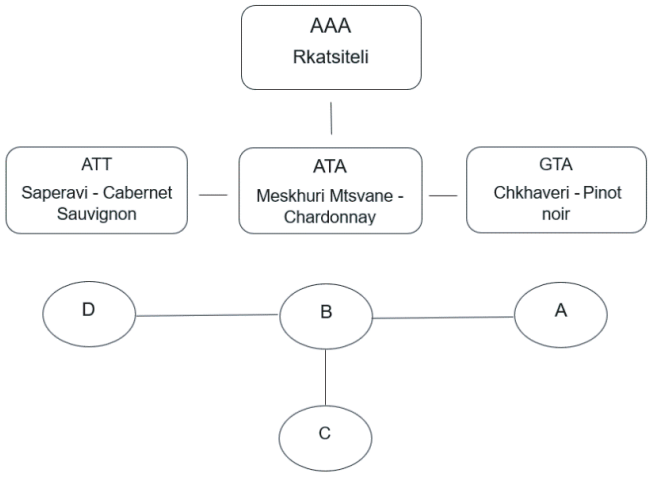

Fig. 1: Joint dates of *Vitis* haplotype network obtained from plastid DNA non-coding regions (upper side) (Beridze *et al.*, 2011; Pipia *et al.*, 2012) and Chlorotypes detected by Arrayo-Garcia *et al.*, (bellow) (Arrayo-Garcia *et al.*, 2006). According to this scheme Chlorotype A correlates with GTA haplotype, Chlorotype B/C with ATA haplotype and Chlorotype D with ATT haplotype.

Interestingly, some coincidences were found between the chlorotypes identified by Arrayo-Garcia *et al.* (2006), and aforementioned haplotypes (Beridze *et al.*, 2011; Pipia *et al.*, 2012; Pipia *et al.*, 2017). In particular, there is certain tendency among chlorotype A and haplotype GTA, chlorotype B/C and haplotype ATA, and chlorotype D with haplotype ATT (Pipia *et al.*, 2012) (Fig. 1). It should be noted that in the study developed by Arroyo-Garcia *et al.*, (2006), Georgian grapevine samples were not included, which may explain the lack of the possible overlapping between chlorotypes and AAA haplotype.

During the last decades, the next-generation plastid DNA genomics has emerged as a powerful and increasingly accessible tool for plant phylogenetics. Phylogenetic relationships between grapevine cultivars determined by complete plastid DNA sequence analysis revealed a high level of identity among members of the same haplogroup, while Georgian grapevine cultivars were practically identical to that of West European cultivars (Tabidze *et al.*, 2014). Based on 95 non-coding plastid regions inferred from BEAST using the random local clock model, the approximate age of aforementioned plastid haplotypes was also determined. According to the Chronogram of the main lineages of Vitaceae, Chkhaveri-Pinot-like genomes date to 5-4 Ma, while for Saperavi, Meskhuri Mtsvane and Rkatsiteli line age is 3-1 Ma (Zecca *et al.*, 2019).

The main goal of the present research was to analyze the whole plastid genome sequence diversity of wild grapevines (*V. vinifera* subsp. *sylvestris*) – from South Caucasus, especially Georgia, Europe, and the Mediterranean basin by using complete chloroplast DNA Illumina sequencing. The research was performed through the following steps: (1) A complete chloroplast DNA sequencing of wild grapevine accessions was achieved; (2) Identification of the haplotype for each analyzed plastid genome; (3) Assessment of plastid genome sequence diversity by identification of SNPs and InDels among the studied genomes.

Material and Methods

All samples from Georgian wild grapevines (*V. vinifera* subsp. *sylvestris*) were collected in the field trips in South-cen-

tral and East Georgia, during the years 2008 to 2010. Other wild grapevine samples from Europe and the Mediterranean basin were received from the National Research Institute for Agriculture, Food and Environment (INRAE), Grapevine collection of Vassal-Montpellier, France (Table 1). Cuttings of grapevine samples were grown to collect young green leaves for DNA extraction. Leaves were ground using liquid nitrogen for homogenization, and DNA extraction was conducted by using CTAB-based protocol (Lodhi *et al.*, 1994). The same extraction method was used for the leaves of Georgian wild grapevine samples collected directly during the field trips.

