Plant Omics Journal POJ

POJ 8(1):55-61 (2015) ISSN:1836-3644

Comparative phylogenetic analysis of phenylpropanoid metabolism genes of legume plants

Neslihan Turgut-Kara*, Özgür Çakır

Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, 34118, Vezneciler, Istanbul, Turkey

***Corresponding author: neslihantk@istanbul.edu.tr**

Abstract

Plant phenylpropanoids contribute various physiological functions in accordance with environmental influences; therefore, most of secondary metabolites are synthesized through phenylpropanoid pathway. In this study, National Center for Biotechnology Information (NCBI) was searched to collect protein sequences that encode legume phenylpropanoid metabolism enzymes homologues. A total of 95 phenylpropanoid metabolism enzymes sequences from several legume species were phyletically analysed to light the way for the evolution characteristics of legume-specific homologues. One of the main emphases of this study was the elucidation of the conserved sequences of phenylpropanoid enzymes, and designing a set of quantitative PCR primers to standardize gene expression experiment for different legumes at once. As a result of the analyses, conserved sequences of phenylpropanoid enzymes were determined, and the sets of real-time PCR primers were generated for 5 main phenylpropanoid genes, phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI). This study will also assist in furthering our understanding of the evolutionary relation of phenylpropanoids between legume species.

Keywords: phenylpropanoid; leguminosae; phylogenetic; qPCR

Introduction

Phenylpropanoids consist of a large group of natural products, which are synthesized by the enzymes in the phenylpropanoid pathway (Holton and Maizeish 1995; Dixon and Steele 1999, Zhang and Liu 2014), and it is one of the most studied biochemical pathways. This pathway originates from the phenylalanine and ends up with the synthesis of a large class of phytochemicals. Phenylalanine ammonia-lyase (PAL) catalyzes the first step of this pathway, it converts the phenylalanine into trans-cinnamic acid. The second step, is the conversion of trans-cinnamic acid to p-coumaric acid by the cinnamate 4-hydroxylase. The last step of the general part of this pathway is the formation of p-coumaroyl CoA from p-coumaric acid and it is catalyzed by 4-coumarate: CoA ligase This compound is the substrate for several branches so the pathway leading to biosynthesis of lignin/lignans, flavonoid/anthocyanins and isoflavonoids (Winkel-Shirley 2001; Dixon 2005; Dixon et al., 2005; Vogt 2010). For biosynthesis of economically and biologically important secondary metabolites flavonoids, the pathway continues with early biosynthetic genes, such as *CHS*, *CHI* (Fig. 1).

Soybean *(Glycine max)*, common bean *(Phaseolus vulgaris* L.), alfalfa *(Medicago sativa)*, and chickpea *(Cicer arietinum)* belong to the Papilionoid subfamily of legumes. All these species are grown for their seeds and protein, oil and starch they contain. Thus, their consumption by human and animals are crucial, because of their high protein, complex carbohydrate, micronutrients, vitamins and dietary fiber content. Also, most of the legume species are rich in phenylpropanoid pathway derivated bioactive secondary metabolites such as flavonoids (Reinprecht et al., 2013). Major areas of future research on legume improvement include developing and improving the content and composition of phenylpropanoids, especially flavonoids. In this respect, to understand exact relationships between legume seed quality, phenypropanoid content and composition**,** more predictable approaches are necessary. For these purposes, using the computational tools could enable to the researchers to understand physical and biochemical properties of proteins (Sivakumar et al., 2007). In this study, we elucidated the conserved sequences of flavonoid targeted phenylpropanoid enzymes for selected species of legumes, and designed a set of quantitative PCR primers by using computational tools to standardize gene expression for different legumes in terms of phenylpropanoid metabolism at once.

Results and Discussion

In this study, five kinds of enzyme amino acid sequences belonging to phenylpropanoid pathway were obtained from NCBI Genbank and analyzed. Homology searches were applied to find conserved amino acid motifs between phenylpropanoid pathway enzymes using the MEME SuiteMotif-based sequence analysis tool. These regions thought to be important for biological functions such as enzyme catalytic sites or prostethic attachment sites.

