Pak. J. Agri. Sci., Vol. 53(4), 827-831; 2016
ISSN (Print) 0552-9034, ISSN (Online) 2076-0906
DOI: 10.21162/PAKJAS/16.3793
http://www.pakjas.com.pk

GENE DUPLICATION ASSESSMENT AND SEQUENCE ANALYSIS OF CHALCONE SYNTHASE GENE IN PLANTS

Ali Askar Nouhi

Sama Technical and Vocational Training College, Islamic Azad University, Firouzabad Branch, Fars, Iran.
***Corresponding author's e-mail: ali_nouhi2002@yahoo.com**

Chalcone synthase gene is a key enzyme in the biosynthesis of *flavonoids and isoflavonoids*. This gene assumes several functions in plants such as protection against UV ray and defense against pathogenic factors. In this study, chalcone synthase gene was studied and evaluated in terms of evolutionary rate in 119 species of creatures such as agricultural plants as eukaryote creatures and bacteria as prokaryotes. The evolutionary analyses were performed using bioinformatics tools. After preparing the phylogenic tree, the evolutionary study and comparison were performed. Also divergent time of chalcone synthase gene was investigated in the agricultural plants and bacteria. This study led to obtaining two general conclusions: First, the evolutionary rate between two prokaryote and eukaryote groups is different over time, and second, considering the calculation of divergent time for the genome of prokaryotes from eukaryotes in the previous works, it is recommended to consider the Chalcone synthase gene duplication time as about 410 million years.
Keywords: Chalcone synthase gene, divergent time, evolution rate, gene duplication, sequence analysis.

INTRODUCTION

Chalcone synthase gene is a key enzyme in the biosynthesis of flavonoids and isoflavonoids and assumes several functions in plants such as protection against Ultra violet ray and defense against pathogenic factors (Huang *et al*., 2008; Jiang *et al*., 2006; Durbin et al., 2000; Koes et al., 1994). It seems that chalcone synthase gene has been created from the enzyme involved in the synthesis of fatty acids (Stafford, 1991) and the structure of Chalcone synthase gene in flowering plants is conserved. Except *Antirrhinum majus* plant which has two introns, others have one conserved intron (Sommer and Saedler 1986; Griesbach et al., 2005; Han et al., 2006; Huang et al., 2004). It has been also determined that compacting enzyme coding gene in the synthesis pathway of fatty acids in *Escherichia coli* bacteria and CHS coding gene have been derived from the same ancestral gene (Verwoert et al., 1992). In recent studies, many genes from CHS family have been studied in different species, which include chalcone synthase in corn (*Zea mays*) (Franken *et al*. 1991) and in Arabidopsis (Saslowsky *et al*., 2000). In some of the plant species, chalcone synthase gene is encoded using a polygenic family, such as *Petunia* (Koes *et al*., 1989) and leguminous family (Ryder *et al*., 1987; Ito *et al*., 1997, Wingerder *et al*., 1989; Durbin et al., 1995; Fukuda et al., 1997; Hauttley et al., 1997). There is also some evidence showing that chalcone synthase gene is re-branching in chalcone synthase gene family (Helariutta *et al*., 1996; Tropf *et al*., 1994). The sites with synonymous and non-synonymous substitutions between chalcone synthase genes in *Denderanthema* family were evaluated by Nei and Gojobori (1986). Some studies were conducted on chalcone synthase gene sequence (CHS) in plants and bacteria and the chalcone synthase genes were aligned to investigate the relations of these genes among the plants by neighbour-joining (NJ) method (Jiang and Cao, 2007; Lei et al., 2010). In another study which was conducted on several plants, phylogeny relations were examined using maximum likelihood (ML) method and based on the nucleotide substitution parameters (Farzad *et al*., 2005). In another work on fungi, divergent time was calculated using nonparametric rate-smoothing (NPRS) and Longly Fitch (LF) algorithms (Kasuga *et al*., 2002). Molecular clock hypothesis and evolutionary rate between chalcone synthase genes in *Dendranthema* family was studied. For this purpose, relative test (Muse and Gaut 1994; Nei and Kumar 2000; Takezaki et al., 1995) was used. In addition, divergent time of chalcone synthase gene in *Dendranthema*family was examined (Cronquitst, 1977; Doolittle et al., 1996). Some tests were also performed for determining the substitution rate of synonymous and non-synonymous sites (Yang *et al*., 2002). Substitution rate in CHS gene was compared with that of other genes and it was specified that substitution rate inside chalcone synthase gene was faster than that of others (Yang *et al*., 2002). Generally, gene duplication over time is interpreted and analyzed as a main mechanism for evolution (Force *et al*., 1999; Ohta, 1993). In the study for estimating gene duplication time considering the lack of access to paleontological information, average synonymous substitution rate of the genes in the plant cell nucleus was used (Yang *et al*., 2002; Muse and Gaut, 1997). This study assumed that rate of evolutionary changes of chalcone synthase gene was equal in both eukaryote and prokaryote