Construction of shotgun genomic DNA libraries and Sequencing

The genomic DNA libraries were constructed using the TruSeq DNA Sample preparation kit (Illumina, San Diego, CA, USA). Genomic DNAs were quantified using the Qubit BR reagents (Qbit 2.0 Fluorometer, Life Technologies). Briefly, 1 µg of DNA was sheared into 300 bp length fragments on a Covaris M220 focused ultrasonicator (CovarisInc) using SonoLabTM 7.1 software for 250 cycles/burst, 20.0 Duty factor, 50.0 Peak power, in screw-cap microtubes. After shearing, the DNA was blunt-ended, 30-end A-tailed and ligated to indexed adapters. The adaptor-ligated genomic DNA was selected by size with AMPure-beads using the gel-free protocol described in the TruSeq DNA Sample Preparation manual. Size selected DNA was amplified by PCR to selectively enrich for fragments that have adapters on both ends. Final-amplified libraries were run on an Agilent bioanalyzer DNA 2100 (Agilent, Santa Clara, CA, USA) to determine the average fragment size and to confirm the presence of DNA according to the expected size range. The libraries were pooled in equimolar concentration, loaded into a flow cell for cluster formation, and sequenced using an Illumina MiSeq platform, considering both ends of the molecules and a total read length. The libraries were sequenced from both ends of the molecules to a total read length of 150 nt from each end. The raw.bcl files were converted into demultiplexed compressed fastq files using Casava 1.8.2 (Illumina). DNA-sequence reactions were performed at the facilities of the National Centre for Disease Control and Public Health, Tbilisi, Georgia.

Assembly of plastid DNA

The total reads were trimmed using the computer program Sickle, a windowed adaptive trimming tool for FASTQ (a text-based format for storing nucleotide sequence) files using quality (https://github.com/najoshi/sickle). The reads were filtered by standard parameters (quality reads—20; cut-off length-20) and reads containing 'N' were discarded. The filtered reads were assembled using the SOAPdenovo2 software program (version 127mer) (Li *et al.*, 2009a; Li *et al.*, 2009b). The reads were first *de novo* assembled into contigs with k-mers 61–89. All contigs were aligned to the reference plastome sequence using BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The large overlapping contigs were merged using EMBOSS 6.3.1: merger (Rice *et al.*, 2000). The annotation of plastid DNA coding genes software was performed using GESEQ v. 1.42 software and nucleotide. BLAST (blastn). The annotated genome sequences are already available at NCBI (Table 2). For the construction of phylogenetic tree the MEGA program was used.

Table 2: The accession codes of the wild grapevine plastoms at NCBI.

Wild Grape	Accession number
Borjomi district (Georgia)	LC494572
Dusheti district (Georgia)	LC495883
Village Qsovrisi (Georgia)	LC523806
Lambrusque Tighzirt n°3 (Algeria)	LC501387
Lambrusque Ain Draham 1 (Tunisia)	LC508115
Lambrusque Beni Hassan S9 Ind 6 (Morocco)	LC510289
Vitis silvestris Rio Turon primera (Malaga, Spain)	LC507100
Teulere pied sauvage (France)	LC495478
Vitis silvestris Mannheim Nr 2 (Germany)	LC523989
Vitis silvestris 3-23 (Hungary)	LC508116
Vigne de Pausanias (Greece)	LC492108

Table 1: List of analyzed wild grapevine samples (*V. vinifera* ssp. *sylvestris*) received from INRA Vassal-Montpellier collection, France.

Accession name and code	Country Provenance	Sex	Identification Status
Lambrusque Tighzirt n°3 (8500Mtp140)	Algeria	-	Natural, wild
Lambrusque Ain Draham 1 (8500Mtp25)	Tunisia	Male	Semi-natural, feral
Lambrusque Beni Hassan S9 Ind 6 (8500Mtp232)	Morocco	-	Natural, wild
Vitis silvestris Rio Turon primera (Malaga) (8500Mtp49)	Spain	Female	Semi-natural, feral
Teulere pied sauvage (0Mtp1323)	France	Female	Natural, wild
Vitis silvestris Mannheim Nr 2 (8500Mtp144)	Germany	Male	Natural, wild
Vitis silvestris 3-23 (8500Mtp45)	Hungary	Hermaphrodite	Semi-natural, feral
Vigne de Pausanias (8500Mtp141)	Greece	Male	Natural, wild

Results and Discussion

Haplotype and size of analyzed plastoms

Eleven accessions covering eight samples from different places of Europe and the Mediterranean basin, and 3 samples from Georgia were analyzed by Illumina sequencing. Table 3 shows the list of sequenced grapevine chloroplast DNAs evaluated in this study, indicating the haplotype and genome size of these samples. It was found that most of the European and Mediterranean basin wild grapevine samples presented the GTA haplotype, except for the wild grapevine sample from Greece, which showed the ATA haplotype. The three samples from Georgia were identified as AAA haplotype, in concordance with our previous studies (Pipia *et al.*, 2012; Pipia *et al.*, 2015).

