

Amino Acids Sequence Based In Silico Analysis of RuBisCO (Ribulose-1, 5 Bisphosphate Carboxylase Oxygenase) Proteins in Some *Carthamus L. ssp.*

Emre SEVINDIK

Adnan Menderes University, Faculty of Agriculture, Department of Agricultural Biotechnology, South Campus, Cakmar, Aydin, Turkey; ph.d-emre@hotmail.com

Abstract

RuBisCO is an important enzyme for plants to photosynthesize and balance carbon dioxide in the atmosphere. This study aimed to perform sequence, physicochemical, phylogenetic and 3D (three-dimensional) comparative analyses of RuBisCO proteins in the *Carthamus* spp. using various bioinformatics tools. The sequence lengths of the RuBisCO proteins were between 166 and 477 amino acids. Their molecular weights (Mw) ranged from 18,711.47 to 52,843.09 Da; the most acidic and basic protein sequences were detected in *C. tinctorius* ($pI = 5.99$) and in *C. tenuis* ($pI = 6.92$), respectively. The extinction coefficients of RuBisCO proteins at 280 nm ranged from 17,670 to 69,830 $M^{-1} cm^{-1}$, the instability index (II) values for RuBisCO proteins ranged from 33.31 to 39.39, while the GRAVY values of RuBisCO proteins ranged from -0.313 to -0.250. The most abundant amino acid in the RuBisCO protein was Gly (9.7%), while the least amino acid ratio was Trp (1.6 %). The putative phosphorylation sites of RuBisCO proteins were determined by NetPhos 2.0. Phylogenetic analysis revealed that RuBisCO proteins formed two main clades. A RAMPAGE analysis revealed that 96.3%-97.6% of residues were located in the favoured region of RuBisCO proteins. To predict the three dimensional (3D) structure of the RuBisCO proteins PyMOL was used. The results of the current study provide insights into fundamental characteristic of RuBisCO proteins in *Carthamus* spp.

Keywords: *Carthamus*, RuBisCO, in silico analysis, physicochemical analysis, phylogenetic

Introduction

Photosynthesis is arguably the most important energy conversion process on earth because the chemical energy it yields is the base of food chains that sustain the overwhelming majority of other life forms. Plants utilize atmospheric CO₂ to liberate oxygen and synthesize carbohydrates during photosynthesis. It is an event where radiant energy of sunlight is utilized to convert carbon dioxide into photosynthetic by products (Naeem *et al.*, 2013). Ribulose-1,5-bisphosphate carboxylase/oxygenase (EC4.1.1.39, RuBisCO), transforming the carbon dioxide and ribulose-1,5-bisphosphate (RuBP) into two molecular 3-phosphoglyceric acid, catalyzes the first reaction of carbon dioxide fixation in photosynthetic dark reaction. In addition, RuBisCO catalyses the reaction of oxygen and RuBP to phosphoglyceric acid and phosphoglycolic acid, which is the first reaction of photorespiration. Therefore, RuBisCO is the key enzyme deciding the photosynthetic efficiency by regulating photosynthesis and photorespiration. Based on the dissimilarity of the primary and quaternary structures, the RuBisCOs can be

partitioned into three types: I form exists in higher plants and most prokaryotes, consisting of eight large subunits (50~60 kD) and eight small subunits (12~18 kD), presenting square symmetry structure (L8S8) (Andersson *et al.*, 1989); II form was discovered in purple non-sulfur photosynthetic bacteria, and composed of only two large subunits (L₂); III form was dug out in *Thermococcus kodakarensis* lately by Kitano (Kitano *et al.*, 2001), likewise formed with only large subunits, and no small subunit, appearing structure of (L₂)₅ (Zhang *et al.*, 2011). The large subunits of RuBisCO (rbcL) are encoded in the chloroplast, whereas the small subunits (rbcS) are encoded in the nucleus. Although both subunits are important for the functionality of the protein, the transcript levels of rbcS have been considered as a factor determining the level of RuBisCO in plants (Mukherjee *et al.*, 2015). The rbcS promoter has therefore been a popular genetic element in molecular studies (Avci and Tezcan, 2016).

