



Organelar genome evolution in *Ribes* L.: hotspots of mutation, sequence transfer, codon usage, and phylogenetic context

Karolina Kłubowicz¹ · Jakub Sawicki¹ · Łukasz Pauksto¹ · Kamil Ciborowski¹ · Mateusz Maździarz¹ · Katarzyna Krawczyk¹

Received: 10 June 2024 / Revised: 4 October 2024 / Accepted: 24 November 2024 / Published online: 4 December 2024
© The Author(s) 2024

Abstract

Ribes genus from the Grossulariaceae family, have been used in the food industry and for medicinal purposes for years. Despite the valuable medicinal values and the widespread occurrence of *Ribes*, many species remain understudied. One of these is the alpine currant *Ribes alpinum* L., which occurs natively in Europe. The study aimed to thoroughly understand and characterize the organelle genomes of *R. alpinum*, conduct their comparative analysis with available data for other representatives of Grossulariaceae and, in the case of the chloroplast genome, also with *Saxifraga insolens* Irmsch., a representative of the closely related Saxifragaceae family. The analysis of plastome interspecific variation revealed several hotspots of variation which could be useful as potential phylogenetic markers or specific DNA barcodes. The phenomenon of sequence transfer between the chloroplast and mitochondrial genomes and the relative synonymous codon usage between the plastome and mitogenome were also analyzed. A phylogenetic inference based on the chloroplast genome sequence questioned earlier reports of a close relationship of *R. alpinum* to *R. rubrum* and confirmed its affiliation with an early-diverged phylogenetic line together with *R. fasciculatum*. Analysis of hotspots of variation in chloroplast genome allowed us to indicate petN-psbM and trnS-trnG as high-utility barcode sequences, competitive with standard barcodes and the ycf1 gene.

Keywords *Ribes alpinum* · Mitogenome · Plastome · Gene transfer · Codon usage · Phylogeny

Introduction

Currants (*Ribes* L., Grossulariaceae) have long been used in the food industry and traditional medicine for their valuable properties. Currant fruits are characterized by high vitamin content compared to other plants. Black currant (*Ribes nigrum*), the most widely used and best described in the industry, can contain up to 250 mg of vitamin C per 100 g of fresh fruit, fully meeting an adult's daily requirement for this vitamin. In addition, it has been shown that seeds of many species can be used to extract unsaturated fatty acids, with black currant having a rich fatty acid profile,

including linolenic acid and alpha-linolenic acid, which may positively affect immunity and have anti-inflammatory properties. Moreover, these plants contain high levels of compounds beneficial to human health, such as anthocyanins, flavonoids, and phenolic acids, which act as antioxidants and protect cells from oxidative stress.

The family Grossulariaceae includes more than 160 currant species found worldwide. Despite the valuable medicinal properties and widespread occurrence of *Ribes*, many species remain understudied. One such species is the alpine currant (*Ribes alpinum*), which occurs natively in Europe, mainly in its northern and central regions in mountainous areas. While this shrub is primarily ornamental, it also holds potential for the food sector. Its use in traditional medicine, particularly for making infusions, has also been reported.

So far, various infrageneric classifications of *Ribes* have been proposed in the literature which varied - from four to even eight. The latest classifications recognize two subgenera and up to six sections: *Ribesia*, *Ribes griffithii*, *Coreosma*, *Grossularia*, *Berisia*, and *Hemibotrya*. Alpine

Communicated by G. G. Vendramin

✉ Katarzyna Krawczyk
katarzyna.krawczyk@uwm.edu.pl

¹ Department of Botany and Evolutionary Ecology, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland

currant is placed in the *Berisia* section, along with *Ribes glaciale* (Zhang et al. 2023a). This division was partially confirmed by the results of studies on the phylogeny of currants based on the variation of the nuclear ITS region. This study, covering 66 species of the genus *Ribes*, is the most extensive study of the phylogeny of this genus to date. However, a recent attempt to reconstruct the phylogenetic relationships within the *Ribes* genus, based on complete chloroplast genome variation, questioned the validity of this classification, showing a very fragmentary, because it concerns only eight species, but nevertheless significantly different picture of the phylogeny of currants (Sun et al. 2023). Considering the phylogenetic informativeness of the complete chloroplast genome compared to the information contained in the ITS region variability, and the observed incongruence between the phylogeny pictures obtained from these two data sets, allows us to hypothesize that the phylogeny of *Ribes* based on ITS requires verification. *Ribes alpinum* was not included in the recent phylogenomic studies and to date, neither the complete chloroplast genome nor the mitogenome of this species has been described.

Understanding the structure of organellar genomes in plants is crucial for uncovering evolutionary breakthroughs and studying the functions of specific plant genes. Advances in sequencing technology are making it easier to explore the structure of new plastomes and mitochondrial genomes of plants. The frequent use of chloroplast genomes in phylogenetic inference is dictated by the fact that they are haploid structures, usually inherited uniparentally (Palmer 1985; Li et al. 2015; Chung et al. 2023). In addition, they are characterized by their relatively small size, as in land plants the chloroplast genome reaches a size of 120 to 160 kb (Zhang et al. 2023) and contains about 100 genes, but this number can vary between species (Wang et al. 2024a). The lack of recombination between molecules and the slow rate of mutations are other features of the chloroplast genome predestinate it to be used not only in phylogenetics, but also in the study on biodiversity, the development of markers for molecular breeding and DNA barcoding (Palmer 1985; Dobrogojski et al. 2020; Tan et al. 2020).

The size of the mitogenome varies considerably between species, covering a range from 66 kb in *Viscum scurruloideum* (Caryophyllaceae) to as much as 11,300 kb in *Silene conica* (Caryophyllaceae) (Sloan et al. 2012). They contain up to dozens protein-coding genes - the number usually ranges from several or so in *Viscum minimum* to more than 50 in *Marchantia polymorpha*; (Petersen et al. 2015; Wang et al. 2024a). Mitochondria are responsible for plant development and reproduction. They control, among other processes, the plant's response to stress, apoptosis or performing defense functions against microorganisms (Zervas et al. 2019). They undergo independent replication and

encode many essential peptides. In most plant species, mitochondrial genome sequences undergo slow evolutionary change and are relatively rarely affected by mutations, making them good candidates for molecular studies (Gualberto et al. 2014; Wynn and Christensen 2019). Genome research (including nucleotide variation) can provide a lot of information about the evolution of mitochondria, encoded traits or interactions between the nucleus and cytoplasm, making the genome of these organelles exceptionally interesting (Kitazaki and Kubo 2010). The phenomenon of gene transfer between organelles is also being reported (Wang et al. 2007), which needs to be better understood for the potential use of mitochondrial sequences in molecular evolution studies. The growing number of sequenced and described mitogenomes indicates the transfer of genes from the chloroplast to the mitochondrion, and from the mitochondrion to the nuclear genome (Cusimano and Wicke 2016; Filip and Skuza 2021). The most common is gene transfer from the chloroplast to the mitochondrion, where cpDNA sequences can account for 1 to 12% of the entire mitochondrial genome (Mower et al. 2012).

Given the insufficient knowledge of the organellar genomes of *Ribes alpinum*, conducting a comparative analysis of these genomes with other known genomes from the *Ribes* genus is essential. Such research would not only fill the gaps in knowledge by clarify the phylogenetic relationships of *Ribes alpinum* within the *Ribes* genus but also provide insights into the phenomenon of plastid-to-mitochondrial gene transfer. Both chloroplast and mitochondrial genome in detail described in the study will be a valuable resource as a genetic basis for further studies, including phylogenomics, DNA barcoding, population genomics, adaptive evolution and many others.

Materials and methods

Plant material and DNA extraction

The fresh leaves of alpine currant (*Ribes alpinum* L.) were collected on the territory of north-eastern Poland (53°45'32.8"N 20°27'26.1"E). The material was frozen at -20 °C until DNA extraction. 0.2 g of sample was grated in a homogenizer VWR Mini Bead Mill in order to grind the material. Total DNA was isolated using DNeasy Plant Mini Kit (250) (Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA concentration was measured using fluorometric dyes with the Qubit dsDNA HS Assay Kit (Invitrogen) following the manufacturer's instructions on Qubit 4. DNA concentration in the sample was 36,2 ng/μl. The purity of the DNA extracts was measured on Cary 60 UV-Vis (Agilent) at a wavelength 230, 260 and 280 nm,

where the 260/280 nm ratio was 1,675 and the 260/230 ratio at 1,99.

