Plant Diversity of Central Asia

PDCA ISSN 2181-2403
PDCA ISSN 2181-239X

RESEARCH ARTICLE

Plastid genomes of four species of *Iris* from subgenus *Scorpiris*

Elyor A. Ortikov[1], Mamura B. Kurbonalieva[1,3,4,5], Kumush B. Alieva[1], Sabina A. Khudoyberdieva[1], Temur N. Asatulloev[3,4,5,6], Ziyoviddin O. Yusupov[1,2,3,4,*] Feruza U. Mustafina[1,2*]

[1]*Institute of Botany, Academy of Sciences, Tashkent 100125, Uzbekistan*
[2]*Tashkent Botanical Garden named after F.N. Rusanov, Tashkent, Uzbekistan*
[3]*Yunnan International Joint Laboratory for Biodiversity of Central Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China.*
[4]*CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China.*
[5]*Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China.*
[6]*University of Chinese Academy of Sciences, Beijing, 100864, China*

✉ yusupov@mail.kib.ac.cn; mustafinaferuza@yahoo.com

ABSTRACT

The taxonomic complexity of *Iris* (Iridaceae) is high, and its molecular phylogeny remains largely unknown. With the growing availability of full plastid genomes of species representing major taxonomic groups within *Iris*, it will become possible to produce a robust phylogenetic tree and reconstruct the evolutionary history of the genus. With this objective in mind, we sequenced and analyzed the chloroplast genomes of four species from subgenus *Scorpiris* native to Uzbekistan: *Iris austrotschatkalica*, *I. pseudocapnoides*, *I. victoris* and *I. hippolyti*. The phylogenetic tree obtained agreed well with those produced earlier with a limited number of chloroplast markers. We suggest as the most useful for phylogenetic analyses of *Iris* the following loci that show high variability and lack of loop structures: *atpF*/*atpH*, *rps15*/*ycf1*, *atpA*/*atpF*, *trnL-UAA*/*trnF-GAA*, *trnG-UCC*/*trnR-UCU*, and *psaA*/*ycf3*.

Key words: chloroplast genome, nucleotide divergence, phylogenic markers, phylogeny, *Iris*

Introduction

Iris L., one of the most diverse genera of Iridaceae, is widespread in the Northern Hemisphere. It comprises ca. 250–300 species (Goldblatt et al. 1998; Crespo et al. 2015). According to Mathew (1989) and can be subdivided, in part according to their underground storage organs. into six subgenera. One of them, subg. *Scorpiris* Spach (synonym Juno Tratt.), known colloquially as 'the junos', is a distinct group of irises having bulbs with papery tunics and usually fleshy, sometimes thick, storage roots during the dormant stage. With approximately 55 species distributed in central and southwestern Asia, the Caucasus, and the Mediterranean, this is the largest subgenus in *Iris* (Mathew, 2001).

http://doi.org/10.54981/PDCA/vol2_iss2/a4

Received: 13 March 2023; Accepted: June 2023 Academic editor: Hyeok Jae Choi
Plant Diversity of Central Asia (2023) 2(2): 102–123

DNA-based phylogenies of *Iris* have been assessed using various plastid DNA regions (Tillie et al. 2001; Wilson 2004, 2009, 2011; Ikinci et al. 2011; Mavrodiev et al. 2014). Those studies showed that subgenus *Scorpiris* is monophyletic and deeply embedded within Iris and therefore should not be considered to be the distinct genus, *Juno*. However, those conclusions still require confirmation because in none of the studies did the number of plastid regions exceed ten, with the total chloroplast genome coverage being less than 8300 bp.

In this study, we sequenced the plastome genomes of four species of *Scorpiris*: *Iris austrotschatkalica* Tojibaev, F.Karimov & Turgunov, *I. victoris* F.O.Khass., Khuzhan. & Rakhimova in Stapfia, *Iris pseudocapnoides* Rukšāns, *I. hippolyti* (Vved.) Kamelin. The species names follow Sennikov et al (2023). Our objectives were to i) characterize the *Scorpiris* chloroplast genome, ii) check the phylogenetic position of the species of *Scorpiris*, and ii) determine the most variable regions useful for future studies of Iris phylogeny.

Materials and Methods

DNA extraction

DNA was extracted from the leaves of the specimens collected in the field (Fig. 1). Vouchers were deposited at TASH. DNA extraction followed Doyle and Doyle (1990). The product was checked by agarose gel (1.3%) electrophoresis with a DNA ladder (Thermo Scientific, Waltham, MA, U.S.A). DNA quality and its concentration were checked with a spectrophotometer NanoDrop 2000 (Thermo Scientific, U.S.A.).

Sequencing, genome assembly and annotation

DNA was sheared to construct a 350 bp (insert size) sequence library with the Genomic DNA Sample Prep Kit (Illumina) according to the manufacturer's protocol and sequenced using 150 paired-end reads on the Illumina HiSeq 4000 at Beijing Novogene Bioinformatics Technology Co., Ltd, China. For raw data processing, the Next Generation Sequencing (NGS) QC Tool Kit with default settings (Patel and Jain 2012) was used. The resulting clean reads were assembled in NovoPlasty version 3.8.3 (Dierckxsens et al. 2017). The plastome sequence of Iris gatesii (KM014691.1) was used as the reference genome. Gene annotation was performed in Geneious v.10.0.2 by the reference. The start and stop codons, as well as intron/exon boundaries for genes were checked manually (Kearse et al. 2012). Gene nomenclature follows the Chloroplast Genome Database (http://chloroplast.cbio.psu.edu; Cui et al. 2006). The generated chloroplast genome annotation files for *I. austrotschatkalica* (NC068838), *I. pseudocapnoides* (NC068839), *I. victoris* (NC068840) and *I. hippolyti* (OK138594) were converted into feature table files with GB2squine (Lehwark and Greiner 2019) and submitted to GenBank via BankIt. The genome circular maps with nomenclature and arrangement of the plastome's major structural components were drawn by OrganellarGenomeDRAW version 1.3.1 (Lohse et al. 2013). The inverted repeats (IR) - single copy (SC) boundaries of the four species of Scorpiris were compared with those of six species representing other subgenera of *Iris* (subfamily Iridoideae) and of a representative of subfamily Ixioideae *Crocus sativus* L.

Genome structure and codon usage analyses

The GC content and codon usage were examined using MEGA6 v. 6.06 (Tamura et al. 2013). Complete chloroplast genomes were aligned using Mauve v. 2.3.1 (Darling et al. 2010). The IR/SSC and IR/LSC boundary analyses among species of *Scorpiris* were performed in Geneious v.10.0.2 (Kearse et al. 2012). Aligned protein coding gene sequences and intergenic spacer (IGS) regions were extracted for further analyses. Nucleotide diversity and Ka/Ks ratio of protein coding regions were analyzed using DnaSP v. 5.10.01 (Librado and Rozas 2009) and MEGA6. Each protein coding region, intergenic spacer, and intron were compared using MUSCLE (Edgar 2004). SNP finder program

integrated in Geneious v. 10.3 (Kearse et al. 2012) was used to detect polymorphic sites (Ts and Tv).

Repeat sequence analysis

Simple sequence repeats (SSRs) were detected with Simple Sequence Repeat Locator (da Maia et al. 2008) using a size threshold ≥10 bp and length of repeated sequence from 2 to 6. Repeats were identified by Tandem Repeats Finder v. 4.07b (Benson 1999) and REPuter (Kurtz et al. 2001). The following settings for detecting direct and inverted (palindromic) repeats were used: (1) hamming distance equal to 3; (2) minimal repeat size of 30 bp; and (3) sequence identity ≥90%.

Phylogenetic Analysis

The dataset for phylogenetic analysis included plastomes of 4 newly sequenced species of *Scorpiris plus* sequences of 28 species of *Iris* and *Sisyrinchium angustifolium* (outgroup) downloaded from NCBI. The sequences were aligned using MAFFT (Katoh & Standley 2013). The phylogenetic tree was constructed based on 65 protein coding genes (CDS). After extraction, the protein coding gene sequences were aligned using MUSCLE v.3.6 (Edgar 2004) and then concatenated as a super locus of coding DNA sequences using SequenceMatrix v.1.8 (Vaidya et al. 2011). Maximum parsimony (MP) was performed using PAUP* version 4.10 (Swofford 2003). All characters were equally weighted, gaps were treated as missing and character states were treated as unordered. A heuristic search was performed with TBR branch swapping and the Multrees option, and random stepwise addition with 1000 replications. All analyses used GTR + G as the best-fitting model of nucleotide substitutions selected in jModelTest v.2.1.4 (Darriba 2012) under the Akaike information criterion (AIC). For the Maximum likelihood (ML) inference we used RAxML v.8.0 (Stamatakis 2014) with 1000 bootstrap replicates. Bayesian analysis was performed in MrBayes v.3.2 (Ronquist 2011) with 10.000 bootstrap replicates.

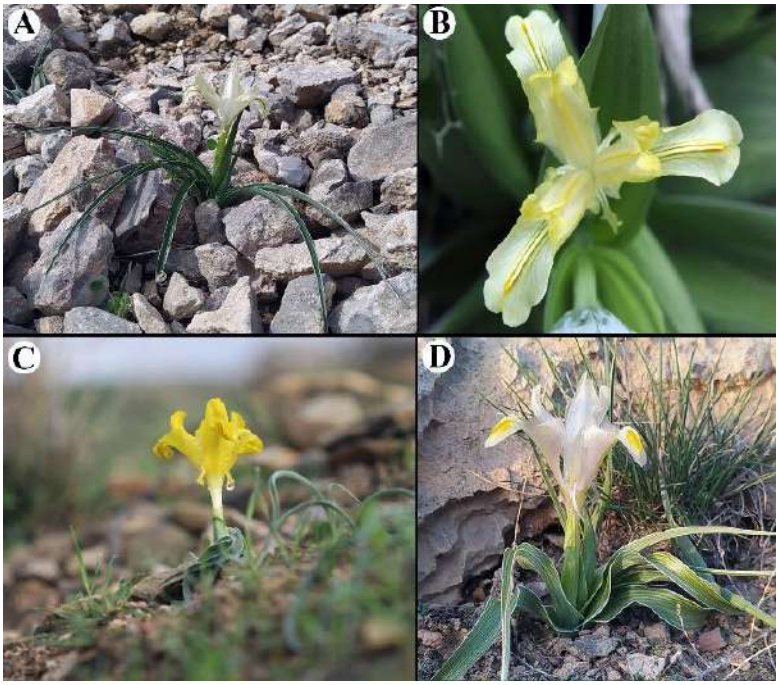

Fig. 1 A-D. The study species. *Iris austrotschatkalica*, *I. pseudocapnoides*, *I. victoris* and *I. hippolyti*, respectively. Photos by E. Ortikov

Results

General plastome features

The plastomes of four species of *Scorpiris* were identical in their structural organization, gene content, and gene arrangement (Table 1, Fig. 2). Among them, the plastome of *I. austrotschatkalica* was the smallest (150,887 bp) and of *I. pseudocapnoides* the largest (151,393 bp). The gene sets were identical, each with 4 unique rRNA genes, 30 unique tRNA genes, and 79 unique protein coding genes. Each plastome contained 133 complete coding regions including duplicated genes in the IR and open reading frames (ORFs). The total GC content in all species of *Scorpiris* was 38%. The GC content of the IR, LSC and SSC regions is presented in Table 1. Within protein coding regions of *I. austrotschatkalica* and *I. pseudocapnoides*, the GC content was higher at the first codon position (62%), whereas in *I. victoris* and *I. hippolyti* the GC content was higher at the third codon position (62.2% and 62.9% correspondingly).

