



# The first draft genome sequence of Russian olive (*Elaeagnus angustifolia* L.) in Iran

Leila Zirak, Reza Khakvar\* and Nadia Azizpour

Department of Plant Protection, Faculty of Agriculture, University of Tabriz, 5166616471, Tabriz, Iran

**Abstract:** Russian olive (*Elaeagnus angustifolia* L.) is a native tree species of Iran and the Caucasus region growing in both wild habitats and cultivated settings. The area under cultivation of this tree has been increasing in recent years due to its ability to withstand drought and soil salinity. Revealing the complete genome of this tree holds great importance. To achieve this, a local cultivar of Russian olive was sampled from the northwest region of Iran for whole genome sequencing using the Illumina platform resulting in approximately 6GB of raw data. A quality check of the raw data indicated that approximately 45,011,388 read pairs were obtained from sequences totalling around  $6.7 \times 10^9$  bp with CG content of 31%. To assemble the genome of the Russian olive tree, the raw data was aligned to a reference sequence of the jujube (*Ziziphus jujuba*) genome, which is the taxonomically closest plant to the Russian olive. Assembly of alignments yielded a genome size of 553,696,299 bp consisting of 339,701 contigs. The N50 value was 5,300 with an L50 value of 24,921 and GC content of the Russian olive genome was 31.5%. This research represents the first report on the genome of the Iranian cultivar of the Russian olive tree.

**Keywords:** Russian olive (*Elaeagnus angustifolia* L.), Genome, WGS, Iran

**Citation:** Zirak, L., Khakvar, R., Azizpour, N. (2024). The first draft genome sequence of Russian olive (*Elaeagnus angustifolia* L.) in Iran. *Genetic Resources* 5 (9), 29–35. doi: [10.46265/genresj.WAOT8693](https://doi.org/10.46265/genresj.WAOT8693).

© Copyright 2024 the Authors.

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Introduction

The Russian olive (*Elaeagnus angustifolia* L.) is a deciduous tree growing abundantly in several areas in Iran (Mozaffarian, 2009). It belongs to the Elaeagnaceae family and is native to western and central Asia including Iran, southern Russia, Turkey, Kazakhstan (Lamers and Khamzina, 2010) and China (Huang *et al.*, 2010). Recently, it has also been cultivated in North America (Mineau *et al.*, 2012). The Russian olive is a drought-resistant species and plays an important ecological role in Iran's dry climate. It is widely cultivated in Iran but also grows wild. Since 97% of lands in Iran are arid or semi-arid, many artificial afforestation and urban green space projects have been devised, especially in the drainage basin of Lake Urmia, which is at risk of drying out completely, with Russian olive being a key species in these efforts (Tabatabaei, 2010). About 10,000 ha of arid or semi-arid lands are cultivated with Russian olive tree

in East Azerbaijan province in the northwest, the most important Russian olive cultivation area in Iran (Mozaffarian, 2009).

The whole genome of a local cultivar of Russian olive (Tabriz cultivar) was sequenced using the next-generation sequencing (NGS) method. Prior to this research, there was no available information on the genome of this species. Due to the importance of the Russian olive in the construction of artificial forests and urban green spaces in the northwest of Iran, the information obtained from sequencing the genome could help characterize the Iranian cultivar. The full genome sequence obtained in this research is the first report for the Russian olive tree genome worldwide.

## Materials and methods

### Plant sampling and DNA extraction

For sampling, a tree was randomly selected as a representative sample of this cultivar (*E. angustifolia* cultivar Tabriz) (Figure 1) from the Eynali artificial afforestation area located in Tabriz city in the northwest

\*Corresponding author: Reza Khakvar  
([khakvar@tabrizu.ac.ir](mailto:khakvar@tabrizu.ac.ir); [khakvar@gmail.com](mailto:khakvar@gmail.com))

of Iran. About 5g of leaf tissues were separated, crushed using liquid nitrogen and then used for DNA extraction (Murray and Thompson, 1980). Extracted DNA was dissolved in 100ml distilled sterile water and stored at  $-20^{\circ}\text{C}$ .

