International Journal of Science, Environment and Technology, Vol. 5, No 4, 2016, 1968 – 1979

ISSN 2278-3687 (O)
2277-663X (P)

IN SILICO ANALYSIS OF CHALCONE SYNTHASE 1 PROTEIN SEQUENCES FROM DIFFERENT PLANT SPECIES

E. Ramprasad[1*], G. Rakesh[2], Ch. V. Durga Rani[3], S. Vanisri[4] and MNV Prasad Gajula[5]

[1, 3, 4&5]Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad-500030, India

[2]Department of Genetics and Plant Breeding, PJTSAU, Hyderabad-500030, India

E-mail: rambiotech100@gmail.com (**Corresponding Author*)

Abstract: Chalcone synthase 1 protein is the most well-known representative of plant polyketide synthase superfamily. It provides the starting materials for a diverse set of metabolites (flavonoids) which have different and important roles in flowering plants, such as providing floral pigments like anthocyanin, antibiotics, UV protectants and insect repellents. A total of 16 full lengths of chalcone synthase 1 protein sequences from different plant species available in uniport were evaluated by using bioinformatics tools to investigate physico-chemical properties, secondary structure prediction, putative phosphorylation sites, conserved motif search and phylogenetic tree construction. Physicochemical analysis offers data such as pI, EC, Al, GRAVY and II about these sequences and the results showed that all chalcone synthase 1 protein sequences are acidic, hydrophilic, thermo stable, having some intracellular portion. The secondary structure of the protein sequences were also predicted using SOPMA server. It was observed that alpha helix was predominant, followed by random coil, extended strand and least beta turn was found. Putative phosphorylation sites were also identified which are found to be conserved in plant species and the results showed that the most abundant phosphorylation site is serine residues in chalcone synthase 1 protein sequences. Conserved protein motifs subjected to MEME to obtain the best possible matches. The phylogenetic tree represented three major clusters and chalcone synthase 1 protein sequences of plant species belongs to same family clustered together. The obtained results could be used for further *in silico* analysis and homology modeling studies.

Keywords: chalcone synthase 1 protein, polyketide synthase, *in silico,* and homology modeling.

Abbreviations:

pI	:	Isoelectric point
EC	:	Extinction coefficient
Al	:	Aliphatic index
GRAVY	:	Grand average of hydropathy
II	:	Instability index

Introduction

In nature plants are exposed to a variety of biotic and abiotic stresses. Viruses, bacteria, fungi, nematodes and other pests attacking plants are biotic stresses, while light, temperature, wounding, drought, etc. are abiotic stresses. During stress conditions plant is

*Received June 10, 2016 * Published Aug 2, 2016 * www.ijset.net*

expressing a number of genes as part of its defense. Among which, CHS (chalcone synthase) is quite commonly induced gene under different forms of stresses like UV, wounding, herbivory and microbial pathogens. It results in the production of some of the compounds which are having antimicrobial activity (phytoalexins), insecticidal activity and antioxidant activity or quench UV light directly or indirectly (1). CHS expression is also involved in the salicyclic acid defense pathway (2). In recent studies there is growing evidence that chalcone synthase is closely related to other plant-specific polyketide synthases, including stilbene synthase (STS) (3,4), acridone synthase (ACS) (5), bibenzyl synthase (BBS) (6), 2-pyrone synthase (2PS) (7) and phlorisovalerophenone synthase (PVPS) (8). It is also observed that even very few amino acid changes in the chalcone synthase proteins may result in shifts in enzyme function. So it is very important to study the structural and functional properties of these enzymes in order to know the changes that are happened in these enzymes in due course of evolution or during speciation. In this context, the present study was undertaken to study the structural and functional properties of CHS enzymes by taking chalcone synthase 1 protein sequences for analysis using bio-computational tools.

Computational tools provide an opportunity to researchers for understanding the physico-chemical and structural properties of proteins that can be obtained from different sources (9). These tools analyze protein sequences for number of amino acid, sequence length and the physico-chemical properties of proteins such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, prediction and characterization of protein structure. Phylogenetic analysis is also powerful tool for addressing many different evolutionary questions (10). The identification of one gene or one protein and its homologs in several species indicates that these genes or proteins diverged from their common ancestor. This can aid in our understanding of the evolution of species and could serve to develop model systems for studying gene functions in the future. In this paper, *in silico* analysis and characterization studies on 16 chalcone synthase 1 protein sequences of different plant species were studied.