The Asteraceae family is considered to be the largest known plant family in the World (Nylander and Anderberg, 2015) and comprises approximately 23,000 species in 1,535

genera (Öztürk and Çetin, 2013). The family contains food plants, raw material resources, medical and medicinal plants, tender and succulent plants, wild weeds and poisonous plant (Süslü *et al.*, 2010). Acquisition of esculents such as honey and acquisition of cooking oil from this family is used in many fields such as pharmaceutical industry. In addition, many of the species are cultivated as ornamental plants in Asteraceae (Paksoy *et al.*, 2016). *Carthamus* L. is a genus belonging to the tribe Cynareae of the Asteraceae family. The eastern part of the Mediterranean region is regarded as the original centre of this genus. The genus includes about 25 species, distributed from Spain and North Africa across the Middle East to northern India (Ashri, 1960; Yue *et al.*, 2013). The oil of *Carthamus tinctorius* plant from *Carthamus* species is found suitable for biodiesel manufacturing and used as an industrial good. Carthamin matter obtained from safflowers is of importance as a raw material for natural dye (Nagaraj *et al.*, 2001). In addition, the plant itself is a valuable ornamental plant used as green fence and dry flower. Its residue is a valuable feedstuff and handles are used as fuel. Furthermore, it is an oil plant, with valuable properties which are rapidly increasing in the world, due to its high concordance with arid regions and high quality oil obtained from the seeds of the plant. (Yildirim *et al.*, 2005; Uysal *et al.*, 2006). In this study, RuBisCO was analysed in economically valuable *Carthamus* species with respect to physicochemical, phylogenetic and 3D structure properties utilizing bioinformatic tools.

Materials and Methods

The RuBisCO protein sequences of *Carthamus* species were retrieved in FASTA format from the National Centre for Biotechnology Information (NCBI: <https://www.ncbi.nlm.nih.gov/protein>). The NCBI and UNIPROT accession numbers of the selected proteins are given in Table 1. The physicochemical analysis and amino acid contents of the proteins were analysed by ExPASy's ProtParam (<http://web.expasy.org/protparam/>) to determined their isoelectric point (*pI*), molecular weight (Mw), total number of positive (+R), and negative (-R) residues, extinction coefficient (EC), instability index (II), aliphatic index (AI), and GRAVY (grand average of hydrophathy) values. The putative phosphorylation sites of the RuBisCO proteins were determined by NetPhos 2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>). All the protein sequences were aligned using ClustalW

(Thompson, 1994) and MEGA 6.0 (Tamura *et al.*, 2013). A phylogenetic tree of *Carthamus* species RuBisCO proteins were constructed using maximum likelihood method with MEGA 6.0 and bootstrap values were performed with 1000 replicates. To predict the 3D structure of the RuBisCO proteins, homology models were performed using the following method options in PSIPRED v3.3 (<http://bioinf.cs.ucl.ac.uk/psipred/>): (Predict Secondary Structure), BioSerf v2.0 (Automated Homology Modelling by Homology).

The results were checked and verified by a Ramachandran plot analysis in RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>), which determined the best predicted models. Finally, 3D comparative analyses were performed using PyMOL (TM) (Schrodinger, LLC).

Results

The physicochemical analysis of the predicted RuBisCO protein was performed using ExPASy ProtParam and results were shown in Table 1. The amino acid sequence length ranged from 166 to 477. The shortest amino acid sequence was in *C. tenuis* (166 aa.), while the longest amino acid sequence was in *C. tinctorius* (477 aa.) species. The minimum and maximum molecular weights (Mw) were 18711.47 and 52843.09 Da, respectively. The most acidic and basic protein sequences were in *C. tinctorius* (*pI* = 5.99) and *C. tenuis* (*pI* = 6.92), respectively. The extinction coefficients of RuBisCO proteins at 280 nm ranged from 17,670 to 69,830 M⁻¹ cm⁻¹. The higher extinction coefficients were detected in *C. duvauxii*, *C. oxyacanthus* and *C. tinctorius* while the lowest extinction coefficient was in *C. tenuis*. The instability index (II) values for the RuBisCO proteins ranged from 33.31 to 39.34. The aliphatic index (AI) of proteins of thermophilic bacteria was found to be higher and the index could be used as a measure of thermostability of proteins. This index is directly related to the mole fraction of Ala, Ile, Leu and Val in the protein (Idicula-Thomas and Balaji, 2005). The AI values ranged from 77.31 to 79.17. The highest AI values were seen in *C. duvauxii* and in *C. oxyacanthus*, while the lowest AI value was seen *C. turkestanicus* (Table 1). As expected, when the total ratios of aliphatic amino acid contents were compared in the proteins with the lowest and highest AI values, it was observed that AI values increased when aliphatic amino acid