Table 1. Estimated number of synonymous (below diagonal) substitution per site.

1.Arthroderma_otae	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***
2.Bdellovibrio_bacteriovorus_HD100	*0.658*												
3.Bdellovibrio_bacteriovorus_str_Tiberius	*0.662*	*0.262*											
4.Deinococcus_deserti	*0.628*	*0.631*	*0.677*										
5.Halobacillus_halophilus	*0.702*	*0.665*	*0.639*	*0.641*									
6.Mycobacterium_africanum	*0.646*	*0.619*	*0.653*	*0.568*	*0.644*								
7.Ralstonia_solanacearum	*0.670*	*0.611*	*0.618*	*0.573*	*0.627*	*0.565*							
8.Arabidopsis_lyrata_subsp_lyrata	*0.653*	*0.682*	*0.679*	*0.563*	*0.717*	*0.584*	*0.674*						
9.Arabidopsis_thaliana	*0.661*	*0.681*	*0.671*	*0.566*	*0.698*	*0.598*	*0.688*	*0.090*					
10.Emiliania_huxleyi	*0.576*	*0.596*	*0.626*	*0.523*	*0.747*	*0.574*	*0.610*	*0.597*	*0.597*				
11.Medicago_truncatula_chromosome4	*0.626*	*0.706*	*0.650*	*0.671*	*0.649*	*0.690*	*0.628*	*0.590*	*0.617*	*0.707*			
12.Oryza_brachyantha	*0.630*	*0.658*	*0.711*	*0.408*	*0.709*	*0.536*	*0.596*	*0.507*	*0.504*	*0.527*	*0.669*		
13.Ricinus_communis_scf_1106159295382	*0.568*	*0.635*	*0.601*	*0.676*	*0.704*	*0.635*	*0.680*	*0.663*	*0.653*	*0.678*	*0.506*	*0.588*	
14.Zea_mays	*0.600*	*0.586*	*0.612*	*0.444*	*0.670*	*0.527*	*0.590*	*0.399*	*0.416*	*0.538*	*0.688*	*0.358*	*0.642*

groups. This hypothesis was investigated by different parameters and analyzed in bioinformatics software. Gene duplication time was also calculated.

MATERIALS AND METHODS

CHS gene was studied in different eukaryote and prokaryote species using the Genbank (www.ncbi.nlm.nih.gov). The 119 species containing CHS gene of eukaryotes (17 plant species and 2 fungal species) and prokaryotes (100 bacteria species) were selected. Gene sequence was prepared in FASTA Format. In this study, chalcone synthase-like genes were not used and only CHS genes were used to compare the evolutionary changes between the eukaryotes and prokaryotes. Also, full sequence of genes including sequence of coding region and intron was used in this study. Two files containing CHS gene sequence were separately prepared for the eukaryotes and prokaryotes creatures and investigated from the evolutionary viewpoint.

***Bioinformatics analyses*:**Pair-wise and multiple sequence alignments were done in MEGA 6.0 Tamura and Kumar(2013). Mean intra-group distance in both of the groups was prepared based on ML model 100 times of re-sampling and transition-transversion substitution pattern. Mean distance was calculated for both eukaryote and prokaryote groups. Nucleotide substitution was also compared based on different models including F81 (Felsenstein, 1981), JC69 (Jukes and Cantor, 1969), k80 (Kimura, 1980), HKY85 (Hasegawa et al., 1985), and GTR (Tavare, 1986) in DAMBE software (Xia, 2013) and the best model was selected. Nucleotide substitution using ML transition-transversion bias parameter was calculated based on the two-parameter method by Kimura (1980) for both eukaryote and prokaryote groups. Gamma parameter was calculated based on Tamura and Nei's (1993) pattern for eukaryotes and prokaryotes. Molecular clock hypothesis was calculated for both groups based onML model using Tamura and Nei's (1993) method and statistically studied at $\rho \leq 0.05$. This hypothesis was commonly calculated based on ML method for 119 species (eukaryotes and prokaryotes) using different models. Phylogenic tree was commonly prepared based on ML model and Tamura and Nei's (1993) method with 100 re-sampling time for both species (Fig. 1). Nucleotide frequency of the 6 species of prokaryotes (including 3 bacteria) and eukaryotes (including 3 plants) representing 119 species was studied in DAMBE (Xia, 2013) program using Chi-square test. Heterogeneity test was conducted for the nucleotide frequency by Chi-square test.