Material and Methods

Sequence retrieval

Expasy (uniprot KB) that provides protein sequences and annotation data (11) was used to retrieve the chalcone synthase 1 protein sequences. These were downloaded in FASTA format to be used for further analysis (http://www.uniprot.org).

Physio-chemical characterization

For Physio-chemical characterization, theoretical Isoelectric Point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average of hydropathy (GRAVY) were computed using the Expasy Protparm server (12) (http://us.expasy.org/tools/protparam.html).

Secondary structure prediction

SOPMA tool (Self-Optimized Prediction Method with Alignment) (13) was applied to extract the information regarding the secondary structures that consist of Alpha helix, Extended strand, Beta turn and Random coil.

Putative phosphorylation sites and motif search

The amino acid sequences of the selected plants were analyzed for the putative phosphorylation sites at the NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/) (14). Analysis of domain and conserved protein motifs was performed using MEME (http://meme.sdsc.edu/meme/meme.html) (15).

Phylogenetic analysis

For phylogenetic analysis, multiple sequence alignment (16) was performed by using ClustalW2 sequence alignment program and the phylogenetic tree was constructed by using the Neighbor- Joining (NJ) method in tree view software (Fig. 1).

Results and discussion:

Chalcone synthase 1 protein sequences of 16 plant species (Table 1) were analyzed in this study and corresponding protein sequences were collected from Uniport (http://www.uniprot.org/). Physio-chemical properties were examined to find differences between seventeen chalcone synthase 1 protein sequences using Expasy's ProtParam tool (Table 2). The isoelectronic point is the pH at which the protein does not migrate in an electric field. It plays an important role in protein purification. The computed pI value that was less than 7 (pI<7) indicates that proteins were considered as acidic or greater than 7 (pI>7) reveals that proteins were basic in character. The pI value of all the sequences under study having less than 7 (pI<7) reveals that these proteins were acidic in nature. Total numbers of negatively charged residues are higher than the total number of positively charged residues implies that these proteins are having intracellular portion. The extinction coefficient (EC) indicates that the amount of light being absorbed proteins at a certain wavelength. Extinction coefficient of chalcone synthase 1 protein sequences at 280 nm was ranged from 31775 to 38765 $M^{-1}cm^{-1}$. The high extinction coefficient of chalcone synthase 1

protein sequences (CHS1_GERHY, CHS1_SOYBN CHS1_RUTGR and CHS1_TRISU) indicates the presence of high concentration of cysteine, tryptophan and tyrosine. These amino acids (cysteine, tryptophan and tyrosine) are considered to be an important parameter in the calculation of extinction coefficient of proteins (17). The instability index is used to determine whether it will be stable in a test tube. If the index is less than 40, it is probably stable in the test tube. If the value is greater than 40, it is probably not stable (18). The instability index value for the Chalcone synthase 1 protein sequences were found to be ranging from 33.49 to 42.09. The results imply of present study revealed that most of the protein sequences under study are stable (values less than 40) except CHS1_ORYSJ, CHS1_ORYSI, CHS1_CAMSI, CHS1_GERHY, CHS1_CICAR and CHS1_MEDSA are unstable proteins (values more than 40) (Table 2). The aliphatic index of a protein is a measure of the relative volume occupied by aliphatic side chain of the following amino acids *viz.,* alanine, valine, leucine and isoleucine. The aliphatic index values of chalcone synthase 1 protein sequences ranging from 85.84 to 92.85. The very high aliphatic index of all chalcone synthase 1 protein sequences supports the view that chalcone synthase 1 protein sequences may be stable for a wide range of temperatures. The GRAVY value for protein is calculated as the sum of hydropathy values of all the amino acids. A hydropathy scale which is based on the hydrophobic and hydrophilic properties of the 20 amino acids is used in the study (19). GRAVY values of chalcone synthase 1 protein sequences were ranged from -0.040 to -0.117. The very low GRAVY index of chalcone synthase 1 protein sequences CHS1_CITSI, CHS1_DAUCA and CHS1_SOYBN implies that these chalcone synthase 1 protein sequences could result in a better interaction with water and their negative GRAVY score indicate that they are hydrophilic in nature.

The secondary structure of the protein sequences were predicted using SOPMA server (Table 3). It was observed that alpha helix was predominant, followed by random coil, extended strand and least beta turn was found.

