Plastid super-barcoding in the pan-plastomic era: dense sampling of Pulsatilla patens plastomes exposes the limits of species delimitation in reticulate complexes

Monika Szczecińska

monika.szczecinska@uwm.edu.pl

University of Warmia and Mazury in Olsztyn

Mateusz Maździarz

University of Warmia and Mazury in Olsztyn

Kamil Koczwara

University of Warmia and Mazury in Olsztyn

Aniela Garbowska

University of Warmia and Mazury in Olsztyn

Jakub Sawicki

University of Warmia and Mazury in Olsztyn

Research Article

Keywords: chloroplast genome, DNA barcoding, introgression, Pleistocene, reticulate evolution, Pulsatilla patens, super-barcoding

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9672477/v1

Additional Declarations: No competing interests reported.

Abstract

Background.

Whole-plastome "super-barcoding" is widely applied to resolve taxonomic uncertainties in complex plant groups. In lineages prone to reticulate evolution, such as the tribe Anemoneae, sparse sampling can overestimate species-level diagnostic power. We evaluated the utility of complete chloroplast genomes for the identification of *Pulsatilla patens* by expanding the sampling to 33 plastomes within a broader genus-level dataset (49 plastomes), aiming to distinguish true species divergence from deep population structure.

Results.

We identified high genomic variability in *P. patens*, with a total of 2,127 mutations (1,360 indels and 767 SNPs). Variation was strongly clustered, with hotspots concentrated in *ycf1*, *rpoC2*, and *ndhF*. Despite this richness, dense sampling led to the collapse of taxonomic resolution. *P. patens* was resolved as paraphyletic, with *P. vernalis* specimens nested within its variation. After excluding introgressed individuals and divergent subspecies, the number of fixed diagnostic nucleotides for *P. patens* s.s. dropped to merely two. Phylogenetic structure showed a mosaic pattern (LMEM analysis) that reflected ancient Pleistocene connectivity rather than contemporary geography or habitat type.

Conclusions.

Our results demonstrate a "super-barcoding paradox" in *Pulsatilla*: increasing data volume combined with dense sampling eliminates apparent diagnostic markers due to pervasive historical introgression and shared ancestral polymorphism. In reticulate complexes, the plastome functions as a record of ancient connectivity (the "ghost" of the Mammoth Steppe) rather than a straightforward taxonomic barcode. Practical barcoding should target identified hotspots but must be integrated with nuclear markers to reflect contemporary isolation and guide effective conservation management.

Introduction

Over the past two decades, DNA barcoding has advanced plant taxonomy, supporting cryptic species identification and providing tools for conservation genetics, forensic botany, and environmental genomics [1, 2]. Unlike the mitochondrial COI marker in animals, however, a single universal plant barcode remains elusive [3, 4]. Standard loci often lack the discriminatory power required to delimit evolutionarily young lineages or closely related taxa, where incomplete lineage sorting mimics species divergence. To address these limitations, contemporary research increasingly employs complete chloroplast genomes as "super-barcodes" [5, 6]. Advances in next-generation sequencing (NGS) have made plastome assembly cost-effective and accessible, providing a conserved yet variable source of

data (typically 110–160 kbp) that exceeds the variability of standard markers. Using the entire plastome also circumvents issues associated with gene deletion or PCR failure that frequently impede traditional single-locus approaches [7].

Two genomic strategies have been deployed to address taxonomic challenges within the genus *Pulsatilla*: whole-plastome super-barcoding and the targeted use of specific mutational hotspots, such as the highly variable *ycf1* gene [8] or the *ndhD–ccsA* intergenic spacer [9]. These approaches are typically validated using limited datasets, often comprising only one or two reference individuals per species. Such sparse sampling overlooks the structural complexities of Anemoneae plastomes, including inversions and shifts in Inverted Repeat (IR) boundaries [9, 10], and fails to account for substantial intraspecific variation. Phylogenies based solely on typological plastid data may therefore yield misleading topologies due to processes such as introgression [11], arguing for a shift from simple barcoding to a pan-plastomic perspective.

Pan-plastomic studies, which analyse complete genomes across multiple populations and individuals, are essential for distinguishing true species-level divergence from population structure [12, 13]. This distinction is particularly critical for rare and endangered species, where understanding genetic diversity and local adaptation is a prerequisite for effective conservation strategies [14, 15].

To test the limits of plastid super-barcoding empirically, we selected the Eastern Pasqueflower (*Pulsatilla patens* (L.) Mill.) as a model. This conservation-priority taxon, strictly protected across Europe under the Bern Convention and the Habitats Directive [16, 17], is an ideal candidate for pan-plastomic analysis due to its severe population fragmentation and documented genetic erosion caused by habitat loss. Once widespread across Europe, *P. patens* has undergone drastic decline following the cessation of traditional land management practices, such as grazing and litter raking, leaving a fragmented landscape of isolated populations well suited to testing intraspecific genomic variation [18–20].

Earlier population genetic analyses based on nuclear microsatellites (SSR) revealed a complex genetic structure across the European range of *P. patens*, identifying significant isolation between populations. These studies distinguished three discrete genetic lineages: two "Baltic" clusters and a divergent "Southern" group with reduced allelic richness, likely a relic of historical fragmentation [21]. Preliminary plastome studies suggested that this strong nuclear structure might be mirrored by plastid variability [9], but were constrained by sparse sampling regimes [22]. It remains unverified whether deep intraspecific divergence in chloroplast genomes driven by this geographical isolation could exceed interspecific distances and blur species boundaries, confounding standard barcoding identification.

To resolve this uncertainty and bridge the gap between population genetics and phylogenomics, we expanded both taxon and sample coverage and evaluated the robustness of plastid super-barcodes. We analysed 49 complete chloroplast genomes belonging to 12 *Pulsatilla* species, including 33 plastomes of *P. patens* representing the majority of its Central European range and covering diverse habitats, from xerothermic grasslands to pine forests [21]. We hypothesised that dense intraspecific sampling would reveal high levels of isolation-driven variation that might overlap with interspecific divergence, thereby

challenging the utility of standard "diagnostic nucleotides" and simple distance-based identification. We aimed to (1) determine whether super-barcoding can universally identify closely related species despite profound population structure, and (2) quantify how dense sampling affects the stability and number of diagnostic markers. The results provide a critical assessment of the resolution limits of genomic methods in plant taxonomy and identify the evolutionary mechanisms responsible for incongruence between phylogenetic trees and traditional species boundaries.

Materials and Methods

Plant material

The plant material analysed comprised 49 complete chloroplast genomes (plastomes) representing 12 taxa from the genus *Pulsatilla* and the outgroup *Anemone cylindrica* (Table 1, Supplementary). The focal species was *P. patens*, for which we assembled a dense intraspecific dataset of 33 plastomes for pan-plastomic analysis. Sampling encompassed populations across Central Europe inhabiting both open xerothermic grasslands and shaded pine forests as well as previously delineated genetic lineages (including the relic "Southern" group) and the Asian subspecies *P. patens* subsp. *multifida*. The final dataset comprised 32 plastomes newly sequenced for this study (31 *P. patens* and one *P. × hackelii*), 6 sequences derived from our previous research, and 11 reference genomes retrieved from public databases. representing the remaining *Pulsatilla* taxa (*P. vernalis*, *P. pratensis*, *P. grandis*, *P. dahurica*, *P. chinensis*, *P. saxatilis*, *P. campanella*, *P. alpina*, *P. occidentalis*, *P. patens* subsp. *multifida*) and the outgroup *A. cylindrica*.

Detailed information on formal plant identification, legal collection permits, and voucher specimens is provided in the Ethics approval and consent to participate section.

