Computational analysis of common bean (Phaseolus vulgaris L., genotype BAT93) lycopene β-cyclase and β-carotene hydroxylase gene’s cDNA

Subhash Janardhan Bhore1,2*, Kassim Amelia1,2, Edina Wang3, Sindhuja Priyadharsini2 & Farida Habib Shah1,3

1Molecular Biology Division, Melaka Institute of Biotechnology, Lot 7, Melaka International Trade Centre City, 75450 Ayer Keroh, Melaka, Malaysia; 2Department of Biotechnology, Faculty of Applied Sciences, AIMST University, Bedong-Semeling Road, Bedong, 08100, Kedah, Malaysia; 3Novel Plants Sdn. Bhd., 27C Jln Petaling Utama 12, 7.5 miles Old Klang Road, 46000 Petaling Jaya, Malaysia; Subhash Janardhan Bhore – Email: subhashbhore@gmail.com; Ph: +60-4-429 8176; Fax: +60-4-429 8109; *Corresponding author

Received December 09, 2012; Accepted December 21, 2012; Published February 21, 2013

Abstract:
The identification of genes and understanding of genes’ expression and regulation in common bean (Phaseolus vulgaris L.) is necessary in order to strategize its improvement using genetic engineering techniques. Generation of expressed sequence tags (ESTs) is useful in rapid isolation, identification and characterization of the genes. To study the gene expression in P. vulgaris pods tissue, ESTs generation work was initiated. Early stage and late stage bean-pod tissues cDNA libraries were constructed using CloneMiner cDNA library construction kit. In total, 5972 EST clones were isolated using random method of gene isolation. While processing ESTs, we found lycopene β-cyclase (PvLCY-β) and β-carotene hydroxylase (PvCHY-β) gene’s cDNA. In carotenoid biosynthesis pathway, PvLCY-β catalyzes the production of carotene; and PvCHY-β is known to function as a catalyst in the production of lutein and zeaxanthin. To understand more about PvLCY-β and PvCHY-β, both strands of both cDNA clones were sequenced using M13 forward and reverse primers. Nucleotide and deduced protein sequences were analyzed and annotated using online bioinformatics tools. Results showed that PvLCY-β and PvCHY-β cDNA s are 1639 and 1107 bp in length, respectively. Analysis results showed that PvLCY-β and PvCHY-β gene’s cDNA contains an open reading frame (ORF) that encodes for 502 and 305 amino acid residues, respectively. The deduced protein sequence analysis results also showed the presence of conserved domains needed for PvLCY-β and PvCHY-β functions. The phylogenetic analysis of both PvLCY-β and PvCHY-β proteins showed it’s closeness with the LCY-β and CHY-β proteins from Glycine max, respectively. The nucleotide sequence of PvLCY-β and PvCHY-β gene’s cDNA and it’s annotation is reported in this paper.

Keywords: Expressed sequence tags, Genetic engineering, Health, Human population, Malaysia, Natural products, Nutrition, Phaseomics, Proteins, Vegetables.

Background:
We certainly do not know how many people are malnourished; but, FAO report indicates that there are about 925 million undernourished people in the world [1]. The animal products (eggs, meat, milk, etc.) are a source of dietary proteins; but, proteins are usually derived from legumes (plants from the bean and pea family) especially by poor people [2]. There are thousands of legume species, but common beans (Phaseolus vulgaris L.) are cultivated on the large scale. By understanding the importance of P. vulgaris, the Phaseomics international
consortium was developed to establish the necessary framework of knowledge and materials for the advancement of bean genomics, transcriptomics, and proteomics; and the main goal of it is to help in generating new common bean varieties suitable and desired by farmers and consumers [3]. As a part of the international consortium for Phaseolus genomics [3], research work on generation of P. vulgaris expressed sequence tags (ESTs) was initiated at Melaka Institute of Biotechnology, Malaysia.

The randomly isolated anonymous cDNA clones (on a large scale) are treated as ESTs and used extensively in the gene’s expression and regulation studies [4]. The generated ESTs data is also used in the evaluation of the genomes for genes content and its structure, in comparative gene expression analysis between different plant tissues using computational tools [5], and in discovery of new and novel genes [6].

