Chloroplast genomes of *Vitis sylvestris* Gmel. samples from Damanskaya population of the Krasnodar region

Alexander Milovanov[1,*], *Daria* Savenkova[1], *Anastasia* Elisyutikova[1], *Vladimir* Khachumov[2], and *Leonid* Troshin[3]

[1]Laboratory of Plant and Animal genetics, Kuban State Agrarian University named after I.T. Trubilin, 350044 Kalinina 12, Krasnodar, Russian Federation
[2]Academy of biology and biotechnology, Southern Federal University, 344090 Rostov-on-Don, Stachki Av., 194/1, Russian Federation
[3]Department of Viticulture, Kuban State Agrarian University named after I.T. Trubilin, 350044 Kalinina 12, Krasnodar, Russian Federation

Abstract. This article presents the structures of chloroplast genomes of three *Vitis sylvestris* Gmel. samples from Damanskaya population of the Krasnodar Territory. An expedition to the place where wild forest grapevine grows was made. After that, the selected leaves served as a source of chloroplasts. DNA was isolated from them and DNA libraries were prepared and sequenced. Genome assembly was carried out after selection of the most suitable reference sample at coverage of 35.0x. The genomes ranged in size from 159,900 to 160,887. Aligned chloroplast genomes were annotated with GeSeq and GeneMark.hmm. OGDRAW was used to visualize the structure of the genomes. GenBank search allowed to determine their belonging to *V. sylvestris* species. At the same time, a comparison of the genomes with each other showed the presence of minor differences in their structure.

1 Introduction

Cultivated grapevine (*Vitis vinifera* L.) is one of the most widespread agricultural crops in the world. The domestication of grapevine began in the Middle East and later spread throughout the earth [1, 2]. At present, this species is represented by *Vitis vinifera* subsp. *vinifera* and *Vitis vinifera* subsp. *sylvestris*. *V. vinifera* subsp. *sylvestris*, which is distributed throughout the Mediterranean region, is considered to be the ancestor of cultivated grapevine [3]. Both subspecies have variations in their morphological structure [4, 5], and are also distributed in different climatic conditions.

In order to compare and establish the relationship between wild and cultivated grapevine, studies using various kinds of phenotypic [6] and genetic [7] markers have already been carried out. However, the results obtained using nuclear markers can be misleading due to an active exchange of genetic material between wild samples growing in

* Corresponding author: milovanov1991@mail.ru

the forest and cultivated ones [8]. In particular, there are barely any untouched forests left in Russia. This is especially true for such agronomically important regions as the Krasnodar Territory. At the same time, there are wild forest grapevine centers of growth recognized in this area [9]. The identified forms were characterized as *Vitis vinifera* subsp. *sylvestris* representatives [10]. At the same time, it is important to note that the geographical dispersion of these centers of growth (in the Krasnodar Territory and the Republic of Adygea) made it possible to preserve their phenotypic diversity. Also, a study of genotypic diversity within and between populations was conducted using microsatellite markers [11]. The results indicate that there may be admixtures within populations and support the claim of pollination with locally cultivated grapevine varieties.

Studying the chloroplast genomes of wild and cultivated grapevine may provide a better understanding of its origins [12]. In general, study and comparison of microsatellite sequences of chloroplast genomes with each other gives an understanding of belonging to one or another chlorotype, inherited through the female line [13]. However, examination of only individual microsatellite sequences provides a limited view in contrast to the sequencing of the entire chloroplast genome [14]. Therefore, in our study we performed chloroplast genome sequencing of three representatives from Damanskaya population of the Krasnodar Territory.

2 Materials and methods

The coordinates of the place of growth published earlier were used to collect plant material [10]. After that, the leaves in coolers were transported to the place where chloroplasts were isolated. Chloroplasts isolation was carried out in accordance with the published method [15]. DNA isolation from chloroplasts was carried out using DNeasy Blood & Tissue Kit (QIAGEN). The quantity and quality of the isolated DNA were determined using the Qubit 4 Fluorometer (Thermo Fisher Scientific).

