Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complete chloroplast genome of Ethiopian endemic Aloe weloensis: comparative genomics and phylogenetic placement

Belay Melese Wolde
Wuhan Botanical Garden
Shuai Peng
Wuhan Botanical Garden
Caifei Zhang
Wuhan Botanical Garden
Emmanuel Nyongesa Waswa
Wuhan Botanical Garden
Alazar Fasika Munushe
Wuhan Botanical Garden
Jacinta Katunge Kawenze
Wuhan Botanical Garden
Johar Zemede Belay
Wuhan Botanical Garden
Tegenu Mekuria Tesema
Wuhan Botanical Garden
Sebsebe Demissew
Addis Ababa University
Qing-feng Wang
Wuhan Botanical Garden
Guang-Wan Hu
guangwanhu@wbgcas.cn

Wuhan Botanical Garden

Research Article

Keywords: Aloe weloensis, Chloroplast genome, Phylogenomics, IR expansion, RNA editing, Hypervariable loci

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11123752/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Complete chloroplast genome of Ethiopian endemic *Aloe weloensis*: comparative genomics and phylogenetic placement

Belay Melese Wolde[a,b,c], Shuai Peng[a,b,c], Caifei Zhang[a,b,c,h], Emmanuel Nyongesa Waswa[a,b,c], Alazar Fasika Munushe[a,b,c], Jacinta Katunge Kawenze[a,b,c], Johar Zemede Belay[a,b,c], Tegenu Mekuria Tesema[a,b,c], Sebsebe Demissew[e,f], Qing-feng Wang[a,b,c,h], and Guang-Wan Hu[a,b,c,h,g]*

[a.] *CAS Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China*.

[b.] *Sino-Africa Joint Research Centre, Chinese Academy of Sciences, Wuhan 430074, China.*

[c.] *University of Chinese Academy of Sciences, Beijing 100049, China.*

[e]. *Addis Ababa University, P.O. Box 3434, Addis Ababa, Ethiopia.*

[f.] *National Herbarium (ETH), P.O. Box 3434, Addis Abeba, Ethiopia.*

[g.] *Hubei Jiangxia Laboratory, Wuhan, 430200, China.*

h. Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, China

*Corresponding Author. E-mail: guangwanhu@wbgcas.cn

Abstract

Aloe weloensis is a narrowly distributed, habitat-threatened endemic species of the Ethiopian Welo highlands. This species has been utilized as a source of medicine for treating diverse ailments, such as ear infections, headaches, rheumatism, stomachaches, wounds, and malaria. However, lack of molecular data has prevented the accurate identification and refinement of taxonomic relationships within *Aloe* L. In this study, we sequenced, assembled and annotated the chloroplast genome of *A. weloensis.* The plastome was 153,001 bp in size, and exhibited a typical quadripartite structure with a Large Single-Copy region (LSC = 83,613 bp), a Small Single-Copy region (SSC = 16,192 bp), and a pair of Inverted Repeats (IR = 26,598 bp). The genome encoded 130 genes encompassing (84 protein-coding genes, 38 tRNAs, and 8 rRNAs). Comparative analysis within the ten studied *Aloe* species revealed a highly conserved plastome architecture. The highest divergence was observed in three coding genes (*matK, psbM, and rpl22*) and three non-coding intergenic regions (*petN_psbM, petD_rpoA, and trnQ-UUG_psbK*), which are proposed as candidate markers for species identification and population studies. In addition, analysis of Simple Sequence Repeats (SSRs) and Long Sequence Repeats (LSRs) revealed 1,022 and 540 repeats, respectively, which also serve as valuable genetic markers that promote genomic variation and rearrangements. The RNA editing prediction identified 734 candidate C-to-U and U-to-C events across 62 sites, most often shifting residues from hydrophilic to hydrophobic. Phylogenomic reconstruction based on 64 protein-coding genes across 21 taxa placed *A. weloensis* within a well-supported clade (BS = 94, PP = 1) alongside *A. perfoliata* and *A. marlothii*, consistent with monophyletic *Aloe*. This study provides valuable genomic resources and candidate molecular markers for conservation genetics, future evolutionary studies, development of molecular markers, and breeding programs.

Keywords: *Aloe weloensis*, Chloroplast genome, Phylogenomics, IR expansion, RNA editing, Hypervariable loci

Introduction

The family Asphodelaceae comprises over 700 species in 15 genera, including the charismatic leaf succulent *Aloe* genus and related alooid genera (Nyffeler & Eggli, 2010). *Aloe* comprises about 591 plant species (Woudstra et al., 2025). It is characterized by rosette-forming, fleshy leaves and tubular flowers, with growth forms ranging from stemless herbs to small trees (Chase et al., 2016; Edwards et al., 1997). All species are native to Africa, Madagascar, the Arabian Peninsula, and the Indian Ocean islands (Bachman et al., 2020). Ethiopia represents a probable centre of diversity, hosting 46 species and three subspecies, 67.3% of which are endemic, distributed across three centres of endemism: the northern and central highlands (15 species), the eastern highlands and lowlands (7 species), and the southern highlands, lowlands, and Rift Valley (11 species) (Belayneh et al., 2020; Demissew et al., 2001; Hedberg & Edwards, 1995) (Figure 1, Table S1). According to the IUCN Red List, six Ethiopian species: *A. harlana* and *A. yavellana* (Endangered), *A. tewoldei* and *A. pubescens* (Near Threatened), and *A. retrospiciens* and *A. rugosifolia* (Vulnerable), have been assessed, while many others remain data-deficient (IUCN, 2025).

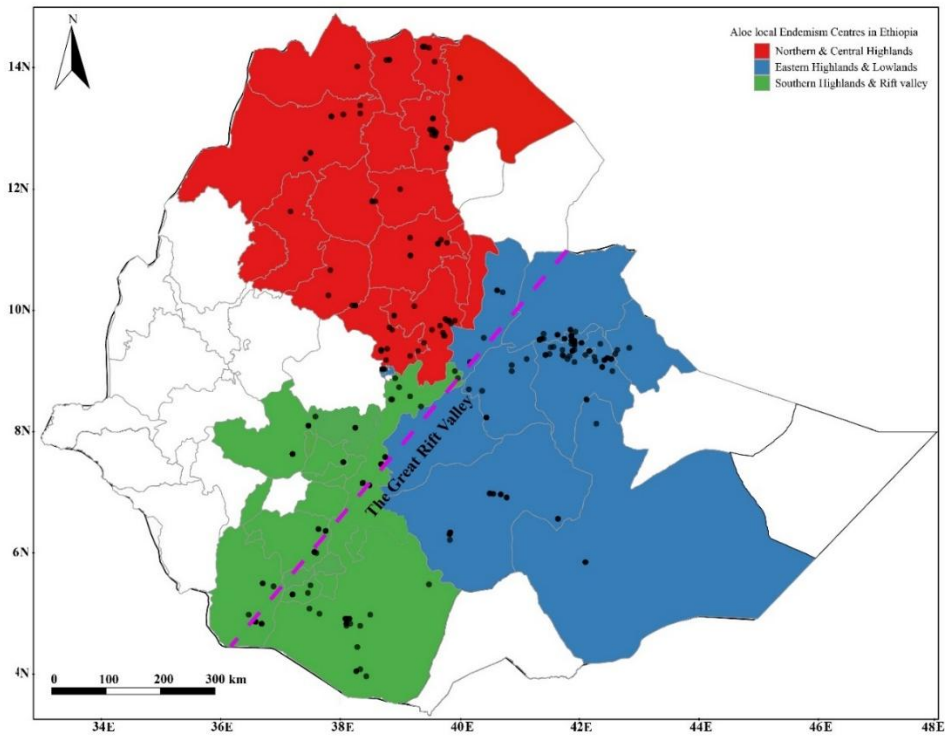


Figure 1. The three local centres of *Aloe* endemism recognised in Ethiopia (Geographic Data Source: GBIF).

Beyond their ecological significance, *Aloe* species are highly valued for ornamental and medicinal applications, driven by bioactive compounds including alkaloids, anthraquinones, and flavonoids (Cock, 2015; Kurniawan et al., 2025). Across their native range, they are widely used to treat infections, digestive disorders, and wounds (Belayneh et al., 2020; Bjora et al., 2015; Grace et al., 2008; Kumar et al., 2017). Species, such as *A. vera*, have demonstrated additional anti-inflammatory, antidiabetic, anticancer, and antimicrobial properties (Sánchez et al., 2020). Despite the extensive ethnobotanical importance, genetic resources for the genus remain limited, leaving the molecular basis for these traits largely unexplored. Genomic studies are essential in clarifying these mechanisms and guiding breeding efforts. For instance, the *A. vera* genome, has revealed drought-adaptive strategies, including crassulacean acid metabolism (CAM) photosynthesis and circadian regulation (Jaiswal et al., 2021), while plastome studies in other taxa have provided resolved phylogenies and crop improvement (Hu et al., 2022; Talkah et al., 2024).

Among the narrowly distributed Ethiopian endemic species is *Aloe weloensis* Sebsebe, a perennial succulent characterized by a solitary or clump-forming, acaulescent (stemless) habit. It forms dense rosettes of thin, soft, spotted leaves that range from vibrant green to deep green, growing up to 45 cm long and exuding yellow latex when cut (Demissew et al., 2011). Mature plants bear marginal spines and produce a slender, pinkish-red inflorescence that can reach 150 cm in height, with ovate bracts and cylindrical racemes bearing bright pinkish-red perianths; the fruit is a capsule approximately 20 mm long (Figure 2). Locally known as Eret tafa, it has documented antioxidant, antibacterial, and antimalarial activities (Mengesha, 2024; Nigussie & Wale, 2022; Teka et al., 2020) and is traditionally used to treat ear infections, headaches, rheumatism, stomachaches, wounds, and malaria (Chekole, 2017), as well as providing essential oils for jute sack production (Oda & Erena, 2017). The species is restricted to rocky outcrops and cliffs in the Welo upland floristic region of northeastern Ethiopia, at elevations between 2,000 and 3,000 m, a habitat that has likely shaped its distinct evolutionary trajectory.

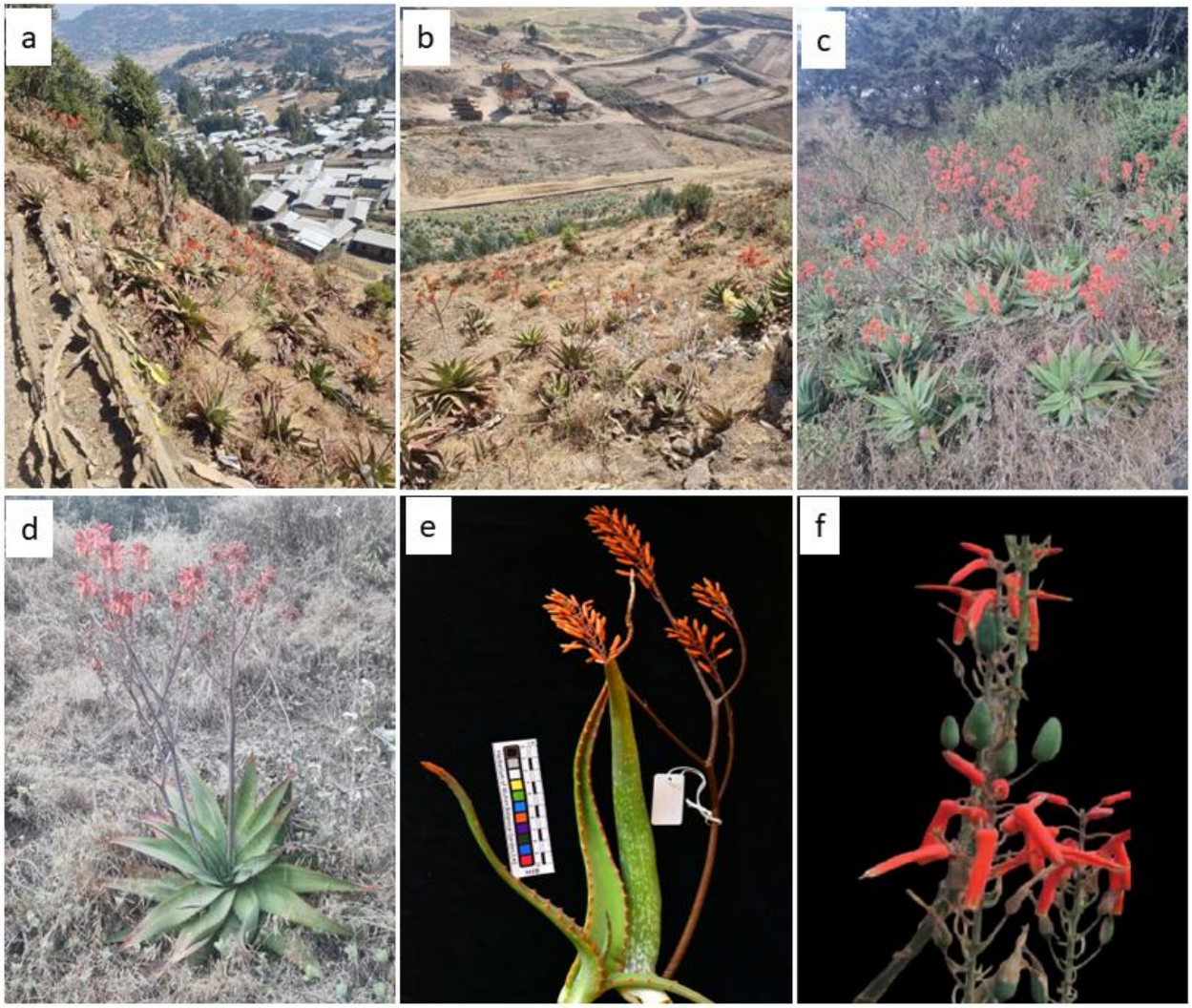


Figure 2. *Aloe weloensis* (a and b): A natural habitat of *A. weloensis* showing urban encroachment and habitat loss around Aziwa, Dessie (c and d): A habit in its natural habitat in Tossa Mountain, Dessie (e and f): fresh specimen, inflorescence and infructescence with flowering buds and fruits (photograph shot by Belay)

The conservation status of *A. weloensis* is of growing concern. As a naturally rare and narrowly distributed species, it faces multiple pressures that heighten its risk of decline. Foremost among these is habitat loss driven by urbanization and the expansion of human settlements (Guildford Cycads, 2025). Ongoing land conversion around Aziwa for infrastructure, including the construction of a waste treatment plant and new housing developments, is steadily reducing the extent of suitable habitat (Figure 2a-b). Such fragmentation can isolate populations, restrict gene flow, and increase vulnerability to environmental changes and disease. Climate change, marked by unpredictable rainfall and rising temperatures, exacerbates these issues, potentially disrupting the fragile balance of its ecosystem and pushing this rare species closer to extinction (Guildford

Cycads, 2025). Tossa Mountain and Harego area have retained more natural integrity than Aziwa, largely because their rugged terrain is less amenable to farming and human habitation (Figure 2c-d). Despite this ethnobotanical and ecological significance, plastid genomic data for *A. weloensis* remain unavailable, limiting independent, plastid-based evaluation of its phylogenetic placement.

