

Chloroplast genomes of *Vitis* sp. samples from Nechaevsk population of the Krasnodar region

Andrej Zvyagin², Daria Savenkova^{1,*}, Anastasia Elisyutikova¹, and Leonid Troshin²

¹Laboratory of Plant and Animal genetics, Kuban State Agrarian University named after I.T. Trubilin, 350044 Kalinina 13, Krasnodar, Russian Federation

²Department of Viticulture, Kuban State Agrarian University named after I.T. Trubilin, 350044 Kalinina 13, Krasnodar, Russian Federation

Abstract. The presented study demonstrates the results of sequencing and analysis of wild grape plastomes assemblies located in the Nechaevsk population. Previously, the habitats of the wild forest grape *V. vinifera* subsp. *silvestris* were explored as an outcome of expeditions. As a result of repeated expeditions, samples were taken and used to isolate chloroplast DNA. Assembly and annotation of plastomes showed the presence of DNA encoding sequences. At the same time, the structure of the obtained plastomes varied. These expressed in the presence of DNA polymorphisms. Comparison of the aligned sequences made it possible to reveal the presence of the species *V. vinifera* subsp. *silvestris* in the population, as well as hybrid specimens. This is also confirmed by the result of determining the taxonomy after depositing the archives of readings in the NCBI database.

1 Introduction

The study of time and place of origin of cultivated grapevine and its domestication is one of the most important questions of modern ampelography. The exact time and place of the domestication event is still in question, while it is known that the cultivated grape (*Vitis vinifera* L.) itself originated from the wild forest grape *Vitis sylvestris* Gmel. There are two main theories about the domestication of grapes and the fundamental difference in them lies in the number of places for this domestication: one [1, 2] or several [3, 4]. However, the differences in form and genetic variability are more indicative of the use of different sources of genetic variability in grape breeding. The wild forest grape itself is widespread in Eurasia, which is represented by five chlorotypes [5]. At the same time, one of the most genetically diverse regions is the Caucasus, where all five chlorotypes were found among representatives of *Vitis sylvestris* Gmel. [5].

The process of domestication of the grapevine took place in the Middle East on the territory between the Black Sea and Central Asia [6]. The area meets the characteristics of a primary center of cultural domestication, including an agro-climatic predisposition for grape cultivation, a crossroads of trade flows, and the presence of socio-cultural trends. In general, archeo-botanical evidence indicates the emergence of viticulture around 6000–

* Corresponding author: dasha_19.99s@mail.ru

5800 BC. [7]. However, there are also difficulties in identifying of ancient remains of plants (*sylvestris* and *vinifera*), but the findings of grape seeds and tartaric acid residues indicates the practice of winemaking [8]. At the moment, the huge genetic variability available to us can be attributed to the selection, vegetative propagation and mutations during the evolution and cultivation of *Vitaceae* species. In addition, on the basis of anthropological conditions, historical and ampelographic data, hypotheses about additional centers of grape domestication were put forward. It is assumed that a secondary center of domestication arose in the Greek region closest to the Caucasus. After, this process was repeated during the colonization of Southern Italy and Sicily by the Greeks, and then during the Roman colonization of Southeastern Iberia [9].

Of course, grapevine traveled along with the settlers. In view of this, tracking the selection carried out by people of antiquity is possible with the help of chloroplast and mitochondrial DNA. It spreads from the mother plant to the offspring and, accordingly, helps to understand which ancestral line is. Of great interest is the study of the genetic diversity of wild grapes that have been used and can be used for crossing, as well as the determination of the places of initial domestication [3, 10, 11]. The high morpho-genetic variability among varieties in the Caucasian region also indicates the use of various morphotypes in breeding, and possibly wild forest grapes. In view of the fact that grape varieties interbreed well with each other, it is very important to know the origin of the maternal line. Therefore, in this study, we set the goal of studying the structure of chloroplast genomes of the wild-growing grapes of the Nechaevsk population of the Krasnodar region and comparing them with each other and those already known to clarify their phylogenetic position.

2 Materials and methods

The coordinates of wild grapes were previously published [12]. They were used for expeditions and biomaterial collection. Adult grape leaves were used for further chloroplasts isolation by a previously published protocol [13]. DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN). The quantity and quality of isolated DNA were determined using a Qubit 4 Fluorometer (Thermo Fisher Scientific). The Nextera DNA Flex Library Prep Kit (Illumina) was used to prepare DNA libraries. Sequencing was performed on a MiSeq instrument (Illumina).