The DNA library preparation and sequencing

To prepare the DNA library about 500 ng of gDNA of *Ribes alpinum* was used. The library was constructed using Ligation sequencing kit V14 (LSK-SQK114) and NEBNext® Companion Module for ONT Ligation Sequencing (New England Biolabs) following the manufacturer's protocol. Sequencing was performed on the MinION Mk1C portable device using Flow Cell R10.4.1 (ONT). To ensure the best quality of the sequences, the raw reads were basecalled in Dorado v 0.4.3 using the SUP model.

Assembly and annotation

Reads were assembled into contigs using the Flye v1.9.1. The contigs were mapped to the *Ribes nigrum* reference genome (chloroplast genome: NC_080517.1, mitochondrial genome: NC_082306.1) in Geneious Prime 2023.2. software (Biomatters, Auckland, NZ). Sequences that resulted from mapping were then verified manually. Genes were annotated onto the complete genomes using the annotation transfer option in Geneious Prime. Analysis of the presence of structural variants of plastomes within analyzed species was carried out using the Cp-Hap pipeline (Wang et al. 2019).

Visualization and comparative genomic analysis

A physical map of *Ribes alpinum* genomes was obtained using OGDRAW (Greiner et al. 2019). The precise boundaries of individual chloroplast genome elements (large copy region, small copy region, inverted repeats) were obtained using IR Scope (Amiryousefi et al. 2018). The reference genome of *Ribes nigrum* (GenBank: OR296460.1) was visualized on a genome map (Figure S1).

To verify the structural homology of the *R. alpinum* mitogenome with the previously described mitogenome of its close relative *R. nigrum*, the sequences of these two species were aligned by MashMap3 (Kille et al. 2023) tool implemented in SYNY software (Julian and Pombert 2024). The synteny of these two genomes was visualized using circos (Krzywinski et al. 2009), being also part of SYNY. The potential gene transfers between plastome and mitogenome of *R. alpinum* have been revealed by Blastn (Camacho et al. 2009) tools with cut-offs: e-value < 10⁻⁵ and minimum alignment score > 100. The organelle sequence transfers were highlighted on circos plot.

The ratio of the Relative Synonymous Codon Usage (RSCU) was calculated by seqinr 4.2 R package (Charif

and Lobry 2007). The RSCU algorithm determines the observed/expected frequency of each synonymous codon encoding a particular amino acid. The expected RSCU metric assumes a uniform distribution across all codons for each amino acid. The RSCU values for each species were computed independently. The distribution and correlation of RSCU values within and between organelle genomes were calculated and visualized by smplot2 (Min and Zhou 2021) and ggplot2 (Wickham 2011) packages in R.

Plastome diversity and diagnostic nucleotides

Diagnostic nucleotides and nucleotide variation (π diversity) were analyzed to determine variation in the genus *Ribes*. For further analysis alignment of 11 sequences of chloroplast genomes were used (*R. alpinum*, *R. rubrum*, *R. rubrum* var. *alba*, *R. glaciale*, *R. mandshuricum*, *R. nevadense*, *R. nigrum*, *R. uva-crispa*, *R. roezlii*, *R. odoratum*, *R. fasciculatum* var. *chinense*), which was 133,725 bp long. The spider R package was applied to calculate diagnostic nucleotides, which are positions in the sequence alignment that are unique to a given species and allow it to be distinguished from other species studied. Both diagnostics nucleotides and nucleotide diversity (π) were calculated in a 500 bp sliding window, as well as for individual coding and non-coding regions. The computations of both divergence parameters were focused on the LSC, SSC and one IR region. The *Ribes* alignment was also used to identify the SNPs (Single Nucleotide Polymorphism) and indels variant using Variant Calling subprogram in Geneious Prime. All divergence variables were visualized using circos (Krzywinski et al. 2009).

Phylogenetic analysis

To determine the phylogenetic position of *Ribes alpinum*, all available chloroplast genomes of currant species were used for analysis (*R. rubrum*, *R. rubrum* var. *alba*, *R. glaciale*, *R. mandshuricum*, *R. nevadense*, *R. nigrum*, *R. uva-crispa*, *R. roezlii*, *R. odoratum*, *R. fasciculatum* var. *chinense*) belonging to the monotypic *Grossulariaceae* family. *Saxifraga insolens* Irmsch., belonging to the Saxifragaceae family, most closely related to *Grossulariaceae*, was used as an outgroup. An alignment was obtained with the Geneious Prime 2023.2.1 software with a length of 175,815 bp. For the actual analysis, one of the genome's repeat fragments (IRa) was trimmed, which shortened the alignment length to 148,990 bp. The tree was calculated using the ML method with the IQtree program. Bootstrap values above 50 were considered significant. The final visualization was created in R Studio (4.2.3.) with ggtree and ggplot2 R environment packages (Yu et al. 2017).

Results

The characteristics of *Ribes alpinum* plastome

For analysis, chloroplast sequences were downloaded from the GenBank database, including ten representatives of the genus *Ribes* (all available as of 26.01.2024) and *Saxifraga insolens* as an outgroup (Table S1). The assembled plastome of *Ribes alpinum* is 157,351 bp long and its structure was typical for other known chloroplast genomes (Figs. 1 and

2). The number of genes was 131, of which 86 are protein-coding genes, 37 tRNAs and eight rRNAs. The GC content of each sequence element was: 36.2% GC in the long single-copy fragment (LSC), 33.3% in the short single-copy fragment (SSC), and 43.1% in the repeat fragments (IRa and IRb). IRa had seven protein-coding genes, four rRNAs and seven tRNAs, while IRb had eight protein-coding genes, four rRNAs and seven tRNAs. Both fragments shared the *ycf1* gene with SSC. IRb shared with LSC the *rps19* gene, which was not present in IRa. The order of genes on the

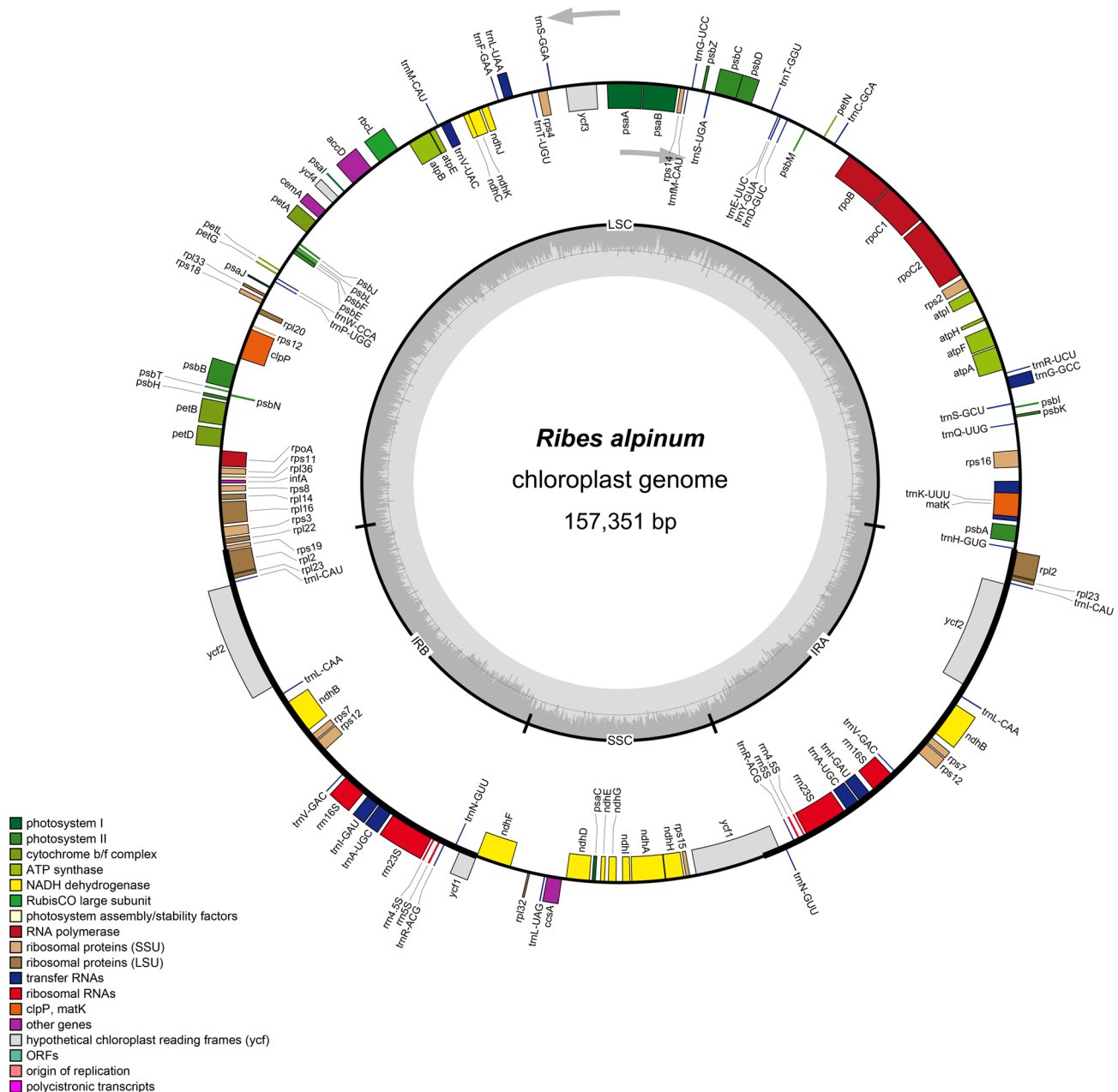


Fig. 1 Map of *Ribes alpinum* chloroplast genome. Genes outside and inside the circle are transcribed clockwise and counterclockwise, respectively. The legend at the bottom left indicates functions and

names of genes belonging to different functional groups. Gray bars in the smaller circle show the GC content (%)

