Journal of Asia-Pacific Biodiversity xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Asia-Pacific Biodiversity

journal homepage: http://www.elsevier.com/locate/japb

Short Communication

First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae

Karimov Bobur [a], Toshtemirov Jorabek [a], Muzafarova Sarviniso [b], Yusupov Ziyoviddin [a,*]

[a] Institute of Botany, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan
[b] Jizzakh State Pedagogical University, Jizzakh, Uzbekistan

ARTICLE INFO

Article history:
Received 2 August 2024
Received in revised form
1 September 2024
Accepted 1 September 2024
Available online xxx

Keywords:
Lamiaceae
medicinal plant
phylogeny
plastid genome
Thymus seravschanicus

ABSTRACT

This study presents the complete chloroplast genome sequence of *Thymus seravschanicus*, a medicinally important species from Central Asia. The 151,712 bp genome features a typical quadripartite structure and 114 unique genes. Comparative analysis of eight *Thymus* chloroplast genomes identified hotspots of nucleotide diversity, including the SSC region, *ycf1* gene, and intergenic spacers, which can be exploited for phylogenetic and population genetic marker development. Phylogenomic analysis of 33 Lamiaceae species strongly supports the monophyly of major clades, including the genus *Thymus*. This work provides a valuable reference for molecular studies of *T. seravschanicus* and contributes to our understanding of chloroplast evolution in the Mint family.

© 2024 National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA), Publishing Services by Elsevier. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

The *Thymus* genus, comprising around 350 species, is part of the *Mentheae* tribe within the Lamiaceae family (specifically, the *Nepetoideae* subfamily) (Bartolucci et al., 2013; Harley, 2004; Song et al., 2023). Characterized by their aromatic, slender-branched subshrubs, *Thymus* species are found across Eurasia, northern Africa, and southern Greenland (Bartolucci et al., 2013; Song et al., 2023).

The extensive application of *Thymus* species in traditional medicine is based on their pharmacological properties has been documented from ancient times. These plants have also been used as aromatic herbs and spices. Therefore, *Thymus* species are of great interest to the pharmaceutical, food, and cosmetic industries (Bower et al., 2016; Nieto, 2020). *Thymus seravschanicus* Klokov is native to Central Asia and a source of bioactive compounds (Zhumakanova et al., 2021). The complete chloroplast genome of the vascular plants has been accumulated speedily but the chloroplast genome of the *Thymus seravschanicus* has not been reported.

In this study, we report the complete chloroplast genome sequence of *T. seravschanicus* using Illumina sequencing. The genome structure and gene content were characterized and compared with other Lamiaceae species. A phylogenetic analysis was also conducted to infer the evolutionary relationships within the family. This study provides the first complete chloroplast genome of *T. seravschanicus*, which will serve as a valuable genetic resource for future research.

Material and methods

Plant material

Fresh leaves of *T. seravschanicus* were collected from Namangan region, Uzbekistan (41.40523, 71.69703) in June 2024. Specimen identification was verified by taxonomists at the National Herbarium of Uzbekistan and voucher specimens (accession number NAM001) were deposited.

DNA extraction and chloroplast genome sequencing and assembly

Genomic DNA was extracted from leaf samples using a kit from Tiangen, following the provided instructions. The DNA was then prepared for sequencing using a kit from NEB, which involved fragmenting the DNA into pieces around 350 base pairs, adding adapters, and amplifying the fragments through PCR. The resulting

* Corresponding author. ORCID.: 0000-0003-2278-542X.
E-mail address: ziyo-nur87@mail.ru (Y. Ziyoviddin).