FastQC was used for quality control [14]. Adapter sequences were removed using Trimmomatic v.0.39 [15]. Preliminary assembly of genomes was carried out in UGENE [16] with the built-in SPAdes v.3.15.3 algorithm [17]. Then reads were aligned to different genomes to select the most efficient one using the BWA program [18]. The chloroplast genome of *V. vinifera* PN40024 (GenBank number NC_007957.1) was used as a reference to assemble studied plastomes. To compare the results among themselves and against already available data the NCBI BLAST algorithm was used [19]. The search for possible CDS was done with the GeneMark.hmm [20]. Genomes were annotated using GeSeq [21] and visualized by OGDRAW [22].

3 Results and Discussion

As a result of the work, the genomes of 11 representatives of the Nechaevsk wild grape population were sequenced and assembled. The resulting reads were aligned to the reference sequence (GenBank number NC_007957.1), annotated, and deposited with NCBI as read archives. The results are represented in the Table 1.

Table 1. Results of chloroplast genomes assembly and annotation.

Sample	GC content	Coverage	CDS number	GenBank number	Taxonomy by NCBI
7.1_Nechajevsk	37.4%	25.0x	159	SRR23815006	<i>V. vinifera</i> subsp. <i>sylvestris</i>
8.1_Nechajevsk	37.4%	25.0x	159	SRR23815005	<i>Vitis</i> hybrid cultivar
9.1_Nechajevsk	37.4%	25.0x	159	SRR23815004	<i>Vitis</i> hybrid cultivar
10.1_Nechajevsk	37.4%	25.0x	156	SRR23815003	<i>Vitis</i> hybrid cultivar
11.1_Nechajevsk	37.4%	25.0x	159	SRR23815017	<i>V. vinifera</i> subsp. <i>sylvestris</i>
12.1_Nechajevsk	37.4%	25.0x	159	SRR23815016	<i>Vitis</i> hybrid cultivar
13.1_Nechajevsk	37.4%	25.0x	159	SRR23815015	<i>Vitis</i> hybrid cultivar
14.1_Nechajevsk	37.4%	25.0x	159	SRR23815014	<i>Vitis</i> hybrid cultivar
15.1_Nechajevsk	37.4%	25.0x	159	SRR23815013	<i>V. vinifera</i> subsp. <i>sylvestris</i>
16.1_Nechajevsk	37.4%	25.0x	159	SRR23815012	<i>Vitis</i> hybrid cultivar
17.1_Nechajevsk	37.4%	25.0x	158	SRR23815011	<i>V. vinifera</i> subsp. <i>sylvestris</i>

The results of sequencing, assembling and annotation of wild grape plastomes showed the presence of all common coding sequences, as well as introns. In view of the alignment to the reference genome of the obtained reads, the total length of the plastomes was the same. The average reading depth was 25x. It should be noted that despite the large amount of information received, there were gaps among the aligned reads, which most likely indicates that some sections were read better, while others were not. Depositing archives of reads into the database also made it possible to establish the species affiliation based on the output of the NCBI SRA Taxonomy Analysis Tool (STAT) [23]. Overall, the STAT results are consistent with the NCBI BLAST results, as some of the plastomes were identified as *V. vinifera* subsp. *sylvestris*, while taxonomic identification of other accessions was difficult. However, accessions 7.1_Nechajevsk, 11.1_Nechajevsk, 15.1_Nechajevsk and 17.1_Nechajevsk were identified as belonging to the wild forest grape species. Also, the resulting DNA sequences and plastomes of the samples were visualized (Fig. 1 and Fig.2).