Using the NetPhos 2.0 Server the putative phosphorylation sites were identified for different plant species (Table 4). The output score was given in 0.000-1.000 range and the score above the threshold (0.500) shows the confidence rate of true phosphorylation site by the server. Several putative phosphorylation sites are completely conserved in plant species and interestingly more phosphorylation sites were found in CHS1_ORYSJ, CHS1_ORYSI, CHS1_PEA, CHS1_SECCE, CHS1_SOLLC and CHS1_RUTGR.

The evolutionary relationships between the plants were evaluated by phylogenetic analysis of the aligned amino acids sequence of Chalcone synthase 1 protein sequences with neighbor-joining (NJ) method (Fig. 1). Similar studies were conducted by (20, 21 and 19) to address the questions related to evolution. Convergence and divergence are two essential phylogenetic properties, which can be useful to identify the closely as well as distantly related group containing plant chalcone synthase 1 protein sequences. The minimum degree of divergence was found between *Oryza sativa* subsp. *japonica* (Rice) and *Oryza sativa* subsp. *indica* (Rice), while the maximum degree of divergence was found between *Oryza sativa* subsp. *japonica* (Rice) and *Daucus carota* (Wild carrot). The phylogenetic tree showed that there are three major clusters and same family plant chalcone synthase 1 protein sequences are grouped together like *Oryza sativa* subsp. *japonica* (Rice), *Oryza sativa* subsp. *indica* (Rice), *Hordeum vulgare* (Barley), *Secale cereale* (Rye) and *Sorghum bicolor* (Sorghum) (*Sorghum vulgare*). This finding suggests that Chalcone synthase 1 protein sequences are conserved and they are evolved from a common ancestor. In conclusion, *in silico* sequence analysis of chalcone synthase 1 protein sequences showed that these sequences taken from different organisms related together evolutionarily as they pocess conserved regions in their protein sequences. These findings will be helpful to further study on chalcone synthase 1 protein functions at molecular or structural levels and also useful in homology modeling studies.

Acknowledgements

We express our gratitude to the Department of Biotechnology, Government of India for providing fellowship to the seniour author and to the Professor Jayashankar Telangana State Agricultural University, Hyderabad for providing the opportunity to work on the above topic.

References

[1] Dao, T.T.H., Linthorst, H.J.M and Verpoorte, R. Chalcone synthase and its functions in plant resistance. Phytochem Rev. 2011, 10: 397–412.

[2] Tohge T, Yonekura-Sakakibara K, Niida R, Wantanabe-Takahasi A, Saito K. "Phytochemical genomics in Arabidopsis thaliana: A case study for functional identification of flavonoid biosynthesis genes". Pure and Applied Chemistry. 2007, 79 (4): 811–23.

[3] Schoppner A, Kindl H. Purification and properties of a stilbene synthase frominduced cell suspension cultures of peanut. J Biol Chem. 1984, 259:6806 –6811.

[4] Schroder J. A family of plant-specific polyketide synthases: Facts and predictions. Trends Plant Sci. 1997, 2:373–378.

[5] Lukacin R, Springob K, Urbanke C, Ernwein C, Schro¨ der G, Schro¨ der J, Matern U Native acridone synthase I and II from Ruta graveolens L. form homodimers. FEBS Lett. 1999, 448:135–140.

[6] Preisig-Muller R, Gnau P, Kindl H The inducible 9, 10-dihydrophenanthrene pathway: Characterization and expression of bibenzyl synthase and S-adenosylhomocysteine hydrolase. Arch Biochem Biophys. 1995, 317:201–207.

[7] Eckermann S, Schroder G, Schmidt J, Strack D, Edrada RA, Helariutta Y, Elomaa P, Kotilainen I, Proksch P, Teeri TH, Schroder J. New pathway to polyketides in plants. Nature. 1998, 396:390–397.

[8] Paniego NB, Zuurbier KWM, Fung SY, Van der Heijden R, Scheffer JJC, Verpoorte R Phlorisovalerophenone synthase, a novel polyketide synthase fromhop (*Humulus lupulus* L.) cones. Eur J Biochem. 1999, 262:612–616.

[9] Sivakumar K., Balaji S. and Radhakrishnan G. In silico characterization of antifreeze proteins using computational tools and servers. The Journal of Chemical Sciences, 2007, Vol. 119, No. 5: 571–579.

[10] Elliott, M.B., D.M. Irwin and E.P. Diamandis, In silico identification and Bayesian phylogenetic analysis of multiple new mammalian kallikrein gene families. Genomics, 2006, 88: 591-599.