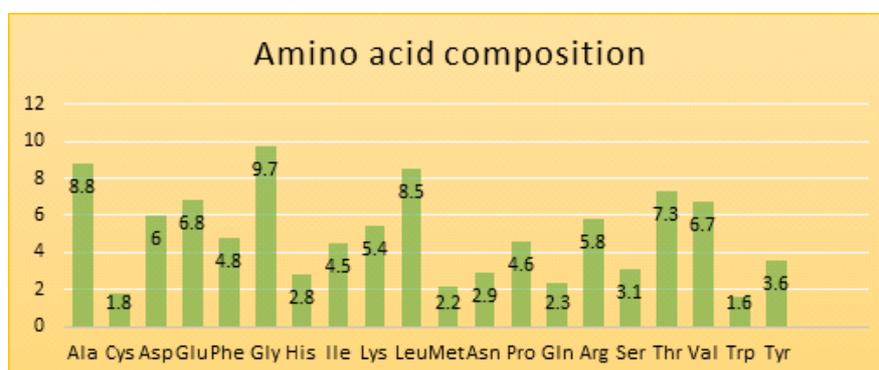


Fig. 1. The average amino acid composition of RuBisCO proteins from *Carthamus* ssp.

Table 1. The physiochemical properties of RuBisCO proteins from *Carthamus* spp.

Taxa	NCBI Accession number	UNIPROT Accession number	Sequence length (aa)	Mw	pI	-R	+R	EC	II	Al	GRAVY	Pfam Domain
<i>C. duvauxii</i>	AGL79759.1	R4QUQ8	472	undefined	-	60	53	69830	37.18	79.17	-0.254	N-terminal domain, Catalytic domain
<i>C. turkestanicus</i>	AGL79770.1	R4QWZ6	472	undefined	-	60	52	66850	36.25	77.31	-0.253	
<i>C. oxyacanthus</i>	AGL79769.1	R4QUR7	472	52241.43	6.03	60	52	69830	38.20	79.17	-0.250	
<i>C. tenuis</i>	ANN24277.1	A0A193CHY6	166	18711.47	6.92	21	21	17670	33.31	77.59	-0.313	
<i>C. tinctorius</i>	YP_009271914.1	A0A0K0N7H7	477	52843.09	5.99	61	53	69830	39.39	78.95	-0.263	

Note: Mw-molecular weight; pI-isoelectric point; +R-total number of positive and negative (-R) residues; EC-extinction coefficient; II-instability index; Al-aliphatic index; GRAVY-grand average of hydrophathy.

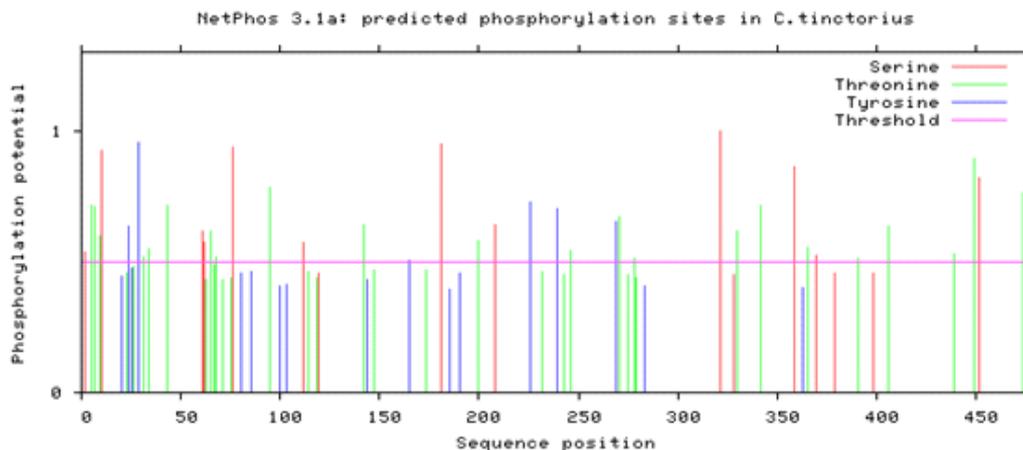


Fig. 2. RuBisCO proteins in *Carthamus* spp. determined by a score above a threshold of 0.5