Polymorphism of analyzed complete plastoms: SNPs and Indels

The GTA haplotype plastid genome was analyzed revealing a complete sequence identity of wild grapevine samples from Algeria, Germany, and Hungary in relation to the reference Maxxa chloroplast DNA (GenBank: DQ424856.1), and some nucleotide polymorphisms were identified in other wild grapevine samples. Furthermore, the sequence analysis of samples presenting the GTA haplotypes showed the existence of six SNPs in coding gene regions, and three of them present non-synonymous substitutions, suggesting possible amino acidic alterations (Table 4). Interestingly, these three SNPs were found in protein-coding regions of *the mat*K *and ndh*F *genes*. In case of the samples Lambrusque Beni Hassan S9 Ind6 (Morocco) and Vitis silvestris Rio Turon primera (Malaga, Spain), a non-synonymous A/C substitution was observed at position 3,262 bp, located in the *mat*K gene region, according to the reference Maxxa DNA. This point mutation causes the change of Y (Tyrosine) amino acid by D (Aspartic Acid). In addition, the A/T missense mutation identified in Lambrusque BeniHassan S9 Ind 6 (Morocco) is located in the protein-coding region of the *ndh*F gene, position 115,958 bp according to the Maxxa genome, and it causes the substitution of amino acid I (Isoleucine) by K (Lysine). Also, the analysis of other point mutations found at the *clp*P and *rpl*22 gene region do not suggest any amino-acidic substitutions effects.

The *mat*K gene is used for the barcoding of different genome sequences. Considering that it has been reported PCR amplification data and DNA-sequence reactions of three genetic regions, *rbc*L, *mat*K and *trn*H-*psb*A from ancient grape DNA sample from Southern Italy (Gismondi *et al.*, 2016), we compared this ancient Italian grapevine's DNA with the plastid

Table 3: Complete chloroplast genome's lengths and haplotypes of analyzed wild grapevine samples.

Wild Grape accession	Country of Origin	DNA Length (bp)	Nucleotide positions (bp) according to Maxxa			Haplotype
			205	86715	86721	
Borjomi	Georgia	160.928	A	A	A	AAA
Dusheti	Georgia	160.928	A	A	A	AAA
Qsovrisi	Georgia	160.928	A	A	A	AAA
Lambrusque Tighzirt n°3	Algeria	160.928	G	T	A	GTA
Lambrusque Ain Draham 1	Tunisia	160.926	G	T	A	GTA
Lambrusque Beni Hassan S9 Ind 6	Morocco	160.926	G	T	A	GTA
Vitis silvestris Rio Turon primera	Spain	160.926	G	T	A	GTA
Teulere pied sauvage	France	160.929	G	T	A	GTA
Vitis silvestris Mannheim Nr 2	Germany	160.928	G	T	A	GTA
Vitis silvestris 3-23	Hungary	160.928	G	T	A	GTA
Vigne de Pausanias	Greece	160.909	A	T	A	ATA

Table 4: List of intravarietal SNPs and amino acid substitutions found in coding gene regions among representatives of GTA haplotype. The samples are named according to their country source and genome positions are given according to the reference Maxxa (GenBank: DQ424856.1) genome.

Nucleotide Position (bp)	Genome Locus	Algeria	Tunisia	Morocco	Spain	France	Germany	Hungary	Maxxa	Amino Acid Substitutions
3263	*mat*K gene	A	A	A/C	A/C	A	A	A	A	Y – D
76616	*clp*P gene	T	T	T/G	T/G	T	T	T	T	syn
88790	*rpl*22 gene	A	A	A/C	A	A	A	A	A	syn
115958	*ndh*F gene	A	A	A/T	A	A	A	A	A	I – K

genome sequences of European and Mediterranean wild grapevines. Our results showed that grape accessions included in this study revealed some similarity with the ancient DNA. Ancient *mat*K and *rbc*L nucleotide sequence analysis just showed 3 and 7 mutations respectively, while *trn*H-*psb*A genes displayed 11 SNPs and one 11bp deletion.

Regarding the plastid haplotypes analyzed, in the case of the ancient DNA, our results coincide with the data reported by Gismondi *et al.*, (2016), with one exception: We found one additional SNP in *trn*H-*psb*A gene region. The sequence alignment also revealed the presence of nine gaps: a single nucleotide insertion/deletion present in all of the analyzed grapevine samples exhibiting the GTA haplotype (Table 5). In addition, the wild Greece sample (Vigne de Pausanias), representative of ATA haplotype plastid genome, was compared with the Georgian Meskhuri Mtsvane (ATA haplotype), and six InDels (5 deletion and one insertion) were identified (Table 6).