Chloroplasts originated from an ancient cyanobacterial endosymbiosis (Zhang et al., 2025; Zou et al., 2025) and retain their own genome, markedly smaller than the nuclear genome and typically inherited uniparentally (Huang et al., 2020). In most angiosperms the plastome is a circular molecule with a conserved quadripartite structure comprising a Large Single-Copy (LSC) region and a Small Single-Copy (SSC) region separated by two Inverted Repeats (IRa and IRb), although several lineages, including some Fabaceae and algae, have lost these repeats entirely (Duan et al., 2021; Zou et al., 2025). Plastome size typically ranges from 120 to 160 kb and encodes 101–130 genes, primarily involved in photosynthesis, transcription, and translation (Cauz-Santos, 2025; Zou et al., 2025).

Despite the generally conserved gene content and low mutation rates (Parks et al., 2009), chloroplast genomes are not static since they may undergo gene loss, genomic rearrangements, and horizontal transfer, reflecting adaptive evolutionary histories (Gao et al., 2023; Liu et al., 2023). The Inverted Repeat regions are particularly important for maintaining plastome stability, and their expansion or contraction provides key phylogenetic signals (Cauz-Santos et al., 2020). Collectively, these features render the chloroplast genome a robust resource for phylogenetics, species identification, and conservation planning (Song et al., 2022; Su et al., 2025).

Chloroplast genomes have emerged as a powerful tool for addressing such genomic gaps, with established applications in systematics, germplasm conservation, genetic diversity assessment, and hybridization detection (Straub et al., 2012). Advances in high-throughput sequencing, particularly Illumina platforms, have made chloroplast genome assembly rapid and cost-effective, enabling large-scale comparative studies of genetic relationships and evolutionary patterns across plant groups (Dong et al., 2022; Luo et al., 2021).

In this study, we sequence and analyze the complete chloroplast genome of *A. weloensis*, a geographically and ecologically distinct, high-elevation endemic species. Through comparative sequence analysis, we reveal conserved and lineage-specific features, including inverted repeat boundary dynamics and codon usage bias, and reconstruct a phylogeny of 21 taxa, comprising 10

Aloe species and their closest relatives in Asphodelaceae and Amaryllidaceae, to resolve the taxonomic placement of *A. weloensis*. Specifically, we aim to: (1) characterize genomic structural variation, with particular attention to the evolutionary dynamics of IR boundaries and gene content; (2) screen for hypervariable coding and intergenic regions and simple sequence repeats (SSRs) to develop high-resolution molecular markers for species identification and germplasm conservation; (3) Assess the RNA editing patterns as candidate signatures of lineage-specific adaptation to the species high-elevation habitat; and (4) reconstruct a robust phylogeny to resolve the precise evolutionary placement of *A. weloensis* within *Aloe* clade. These findings provide essential genomic resources to support future taxonomic, evolutionary, and conservation research on this endangered endemic species.

2. Materials and methods

2.1. Taxon sampling and study site

Fresh, disease-free leaf samples of *A. weloensis* were collected in their natural habitats in February 2025 from Harego, southeast of Dessie (11°05′28″N, 39°39′58″E; 2,205 m a.s.l.; Figure 3), within the Welo upland floristic region of northeastern Ethiopia. Samples were dried in silica gel prior to DNA extraction. Identification of the plant material was done by Prof. Sebsebe Demissew, and a voucher specimen (SAJITBM0004) was deposited at the National Herbarium (Addis Ababa University).

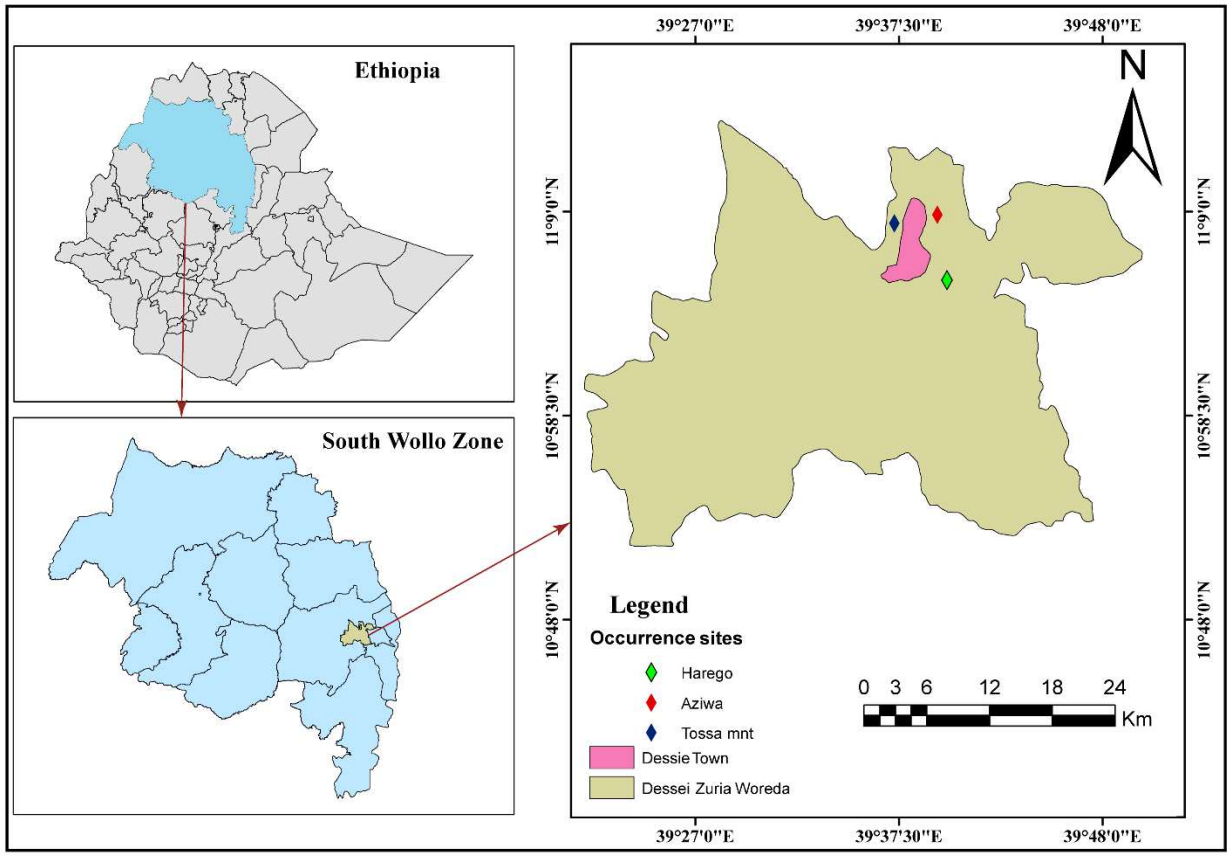


Figure 3. Map of occurrence sites of *Aloe weloensis*

The study area experiences a bimodal rainfall pattern, with the main wet season (Kiremt) from June to September and a shorter rainy period (Belg) from March to May. Mean annual temperature is 15.4°C, with absolute maxima and minima of 24.4°C and 7.4°C, respectively; total annual rainfall averages 717 mm (Figure 4).

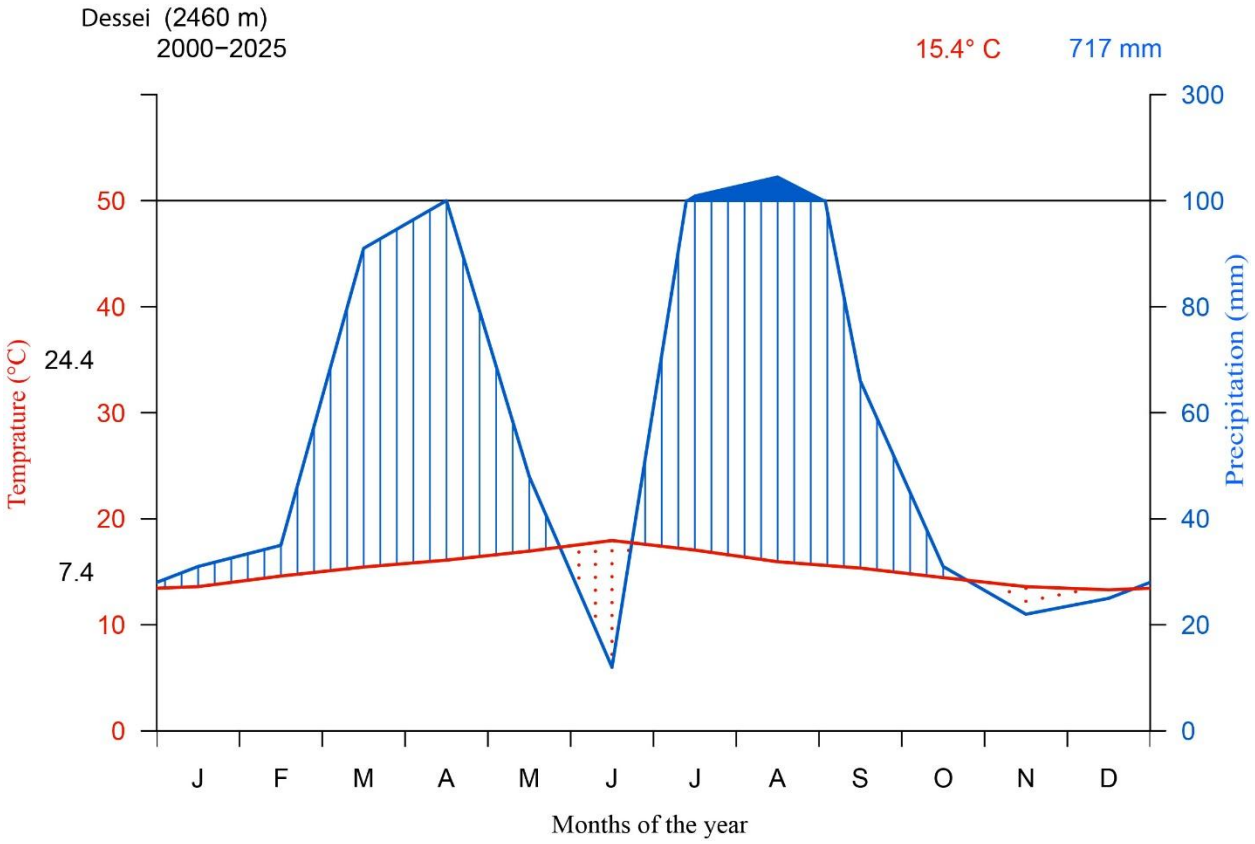


Figure 4. Climadiagram of Dessie, Ethiopia, showing monthly temperature and precipitation patterns from 2000-2025 (Climate data obtained from the NASA POWER site)

2.2. DNA extraction and sequencing

Total genomic DNA was extracted using the DNAsecure Plant Kit (Tiangen, China) following the manufacturer's protocol. DNA integrity was assessed by agarose gel electrophoresis, purity was measured using a NanoDrop spectrophotometer (OD260/280 ratio), and concentration was quantified with a Qubit 2.0 fluorometer. Sequencing libraries were constructed from qualified DNA samples using the MGIEasy Universal DNA Library Preparation Kit. Genomic DNA was fragmented by ultrasonication, followed by end repair, 5′-phosphorylation, 3′-A-tailing, adapter ligation, and PCR amplification. Library insert size was verified on an Agilent Bioanalyzer, and effective library concentration (>2 nM) was quantified by qPCR. Libraries were sequenced on the DNBSEQ-T7 platform. Raw reads were processed with fastp (Chen et al., 2018) using default parameters to remove low-quality reads, and screened for contamination prior to assembly.

2.3. Genome assembly and annotation

Clean reads were assembled using GetOrganelle v1.7.2 (https://github.com/kinggerm/GetOrganelle) (Jin et al., 2020) with the parameters -w 0.6 -o -R 15 -k 21,45,65,85,105 -F embplant_pt. Assembly graphs were visualized in Bandage to assess connectivity and completeness (Figure S1). Annotation was performed using PGA (Qu et al., 2019) and CPGAVAS2 (http://47.96.249.172:16019/analyzer/annotate) (Shi et al., 2019), with *A. vera* (NC035506) as the reference. Manual curation to verify gene boundaries and correct frameshifts was performed using Geneious Prime v2025.1.2 (https://www.geneious.com). Circular genome maps were visualized using an online program Chloroplot (https://irscope.shinyapps.io/Chloroplot/) (Zheng et al., 2020). For comparative analyses, raw reads for six additional *Aloe* species were downloaded from the China National GeneBank (CNGB) (https://db.cngb.org/data_resources/project/CNP0002108/) under project accession CNP0002108 (Kurniawan et al., 2025). These reads were reassembled using the same pipeline as for *A. weloensis* to ensure dataset consistency. Additional assembled chloroplast genomes, including *A. vera*, were obtained from NCBI (https://www.ncbi.nlm.nih.gov/) as references for annotation and comparative analysis (Table 1).

2.4. Structural variation analyses

Chloroplast simple sequence repeats (SSRs) were identified using the MicroSAtellite identification tool web (MISA) (http://pgrc.ipk-gatersleben.de/misa/) (Li et al., 2025) with thresholds of 10, 5, 4, 3, 3, and 3 repeats for mono- to hexa-nucleotide motifs, respectively (Xu et al., 2024). Long repeats (palindromic, complement, forward, and reverse) were detected using REPuter web server at (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (Kurtz et al., 2001) with a minimum repeat size of 30 bp, a maximum of 90 (Feng et al., 2023).

2.5. Comparative genomic analyses

The Mauve multiple genome alignment method was used to detect rearrangements and co-linearities in the chloroplast genomes of *A. weloensis* and related species with default parameters. The program mVISTA in Shuffle-LAGAN mode (Frazer et al., 2004) (https://genome.lbl.gov/vista/mvista/submit.shtml), was used to compare the complete cp genomes of the selected *Aloe* species with one assembled genome as the reference. Besides, expansion and contraction of inverted repeat (IR) boundaries at the four junctions (JLB-LSC/IRB, JSB-IRB/SSC,

JSA-SSC/IRA, and JLA-IRA/LSC) among cp genomes of *Aloe* species were visualised using the IRscope online web tool (https://irscope.shinyapps.io/irapp/).

