

To Cite: Yılmaz A,2021. The Evaluations and Comparisons of Nuclear and Chloroplast DNA Regions Based on Species Identification and Phylogenetic Relationships of *Crocus* L. Taxa. Journal of the Institute of Science and Technology, 11(2): 1504-1518.

The Evaluations and Comparisons of Nuclear and Chloroplast DNA Regions Based on Species Identification and Phylogenetic Relationships of *Crocus* L. Taxa

Aykut YILMAZ^{1*}

ABSTRACT: The genus *Crocus* L. (Iridaceae) consists of about 200 species distributed from western Europe and north west Africa to western China. The genus is taxonomically very problematic and day after day it becomes more difficult to understand the relationships between *Crocus* species because of the identification of new species. Recently, the studies based on DNA sequences are frequently used for the solution of taxonomic problems, the determination of species diversity and especially the evaluation of phylogenetic relationships among the studied taxa. The use of the regions belonging to nuclear DNA and especially chloroplast genome and the combinations of these regions are frequently recommended in many studies. However, there is no information about suitable barcoding regions that can be used for all plant groups. This study focuses on determination of the most successful barcoding regions, understanding the species discrimination abilities and finally evaluation of phylogenetic relationships among the *Crocus* taxa using one region from nuclear DNA (ITS1-5.8S rRNA gene-ITS2) and four regions from chloroplast DNA (1. trnK-matK IGS, 2. partial psbA gene-psbA/trnH IGS-partial trnH gene, 3. rpoC1 gene, 4. partial trnL gene-trnL/trnF IGS-partial trnF gene). Sequence informations for related DNA regions were obtained from NCBI. As a result, it can be stated that the region which is only covering ITS1-5.8S rRNA gene-ITS2 separated the all taxa in species level and it is more beneficial for the phylogenetically evaluation of *Crocus* taxa. Furthermore, dendrogram show taxonomic problems belonging to some species in the genus such as *C. biflorus* taxa.

Keywords: *Crocus*, nuclear DNA, chloroplast DNA, *C. biflorus*

¹ Aykut YILMAZ ([Orcid ID: 0000-0002-0327-8388](https://orcid.org/0000-0002-0327-8388)), Uşak Üniversitesi, Fen Edebiyat Fakültesi, Moleküler Biyoloji ve Genetik Bölümü, Uşak, Türkiye

*Corresponding Author: Aykut YILMAZ, e-mail: aykut.yilmaz@usak.edu.tr

INTRODUCTION

The genus *Crocus* L. is a member of the family Iridaceae which is presented by about 92 genera and 2000 species (Candan, 2015; Addam et al., 2019). The genus *Crocus* containing popular ornamentals and Saffron provided from *C. sativus* which is known as the most expensive spices by weight of the world is represented by about 200 species ranging from western Europe and north west Africa to western China with the center of species diversity in the Balkan Peninsula and Turkey (Mathew, 1982; Harpke et al., 2016; Saxena, 2016; Roma-Marzio et al., 2018). The genus was divided into two subgenera based on Mathew's (1982) classification, the monotypic subgenus *Crociris* that is containing only *C. banaticus* and the subgenus *Crocus* divided into section *Crocus* and section *Nudiscapus*. After that each section is again divided into totally 15 series (section *Crocus*: six series and section: *Nudiscapus*: nine series).

The genus *Crocus* is taxonomically very problematic and very hard to understand the relationships among the species. There are several reasons of this.

Firstly, new species are continuously determined, however these new taxa especially within the section *Nudiscapus* are not assigned to any group within the section (Kerndorff et al., 2016).

Secondly, there are species exhibiting intermediate or variable characters in the some populations containing two closely related species because of introgression from one species to another as a result of hybridization between them and repeated backcrossing (Harrison and Larson, 2014; Kerndorff et al., 2016).

Thirdly, the species identification and classification based on morphological data can be reasons systematic problems and misclassifications due to the variations in morphological characters caused by environmental factors (Jonah et al., 2011; Erol et al., 2014). Many molecular studies revealed the reality of this situation and changed the taxonomical situations of some species on the genus *Crocus* (Petersen et al., 2008; Harpke et al., 2014, 2015).

Another situation is the changes frequently observed in the chromosome number within the genus (from $2n=6$ to $2n=70$) (Brighton et al., 1973; Uslu et al., 2012; Harpke et al., 2013). Even these changes in the chromosome number are observed in the same species belonging to different localities. Uslu et al. (2012) states that the endemic taxon for Turkey, chromosome number of *C. abantensis* was found $2n = 8$, but previously chromosome number of this taxon was reported as $2n = 16$ by Baytop et al. (1975) and Mathew (1984). Similarly, although chromosome number of *C. ancyrensis* was reported as $2n = 10$ by Karasawa (1950) and Mathew (1984), chromosome number was found as $2n = 6$ by Uslu et al. (2012). All of these makes the difficult to understand the genus taxonomically and makes species identification doubtful.

Finally, phylogenetic analyse results based on Harpke et al. (2013) show that subspecies of *C. biflorus* occur in several series and anymore subspecies status can not be maintained (Harpke et al. 2016).

Recently, DNA based studies are frequently used in the solution of all these taxonomic problems and species identification (Petersen et al., 2008; Erol et al., 2011; Harpke et al., 2013; Mosolygo-Lukacs et al., 2016). Especially, DNA barcoding studies based on short DNA sequence information has gained momentum with advance of sequence analysis techniques. DNA barcoding is frequently used for the solution of stated taxonomic problems, determination of species diversity and evaluation of phylogenetic relationships. For this aim, the regions belonging to genomic DNA and plastid DNA that contain sufficient sequence variation are used to understand the phylogenetic relationships among the taxa.

Internal transcribed spacers (ITS) between rDNA genes in genomic DNA are frequently preferred by plant systematists (Sramko, 2008; Harpke et al., 2013; Sramko et al., 2014).

Especially, the regions and the region combinations belonging to chloroplast genome such as matK, trnK, rbcL, trnH-psbA, trnL-trnF and atpB-rbcL are used to solve problems and understand the phylogenetic relationships (Harpke et al., 2013, 2014). Nevertheless, there is still no information about the suitable barcoding region that can be used in all plant groups.