Some of the analyzed wild grapevine genomes (*i.e.*, from Algeria, France, Germany, see Table 1), showed high identity with the reference Maxxa genome sequence, and no significant modifications were found. Therefore, this finding raises the question: are these samples true wild species or not?

The identification of true wild grapevines is quite a complex issue. Generally, there are three main step-procedures applied to identify the "true" wild grapevines: 1) Ecology or taking into consideration the location, e.g., distance from vineyards, etc.; 2) Morphology or ampelographic observations to distinguish individuals from wild *Vitis sp.* (hybrids, rootstocks) and then from *V. vinifera* cultivars; and 3) Genetic profiling using about 20 SSR markers to detect parentages, to exclude half-breeds with *sativa* and perform analysis of diversity to evaluate the position of wild individuals among the whole *V. vinifera* sample (*sativa* + other *sylvestris*) (Zdunic *et al.* 2017; Laucou *et al.* 2011). Considering these criteria, samples from Table 1 are true wilds, and there is no evidence of introgression from *sativa* gene pool. A high degree of *sativa* introgression into *sylvestris* can be easily detected by ampelography; however, in the case of medium degree of integration, nuclear SSR markers can be helpful detecting the introgression. In

Table 5: List of Gaps found among representatives of GTA haplotype. Positions are given according to the reference Maxxa (GenBank: DQ424856.1) genome.

Nucleotide Position	Genome Locus	Algeria	Tunisia	Morocco	Spain	France	Germany	Hungary	Maxxa
4423	Intergenic *trn*K-*rps*16	-	-	+ A insertion	+ A insertion	-	-	-	-
14789	Intergenic *atp*F-*atp*H	T	T	- T deletion	- T deletion	T	T	T	T
32585	Intergenic *psb*M-*trn*D	T	- T deletion	T	T	T	T	T	T
39588	Intergenic *psb*Z- *trn*G	T	T	- T deletion	- T deletion	T	T	T	T
76250	Intron *clp*P gene	-	-	+ A insertion	+ A insertion	-	-	-	-
82718	Intergenic *pet*D- *rpo*A	A	- A deletion	A	A	A	A	A	A
89221	Intergenic *rps*19- *rpl*2	-	-	-	-	+ G insertion	-	-	-
123308	Intergenic *ndh*D- *psa*C	A	A	- A deletion	- A deletion	A	A	A	A
123309	Intergenic *ndh*D- *psa*C	A	A	- A deletion	- A deletion	A	A	A	A

Table 6: Gaps found by comparative analysis of wild grapevine of Greece (Vigne de Pausanias) and Georgian cultivar Meskhuri Mtsvane (ATA Haplotype). Positions and genome regions are presented according to Meskhuri Mtsvane reference genome (GenBank: AB856291).

Nucleotide Position (bp)	Genome Locus	Vigne de Pausanias (Greece)	Meskhuri Mtsvane Reference genome
5416	Intron rpl16	- C (deletion)	C
51762	Intergenic *trn*L-*trn*F	+T (Insertion)	-
63110	Intergenic *acc*D-*psa*I	-T (deletion)	T
89180	Intergenic *rps*19-*rpl*2	-T (deletion)	T
160824	Intergenic *rpl*23-*trn*H	-A (deletion)	A
160825	Intergenic *rpl*23-*trn*H	-A (deletion)	A

the case of a low degree of introgression, morphology and SSR markers are not powerful enough, and it is recommended to sequence other DNA regions to detect introgressions.

The numerous coincidences observed between the wild grapevine genomes and the reference Maxxa genome sequence could be explained considering that the chloroplast DNA regions we analyzed have a high frequency in the *sativa* gene pool, they can also be found in the *sylvestris'* gene pool. Interestingly, the origin of this pattern would have occurred in *vinifera* species before the separation *sylvestris/sativa*, and later more selected in the *sativa* pool.

Plastid genome diversity of Georgian wild grapevine samples

Concerning the plastid genome diversity observed among the three Georgian samples, two of them were collected in the Eastern part of Georgia (village Qsovrisi, Mtskheta district and Dusheti district area), while the third one in the South-central part of Georgia (Borjomi district), all of them presenting the AAA haplotype (Table 3). Sequence alignments revealed a complete identity among the plastoms of 'Qsovrisi' and 'Dusheti' samples with the reference Georgian cultivar Rkatsiteli (AAA haplotype), and in the case of 'Borjomi' samples, two 4-bp long inversions were detected in both copies of inverted repeat, representing *ycf*2-*trn*L-CAA and *trn*L-*CAA*-*ycf*2 intergenic regions in positions 99.218-99.221 and 150.840-150.843, respectively.

