



The plastid genome of the critically endangered *Valeriana trinervis* (= *Centranthus trinervis*) and insights from comparison with other *Valeriana* plastomes (Caprifoliaceae)

Daniele De Luca¹ · Olga De Castro^{2,3}

Received: 3 March 2025 / Accepted: 27 August 2025 / Published online: 10 September 2025
© The Author(s) 2025

Abstract

Main conclusion The first complete plastid genome of the critically endangered species *Valeriana trinervis* was sequenced, assembled and compared with other published *Valeriana* plastomes.

Abstract In this study, we assembled the plastid genome of the critically endangered, endemic species *Valeriana trinervis* (= *Centranthus trinervis*) and compare it with all published plastomes of *Valeriana*. We found not only differences in the inverted repeats boundaries, in the type and abundance of repeats, but also similarities in codon usage and microsatellite numbers. We detected non-canonical start codons in several genes and identified variation in several regions that could be useful for phylogenetic and phylogeographic studies. The phylogenetic tree inference based on both full plastomes and coding sequence data indicated that *V. trinervis* is sister to all Eurasian *Valeriana* accessions confirming the phylogenetic position recently investigated. This is the first plastome available for a species of the Mediterranean clade of *Valeriana* previously known as *Centranthus*, and it adds further data to understand the evolution and diversification of this systematically debated genus.

Keywords Mediterranean endemic plant · Phylogenetic relationships · Plastome evolution · Whole-Genome Sequencing

Abbreviations

IR	Inverted Repeat
LSC	Large Single Copy
RSCU	Relative Synonymous Codon Usage
SSC	Small Single Copy
SSR	Simple Sequence Repeat

Introduction

The traditional approach to plant molecular phylogenetics involves the amplification and Sanger sequencing of several variable markers (e.g. intergenic spacers, introns or genes) from different taxa (Dong et al. 2012). The plastid genome (plastome) has been always favoured over the nuclear and mitochondrial ones as source of informative molecular markers due to its small genome size (120–160 kb; Xiao-Ming et al. 2017), high copy number per cell (Morley and Nielsen 2016; Sakamoto and Takami 2018), a generally conserved structure and organization (Wicke et al. 2011), and a mutation rate suitable to phylogenies at different taxonomic levels (Gitzendanner et al. 2018). Plastid DNA sequences have been used at small scale to investigate genetic variation and gene flow within and between species (De Castro et al. 2012; Hajiahmadi et al. 2013) and for species identity (Nock et al. 2011; Dong et al. 2012; De Castro et al. 2017), and to a bigger scale to investigate relationships between genera, families, and larger groups (Li et al. 2021; Ulaszewski et al. 2021). Despite these advantages, the failure to find a universal

Communicated by Dorothea Bartels.

✉ Daniele De Luca
daniele.deluca@unisob.na.it; daniele.deluca088@gmail.com

✉ Olga De Castro
odecastr@unina.it; odecastr@gmail.com

¹ Department of Humanities, University of Naples Suor Orsola Benincasa, Via Santa Caterina da Siena 37, 80132 Naples, Italy

² Department of Biology, University of Naples Federico II, Via Cinthia 26, 80126 Naples, Italy

³ Botanical Garden of Naples, University of Naples Federico II, Via Foria 223, 80139 Naples, Italy

barcode for plants or a standard set of gene markers for phylogeographic or phylogenetic studies (Cowan et al. 2006; Roy et al. 2010) has pushed the researchers towards the amplification and sequencing of larger regions (sometimes referred to as “superbarcodes”, e.g. Wu et al. 2021) or obtaining complete plastid genome sequences (Hollingsworth et al. 2016; Ji et al. 2019). Accomplices of this trend were the recent advances in sequencing technology, combined with the reduction of costs associated with high-throughput production and sequencing of DNA libraries (Tonti-Filippini et al. 2017). Although several protocols are available to enrich or amplify plastid DNA from total genomic DNA (Twyford and Ness 2017; Takamatsu et al. 2018) before carrying out high-throughput sequencing, the common procedure is to get several plastid sequences from genome-wide approaches (Dodsworth et al. 2019) or to sequence the whole plant genomes and then using bioinformatics pipelines to assemble the plastome (Freudenthal et al. 2020). In this context, a plastome can be assembled from virtually any plant genome sequencing project with adequate coverage.

In this study, we used the raw sequence data generated from a whole-genome sequencing project aimed at finding microsatellite markers for population genetic analyses in the critically endangered plant *Valeriana trinervis* Viv. (= *Centranthus trinervis* (Viv.) Bég.; Valerianoideae, Caprifoliaceae) (Di Iorio et al. 2024; De Castro et al. 2025a) to assemble and annotate the plastome of this species and compare it with those of other species of *Valeriana* L. Indeed, the phylogenetic history of Valerianoideae (Caprifoliaceae) is problematic, and several taxonomic treatments have been proposed over the years, often involving the lumping of *Centranthus* DC. in *Valeriana* (Christenhusz et al. 2018) or their recognition as distinct genera (e.g. Weberling 1961; Bell and Donoghue 2005). A recent study by De Castro et al. (2025b) aimed to test the monophyly of *Centranthus* and its relationships with *Valeriana* using a comprehensive dataset, implementing Sanger sequencing of the nuclear rRNA cistron and 10 plastid molecular markers. The results obtained identified *Centranthus* as a monophyletic group within *Valeriana* and suggested maintaining it as an infrageneric taxon under *Valeriana*. In fact, considering the position of *Centranthus*, its clade appears to be nested within a broader *Valeriana* with strong support, and given its relationship with *V. longiflora* Wilk., the separation of *Centranthus* from *Valeriana* does not seem well-supported at the present state of knowledge (De Castro et al. 2025b). Therefore, in this study we will use the accepted name of *V. trinervis* over *C. trinervis* (POWO 2025). In light of the above, the present work adds new, full plastome data to better understand the relationships and plastid features in these taxa.

Materials and methods

Plastome assembly and annotation

To characterise the plastome of the critically endangered *V. trinervis*, we used the Whole-Genome Sequencing (WGS) data published by us in Di Iorio et al. (2024) and available at NCBI SRA under the BioProject PRJNA1031832. Further details on sampling, DNA extraction, and library preparation are provided in the same paper. The plastome was de novo assembled using NOVOPlasty v4.3.5 (Dierckxsens et al. 2017; <https://github.com/ndierckx/NOVOPlasty>) by setting the configuration file with the following parameters: genome range = 120,000–180,000, k-mer = 23, 6 Gb of memory, read length = 150, and insert size = 400. The plastome sequence of *V. officinalis* L. isolate B267 (GenBank accession number: NC_045052) was used as seed input. The assembly quality was evaluated as sequencing (coverage) depth (i.e. the number of sequencing reads covering each base position) following the procedure described by Ni et al. (2023), and as assembly coverage (i.e. the number of bases of all mapped reads divided by the number of bases of the reference). The circular plastid genome was first annotated with GeSeq (Tillich et al. 2017) on the CHLOROBX platform (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) at default parameters. The resulting annotations were further checked by comparing them with the results of two other annotation software, PGA (Qu et al. 2019) and CPGAVAS2 (Shi et al. 2019). These three tools differ in the approach used for gene identification, specifically using BLAT-based homology searches (GeSeq), BLAST search approach (PGA), and integrating the results of multiple tools like blastn, blastx, protein2genome, and est2genome (CPGAVAS2). The results of these three tools were annotated, and then tRNA sequences were further verified using the tRNAscan-SE server v2.0 (Chan et al. 2021), and protein-coding sequences by translating the DNA sequences into amino acid sequences in MEGA X (Kumar et al. 2018) using the “Plant plastid” genetic code, followed by a blastp search (Altschul et al. 1997). In case of discordant annotations, we manually verified the start and stop codons, and checked intron and exon boundaries with the help of blast searches against closely related species. A physical map of the annotated plastome of *V. trinervis* was generated using the CPGView tool (Liu et al. 2023), which was also used as further check for annotations.

Comparative analysis of plastid genomes

For comparative analyses of plastomes, we downloaded from NCBI the sequences (in.fasta and.gb formats) and the annotation (in.gff3 format) of the following species

of *Valeriana*, for which complete plastome data were available: *V. dageletiana* Nakai ex F.Maek (PQ280047), *V. fauriei* Briq. (MW123075 and NC_065841), *V. jatamansi* Jones ex Roxb. (NC_067975), and *V. officinalis* (MW788537, NC_045052, and PQ030840). The plastid sequence NC_065841 was from Park et al. (2021), MW788537 from Li et al. (2024a, b), and NC_045052 from Wang et al. (2020a, b); all the others were not linked to any published or in press paper. Multiple accessions for the same species were kept in case of samples from different countries or when it was not possible to locate the samples with precision. The mVISTA program (Frazer et al. 2004) was used to align the above-mentioned plastid sequences in the shuffle-Lagan mode (Brudno et al. 2003) and the output was visualised with the VISTA plot tool (Mayor et al. 2000) using the plastome of *V. trinervis* from this study as reference and the default visualisation settings. To assess the degree of similarity between plastomes, we calculated the pair-wise percentage of identity using blastn (Altschul et al. 1990). Later, a second analysis in mVISTA was carried out to align the plastomes of *V. trinervis* and *Valeriana* spp. to other Valerianoideae for subsequent phylogenetic analyses: *Fedia cornucopiae* (L.) Gaertn. (= *Valeriana cornucopiae* L.) (NC_065839), *Nardostachys jatamansi* (D.Don) DC. (NC_054306), *Patrinia heterophylla* Bunge (NC_045047), and *Valerianella locusta* (L.) Laterr. (= *Valeriana locusta* L.) (NC_065842). The sequences NC_054306, NC_065839, and NC_065842 were from Park et al. (2021), while NC_045047 from Wang et al. (2020a, b).

The CPJSDraw software (Li et al. 2023) was used to visualise the junction sites of *V. trinervis* and other *Valeriana* spp. plastid genomes. Before running the software, to avoid misleading conclusions caused by non-standardised annotations, we compared the annotations of *Valeriana* plastomes downloaded from NCBI with the ones generated by CPGA-VAS2, GeSeq, and PGA. To identify hotspots of variation between the above-mentioned taxa, we calculated the nucleotide diversity (π) in DnaSP v6.12.03 (Rozas et al. 2017) using a window length of 500 and a step size of 250 bp. In addition, π values were also calculated for the LSC (Large Single Copy), SSC (Small Single Copy), and IR (Inverted Repeat) regions.

