



OPEN ACCESS

EDITED BY

Angelo Gismondi,
University of Rome Tor Vergata, Italy

REVIEWED BY

Clizia Villano,
University of Naples Federico II, Italy
Olfa Saddoud Debbabi,
National Gene Bank of Tunisia, Tunisia

*CORRESPONDENCE

Nunzio D'Agostino
✉ nunzio.dagostino@unina.it
Ayed M. Al-Abdallat
✉ a.alabdallat@ju.edu.jo

RECEIVED 16 June 2025

ACCEPTED 20 August 2025

PUBLISHED 09 September 2025

CITATION

D'Agostino N, Fruggiero I, Maisto A,
Taranto F, Al-Kilani MA and Al-Abdallat AM
(2025) Exploring the plastome diversity of
fifteen centuries-old olive trees (*Olea
europaea* L.) from Jordan: insights and
implications for conservation.
Front. Plant Sci. 16:1647776.
doi: 10.3389/fpls.2025.1647776

COPYRIGHT

© 2025 D'Agostino, Fruggiero, Maisto, Taranto,
Al-Kilani and Al-Abdallat. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Exploring the plastome diversity of fifteen centuries-old olive trees (*Olea europaea* L.) from Jordan: insights and implications for conservation

Nunzio D'Agostino^{1*}, Ivan Fruggiero¹, Alessandro Maisto¹,
Francesca Taranto², Mazen A. Al-Kilani³
and Ayed M. Al-Abdallat^{4*}

¹Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy, ²Institute of Biosciences and Bioresources, National Research Council (CNR-IBBR), Bari, Italy, ³National Center for Agriculture Research (NARC), Amman, Jordan, ⁴Department of Horticulture and Crop Science, School of Agriculture, The University of Jordan, Amman, Jordan

The olive tree (*Olea europaea* L.) holds exceptional ecological, cultural, and economic significance in the Mediterranean Basin. Understanding its genetic diversity is critical for conservation, breeding, and authentication of olive cultivars. While nuclear genome analyses have elucidated much of the species' genetic structure, chloroplast genome sequencing provides complementary insights, particularly in tracing maternal lineages, uncovering domestication pathways, and identifying cryptic genetic variation. In this study, we investigated the plastome diversity of fifteen centuries-old olive trees from Jordan through reference-guided assembly and comparative analysis using the FARGA cultivar plastome as a reference. Despite overall genomic conservation, nucleotide diversity analyses revealed several polymorphic hotspots—most notably within the *psbM* and *ycf1* genes and the *atpB-rbcL* intergenic spacer. Structural variation, including simple sequence repeats and tandem repeats, highlighted intra-population diversity. One sample (TF - 3) exhibited heteroplasmy, suggesting a biological origin that warrants further investigation. Phylogenetic reconstruction grouped most samples within the Mediterranean E1 lineage, with TF - 3 and a few others forming distinct clusters. Comparisons with nuclear genotyping data demonstrated both congruence and divergence, emphasizing the value of a dual-genome approach. This study reinforces the utility of plastome sequencing in varietal identification, conservation genetics, and evolutionary studies, and contributes novel genomic resources for Jordanian olive germplasm.

KEYWORDS

DNA barcoding, super-barcode, plastid DNA (cpDNA), genetic diversity, conservation of biodiversity

Introduction

The concept of DNA barcoding involves using a short genetic sequence from a standardized region of the genome to identify and differentiate between species (international barcode of life, <https://ibol.org>). This method serves as a molecular tool for biodiversity assessment, conservation efforts, and the study of evolutionary relationships. By comparing these sequences among different organisms, researchers can establish distinct genetic signatures, or “barcodes,” that facilitate species identification (Kress and Erickson, 2008).

Early plastid DNA (cpDNA) barcoding primarily relied on a single locus, often resulting in limited resolution and accuracy in species identification. However, advancements in the field have led to the adoption of several key loci recognized for their effectiveness in species identification and phylogenetic studies. The most used loci include *trnH-psbA* intergenic spacer, *maturase K* (*matK*), *ribulose biphosphate carboxylase/oxygenase large subunit* (*rbcl*), *psbA-trnH* intergenic spacer, *NADH dehydrogenase subunit F* (*ndhF*), *atpF-atpH* intergenic spacer, *RNA polymerase subunit B* (*rpoB*), and *RNA polymerase subunit B* (*rpoC1*) (Dong et al., 2012). Combining these loci through multi-locus barcoding enhances species identification and phylogenetic analyses by exploiting the unique strengths of each marker (Fazekas et al., 2008; Hollingsworth et al., 2011). Next-generation sequencing (NGS) has further transformed cpDNA barcoding by enabling the analysis of entire plastid genomes, a technique referred to as super-barcoding (Li et al., 2015). This comprehensive genetic framework for species identification addresses the limitations of single- or multi-locus methods, which can sometimes struggle to resolve closely related taxa.

Plastome sequencing is particularly important for the olive tree (*Olea europaea* L.) for several interconnected reasons. First, it enables the detection of genetic variations not easily identifiable through nuclear genomic approaches, offering valuable insights into the phylogenetic relationships among olive subspecies and cultivars. This in-depth genetic understanding supports both conservation and breeding programs, enabling a better assessment of genetic structure and diversity across cultivated and wild olive populations (Besnard et al., 2013; Dong et al., 2022). Such insights shed light on the domestication processes and migration patterns of olive species, especially within the Mediterranean basin, where the olive tree holds substantial ecological and cultural significance (Bazakos et al., 2023; Al-Kilani et al., 2024). Understanding these evolutionary dynamics can guide conservation efforts and ensure the maintenance of valuable genetic resources (Díez et al., 2016; Rallo et al., 2018).

Plastome analysis is valuable for varietal identification, especially in efforts to trace the maternal lineage of cultivars, clonal lines, or wild populations (Besnard et al., 2002a; Mariotti et al., 2023). In *Olea europaea*, accurate varietal identification plays a crucial role in ensuring the genetic authenticity and traceability of

olive oil production chains. It safeguards traditional and endemic germplasm and supports the implementation of quality certification schemes, such as Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI) (Pérez-Jiménez et al., 2013; Sion et al., 2021). Recent studies have demonstrated the efficacy of super-barcoding in various plant groups, establishing it as a powerful tool for systematic biology (e.g. Krawczyk et al., 2018; Chen et al., 2023). Additionally, plastome-derived super-barcodes enable precise differentiation among cultivars, thereby enhancing quality control, preventing frauds, and supporting accurate product labeling in the marketplace (Ganopoulos et al., 2013; Kumar et al., 2011). Finally, plastome-based analyses contribute significantly to the conservation of genetic diversity by identifying unique or endangered cultivars, guiding targeted preservation strategies, and reinforcing the long-term resilience of olive agroecosystems.

Early studies, such as Amane et al. (2000), investigated Moroccan olive populations and identified four chlorotypes, emphasizing distinct genetic lineages in wild olives. Building on this, Besnard and Bervillé (2002) compared RFLP and PCR-based methods for cpDNA analysis, demonstrating the superior reliability of PCR-based approaches. Subsequent studies using PCR-RFLP and microsatellite markers identified 15 chlorotypes in the olive complex and constructed a cpDNA phylogenetic tree with five geographically distinct clades: clade A in Central and Southern Africa, clade C in Asia, clade M in North-West Africa, and clades E1 and E2 in the Mediterranean Basin. Cultivated olives clustered with Mediterranean and Saharan wild forms (E1 and E2), showing strong genetic differentiation between eastern and western Mediterranean olives, likely due to distinct glacial refugia. An eastern chlorotype, dominant in cultivars, appears to have been spread westward by humans (Besnard et al., 2002b). Later, Besnard et al. (2011) sequenced eight cpDNA genomes, identifying 22 haplotypes across Mediterranean olive lineages and revealing low nucleotide divergence, particularly in cultivated olives. Mariotti et al. (2010) identified new polymorphic cpDNA regions for improved cultivar differentiation, and Pérez-Jiménez et al. (2013) developed cpDNA markers for varietal tracing in olive oils. In 2015, Noman et al. (2015) advanced the field by creating cpDNA markers for varietal identification, supporting on-farm conservation and preserving authentic olive varieties. Finally, Mariotti et al. (2023) provided deeper insights into domestication by exploring genealogical relationships within olive species.