Nucleotide diversity (π) was calculated for the 73 protein-coding gene alignments and 64 intergenic spacer alignments across the ten sampled *Aloe* plastomes, following Nei & Li, (1979). Within each individual locus alignment, a sliding window analysis (window length 100 bp, step size 25 bp) was performed using DnaSP v6 (Rozas et al., 2017). For each locus class, π values were sorted in descending order, and loci were classified as hypervariable when their nucleotide diversity exceeded the mean (μ) by more than 1.5 standard deviations (σ), i.e., $\pi > \mu + 1.5\sigma$, a threshold corresponding to approximately the top 6.7% most variable loci in the dataset. This criterion, following the approach of He et al., (2026), was applied uniformly to both the coding-gene and intergenic-spacer datasets to ensure comparability across locus classes. Nucleotide diversity (π) was calculated as the average number of nucleotide differences per site between two sequences: $\pi = (\sum_{i<j} k_{ij}) / (n(n-1)/2 \times L)$ Where: π = nucleotide diversity, k_{ij} = number of nucleotide differences between sequences i and j, n = number of sequences, L = total number of sites analysed, $\Sigma_{i<j}$ = sum over all pairwise sequence comparisons.

2.6. Codon usage and RNA editing analysis

Relative synonymous codon usage (RSCU) was analyzed using MEGA 11 (Tamura et al., 2021) and visualized as heatmaps in OriginPro (Yao et al., 2023). RSCU > 1 indicates preferential codon usage, RSCU < 1 indicates avoidance, and RSCU = 1 indicates no bias. RNA editing events were predicted using PREPACT 3.0 (http://www.prepact.de/prepact-main.php) (Lenz et al., 2018). with *Arabidopsis thaliana* (NC000932) as the reference, using default BLASTx settings (e-value cutoff 0.001, 30% filter threshold) for both forward (C-to-U) and reverse (U-to-C) predictions.

2.7. Phylogenetic Analysis

Phylogenetic relationships were reconstructed using both Maximum Likelihood (ML) and Bayesian Inference (BI) methods. Protein-coding genes were extracted from 21 species, including 10 *Aloe* species, 10 additional Asphodelaceae outgroups, and *Lycoris hunanensis* (Amaryllidaceae) as the distant outgroup, using PhyloSuite v1.2.3 (Zhang et al., 2020). After removing duplicate copies, 64 shared protein-coding genes were used for phylogenetic reconstruction. These were aligned with MAFFT v7.505 (Katoh & Standley, 2013), using the FFT-NS-2 strategy. Gap sites were removed with trimAl v1.2rev57 (Capella-Gutiérrez et al., 2009)

using the "automated1" command, and ambiguously aligned regions were eliminated with Gblocks 0.91b (Talavera & Castresana, 2007) with parameters: minimum 2 sequences for conserved/flank positions, maximum 8 contiguous non-conserved positions, minimum block length 10, and gap allowances. The best-fit substitution models were determined using ModelFinder v2.2.0 (Kalyaanamoorthy et al., 2017), with the AIC criterion for ML inference and BIC for BI inference. The ML trees were inferred in IQ-TREE v2.2.0 (Nguyen et al., 2015) under the GTR+I+G4+F model with 10,000 ultrafast bootstraps replicates (Minh et al., 2013). The BI tree was inferred using MrBayes v3.2.7a (Ronquist et al., 2012) under the JC model with two parallel runs of 2,000,000 generations, sampling every 1,000 generations, discarding the first 25% as burn-in. Trees were visualized using FigTree v1.4.4 (Rambaut, 2018) (https://tree.bio.ed.ac.uk/software/figtree/) and edited in Adobe Illustrator.

3. Results

3.1. General features of the chloroplast genome

The complete chloroplast genome of *A. weloensis* was assembled as a single circular molecule of 153,001 bp (GenBank accession: QB012994), exhibiting the typical angiosperm quadripartite structure (Table 1): a Large Single-Copy region (LSC = 83,613 bp), a Small Single-Copy region (SSC = 16,192 bp), and a pair of Inverted Repeats (IR = 26,598 bp each). Among the ten *Aloe* plastomes compared here, total genome size ranged from 152,364 bp (*A. arborescens*) to 154,127 bp (*A. barberae*), placing *A. weloensis* near the middle of the observed range. The LSC, IR, and SSC regions varied slightly across *Aloe* genomes (Table 1). *Aloe maculata* possessed the smallest SSC region (15,892 bp) and the largest IRs (26,858 bp), whereas *A. tenuifolia* exhibited the opposite pattern, with the largest SSC (17,004 bp) and the smallest IRs (26,565 bp).

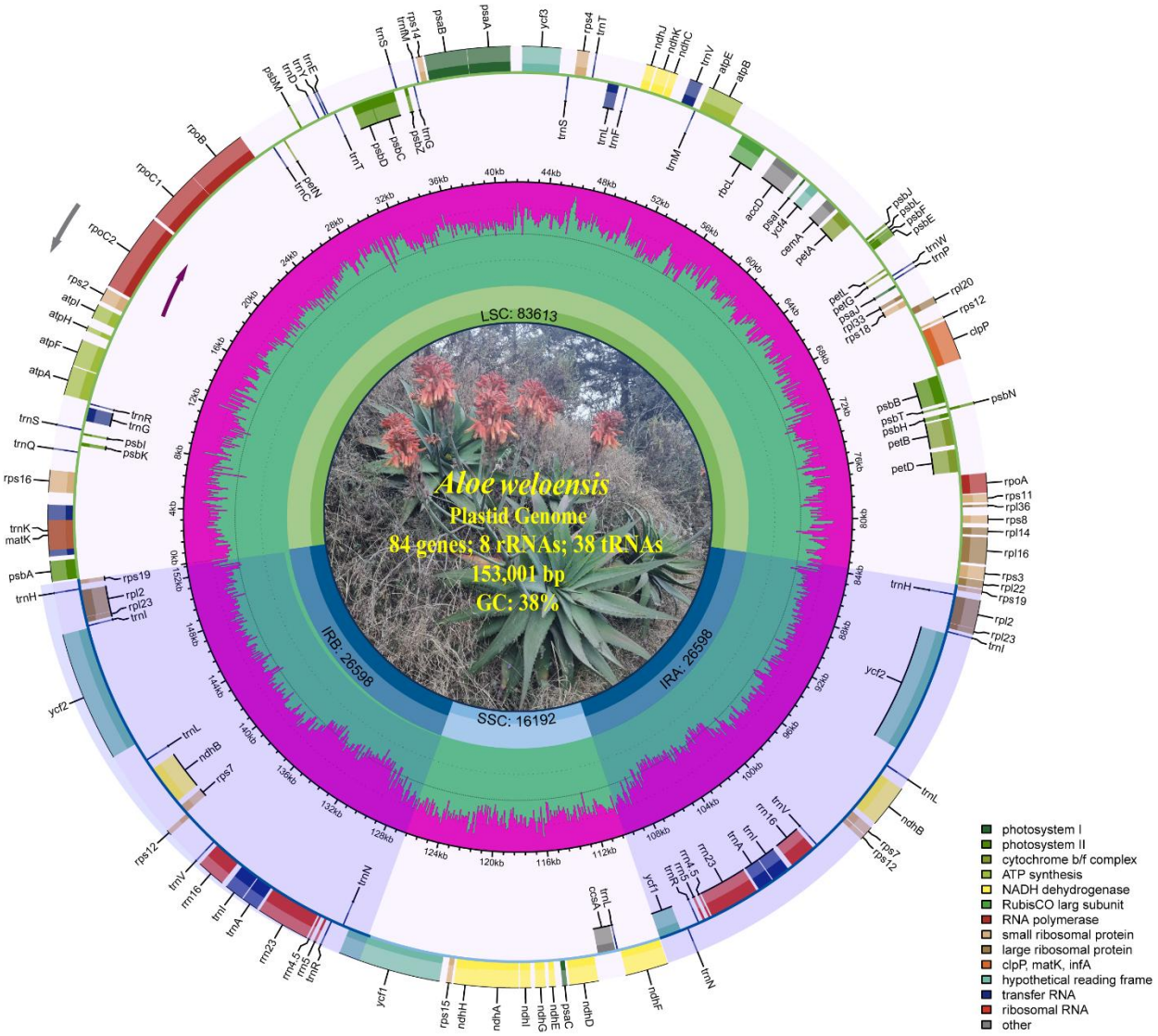


Figure 5. Circular chloroplast genome map of *Aloe weloensis*. Genes on the inner and outer circles are transcribed clockwise and counterclockwise, respectively. The functional group of each gene is represented in different colours.

3.2. Gene content and classification

All *Aloe* plastomes contained two copies of IR regions, ranging from 26,565 bp (*A. tenuifolia*) to 26,858 bp (*A. maculata*). The IRs flanked the SSC and LSC regions, which ranged from 15,151 to 17,004 bp and 83,002 to 84,010 bp, respectively (Table 1), while GC content ranged from 37.6% to 37.7%, and AT content from 62.3% to 62.4%.

Table 1. The characteristics of complete plastomes of *Aloe*. The new sequence is marked with an asterisk (*)

Genome features	Genome size (bp)	LSC size (bp)	SSC size (bp)	IR size (bp)	Total genes	PCG	tRNA genes	rRNA genes	AT (%)	GC (%)	Accession No.
*A. weloensis**	153,001	83613	16192	26598	130	84	38	8	62.4	37.6	QB012994*
A. vera	152,875	83504	16177	26597	131	85	38	8	62.3	37.7	NC035506
A. maculata	153,175	83567	15892	26858	131	85	38	8	62.4	37.6	NC035505
A. arborescens	152,364	83002	16182	26590	128	82	38	8	62.3	37.7	NC066013
A. excelsa	152754	83413	15151	26595	131	85	38	8	62.3	37.7	CNS0409471
A. barberae	154,127	84010	16959	26579	131	85	38	8	62.4	37.6	CNS0409470
A. glauca	152,383	83028	16179	26588	131	85	38	8	62.3	37.7	CNS0409474
A. marlothii	152,757	83415	16152	26595	131	85	38	8	62.3	37.7	CNS0409472
A. perfoliata	152,797	83462	16175	26580	131	85	38	8	62.4	37.6	CNS0409473
A. tenuifolia	154,103	83969	17004	26565	131	85	38	8	62.4	37.6	CNS0409475

Large single-copy (LSC), small single-copy (SSC), inverted repeat (IR), protein-coding gene (PCG)

We identified 131 functional genes in most *Aloe* species, with exceptions in *A. weloensis* (130 genes) and *A. arborescens* (128 genes). The gene complement comprised 85 protein-coding genes, with *A. weloensis* and *A. arborescens* possessing 84 and 82, respectively, while all species shared the same set of 38 tRNA and 8 rRNA genes (Figure 5, Table 1).

The genes of *A. weloensis* were categorized into five groups based on biological function: (1) 45 genes associated with photosynthesis; (2) 29 genes involved in self-replication, (3) 46 RNA genes comprising 38 tRNA and 8 rRNA, (4) 5 genes of other known functions, and (5) 5 genes of unknown function (Table 2). Among these, 20 genes were duplicated in the IR regions, comprising eight protein-coding genes (*ycf1*, *rpl2*, *rps7*, *ndhB*, *ycf2*, *rpl23*, *rps19*, and *rps12*), four rRNA genes, and eight tRNA genes (*trnN-GUU*, *trnR-ACG*, *trnA-UGC*, *trnI-GAU*, *trnV-GAC*, *trnL-CAA*, *trnI-CAU*, and *trnH-GUG*). The *rps12* gene is trans-spliced, with its 5′ end located in the LSC region and its 3′ end duplicated within the IR regions. The complete set of genes encoded by the *A. weloensis* plastomes is listed in Table 2.

Table 2. A list of genes found in the chloroplast genome of *A. weloensis*

Category of Genes	**Group of Genes**	**Genes**	**Number of Genes**
Genes for photosynthesis (45)	photosystem I	*psaA, psaB, psaC, psaI, psaJ*	5
	photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3***	15
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*	6
	Subunits of cytochrome b_6f complex	*petA, petB*, petD*, petG, petL, petN*	6
	Large subunit of Rubisco	*rbcL*	1
	Subunits of NADH dehydrogenase	*NdhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12
Self-replication (29)	Small subunit of ribosome	*rps11, rps12$^{\beta}$(2), rps14, rps15, rps16*, rps18, rps19(2), rps2(2), rps4, rps7(2), rps8,*	15
	Large subunit of ribosome	*rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl33, rpl36*	10
	DNA-dependent RNA polymerase	*rpoA, rpoB, rpoC1*, rpoC2*	4
RNA genes (46)	ribosomal RNA	*rrn5(2), rrn4.5(2), rrn16(2), rrn23(2)*	8
	tRNA genes	*trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC*, trnH-GUG(2), trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA*	38
Other genes (5)	Maturase	*matK*	1
	Envelope membrane protein	*cemA*	1
	Subunit of acetyl-CoA-carboxylase	*accD*	1
	C-type cytochrome synthesis gene	*ccsA*	1
	Protease	*ClpP***	1
Genes of unknown functions (5)	conserved reading frames	*ycf1(2), ycf2(2), ycf4*	5
Total			130

Note: *, indicates a gene with one intron, **, gene with two introns, $^{\beta}$, is a trans-spliced gene

Most genes lack introns, but two specific genes, the *clpP* and *ycf3*, are interrupted by two introns. Other 15 genes, *trnK-UUU, trnG-UCC, trnL-UAA*, *trnV-UAC, trnl-GAU, trnA-UGC, petB, petD, atpF, ndhB, ndhA, rpl2, rpl16, rps16*, and *rpoC1*, contain only a single intron (Table 3). Among

these, *trnK-UUU* harboured the largest intron, at 2,609 bp. Eleven intron-containing genes were located in the LSC region, one in the SSC region, and the remaining genes in the IR regions.

Table 3. The lengths of introns and exons for the split genes in *A. weloensis*

Genes	Region	ExonI (bp)	IntronI (bp)	ExonII (bp)	IntronII (bp)	ExonIII (bp)
trnK-UUU	LSC	37	2609	35		
rps16	LSC	39	887	213		
trnG-UCC	LSC	23	685	49		
atpF	LSC	160	815	407		
rpoC1	LSC	430	726	1634		
ycf3	LSC	124	745	230	768	153
trnL-UAA	LSC	37	521	50		
trnV-UAC	LSC	39	595	37		
clpP	LSC	71	833	291	673	250
petB	LSC	6	792	651		
petD	LSC	9	720	525		
rpl16	LSC	9	978	402		
rpl2(2)	IRA; IRB	394	661	428		
ndhB(2)	IRA; IRB	775	695	758		
trnI-GAU(2)	IRA; IRB	42	937	35		
trnA-UGC(2)	IRA; IRB	38	806	35		
ndhA	SSC	553	1094	539		

(2) indicates identical genes that are duplicated in the IR regions.