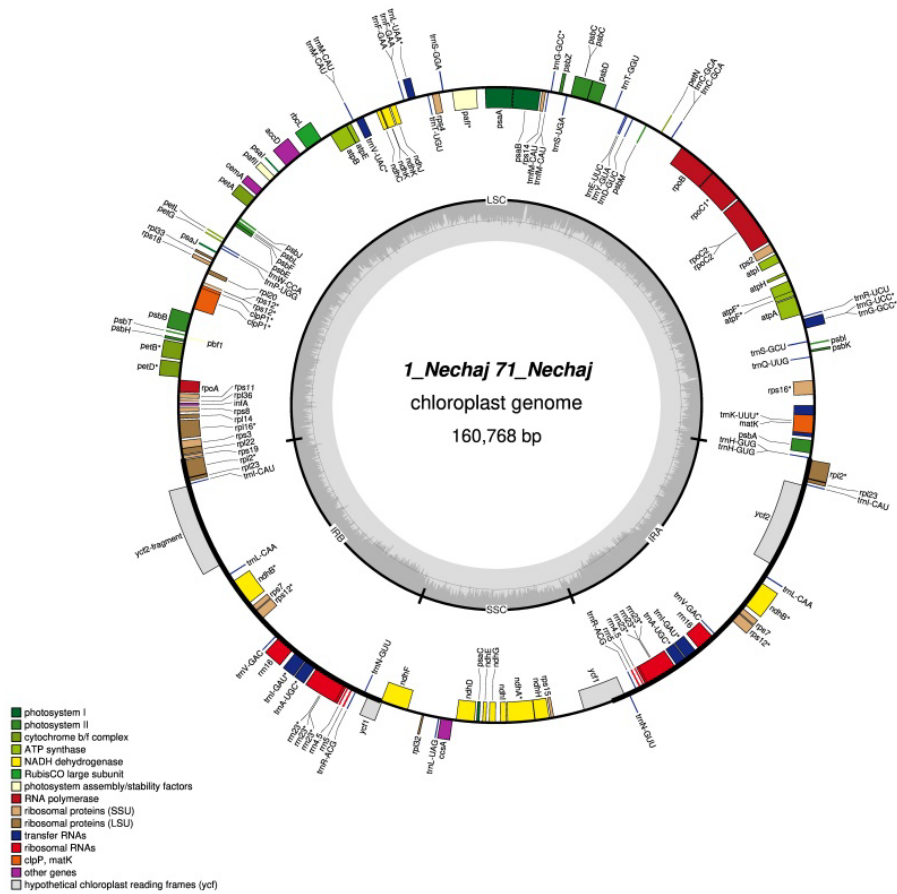


Fig. 1. Visualization of 7.1_Nechajevsk genome *V. vinifera* subsp. *sylvestris*.

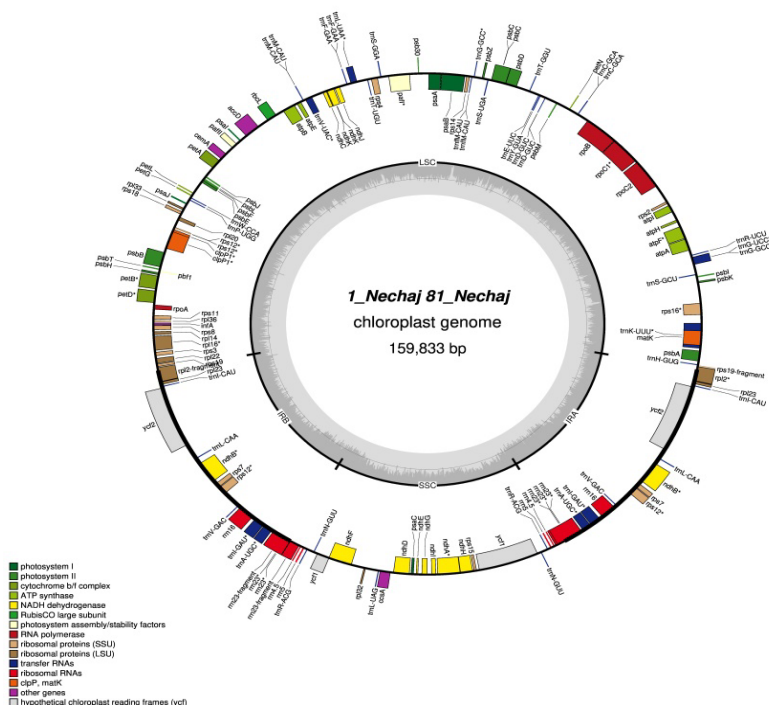


Fig. 2. Visualization of 8.1_Nechajevsk genome *Vitis* hybrid cultivar.

As can be seen from the two figures, a graphical comparison of them allows us to conclude that, although the genome of the 8.1_Nechajevsk sample was assembled with small gaps, its complete sequence also allows us to identify all the main genes. At the same time, it is interesting that the results of alignment of the full sequence in NCBI BLAST and STAT showed that this sample belongs to hybrids rather than forest grapes. This is interesting because chloroplasts in grapes are inherited maternally. In view of this, it can be concluded that some accessions can be introduced into the Nechaevsk population from outside, and not be descendants of the local wild grapes. For example, this is also directly indicated by the fact that one third of the readings according to the results of the taxonomic analysis of the readings belong to the species *Vitis riparia*. Alignment of DNA sequences among themselves revealed the presence of obvious similarities and differences. The differences are in the presence of length polymorphism and single nucleotide polymorphisms. If the first can be explained by insufficient reading depth, then the second may be of interest for a more detailed study.

4 Conclusion

As a result of the work, plastome maps of 11 representatives of the Nechaevsk population of wild forest grapes were created. The results obtained were deposited with the NCBI SRA under the numbers SRR23815006, SRR23815005, SRR23815004, SRR23815003, SRR23815017, SRR23815016, SRR23815015, SRR23815014, SRR23815013, SRR23815012 and SRR23815011 and are available to everyone. After assembly, the plastomes showed differences in nucleotides composition, but not in GC-content. The number of annotated coding sequences was the same for most samples (159), except for

10.1_Nechajevsk (156) and 17.1_Nechajevsk (158). At the same time, such a difference in structure did not prevent them from determining their taxonomic affiliation. Thus, it was revealed that most likely this population is a mixture of several species (or hybrid forms). Despite this, four individuals 7.1_Nechajevsk, 11.1_Nechajevsk, 15.1_Nechajevsk and 17.1_Nechajevsk are representatives of *V. vinifera* subsp. *sylvestris*.