Peer review under responsibility of National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA).

https://doi.org/10.1016/j.japb.2024.09.001

pISSN2287-884X eISSN2287-9544/© 2024 National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA), Publishing Services by Elsevier. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article as: Bobur K et al., First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae, Journal of Asia-Pacific Biodiversity, https://doi.org/10.1016/j.japb.2024.09.001

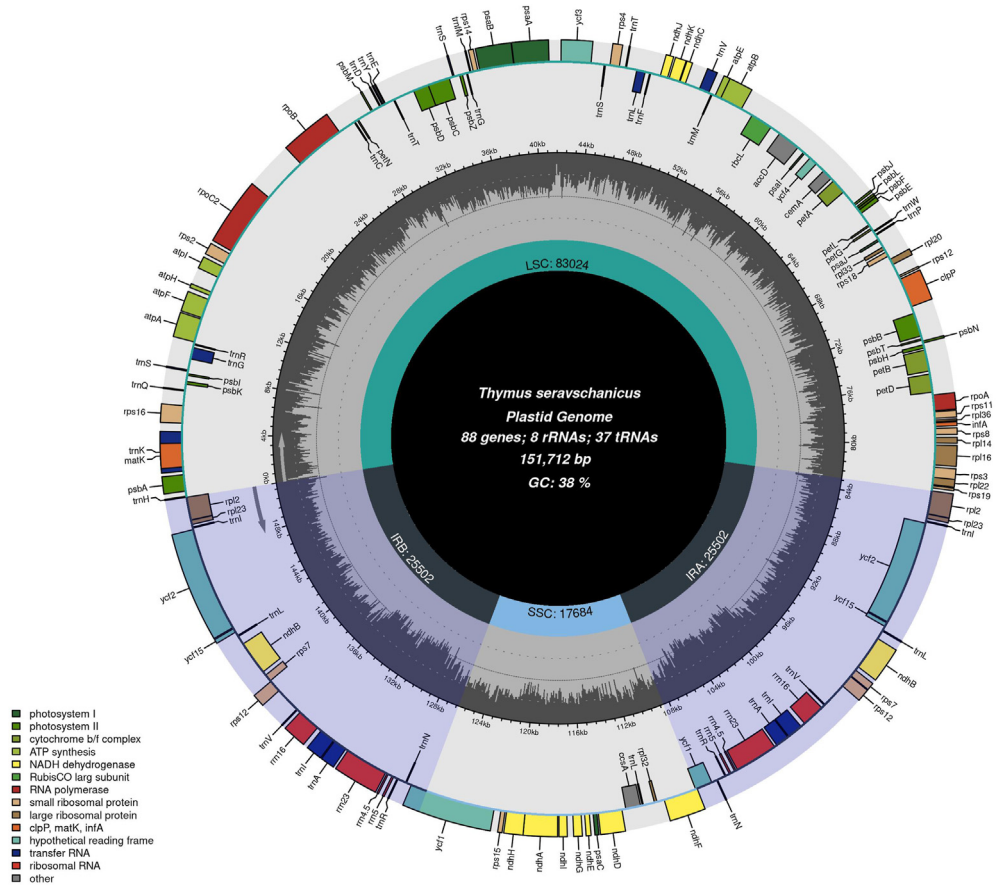


Figure 1. The circular map of the chloroplast genome of *Thymus seravschanicus*. Genes of different functional groups are color-coded. Genes drawn inside of the circle are transcribed clockwise, while those outside are transcribed anticlockwise. The inverted repeat regions (IRa and IRb), which are separated by the large single copy (LSC) region and the small single copy (SSC) region, are denoted.

libraries were purified and evaluated for quality using the Agilent 5400 system, and their concentration was determined. The qualified libraries were then sequenced on an Illumina platform at Novogene Bioinformatics Technology Co. The resulting clean data was assembled using Novoplasty software (Dierckxsens et al., 2017), and the genes were annotated using Geneious software, with the genome of *Thymus mongolicus* (MN482117) as a reference. The annotation was manually reviewed to ensure accuracy, including the start and stop codons, and the boundaries of introns and exons for protein-coding genes (Kearse et al., 2012).

Circular maps of the *T. seravschanicus* cp genomes were obtained using OGDRAW (Zheng et al., 2020). The genome sequence data that support the findings of this study are openly available in GenBank of NCBI under accession no. PQ115142 The associated BioProject, Bio-Sample, and SRA, numbers are PRJNA1140891, SAMN42866766, and SRR30012683 respectively.

Nucleotide diversity

To assess the nucleotide variability (Pi) of the entire cp genome, the DnaSP v.6.12.03 software (Rozas et al., 2017) was utilized. A window length of 1000 base pairs and a step size of 500 base pairs were used for this analysis.