Preparation of libraries for sequencing was carried out using the Nextera DNA Flex Library Prep Kit (Illumina). Sequencing was performed on the MiSeq (Illumina). FastQC was used for quality control [16]. Adapter sequences were removed using Trimmomatic v.0.39 [17]. Preliminary assembly of genomes was carried out in UGENE [18] with the built-in SPAdes v.3.15.3 algorithm [19]. Then reads were alligned to different genomes to select the most efficient one using the BWA program [20]. Therefore, we chose the chloroplast genome of *V. vinifera* PN40024 (GenBank number NC_007957.1) as a reference. To compare the results among themselves and against already available data the NCBI BLAST algorithm was used [21]. The search for possible CDS was done with the GeneMark.hmm [22]. Genomes were annotated using GeSeq [23] and visualized by OGDRAW [24].

3 Results and Discussion

As a result of the work, we obtained chloroplast genomes of three representatives of wild grapevine from Damanskaya population of the Krasnodar Territory. The sequencing results are presented in Table 1. The results of genome annotation are displayed graphically in Figures 1-3.

Table 1. Results of chloroplast genomes assembly and annotation.

Sample	Total length	GC content	Coverage	CDS number	Nucleotide BLAST result	GenBank number
1.1 Damanka k.1.1	160887 b.p.	37%	35.0x	86	*Vitis vinifera* subsp. *sylvestris* INRA	LC501387.1
2.1 Damanka k.1.2	159900 b.p.	37%	35.0x	97	*Vitis vinifera* subsp. *sylvestris* INRA	LC501387.1
3.1 Damanka k.1.1	160873 b.p.	38%	35.0x	87	*Vitis aestivalis* cultivar Norton	KT997470.1