3.3. Contraction and expansion of IR regions

We compared the IR border regions across ten *Aloe* species, focusing on the four junction sites: LSC/IRb (JLB), IRb/SSC (JSB), SSC/IRa (JSA), and IRa/LSC (JLA), as well as adjacent genes (Figure 6). At the JLB (LSC/IRb), the *rps19* gene extends 46-86 bp into the IRb region (Figure 6), resulting in a truncated *rps19* pseudogene (*Ψrps19*) at the JLA (IRa/LSC) junction. The duplicated *rps19* fragments vary in length from 210 to 285 bp at both JLB and JLA. The *rpl22* gene, located entirely within the LSC adjacent to *rps19* at JLB, lies within 1-80 bp away from of the boundary. Near the duplicated *rps19* pseudogene at JLA, the photosystem II-associated gene *psbA* maintains a conserved position within the LSC, 49-88 bp from the IRa boundary, with a relatively constant size (1,062 bp) across all species.

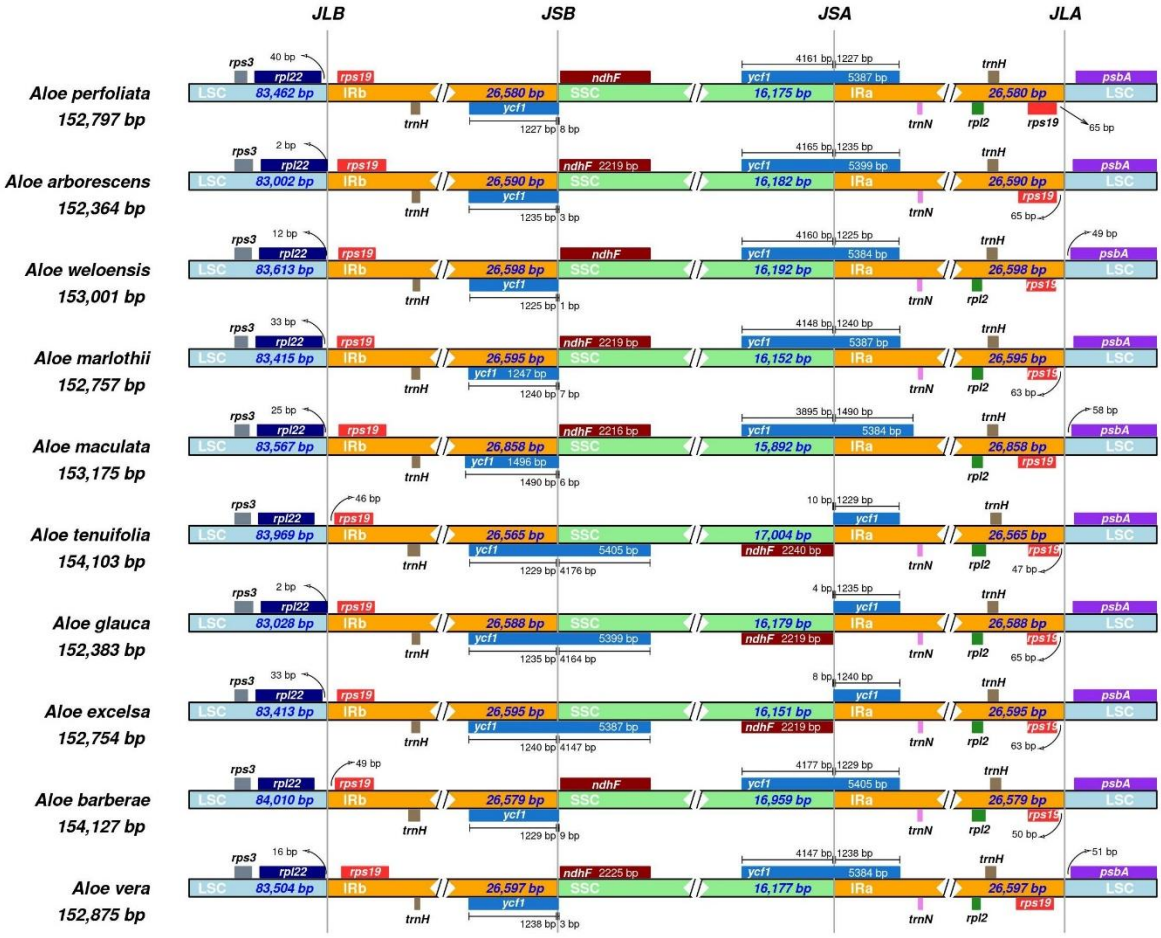


Figure 6. Structural variations and boundary analyses among the quadripartite regions (JLB, JSA, JSB, and JLA) of the *A. weloensis* chloroplast genome and related species.

At the JSB junction, the *ycf1* gene extends across the boundary into the SSC, generating a Ψ*ycf1* of variable length, ranging from 1,225 bp (*A. weloensis*) to 1,490 bp (*A. maculata*). The functional *ycf1* copy at the JSA (SSC/IRa), spanning from the SSC into IRa, varies in length from 3,895 bp (*A. maculata*) to 4,177 bp (*A. barberae* and *A. tenuifolia*), showing the greatest size variation across all junctions and identifying JSA as the most dynamic boundary in the genus. The *trnN* gene is located 1,582–1,847 bp from the JSA boundary, with the maximum distance recorded in *A. maculata* (1,847 bp) and the shortest distance in *A. weloensis* (1,582 bp). Overall, the LSC/IR boundaries are relatively stable, whereas the SSC/IR boundaries vary more substantially among species, with IR expansion corresponding to SSC contraction (Figure 6).

3.4. Comparative genomic analysis

The chloroplast genomes of four *Aloe* species (*A. vera*, *A. weloensis*, *A. arborescens*, and *A. maculata*) were compared using mVISTA, with *A. vera* as the reference (Figure 7a), to evaluate genomic conservation and divergence. Coding regions (exons, represented in purple) were broadly conserved, maintaining high sequence identity, whereas non-coding intergenic regions and untranslated regions (represented in cyan and pink as UTRs and CNS, respectively) exhibited substantially greater divergence. The IR regions showed the highest sequence identity while the LSC and SSC regions displayed greater variability, particularly within non-coding intergenic spacers, suggesting relaxed selective constraint in these regions (W. Shi et al., 2023). A particularly pronounced, species-specific divergence was observed in *A. arborescens* around the *petN–psbM* region (~28 kb), where sequence identity dropped sharply relative to the other three species compared. This localized loss of similarity is consistent with the additional gene losses reported for *A. arborescens*, including *psbN, rpl2, rps12,* and *ycf3*.

Several coding genes were notably more variable, particularly *ycf1*, *ndhA*, and *ndhF*, which have been consistently identified as mutational hotspots across Asphodelaceae and related families (Chen et al., 2025; Henriquez et al., 2020). Among these, *ycf1* stands out as it spans the SSC/IR boundary and has been widely proposed as a universal plant barcode owing to its balance of moderate conservation and sufficient polymorphism (Dong et al., 2022; Mehmood et al., 2024). The *ndh* gene complex likewise showed divergence in the SSC region, particularly for *ndhF*, suggesting relaxed selective pressure in this locus within the genus (Jin et al., 2020) (Figure 7a). Despite morphological differences among *Aloe* species, their chloroplast genomes remain highly conserved structurally, with variation concentrated in hypervariable loci. These hypervariable regions (identified by dips in the identity tracks) are strong candidates for developing phylogenetic markers and for DNA barcoding applications in *Aloe*, supporting the utility of plastome data for resolving phylogenetic relationships within the genus.

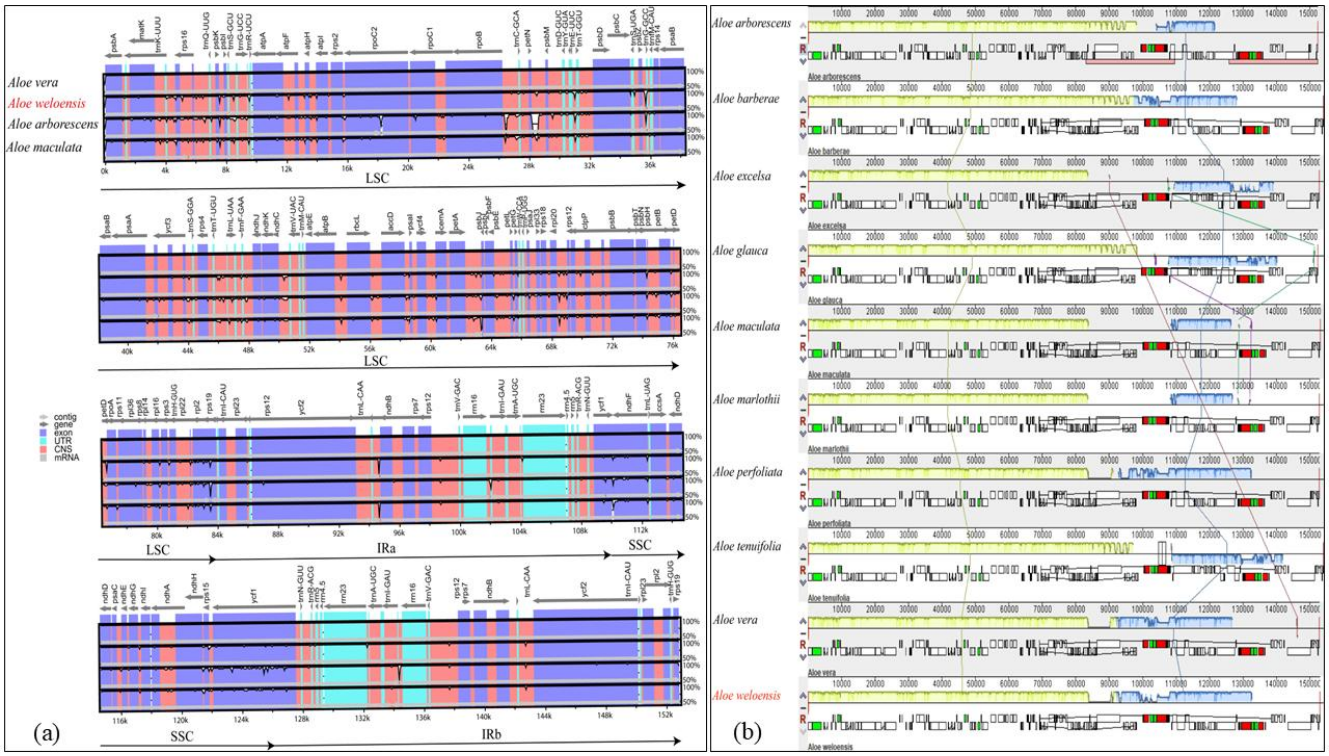


Figure 7. Comparing chloroplast genomes across *Aloe*. (a) Sequence alignment of four *Aloe* chloroplast genomes visualized using mVISTA, with *A. vera* as the reference. The graph plots sequence identity (50-100%) against the ~152 kb genome coordinates. (b) Collinearity analysis of ten *Aloe* species performed using Mauve. Colored blocks represent locally collinear blocks (LCBs) of homologous sequences, and connecting lines indicate the position of corresponding LCBs across adjacent genomes.

To further assess structural conservation, a mauve alignment of ten *Aloe* species was generated (Figure 7b). Within the LSC region, all ten genomes showed a single, uninterrupted, identically ordered locally collinear block (LCB) spanning approximately 0-83 kb, with no evidence of inversion or rearrangement in this portion of the plastome. In contrast, the IR/SSC boundary region (~90-150 kb) showed clear evidence of local block reordering among several taxa, most notably between *A. barberae*, *A. excelsa*, *A. glauca*, *A. maculata*, and *A. marlothii*, consistent with the IR expansion/contraction patterns already identified in Figure 6. Gene content and order remained highly stable across all ten genomes, and sequence identity within the LCBs was consistently high; overall, the alignment indicates that *Aloe* plastome architecture is strongly conserved at the whole-genome scale, with structural variation concentrated specifically at the IR boundaries rather than distributed broadly across the genome.

3.5. Codon usage bias analysis

We plotted the heatmap of relative synonymous codon usage (RSCU) in ten *Aloe* chloroplast genomes to assess codon preference patterns (Figure 8a). Among the 65 codons analyzed, 30 exhibited RSCU < 1 (low usage), 33 exhibited RSCU > 1 (high usage), and two codons (UGG and AUG) showed no preference (RSCU = 1). Codon AGA had the highest RSCU value (1.92), whereas CGC had the lowest (0.45) (Figure 8a, Table S2). Of the 33 preferentially used codons, 27 terminated in A or U, whereas only 5 ended in C or G, indicating a strong AT-biased codon usage in the *A. weloensis* plastome, a pattern consistent with previous reports in *Albizia julibrissin* (Zhang et al., 2021) and *Plantago asiatica* (Wu et al., 2023).