Polymorphism of analyzed complete plastoms: InDels

The comparison of the chloroplast DNAs sequences revealed the presence of four long size InDels in the plastid genomes of *Vitis vinifera* cultivars, highly informative for the phylogenetic and phylogeographic analysis. The 33-bp duplication in *rps*16-*trn*Q-UUG intergenic spacer and 54-bp deletion in the *trn*C-*GCA*-*pet*N intergenic spacer were detected in all the sequenced *Vitis vinifera* cultivars exhibiting AAA, ATT, and ATA haplogroups, excepting in the case of the cultivars of the GTA haplotype. In addition, two 18-bp duplications were found, the former located at the *trn*S-GCU-*trn*G-GCC intergenic spacer of the ATT haplogroup cultivars, and the latter, in the *rpo*C2 gene of Georgian cultivar Rkatsiteli (AAA haplotype), resulting in 6-amino acid-long peptides (TLLNRN in the rpoC2 protein) (Tabidze *et al.*, 2014), and also observed in the wild grape accessions considered in this study (Table 7). Considering the results observed in the wild grapevine samples from Europe and Mediterranean basin, the sample from Greece (Vigne de Pausanias) differs from the others in relation to two of the aforementioned InDels, 33-bp duplication in *rps*16-*trn*Q-UUG intergenic spacer and 54-bp deletion in the *trn*C-*GCA*-*pet*N intergenic spacer. However, the three InDels typically related to the *Vitis vinifera* ATA haplotype were also observed in the Georgian wild grapevine samples.

Phylogenetic Analysis

A total of 14 grapevine accessions covering 11 *Vitis sylvestris* samples plus three *Vitis vinifera* cultivars, representing the reference chloroplast DNA for the GTA, ATA and AAA haplotypes were subject to maximum likelihood (ML) analysis by the MEGA program. Fig. 2 shows the genetic relationship among the European, the Mediterranean basin and Georgian wild grapevine samples (*Vitis vinifera* subsp. *sylvestris*) based on the complete chloroplast DNA sequence analysis. The phylogenetic tree revealed three basic clades, in concordance to the haplotype distribution of the analyzed samples. Two groups were observed among the GTA haplotype samples, the former included by plastid genomes of wild grapevines from Algeria, France, Hungary and Germany, while in the latter, grapevine samples from Tunisia, Morocco and Spain were observed. Interestingly, the wild grapevine from Greece (Vigne de Pausanias) formed a separate clade along with the Georgian cultivar Meskhuri Mtsvane, used as reference chloroplast DNA for the ATA haplogroup samples. The third clade represents a well-supported group of grapevine accessions, including the Georgian wild samples as well as the Georgian cultivar Rkatsiteli (AAA haplotype).

Table 7: Long size indels found in a set of wild grapevine plastoms.

Nucleotide position (bp)	Locus	Indel Length	Indel Type	Wild grape samples	Haplotype
6.658-6.691	Intergenic *rps*16-*trn*Q-UUG	33 bp	Duplication	Vigne de Pausanias (Greece)	ATA
				Borjomi district (Georgia)	AAA
				Dusheti district (Georgia)	AAA
				Village Qsovrisi (Georgia)	AAA
30.133-30.186	Intergenic *trn*C-GCA – *pet*N	54 bp	Deletion	Vigne de Pausanias (Greece)	ATA
				Borjomi district (Georgia)	AAA
				Dusheti district (Georgia)	AAA
				Village Qsovrisi (Georgia)	AAA
19,527-19,544	*rpo*C2 gene	18 bp (6 amino acid)	Duplication	Borjomi district (Georgia)	AAA
				Dusheti district (Georgia)	AAA
				Village Qsovrisi (Georgia)	AAA

