UDC 575
https://doi.org/10.2298/GENSR1901069S
Original scientific paper

COMPARATIVE AND PHYLOGENETIC ANALYSIS OF RUBISCO LARGE SUBUNIT (*rbc*L) PROTEINS IN SOME *Sideritis* L. (LAMIACEAE) SPECIES: A BIOINFORMATIC APPROACH

Emre SEVİNDİK

Faculty of Agriculture, Department of Agricultural Biotechnology, Adnan Menderes University, South Campus, Cakmar, Aydin, Turkey

Sevindik E. (2019): *Comparative and phylogenetic analysis of rubisco large subunit (rbcl) proteins in some Sideritis L. (Lamiaceae) species: a bioinformatic approach*.- Genetika, Vol 51, No.1, 69-80.

The large subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein, which plays an important role in the photosynthesis reaction, are encoded by the chloroplast genome. *Sideritis* L., a medical and aromatic plant group, belongs to Lamiaceae family. In this study, we performed sequence, physicochemical, phylogenetic and three-dimensional (3D) bioinformatic analyses of RuBisCO large subunit (*rbc*L) proteins in the *Sideritis* ssp. using various bioinformatics tools. Physicochemical analyzes were performed by ExPASy - ProtParam. The putative phosphorylation sites of the *rbc*L proteins were determined by NetPhos 2.0 and NetPhos 3.1. Phylogenetic analyses were performed with the MEGA 6.0 software. To estimate 3D protein structures, PyMol program was used. At the end of the study, it was found that the amino acid number of stilbene synthase proteins ranged between 171 and 456, molecular weight ranged between 19002.67 and 50420.44 Da, instability index ranged between 27.30 to 40.70 and GRAVY values ranged between -0.394 to -0.226. While the highest average amino acid rate in the *rbc*L proteins was Gly (10.00%), the lowest amino acid ratio (1.4%) was determined as Trp. In phylogenetic analyses performed using protein sequences, maximum likelihood (ML) tree consisted of 2 large clades. Pairwise distance analysis based on *Sideritis* species' *rbc*L protein sequences was performed using MEGA 6.0. The lowest pairwise

Corresponding author: Emre Sevindik, Faculty of Agriculture, Department of Agricultural Biotechnology, Adnan Menderes University, South Campus, Cakmar, Aydin, Turkey, E-mail: ph.d-emre@hotmail.com

distance was 0.000, while the highest pairwise distance was 0.024. When the estimated 3D structures of the proteins were examined, the Gly residue, which plays an important role in the structure and function of the proteins, was detected as the least in *S. libanotica subsp. kurdica* species while it was the most abundant residue in *S. cretica subsp. spicata*. The results of our study provide insights into fundamental characteristics of *rbc*L proteins in *Sideritis* taxa.

Keywords: *Sideritis*, RuBisCO, *rbc*L, phylogenetic analysis, bioinformatic analysis

INTRODUCTION

The family Lamiaceae is of great importance in a variety of fields such as medical, food, cosmetic and perfumery since its most members are rich in essential oils, aromatic oils and various secondary metabolites (KAHRAMAN *et al*., 2009; BAŞER, 1993). The genus *Sideritis* of Lamiaceae family has more than 150 species distributed in the temperate and tropical regions of the Northern Hemisphere (KAYA *et al*., 2015). *Sideritis* species, which have become popular in recent years, are used as herbs for preparing tea or aromatic properties in local cuisines. They are sold in various shops and are markets as mountain teas, mountain stream, sage, yaylaçayi, malotira, de Puerto, rabo de gato or zaharena in various Mediterranean countries (TÜRKMENOĞLU *et al*., 2015). Photosynthesis can be defined as the reduction of CO_2 by green plants in the catalysis of specific pigment molecules by the help of light energy in the presence of water to form various organic substances. In this system, solar energy is stored as chemical energy in organic matter (TÜRK and ÇELIK, 2006). RuBisCO (ribulose-1,5-bisphosphate carboxylase / oxygenase, EC 4.1.1.39) acts as the major gateway for the inorganic carbon to enter metabolic pathways in most ecosystems, and therefore the importance of supporting life is unique (KAPRALOV *et al.,* 2012). Functional RuBisCO has low catalytic efficiency, which requires all plants to produce large amounts of enzyme (SPREITZER and SALVUCCI, 2002; TABITA *et al.*, 2007). As a result, RuBisCO has been labeled as the most abundant protein on earth and can be found in autotrophic organisms including bacteria, algae and plants with an estimated contribution of over half of the total soluble proteins in plant leaves (ANDERSON and BACKLUND, 2008; UDENIGWE *et al*., 2013). In this study, large subunit of RuBisCO (*rbc*L) was analyzed in some medicinal plant (*Sideritis*) species with respect to physicochemical, phylogenetic and 3D structure properties utilizing bioinformatics tools.