### Next-generation sequencing analyses

About  $200\mu\text{l}$  of the DNA solution, with a total amount of  $10\mu\text{g}$  DNA, extracted from Russian olive leaf samples was purified and used for library preparation. The concentration and purity of DNA was determined using a Qubit fluorometer. The concentration of purified DNA was  $43.20\text{ng}/\mu\text{l}$  and evaluated appropriately for the whole genome sequencing process. The Illumina 1.9 Novaseq 6000 platform was used to generate paired-end libraries by Novogene (Beijing, China). Finally, about 6GB of raw data was obtained. In total, 45,011,388 read pairs in about  $6.7 \times 10^9$ bp sequences with GC content of 31% were obtained from Russian olive genome sequencing. Each raw read length was 150bp and the insert size was 350bp.

### Genome qualification, reference genome preparation and sequence alignment

The quality of Illumina raw data was checked by FastQC software version 0.73 (Brown et al, 2017). For reference genome preparation, the common jujube (*Ziziphus jujuba* (2n=24)) genome, with 405,637Mbp size and GC content of 33.084% comprising of 12 full-length chromosomes submitted in the NCBI genome database (accession numbers NC\_063287, NC\_063288, NC\_063289, NC\_063290, NC\_063291, NC\_063292, NC\_063293, NC\_063294, NC\_063295, NC\_063296, NC\_063297 and NC\_063298), a 365,812bp mitochondrion sequence (CM036902) and a 161,185bp chloroplast sequence (CM036903), was used Yang et al (2023). The NGS raw data was aligned to the reference genome using Bowtie2 software version 2.5.0 (Langmead and Salzberg, 2012).

### Genome assembly

For the genome preparation, all aligned reads which resulted from alignment analysis, were used for the assembly by metaSPAdes software version 3.15.4 (Nurk et al, 2017). For nucleotide sequence clustering and to improve the performance of sequence analyses, all contigs were clustered using CD-HIT-EST software version 4.8.1 (Fu et al, 2012).

### Genome annotation

To characterize proteins related to the Russian olive genome, contigs were annotated using InterProScan functional annotation software version 5.59-91.0 (Jones et al, 2014). The annotation results using InterProScan are summarized in Supplemental Table 1. In addition, for genes and proteins sequence prediction, all contigs were subjected to another annotation method using GhostKOALA tool of the KEGG (Kyoto Encyclope-

dia of Genes and Genomes) database (<https://www.kegg.jp/ghostkoala/>).

## Results

### Russian olive genome information

Since no Russian olive genome has been submitted to the NCBI genome database so far, the NGS raw data were aligned to the reference genome prepared with the common jujube (*Z. jujuba*) genome. The jujube tree is a species taxonomically close to the Russian olive and its genome is available in the NCBI genome database. The mapping rate was 96.4%. Assembly of aligned reads resulted in a genome in contig level with 553,696,299bp size consisting of 339,701 contigs with  $N50 = 5,300\text{bp}$ ,  $L50 = 24,921\text{bp}$  and GC content of 31.5%. The genome coverage was 442.0x. Finally, the *E. angustifolia* cultivar Tabriz genome was deposited in the NCBI genome database under the whole genome accession number JAIFOS000000000, BioProject accession number PRJNA744085, BioSample accession number SAMN20079343 and Assembly accession number GCA 019593565.

### Genome annotation results using InterProScan

For genome annotation, two methods were used. Initially, functional analysis of proteins and nucleotides was conducted using InterProScan software, which classifies them into families and predicts domains and important sites. To classify proteins, InterProScan uses predictive models, known as signatures, provided by several different databases. According to InterProScan results, a total of 496,838 proteins were predicted in the Russian olive genome. Among all proteins, 106,757 proteins shared consensus disorder prediction. The intrinsic disorder (ID) is recognized as an important feature of protein sequences. The consensus-based prediction of disorder in protein was done using the MobiDB-lite method which has been integrated with the InterPro database (Necci et al, 2017). About 1,454 proteins remained uncharacterized (Supplemental Table 1).

### Genome annotation results using GhostKOALA

The assembled genome was subjected to annotation using the GhostKOALA server to characterize individual gene functions. The protein sequences that were used for GhostKOALA analysis were provided using the MetaGenMark online web tool (Zhu et al, 2010). The KEGG GENES database (Kanehisa et al, 2016) searches indicated that 54,162 proteins (about 17% of whole proteins) acquired original KO numbers, 178,897 proteins acquired second-best KO numbers and 85,713 proteins could not be matched with any characterized proteins and were therefore considered as uncharacterized proteins. The GhostKOALA annotation results are summarized in Table 1.