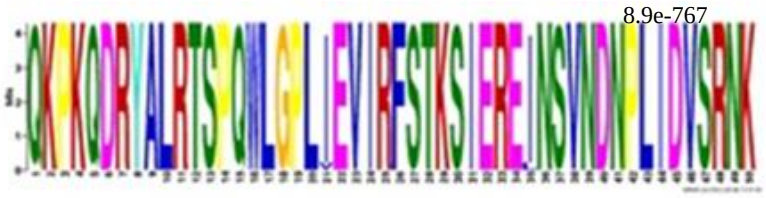
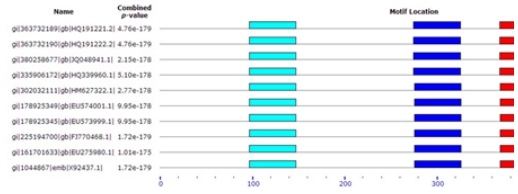
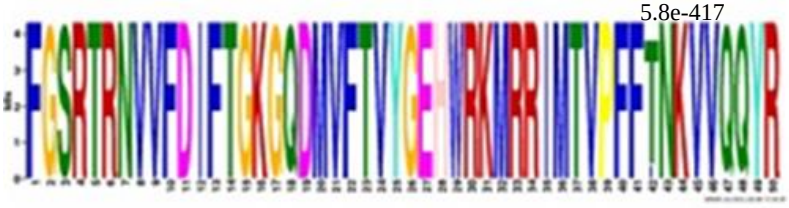
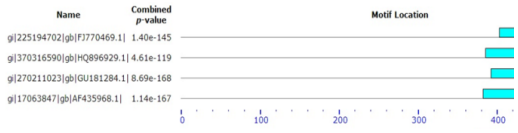
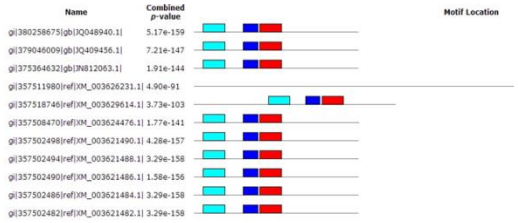
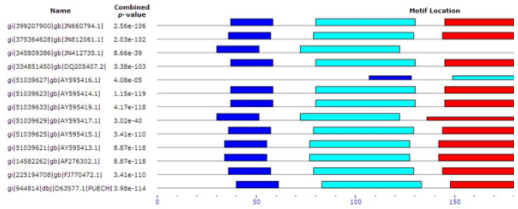
A total of 95 phenylpropanoid-pathway enzyme's amino acid sequences were obtained from different legume plant species. The majority of these sequences belonged to two model leguminous species, soybean and alfalfa. As a result of

Table 1. List of used legume protein sequences and their lengths.

Enzyme Name	Species and accession numbers	Count of Sequence	Shortest Sequence	Longest Sequence	Average Length
PAL	*Acacia mangium x Acacia auriculiformis (DQ377806)* *Astragalus membranaceus (EF567076)* *Astragalus mongholicus (AY986506)* *Glycine max (GQ220305, NM001250027)* *Glycine soja (FJ483834)* *Lotus japonicus (AB283031, AB283035-6, AB283040)* *Phaseolus vulgaris (M11939)* *Pisum sativum (D10001)* *Robinia pseudoacacia (EU650628)* *Stylosanthes humilis (L36822)* *Trifolium pratense (DQ073808-10)*	17	505	732	705,1
C4H	*Acacia auriculiformis x Acacia mangium (EU275980)* *Astragalus membranaceus (JQ048941)* *Astragalus mongholicus (HQ339960)* *Glycine max (FJ770468, X92437* *Leucaena leucocephala (HQ191221-2)* *Medicago truncatula (HM627322)* *Trifolium pratense (EU573999, EU574001)*	10	505	553	515,1
4CL	*Amorpha fruticosa (AF435968)* *Galega orientalis (GU181284)* *Glycine max (FJ770469, HQ896929)*	4	540	562	548
CHS	*Acacia confusa (JN812063)* *Arachis hypogaea (AY192572)* *Astragalus membranaceus (JQ048940, HQ840676-8)* *Caragana arborescens (HQ840679)* *Clitoria ternatea (AB185897)* *Glycine max (FJ770471)* *Glycyrrhiza inflata (EU706287)* *Glycyrrhiza uralensis (EF026979, HQ840673-5)* *Lupinus albus (EF640320)* *Lupinus luteus (AY289103-4,DQ507388-9)* *Medicago sativa (L029015, U01018, U01021)* *Medicago truncatula (XM003591973-4, XM003591996-7, XM003595597, XM003601599, XM003601600, XM003601819, XM003607450, XM003610803-5, XM003610807-8, XM003621482, XM003621484, XM003621486, XM003621488, XM003621490, XM003624476, XM003626231,XM003629614)* *Onobrychis viciifolia (AF026258)* *Pueraria candollei (JQ409456)* *Pueraria montana (D10223)* *Senna alata (AF358430-2* *Senna tora (EU430077)*	51	269	1317	419
CHI	*Acacia confusa (JN812061)* *Arachis hypogaea (JN412735, JN660794)* *Astragalus mongholicus (DQ205407)* *Glycine max (AF276302, AY595413-7, AY595419, FJ770472)* *Pueraria lobata (D63577)*	13	209	281	228,9