The goal of this work is to explore the plastome diversity of fifteen centuries-old olive trees from Jordan, providing insights and implications for conservation. These trees were previously genotyped at the nuclear level (Al-Kilani et al., 2024), and here we utilize cpDNA analysis to further elucidate relationships among the samples. Additionally, this study aims to demonstrate the discriminatory power of the complete plastid genome in detecting and identifying mutation hotspots, offering new perspectives for research on biodiversity, phylogeny, and biogeography. Together, these efforts contribute to a deeper understanding of genetic heritage and inform strategies for the preservation of these centuries-old trees.

Materials and methods

Plant material, DNA isolation, and sequencing

Leaf samples were collected from fifteen olive trees belong to *Olea europaea* subsp. *europaea*, each estimated to be centuries-old, based on trunk diameter measurements taken at 130 cm above ground level, following the methodology outlined by Arnan et al. (2012). These samples were sourced from various regions across Jordan (see Table 1).

Total genomic DNA (gDNA) was extracted using the Qiagen DNeasy Plant Mini Kit. The integrity of the extracted DNA was assessed through agarose gel electrophoresis, and its concentration was measured using a Qubit® 2.0 fluorimeter (Thermo Scientific).

Approximately 100 ng of gDNA per sample was used as input material for DNA sample preparation. Sequencing libraries were constructed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA), in accordance with the manufacturer’s instructions. Index codes were incorporated to distinguish sequences belonging to each sample.

In brief, the DNA samples were fragmented to an average size of 350 bp using sonication. The resulting DNA fragments were then end-polished, A-tailed, and ligated with full-length adapters suitable for Illumina sequencing. Following ligation, PCR amplification was

performed to enrich the library. Finally, the PCR products were purified using the AMPure XP system, and the resulting libraries were analyzed for size distribution with an Agilent 2100 Bioanalyzer and quantified via real-time PCR.

Plastome sequencing, assembling, and annotation

DNA samples were sequenced using the Illumina NovaSeq 6000 platform, generating 2 × 150 paired-end reads. Quality control of the raw reads was performed with FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and summarized using MultiQC v1.12 (<https://seqera.io/multiqc/>). Subsequently, Trimmomatic v0.40-rc1 (Bolger et al., 2014) was utilized to eliminate Illumina technical sequences, remove low-quality sequences, trim reads from both ends, filter out any reads that were too short after trimming. The following parameters were applied: adapters.fa:2:20:10:5, LEADING:30, TRAILING:30, SLIDINGWINDOW:10:30, and MINLEN:75.

Reference-based assembly was conducted using NOVOPlasty (Dierckxsens et al., 2016), with the plastome sequence deposited in GenBank under accession LR743800.1 serving as the reference genome. The *matk* gene sequence was used as the seed for the assembly process.

TABLE 1 Collection details for the 15 olive trees (*Olea europaea* subsp. *europaea*) investigated in this study, including locations, regions, area names, and climatic zones.

Tree identification code	Location	Region	Area	Latitude	Longitude	Elevation	Physiography	Habitat	Micro-environment
AJ-24	North Jordan	Ajloun	Anjara	32.3060110	35.7525660	877	Plain	Arable	Farmer field
BA-40	Central Jordan	Balaqa	ArRumma	32.1295181	35.6794210	725	Mountain	Forest	Forest margin
BA-42	Central Jordan	Balaqa	ArRumma	32.1295181	35.6794210	725	Mountain	Forest	Forest margin
BA-66	Central Jordan	Balaqa	ArRumma	32.1296770	35.6798990	708	Mountain	Forest	Forest
BA-74	Central Jordan	Balaqa	ArRumma	32.1296310	35.6798770	713	Hill	Forest	Forest margin
BA-75	Central Jordan	Balaqa	ArRumma	32.1296250	35.6798650	709	Hill	Forest	Forest margin
IR-55	North Jordan	Irbid	Kufr-Asad	32.5844280	35.7238330	345	Plain	Arable	Farmer field
IR-61	North Jordan	Irbid	Bani Kana	32.7151360	35.8086270	267	Valley	Rangeland	Rock face
IR-62	North Jordan	Irbid	Bani Kana	32.7151100	35.8086010	267	Valley	Rangeland	Rock face
JA-56	North Jordan	Jarash	Borma	32.2096010	35.7791730	445	Mountain	Forest	Forest
JA-71	North Jordan	Jarash	Borma	32.2158990	35.7991260	470	Mountain	Forest	Forest
JA-72	North Jordan	Jarash	Borma	32.2158860	35.7991220	470	Mountain	Forest	Forest
MN-14	South Jordan	Ma'an	Wadi Musa	29.3358760	35.2861220	1101	Plain	Desert	Sand bank
TF-1	South Jordan	Tafilah	Aema	30.8781990	35.5965200	870	Upland	Arable	Farmer field
TF-3	South Jordan	Tafilah	Aema	30.8781990	35.5965200	870	Upland	Arable	Farmer field

Chloroplast genome annotation was conducted using the GeSeq tool (Tillich et al., 2017), with gene structures manually curated based on the structural annotations of reference sequence LR743800.1.

The circular plastome map of sample AJ - 24 was generated using the online tool OGDRAW (Greiner et al., 2019).

Chloroplast genome sequences and annotations generated in this study are available in GenBank under accession numbers PQ676988-PQ677002.

Identification and characterization of sequence variations

Nucleotide variability among olive cpDNAs was analyzed utilizing a variety of bioinformatic tools.

To enable comparative analysis, multiple sequence alignment was performed using the MAFFT program (Katoh and Standley, 2013). The analysis included the plastome of *Olea europaea* subsp. *europaea* cultivar FARGA (accession LR743800.1) along with the following reference genomes deposited in GenBank: NC_015401.1 (*Olea europaea* var. *sylvestris*), MW072292.1 (*Olea europaea* subsp. *europaea* cultivar Mehras), and CM032646.1 (*Fraxinus excelsior*).

RAxML (Stamatakis, 2014) was employed to construct a maximum-likelihood (ML) tree with 1,000 rapid bootstrap inferences, using a generalized time reversible (GTR) substitution model and the Gamma model of rate heterogeneity. In parallel, a neighbor-joining (NJ) tree was generated in MEGA 11 (Tamura et al., 2021; <https://www.megasoftware.net>) with 1,000 bootstrap replicates. The plastome of *F. excelsior* was designated as the outgroup for both analyses. Tree visualizations were performed using FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

The identification of microsatellites was performed using the MISA web tool (Beier et al., 2017), available at <https://webblast.ipkgatersleben.de/misa/>. The unit size and minimum repeat parameters were set as follows: 1/10, 2/6, 3/5, 4/5, 5/5, and 6/5. Additionally, tandem repeats were detected using the Tandem Repeat Finder (TRF) web tool, accessible at <https://tandem.bu.edu/trf/home>, with default settings applied (Benson, 1999).