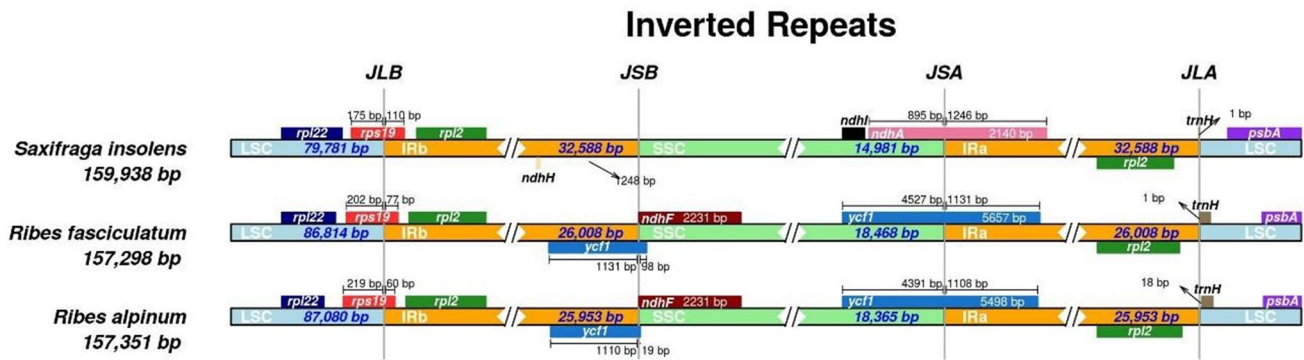


Fig. 2 IRplot presenting the comparison of LSC, SSC and IR regions boundaries in the two *Ribes* chloroplast genomes and one *Saxifragales* genome. JLB, JSB, JSA and JLA mark the boundaries between LSC and IRb, IRb and SSC, SSC and IRa, IRa and LSC, respectively

genome was the same as in the other plastomes described in the genus *Ribes*.

The exact boundaries of the chloroplast genome elements of *Ribes alpinum*, the currant representative *R. fasciculatum* var. *chinense* (GenBank: MH191388.1) and *Saxifragales* representative – *Saxifraga insolens* (GenBank: ON720893.1) are shown in Fig. 2.

Ribes alpinum and *R. fasciculatum* var. *chinense* had identical genes lying on the boundaries of JLB, JSB and JSA - they differed only slightly in the number of base pairs. *Saxifraga insolens* was visibly different from the visualized *Ribes*. In both currant species, the IRb and SSC repeat region shared the *ycf1* gene (*R. alpinum*: 1110 bp belonged to IRb, while 19 bp belonged to SSC), while in *S. insolens* no gene was present at the JSB boundary. The short single-copy region of SSC and IRa in *Ribes* shared the *ycf1* gene, while in *S. insolens* it was the *ndhA* gene.

Hotspots of variation

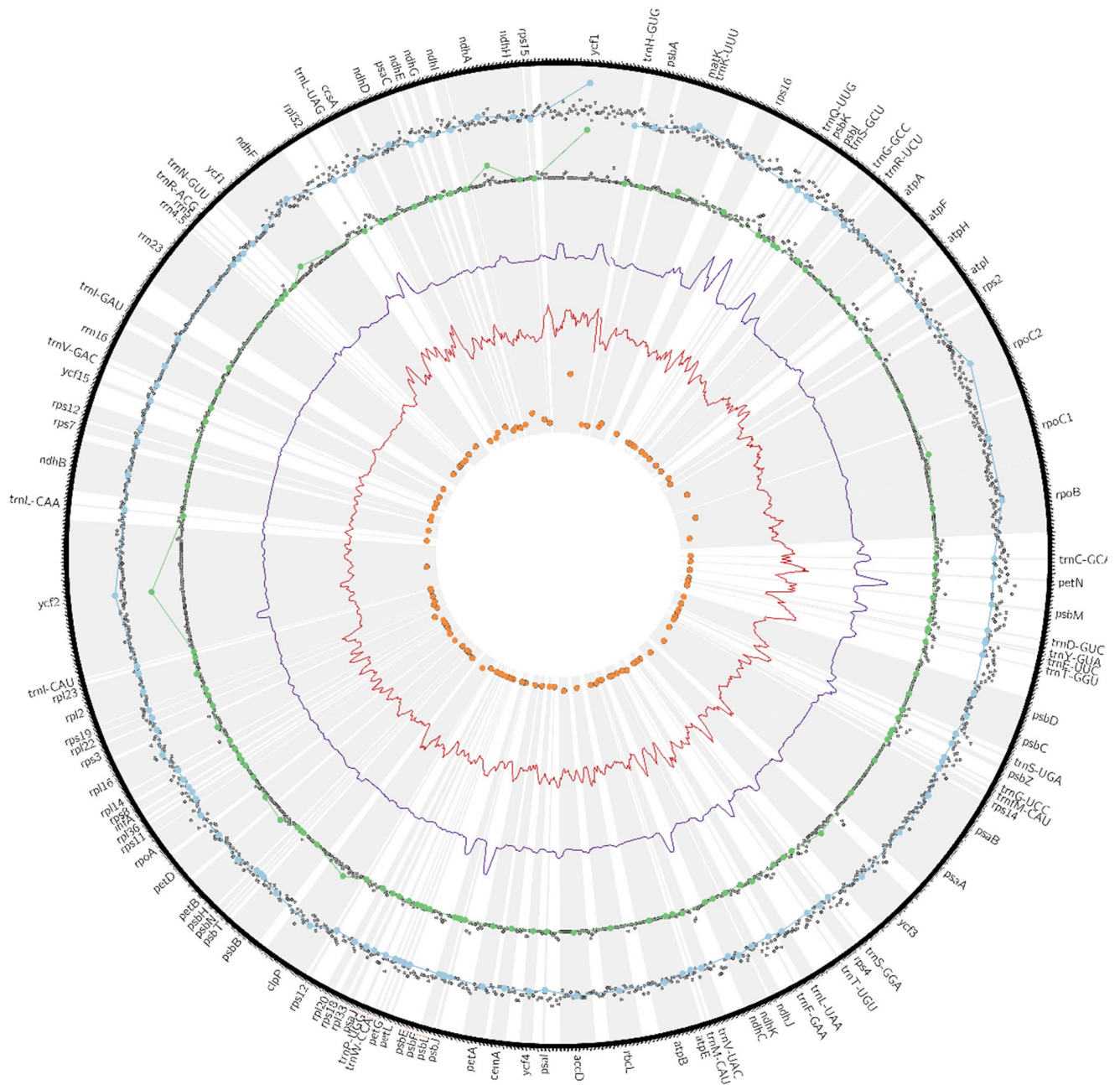
The inverted repeat region (IR; the fragment from the *rps19* gene to *ycf1*) showed the lowest genome-wide changes for all analyses performed except for the *ycf2* gene, where one peak was observed. The highest number of SNPs was observed inside two genes: *ycf1* (228) and *rpoC2* (112) (Fig. 3, Table S3). When analyzing the number of indels (Fig. 3, Table S3), the genes *ycf1* (304), *ndhA* (95) and *ycf2* (159), belonging to the IR region, showed the highest analysis values. The results of the nucleotide variation analysis (Fig. 3) showed the highest peaks mainly in the intergenic regions, such as *ndhF-rpl32*, *petN-psbM*, *trnT-GGU-psbD*, *ndhD-ccsA* or *ycf3-trnS-GGA* and on one gene (*ycf1*). The highest π value was obtained in the *ycf1* gene region (0,01942). For *R. alpinum* (Fig. 3, Table S4, S5), several interesting spots were detected in diagnostic nucleotide analysis on the chloroplast genome. Particularly high peaks were noted in the intergenic regions of the *rps16-trnQ-UUG* (two peaks – 35 and 41, respectively), *trnS-GCU-trnG-UCU* (27), *petN-psbM*

(30), *petA-psbJ* (28) and *rpl32-trnL-UAG* (25). Two characteristic peaks appeared on the *ycf1* gene, amounting to 14 and 16, respectively. The last analysis showed the mix of indels and SNPs (Fig. 3, Table S3), which are positions in the alignment where the observed mutation is of substitution or InDel type. This type of variability again was the highest in the *ycf1* gene (32). A total of 4596 diagnostic positions in alignment were observed, of which 1420 allowed the identification of *R. alpinum* from other *Ribes* representatives.