The codon usage (i.e. the relative frequency of synonymous codons used to encode amino acids) for all protein-coding sequences was calculated by estimating the relative synonymous codon usage (RSCU) values for each codon. This index refers to the ratio of the actual usage frequency of a particular codon to the expected frequency when it is used without bias. A RSCU = 1 indicates no preference for that codon usage pattern, a RSCU > 1 indicates a preference (codons used more frequently than expected), while a RSCU < 1 indicates unpreferred codons. Codons with

RSCU ≥ 1.5 and ≤ 0.5 are considered as high-frequency and low-frequency codons, respectively. The RSCU was calculated in the software MEGA X (Kumar et al. 2018), setting the genetic code to “Plant plastid”, for *V. trinervis* and other *Valeriana* spp.; for the latter taxa, we downloaded the coding sequences (CDSs) from GenBank using the accession numbers.

Characterisation of plastid repeat sequences and simple sequence repeats (SSRs)

The program REPuter (Kurtz et al. 2001; https://bibiserv.cebitec.uni-bielefeld.de/reputer?id=reputer_view_submission) was used to identify and locate forward, reverse, complemented, and reverse complemented repeats in the plastid genomes of *V. trinervis* and its congeners. For repeat identification, the following constraints were set: minimum repeat size of 12 bp, Hamming distance of 3 (which roughly corresponds to sequence identity $\geq 90\%$), and significance scores (E-values) $\geq 1e-05$. Short tandem repeats (STRs) ≤ 6 bp were predicted using MISA (Beier et al. 2017) with the default parameters, i.e. repeat units ≥ 10 for mononucleotide SSRs, ≥ 6 for dinucleotide, and ≥ 5 for tri-, tetra-, penta-, and hexanucleotide SSRs; long tandem repeats (LTRs) ≥ 7 bp were predicted using Tandem Repeats Finder (TRF; Benson 1999) at the following parameters: match weight = 2, mismatch penalty = 7, delta = 7, minimum score to report a repeat (PM) = 80, minimum period size (PI) = 10, maximum period size = 50, maximum overlap between sequences = 500, and -f -d -m options activated.

Phylogenetic inference

To assess the phylogenetic position of *V. trinervis* in respect to other species of *Valeriana* and other members of the subfamily Valerianoideae for which plastome data were available (*Fedia* Adans, *Nardostachys* DC., *Patrinia* Juss., and *Valerianella* Mill.), we carried out a Maximum Likelihood (ML) and Bayesian (BI) tree inference. First, the best evolutionary model for both full plastome data and only CDS data was selected in ModelFinder (Kalyaanamoorthy et al. 2017) implemented in IQ-Tree v2.3.6 (Minh et al. 2020) with the Bayesian Information Criterion (BIC; Stone 1979) at the settings -m MFP + MERGE and -mset mrbayes; a different partition was created for coding and non-coding regions, and for the former the partition scheme took into account the three different codon positions. Then, the ML analysis was carried out in IQ-Tree software using 1000 ultrafast bootstrap replicates (Hoang et al. 2018), while the BI tree inference was done in MrBayes v3.2.7 (Ronquist et al. 2012), running two replica runs and four chains for 1,000,000 generations and sampling chains every 1000 steps. Convergence and effective sample sizes (ESS) for all parameters of BI analysis

were checked in Tracer v1.7 (Rambaut et al. 2018), and the latter was considered acceptable when > 200 . In addition, the first 10% of samples was discarded as burn-in. Phylogenetic trees were visualised using FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree>). The plastome of *P. heterophylla* was chosen as outgroup.

Results

Assembly of the plastome of *V. trinervis* and its characteristics

The total number of reads utilised for the assembly of the plastome was 9,998,176, of which 354,454 were successfully aligned to the assembly graph, and 240,782 were incorporated into the final assembly with an average coverage of $469\times$. The plot of the sequencing depth (Supp. Information S1 – Suppl. Fig. S1) indicated that the aligned reads had a minimum sequencing depth of $353\times$ for the full-length plastome; the average sequencing depth ($7760\times$) was not markedly different from the maximum sequencing depth ($7998\times$), indicating no differences in the coverage of bases with the increase of contig size. The plastome of *V. trinervis* was circularised into a sequence of 156,610 bp, showing the typical quadripartite structure composed by a large single copy region (LSC), a small single-copy region (SSC) and a pair of inverted sequences in the opposite orientation (IRa and IRb) separating the two single-copy regions (Fig. 1). The LSC region was 85,824 bp long, the SSC was 15,476 bp long, and the inverted repeat region (IRa + IRb) was 55,310 bp long. The plastome sequence is available in GenBank under accession number PV026107.

A total of 136 gene regions were annotated, corresponding to 113 unique (not duplicated) genes distributed as follows: 76 protein-coding genes, 4 rRNAs, 28 tRNAs, and 5 genes of unknown function (Table 1). Among the duplicated genes were *ccsA*, *ndhB*, *ndhD*, all the rRNAs (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*), *rps7*, *rps12*, the tRNAs *trnA*-UGC, *trnH*-GUG, *trnI*-GAU, *trnL*-CAA, *trnL*-UAG, *trnN*-GUU, *trnR*-ACG, *trnV*-GAC, and the genes with unknown function/ open reading frames *ycf2* and *ycf15*; the tRNA-coding genes *trnM*-CAU and *trnT*-GGU were found in three copies. Introns were annotated in 17 genes: 15 of them contained one intron, while two (*clpP* and *ycf3*) two introns. The *rps12* gene, encoding the ribosomal small subunit 12, was transcribed into three exons, of which one separated by around 28 kb from the other two and with a complement region far around 81 kb; two copies were annotated in the plastome, one of 345 bp and another one of 369 bp. The shorter copy (8 aa less) ended with the unusual triplet AGT. We exclude mis-annotations because the following 24 bp or 8 aa had no similarity with the full sequences of other *Valeriana* spp. or

the other copy in *V. trinervis*. The *ndhD* gene was duplicated and presented a non-canonical starting codon in both copies: ACG instead of AUG. In addition, the two copies were different in size (912 bp and 1509 bp).

The investigation of codon usage preferences showed a high preference ($RSCU \geq 1.5$) for the following 12 codons: AAU(N), ACU(T), AGA(R), CAU(H), GAU(D), GCU(A), GGA(G), GUU(V), UAA(*), UAU(Y), UCU(S), and UUA(L). Instead, 11 codons were less abundant, indicating negative codon usage bias ($RSCU \leq 0.5$): AAC(N), ACG(T), AGC(S), CGC(R), CGG(R), CUC(L), CUG(L), GAC(D), GCG(A), GGC(G), and UAC(Y). Two codons, AUG(M) and UGG(W), were considered neutral, being used approximately equally to all synonymous codons (Supp. Information S2–Suppl. Table S1).

The repeat elements detected beyond the chosen threshold were 646, of which 558 (86.4%) as forward repeats, 86 (13.3%) as palindromic repeats, and 2 (0.3%) as reverse repeats. Regarding microsatellite regions, we found 39 SSRs, of which five in compound formation, mostly localised in the LSC (19) and SSC (10) regions. Only five repeats (four mono- and one dinucleotides) were observed in each IR region (Supp. Information S2–Suppl. Table S2). The mono-repeat A/T was the most abundant repetition, accounting for the 87.2% (34/39) of all microsatellites (Supp. Information S2–Suppl. Table S3), followed by the dinucleotide repeats AT/AT (7.7%; 3/39), and the mononucleotide repeats G/C and AC/GT (both 2.6%; 1/39). Most of cpSSRs were located in intergenic regions, except for six that were found within genes (*rpoB*, *rpoC1*, *ycf1*, and *ycf2*) and other six within introns (in *atpF*, in the second intron of *clpP*, in *rpl16*, *rpoC1*, *trnG*-UCC, and *trnI*-GAU).

Comparative genomics analysis with other *Valeriana* spp.

The plastome of *V. trinervis* was 569 bp longer than the longest plastome of *V. officinalis* (PQ030840) and ~6 kb longer than the shortest plastome (*V. officinalis* voucher SN0759, MW788537) (Table 2). Such length differences mostly occurred in conserved non-coding sequences (CNSs) (Supp. Information S1 – Suppl. Fig. S2). The total GC content of the analysed plastomes was comparable across the taxa investigated, setting at 38%. Some differences for GC% were observed in the SSC and IR regions, with *V. officinalis* vouchers B267 and SN0759 showing a higher GC content for IRs and SSC and IRs, respectively (Table 2). In terms of overall similarity, the plastome of *V. trinervis* was identical at the 96–97% to the other *Valeriana* plastomes (Supp. Information S2 – Suppl. Table S4).