DNA extraction

Total genomic DNA for short-read sequencing was extracted from 1 cm² of fresh leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Tissue was homogenised with silica beads in a MiniBead-Beater for 50 s and processed following the manufacturer's protocol. For long-read nanopore sequencing, high-molecular-weight DNA was isolated from a single leaf using a modified CTAB protocol [23] supplemented with PVP (polyvinylpyrrolidone) to remove phenolics. DNA quantity was estimated using a Qubit fluorometer with the Qubit dsDNA BR Assay Kit (Invitrogen, Carlsbad, NM, USA). DNA quality was assessed by electrophoresis on a 0.5% agarose gel stained with Eurx Simple Safe (Eurx, Gdańsk, Poland). Prior to long-read sequencing, the extracted DNA was further purified using the Genomic DNA Clean and Concentrator kit (Zymo Research, Irvine, CA, USA).

Short- and long-read sequencing

Genomic libraries for short-read sequencing were constructed using the TruSeq DNA PCR-Free Kit (Illumina, San Diego, CA, USA). Sequencing was performed on the NovaSeq 6000 platform (Illumina) at Macrogen Inc. (Seoul, Korea) to generate 150 bp paired-end reads with a 350 bp insert size. Long-read libraries were prepared using the ONT Ligation Sequencing Kit (SQK-LSK114) and the NEBNext Companion Module. Sequencing was conducted on MinION Mk1C or PromethION P2 Solo devices using R10.4.1 flow cells. Basecalling was performed using Dorado v0.5.1 in SUP stereo-duplex mode; only reads with a Phred quality score (Q) ≥ 20 were retained for downstream analyses.

Plastome assembly and annotation

Short-read data were assembled using NOVOPlasty v.4.3.3 [24] with default settings, using previously published *P. patens* plastomes as seeds [9, 22]. High-quality long reads were assembled using Flye v.2.9.6 [25] with the --meta and --nano-hq options and a minimum overlap of 5 kbp. Assembled plastomes were manually verified and circularised in Geneious Prime 2023.2.1 (Biomatters, Auckland, New Zealand). To correct potential misassemblies—particularly across LSC inversions—and to verify IR boundaries (JLA/JLB/JSA/JSB), reads were re-mapped to the circularised genomes using the Geneious mapping algorithm (minimum overlap = 80 bp; minimum identity = 96%). Annotation was performed using the Transfer Annotation tool in Geneious Prime, with the *P. patens* reference genome (GenBank acc. no. KR297057.2) as a template. Gene models were manually curated, particularly for *ycf1* and *ndhF*, to account for known boundary shifts in the tribe Anemoneae.

Phylogenetic analysis, nucleotide diversity, and diagnostic nucleotides

A total of 49 chloroplast sequences were analysed: 32 newly sequenced for this study, 6 previously published by our team, and 11 retrieved from public databases (Supplementary Table S1). One copy of the Inverted Repeat (IRb) was removed from all sequences to avoid overweighting. Multiple sequence alignment was performed in Geneious Prime 2023.2.1 using MAFFT (auto algorithm; scoring matrix 200 PAM/k = 2; gap open penalty 1.53; offset 0.123). The resulting alignment was used to reconstruct a maximum likelihood (ML) phylogenetic tree with IQ-TREE v.2.3.6 [26]. The best-fit substitution model was selected by ModelFinder [27] according to the Bayesian Information Criterion (BIC), and branch support was assessed using bootstrap with 1,000 replicates [28]. The tree was rooted with *Anemone cylindrica*.

Diagnostic nucleotides (DNs) distinguishing *Pulsatilla* species (*P. patens*, *P. vernalis*, *P. patens* subsp. *multifida*, *P. pratensis*, *P. grandis*, *P. dahurica*, *P. chinensis*, *P. saxatilis*, *P. campanella*, *P. alpina*, *P. occidentalis*, and the outgroup *A. cylindrica*) were identified using the spider v.1.5.0 R package [29]. A diagnostic nucleotide was defined as a site where (i) all individuals of a focal taxon shared a single, non-ambiguous base, (ii) this base was absent in all other taxa, and (iii) the site contained no missing data or gaps. This definition implicitly assumes monophyly of the focal taxon; in para- or polyphyletic complexes the metric loses its conventional interpretation, a caveat we address explicitly in the Results. Nucleotide

diversity (π) was computed using the PopGenome v.2.7.7 R library [30] for defined groups: Forest, Meadow, South, Other, all *P. patens*, and all species. Both diagnostic nucleotide and diversity analyses were conducted using 500-bp sliding windows with a 100-bp step size. SNPs and indels were identified in Geneious Prime 2025.0.2 (minimum variant frequency = 0.02). Hotspots were defined as windows in the top decile of variant density.

Demographic history reconstruction

To estimate divergence times, the bootstrapped phylogenetic tree (.nwk) was imported into MEGA 11 [31]. Divergence times were calculated with the RelTime–Branch Lengths method [32, 33]. For divergence-time calibration only, we incorporated additional taxa not used elsewhere in the analyses, applying two calibration constraints with normal calibration densities at the corresponding nodes [34].

The first calibration was set at the node representing the divergence between *Anemone tomentosa* and the clade comprising all remaining sampled taxa (predominantly *Pulsatilla* spp.), with a divergence time of 14.74 Ma and a 95% confidence interval of 12.49–17.00 Ma based on previous estimates [35, 36]. The second calibration was assigned to the node indicating the divergence between the clade containing *Pulsatilla campanella* and *P. saxatilis* and the clade containing all other *Pulsatilla* spp., with a time of 4.29 Ma following Xue et al. [36]. *Clematis brevicaudata* was used as the outgroup to anchor the deepest node in this calibrated analysis, while *A. cylindrica* served as the outgroup in the main ML and DN analyses.

To reconstruct changes in effective population size (*Ne*) over time, a Bayesian Skyline Plot (BSP) analysis was conducted in BEAST v.2.7.8 [37]. The GTR + G+I substitution model was selected based on best-fit criteria. A piecewise-constant skyline model with 10 groups was employed to balance resolution against parameter overfitting (n = 38). We applied a strict molecular clock; this choice was supported by relatively low rate variation among lineages within the focal clade and by the modest temporal depth of the radiation (Pleistocene–Pliocene), where the assumption of clock-like substitution rates is generally appropriate for plastid data. Markov Chain Monte Carlo (MCMC) chains were run for 50 million generations, with sampling every 5,000 steps. Convergence (Effective Sample Size, ESS > 200) was assessed using the tracerer R package [38] following a 10% burn-in.

Longest maximum exact matches

LMEM analysis was performed to identify shared genomic regions and potential transfer events. *P. patens* sequences were classified into four groups according to habitat or geographic origin: Meadow, South, Forest, and "Other" (Chinese subsp. *multifida* populations). *P. vernalis* was included based on phylogenetic placement. Condensed de Bruijn graph-compatible files were generated using Cuttlefish v.2.2.0 [39] with a k-mer size of 31. LMEMs were subsequently identified using Graphite v.2.0.0 [40] with a minimum match length of 100 bp. Visualisations were prepared in R using phytools for tree plotting and ggplot2 v.3.5.1 for other graphics.

Results

Phylogenetic and demographic history reconstruction

The maximum likelihood (ML) tree, based on the alignment of complete plastome sequences, reconstructed phylogenetic relationships within the genus, resolving *Pulsatilla alpina* and *P. occidentalis* as the earliest diverging lineages, followed by *P. campanella* and *P. saxatilis* (Fig. 1). The next strongly supported clade comprised five species: *P. pratensis*, *P.* × *hackelii*, and *P. grandis*, forming a sister group to the Asian *P. chinensis* and *P. dahurica*.