To analyze usefulness of particular chloroplast regions in discrimination of currant species, the number of diagnostic positions in the alignment was verified. Three standard barcodes: *matK*, *rbcL* and *trnH-psbA* were compared with nine regions identified as hotspots of variation in *Ribes*. Considering the number of diagnostic characters, the number of species for which it was impossible to identify based on a given sequence, as well as the length of the analyzed fragments, the standard *matK* and *rbcL* barcodes turned out to be of little use in identifying currant species (Table S6). The advantage of *trnH-psbA* is its short length, but its inter-species variability is relatively low. Considering the sum of diagnostic characters calculated for all the species, the *ycf1* gene had the greatest power of species discrimination (524 diagnostic characters), but it is a very long gene with more than 5.7 kb. The next regions with high species identification efficiency were *petN-psbM* and *trnS-trnG*, having 354 and 333 diagnostic characters, respectively.

Codon usage (RSCU)

RSCU analysis (The Relative Synonymous Codon Usage) allows comparison of the frequency of codons used to encode individual amino acids between the chloroplast and mitochondrial genomes (Fig. 4, Table S7). The highest frequency was shown by isoleucine encoded by codon att (1072) in both the mitochondrial and chloroplast. 31 codons reached an RSCU value above 1, indicating a higher preference in usage, and as many reached an RSCU value less



of InDels per 100 bp, green dots – number of InDels in genic and intergenic regions). The next layers show the amount of diagnostic nucleotides (color purple), nucleotide diversity (color red) and mix of indels and SNPs (orange dots), respectively

Characteristics of *Ribes alpinum* mitochondrial genome

 Springer

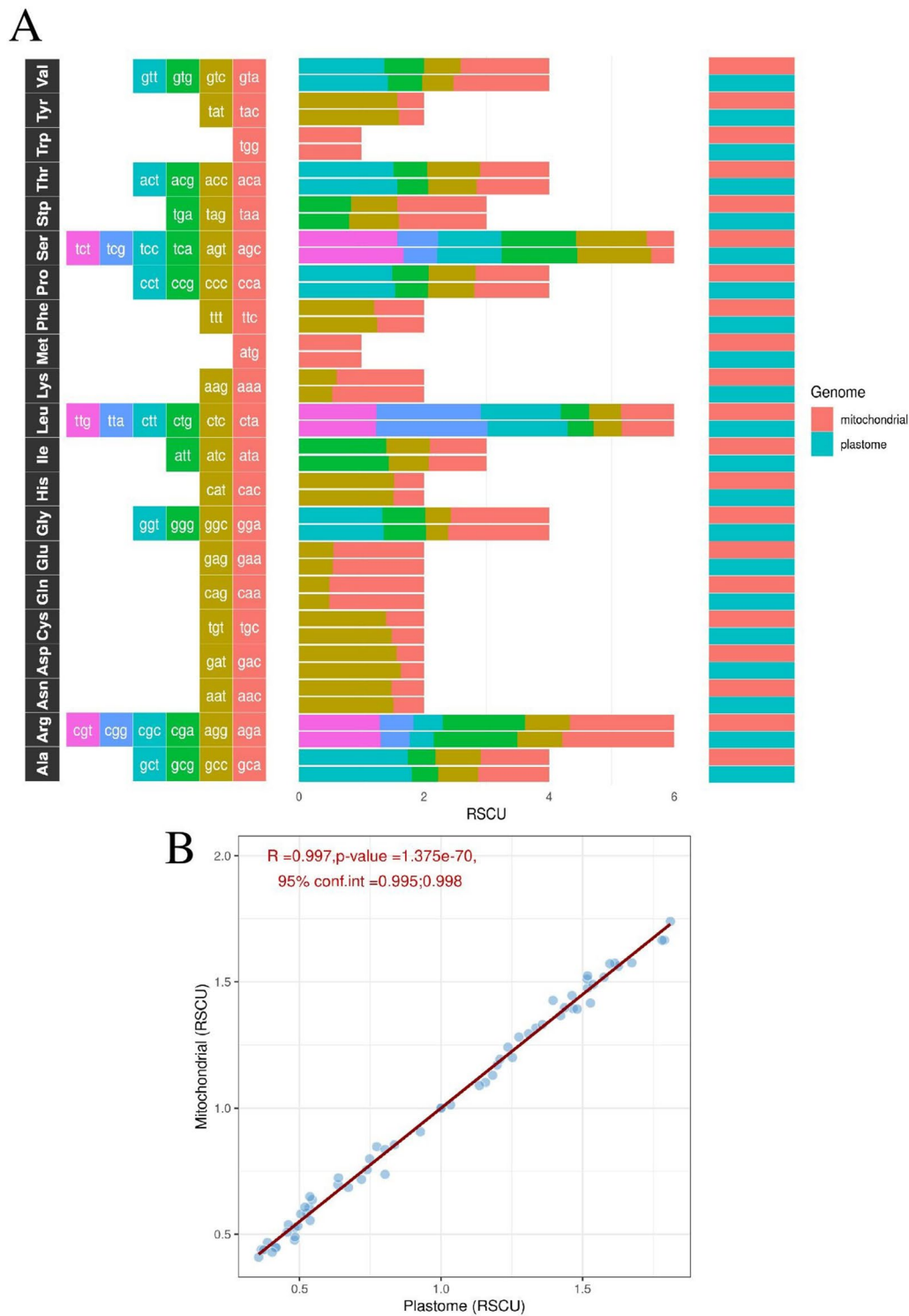


Fig. 4 **A.** The Relative Synonymous Codon Usage (RSCU) plot. For every amino acid a nucleotide triplet found in analysis is implemented. In the middle is presented a codon usage for the plastome or mitochondrial genome. **B.** The correlation between plastome (axis x) and

mitochondrial genome (axis y). The red line represents the Pearson correlation coefficient. The upper left corner shows the values of R, p-value, and confidence interval (conf.int)