Consequently, the use of different regions belonging to genomic DNA and especially chloroplast DNA in various plant groups is very important to understand the species discrimination ability and to evaluate of phylogenetic relationships among the species. For this aim, in this study, the region which is covering ITS1-5.8S rRNA gene-ITS2 belonging to genomic DNA and the regions which are covering trnK-matK intergenic spacer (IGS), partial psbA gene-psbA/trnH intergenic spacer (IGS)-partial trnH gene, rpoC1 gene and partial trnL gene-trnL/trnF intergenic spacer (IGS)-partial trnF gene belonging to chloroplast DNA were used to compare the species discrimination abilities with each other and finally to evaluate the phylogenetic relationships of the *Crocus* taxa.

MATERIALS AND METHODS

Sequence informations related to studied taxa were obtained from National Centre of Biotechnology Information (NCBI). One region from the nuclear genome (ITS1-5.8S rRNA gene-ITS2) and four regions from chloroplast DNA (1. trnK-matK IGS, 2. partial psbA gene-psbA/trnH IGS-partial trnH gene, 3. rpoC1 gene, 4. partial trnL gene-trnL/trnF IGS-partial trnF gene) were preferred.

The sequence informations of 51 taxa for the ITS1-5.8S rRNA gene-ITS2 region belonging to nuclear DNA were provided. Similarly, the sequence informations of four regions belonging to chloroplast DNA (18 taxa for trnK-matK IGS, 36 taxa for partial psbA gene-psbA/trnH IGS-partial trnH gene, 109 taxa for rpoC1 gene, 99 taxa for partial trnL gene-trnL/trnF IGS-partial trnF gene) were provided separately. After the sequence informations of the determined taxa were provided for each studied region, these sequences belonging to determined taxa for each studied regions were aligned. In other words, multiple sequence alignments were separately performed by using Molecular Evolutionary Genetics Analysis (MEGA).

Then, different analysis using alignment sequence informations such as the probabilities of substitution from one base to another base, transition/transversion ratios for purines-pyrimidines and overall, nucleotide frequencies for each studied region were performed.

DNA regions used in this study, taxon numbers for each examined DNA regions, alignment lengths (bp), variable sites showing the sequence differences among taxa, transitional and transversional substitutions (%), transition/transversion rates for purines-pyrimidines-overall (%) and nucleotide frequencies (%) for each examined DNA region giving the informations about sequences of taxa were computed and showed in Table 6.

Finally, Neighbour-joining dendrograms that bootstrap values are reported on branches were obtained to evaluate the phylogenetic relations among taxa and to compare the studied DNA regions according to species identification ability. For this aim, totally five dendrogram that consist of four regions from chloroplast DNA and one region from nuclear DNA were created using MEGA X.

All positions containing gaps and missing data for each studied barcoding regions were eliminated with complete deletion option of program in the determination of variable sites and in other analysis such as substitutions from one base to another base, transition/transversion ratios for purines-pyrimidines, and dendrograms.

RESULTS AND DISCUSSION

Analysis Results for The Region Which is Covering ITS1-5.8S rRNA gene-ITS2

All taxa provided sequence informations from NCBI were preferred from Turkey' location (Harpke, 2012; Harpke et al., 2013; Erol et al., 2015). Totally, sequence datas of 51 taxa containing two spacer regions (ITS1 and ITS2) giving useful information in species level and one rRNA gene (5.8S rRNA) between spacer regions were used for analysis. Firstly; sequence alignments were performed by using Molecular Evolutionary Genetics Analysis (MEGA X) program. The sites with missing/ambiguous data and gaps in the determination of alignment length and variable sites among studied *Crocus* taxa were excluded for effective analyses. After these regions were excluded, alignment length for studied taxa was established as 526 bp. Besides, this DNA region has 118 variable sites for 51 taxa.

The probability of substitution from one base to another base were computed and shown in Table 1. This analysis involved nucleotide sequences belonging to 51 *Crocus* taxa and similarly, all positions containing gaps and missing data were eliminated with the complete deletion option in the program. In this table, transitional substitutions as bold and transversional substitutions as italics was shown. The rate of transitional substitutions with 72.39 % is higher than the rate of transversional substitutions. The highest base substitutions are observed in the rate of 35.04 % from T to C and 19.38 % from C to T, respectively (Table 1).

In addition to the rate of base substitutions, transition/transversion rate for purines and pirimidines was created (Table 6). According to this table, transition/transversion rate of pyrimidines (k_2) with 8.04 is the higher than purines (k_1) having 2.55 value. Moreover, overall transition/transversion rate (R) is 2.39 ($R = [A * G * k_1 + T * C * k_2] / [(A + G) * (T + C)]$).

As the another analysis, the nucleotide frequencies for *Crocus* taxa are analysed. The nucleotide frequencies are 19.02 % (A), 17.45 % (T/U), 31.55 % (C), and 31.98 % (G). As a result, it is observed that the percentage of C and G bases (63.53 %) for related *Crocus* taxa of the region which is covering ITS1-5.8S rRNA gene-ITS2 is higher than the percentage of A and T/U bases.

Table 1. The probability of substitution (r) from one base (row) to another base (column) for ITS1-5.8S rRNA gene-ITS2 region.

	A	T	C	G
A	-	<i>2.41</i>	<i>4.36</i>	11.27
T	<i>2.63</i>	-	35.04	<i>4.42</i>
C	<i>2.63</i>	19.38	-	<i>4.42</i>
G	6.7	<i>2.41</i>	<i>4.36</i>	-

Finally, Neighbor-Joining (NJ) dendrogram was drawn to understand the species discrimination ability for studied region and evaluate the phylogenetic relationships of *Crocus* taxa (Figure 1). Furthermore, dendrograms that was separately drawn for each regions were used in the comparison of the DNA sequences belonging to four region of cpDNA and one region of nuclear DNA and in determination the importance in *Crocus* taxa. The evolutionary distances in phylogenetic tree were

computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and also branch lengths in phylogenetic tree infer the evolutionary distances.