A central result regarding the limits of super-barcoding is the structure of the *P. patens* complex. The analysis showed that this species does not form a monophyletic group. Instead, all analysed *P. patens* individuals (including *P. vernalis*) clustered within a single broad clade (100% bootstrap support), within which relationships remained largely unresolved. An exception is the subspecies *P. patens* subsp. *multifida* (populations from China), which formed a distinct, well-defined lineage.

Within European *P. patens* s.s. populations, species boundaries were blurred: the two analysed *P. vernalis* specimens did not form a distinct clade but were nested within the variation of *P. patens*, forming a series of poorly supported groups with it (Figs. 1, 2). Within this mixed complex, only three smaller clades obtained significant statistical support (BS > 70%). The first is a fully supported group (100% BS) formed by specimens P3_1, P9_4, P9_5, P18_1, P19_2, P20_1, and P21_1. The second cluster, supported by a bootstrap value of 75%, comprises specimens P19_1, P22_1, P26_1, and P29_1. The third well-supported group contains all analysed individuals from populations P4 and P8, together with specimen P9_3.

The tree topology does not reflect either geographical or habitat structure. Apart from populations P4 and P8, individuals from distant populations (e.g. P5 and P13) grouped together, with only samples from P5 forming a population-specific cluster (63% BS support). Population P9, one of the largest populations and situated in an anthropogenically disturbed habitat, showed the highest haplotype diversity, with five individuals distributed across three distinct clusters in the phylogenetic tree. The clustering of sequenced haplotypes was uncorrelated with habitat type—both meadow and forest populations formed mixed, paraphyletic clades—suggesting that adaptations to local conditions are not yet reflected in the plastome.

Molecular dating placed the origin of the studied *P. patens* complex at the Pliocene–Pleistocene boundary. The deepest divergence event, estimated at approximately 2.95 Ma, marked the separation of the reference *P. patens* subsp. *multifida* lineage from the common ancestor of the remaining populations. A subsequent split occurred in the Early Pleistocene (~ 2.06 Ma), segregating the main *P. patens* group from a distinct evolutionary lineage comprising *P. vernalis* and sample P24. Within this latter clade, the divergence is comparatively recent: the lineage leading to P24 separated from *P. vernalis* in the Middle Pleistocene, ~ 0.45 Ma. This finding underscores the deep evolutionary distinctiveness of the *P.*

vernalis+P24 lineage relative to the core *P. patens* complex, which began its independent radiation later, ~ 1.68 Ma.

The demographic history of *P. patens*, reconstructed by integrating the dated phylogenetic tree with Bayesian Skyline Plot (BSP) analysis, reveals fluctuations in effective population size (*Ne*) throughout the Middle and Late Pleistocene (Fig. 2). With a median root height estimated at ~ 865,574 years ago, the species experienced two primary expansion peaks. The first rapid growth phase occurred between 850,000 and 700,000 years ago, peaking at ~ 650,000 years ago before a regression that lasted until ~ 500,000 years ago. A second major expansion followed, stabilising between 300,000 and 250,000 years ago, during which *Ne* reached its maximum, exceeding 10^6 individuals. A systematic, accelerating demographic contraction began ~ 200,000 years ago and became most pronounced over the past 100,000 years. This recent decline likely reflects the species' response to climatic shifts, the resulting forest encroachment on its preferred open steppe habitats, and modern habitat fragmentation. The widening of the 95% highest posterior density (HPD) intervals near the present is characteristic of coalescent models and reflects the inherent uncertainty in estimating recent *Ne* from a finite number of contemporary sequences.

Longest Maximum Exact Matches (LMEM)

The visualisation of LMEMs (Fig. 3) reveals a complex, mosaic structure of plastomes within the studied *P. patens* complex, indicating no strict correlation between plastid haplotype and habitat type or taxon. The plot shows the distribution of conserved sequence blocks along genome coordinates, where colours correspond to defined habitat groups: Forest (green), Meadow (blue), South (red), and Other/multifida (yellow). The most distinct pattern was observed for individual P24_1, whose LMEM profile shows nearly complete convergence with *P. vernalis* reference sequences (P. vernalis_1 and P. vernalis_2), sharing the longest fragment of 64,598 nt. These matches cover ~ 99% of the P24_1 plastome, providing visual and quantitative evidence for chloroplast capture and justifying the reclassification of this sample to the *P. vernalis* group in subsequent analyses. The visualisation confirms close plastid affinity between *P. vernalis* and the "Southern" lineage of *P. patens*, as *P. vernalis* specimens carry a substantial proportion of fragments characteristic of the southern group (red), with the longest shared fragment measuring 45,130 bp. The Asian subspecies *P. patens* subsp. *multifida* (yellow) showed the highest LMEM diversity, possessing between 168 and 218 unique matches. Insertions characteristic of the European "Forest" genotype (green) were identified within these genomes, with the longest fragment of 11,741 nt, suggesting the retention of ancestral polymorphisms despite geographic isolation. Most European *P. patens* individuals showed a mosaic structure combining elements attributed to different habitats; for example, individual P3_1 carried LMEMs associated with both meadow (9,176 bp) and forest (5,480 bp) environments. A particular case of incongruence is individual P8_1: despite originating from a meadow population, its LMEM profile is dominated by fragments characteristic of the forest genotype (green bars), confirming the permeability of genetic barriers and the lack of a simple association between ecological morphotype and plastid haplotype.

Nucleotide diversity, SNPs, and diagnostic nucleotides

Analysis of nucleotide diversity within *P. patens* revealed high sequence variability, with a total of 2,127 mutations (1,360 indels and 767 SNPs), of which 145 were nonsynonymous changes (Fig. 4). This variation was strongly clustered in specific regions (hotspots), notably *ycf1*, *rpoC2*, and *ndhF* (Fig. 5). Despite this high level of polymorphism, the analysis of diagnostic nucleotides (DN) demonstrated an absence of unique markers for the species *sensu lato*. Because *P. patens* is paraphyletic with respect to *P. vernalis*, the standard DN definition cannot be applied without first resolving the species circumscription. Only after excluding the introgressed individual (P24_1) and the distinct subspecies *multifida* did the number of identified diagnostic nucleotides for *P. patens* s.s. recover, and even then it remained at merely 2 (Supplementary Fig. S1). The vast majority of the 2,127 detected mutations therefore represent intraspecific polymorphisms or variation shared with *P. vernalis*, rather than fixed interspecific differences.

Discussion

An ideal DNA barcode for species delimitation must combine low intraspecific variation with sufficient interspecific variation to ensure successful identification [6, 41]. In the era of Sanger sequencing, single- and multilocus barcodes were tested for many groups of plants with variable success. The advent of "super-barcodes" based on complete plastid sequences enabled the identification of even closely related and relatively young species where traditional barcodes fail [5, 6, 42]. Most studies, however, rely on single or 1–3 specimens per species, which leads to a systematic overestimation of diagnostic power.

This reliance creates a "super-barcoding paradox" that becomes visible only under dense sampling. *Pulsatilla* species were among the first for which the number of diagnostic nucleotides in the plastome was estimated, suggesting high molecular distinctness. Earlier analyses based on only two individuals per species reported hundreds of diagnostic markers, ranging from 56 for *P. vernalis* to 432 for *P. pratensis*, and 49 substitutions plus 18 indels for *P. patens* itself [9]. In contrast, our pan-plastomic analysis of 33 *P. patens* plastomes critically re-evaluates these estimates. Although genomic variability proved to be very high, we identified 2,127 mutations that did not translate into increased taxonomic resolution; instead, resolution collapsed. The number of fixed diagnostic nucleotides for *P. patens* s.s. (after excluding introgressants) dropped from dozens to merely two. In complexes characterised by reticulate evolution, increasing the volume of data combined with dense sampling does not solve identification problems but exacerbates them by eliminating apparent diagnostic markers. The issue is not a lack of variation but the fact that this genomic richness is overwhelmingly intraspecific polymorphism or shared ancestral variation.