Table 2. The RuBisCO proteins with aliphatic index (Al) values and their corresponding number of aliphatic residues

Aminoacids	<i>C. duvauxii</i>	<i>C. turkestanicus</i>	<i>C. oxyacanthus</i>	<i>C. tenuis</i>	<i>C. tinctorius</i>
Ala (Alanine)	9.1	8.7	9.1	7.8	9.0
Ile (Isoleucine)	4.4	4.4	4.4	4.8	4.4
Leu (Leucine)	8.5	8.3	8.5	9.0	8.4
Val (Valine)	6.8	6.6	6.8	5.4	6.9
Total	28.8	28.0	28.8	27.0	28.7

Table 3. Putative phosphorylation residue in RuBisCO protein sequences of *Carthamus* spp. with a score above 0.8

Taxa	Serine	Threonine	Tyrosine	Total
<i>C. duvauxii</i>	7	5	5	17
<i>C. turkestanicus</i>	7	5	4	16
<i>C. oxyacanthus</i>	7	6	5	18
<i>C. tenuis</i>	1	2	2	5
<i>C. tinctorius</i>	7	7	5	19

content increased (Table 2). The GRAVY values of RuBisCO proteins ranged from -0.313 to -0.250 (Table 1). A very low GRAVY value implies that the protein is more soluble in water than those with a higher GRAVY index. The most abundant amino acid ratio in the RuBisCO protein was Gly (9.7%), while the minimum amino acid ratio in the RuBisCO protein was Trp (1.6%) (Fig. 1). The putative phosphorylation sites were determined using the NetPhos 2.0 server based on a score above 0.8 (Table 3). The most phosphorylated site was found in *C. tinctorius* (Fig. 2). The confidence rates that these were true phosphorylation sites were above the threshold (0.5) and output score was given in a 0.0-0.1 range.

Phylogenetics has become a powerful tool and a starting point in many areas of biology, such as developmental genetics, genomics, taxonomy and biogeography (Wei *et al.*,

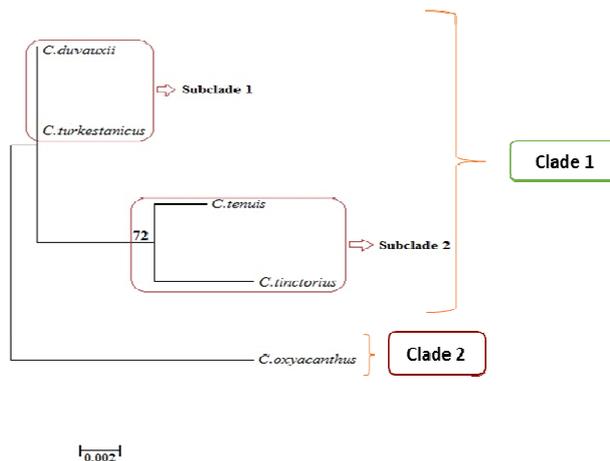


Fig. 3. Phylogenetic tree of *Carthamus* spp. RuBisCO proteins constructed using Maximum Likelihood method with MEGA 6.0

2014). In recent years, morphological, protein and DNA based markers have been developed for plant phylogenetic studies (Sun *et al.*, 2015). Previously studies, RAPD (Vilatersana *et al.*, 2005; Mahasi *et al.*, 2009), ISSR (Golkar *et al.*, 2011; Yaman *et al.*, 2014), SSR (Kiran *et al.*, 2015),