this study conserved motif number, locations (Table 2), and phylogenetic trees (Fig. 2) of the investigated proteins were determined. Also, the designing of qPCR primers for five phenylpropanoid genes were made by submitting the conserved amino acid sequences to Conserved Primers 2.0 tool (Table 3).Comparative analysis of the complete amino acid sequences of the phenylpropanoid pathway proteins done by using BLOCKS and MEME software are in agreement with those of the presented phylogenetic analysis, since several family specific conserved motifs. These analysis were suggested that all tested enzyme proteins have strong conservation for the motifs but the most conserved one was C4H (Table 2).Neighbor-Joining method was subjected to infer the evolutionary history (Saitou and Nei 1987). The optimal tree with the sum of branch lengths were on the Figure 2. The trees were scaled by using the branch lengths in the same units as the evolutionary distances. Poisson correction method was applied to compute the evolutionary distances (Zuckerkandl and Pauling 1965) and they are in the units of the number of amino acid substitutions per site. PAL analysis involved 17 amino acid sequences and a total of 504 positions in the final dataset. C4H analysis involved 10 amino acid sequences and a total of 444 positions in the final dataset. 4CL analysis involved 4 amino acid sequences and a

Table 2. Combined block diagrams on first column and MEME suite motif logos of conserved motifs identified in legume phenylpropanoids. The height of a letter indicates its relative frequency at the given position (x-axis) in the motifs. The rightmost column shows the *p*-value of the motif calculated against random background sequences.

Enzyme name	Combined block diagrams	Discovered motif	*p*-value
PAL			8.9e-767
C4H			5.8e-417
4CL			9.3e-073
CHS			2.9e-2523
CHI			4.1e-256

total of 540 positions in the final dataset. CHS analysis involved 55 amino acid sequences and a total of 268 positions in the final dataset. CHI analysis involved 13 amino acid sequences and a total of 209 positions in the final dataset. Totally, the gaps and missing data were screened out for all positions. MEGA6 was used to perform evolutionary analyses (Tamura et al., 2013). According to neighbor-joining algorithim groups and subgroups were defined for per proteins phylogenetic tree (Fig. 2). Convergence and divergence are two essential phylogenetic properities, which can be useful to identify the closely as well as distantly related group of legume phenylpropanoid proteins. This finding suggests that all tested enzymes are conserved and all of the analyzed sequences evolved from a common ancestral enzyme. In conclusion, *in silico* sequence analysis of phenylpropanoid proteins showed that these higher plants have been related together evolutionarily and conserved regions. *In silico* analysis of these protein sequences would contribute to a better understanding of divergence of these proteins in other plants. For testing designed conserved qPCR primers for *PAL, C4H, 4CL, CHS, CHI* or *actin* qPCR reactions were carried out, and 55°C was found as the optimum and associate annealing temperature for all tested enzyme gene (Table 3 and Table 4). These analyses led to a classification of the phenylpropanoid enzyme amino acid sequences according to closeness degree.In conclusion, this study will help us about the understanding of the evolutionary features of the legume phenylpropanoids. In addition, phylogenetic and comparative analysis of legume phenylpropanoids would be useful to analyse flavonoid