Phylogenetic analysis

The study used a total of 30 chloroplast genomes from the Lamiaceae family, along with *Callicarpa rubella*, *Callicarpa kochiana*, and *Prostanthera wilkieana* as outgroups, for a total of 32 chloroplast genome sequences downloaded from NCBI GenBank. To investigate the phylogenetic relationships, the complete chloroplast genomes were aligned using the MAFFT program (Katoh and

Table 1. List of genes annotated in the chloroplast genomes of *T. seravschanicus*.

Group of genes	Name of genes
Photosystem I	*psaA, B, C, I, J, ycf3*[†], *ycf4*
Photosystem II	*psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z*
Cytochrome b6/f	*petA, B*[*], *D*[*], *G, L, N*
ATP synthase	*atpA, B, E, F*[*], *H, I*
Rubisco	*rbcL*
NADH oxidoreductase	*ndhA*[*], *B*[*,‡], *C, D, E, F, G, H, I, J, K*
Large subunit ribosomal proteins	*rpl2*[*,‡], *14, 16*[*], *20, 22, 23*[‡], *32, 33, 36*
Small subunit ribosomal proteins	*rps2, 3, 4, 7*[‡], *8, 11, 12*[*,‡], *14, 15, 16*[*], *18, 19*
RNA polymerase	*rpoA, B, C1*[*], *C2*
Unknown function protein coding gene	*ycf1*[‡], *ycf2*[‡], *ycf15*[‡]
Other genes	*accD, ccsA, cemA, clpP*[†], *infA, matK*
Ribosomal RNAs	*rrn4.5*[‡], *rrn5*[‡], *rrn16*[‡], *rrn23*[‡]
Transfer RNAs	*trnA-UGC*[*,+], *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU*[‡], *trnI-GAU*[*,‡], *trnK-UUU, trnL-CAA*[‡], *trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU*[‡], *trnP-UGG, trnQ-UUG, trnR-ACG*[‡], *trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC*[‡], *trnV-UAC, trnW-CCA, trnY-GUA*

[*] The gene containing a single intron.
[†] Gene containing two introns.
[‡] Two gene copies in IRs.

Please cite this article as: Bobur K et al., First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae, Journal of Asia-Pacific Biodiversity, https://doi.org/10.1016/j.japb.2024.09.001

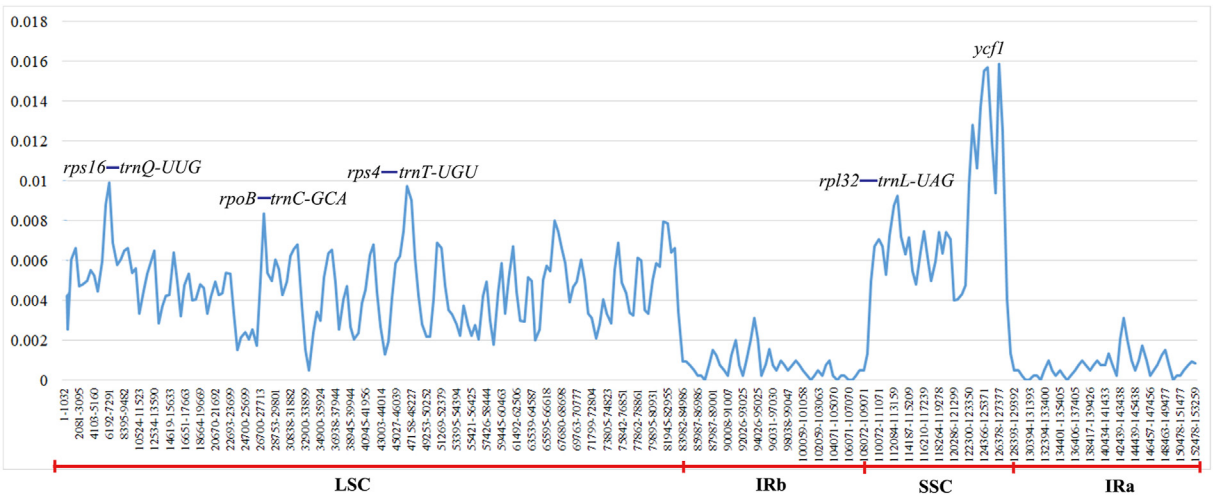


Figure 2. Nucleotide diversity of genomes of seven species and one subspecies.