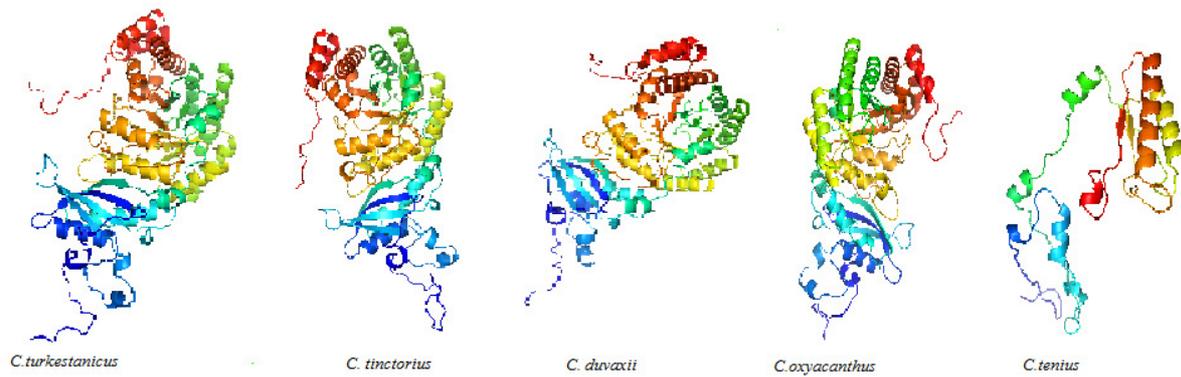


Fig. 4. The comprehensive 3D structures of *Carthamus* ssp. RuBisCO proteins. The visual data were obtained from PSIPRED software and analysed using PyMOL

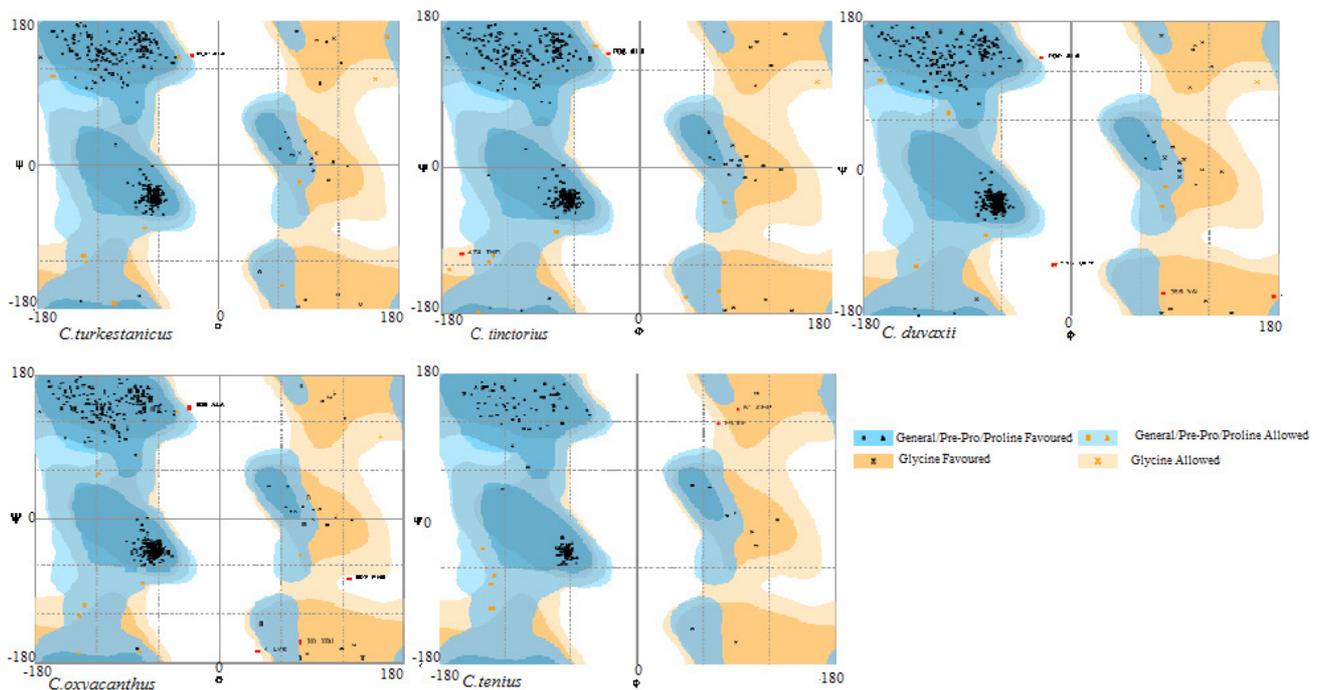


Fig. 5. Ramachandran plots of *Carthamus* ssp. RuBisCO proteins by using Rampage server