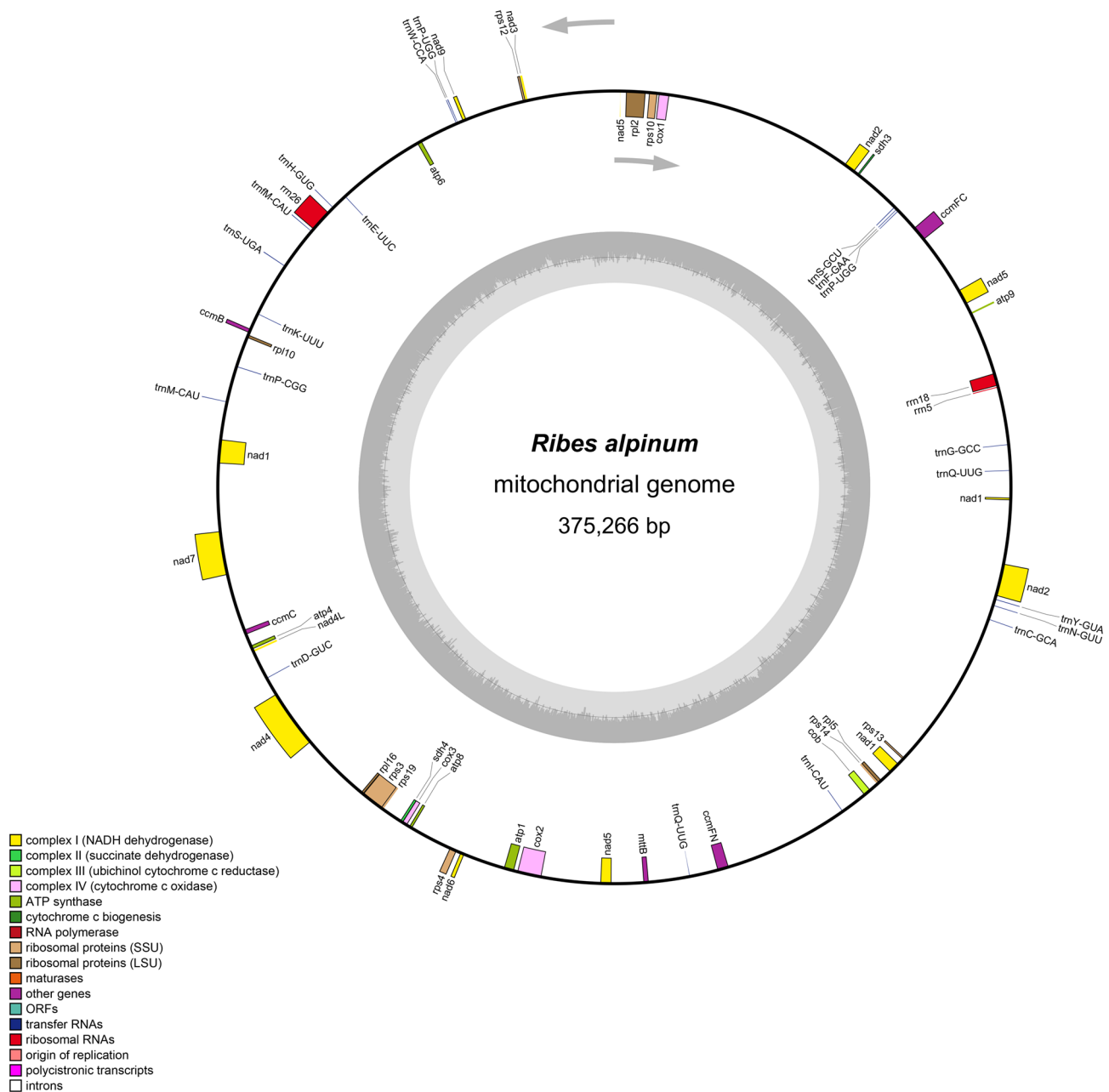


Fig. 5 Gene map of the mitochondrion of *Ribes alpinum*. Genes inside and outside the outer circle are transcribed in counterclockwise and clockwise directions, respectively. The genes are color-coded based on their function. The inner circle visualizes the G/C content

were protein-coding genes, 20 genes were tRNAs and three genes were rRNAs. Their order and location are illustrated in Fig. 5 and precisely described in Table S2.

The analysis of structural homology between the mitogenomes of two currant species revealed that *R. alpinum* mitogenome structure is rearranged in comparison with *R. nigrum*. The location of the homologous regions of the compared mitogenomes is illustrated in Fig. 6. The differences between the analyzed mitogenomes concern not only the order of gene arrangement, but also their orientation. In

comparison to *R. nigrum* mitogenome, some genes are transcribed in the opposite direction - these are protein-coding genes *atp9*, *nad3*, *rps12*, *rrn26*, *ccmB*, *rpl10*, *nad1*, *rps7*, *matR*, *nad7*, *rps4*, *nad6*, *ccmFN*, *cob*, *rps14*, *rpl5*, *rps13*, *trnQ-UGG*, *trnG-GCC*, *trnE-UUC*, *trnH-GUG*, *trnM-CAU*, *trnS-UGA*, *trnK-UUU*, *trnP-CGG*, *trnM-CAU*, *trnD-GUC*, *trnL-CAU* and *rrn26*. The *rps1* gene is absent from the composition of the *R. alpinum* genome.

The evaluation of homologous sequences in mitochondrial and chloroplast genomes of *R. alpinum* indicated 11