Single nucleotide variants (SNVs) were extracted from a multiple sequence alignment using the *snp-sites* tool (Page et al., 2016), with the plastome LR743800.1 as the reference sequence. A custom Python script was then developed to determine whether the SNVs were located within genic or intergenic regions.

Nucleotide diversity (π) was assessed using a sliding window approach. The π calculations were performed iteratively with the MSA2pi tool (<https://github.com/baoxingsong/MSA2pi>), which lacks support for overlapping window configurations. To overcome this limitation, a custom Python script was developed to segment the multiple sequence alignment and automate the execution of MSA2pi. Additionally, a second Python script was created to include both single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) in the diversity calculation. The analysis utilized a window size of 1000 base pairs (bp) and a step size of 500 bp. The results were stored in a comma-separated values

(CSV) file for subsequent analysis. Graphical representations of the nucleotide diversity data were produced using Python, with the *pandas* and *matplotlib.pyplot* libraries for visualization.

Results

Plastome size and organization

A reference-guided assembly was conducted using the cpDNA of the FARGA cultivar (LR743800.1) as the reference to obtain the complete plastome for all 15 samples. Ten samples exhibited an identical genome size of 155,886 bp, while the remaining five showed variations of one or two base pairs, either larger or smaller (Table 2). Notably, the sequence of sample MN - 14 is identical to that of JA - 72 and IR - 62. As expected, all genotypes exhibited the characteristic quadripartite structure typical of angiosperms, consisting of a pair of inverted repeats (IR) of 25,741 bp, flanked by two single-copy regions: a large single copy (LSC) ranging from 86,610 bp to 86,614 bp, and a small single copy (SSC) of 17,791 bp (Table 2). No variation in guanine/cytosine (GC) content was observed among the genotypes, with all showing a consistent GC content of 37%.

The plastid DNA contains 125 unique, non-redundant genes, including 80 protein-coding genes, 30 *tRNA* genes, and 4 *rRNA* genes. Of the *tRNA* genes, 7 are located within the inverted repeat (IR) regions and are duplicated, as are the *rRNA* genes. The LSC region contains 60 protein-coding genes, 7 of which are multi-exon, and 22 *tRNA* genes, 4 of which have two exons. The SSC region includes 11 protein-coding genes, one with two exons, and a single *tRNA* gene.

The IR regions feature 8 protein-coding genes, including a truncated copy of *ycf1*, 7 *tRNA* genes (2 with two exons), and the 4 *rRNA* genes. The *rps12* gene is trans-spliced, with exons 2 and 3 in the IR. Additionally, *rps19* spans the LSC-IRb junction (Figure 1).

Nucleotide variability and phylogenetic insights

Figure 2A provides an overview of nucleotide variability. Of the 158 variable loci, 89 were located in intergenic regions, while the remaining 69 were found within 23 genes (see Supplementary Table 1). SNP counts varied across samples, ranging from 64 in JA - 56 to 136 in TF - 3 (Figure 2A). The large number of SNPs observed in the TF3 sample is attributed to loci in a heterozygous state. We investigated whether these loci are the result of sequencing errors or due to heteroplasmy. As shown in Supplementary Figure 1, a large percentage of sequence reads confirm the presence of the alternative allele, supporting the hypothesis that heteroplasmy is the underlying cause.

Fifty-four simple sequence repeats (SSRs) were identified, including 4 compound repeats (Figure 2B). Fourteen of these microsatellites were located within 8 genes, with *rpoC2* containing 4 SSRs, *clpP* having 3 SSRs, and *rpl16* containing 2 SSRs.

TABLE 2 Plastome features of the olive tree samples sequenced in this study.

Tree identification code	GenBank acc.	Size (base pairs)			
		Total	LSC	SSC	IR
AJ-24	PQ676988	155887	86612	17791	25742
BA-40	PQ676989	155885	86610	17791	25742
BA-42	PQ676990	155886	86611	17791	25742
BA-66	PQ676991	155886	86611	17791	25742
BA-74	PQ676992	155886	86611	17791	25742
BA-75	PQ676993	155886	86611	17791	25742
IR-55	PQ676994	155888	86613	17791	25742
IR-61	PQ676995	155886	86611	17791	25742
IR-62	PQ676996	155886	86611	17791	25742
JA-56	PQ676997	155885	86610	17791	25742
JA-71	PQ676998	155886	86611	17791	25742
JA-72	PQ676999	155886	86611	17791	25742
MN-14	PQ677000	155886	86611	17791	25742
TF-1	PQ677001	155886	86611	17791	25742
TF-3	PQ677002	155884	86609	17791	25742

Mononucleotide repeats (adenosine or thymine) were the most common type of microsatellite. Twenty-three loci exhibited variable numbers of repetitive units (Figure 2B; Supplementary Table 2).

A total of 25 tandem repeats were identified, with 11 located in the IR regions (Figure 2C). Their sizes ranged from 9 to 36 bases, and repeat periods varied between 2 and 7.4 bases. Among these, 10 tandem repeats were found in intergenic regions, while the remaining 15 were associated with four genes (Figure 2C), with nine specifically localized within the *ycf2* gene.

Notably, a tandem repeat with a period size of 12 bases and a copy number of 3.2 was absent in the reference genome but present in all sequenced samples. Conversely, a unique tandem repeat with a period size of 18 bases and a copy number of 2 was found exclusively in the reference sequence LR743800.1 (Supplementary Table 3).

The nucleotide diversity (π) analysis revealed a high degree of sequence similarity across the olive plastomes while identifying three main polymorphic hotspots: two within genic regions (*psbM* and *ycf1*) and one in the intergenic *atpB-rbcL* spacer. Notably, *psbM* and *atpB-rbcL* exhibited high SNP variability, whereas *ycf1* showed a higher frequency of indels (Figure 3). Specifically, the *ycf1* protein in the reference plastome is 1,801 amino acids long, whereas in all analyzed samples, it extends to 1,876 amino acids. The pairwise global alignment highlights a 75-amino acid insertion in the *ycf1* protein of AJ - 24. A screenshot of this alignment, comparing the *ycf1* protein in the reference sequence (LR743800.1) and AJ - 24, is provided in Supplementary Figure 2.

Two distinct algorithms were utilized to investigate the phylogenetic relationships among the samples under study. Figure 4 illustrates the phylogenetic trees generated using the Maximum Likelihood (ML) algorithm (Figure 4A) and the

Neighbor Joining (NJ) method (Figure 4B). These trees provide a visual representation of the evolutionary relationships among the analyzed samples, highlighting both genetic similarities and differences. Bootstrap values placed near the nodes indicate the level of statistical support for each grouping, with higher values suggesting stronger confidence in the tree's structure. In the ML-based phylogenetic tree (Figure 4A), a distinct cluster is formed by samples JA - 56, BA - 75, JA - 71, and TF - 3, which group together with a bootstrap value of 100, signifying a high degree of genetic similarity among them. The NJ-based phylogenetic tree (Figure 4B) shows that the majority of the Jordanian samples are positioned in the upper portion of the tree, implying a closer evolutionary relationship between them. The reference sequences, LR743800.1 and NC_015401.1, are found in the lower section, confirming their role as benchmark sequences for comparison against the analyzed samples. CM032646.1 is designated as an outgroup and is located at the base of both trees. This outgroup serves to root the tree and provides a reference point for determining the evolutionary direction.