AFLP (Zhang *et al.*, 2006), nuclear SACPD gene and chloroplast trnL–trnF IGS region (Sasanuma *et al.*, 2008) molecular markers were used in studies on the genetic diversity and phylogenetic analyses of *Carthamus* plants. For phylogenetic analysis, MEGA 6.0 software was used. Maximum likelihood with bootstrap analysis was constructed in order to identify the relationships among *Carthamus* ssp. A phylogenetic analysis was performed with the RuBisCO protein sequences, which were found to comprise two main clades. Clade 1, was divided into 2 subclades; subclade 1 consists of *C. duvaxii* and *C. turkestanicus*. This subclade does not support bootstrap analysis. Subclade 2 consists of *C. tenuis* and *C. tinctorius*. This subclade received support of 72% bootstrap value. Clade 2 only consists of *C. oxyacanthus* (Fig. 3). The three-dimensional structure of the RuBisCO proteins were constructed using the PyMOL program and alpha helix, beta helix structures were demonstrated (Fig. 4). The three-dimensional structure of the proteins, contributes to the

understanding of protein function and active regions and facilitating drug design (Filiz and Koç, 2014). In the model validation, the Ramachandran plot analysis using the RAMPAGE server showed that 97.6%, 97.5%, 97.4%, 97.00% and 96.3% were in the favoured region; 2.2%, 2.1%, 1.7%, 2.1 % and 2.4 % in the allowed region; and 0.2 %, 0.4%, 0.9%, 0.9% and 1.2% in the outlier region in *C. turkestanicus*, *C. tinctorius*, *C. duvaxii*, *C. oxyacanthus*, and *C. tenuis* respectively (Fig. 5), indicating that the 3D models were fairly good in quality.

Conclusions

In this study, a silico analysis was carried out using bioinformatic tools such as ExPASy's ProtParam, MEGA 6.0, NetPhos 2.0, PSIPRED v3.3, RAMPAGE, and PyMOL of the RuBisCO protein in *Carthamus* ssp. The results of this study pave the way for further researches into the RuBisCO protein in different plant species.

