Int. Jr. Bioinformatic and Biological Sci. : v.2 n.3 and 4, p. 165-177. Dec. 2014

DOI No. 10.5958/2321-7111.2014.00014.6

Computational and evolutionary insights into anthocyanin biosynthesis genes between Solanaceae and Poaceae

Aumreetam Dinabandhu*

National Research Centre on Plant Biotechnology; Pusa Campus, New Delhi-110012, India.

**Corresponding author: aumreetam@gmail.com*

ABSTRACT

Six common anthocyanin pathway genes viz. Chalcone synthase (CHS), Chalcone isomerase (CHI), Flavanone 3-hydroxylase (F3H), Dihydrofolate reductase (DFR), Anthocyanindin synthase (ANS) and UDPG-flavonoid 3-glucosyl transferase (UF3GT) were analyzed by performing comparative analysis between different species of solanaceae and poacea families. Sequence conservation between different species of these two families was highest in *CHS* and lowest in *CHI*. The number of polymorphic sites between different species of these two families was highest in *ANS* and lowest in *F3H*. McDonald–Kreitman analysis showed that non-synonymous changes between different species of these two families were higher for downstream enzymes: *UF3GT*, *DFR* and *ANS* than upstream enzymes: *CHS, CHI* and *F3H*. It was also observed that nucleotide diversity between these two families was highest in *CHI* but lowest in *F3H*. From the evolutionary analysis it was concluded that that *CHI* might have undergone ancient duplication and subsequent divergence during evolution and upstream enzymes CHS and CHI evolves very slowly than downstream enzymes DFR, ANS and UF3GT not because of their mutation rate but because of the selective constraint between different species of solanaceae and poaceae.

Keywords: anthocyanins, computational, enzymes, evolution, poaceae, solanaceae

Anthocyanins are ubiquitously present in higher plants. The anthocyanin biosynthetic pathway is one of the well studied pathways in plants. The sequential evolution of enzymes in the pathway made it as a model pathway to understand evolutionary processes. Duplication and divergence of genes from the pathway has led to the emergence of new classes of compounds (Rausher, 2006). The green algae from which land plants have been derived do not seem to have any anthocyanin enzymes. Bryophytes (mosses) also do not have anthocyanins but they contain chalcones, flavonols, and flavones (Markham, 1988). They contain only three upstream enzymes such as Chalcone synthase (CHS), Chalcone

isomerase (CHI) and Flavanone 3-hydroxylase (F3H) that seem to have recruited simultaneously and presumed to have come from duplication of genes involved in primary metabolism (Winkel-Shirley, 2001). Ferns are known to produce only leucoanthocyanindins such as kaempferol, quercetin, and myricetin. The enzyme Dihydrofolate reductase (DFR) converts substrate dihydroflavonols to leucoanthocyanindins though the presence of DFR in ferns is yet to be confirmed (Rausher, 2006). Finally, in the group of higher plants, by the addition of two more enzymes Anthocyanindin synthase (ANS) and UDPG-flavonoid glucosyl transferase (UF3GT) which convert leucoanthocyanindin to anthocyanidins and anthocyanidins to anthocyanins respectively, gymnosperms and angiosperms produce anthocyanins (Rausher, 2006).

The solanaceae family comprises about 95 genera and at least 2,400 species. Economically, this is the third most important plant taxon and most valuable in terms of vegetable crops. It contains most variable of crop species in terms of agricultural utility. In addition to their role as important food sources, many solanaceous species have a role as scientific model plants such as potato for tuber development, petunia for the analysis of anthocyanin pigments, and tomato and tobacco for plant defense (Mueller 2005). This is the first family of flowering plants for which comparative mapping was conducted (Tanksley *et al.*, 1988). Comparative genome analysis among two important solanaceous species tomato and potato showed that they differ in only five paracentric inversions (Tanksley *et al.*, 1992), whereas the tomato and pepper genomes differ in numerous rearrangements including several translocations as well as both pericentric and paracentric inversions (Prince *et al.,* 1993). Comparative genomics research is presently gaining momentum in solanaceae due to availability of genome sequence data for several species. Potato genome has been sequenced and it is available at public domain (http:// www.potatogenome.net/) and the tomato genome sequencing is under way and is available at http://solgenomics.net/. Within *Solanum* genera, tomato and potato are closely related and both are members of phylogenetically similar group of species. Sequence comparison between different organisms is one of the keys for the research of orthologous genes. To understand the organization and the evolution of genomes the whole genome comparison and synteny blocks were used (Wu *et al*., 2006). Exceptionally high level of conservation of genome organization at the macro and micro levels makes solanaceae family a model plant to explore the basis of phenotypic diversity and adaptation to natural and agricultural environments (Wu *et al*., 2006; Mueller 2005). The Poacea, one of the largest families in plant kingdom comprises about 700 genera and 11,000 species (Tzvelev 1989). Both are having highly economic importance by providing livelihood to the whole world. Many of the grass domesticates have undergone rather modest decrease in diversity relative to their wild relatives (Buckler *et al*., 2002). In domesticated maize, the diversity is roughly 30% below compared to