MATERIAL AND METHODS

The RuBisCO protein sequences of *Sideritis* species were obtained from the National Center for Biotecnology Information (NCBI: https://www.ncbi.nlm.nih.gov/protein) in FASTA format. The physiochemical properties of *rbc*L proteins including isoelectric point (pI), molecular weight (MW), total number of positive (R) and negative (-R) residues, extinction coefficient (EC), and GRAVY values were identified by ExPASy - ProtParam. (http://web.expasy.org/protparam/) (GASTEIGER *et al.,* 2005). The putative phosphorylation sites of the *rbc*L proteins were detected by NetPhos 2.0 and NetPhos 3.1 (http://www.cbs.dtu.dk/services/NetPhos/) (BLOM *et al*., 1999). The *rbc*L protein sequences were aligned using MEGA 6.0 (TAMURA *et al*., 2013). The phylogenetic tree of the *rbc*L proteins of *Sideritis* species was generated using the maximum likelihood method with MEGA 6.0 and the bootstrap values were performed with 1000 replicates. Pairwise distance analysis based on

*Sideritis rbc*L proteins' sequence set was performed with MEGA 6.0. All domains and conserved protein motifs were analyzed using a protein BLAST in the MEME (http://meme-suite.org/doc/fasta-format.html) (BAILEY *et al.*, 2009). To estimate the 3D structure of *rbc*L proteins, homology models were performed in PSIPRED v3.3 using the following method alternatives (http://bioinf.cs.ucl.ac.uk/psipred/) (BUCHAN *et al.*, 2013). The results were checked and verified with a Ramachandran plot analysis in RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (LOVELL *et al.*, 2003), that determined the best predictive models. 3D comparative analyzes were performed using PyMOL.

RESULTS AND DISCUSSION

RuBisCO large subunit protein sequences were aligned using MEGA 6.0 software. In all *Sideritis* species, the residues corresponding to the region between 23-191 amino acids were conserved, (except three residues: P 47, V 137 and T 139) (Fig 1). The physicochemical analyses of the predicted *rbc*L proteins were performed using ExPASy - ProtParam and the results are shown in Table 1. The amino acid sequence length ranged from 171 to 456 amino acids. The shortest amino acid sequence was in *Sideritis vulcanica* (171 amino acids), while the longest amino acid sequences were in *Sideritis dasygnaphala, Sideritis gomeraea, Sideritis hyssopifolia, Sideritis macrostachys, Sideritis montana, Sideritis romana* and *Sideritis syriaca* (456 amino acids). The minimum and maximum molecular weights (MW) were 19002.67 and 50420.44 Da, respectively. The most acidic protein sequence (*pI*=6.46) was detected in 7 species while the most basic protein sequence (*pI=8.31*) was found only in *Sideritis libanotica* subsp. *kurdica* (Table 1).

Table 1. The physiochemical properties of rbcL proteins from Sideritis species

Taxa	NCBI Accession Number	Sequence Lenght (aa)	Mw	*pI*	-R	+R	EC	II	Al	GRAVY
S.gomerae subsp. *gomerae*	AIE90067.1	203	22765.92	6.92	25	25	36120	33.54	74.93	-0.394
Sideritis lotsyi	AIE90069.1	203	22765.92	6.92	25	25	36120	33.54	74.93	-0.394
Sideritis dasygnaphala	AAM33270.1	456	50378.37	6.46	55	51	65945	40.70	81.32	-0.244
Sideritis gomeraea	AAM33271.1	456	50378.37	6.46	55	51	65945	40.70	81.32	-0.244
Sideritis libanotica subsp. *kurdica*	APP91296.1	189	21047.05	8.31	22	24	30620	27.30	78.94	-0.316
Sideritis sipylea	AEX55433.1	174	19380.11	6.92	20	20	29130	30.31	81.26	-0.280
Sideritis hyssopifolia	AAM33272.1	456	50382.36	6.46	55	51	65945	39.43	81.54	-0.237
Sideritis vulcanica	APP91297.1	171	19002.67	6.92	20	20	27640	29.94	80.41	-0.296
Sideritis macrostachys	AAM33273.1	456	50378.37	6.46	55	51	65945	40.70	81.32	-0.244
Sideritis montana	AAM33274.1	456	50420.44	6.46	55	51	65945	40.02	80.68	-0.232
Sideritis cretica subsp. *spicata*	BAO57028.1	442	49006.84	6.60	53	50	65820	38.86	82.10	-0.242
Sideritis romana	AAM33275.1	456	50408.44	6.46	55	51	65945	39.25	82.39	-0.226
Sideritis syriaca	AAM33276.1	456	50382.36	6.46	55	51	65945	39.43	81.54	-0.237

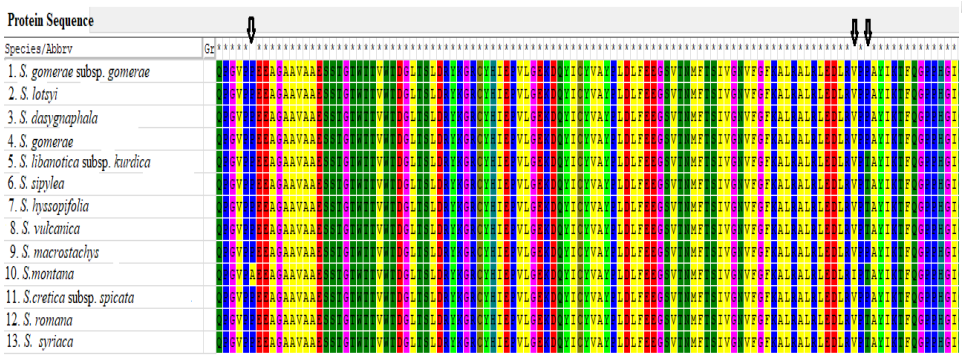

Fig. 1. Multiple sequence alignment of *rbc*L protein sequences from amino acid residues (23-191 aa)

