

## Characterization of chloroplast genome and phylogenetic analysis of the *Elaeagnus* hybrid variety *E. 'viridis'* (Elaeagnaceae)

Naiwei Li<sup>a</sup>, Hong Chen<sup>a,b</sup>, Yanwei Zhou<sup>a,b</sup>, Xinran Chong<sup>a,b</sup>, Ting Zhou<sup>a,b</sup> and Fan Zhang<sup>a,b</sup>

<sup>a</sup>Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China; <sup>b</sup>Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Nanjing, China

### ABSTRACT

*Elaeagnus 'viridis'*, an artificial hybrid of *E. macrophylla* (♂) Thunb. (1784) × *E. pungens* (♀) Thunb. (1784), is known for its economic and ecological value. In this study, we sequenced and assembled the whole chloroplast (cp) genome of *E. 'viridis'*. The results showed that its cp genome was 152,284 bp long, showing a typical quadripartite structure and containing a large single-copy region (LSC, 82,299 bp), a small single-copy region (SSC, 18,239 bp), and a pair of inverted repeats (IRs, 51,746 bp). The cp genome contains 132 genes, including 86 protein-coding genes (PCGs), 38 tRNA genes, and 8 rRNA genes. Phylogenetic analysis based on 66 common PCGs revealed that *E. 'viridis'* is most closely related to its maternal parent *E. pungens*. The chloroplast genomic information reported in this study will shed some useful light for further genetic studies in the genus *Elaeagnus*.

### ARTICLE HISTORY

Received 30 July 2023  
Accepted 2 December 2023

### KEYWORDS

*Elaeagnus 'viridis'*; complete chloroplast genome; phylogenetic analysis

### Introduction

The genus *Elaeagnus* is mainly found in temperate and subtropical regions. Due to its economic and ecological value, it has been widely used as an ornamental, medicinal, and wild fruit tree resource, while also being used as pioneer tree species for soil curing and conversion (Kim et al. 2020; Lu et al. 2022). *E. macrophylla* and *E. pungens*, evergreen shrubs of 3–5 m in height, are frequently found in southern China (e.g. Jiangsu, Zhejiang) (Sun and Lin 2010). Their fruits and leaves/stems are used as health foods and in folk medicine, respectively (Shang et al. 2017). Most previous studies on the genus *Elaeagnus* focused on investigating its chemical and nutritional compositions, and its potential for reproduction breeding and as a plant resource (Ge et al. 2013), with only a few studies concentrating on its molecular genetic diversity. To study the mechanism of formation of important traits, we constructed hybrid populations by making artificial crosses with *E. macrophylla* and *E. pungens* as parents. The chloroplast (cp) genome is often used in species delimitation and phylogenetic analysis due to its uniparental heritability and lower substitution rate compared to the nuclear genome (Wei et al. 2020). In this study a hybrid variety *E. 'viridis'* from *E. macrophylla* × *E. pungens* has been selected for cp genome studies, which aimed at investigating the mode of inheritance of the genus *Elaeagnus*, as well as assisting with molecular marker screening.

### Materials and methods

#### Plant materials

Fresh leaves of two-year-old *E. 'viridis'* seedlings were collected from the Nanjing Botanical Garden, Memorial Sun Yatsen (118°49'55"E, 32°3'32"N), Nanjing, China (Figure 1). A specimen was deposited at the Institute of Botany, Jiangsu Province, and the Chinese Academy of Science (<https://www.cvh.ac.cn>, and Li NW, [linaiwei@jib.ac.cn](mailto:linaiwei@jib.ac.cn)) under the voucher number NBGJIB-*Elaeagnus*-0001.


#### DNA extraction and sequencing

Total DNA was extracted from the fresh leaves using the CTAB method (Doyle and Doyle 1987) and a sequencing library was constructed using the NEBNext®Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, England) according to the manufacturer's instructions. The library was sequenced using the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads length.

#### Genome annotation

*De novo* cp genome assembly were performed using NOVOPlasty software with default parameters (Dierckxsens et al. 2017). The cp genes were annotated using the online GeSeq tool (Tillich et al. 2017), and protein-coding genes

**CONTACT** Fan Zhang  [mumizhongfeng@126.com](mailto:mumizhongfeng@126.com)  Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2023.2292156>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



**Figure 1.** Two-year-old *E. 'viridis'* seedlings. The leaves of *E. 'viridis'* were ovoid and dark green, with dense white spots on the leaf surface (photographed by Fan Zhang, in Nanjing Botanical Garden, mem. Sun Yat-sen (118°49'55"E, 32°3'32"N), Nanjing, China).