The mVISTA alignment of *V. trinervis* and the other *Valeriana* plastomes (Supp. Information S1–Suppl. Fig. S2, and Supp. Information S3–Suppl. Dataset S1) showed an overall

Valeriana trinervis

plastid genome

156,610 bp

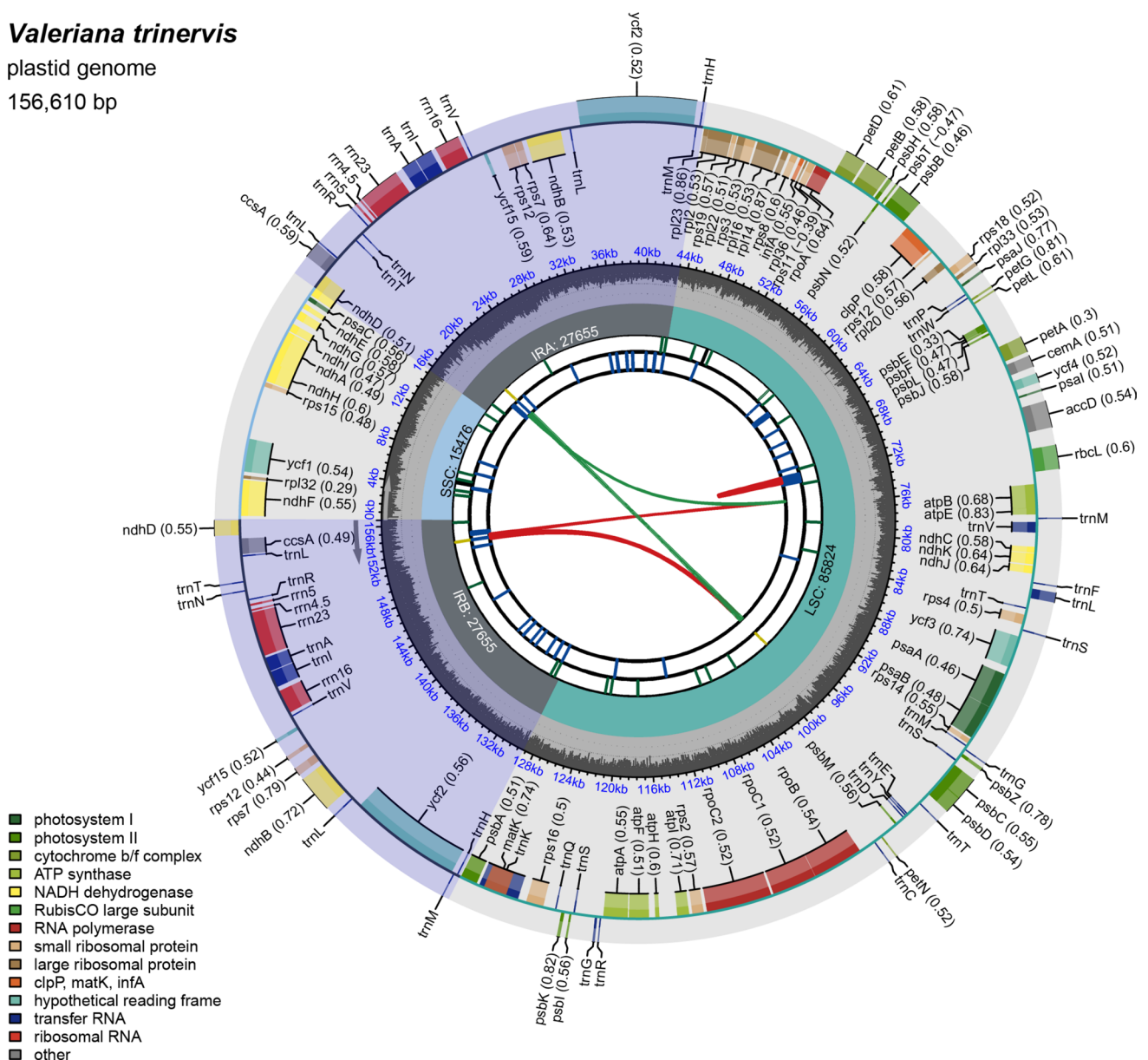


Fig. 1 Map of the plastid genome of *Valeriana trinervis* (Caprifoliaceae). In the six tracks from the centre outward are indicated: (i) dispersed repeats: in green palindromic and in red direct repeats; (ii) long tandem repeats (short blue bars); (iii) short tandem repeats

or microsatellite sequences (short bars in black=complex repeats; green=mono-repeats; yellow=di-repeats); (iv) small single-copy (SSC), inverted repeat (IRA and IRB), and large single-copy (LSC) regions; (v) GC content along the genome; (vi) genes

conserved structure, with most of the variation present not only in non-coding regions (e.g. *accD-rbcL*, *clpP* introns, *ndhI-ndhG*, *psbJ-petA*, *trnT-GGU-trnL-UAG*, *trnV-UAC-ndhC*, *ycf1-rps15*) but also in coding ones (e.g. *accD*, *petD*, *rpoC1*, *rps18*, and *ycf1*).

Our check of *Valeriana* annotations following the procedure described in Materials and Methods enabled to properly identify non-canonical start codons for the *ndhD* gene and to annotate this gene in the plastome of *V. jatamansi*. Annotations in.gb format from CPGAVAS2 are available in

Supp. Information S3 – Suppl. Datasets S2–S8. All *Valeriana* plastomes from literature contained two copies of the *ndhD* gene, except those of *V. officinalis* vouchers B267 and SN0759, which presented just one copy. As for *V. trinervis*, the two copies of *ndhD* of other *Valeriana* spp. had different length (Supp. Information S1 – Suppl. Fig. S3). The map of IR boundaries (Supp. Information S1 – Suppl. Fig. S3) showed a rather conserved gene order at the inverted repeats boundaries of *V. trinervis* and most of *Valeriana* spp. here considered; however, some exceptions also occurred. For

Table 1 List of the genes annotated in the plastome of *Valeriana trinervis*. *Gene with one intron; **Gene with two introns; (×n) indicates the number of copies

Functional category	Localization/ Function	Name	Total
Photosynthesis	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>	6
	Cytochrome b/f complex	<i>petA</i> , <i>petB*</i> , <i>petD*</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
	NADH dehydrogenase	<i>ndhA*</i> , <i>ndhB*</i> (×2), <i>ndhC</i> , <i>ndhD</i> (×2), <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	11
	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>	5
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbG</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	16
	Rubisco large subunit	<i>rbcL</i>	1
Self-replication	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>	4
	Ribosomal RNAs	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)	4
	Ribosome large subunits	<i>rpl2*</i> , <i>rpl14</i> , <i>rpl16*</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> , <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	9
	Ribosome small subunits	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12*</i> (×2), <i>rps14</i> , <i>rps15</i> , <i>rps16*</i> , <i>rps18</i> , <i>rps19</i>	12
	Transfer RNAs	<i>trnA</i> -UGC* (×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -GCC, <i>trnG</i> -UCC*, <i>trnH</i> -GUG (×2), <i>trnI</i> -GAU* (×2), <i>trnK</i> -UUU, <i>trnL</i> -CAA (×2), <i>trnL</i> -UAA*, <i>trnL</i> -UAG* (×2), <i>trnM</i> -CAU (×3), <i>trnN</i> -GUU (×2), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG (×2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU (×3), <i>trnT</i> -UGU, <i>trnV</i> -GAC (×2), <i>trnV</i> -UAC*, <i>trnW</i> -CCA, <i>trnY</i> -GUA	28
Other genes	Acetyl-CoA carboxylase	<i>accD</i>	1
	C-type cytochrome synthesis gene	<i>ccsA</i> (×2)	1
	Envelope membrane protein	<i>cemA</i>	1
	Maturase	<i>matK</i>	1
	Protease	<i>clpP**</i>	1
	Translation initiation factor	<i>infA</i>	1
	Hypothetical open reading frames	<i>ycf1</i> , <i>ycf2</i> (×2), <i>ycf3**</i> , <i>ycf4</i> , <i>ycf15</i> (×2)	5

instance, the LSC/IRb boundary (JLB) was found between *psbA* and *trnH* genes in all taxa except *V. officinalis* B267, in which *trnM* replaced *trnH*. In *V. officinalis* accession B267, *psbA* was 482 bp distant from the beginning of IRb against the 61–76 bp distance observed in the other taxa. On the other side of the boundary and in the same accession, *trnM* was 218 bp distant from the beginning of IRb, while in all the other taxa *trnH* was 122–159 bp distant (Suppl. Information S1 – Suppl. Fig. S3). The junction SB (IRb-SSC) was characterised by *ndhD* and *ndhF* genes in all taxa except for two accessions of *V. officinalis*, B267 and SN0759, where *trnN* was found instead of *ndhD*. The latter gene was found within the JSB in *V. dageletiana*, the two accessions of *V. fauriei* and in one accession (PQ030840) of *V. officinalis*; in *V. trinervis*, *ndhD* was located at 79 bp from the JSB, while in *V. jatamansi* it was exactly at the boundary. The highest distance between a gene and the end of IRb was observed in *V. officinalis* B267 and SN0759 (269 bp and 432 bp, respectively). In the junction SSC-IRa (JSA) the genes *psaC* and *ndhD* were present in all taxa except for the two above-mentioned accessions of *V. officinalis*, which had *trnN* instead of *ndhD*. At this junction, *ndhD* was within the boundary in all the taxa in which it was present except

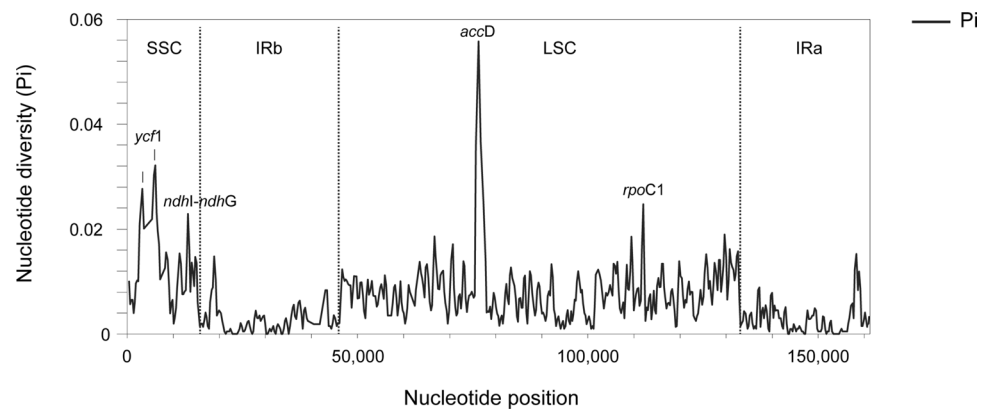
for *V. fauriei* PDBK, where it was located at 150 bp from the junction, within the IRa. The gene *psaC* was at only 337 bp from the beginning of IRa, against the 528–654 bp distance observed in the other taxa. Finally, the IRa-LSC boundary (JLA) was between *trnH* and *rpl23* in *V. trinervis* and all *Valeriana* accessions excluding *V. officinalis* B267, where *trnM* replaced *trnH*, and it was located at the highest distance (218 bp against 122–159 bp) from the beginning of LSC in respect to other taxa (Suppl. Information S1 – Suppl. Fig. S3). Regarding expansions and contractions of IRs, no relevant differences were generally observed between *V. trinervis* (55,154 bp) and most of *Valeriana* taxa (53,662–55,600) except for two accessions of *V. officinalis*, B267 and SN0759, which showed a marked reduction in size, with 47,598 bp and 46,726 bp, respectively (Table 2, and Suppl. Information S1 – Suppl. Fig. S3).