***Gene duplication*:** Considering the drawing of phylogenic tree based on ML method (Fig. 1), point A (node A) which is shown using an arrow in the figure, was considered as the point at which the divergence of CHS gene occurred between the eukaryotes and prokaryotes. Also, 14 species of the eukaryote and prokaryote creatures located in the phylogenic tree were selected and the distance matrix was prepared based on synonymous-non-synonymous method for these 14 species which included eukaryotes and prokaryotes (7 species) (Table 1). Mean distance of the eukaryote and prokaryote species was separately calculated.

Duplication time of CHS gene was separately estimated based on the mean distance in the eukaryotes and prokaryotes. Gene duplication time was also estimated based on the mean distance of eukaryotes and prokaryotes.

RESULTS

Mean distance in the prokaryotic group with 100 nucleotide sequences and eukaryotic group with 19 nucleotide sequences was calculated as 1.81 and 0.745, respectively. Different substitution models were compared with each other using χ^2-test at $\rho \leq 0.05$. Considering the calculations, GTR model as the best model was selected because it was found to have minimum theoretic information indices (Table 2).

Two F81 and JC69 models were compared with each other and the p-value was calculated ($\rho = 0.0001$) and in the comparison of other models, the p-value was calculated ($\rho = 0$). ML transition-transversion bias was calculated as 0.71 and 1.09 for prokaryotes and eukaryotes based on the two-parameter method of Kimura (1980), respectively. Gamma parameter was calculated as 1.2732 for prokaryotes and

Table 2. Finding the best nucleotide substitution model.

Model	NumParam	lnL	AIC	AICc	BIC
GTR	GTR	-5464.573	10945.147	10947.673	10962.664
TN93	TN93	-5497.818	11005.637	11006.637	11016.585
T92	T92	-5542.210	11088.420	11088.610	11092.799
HKY85	HKY85	-5530.916	11069.832	11070.488	11078.591
F84	F84	-5519.487	11046.974	11047.630	11055.733
F81	F81	-5601.595	11209.190	11209.577	11215.759
K80	K80	-5571.652	11145.305	11145.368	11147.495
JC69	JC69	-5612.662	11225.324	11225.324	11225.324

NOTE: Nucleotide frequencies; T= 0.16437; C =0.27311; A =0.16361; G =0.39891

0.6065 for eukaryotes based on the substitution pattern by Tamura and Nei (1993). Molecular clock hypothesis was also calculated for both groups based onML model and Tamura and Nei's (1993) method. This hypothesis was calculated in the prokaryotes (ρ=1.3234763504256E-134) and rejected at ρ≤0.05. Moreover, this hypothesis was calculated in the eukaryotes(ρ=1.21672141647735E-144) and rejected at ρ≤0.05. Nucleotide frequency was investigated in 6 species of eukaryotes and prokaryotes using DAMBE (Xia, 2013) program. $\chi 2$ distribution was applied to study the nucleotide frequency for these 6 species. In most of the cases, nucleotide frequency distribution was significantly different from the normal frequency (0.25) for each of the A, C, G, and T nucleotides. $\chi 2$- test was run to study the heterogeneity of nucleotide frequency and, considering the calculated p-value, the null hypothesis of non-heterogeneity of the sequences was rejected at ρ≤0.05. Nucleotide substitution rate was calculated based on Tamura and Nei's (1993) pattern for prokaryotes and two nucleotides of A and G had maximum substitution rates compared with others (13.7). In the eukaryotes, substitution rate was calculated based on this pattern and maximum substitution rate was between nucleotides C and (T,U) (15.17). Gene duplication time was also calculated based on the substitution rate in each site per year (5×10^{-9}) (Li 1997) considering the mean distance in the prokaryotes (2.16) and eukaryotes (6.04). Therefore, CHS gene duplication time between the prokaryotes and eukaryotes was estimated to be about 216 million years based on the mean distance in the prokaryotes and about 600 million years based on the mean distance in the eukaryotes. For the mean distance in the prokaryotes and eukaryotes (4.1), CHS gene duplication time was estimated as 410 million years.