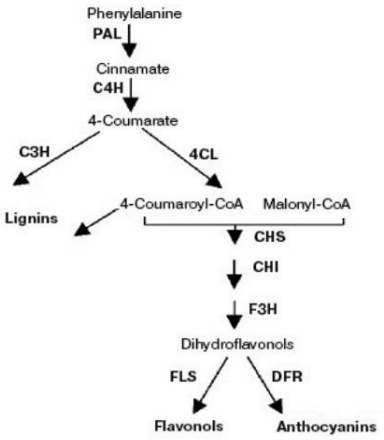
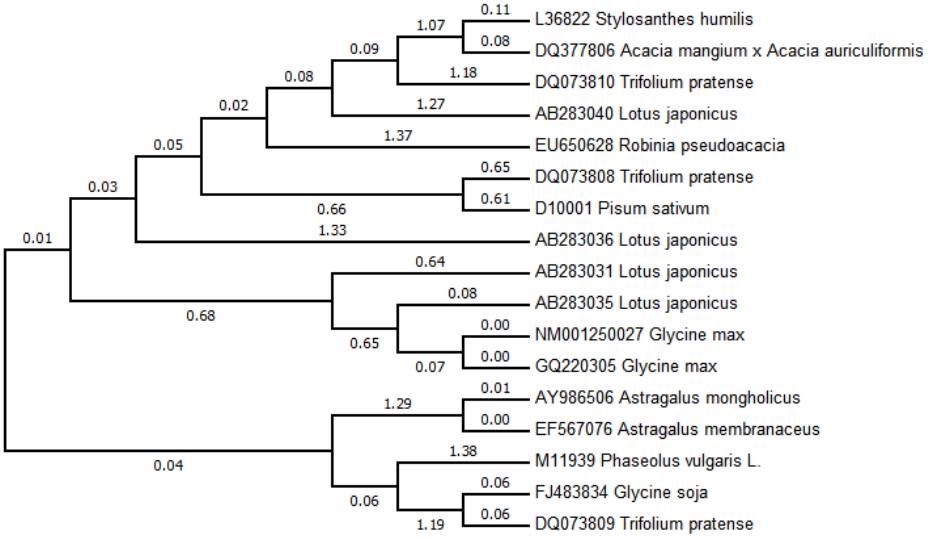


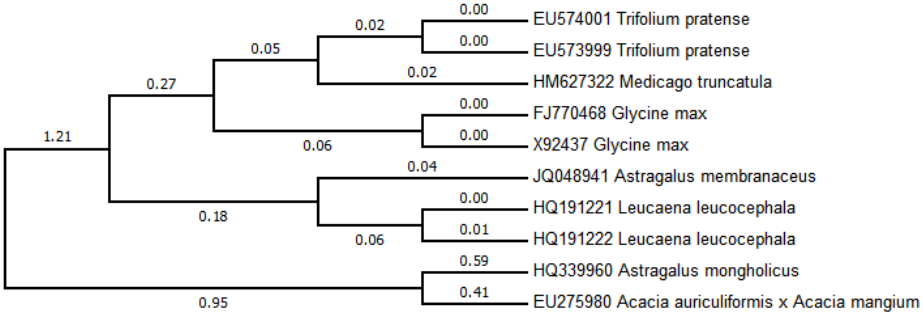
Fig 1. General phenylpropanoid pathway. PAL phenylalanine ammonia-lyase, C4H cinnamate 4-hydroxylase, C3H *p*coumarate 3-hydroxylase, 4CL 4-coumaroyl CoA ligase, CHS chalcone synthase, CHI chalcone isomerase, F3H flavanone-3-hydroxylase, DFR dihydroflavonol 4-reductase, FLS flavonol synthase (adopted and revised from Singh et al., 2009).