To this point, we have generated 5972 ESTs; and annotated ESTs were deposited into ESTs database hosted by National Center for Biotechnology Information (NCBI) GenBank / DDBJ / EMBL (our unpublished work). While processing and analysing generated ESTs, we found lycopene β-cyclase and β-carotene hydroxylase gene's cDNA [12, 13]. The source of lycopene β-cyclase and β-carotene hydroxylase cDNA is P. vulgaris; hence, lycopene β-cyclase and β-carotene hydroxylase cdNA's were designated as PvLCY-β and PvCHY-β, respectively. In carotenoids biosynthesis pathway, PvLCY-β catalyzes the production of carotene (α-carotene and β-carotene) [12, 14]; and PvCHY-β is known to function as a catalyst in the production of lutein and zeaxanthin [13].

Due to antioxidant properties of carotenes (β-carotene), several health benefits associated with its consumption are reported elsewhere [15]. Similarly, the benefits of lutein and zeaxanthin consumption are reported by many researchers; and their reports are reflecting the importance of these (carotenes, lutein and other health benefits associated with its consumption are reported by many researchers; and their reports are reflecting the importance of these (carotenes, lutein and zeaxanthin) natural products in human health [16-22].

Both, PvLCY-β and PvCHY-β cDNA clones do have potential applications in genetic engineering of P. vulgaris and other plants. That is why, both clones were fully sequenced. These two cDNA clones could be used in manipulating P. vulgaris and level of carotene, lutein and zeaxanthin could be elevated. Hence, in order to understand more about PvLCY-β and PvCHY-β, their cDNA clones were analysed and annotated. The nucleotide and deduced protein sequence of PvLCY-β and PvCHY-β gene's cDNA are analyzed and annotated in this study using computational tools. The nucleotide sequence of PvLCY-β and PvCHY-β gene's cDNA and its annotation is reported in this paper.

Methodology:
Plant Materials:
The seeds of P. vulgaris genotype BAT93 were kindly provided by Patricia Lariguet, Laboratoire de Biologie Moléculaire des Plantes Supérieures, Department of Plant Biology, University of Geneva, Geneva, Switzerland. Seeds were germinated in soil obtained from a nursery (Melaka, Malaysia), and seedlings were maintained to grow in the open area at Melaka Institute of Biotechnology, Malaysia.

The PvLCY-β and PvCHY-β cDNA clones isolation
The PvLCY-β and PvCHY-β cDNA clones were identified from the ESTs generated using random method of gene isolation [7, 8, 23]. The cDNA clone encoding PvLCY-β was isolated from 20-day-old [days after anthesis (DA)] bean-pod-tissue cDNA Entry Library; and the cDNA clone encoding PvCHY-β was isolated from 5-day-old bean-pod tissue cDNA Entry Library. The cDNA libraries were constructed (our unpublished data) using ‘CloneMiner cDNA library construction’ procured from Invitrogen Corporation.

Plasmid DNA isolation
The individual cultures of Escherichia coli strain DH5α cells harbouring recombinant plasmids with PvLCY-β and PvCHY-β cDNA clones were cultivated in 10 ml LB medium supplemented with 40µg/ml Kanamycin. Cultures were incubated in dark at 37°C, 160 rpm for 18 h. From harvested E. coli cells, plasmid DNA was isolated using Wizard® Plus SV Minipreps DNA purification system, a commercial kit (Promega).

Nucleotide sequencing
Purified plasmid DNA was used in sequencing reactions. Both strands of both PvLCY-β and PvCHY-β cDNA clones were sequenced using M13 (Forward) [5'-GTAAACGACGGCCAGC-3'] and M13 (Reverse) [5'-GGATAA CAA TTT CAC AAC CAG-3'] primers.

cDNA and deduced protein sequence analysis
For both PvLCY-β and PvCHY-β cDNA clones, the nucleotide sequence of plus (+) and minus (-) strands were aligned using Blast (bl2seq) program available at NCBI [http://blast.ncbi.nlm.nih.gov/]. The 5' and 3' ends of the cDNA sequences were edited to eliminate adaptor and vector sequences. The final cDNA sequences were analyzed using online bioinformatics tools.