The boundary positions of IR-single copy of the four species of *Scorpiris*, four species of *Limniris*, two species of *Pardanthopsis* of *Iris* L. and one species of *Crocus* L. – *C. sativa* are shown in Fig. 3. In all species studied, the IRb/LSC boundary fell between *rpl22* and *rps19*, specifically in *ycf1*, creating *ycf1* pseudogenes.

Identification of highly variable regions

The average sequence diversity for the four species of *Scorpiris* was 0.4% for the whole genome, and 0.78, 0.40 and 0.10% for SSR, LSC and IR, respectively. Diversity of 83 protein coding genes is presented in (Table S1. Fig. 4). Average Ka/Ks ratios for those genes were 1.14 across entire plastomes, 1.12 for the Irs, and 1.11 for the LSC and 1.20 for the SSC regions. Diversity values of protein coding genes ranged 0 – 1.4%. Genes showing the highest diversity were rpl22 (1.3%), rpl32 (1.2%), rpl16 (1.1%), and ycf2 (0.9%). Genes *rpoC2, accD, ycf2*, and *ycf1* exhibited length differences in pairwise comparisons due to the presence of multiple indels (15, 18, 45, and 87 bp, respectively).

In total, 113 intergenic spacer regions with the length > 10 bp were detected in the plastomes of *Scorpiris* (Table S2). Their diversity values ranged from 0 – 17.1%. The most divergent IGS regions were *rps16/trnQ-UUG* (17.1%), *atpF/atpH* (14.8%), *rps3/rpl22* (13.6%), and *trnK-UUU/rps16* (11.7%) in the LSC, *rrn4.5/rrn5* (2.6%) in the IR, and rps15/ycf1 (11.3%) in the SSC. Small and medium size insertions and deletions prevailed over long length indels in the IGS regions.

In 18 introns, diversity values ranged 0-2.3% (Table S3). The most divergent introns were detected in genes *clpP* (2.3%), *rpl16* (2.1%) and *petB* (2%) in the LSC and ndhA (1.3%) in the IR.

The data on polymorphic sites in the four plastomes of *Scorpiris* are presented in Tables S1 – S3. The total number of polymorphic sites in the protein coding regions was 572 (24,335 and 213 in the IR, LSC and SSC, respectively), 634 (16,522, and 96 in the IR, LSC and SSC, respectively) in the intergenic spacer regions, and 149 (3,131 and 15 in the IR, LSC and SSC, respectively) in the introns. The ratio Ts:Tv was 79:21, 64:35 and 66:34, respectively, for the three types of genetic regions. The total number of polymorphic sites in all regions of the plastome was 1355 (43 in the IR, 988 in LSC, and 324 in SSC), representing 70.5 % Ts and 29.5% Tv.

Overall, the highest nucleotide variability showed the following genetic regions: *rps16/trnQ-UUG, atpF/atpH, rps3/rpl22, trnK-UUU/rps16, rps15/ycf1, petA/psbJ, /psbA, trnG-GCC/trnfM-CAU, trnR-UCU/atpA, atpA/atpF, rps8/rpl14, trnL-UAA/trnF-GAA, trnG-UCC/trnR-UCU* and *psaA/ycf3* (Table 2).

After identification of the 14 most variable regions, their structure was analyzed for the purposes of primer design (Tables S1-S3 and Fig. 4). In eight out of 14 we detected hairpin structures created by palindromic repeats (Table 2 and Fig. 5).

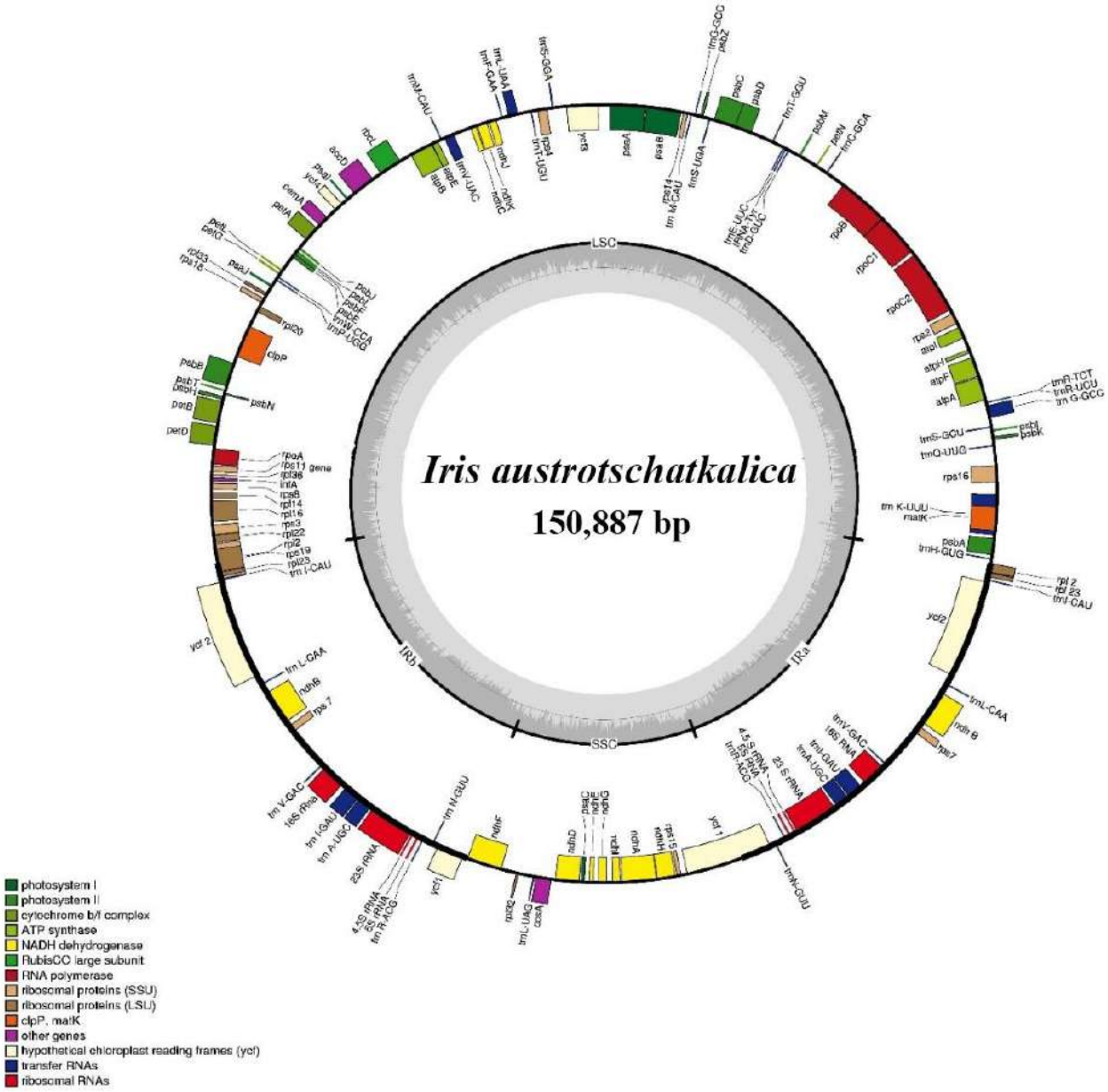

Fig. 2. Circular plastome map of *Iris austrotschatkalica*. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. Genes transcribed clockwise are shown inside of the circle; genes transcribed counter-clockwise are shown outside of the circle. The internal gray circle indicates GC content and the thin circular line marks the 50% threshold.