The extinction coefficients of *rbc*L proteins at 280 nm ranged from 27640 to 65945 M^{-1} cm^{-1}. The instability index (II) values for the *rbc*L proteins ranged from 27.30 to 40.70. The aliphatic index (AI) of the proteins of thermophilic bacteria has been 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 in our study ranged from 74.93 to 82.39. (Table 2).

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

Taxa	Ala (A)	Ile (I)	Leu (L)	Val (V)
Sideritis gomerae subsp. *gomerae*	6.9	3.4	8.9	6.9
Sideritis lotsyi	6.9	3.4	8.9	6.9
Sideritis dasygnaphala	9.4	4.6	8.8	6.8
Sideritis gomerae	9.4	4.6	8.8	6.8
Sideritis libanotica subsp. *kurdica*	7.9	3.7	9.0	7.4
Sideritis sipylea	7.5	4.0	9.8	6.9
Sideritis hyssopifolia	9.6	4.6	8.8	6.8
Sideritis vulcanica	7.6	4.1	9.4	7.0
Sideritis macrostachys	9.4	4.6	8.8	6.8
Sideritis montana	9.6	4.8	8.3	6.8
Sideritis cretica subsp. *spicata*	8.8	4.8	8.8	7.0
Sideritis romana	9.6	4.8	8.8	6.8
Sideritis syriaca	9.6	4.6	8.8	6.8

The GRAVY values of *rbc*L proteins ranged from -0.394 to -0.226 (Table 1). While the highest average amino acid rate in the *rbc*L proteins was Gly (10.00%), the lowest amino acid ratio (1.4%) was determined as Trp. The putative phosphorylation sites were determined using the NetPhos 2.0 and NetPhos 3.1 server (Table 3). While the mostly phosphorylated sites were determined in *Sideritis hyssopifolia* and *Sideritis syriaca*, the least phosphorylated sites were

detected in *Sideritis gomerae* subsp. *gomerae*, *Sideritis lotsyi*, *Sideritis libanotica* subsp. *kurdica*, *Sideritis sipylea*, and *Sideritis vulcanica* (Table 3; Fig 2).

Table 3. Putative phosphorylated residues in rbcL protein sequences of Sideritis species with a score above 0.8

Taxa	Ser	Thr	Tyr	Total
Sideritis gomerae subsp. *gomerae*	3	2	2	7
Sideritis lotsyi	3	2	2	7
Sideritis dasygnaphala	7	4	5	16
Sideritis gomerae	7	4	5	16
Sideritis libanotica subsp. *kurdica*	3	2	2	7
Sideritis sipylea	3	2	2	7
Sideritis hyssopifolia	8	4	5	17
Sideritis vulcanica	3	2	2	7
Sideritis macrostachys	7	4	5	16
Sideritis montana	7	4	5	16
Sideritis cretica subsp. *spicata*	6	3	5	14
Sideritis romana	7	4	5	16
Sideritis syriaca	8	4	5	17

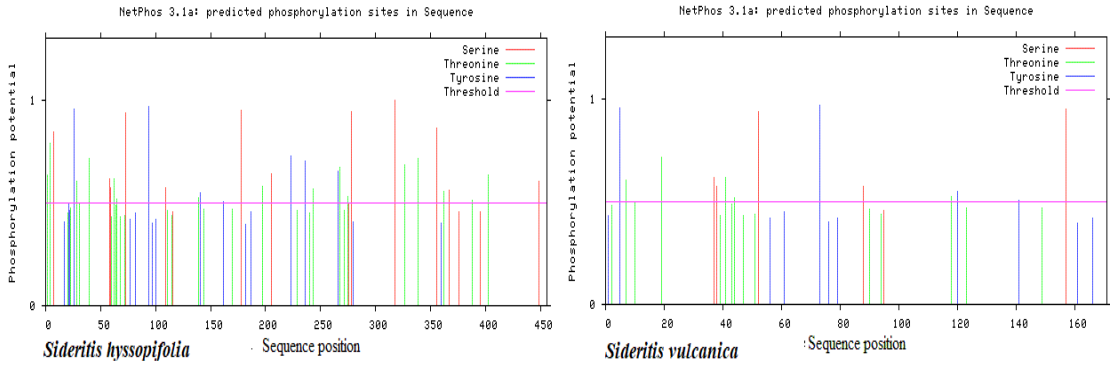

Fig. 2. *rbc*L proteins in *Sideritis hyssopifolia* and *Sideritis vulcanica* determined by a score above a threshold of 0.5