Figure 1. The Russian olive cultivar Tabriz subject to NGS analyses in this study.

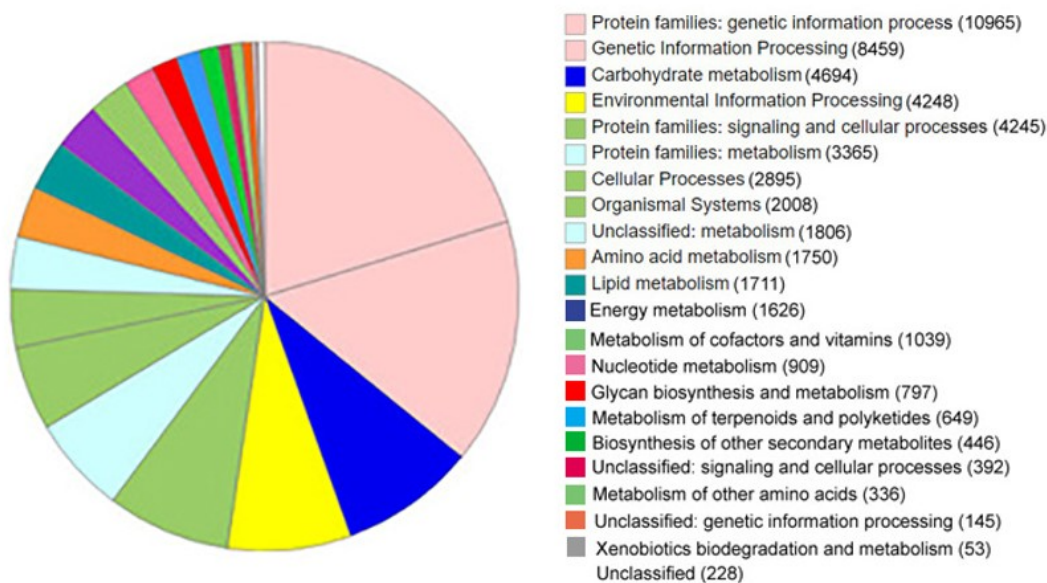


Figure 2. Functional category and pathways for predicted proteins of Russian olive tree prepared with GhostKOALA.

**Table 1.** Functional classification predictions of proteins annotated in the Russian olive genome based on KEGG BRITE classification.

Pathway	Protein description	No. of predicted proteins
Genes and proteins	Ribosomal Proteins	153
	RNA polymerases	34
	DNA polymerases	25
	Aminoacyl-tRNA synthetases	28
	Enzymes of 2-oxocarboxylic acid metabolism	33
	Dioxygenases	2
	Photosynthetic and chemosynthetic capacities	4
Orthologs, modules and networks	KEGG Orthology (KO)	8,957
Protein families: metabolism	Enzymes	3,712
	Protein kinases	315
	Protein phosphatases and associated proteins	204
	Peptidases and inhibitors	379
	Glycosyltransferases	148
	Lipopolysaccharide biosynthesis proteins	25
	Peptidoglycan biosynthesis and degradation proteins	29
	Lipid biosynthesis proteins	67
	Polyketide biosynthesis proteins	7
	Prenyltransferases	28
	Amino acid-related enzymes	64
	Cytochrome P450	67
	Photosynthesis proteins	56
Protein families: genetic information processing	Transcription factors	410
	Transcription machinery	229
	Messenger RNA biogenesis	308
	Spliceosome	256
	Ribosome	154
	Ribosome biogenesis	259
	Transfer RNA biogenesis	187
	Translation factors	84
	Chaperones and folding catalysts	165
	Membrane trafficking	787
	Ubiquitin system	441
	Proteasome	52
	DNA replication proteins	132
	Chromosome and associated proteins	678
DNA repair and recombination proteins	309	
Mitochondrial biogenesis	269	

Continued on next page

Table 1 continued

Pathway	Protein description	No. of predicted proteins
Protein families: signaling and cellular processes	Transporters	595
	Secretion system	79
	Two-component system	40
	Cilium and associated proteins	181
	Cytoskeleton proteins	222
	Exosome	333
	G protein-coupled receptors	130
	Cytokine receptors	9
	Pattern recognition receptors	6
	Nuclear receptors	14
	Ion channels	125
	GTP-binding proteins	68
	Cytokines and growth factors	21
	Cell adhesion molecules	50
	CD molecules	55
	Proteoglycans	16
	Glycosaminoglycan binding proteins	46
	Glycosylphosphatidylinositol (GPI)-anchored proteins	21
	Lectins	23
	Domain-containing proteins not elsewhere classified	241
Other proteins	65	



An overview of putative functions of annotated proteins is given in [Figure 2](#).