It can be stated as a result of the examination of NJ tree provided from ITS1-5.8S rRNA gene-ITS2 that the sequence informations of the DNA region interested for *Crocus* taxa are sufficient for the species separation and identification. Furthermore, it is observed that NJ tree identified all taxa in distinct branch as compatible with their subspecies and separated the *Crocus* taxa to two main groups except for *C. carpetanus* showing the most distinction among the studied taxa according to NJ dendrogram (Figure 1).

Analysis Results for The Region Which is Covering trnK-matK Intergenic Spacer (IGS)

The sequence informations of 18 *Crocus* taxa were provided from NCBI (Harpke, 2018). DNA sequences belonging to 18 taxa were aligned by using MEGA X program. The missing/ambiguous data and gaps in the sequences for studied taxa were excluded for effective analyses. Alignment length of *Crocus* taxa was established as 599 bp and 10 variable site were determined among taxa.

Table 2 show the probability of substitution from one base to another base. Transitional and transversional substitutions in the table are shown as bold and as italics, respectively.

While the rate of transitional substitutions is 35.48 %, the rate of transversional substitutions is 64.52 % (Table 2, 6). In other words, transversional substitutions are higher than the transitional substitutions on the contrary of the region which is covering ITS1-5.8S rRNA gene-ITS2. The highest base substitutions are observed between C and T bases (Table 2).

Table 6 show the transition/transversion rate for purines and pyrimidines belonging to trnK-matK IGS region. Transition/transversion rate for purines, pyrimidines and overall is 0.39, 1.73 and 0.48, respectively.

The nucleotide frequencies for studied *Crocus* taxa are 35.11 % (A), 35.78 % (T/U), 16.92 % (C), and 12.19 % (G). As a result of this, it can be stated that the percentage of A and T/U bases with 70.89 % is higher than G and C bases (29.11 %) on the contrary of the region which is covering ITS1-5.8S rRNA gene-ITS2 (Table 6).

Finally, Neighbor-Joining (NJ) dendrogram was drawn for:

Firstly to determine the species discrimination ability for *Crocus* taxa of the trnK-matK IGS region belonging to cpDNA,

Secondly to detect the best region for the evaluation of phylogenetic relations among *Crocus* taxa (Figure 2).

NJ tree provided from the trnK-matK IGS show that related DNA region has insufficient sequence information with only 10 variable sites for species identification and phylogenetic relations, although some species were taxonomically separated from each other.

Table 2. The probability of substitution (r) from one base (row) to another base (column) for trnK-matK intergenic spacer (IGS) region.

	A	T	C	G
A	-	<i>11.54</i>	<i>5.46</i>	1.55
T	<i>11.33</i>	-	9.46	<i>3.93</i>
C	<i>11.33</i>	20	-	<i>3.93</i>
G	4.47	<i>11.54</i>	<i>5.46</i>	-

Analysis Results for The Region Which is Covering Partial psbA gene-psbA/trnH IGS-Partial trnH gene

The sequence informations belonging to 36 *Crocus* taxa for the related DNA region were obtained from NCBI (Petersen et al., 2008; Schori et al., 2018). DNA sequences obtained from NCBI were aligned by MEGA X program. DNA sequences with the missing/ambiguous data and gaps for 36 taxa were excluded for other analyses. After these regions were excluded, alignment length and variable sites for studied taxa were determined as 529 bp and 30 site, respectively (Table 6).

When the probability of substitution (r) from one base to another base is examined for the regions showing variation, the rate of transitional substitutions with 33.03 % appears to be lower than transversional substitutions (Table 3, 6). In other words, the most of the variations in the DNA sequences for 36 taxa studied are caused by base substitutions between the different base groups belonging to purine (A and G) and pyrimidine (C and T) (Table 3). The base substitutions from C to T is observed in highest rate with 11.26 % (Table 3).

The transition/transversion rates are 0.95 for purines and 1.01 for pyrimidines (Table 6). The overall transition/transversion rates is 0.45.

The nucleotide frequencies for the region which is covering partial psbA gene-psbA/trnH IGS-partial trnH gene are 31.18 % (A), 33.14 % (T/U), 16.40 % (C) and 19.29 % (G). The nucleotide frequencies for A and T/U is higher with 64.32 % than the percentage of G and C (Table 6).

Finally, Neighbor-Joining (NJ) dendrogram to show phylogenetic relations among *Crocus* taxa and to determine the importance for DNA barcoding of partial psbA gene-psbA/trnH IGS-partial trnH gene region was provided using the Maximum Composite Likelihood method (Tamura et al., 2004). The sequence informations for 30 sites showing variation among taxa studied were used for the comparisons of taxa. NJ dendrogram show to us that sequence variations for barcoding region related are not enough for identification of all species although it seperated some *Crocus* taxa from each other (Figure 3).

Table 3. The probability of substitution (r) from one base (row) to another base (column) for partial psbA gene-psbA/trnH intergenic spacer (IGS)-partial trnH gene region.

	A	T	C	G
A	-	<i>11.1</i>	<i>5.49</i>	6.19
T	<i>10.44</i>	-	5.57	<i>6.46</i>
C	<i>10.44</i>	11.26	-	<i>6.46</i>
G	10.01	<i>11.1</i>	<i>5.49</i>	-

Analysis Results for The Region Which is Covering rpoC1 gene

Sequence informations of 109 *Crocus* taxa for rpoC1 gene belonging to cpDNA are obtained from NCBI (Petersen et al., 2008). Sequence informations of the *Crocus* taxa provided from NCBI for the region which is covering rpoC1 gene are used and analysed. DNA sequences of all taxa were aligned using MEGA X program and the sequences containing the missing/ambiguous data and gaps for 109 *Crocus* taxa were excluded for effective analysis. After the exclusion of missing datas, alignment length was determined as 503 bp and 21 variable sites were observed among the *Crocus* taxa studied (Table 6).

The rates of nucleotide substitutions are shown in Table 4 as transitional substitutions and transversional substitutions in bold and italics, separately. It can be stated that transitional substitutions with the rate of 85.01 % is very high in comparison to transversional substitutions (Table 4, 6). The

highest base substitutions are observed between G-A and than A-G with 36.14 % and 31.13 %, respectively (Table 4). Transversional base substitutions have quite low rate with 14.99 % of total base substitutions (Table 4, 6).