Molecular analysis of genetic variation among individuals of a population can provide a means of monitoring the genetic diversity of a declining population and assess genetic consequences of fragmentation on remaining populations (AL-QURANINY *et al.*, 2014). In recent years, many marker techniques based on morphological, protein and DNA have been developed and are used in plant phylogenetic studies. The application of these molecular marker techniques has been used to examine and analyze the genome-wide variability (SUN *et al.*, 2015). In

previous studies utilizing RAPD (VAZQUEZ *et al.*, 1999), AFLP (NEMLI *et al.*, 2014), nuclear ribosomal ITS sequences (BARBER *et al.*, 2002; DÜNDAR *et al*., 2013), *trn*L intron and *trn*T–*trn*L intergenic spacer regions of chloroplast DNA (BARBER *et al*., 2002), molecular markers were used to document the genetic diversity and phylogenetic analyses of *Sideritis* species. *rbc*L protein sequences from 13 *Sideritis* species were used in our phylogenetic analysis. According to the results, two clades were observed in the phylogenetic tree generated using Maximum Likelihood method with MEGA 6.0 (Fig 3).

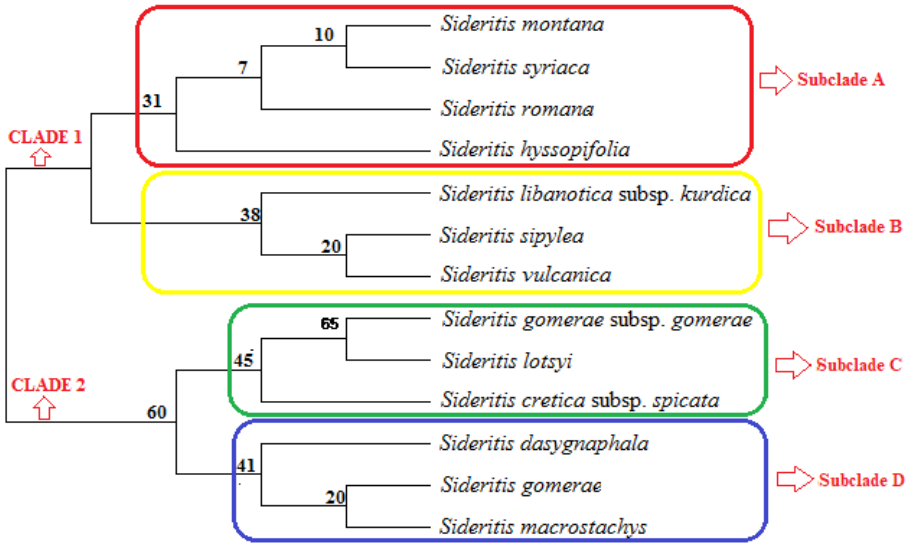

Fig. 3. Phylogenetic tree of *Sideritis* ssp. *rbc*L proteins constructed using Maximum likelihood method with MEGA 6.0

Clade 1 consists of two subclades; subclade A and subclade B. Subclade A is comprised of *S. montana*, *S. syriaca*, *S. romana* and *S. hyssopifoli.* In a study based on nrDNA ITS sequence analysis (BARBER *et al.*, 2002), *S. montana* and *S. romana* have also appeared in the same clade. However, BARBER *et al.* (2002) also study *trn*L intron and *trn*T–*trn*L intergenic spacer regions of chloroplast DNA. In their study, *S. montana* and *S.romana* were found in different clades (BARBER *et al.*, 2002). Subclade B is comprised of *S. libanotica* subsp. *kurdica*, *S.sipylea* and *S. vulcanica*. This Subclade's members belong to the *Empedoclia* section taxonomically. In the phylogenetic study using nrDNA ITS sequences (DÜNDAR *et al*., 2013), *S. montana* and *S. syriaca* species were grouped in the same clade. In our study, these two species formed a monophyletic group (Fig 3). Clade 2 consisted of two subclades; subclade C and subclade D. Subclade C included *S. gomerae* subsp. *gomerae*, *S. lotsyi* and *S. cretica* subsp. *spicata* while subclade D had *S. dasygnaphala*, *S. gomerae* and *S. macrostachys* (Figure 3). Using nrDNA ITS sequences, *trn*L intron and *trn*T-*trn*L intergenic spacer regions of chloroplast DNA, BARBER *et al*. (2002) determined *S. gomerae* subsp. *gomerae* and *S. macrostachys* species

in the same clade. In taxonomic classification, these two species belong to section *Marrubiastrum*. DÜLGEROĞLU (2017) has reported morpho-phylogenetic analysis of *Sideritis* taxa using 26 morphological characters. In his study (DÜLGEROĞLU, 2017), *S. montana* and *S. romana* were grouped in the same branch. These two species were grouped in subclade A in our study (Fig 3). He also detected *S. gomerae*, *S. macrostachys* and *S. cretica* in the same branch. Similarly in our study, these two species were grouped in clade 2 (Fig 3). DÜLGEROĞLU (2017) also detected *S. syriaca* and *S.sipylea* in different branches as in our findings (Fig 3). Pairwise distance analysis based on *Sideritis* species' *rbc*L protein sequences was performed using MEGA 6.0. The lowest pairwise distance was 0.000, while the highest pairwise distance was 0.024 (Table 4). A total of six conserved motifs were detected in *rbc*L protein sequences (Fig 4; Table 5).