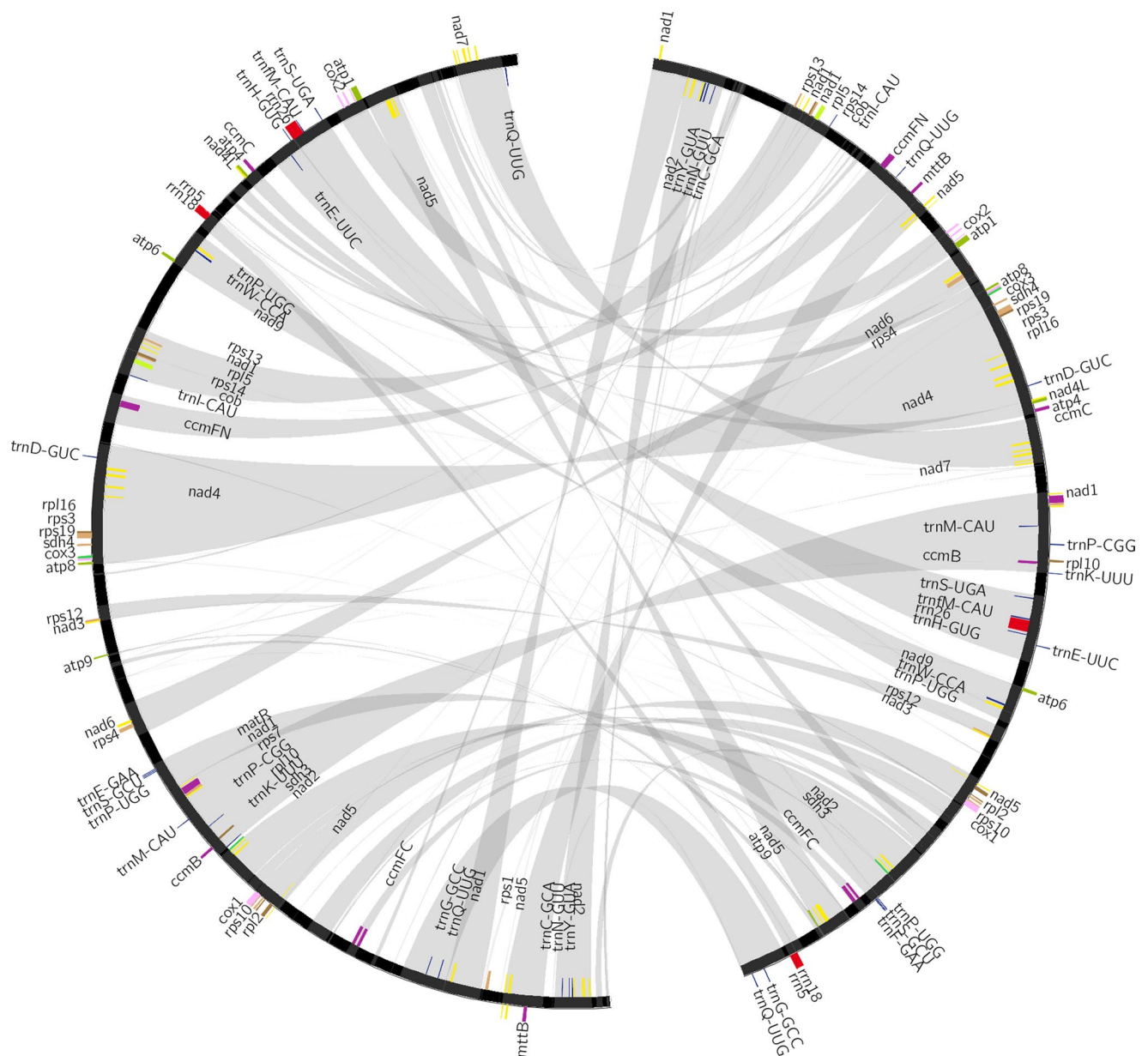


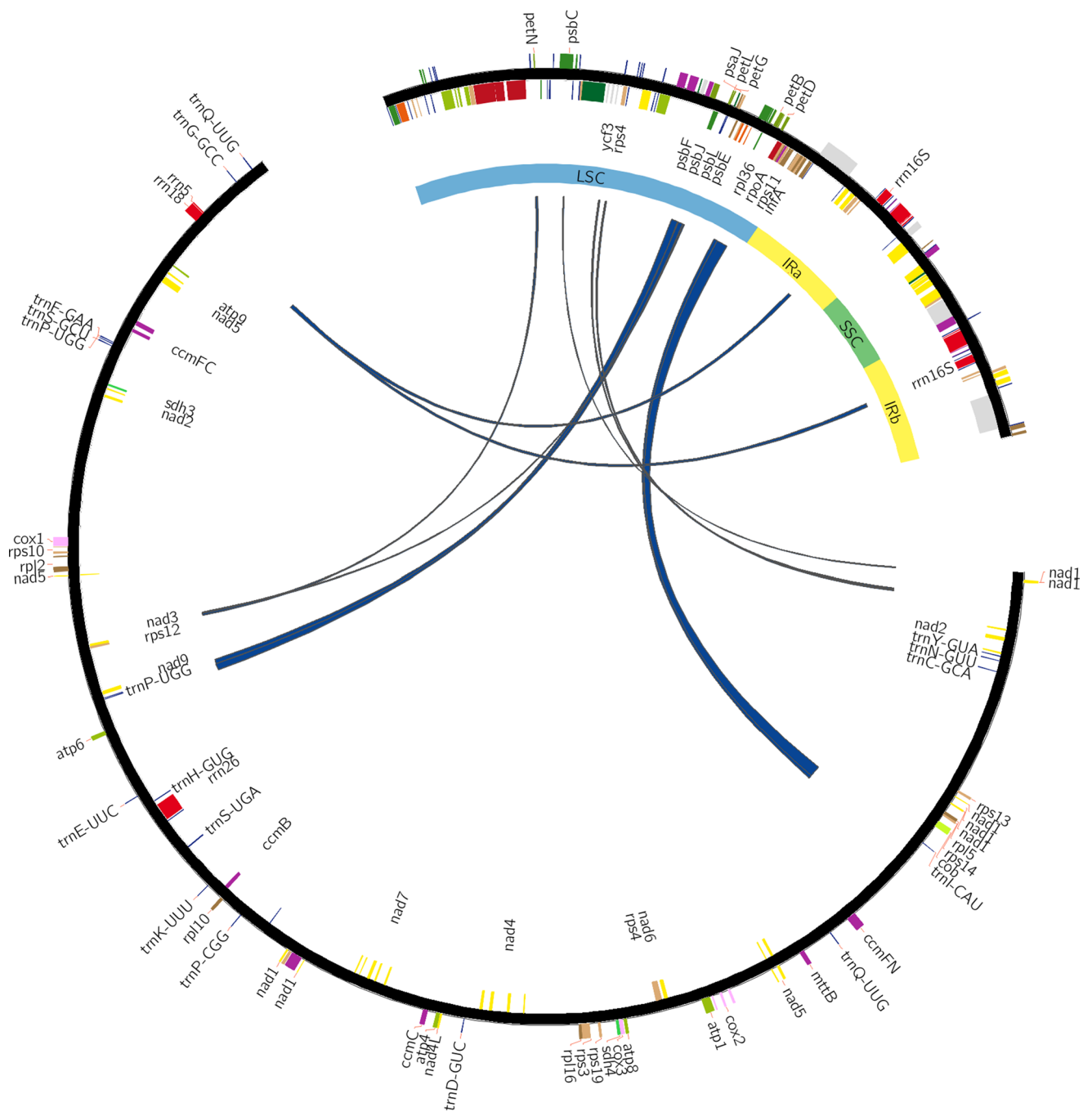
Fig. 6 Comparison of *Ribes alpinum* (right) and *R. nigrum* (left) mitochondrial genome structure and gene order. Genes inside and outside the outer circle are transcribed in counterclockwise and clockwise directions, respectively. Gray ribbons depict the homologous sequences between the analyzed species

regions with high identity and blast score values (Table S8), whose total length is 9,499 bp, which constitutes 2,53% of the mitochondrial genome. Their length varies from 147 to 2,931 bp and they cover at least partially from one to five coding regions. Nine of them are located in the LSC region while two (copies of *rrn16S*) in the inverted repeat (Fig. 7). Identity of homologous regions ranged from 73,87% to as much as 99,08% (Table S8). Homology of genome fragments located in LSC with mitogenome sequences probably indicate sequence transfer events between organelles. However, similarity of *rrn16S* genes from IRa and IRb regions

to *rrn18S* in mitogenome seems to result directly from the function of these genes and does not necessarily indicate the MTPT (mitochondrial-plastid) transfer event.

Phylogenetic analysis

The result of the plastome-based phylogenetic analysis is a well-resolved tree with very high degrees of reliability (Fig. 8). All nodes reached a bootstrap value of 100%, indicating the high significance of the clades with the exceptions are two branches that achieved a value of 77%. The



newly sequenced alpine currant plastome was classified into one clade together with *R. fasciculatum* var. *chinense*, with which it forms a clade sister to the other analyzed species of the genus *Ribes* L. The exact bootstrap values and species affinities are shown in the cladogram (Figure S2), attached in a supplement.

The development of molecular biology and bioinformatics tools has allowed for enormous progress in the sequencing of plant genomes (Kersey 2019); (Wang et al. 2024b). This makes it possible to learn about the evolutionary course of given groups of plants and the development of biology and phylogenetics thanks to various analyses.

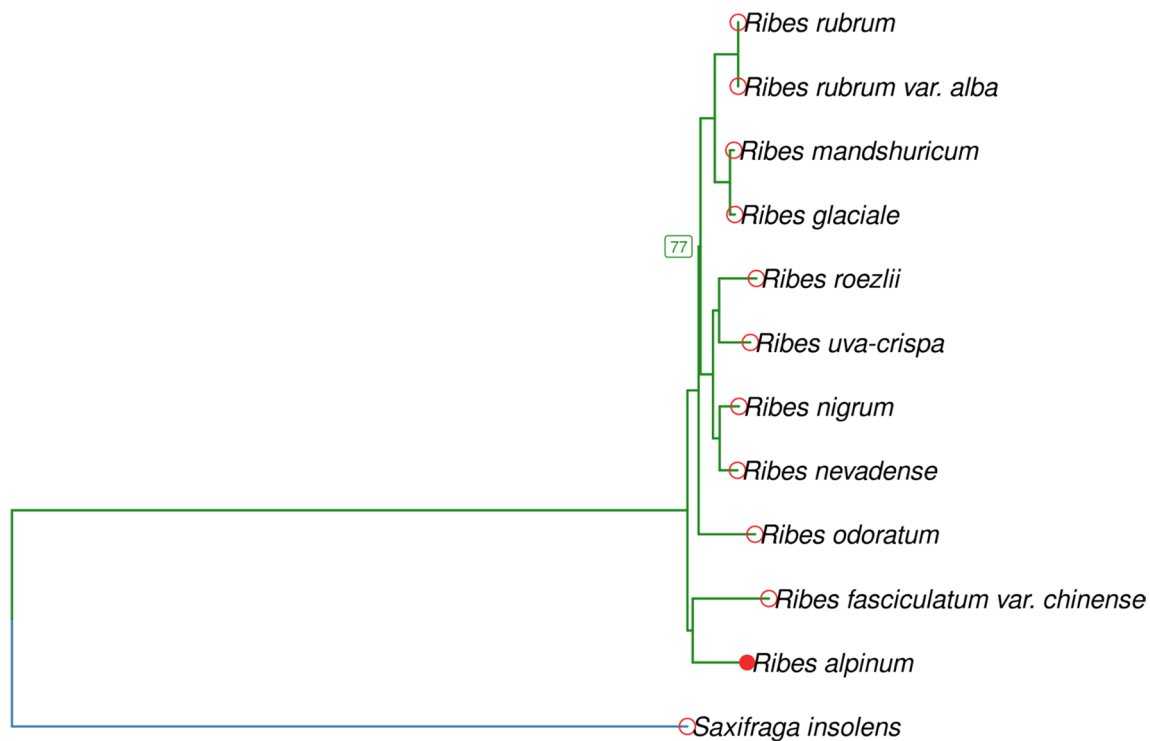


Fig. 8 Phylogenetics relationships of 12 species based on chloroplast genomes obtained using ML analysis. Bootstrap value lower than 100 is written at the node. Visualization performed in R Studio using the ggtree package

Research conducted on the chloroplast genome of plants from the genus *Saxifragales*, which also includes the *Grosulariaceae* family, represented by the genus *Ribes*, has shown high conservation of plastid genomes due to similar length, order and arrangement of genes and GC content (Li et al. 2019; Chen et al. 2023). Ten complete chloroplast sequences of *Ribes* have been described so far. The length of the chloroplast genome of *R. alpinum*, despite minor differences in the length of some regions, does not differ significantly from data from the chloroplast genomes of other species (~157,000 bp) (Sun et al. 2023; Wang et al. 2024a). The GC content (~38%), gene order and genome composition is comparable to other *Ribes* species, which indicates its conservativeness.

The performed functional analysis identified the IR region as more conserved than the LSC and SSC regions, which coincides with data from analyzes on other species of the genus *Saxifragales* (Wu et al. 2020; Sun et al. 2023). Structural heteroplasmy of the plastome has been discovered in species within angiosperms, gymnosperms, and pteridophytes using long-read sequencing (Wang et al. 2019). However, not in the case of *R. alpinum*, which possesses only one plastome structural variant. Despite the high conservation of the chloroplast genome in terms of its structure, composition and gene arrangement, comparison of plastome sequences in *Ribes* showed some regions in the genome with relatively high genetic variability. Part of observed

variable sites in the analyzed alignment are species-specific mutations enabling the successful molecular identification of *R. alpinum* from other representatives of the genus. However, the observed number of 1420 diagnostic positions in the alignment is not a very high value compared to the results of research on the variability of Poales, where the reported highest average value of observed molecular diagnostic characters among the 96 analyzed genera was 3071 (*Eriocaulon*, *Eriocaulaceae*) (Krawczyk et al. 2023).