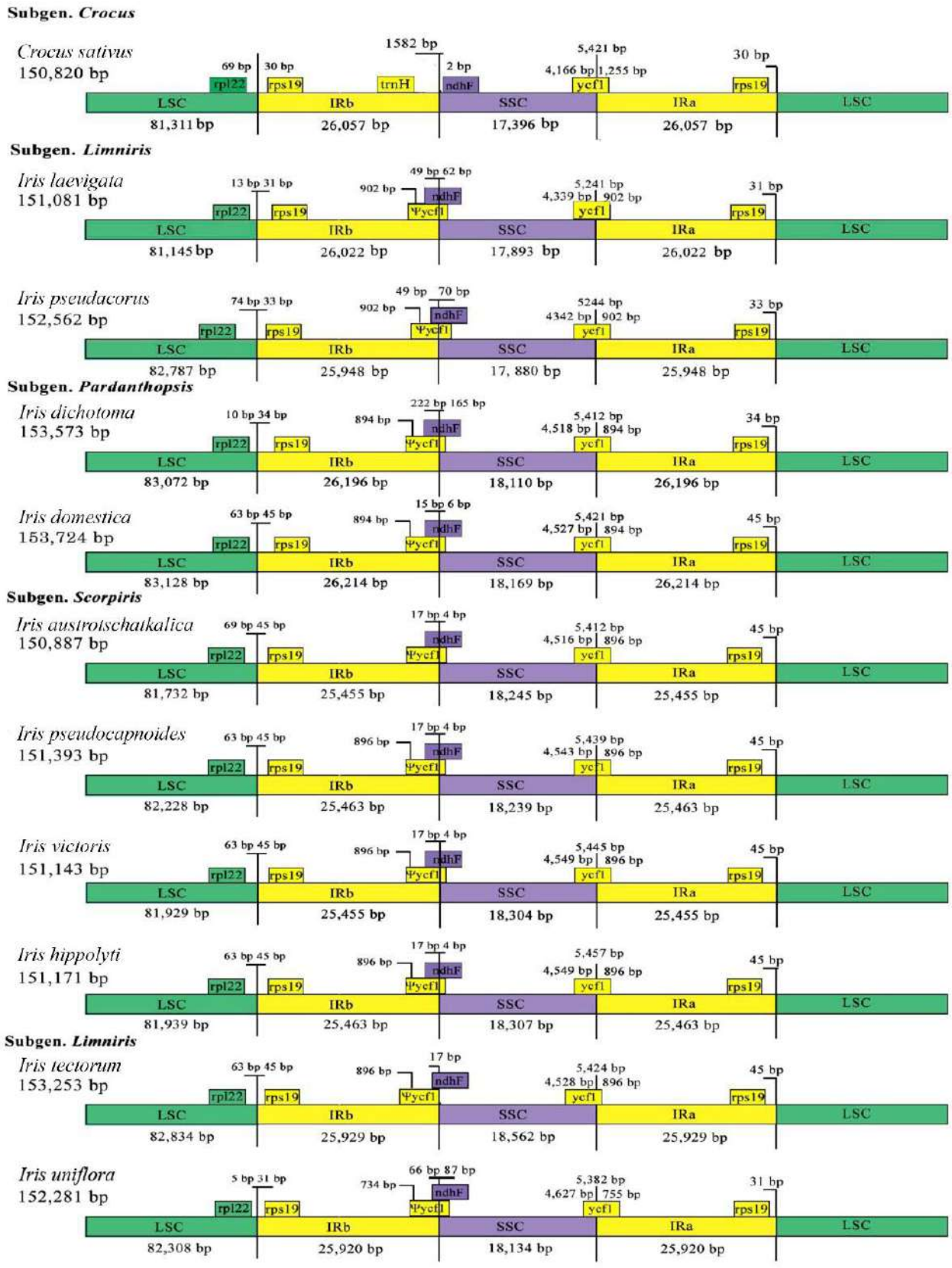

Fig. 3. Schematic comparison of the large single-copy (LSC), inverted repeat (IRa, IRb), and a small single-copy (SSC) regions among one Ixioideae (*Crocus sativus*) and 10 Irioideae plastid genomes, showing the genes overlapping or flanking IR single-copy junctions (shown sizes do not correspond to the actual sizes). Pseudogenes are denoted by ψ (i.e. ψycf1).

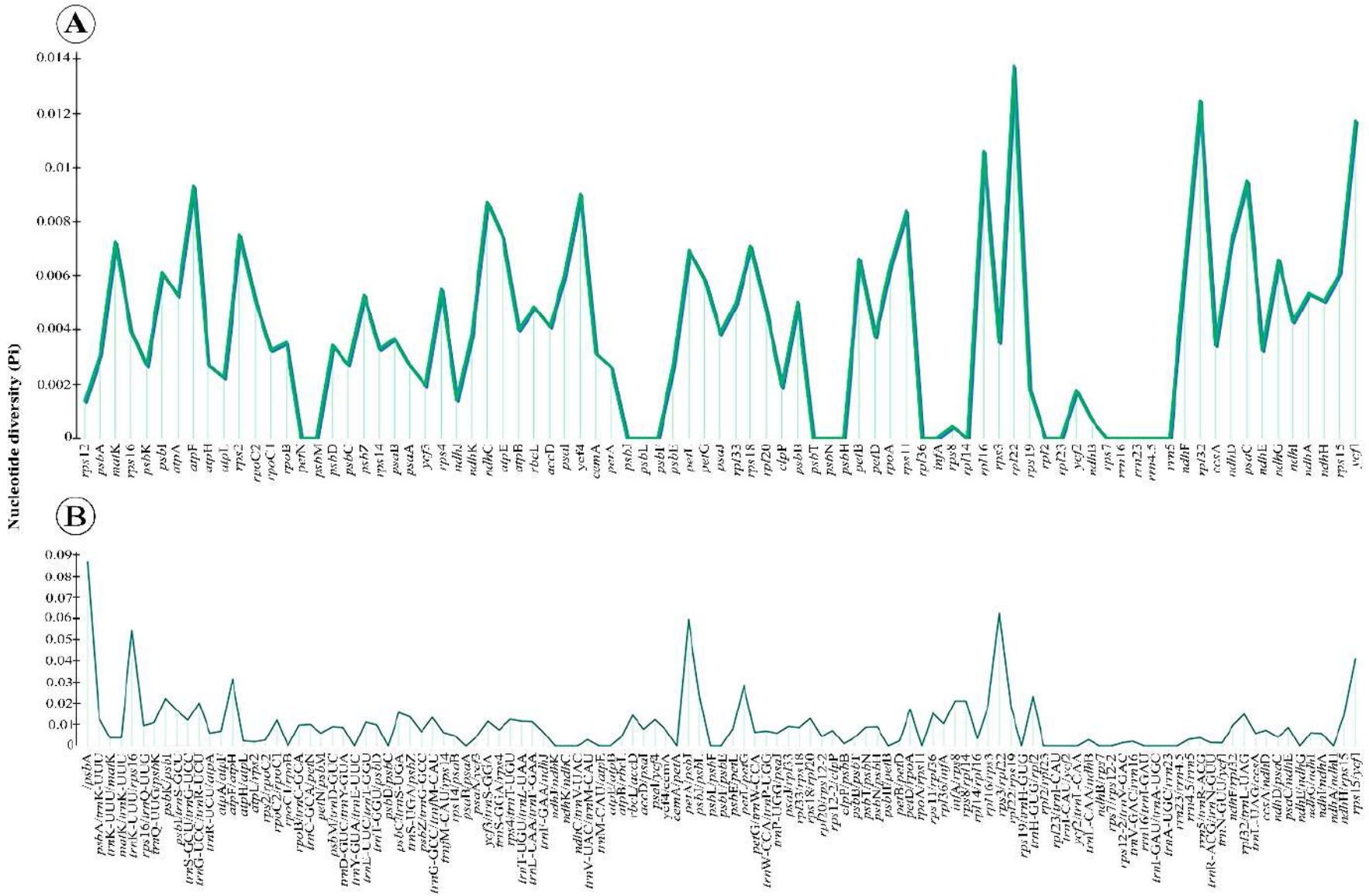

Fig. 4. Nucleotide diversity (Pi) of coding regions (A) and non-coding regions (B).

Table 1 Comparison of major structural features of the studied plastomes

Feature	*I. austrotschatkalica*	*I. pseudocapnoides*	*I. victoris*	*I. hippolyti*
Accession number	NC068838	NC068839	NC068840	OK138594
Entire plastome size (bp)	150,887	151,393	151,143	151,171
IR size (bp)	25,455	25,463	25,455	25,463
LSC region size (bp)	81,732	82,228	81,929	81,938
SSC region size (bp)	18,245	18,239	18,304	18,307
Number of coding regions	133	133	133	133
Number of genes	113	113	113	113
Number of protein-coding genes	79	79	79	79
Number of genes duplicated in the IR	20	20	20	20
Number of open reading frames	4	4	4	4
Number of pseudo- genes	1 (*ycf1*)	1 (*ycf1*)	1 (*ycf1*)	1 (*ycf1*)
Number of *tRNA* genes	30	30	30	30
Number of *rRNA* genes	4	4	4	4
Number of genes with intron(s)	20	20	20	20
GC content of whole genome (%)	38	38	38	38
GC content (%) LSC/ IR/ SSC	36.3/43/31.6	36.3/43.1/31.5	36.3/43.1/31.5	36.3/43/31.6
GC content in codon regions (%), 1/2/3	38.5/37.1/38.6	37.8/37.8/38.4	38.5/38.3/37.3	37.5/38.9/36.7

Table 2. The most divergent regions in protein coding genes, intergenic sequences and introns of the studied species

#		Aligned Length, bp	Indel, bp	Number of polymorphic sites	Nei nucleotide diversity	Diversity (%)	Notes
1	*rps*16/*trn*Q-UUG	879	1;206;1	13	0,00967	17.1	24 bp palindromic repeat
2	*atp*F/*atp*H	375	1;99;2		0.03152	14.8	No hairpin structure.
3	*rps*3/*rpl*22	68	1;1;26;6	10	0.0625	13.6	Two 20 bp forward repeats in *I. pseudacapnoides*
4	*trn*K-UUU/*rps*16	662	2;1;11;120;2;1;6;3	12	0.0546	11.7	20 bp palindromic repeat in *I. austrotschatkalica* and *I. pseudacapnoides*
5	*rps*15/*ycf*1	495	2;25,13,4;1;4;5;2;2,1;36	29	0.04088	11.3	No hairpin structure
6	*pet*A/*psb*J	846	5;1;7;134	42	0.05971	11.4	Two dispersed 21 bp palindromic repeats
7	*psb*A	77	0	13	0.08658	8.7	20 bp palindromic repeat
8	*trn*G-GCC/*trn*fM-CAU	268	7;1;20;6	10	0.01353	7.8	Two 20 bp forward repeats in *I. pseudacapnoides*
9	*trn*R-UCU/*atp*A	124	1;1;13	3	0.00612	7	Two 13 bp forward repeats in *I. hippolyti*
10	*atp*A/*atp*F	77	7	1	0.00714	6.8	No hairpin structure
11	*rps*8/*rpl*14	187	20	10	0.0625	6.8	Two forward repeats of 20 bp in *I. pseudacapnoides*
12	*trn*L-UAA/*trn*F-GAA	389	2;5;35	7	0.01153	6.7	No hairpin structure
13	*trn*G-UCC/*trn*R-UCU	142	6;4;3	5	0.0202	5.6	No hairpin structure
14	*psa*A/*ycf*3	637	6;1;2;4;24;4	5	0.00473	4.4	No hairpin structure