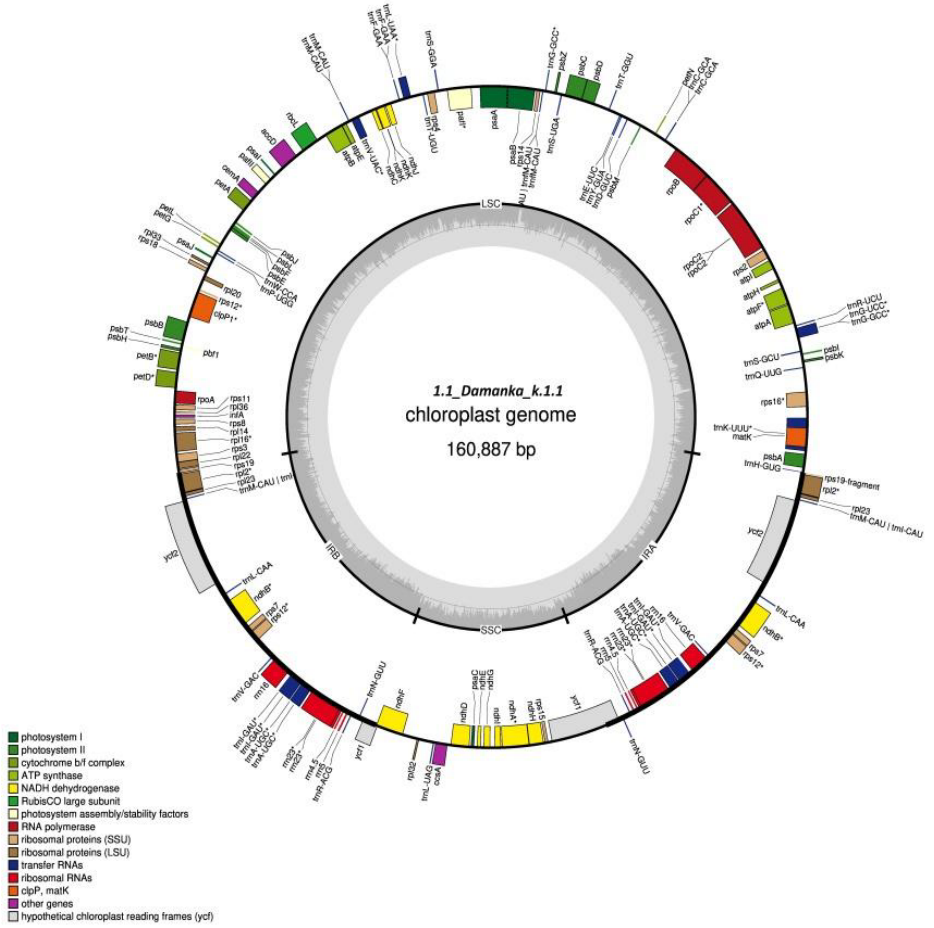

Fig. 1. Visualization of 1.1 Damanka k.1.1 genome.

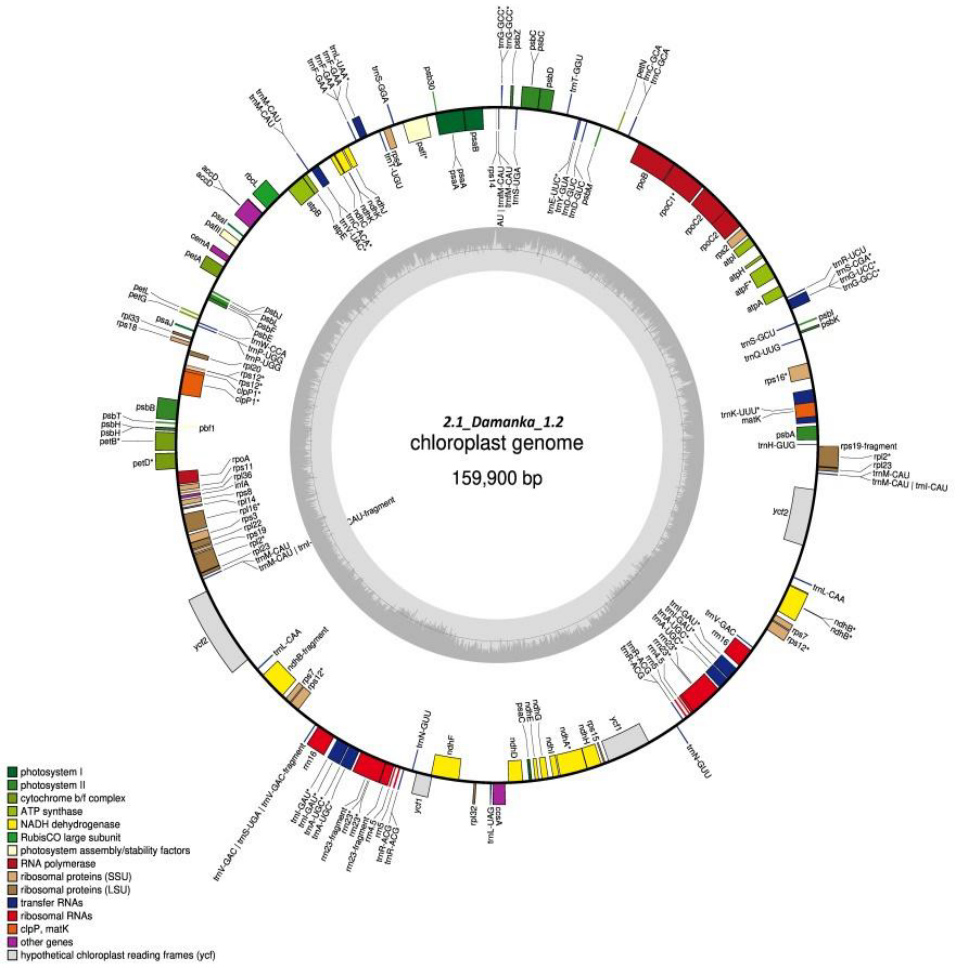

Fig. 2. Visualization of 1.1 Damanka k.1.2 genome.