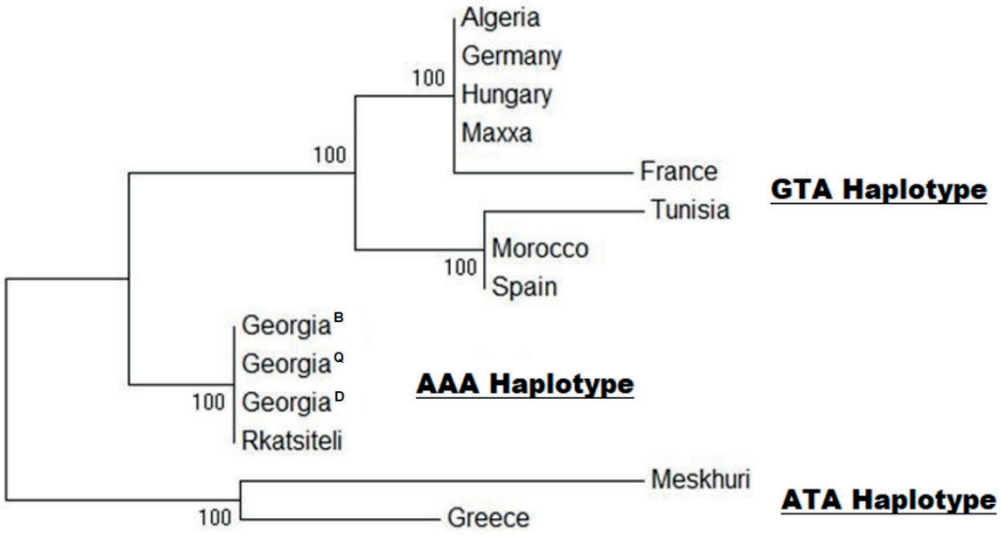

Fig. 2: Complete chloroplast genome phylogeny of Georgian, European and Mediterranean wild grapevine (*Vitis vinifera* ssp. *sylvestris*) conducted by maximum likelihood (ML) analysis of MEGA software. The samples are named according to their country source. In the case of Georgian wild grapevines GeorgiaQ and GeorgiaD represent East Georgian *sylvestris*, from village Qsovrisi and Dusheti district, GeorgiaB corresponds to south-central Georgian *sylvestris* from Borjomi district.

Conclusion

In this report, complete chloroplast DNA sequencing of eleven *Vitis sylvestris* accessions was performed. The obtained results can be formulated as follows: 1) Among the European and Mediterranean wild grapevines, the GTA haplotype prevails and only the Greek sample (Vigne de Pausanias) represents the ATA haplotype; 2) In three Georgian wild grapevine samples the AAA haplotype is present what is consistent with our previous study that the majority of Georgian wild grapevine samples represent the genetically unique AAA haplotype; 3) In the plastid genomes of wild grapevines from Morocco and Spain, presenting the GTA haplotypes, the existence of 6 SNPs in coding genes regions were found. Among them three present non-synonymous substitutions, suggesting possible amino acidic alterations; 4) Some SNPs and gaps were found in all wild grapevine plastoms analyzed in this research and long size InDels (33 bp duplication at the locus of intergenic *rps*16-*trn*Q-UUG, 54 bp deletion at the position of Intergenic *trn*C-*GCA*-*pet*N and 18 bp duplication at the position of the *rpo*C2 gene) were detected in the wild grapevine samples of Greece and Georgia; 5) Phylogenetic analyses of all the analyzed samples of wild grapevine genomes had shown that plastome distribution on the phylogenetic tree reflects both geographical distribution and haplotypes of each plastome. The results of the present research are important for the understanding of the genomic compositions of wild grapevine plastoms and the study of genetic relationships between wild and cultivated grapevine samples from different geographical locations, which can shed light on the molecular bases of the grapevine domestication processes in general. It should be noted that for better understanding of the aforementioned issues it is necessary to sequence more wild grapevine genomes to clarify and analyze the substitution rates within the grape family, which can be coped with in future research.

Acknowledgements

We thank C. Marchal and S. Lalet from INRAE Vassal-Montpellier collection for providing *sylvestris* plant material.

Conflicts of interest

The authors declare that they do not have any conflicts of interest.

References

De Andres, M. T., Benito, A., Pérez-Rivera, G., Ocete, R., Lopez, M. A., Gaforio, L., Muñoz-Organero, G., Cabello, F., Martínez-Zapater, J. M., Arroyo-García, R., 2012: Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. Molecular Ecology. 21(4): 800-816, DOI: 10.1111/j.1365-294X.2011.05395.x.

Aradhya, M. K., Dangl, G. S., Prins, B. H., *et al.* 2003: Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genetics Research, 81: 179–192, DOI: 10.1017/S0016672303006177.

Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M. A., Arnold, C., Ergul, A., Söylemezoğlu, G., Uzun, H. I., Cabello, F. *et al.*, 2006: Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. Molecular Ecology. 15 (12), 3707–3714 DOI: 10.1111/j.1365-294X.2006.03049.x.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., Saunders, N. C., 1987: Intraspecific Phylogeography: The Mitochondrial DNA BRIDGe between Population Genetics and Systematics. Annual Reviews of Ecology and Systematics, 18, 489-522, DOI: 10.1146/annurev.es.18.110187.002421.

Avise, J. C., 2000: Phylogeography. The History and Formation of Species. Harvard University Press.

Beridze, T., Pipia, I., Beck, J., Hsu, S.-C., Gamkrelidze, M., Gogniashvili, M., Tabidze, V., This, P., Bacilieri, R., Gotsiridze, V., *et al.*, 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*, 91–96.

De Lorenzis, G., Chipashvili, R., Failla, O., Maghradze, D., 2015: Study of genetic variability in *Vitis vinifera* L. germplasm by high-throughput Vitis18kSNP array: the case of Georgian genetic resources. BMC Plant Biology Journal. 15: 154. DOI: 10.1186/s12870-015-0510-9.

Ekhvaia, J., Gurushidze, M., Blattner, F. R., and Akhalkatsi, M., 2014: Genetic diversity of *Vitis vinifera* in Georgia: relationships between local cultivars and wild grapevine, V. *vinifera* L. subsp. *sylvestris*. Genetic. Resources and Crop Evolution, 61, 1507–1521. DOI: 10.1007/s10722-014-0125-2.

Gismondi, A., Di Marco, G., Martini, F., Sarti, L., Crespan, M., Martínez-Labarga, C., Rickards, O., Canini, A., 2016: Grapevine carpological remains revealed the existence of a Neolithic domesticated *Vitis vinifera* L. specimen containing ancient DNA partially preserved in modern ecotypes. Journal of Archeological Science: 69, 75-84. DOI: 10.1016/j.jas.2016.04.014.

Grassi, F., Labra, M., Imazio, S., Spada, A., Sgorbati S., Scienza A., Sala F., 2003: Evidence of a secondary grapevine domestication centre detected by SSR analysis. Theoretical and Applied Genetics, 107:1315–1320, DOI: 10.1007/s00122-003-1321-1.

Grassi, F., Arroyo-Garcia, R., 2020: Editorial: Origins and Domestication of the Grape. Frontiers in Plant Science, 11:1176. DOI: 10.3389/fpls.2020.01176.

Imazio, S., Maghradze, D., Lorenzis, G., Bacilieri, R., Laucou, V., This, P., *et al.* 2013: From the cradle of grapevine domestication: molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. Tree Genetics & Genomes 9, 641–658. DOI: 10.1007/s11295-013-0597-9.

Jackson, R., 2014: Wine Science, Principles and Application, 4th edn (New York, NY, USA: Academic Press).

Laucou, V., Lacombe, T., Dechesne, F., Siret, R., Bruno, J.P., Dessup, M., Dessup, T., Ortigosa, P., Parra, P., Roux, C., Santoni, S., Vares, D., Peros, J.P., Boursiquot, J.M., This, P., 2011: High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theoretical and Applied Genetics 122(6):1233–1245. DOI: 10.1007/s00122-010-1527-y.

Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A.-F., Bérard, A., *et al.* 2018: Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. PLoS One 13:e0192540. DOI: 10.1371/journal.pone.0192540.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009a: 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16): 2078-2079. DOI: 10.1093/bioinformatics/btp352.

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen K., Wang J., 2009b: SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25 (15): 1966-1967. DOI: 10.1093/bioinformatics/btp336.

Lodhi, A., Ye, G., Weeden, N.F., Reisch, B.I., 1994: A Simple and Efficient Method for DNA Extraction from Grapevine Cultivars and Vitis Species. Plant Molecular Biology Reporter, 12: 6-13. DOI: 10.1007/BF02668658.

Lopes, M. S., Mendonça, D., Rodrigues dos Santos, M., Eiras-Dias, J. E., da Câmara Machado, A., 2009: New insights on the genetic basis of Portuguese grapevine and on grapevine domestication. Genome 52(9):790-800. DOI: 10.1139/g09-048.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, G. J., Buckler, E., Doebley J., 2002: A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences of the United States of America., 99, pp. 6080-6084, DOI: 10.1073/PNAS.052125199.

McGovern, P. E., Rudolph, H. M., 1996: The analytical and archaeological challenge of detecting ancient wine: two case studies from the ancient Near East. In: McGovern P. E., Fleming S. J., Katz S. H., editors. The origins and ancient history of wine. New York: Gordon and Breach, pp. 57–67.