The graph of nucleotide diversity (Pi) analysis carried out in DnaSP software (Fig. 2) indicated that the regions with the highest mean Pi values were the SSC (Pi=0.012) and LSC (0.008), respectively; the IR region had a nucleotide diversity of 0.002. Within these regions, the genes/non-coding sequences with the highest Pi values were *accD*, *ndhI*-*ndhG*, *rpoC1*, and *ycf1*.

Table 2 Characteristics of the plastome of *Valeriana trinervis* and comparisons with other *Valeriana* spp. plastomes. N.a. stands for “not available”; *, according to POWO (2025)

	<i>V. trinervis</i>	<i>V. dageletiana</i>	<i>V. fauriei</i>	<i>V. fauriei</i> PDBK	<i>V. jatamansi</i>	<i>V. officinalis</i>	<i>V. officinalis</i> B267	<i>V. officinalis</i> SN0759
GenBank accession number	PV026107 (this study)	PQ280047	NC_065841	MW123075	NC_067975	PQ030840	NC_045052	MW788537
Geographic location	Corsica (France)	South Korea	n.a.	n.a.	China	Kazakhstan	China	China
Native geographic distribution*	Corsica (France)	Europe	Russian Far East to temperate East Asia	Russian Far East to temperate East Asia	Himalaya to W. & Central China and N. Myanmar	Europe to NW. Iran	Europe to NW. Iran	Europe to NW. Iran
Total length (bp)	156,610	155,311	155,302	154,432	155,138	156,041	151,505	150,612
%GC	38.15	38.34	38.35	38.36	38.34	38.26	30.45	38.49
SSC length (bp)	15,476	15,195	15,159	15,199	15,108	15,250	16,288	18,339
SSC %GC	32.67	32.6	32.59	32.65	32.65	32.56	32.53	33.1
LSC length (bp)	85,824	85,478	85,541	85,295	86,368	85,191	87,619	85,547
LSC %GC	36.55	36.71	36.73	36.72	36.74	36.66	36.54	36.76
IRs length (bp)	55,310	54,728	54,602	53,938	53,662	55,600	47,598	46,726
IR %GC	42.2	42.47	42.5	42.56	42.5	42.25	43.97	43.77

Fig. 2 Nucleotide diversity of *Valeriana trinervis* and other *Valeriana* plastomes



The codon usage in *V. trinervis* was comparable to other *Valeriana* species (Supp. Information S1 – Suppl. Fig. S4). The codons GCU, UCU, and UUA, respectively corresponding to the amino acids alanine, serine, and leucine, were the ones with the highest codon bias (approaching RSCU = 2). On the contrary, the codons ACG (serine), CAC (histidine), CGC (arginine), CUG and CUC (leucine), and UAC (tyrosine) showed negative bias (were less used). The preferred stop codon was UAA (RSCU ~ 1.5), while the canonical start codon AUG (methionine) showed no usage bias (Supp. Information S1 – Suppl. Fig. S4). Overall, approximately the 47% of codons showed preferences in usage (RSCU > 1).

Non-canonical start codons were observed in the *ndhD*, *psbC*, *psbL*, and *rps19* genes. Specifically, the *ndhD* gene presented the codon ACG (threonine) instead of ATG (methionine) as start codon in all the taxa analysed except for *V. officinalis* voucher SN0759, which presented GTG as start codon; this gene was not annotated in the plastome of *V. jatamansi*. The *psbC* gene showed the non-canonical start codon GTG only in *V. officinalis* accession PQ030840. In all the other taxa it started with ATG (methionine). The start codon for *psbL* was ACG instead of ATG in all the taxa except *V. officinalis* voucher SN0759 (TTG); it was not annotated in *V. jatamansi*. Finally, the *rps19* gene presented

the canonical start codon ATG only in *V. trinervis* and *V. jatamansi*, but was replaced by GTG in all the other taxa.

Regarding repeated regions, the plastome of *V. trinervis* contained the highest proportion of forward repeats (86.4%)

and the lowest of palindromic repeats (13.3%) (Fig. 3a). Within the other *Valeriana* species there was high heterogeneity in the proportion of different classes of repeat elements, which was generally not consistent with taxonomy.

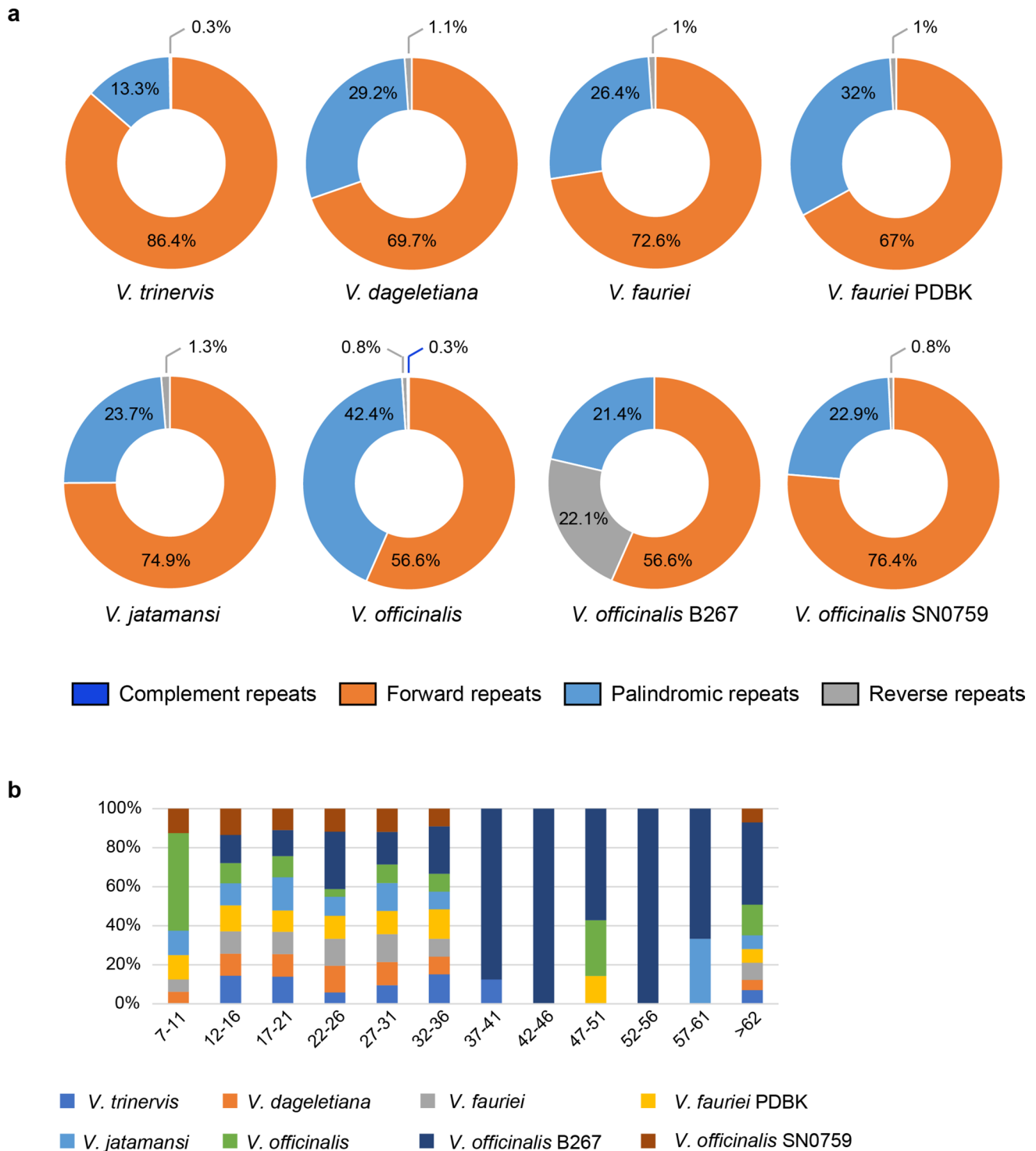


Fig. 3 Repeat elements in the plastomes of *Valeriana trinervis* and other *Valeriana* species. **a** Complement, forward, palindromic, and reverse repeats. **b** Long tandem repeats (> 7 bp)

This was particularly evident in the *V. officinalis* group, where reverse repeats were 0.3% and 0.8% in and *V. officinalis*, respectively, and 22.1% in, while palindromic repeats were 21.4% and 22.9% in *V. officinalis* voucher B267 and voucher SN0759, and 42.4% in *V. officinalis* (Fig. 3a). The sizes of long tandem repeats (LTRs) common to all taxa ranged from 12 to 36 bp (Fig. 3b, Suppl. Information S2 – Suppl. Tables S5–12). *Valeriana officinalis* B267 lacked LTRs < 12 bp, but was the only taxon having them in all the size ranges considered, including exclusive sizes (e.g. for repeat classes of 42–46 bp and 52–56 bp). LTRs in the sizes between 37 and 61 bp were present in a few taxa like *V. trinervis* (37–41), *V. fauriei* and *V. officinalis* (47–51 bp), *V. jatamansi* (57–61 bp), and the above-mentioned *V. officinalis* B267. Long tandem repeats > 62 bp were present in all taxa but in scattered and mostly non-overlapping size classes (Suppl. Information S2 – Suppl. Tables S5–12).