PAL



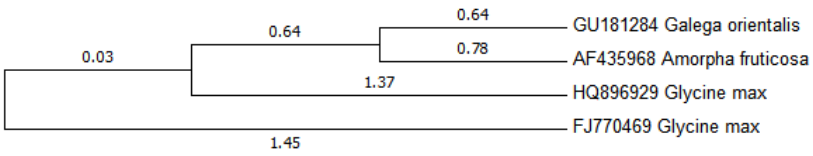
The sum of branch length = 0.58364015

C4H



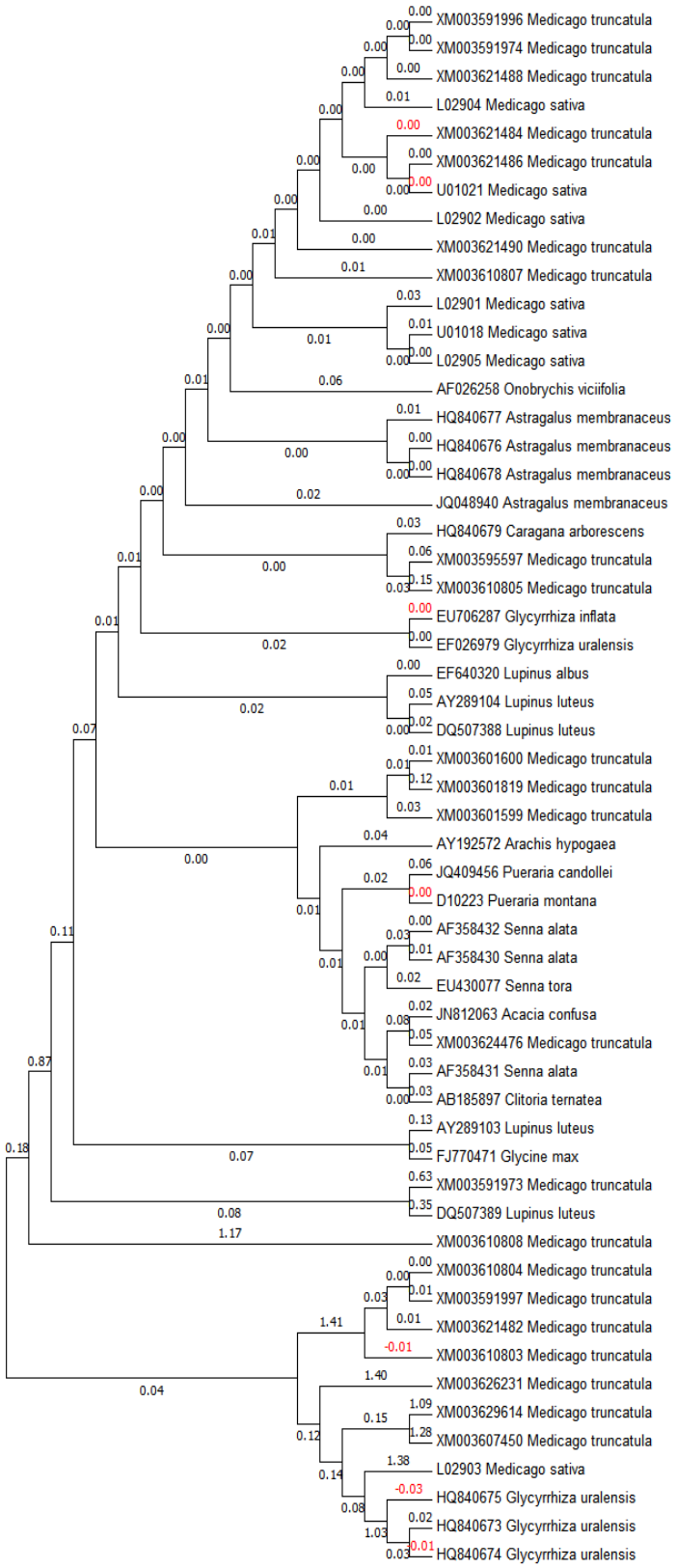
The sum of branch length = 3.86957703

4CL



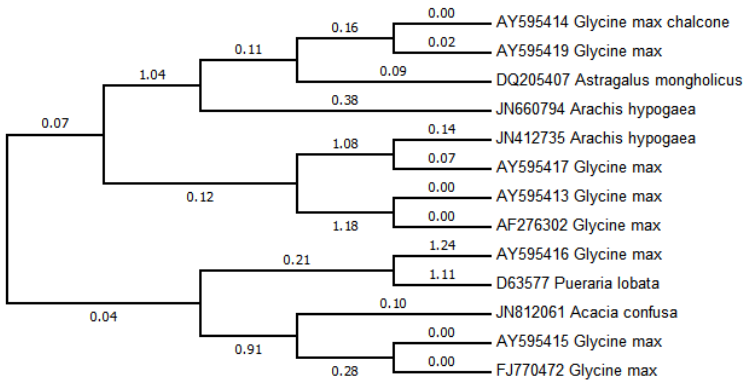
The sum of branch length = 4.91489840

CHS



The sum of branch length = 13.08881692

CHI



The sum of branch length = 8.35468415

Fig 2. Neighbor-joining method based phylogenetic trees for the phenylpropanoid enzymes of legume species (Trees were constructed according to conserved sites of the protein by using MEGA6).

Table 3. Sets of qPCR primers were generated for 5 phenylpropanoid genes and housekeeping gene actin (Tm for all primers 55°C).

Primer Name	Forward primer (F) Reverse primer (R)	Product size (bp)
Legume_*PAL*	F: 5'-GCAATGGCTTGGTCCTCTTA-3' R: 5'-CCATGCAAAGCCTTGTTTCT-3'	120
Legume_*C4H*	F: 5'-GCCAAAGAAGTCCTCCACAC-3' R: 5'-CCGTAGACGGTGAAGACCAT-3'	110
Legume_*4CL*	F: 5'-CGGCGTTACCATTCTCATC-3' R: 5'-ACTTGTTGTGCAACGCTTGT-3'	91
Legume_*CHS*	F: 5'-GGTGCACGTGTTCTTGTTGT-3' R: 5'-GCTTGGCCCACAAGACTATC-3'	92
Legume_*CHI*	F: 5'-TGGAGAAGGTGATTAGGCTTG-3' R: 5'-TTTGTCATCAGCAGCCAAAC-3'	104
Actin	F: 5'-TCAAGACGAAGGATG-3' R: 5'-TTGGATTCTGGTGAT-3'	100

Table 4. The trial qPCR test results of the designed qPCR primers and housekeeping gene actin at annealing temperature 55°C.

	Ct values ± SE			
	Soybean *(Glycine max)*	Common bean *(Phaseolus vulgaris* L.*)*	*Alfalfa (Medicago sativa)*	*Chickpea (Cicer arietinum)*
Actin	21.49±1.52	18.5±0.12	22.94±2.97	19.27±1.09
PAL	29.94±0.21	26.84±1.72	30.04±0.44	22.50±2.65
C4H	28.46±0.32	26.84±2.34	25.38±1.01	27.23±0.45