In rapidly radiating groups, whole-plastome data frequently expose a continuum of shared ancestral polymorphism or introgression that was previously masked by the low resolution of single loci. The pattern is not unique to *Pulsatilla*; it mirrors observations in other European genera prone to reticulate evolution. In willows (*Salix*), pervasive chloroplast capture renders plastid phylogenies geographically

structured rather than taxonomically coherent [43]. In European white oaks (*Quercus*), plastid haplotypes reflect shared post-glacial migration routes across species boundaries rather than species identity [44, 45].

The failure of super-barcoding in *P. patens* differs fundamentally from cases driven by low mutational rates. Low plastome mutation rates can impede molecular species delimitation in diverse plant groups, such as the olive tree complex (*Olea europaea*) [46] and the conifer genus *Araucaria* [47], and also in the yam genus *Dioscorea* [48] and spruce species, especially within the *Picea glauca* complex [49]. The challenge extends to higher taxonomic units such as complex thalloid liverworts (Marchantiales) [50, 51], although even here very low intraspecific variation can still enable molecular identification [52]. A low plastome mutation rate is not the problem in Ranunculaceae, a family known for diversity in both species composition and plastome structure [9, 36]. Our results show extensive variation within *P. patens* —2,127 mutations—confirming that the problem is not a lack of variation, but a lack of diagnostically structured variation.

This lack of diagnostic structure stems directly from paraphyly and a complex history of hybridisation. The phylogenetic analyses nested *P. vernalis* inside the *P. patens* clade, resolving the latter as paraphyletic [9, 53, 54]. This molecular result contrasts with morphology, as *P. patens* and *P. vernalis* are exceptionally well-defined and distinct species: *P. vernalis* produces leathery, evergreen leaves, while *P. patens* has palmate–trisectate, deciduous leaves. This discordance is captured explicitly in our study by individual P24_1 from Slovakia. The geographic origin of this introgressed specimen is particularly noteworthy in an evolutionary context. During the Pleistocene glaciations, the Carpathian Mountains and the adjacent Pannonian Basin served as major glacial refugia. Retreating open-habitat species were forced into close physical proximity within these ice-free enclaves, creating secondary contact zones [53, 55]. Although morphologically identified as *P. patens*, our LMEM analysis revealed that its plastome shares ~ 99% identity with *P. vernalis* reference sequences, providing empirical evidence for complete chloroplast capture. This Slovakian population acts as a genetic time capsule from a period when these taxa formed a permeable syngameon. This pattern suggests that *Pulsatilla* likely functions as a syngameon, where distinct morphological species are maintained by strong selection on nuclear genes governing phenotype, despite a permeable barrier to organellar gene flow. The chloroplast genome is inherently more susceptible to such introgression and complete capture due to its uniparental inheritance and lack of recombination. These traits result in a smaller effective population size (*Ne*) compared to the nuclear genome, accelerating the fixation of alien plastomes through genetic drift, particularly during periods of demographic fluctuation.

Such reticulate evolution is frequently noted in the genus *Pulsatilla*, supporting the hypothesis of historical introgression [11, 56–58]. One of the most well-documented cases is the hybridisation between *P. patens* and *P. pratensis*, resulting in the hybrid *P.* × *hackelii*, which exhibits intermediate morphological characteristics. Another example is *P.* × *yanbianensis*, arising from the natural crossing of *P. dahurica* and *P. cernua* [59]. Beyond documented hybrids, our LMEM analysis confirms pervasive genomic mosaicism within *P. patens* itself. Individuals from distant populations or different habitats

often group together; for example, individual P3_1 carries LMEMs associated with both meadow and forest environments, while individual P8_1, originating from a meadow population, is dominated by forest-genotype fragments. The lack of coherence between plastid genotype and habitat further underscores the decoupling of cytoplasmic markers from ecological adaptation.

The lack of a clear phylogeographic structure and the collapse of diagnostic resolution under dense sampling in *P. patens* can be explained by the species' evolutionary history within the Pleistocene Mammoth Steppe biome. Our demographic reconstruction (BSP) identified major population expansions during periods of global cooling, which align with the maximum availability of open steppe-like landscapes across Eurasia. This vast, continuous ecological corridor likely facilitated high levels of gene flow and panmixia among ancestral lineages, as described in the classic "Mammoth Steppe" model [60, 61]. The plastomes of contemporary *P. patens* populations therefore act as a "genomic fossil record", retaining the signature of this ancient connectivity. The pervasive shared ancestral polymorphism and evidence of chloroplast capture from *P. vernalis* suggest that what we observe today are fragmented remnants of a once-unified syngameon that spanned the Mammoth Steppe.

This deep temporal signal stands in direct contradiction to the contemporary phylogeographic structure (Baltic 1, Baltic 2, and Southern groups) revealed by nuclear SSR and ISJ markers [21, 62]. The discordance illustrates a temporal stratification of the genetic signal: while the highly polymorphic nuclear genome captures the recent history of post-glacial fragmentation into discrete refugia and habitat islands [63], the plastome retains the "ghost" of ancient panmixia. Reliance on plastid barcodes alone for *P. patens* therefore fails to distinguish species and obscures the contemporary genetic structure relevant for conservation management.

This conflict has direct implications for barcoding markers such as *ycf1*. While *ycf1* harboured the most coding-sequence changes, its use in *Pulsatilla* is problematic because it spans the IR/SSC boundary, complicating amplification [9]. Even if technical issues were resolved, the *ycf1* sequence would likely yield the same paraphyletic signal as the rest of the plastome, failing to distinguish *P. patens* from *P. vernalis* due to shared ancestral polymorphism. Our pan-plastomic study shows that increasing sampling density eliminates super-barcode resolution, as high plastome variation is rendered useless for species diagnostics by a complex evolutionary history.

Conclusions

Dense intraspecific sampling redefines what plastid super-barcoding can deliver in reticulate plant complexes. Across 33 plastomes of *Pulsatilla patens* and 49 plastomes of the genus, we recovered 2,127 polymorphisms but only two fixed nucleotides that diagnose *P. patens* s.s. once the introgressed Slovakian individual (P24_1) and the divergent Asian subspecies *multifida* are removed. *P. patens* resolves as paraphyletic with respect to *P. vernalis*, and P24_1 carries a nearly complete *P. vernalis* plastome inside a morphologically typical *P. patens*, recording a chloroplast capture event preserved in a former Carpathian–Pannonian glacial contact zone.

We interpret this pattern as a super-barcoding paradox specific to reticulate complexes: increasing both data volume and sampling density does not strengthen species identification but rather eliminates it. The plastome of *P. patens* preserves the signature of Pleistocene panmixia across the Mammoth Steppe rather than contemporary species boundaries or the post-glacial population structure recovered from nuclear SSR data. Plastid super-barcodes therefore cannot stand alone in this group, and diagnostic-marker counts reported from sparse-sampling studies need reassessment under denser intraspecific coverage.

For *Pulsatilla* and other rapidly radiating, reticulate plant lineages, the operational recommendation is concrete. Plastid hotspots such as *ycf1*, *rpoC2*, and *ndhF* remain useful for rapid molecular screening and for identifying major lineages, but reliable species delimitation and evidence-based conservation of fragmented *P. patens* populations require nuclear genomic data alongside the plastome.