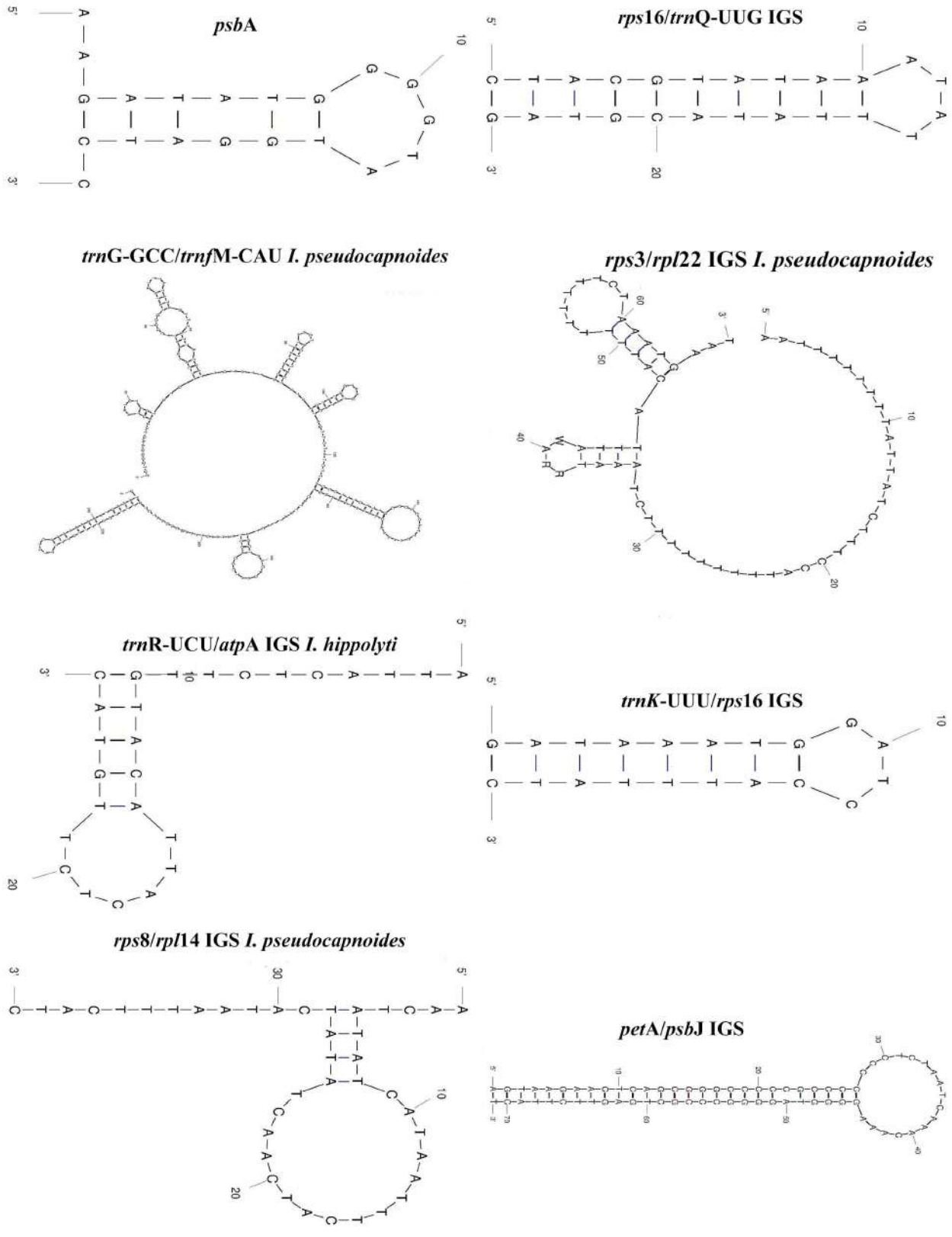

Fig. 5. Location of hairpin loops within *Iris* chloroplast genome.

Repeat analysis

The number of short motif SSRs with lengths ≥10 bp was 35, 32, 34 and 37 for *I. austrotschatkalica, I. pseudocapnoides, I. victoris*, and *I. hippolyti*, respectively. Repeat lengths of A/T class of mononucleotides of *I. austrotschatkalica, I. hippolyti, I. victoris and I. pseudocapnoides* were 25, 29, 23 and 21, respectively. For other types of repeat classes repeat lengths ranged from 0 to 3 (Fig. 6). The distribution of the detected repeats among 5 classes (tandem, forward, reverse, complementary and palindromic) are shown in Figure 5. In *I. austrotschatkalica*, a poly-A nucleotide repeat of 16 bp was found in the ndhF and rpl32 IGS in SSR. Poly-C nucleotide repeat of 10 bp was found only in I. austrotschatkalica and did not occur in plastomes of I. pseudocapnoides, I. victoris or I. hippolyti. Poly-G nucleotide repeats of ≥10 bp length were not detected in surveyed plastomes.

Dinucleotide AT repeats with the lengths of 10 bp were found in *I. austrotschatkalica, I. victoris*, and *I. hippolyti*, and with lengths of 10-12 bp in *I. pseudocapnoides*. One GA dinucleotide repeat with a length of 10 bp was detected in plastomes of *I. austrotschatkalica* and *I. victoris*, and one TC nucleotide repeat with a length of 10 bp was in *I. victoris* and in *I. hippolyti*.

The plastomes of *I. austrotschatkalica, I. pseudocapnoides, I. victoris* and *I. hippolyti* were surveyed to detect large tandem repeats with repeat size ≥ 30 bp and sequence identity >90%. Repeat units were repeated 2 – 4 times, with the majority occurring in the LSC and IRa regions. Among the five uniquely occurring large tandem repeats identified in Iris, most were palindromic dispersed. Two repeats forming 50 bp tandem had identical features and were located within the same genomic regions in four plastomes, i.e. in IGS region of *petN-psbM*.

Phylogenetic analysis

In our phylogenetic tree (Fig. 7), the four species of *Scorpiris* comprised clade A, which was imbedded in clade B, whose other members were two species from subg. *Crossiris* (=*Evansia*) *I. japonica* (=*Evansia chinensis*) and *I. tectorum* (=*Evansia tectorum*). The latter clade was sister to clade C composed of representatives of subg. *Iris* (sections *Oncocyclus* and *Iris*), *I. domestica* (=*Belamcanda chinensis*) and *I. dichotoma* (=*Pardanthopsis dichotoma*). These three monophyletic clades (A, B and C) were highly supported (bootstrap 100%).

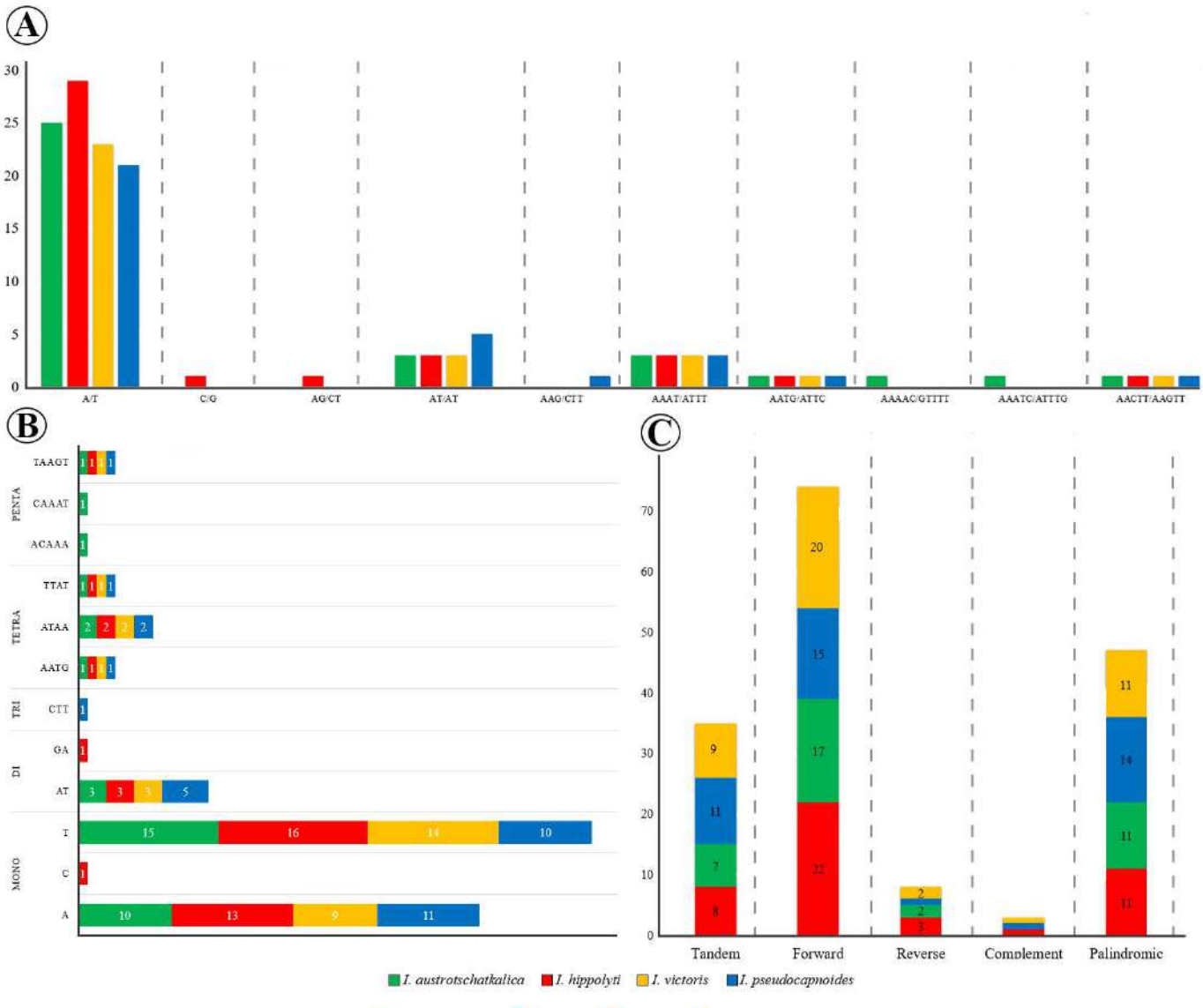

Fig. 6. Statistics of repeat elements in the plastomes of four studied *Scorpiris* species. A. Frequency of identified SSR motifs in 4 plastomes. B. Number of different simple sequence repeat (SSR) types detected in 4 plastomes. C. Number of tandem, forward, reverse, complement and palindromic repeats.