## Discussion

### Ecological importance of the Russian olive tree in Iran

The Russian olive is a long-lived tree that can live up to 100 years and tolerate a wide range of hard environmental conditions such as severe drought, flood and high salinity or alkalinity of the soils ([Asadiar et al, 2013](#)). This tree produces edible fruits with high medicinal properties. Russian olive fruits have antioxidant activities and anti-inflammatory properties. Fruit kernel powder is used in the treatment of acute and chronic inflammations, such as arthritis ([Tabatabaei, 2010](#); [Wang et al, 2013](#)). The climate of Iran is mostly arid or semi-arid and is strongly affected by depleting water resources, as a result of rising demand, salinization, ground water overexploitation and increasing drought frequency. Therefore, plants that could withstand harsh environmental conditions and have low water consumption have been considered for cultivation in several regions. The Russian olive is growing as a wild plant in all areas with a dry climate in Iran; however, it also serves as the main species in many artificial forestation projects. The climatic and ecological benefits provided by the Russian olive in Iran underline the importance of exploring the genomic characteristics of its Iranian cultivar.

### Genomic characteristics of the Russian olive cultivar Tabriz

Before this study, no information about the Russian olive genome was available. Therefore, the common jujube genome was used for the Russian olive genome preparation, since it is the closest taxonomical relative to the Russian olive and its genome is available in NCBI. The common jujube belongs to the Rhamnaceae family, which along with the Elaeagnaceae family belongs to the Rosales order. The genome of *Z. jujuba* comprises 12 chromosomes with an average size of 405.637Mb and GC content of about 33%. Alignments of NGS reads obtained from Russian olive to the jujube whole genome sequence resulted in a 553,696,299bp genome composed of 339,701 contigs. The GC content of the new genome was 31.5% which was nearly the same as the jujube genome GC content.

Annotation analysis was accomplished by several programmes. At last, two methods based on the online GhostKOALA web server and InterProScan software were found suitable for Russian olive genome annotation. The genome functional annotation using online KEGG mapper reconstruction resulted in 3,186 proteins for metabolism pathways in the genome including 647 involved in carbohydrate metabolism pathways, 345 in energy metabolism pathways, 381 proteins for lipid metabolism pathways, 185 proteins for nucleotide metabolism pathways, 686 proteins for

amino acid metabolism pathways, 258 proteins for glycan biosynthesis and metabolism pathways, 296 proteins for metabolism of cofactors and vitamins pathways, 122 proteins for metabolism of terpenoids and polyketides pathways, 144 proteins for biosyntheses of other secondary metabolites pathways and 122 proteins for xenobiotics biodegradation and metabolism pathways.

Also, for the transcription and translation systems, 35 RNA polymerases, 34 basal transcription factors, 105 spliceosomes, 124 ribosomes, 33 proteins for aminoacyl-tRNA biosynthesis pathways, 83 nucleocytoplasmic transport proteins, 58 mRNAs surveillance and 62 ribosomes of biogenesis in eukaryotes were obtained.

The folding, sorting and degradation systems included 31 export proteins, 38 proteasomes, 70 RNA degradation proteins and 237 other proteins. The replication and repair systems include 43 DNA replication proteins, 34 base excision repair proteins, 42 nucleotide excision repair proteins, 54 homologous recombination proteins and 11 non-homologous end-joining proteins. The membrane transport includes 114 ABC transporters and 9 proteins for phosphotransferase system PTS. Also, 1600 signal transduction proteins, 497 proteins for cell growth and death pathways, 272 cellular community proteins and 8,494 other proteins exist in the genome.

## Conclusion

The Russian olive is an ecologically important tree serving as vegetation in Iran's arid climate. It is also known as an important medicinal plant in Iranian traditional medicine. However, genetic information about this species remains sparse. In this research, we describe the genome of an Iranian cultivar of the Russian olive by using the jujube genome as a reference, since it is the closest species with a characterized genome. As a result, a full-size genome with 553.7Mb size in contig level was obtained, which can provide the foundations for the chromosomal sequence of this species. Russian olive is one of the most important horticultural tree species in the northwest of Iran and its genome characterization serves as a key step towards broader research to characterize the genome of other important plant species of Iran.