Table 4. Pairwise distance among Sideritis species' rbcL proteins obtained using MEGA 6.0

Taxa	1	2	3	4	5	6	7	8	9	10	11	12	13
S. gomerae subsp. *gomerae*	-												
S. lotsyi	0.000	-											
S. dasygnaphala	0.006	0.006	-										
S. gomerae	0.006	0.006	0.000	-									
S. libanotica subsp. *kurdica*	0.012	0.012	0.012	0.012	-								
S. sipylea	0.012	0.012	0.012	0.012	0.006	-							
S. hyssopifolia	0.012	0.012	0.006	0.006	0.006	0.006	-						
S. vulcanica	0.012	0.012	0.012	0.012	0.006	0.006	0.006	-					
S. macrostachys	0.006	0.006	0.000	0.000	0.012	0.012	0.006	0.012	-				
S. montana	0.024	0.024	0.018	0.018	0.018	0.018	0.012	0.018	0.018	-			
S. cretica subsp. *spicata*	0.006	0.006	0.006	0.006	0.012	0.012	0.012	0.012	0.006	0.024	-		
S. romana	0.012	0.012	0.006	0.006	0.006	0.006	0.000	0.006	0.006	0.012	0.012	-	
S. syriaca	0.012	0.012	0.006	0.006	0.006	0.006	0.000	0.006	0.006	0.012	0.012	0.000	-

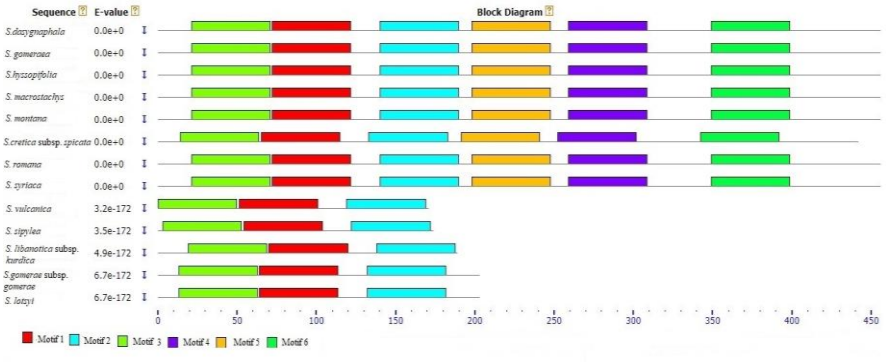

Fig. 4. Combined block diagrams of the conserved protein motifs in *rbc*L proteins belongs to *Sideritis* ssp. determined by MEME.

Table 5. The most conserved motifs in rbcL proteins of the Sideritis species

Motif	Width	Best possible match	Site Number
1	50	SLDRYKGRCYHIEPVLGEKDQYICYVAYPLDLFEEGSVTNMFTSIVGNVF	13
2	50	YIKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECL	13
3	50	YTPEYETKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTTVWTDGL	13
4	50	PIVMHDYLTGGFTANTSLSHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFR	8
5	50	DDENVNSQPFMRWRDRFLFCAEAIYKAQAETGEIKGHYLNATAGTCEEMM	8
6	50	FIEKDRSRGIYFTQDWVSLPGVIPVASGGIHVWHMPALTEIFGDDSVLQF	8

Motifs 1, 2, 3, 4, 5, and 6 were observed in eight protein sequences including *S. dasygnaphala, S. gomerae, S. hyssopifoli, S. macrostachys, S. montana, S. cretica* subsp. *spicata, S. romana* and *S. syriaca*. Motifs 1, 2 and 3 were found in *S. vulcanica, S. sipylea, S. libanotica* subsp. *kurdica, S. gomerae* subsp. *gomerae and S. lotsyi.* In the model validation, the Ramachandran plot analysis conducted through the RAMPAGE server revealed that 92.9-97.1% of the residues were found in the favored regions, 1.7-6.0% of the residues were located in the allowed regions, and 0-1.2% of the residues were found in the outlier regions. Two *Sideritis rbc*L proteins revealed the results of RAMPAGE (Fig 5), indicating that the 3D models were fairly good in quality. The three-dimensional structures of the *rbc*L proteins were constructed using the PyMOL program. Gly residue is unique among amino acids since it has a hydrogen atom as its variable (R) group. Conformation has more freedom, so that it can provide flexibility for adjacent residues. Therefore, it is not surprising that Gly plays a special role in the structure and function of enzymes (YAN and SUN, 1997). In this study, the highest amount of Gly residues was detected in *S. cretica* subsp. spicata, while the lowest amount of Gly residues was detected in *S. libanotica* subsp. *kurdica* (Fig 6).