its closest wild relative (Wang *et al,* 1999). Grasses were thought to have evolved around 55 million years ago and the age of Solanceae is about ~40 million years. It has been reported by El-Sayed *et al*., (2008) that Purple corn (*Zea mays*), a member of poacea family, contains high concentration of anthocyanins (1277µg/g); much higher than other anthocyanin-rich sources whereas some members of solanaceae contain lesser or no anthocyanins such as cultivated tomatoes.

Enormous variations with respect to color and pigmentation patterns are observed between plant species. Anthocyanins are major pigments that contribute colors to different parts of the plants and help to attract pollinators. Any genetic alerations in structural and regulatory genes of anthocyanin biosynthesis will lead to variation in anthocyanin content, color and pigment patterns. Rausher *et al*. (1997) reportd that adaptive evolution of flower color governed more by regulatory genes because regulatory genes in anthocyanin pathway tend to evolve faster than the structural genes they regulate. Espley *et al*. (1997) reported that red colouration in apple fruit is due to the activity of a MYB transcription factor, MdMYB10 which regulate anthocyanins. Dias *et al*. (2003) reported that after duplication of maize *R2R3 Myb* genes, divergence of C-terminal regions of these genes occurred due to accumulation of substitutions during evolution and showed that divergent C termini of these R2R3 MYB proteins were subjected to purifying selection. Phylogenetic and evolutionary analyses of CHS gene family of *Viola cornuta* with maize, petunia, grape and Ipomoea have been studied by Farzad *et al*. (2005). Extensive research on the evolution of the CHS gene family has been conducted in *Petunia.* Rausher *et al.* (1999) compared the six core anthocyanin structural genes *CHS, CHI, F3H, DFR, ANS* and *UF3GT* in maize sequences to *Ipomoea* and *Antirrhinum* and found more than fivefold difference of nonsynonymous substitution rate between *CHS* and *UF3GT.* Durbin *et al*., (2000) suggested that different duplicate copies of *CHS* in Morning glory have acquired specialized functional roles over the course of evolution. Durbin *et al.*, (2000) found that Morning glory contains a six-member gene family exhibiting sequence similarity to chalcone synthases from other plants. Three of these copies CHS-A, CHS-B and CHS-C were unable to convert 4-coumaroyl-CoA and malonyl-CoA to naringenin chalcone, two of these copies CHS-A, CHS-B were capable of catalyzing the condensation reaction while the sixth copy was a pseudogene. Hof *et al*. (2008) compared the *CHS* gene lineage tree and the plastid species phylogeny of *Viola* and indicated that the different CHS copies in *Viola* were the products of both recent and more ancient duplications. Lu and Rausher (2003) analyzed the rate variation of anthocyanin genes and concluded that downstream enzymes evolve faster than upstream enzymes. This might be due to reduced selective constraint of downstream enzymes than upstream genes. They also observed the variation in synonymous and nonsynonymous substitution rates in *CHS, ANS*, and *UF3GT.* The same *UF3GT* showed higher amino acid substitution rate, *ANS* an intermediate