Abbreviations

BIC
Bayesian Information Criterion
BS
bootstrap support
BSP
Bayesian Skyline Plot
CTAB
cetyltrimethylammonium bromide
DN
diagnostic nucleotide(s)
DNA
deoxyribonucleic acid
ESS
Effective Sample Size
HPD
Highest Posterior Density
indel
insertion/deletion
IR
Inverted Repeat
IRb
Inverted Repeat copy B
ISJ
Intron—Splice Junction (marker)
JLA / JLB / JSA / JSB
junctions between LSC—IRA, LSC—IRB, SSC—IRA, SSC—IRB

LMEM
Longest Maximum Exact Match(es)
LSC
Large Single Copy region
Ma
million years ago
MCMC
Markov Chain Monte Carlo
ML
Maximum Likelihood
Ne
effective population size
NGS
next—generation sequencing
ONT
Oxford Nanopore Technologies
PCR
polymerase chain reaction
PVP
polyvinylpyrrolidone
Q
Phred quality score
SNP
single nucleotide polymorphism
SRA
Sequence Read Archive
SSC
Small Single Copy region
SSR
simple sequence repeat

Declarations

Ethics approval and consent to participate

Field sampling of *Pulsatilla patens* (L.) Mill., a strictly protected species in Poland under the Bern Convention and the European Union Habitats Directive (Council Directive 92/43/EEC), was conducted and formally identified by Monika Szczecińska. Collection of the Polish populations was performed under the explicit permission of the General Directorate for Environmental Protection (Generalna Dyrekcja Ochrony Środowiska) (permit no. DOP-PN-4102-147/8810/10/25), relevant Regional Directorates for Environmental Protection (Regionalne Dyrekcje Ochrony Środowiska), and the directors

of the Biebrza, Białowieża, and Wigry National Parks. The remaining populations from outside Poland were collected by G.S. with the assistance of L. Bartha (Cluj-Napoca), B. Lesku (Debrecen), B.A. Lukács (Debrecen), M. Russin (Kiev), I. Sramkóné Gáspár (Debrecen), R. Šuvada (Rožnava), A. Szabó (Cluj-Napoca), and V. Virók (Aggtelek). These plants were primarily collected from protected areas; the non-destructive sampling of a 1–2 cm^2 leaf segment was done with the consent of the local nature conservation authorities in the presence of an official representative. Voucher specimens for the analyzed populations have been deposited at the Herbarium of the Department of Botany and Evolutionary Ecology, Faculty of Biology and Biotechnology, University of Warmia and Mazury in Olsztyn (OLS).

Material derived from public databases was not subject to additional permits. The collection of all plant material complied with the relevant institutional, national, and international guidelines and legislation , including the Convention on Biological Diversity and the Nagoya Protocol on Access and Benefit-sharing. While *Pulsatilla patens* is not listed under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) appendices, all activities adhered strictly to international conservation standards. No human subjects or animal experimentation was involved in this study

Consent for publication

Not applicable.

Availability of data and materials

The 32 newly assembled *Pulsatilla* plastid genome sequences generated in this study have been deposited in the NCBI GenBank database under accession numbers **(**Supplementary Table S1**)**.

Competing interests

The authors declare that they have no competing financial or non-financial interests. The funding organization (National Science Centre, Poland) is a public governmental agency for basic research and had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript

Funding

This work was funded by the National Science Centre, Poland grant no. DEC- 2017/01/X/NZ8/00880.

Authors' contribution

Conceptualization, M. S.; methodology, J.S., M.S; software, M.M, K.K., J.S.; formal analysis, M.S. J.S. A.G, K.K, M.M.; resources, M.S.; data curation, K.K.; writing original draft preparation, M.S., J.S., K.K., M.M.,A.G.; writing review and editing, MS.; visualization, K.K., J.S., M.M; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript

Author Contribution

Monika Szczecińska, Email: monika.szczecinska@uwm.edu.pl.

Mateusz Maździarz, Email: mateusz.mazdziarz@uwm.edu.pl

Kamil Koczwara, Email: kamil.koczwara@uwm.edu.pl

Aniela Garbowska, Email; aniela.garbowska@doktorant.uwm.edu.pl

Jakub Sawicki, Email; jakub.sawicki@uwm.edu.pl

References

1. Hebert PDN, Cywińska A, Ball SL, de Waard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci. 2003;270(1512):313–21.
2. Kress WJ, Plant. DNA barcoding: from basics to applications. Appl Plant Sci. 2017;5(10):1700067.
3. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci U S A. 2009;106(31):12794–7.
4. Hollingsworth PM, Graham SW, Little DP. Choosing and using a plant DNA barcode. PLoS ONE. 2011;6(5):e19254.
5. Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, et al. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. Am J Bot. 2012;99(2):320–9.
6. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Rev. 2015;90(1):157–66.
7. Huang CY, Ayliffe MA, Timmis JN. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol. 2005;138(3):1723–32.
8. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S. *ycf1*, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5:8348.
9. Szczecińska M, Sawicki J. Genomic resources of three *Pulsatilla* species reveal evolutionary hotspots, species-specific sites and variable plastid structure in the family Ranunculaceae. Int J Mol Sci. 2015;16(9):22258–79.
10. Hoot SB, Meyer KM, Manning JC. Phylogeny and reclassification of *Anemone* (Ranunculaceae), with an emphasis on austral species. Syst Bot. 2012;37(1):139–52.
11. Szczecińska M, Łazarski G, Bilska K, Sawicki J. The complete plastid genome and nuclear genome markers provide molecular evidence for the hybrid origin of *Pulsatilla* x *hackelii* Pohl. Turk J Bot. 2017;41(4):329–37.

12. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009;7:84.
13. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, et al. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J. 2011;9(3):328–33.
14. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
15. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot. 2018;105(3):291–301.
16. Council of Europe. Convention on the conservation of European wildlife and natural habitats (Bern Convention). Bern, Switzerland. 1979. Available from: https://www.coe.int/en/web/bern-convention
17. European Communities. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (Habitats Directive). Official J Eur Union. 1992;L 206:7–50.
18. Wójtowicz W. Biologia, wymagania siedliskowe i możliwości uprawy zachowawczej *Pulsatilla patens* (L.) Mill. Biul Ogr Bot Muz Zbiorów. 2000;9:45–54.
19. Zych M. Krajowy plan ochrony: sasanka otwarta *Pulsatilla patens* (L.) Mill. Warszawa: Ministerstwo Środowiska; 2007. pp. 1–33.
20. Chmura D. Zagrożenia lokalnych populacji sasanki otwartej *Pulsatilla patens* na przykładzie stanowiska na Sodowej Górze w Jaworznie. Chrońmy Przyr Ojcz. 2003;59(5):14–27.
21. Szczecińska M, Sramko G, Wołosz K, Sawicki J. Genetic diversity and population structure of the rare and endangered plant species *Pulsatilla patens* (L.) Mill in East Central Europe. PLoS ONE. 2016;11(3):e0151730.
22. Szandar K, Sawicki J, Paukszto Ł, Krawczyk K, Szczecińska M. Are the organellar genomes useful for fine scale population structure analysis of endangered plants? - A case study of *Pulsatilla patens* (L.) Mill. Genes. 2023;14:1–16.
23. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 1991;12:13–5.
24. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4):e18.
25. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17(11):1103–10.
26. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530–34.
27. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
28. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35(2):518–22.