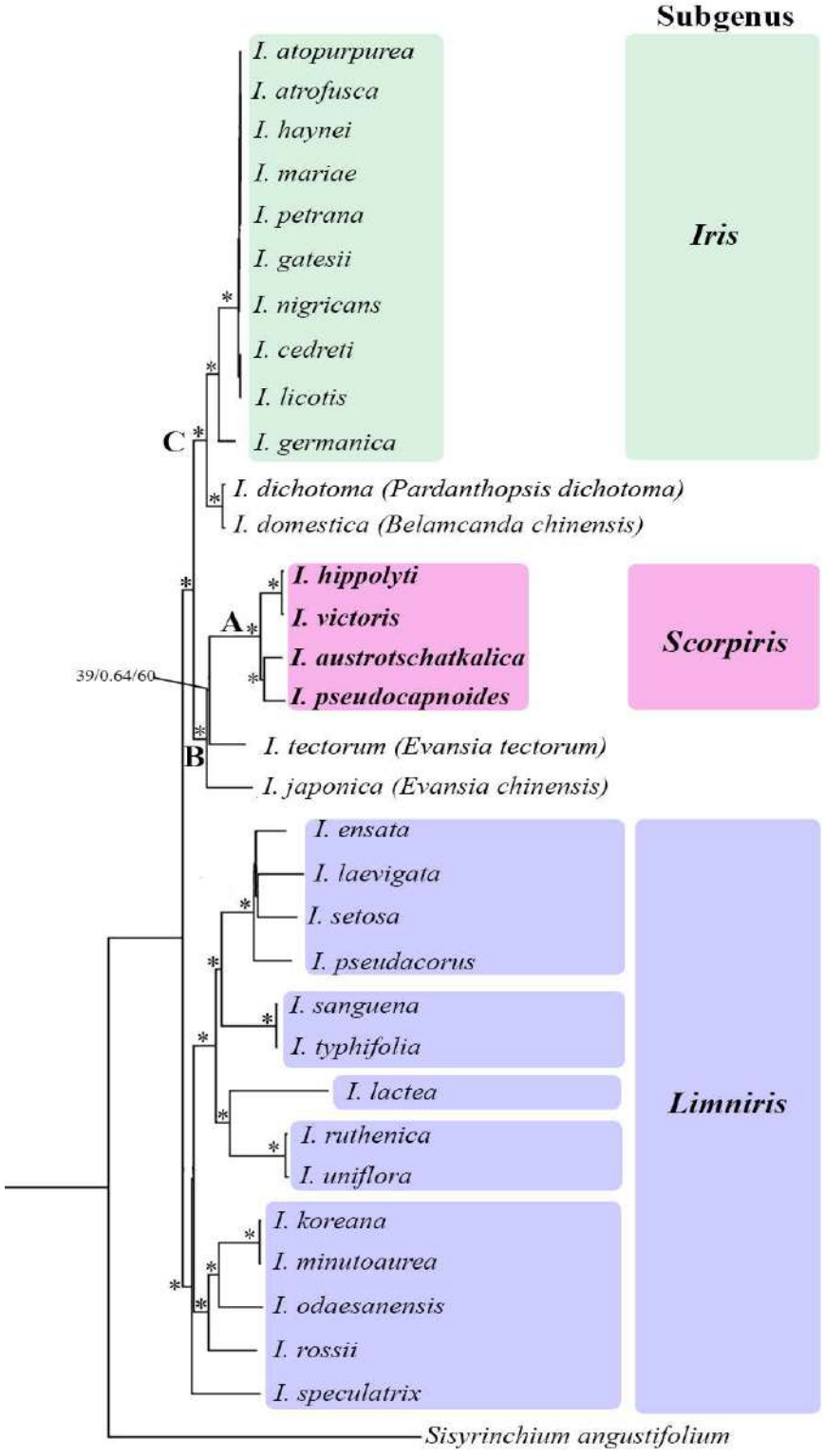

Fig. 7. Phylogenic tree of 32 species of *Iris* produced using maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI). The asterisk indicates that the support of a branch is 100/100/1.0

Discussion

In our phylogenetic tree, the four species of Scorpiris formed a compact clade, which was imbedded into a larger clade whose other members were two rhizomatous species from subg. *Crossiris* (=*Evansia*). A clade sister to this one comprised rhizomatous species from subg. *Iris* (sections *Oncocyclus* and *Iris*), *Belamcanda* and *Pardanthopsis*. The group of *Scorpiris* irises in our study was represented by only four species, therefore any conclusions regarding phylogeny of this group should be treated as preliminary. Nevertheless, our results are consistent with the molecular phylogeny reported by Wilson (2011), Guo and Wilson (2013) and Mavrodiev et al. (2014) based on different plastid fragments and support the assertion of Mavrodiev et al. (2014) and Crespo et al. (2015) that the genus *Iris* is artificial and the current circumscription of the genus requires thorough revision. Although we sequenced only four plastomes and detailed reconstruction of the phylogeny of *Scorpiris* is a matter for future research, it is clear that the *Scorpiris* irises have a long history as a distinct group of species within the currently circumscribed genus *Iris*.

After comparison of the four species of Scorpiris for 83 protein coding regions, 113 IGS regions and 18 introns, we identified 14 loci with high nucleotide variability: *rps16/trnQ-UUG, atpF/atpH, rps3/rpl22, trnK-UUU/rps16, rps15/ycf1, petA/psbJ, /psbA, trnG-GCC/trnfM-CAU, trnR-UCU/atpA, atpA/atpF, rps8/rpl14, trnL-UAA/trnF-GAA, trnG-UCC/trnR-UCU*, and *psaA/ycf3*. The most useful regions for phylogenetic inference are those which can be easily amplified, are approximately 700-1500 bp in size and show high variability (Shaw et al. 2005). In addition, the candidate loci should not have hairpin structures. Presence of loop structures can significantly reduce the robustness of the phylogenetic hypothesis inferred from the data set and deletion of the loop sites is strongly recommended (Quandt et al. 2003; Quandt & Stech 2004). Therefore, usage of 8 loci from the above list for which hairpin structures have been detected (*rps16/trnQ-UUG, rps3/rpl22, trnK-UUU/rps16, petA/psbJ, psbA, trnG-GCC/trnfM-CAU, trnR-UCU/atpA, rps8/rpl14*), should be avoided. Thus, we suggest as the most useful for phylogenetic analyses of Iris the following loci that show high variability and lack of loop structures: *atpF/atpH*, *rps15/ycf1*, *atpA/atpF*, *trnL-UAA/trnF-GAA*, *trnG-UCC/trnR-UCU*, psaA/ycf3. Earlier, two of these loci were recommended for phylogenetic inference based on an analysis of nucleotide diversity in *Iris*: *rps15-ycf1* (Feng et al. 2022) and *trnG-UCC-trnR-UCU* (Zhang et al. 2021).

Acknowledgements

This research was conducted within the state program PFI-5 “The Tree of Life: monocots of Uzbekistan”. We would like to express gratitude to Tashkent Botanical Garden, named after F.N. Rusanov, for support of this research within the framework of the state program of the biotechnology laboratory.

Conflict of Interest Statement

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

References

Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research* 27: 573–580.

Crespo MB, Martinez-Azorin M, Mavrodiev E (2015) Can a rainbow consist of a single colour? A new comprehensive generic arrangement of the ‘*Iris sensu latissimo*’ clade (Iridaceae), congruent with morphology and molecular data. *Phytotaxa* 232: 1–78.

Cui L, Veeraraghavan N, Richter A, Wall K, Jansen RK, Leebens-Mack J, Makalowska I, dePamphilis CW (2006) *ChloroplastDB: The Chloroplast Genome Database. Nucleic Acids Research* 34 (Database issue): D692–D696.

da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008) SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. *International Journal of Plant Genomics* 41: 2696.

Darling AE, Mau B, Perna NT (2010) Progressive Mauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLOS One* 5: e11147.

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Feng JL, Wu LW, Wang Q, Pan YJ, Li BL, Lin YL, Yao H (2022). Comparison analysis based on complete chloroplast genomes and insights into plastid phylogenomic of four *iris* species. *BioMed Research International* 2194021.

Goldblatt P, Manning JC, Rudall P (1998) Iridaceae. In: Kubitzi K (ed.) The families and genera of flowering plants. Springer, Berlin, Heidelberg and New York, pp. 295–333

Guo J, Wilson CA (2013) Molecular phylogeny of crested *Iris* based on five plastid markers (Iridaceae). Systematic Botany 38: 987–995.

Ikinci N, Hall T, Lledo MD, Clarkson JJ, Tillie N, Seisums A, Saito T, Harley M, Chase MW (2011) Molecular phylogenetics of the Juno irises, *Iris* subgenus *Scorpiris* (Iridaceae), based on six plastid markers. *Botanical Journal of the Linnean Society* 167: 281–300.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29: 4633–4642.

Lehwark P, Greiner S (2019) GB2sequin – A file converter preparing custom GenBank files for database submission. *Genomics* 111: 759–761.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41: W575–W581.

Mathew B, (1989) The *Iris*, ed. 2. Timber Press, Portland, OR, USA

Mavrodiev EV, Martínez-Azorín M, Dranishnikov P, Crespo MB (2014) At least 23 genera instead of one: The case of *Iris* L. s.l. (Iridaceae). *PLOS One* 9: e106459.

Patel RK, Jain M (2012) NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PloS ONE* 7: e30619.

Quandt D, Muller K, Huttunen S (2003) Characterisation of the Chloroplast DNA *psbT-H* Region and the Influence of Dyad Symmetrical Elements on Phylogenetic Reconstructions. *Plant Biology* 5: 400–410.

Quandt D, Stech M (2004) Molecular evolution of the *trnT-UGU-trnF-GAA* region in Bryophytes. *Plant Biology* 6: 545–554.

Reeves G, Chase MW, Goldblatt P, Rudall P, Fay MF, Cox AV, Lejeune B, Souza-Chies T (2001) Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany* 88: 2074–2087.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II : Relative utility of 21 noncoding

chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Sennikov A, Khassanov F, Ortikov E, Kurbonaliyeva M, Tojibaev KS (2023) The genus *Iris* L. s. l. (Iridaceae) in the Mountains of Central Asia biodiversity hotspot. *Plant Diversity of Central Asia* 2: 1–104

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular *Biology and Evolution* 30: 2725–2729.

Tillie N, Chase MW, Hall T (2001) Molecular studies in the genus *Iris* L.: a preliminary study. *Annali di Botanica* 1: 105–112.

Wilson CA (2009) Phylogenetic relationships among the recognized series in *Iris* Section *Limniris*. *Systematic Botany* 34: 277–284.