[11] Jain E, A Bairoch, S Duvaud, I Phan, N Redaschi, BE Suzek, MJ Martin, P, McGarvey, E Gasteiger. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics, 2009, 10.

[12] Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch. Protein Identification and Analysis Tools on the ExPASy Server, (In) John M.Walker (ed): The Proteomics Protocols Handbook, Humana Press. 2005, 571-607.

[13] Geourjon C, G Deleage. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 1995, 11, pp. 681-684.

[14] Blom N., Gammeltoft S. and Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Biology, 1999, Vol: 294 (5) pp: 1351-1362.

[15] Timothy L. Bailey and Charles Elkan. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, 1994, pp. 28-36, AAAI Press, Menlo Park, California,.

[16] Higgins, D.G., Thompson, J.D and Gibson, T.J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymology*. 1996, 266: 383-402.

[17] Kumar N. and Bhalla T.C. In silico analysis of amino acid sequences in relation to specificity and physiochemical properties of some aliphatic amidases and kynurenine formamidases. Journal of Bioinformatics and Sequence Analysis, 2011, Vol. 3(6), pp. 116-123.

[18] Guruprasad K., Reddy B.V.P. and Pandit M.W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering Design and Selection, 1990, 4: 155-164.

[19] Filiz, E and Koc, I. In silico analysis of dicer-like protein (DCLs) sequences from higher plant species. IUFS Journal of Biology. 2013, 72(1):53-63.

[20] Zhang H., Kolb F.A., Jaskiewicz L., Westhof E. and Filipowicz W. Single processing center models for human Dicer and bacterial RNase III. Cell, 2004, 118: 57–68.

[21] Liu Q., Feng Y. and Zhu Z. Dicer-like (DCL) proteins in plants. Functional & Integrative Genomics, 2009, 9: 277–286.

Table 1 Details of chalcone synthase 1 protein sequences from different species

Entry	Entry name	Organism	Length (No. of A.A)
Q2R3A1	CHS1_ORYSJ	*Oryza sativa* subsp. *japonica* (Rice)	398
A2ZEX7	CHS1_ORYSI	*Oryza sativa* subsp. *indica* (Rice)	398
Q01286	CHS1_PEA	*Pisum sativum* (Garden pea)	389
P13416	CHS1_SINAL	*Sinapis alba* (White mustard) (*Brassica hirta*)	395
Q9XJ58	CHS1_CITSI	*Citrus sinensis* (Sweet orange) (*Citrus aurantium* var. *sinensis*)	389
P48386	CHS1_CAMSI	*Camellia sinensis* (Tea)	389
P48390	CHS1_GERHY	*Gerbera hybrida* (Daisy)	398
P26018	CHS1_HORVU	*Hordeum vulgare* (Barley)	398
Q9SML4	CHS1_CICAR	*Cicer arietinum* (Chickpea) (Garbanzo)	389
Q9ZS41	CHS1_DAUCA	*Daucus carota* (Wild carrot)	389
P30073	CHS1_MEDSA	*Medicago sativa* (Alfalfa)	389
P53414	CHS1_SECCE	*Secale cereale* (Rye)	392
P23418	CHS1_SOLLC	*Solanum lycopersicum* (Tomato) (*Lycopersicon esculentum*)	389
P24826	CHS1_SOYBN	*Glycine max* (Soybean) (*Glycine hispida*)	388
Q9XGX2	CHS1_SORBI	*Sorghum bicolor* (Sorghum) (*Sorghum vulgare*)	401
Q9FSB9	CHS1_RUTGR	*Ruta graveolens* (Common rue)	393
P51083	CHS1_TRISU	*Trifolium subterraneum* (Subterranean clover)	389

Table 2 Details of Physiochemical Properties of chalcone synthase 1 protein sequences from different species