Discussion

Plastome features and their barcoding potential

Using complete plastome sequences as DNA barcodes offers a groundbreaking method for identifying and characterizing *Olea europaea* species and cultivars. Traditional barcoding, based on short genetic markers, often fails to distinguish closely related taxa. Advances in next-generation sequencing now allow for full plastid

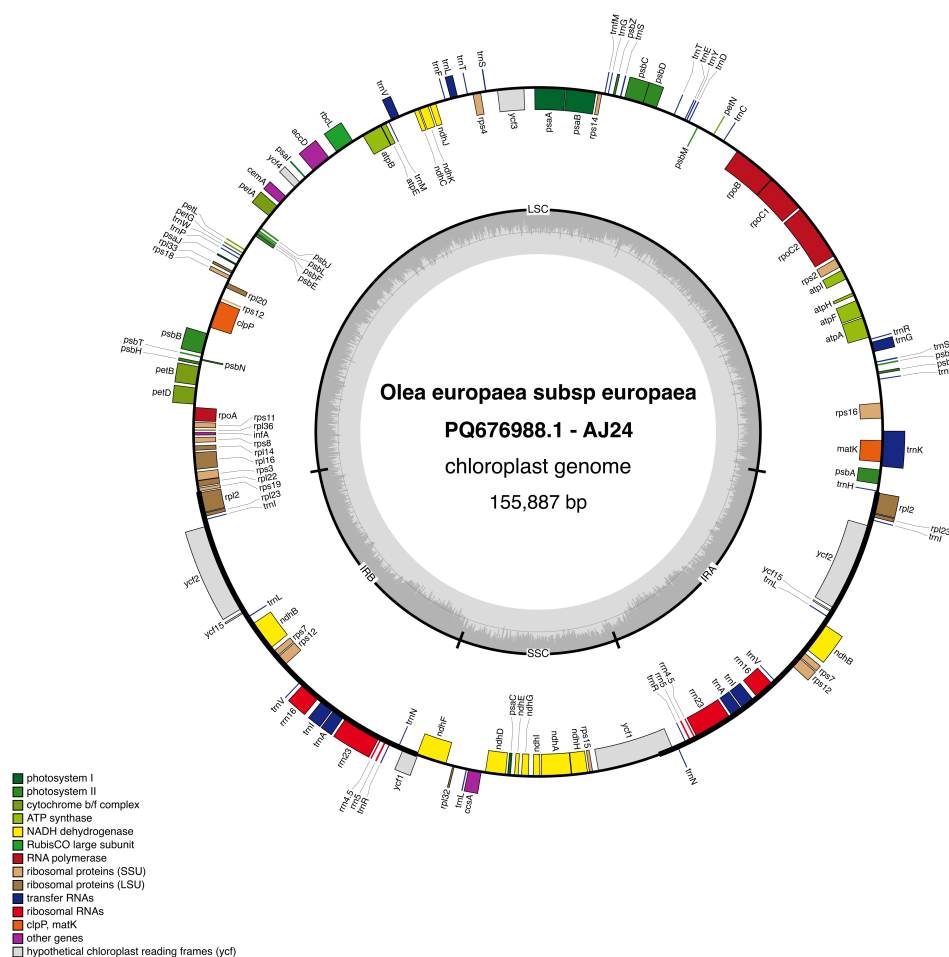


FIGURE 1

Map of the AJ - 24 chloroplast genome. Genes located inside the outer circle are transcribed in the clockwise direction, while those outside are transcribed counterclockwise. Different colors represent genes classified into various functional groups. The inner circle highlights the GC content, with the 50% threshold marked for reference. The genome is divided into the inverted repeat (IR), large single-copy (LSC), and small single-copy (SSC) regions, as indicated by their respective labels.

genome analysis, improving resolution in complex cases (Krawczyk et al., 2018; Song et al., 2020; Li et al., 2015; Coissac et al., 2016).

Olive plastomes, which exhibit relatively low mutation rates and highly conserved structures, provide a robust framework for species identification and phylogenetic studies. The use of plastomes as barcodes leverages the unique advantages of these genomic regions, allowing for the identification of distinct genetic signatures associated with different olive subspecies and cultivars. This enhanced resolution is particularly beneficial in regions with rich biodiversity, such as the Mediterranean basin, where olives have been cultivated for millennia (Besnard et al., 2018).

This method can detect cryptic diversity among phenotypically similar cultivars, aiding in taxonomy, conservation, and the understanding of evolutionary relationships (Gostel and Kress, 2022; Kartzinel et al., 2025).

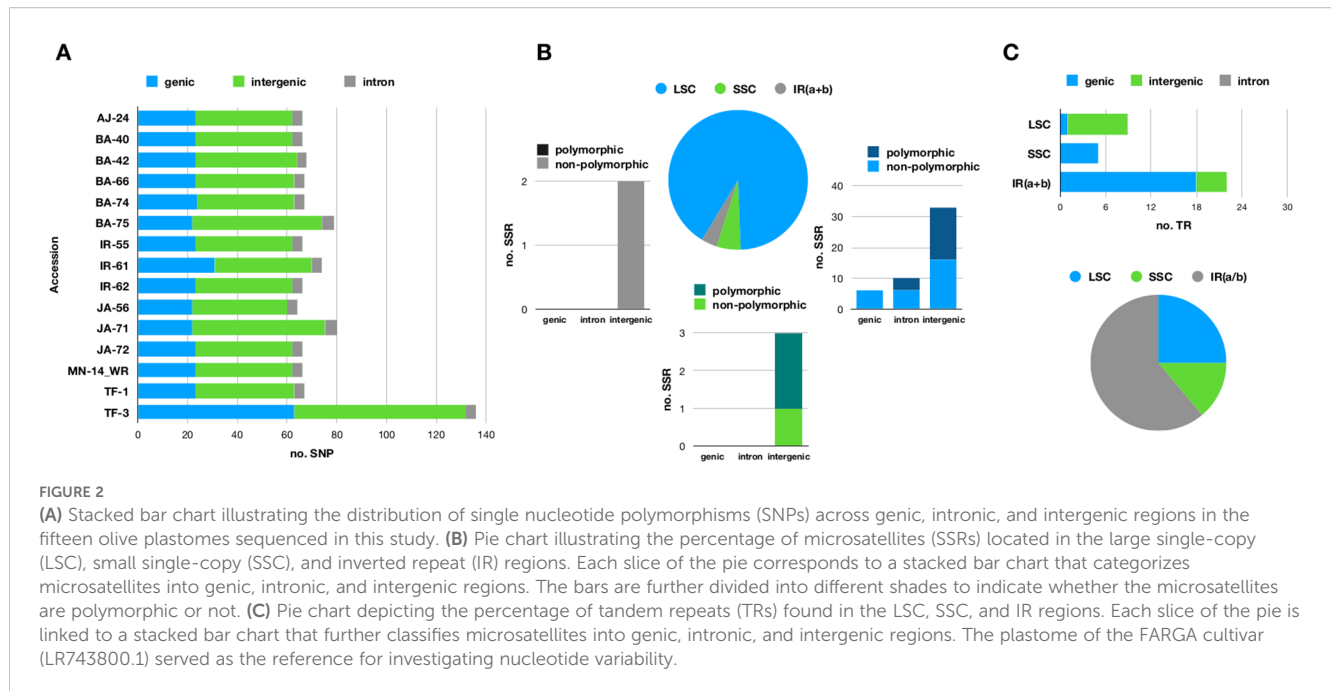
Challenges include the need for comprehensive plastome databases and standardized protocols to ensure reproducibility across studies (Freudenthal et al., 2020; Turudić et al., 2021; Heinze, 2021).

Moreover, plastid diversity in cultivated olives is limited, with strong plastome-mitochondrial linkage indicating minimal recurrent organellar mutations (Mariotti et al., 2010).

Genomic variation and polymorphic hotspots

SNP identification reveals a higher degree of variability in non-coding regions compared to coding regions. The number of SNPs varied across samples (Supplementary Table 1), which may reflect population-specific variations or the influence of environmental factors.