In addition to transitional and transversional nucleotide substitutions, transition/transversion rates for purines (k_1), pyrimidines (k_2) and overall (R) was showed for rpoC1 genes belonging to 109 *Crocus* taxa (Table 6). The transition/transversion rate is 17.43 for purines, 4.88 for pyrimidines and 5.66 for overall.

The nucleotide frequencies for the rpoC1 gene are 27.65 % for A, 30.79 % for T/U, 17.73 % C and 23.82 % for G. It can be stated that the percentage of A and T/U bases for the rpoC1 genes belonging to 109 *Crocus* taxa studied is higher with 58.44% than the percentage of G and C bases (Table 6).

Despite NJ dendrogram separated the *Crocus* taxa to groups, it is observed that many *Crocus* taxa were not separated from each other in species level owing to insufficient variation in DNA sequences for rpoC1 gene (Figure 4). The sequence variations in only 21 nucleotide for 109 taxon were determined. In other words, it can be stated that the DNA sequence of rpoC1 gene for *Crocus* species is highly preserved and however the existing nucleotide changes in these region are meaningful to evaluate the phylogenetic relationships among *Crocus* species.

Table 4. The probability of substitution (r) from one base (row) to another base (column) for rpoC1 gene region.

	A	T	C	G
A	-	2.31	1.33	31.13
T	2.07	-	6.48	1.79
C	2.07	11.26	-	1.79
G	36.14	2.31	1.33	-

Analysis Results for The Region Which is Covering Partial trnL gene-trnL/trnF IGS-Partial trnF gene

DNA sequence informations of the region which is covering partial trnL gene-trnL/trnF IGS-partial trnF gene for 99 *Crocus* taxa were obtained from NCBI (Harpke, 2012; Harpke et al., 2013; Harpke, 2014; Harpke et al., 2015). DNA sequences for trnL/trnF regions belonging to 99 *Crocus* taxa were aligned using MEGA X program and then the missing/ambiguous data and gaps within region examined were excluded to provide more accurate information in analysis. Alignment length was determined as 507 bp after the missing/ambiguous data and gaps were excluded. Furthermore 56 variable sites were observed for trnL/trnF regions belonging to *Crocus* taxa studied. The most variable site among the all studied DNA regions was determined in the region which is covering partial trnL gene-trnL/trnF IGS-partial trnF gene after the ITS1-5.8S rRNA gene-ITS2 which contain 118 variable nucleotide (Table 6). Table 5 shows the probability of nucleotide substitution for trnL-trnF region. Transitional substitutions which is showed as bold are higher with the rate of 61.13 % than transversional substitutions (Table 5, 6). Moreover, the highest nucleotide substitutions occur from G to A with 24.97 % and then from C to T with 15.19 %, respectively (Table 5). The transition/transversion rates are 3.45 for purines (k_1), 2.76 for pyrimidines (k_2) and 1.45 for overall (R) (Table 6). The nucleotide frequencies for the DNA region related of 99 *Crocus* are 37.19 % (A), 28.27 % (T/U), 16.53 % (C) and 18 % (G). It

The Evaluations and Comparisons of Nuclear and Chloroplast DNA Regions Based on Species Identification and Phylogenetic Relationships of *Crocus* L. Taxa

can be stated that A and T/U nucleotide frequencies for the DNA region containing trnL gene-trnL/trnF IGS-partial trnF gene is much higher with 65.46 % than G and C nucleotides (Table 6). NJ dendrogram for the separation of species according to evolutionary distances and for the grouping the taxa examined according to phylogenetic relations was prepared (Figure 5). Although the sequence variations observed in 56 nucleotide for trnL gene-trnL/trnF IGS-partial trnF gene are secondly the highest among the regions studied, it is insufficient to identify for many species belonging to the genus *Crocus*, besides some species separated in dendrogram. In other words, NJ dendrogram shows that species separation ability of the DNA region studied for *Crocus* species is low although it has the highest variation rate among the cpDNA regions studied (Figure 5, Table 6).

Table 5. The probability of substitution (r) from one base (row) to another base (column) for partial trnL gene-trnL/trnF IGS-partial trnF gene region.

	A	T	C	G
A	-	5.5	3.21	12.09
T	7.23	-	8.88	3.5
C	7.23	15.19	-	3.5
G	24.97	5.5	3.21	-

Table 6. The comparisons of all studied DNA barcoding regions.

DNA regions	Taxon (number)	Alignment length (bp)	Variable site	Transitional substitutions (%)	Transversional substitutions (%)	Transition/Transversion rate			Nucleotide freq. (%)	
						Purines	Pyrimidines	Overall	A+T/U	G+C
						(k ₁)	(k ₂)	(R)		
ITS1-5.8SrRNA-ITS2	51	526	118	72.39	27.61	2.55	8.04	2.39	36.47	63.53
trnK-matK IGS	18	599	10	35.48	64.52	0.39	1.73	0.48	70.89	29.11
psbA-psbA/trnH IGS-trnH	36	529	30	33.03	66.97	0.95	1.01	0.45	64.32	35.68
rpoC1	109	503	21	85.01	14.99	17.43	4.88	5.66	58.44	41.56
trnL-trnL/trnF IGS- trnF	99	507	56	61.13	38.87	3.45	2.76	1.45	65.46	34.54
Total	313									

The fundamental aim of this study is to determine the species identification and separation ability of different DNA regions belonging to nuclear and chloroplast genome, to find out the most successful barcoding region and to most accurately classify the species belonging to the genus *Crocus* to contribute to the solution of taxonomic problems. For this aim, four regions belonging to chloroplast genome and ITS1-5.8SrRNA-ITS2 region of nuclear genome were preferred as barcoding regions and examined. Totally 313 taxa from the genus *Crocus* for five barcoding regions were analysed in this study (Table 6).

Alignment lengths of barcoding regions used were ranged between 503 bp and 599 bp. As a result of comparison of the barcoding regions according to alignment length, it can be stated that DNA lengths of regions preferred are compatible with each other and convenient for barcoding.