Considering mean distance in the prokaryote (1.81) and eukaryote (0.745) groups, it seems that there were more differences in terms of nucleotide sequence in the prokaryotes. This issue showed that the eukaryote group underwent less variation during the evolution of nucleotide sequence and sequence similarity of CHS gene was better maintained over time.

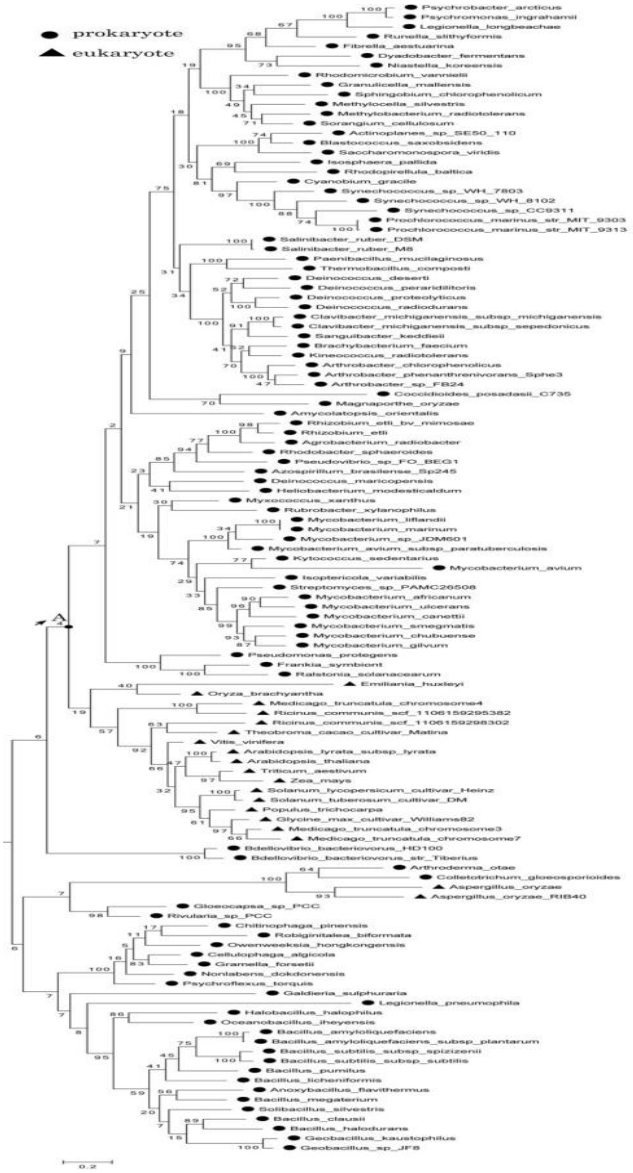

Figure1. Phylogenetic tree constructed by using ML method for Chalcone synthase gene for 119 species of plants and bacteria

DISCUSSION

Different substitution models were compared with each other and GTR model was selected as the best substitution model considering its value of theoretic information indices. According to the calculation of ML transition-transversion bias for both groups, nucleotide displacement value for CHS gene was higher in the eukaryotes. Based on the value of gamma parameter, there was less correlation between CHS gene sequence and nucleotide frequency of the eukaryotes. In other words, it can be said that each CHS gene sequence in the eukaryotes was more likely to have its own special nucleotide frequency. Considering the calculation of molecular clock hypothesis in both groups, no equal evolutionary rate existed forCHS gene in these groups and each group had its own evolutionary rate. Considering the results of $\chi 2$-test for studying the heterogeneity of nucleotide frequency in the 6 species representing 119 species, as expected, there was a significant difference in terms of CHS gene sequence between the prokaryotes and eukaryotes, which indicated the difference in the CHS gene sequence between them. Based on the calculated substitution rate, it can be inferred that, first, maximum substitution was related to purine bases in the prokaryotes and to pyrimidine bases in the eukaryotes. Second, transition value was higher than transversion value. CHS gene duplication time in the flowering plants was estimated to be almost 60 million years (Kasuga *et al*., 2002); but, no studies have calculated the CHS gene duplication time between prokaryotes and eukaryotes thus far. In the present work, CHS gene duplication time between the prokaryotes and eukaryotes was estimated as 410 million years. Considering the divergent time of genome of prokaryotes from eukaryotes (Gu, 1997), it is recommended to consider the calculated time in this study as the CHS gene duplication time.