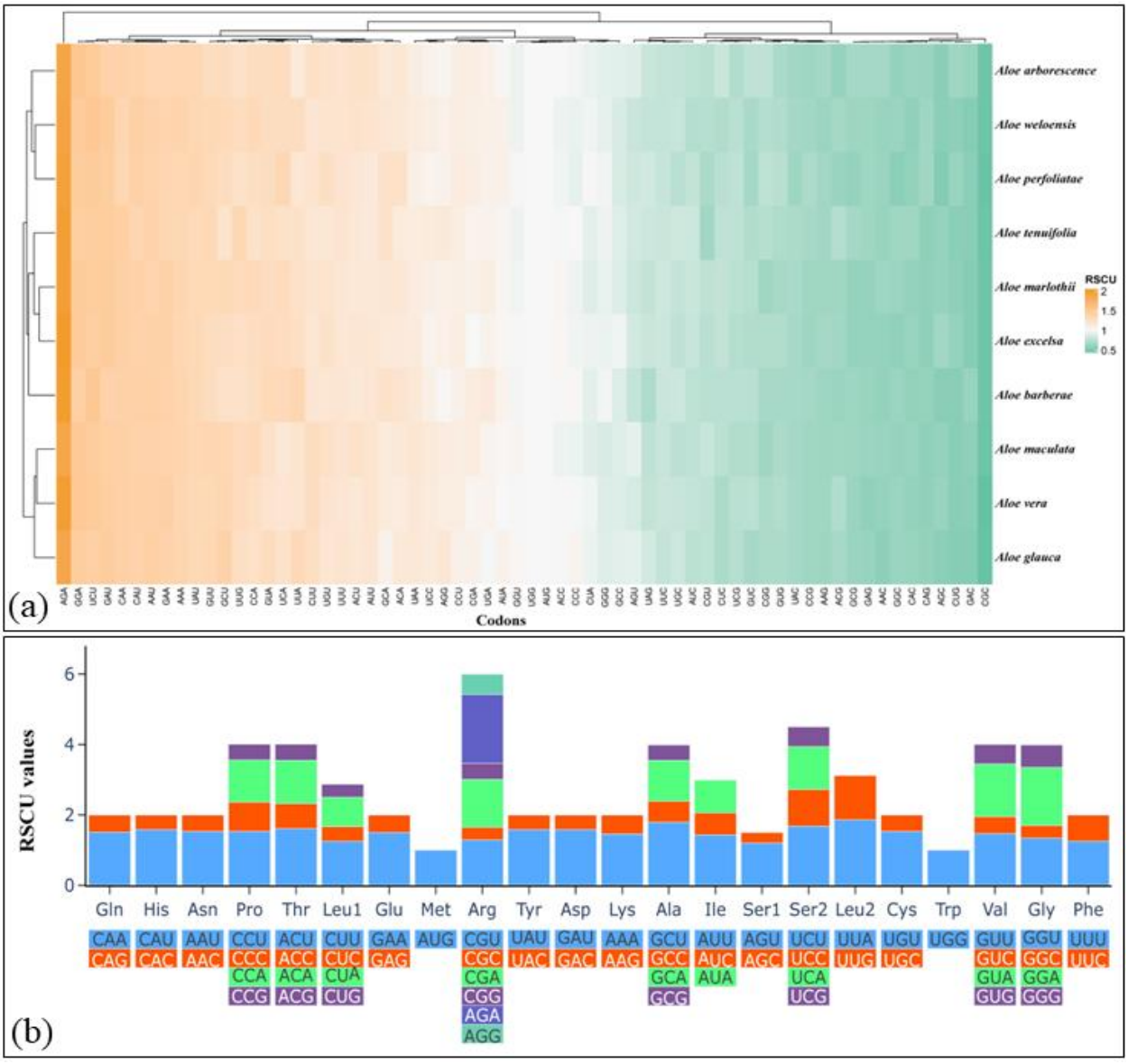


Figure 8. Codon usage analysis of *Aloe* chloroplast genomes. (a) Heatmap showing relative synonymous codon usage (RSCU) across ten *Aloe* species based on hierarchical clustering; orange indicates preferred codons (RSCU > 1), dark green indicates less preferred codons (RSCU < 1). (b) Codon composition for the 20 amino acids and stop codons in the *A. weloensis* plastome, with colours representing RSCU values.

In plant chloroplast genomes, codon usage is largely shaped by the base composition of the third codon position, which is subject to selective pressure due to its degenerate nature (Morton, 2003). The *Aloe weloensis* chloroplast genome contained 51,000 codons, with leucine (Leu; 5,091) the most frequent and tryptophan (Trp; 710) the least frequent (Table S2). The RSCU values for the 64 codons including stop codons are illustrated in Figure 8b. Among the amino acids, arginine (Arg) and serine (Ser) had the highest total RSCU values (6 and 5, respectively), indicating strong preferences for CGU/CGA/AGA and UCU/UCC/UCA, respectively. Proline (Pro), threonine (Thr), alanine (Ala), valine (Val), and glycine (Gly) each had a total RSCU of 4, with their preferred codons being CCU/CCA, ACU/ACA, GCU/GCA, GUU/GUA, and GGU/GGA (Figure 8).

Additionally, most remaining amino acids displayed a distinct bias toward A- or U-ending codons: asparagine (Asn) preferred AAU, aspartate (Asp) GAU, cysteine (Cys) UGU, glutamine (Gln) CAA, glutamic acid (Glu) GAA, histidine (His) CAU, lysine (Lys) AAA, phenylalanine (Phe) UUU, and tyrosine (Tyr) UAU. Isoleucine (Ile) was encoded by three codons, yet only AUU exhibited an RSCU value greater than 1 (1.4). Methionine (Met) and tryptophan (Trp) were encoded by AUG and UGG, respectively, resulting in an RSCU of 1 for both, indicating no synonymous preference.

3.6. Repetitive sequence analysis

Repetitive sequences serve as valuable genetic markers in population research and may promote genomic variation and rearrangements (Su et al., 2025). In the repeat structure analysis, four types of Long Sequence Repeats (LSRs), including palindromic, forward, complementary, and reverse elements, were identified in the studied *Aloe* cp genomes. Among these, palindromic repeats were the most abundant (48%), followed by forward repeats (38%), whereas reverse and complement repeats were comparatively rare, representing only 9% and 5%, respectively (Figure 9b). The highest total number of repeats was recorded in the *A. weloensis*, with 89 repeats, followed by *A. perfoliata* (76*)* and *A. vera* (56) (Figure 9a, Table S4).

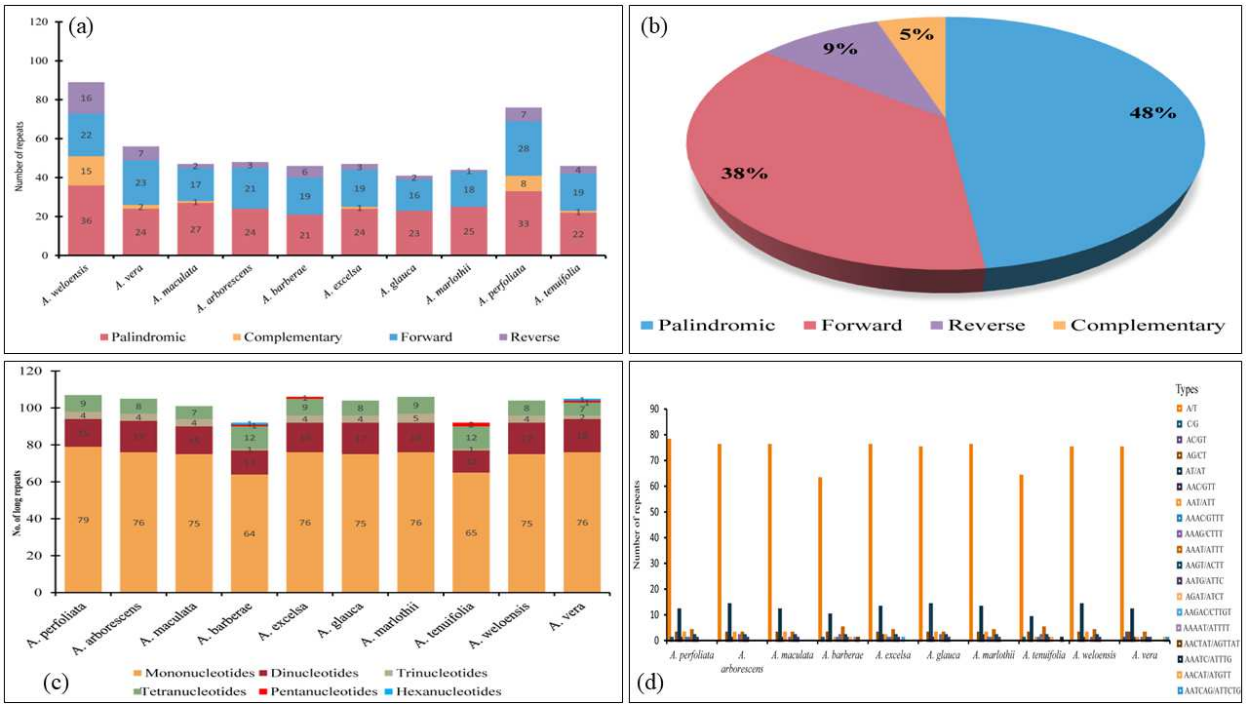


Figure 9. Analysis of simple sequence repeats (SSRs) and long sequence repeats (LSRs) in the chloroplast genomes of ten *Aloe* species. a. Number of long repeat types across the species; b. Frequency of different repeat types; c. Abundance of different SSR classes; and d. counts of individual SSR motif types.

The microsatelite analysis revealed a total of 1,022 simple sequence repeats (SSRs), with repeat types ranging from mono- to hexanucleotides (Figure 9c). Among the species, *A. perfoliata* showed the highest SSR count (107), followed by *A. excelsa* and *A. marlothii* (106 each). In contrast, *A. tenuifolia* and *A. barberae* exhibited the lowest numbers (92 each), while *A. weloensis* showed an intermediate SSR count of (104) (Figure 9c, Table S3). The most abundant nucleotide contents across all studied genomes were mononucleotide repeats, specifically A/T motifs, followed by di-, tetra-, tri-, and pentanucleotides, with hexanucleotides being the least frequent (Figure 9c).

3.7. Nucleotide diversity analysis

Nucleotide diversity analysis revealed variable levels of sequence divergence among coding genes and intergenic spacers across ten *Aloe* plastomes. Protein-coding genes showed π values ranging from 0 to 0.013, with ten hypervariable loci ($\pi > \mu + k\sigma$, $\pi > 0.0045 + 1.5 \times 0.0027$, $\pi > 0.0085$): *rpl22*, *psbM*, *ndhF*, *rpl16*, *matK*, *ndhD*, *rpl20*, *psbN*, *atpF*, and *rpoA*. These loci were distributed across the LSC (eight) and SSC (two) regions. The average nucleotide diversity (μ) of coding regions was 0.0045 (Figure 10a).

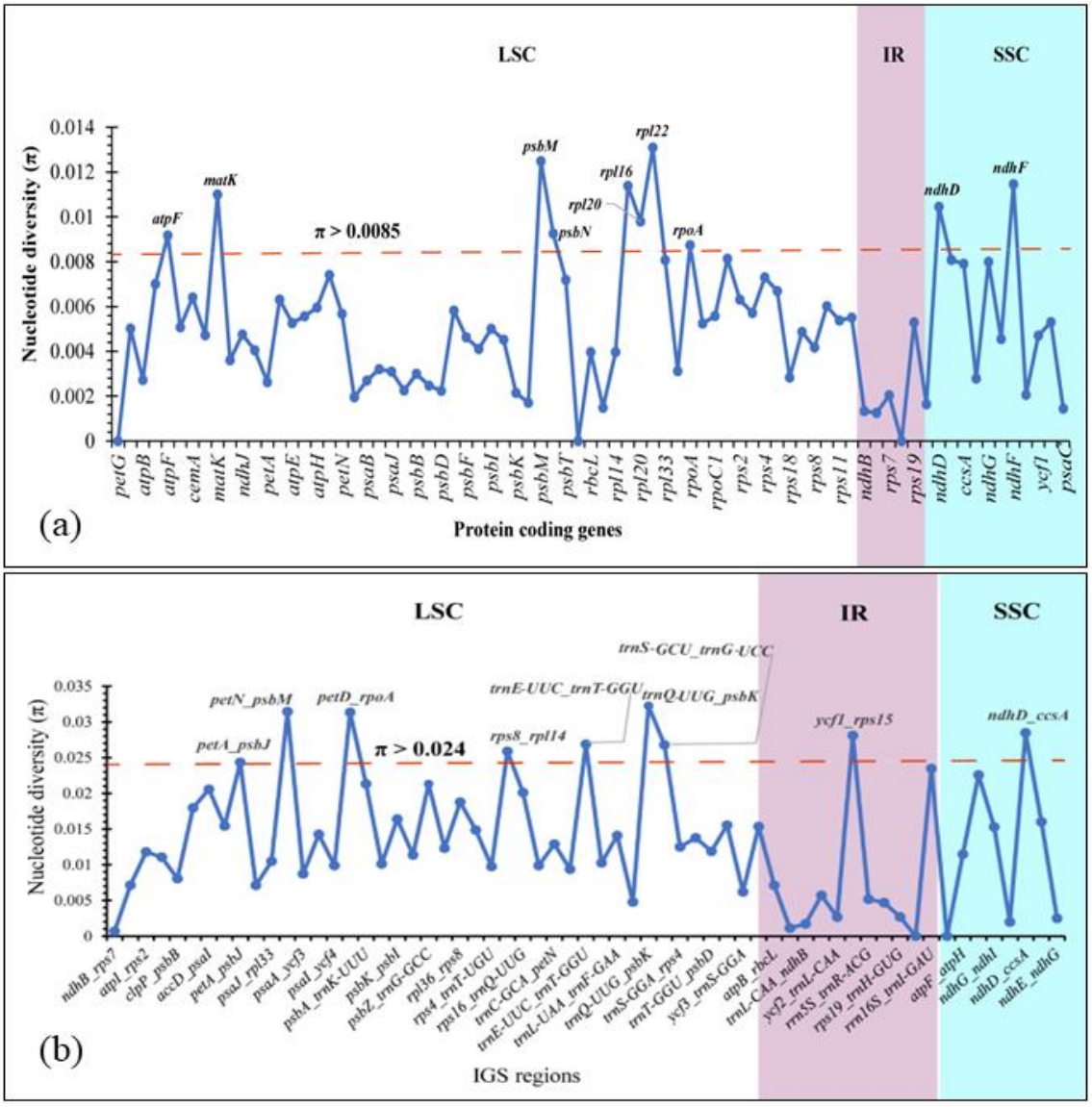


Figure 10. Sliding-window analysis showing nucleotide diversity across 10 *Aloe* chloroplast genomes. (a) Nucleotide diversity (π) of 73 protein-coding genes, (b) Nucleotide diversity (π) of 64 intergenic spacer (IGS) regions.

Intergenic spacers exhibited higher levels of variation, with π values ranging from 0 to 0.032. Nine hypervariable intergenic regions ($\pi > 0.011 + 1.5 \times 0.0084$, $\pi > 0.024$) were identified, most located in the LSC region (*trnQ-UUG_psbK, petN_psbM, petD_rpoA, trnE-UUC_trnT-GGU, trnS-GCU_trnG-UCC, rps8_rpl14*, and *petA_psbJ*), and two in the SSC (*ndhD_ccsA*) and IR (*ycf1_rps15)* (Figure 10b). The average π value of noncoding regions (0.011) was markedly higher than that of protein-coding genes, consistent with the expected relaxation of selective constraints

in non-coding regions (Shaw et al., 2007). This pattern suggests that IGS regions may provide greater phylogenetic resolution at lower taxonomic levels, particularly within the LSC region, where most hypervariable loci are concentrated.

3.8. RNA editing patterns

RNA editing is a molecular process by which protein-coding genes alter genetic information in RNA molecules, primarily through the addition or deletion of single nucleotides (Rodrigues et al., 2017). This can alter amino acid encoding and change the protein's primary, secondary, and tertiary structures, which may be essential for its function. In this study, we used forward (C to U) and reverse (U to C) editing in BLASTX, and a total of 734 potential RNA editing events, comprising 395 forward and 339 reverse, were predicted across 62 RNA editing sites in *A. weloensis* cp genomes (Table S5). They showed no variation in the number of editing sites or in amino acid conversion across genomes from different regions. However, no editing events were found in 16 genes, including *psaC*, *rps7*, *rpl36*, *psbH*, *psbN*, *clpP*, *rps12*, *psaJ*, *psbE*, *psbL*, *psaI*, *rbcL*, *psbD*, *petN*, *atpH*, and *psbI*. The most abundant conversion type was phenylalanine to leucine (F-L) (11 %) change, followed by serine to leucine (S-L), accounting for 10 %, while stop codon to arginine (* to R), Leucine to Serine (L to S) and Proline to Phenylalanine (P to F) showed the least conversions (Table S6).