When variable sites that is being the most important character in species identification and phylogenetic relationships for each barcoding regions are examined, it is observed that the region which is covering ITS1-5.8SrRNA-ITS2 has the highest variable site with 118 nucleotide. Thereafter, the barcoding regions containing the highest variable site are trnL-trnL/trnF IGS-trnF with 56 nucleotide, psbA-psbA/trnH IGS-trnH with 30 nucleotide, rpoC1 with 21 nucleotide and trnK-matK IGS with 10 nucleotide, respectively (Table 6). Especially, it is observed that the barcoding regions belonging to

chloroplast genome show low variations in DNA sequences. In other words, it can be stated that sequence information of cp DNA regions used for barcoding are highly conserved. However, although the ITS sequences between rRNA genes in nuclear DNA show more variation and seem to be more successful in taxonomically evaluation of species in comparison with other regions (Alvarez and Wendel, 2003; Hürkan, 2017), it has disadvantage due to being hundreds or thousands of ITS copies in a typical plant genome and the use of a single locus in barcoding (Coleman, 2003).

Although the number of variable sites expressing the nucleotide substitutions among the taxa studied is very important, these sequence substitutions must also be meaningful in identification and phylogenetic relationships of species. trnL-trnL/trnF IGS-trnF has the highest variation rate with 56 nucleotide among the cpDNA regions examined but it has been determined that this region are insufficient to separate and phylogenetically evaluate the most of taxa studied. The main reason for this is that the existing variations do not represent all taxa and therefore cannot reveal the differences between taxa.

When the substitutions from one base to another base which is the most important reason of sequence variations are examined for each barcoding regions, it can be stated that the most of the variations are caused by transitional base substitutions except barcoding regions which are covering trnK-matK IGS and psbA-psbA/trnH IGS-trnH. Furthermore overall transition/transversion rates for the most regions studied are higher than one.

When the nucleotide types in the DNA sequences of the each barcoding regions studied are examined, it is observed that the percentages of A+T/U nucleotides are higher than G+C nucleotides for all cpDNA regions studied except ITS1-5.8SrRNA-ITS2 region belonging to nuclear DNA. Denk and Grimm (2010) states in their study on the oaks that guanine-rich region within the 5' ITS1 region can be difficult to sequence and moreover, gene bank sequences may include missing data because of this region. For this reason, Denk and Grimm recommend to re-run the sequencing to guarantee at least one completely sequenced ITS clone per individual.

In the comparisons according to NJ dendrogram of the DNA regions studied; only the region which is covering ITS1-5.8SrRNA-ITS2 belonging to nuclear DNA separated the all taxa in species level unlike four regions of cpDNA. *C. carpetanus* from series *Carpetani* of section *Nudiscapus* showed the highest diversity according to phylogenetic tree. Similarly, *C. carpetanus* and *C. nevadensis* from series *Carpetani* of section *Nudiscapus* were grouped in the separate branch from other species in the study related to combined sequences of the ITS and trnL-F regions by Harpke et al. (2013). Moreover, Petersen et al. (2008) in the study derived from the analysis of five plastid regions showed that *C. carpetanus* and *C. nevadensis* have the highest diversity in the comparison to the other *Crocus* species. After the *C. carpetanus*, NJ dendrogram separated the *Crocus* taxa into two main groups (Figure 1). I. main group was separated to two groups containing the taxa belonging to section *Nudiscapus* and section *Crocus* (Figure 1). In other words, it can be stated that NJ dendrogram separated the species from each other and it grouped the taxa as sectional. However, new species are constantly determined in the genus *Crocus* and these new species are not assigned to any series within the section (Erol et al., 2015). This is the problematic situation for the evaluation of taxonomic relationships within genus, in addition to taxa containing intermediate or variable characters caused by introgression and backcrossing. II. main group in dendrogram belonging to Figure 1 show the reality of this situation. For example, subspecies of *C. biflorus* belonging to section *Nudiscapus* occur in distinct clade on the dendrogram. This result supports the articles stated that subspecies status for *C. biflorus* is incorrect (Petersen et al., 2008, Harpke et al.,

The Evaluations and Comparisons of Nuclear and Chloroplast DNA Regions Based on Species Identification and Phylogenetic Relationships of *Crocus* L. Taxa

2013). In other words, the term “subspecies” which was brought into the genus taxonomy by Mathew (1982) can not be maintained any more (Harpke et al., 2016).