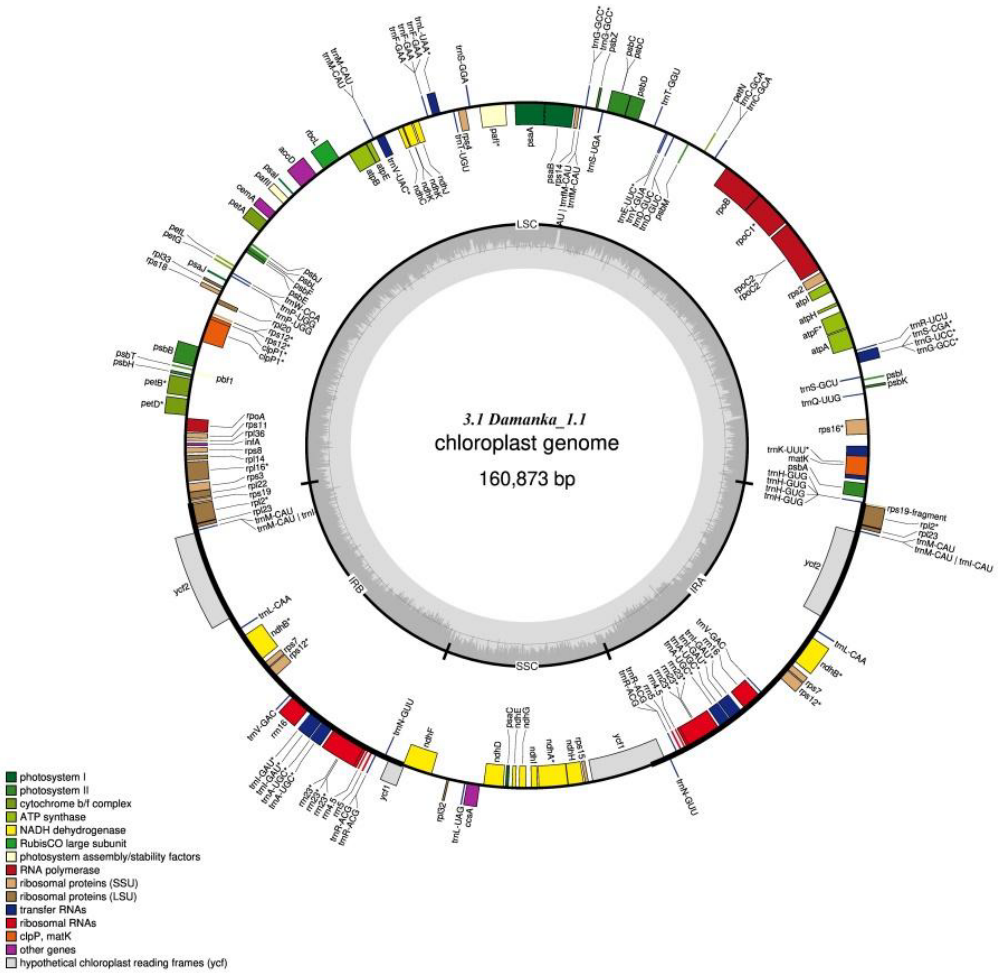

Fig. 3. Visualization of 3.1 Damanka k.1.1 genome.

Table 1 shows that the size of the obtained genomes varied within 160 kb, which indicates a high coverage of the genome as a whole. At the same time, coverage in terms of the number of reads per region for each of the samples was 35.0x. A search in GenBank showed that two out of three representatives of the population belong to the *V. sylvestris*. At the same time, the sample 3.1 Damanka k.1.1 was identified as a representative of a different species. This is interesting in view of the fact that phenotypically all samples of the Damanskaya population are similar [25]. Also, indirectly, this can be confirmed by the fact that *Vitis vinifera* subsp. *sylvestris* INRA presented below in the list of nBLAST results with a small difference of 0.01%. In order to test this assumption, we compared the 1.1 Damanka k.1.1 and 3.1 Damanka k.1.1 genomes with each other (Figure 4).