A total of 28 amino acid conversion types were identified. Editing occurred predominantly at the first nucleotide position (700 events; 95.4%), with 17 events (2.3%) at the second position and 17 events (2.3%) at both positions. Conversions such as P to S, S to P, F to L, L to F, H to Y, Y to H, W to R, R to W, R to C, C to R, and * to R resulted from first-position editing, whereas P to F and F to P required editing at both positions. In contrast, P to L, S to F, F to S, A to V, V to A, T to I, I to T, M to T, and T to M conversions resulted solely from second-position editing (Table S6). Silent RNA edits were identified in the *ycf1* transcript, where codon changes (UUA→CCA, UUG→CCG, and CUA→UCA) did not alter the corresponding amino acid.

The RNA editing events triggered corresponding codon changes. Among them, 281 events switched amino acids from hydrophilic to hydrophobic, with the most frequent being S to L (32%), followed by T to I (18.5%) and S to F (14%) (Table S7). Conversely, 182 events switched from hydrophobic to hydrophilic, with I to T the most common (28%) and W to R and C to R the least

common (3% each). A total of 271 events maintained hydrophobicity, with F to L (30%) being the most frequent, followed by V to A (14%) and L to F (13.5%).

3.9. Phylogenetic analysis

Phylogenetic reconstruction based on 64 shared protein-coding genes from 21 plastomes resolved a well-supported topology for *Aloe* and its close relatives (Figure 9). The ten sampled *Aloe* species formed a single, strongly supported clade (BS = 100%, PP = 1) to the exclusion of *Aloidendron pillansii* and the remaining Asphodelaceae outgroups (*Bulbine bulbosa*, five *Eremurus species*, *Kniphofia uvaria*, *Xanthorrhoea preissii*, and *Hemerocallis fulva*), with *Lycoris hunanensis* (Amaryllidaceae) as the most distant outgroup. Nearly all internal nodes showed maximum bootstrap support (BS = 100%), except for *A. perfoliata* placement (BS = 79%, PP = 0.98) and the clade uniting *A. weloensis, A. perfoliata,* and *A. marlothii* (BS = 94%, PP = 1). *Aloe weloensis* was recovered in a moderately supported clade alongside *A. perfoliata* and *A. marlothii*, indicating a close phylogenetic affinity among these three species.

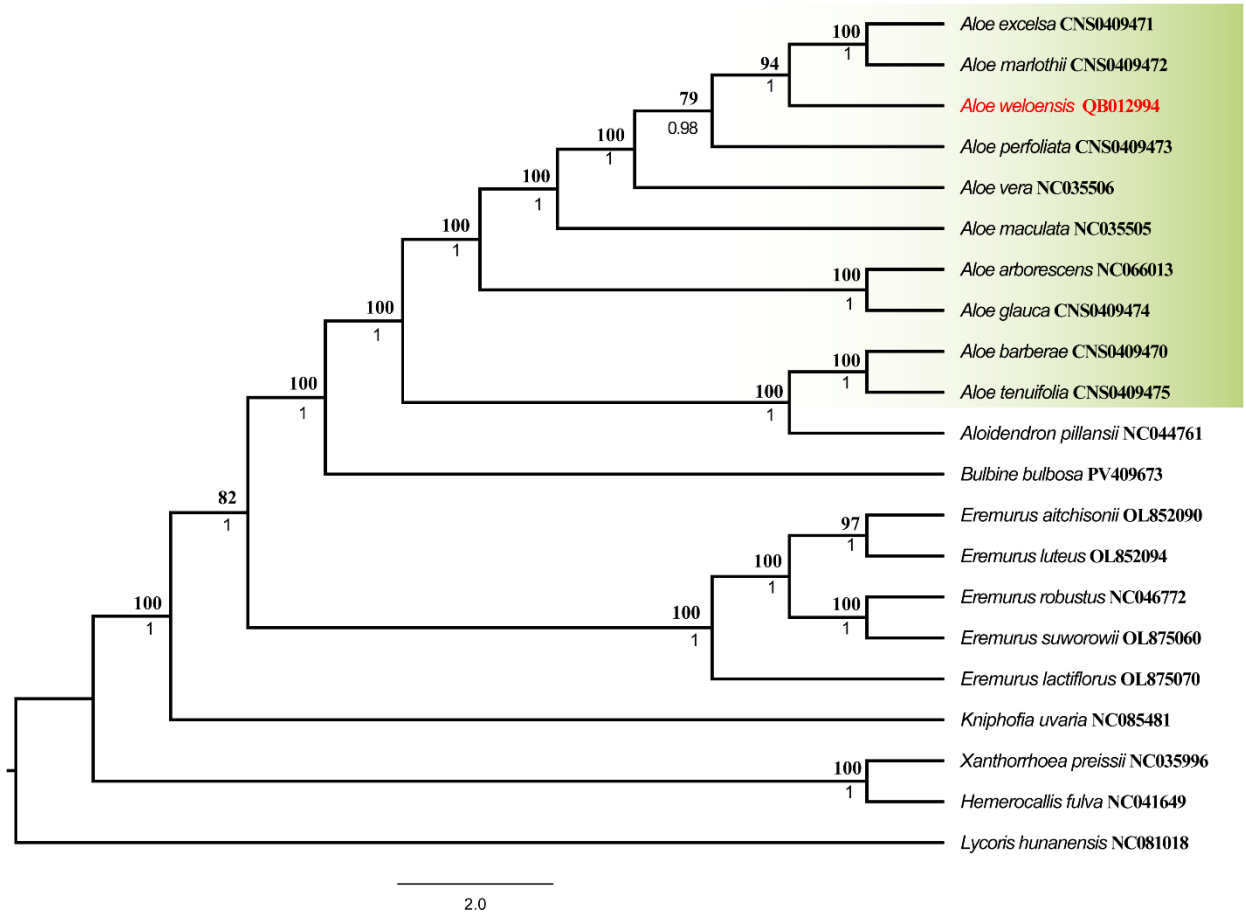


Figure 11. Maximum likelihood (ML) and Bayesian inference (BI) phylogeny of *Aloe* and related Asphodelaceae species based on 64 shared protein-coding genes. The numbers on the nodes indicate ML bootstrap percentages (top) and BI posterior probabilities (bottom). *Lycoris hunanensis* (Amaryllidaceae) was used as the outgroup.

4. Discussions

4.1. Genomic architecture and structural conservation

The newly sequenced plastome of *A. weloensis* (153,001 bp) is comparable in size to those previously reported for *Aloe* species (152,364–154,127 bp; Kurniawan et al., 2025). Like most angiosperm plastomes, it exhibits the typical quadripartite structure and gene content characteristic of the genus (Table 1). This size range is also consistent with data from other Asphodelaceae genera, including *Eremurus* (153,782–155,647 bp) and *Hemerocallis* (~156,000 bp) (Kurniawan et al., 2025; Makhmudjanov et al., 2023). Such findings reinforce the widely accepted view that angiosperm chloroplast genomes are highly conserved in structure, organization, and gene content, with minor size differences largely attributable to evolutionary events (Abdullah et al., 2020; Daniell et al., 2016; Ran et al., 2020). The slight size variation observed across species, primarily driven by differences in SSC and IR lengths, supports the hypothesis that IR boundary shifts are a primary contributor to genome-size heterogeneity in angiosperm plastomes (Abdullah et al., 2025). This pattern aligns with observations in other Asphodelaceae genera, including *Eremurus* (153,782–155,482 bp) (Makhmudjanov et al., 2023) and *Hemerocallis* (~156,000 bp) (Wu et al., 2024), suggesting that IR dynamics represent a shared evolutionary feature across the family.

The IR-boundary measurements agree closely with the results of Kurniawan et al. (2025), including the distance of *rpl22* from the LSC/IRb boundary and the length of the functional *ycf1* copy at the SSC/IRa junction, lending independent support for the stability of these features across the genus. The GC content patterns were similar to those in most angiosperms, with the IR region showing the highest GC content due to the presence of GC-rich rRNA genes (Makhmudjanov et al., 2023; Mower et al., 2019).

The microsatellites (SSRs) constitute an important class of repetitive elements in chloroplast genomes. Owing to their high polymorphism, abundance, and codominant inheritance, SSRs are widely applied as effective molecular markers in genetic diversity, population structure, and phylogeographic studies (Ni et al., 2016).

Repetitive elements play important roles in plastome evolution, including the regulation of gene expression, the facilitation of recombination, and the maintenance of genome organization (Zhang et al., 2024). Across the ten *Aloe* plastomes, 540 LSRs were identified, predominantly palindromic (48%) and forward repeats (38%), with reverse and complement repeats comparatively rare (9% and 5%, respectively). *Aloe weloensis* harbored the highest total repeat count among the sampled species (89), followed by *A. perfoliata* (76) and *A. vera* (56), suggesting that repeat accumulation is not uniform across the genus and may reflect lineage-specific differences in replication slippage or recombination activity (Su et al., 2025). Codon usage in the *A. weloensis* plastome showed a pronounced bias toward A/T-ending codons, with 27 of the 33 preferentially used codons terminating in A or U, a pattern consistent with the overall AT-richness of the genome and with codon usage patterns reported in other angiosperm plastomes, including *Albizia julibrissin* (Zhang et al., 2021) and *Plantago asiatica* (Wu et al., 2023). This bias is generally attributed to a combination of mutational pressure at the AT-rich third codon position and weak translational selection acting on highly expressed plastid genes (Morton, 2003), and is consistent with the pattern reported for other *Aloe* plastomes (Kurniawan et al., 2025).

4.2. Comparative structural analysis

Sequence-level comparison against *A. vera, A. arborescens,* and *A. maculata* using mVISTA confirmed that coding regions of the *A. weloensis* plastome are highly conserved, with divergence concentrated in a small number of loci (*ycf1, ndhA, ndhF*) and in non-coding intergenic spacers, consistent with the general pattern of purifying selection acting more strongly on coding than non-coding plastid sequence (Shaw et al., 2007). Notably, sequence identity was highest within the IR regions and lowest within the LSC and SSC, reflecting the well-documented role of biased gene conversion between IR copies in constraining divergence between IR copies relative to single-copy regions (Khakhlova & Bock, 2006). A pronounced, species-specific divergence was observed in *A. arborescens* around the *petN-psbM* region, consistent with the additional gene losses (*psbN, rpl2, rps12, ycf3*) reported for this species relative to other sampled *Aloe* taxa (Kurniawan et al., 2025). Whole-genome collinearity analysis using Mauve showed that all ten sampled genomes shared a single, uninterrupted, identically ordered locally collinear block across the entire LSC region, with no evidence of inversion, while the IR/SSC boundary region showed local block reordering among several taxa, consistent with the IR expansion/contraction dynamics described above rather than large-scale genomic rearrangement. *Aloe arborescens* also showed internal

repeat regions flagged in this portion of the genome, not detected in the other nine species, further reinforcing this taxon's distinct structural profile within the genus.

4.3. Nucleotide diversity and candidate molecular markers

Consistent with the general pattern of relaxed selective constraint on non-coding plastid sequence, average nucleotide diversity was substantially higher across intergenic spacers ($\mu = 0.011$) than protein-coding genes ($\mu = 0.0044$), and both loci classes showed a small subset of consistently divergent regions rather than a uniform elevation in variability across the genome (Shaw et al., 2007). Using a threshold of $\pi > \mu + 1.5\sigma$, following the approach of He et al. (2026), ten hypervariable coding genes (*rpl22, psbM, ndhF, rpl16, matK, ndhD, rpl20, psbN, atpF,* and *rpoA*) and nine hypervariable intergenic regions (including *trnQ-UUG-psbK, petN-psbM, petD-rpoA,* and *trnE-UUC-trnT-GGU*) were identified, with the majority concentrated within the LSC region.

The presence of *matK* among the hypervariable coding loci is notable, as it is one of the two core plant DNA barcoding markers (Hollingsworth et al., 2011) and shows here that it retains meaningful resolving power primarily within *Aloe*. Several intergenic hotspots, such as *trnQ-UUG-psbK* and *petN-psbM* do not overlap with the five hotspots reported by Kurniawan et al. (2025), which may reflect differences in taxon sampling, the addition of the phylogenetically isolated *A. weloensis* lineage, or differences in the underlying calculation approach. Together, these loci, spanning both coding and non-coding sequences, constitute a candidate marker panel with direct utility for species-level identification, phylogeographic sampling, and future population genetic assessment of *A. weloensis* and its close relatives. The highly repeatable and polymorphic SSRs in the plant chloroplast genome are typically utilized as molecular markers for genetic variation and species identification (Khabbazi, 2026). As a result, the SSRs in the *Aloe* chloroplast genome were AT-rich, consistent with the overall AT abundance of the chloroplast sequence (Figure 9d). This discovery has been observed in other chloroplast genomes, with the hypothesis that it is due to the greater mutability of A/T relative to G/C base pairs (Xie et al., 2018).

4.4. RNA editing as a source of proteomic diversification

This study provides the first characterization of RNA editing patterns in the genus *Aloe.* A total of 734 candidate editing events were predicted across 62 gene loci in *A. weloensis*, predominantly C-to-U conversions (395 events) with a smaller contribution from reverse U-to-C editing (339 events), consistent with the general pattern reported across land plant plastomes (Rodrigues et al.,

2017). Editing occurred overwhelmingly at the first codon position (95.4% of events), and the resulting amino acid substitutions most frequently increased hydrophobicity (281 events) or left it unchanged (303 events), while a smaller number decreased hydrophobicity (182 events). This pattern is consistent with the proposed role of RNA editing in restoring evolutionarily conserved, functionally important hydrophobic residues at the protein level that are not directly encoded at the DNA level, a mechanism particularly relevant for membrane-associated photosynthetic complexes, where hydrophobicity at key residues is often required for correct protein folding and insertion (Jiang et al., 2012). Notably, editing frequency did not track sequence-level conservation: *ycf2*, one of the least variable coding genes in the genus-wide nucleotide diversity analysis, carried by far the highest number of predicted editing sites (165) of any gene in the *A. weloensis* plastome, nearly twice that of the next most heavily edited locus, *ycf1* (80 sites), followed by *rpoC2* (52), *matK* (50), *ndhF* (37), and *ndhB* (34) (Table S5). This decoupling between DNA-level conservation and RNA-level editing frequency suggests that post-transcriptional modification and nucleotide substitution operate under largely independent constraints at this locus, consistent with the proposed role of RNA editing in maintaining protein function despite strong purifying selection on the underlying coding sequence (Jiang et al., 2012). The identification of three silent edits within the *ycf1* transcript (UUA→CCA, UUG→CCG, and CUA→UCA), which alter codon sequence without changing the encoded amino acid, further illustrates that not all editing events are functionally consequential, and that editing-site conservation should not be assumed to universally reflect selective maintenance of specific amino acids. The absence of editing at 16 otherwise well-conserved genes, including *rbcL*, *psbD*, and several small photosystem subunits, likely reflects a genuine absence of C-to-U mismatches at the genomic level in this lineage rather than a detection artifact.