Wilson CA (2004) Phylogeny of *Iris* based on chloroplast *matK* gene and *trnK* intron sequence data. *Molecular Phylogenetics and Evolution* 33: 402–412.

Wilson CA (2011) Subgeneric classification in *Iris* re-examined using chloroplast sequence data. *Taxon* 60: 27–35.

Zhang G, Han Y, Wang H, Wang Z, Hongxing X, Sun M (2021) Phylogeography of *Iris loczyi* (Iridaceae) in Qinghai–Tibet Plateau revealed by chloroplast DNA and microsatellite markers. *AoB Plants* 13: plab070.

Table S1. Comparison of protein coding genes in plastomes of *I. austrotschatkalica* (*I.a.*), *I. pseudocapnoides* (*I.p.*), *I. victoris* (*I.v.*), and *I. hippolyti* (*I.h.*)

Gene	Size (bp)				Aligned length (bp)	Indel (bp)	Number of indel events	Number of polymorphic sites			Nucleotide diversity	*Ks*	*Ka*	*Ka*/*Ks*	Identity (%)	Diversity (%)
	(*I.a.*)	(*I.p.*)	(*I.v.*)	(*I.h.*)				Total	Ts	Tv						
LSC																
*rps*12	369	369	369	369	369	0	0	1	0	1	0.00136	0.00685	0	0	99.7	0.3
*psb*A	1062	1062	1062	1062	1062	0	0	6	6	0	0.00314	0	0.00417	-	99.7	0.3
*mat*K	1566	1569	1566	1566	1569	3	1	19	16	3	0.00724	0.00785	0.00713	0.90871	99.2	0.8
*rps*16	252	252	252	252	252	0	0	2	2	0	0.00397	0	0.005	-	99.6	0.4
*psb*K	186	186	186	186	186	0	0	1	1	0	0.00269	0	0.00365	-	99.7	0.3
*psb*I	110	110	110	110	110	0	0	1	1	0	0.00612	0	0.00807	-	99.4	0.6
*atp*A	1524	1524	1524	1524	1524	0	0	14	11	3	0.00525	0	0.00668	-	99.5	0.5
*atp*F	555	555	555	555	555	0	0	9	7	2	0.00931	0.01082	0.00912	0.84258	99.1	0.9
*atp*H	246	246	246	246	246	0	0	1	1	0	0.00271	0	0.00353	-	99.7	0.3
*atp*L	744	744	744	744	744	0	0	3	2	1	0.00224	0.0042	0.00175	0.41667	99.8	0.2
*rps*2	711	711	711	711	711	0	0	10	9	1	0.0075	0.00338	0.0087	2.57397	99.2	0.8
*rpo*C2	4119	4119	4104	4104	4119	9;6	2	38	33	5	0.00508	0.00468	0.0052	1.11111	99.3	0.7
*rpo*C1	2049	2049	2049	2049	2049	0	0	12	10	2	0.00325	0.0028	0.00338	1.20833	99.7	0.3
*rpo*B	3213	3213	3213	3213	3213	0	0	20	20	0	0.00353	0	0.00445	-	99.6	0.4
*pet*N	90	90	90	90	90	0	0	0	0	0	0	0	0	-	100	0
*psb*M	105	105	105	105	105	0	0	0	0	0	0	0	0	-	100	0
*psb*D	1062	1062	1062	1062	1062	0	0	7	6	1	0.00345	0.002	0.00393	1.96667	99.7	0.3
*psb*C	1422	1422	1422	1422	1422	0	0	7	6	1	0.0027	0	0.00353	-	99.7	0.3
*psb*Z	189	189	189	189	189	0	0	2	2	0	0.00529	0	0.007	-	99.5	0.5
*rps*14	303	303	303	303	303	0	0	2	2	0	0.0033	0	0.0042	-	99.7	0.3
*psa*B	2205	2205	2205	2205	2205	0	0	14	13	1	0.0037	0	0.0048	-	99.6	0.4
*psa*A	2253	2253	2253	2253	2253	0	0	11	11	0	0.00266	0	0.00345	-	99.7	0.3
*ycf*3	510	510	510	510	510	0	0	2	2	0	0.00196	0	0.00255	-	99.8	0.2
*rps*4	606	606	606	606	606	0	0	6	4	2	0.0055	0.0187	0.00147	0.07843	99.4	0.6
*ndh*J	477	477	477	477	477	0	0	1	1	0	0.0014	0	0.0018	-	99.9	0.1
*ndh*K	741	741	741	741	741	0	0	5	3	2	0.00382	0	0.00492	-	99.6	0.4
*ndh*C	363	363	363	363	363	0	0	6	5	1	0.00872	0.00945	0.00875	0.92593	99.1	0.9
*atp*E	405	405	405	405	405	0	0	5	5	0	0.00741	0.0058	0.008	1.37931	99.3	0.7
*atp*B	1497	1497	1497	1497	1497	0	0	11	10	1	0.00401	0	0.00515	-	99.6	0.4
*rbc*L	1440	1440	1440	1440	1440	0	0	14	10	4	0.00486	0.00413	0.0051	1.23487	99.5	0.5
*acc*D	1449	1467	1467	1467	1467	18	1	12	11	1	0.00414	0.00735	0.00325	0.44218	99	1
*psa*I	111	111	111	111	111	0	0	1	1	0	0.00601	0	0.00773	-	99.4	0.6
*ycf*4	555	555	555	555	555	0	0	9	6	3	0.00901	0.00842	0.0093	1.10451	99.1	0.9
*cem*A	690	690	690	690	690	0	0	4	4	0	0.00314	0	0.00398	-	99.7	0.3
*pet*A	963	963	963	963	963	0	0	5	4	1	0.0026	0	0.00325	-	99.7	0.3
*psb*J	123	123	123	123	123	0	0	0	0	0	0	0	0	-	100	0
*psb*L	117	117	117	117	117	0	0	0	0	0	0	0	0	-	100	0
*psb*F	120	120	120	120	120	0	0	0	0	0	0	0	0	-	100	0
*psb*E	252	252	252	252	252	0	0	1	1	0	0.00265	0	0.00353	-	99.7	0.3
*pet*L	96	96	96	96	96	0	0	1	1	0	0.00694	0	0.00953	-	99.3	0.7
*pet*G	114	114	114	114	114	0	0	1	1	0	0.00585	0	0.00753	-	99.4	0.6
*psa*J	129	129	129	129	129	0	0	1	1	0	0.00388	0	0.00515	-	99.6	0.4
*rpl*33	201	201	201	201	201	0	0	2	2	0	0.00498	0.01202	0.00325	0.27038	99.5	0.5
*rps*18	306	306	306	306	306	0	0	4	2	2	0.00708	0.0075	0.00712	0.94889	99.3	0.7
*rpl*20	354	354	354	354	354	0	0	3	2	1	0.00471	0.05763	0.0018	0.03123	99.5	0.5

*clp*P	612	612	612	612	612	0	0	2	1	1	0.00191	0	0.00242	-	99.8	0.2
*psb*B	1527	1527	1527	1527	1527	0	0	13	11	2	0.00502	0.00193	0.00598	3.10017	99.5	0.5
*psb*T	108	108	108	108	108	0	0	0	0	0	0	0	0	-	100	0
*psb*N	132	132	132	132	132	0	0	0	0	0	0	0	0	-	100	0
*psb*H	222	222	222	222	222	0	0	0	0	0	0	0	0	-	100	0
*pet*B	654	654	654	654	654	0	0	8	7	1	0.00663	0.0034	0.00767	2.2549	99.3	0.7
*pet*D	489	489	489	489	489	0	0	5	3	2	0.00375	0	0.00468	-	99.6	0.4
*rpo*A	780	780	780	780	780	0	0	9	7	2	0.00641	0.00323	0.00728	2.2549	99.4	0.6
*rps*11	417	417	417	417	417	0	0	6	6	0	0.00839	0	0.0108	-	99.2	0.8
*rpl*36	114	114	114	114	114	0	0	0	0	0	0	0	0	-	100	0
*inf*A	234	234	234	234	234	0	0	0	0	0	0	0	0	-	100	0
*rps*8	399	399	399	399	399	0	0	3	1	2	0.000418	0.0082	0.0032	0.39024	99.6	0.4
*rpl*14	369	369	369	369	369	0	0	0	0	0	0	0	0	-	100	0
*rpl*16	408	408	408	408	408	0	0	7	5	2	0.01062	0	0.01368	-	98.9	1.1
*rps*3	657	657	657	657	657	0	0	3	3	0	0.00355	0	0.00443	-	99.6	0.4
*rpl*22	309	309	303	303	309	6	1	5	4	1	0.01375	0.01153	0.01182	1.02486	98.6	1.4
total	**43385**	**43406**	**43382**	**43382**	**43406**	**43**	**6**	**335**	**283**	**52**	**0.0040**	**0.0033**	**0.0045**	**1.1122**	**99.6**	**0.4**
IR																
*rps*19	279	279	279	279	279	0	0	1	1	0	0.00179	0	0.0023	-	99.8	0.2
*rpl*2	822	822	822	822	822	0	0	0	0	0	0	0	0	-	100	0
*rpl*23	258	258	258	258	258	0	0	0	0	0	0	0	0	-	100	0
*ycf*2	6879	6879	6870	6870	6912	30;3;3;36	4	21	20	1	0.00175	0.0016	0.0018	1.125	99.1	0.9
*ndh*B	1533	1533	1533	1533	1533	0	0	2	2	0	0.00065	0	0.00085	-	99.9	0.1
*rps*7	468	468	468	468	468	0	0	0	0	0	0	0	0	-	100	0
*rrn*16	1491	1491	1491	1491	1491	0	0	0	0	0	0	0	0	-	100	0
*rrn*23	2811	2811	2811	2811	2811	0	0	0	0	0	0	0	0	-	100	0
*rrn*4.5	103	103	103	103	103	0	0	0	0	0	0	0	0	-	100	0
*rrn*5	121	121	121	121	121	0	0	0	0	0	0	0	0	-	100	0
total	**14765**	**14765**	**14756**	**14756**	**14798**	**72**	**4**	**24**	**23**	**1**	**0.0004**	**0.0002**	**0.0005**	**1.1250**	**99.9**	**0.1**
SSC																
*ndh*F	2214	2214	2214	2214	2214	0	0	26	19	7	0.00647	0.0026	0.00753	2.89744	99.4	0.6
*rpl*32	174	174	174	174	174	0	0	4	3	1	0.01245	0.03487	0.00737	0.21126	98.8	1.2
*ccs*A	966	966	966	966	966	0	0	6	5	1	0.00345	0.0025	0.00373	1.49333	99.7	0.3
*ndh*D	1506	1506	1506	1506	1506	0	0	19	13	6	0.00719	0.003	0.00848	2.82778	99.3	0.7
*psa*C	246	246	246	246	246	0	0	4	4	0	0.00949	0	0.01248	-	99.1	0.9
*ndh*E	306	306	306	306	306	0	0	2	2	0	0.00327	0	0.00422	-	99.7	0.3
*ndh*G	531	531	531	531	531	0	0	6	5	1	0.00659	0.01595	0.00437	0.27377	99.3	0.7
*ndh*I	543	543	543	543	543	0	0	4	3	1	0.0043	0.00987	0.0028	0.28369	99.6	0.4
*ndh*A	1092	1092	1092	1092	1092	0	0	10	10	0	0.00534	0.00502	0.00548	1.0923	99.5	0.5
*ndh*H	1182	1182	1182	1182	1182	0	0	11	10	1	0.00508	0.0051	0.0051	1	99.5	0.5
*rps*15	273	273	273	273	273	0	0	3	2	1	0.00611	0	0.00773	-	99.4	0.6
*ycf*1	5412	5439	5445	5457	5481	9;6;3; 21;6; 9,3;12; 12;3;3	11	118	71	47	0.01175	0.0145	0.01115	0.76897	98	2
total	**14445**	**14472**	**14478**	**14490**	**14514**	**87**	**12**	**213**	**147**	**66**	**0.00679**	**0.00778**	**0.00670**	**1.20539**	**99.2**	**0.72**
TOTAL	**72595**	**72643**	**72616**	**72628**	**72718**	**202**	**21**	**572**	**453**	**119**	**0.00375**	**0.00375**	**0.00389**	**1.14754**	**99.6**	**0.4**