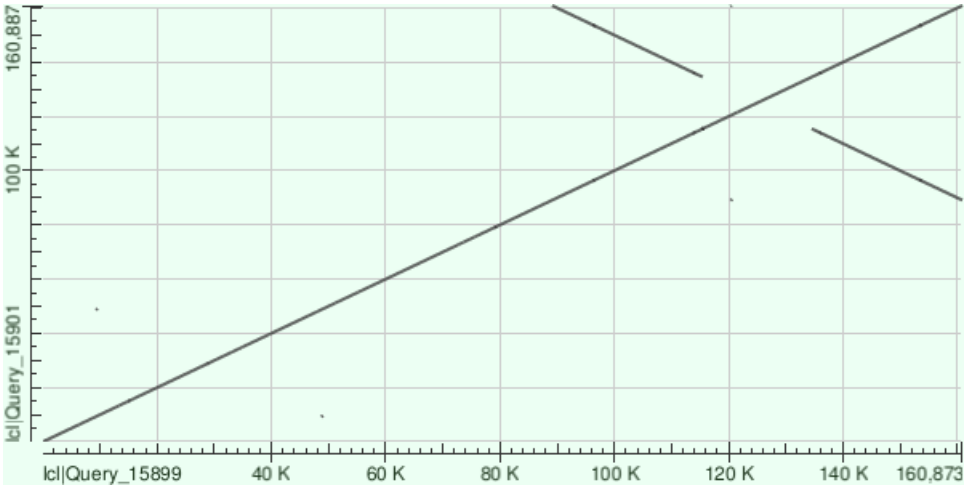

Fig. 4. Comparison of 1.1 Damanka k.1.1 and 3.1 Damanka k.1.1 genomes.

As we can see from Figure 4, in general the structure of chloroplast genomes is similar, which is also indicated by the results displayed in percentages: 99% coverage and 99.86% identity. It was unexpected that the comparison between the first and second samples gave lower numbers: 99% coverage and 99.17% identity. However, this can be explained by the fact that the assembly size of the chloroplast genome of sample 1.1 Damanka k.1.2 is smaller. On the other hand, this can also be explained by some polymorphism of the genomes, for example, when comparing LC510289.1 and LC494572.1, the coverage is 99%, while the identity is 99.94%. Therefore, given the previous phenotyping, all three samples can be identified as *Vitis vinifera* subsp. *sylvestris* Gmel.

4 Conclusion

As a result of the work, chloroplast genomes of wild forest grapevine from Damanskaya population were sequenced for the first time. With coverage of 35.0x, the studied genomes were assembled based on the alignment of reads to the reference DNA sequence. It is interesting to note that we performed alignment for various deposited genomes, including *V. sylvestris*, as a result, alignment for the chloroplast genome of *V. vinifera* showed the greatest efficiency. CDSs were determined using GeneMark.hmm. Visualization of each genome was made, displaying both similarity and some differences in structure. The determination of the systematic position of the three genotypes by comparison with GenBank showed that all chloroplast genotypes were inherited from the mother plant, which is *V. sylvestris*. Together with the previous results of phenotyping, the obtained data give grounds to attribute the studied genotypes to *V. sylvestris*.