The number of microsatellite repeats (SSRs) in *V. trinervis* (39) was identical to *V. fauriei* voucher PDBK (39), and not very different from *V. officinalis* (38) and *V. dageletiana*, *V. fauriei*, and *V. officinalis* SN0759 (38); the lowest number of SSRs was observed in *V. officinalis* B267 (30) (Table 3). The motif A/T was the most recurrent in all the plastid genomes here investigated, with a length in the range of 10–17 repeats in all the taxa except *V. officinalis*, where the longest repeat was of 19 bp. *Valeriana trinervis* was the only species showing the AC/GT dinucleotide motif, while the tri-nucleotide motif AAT/ATT was only found in *V. officinalis* B267 and presented the longest repetition range (24–31). The G/C mono-repeat was found only in four taxa and was shortest in *V. trinervis* (10 bp against 12–14 in the other taxa). The AT/AT dinucleotide repeat, common to all taxa, showed the longest repetition range in *V. trinervis* (6–10 against the common 6–7 range of other *Valeriana* species).

The topology of Maximum likelihood and Bayesian phylogenetic trees was identical, as well as the topology of these trees when considering the full plastome sequences (tree on the left in Fig. 4) and the CDS-based trees (tree on the right in the same figure). All nodes were supported with a 100% bootstrap value or 1 posterior probability, except for the node

connecting *V. fauriei* voucher PDBK (MW123075) and *V. officinalis* voucher SN0759 (MW788537), which had a bootstrap support of 88 in the plastome-based ML tree and 94 in the CDS-based one. Whatever the methodology and dataset used, *V. trinervis* was sister to all the other *Valeriana* spp. with maximum support (Fig. 4). The samples of *V. officinalis* and *V. fauriei* turned out to be not monophyletic.

Discussion

In this study, we assembled, annotated, and deposited the plastid genome of the critically endangered species *V. trinervis*, a taxon previously included in the problematic *Centranthus* s.l. group (De Castro et al. 2025b). This species is found exclusively in a single location in Corsica, and has an estimated population size of less than 150 individuals (Revaka et al. 2012–2017). The plastome was fully assembled from raw sequence data originally produced to obtain a genome-wide set of nuclear microsatellite markers for population genetics studies in this species and other congeners (Di Iorio et al. 2024; De Castro et al. 2025a). By subsampling the 4.56% of reads present in the raw data, NOVOPlasty was capable of circularising a plastome with an assembly coverage of 469× using 6 Gb of RAM; the sequencing depth was overall high (average value of 7760×). In line with available literature on plastome assembly strategies (reviewed in Twyford and Ness 2017) we confirmed that these “untargeted” whole-genome sequencing (WGS) approaches can be suitable for plastome assembly, providing that data have adequate depth and that did not undergo size selection with specific restriction enzymes (e.g. in GBS or RAD approaches).

Centranthus is a problematic genus into an even more problematic family, the Valerianaceae, whose taxonomic validity is currently debated in favour of Caprifoliaceae (APG 2016; POWO 2025). The most recent accepted treatment (Christenhusz et al. 2018) sees the inclusion of some Valerianoideae (*Centranthus*, *Fedia*, *Valeriana*, and *Valerianella*) except *Patrinia* and *Nardostachys* into a large *Valeriana* “super-genus” (De Castro et al. 2025b). As outlined in

Table 3 Number of cpSSRs in the plastome of *Valeriana trinervis* and of the others *Valeriana* spp. The estimates consider sequence complementarity. In parentheses is reported the minimum and maximum repeat length

	Tot SSRs	A/T	G/C	AC/GT	AT/AT	AAT/ATT
<i>V. trinervis</i>	39	34 (10–17)	1 (10)	1 (8)	3 (6–10)	0
<i>V. dageletiana</i>	37	32 (10–17)	2 (12–14)	0	3 (6–7)	0
<i>V. fauriei</i>	37	32 (10–17)	2 (12–14)	0	3 (6–7)	0
<i>V. fauriei</i> PDBK	39	36 (10–17)	0	0	3 (6–7)	0
<i>V. jatamansi</i>	36	34 (10–17)	1 (12)	0	1 (8)	0
<i>V. officinalis</i>	38	36 (10–19)	0	0	2 (6–7)	0
<i>V. officinalis</i> B267	30	25 (10–17)	0	0	3 (6–7)	2 (24–31)
<i>V. officinalis</i> SN0759	37	34 (10–17)	0	0	3 (6–7)	0

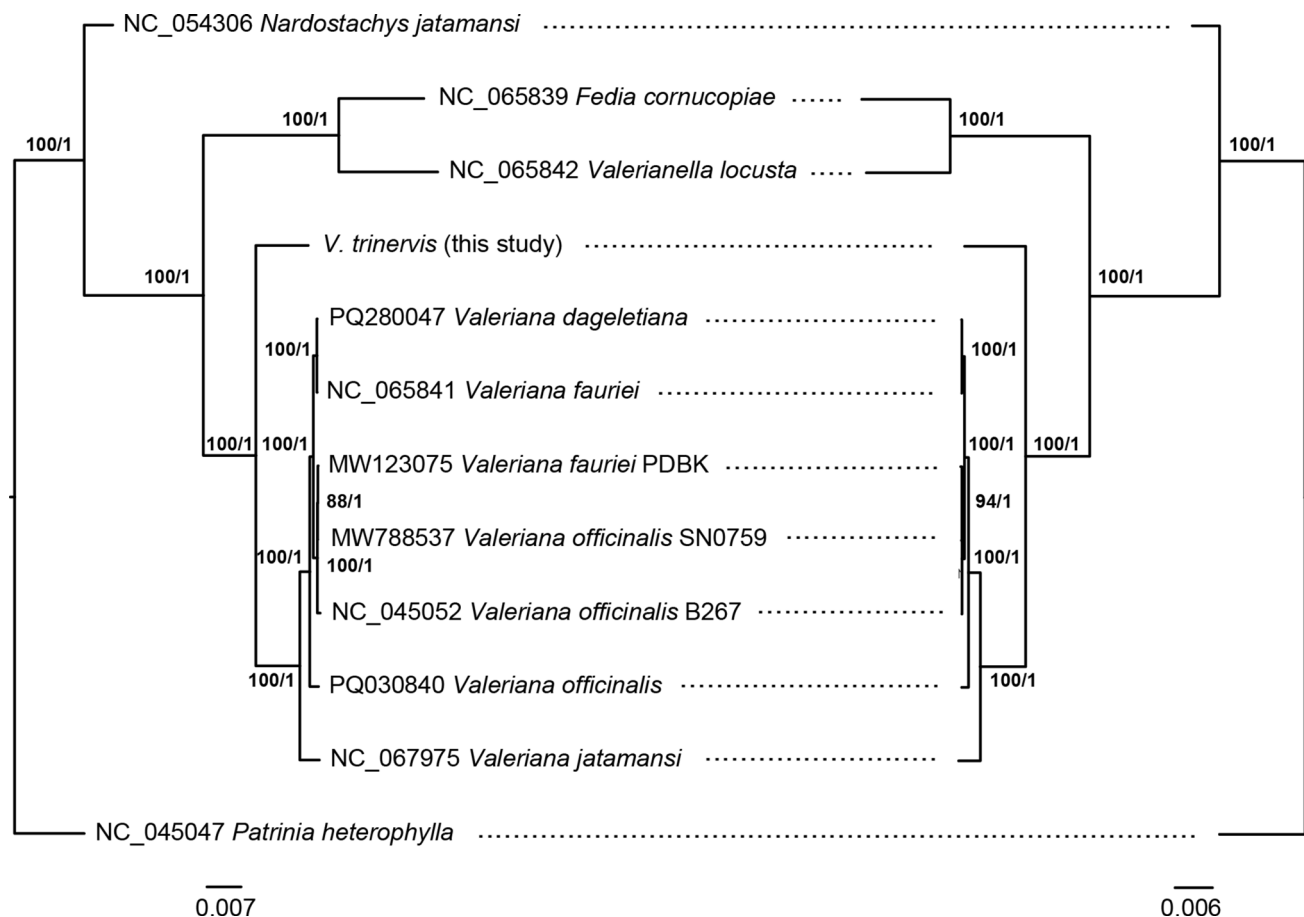


Fig. 4 Maximum likelihood and Bayesian phylogenetic trees based on the full plastome (left) and coding sequences data (right). Numbers at nodes refer indicate support to that relationship in terms of bootstrap (before /) and posterior probability (after /)

De Castro et al. (2025b), this choice is premature on account of the current phylogenetic evidence, which would benefit from a wider taxon sampling, and considering the distinct morphological traits of some members. All the phylogenetic studies produced so far (e.g. Hidalgo et al. 2004, 2010; Bell et al. 2012, 2015) agree in retrieving *Patrinia* and, in turn, *Nardostachys*, as sister to all the other Valerianoideae; in turn, *Fedia* and *Valerianella* are sister to *Centranthus* and *Valeriana*. On the contrary, recent (De Castro et al. 2025b) and past (Hidalgo et al. 2004, 2010; Bell et al. 2015) phylogenetic studies support *Centranthus* as synonym of *Valeriana*. As consequence, we considered *C. trinervis* as any other *Valeriana*, under the name *V. trinervis*.

The full plastome and CDS-based Bayesian and Maximum likelihood trees here inferred confirm this topology with high support (Fig. 4) and retrieve *V. trinervis* as sister to the Eurasian *Valeriana* species for which plastome sequences were available. Regarding the relationships among the other *Valeriana* species, for the purposes of this study we highlight that there is no monophyly for *V.*

fauriei and *V. officinalis* samples; these differences also reflect in other characteristics other than sequence differences (e.g. repeat type, see below) and could be due to real differences in samples or erroneous taxonomic determination. Indeed, other studies (Fujii et al. 2021; Bączek et al. 2022) have found intraspecific diversity in these taxa, that could be explained with the wide geographic distribution. From a phylogenetic perspective, we cannot argue more about the relationships among *Valeriana* plastomes because most of taxa retrieved from literature and included in this study (*V. fauriei*, *V. jatamansi*, and *V. officinalis*) are all in the so called “Eurasian group” (Hidalgo et al. 2004; Bell and Donoghue 2005; De Castro et al. 2025b) and representatives of other groups across Europe and America are missing. In this sense, our study represents a first step towards a better comprehension of plastome diversity in *Valeriana sensu lato*. About *V. trinervis*, the position in the phylogenetic trees is a preliminary but good representation of some distinctive features observed in respect to the plastomes of other *Valeriana* species.