The similarity searches were performed using blast programs (BlastN and BlastP) available at NCBI. Online bioinformatics tools available at justBio [http://www.justbio.com/] were used to deduce the protein sequence, to find out the general features of PvLCY-β and PvCHY-β cDNA and deduced protein sequences. The EMBOSS Water - Pairwise Sequence Alignment [http://www.ebi.ac.uk/Tools/emboss/align/] was used to compare cDNA and deduced protein sequences to find out similarity% with their counterparts from other species. Guanine and cytosine (GC %) content calculation was carried out by using ‘DNA / RNA base composition calculator’. Alignment of multiple protein (amino acids) sequences was carried out using multiple sequence alignment by ClustalW program, and the phylograms were constructed using BioEdit and TreeView programs [24, 25]. Proteins sequences were aligned by using CLUSTAL 2.1 multiple sequence alignment program to find out conserved residues in both PvLCY-β and PvCHY-β deduced proteins.
Discussion:

**PvLCY-β and PvCHY-β cDNA clones isolation**

The full-length PvLCY-β and PvCHY-β cDNA clones were isolated from 20-day-old and 5-day-old bean-pod-tissues cDNA libraries, respectively. The isolated PvLCY-β and PvCHY-β cDNA clones were designated as PvLCY-β and PvCHY-β to indicate their precise identity and the source of the plant to which they belong.

**Nucleotide sequencing**

Both, sense (+) and antisense (-) strands of both cDNA clones were sequenced where M13 forward and M13 reverse primers were used. After elimination of the vector and adaptor sequence, the sequence of sense and antisense strand of individual cDNA was compared using blast (bl2seq) program. Analysis of the results showed that PvLCY-β and PvCHY-β cDNAs are 1639 and 1107 bp in length, respectively.

Figure 1: Nucleotide and deduced amino acid sequences of *Phaseolus vulgaris* lycopene β-cyclase (PvLCY-β) cDNA clone. Open reading frame (ORF) and 3’ non-coding region of cDNA are shown in capital and small letters, respectively. The deduced amino acid sequence is given below the nucleotide sequence, and numbered at both ends of each sequence line. The ORF encodes for a protein of 502 amino acid residues (blue). Amino acid residues are numbered beginning with the initial Methionine (M) till last Glutamic acid (E) residue. Initiation and termination codons are shown in green and red colour, respectively. * represents the termination codon. This cDNA clone was isolated from *P. vulgaris* 20-day-old pods tissue cDNA library.

**cDNA and Deduced Protein Sequence Analysis**

The identity of both cDNA clones was confirmed by analyzing finalized respective cDNA sequence and its deduced amino acid sequence. Annotated nucleotide sequences of both PvLCY-β and PvCHY-β cDNA were deposited in GenBank/ DDBJ/ EMBL under the accession numbers...
HQ199604 and JN255133, respectively. Annotated general features of cDNA nucleotide and protein sequences are summarized in Table 1 (see supplementary material); and nucleotide sequence of PvLCY-β and PvCHY-β cDNA along with its deduced amino acid sequence is shown in Figure 1 & Figure 2, respectively.

Figure 2: Nucleotide and deduced amino acid sequences of Phaseolus vulgaris beta-carotene hydroxylase (PvCHY-β) cDNA clone. Open reading frame (ORF) and non-coding regions of cDNA are shown in capital and small letters, respectively. The deduced amino-acid sequence is given below the nucleotide sequence, and numbered at both ends of each sequence line. The ORF encodes for a protein of 305 amino acid residues (blue). Amino acid residues are numbered beginning with the initial Methionine (M) till the last Serine (S) residue. Initiation and termination codons are shown in green and red colour, respectively. *represent the termination codon. This cDNA clone was isolated from P. vulgaris 5-day-old-pods tissue cDNA library.

The similarity% of both PvLCY-β and PvCHY-β cDNA nucleotide and deduced protein sequence with their counterparts from other species are shown in Table 2 & Table 3 (see supplementary material), respectively. The amino acid Sequence analysis results showed that both PvLCY-β and PvCHY-β proteins are Leucine (L) rich (Supplementary Figure 1 & 2). The comparison of the PvLCY-β protein with its counterparts from other species showed that 217 (out of 502) residues (43.23%) are fully conserved. But, in case of the PvCHY-β protein, results showed that only 67 (out of 305) residues (21.97%) are fully conserved. The consecutive search for conserved domains in PvLCY-β and PvCHY-β protein sequences resulted in the detection of their conserved domains, and the results are summarised in Table 4 (see supplementary material). The phylograms were constructed in order to understand phylogenetic relationship of PvLCY-β and PvCHY-β proteins with their counterparts from other species. The phylograms for PvLCY-β and PvCHY-β proteins are shown in Figure 3 & Figure 4, respectively.