Acknowledgements

This work was supported by the grant of the President of the Russian Federation for young scientists № MK-2070.2022.5

References

1. H.P. Olmo, *Origin and Distribution of Grapes*, In: Simmonds, N.W., Ed., *Evolution of Crop Plants*, (Longman and New York, 1976) 294-298.
2. P. E. MCGOVERN, *The search for the origins of Viniculture*, In: Princeton Univ. Press (Eds.). *Ancient Wine*, 96-122 (2003).
3. R. Arroyo-García, L. Ruiz-García, L. Bolling, R. Ocete, M. A. López, C. Arnold, J. M. Martínez-Zapater, *Molecular ecology*, **15**(12), 3707-3714 (2006) DOI: 10.1111/j.1365-294X.2006.03049.x
4. R. A. A. García, E. Revilla, The current status of wild grapevine populations (*Vitis vinifera* ssp *sylvestris*) in the Mediterranean basin, *Agricultural and Biological Sciences "The Mediterranean Genetic-Code-Grapevine and Olive*, 51-72 (2013) DOI: 10.5772/52933
5. F. Grassi, M. Labra, S. Imazio, R. O. Rubio, O. Failla, A. Scienza, F. Sala, *Conservation Genetics*, **7**(6), 837-845 (2006) DOI: 10.1007/s10592-006-9118-9
6. D. Zohary, M. Hopf, *Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley* (No. Ed. 3) Oxford university press (2000).
7. P. McGovern, M. Jalabadze, S. Batiuk, M. P. Callahan, K. E. Smith, G. R. Hall, D. Lordkipanidze, *Proceedings of the National Academy of Sciences*, **114**(48), E10309-E10318, (2017) DOI: 10.1073/pnas.1714728114
8. D. Q. Fuller, C. J. Stevens, *Vegetation history and archaeobotany*, **28**, 263-282 (2019) DOI: 10.1007/s00334-019-00727-4
9. G. Forni, *Vitis*, **51**, 27-38 (2012)
10. K. M. Sefc, H. Steinkellner, F. Lefort, R. Botta, A. da Câmara Machado, J. Borrego, J. Glössl, *American Journal of Enology and Viticulture*, **54**(1), 15-21 (2003) DOI: 10.5344/ajev.2003.54.1.15
11. J. Ekhvaia, M. Akhalkatsi, *Flora-Morphology, Distribution, Functional Ecology of Plants*, **205**(9), 608-617 (2010) DOI: 10.1016/j.flora.2009.08.002
12. L.P. Troshin, *Polythematic network electronic scientific journal of the Kuban State Agrarian University*, **72**, 312-330 (2011) URL: <https://cyberleninka.ru/article/n/morfometriya-listiev-kubanskih-dikorastuschih-lian-vinograda-1>
13. S. Triboush, N. Danilenko, O. A Davydenko, *Plant Molecular Biology Reporter*, **16**, 183 (1998).
14. S. Andrews, *FastQC: a quality control tool for high throughput sequence data* (2010)

15. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, **30** (15), 2114-2120, (2014). DOI: 10.1093/bioinformatics/btu170
16. K. Okonechnikov, O. Golosova, M. Fursov, Ugene Team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28** (8), 1166-1167 (2012) DOI: 10.1093/bioinformatics/bts091
17. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, P. A. Pevzner, *J. Comp. Biol.*, **19** (5), 455-477 (2012) DOI: 10.1089/cmb.2012.0021
18. H. Li, R. Durbin, *Bioinformatics*, **25**(14), 1754-1760 (2009) DOI: 10.1093/bioinformatics/btp324
19. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T. L. Madden, *Nucleic acids research*, **36**(2), W5-W9 (2008) DOI: 10.1093/nar/gkn201
20. A. V. Lukashin, M. Borodovsky, *Nucleic acids research*, **26**(4), 1107-1115 (1998) DOI: 10.1093/nar/26.4.1107
21. M. Tillich, P. Lehwark, T. Pellizzer, E. S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, *Nucleic acids research*, **45**(W1), W6-W11 (2017)
22. S. Greiner, P. Lehwark, R. Bock, *Nucleic acids research*, **47**(W1), W59-W64, (2019) DOI: 10.1093/nar/gkz238
23. K.S. Katz, O. Shutov, R. Lapoint, M. Kimelman, J.R. Brister, C. O'Sullivan, *Genome Biology*, **22** (1), 1-15 (2021) DOI: 10.1186/s13059-021-02490-0