Repeat elements

Repeat sequences play an important role in generating variation in plastid genomes through point mutations, non-homologous recombination, and DNA polymerase slippage during replication (Ellegren 2002; George et al. 2015). Therefore, they are a valuable and highly utilised source of molecular markers for phylogeography and population genetics studies (e.g. Bucci et al. 2007; Semerikova et al. 2021). In our plastome comparisons, we found a marked difference in the type and abundance of repeat elements. The plastome of *V. trinervis* contains the highest number of forward repeats (86.4%) and the lowest of palindromic repeats (13.3%). On the contrary, in the other *Valeriana* spp. forward repeats ranged from 56.6% to 76.4%, and palindromic repeats from 21.4% to 42.4% and conspecific samples were not comparable in these terms in accordance with the position in the phylogenetic trees. Regarding short tandem repeats, the difference between the analysed plastomes were less marked. The number of total SSRs (39) was in line with *Valeriana* taxa (36–39) except *V. officinalis* voucher B267 that presented the lowest number of SSRs (30) and a unique tri-nucleotide motif (AAT/ATT); instead, *V. trinervis* exhibited a unique AC/GT repeat motif. Similarly, long tandem repeats were comparable between *V. trinervis* and the other *Valeriana* spp. in the size range between 12 and 36 bp, but differences occurred for smaller and larger sizes. *Valeriana officinalis* B267 had unique LTRs (e.g. 42–46 and 52–56 bp) and, generally, larger sizes of repeats not observed in any other taxon. Compared to other Caprifoliaceae (e.g. *Lonicera* L. spp., *Weigela florida* (Bunge) A.DC.), which presented repeat units mainly in the range of 21–50 bp (Fan et al. 2018), we observed smaller LTRs (7–36 bp) in our target species. However, there is no common trend when comparing the number of repeats among congeneric species across plant families. For instance, in *Primula* L. (Primulaceae; Ren et al. 2018) and in *Ligustrum* L. (Oleaceae; Long et al. 2023), high differences in the number of long repeat sequences have been observed, while these differences were neglectable in *Bupleurum* L. (Apiaceae; Huang et al. 2021) and in *Wikstroemia* Spreng (Thymelaeaceae; Zhang et al. 2025).

Codon usage and non-canonical start codons

Codon usage bias, i.e. the non-random use of synonymous codons within a coding sequence, is a universal feature of genomes originating from a balance between mutational patterns and natural selection (Atthey et al. 2017). It plays an important role in a multitude of transcriptional and translational processes by affecting chromatin structure and mRNA folding (Quax et al. 2015; Liu 2020). Codon bias varies not only within and among species, genera and families, but also between genes and sites within a gene into a specific

organism (Parvathy et al. 2022). In our study, 29 codons showed RSCU > 1, indicating a preference in their usage in respect to synonymous codons. This number is line with values (29–32 biased codons) found in other studies considering taxa within the same genus (Chen et al. 2024; Liu et al. 2024) or family (Shi et al. 2023; Li et al. 2024a, b). In addition, the codon preferences of *V. trinervis* and other *Valeriana* spp. were quite similar, suggesting a close relationship between them, as observed in some congeneric taxa (Jiang et al. 2023; Sudmoon et al. 2024).

Some plastid genes also presented non-canonical start codons, a feature observed in several groups from algae to angiosperms (Parvathy et al. 2022; Sadhu et al. 2023). The *ndhD* gene represents an interesting and well-studied case of non-canonical translation initiation, where an ACG codon replaces the canonical one AUG in many species (Martín and Sabater 2010). To carry out the translation processes, in some species *ndhD* transcripts undergo RNA editing to convert ACG to the canonical AUG start codon, while in others the translation starts from the ACG codon itself. For instance, in *Nicotiana tabacum* L. and *N. sylvestris* Speg. partial editing occurs, while in *N. tomentosiformis* Goodsp. the ACG codon remains completely unedited without hampering the normal translation (Zanduetta-Criado and Bock 2004). We found an ACG starting codon in the *ndhD* gene of *V. trinervis* but in none of the other *Valeriana* plastomes available in NCBI; however, we attribute this result to a likely mis-annotation rather than a real absence of non-canonical starting codons. Indeed, after re-annotating the deposited *Valeriana* plastomes with CPGAVAS2, we found non-canonical start codons in the *ndhD* genes of all taxa (see Supp. Information S3 – Suppl. Datasets S2–S8). Therefore, we recommend comparing different annotation tools before making conclusions. We detected also differences in the length of the two copies of *ndhD* in *V. trinervis* and all the other *Valeriana* taxa with two copies (all except *V. officinalis* B267 and SN0759). All the annotations of *ndhD* obtained with CPGAVAS2 were consistent with the gene boundaries found in other species and checked with BLAST searches. In addition, the translation of the nucleotide sequences before and after the detected boundaries contained stop codons or were not recognised as *ndhD* in BLAST searches. Therefore, we are confident that the annotations are correct and this discrepancy could be the result of ongoing pseudogenisation, but further evidence is needed to confirm or not this hypothesis. In addition to *ndhD*, other genes showing non-canonical start codons were *psbC*, *psbL*, and *rps19*. These findings were corroborated by some literature studies (e.g. Kuroda et al. 2007; Chen et al. 2018; Su et al. 2018; Zheng et al. 2019; Liew et al. 2022), but also reported for other genes too (e.g. Wang et al. 2020a,

b; Pei et al. 2022). In general, the preferred start codon was ACG over ATG. The only exceptions were GTG in the *ndhD* gene of *V. officinalis* voucher SN0759, in the *psbC* gene of *V. officinalis* accession PQ030840 and in all *Valeriana* plastomes except *V. jatamansi* and *V. trinervis* for *rps19*, and TTG in the *psbL* of *V. officinalis* voucher SN0759. The reasons behind this non-canonical ATG start for some plastid genes is not fully understood yet, but it is known that the ACG codon can be converted into an AUG start codon through RNA editing (Neckermann et al. 1994). In addition, it has been reported that both ACG and GTG triplets can initiate translation with a strength varying from 15 to 30% of AUG activity, and the relative efficiency of ACG is lower than GUG (Mehdi et al. 1990; Su et al. 2018). We also found a non-canonical stop codon (CTA) for the second copy of *rps12*. As for *ndhD*, this is not explainable with mis-annotation because the translation of nucleotides after this point is not recognised as *rps12* by BLAST searches and comparisons with other *Valeriana* sequences. Therefore, this could be a case of pseudogenisation too. To the best of our knowledge, no cases of AGT as stop codon have been reported in other plants.

Hotspots of variation

One of the reasons behind the assembly of plastid genomes and their comparisons with other species is to find variable regions to be used for phylogenetic and phylogeographic studies. Due to the stable genetic structure, high copy number, absence of (or very rare) recombination, and uniparentally transmission, the plastid genome has been and still is the preferred target for evolutionary studies in plants (Dobrogojski et al. 2020). Several fragments of coding regions, introns, and intergenic spacers, like *atpB-rbcL*, *matK*, *rbcL*, *rps16*, *trnH-psbA*, *trnQ-rps16*, *trnL-F*, *trnS-G*, etc., have been used in phylogenetic and phylogeographic studies (e.g. De Castro et al. 2012, 2020; De Luca et al. 2022, 2024), and some of them (e.g. *matK*, *rbcL*, *trnH-psbA*) have been proposed as markers for plant DNA barcoding (CBOL Plant Working Group 2009; Hollingsworth et al. 2016; De Castro et al. 2017). Through the alignment of *V. trinervis* and the other *Valeriana* plastomes (Fig. 2), we identified several potential markers for phylogenetic studies: the genes *accD*, *petD*, *rpoC1*, *rps18*, and *ycf1*; the intergenic spacers *accD-rbcL*, *ndhI-ndhG*, *psbJ-petA*, and *trnV-UAC-ndhC*; and *ycf1-rps15* and the *clpP* introns, some of which have already been implemented in De Castro et al. 2025a, b (a, b) for *Centranthus* phylogeny and the phylogeography of *Centranthus* taxa belonging to sect. *Nervosae*. Variation in these regions was both in terms of single nucleotide polymorphisms and

indels and could be particularly useful for future studies in this complex family.

Conclusions

In this study, we assembled the complete plastid genome sequence of the critically endangered *V. trinervis* by Illumina sequencing of WGS data. After annotating and reporting its features, we compared it with the plastome of other *Valeriana* species and accessions to identify differences and similarities in gene content, repeat regions, and codon usage. Regarding repeated regions, the plastome of *V. trinervis* contained the highest proportion of forward repeats and the lowest of palindromic repeats compared to *Valeriana* plastomes. On the contrary, the number of microsatellite repeats in *V. trinervis* was similar to all the other *Valeriana* taxa except *V. officinalis* voucher B267. Codon usage was comparable across all plastid coding genes. Some non-canonical start codons were observed and not shared by all taxa. From the phylogenetic point of view, the plastid sequence of *V. trinervis* was sister to all Eurasian *Valeriana* spp. In addition, several markers presented high levels of variation that make them useful candidates for future phylogenetic studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00425-025-04815-w>.

Author contribution DDL conceived and designed the study with the support of ODC. DDL performed the analyses and interpreted the data. ODC obtained the funding. DDL wrote the first draft of the manuscript. ODC curated the draft of the manuscript. All authors have read, revised, and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Data availability The complete plastid genome of *V. trinervis* is available in NCBI GenBank under the accession number PV026107. All the data not published in the text are available as supplementary material on the publisher's website.