(PCGs), transfer RNA (tRNA) genes, and ribosomal RNA (rRNA) genes were predicted separately under default parameters. The Chloroplast Genome Viewer (CPGView) was used to test the accuracy of genome annotation (Liu et al. 2023). The annotated chloroplast genome of *E. 'viridis'* was deposited in the GenBank database under the accession number OM935758.

### Phylogenetic analysis

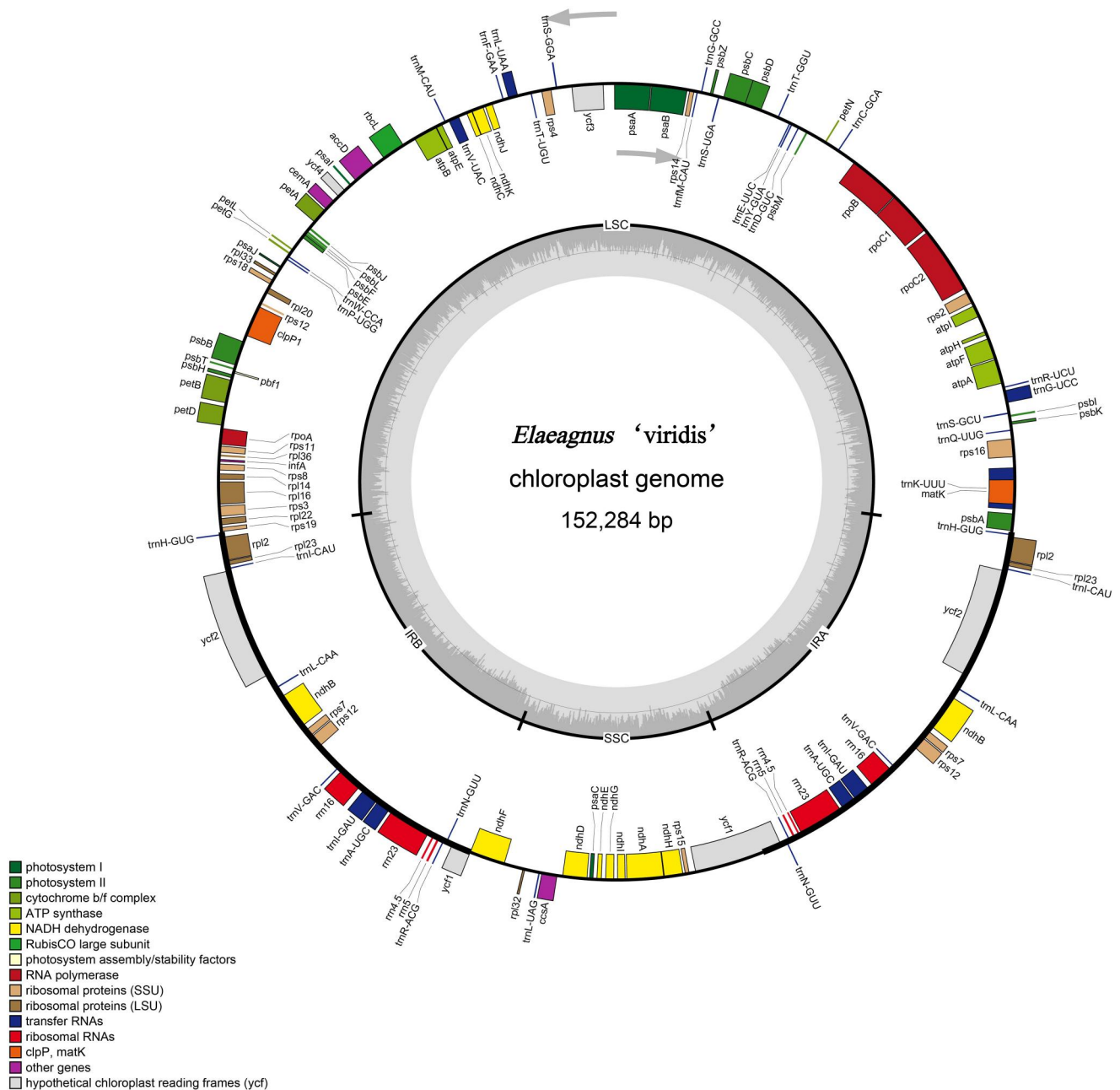
For phylogenetic analysis, the entire cp genomes of 17 species were downloaded from the NCBI database; together with our sequenced cp genome, there were a total of 18 sets of cp genomes. All these 18 species were members of the Elaeagnaceae, with *Barbeya oleoides* being used as outgroups. In total, 66 PCGs, which are commonly present in 18 cp genomes, were used for multiple sequence alignments. The sequences were aligned using MAFFT (Katoh and

Standley 2013), and phylogenetic trees based on the maximum likelihood (ML) method were constructed using IQ-TREE (Minh et al. 2020).