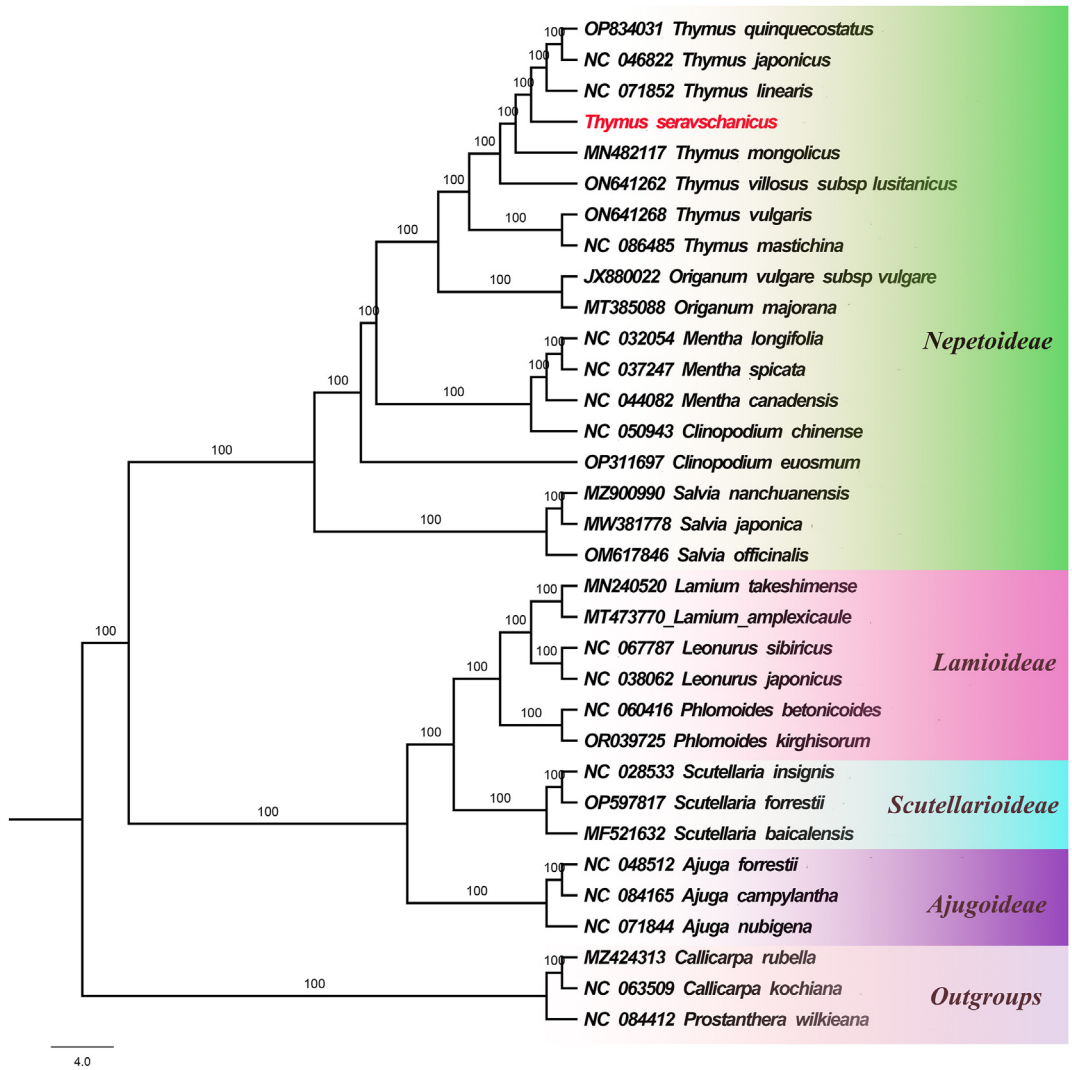


Figure 3. Maximum-likelihood (ML) phylogenetic tree of *Thymus seravschanicus* (in red, PQ115142) and 32 other complete chloroplast genome sequences.