Due to the still small number of known complete plastome sequences for super-barcoding analyses, it is worth looking for highly variable genome fragments to be used as specific barcodes or phylogenetic markers for a given group of plants. A detailed analysis of the genetic diversity of plastomes, taking into account individual functional regions and separately from their functions, allowed us to select regions in the chloroplast region with particular potential as molecular markers, also as DNA barcodes. Variation results from 11 available currant genomes revealed high numbers of variable sites mainly in the intergenic regions of the plastome, especially in *ndhF-rpl32*, *petN-psbM*, *trnT-GGU-psbD*, *ndhD-ccsA*, which is consistent with the analysis results for other *Ribes* species (Sun et al. 2023; Wang et al. 2024a). The gene with the highest variability in studied dataset was *ycf1*, the gene which two regions (*ycf1a* and *ycf1b*), according to literature show the highest nucleotide variability among the chloroplast genomes of angiosperms

(Dong et al. 2015). The above mentioned regions can be used in molecular research on a large group of plants for DNA barcoding.

The analysis of species-specific diagnostic nucleotides distribution in the genome, as well as nucleotide variation analysis, proved that noncoding regions show higher divergence than coding sites. In the chloroplast genome of *R. alpinum*, several particularly variable sites were detected in intergenic regions: *rps16-trnQ*-UGG, *petA-psbJ*, *rpl32-trnL*-UAG, *petN-psbM*, and *trnS-GCU-trnG*-UCU. Three of them: *rps16-trnQ*-UGG, *petA-psbJ* and *rpl32-trnL*-UAG were previously described in the literature as markers with the highest genetic diversity among various plant species (Shaw et al. 2014; Guo et al. 2014). These are some of the most informative regions and have been used to determine intraspecific variability, among others, to determine the phylogenetic structure, course of speciation or evolution (Aguirre-Liguori et al. 2014; Guo et al. 2014). Also the *petN-psbM* intergenic spacer, is a genetic marker previously applied to determine interspecies variability (Adams et al. 2010, 2013). The *trnS-GCU-trnG*-UCU region appeared as one of the hotspots. This region has not been used in studies on Grossulariaceae so far, but has been successfully used in phylogenetic studies, including Fabaceae (Zhang et al. 2009, 2015), Rosaceae (Vafadar et al. 2014) and Scrophulariaceae (Attar et al. 2011).

The highest nucleotide variability expressed with the largest number of SNPs and InDels is characterized by *ycf1*, *ycf2* and *rpoC2*. Research conducted on nine *Ribes* species identified the same genes as the most diverse (Sun et al. 2023). The *ycf2* gene, with one hotspot marked, is the largest known plastome gene and can be used for phylogenetic analysis (Zhong et al. 2019).

The picture of phylogeny based on data from complete chloroplast genomes differs significantly from the results obtained based on ITS regions in terms of the position of *R. alpinum* on the phylogenetic tree. Results based on ITS variability indicated a close relationship between *R. alpinum* and *R. rubrum*, which grouped in a common clade, as well as a more distant relationship with *R. nigrum*. According to these results, *R. fasciculatum* belongs to the early-diverged lineage. In turn, phylogenetic inference based on plastome variability shows that *R. alpinum* groups with *R. fasciculatum* in a clade sister to the other studied *Ribes* representatives, which includes *R. nigrum* and *R. rubrum*. The relatively close relationship of *R. alpinum* to *R. fasciculatum* was also confirmed by Zhang and coauthors (2023) in a study using the Rad-Seq sequencing method.

The mitochondrial genome of *R. alpinum* characterized in detail in the study is the second genome in the *Ribes* genus to be sequenced. Recent research showed that the structure and less often the sequence of plant mitogenomes

are subject to evolution (Wang et al. 2024a). The mitochondrial genome contains many repeated sequences and insertion/deletion sites or single nucleotide polymorphism (SNP) sites (Yang et al. 2022). Comparison of the mitogenome of *R. alpinum* and *R. nigrum* showed high similarity of the GC content, which amounted to 46% in alpine currant and 45.8% in black currant. High concordance of gene content between these two mitogenomes was also observed. In the mitogenome of *R. alpinum*, only the lack of *rps1* was found, the gene responsible for the biosynthesis of ribosomal proteins. As reported in the literature, this gene is not always present in mitochondrial genomes, which is attributed to the transfer of *rps1* to the cell nucleus or its replacement by chloroplast analog genes (Hazle and Bonen 2007). The order of the genes between analyzed mitogenomes highly differs from each other. The location of genes on mtDNA may differ between related species due to the susceptibility of the genome to rearrangements (Gualberto and Newton 2017). Large non-tandem repeats, found in angiosperms, are mainly responsible for the high recombination activity (Zardoya 2020). Research on other groups of plants, including genus *Beta* and *Zea mays*, have already shown changes in the order of gene location among the same genus, which the authors attribute to gene rearrangement (Allen et al. 2007; Darracq et al. 2011).

The high variability of mitogenomes is also influenced by the transfer of sequences originating from the chloroplast genome. So far, in the order Saxifragales, this phenomenon has been described in *Sedum plumbizincicola* (Crassulaceae), where the analysis revealed four fragments that were moved from the chloroplast genome to the mitochondrion (Ding et al. 2022). In turn, research on the same species recently published by other researchers (Tang et al. 2024) showed the presence of 15 homologous regions. In the same study the authors performed MTPT analysis for seven species representing Saxifragales from genus: *Loropetalum*, *Heuchera*, *Paeonia*, *Rhodiola*, *Sedum* and also for *R. nigrum*. The number of observed MTPTs ranged from 15 in *S. plumbizincicola* to 34 in *Loropetalum chinense*. In *R. nigrum* the analysis revealed as much as 20 genome fragments transferred. This result differs significantly from that obtained in our study for *R. alpinum*, which revealed 11 homologs which constitute 2.53% of the mitochondrial genome. It should be noted, however, that the number of regions described is significantly influenced by the adopted methodology - the literature data on *R. nigrum* (Ding et al. 2022) reported transfers of two trna genes and four IGS regions, which were not included in our analysis due to the high probability of obtaining false positive results. After excluding these regions, there remain 14 transfers observed in the black currant mitogenome. Interestingly, in the case of both currant species, the MTPT sequences differ

significantly in content. Of the ten protein coding genes at least partially located within the transfer regions in *R. nigrum* and 18 PCGs in *R. alpinum*, only four are common: *psbE*, *petL*, *ycf3* and *psbC*.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11295-024-01684-4>.

Author contributions All authors contributed to and accepted the final manuscript. Conceptualization: KKr, JS; Formal analysis: KKl, JS, LP, MM, KC; Acquisition of material in the field: KKr; Writing – Original draft: KKl, KKr; Writing – Review and Editing: JS, KKr, LP.

Funding This study was financially supported by Department of Botany and Evolutionary Ecology, University of Warmia and Mazury in Olsztyn.

Data availability Raw data are available in European Nucleotide Archive (ENA) under the accession number PRJEB76467. The genome sequences assembled for this study are available in the NCBI under accession numbers: PP657412.1 (chloroplast genome) and PP716826.1 (mitochondrial genome).

Declarations

Conflict of interest The authors have declared that no competing interests exist.