29. Brown SDJ, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, Vink CJ, et al. Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour. 2012;12(3):562–5.
30. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31(7):1929–36.
31. Kumar S, Stecher G, Tamura K. MEGA: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2024.
32. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A. 2012;109(47):19333–8.
33. Tamura K, Tao Q, Kumar S. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. Mol Biol Evol. 2018;35(7):1770–82.
34. Tao Q, Montoya JB, Kumar S. Data-driven speciation tree prior for better species divergence times in calibration-poor molecular phylogenies. Bioinformatics. 2021;37:102–10.
35. Li Qj, Su N, Zhang L, et al. Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of *Pulsatilla* (Ranunculaceae). Sci Rep. 2020;10:19781.
36. Xue Y et al. Comparative analysis of chloroplast genomes of *Pulsatilla* species reveals evolutionary and taxonomic status of newly discovered endangered species *Pulsatilla saxatilis*. BMC Plant Biol. 2024.
37. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650.
38. Bilderbeek RJC, Etienne RS, babette. BEAUti 2, BEAST2 and Tracer for R. *bioRxiv* [Preprint]. 2018. Available from: https://doi.org/10.1101/271866
39. Khan J, Kokot M, Deorowicz S, Kingsford C. Cuttlefish: fast, parallel, and low-memory graph construction from sequencing reads. Bioinformatics. 2021;38(1):177–86.
40. Bello R, Zomer AL, Deorowicz S, Dutilh BE. Graphite: painting genomes using a colored de Bruijn graph. NAR Genom Bioinform. 2024;6(4).
41. Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ. Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae). Mol Ecol Resour. 2012;12(1):98–108.
42. Ślipiko M, Myszczyński K, Buczkowska K, Bączkiewicz A, Sawicki J. Super-mitobarcoding in plant species identification? It can work! The case of leafy liverworts belonging to the genus *Calypogeia*. Int J Mol Sci. 2022;23(24):15570.
43. Percy DM, Argus GW, Cronk QC, Fazekas AJ, Kesanakurti PR, Burgess KS, et al. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): does this result from a trans-specific selective sweep? Mol Ecol. 2014;23(19):4737–56.
44. Dumoulin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ. Phylogeographic structure of white oaks throughout the European continent. Genetics. 1997;146(4):1475–87.

45. Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, et al. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen. Ecol Manage. 2002;156(1–3):49–74.
46. Besnard G, Hernandez P, Khadari B, Dorado G, Savolainen V. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. BMC Plant Biol. 2011;11:80.
47. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? Mol Ecol Resour. 2015;15(5):1067–78.
48. Magwé-Tindo J, Wieringa JJ, Sonké B, Zapfack L, Vigouroux Y, et al. Complete plastome sequences of 14 African yam species (*Dioscorea* spp). Mitochondrial DNA B Resour. 2019;4(1):74–6.
49. Bouillé M, Senneville S, Bousquet J. Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. Tree Genet Genomes. 2011;7:469–84.
50. Villarreal AJC, Crandall-Stotler BJ, Hart ML, Long DG, Forrest LL. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytol. 2016;209(4):1734–46.
51. Dong S, Zhang S, Zhang L, Wu H, Goffinet B, Liu Y. Plastid genomes and phylogenomics of liverworts (Marchantiophyta): conserved genome structure but highest relative plastid substitution rate in land plants. Mol Phylogenet Evol. 2021;161:1055–7903.
52. Ślipiko M, Myszczyński K, Buczkowska K, et al. Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. BMC Plant Biol. 2020;20:243.
53. Ronikier M, Costa A, Aguilar JF, Feliner GN, Küpfer P, Mirek Z. Phylogeography of *Pulsatilla vernalis* (L.) Mill. (Ranunculaceae): chloroplast DNA reveals two evolutionary lineages across central Europe and Scandinavia. J Biogeogr. 2008;35(9):1650–64.
54. Li Q, Wang X, Wang J, Su N, Zhang L, Ma Y, et al. Efficient identification of *Pulsatilla* (Ranunculaceae) using DNA barcodes and micro-morphological characters. Front Plant Sci. 2019;10:1196.
55. Schmitt T, Varga Z. Extra-Mediterranean refugia: the rule and not the exception? Front Zool. 2012;9(1):22.
56. Zimmermann W. Zur Taxonomie von *Pulsatilla* Miller. Die Kulturpflanze. 1964;12:93–133.
57. Hegi G, Weber HE. Illustrierte Flora von Mitteleuropa. Verlag Paul Parey; 1974.
58. Akeroyd JR. *Pulsatilla* Miller. In: Tutin TG, editor. Flora Europaea. Volume 1, 2nd ed. Cambridge: Cambridge University Press; 1993. pp. 264–6.
59. Lv HZ, So SK, Kim MY. A new hybrid species of *Pulsatilla* (Ranunculaceae): *P. x yanbianensis* H.Z. Lv. Korean J Pl Taxon. 2011;41(4):365–9.
60. Guthrie RD. Frozen Fauna of the Mammoth Steppe: The Story of Blue Babe. Chicago: University of Chicago Press; 1990.

61. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci. 2004;359(1442):183–95.
62. Bilska K, Szczecińska M. Comparison of the effectiveness of ISJ and SSR markers and detection of outlier loci in conservation genetics of *Pulsatilla patens* populations. PeerJ. 2016;4:e2504.
63. Schmitt T. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Front Zool. 2007;4:11.
64. Supplementary.
65. Supplementary Figure S. Number of diagnostic nucleotides recovered for each focal taxon under different sampling and inclusion regimes.
66. Supplementary Table S. List of plastomes analysed in this study, including newly sequenced and publicly available accessions.