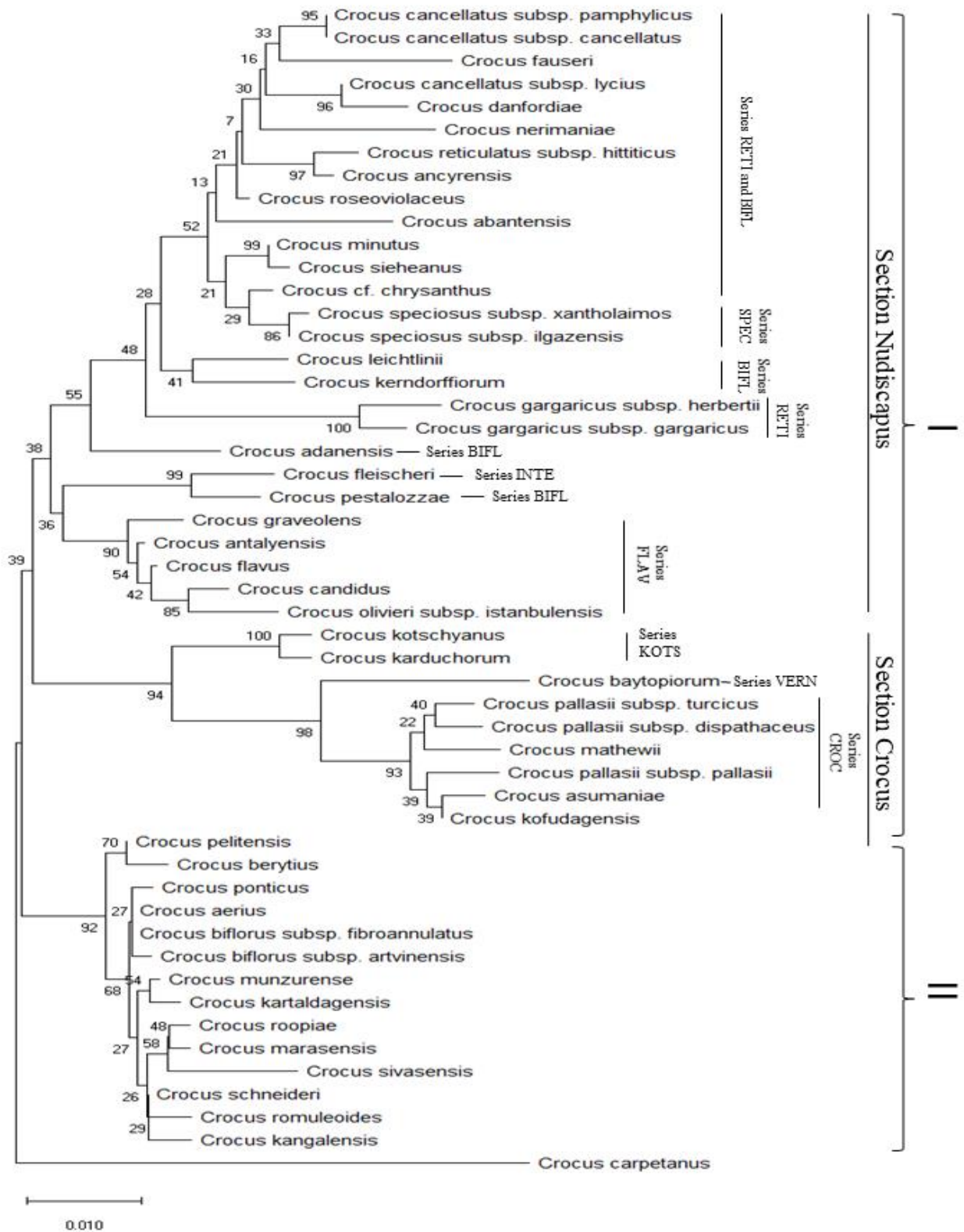


Figure 1. Neighbor-Joining dendrogram provided from ITS1-5.8S rRNA gene-ITS2 region for 51 *Crocus* taxa.

In addition to the separation of taxa as sectional according to NJ dendrogram provided from ITS1-5.8S rRNA gene-ITS2 region, it was observed that some taxa were grouped in different branch based on series. Especially, the taxa belonging to series *Flavi* from section *Nudiscapus* were grouped in same branch (Figure 1). Similar situation was observed in the taxa belonging to series *Crocus* from section *Crocus* (Figure 1).

In this study, four plastid DNA regions were examined with the aim of understanding the species discrimination ability and comparing of species. As a result of the using these barcoding regions, it can be stated that although cpDNA regions separated some species from others and they grouped the taxonomically close species in same or close branch, these regions examined are not enough to separate and evaluate phylogenetically the all studied species. For example, although trnL-trnL/trnF IGS- trnF region among the cpDNA regions had the highest variable site with 56 nucleotide and was examined with 99 taxa, it was observed that this region was not enough in separation of the most of species studied. Similarly, it can be said that psbA-psbA/trnH IGS-trnH region with 30 variable site and rpoC1 region with 21 variable site - 109 taxa examined are not suitable for the species identification alone. It is recommended in many studies that barcoding regions used together and region combinations can provide more effective data for phylogenetic analysis (Chase et al., 2007; Filiz and Koç, 2012). However, studies of Petersen et al. (2008) and Seberg and Petersen (2009) showed that region combinations belonging to plastid DNA are not enough for the separation and identification of *Crocus* species.

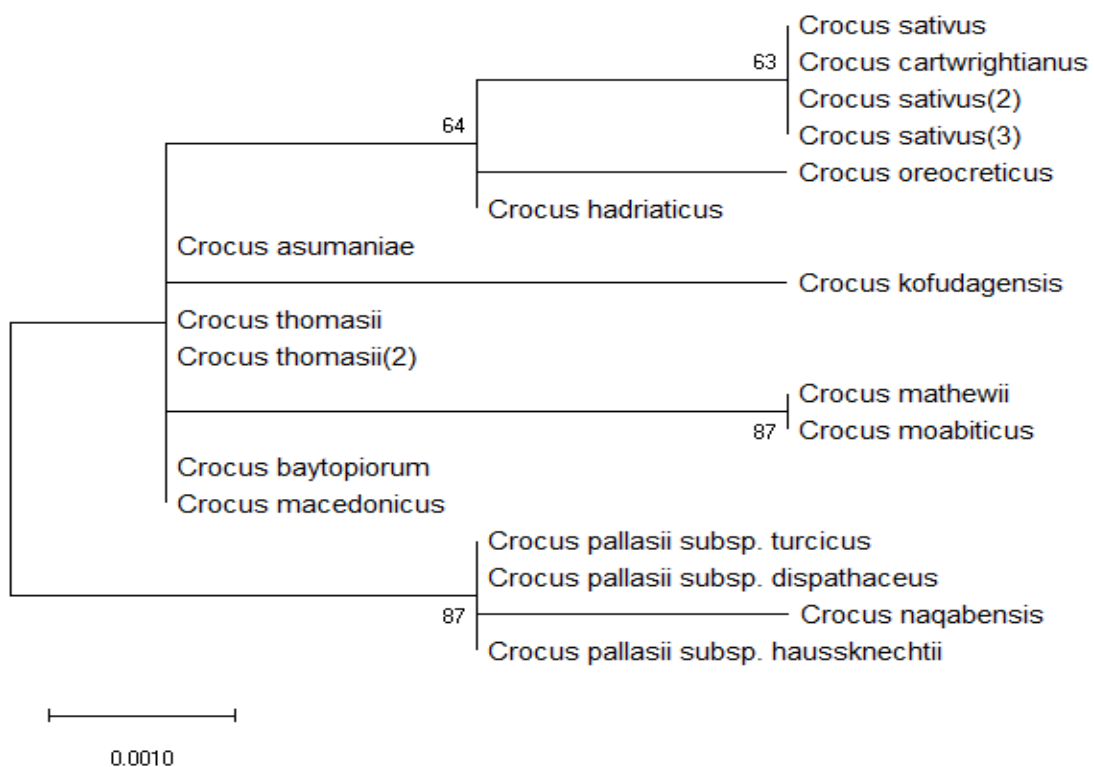


Figure 2. Neighbor-Joining dendrogram provided from trnK-matK intergenic spacer (IGS) region for 18 *Crocus* taxa.