### Results

The raw sequencing data had an average coverage depth of 1695 $\times$ . The assembled complete cp genome of *E. 'viridis'* was a closed single circular DNA (Figure S1), with 152,284 bp in length and an average GC content of 37.04%. The cp genome has a typical quadripartite structure and comprises a pair of inverted repeat (IRa and IRb) regions of 51,746 bp separated by a large single-copy (LSC) region of 82,299 bp and a small single-copy (SSC) region of 18,239 bp (Figure 2). A total of 132 genes were identified in the cp genome, including 38 tRNA genes, 8 rRNA genes, and 86 PCGs. A total of 45 genes belonged to the photosynthesis category, and 71 were related to photosynthesis. The biosynthesis category has six





**Figure 2.** Physical map of *E. 'viridis'* cp genome. Genes transcribed clockwise are drawn inside the circle, while genes transcribed counterclockwise are drawn outside the circle (as indicated by the gray arrows). Different colors are used to show genes with different functions. GC and AT contents are represented by the dark and light gray areas of the inner circle, respectively. LSC: large single-copy region, SSC: small single-copy region, IRA and IRB: a pair of inverted repeats.

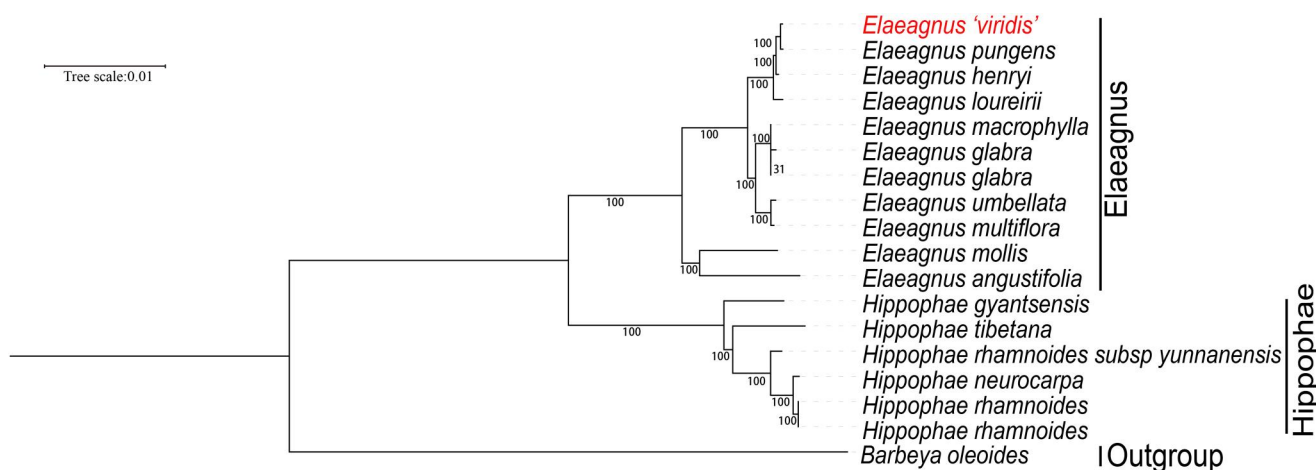
groups, each containing one gene (Table S1). Seventeen genes contain introns, of which 14 (*trnK-UUU*, *rps16*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnL-GAU*, *trnA-UGC*, and *ndhA*) contain a single intron, and three (*rps12*, *ycf3*, and *clpP1*) contain two introns (Table S1). CPGView software was used to identify the gene structures, eleven genes (*rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhA*, and *ndhB*) were identified with a cis-spliced structure (Figure S2). The *rps12* was identified as a trans-spliced gene, containing three exons, with one being located in the LSC region and the other two in the IR regions (Figure S3).