Notably, the TF - 3 sample exhibits heteroplasmy at multiple loci, indicating the possible presence of different plastid genome variants within the same organism. Plastid heteroplasmy, defined as the coexistence of multiple sequence variants in a single cell, challenges the conventional view of plastomes as predominantly homoplasmic due to their clonal inheritance and high copy number. The occurrence of heteroplasmy at different loci within the chloroplast



genome suggests the involvement of underlying biological mechanisms that warrant further investigation. Several factors may contribute to this heteroplasmic state, including biparental inheritance, somatic mutations, and intracellular recombination (Greiner et al., 2015). In the TF - 3 sample, the presence of distinct alleles has been independently confirmed by consistent read coverage, allowing us to rule out sequencing artifacts or assembly errors. Therefore, further studies are needed to elucidate the biological basis of chloroplast heteroplasmy in this case.

Twenty-three microsatellite loci exhibited variable numbers of repetitive units (Supplementary Table 2), which could be crucial for understanding the genetic diversity and stability of these loci across different samples. Notably, a tandem repeat with a size of 18 bases and a copy number of 4.2 was absent in samples BA - 40 and BA - 42, suggesting this repeat may be polymorphic or that these samples have undergone genomic variation leading to its loss. In contrast, a tandem repeat with a period size of 12 bases and a copy number of 3.2 was absent in the reference genome but present in all sequenced samples, highlighting its potential significance or uniqueness in the sequenced populations. Furthermore, a unique tandem repeat with a period size of 18 bases and a copy number of 2 was found exclusively in the reference sequence LR743800.1 (Supplementary Table 3), indicating that this repeat may be specific to the reference genome and could serve as a valuable marker for further comparisons. The nucleotide diversity (π) analysis identified three polymorphic hotspots. The *psbM* gene, encoding a small subunit of photosystem II, is not typically reported as a major hotspot for plastome polymorphism. However, genes involved in photosynthesis can accumulate variations due to environmental selection pressures or lineage-specific mutations.

The high SNP variability observed in *psbM* among olive plastomes could potentially result in structural variations of the PSBM protein, which might influence its function within the photosystem II complex. Assessing the amino acid substitutions

and their locations relative to critical functional domains or interaction sites could provide insight into the possible structural and functional impacts of these variations.

The *atpB-rbcL* intergenic spacer is widely recognized as a polymorphic hotspot in plastomes across various plant species. Intergenic spacers, particularly those between essential photosynthetic genes like *atpB* and *rbcL*, tend to accumulate mutations more readily due to relaxed selective constraints. They are often used as molecular markers in phylogenetic and population genetics studies due to their high variability (Shaw et al., 2007). Finally, the *ycf1* gene is one of the most well-documented polymorphic regions in plastomes. It plays a role in protein translocation and is known for its high evolutionary rate. Studies across diverse angiosperms have consistently identified *ycf1* as one of the most variable plastid genes, making it a valuable marker for phylogenetic and evolutionary studies (Dong et al., 2015). The observed 75-amino acid insertion in AJ - 24 and other Jordanian samples suggests a structural modification, which may have functional or evolutionary significance.

Phylogenetic clustering and lineage differentiation

The phylogenetic analysis revealed distinct clustering patterns among the 15 selected olive trees from Jordan, with the NJ tree including all samples while the ML tree represented 13. Notably, samples IR - 62 and JA - 72 were absent from the ML tree (as they were identical to MN - 14), yet they appeared identical in the NJ analysis. The clustering of the majority of Jordanian olive trees within the E1 lineage, which includes the MEHRAS cultivar (referred to as KEFARI BALADI by Al-Kilani et al., 2024), indicates a significant genetic affinity among these trees. Importantly, no samples were

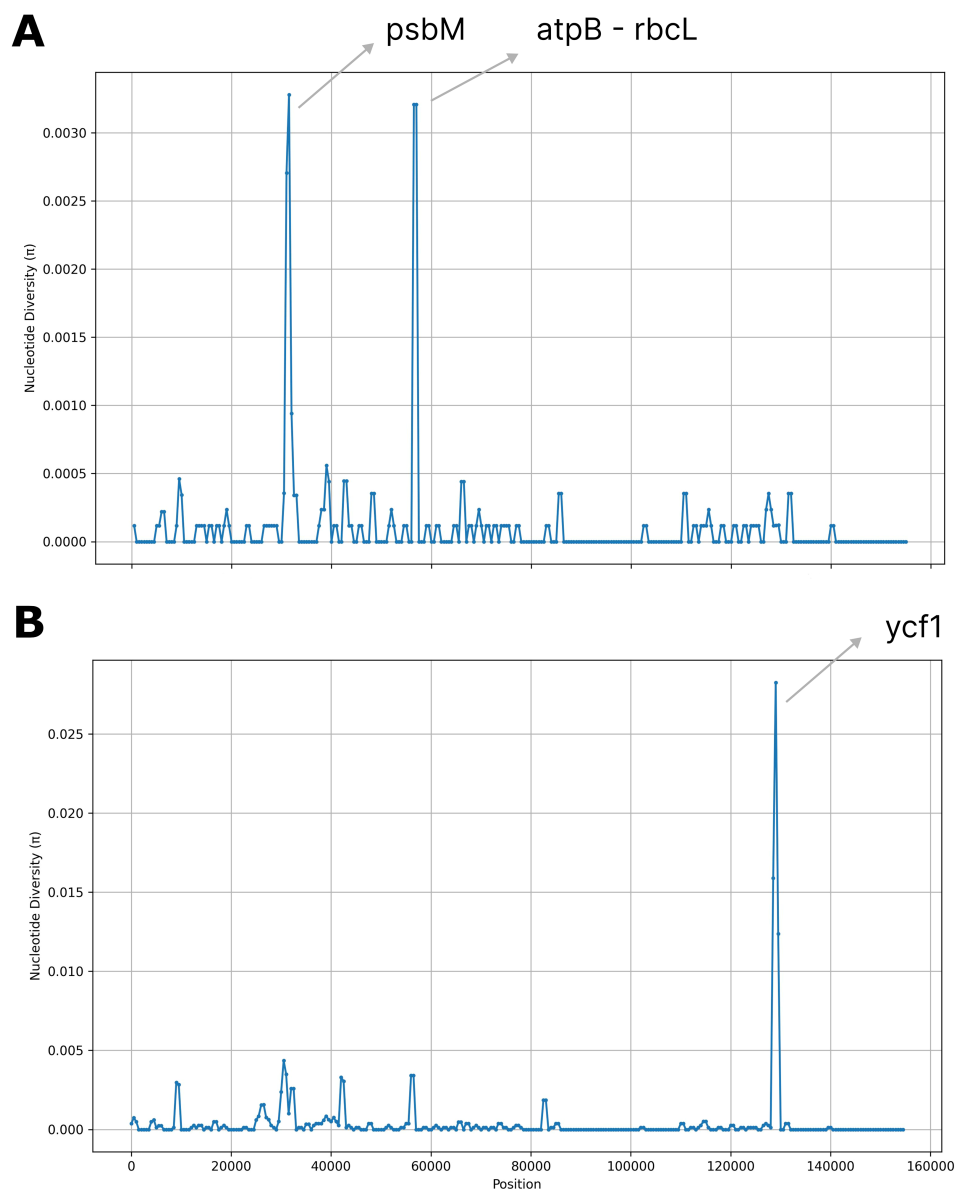


FIGURE 3

Nucleotide diversity (π) values across 17 complete olive plastomes. Regions with high nucleotide variability ($\pi > 0.025$) are highlighted. (A) Single nucleotide polymorphisms (SNPs). (B) Insertions and deletions (indels). Window size = 1000 base pairs (bp); step size = 500 bp.

found to cluster with HAUT ATLAS (E2 lineage) or FARGA (E3 lineage), suggesting a distinct genetic landscape for the olive trees in Jordan.