Ts = transitions; Tv = transversions; *Ks* = the number of synonymous substitutions per synonymous site; *Ka* = the number of nonsynonymous substitutions per nonsynonymous site; LSC = large single copy; IR = inverted repeat; SSC = small single copy.

Table S2. Comparison of intergenic space regions among plastomes of *I. austrotschatkalica* (*I.a.*), *I. pseudocapnoides* (*I.p.*), *I. victoris* (*I.v.*), and *I. hippolyti* (*I.h.*)

IGS	Size (bp)				Aligned length (bp)	Indel (bp)	Number of indel events	Number of polymorphic sites			Nucleotide diversity	Identity (%)	Diversity (%)
	(*I.a.*)	(*I.p.*)	(*I.v.*)	(*I.h.*)				Total	Ts	Tv			
LSC													
/*psb*A	77	77	77	77	77	0	0	13	6	7	0.08658	91.3	8.7
*psb*A/*trn*K-UUU	237	237	237	237	237	0	0	6	4	2	0.01336	98.7	1.3
*trn*K-UUU/*mat*K	252	251	252	252	252	1	1	2	1	1	0.00398	99.4	0.6
*mat*K/*trn*K-UUU	775	774	745	743	777	2;1;1;32	4	5	1	4	0.00382	96.7	3.3
*trn*K-UUU/*rps*16	528	651	647	651	662	2;1;11;120;2;1;6;3	8	12	7	5	0.0546	88.3	11.7
*rps*16/*trn*Q-UUG	878	878	672	672	879	1;206;1	3	13	10	3	0,00967	82.9	17.1
*trn*Q-UUG/*psb*K	331	330	329	329	331	1;1	2	7	6	1	0.01114	98.5	1.5
*psb*K/*psb*I	382	386	385	388	388	2;1;4;2;1	5	18	14	4	0.02231	96.7	3.3
*psb*I/*trn*S-GCU	128	128	128	128	128	0	0	4	2	2	0.01693	98.3	1.7
*trn*S-GCU/*trn*G-UCC	856	889	865	866	890	15;7;15;2;9;1	6	22	15	7	0.01229	95.7	4.3
*trn*G-UCC/*trn*R-UCU	136	138	138	139	142	6;4;3	3	5	4	1	0.0202	94.4	5.6
*trn*R-UCU/*atp*A	110	110	110	123	124	1;1;13	3	1	0	1	0.00612	93	7
*atp*A/*atp*F	77	77	70	70	77	7	1	1	0	1	0.00714	93.2	6.8
*atp*F/*atp*H	275	375	373	373	375	1;99;2	3	11	6	5	0.03152	85.2	14.8
*atp*H/*atp*L	479	483	479	481	485	4;2;1;1;1;3;1	7	2	1	1	0.00245	98.7	1.3
*atp*L/*rps*2	251	251	251	251	251	0	0	1	1	0	0.00199	99.8	0.2
*rps*2/*rpo*C2	180	179	179	179	180	1	1	1	1	0	0.00279	99.4	0.6
*rpo*C2/*rpo*C1	178	178	178	178	178	0	0	4	3	1	0.01217	98.8	1.2
*rpo*C1/*rpo*B	26	26	26	26	26	0	0	0	0	0	0	100	0
*rpo*B/*trn*C-GCA	1212	1201	1208	1206	1216	2;3;2;1;1;1;1;3;1;6	8	21	15	7	0.00988	98.2	1.8
*trn*C-GCA/*pet*N	1070	1089	1099	1100	1100	3;3;1;9;1;2;6;6	8	23	20	3	0.01028	97.3	2.7
*pet*N/*psb*M	897	899	896	894	902	1;1;1;4;1;2	6	10	5	5	0.00616	98.8	1.2
*psb*M/*trn*D-GUC	815	843	854	853	857	12;2;2;1;42	5	15	8	7	0.00916	95.8	4.2
*trn*D-GUC/*trn*Y-GUA	382	383	382	382	383	1	1	6	5	1	0.00873	99	1
*trn*Y-GUA/*trn*E-UUC	59	59	59	59	59	0	0	0	0	0	0	100	0
*trn*E-UUC/*trn*T-GGU	759	744	741	732	768	9;25;15;1;1	5	16	12	4	0.01138	96.2	3.8
*trn*T-GGU/*psb*D	1034	1030	1044	1044	1048	1;1;9;7;5;1;1	7	18	14	4	0.00994	97.7	2.3
*psb*D/*psb*C	-	-	-	-	-	-	-	-	-	-	-	-	-
*psb*C/*trn*S-UGA	148	148	147	147	148	1	1	2	1	1	0.01587	98.9	1.1
*trn*S-UGA/*psb*Z	328	328	328	328	328	0	0	8	7	1	0.01372	98.6	1.4
*psb*Z/*trn*G-GCC	298	298	320	304	320	22;16	2	3	3	0	0.00671	95.5	4.5
*trn*G-GCC/*trn*fM-CAU	243	261	241	241	268	7;1;20;6	4	10	2	8	0.01353	92.2	7.8
*trn*fM-CAU/*rps*14	158	158	158	158	158	0	0	2	0	2	0.00633	99.4	0.6
*rps*14/*psa*B	135	135	135	135	135	0	0	1	1	0	0.00494	99.5	0.5
*psa*B/*psa*A	25	25	25	25	25	0	0	0	0	0	0	100	0
*psa*A/*ycf*3	629	633	606	605	637	6;1;2;4;24;4	6	5	3	2	0.00473	95.6	4.4
*ycf*3/*trn*S-GGA	559	559	558	558	559	1	1	12	7	5	0.01165	98.7	1.3
*trn*S-GGA/*rps*4	287	292	292	292	293	6;1	2	4	3	1	0.00758	98.1	1.9
*rps*4/*trn*T-UGU	316	317	317	317	317	1	1	8	7	1	0.01266	98.5	1.5

*trn*T-UGU/*trn*L-UAA	1007	996	1011	1013	1035	12;2;1;11;4;1;7;9;3	9	20	19	1	0.01166	96.1	3.9
*trn*L-UAA/*trn*F-GAA	349	382	387	387	389	2;5;35	3	7	7	0	0.01153	93.3	6.7
*trn*F-GAA/*ndh*J	695	698	695	695	698	3;2;1	3	9	5	4	0.00648	99	1
*ndh*J/*ndh*K	101	101	101	101	101	0	0	0	0	0	0	100	0
*ndh*K/*ndh*C	-	-	-	-	-	-	-	-	-	-	-	-	-
*ndh*C/*trn*V-UAC	133	128	128	128	133	5	1	5	2	3	0	96.2	3.8
*trn*V-UAC/*trn*M-CAU	164	164	164	164	164	0	0	1	1	0	0.00305	99.7	0.3
*trn*M-CAU/*atp*E	193	192	196	194	196	4;3;2	3	0	0	0	0	98.9	1.1
*atp*E/*atp*B	-	-	-	-	-	-	-	-	-	-	-	-	-
*atp*B/*rbc*L	631	631	631	631	631	0	0	5	4	1	0.00423	99.6	0.4
*rbc*L/*acc*D	733	741	734	734	741	5;3;4	3	18	13	5	0.0144	97.7	2.3
*acc*D/*psa*I	378	374	378	378	378	4	1	6	5	1	0.00802	98.7	1.3
*psa*I/*ycf*4	350	354	354	354	354	4	1	8	6	2	0.01238	98.2	1.8
*ycf*4/*cem*A	282	283	283	283	284	1;1	2	4	2	2	0.00768	98.8	1.2
*cem*A/*pet*A	214	223	223	223	223	9	1	0	0	0	0	98	2
*pet*A/*psb*J	707	846	833	833	846	5;1;7;134	4	42	30	12	0.05971	88.6	11.4
*psb*J/*psb*L	134	125	125	125	134	9	1	6	4	2	0.024	97.6	2.4
*psb*L/*psb*F	22	22	22	22	22	0	0	0	0	0	0	100	0
*psb*F/*psb*E	10	10	10	10	10	0	0	0	0	0	0	100	0
*psb*E/*pet*L	1184	1198	1184	1184	1202	1;16;2;1;2;1	6	17	11	6	0.00819	98.3	1.7
*pet*L/*pet*G	176	176	176	176	176	0	0	9	6	3	0.02841	97.2	2.8
*pet*G/*trn*W-CCA	105	104	105	105	105	1	1	1	1	0	0.00641	98.9	1.1
*trn*W-CCA/*trn*P-UGG	165	171	171	171	171	6	1	2	2	0	0.00707	97.6	2.4
*trn*P-UGG/*psa*J	220	219	224	225	225	4;2;1	3	2	1	1	0.00609	97.8	2.2
*psa*J/*rpl*33	461	461	454	463	463	9;1;1	3	7	6	1	0.00922	97.8	2.2
*rpl*33/*rps*18	152	152	153	153	153	1	1	2	2	0	0.00877	98.7	1.3
*rps*18/*rpl*20	265	258	258	258	265	6;1	2	6	4	2	0.01292	97.4	2.6
*rpl*20/*rps*12-2	748	748	748	748	748	0	0	6	4	2	0.00468	99.5	0.5
*rps*12-2/*clp*P	138	138	138	138	138	0	0	2	2	0	0.00725	99.3	0.7
*clp*P/*psb*B	433	434	432	432	434	2;1	2	1	0	1	0.00116	99.6	0.4
*psb*B/*psb*T	166	166	166	166	166	0	0	1	1	0	0.00402	99.6	0.4
*psb*T/*psb*N	75	75	75	75	75	0	0	1	1	0	0.00889	99.1	0.9
*psb*N/*psb*H	111	111	111	111	111	0	0	2	2	0	0.00901	99.1	0.9
*psb*H/*pet*B	122	122	122	122	122	0	0	0	0	0	0	100	0
*pet*B/*pet*D	207	207	208	208	208	1	1	1	0	1	0.00242	99.4	0.6
*pet*D/*rpo*A	174	164	173	174	174	1;10	2	5	4	1	0.01738	98	2
*rpo*A/*rps*11	68	68	68	68	68	0	0	0	0	0	0	100	0
*rps*11/*rpl*36	128	128	128	128	128	0	0	4	3	1	0.01563	98.4	1.6
*rpl*36/*inf*A	110	110	110	110	110	0	0	2	1	1	0.01061	98.9	1.1
*inf*A/*rps*8	118	118	118	118	118	0	0	4	4	0	0.02119	97.9	2.1
*rps*8/*rpl*14	167	187	167	167	187	20	0	4	2	2	0.02108	93.2	6.8
*rpl*14/*rpl*16	142	142	142	142	142	0	0	1	1	0	0.00352	99.6	0.4
*rpl*16/*rps*3	155	155	155	155	155	0	0	5	4	1	0.01935	98.1	1.9
*rps*3/*rpl*22	41	60	66	67	68	1;1;26;6	4	10	2	8	0.0625	86.4	13.6
*rpl*22/*rps*19	114	108	108	108	114	6	1	4	0	4	0.01852	95.5	4.5
total	**27558**	**28070**	**27786**	**27790**	**28345**	**1354**	**163**	**522**	**357**	**166**	**0.011876**	**97.19**	**2.81**