Acknowledgements

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

References

1. Y. Zhou, A. Minio, M. Massonnet, E. Solares, Y. Lv, T. Beridze, B. S. Gaut, Nature plants, **5**(9), 965-979 (2019) DOI: https://doi.org/10.1038/s41477-019-0507-8
2. F. Grassi, G. De Lorenzis, International Journal of Molecular Sciences, **22**(9), 4518 (2021) DOI: 10.1007/s00122-003-1321-1
3. X. Guan, S. Essakhi, H. Laloue, P. Nick, C. Bertsch, J. Chong, Plant Pathology, **65**(2), 273-284 (2016) DOI: 10.1111/ppa.12405
4. J. Ekhvaia, M. Akhalkatsi, Functional Ecology of Plants, **205**(9), 608-617 (2010) DOI: 10.1016/j.flora.2009.08.002
5. A. Naqinezhad, E. Ramezani, M. Djamali, A. Schnitzler, C. Arnold, Journal of Forestry Research, **29**(6), 1757-1768, (2018). DOI: 10.1007/s11676-017-0549-6
6. S. Barth, A. Forneck, F. Verzeletti, R. Blaich, F. Schumann, Genetic resources and crop evolution, **56**(8), 1171-1181 (2009) DOI: 10.1007/s10722-009-9443-1
7. M. T. De Andrés, A. Benito, G. Pérez-Rivera, R. Ocete, M. A. Lopez, L. Gaforio, R. ARROYO-GARCÍA, Molecular Ecology, **21** (4), 800-816, (2012) DOI:10.1111/j.1365-294X.2011.05395.x
8. M. Di Vecchi-Staraz, V. Laucou, G. Bruno, T. Lacombe S. Gerber, T. Bourse, P. This, Journal of Heredity, **100**(1), 66-75 (2009) DOI: 10.1093/jhered/esn084
9. L.P. Troshin, Polythematic network electronic scientific journal of the Kuban State Agrarian University, **81**, 561-580 (2012) URL: https://cyberleninka.ru/article/n/ampelograficheskaya-i-selektsionnaya-nauchno-issledovatelskaya-rabota-kubanskogo-gosagrouniversiteta
10. 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
11. A. S. Zvyagin, L. P. Troshin, Polythematic network electronic scientific journal of the Kuban State Agrarian University, **58**, 376-387 (2010) URL: https://cyberleninka.ru/article/n/issledovanie-dikogo-vinograda-vitis-silvestris-gmel-na-severnom-kavkaze
12. M. Rogalski, L. do Nascimento Vieira, H. P. Fraga, M. P. Guerra, Frontiers in plant science, **6**, 586 (2015) DOI: 10.3389/fpls.2015.00586
13. R. Arroyo-Garcна, L. Ruiz-Garcia, L. Bolling, R. Ocete, M. A. López, C. Arnold, J. M. Martinez-Zapater, Molecular ecology, **15**(12), 3707-3714 (2006) DOI: 10.1111/j.1365-294X.2006.03049.x
14. S. A. Olejniczak, E. Łojewska, T. Kowalczyk, T. Sakowicz, Planta, **244**(3), 517-527 (2016) DOI: 10.1007/s00425-016-2551-1
15. S. Triboush, N. Danilenko, O. A Davydenko, Plant Molecular Biology Reporter **16**, 183 (1998)
16. S. Andrews, FastQC: a quality control tool for high throughput sequence data (2010)
17. A. M. Bolger, M. Lohse, B. Usadel, Bioinformatics, **30** (15), 2114-2120 (2014) DOI: 10.1093/bioinformatics/btu170
18. K. Okonechnikov, O. Golosova, M. Fursov, Bioinformatics. **28** (8), 1166-1167 (2012) DOI: 10.1093/bioinformatics/bts091
19. 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

20. H. Li, R. Durbin, Bioinformatics, **25**(14), 1754-1760 (2009) DOI:10.1093/bioinformatics/btp324
21. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T. L. Madden, Nucleic acids research, **36**(2), W5-W9 (2008) DOI: 10.1093/nar/gkn201
22. A. V. Lukashin, M. Borodovsky, Nucleic acids research, **26**(4), 1107-1115 (1998) DOI: 10.1093/nar/26.4.1107
23. M. Tillich, P. Lehwark, T. Pellizzer, E. S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, Nucleic acids research, **45**(W1), W6-W11 (2017)
24. S. Greiner, P. Lehwark, R. Bock, Nucleic acids research, **47**(W1), W59-W64 (2019) DOI: 10.1093/nar/gkz238
25. L.P. Troshin, Polythematic network electronic scientific journal of the Kuban State Agrarian University, **70**, 506-536 (2011) URL: https://cyberleninka.ru/article/n/morfometricheskiy-analiz-listovoy-ampelograficheskoy-informatsii