INSIGHT INTO CHLOROPLAST GENOME STRUCTURAL VARIATION OF ENDANGERED SPECIES OF THE GENUS *TULIPA* L. IN KAZAKHSTAN

D. Tussipkan[1], M.B. Ramazanova[1], S.A. Manabayeva[1,2]

[1]National Center for Biotechnology
Kazakhstan, 010000, Astana, Qorghalzhyn Hwy. 13/5
[2]L.N.Gumilyov Eurasian National University
Kazakhstan, 010000, Astana, Satpayev St. 2
e-mail: manabayeva@biocenter.kz

The *Liliaceae* family containes about 250 genera and 3500 species distributed all over the world. Among the genera of this family, the genus *Tulipa*. L belongs to the subfamily *Lilioideae* and tribe *Lilieae*, the center of diversity of the genus is in the Pamir and Hindu Kush mountains and the steppes of Kazakhstan. About 150 species of the genus *Tulipa* grow and develop well in the geographical regions of Asia, Europe, and North Africa. In Central Asia the genus *Tulipa* is represented by 63 species. A total of 34 species belonging to the three subgenera *Tulipa*, *Eriostemones*, and *Orythia* have been described in the Kazakhstan. Eighteen species are listed in the Red Book and are protected by the state.

The present study is the first report on typical characteristics of the complete chloroplast genome sequences of *T. alberti, T. kaufmanniana, T. greigii,* and *T. dubia* from Kazakhstan. These species are listed in the Red Book and are protected at the state level.

Comparative analyses revealed that the complete genomes of four species showed high conservation in various parameters, including whole genome size (152,006 bp - 152,382 bp), total GC content (36.58-36.62), gene number (130-131), and intron length (540 bp- 2620 bp) of specific genes. Despite this conservation, there was remarkable nucleotide diversity among the complete genomes (Pi=0,54058). We identified sources of inversion variation in the chloroplast structure of the SSC and LSC regions, leading to the identification of four chloroplast structure haplotypes (haplotype A and B in SSC region, haplotype I and haplotype II in the LSC region). In addition, we found a total of 159 SSR loci in *T. alberti*, 158 in *T. kaufmanniana*, 174 in *T. greigii,* and 163 in *T. dubia*, with nine genes (*petB*, *ndhA*, *rpl2*, *ycf2*, *clpP1*, *ycf1*, *atpF*, *rpoC1*, and *pafI*) showing effective polymorphic SSRs ($Pi \geq 0,0431$) suitable as molecular markers for identification *Tulipa* species. Codon usage analysis revealed conserved preferences within the genus *Tulipa*. *T. albertii* and *T. dubia* species showed only one overrepresented codon, whereas seven codons were found to be significantly overrepresented ($RSCU>1.6$) in *T. greigii* and *T. kaufmanianna*. Species-specific trends were also observed, with *T. albertii* and *T. dubia* showing fewer underrepresented codons ($RSCU<0.6$) than *T. greigii* and *T. kaufmanianna*. Phylogenetic analysis was performed on the whole chloroplast genome sequences of four species of the present study together with 15 *Tulipa* species from NCBI. Based on the Tajima's neutrality test (), all species were clustered into three major groups, with *T. dubia* having the closest relationship with *T. gesneriana* (ON041137) with a pairwise distance of 0,0019. The chloroplast genomes of *T. alberti*, *T. kaufmanniana*, and *T. greigii* from Kazakhstan formed a distinct group, indicating increased nucleotide variation within species based on ecological region. Overall, the complete chloroplast genome sequences of these four *Tulipa* species provide valuable information for future studies and molecular marker design.