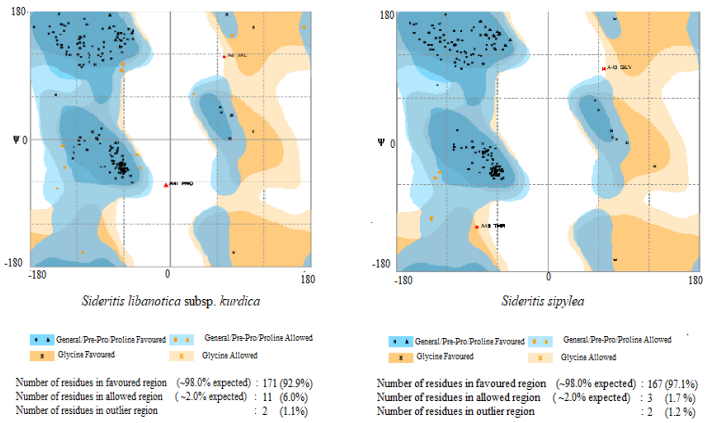

Fig. 5. The ramachandram plots of two *Sideritis rbc*L proteins produced using RAMPAGE according to the results of PSIBRED. These *rbc*L proteins have the lowest (A) and highest (B) amount of residues located in the favored regions respectively.

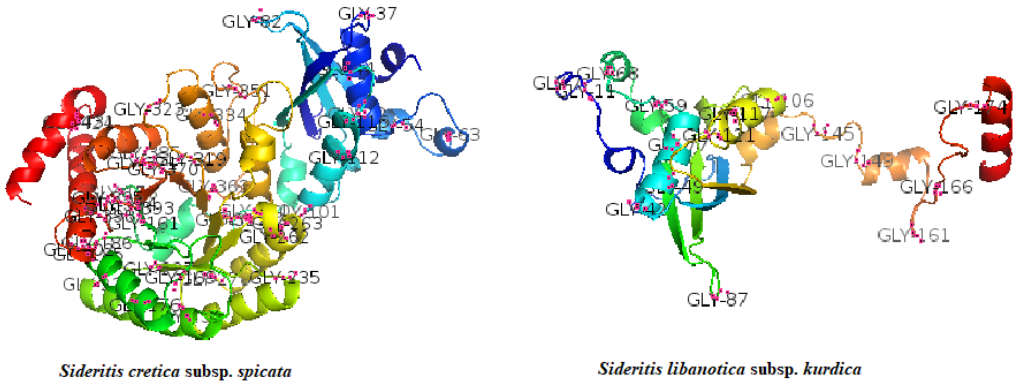

Fig. 6. Conserved Gly residues in the *S. cretica* subsp. *spicata* and *Sideritis libanotica* subsp. *kurdicarbc*L protein sequences

However, in our study PHE 217-CYS 218 peptide was also detected. The dipeptide is present in *rbc*L derived from various photosynthesizing plant-based foods and edible microalgae, and was predicted to be bioactive; the two amino acid residues, PHE and CYS can make the peptide an antioxidant (JE *et al.*, 2015). This dipeptide was detected in the *rbc*L proteins of *Sideritis* species (Fig 7).

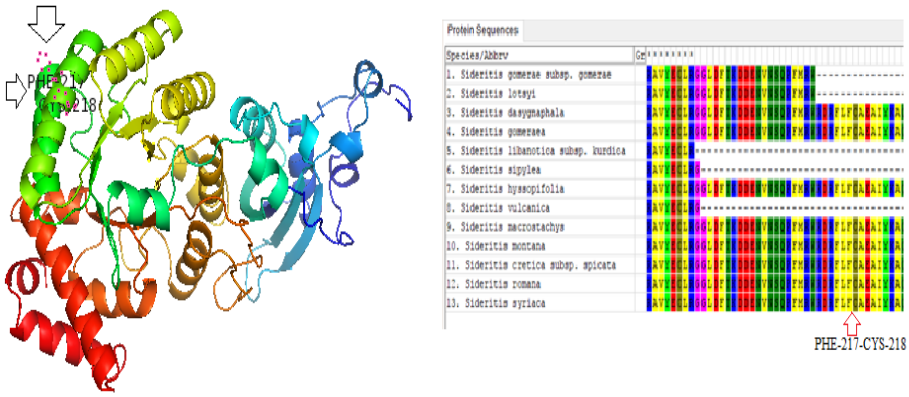

Fig. 7. The location of the *Sideritis montana* PHE-217-CYS-218 peptide

Using RuBisCO protein sequences, numerous bioinformatics and in silico studies have been performed before. AVCI and TEZCAN (2016) have performed a genome-wide analysis of RuBisCO proteins belonging to some Asteraceae species. SUBRAMANI and HWA (2010) have performed in silico studies of RuBisCO proteins of some C3 plants belonging to the Poaceae family. NAEEM *et al.* (2012) have performed in silico studies of RuBisCO proteins of *Triticum*

aestivum. ZHANG *et al*. (2011) performed bioinformatic analysis of RuBisCO proteins of different plants.

CONCLUSION

Bioinformatic analyses are made to understand the structures of proteins and the functions of proteins using computer programs. In this study, bioinformatic and phylogenetic analyses of the *rbc*L protein in *Sideritis* species were carried out using bioinformatics tools such as ExPASy - ProtParam, MEGA 6.0, NetPhos 2.0, NetPhos 3.1, PSIPRED v3.3, RAMPAGE, and PyMOL. The results will be a useful source for the future biotechnology and bioinformatics analyzes on RuBisCO or other proteins.