McGovern, P. E., 2003: Ancient Wine. Princeton, NJ: University Press.

McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M. P., Smith, K. E., Hall, G. R., Kvavadze, E., Maghradze, D., Rusishvili, N., Bouby, L., *et al.* 2017: Early Neolithic wine of Georgia in the South Caucasus. Proceedings of the National Academy of Sciences of the United States of America. 114 (48), E10309–E10318 DOI: 10.1073/pnas.1714728114.

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., Buckler, E. S., 2011: Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences *(*PNAS*)*, 108 (9) 3530-3535, DOI: 10.1073/pnas.1009363108.

Olsen, K. M., Schaal, B. A., 1999: Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences of the United States of America, 96, 5586-5591. DOI: 10.1073/pnas.96.10.5586.

Olsen, K. M., Purugganan, M. D., 2002: Molecular evidence on the origin and evolution of glutinous rice. Genetics, 162, 941-950.

Pipia, I., Gogniashvili, M., Tabidze, V., Beridze, T., Gamkrelidze, M., Gotsiridze, V., Melyan, G., Musayev, M., Salimov, V., Beck, J., Schaal, B., 2012: Plastid DNA sequence diversity in wild grapevine samples (*Vitis vinifera* subsp. *sylvestris*) from the Caucasus region. Vitis 51, 119–124. DOI: 10.5073/vitis.2012.51.119-124.

Pipia, I., Gogniashvili, M., Tabidze, V., Kotorashvili, A., Kotaria, N., 2015: Plastid DNA sequence diversity in wild grapevine

samples (*Vitis vinifera* ssp. *sylvestris*) from the Caucasus region: complete plastid genome sequences of three Georgian samples. XI International Conference on Grapevine Breeding and Genetics, Acta Horticulturae. 1082, 213–216 DOI: 10.17660/ActaHortic.2015.1082.29.

Pipia, I., Kunelauri, N., Gogniashvili, M., Kotaria, N., Kotorashvili, A., Lacombe, T., Tabidze, V., 2017: Complete plastid genomes of South Caucasian, European and Mediterranean Basin wild grapevines (*Vitis vinifera* subsp. *sylvestris*). Acta Horticulturae. 1188, 351–356 DOI: 10.17660/ActaHortic.2017.1188.47.

Rice, P., Longden, I., Bleasby, A., 2000: EMBOSS: the European molecular biology open software suite. Trends in Genetics 16 (6):276-277.

Sargolzaei, M., Rustioni, L., Cola, G., Ricciardi, V., Bianco, P. A., Maghradze, D., Failla, O., Quaglino, F., Toffolatti, S. L., De Lorenzis, G., 2021: Georgian Grapevine Cultivars: Ancient Biodiversity for Future Viticulture. Frontiers in Plant Science, 12:630122. DOI: 10.3389/fpls.2021.630122.

Singh, S. P., 2001: Broadening the genetic base of common bean cultivars. A Review. Crop Science, 41, 1659-1675. DOI: 10.2135/cropsci2001.1659.

Tabidze, V., Baramidze, G., Pipia, I., Gogniashvili, M., Ujmajuridze, L., Beridze, T., Hernandez, A. G., Schaal, B., 2014: The complete chloroplast DNA sequence of eleven grape cultivars. Simultaneous resequencing methodology. Journal International des Sciences de la Vigne et du Vin 48 (2), 99–109 DOI: 10.20870/oeno-one.2014.48.2.1568.

Zdunic, G., Maul, E., Eiras Dias, J. E. J., Munoz Organero, G., Carka, F., Maletic, E., Savvides, S., Jahnke, G. G., Nagy, Z. A., Nikolic, D., Ivanisevic, D., Beleski, K., Maras, V., Mugosa, M., Kodzulovic, V., Radic, T., Hancevic, K., Mucalo, A., Luksic, K., Butorac, L., Maggioni, L., Schneider, A., Schreiber, T., Lacombe, T., 2017: Guiding principles for identification, evaluation and conservation of *Vitis vinifera* subsp. *sylvestris*. Vitis 56 (3) pp.127–131. DOI: 10.5073/vitis.2017.56.127-131.

Zecca, G., Grassi, F., Tabidze, V., Pipia, I., Kotorashvili, A., Kotaria, N., Beridze, T., 2019: Dates and rates in grape's plastomes: evolution in slow motion. Current Genetics 66:123-140. DOI: 10.1007/s00294-019-01004-7.

Zohary, D., Hopf, M., 2000: Domestication of Plants in the Old World. Oxford University Press, Oxford UK.