Figures

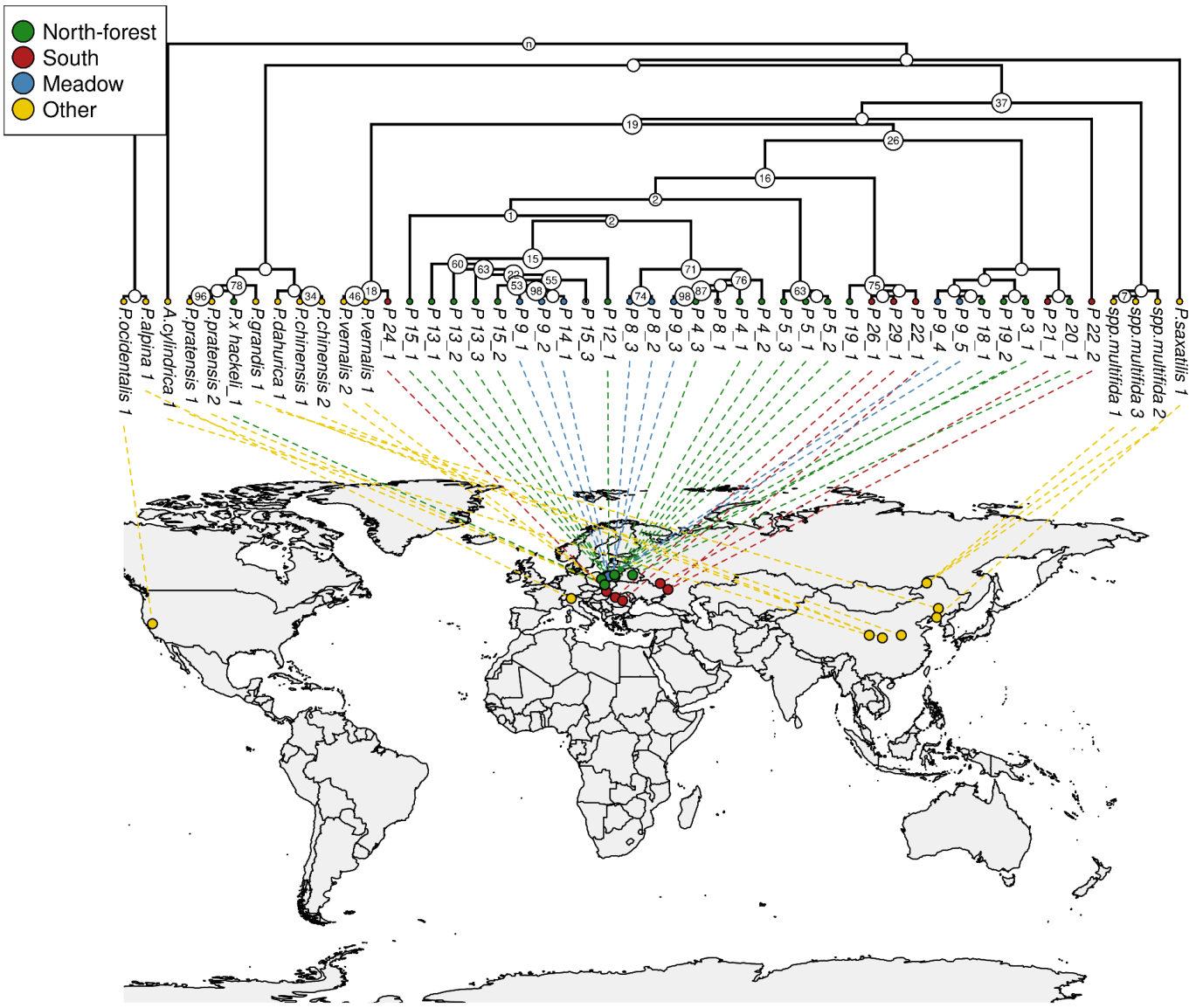

Figure 1

Maximum Likelihood (ML) phylogenetic tree of *Pulsatilla* species mapped onto their geographic distribution. The tree is based on complete chloroplast genome alignments. Dot colors denote habitat type or lineage: green – forest; blue – meadow; red – southern populations; yellow – other (*P. patens* subsp. *multifida*). Bootstrap support values below 100% are indicated at the nodes.

Pulsatilla patens ssp. *patens* in natural habitat

Figure 2

Demographic history reconstruction of *Pulsatilla patens* based on whole-plastome data. The Bayesian Skyline Plot (BSP) depicts changes in effective population size (Ne) over time. The solid black line represents the median estimate of Ne, while the blue shaded region indicates the 95% Highest Posterior Density (HPD) intervals. The x-axis represents time in millions of years before present (Ma), spanning the Miocene, Pliocene, and Pleistocene epochs. To the right, the time-calibrated phylogenetic tree aligns branching events with the inferred demographic fluctuations, highlighting the correlation between lineage divergence and population expansions during the Pleistocene.

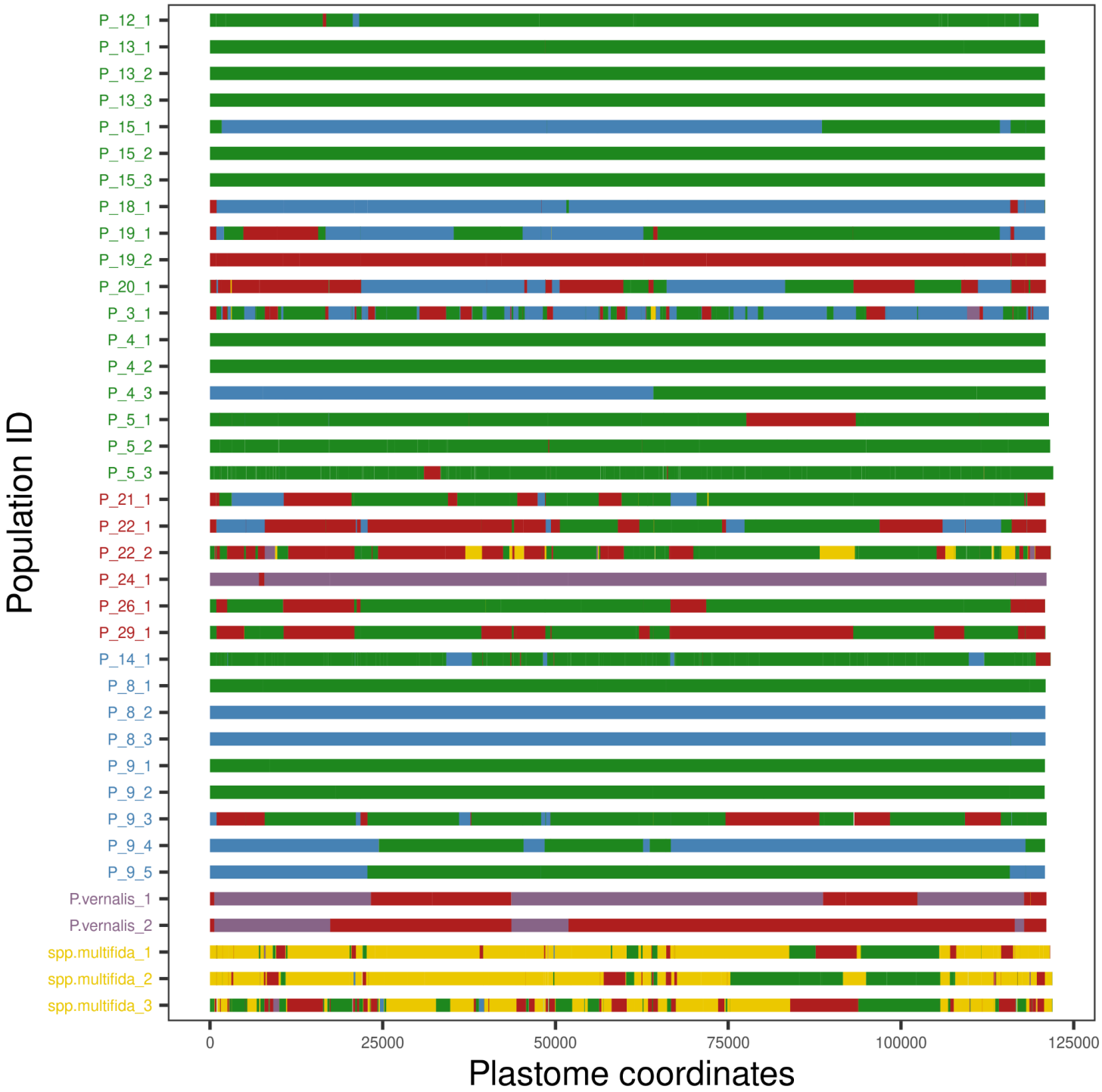

Figure 3

Visualization of Longest Maximum Exact Matches (LMEMs) across *Pulsatilla* plastomes. Horizontal bars represent genomic coordinates of analyzed individuals. Colors denote the habitat affinity of identified LMEM regions: green – Forest; blue – Meadow; red – Southern; yellow – *P. patens* subsp. *multifida*. The distinct color blocks within single individuals (e.g., P_3_1, P_8_1) illustrate the mosaic nature of the plastomes, while individuals like P_24_1 show complete capture of a divergent genotype.