Declarations

Conflict of interest. The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- APG. The Angiosperm Phylogeny Group. Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn* 181:1–20. <https://doi.org/10.1111/boj.12385>
- Atthey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C (2017) A new and updated resource for codon usage tables. *BMC Bioinformatics* 18:391. <https://doi.org/10.1186/s12859-017-1793-7>
- Bączek KB, Kosakowska O, Boczkowska M, Bolc P, Chmielecki R, Pióro-Jabucka E, Raj K, Węglarz Z (2022) Intraspecific variability of wild-growing common valerian (*Valeriana officinalis* L.). *Plants* 11:3455. <https://doi.org/10.3390/plants11243455>
- Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
- Bell CD, Donoghue MJ (2005) Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. *Org Divers Evol* 5:147–159. <https://doi.org/10.1016/j.ode.2004.10.014>
- Bell CD, Kutschker A, Arroyo MT (2012) Phylogeny and diversification of Valerianaceae (Dipsacales) in the southern Andes. *Mol Phylogenet Evol* 63:724–737. <https://doi.org/10.1016/j.ympev.2012.02.015>
- Bell CD, Calderon G, Gonzalez L, Scholz A, Liede-Schumann S (2015) Resolving relationships within Valerianaceae (Dipsacales): new insights and hypotheses from low-copy nuclear regions. *Syst Bot* 40:327–335. <https://doi.org/10.1600/036364415X686611>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>
- Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S (2003) Glocal alignment: finding rearrangements during alignment. *Bioinformatics* 19(S1):i54–i62. <https://doi.org/10.1093/bioinformatics/btg1005>
- Bucci G, González-Martínez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alfá R, Vendramin GG (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol Ecol* 16:2137–2153. <https://doi.org/10.1111/j.1365-294X.2007.03275.x>
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 106:12794–12797. <https://doi.org/10.1073/pnas.0905845106>
- Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* 49:9077–9096
- Chen N, Sha LN, Dong ZZ, Tang C, Wang Y, Kang HY, Zhang HQ, Yan XB, Zhou YH, Fan X (2018) Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. *Gene* 640:86–96. <https://doi.org/10.1016/j.gene.2017.10.009>
- Chen H, Zhang X, Zhang G, Zhang Z, Ma G, Sun Z, Liu C, Huang L (2024) The complete chloroplast genome sequence of *Nepeta bracteata* and comparison with congeneric species. *Gene* 893:147919. <https://doi.org/10.1016/j.gene.2023.147919>
- Christenhusz MJM, Fay MF, Byng JM (2018) The global flora. GLOVAP Nomenclature Part 1, vol. 4. Plant Gateway Ltd, Bradford
- Cowan RS, Chase MW, Kress WJ, Savolainen V (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55:611–616. <https://doi.org/10.2307/25065638>
- De Castro O, Brullo S, Colombo P, Jury S, De Luca P, Di Maio A (2012) Phylogenetic and biogeographical inferences for *Pancratium* (Amaryllidaceae), with an emphasis on the Mediterranean species based on plastid sequence data. *Bot J Linn Soc* 170:12–28. <https://doi.org/10.1111/j.1095-8339.2012.01268.x>
- De Castro O, Comparone M, Di Maio A, Del Guacchio E, Menale B, Troisi J, Aliberti F, Trifuoggi M, Guida M (2017) What is in your cup of tea? DNA verity test to characterize black and green commercial teas. *PLoS ONE* 12:e0178262. <https://doi.org/10.1371/journal.pone.0178262>
- De Castro O, Innangi M, Menale M (2020) Message in a bottle: the Mediterranean Sea currents acted as protagonists in shaping the distribution of the sea daffodil (*Pancratium maritimum*, Amaryllidaceae). *Bot J Linn Soc* 194:207–220. <https://doi.org/10.1093/botlinnean/boaa037>
- De Castro O, Menale B, Piazza C, Bacchetta G, Del Guacchio (2025b) Can we rescue *Centranthus* (Caprifoliaceae, Valerianoideae) from the *Valeriana* sea? *Bot J Linn Soc*: boaf019. <https://doi.org/10.1093/botlinnean/boaf019>
- De Castro O, Piazza C, Di Iorio E, Bacchetta G, Menale B (2025a) Impact of barriers on Cyno-Sardinian endemisms: a comparative study of population genetics and phylogeography within taxa of *Centranthus* sect. *Nervosae* (Caprifoliaceae). *Plant Biol* 27:362–377. <https://doi.org/10.1111/plb.13775>
- De Luca D, Del Guacchio E, Conti F, Iamónico D, Caputo P (2022) Relationships within *Mcneillia* indicate a complex evolutionary history and reveal a new species of *Minuartiella* (Caryophyllaceae, Alsinoideae). *Plants* 11:2118. <https://doi.org/10.3390/plants11162118>
- De Luca D, Del Guacchio E, Cennamo P, Minutillo F, Bernardo L, Caputo P (2024) Genetics and distribution of the Italian endemic *Campanula fragilis* Cirillo (Campanulaceae). *Plants* 13:3169. <https://doi.org/10.3390/plants13223169>
- Di Iorio E, De Luca D, Menale B, Piazza C, Bacchetta G, De Castro O (2024) Genome-wide characterization of nuclear microsatellite markers for *Centranthus trinervis* (= *Valeriana trinervis*) and cross-amplification in five congeneric taxa. *Plant Mol Biol Rep* 42:743–751. <https://doi.org/10.1007/s11105-024-01454-5>
- Dierckx N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18–e18. <https://doi.org/10.1093/nar/gkw955>
- Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review. *Acta Physiol Plant* 42:98. <https://doi.org/10.1007/s11738-020-03089-x>
- Dodsworth S, Pokorny L, Johnson MG, Kim JT, Maurin O, Wickett NJ, Forest F, Baker WJ (2019) Hyb-seq for flowering plant systematics. *Trends Plant Sci* 24:887–891. <https://doi.org/10.1016/j.tplants.2019.07.011>
- Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7:e35071. <https://doi.org/10.1371/journal.pone.0035071>
- Ellegren H (2002) Microsatellite evolution: a battle between replication slippage and point mutation. *Trends Genet* 18:70. [https://doi.org/10.1016/S0168-9525\(02\)02631-8](https://doi.org/10.1016/S0168-9525(02)02631-8)

- Fan WB, Wu Y, Yang J, Shahzad K, Li ZH (2018) Comparative chloroplast genomics of Dipsacales species: insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front Plant Sci* 9:689. <https://doi.org/10.3389/fpls.2018.00689>
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32(S2):W273–W279. <https://doi.org/10.1093/nar/gkh458>
- Freudenthal JA, Pfaff S, Terhoeven N, Korte A, Ankenbrand MJ, Förster F (2020) A systematic comparison of chloroplast genome assembly tools. *Genome Biol* 21:254. <https://doi.org/10.1186/s13059-020-02153-6>
- Fujii T, Mori T, Tatsuo Y, Takao Y, Fujino H, Tsuchida T, Minami M (2021) Identification of *Valeriana fauriei* and other Eurasian medicinal valerian by DNA polymorphisms in *psbA-trnH* intergenic spacer sequences in chloroplast DNA. *J Nat Med* 75:699–706. <https://doi.org/10.1007/s11418-021-01488-z>
- George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr Genet* 61:665–677. <https://doi.org/10.1007/s00294-015-0495-9>
- Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE (2018) Plastome phylogenetics: 30 years of inferences into plant evolution. *Adv Bot Res* 85:293–313. <https://doi.org/10.1016/bs.abr.2017.11.016>
- Hajiahmadi Z, Talebi M, Sayed-Tabatabaei BE (2013) Studying genetic variability of pomegranate (*Punica granatum* L.) based on chloroplast DNA and barcode genes. *Mol Biotechnol* 55:249–259. <https://doi.org/10.1007/s12033-013-9676-2>
- Hidalgo O, Garnatje T, Susanna A, Mathez J (2004) Phylogeny of Valerianaceae based on *matK* and ITS markers, with reference to *matK* individual polymorphism. *Ann Bot* 93:283–293. <https://doi.org/10.1093/aob/mch042>
- Hidalgo O, Mathez J, Garcia S, Garnatje T, Pellicer J, Valles J (2010) Genome size study in the Valerianaceae: first results and new hypotheses. *J Bot* 2010:797246. <https://doi.org/10.1155/2010/797246>
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2018) UFboot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
- Hollingsworth PM, Li DZ, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. *Philos Trans R Soc Lond B Biol Sci* 371:20150338. <https://doi.org/10.1098/rstb.2015.0338>
- Huang R, Xie X, Li F, Tian E, Chao Z (2021) Chloroplast genomes of two Mediterranean *Bupleurum* species and the phylogenetic relationship inferred from combined analysis with East Asian species. *Planta* 253:1–17. <https://doi.org/10.1007/s00425-021-03602-7>
- Ji Y, Liu C, Yang Z, Yang L, He Z, Wang H, Yang J, Yi T (2019) Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Mol Ecol Res* 19:1333–1345. <https://doi.org/10.1111/1755-0998.13050>
- Jiang D, Cai X, Gong M, Xia M, Xing H, Dong S, Tian S, Li J, Lin J, Liu Y, Li HL (2023) Complete chloroplast genomes provide insights into evolution and phylogeny of *Zingiber* (Zingiberaceae). *BMC Genomics* 24:30. <https://doi.org/10.1186/s12864-023-09115-9>
- Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kuroda H, Suzuki H, Kusumegi T, Hirose T, Yukawa Y, Sugiura M (2007) Translation of *psbC* mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine-Dalgarno sequence in tobacco chloroplasts. *Plant Cell Physiol* 48:1374–1378. <https://doi.org/10.1093/pcp/pcm097>
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) Reputer: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642. <https://doi.org/10.1093/nar/29.22.4633>
- Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, Jin J-J, Zeng C-X, Wang H, Yu W-B, Zhang R, van der Bank M, Olmstead RG, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Yi T-S, Li DZ (2021) Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol* 19:232. <https://doi.org/10.1186/s12915-021-01166-2>
- Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X (2023) CPJSDraw: analysis and visualization of junction sites of chloroplast genomes. *PeerJ* 11:e15326. <https://doi.org/10.7717/peerj.15326>
- Li Y, Sun Z, Jin Z, Li J (2024a) Genome sequencing provides insights into Caprifoliaceae genome evolution and the mechanism underlying second blooming phenomenon in *Heptacodium miconioides*. *Hortic Plant J* 10:897–910. <https://doi.org/10.1016/j.hpj.2022.11.008>
- Li DM, Pan YG, Liu HL, Yu B, Huang D, Zhu GF (2024b) Thirteen complete chloroplast genomes of the Costaceae family: insights into genome structure, selective pressure and phylogenetic relationships. *BMC Genomics* 25:68. <https://doi.org/10.1186/s12864-024-09996-4>
- Liew YJM, Chua KO, Yong HS, Song SL, Chan KG (2022) Complete chloroplast genome of *Boesenbergia rotunda* and a comparative analysis with members of the family Zingiberaceae. *Rev Bras Bot* 45:1209–1222. <https://doi.org/10.1007/s40415-022-00845-w>
- Liu Y (2020) A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun Signal* 18:145. <https://doi.org/10.1186/s12964-020-00642-6>
- 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* 23:694–704. <https://doi.org/10.1111/1755-0998.13729>
- Liu L, Li H, Li J, Li X, Hu N, Sun J, Zhou W (2024) Chloroplast genomes of *Caragana tibetica* and *Caragana turkestanica*: structures and comparative analysis. *BMC Plant Biol* 24:254. <https://doi.org/10.1186/s12870-024-04979-9>
- Long L, Li Y, Wang S, Liu Z, Wang J, Yang M (2023) Complete chloroplast genomes and comparative analysis of *Ligustrum* species. *Sci Rep* 13:212. <https://doi.org/10.1038/s41598-022-26884-7>
- Martín M, Sabater B (2010) Plastid *ndh* genes in plant evolution. *Plant Physiol Biochem* 48:636–645. <https://doi.org/10.1016/j.plaphy.2010.04.009>
- Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16:1046–1047. <https://doi.org/10.1093/bioinformatics/16.11.1046>
- Mehdi H, Ono E, Gupta KC (1990) Initiation of translation at CUG, GUG, and ACG codons in mammalian cells. *Gene* 91:173–178. [https://doi.org/10.1016/0378-1119\(90\)90085-6](https://doi.org/10.1016/0378-1119(90)90085-6)
- 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* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>