4CL	33.39±1.66	31.74±1.55	30.00±3.21	29.59±1.83
CHS	28.15±2.11	31.58±1.98	31.44±1.33	29.51±0.51
CHI	31.12±1.69	30.75±3.02	29.91±2.51	30.24±1.67

pathway gene expression in one time via the designed set of qPCR primers.

Materials and Methods

Retrieval of sequences and screening of the databases

Legume specific sequences were retrieved from publicly available online data resource; NCBI (http://www.ncbi.nih.gov). All sequences were downloaded in FASTA format and used for further analysis. First, all databases were screened for related phenylpropanoid gene. Then only Leguminosae family members CDS sequences were used for the homology analysis (Table 1). Blast algorithm was used for the initial screening (Altschul et al., 1990).

Multiple alignment, phylogenetic tree construction, and primer design

To discover the protein motifs, a total of 95 phenylpropanoid enzyme amino acid sequences were multiply aligned by utilizing the MEME SuiteMotif-based sequence analysis tool (Timothy et al., 2009). The following parameter settings were

used: distribution of motifs, zero or one per sequence; maximum number of motifs to find, 3; minimum width of motif 6; maximum width of motif, 50. After that, each gene group was multiply aligned within its own group, and motifs were determined and used for the construction of phylogenetic trees. The Neighbor-Joining trees (Saitou and Nei 1987) were constructed with MEGA6 (Tamura et al., 2013) program after the multiple alignment of each subfamily of genes. Primers were designed with Conserved Primers 2.0 tool (You et al., 2009).