IR														
*rps*19/trnH-GUG	128	128	128	128	128	0	0	0	0	0	0	0	100	0
trnH-GUG/rpl2	43	43	43	43	43	0	0	2	1	1	0.02326	97.7	2.3
rpl2/*rpl*23	18	18	18	18	18	0	0	0	0	0	0	100	0
*rpl*23/*trn*I-CAU	139	139	139	139	139	0	0	0	0	0	0	100	0
*trn*I-CAU/*ycf*2	97	97	97	97	97	0	0	0	0	0	0	100	0
*ycf*2/*trn*L-CAA	330	326	330	330	330	4	1	0	0	0	0	99.4	0.6
*trn*L-CAA/*ndh*B	560	560	560	568	568	8	1	3	1	2	0.00298	99	1
*ndh*B/*rps*7	128	128	128	128	128	0	0	0	0	0	0	100	0
*rps*7/*rps*12-2	58	58	58	58	0	0	0	0	0	0	0	100	0
*rps*12-2/*trn*V-GAC	1794	1811	1806	1806	1811	5;12	2	5	4	1	0.00149	99.4	0.6
trnV-GAC/*rrn*16	226	226	226	226	226	0	0	1	1	0	0.00221	99.8	0.2
*rrn*16/*trn*I-GAU	304	304	304	304	304	0	0	0	0	0	0	100	0
*trn*I-GAU/*trn*A-UGC	69	69	69	69	69	0	0	0	0	0	0	100	0
*trn*A-UGC/*rrn*23	145	145	145	145	145	0	0	0	0	0	0	100	0
*rrn*23/*rrn*4.5	98	99	98	98	99	1	1	0	0	0	0	99.5	0.5
*rrn*4.5/*rrn*5	212	222	222	222	222	10	1	1	1	0	0.00314	97.4	2.6
*rrn*5/*trn*R-ACG	255	255	255	255	255	0	0	2	1	1	0.00392	99.6	0.4
*trn*R-ACG/*trn*N-GUU	309	292	300	300	309	8;9	2	1	0	1	0.00171	97	3
*trn*N-GUU/*ycf*1	325	325	325	325	325	0	0	1	0	1	0.00154	99.8	0.2
total	**5238**	**5245**	**5251**	**5259**	**5216**	**57**	**8**	**16**	**9**	**7**	**0.00211**	**99.4**	**0.6**
SSC														
*ndh*F/*rpl*32	888	899	888	893	915	1;5;5;5;5;2,1;12,5;9,4	11	20	9	11	0.00978	96.2	3.8
*rpl*32/*trn*L-UAG	851	745	829	828	854	82;1;18,13;9;1;2,1	8	30	15	15	0.01507	96.1	3.9
*trn*L-UAG/*ccs*A	84	84	84	84	84	0	0	1	1	0	0.00595	99.4	0.6
*ccs*A/*ndh*D	204	204	211	211	211	7	1	3	3	0	0.00735	97.1	2.9
*ndh*D/*psa*C	123	123	119	119	123	4	1	1	1	0	0.0042	97.4	2.6
*psa*C/*ndh*E	325	331	334	334	334	1;6;2	3	5	1	4	0.00872	97.7	2.3
*ndh*E/*ndh*G	193	193	193	193	193	0	0	0	0	0	0	100	0
*ndh*G/*ndh*I	265	266	266	266	266	1	1	3	3	0	0.00629	99.2	0.8
*ndh*I/*ndh*A	85	85	85	85	85	0	0	1	0	1	0.00588	99.4	0.6
*ndh*A/*ndh*H	1	1	1	1	1	0	0	0	0	0	0	100	0
*ndh*H/*rps*15	115	115	115	115	115	0	0	3	0	3	0.01449	98.6	1.4
*rps*15/*ycf*1	444	490	479	466	495	2;25,13,4;1;4;5;2;2,1;36	11	29	9	20	0.04088	88.7	11.3
total	**3578**	**3536**	**3604**	**3595**	**3676**	**297**	**36**	**96**	**42**	**54**	**0.009884**	**97.5**	**2.5**
TOTAL	**36374**	**36851**	**36641**	**36644**	**37237**	**1708**	**207**	**634**	**408**	**227**	**0.007959**	**98.0**	**2.0**

Ts = transitions; Tv = transversions; LSC = large single copy; IR = inverted repeat; SSC = small single copy; "-" = no integenic sequence between *psb*D and *psb*C genes.

Table S3. Comparison of introns among plastomes of *I. austrotschatkalica* (*I.a.*), *I. pseudocapnoides* (*I.p.*), *I. victoris* (*I.v.*), and *I. hippolyti* (*I.h.*)

Region	Intron	Size (bp)				Aligned length (bp)	Indel (bp)	Number of indel events	Number of polymorphic sites			Nucleotide diversity	Identity (%)	Diversity (%)
		(*I.a.*)	(*I.p.*)	(*I.v.*)	(*I.h.*)				Total	Ts	Tv			
LSC														
	trnK-UUU	2593	2594	2563	2561	2598	1;3;2,1;1;32	6	28	18	10	0.00593	98.5	1.5
	*rps*16	856	856	850	850	860	3;1;4;6	4	11	8	3	0.00749	98.2	1.8
	*trn*G-UCC	673	673	669	669	674	5;1	2	8	6	2	0.00624	98.8	1.2
	*atp*F	764	763	765	765	767	2;2;1	3	10	3	7	0.00722	98.9	1.1
	*rpo*C1	730	731	718	718	734	1;1;2;13,3	5	3	0	3	0	98.2	1.8
	*ycf*3	727	728	730	730	733	1;1;2;5	4	6	4	2	0.0046	98.7	1.3
	*ycf*3	712	711	711	711	712	1	1	8	6	2	0.00609	99.3	0.7
	*trn*L-UAA	399	398	399	399	399	1	1	3	3	0	0.00461	99.4	0.6
	*trn*V-UAC	601	602	601	601	602	1	1	5	4	1	0.00471	99.4	0.6
	*clp*P	659	666	662	667	668	8,4;1;1;2,1	6	15	9	6	0.01016	97.7	2.3
	*clp*P	797	797	789	789	798	2,1;2,1;5	5	1	1	0	0.00253	99.2	0.8
	*pet*B	746	746	737	737	747	1;4;6	3	14	14	0	0.01019	98	2
	*pet*D	754	756	750	750	757	1;5;2	3	9	6	3	0.00712	98.6	1.4
	*rpl*16	869	877	878	878	884	5;1;3;12,6	5	10	2	8	0.00618	97.9	2.1
total		**11880**	**11898**	**11822**	**11825**	**11933**	**171**	**49**	**131**	**84**	**47**	0.0059	98.6	1.4
IR														
	*rpl*2	664	664	664	664	664	0	0	0	0	0	0	100	0
	*ndh*B	699	699	695	695	699	4	1	1	1	0	0.00096	99.5	0.5
	*rps*12-2	541	541	541	541	541	0	0	0	0	0	0	100	0
	*trn*I-GAU	946	946	945	945	946	1	1	0	0	0	0	99.9	0.1
	*trn*A-UGC	808	809	809	809	809	1	1	2	2	0	0.00124	99.8	0.2
total		**3658**	**3659**	**3654**	**3654**	**3659**	**6**	**3**	**3**	**3**	**0**	0.00044	99.84	0.16
SSC														
	*ndh*A	1055	1064	1055	1055	1065	2;1;8	3	15	11	4	0.00822	98.7	1.3
total		**1055**	**1064**	**1055**	**1055**	**1065**	**11**	**3**	**15**	**11**	**4**	**0.00822**	**98.7**	**1.3**
TOTAL		**16593**	**16621**	**16531**	**16534**	**16657**	**188**	**55**	**149**	**98**	**51**	**0.00486**	**99.1**	**0.9**

Ts = transitions; Tv = transversions; LSC = large single copy; IR = inverted repeat; SSC = small single copy.