Investigation of different evolutionary parameters and the analyses conducted using bioinformatics software led to obtaining two general conclusions: First, the evolutionary rate between two prokaryote and eukaryote groups is different over time, and second, considering the calculation of divergent time for the genome of prokaryotes from eukaryotes in the previous works, it is recommended to consider the CHS gene duplication time as about 410 million years.

***Acknowledgments*:** I would like to thank Dr. Xia (University of Ottawa) and Dr. Yang (University College London) about their guides. I would also like to thank all scientists whose studies have been used in this work. Cost of this work is provided by own and no cost is provided by university.

REFERENCES

Cronquist, A. 1977. The compositae revisited. Brittonia 29:137-153.

Doolittle, R.E., D.I. Feng, S. Tsang, G.Cho and E.Little. 1996. Determining divergence times of the majorkingdoms of living organisms with a protein clock. Sci.271:470-477.

Durbin, M.L., G.H. Learn, G.A. Huttley and M.T. Clegg. 1995. Evolution of the chalcone synthase gene family in the genus Ipomoea. Proc.Nat. Acad. Sci. USA. 92:3338-3342.

Durbin, M.L., B. Mc Caig and M.T. Clegg. 2000. Molecular evolution of the Chalcone synthase multigene family in the morning glory genomic. Plant Mol. Biol. 42:79-92.

Farzad, M., D.F. Soria-Hernanz, M. Altura, M.B. Hamilton, M.R. Weiss and H.G. Elmendorf. 2005. Molecular evolution of the Chalcone synthase gene family and identification of the expression copy in flower petal tissue of *Viola cornuta*. Plant Sci. 168:1127-1134.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan and J. Postlethwait. 1999. Preservation of duplication gene by complementary, degenerative mutations. Genetics 151:1531-1545.

Franken, P., U. Niesbach-Klosgen, U. Weydemann, L. Marechal-Drouard, H. Saedler and U. Wienand. 1991. The duplicated chalcone synthase gene C_2 and Whp (white pollen) of *Zea mays* are independently regulated; evidence for translational control of Whp expression by the anthocyanin intensifying gene in. EMBO J. 10:2605-2612.

Fukuda-Tanaka, S., A. Hoshino, Y. Hashimoto, Y. Habu, M. Hasebe and S. Lida. 1997. Identification of new chalcone synthase gene for flower pigmentation in the Japanese and common morning glories. Plant Cell Physiol. 38:754-758.

Griesbach, R.J. and R.M. Beek. 2005. Sequence analysis of the Chalcone synthase gene intron in four *Petunia jussiea* Taxa. J. Amer. Soc. Hort. Sci. 130:360-365

Gu, X. 1997. The Age of the Common ancestor of eukaryotes and prokaryotes: Statistical iferences. Mol. Biol. Evol. 14:861-866

Han, Y.Y., F. Ming, W. Wang and J. W Wang. 2006. Molecular evolution and functional specialization of Chalcone synthase super family from *Phalaenopsis orchid*. Genetica 128:429-438.

Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160-174.

Helariutta, Y., M. Kotilainen, P. Elomaa, N. Kalkkinen, K. Bremer, T.H. Teeri and V.A. Albert. 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification, Proc. Nat. Acad. Sci. USA. 93:9033-9038.

Huang, J., Q. Liu-Jia, Y. Ji, Y. Hao and G. Hong-Ya. 2004. A

preliminary study on the origin and evolution of Chalcone synthase (CHS) gene in angiosperms. Acta.Botanica Sinia 46:10-19.

Hauttley, G.A., M.L Durbin, D.E Glover and M.T Clegg. 1997. Nucleotide polymorphism in the chalcone synthase: A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. Mol. Ecol. 6:549-558.

Ito. M., Y. Ichinose., H. Kato., T. Shiraishi and T. Yamada. 1997. Molecular evolution and functional relevance of the chalcone synthase gene of pea. Mol. Gen. Genet. 255: 28-37

Jiang, C., C.K Schommer, S.Y Kim and D.Y Suh. 2006. Cloning and characterization of Chalcone synthase from the moss *Physcomitrella patens*. Phytochemistry 67:2531-2540.