Plant materials

Soybean *(Glycine max)*, common bean *(Phaseolus vulgaris* L.), alfalfa *(Medicago sativa)*, and chickpea *(Cicer arietinum)* seeds were germinated in perlite. After 1 week, first leaves of these plantlets were harvested, immediately frozen by liquid nitrogen, and keep in -80°C until RNA isolation.

qPCR analysis

For testing designed conserved qPCR primers for *PAL*, *C4H*, *4CL*, *CHS*, *CHI*, and *actin* which were determined after bioinformatic analysis, qPCR reaction was carried out. For this experiment, total RNA was isolated using the Favorgen Plant RNA isolation kit (FAPGK001, Favorgen Biotech Corp., Kaohsiung, Taiwan, ROC), with the application of DNaseI (EN0521, Thermo Scientific, Fermentas Inc, Glen Burnie, MD, USA) treatment during the isolation process. The Fermentas kit (K1612, Glen Burnie,MD, USA) was used for the synthesis of first-strand cDNA, which served as a template for qPCR. qPCR was carried out with the DyNAmoe HS SYBR wGreen qPCR Kit (Thermo Scientific) following the manufacturer's instructions. Reaction mixtures consisting of 12.5ml SYBR Green master mix, 0.2mM each of forward and reverse primers and 3 µl cDNA were brought to 25ml with MQ water. The qPCR conditions were set as follows: 1 cycle of 10 min at 95°C, 40 cycles of 10 s at 94°C, 30 s at 55°C and 30 s at 72°C. The final step included a gradual temperature increase from 55°C to 95°C at the rate of 1°C/10 s to enable melting curve data collection. A non-template control was run alongside, and the reference (actin) and the target genes were included with every assay. Amplification specificity of each reaction was verified by melting curve analysis. Expressions were normalized against the reference gene. Ct values was determined according to the MX3000 Stratagene software (Agilent Technologies, Inc., Santa Clara,USA). Triplicate qPCR experiments were conducted for each sample.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgements

This study was financially supported by the Research Fund of the Istanbul University (Project numbers 35180 and 37866).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215: 403–410.

Dixon RA (2005) Engineering of plant natural product pathways. Curr Opin Plant Biol. 8: 329–336.

Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids—a gold mine for metabolic engineering. Trends Plant Sci 4: 394–400.

Dixon RA, Xie DY, Sharma SB (2005) Proanthocyanidins: a final fronter in flavonoid research. New Phytol 165: 9–28.

Holton TA, Maizeish EC (1995). Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell. 7: 1071–1083.

Reinprecht Y, Yadegari Z, Perry GE, Siddiqua M, Wright LC, McClean PE, Pauls KP (2013) *In silico* comparison of genomic regions containing genes coding for enzymes and transcription factors for the phenylpropanoid pathway in *Phaseolus vulgaris* L. and *Glycine max* L. Merr. Front Plant Sci. 4: 317.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4 (4): 406-425.

Singh K, Kumar S, Rani A, Gulati A, Ahuja P S (2009) Phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. Funct Integr Genom. 9:125–134.

Sivakumar K, Balaji S, Radhakrishnan G (2007) *In silico* characterization of antifreeze proteins using computational tools and servers. J Chem Sci. 119 (5): 571–579.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30: 2725-2729.

Timothy L, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37: 202–208.

Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant. 3: 2-20.

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126(2): 485-493.

You FM, Huo N, Gu YQ, Lazo GR, Dvorak J, Anderson OD (2009) ConservedPrimers 2.0: A high-throughput pipeline for comparative genome referenced intron-flanking PCR primer design and its application in wheat SNP discovery. BMC Bioinformatics. 10: 331.

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving Genes and Proteins, Academic Press, New York, pp. 97-166.

Zhang X, Liu CJ (2014) Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. Mol Plant. 134 [Epub in press]