rate, *CHS* the lowest rate by experimenting six *Ipomoea* species. Rausher *et al*. (2008) concluded that difference in the rate of nonsynonymous substitution between upstream enzymes ANS and UFGT and down srteam enzyme CHS is not because of rate of adaptive substitution but because of selective constraint. Oberholzer (2000) reported that a new copy of CHS gene is evolved into the poaceae genome 15 to 25 million years interval and the rate of protein evolution is increased after duplication of the gene. Clegg *et al*. (1997) analysed three nuclear gene families chalcone synthase (*CHS*), ribulose-1,5-bisphosphate carboxylase (*RBC*S), and the gene alcohol dehydrogenases (*ADH*) and found that duplication and divergence in function appears to be relatively common for *CHS* genes in evolution of flowering plants. They also suggested that adaptive evolution has played an important role in driving divergence following gene duplication events and recombination is a pervasive force at all levels of plant evolution. Additionally, the nucleotide diversity across a genome is the source of most of the phenotypic variation (Buckler *et al*., 2002) although plant lineages differ in mutation rates, research has yet to show the connection between the mutation rate and extent of gene diversity (Muse, 2000). Nucleotide diversity provides a snap shot of evolution at its most basic level. Nucleotide diversity reflects a rich history of selection, migration, recombination, and mating systems (Muse, 2000). Balancing selection and/or frequency-dependent selection may also play an important role in increasing diversity at specific loci within a genome (Oleksyk *et al*., 2001). Comparative sequence analysis is a valuable tool to determine the nature of the evolved functions that make one species different from another (Bennetzen, 2000). Though different crops species of solanaceae and poaceae are scientifically and economically very important, the comparative analysis with respect to anthocyanin pathway genes between these two families has not been reported yet. Therefore, comparative sequence analysis between different species of solanaceae and poaceae will unravel the evolutionary relationship between these two important families.

Materials and Methods

The released tomato genome was downloaded from and anthocyanin genes were predicted by using BLAST (http://blast.ncbi.nlm.nih.gov/ Blast.cgi) and FGENESH (www.softberry.com) gene prediction programme. The predicted genes were BLAST searched against NCBI non redundant (nr) database (http://blast.ncbi.nlm.nih.gov/ Blast.cgi) to retrieve orthologs from monocots (poaceae) and dicots (solanaceae) present in whole plant kingdom. Synonymous and non-synonymous substitution rate (McDonald and Kreitman analysis- McDonald and Kreitman (1991), DNA divergence and conserved DNA regions, codon usage bias were calculated by using DnaSP version-5.10 software (www.ub.edu/dnasp/). Ka (the number of nonsynonymous substitutions per nonsynonymous

site), and Ks (the number of synonymous substitutions per synonymous site) for different species between solanaceae and poaceae was also carried out separately (Nei and Gojobori, 1986). The accession numbers of six common genes are chosen for comparative analysis between Solanaceae and Poaceae and obtained from NCBI database are: *Solanum lycopersicum*- all predicted genes*; Solanum tuberosum*- HQ659493.1, HQ659498.1, HQ659496.1, AF449422.1, HQ337900.1, DQ106850.1; *Petunia hybrida*- X04080.1, X14589.1, X60512.1, AF233639.1, AB027454.1; *Solanum melongena*- EU809469.1; *Zea mays*- NM001148774.1, *Zea mays*- NM001148774.1, NM001157980.1, NM001112225.1, NM00158995.1, X55314.1, NM001112416.1; *Oryza sativa*- AB000801.2, AF474922.1, XM474226.1, AB003496.1, Y07955.1, AY625694.1; *Triticum aestivum*- AY286097.1, DQ233636.1, AY373831.1, AB247917.1; *Hordium vulgare*- AF474923.1, X15694.1.

Results and Discussion

DNA sequence conservation calculated by DnaSP software between different species of solanaceae and poaceae was found highest (C value- 0.52 and conservation threshold: CT value -0.61) in *CHS* and lowest in *CHI* (Table 1). Number of polymorphic sites was highest (641) in *CHI* and lowest (294) in *F3H*. McDonald-Kreitman analysis showed that the synonymous changes between the species belonging to both solanaceae and poaceae were higher in downstream enzymes DFR, ANS and UF3GT and lower in upstream enzymes CHS, CHI and F3H. The number of segregating sites for solanaceae was highest in *CHS* (283) and lowest in ANS (74). The Ks and Ka were also analysed separately for different species of both the families. Between potato and petunia Ks for CHI was highest whereas between tomato and petunia Ks was lowest. Similar results were obtained for other species and given in Table1. Ks for CHS, DFR and ANS was not available between most of the species of solanaceae and poaceae. The nucleotide diversity between solanaceae and poaceae was higher in *CHI,* UF3GT, DFR and ANS and lower in *F3H* and CHS (Figure 1).