Among the analyzed samples, TF - 3 stands out for its marked divergence from the other Jordanian samples. This distinction is particularly intriguing given TF - 3's location in southern Jordan, a region geographically isolated from the northern areas. This divergence warrants further investigation, as it could offer valuable insights into localized adaptations or historical events that contributed to genetic separation. Interestingly, TF - 3 formed a distinct cluster with JA - 56, BA - 75, and JA - 71 in both phylogenetic trees, clearly separating them from the remaining samples. In contrast, nuclear genotyping of the same samples in a previous study (Al-Kilani et al., 2024) placed them within a larger cluster that encompassed the majority of genetically diverse trees from Jordan.

Interestingly, the predominant cultivars KEFARI BALADI (MEHRAS), KANABSI (MN - 14, Wadi Rum), and BA - 66 were found to be genetically identical, aligning with previous studies indicating their classification within the E1 lineage. This observation highlights the need for nuclear-level genetic analysis to further differentiate these cultivars, as plastid data alone may not capture the full extent of their genetic diversity.

These findings are consistent with the broader history of olive domestication, which began in the Eastern Mediterranean—particularly the Levant, including present-day Jordan—before spreading westward through trade and migration (Besnard et al., 2013). Eastern olive germplasm played a key role in shaping cultivated varieties across the Mediterranean, as reflected in the widespread E1 plastid lineage (Julca et al., 2020). Secondary

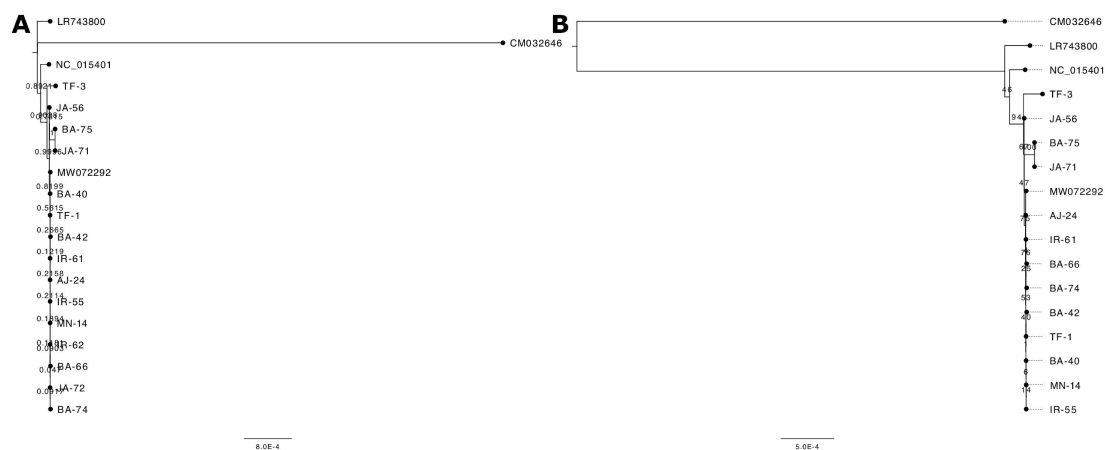


FIGURE 4

Phylogenetic trees constructed using (A) the Neighbor-Joining method and (B) the Maximum Likelihood method. The analysis included 15 Jordanian olive samples alongside reference sequences: LR743800.1 (*Olea europaea* subsp. *europaea*), NC_015401.1 (*Olea europaea* var. *sylvestris*), MW072292.1 (*Olea europaea* subsp. *europaea*, cultivar MEHRAS), and CM032646.1 (*Fraxinus excelsior*), with *F. excelsior* serving as the outgroup. Bootstrap support values are indicated on the branches. The tree in panel (B) contains two fewer samples, as their sequences were identical to MN - 14 and were removed from the analysis.

domestication events through admixture with local wild populations further contributed to modern olive diversity.

Recent genomic studies underscore this complexity, revealing recurrent admixture (Julca et al., 2020), crop-to-wild gene flow in the west (Zunino et al., 2023), and genetic links across Mediterranean lineages (Bazakos et al., 2023; Mariotti et al., 2023), supporting a multi-regional domestication model.

However, data on historical gene flow involving Jordanian germplasm remain limited. This highlights the need for expanded nuclear and plastome analyses and regional phylogeographic studies to clarify Jordan's role in Mediterranean olive diversity. Future work will integrate plastid, nuclear, and morphological data to address this gap.

Integrating nuclear and plastome genotyping for evolutionary, agricultural, and conservation insights

All samples analyzed in this study were previously subjected to nuclear genotyping (Al-Kilani et al., 2024). Nuclear genotyping successfully separated Jordanian centuries-old olive trees into distinct sub-groups and identified 73 previously unreported olive genotypes, 15 of which were used in this study. In contrast, plastome genotyping grouped together some genotypes that had been separated by nuclear markers, while also revealing clear variation among others. The integration of both nuclear and plastid genomic data enables researchers to capitalize on their complementary inheritance patterns, mutation rates, and functional roles, providing a more comprehensive framework for investigating evolutionary relationships, population structure, and adaptive traits. By combining these two genomic approaches, this study not only contributes valuable genomic resources but also

demonstrates the power of integrated genotyping to deepen our understanding of plant genetics and evolutionary dynamics.

One of the key findings that underscores this integrative approach is the enhanced phylogenetic resolution achieved by combining nuclear and plastome genotyping. Each genome offers distinct insights into evolutionary history: nuclear DNA, inherited biparentally and subject to recombination, reflects recent divergence and gene flow, while plastome DNA, with its uniparental inheritance and lower mutation rate, is better suited for tracing deep evolutionary lineages and maternal ancestry. This dual-genome approach thus enables the reconstruction of both recent and ancient evolutionary pathways, which could otherwise remain ambiguous if either genome were studied in isolation.

Furthermore, this study underscores the utility of combined genotyping in elucidating cytoplasmic-nuclear interactions, which play a significant role in plant fitness and adaptability. Cytoplasmic genomes, such as those of chloroplasts, encode essential components of photosynthesis and other metabolic pathways, often interacting with nuclear-encoded factors to produce key adaptive traits. For instance, environmental stress responses and energy metabolism are often the product of complex interactions between nuclear and chloroplast genes. By analyzing both genomes simultaneously, researchers can better understand these interactions and identify specific nuclear-plastid genotype combinations that contribute to environmental resilience. This insight has practical applications in agriculture, particularly in the development of crops with enhanced tolerance to abiotic stresses such as drought or high temperatures (Zhang et al., 2020).

Additionally, the combined approach offers significant advantages for studying population structure and gene flow in plant species.

The evolutionary history and domestication of the olive have been extensively studied using both nuclear and plastid DNA markers. Plastome genotyping, which follows maternal

inheritance, has been instrumental in identifying distinct chloroplast lineages and tracing seed dispersal patterns across the Mediterranean Basin. For example, two main oleaster (wild Mediterranean olive) genepools have been detected in the Western and Eastern Mediterranean, with specific plastid lineages showing geographic differentiation (Besnard et al., 2018).

In contrast, nuclear genotyping, reflecting biparental inheritance, reveals pollen-mediated gene flow and tends to show weaker geographic structure due to recurrent gene flow and metapopulation dynamics influenced by environmental changes. This complementary use of plastid and nuclear markers helps clarify olive biogeography, domestication, and diversification processes (Besnard et al., 2018).