Entry	Entry name	Length (No. of A.A)	M.wt	pl	-R	+R	EC	II	Al	GRAVY
Q2R3A1	CHS1_ORYSJ	398	43264.6	5.85	48	41	35785	41.61	86.78	-0.076
A2ZEX7	CHS1_ORYSI	398	43237.6	5.85	48	41	35785	42.09	86.78	-0.070
Q01286	CHS1_PEA	389	42547.1	5.97	48	43	35785	34.04	90.51	-0.088
P13416	CHS1_SINAL	395	42570.2	5.97	49	44	35785	35.90	92.75	-0.088
Q9XJ58	CHS1_CITSI	389	43494.3	6.38	46	43	37275	35.01	90.90	-0.040
P48386	CHS1_CAMSI	389	43532.1	6.08	48	42	35785	40.10	87.76	-0.094
P48390	CHS1_GERHY	398	42873.5	6.05	48	44	38765	40.58	92.49	-0.084
P26018	CHS1_HORVU	398	42707.2	6.72	46	45	32805	34.12	91.98	-0.101
Q9SML4	CHS1_CICAR	389	42766.5	6.12	47	43	38765	41.21	91.52	-0.063
Q9ZS41	CHS1_DAUCA	389	42889.6	6.42	48	46	37400	36.40	92.85	-0.053
P30073	CHS1_MEDSA	389	43005.5	5.97	49	43	35785	40.58	85.84	-0.117
P53414	CHS1_SECCE	392	42993.5	6.03	50	44	32805	33.49	91.59	-0.075
P23418	CHS1_SOLLC	389	42552.3	6.72	45	44	35910	39.38	94.24	-0.060
P24826	CHS1_SOYBN	388	42516.3	6.21	45	42	38765	38.86	91.96	-0.055
Q9XGX2	CHS1_SORBI	401	43745.2	6.23	48	44	31775	36.57	86.11	-0.115
Q9FSB9	CHS1_RUTGR	393	42802.5	6.22	46	43	38765	38.53	90.98	-0.094
P51083	CHS1_TRISU	389	42741.4	5.90	48	43	38765	37.15	91.75	-0.083

Table 3 Details of secondary structures of chalcone synthase 1 protein sequences from different species

Entry	Entry name	Alpha helix	Extended strand	Beta turn	Random coil
Q2R3A1	CHS1_ORYSJ	43.72%	18.34%	12.56%	25.38%
A2ZEX7	CHS1_ORYSI	43.97%	18.09%	12.56%	25.38%
Q01286	CHS1_PEA	41.39%	17.99%	10.80%	29.82%
P13416	CHS1_SINAL	41.65%	17.48%	10.80%	30.08%

Q9XJ58	CHS1_CITSI	38.44%	20.35%	11.06%	30.15%
P48386	CHS1_CAMSI	42.21%	18.34%	12.06%	27.39%
P48390	CHS1_GERHY	39.33%	19.54%	10.54%	30.59%
P26018	CHS1_HORVU	41.13%	19.79%	12.08%	26.99%
Q9SML4	CHS1_CICAR	42.16%	18.77%	9.25%	29.82%
Q9ZS41	CHS1_DAUCA	42.49%	18.32%	10.69%	28.50%
P30073	CHS1_MEDSA	43.11%	18.11%	10.97%	27.81%
P53414	CHS1_SECCE	39.49%	18.73%	12.91%	28.86%
P23418	CHS1_SOLLC	39.85%	18.51%	11.31%	30.33%
P24826	CHS1_SOYBN	39.18%	19.33%	11.08%	30.41%
Q9XGX2	CHS1_SORBI	41.15%	18.45%	10.47%	29.93%
Q9FSB9	CHS1_RUTGR	37.28%	19.54%	11.83%	31.36%
P51083	CHS1_TRISU	41.13%	18.51%	10.03%	30.33%

Table 4 Putative phosphorylation residues in chalcone synthase 1 protein sequences from different species

Entry	**Entry name**	**Putative phosphorylation residues**		
		Serine	**Threonine**	**Tyrosine**
Q2R3A1	CHS1_ORYSJ	10	8	2
A2ZEX7	CHS1_ORYSI	12	7	1
Q01286	CHS1_PEA	10	8	2
P13416	CHS1_SINAL	8	6	2

Q9XJ58	CHS1_CITSI	9	6	2
P48386	CHS1_CAMSI	6	3	1
P48390	CHS1_GERHY	8	5	2
P26018	CHS1_HORVU	9	7	2
Q9SML4	CHS1_CICAR	8	4	1
Q9ZS41	CHS1_DAUCA	9	6	2
P30073	CHS1_MEDSA	10	5	1
P53414	CHS1_SECCE	10	8	2
P23418	CHS1_SOLLC	8	9	3
P24826	CHS1_SOYBN	7	5	1
Q9XGX2	CHS1_SORBI	10	6	2
Q9FSB9	CHS1_RUTGR	11	7	2
P51083	CHS1_TRISU	8	8	2

Table 5 Different motifs commonly observed in chalcone synthase 1 protein sequences with best possible match amino acid sequences.