Table 1. Computational analysis of six common anthocyanin pathway genes between solanaceae (Pi1) and poaceae (Pi2) families

(1) Comparative analysis of consrved DNA regions, DNA divergence between Solanaceae (p1) and Poaceae (p2)						
Parameters	**CHS**	**CHI**	**F3H**	**DFR**	**ANS**	**UF3GT**
Selected region	1-1779	1-995	1-1807	1-1226	1-1439	1-1440
Total no. of analyzed sites	1165	676	1204	1126	1233	1325

Sequence conservation (C)	0.52	0.05	0.409	0.433	0.376	0.303
Conservation thresold (CT)	0.61	0.15	0.5	0.53	0.47	0.4
No. of polymerphic sites of Pi1	333	406	87	116	89	102
No. of polymerphic sites of Pi2	245	432	139	319	371	304
Total no. of polymerphic sites between Pi1and Pi2	557	530	294	580	641	394
No of mutations of Pi1	382	509	90	120	90	106
No of mutations of Pi2	276	517	146	344	407	342
Total no of mutations between Pi1 and Pi2	823	922	379	756	841	595
No of shared mutations between p1 and Pi2	39	147	8	18	13	14
Avg. no. of nuc. subs. per site between Pi1 and Pi2 (Dxy)	0.34013	0.56726	0.31939	0.42681	0.47067	0.60162
No. of net nuc. subs. per site between Pi1 and Pi2 (Da)	0.182.55	0.09743	0.21078	0.28438	0.32188	0.30509
Nuc. divesity of Pi1	0.16609	0.37276	0.08322	0.07478	0.05566	0.14444
Nuc. divesity of Pi2	0.14907	0.56691	0.13399	0.21008	0.24192	0.44861
Nuc. divesity between Pi1 and Pi2 (Pi (t))	0.26311	0.49776	0.23507	0.31305	0.34192	0.47958
Avg. no. of synonymous. sites	281.1	107.9	166.97	235.53	249.5	109.56
Avg. no. of non synonymous. sites	873.9	363.1	526.03	781.47	797.5	334.44
Avg. no. of synon. Subs. per site between Pi1 and Pi2	0.66084	0.55996	0.12977	0.60423	0.61658	0.52425
Avg. no. of non synon. Subs. Per site between Pi1 and Pi2	0.13475	0.48096	0.26615	0.21956	0.25147	0.47039
Tajima' D (nonsyn/syn) ratio	-1.06504	-0.02764	0.47711	1.11838	1.267711	0.79533
(2) McDonald–Kreitman analysis						
No. of sites	1779	995	1807	1226	1439	1440

No. of codons analysed	364	55	228	323	319	117
Segregating sites of Pi1	283	100	81	97	74	69
Segregating sites of Pi2	214	113	127	258	263	189
Segregating sites between Pi1 and Pi2	389	143	178	308	304	219
Total no. of nonsyn. changes Pi1	98	83	77	32	21	51
Total no. of syn. changes Pi1	221	28	7	68	54	21
Total no. of nonsyn. changes Pi2	61	87	108	145	146	147
Total no. of syn. changes Pi2	179	40	23	130	135	59
Total no. of nonsyn. changes between Pi1 and Pi2	148	153	160	171	166	192
Total no. of syn. changes between Pi1 and Pi2	378	57	30	188	180	76
Fixed diff.of nonsyn. changes between Pi1 and Pi2	59	12	123	136	147	78
Fixed diff.s of syn. changes between Pi1 and Pi2	137	7	24	148	180	49
Neutrality Index (NI)	0.909	1.566	1.158	0.99	1.129	1.587
Alfa value	0.091	-0.566	-0.158	0.1	-0.129	-0.587
G value	0.268	0.774	0.241	0.004	0.618	4.096
Fisher's exact test. P-value (two tailed)		0.424102	0.654285	1.0	0.440424	0.048906
3) Codon usage bias between Pi1 and Pi2						
GC content at (synonymous) third codon positions (GC3s)	0.571	0.516	0.642	0.655	0.656	0.534
Effective number of codons (ENC)	48.049	55.111	52.988	46.593	43.788	43.989
Codon Bias Index (CBI)	0.428	0.314	0.295	0.456	0.494	0.465
Scaled Chi-square using Yates' correction (SChi2)	0.362	0.194	0.179	0.406	0.524	0.391