4.5. Phylogenomic placement and taxonomic implications

Phylogenomic reconstruction using both Maximum Likelihood and Bayesian Inference recovered *Aloe* as a strongly supported monophyletic clade (BS = 100%, PP = 1), excluding *Aloidendron pillansii*, the only segregate alooid genus included in our outgroup sampling, along with the remaining Asphodelaceae and Amaryllidaceae outgroups, consistent with the narrower, post-2020 circumscription of *Aloe* sensu stricto following the segregation of *Aloidendron*, *Aloiampelos*, and *Kumara* as independent genera (Grace et al., 2015; Klopper et al., 2020) and with the monophyly of *Aloe* relative to these segregates recently demonstrated on a much larger nuclear dataset

(Woudstra et al., 2025). *Aloe weloensis* was placed with moderate support (BS = 94%, PP = 1) alongside *A. perfoliata* and *A. marlothii*, indicating a close phylogenetic affinity among these three species despite their considerable geographic separation. This plastid-based topology, however, should be interpreted alongside independent nuclear evidence. A recent nuclear phylogenomic analysis of the alooid clade using target-capture sequencing (Woudstra et al., 2025) also sampled *A. weloensis* and placed it, together with *A. citrina* and *A. congdonii*, at some phylogenetic distance from *Aloe* sect. *Pictae*, a section that includes several taxa closely allied with *A. weloensis's* plastid-inferred neighbors in the present study. Discordance between plastid and nuclear topologies is not unusual within *Aloe* and has been attributed elsewhere in the alooid clade to incomplete lineage sorting or historical hybridization (Malakasi et al., 2019). Given that chloroplast DNA is maternally inherited and non-recombining, the plastid tree presented here reflects a single maternal lineage history rather than the full, potentially reticulate evolutionary history of the species, and should be treated as a robust hypothesis for the immediate relatives of *A. weloensis* rather than a definitive resolution of its taxonomic placement. Future work incorporating nuclear markers, ideally cross-referenced against the (Woudstra et al., 2025) dataset in which *A. weloensis* is already represented would be significant to reconcile these two independent lines of evidence.

4.6. Implications for conservation genetics

Beyond its phylogenetic and structural findings, this study provides the first genomic resource for a species already facing documented habitat pressure from urban and agricultural expansion in the Welo highlands. The identified hypervariable loci and SSR markers offer an actionable panel for future population-level genotyping, enabling assessment of genetic diversity and structure among the geographically disjunct localities where *A. weloensis* is currently known to persist. The present study, however, is based on a single individual and cannot itself capture intraspecific genetic diversity or the degree of isolation among these subpopulations, information that would be directly relevant to prioritizing populations for in situ protection or ex situ seed banking. Extending this work through population-level resequencing across the species' known range, together with wet-lab validation of the candidate SSR and hypervariable markers proposed here, represents the logical next step for translating these genomic resources into concrete, evidence-based conservation action for this narrowly distributed Ethiopian endemic species.

5. Conclusion

This study reports the first complete chloroplast genome of *A. weloensis*, a narrowly distributed, habitat-threatened endemic of the Ethiopian Welo highlands. At 153,001 bp, the genome exhibits the conserved quadripartite structure and gene content typical of the genus, encoding 130 genes with minor, IR-boundary-driven size variation relative to its congeners. Comparative analysis across ten *Aloe* species identified a consistent set of hypervariable coding genes (*rpl22, psbM, ndhF, rpl16, matK,* and *ndhD*, among others) and intergenic spacers (including *trnQ-UUG–psbK, petN–psbM,* and *petD–rpoA*), which we propose as candidate markers for species identification, population-level studies, and future germplasm conservation efforts.

The RNA editing analysis, the first conducted for this genus, is consistent with a role in restoring conserved, functionally important residues in chloroplast-encoded proteins. Phylogenomic reconstruction using both maximum likelihood and Bayesian approaches placed *A. weloensis* within a well-supported clade alongside *A. perfoliata* and *A. marlothii*, consistent with a monophyletic *Aloe*, though this plastid-based topology should be considered alongside independent nuclear phylogenomic evidence as further taxon sampling becomes available. Beyond its phylogenetic and evolutionary value, this genome provides an immediately usable reference and a validated panel of candidate molecular markers for a species already facing documented habitat pressure from urban and agricultural expansion in the Welo highlands. Extending this work to multiple individuals across the species known localities represents the logical future prospect for translating these genomic resources into concrete, population-informed conservation actions.

Supplementary information

The supplementary information files include the tables and figures used to support the results can be downloaded at: [-]. Table S1: The three local centres of *Aloe* endemism recognised in Ethiopia (Geographic Data Source accessed from GBIF). Table S2. Relative synonymous codon usage (RSCU) of *A. weloensis* chloroplast genome. Table S3. Simple Sequence Repeats (SSRs). Table S4. Long Sequence Repeats (LSRs). Table S5. RNA editing genes and events in *A. weloensis* complete chloroplast genomes dataset. Table S6. Amino acid conversions in the chloroplast genomes of *A. woleonsis*. Table S7: Amino acid conversions among *A. weloensis* complete plastomes.

Author contributions

G.W.H: Conceptualization, Funding acquisition, Project administration, Supervision, validation, Resources, reviewing and proofreading. **B.M.W**: Plants sample collection, Writing-original draft, Methodology, data curation, validation and Formal analysis. **S.D**: Supervision, Specimen Identification, reviewing and proofreading, **S.P** and **C.Z:** Supervision, Conceptualization and Validation, **J.Z.B**, **T.M.T**, **A.F.M**, and **J.K.K**: Reviewing and proofreading, **E.N.W**: Software, methodology and visualization, **Q.F.W**: Resources, supervision, and validation. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by grants from the Key Research and Development Project of Hunan Province's Science and Technology Innovation Program (2025WK2012), the Key Project at Central Government Level: The Ability Establishment of Sustainable Use for Valuable Chinese Medicine Resources (2060302), the National Natural Science Foundation of China (32270228), the Project for Scientists in Biological Taxonomy of the Chinese Academy of Sciences (CAS-TAX-24-39), the Biological Resources Program, Chinese Academy of Sciences (KFJ-BRP-017-098), and the Sino-Africa Joint Research Center, CAS (SAJC202401).

Data availability

The complete chloroplast genomes of *A. weloensis* were deposited at the GenBank repository under accession number QB012994.

Statements and declarations of competing interest

The authors declare no conflicts of interest.

References

Abdullah, Haram, D., Yan, R., Ma, M., Naz, I., Ullah, R., Shah, S. A., Ahmed, I., Heidari, P., & Tian, X. (2025). Inverted repeats dynamics shape Asclepiadoideae (Apocynaceae) Chloroplast genomes: effects on genome size, gene content, structural arrangement, and mutation rate. *BMC Genomics*, *26*(1). https://doi.org/10.1186/s12864-025-11839-9

Bachman, S. P., Wilkin, P., Reader, T., Field, R., Weber, O., Nordal, I., & Demissew, S. (2020). Extinction risk and conservation gaps for *Aloe* (Asphodelaceae) in the Horn of Africa. *Biodiversity and Conservation*, *29*(1), 77–98. https://doi.org/10.1007/S10531-019-01870-0

Belayneh, A., Demissew, S., Bussa, N. F., & Bisrat, D. (2020). Ethno-medicinal and bio-cultural importance of *Aloe*s from south and east of the Great Rift Valley floristic regions of Ethiopia. *Heliyon*, *6*(6). https://doi.org/10.1016/j.heliyon.2020.e04344

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/BIOINFORMATICS/BTP348

Cauz-Santos, L. A. (2025). Beyond conservation: the landscape of chloroplast genome rearrangements in angiosperms. *New Phytologist*, *247*(6), 2571–2580. https://doi.org/10.1111/NPH.70364

Cauz-Santos, L. A., da Costa, Z. P., Callot, C., Cauet, S., Zucchi, M. I., Bergès, H., van den Berg, C., & Vieira, M. L. C. (2020). A Repertory of Rearrangements and the Loss of an Inverted Repeat Region in Passiflora Chloroplast Genomes. *Genome Biology and Evolution*, *12*(10), 1841–1857. https://doi.org/10.1093/GBE/EVAA155

Chekole, G. (2017). Ethnobotanical study of medicinal plants used against human ailments in Gubalafto District, Northern Ethiopia. *Journal of Ethnobiology and Ethnomedicine 2017 13:1*, *13*(1), 55-. https://doi.org/10.1186/S13002-017-0182-7

Chen, K., Liu, Y. C., Huang, Y., Wu, X. K., Ma, H. Y., Peng, H., Li, D. Z., & Ma, P. F. (2025). Reassessing the phylogenetic relationships of Pseudosorghum and Saccharinae (Poaceae) using plastome and nuclear ribosomal sequences. *Plant Diversity*, *47*(3), 382–393. https://doi.org/10.1016/j.pld.2025.03.002

Chen S, Z. Y. C. Y. et al. (2018). *fastp: an ultra-fast all-in-one FASTQ preprocessor*.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, *34*(17), i884–i890. https://doi.org/10.1093/BIOINFORMATICS/BTY560

Cock, I. E. (2015). The genus *Aloe*: Phytochemistry and therapeutic uses including treatments for gastrointestinal conditions and chronic inflammation. *Progress in Drug Research*, *70*, 179–235. https://doi.org/10.1007/978-3-0348-0927-6_6

Daniell, H., Lin, C. S., Yu, M., & Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology 2016 17:1*, *17*(1), 134-. https://doi.org/10.1186/S13059-016-1004-2

Demissew, S., Friis, I., Awas, T., Wilkin, P., Weber, O., Bachman, S., & Nordal, I. (2011). Four new species of *Aloe* (Aloaceae) from Ethiopia, with notes on the ethics of describing new taxa from foreign countries. *Kew Bulletin*, *66*(1), 111–121. https://doi.org/10.1007/S12225-011-9263-2

Dong, S., Zhou, M., Zhu, J., Wang, Q., Ge, Y., & Cheng, R. (2022). The complete chloroplast genomes of Tetrastigma hemsleyanum (Vitaceae) from different regions of China: molecular structure, comparative analysis and development of DNA barcodes for its geographical origin discrimination. *BMC Genomics*, *23*(1), 620. https://doi.org/10.1186/S12864-022-08755-7

Duan, L., Li, S. J., Su, C., Sirichamorn, Y., Han, L. N., Ye, W., Lôc, P. K., Wen, J., Compton, J. A., Schrire, B., Nie, Z. L., & Chen, H. F. (2021). Phylogenomic framework of the IRLC legumes (Leguminosae subfamily Papilionoideae) and intercontinental biogeography of tribe Wisterieae. *Molecular Phylogenetics and Evolution*, *163*, 107235. https://doi.org/10.1016/J.YMPEV.2021.107235

Edwards, S., Demissew, S., & Hedberg, I. (1997). *Flora of Ethiopia and Eritrea Hydrocharitaceae to Arecaceae*. 612. https://books.google.com/books/about/Flora_of_Ethiopia_and_Eritrea.html?eng, L.,

Zhao, G., An, M., Wang, C., & Yin, Y. (2023). Complete chloroplast genome sequences of the ornamental plant Prunus cistena and comparative and phylogenetic analyses with its closely related species. *BMC Genomics*, *24*(1), 1–13. https://doi.org/10.1186/S12864-023-09838-9

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., & Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Research*, *32*(Web Server issue). https://doi.org/10.1093/NAR/GKH458

Gao, C., Li, T., Zhao, X., Wu, C., Zhang, Q., Zhao, X., Wu, M., Lian, Y., & Li, Z. (2023). Comparative analysis of the chloroplast genomes of Rosa species and RNA editing analysis. *BMC Plant Biology 2023 23:1*, *23*(1), 318-. https://doi.org/10.1186/s12870-023-04338-0

Grace, O. M., Buerki, S., Symonds, M. R. E., Forest, F., Van Wyk, A. E., Smith, G. F., Klopper, R. R., Bjorå, C. S., Neale, S., Demissew, S., Simmonds, M. S. J., & Rønsted, N. (2015). Evolutionary history and leaf succulence as explanations for medicinal use in *Aloe*s and the global popularity of *Aloe* vera. *BMC Evolutionary Biology 2015 15:1*, *15*(1), 29-. https://doi.org/10.1186/S12862-015-0291-7

Grace, O. M., Klopper, R. R., Smith, G. F., Crouch, N. R., Figueiredo, E., Rønsted, N., & van Wyk, A. E. (2013). A revised generic classification for *Aloe* (Xanthorrhoeaceae subfam. Asphodeloideae). *Phytotaxa*, *76*(1), 7-14–17–14. https://doi.org/10.11646/PHYTOTAXA.76.1.2

Grace, O. M., Simmonds, M. S. J., Smith, G. F., & van Wyk, A. E. (2008). Therapeutic uses of *Aloe* L. (Asphodelaceae) in southern Africa. *Journal of Ethnopharmacology*, *119*(3), 604–614. https://doi.org/10.1016/J.JEP.2008.07.002

Guildford Cycads. (2025). *Aloe weloensis: The Welo Aloe – Ethiopian Endemic*. https://www.guildfordcycads.com.au/*Aloe*-weloensis-the-welo-*Aloe*-ethiopianendemic/

He, Z. L., Xiao, J. F., Zhao, F., Sukhorukov, A. P., & Xiang, C. L. (2026). Comparative chloroplast genomics of Limeum (Limeaceae) species: New insights into genome evolution and phylogenomic implications. *Korean Journal of Plant Taxonomy*, *56*(3), 224–242. https://doi.org/10.11110/kjpt.2026.56.3.224

Henriquez, C. L., Abdullah, Ahmed, I., Carlsen, M. M., Zuluaga, A., Croat, T. B., & McKain, M. R. (2020). Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics*, *112*(3), 2349–2360. https://doi.org/10.1016/J.YGENO.2020.01.006

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLOS ONE*, *6*(5), e19254. https://doi.org/10.1371/JOURNAL.PONE.0019254

Hu, G., Wu, Y., Guo, C., Lu, D., Dong, N., Chen, B., Qiao, Y., Zhang, Y., & Pan, Q. (2022). Haplotype Analysis of Chloroplast Genomes for Jujube Breeding. *Frontiers in Plant Science*, *13*, 841767. https://doi.org/10.3389/FPLS.2022.841767