- Morley SA, Nielsen BL (2016) Chloroplast DNA copy number changes during plant development in organelle DNA polymerase mutants. *Front Plant Sci* 7:57. <https://doi.org/10.3389/fpls.2016.00057>
- Neckermann K, Zeltz P, Igloi GL, Kössel H, Maier RM (1994) The role of RNA editing in conservation of start codons in chloroplast genomes. *Gene* 146:177–182. [https://doi.org/10.1016/0378-1119\(94\)90290-9](https://doi.org/10.1016/0378-1119(94)90290-9)
- Ni Y, Li J, Zhang C, Liu C (2023) Generating sequencing depth and coverage map for organelle genomes. *protocols.io*. <https://doi.org/10.17504/protocols.io.4r3l27jkg1y/v1>
- Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J* 9:328–333. <https://doi.org/10.1111/j.1467-7652.2010.00558.x>
- Park S, Jun M, Park S, Park S (2021) Lineage-specific variation in IR boundary shift events, inversions, and substitution rates among Caprifoliaceae s.l. (Dipsacales) plastomes. *Int J Mol Sci* 22:10485. <https://doi.org/10.3390/ijms221910485>
- Parvathy ST, Udayasuriyan V, Bhadana V (2022) Codon usage bias. *Mol Biol Rep* 49:539–565. <https://doi.org/10.1007/s11033-021-06749-4>
- Pei J, Wang Y, Zhuo J, Gao H, Vasupalli N, Hou D, Lin X (2022) Complete chloroplast genome features of *Dendrocalamus farinosus* and its comparison and evolutionary analysis with other Bambusoideae species. *Genes* 13:1519. <https://doi.org/10.3390/genes13091519>
- POWO (2025) Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <https://powo.science.kew.org/>. Retrieved 21 February 2025
- Qu XJ, Moore MJ, Li DZ, Yi TS (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:50. <https://doi.org/10.1186/s13007-019-0435-7>
- Quax TE, Claassens NJ, Söhl D, van der Oost J (2015) Codon bias as a means to fine-tune gene expression. *Mol Cell* 59:149–161. <https://doi.org/10.1016/j.molcel.2015.05.035>
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ren T, Yang Y, Zhou T, Liu ZL (2018) Comparative plastid genomes of *Primula* species: sequence divergence and phylogenetic relationships. *Int J Mol Sci* 19(4):1050. <https://doi.org/10.3390/ijms19041050>
- Revaka M-A, Piazza C, Hugot L (2012–2017) Plan national d’actions en faveur de *Centranthus trinervis* (Viv.) Bég. - Centranthe à trois nervures -2012–2017. Conservatoire Botanique National de Corse (CBNC). Available from: https://www.ecologie.gouv.fr/sites/default/files/PNA_Centranthe-a-3-nervures_2013-2017.pdf. Accessed February 2025
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
- Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Datt B, Bag SK, Singh PK, Nair NK, Husain T, Tuli R (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian berberis species. *PLoS ONE* 5:e13674. <https://doi.org/10.1371/journal.pone.0013674>
- Rozas J, Ferrer-Mata A, Sánchez-Del Barrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sadhu L, Kumar K, Kumar S, Dass A, Pathak R, Bhardwaj A, Pandey P, Van Cuu N, Rawat BS, Reddy VS (2023) Chloroplasts evolved an additional layer of translational regulation based on non-AUG start codons for proteins with different turnover rates. *Sci Rep* 13:896. <https://doi.org/10.1038/s41598-022-27347-9>
- Sakamoto W, Takami T (2018) Chloroplast DNA dynamics: copy number, quality control and degradation. *Plant Cell Physiol* 59:1120–1127. <https://doi.org/10.1093/pcp/pcy084>
- Semerikova SA, Isakov IY, Semerikov VL (2021) Chloroplast DNA variation and phylogeography of pedunculate oak *Quercus robur* L. in the eastern part of the range. *Russ J Genet* 57:47–60. <https://doi.org/10.1134/S1022795421010130>
- Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res* 47(W1):W65–W73. <https://doi.org/10.1093/nar/gkz345>
- Shi W, Hu S, Song W, Huang Y, Shi C, Wang S (2023) Uncovering the first complete chloroplast genomics, comparative analysis, and phylogenetic relationships of the medicinal plants *Rhamnus cathartica* and *Frangula alnus* (Rhamnaceae). *Physiol Mol Biol Plants* 29:855–869. <https://doi.org/10.1007/s12298-023-01331-7>
- Stone M (1979) Comments on model selection criteria of Akaike and Schwarz. *JR Stat Soc B* 41:276–278. <https://doi.org/10.1111/j.2517-6161.1979.tb01084.x>
- Su Y, Huang L, Wang Z, Wang T (2018) Comparative chloroplast genomics between the invasive weed *Mikania micrantha* and its indigenous congener *Mikania cordata*: structure variation, identification of highly divergent regions, divergence time estimation, and phylogenetic analysis. *Mol Phylogenet Evol* 126:181–195. <https://doi.org/10.1016/j.ympev.2018.04.015>
- Sudmoon R, Kaewdaungdee S, Ho HX, Lee SY, Tanee T, Chaveerach A (2024) The chloroplast genome sequences of *Ipomoea alba* and *I. obscura* (Convolvulaceae): genome comparison and phylogenetic analysis. *Sci Rep* 14:14078. <https://doi.org/10.1038/s41598-024-64879-8>
- Takamatsu T, Baslam M, Inomata T, Oikawa K, Itoh K et al (2018) Optimized method of extracting rice chloroplast DNA for high-quality plastome resequencing and de novo assembly. *Front Plant Sci* 9:266. <https://doi.org/10.3389/fpls.2018.00266>
- 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. <https://doi.org/10.1093/nar/gkx391>
- Tonti-Filippini J, Nevill PG, Dixon K, Small I (2017) What can we do with 1000 plastid genomes? *Plant J* 90:808–818. <https://doi.org/10.1111/tpj.13491>
- Twyford AD, Ness RW (2017) Strategies for complete plastid genome sequencing. *Mol Ecol Resour* 17:858–868. <https://doi.org/10.1111/1755-0998.12626>
- Ulaszewski B, Jankowska-Wróblewska S, Świło K, Burczyk J (2021) Phylogeny of *Maleae* (Rosaceae) based on complete chloroplast genomes supports the distinction of *Aria*, *Chamaemespilus* and *Torminalis* as separate genera, different from *Sorbus* sp. *Plants* 10:2534. <https://doi.org/10.3390/plants10112534>
- Wang HX, Liu H, Moore MJ, Landrein S, Liu B, Zhu ZX, Wang HF (2020a) Plastid phylogenomic insights into the evolution of the Caprifoliaceae s.l. (Dipsacales). *Mol Phylogenet Evol* 142:106641. <https://doi.org/10.1016/j.ympev.2019.106641>
- Wang H, Park SY, Song SH, San ML, Kim YC, Ham TH, Kim DY, Kim TS, Lee J, Kwon SW (2020b) Analysis of complete chloroplast genome sequence of Korean landrace *Cymbidium goeringii*. *3 Biotech* 10:1–11
- Weberling F (1961) Die infloreszenzen der Valerianaceen und ihre systematische Bedeutung. Akademie der Wissenschaften und der Literatur, Mainz. *Abh Akad Wiss Lit Mainz, Math.-Naturw.* 1(5): 155–281
- Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land

- plants: gene content, gene order, gene function. *Plant Mol Biol* 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4>
- Wu L, Wu M, Cui N, Xiang L, Li Y, Li X, Chen S (2021) Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chin Med* 16:52. <https://doi.org/10.1186/s13020-021-00460-z>
- Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, Jing L, Yan S, Weihua Q, Lifang Z, Yunlian C, Qingwen Y (2017) Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep* 7:1555. <https://doi.org/10.1038/s41598-017-01518-5>
- Zanduetta-Criado A, Bock R (2004) Surprising features of plastid *ndhD* transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon. *Nucleic Acids Res* 32:542–550. <https://doi.org/10.1093/nar/gkh217>
- Zhang C, Li J, Yan F, Wang Z, Zeng X, Zhang J (2025) Comparative analysis of the complete chloroplast genome of seven *Wikstroemia* taxa (Thymelaeaceae) provides insights into the genome structure and phylogenetic relationships. *Planta* 261:40. <https://doi.org/10.1007/s00425-025-04611-6>
- Zheng X, Ren C, Huang S, Li J, Zhao Y (2019) Structure and features of the complete chloroplast genome of *Melastoma dodecandrum*. *Physiol Mol Biol Plants* 25:1043–1054. <https://doi.org/10.1007/s12298-019-00651-x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.