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

References

- Adams RP, Rumeu B, Nogales M, Fontinha SS (2010) Geographic Variation And Systematics Of *Juniperus Phoenicea* L. From Madeira And The Canary Islands. Snps From nrDNA and petN-psbM
- Adams RP, Boratynski A, Arista M et al (2013) Published on-line www
- Aguiar-Liguori JA, Scheinvar E, Eguarte LE (2014) Gypsum soil restriction drives genetic differentiation in *Fouquieria shrevei* (Fouquieriaceae). *Am J Bot* 101:730–736. <https://doi.org/10.3732/ajb.1400031>
- Allen JO, Fauron CM, Minx P et al (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177:1173–1192. <https://doi.org/10.1534/genetics.107.073312>
- Amiryousefi A, Hyvönen J, Poczar P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030–3031. <https://doi.org/10.1093/bioinformatics/bty220>
- Attar F, Riahi M, Daemi F, Aghabeigi F (2011) Preliminary molecular phylogeny of eurasian Scrophularia (Scrophulariaceae) based on DNA sequence data from trns-trng and ITS regions. *Plant Biosystems - Int J Dealing all Aspects Plant Biology* 145:857–865. <https://doi.org/10.1080/11263504.2011.590826>
- Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10. <https://doi.org/10.1186/1471-2105-10-421>
- Charif D, Lobry JR (2007) SeqinR 1.0–3: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis
- Chen X, Wu X, Zhang J et al (2023) Characterization of the complete mitochondrial genome of *Elymus magellanicus* (É.Desv.) Á.Löve (Poaceae, Pooideae). *Mitochondrial DNA B Resour* 8:795–798. <https://doi.org/10.1080/23802359.2023.2238931>
- Chung KP, Gonzalez-Duran E, Ruf S et al (2023) Control of plastid inheritance by environmental and genetic factors. *Nat Plants* 9:68–80. <https://doi.org/10.1038/s41477-022-01323-7>
- Cusimano N, Wicke S (2016) Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytol* 210:680–693. <https://doi.org/10.1111/nph.13784>
- Darracq A, Varré JS, Maréchal-Drouard L et al (2011) Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. *Genome Biol Evol* 3:723–736. <https://doi.org/10.1093/gbe/evr042>
- Ding H, Bi D, Zhang S et al (2022) The mitogenome of sedum plumbizincicola (Crassulaceae): insights into RNA editing, lateral gene transfer, and phylogenetic implications. *Biology (Basel)* 11. <https://doi.org/10.3390/biology11111661>
- Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review. *Acta Physiol Plant* 42
- Dong W, Xu C, Li C et al (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep* 5:8348. <https://doi.org/10.1038/srep08348>
- Filip E, Skuza L (2021) Horizontal gene transfer involving chloroplasts. *Int J Mol Sci* 22
- Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47:W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Gualberto JM, Newton KJ (2017) Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. <https://doi.org/10.1146/annurev-arplant-043015>
- Gualberto JM, Milesina D, Wallet C et al (2014) The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 100:107–120
- Guo GY, Yang RW, Ding CB et al (2014) Phylogenetic relationships among *Leymus* and related diploid genera (Triticeae: Poaceae) based on chloroplast trnQ-rps16 sequences. *Nord J Bot* 32:658–666. <https://doi.org/10.1111/j.1756-1051.2013.00233.x>
- Hazle T, Bonen L (2007) Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes. *Gene* 405:108–116. <https://doi.org/10.1016/j.gene.2007.09.019>
- Julian AT, Pombert J-F (2024) SYNy: a pipeline to investigate and visualize collinearity between genomes
- Kersey PJ (2019) Plant genome sequences: past, present, future. *Curr Opin Plant Biol* 48:1–8
- Kille B, Garrison E, Treangen TJ, Phillippy AM (2023) Minners are a generalization of minimizers that enable unbiased local Jaccard estimation motivation: the Jaccard similarity on k-mer sets
- Kitazaki K, Kubo T (2010) Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. *J Bot* 2010:1–12. <https://doi.org/10.1155/2010/620137>

- Krawczyk K, Pauksto Ł, Małdziarz M, Sawicki J (2023) The low level of plastome differentiation observed in some lineages of Poales hinders molecular species identification. *Front Plant Sci* 14. <https://doi.org/10.3389/fpls.2023.1275377>
- Krzywinski M, Schein J, Birol I et al (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Li X, Yang Y, Henry RJ et al (2015) Plant DNA barcoding: from gene to genome. *Biol Rev* 90:157–166. <https://doi.org/10.1111/BRV.12104>
- Li Y, Jia L, Wang Z et al (2019) The complete chloroplast genome of *Saxifraga sinomontana* (Saxifragaceae) and comparative analysis with other Saxifragaceae species. *Revista Brasileira De Bot* 42:601–611. <https://doi.org/10.1007/s40415-019-00561-y>
- Min SH, Zhou J (2021) Smplot: an R Package for easy and elegant data visualization. *Front Genet* 12:802894. <https://doi.org/10.3389/FGENE.2021.802894/BIBTEX>
- Mower JP, Sloan DB, Alverson AJ (2012) Plant mitochondrial genome diversity: The genomics revolution. In: *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics*. Springer-Verlag Vienna, pp 123–144
- Palmer JD (1985) Chloroplast DNA and molecular phylogeny. *BioEssays* 2:263–267. <https://doi.org/10.1002/BIES.950020607>
- Petersen G, Cuenca A, Möller IM, Seberg O (2015) Massive gene loss in mistletoe (*Viscum*, Viscaceae) mitochondria. *Sci Rep* 5. <https://doi.org/10.1038/srep17588>
- Shaw J, Shafer HL, Rayne Leonard O et al (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot* 101:1987–2004. <https://doi.org/10.3732/ajb.1400398>
- Sloan DB, Alverson AJ, Chuckalovcak JP et al (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* 10. <https://doi.org/10.1371/journal.pbio.1001241>
- Sun X, Zhan Y, Li S et al (2023) Complete chloroplast genome assembly and phylogenetic analysis of blackcurrant (*Ribes nigrum*), red and white currant (*Ribes rubrum*), and gooseberry (*Ribes uva-crispa*) provide new insights into the phylogeny of Grossulariaceae. *PeerJ* 11. <https://doi.org/10.7717/peerj.16272>
- Tan W, Gao H, Jiang W et al (2020) The complete chloroplast genome of *Gleditsia sinensis* and *Gleditsia japonica*: genome organization, comparative analysis, and development of taxon specific DNA mini-barcodes. *Sci Rep* 10. <https://doi.org/10.1038/s41598-020-73392-7>
- Tang P, Ni Y, Li J et al (2024) The complete mitochondrial genome of *Paeonia lactiflora* pall. (Saxifragales: Paeoniaceae): evidence of gene transfer from Chloroplast to mitochondrial genome. *Genes (Basel)* 15. <https://doi.org/10.3390/genes15020239>
- Vafadar M, Osaloo SK, Attar F (2014) Molecular phylogeny of the genus *Amygdalus* (Rosaceae) based on nrDNA ITS and cpDNA trns-trng sequences. *Turk J Bot* 38:439–452. <https://doi.org/10.3906/bot-1303-46>
- Wang D, Wu YW, Shih ACC et al (2007) Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Mol Biol Evol* 24:2040–2048. <https://doi.org/10.1093/molbev/msm133>
- Wang W, Lanfear R, Gaut B (2019) Long-reads Reveal that the Chloroplast Genome exists in two distinct versions in most plants. *Genome Biol Evol* 11:3372–3381. <https://doi.org/10.1093/gbe/evz256>
- Wang J, Kan S, Liao X et al (2024a) Plant organellar genomes: much done, much more to do. *Trends Plant Sci*. <https://doi.org/10.1016/j.tplants.2023.12.014>
- Wang J, Kan S, Liao X et al (2024b) Plant organellar genomes: much done, much more to do. *Trends Plant Sci*. <https://doi.org/10.1016/j.tplants.2023.12.014>
- Wickham H (2011) ggplot2. *Wiley Interdiscip Rev Comput Stat* 3:180–185. <https://doi.org/10.1002/WICS.147>
- Wu Z, Liao R, Yang T et al (2020) Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-07045-4>
- Wynn EL, Christensen AC (2019) Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3: genes*. *Genomes Genet* 9:549–559. <https://doi.org/10.1534/g3.118.200948>
- Yang J, Ling C, Zhang H et al (2022) A Comparative Genomics Approach for Analysis of Complete Mitogenomes of five Actinidiaceae plants. *Genes (Basel)* 13. <https://doi.org/10.3390/genes13101827>
- Yu G, Smith DK, Zhu H et al (2017) Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zardoya R (2020) Recent advances in understanding mitochondrial genome diversity. *F1000Res* 9
- Zervas A, Petersen G, Seberg O (2019) Mitochondrial genome evolution in parasitic plants. *BMC Evol Biol* 19. <https://doi.org/10.1186/s12862-019-1401-8>
- Zhang M, Fritsch PW, Cruz BC (2009) Phylogeny of *Caragana* (Fabaceae) based on DNA sequence data from rbcL, trnS-trnG, and ITS. *Mol Phylogenet Evol* 50:547–559. <https://doi.org/10.1016/j.ympev.2008.12.001>
- Zhang ML, Wen Z, Bin, Hao XL et al (2015) Taxonomy, phylogenetics and biogeography of *Chesneya* (Fabaceae), evidenced from data of three sequences, ITS, trnS-trnG, and rbcL. *Biochem Syst Ecol* 63:80–89. <https://doi.org/10.1016/j.bse.2015.09.017>
- Zhang B, Yu Z, Xu Z, Zheng B (2023) A phylogenetic and morphological evolution study of *Ribes* L. in China using RAD-Seq. *Plants* 12. <https://doi.org/10.3390/plants12040829>
- Zhong Q, Yang S, Sun X et al (2019) The complete chloroplast genome of the Jerusalem artichoke (*Helianthus tuberosus* L.) and an adaptive evolutionary analysis of the ycf2 gene. *PeerJ*. <https://doi.org/10.7717/peerj.7596>

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