Received, February 13th, 2018
Accepted November 18th, 2018

REFERENCES

AL-QURAINY, F., A.R.Z., GAAFAR, S., KHAN, M., NADEEM, A.A., AL-AMERI, M., TARROUM (2014): Genetic Diversity in *Breonadia salicina* Based On Intra-Species Sequence Variation of Chloroplast DNA Spacer Sequence. Pak. J. Bot., *46*(2): 599-604.

ANDERSSON, I., A., BACKLUND (2008): Structure and function of RuBisCO. Plant Physiol. Biochem., *46*:275–291.

AVCI, M.K., E., TEZCAN (2016): Erratum to: Genome-wide identification and comparative structural analysis of RuBisCO proteins in the asteraceae. Hortic. Environ. Biotech., *5*(57): 529-529.

BAILEY, T. L., M., BODEN, F.A., BUSKE, M., FRITH, C.E., GRANT, L., CLEMENTI, ... & W.S., NOBLE (2009): MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res., *37*(2): 202-208.

BARBER, J.C., J., FRANCISCO-ORTEGA, A., SANTOS-GUERRA, K.G., TURNER, R.K., JANSEN (2002): Origin of Macaronesian *Sideritis* L. (Lamioideae: *Lamiaceae*) inferred from nuclear and chloroplast sequence datasets. Mol. Phylogenet. Evol., *23*(3): 293-306.

BAŞER, K.H.C. (1993): Essential Oils of Anatolian Labiateae: A Profile. Acta Horticul., *333*:217-238.

BLOM, N., S., GAMMELTOFT, S., BRUNAK (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol., *294*(5): 1351-1362.

BUCHAN, D.W., F., MINNEC, T.C., NUGENT, K., BRYSON, D.T., JONES (2013): Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res., *41*(1): 349-357.

DÜLGEROĞLU, C. (2017): A preliminary intra phylogeny of the genus *Sideritis* by morphology. IJAER, *3*(5): 3901-3909

DÜNDAR, E., E. AKCICEK, T., DIRMENCI, Ş., AKGÜN (2013): Phylogenetic analysis of the genus *Stachys* sect. *Eriostomum* (Lamiaceae) in Turkey based on nuclear ribosomal ITS sequences. Turk. J. Bot., *37*(1): 14-23.

GASTEIGER, E., C., HOOGLAND, A., GATTIKER, S., DUVAUD, M.R., WILKINS, R.D., APPEL, A., BAIROCH (2005): Protein identification and analysis tools on the ExPASy server, In: John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, 571–607.

IDICULA-THOMAS, S., P.V., BALAJI (2005): Understanding the relationship between the primary structure of proteins and their amyloidogenic propensity: clues from inclusion body formation. Prot. Eng. Des. Sel., *18*:175-180.

JE, J.Y., Y.S., CHO, M., GONG, C.C., UDENIGWE (2015): Dipeptide Phe-Cys derived from in silico thermolysin-hydrolysed RuBisCO large subunit suppresses oxidative stress in cultured human hepatocytes. Food Chem., *171*: 287-291.

KAHRAMAN, A., F., CELEP, M., DOĞAN (2009): Morphology, anatomy and palynology of *Salvia indica* L. (Labiatae). World Appl. Sci., *6*(2): 289-296.

KAPRALOV, M.V., J.A.C., SMITH, D.A., FILATOV (2012): Rubisco evolution in C4 eudicots: an analysis of Amaranthaceae sensu lato. PloS one, *7*(12): e52974.

KAYA, M.D., E.G., KULAN, G., GÜMÜŞÇÜ, A., GÜMÜŞÇÜ (2015): Factors Influencing Germination Performance of Four Endemic *Sideritis* Species of Turkey. JAS, *21*(3): 406-413.

LOVELL, S.C., I.W., DAVIS, W.B., ARENDALL, P.I.W., BAKKER, J.M., WORD, M.G., PRISANT, J.S., RICHARDSON, D.C., RICHARDSON (2003): Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics, *50*(3): 437-450.

NAEEM, M.K., S., RAUF, H., IQBAL, M.K., NAWAZ SHAH, A., MIR (2012): In silico studies of C3 metabolic pathway proteins of wheat (*Triticum aestivum*). BioMed Res. Int., *2013*: 1-7.

NEMLI, S., U., SUBASI, V., EROGLU, S.G., SENOL, M.B., TANYOLAC (2014): High Levels of Genetic Variation as Detected by AFLP in *Sideritis tmolea* from Western Turkey. Turk. J. Field Crops, *19*(2): 238-245.

SPREITZER, R.J., M.E., SALVUCCI (2002): Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. Ann. Rev. Plant Biol., *53*:449–475.

SUN, Y.L., H.M., KANG, S.H., HAN, Y.C., PARK, S.K., HONG (2015): Taxonomy and Phylogeny of the Genus *Citrus* Based on the Nuclear Ribosomal DNA ITS Region Sequences. Pak. J. Bot., *47*(1):95-101.