The cp genomes of 11 species of *Elaeagnus* and 6 species of *Hippophae* were reported previously (Kim et al. 2020; Zhao et al. 2020). Thus, we constructed a phylogenetic tree based

on 66 common protein-coding genes. The results showed that *E. 'viridis'* is closely related to *E. pungens* with 100% bootstrap support (Figure 3).

## Discussion and conclusions

In this study, the complete cp genome of *E. 'viridis'* was sequenced and assembled. Sequence analysis showed that the cp genome of *E. 'viridis'* was 152,284 bp long, and contained 132 genes, including 86 PCGs, 38 tRNAs, and 8 rRNAs. Phylogenetic analysis showed that the cpDNA of *E. 'viridis'* was clustered with its maternal parent. The most common mode of cpDNA inheritance in angiosperms is known to be



**Figure 3.** ML phylogenetic tree of *E. 'viridis'* with 16 related species based on 66 PCGs. Numbers on the nodes are bootstrap values from 1000 replicates. The cp sequences of *berchemiella wilsonii* and *zizphus jujuaba* were set as outgroups. The following sequences were used: *Elaeagnus 'viridis'* OM935758, *E. pungens* MW145133, *E. henryi* NC\_06127, *E. loureirii* MH394425, *E. macrophylla* NC\_028066, *E. glabra* LC522137, *E. glabra* MN306571, *E. umbellata* LC522506, *E. multiflora* LC522136, *E. mollis* NC\_036932, *E. angustifolia* NC\_040992, *Hippophae gyantsensis* NC\_044478, *H. tibetana* MN643620, *H. rhamnoides subsp yunnanensis* MK552376, *H. rhamnoides* NC\_035548, *H. rhamnoides* KY794808, *Barbeya oleoides* NC\_040984 (Liu et al. 2019; Wang et al. 2017; Zhao et al. 2020).

matrilineal (Li et al. 2013). However, there are still some angiosperms or individuals that have been reported to be patrilineally inherited (Chat et al. 1999; Ellis et al. 2008). Our study suggests that *E. 'viridis'* is maternally inherited. Although a single sample is not representative of the inheritance of cpDNA in the genus *Elaeagnus*, it provided evidence for the study of this genus. Comparative analyses of the chloroplast genomes show that *E. 'viridis'* is more similar to its female parent (Table S2). However, there are some differences. For the first, the number of genes annotation is different. *E. pungens* (MW145133) has 131 genes annotated because of the absence of *rps16* between *matK* and *trnQ-UUG* (Table S2). We analyzed and compared the sequence between *matK* and *trnQ-UUG* at NCBI, and concluded that this should contain *rps16*. Secondly, the length of their genome sequences (including the LCS, SSC, and IR regions) is different. We hypothesize that the cause of this discrepancy may be due to natural variation, or caused by differences between the female parent of *E. 'viridis'* and the previous test sample (MW145133). In conclusion, our study provided a useful resource for the development of molecular markers for *Elaeagnus* species identification and breeding.

### Authors contributions

Naiwei Li was mainly responsible for the design of the experiment; Fan Zhang and Xinran Chong were mainly responsible for the writing and revising of the paper; Naiwei Li and Hong Chen had an important role in the introduction and breeding of the plant materials; Yanwei Zhou was responsible for plant growth management and sample collection; Ting Zhou participated in the sample preparation work and was responsible for genomic DNA extraction; Fan Zhang and Yanwei Zhou analyzed and interpreted the data; All authors are accountable for all aspects of the work and approved the final version of the paper.

### Ethical approval

The plant material (*Elaeagnus pungens* Thunb.) used in this study is not a protected plant, therefore, no permission or approval was required. The collection of the plant samples was conducted according to the

guidelines provided by Nanjing botanical garden Memorial Sun Yat-sen and did not harm the growth of plants.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work was supported by the Public Science and Technology Research Funds Project of Oceans [grant number: 201505023], the Natural Science Foundation of Jiangsu Province [grant no. BK20201244], and the Natural Science Foundation of Jiangsu Province [grant no. BK20210163]

### Data availability statement

The data of this study are available in GenBank of NCBI (<https://www.ncbi.nlm.nih.gov>) under accession No. OM935758. The associated BioProject, SRA, and BioSample numbers are PRJNA813133, SRR18238883, and SAMN26449928, respectively.