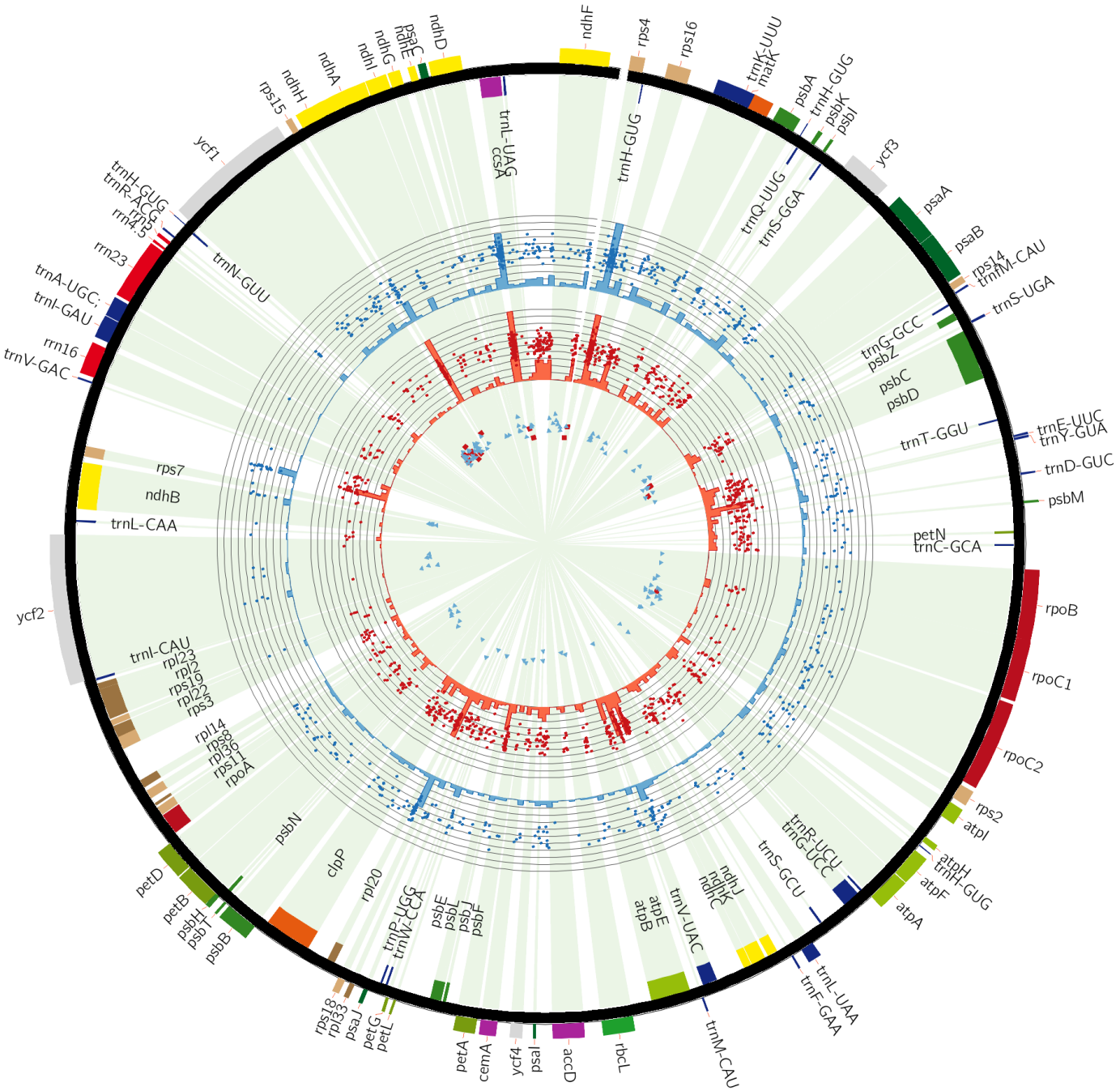

Figure 4

Genomic variation across *Pulsatilla*plastomes. Circular distribution of genomic variation in *P. patens* plastomes. Tracks from outer to inner: (1) gene annotations (green); (2) SNP density (500-bp bins) and individual positions (blue dots); (3) indel density and individual positions (red dots); (4) non-synonymous mutations in coding sequences (blue triangles: SNPs; red squares: indels).

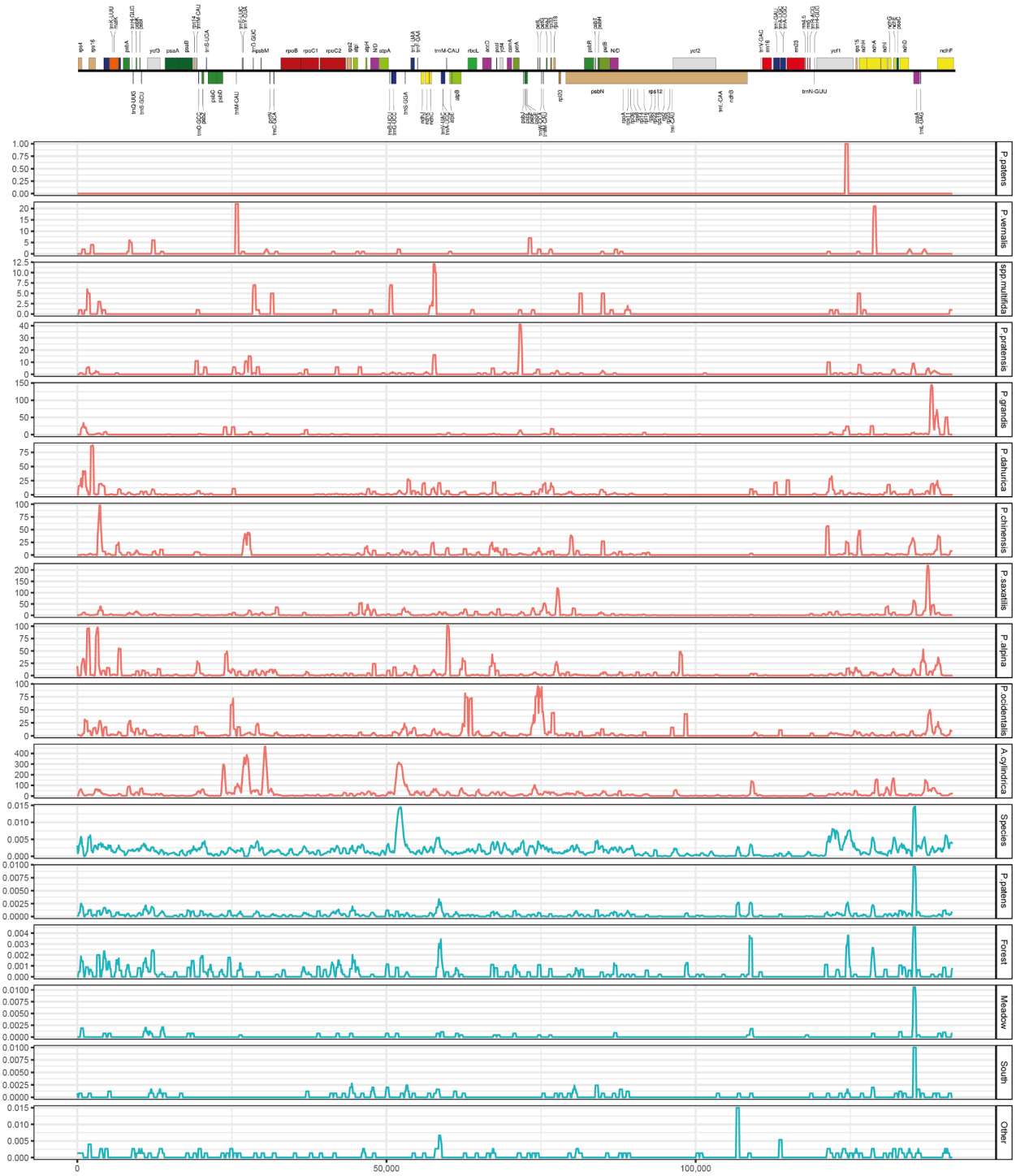

Figure 5

Nucleotide diversity (π; blue) across *Pulsatilla*plastomes (500-bp windows). Red ticks mark positions of diagnostic nucleotides for each dataset. The *P. vernalis* panel includes specimen P24_1 (cp-assigned to *P. patens*).

Supplementary Files

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

- Table1..xlsx
- Figures1Supplementary.png