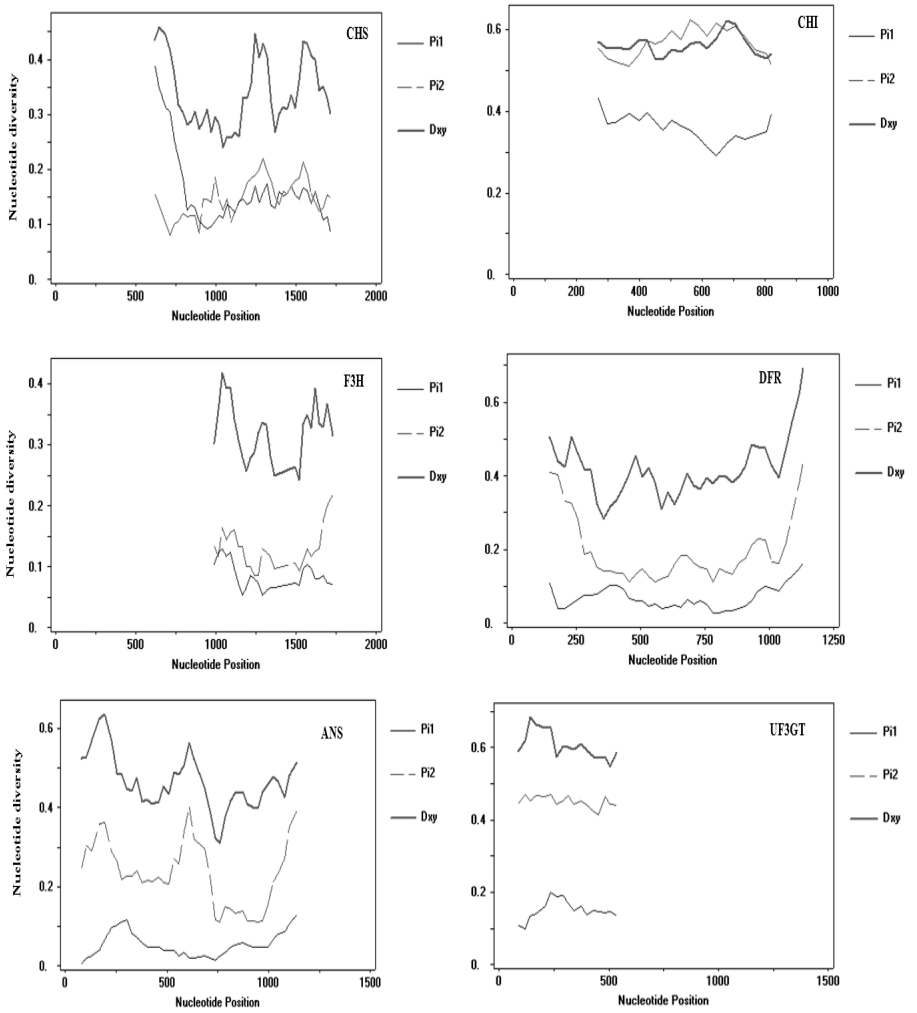

Figure 1. Nucleotide diversity of anthocyanin pathway genes between solanaceae (Pi1) and poaceae (Pi2) families. Dxy is the average number of nucleotide substitutions per site between Pi1 and Pi2. Genes are shown at top right corner of each box. X-axis shows the nucleotide position and Y-axis shows the nucleotide substitution per site.