Jiang, M. and J.S. Cao. 2007. Chalcone synthase gene. Chines J. Cell Biol. 29:525-529.

Jukes, T.H. and C.R Cantor. 1969. Evolution of Protein Molecules. New York: Academic Press. pp.21-132.

Kasuga, T., T.J Whiter and J.W Taylor. 2002. Estimation of nucleotide substitution rate in euromycete Fungi. Mol. Biol. Evol. 19:2318-2324.

Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, J. Mol. Evol. 16:110-120.

Koes, R.E., C.E. Spelt, P.J.M Van Den Elzen and J.N.M. Mol. 1989. Cloning and molecular characterization of the Chalcone synthase multi gene family of *Petunia hybrida*. Gene 81:245-257.

Koes, R.E., F. Quattrocchio and J.N.M. Mol. 1994. The flavonoid biosynthetic pathway in plants function and evolution. Bio Essays 16:123-132.

Lei, W., S.H Tang, K.M. Luo and M. Sun. 2010. Molecular cloning and expression profiling of a Chalcone synthase gene from hairy root cultures of *Scutellaria viscidula* Bung. Genetic and Mol. Biol. 33:285-291.

Li, W.H. 1997. Molecular evolution. Sinauer Associates Sanderland, Mass pp.177-213.

Muse, S.V. and B.S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rate with application to the chloroplast genome. Mol. Biol. Evol. 11:715-724.

Muse, S.V. and B.S. Gaut. 1997. Comparing patterns of nucleotide substitution rates among chloroplast loci using the relative ratio test. Genetics 146:393-399.

Nei, M. and T. Gojobori. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.

Nei, M. and S. Kumar. 2000. Molecular clocks and linearized trees, pp.187-206. In: Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Ohta, T. 1993. Pattern of nucleotide substitution in growth hormone-prolactin gene family: A paradigm for evolution by gene duplication. Genetics 134:1271-1276.

Ryder, T.B., S.A. Hedrick, J.N. Bell, X. Liang, S.D. Clouse and C.J. Lamb. 1987. Organization differential activation of a gene family encoding the plant defence enzyme chalcone synthase in *Phaseoluse vulgaris*. Mol. Gen. Genet. 210:219-233.

Saslowsky, D.E., C.D. Dana and B. Winkel-Shirley. 2000. An allelic series for the chalcone synthase locus in Arabidopsis. Gene 255:127-138.

Sommer, H. and H. Saedler. 1986. Structure of the chalcone synthase gene of *Antirrhinum majus*. Mol. Gen. Genet. 202:429-434.

Stafford, H.A. 1991. Flavonoid evolution an enzymic approach. Plant Physiol. 96:680-685.

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

Takezaki, T., A. Rzhetsky and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized tree. Mol. Biol. Evol. 12:823-833.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitution in the control region of mitochondrial DNA in human and chimpanzees. Mol. Biol. Evol. 10:512-526.

Tavare, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences: Lectures on Mathematics in the Life Sciences. Amer. Math. Soc. 17:57-86.

Tropf, S., T. Lanz, S.A. Renzing, J. Schroder and G. Schroder. 1994. Evidence that stiben synthase have developed from chalcone synthase several times in the course of evolution. J. Mol. Evol. 38:610-618.

Verwoert, I.I., E.C. Verbree, K.H. Van der Linden, H.J.J. Ni Jkamp and A.R. Stuitje. 1992. Cloning nucleotide and expressin of the *Escherichia coli* fab D gene, encoding malonyle coenzyme A-acyl carrier protein transacylase. J. Bacteriol. 174:2851-2857.

Wingender, R., R.H. Rohrig, C. Horicke, D. Wing and J. Schell. 1989. Differential regulation of soybean chalcone synthase gene in plant defence. Symbiosis and upon environmental stimuli. Mol. Gen. Genet. 218:315-322.

Xia, X. 2013. DAMBE 5: A comprehensive software package for data analysis in molecular biology and evolution. Mol. Biol. Evol. 30:1720-1728.

Yang, J., J. Huang, H. Gu, Y. Zhong and Z. Yang. 2002. Duplication and adaptive evolution of the Chalcone synthase gene of Dendranthema (Asteraceae). Mol. Biol. Evol. 19:1752-1759