The understanding of the identified genes, their expression patterns and regulation is crucial in order to strategize the approach to manipulate any biosynthesis pathway of interest in the plants. For the suppression of a gene expression, partial sequence of that gene can be utilized to induce post-transcriptional gene silencing (PTGS) (26-28). However, the full
length gene or its cDNA is required for its over-expression in order to increase either the production of desired vital proteins or natural products [29]. Therefore, understanding of gene of interest and its cDNA is prerequisite before it can be used in recombinant DNA (rDNA) technology to manipulate genetically, any plant of interest or organism [30].

The main goal of this study was to annotate PvLCY-β and PvCHY-β gene’s cDNA and deduced respective protein (amino acid sequence). The PvLCY-β cDNA clone was identified in 20-day-old-pod tissue cDNA library, and it indicates that PvLCY-β is expressed in bean’s 20-day-old developing pod tissue. However, the PvCHY-β cDNA clone was identified in 5-day-old-pod tissue cDNA library; and it reflects that PvCHY-β is expressed in bean’s 5-day-old developing pod tissue. However, the level of both gene’s expression, pattern of expression, and tissue-specificity is not clear at this moment as we have not characterised these two gene’s expression. It can be done by using either Northern hybridization technique or microarray technique [31].

The GC content in PvLCY-β and PvCHY-β cDNAs is 42% and 46%, respectively. The GC% in both PvLCY-β and PvCHY-β cDNAs is significantly higher than that of the GC% (39.4 %) reported in nuclear DNA of the broad bean [32].

**Figure 3:** The phylogram showing phylogenetic relationship of common bean (Phaseolus vulgaris L.) lycopene β-cyclase (PvLCY-β) protein with its counterparts from other species. Available 35 full-length PvLCY-β protein sequences were retrieved from NCBI database (see supplementary Table 2). The location of PvLCY-β protein in phylogram is shown in a pink box.

**Figure 4:** The phylogram showing phylogenetic relationship of common bean (Phaseolus vulgaris L.) β-carotene hydroxylase (PvCHY-β) protein with its counterparts from other species. Available 16 full-length PvCHY-β protein sequences were retrieved from NCBI database (see supplementary Table 3). Location of PvCHY-β protein in phylogram is shown in a pink box.

The LCY-β protein of G. max showed the maximum similarity (95.2%) with inferred PvLCY-β protein. Whereas, LCY-β protein from Cryptomeria japonica and Taxodium distichum showed less similarity (83.8%) with inferred PvLCY-β protein. Both, Cryptomeria japonica and Taxodium distichum are species that belong to gymnosperms; the relatively low level of PvLCY-β similarity with LCY-β from Gymnosperm members is in line with evolution in plant species [33]. Interestingly, LCY-β protein of Salicornia europaea and Crocus sativus showed lowest
similarity with PvLCY-β. While analysing PvCHY-β, we noticed that Glycine max CHY-β protein shows the maximum (78%) similarity with inferred PvCHY-β protein. On the contrary, M.uriella zofingiensis CHY-β protein showed less (58%) similarity with PvCHY-β protein. M. uriella zofingiensis is an algal member, and the relatively low level of similarity between PvCHY-β and CHY-β from M. zofingiensis is along the lines of the evolution in plants. These results are similar to the results reported by Bhore et al. [23]. Both, PvLCY-β and PvCHY-β proteins showed highest similarity with LCY-β and CHY-β of G. max, respectively. This makes logical sense because both P. vulgaris and G. max belong to the same family, Fabaceae [34].

The PvCHY-β protein contains conserved domain for beta-carotene hydroxylase which is a member of the Fatty acid hydroxylase superfamily (Accession No: cl01132) [35]. In PvLCY-β, two main conserved domains named, NAD_B, Rossmann super family and lycopene beta cyclase were detected. The NAD_B domain does exist in numerous dehydrogenases of metabolic pathways; for example glycolysis. The lycopene cyclase family protein conserved domain was detected in PvLCY-β, and lycopene beta and epsilon cyclases are part of this protein family [36-38].