Please cite this article as: Bobur K et al., First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae, Journal of Asia-Pacific Biodiversity, https://doi.org/10.1016/j.japb.2024.09.001

Standley, 2013). Maximum likelihood (ML) phylogenetic analysis was then performed using RAxML with 1,000 bootstrap replicates to assess the support for the inferred relationships (Stamatakis, 2014). The best-fitting model of nucleotide substitutions was determined using Akaike Information Criterion (AIC) in JModelTest V2.1.4, which identified the GTR+G model as the most appropriate (Darriba et al., 2012).

Result and discussion

General feature of the cp genome

The complete chloroplast genome sequence of *T. seravschanicus* was found to be 151,712 base pairs in length, exhibiting the typical four-part structure common among plant chloroplast genomes (Figure 1). This structure includes a large single copy region of 83,024 bp, a small single copy region of 17,684 bp, and two inverted repeat regions each consisting of 25,502 bp. The overall GC content across the chloroplast genome was calculated to be 38%, indicating its composition is rich in adenine and thymine like many other plant chloroplasts.

A total of 114 unique genes were identified, comprising 80 protein-coding genes, 30 transfer RNA genes, and 4 ribosomal RNA genes. Nineteen genes had duplicates located in the inverted repeat regions, while 14 genes contained introns, with 12 having one intron and two genes - *ycf3* and *clpP* - containing two introns each (Table 1).

Nucleotide diversity

In this study, we conducted a comprehensive analysis of the nucleotide diversity in the chloroplast genomes of seven species and one subspecies (*T. seravschanicus, T. quinquecostatus, T. japonicas*, *T. mongolicus, T. mastichina, T. villosus* and *T. villosus* subsp. *lusitanicus*) within the *Thymus* genus. Our results revealed a notable variation in the diversity of nucleotides across the chloroplast genome, with the nucleotide diversity index (Pi) ranging from 0 to 0.016 (Figure 2). Interestingly, we found that the Single Copy Region (SSC) exhibited the highest level of nucleotide diversity, whereas the Large Single Copy (LSC) and Inverted Repeat (IR) regions showed the lowest levels of diversity. Furthermore, our analysis identified specific genes and regions within the chloroplast genome that exhibited high levels of nucleotide diversity, including *ycf1*, *rpl32-trnL-UAG*, *rps4-trnT-UGU*, *rpoB-trnC-GCA*, and *rps16-trnQ-UUG*. These regions are likely to be hotspots of genetic variation and may play important roles in the evolution and adaptation of *Thymus* species. The identification of these high-diversity regions can provide valuable insights into the chloroplast genome evolution and may have implications for the development of novel molecular markers for phylogenetic and population genetic studies in *Thymus.*

Phylogenetic relationship

The analysis of the complete chloroplast genome sequences resulted in the construction of a maximum likelihood phylogenetic tree with 31 well-supported branches, as indicated by the exceptionally high bootstrap values of 100%. This strongly corroborates the evolutionary relationships inferred from the data (Figure 3). Within the plant family Lamiaceae (mints), the subfamilies *Nepetoideae*, *Scutellarioideae*, *Lamioideae*, and *Ajugoideae* were found to be distinct clades (evolutionary lineages) in the phylogenetic tree. The *Scutellarioideae* and *Lamioideae* subfamilies were recovered as monophyletic groups, meaning they each form their own distinct evolutionary lineages. The *Ajugoideae* subfamily was identified as a sister group to the *Scutellarioideae* and *Lamioideae* subfamilies, indicating a close evolutionary relationship between these three subfamilies. The monophyly of the subfamily *Nepetoideae* was strongly supported which supports other studies (Kim et al., 2020). Within the *Nepetoideae* subfamily, the species *Thymus seravschanicus* and 7 other taxa formed a clade. The genus *Thymus* is a monophyletic group. This means that the species currently classified within *Thymus* all share a common evolutionary origin. This aligns with a recent study, JH Song et al supported the monophyly of this genus with three taxa of *Thymus* (Song et al., 2023). Additionally, within the *Nepetoideae* subfamily, the genus *Salvia* was identified as a sister group to the genera *Thymus*, *Origanum, Mentha*, and *Clinopodium*. This indicates a close evolutionary relationship between these genera within the *Nepetoideae* subfamily.