Fixed difference of synonymous and nonsynonymous changes between solanaceae and poaceae was found highest in ANS and lowest in CHI. Effective number of codons (ENC) between solanaceae and poaceae was found higher in upstream enzymes and lower in downstream enzymes. Multiple sequence alignment of amino acid sequences of anthocyanin pathway genes showed that CHS was highly conserved within Solanaceae and Poaceae and CHI was least conserved (Figure 2a and 2b). UF3GT was less conserved than DFR and ANS across the families.

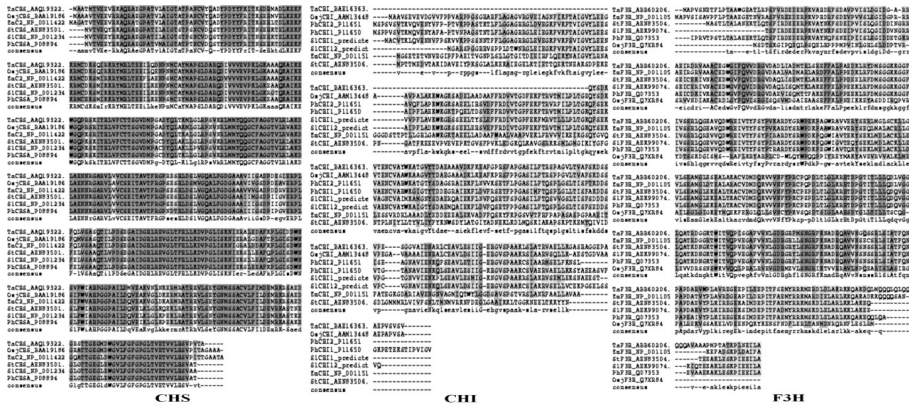

Figure 2a. Multiple sequence alignment showing conserved amino acid sequences of anthocyanin pathway genes- CHS, CHI and F3H within solanaceae and poaceae families. The color code indicates Green - conserved, Yellow- identical, Cyan- similar, and White- different domains.

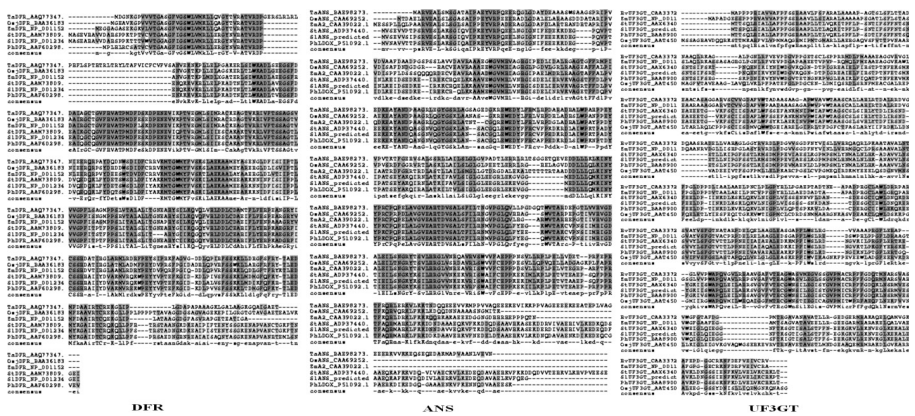

Figure 2b. Multiple sequence alignment showing conserved amino acid sequences of anthocyanin pathway genes- DFR, ANS and UFGT within solanaceae and poaceae. Represented color of the residues is: Green - conserved, Yellow- identical, Cyan-similar, and White- different domains.

Solanaceae and poaceae families were chosen to perform comparative analysis because some of the cultivated species in both families have lost anthocyanins from their wild ancestors though both families are scientifically and economically very important. Furthermore, minimum three species were chosen to represent each family because one species to represent a family might lead to error in sampling even though nucleotide sequence of a gene between different species within a family is highly conserved. Orthologs of *ANS* and *UF3GT* were taken