Figure 5. Neighbor-Joining dendrogram provided from partial trnL gene-trnL/trnF IGS-partial trnF gene region for 99 *Crocus* taxa.

CONCLUSION

As a result, it can be said that this study shows the regions which contribute more in the evaluation of *Crocus* species. In addition, it is recommended as a result of the data obtained from this study and the previous studies in this field that the using of the whole chloroplast genome could provide enormous data in the solution of taxonomic problems and evaluation of phylogenetic relations.

ACKNOWLEDGMENTS

The author would like to thank to NCBI for sequence informations used in this study and the authors sharing the sequence informations in NCBI.

APPENDIX (DNA regions and their gene bank accession numbers)

ITS1-5.8S rRNA gene-ITS2; LN864707, HE664013, HE663993, HE801128, HE801129, LN864719, LN864718, LN864717, LN864716, LN864715, LN864714, LN864713, LN864712, LN864711, LN864710, LN864708, LN864706, HE664019, HE664015, HE664011, HE664010, HE664009, HE664005, LS398415, LS398378, HE664000, HE801124, HE801120, HE663995, HE663992, HE663988, HE663987, HE663984, HE663983, HE663981, HE663979, HE663977, HE663974, HE801159, HE801157, HE801154, HE801153, HE801151, HE801138, HE801141, HE801133, LS398367, LS398391, HE801071, LS398388, LS398370

trnK-matK intergenic spacer (IGS); LS398342, LS398306, LS398319, LS398316, LS398298, LS398295, LS398339, LS398336, LS398334, LS398335, LS398333, LS398326, LS398323, LS398321, LS398318, LS398315, LS398312, LS398304

partial psbA gene-psbA/trnH intergenic spacer (IGS)-partial trnH gene; EU257491, MG947147, MG947146, MG947145, MG947144, MG947143, MG947142, MG947141, MG947077, KY645945, EU110163, EU110148, EU110140, EU110134, EU110115, EU110107, EU110235, EU110234, EU110232, EU110229, EU110227, EU110106, EU110109, EU110116, EU110122, EU110172, EU110168, EU110162, EU110160, EU110153, EU110194, EU110191, EU110181, EU110218, EU110211, EU110201

rpoC1; EU110533, EU110523, EU110573, EU110558, EU110525, EU110517, EU110645, EU110643, EU110642, EU110641, EU110640, EU110639, EU110638, EU110636, EU110635, EU110634, EU110633, EU110632, EU110631, EU110630, EU110629, EU110628, EU110627, EU110626, EU110625, EU110624, EU110623, EU110622, EU110621, EU110620, EU110619, EU110618, EU110617, EU110616, EU110615, EU110614, EU110613, EU110611, EU110610, EU110609, EU110608, EU110607, EU110606, EU110604, EU110603, EU110602, EU110601, EU110600, EU110599, EU110598, EU110597, EU110596, EU110592, EU110591, EU110590, EU110589, EU110588, EU110587, EU110586, EU110585, EU110584, EU110583, EU110582, EU110581, EU110580, EU110579, EU110578, EU110577, EU110576, EU110575, EU110572, EU110571, EU110570, EU110569, EU110568, EU110567, EU110566, EU110565, EU110564, EU110563, EU110562, EU110559, EU110557, EU110556, EU110555, EU110554, EU110553, EU110552, EU110551, EU110549, EU110548, EU110545, EU110543, EU110542, EU110541, EU110540, EU110538, EU110537, EU110536, EU110534, EU110532, EU110531, EU110528, EU110526, EU110521, EU110520, EU110519, EU110516, EU257489

partial trnL gene-trnL/trnF IGS-partial trnF gene; LM993604, LM993597, LT991689, LT991667, LT991664, LT991654, LT991646, LT991643, HE864211, HE864207, HE864229, LM993638, LM993633, LM993589, HG518229, HG518218, HG518217, HG518213, HG518211, HG518209, HG518207, HG518205, HG518201, HG518200, HE864281, HE864280, HE864278, HE864277, HE864276, HE864274, HE864273, HE864272, HE864271, HE864270, HE864268, HE864267, HE864266, HE864265, HE864264, HE864263, HE864261, HE864259, HE864256, HE864255, HE864254, HE864253, HE864252, HE864250, HE864249, HE864248, HE864247, HE864246, HE864244, HE864243, HE864242, HE864241, HE864239, HE864238, HE864236, HE864235, HE864234, HE864233, HE864231, HE864230, HE864228, HE864224, HE864223, HE864222, HE864219, HE864218, HE864216, HE864215, HE864213, HE864209, HE864205, HE864204, HE864203, HE864201, HE864199, HE864194, HE864189, HE864188, HE864186, HE864184, HE864181, HE864179, HE864175, HE864171, HE864170, HE864169, HE864168, HE864167, HE864166, LT991688, LT991683, LT991673, LT991671, LT991669, LT991665

Conflict of Interest

I declare that there is no conflict of interest during the planning, execution and writing of the article.

Author's Contributions

I hereby declare that the planning, execution and writing of the article was done by me as the sole author of the article.