Conclusion

This study presented the complete chloroplast genome sequence of *Thymus seravschanicus*, an important medicinal plant species from Central Asia. The cp genome was 151,712 bp in size with a typical quadripartite structure and 114 unique genes.

Analysis of nucleotide diversity across eight Thymus chloroplast genomes identified hotspots of variation, including the SSC region, *ycf1* gene, and several intergenic spacers. These variable regions can serve as candidates for developing phylogenetic and population genetic markers in *Thymus*.

A chloroplast phylogenomic analysis using 33 species strongly supported the monophyly of major clades within Lamiaceae, including the subfamilies *Nepetoideae, Scutellarioideae, Lamioideae,* and *Ajugoideae*. The genus *Thymus* also formed a well-supported monophyletic group.

The cp genome of *T. seravschanicus* provides an important reference for further molecular studies of this medicinal species and genetic diversity within the genus *Thymus*. It also contributes valuable insights into chloroplast evolution across the mint family Lamiaceae. Overall, this work expands our understanding of plant biology and establishes key genetic resources for future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

CRediT authorship contribution statement

Karimov Bobur: Writing – review & editing, Writing – original draft, Software, Methodology. **Toshtemirov Jorabek:** Writing – review & editing, Writing – original draft, Methodology. **Muzafarova Sarviniso:** Writing – review & editing. **Yusupov Ziyoviddin:** Writing – review & editing, Formal analysis, Conceptualization.

Acknowledgment

This research was funded by the project titled “Molecular-genetic identification of medicinal plant species in the flora of Uzbekistan and Belarus using DNA markers” (FL-7923051878). Thank you, anonymous reviewers, for your comments. We appreciate the time and effort you put into reviewing our work.

References

Bartolucci F, Peruzzi L, Passalacqua N. 2013. Typification of names and taxonomic notes within the genus *Thymus* L.(Lamiaceae). *Taxon* 62:1308–1314.

Bower A, Marquez S, de Mejia EG, et al. 2016. The health benefits of selected culinary herbs and spices found in the traditional Mediterranean diet. *Critical Reviews in Food Science and Nutrition* 56:2728–2746.

Please cite this article as: Bobur K et al., First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae, Journal of Asia-Pacific Biodiversity, https://doi.org/10.1016/j.japb.2024.09.001

ARTICLE IN PRESS

Darriba D, Taboada GL, Doallo R, et al. 2012. jModelTest 2: more models, new heuristics and high-performance computing. *Nature Methods* 9:772.
Dierckxsens N, Mardulyn P, Smits GJNar. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45:e18.
Harley RJ. 2004. *The families and genera of vascular plants*, 167.
Katoh K, Standley D. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.
Kearse M, Moir R, Wilson A, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
Kim HT, Pak JH, Kim JS. 2020. The complete chloroplast genome sequence of *Thymus quinquecostatus* var. *japonicus* (Lamiaceae), an endemic to Ullenung Island of Korea. *Mitochondrial DNA Part B* 5:2401–2402.
Nieto GJP. 2020. A review on applications and uses of thymus in the food industry. *Plants* 9:961.
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34: 3299–3302.
Song JH, Yang S, Choi G, et al. 2023. Characterization of the complete chloroplast genome of the important herbal medicinal plant, *Thymus quinquecostatus* (Lamiaceae). *Journal of Asia-Pacific Biodiversity* 16:140–143.
Stamatakis AJ. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
Zheng S, Poczai P, Hyvönen J, et al. 2020. Chloroplot: an online program for the versatile plotting of organelle genomes. *Frontiers in Genetics* 11:576124.
Zhumakanova BS, Korona-Głowniak I, Skalicka-Woźniak K, et al. 2021. Phytochemical fingerprinting and in vitro antimicrobial and antioxidant activity of the aerial parts of *Thymus marschallianus* Willd. and *Thymus seravschanicus* Klokov growing widely in southern Kazakhstan. *Molecules* 26:3193.

Please cite this article as: Bobur K et al., First complete chloroplast genome of *Thymus seravschanicus*: Insights into genome structure and phylogenetic relationships within Lamiaceae, Journal of Asia-Pacific Biodiversity, https://doi.org/10.1016/j.japb.2024.09.001