The complementary insights gained from nuclear and plastome marker systems deepen our understanding of olive biogeography, domestication, and diversification. By distinguishing between pollen- and seed-mediated gene flow, researchers can better assess genetic connectivity within populations—a critical factor for effective conservation strategies, particularly in fragmented habitats where gene flow may be limited. Maternal lineages tracing via plastome genotyping further supports the identification and preservation of distinct genetic lineages, helping preserve the adaptive potential of vulnerable olive populations.

Finally, our study demonstrates the broader value of combined nuclear and plastome data for conservation genetics and biodiversity studies. This dual-genome approach ensures the preservation of genetic diversity across both biparental and uniparental lineages, which is essential for maintaining a species' adaptive potential and resilience in the face of environmental changes.

While the limited sample size ($n=15$) poses constraints on the generalizability of our findings, it reflects the realities of working with centuries-old, protected trees and logistical limitations. Nonetheless, these samples were strategically selected to represent key genetic clades identified in previous research (Al-Kilani et al., 2024), ensuring meaningful insight into the Jordanian olive genetic pool. We are actively working to conserve Jordan's ancient olive germplasm through the establishment of a national gene bank. This initiative will support broader sampling and more comprehensive genetic studies, enabling deeper exploration of gene flow and lineage connections between Jordan and other Mediterranean regions.

Conclusion

Overall, plastome sequencing offers a comprehensive approach to enhancing our understanding of olive tree genetics, supporting conservation efforts, and contributing to the agricultural and economic sustainability of olive cultivation.

In conclusion, the application of olive plastomes as barcodes represents a significant advancement in the fields of botany, conservation, and agriculture. By harnessing the power of complete plastid genome sequencing, researchers can enhance species identification, uncover genetic diversity, and support conservation and breeding initiatives. The integration of plastome barcoding into olive research has the potential to transform our

understanding of olive biodiversity and strengthen the sustainability of olive cultivation in a rapidly changing world.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, PQ676988–PQ677002 <https://www.ncbi.nlm.nih.gov/sra/>, PRJNA1182319.

Author contributions

ND: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Visualization. IF: Formal analysis, Investigation, Visualization, Writing – review & editing. AM: Formal analysis, Investigation, Visualization, Writing – review & editing. FT: Data curation, Validation, Writing – review & editing. MA: Data curation, Writing – review & editing. AA: Conceptualization, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, and/or publication of this article.

Acknowledgments

We would like to extend our heartfelt thanks to Francesca Bencivenga for her dedicated work in assembling and annotating the plastomes as part of her Master's thesis and Shireen Hasan for her technical assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer CV declared a shared affiliation with the authors ND'A, IF and AM to the handling editor at the time of review.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1647776/full#supplementary-material>

References

- Al-Kilani, M. A., Taranto, F., D'Agostino, N., Montemurro, C., Belaj, A., Ayoub, S., et al. (2024). Evaluation of genetic diversity among olive trees (*Olea europaea* L.) from Jordan. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1437055
- Amane, M., Ouazzani, N., Lumaret, R., and Debain, C. (2000). 'Chloroplast-DNA variation in the wild and cultivated olives (*Olea europaea* L.) of Morocco'. *Euphytica* 116, 59–64. doi: 10.1023/A:1004025431960
- Arnan, X., López, B. C., Martínez-Vilalta, J., Estorach, M., and Poyatos, R. (2012). 'The age of monumental olive trees (*Olea europaea*) in northeastern Spain'. *Dendrochronologia* 30, 11–14. doi: 10.1016/j.dendro.2011.02.002
- Bazakos, C., Alexiou, K. G., Ramos-Onsins, S., Koubouris, G., Tourvas, N., Xanthopoulou, A., et al. (2023). 'Whole genome scanning of a Mediterranean basin hotspot collection provides new insights into olive tree biodiversity and biology'. *Plant J.* 116, 303–319. doi: 10.1111/tpj.16270
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). 'MISA-web: A web server for microsatellite prediction'. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Benson, G. (1999). 'Tandem repeats finder: A program to analyze DNA sequences'. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Besnard, G., and Bervillé, A. (2002). 'On chloroplast DNA variations in the olive (*Olea europaea* L.) complex: comparison of RFLP and PCR polymorphisms'. *Theor. Appl. Genet.* 104, 1157–1163. doi: 10.1007/s00122-001-0834-8
- Besnard, G., Hernandez, P., Khadari, B., Dorado, G., and Savolainen, V. (2011). 'Genomic profiling of plastid DNA variation in the mediterranean olive tree'. *BMC Plant Biol.* 11, 80. doi: 10.1186/1471-2229-11-80
- Besnard, G., Khadari, B., Baradat, P., and Bervillé, A. (2002a). 'Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.)'. *Theor. Appl. Genet.* 105, 139–144. doi: 10.1007/s00122-002-0868-6
- Besnard, G., Khadari, B., Baradat, P., and Bervillé, A. (2002b). '*Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism'. *Theor. Appl. Genet.* 104, 1353–1361. doi: 10.1007/s00122-001-0832-x
- Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, A., Arrigo, N., et al. (2013). 'The complex history of the olive tree: from late quaternary diversification of Mediterranean lineages to primary domestication in the northern levant'. *Proc. R. Soc. B: Biol. Sci.* 280, 20122833. doi: 10.1098/rspb.2012.2833
- Besnard, G., Terral, J.-F., and Cornille, A. (2018). 'On the origins and domestication of the olive: A review and perspectives'. *Ann. Bot.* 121, 587–885. doi: 10.1093/aob/mcy002
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). 'Trimmomatic: A flexible trimmer for illumina sequence data'. *Bioinformatics* 30, 2114–2205. doi: 10.1093/bioinformatics/btu170
- Chen, H., Chen, H., Wang, B., and Liu, C. (2023). 'Conserved chloroplast genome sequences of the genus *clerodendrum* linn. (Lamiaceae) as a super-barcode'. *PLoS One* 18, e02778095. doi: 10.1371/journal.pone.0277809
- Coissac, E., Hollingsworth, P. M., Laverne, S., and Taberlet, P. (2016). 'From barcodes to genomes: extending the concept of DNA barcoding'. *Mol. Ecol.* 25, 1423–1285. doi: 10.1111/mec.13549
- Dierckx, N., Mardulyn, P., and Smits, G. (2016). 'NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data'. *Nucleic Acids Res.* gkw955. doi: 10.1093/nar/gkw955
- Diez, C. M., Moral, J., Barranco, D., and Rallo, L. (2016). Genetic diversity and conservation of olive genetic resources. *Springer Nature Link*. 337–356. doi: 10.1007/978-3-319-25954-3_10
- Dong, W., Liu, J., Yu, J., Wang, L., and Zhou, S. (2012). 'Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding'. *PLoS One* 7, e350715. doi: 10.1371/journal.pone.0035071
- Dong, W.-P., Sun, J.-H., Liu, Y.-L., Xu, C., Wang, Y.-H., Suo, Z.-L., et al. (2022). 'Phylogenomic relationships and species identification of the olive genus *Olea* (Oleaceae)'. *J. Syst. Evol.* 60, 1263–1805. doi: 10.1111/jse.12802
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). 'Ycf1, the most promising plastid DNA barcode of land plants'. *Sci. Rep.* 5, 83485. doi: 10.1038/srep08348
- Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., et al. (2008). 'Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well'. *PLoS One* 3, e28025. doi: 10.1371/journal.pone.0002802
- Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., and Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biol.* 21, 254. doi: 10.1186/s13059-020-02153-6
- Ganopoulos, I., Bazakos, C., Madesis, P., Kalaitzis, P., and Tsaftaris, A. (2013). Barcode DNA high-resolution melting (Bar- HRM) analysis as a novel close-tubed and

SUPPLEMENTARY TABLE 1

Single-nucleotide polymorphisms (SNPs) identified in the 15 olive chloroplast genomes analyzed, using FARGA (LR743800.1) as the reference plastome. The figure presents SNP alleles per sample and their distribution across genic and intergenic regions of the plastome.