MOTIF	WIDTH	BEST POSSIBLE MATCH
1	50	EWGQPKSKITHLIFCTTSGVDMPGADYQLTKMLGLRPSVKRYMMYQQGCF
2	50	DYYFRITNSEHMTELKEKFKRMCDKSMIRKRYMHLTEEILKENPNMCAYM
3	50	WNSIFWIAHPGGPAILDQVEAKLGLKPEKMRATRHVLSEYGNMSSACVLF

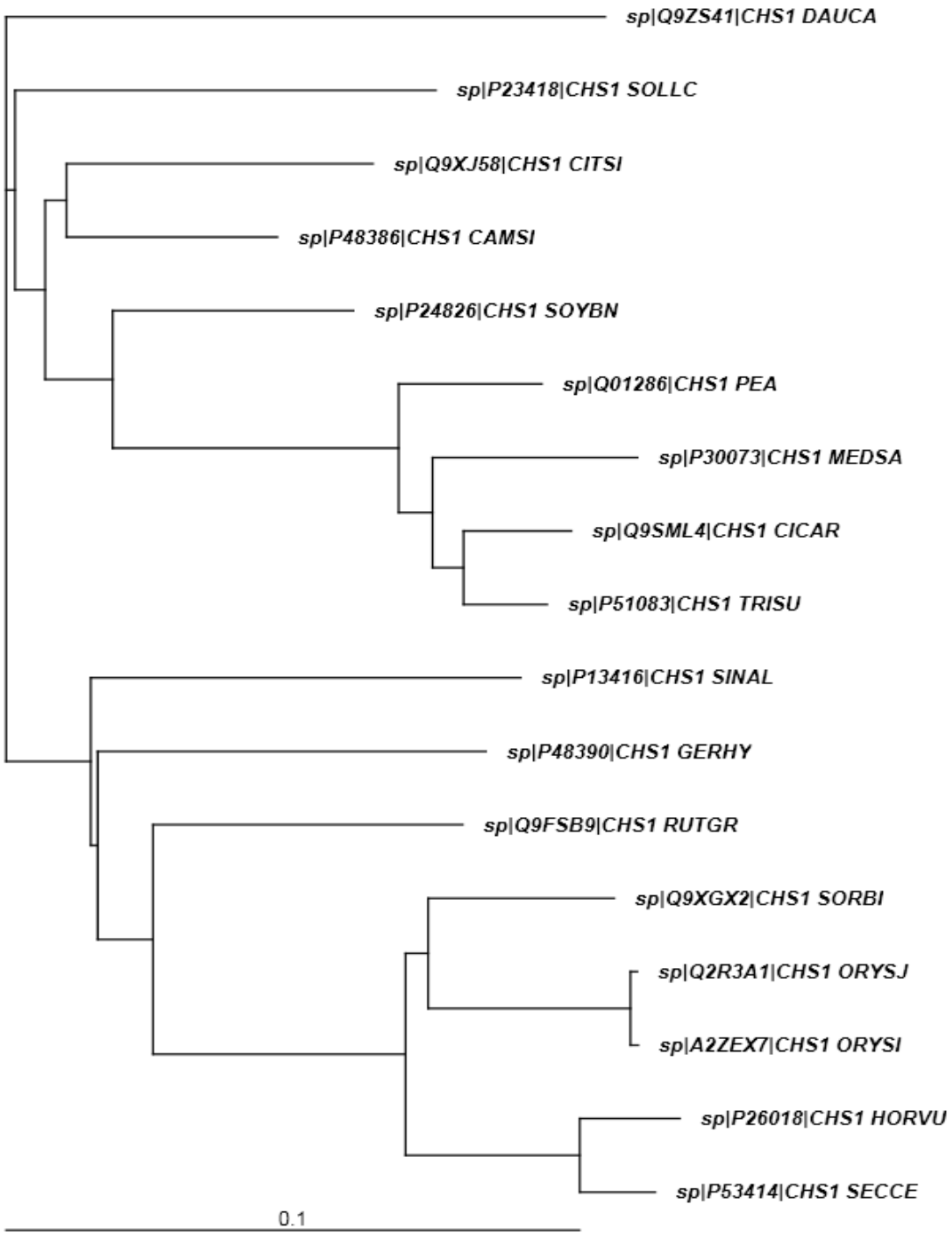

Fig. 1 The phylogenetic tree of chalcone synthase 1 protein sequences from some plants species