References

- Andersson I, Knight S, Schneider G, Lindqvist Y, Lundqvist T, Brändén CI, Lorimer GH (1989). Crystal structure of the active site of ribulose-1,5- bisphosphate carboxylase. *Nature* 337(19):229-234.
- Ashri A, Knowles PF (1960). Cytogenetics of safflower (*Carthamus L.*) species and their hybrids. *Agronomy Journal* 52:11-17.
- Avcı MK, Tezcan E (2016). Erratum to: Genome-wide identification and comparative structural analysis of RuBisCo proteins in the Asteraceae. *Horticulture, Environment, and Biotechnology* 57(5):529-529.
- Filiz E, Koç İ (2014). In silico sequence analysis and homology modeling of predicted beta-amylase 7-like protein in *Brachypodium distachyon L.* *Journal of BioScience and Biotechnology* 3(1):61-67.
- Golkar P, Arzani A, Rezaei AM (2011). Genetic variation in safflower (*Carthamus tinctorious L.*) for seed quality-related traits and inter-simple sequence repeat (ISSR) markers. *International Journal of Molecular Sciences* 12(4):2664-2677.
- Idicula-Thomas S, Balaji PV (2005). Understanding the relationship between the primary structure of proteins and their amyloidogenic propensity: clues from inclusion body formation. *Protein Engineering Design and Selection* 18:175-180.
- Kiran BU, Mukta N, Kadirvel P, Alivelu K, Senthilvel S, Kishore P, Varaprasad KS (2015). Genetic diversity of safflower (*Carthamus tinctorius L.*) germplasm as revealed by SSR markers. *Plant Genetic Resources* 1-11. doi:10.1017/S1479262115000295
- Kitano K, Maeda N, Fukui T, Atomi K, Imanaka T, Miki K (2001). Crystal structure of a novel type archaeal rubisco with pentagonal symmetry. *Structure* 9(6):473-481.
- Mahasi MJ, Wachira FN, Pathak RS, Riungu TC (2009). Genetic polymorphism in exotic safflower (*Carthamus tinctorious L.*) using RAPD markers. *Journal of Plant Breeding and Crop Science* 1(1):8-12.
- Mukherjee S, Stasolla C, Brûlé-Babel A, Ayele BT (2015). Isolation and characterization of rubisco small subunit gene promoter from common wheat (*Triticum aestivum L.*). *Plant Signaling and Behavior* 10(2):e989033.
- Naeem MK, Rauf S, Iqbal H, Nawaz Shah MK, Mir A (2013). In silico studies of C3 metabolic pathway proteins of wheat (*Triticum aestivum*). *BioMed Research International* 1-7. <http://dx.doi.org/10.1155/2013/294759>
- Nagaraj G, Devi GN, Srinivas CVS (2001). Safflower Petals and their Chemical Composition. Proc. V. International Safflower Conference, July 23-27, USA.
- Nylinder S, Anderberg AA (2015). Phylogeny of the Inuleae (Asteraceae) with special emphasis on the Inuleae-Plucheinae. *TAXON* 64(1):110-130.
- Öztürk M, Çetin Ö (2013). *Inula tuzgoluensis* (Asteraceae), a new species from Central Anatolia, Turkey. *Turkish Journal of Botany* 37:825-835.
- Paksoy MY, Selvi S, Sevindik E, Uysal H (2016). Soil content in the *Anacyclus L.* (Asteraceae) genus growing in Turkey. *Turkish Journal of Life Sciences* 1(1)8-12.
- Sasanuma T, Sehgal D, Sasakuma T, Raina SN (2008). Phylogenetic analysis of *Carthamus* species based on the nucleotide sequence of the nuclear SACP gene and chloroplast *trn L-trn F* IGS region. *Genome* 51(9):721-727.
- Sun YL, Kang HM, Han SH, Park YC, Hong SK (2015). Taxonomy and phylogeny of the genus *citrus* based on the nuclear ribosomal DNA its region sequence. *Pakistan Journal of Botany* 47(1):95-101.
- Süslü İ, Pehlivan S, Ekni M, Şenel E (2010). Türkiye'nin çeşitli ballarına kaynak oluşturan compositae (Asteraceae) familyasında *Inula* türlerinin polen morfolojilerine istatistiksel bir yaklaşım. [A statistical approach to the pollen morphologies of *Inula* species in the family Compositae (Asteraceae) as a source of various kinds of Turkish honeys]. *TUBAV Bilim Dergisi* 3(2):182-187.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Symposium Series* 22:4673-4680.
- Uysal N, Baydar H, Erbaş S (2006). Determination of agricultural and technological properties of safflower (*Carthamus tinctorius L.*) lines developed from Isparta population. *Süleyman Demirel Üniversitesi Ziraat Fakültesi Dergisi* 1(1):52-63.
- Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N (2005). Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. *Botanical Journal of the Linnean Society* 147(3):375-383.
- Wei W, HongLei L, ZhDuan C (2014) Analysis of plastid and nuclear DNA data in plant phylogenetics-evaluation and improvement. *Life Sciences* 57(3):280-286.
- Yaman H, Tanıkahya-Hacıoğlu B, Arslan Y, Subaşı İ (2014). Molecular characterization of the wild relatives of safflower (*Carthamus tinctorius L.*) in Turkey as revealed by ISSRs. *Genetic resources and crop evolution* 61(3):595-602.
- Yıldırım B, Tunçtürk M, Dede Ö, Okut N (2005). Aspir (*Carthamus tinctorius L.*) de Farklı Azot ve Fosfor Dozlarının Verim ve Kalite Üzerine Etkileri [Effects of different nitrogen and phosphorus doses on yield and quality in Aspir (*Carthamus tinctorius L.*)]. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi* 15(2):113-116.
- Yue S, Tang Y, Li S, Duan JA (2013). Chemical and biological properties of quinochalcone C-Glycosides from the florets of *Carthamus tinctorius*. *Molecules* 18:15220-15254.
- Zhang L, Huang B B, Kai G Y, Guo M L (2006). Analysis of intraspecific variation of Chinese *Carthamus tinctorius L.* using AFLP markers. *Yao Xue Xue Bao* 41(1):91-96.
- Zhang B, Luo L, Zhang X, Li R, Song Y, Zhang D, Wei Y (2011). Bioinformatics analysis on ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunits in different plants. *Molecular Plant Breeding* 2(15). doi: 10.5376/mpb.2011.02.0015.