from *Hordium vulgare* in the place of *Triticum aestivum* and *CHI* from *Solanum melongena* in the place of *Petunia hybrid* unavailability of the sequences in the databases. In our study, McDonald and Kreitman analysis for positive selection showed higher nonsynonymous substitution changes in downstream enzymes DFR (171), ANS (166) and UF3GT (192) than upstream enzymes CHS (148) and CHI (153) and F3H (160). Ka/Ks ratio among each gene between both plant families was also substantially higher for downstream enzymes DFR, ANS and UF3GT. McDonald and Kreitman (1991) proposed a test that compares the synonymous and nonsynonymous variation within and between populations. Rausher *et al*. (1999) concluded from their studies between plant species (morning glory, snapdragon and maize) concluded that in anthocyanin pathway upstream enzymes evolve more slowly than downstream enzymes. Similarly evolutionary studies by Lu and Rausher (2003) in *Ipomoea sp*. concluded that the most upstream enzyme, chalcone synthase-D (CHS-D), evolves more slowly than the two most downstream enzymes, anthocyanidin synthase (ANS) and UDP glucose flavonoid 3-oxy-glucosyltransferase (UFGT). In this study, we observed same situation for upstream enzyme as well as downstream enzymes by studying plant species belonging to solanaceae and poaceae family. This indicates rapid evolution of downstream enzymes than upstream enzymes between solanaceae and poaceae. Between members of solanaceae and poaceae Ks was not available for CHS, DFR and ANS indicating the synonymous substitutions of these genes between those species have been saturated or close to saturation. Therefore, accurate estimation of Ka/Ks was not possible individually for these genes. Mutation rate for slow evolution of upstream enzymes were ruled out because Fisher's exact test was nonsignificant for all genes. Though CHS showed lower nonsynonymous changes, it was also having higher synonymous changes between solanaceae and poaceae. By assuming the mutation rate might not be the exact explanation of rate variation between upstream and downstream enzymes, we conducted Tajima's D test to know whether there is action of purifying selection over these enzymes or not. Tajima'D statistic was significantly negative for CHS and CHI indicating strong purifying selection during course of their evolution. It was higher and positive for downstream enzymes DFR, ANS and UF3GT. To support this interesting evidence, we analysed the conserved DNA regions, polymorphic sites, segregating sites and codon usage bias among these genes between solanaceae and poaceae. The codon use bias significantly differs between solanaceae and poacea species as these two are taxonomically divergent. Except F3H, GC content at third codon position was higher in CHS and CHI than DFR, ANS and UF3GT between two families. Effective number of codons (ENC) was higher in CHS and CHI than others between solanaceae and poacea species. CHS and CHI also showed more codon bias. The ENC and codon usage bias were higher for upstream enzymes. Nucleotide diversity analysis showed that between both plants families CHI, ANS, DFR are UF3GT are more diverse and CHS is least diverse. Multiple sequence

alignment of the aminoacid sequences also showed CHS is highly conserved than others. Rausher *et al*. (2008), studied of 15 Ipomoea *sp.* and concluded that difference in the rate of nonsynonymous substitution between upstream enzymes ANS and UFGT and down srteam enzyme CHS is not because of rate of adaptive substitution but because of selective constraint. From this study, we also concluded that the top branching enzyme CHS evolves very slowly than others because of selective constraint not because of mutation rate as this enzyme takes part not only in production of anthocyanin flavonoid compounds but also in various biotic and abiotic stresses. Any mutation or changes in enzyme will cause detrimental in physiological and ecological function in the plant. Downstream enzymes evolve more rapidly as they are specific in role of producing anthocyanins according to plants need.

Comparative analysis of anthocyanin pathway genes between different species of solanaceae and poaceae will provide molecular evidence underlying the importance of anthocyanin gene evolution. The extensive divergence and conserved sequences of these genes will help to design genetic markers controlling pigment patterns and anthocyanin content. Understanding of the molecular basis of anthocyanin gene evolution is necessary to unravel the diversity in flower color, variety of anthocyanin contents among different parts of the same plants and same organ of different species.

Acknowledgements

The authors thank Indian Council of Agricultural Research and National Research Centre on Plant Biotechnology for the financial assistance. Authors also thank Dr. T. R. Sharma and Dr N. K. Singh for their critical evaluation of the research.

References

Bennetzen JL 2000. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. *The Plant Cell.* **12**: 1021-1029.