SUPPLEMENTARY TABLE 2

Simple sequence repeats (SSRs) identified in the 15 olive chloroplast genomes analyzed, using FARGA (LR743800.1) as the reference plastome. The table presents SSR size, location, and distribution across genic and intergenic regions of the plastome.

SUPPLEMENTARY TABLE 3

Tandem repeats (TRs) identified in the 15 tomato chloroplast genomes using FARGA (LR743800.1) as the reference plastome. The table presents TR copy number, genomic location, and TR distribution across genic and intergenic regions of the plastome.

SUPPLEMENTARY FIGURE 1

100% stacked bar chart illustrating the allele composition at variable sites in the TF - 3 sample. Each bar represents a distinct variable site, with segments indicating the proportion (%) of reference and alternate alleles.

SUPPLEMENTARY FIGURE 2

A screenshot of the protein alignment comparing the ycf1 gene product in the reference sequence (LR743800.1) and the AJ - 24 sample. The pairwise global alignment reveals a distinct 75-amino acid insertion in the ycf1 protein of the AJ - 24 sample, which is absent from the reference sequence, indicating a potential structural divergence.

accurate tool for olive oil forensic use. *J. Sci. Food Agric.* 93, 2281–2286. doi: 10.1002/jsfa.6040

Gostel, M. R., and Kress, W.J. (2022). 'The expanding role of DNA barcodes: indispensable tools for ecology, evolution, and conservation'. *Diversity* 14, 2135. doi: 10.3390/d14030213

Greiner, S., Lehwork, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Greiner, S., Sobanski, J., and Bock, R. (2015). 'Why are most organelle genomes transmitted maternally?'. *BioEssays* 37, 80–945. doi: 10.1002/bies.201400110

Heinze, B. (2021). Sequencing of complete chloroplast genomes. *Methods Mol Biol* 2222, 89–105. doi: 10.1007/978-1-0716-0997-2_5

Hollingsworth, P. M., Graham, S. W., and Little, D. P. (2011). 'Choosing and using a plant DNA barcode'. *PLoS One* 6, e192545. doi: 10.1371/journal.pone.0019254

Julca, I., Marcet-Houben, M., Cruz, F., Gómez-Garrido, J., Gaut, B. S., Diez, C. M., et al. (2020). Genomic evidence for recurrent genetic admixture during the domestication of Mediterranean olive trees (*Olea europaea* L.). *BMC Biol.* 18, 148. doi: 10.1186/s12915-020-00881-6

Kartzinel, T. R., Hoff, H. K., Divoll, T. J., Littleford-Colquhoun, B. L., Anderson, H., Burak, M. K., et al. (2025). 'Global availability of plant DNA barcodes as genomic resources to support basic and policy-relevant biodiversity research'. *Mol. Ecol.* 34, e17712. doi: 10.1111/mec.17712

Katoh, K., and Standley, D. M. (2013). 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability'. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Krawczyk, K., Nobis, M., Myszczyński, K., Klichowska, E., and Sawicki, J. (2018). 'Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: stipa)'. *Sci. Rep.* 8, 19245. doi: 10.1038/s41598-018-20399-w

Kress, W.J., and Erickson, D. L. (2008). 'DNA barcodes: genes, genomics, and bioinformatics'. *Proc. Natl. Acad. Sci.* 105, 2761–2625. doi: 10.1073/pnas.0800476105

Kumar, S., Kahlon, T., and Chaudhary, S. (2011). 'A rapid screening for adulterants in olive oil using DNA barcodes'. *Food Chem.* 127, 1335–1341. doi: 10.1016/j.foodchem.2011.01.094

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., and Chen, S. (2015). Plant DNA barcoding: from gene to genome. *Biol. Rev.* 90, 157–166. doi: 10.1111/brev.12104

Mariotti, R., Belaj, A., de la Rosa, R., Muleo, R., Cirilli, M., Forgione, I., et al. (2023). 'Genealogical tracing of *Olea europaea* species and pedigree relationships of var. *Europaea* using chloroplast and nuclear markers'. *BMC Plant Biol.* 23, 4525. doi: 10.1186/s12870-023-04440-3

Mariotti, R., Cultrera, N. G. M., Díez, C. M., Baldoni, L., and Rubini, A. (2010). 'Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison'. *BMC Plant Biol.* 10, 2115. doi: 10.1186/1471-2229-10-211

Noman, M., Ajmal, W., Khan, M. R., Shahzad, A., and Ali, G. M. (2015). 'Exploitation of concatenated olive plastome DNA markers for reliable varietal identification for on-farm genetic resource conservation'. *Am. J. Plant Sci.* 06, 3045–3745. doi: 10.4236/ajps.2015.619299

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., et al. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genomics* 2, e000056. doi: 10.1099/mgen.0.000056

Pérez-Jiménez, M., Besnard, G., Dorado, G., and Hernandez, P. (2013). 'Varietal tracing of virgin olive oils based on plastid DNA variation profiling'. *PLoS One* 8, e705075. doi: 10.1371/journal.pone.0070507

Rallo, L., Barranco, D., Díez, C. M., Rallo, P., Suárez, M.P., Trapero, C., et al. (2018). "Strategies for olive (*Olea europaea* L.) breeding: cultivated genetic resources and crossbreeding," in *Advances in Plant Breeding Strategies: Fruits* (Cham, Switzerland: Springer International Publishing), 535–600. doi: 10.1007/978-3-319-91944-7_14

Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). 'Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III'. *Am. J. Bot.* 94, 275–885. doi: 10.3732/ajb.94.3.275

Sion, S., Savoia, M. A., Gadaleta, S., Piarulli, L., Mascio, I., Fanelli, V., et al. (2021). 'How to choose a good marker to analyze the olive germplasm (*Olea europaea* L.) and derived products'. *Genes* 12, 14745. doi: 10.3390/genes12101474

Song, F., Li, T., Burgess, K. S., Feng, Y., and Ge, X.-J. (2020). Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China'. *BMC Plant Biol.* 20, 2615. doi: 10.1186/s12870-020-02466-5

Stamatakis, A. (2014). 'RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies'. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Tamura, K., Stecher, G., and Kumar, S. (2021). 'MEGA11: molecular evolutionary genetics analysis version 11'. *Mol. Biol. Evol.* 38, 3022–3275. doi: 10.1093/molbev/msab120

Tillich, M., Lehwork, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–11. doi: 10.1093/nar/gkx391

Turudić, A., Liber, Z., Grdiša, M., Jakše, J., Varga, F., Šatović, Z., et al. (2021). Towards the well-tempered chloroplast DNA sequences. *Plants* 10 (7), 1360. doi: 10.3390/plants10071360

Zhang, Y., Zhang, A., Li, X., and Lu, C. (2020). 'The role of chloroplast gene expression in plant responses to environmental stress'. *Int. J. Mol. Sci.* 21, 60825. doi: 10.3390/ijms21176082

Zunino, L., Cubry, P., Sarah, G., Mournet, P., Bakkali, A. E., Aqbouch, L., et al. (2023). Genomic evidence of genuine wild versus admixed olive populations evolving in the same natural environments in western Mediterranean Basin. *Evol. Biol.* 19, e0295043. doi: 10.1371/journal.pone.0295043