### References

- Chat J, Chalak L, Petit R. 1999. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. *Theor Appl Genet.* 99(1–2):314–322. doi: [10.1007/s001220051238](https://doi.org/10.1007/s001220051238).
- Dierckx N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18. doi: [10.1093/nar/gkw955](https://doi.org/10.1093/nar/gkw955).
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Ellis JR, Bentley KE, McCauley DE. 2008. Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower *Helianthus verticillatus*. *Heredity.* 100(6):574–580. doi: [10.1038/hdy.2008.11](https://doi.org/10.1038/hdy.2008.11).
- Ge Y, Li M, Mei Z, Yang G. 2013. Two new flavonol glycosides from the leaves of *Elaeagnus pungens*. *J Asian Nat Prod Res.* 15(10):1073–1079. doi: [10.1080/10286020.2013.812078](https://doi.org/10.1080/10286020.2013.812078).
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).

- Kim Y, Shin J, Kim D, Lee H, Choi C. 2020. Complete chloroplast genomes of *E. umbellata* Thunb., *E. multiflora* Thunb., *E. macrophylla* Thunb., and *E. glabra* Thunb. (Elaeagnaceae). *Mitochondrial DNA B Resour.* 5(3):2490–2492. doi: [10.1080/23802359.2020.1779142](https://doi.org/10.1080/23802359.2020.1779142).
- Li D, Qi X, Li X, Li L, Zhong C, Huang H. 2013. Maternal inheritance of mitochondrial genomes and complex inheritance of chloroplast genomes in *Actinidia* Lind.: evidences from interspecific crosses. *Mol Genet Genomics.* 288(3–4):101–110. doi: [10.1007/s00438-012-0732-6](https://doi.org/10.1007/s00438-012-0732-6).
- Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: A package for visualizing detailed chloroplast genome structures. *Mol Ecol Resour.* 23(3):694–704. doi:[10.1111/1755-0998.13729](https://doi.org/10.1111/1755-0998.13729).
- Liu J, Gong L-D, Qi L, Liu Z-Y, Niu Y-F, Shi C. 2019. The complete chloroplast genome of *Elaeagnus conferta* Roxb (Elaeagnaceae). *Mitochondrial DNA Part B.* 4(1):2035–2036. doi: [10.1080/23802359.2019.1617074](https://doi.org/10.1080/23802359.2019.1617074).
- Lu Y, Ma Q, Xu X, Wang Z, Andrie I, Savitskaya T, Yuzuak S. 2022. Characterization of the complete chloroplast genome of *Elaeagnus pungens* (elaegnaceae) and phylogeny within elaeagnaceae. *Mitochondrial DNA B Resour.* 7(7):1213–1215. doi: [10.1080/23802359.2022.2090291](https://doi.org/10.1080/23802359.2022.2090291).
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534. doi: [10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015).
- Shang Y-Y, Qin Q, Li M-S, Xu H-Y, Ge Y-B. 2017. Flavonol glycosides from the leaves of *Elaeagnus pungens*. *Nat Prod Res.* 31(9):1066–1072. doi: [10.1080/14786419.2016.1272109](https://doi.org/10.1080/14786419.2016.1272109).
- Sun M, Lin Q. 2010. A revision of *Elaeagnus* L. (Elaeagnaceae) in mainland China. *Journal of Systematics and Evolution.* 48(5):356–390. doi: [10.1111/j.1759-6831.2010.00085.x](https://doi.org/10.1111/j.1759-6831.2010.00085.x).
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45(W1):W6–W11. doi: [10.1093/nar/gkx391](https://doi.org/10.1093/nar/gkx391).
- Wang W-C, Chen S-Y, Zhang X-Z. 2017. Characterization of the complete chloroplast genome of *Elaeagnus mollis*, a rare and endangered oil plant. *Conservation Genet Resour.* 9(3):439–442. doi: [10.1007/s12686-017-0706-4](https://doi.org/10.1007/s12686-017-0706-4).
- Wei XP, Li HJ, Che P, Guo HJ, Zhang BG, Liu HT, Qi YD. 2020. Comparing chloroplast genomes of traditional Chinese herbs *Schisandra sphenanthera* and *S. chinensis*. *Chin Herb Med.* 12(3):247–256. doi: [10.1016/j.chmed.2019.09.009](https://doi.org/10.1016/j.chmed.2019.09.009).
- Zhao K, Wang J, Zhu Z, Shi G, Luo S, Wang H. 2020. Complete plastome sequence of *Elaeagnus glabra* (Elaeagnaceae): an Asian endemic plant species. *Mitochondrial DNA B Resour.* 5(1):288–289. doi: [10.1080/23802359.2019.1702483](https://doi.org/10.1080/23802359.2019.1702483).