Buckler ES and Thornsberry JM 2002. Plant molecular diversity and applications to genomics. *Curr. Opin. Plant Biol.* **5**:107–111

Clegg MT, Cummings PM, Durbin ML 1997. The evolution of plant nuclear genes. Proc. Natl. Acad.. Sci. USA **94**: 7791–7798

Dias AP, Braun EL, McMullen MD, Grotewold E 2003. Recently Duplicated Maize R2R3 Myb Genes Provide Evidence for Distinct Mechanisms of Evolutionary Divergence after Duplication. *Plant Physiol.* **131**: 610–620.

Durbin ML, McCaig B, Clegg MT 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* **42**: 79-92.

El-Sayed MA, Young JC, Rabalski I 2006. Anthocyanin Composition in Black, Blue, Pink, Purple, and Red Cereal Grains. *J Agric. Food. Chem.* **54**: 4696-4704.

Espley R, Brendolise C, Chagne D, Kutty-Amma S, Green S, Volz R, Putterill J, Schouten HJ, Gardiner SE, Hellens RP 2009. Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* **21**: 168-183

Farzad M, David FS, Altura MM, Hamilton MB, Weiss MR, Elmendorf HG 2005. Molecular evolution of the chalcone synthase gene family and identification of the expressed copy in flower petal tissue of Viola cornuta. *Plant Sci.* **168**: 1127-1134

Hof KVD, Berg RGVD, Gravendeel B 2008. Chalcone Synthase Gene Lineage Diversification Confirms Allopolyploid Evolutionary Relationships of European Rostrate Violets. *Mol. Biol. Evol.* **25**(10):2099–2108

Lu Y, Rausher MD 2003. Evolutionary Rate Variation in Anthocyanin Pathway Genes. *Mol. Biol. Evol.* **20**:1844–1853

Markham KR 1988. Distribution of flavonoids in the lower plants and its evolutionary significance. The flavonoids. Advances in Research Since 1980, Harborne, J. B., ed., Chapman and Hall, London, pp. 427-468.

McDonald JH, Kreitman M 1991. Adaptive protein evolution at the ADH locus in Drosophila. Nature **351**: 652-654

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD 2005. The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.* **138**(3):1310-7

Mueller LA, Tanskley SD, Giovannoni JJ 2005. The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL)," Comp Funct. Gen. **6**: 53–158

Muse SV 2000. Examining rates and patterns of nucleotide substitution in plants. *Plant Mol. Biol.* **42**: 25–43

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

Oberholzer V, Durbin ML, Clegg MT 2000. Comparative genomics of chalcone synthase and MYB genes in the fgrass family. *Genes Genet. Syst.* **75**: 1-16

Oleksyk TK, Smith MW and Brien OJS 2010. Genome-wide scans for footprints of natural selection Phil. Trans. R. Soc. B **365**: 185–20

Prince JP, Pochard E and Tanksley SD 1993. Construction of a molecular linkagemap of pepper and a comparison of synteny with tomato. *Genome* **36**: 404–417

Rausher M, Miller RE, Tiffin P 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* **16**: 266–274

Rausher MD 2006. The evolution of flavonoids and their genes. in E Grotewold, ed. The science of flavonoids. *Springer*, Berlin pp: 175–211

Rausher MD, Lu Y, Meyer K 2008. Variation in Constraint Versus Positive Selection as an Explanation for Evolutionary Rate Variation Among Anthocyanin Genes. *J Mol. Evol.* **67**: 137-144

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* **28**: 2731-2739

Tanksley SD, Bernatzky R., Lapitan NLV, Prince JP 1988. Conservation of gene repertoire but not gene order in pepper and tomato. *Pro. Natl. Acad. Sci.* **85**: 6419–6423

Tanksley SD, Ganal MW, Prince JP 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160

Thompson JD, Higgins DG, Gibson TJ 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673-4680

Tzvelev NN 1989. The system of grasses (Poaceae) and their evolution. The Botanical Review **55**: 141-203.

Wang RL, Stec A, Hey J, Lukens L, Doebley J 1999. The limits of selection during maize domestication. *Nature* **398**: 236-238

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD 2008. Sequencing and Comparative Analysis of a Conserved Syntenic Segment in the Solanaceae. Genetics **180**: 391-408

Winkel-Shirley B 2001. Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* **126**: 485–493

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* **174**: 1407–1420.