Huang, S., Ge, X., Cano, A., Salazar, B. G. M., & Deng, Y. (2020). Comparative analysis of chloroplast genomes for five Dicliptera species (Acanthaceae): Molecular structure, phylogenetic relationships, and adaptive evolution. *PeerJ*, *2020*(2), e8450. https://doi.org/10.7717

IUCN. (2025). *IUCN Red List of Threatened Species*. https://www.iucnredlist.org/

Jaiswal, S. K., Mahajan, S., Chakraborty, A., Kumar, S., & Sharma, V. K. (2021). The genome sequence of *Aloe* vera reveals adaptive evolution of drought tolerance mechanisms. *IScience*, *24*(2), 102079. https://doi.org/10.1016/j.isci.2021.102079

Jiang, Y., Fan, S. L., Song, M. Z., Yu, J. N., & Yu, S. X. (2012). Identification of RNA editing sites in cotton (Gossypium hirsutum) chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genetics and Molecular Research: GMR*, *11*(2), 987–1001. https://doi.org/10.4238/2012

Jin, J. J., Yu, W. Bin, Yang, J. B., Song, Y., Depamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*(1), 1–31. https://doi.org/10.1186/S13059-020-02154-5

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/MOLBEV/MST010

Khabbazi, S. D. (2026). Comprehensive analysis of the chloroplast genome in Cannabis sativa L. reveals variations in simple sequence repeats among cultivars. *Scientific Reports 2025 16:1*, *16*(1), 206-. https://doi.org/10.1038/s41598-025-28205-0

Khakhlova, O., & Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal*, *46*(1), 85–94. https://doi.org/10.1111/J.1365-313X.2006.02673

Klopper, R. R., Crouch, N. R., Smith, G. F., & van Wyk, A. E. (2020). A synoptic review of the *Aloe*s (Asphodelaceae, Alooideae) of KwaZulu-Natal, an ecologically diverse province in eastern South Africa. *PhytoKeys 142: 1-88*, *142*, 1–88. https://doi.org/10.3897/PHYTOKEYS.142.48365

Kurniawan, D., Li, L., Liu, M., Aquavita, J. M. W., Napitupulu, M. B., Narsing Rao, M. P., Turhadi, T., Zhang, S., Wei, T., & Sahu, S. K. (2025). Chloroplast genomes of six *Aloe* species: Insights into genetic variation and evolutionary patterns. *Plant Gene*, *43*, 100530. https://doi.org/10.1016/J.PLGENE.2025.100530

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, *29*(22), 4633–4642. https://doi.org/10.1093/NAR/29.22.4633

Lenz, H., Hein, A., & Knoop, V. (2018). Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics 2018 19:1*, *19*(1), 255-. https://doi.org/10.1186/S12859-018-2244-9

Li, J., Wang, H., Wang, L., Wang, X., Jia, L., & Chen, Z. (2025). Comprehensive analysis of the complete chloroplast genome of the cultivated soapberry and phylogenetic relationships of Sapindaceae. *Industrial Crops and Products*, *228*, 120952. https://doi.org/10.1016/J.INDCROP.2025.120952

Liu, F., Chen, N., Wang, H., Li, J., Wang, J., & Qu, F. (2023). Novel insights into chloroplast genome evolution in the green macroalgal genus Ulva (Ulvophyceae, Chlorophyta). *Frontiers in Plant Science*, *14*, 1126175. https://doi.org/10.3389/fpls.2023.1126175

Luo, C., Huang, W., Sun, H., Yer, H., Li, X., Li, Y., Yan, B., Wang, Q., Wen, Y., Huang, M., & Huang, H. (2021). *Comparative Chloroplast Genome Analysis of Impatiens Species (Balsaminaceae) in the Karst Area of China: Insights Into Genome Evolution and Phylogenomic Implications*. https://doi.org/10.21203/rs.3.rs-132878/v1

Makhmudjanov, D., Abdullaev, D., Juramurodov, I., Tuychiev, S., Yusupov, Z., Sun, H., Tojibaev, K., & Deng, T. (2023a). Comparative Analysis and Characterization of Ten Complete Chloroplast Genomes of Eremurus Species (Asphodelaceae). *Forests*, *14*(9), 1709. https://doi.org/10.3390/F14091709/S1

Makhmudjanov, D., Abdullaev, D., Juramurodov, I., Tuychiev, S., Yusupov, Z., Sun, H., Tojibaev, K., & Deng, T. (2023b). Comparative Analysis and Characterization of Ten Complete Chloroplast Genomes of Eremurus Species (Asphodelaceae). *Forests*, *14*(9), 1709. https://doi.org/10.3390/F14091709/S1

Malakasi, P., Bellot, S., Dee, R., & Grace, O. M. (2019). Museomics Clarifies the Classification of Aloidendron (Asphodelaceae), the Iconic African Tree *Aloe*s. *Frontiers in Plant Science*, *10*, 479776. https://doi.org/10.3389/FPLS.2019.01227

Mehmood, F., Li, M., Bertolli, A., Prosser, F., & Varotto, C. (2024). Comparative Plastomics of Plantains (Plantago, Plantaginaceae) as a Tool for the Development of Species-Specific DNA Barcodes. *Plants*, *13*(19). https://doi.org/10.3390/plants13192691

Mengesha, Y. (2024). A Review on *Aloe* weloensis: Unveiling its Phytochemical Composition and Promising Pharmacological Properties in Antioxidant, Antibacterial, and Antimalarial Activities. In *Natural Product Communications* (Vol. 19, Number 9). SAGE Publications Inc. https://doi.org/10.1177/1934578X241285207

Morton, B. R. (2003). The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *Journal of Molecular Evolution*, *56*(5), 616–629. https://doi.org/10.1007/S00239-002-2430-1

Mower, J. P., Ma, P. F., Grewe, F., Taylor, A., Michael, T. P., VanBuren, R., & Qiu, Y. L. (2019). Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple

inversions between a direct and inverted orientation of the rRNA repeat. *New Phytologist*, *222*(2), 1061–1075. https://doi.org/10.1111/nph.15650

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/MOLBEV/MSU300

Ni, L., Zhao, Z., Xu, H., Chen, S., & Dorje, G. (2016). The complete chloroplast genome of Gentiana straminea (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene*, *577*(2), 281–288. https://doi.org/10.1016/j.gene.2015.12.005

Nigussie, G., & Wale, M. (2022). Medicinal plants used in traditional treatment of malaria in Ethiopia: a review of ethnomedicine, anti-malarial and toxicity studies. *Malaria Journal 2022 21:1*, *21*(1), 262-. https://doi.org/10.1186/S12936-022-04264-W

Nyffeler, R., & Eggli, U. (2010). An up-to-date familial and suprafamilial classification of succulent plants. *Bradleya*, *28*, 125–144. https://doi.org/10.25223/BRAD.N28.2010.A1

Oda, B. K., & Erena, B. A. (2017). *Aloe*s of Ethiopia: A Review on Uses and Importance of *Aloe*s in Ethiopia. *International Journal of Plant Biology*, *5*(1), 1059. https://www.researchgate.net/publication/325485776

Parks, M., Cronn, R., & Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, *7*(1), 1–17. https://doi.org/10.1186/1741-7007-7-84/FIGURES/8

Qu, X. J., Moore, M. J., Li, D. Z., & Yi, T. S. (2019). PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods 2019 15:1*, *15*(1), 50-. https://doi.org/10.1186/S13007-019-0435-7

Rambaut. (2018). *FigTree v1.4.4*. https://github.com/rambaut/figtree/releases

Rodrigues, N. F., Christoff, A. P., da Fonseca, G. C., Kulcheski, F. R., & Margis, R. (2017). Unveiling chloroplast RNA editing events using next generation small RNA sequencing data. *Frontiers in Plant Science*, *8*, 285741. https://doi.org/10.3389/FPLS.2017.01686

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/MOLBEV/MSX248

Sánchez, M., González-Burgos, E., Iglesias, I., & Gómez-Serranillos, M. P. (2020). Pharmacological Update Properties of *Aloe* Vera and its Major Active Constituents. *Molecules 2020, Vol. 25, Page 1324*, *25*(6), 1324. https://doi.org/10.3390/MOLECULES25061324

Shaw, J., Lickey, E. B., Schilling, E. E., & 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. *American Journal of Botany*, *94*(3), 275–288. https://doi.org/10.3732/AJB.94.3.275

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 Research*, *47*(W1), W65–W73. https://doi.org/10.1093/NAR/GKZ345

Shi, W., Song, W., Liu, J., Shi, C., & Wang, S. (2023). Comparative chloroplast genome analysis of Citrus (Rutaceae) species: Insights into genomic characterization, phylogenetic relationships, and discrimination of subgenera. *Scientia Horticulturae*, *313*, 111909. https://doi.org/10.1016/J.SCIENTA.2023.111909

Song, W., Chen, Z., He, L., Feng, Q., Zhang, H., Du, G., Shi, C., & Wang, S. (2022). Comparative Chloroplast Genome Analysis of Wax Gourd (Benincasa hispida) with Three Benincaseae Species, Revealing Evolutionary Dynamic Patterns and Phylogenetic Implications. *Genes*, *13*(3), 461. https://doi.org/10.3390/GENES13030461/S1

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, *99*(2), 349–364. https://doi.org/10.3732/AJB.1100335

Su, Y., Wei, W., Han, L., Wen, H., & Lu, H. (2025). Comprehensive analysis of complete Chloroplast genome sequence of Cardiospermum halicacabum L. (Sapindaceae). *Scientific Reports*, *15*(1), 1–12. https://doi.org/10.1038/s41598-025-05679-6

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564–577. https://doi.org/10.1080/10635150701472164

Talkah, N. S. M., Aziz, N. A. K. A., Rahim, M. F. A., Hanafi, N. F. F., Mokhtar, M. A. A., & Othman, A. S. (2024). The chloroplast genome inheritance pattern of the Deli-Nigerian prospection material (NPM) x Yangambi population of Elaeis guineensis Jacq. *PeerJ*, *12*(5), e17335. https://doi.org/10.7717/PEERJ.17335

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/MOLBEV/MSAB120

Teka, T., Awgichew, T., & Kassahun, H. (2020). Antimalarial Activity of the Leaf Latex of *Aloe* weloensis (Aloaceae) against Plasmodium berghei in Mice. *Journal of Tropical Medicine*, *2020*. https://doi.org/10.1155/2020/1397043

Woudstra, Y., Rees, P., Rakotoarisoa, S. E., Klopper, R. R., Smith, G. F., Rønsted, N., & Grace, O. M. (2025). Nuclear phylogenomics reveals strong geographic patterns in the evolutionary history of *Aloe* and related genera (alooids). *Annals of Botany*. https://doi.org/10.1093/AOB/MCAF109

Wu, C. S., & Chaw, S. M. (2014). Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnology Journal*, *12*(3), 344–353. https://doi.org/10.1111/PBI.12141

Wu, J., Gao, Y., Wang, J., Guo, A., Qin, N., Xing, G., & Li, S. (2024). Comparative analysis of chloroplast genome and evolutionary history of Hemerocallis. *Frontiers in Genetics*, *15*, 1433548. https://doi.org/10.3389/FGENE.2024.1433548

Wu, J., Zhang, J., Guo, X., Yu, N., Peng, D., & Xing, S. (2023). Comprehensive analysis of complete chloroplast genome sequence of Plantago asiatica L. (Plantaginaceae). *Plant Signaling & Behavior*, *18*(1), 2163345. https://doi.org/10.1080/15592324.2022.2163345

Xie, D. F., Yu, Y., Deng, Y. Q., Li, J., Liu, H. Y., Zhou, S. D., & He, X. J. (2018). Comparative analysis of the chloroplast genomes of the chinese endemic genus urophysa and their contribution to chloroplast phylogeny and adaptive evolution. *International Journal of Molecular Sciences*, *19*(7). https://doi.org/10.3390/ijms19071847

Xu, Y., Chen, X., Melkonian, M., Wang, S., & Sahu, S. K. (2024). Comparative chloroplast genome analysis of two Desmodesmus species reveals genome diversity within Scenedesmaceae (Sphaeropleales, Chlorophyceae). *Protist*, *175*(6), 126073. https://doi.org/10.1016/J.PROTIS.2024.126073

Yao, S. T., Lu, Y. Bin, Zhang, Z. J., Li, C., Wang, P. F., & Qin, X. M. (2023). The complete chloroplast genome of Chloranthus nervosus Collett ex Hemsl. 1890 (Chloranthaceae). *Mitochondrial DNA Part B*, *8*(11), 1224–1228. https://doi.org/10.1080/23802359.2023.2278818

Zhang, C., Li, W., Wu, Y., Li, S., Hua, B., & Sun, H. (2025). Chloroplast Functionality at the Interface of Growth, Defense, and Genetic Innovation: A Multi-Omics and Technological Perspective. *Plants 2025, Vol. 14, Page 978*, *14*(6), 978. https://doi.org/10.3390/PLANTS14060978

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*, *20*(1), 348–355. https://doi.org/10.1111/1755-0998.13096

Zhang, J., Huang, H., Qu, C., Meng, X., Meng, F., Yao, X., Wu, J., Guo, X., Han, B., & Xing, S. (2021). Comprehensive analysis of chloroplast genome of *Albizia julibrissin* Durazz. (Leguminosae sp.). *Planta 2021 255:1*, *255*(1), 26-. https://doi.org/10.1007/S00425-021-03812-Z

Zhang, L., Meng, Y., Wang, D., He, G.-H., Zhang, J.-M., Wen, J., & Nie, Z.-L. (2024). *RESEARCH Open Access Plastid genome data provide new insights into the dynamic evolution of the tribe Ampelopsideae (Vitaceae)*. https://doi.org/10.1186/s12864-024-10149-w

Zhang, Y., Shi, Y. bo, Wang, H., Huang, X. ci, Yue, J. hua, & Wang, C. zhong. (2024). The complete chloroplast genome of Dianella ensifolia (Asphodelaceae). Mitochondrial DNA Part B: Resources, 9(3), 322–326. https://doi.org/10.1080/23802359.2024.2323003

974 Zheng, S., Poczai, P., Hyvönen, J., Tang, J., & Amiryousefi, A. (2020). Chloroplot: An Online Program
975 for the Versatile Plotting of Organelle Genomes. *Frontiers in Genetics*, *11*, 576124.
976 https://doi.org/10.3389/FGENE.2020.576124

977 Zou, T., Li, D., Zhao, C. Y., & Chen, M. L. (2025). Chloroplast whole genome assembly and
978 phylogenetic analysis of Persicaria criopolitana reveals its new taxonomic status. *Scientific*
979 *Reports*, *15*(1), 1–10. https://doi.org/10.1038/S41598-025-02686-5

980

981

982

983

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Aloesupplementary.docx