Phaseolus vulgaris is a valuable source of proteins in the human diet; and it is important to increase the yield of this essential crop [3, 39]. Several research teams are using GM technology approach to improve yield of the beans [30, 40]. For instance developing P. vulgaris resistant to the herbicide [41] and viral infection [42]. In addition to this, there is a vast scope to modify beans genetically for improving the nutritional quality of its pods and seeds. This type of genetic manipulation is possible, because, rice (Oryza sativa) has been genetically engineered and β-carotene content in it has been increased for use as a source of vitamin A [43]. Similarly, β-carotene content can be increased in beans by over-expression of PvLCY-β in its carotenoid biosynthesis pathway (Supplementary Figure 3). Furthermore, therapeutically beneficial lutein and zeaxanthin content increment in beans is also possible by over-expression of PvCHY-β [44].

Genetic modification of agricultural crop plants to improve yield and nutritional quality is a viable option, and it is absolutely important as far as human wellbeing is concerned [30, 40]. Both, isolated PvLCY-β and PvCHY-β gene's cDNA are reasonably well annotated in this study, and we believe that the available annotated cDNA sequences could be useful in designing the strategy for the construction of transformation vectors. Further research is needed in this line to achieve the ultimate goal of generating new common bean varieties suitable and desired by farmers and consumers.

Conclusion:
This study has annotated the salient features of PvLCY-β and PvCHY-β gene's cDNA clones. The computational analysis of deduced PvLCY-β and PvCHY-β proteins revealed the presence of conserved domains. Furthermore, the comparative analysis of deduced PvLCY-β and PvCHY-β protein sequences with their counterparts from other species revealed the fully conserved amino acid residues. However, further study is required to understand PvLCY-β and PvCHY-β gene's expression and its regulation in bean-pods. Both genes' over-expression in bean-pods can be considered for further research to explore the possibility of nutritional quality improvement of the bean-pods and bean-seeds.

Acknowledgment:
Authors are grateful to the Ministry of Science, Technology and Innovation (MOSTI), Malaysia for research funding [Research Grant Code Number: BSP (M) / BTK / 004 (3)]; and to Patricia Lariguet, Department of Plant Biology, University of Geneva, Geneva, Switzerland for supplying seeds of bean, genotype BAT93.

Conflict of Interests:
Authors attest that there are no conflicts of interest to declare.

References:
[1] http://www.fao.org/
[2] Massev LK, J Nutr. 2003 133: 8625 [PMID: 12612170]
[3] Broughton WJ et al. Plant and Soil. 2003 252: 55
[4] Lange BM et al. Proc Natl Acad Sci U S A. 2010 97: 2934 [PMID: 10717007]
[5] Ronning CM et al. Plant Physiol. 2003 131: 419 [PMID: 12586677]
[6] Ohirogue J & Benning C, Curr Opin Plant Biol. 2000 3: 224 [PMID: 10837264]
[7] Liu J et al. Plant Mol Biol. 1995 29: 685 [PMID: 8541495]
[8] Keith CS et al. Plant Physiol. 1993 101: 329 [PMID: 8278499]
[9] Park YS et al. Plant Physiol. 1993 103: 259 [PMID: 8029332]
[10] Hölte H et al. Plant J. 1993 3: 1051 [PMID: 8281187]
[11] Umehda M & Uchimiyaa H, Plant Physiol. 1994 106: 1015 [PMID: 12232362]
[12] al-Babili S et al. Novartis Found Symp. 2001 236: 219 [PMID: 11387982]
[13] Amponah-Dwamena C et al. J Exp Bot. 2009 60: 3765 [PMID: 19574250]
[14] Rodríguez-Sáz M et al. Appl Environ Microbiol. 2004 70: 5589 [PMID: 15345447]
[15] Jiang YH et al. Plant Sci. 1994 73: 1137 [PMID: 7937475]
[16] Adluri RS et al. Cell Biochem Biophys. 2011 [Epub ahead of print] [PMID: 21960420]
[17] Holton SG et al. Int J Cancer. 2012 131: 1158 [PMID: 22038870]
[18] Kabat GC et al. Am J Clin Nutr. 2009 90: 162 [PMID: 19474140]
[19] Gao YY et al. Br J Nutr. 2012 107: 1746 [PMID: 22273061]
[20] Roberts RL et al. Clin Dermatol. 2009 27: 195 [PMID: 19168000]
[21] Ma L & Lin XM, J Sci Food Agric. 2010 90: 2 [PMID: 20355006]
[22] Polidori MC et al. Dtsch Med Wochenschr. 2012 137: 305 [PMID: 22318482]
[23] Bhore SJ et al. Bioinformation. 2010 5: 104 [PMID: PCMCID: PMC3040485]
[24] Thompson JD et al. Nucleic Acids Res. 1994 22: 4673 [PMID: 7984417]
[25] Page RD et al. Comput Appl Biosci. 1996 12: 357 [PMID: 8902363]
[26] Smith NA et al. Nature. 2000 407: 319 [PMID: 11014180]
[27] Liu Q et al. Plant Physiol. 2002 129: 1732 [PMID: 12177468]
Supplementary material:

Table 1: The general features of Phaseolus vulgaris lycopene beta-cyclase (PvLCY-β) and β-carotene hydroxylase (PvCHY-β) cDNAs and their deduced protein sequence.

| General features | PvLCY-β | PvCHY-β |
|------------------|---------|---------|
| **cDNA sequence** |         |         |
| i. Size, bp      | 1639    | 1107    |
| ii. Molecular weight (Daltons) | 508009 | 341098 |
| iii. 5' UTR, bp  | 0       | 50      |
| iv. Coding sequence | 1509  | 918     |
| v. 3' UTR, bp    | 130     | 139     |
| vi. Stop Codon   | TAA (UAA) | TGA (UGA) |
| vii. G+C content % | 42 % | 46 % |
| **Protein sequence** |         |         |
| viii. Length, amino acids | 502 | 305 |
| ix. Molecular weight (Daltons) | 56469.44 | 34127.52 |
| x. Isoelectric point | 7.16 | 7.90 |

Table 2: Comparison of Phaseolus vulgaris lycopene beta-cyclase (PvLCY-β) cDNA nucleotide and deduced amino acid sequence with LCY-β from monocot, dicot and gymnosperm members.

| Species | GenBank Accession No. | Length compared | Similarity (%) |
|---------|-----------------------|-----------------|----------------|
|         |                       | nucleotide (bp) | amino acid     | nucleotide (bp) | amino acid     |
| **Monocots** |                     |                 |                |                |                |
| Crocus sativus | GQ202143 | 1503 | 497 | 69.9 | 82.1 |
| Narcissus pseudonarcissus | GQ327929 | 1515 | 504 | 70.9 | 86.0 |
| Narcissus tazetta | Q797381 | 1515 | 504 | 71.5 | 85.2 |
| Sandersonia aurantiaca | A489520 | 1696 | 495 | 67.0 | 84.9 |
| **Dicots** |                     |                 |                |                |                |
| Adonis aestivalis | A321534 | 2026 | 502 | 71.3 | 88.4 |
| Arabidopsis lyrata | XM_002884738 | 1953 | 501 | 66.2 | 85.9 |
| Arabidopsis thaliana | XM_011858 | 1964 | 501 | 66.2 | 85.2 |
| Bixa orellana | AJ542888 | 1500 | 499 | 71.5 | 86.5 |
| Brassica napus | HM989801 | 1739 | 493 | 67.9 | 85.9 |
| Brassica rapa | FJ606628 | 1758 | 493 | 67.8 | 85.3 |
| Capsicum annuum | GU085266 | 1497 | 498 | 71.3 | 86.6 |
| Carica papaya | DQ415894 | 1937 | 503 | 69.0 | 90.6 |
| Chrysanthemum x morifolium | A8205041 | 1705 | 498 | 71.3 | 84.3 |
| Citrullus lanatus | EF183521 | 2146 | 504 | 73.9 | 86.8 |
| Citrus maxima | AY217103 | 1678 | 504 | 76.5 | 90.4 |
| Citrus sinensis | DQ496222 | 1542 | 504 | 75.8 | 90.6 |
| Citrus unshiu | AY166796 | 1822 | 504 | 76.6 | 90.4 |
| Citrus x paradisi | JF907401 | 1515 | 504 | 78.6 | 90.4 |
| Cucumis melo | GU457407 | 1871 | 504 | 73.7 | 86.1 |
| Cucurbita moschata | JN559395 | 2121 | 497 | 70.3 | 85.9 |
| Daucus carota | DQ192190 | 2022 | 508 | 71.7 | 88.8 |
| Diospyros kaki | FJ940723 | 2231 | 504 | 73.5 | 89.2 |
| Gentiana lutea | A8017366 | 2147 | 508 | 70.4 | 86.5 |
| Glycinie max | XM_003554083 | 1928 | 507 | 88.0 | 95.2 |
| Ipomoea sp. | A8490055 | 1649 | 501 | 74.1 | 88.5 |
| Lycium barbarum | A906864 | 1527 | 508 | 71.3 | 86.3 |
| Medicago truncatula | XM_003624977 | 1497 | 498 | 83.6 | 91.6 |
| Populus trichocarpa | XM_002308867 | 1566 | 504 | 78.5 | 90.4 |
| Ricinus communis | XM_002513452 | 1968 | 514 | 74.1 | 89.4 |
| Salicornia europaea | AY789516 | 1937 | 498 | 71.1 | 83.8 |
| Solanum lycopersicum | NM_001247297 | 1503 | 500 | 75.0 | 87.6 |
| Tagetes erecta | AF251017 | 1906 | 511 | 67.8 | 86.2 |
| Taraxacum officinale | A8247456 | 1718 | 504 | 69.5 | 87.6 |
| **Gymnosperm members** |         |                 |                |                |                |
| Cryptomeria japonica | A8161829 | 1635 | 525 | 67.5 | 83.8 |
| Taxodium distichum | A8096608 | 1692 | 525 | 67.7 | 83.8 |
Table 3: Comparison of *Phaseolus vulgaris* beta-carotene hydroxylase (PvCHY-β) cDNA nucleotide and deduced amino acid sequence with CHY-β from monocot, dicot and algal members.