SUBRAMANI, B., K.Y., HWA (2010): In silico Analysis for Enhancing the Rubisco Activity among the C3 Plants of *Poaceae* Family. Information Technology Convergence and Services (ITCS), 2nd International Conference on. 1-6. IEEE

TABITA, F.R., T.E., HANSON, H., LI, S., SATAGOPAN, J., SINGH, S., CHAN (2007): Function, structure, and evolution of the RubisCO-like proteins and their RuBisCO homologs. Microbiol. Mol. Biol Rev., *71*:576–599.

TAMURA, K., G., STECHER, D., PETERSON, A., FILIPSKI, S., KUMAR (2013): MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol., *30*: 2725-2729.

TURKMENOGLU, F.P., I,. BAYSAL, S., CIFTCI-YABANOGLU, K., YELEKCI, H., TEMEL, S., PAŞA, E., NURTEN, I., ÇALIS, G., UCAR (2015): Flavonoids from *Sideritis* species: human monoamine oxidase (hMAO) inhibitory activities, molecular docking studies and crystal structure of xanthomicrol. Molecules, *20*(5): 7454-7473.

TÜRK, M., N., ÇELIK (2006): CO_2 Özümlemesinde C-3 ve C-4 Tipi Bitkilerde Fotosentez-Solunum Denge Noktalarının Belirlenmesi. SDÜ Fen Bil Enst Der., *10(*1):48-51.

UDENIGWE, C.C., M., GONG, S., WU (2013): In silico analysis of the large and small subunits of cereal RuBisCO as precursors of cryptic bioactive peptides. Process Biochem., *48*(11): 1794-1799.

VAZQUEZ, J.H.L., F., GOMEZ-MERCADO, J.L.G., GUERRERO, I., RODRIGUEZ-GARCIA, F., GARCH-MAROTO (1999): Genetic relationships and population structure within taxa of the endemic *Sideritis pusilla* (Lamiaceae) assessed using RAPDs. Bot. J. Linn. Soc., *129* (4): 345-358.

YAN, B.X., Y.Q., SUN (1997): Glycine Residues Provide Flexibility for Enzyme Active Sites. J. Biol. Chem., *272*: 3190-3194.

ZHANG, B., L., LUO, X., ZHANG, R., LI, Y., SONG, D., ZHANG, Y. NIE, Y. ZENG, Q. LIAO, Y. WEI (2011): Bioinformatics analysis on ribulose-1, 5- bisphosphate carboxylase/oxygenase large subunits in different plants. Mol. Plant Breed., *2*:101-108

KOMPARATIVNA I FILOGENETSKA ANALIZA RUBISCO VELIKE PROTEINSKE SUBJEDINICE (rbcL) U NEKIM SIDERITIS (LAMIACEAE) VRSTAMA: BIOINFORMATIČKI PRISTUP

Emre SEVİNDİK

Poljoprivredni fakultet, Departman za poloprivrednu biotehnologiju, Adnan Menderes Univerzitet, Južni Kampus, Čakmar, Ajdin, Turska

Izvod

Velike podjedinice proteina ribuloza-1,5-bisfosfat-karboksilaza / oksigenaza (RuBisCO), koje igraju važnu ulogu u fotosinteznoj reakciji, kodirane su genomima hloroplasta. Sideritis L., grupa medicinskih i aromatičnih biljaka, pripada porodici Lamiaceae. U ovom radu smo uradili sekvencionu, fizičko-hemijsku, filogenetsku i trodimenzionalnu (3D) bioinformatičku analizu proteina RuBisCO velike podjedinice (rbcL) kod Sideritis ssp. koristeći razne bioinformatičke alate. Fizičko-hemijske analize su izvršene EkPASi – ProtParam-om. Pretpostavljena mesta fosforilacije rbcL proteina određena su pomoću NetPhos 2.0 i NetPhos 3.1. Filogenetska analiza urađena je MEGA 6.0 softverom. Za procenu 3D proteinske strukture, korišćen je program PyMol. Na kraju istarživanja, utvrđeno je da broj amino kiselina stilbena sintetaza proteina varirao između 171 i 456, molekulska težina se kretala između 19002,67 i 50420,44 Da, indeks nestabilnosti se kretao između 27,30 do 40,70, a GRAVY vrednosti su se kretale između -0,394 do - 0.226. Dok je najveća prosečna stopa aminokiselina u proteinima rbcL bila Gli (10.00%), najniži odnos amino kiselina (1.4%) je određen kao Trp. U filogenetskim analizama izvršenim korišćenjem sekvenci proteina, stablo sa maksimalnom verovatnoćom (ML) se sastojalo od dva velika kladusa. Analiza distance parova zasnovana na Sideritis vrstama rbcL proteinskih sekvenci izvedena je korišćenjem MEGA 6.0. Najmanja udaljenost para bila je 0.000, dok je najveća udaljenost iznosila 0.024. Kada su procenjene 3D strukture proteina ispitane, Gly ostatak, koji igra važnu ulogu u strukturi i funkciji proteina, najmanje je bio prisutan u S. sipilea vrstama dok je ostatak bio najzastupljeniji u S. cretica subsp. spicata vrstama. Rezultati ovog istraživanja pružaju uvid u osnovne karakteristike rbcL proteina u Sideritis taksonima.

Primljeno 13.II.2018.
Odobreno 18. XI. 2018.