REFERENCES

- Addam K, Bou-Hamdan M, Sabbagh N, Takkoush J, Hout K, 2019. *Crocus baalbekensis* K. Addam and M. Bou Hamdan sp. Nov and its three forms (IRIDACEAE), new endemic species and forms from Lebanon, joined the Lebanese Flora. *MOJ Eco Environ Sci.*, 4(2): 75-83.
- Alvarez I, Wendel JF, 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, 29: 417-434.
- Baytop T, Mathew B, Brighton C, 1975. Four New Taxa in Turkish *Crocus* (Iridaceae). *Kew Bulletin*, 30(2): 241-246.
- Brighton CA, Mathew B, Marchant CJ, 1973. Chromosome counts in the genus *Crocus* (Iridaceae). *Kew Bulletin*, 28: 451-464.
- Candan F, 2015. Morphological and Leaf Anatomical Investigations on 2 Yellow Flowered Endemic Taxa of *Crocus* L. (*Crocus ancyrensis*, *Crocus sieheanus*) from Turkey. *International Journal of Agriculture, Forestry and Fisheries*, 3(3): 93-98.
- Chase MW, Cowan RS, Hollingsworth PM et al., 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon*, 56(2): 295-299.
- Coleman AW, 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.*, 19(7): 370-375.
- Denk, T. and Grimm, G.W. 2010. The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers. *Taxon*. 59(2): 351-366.
- Erol O, Şık L, Kaya HB, Tanyolaç B, Kuçuker O, 2011. Genetic diversity of *Crocus antalyensis* B. Mathew (Iridaceae) and a new subspecies from southern Anatolia. *Plant Syst Evol.*, 294: 281-287.
- Erol O, Kaya HB, Şık L, Tuna M, Can L, Tanyolaç MB, 2014. The genus *Crocus*, series *Crocus* (Iridaceae) in Turkey and 2 East Aegean islands: a genetic approach. *Turk J Biol.*, 38: 48-62.
- Erol O, Harpke D, Yıldırım H, 2015. A new *Crocus* L.(Iridaceae) species from SE Turkey, based on morphological and molecular data. *Phytotaxa*, 239(3): 223-232.
- Fılız E, Koç I, 2012. Bitkilerde DNA Barkodları. *AKÜ FEBİD*, 12: 53-57.
- Harpke D, 2012. *Taxonomy and Evolutionary Biology*, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, D-06466 Gatersleben/Germany.
- Harpke D, Meng S, Rutten T, Kerndorff H, Blattner FR, 2013. Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: ancient hybridization and chromosome number evolution. *Molecular Phylogenetics and Evolution*, 66(3): 617-627.
- Harpke D, 2014. *Taxonomy and Evolutionary Biology*, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, D-06466 Gatersleben/Germany.

- Harpke D, Peruzzi L, Kerndorff H, Karamplianis T, Constantinidis T, Randelović V, Randelović N, Jušković M, Pasche E, Blattner FR, 2014. Phylogeny, geographic distribution and new taxonomic circumscription of the *Crocus reticulatus* species group (Iridaceae). Turkish Journal of Botany, 38: 1182-1198.
- Harpke D, Carta A, Tomović G, Randelović V, Randelović N, Blattner FR, Peruzzi L, 2015. Phylogeny, karyotype evolution and taxonomy of *Crocus* series *Verni* (Iridaceae). Plant Systematics and Evolution, 301: 309-325.
- Harpke D, Kerndorff H, Pasche E, Peruzzi L, 2016. Neotypification of the name *Crocus biflorus* Mill. (Iridaceae) and its consequences in the taxonomy of the genus. Phytotaxa, 260(2): 131-143.
- Harpke D, 2018. IPK-Gatersleben, Taxonomy and Evolutionary Biology, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, D-06466 Gatersleben/Germany.
- Harrison RG, Larson EL, 2014. Hybridization, Introgression, and the Nature of Species Boundaries. Journal of Heredity, 105: 795-809.
- Hurkan K, 2017. Karasal bitkilerde DNA barkodlama: Bazı DNA barkod bölgelerinin incelenmesi. The International Journal of Innovative Approaches in Science Research, vol. 1, no. 1.
- Jonah PM, Bello LL, Lucky O, Mıdau A, Moruppa SM, 2011. Review: the importance of molecular markers in plant breeding programmes. Global Journal of Science Frontier Research, 11: 0975-5896.
- Karasawa K, 1950. Note on the Cytology of *Crocus*. Genetica, 25(1): 188-192.
- Kerndorff H, Pasche E, Harpke D, 2016. The Genus *Crocus* (Liliiflorae, Iridaceae): Taxonomical Problems and How to Determine a Species Nowadays? STAPFIA, 105: 42-50.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K, 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution, 35: 1547-1549.
- Mathew B, 1982. The *Crocus*: a revision of the genus *Crocus*. BT Batsford Ltd. London 127 pp.
- Mathew B, 1984. *Crocus* L. in Davis PH, Flora of Turkey and The East Aegean Islands. Edinburgh University Press, Edinburgh, 8: 413-438.
- Mosolygo-Lukacs A, Sramko G, Barabas S, Czeglédi L, Javor A, Molnar VA, Suranyi G, 2016. Molecular genetic evidence for allotetraploid hybrid speciation in the genus *Crocus* L. (Iridaceae). Phytotaxa, 258(2): 121-136.
- NCBI, National Centre of Biotechnology Information, <https://www.ncbi.nlm.nih.gov/genbank>.
- Petersen G, Seberg O, Thorsøe S, Jørgensen T, Mathew B, 2008. A phylogeny of the genus *Crocus* (Iridaceae) based on sequence data from five plastid regions. Taxon, 57(2): 487-499.
- Roma-Marzio F, Harpke D, Peruzzi L, 2018. Rediscovery of *Crocus biflorus* var. *estriatus* (Iridaceae) and its taxonomic characterisation. Italian Botanist, 6: 23-30.
- Saxena RB, 2016. Botany, Taxonomy and Cytology of *Crocus* Series *Intertexti*. Indo. American Journal of Pharmaceutical Sciences, 3(4): 321-323.
- Schori M, Emch K, Kıtko AR, Shinwari ZK, Showalter AM, 2018. National Germplasm Resources Laboratory, Agricultural Research Service, USDA, 10300 Baltimore Ave., Beltsville, MD 20705 USA.
- Seberg O, Petersen G, 2009. How many loci does it take to DNA barcode a *Crocus*? PLoS One, 4(2): E4598.
- Sramko G, 2008. Sequence variability of the nrITS in the *Ophrys fuciflora* species-complex of the Mediterranean bee-orchid (*Ophrys* L.) genus. Department of Botany. University of Debrecen, Debrecen.
- Sramko G, Attila MV, Hawkins JA, Bateman RM, 2014. Molecular phylogeny and evolutionary history of the Eurasiatic orchid genus *Himantoglossum* s.l. (Orchidaceae). Annals of Botany, 114(8): 1609-1626.
- Tamura K, Nei M, Kumar S, 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101: 11030-11035.
- Uslu E, Babaç MT, Yılmaz A (2012). Karyological Studies On Some *Crocus* L. Taxa From Turkey. Caryologia: Caryologia: International Journal of Cytology, Cytosystematics and Cytogenetics, 65(1): 7-10.