## Supplemental data

[Supplemental Table 1](#). Uncharacterized annotated proteins of Russian olive

## Author contribution

Leila Zirak executed both the laboratory experiments and the subsequent bioinformatics analyses; Reza Khakvar, who supervised the project, was responsible for the validation of all experimental data and contributed to the editing of the final manuscript draft; Nadia Azizpour provided assistance in the collection of samples

and the extraction of DNA, and in the composition of the initial manuscript draft.

### Conflict of interest statement

The authors declare no conflicts of interest.

### References

- Asadiar, L. S., Rahmani, F., and Siami, A. (2013). Assessment of genetic diversity in the Russian olive (*Elaeagnus angustifolia*) based on ISSR genetic markers. *Revista Ciencia Agronomica* 44(2), 310–316. doi: <https://doi.org/10.1590/S1806-66902013000200013>
- Brown, J., Pirrung, M., and Mccue, L. A. (2017). FQC dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33(19), 3137–3139. doi: <https://doi.org/10.1093/bioinformatics/btx373>
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23), 3150–3152. doi: <https://doi.org/10.1093/bioinformatics/bts565>
- Huang, Z., Liu, M., Chen, B., Uriankhai, T., Xu, C., and Zhang, M. (2010). Distribution and interspecific correlation of root biomass density in an arid *Elaeagnus angustifolia* - *Achnatherum splendens* community. *Acta Ecologica Sinica* 30(1), 45–49. doi: <https://doi.org/10.1016/j.chnaes.2009.12.008>
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., Mcanulla, C., Mcwilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9), 1236–1240. doi: <https://doi.org/10.1093/bioinformatics/btu031>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic acids research* 44, 457–462. doi: <https://doi.org/10.1093/nar/gkv1070>
- Lamers, J. P. and Khamzina, A. (2010). Seasonal quality profile and production of foliage from trees grown on on degraded cropland in arid Uzbekistan, Central Asia. *Journal of Animal Physiology and Animal Nutrition* 94, 77–85. doi: <https://doi.org/10.1111/j.1439-0396.2009.00983.x>
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie. 2. *Nature Methods* 9, 357–359. doi: <https://doi.org/10.1038/nmeth.1923>
- Mineau, M. M., Baxter, G. V., Marcarelli, A. M., and Minshall, G. W. (2012). An invasive riparian tree reduces stream ecosystem efficiency via a recalcitrant organic matter subsidy. *Ecology* 93(7), 1501–1508. doi: <https://doi.org/10.1890/11-1700.1>
- Mozaffarian, V. (2009). Trees and shrubs of Iran (Iran Farhang Pree), 2nd edition, 1054p.
- Murray, M. G. and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8(19), 4321–4326. doi: <https://doi.org/10.1093/nar/8.19.4321>
- Necci, M., Piovesan, D., Dosztányi, Z., and Tosatto, S. C. E. (2017). MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* 33(9), 1402–1404. doi: <https://doi.org/10.1093/bioinformatics/btx015>
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research* 27, 824–834. doi: <https://doi.org/10.1101/gr.213959.116>
- Tabatabaei, M. (2010). Senjed = *Elaeagnus angustifolia* L. (Elaeagnaceae) (Tehran, Iran: SANA Publication), 92p.
- Wang, Y., Guo, T., Li, J. Y., Zhou, S. Z., Zhao, P., and Fan, M. T. (2013). Four flavonoid glycosides from the pulps of *Elaeagnus angustifolia* and their antioxidant activities. *Advanced Materials Research* 756, 16–20. doi: <https://doi.org/10.2991/iccia.2012.415>
- Yang, M., Han, L., Zhang, S., Dai, L., Li, B., Han, S., Zhao, J., Liu, P., Zhao, Z., and Liu, M. (2023). Insights into the evolution and spatial chromosome architecture of jujube from an updated gapless genome assembly. *Plant Community* 4(6), 100662. doi: <https://doi.org/10.1016/j.xplc.2023.100662>
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research* 38(12), e132. doi: <https://doi.org/10.1093/nar/gkq275>