| Species            | GenBank Accession No. | Length compared | Similarity (%) |
|--------------------|-----------------------|----------------|----------------|
|                    |                       | nucleotide (bp)| amino acid      | nucleotide level | amino acid level |
| **Monocots**       |                       |                |                |                |                |
| Crocus sativus     | AJ416711              | 918            | 305            | 59.8           | 72.0            |
| Elaeis oleifera    | EU057623              | 1414           | 325            | 51.1           | 68.2            |
| Zea mays           | AY844956              | 1392           | 308            | 49.0           | 69.2            |
| **Dicots**         |                       |                |                |                |                |
| Adonis aestivalis  | EF120636              | 1187           | 309            | 58.7           | 73.5            |
| Brassica napus     | EF026098              | 903            | 300            | 64.9           | 74.3            |
| Capsicum annuum    | Y09225                | 1112           | 315            | 61.8           | 70.8            |
| Citrus sinensis    | DQ228870              | 936            | 311            | 71.1           | 75.3            |
| Coffea arabica     | DQ157169              | 933            | 310            | 62.5           | 70.4            |
| Dacus carota       | DQ192193              | 1231           | 309            | 58.9           | 68.6            |
| Diospyros kaki     | FJ790215              | 1365           | 312            | 61.3           | 70.1            |
| Gentiana lutea     | A0027187              | 1159           | 320            | 50.4           | 73.2            |
| Glycine max        | AY575953              | 1164           | 334            | 70.7           | 78.0            |
| Solanum lycopersicum | Y14809              | 1125           | 309            | 60.8           | 73.6            |
| Vitis vinifera     | AF499108              | 900            | 299            | 62.1           | 72.0            |
| **Algal members**  |                       |                |                |                |                |
| Chlamydomonas reinhardtii | XM_001698646 | 1757           | 297            | 48.0           | 60.3            |
| Muriella zofingiensis | EU016205               | 1545           | 299            | 45.6           | 58.0            |

Table 4: The conserved domains detected in *Phaseolus vulgaris* lycopene beta-cyclase (PvLCY-β) and beta-carotene hydroxylase (PvCHY-β) protein sequences.

| No | Conserved Domain [GenBank’s code] | Location within protein (amino acids) |
|----|-----------------------------------|---------------------------------------|
| 1  | Lycopene cyclase protein [pfam05834] | 85 - 480                              |
| 2  | Lycopene cyclase [TIGR01789]       | 87 - 482                              |
| 3  | Lycopene beta cyclase [PLN02463]   | 58 - 502                              |
| 4  | Lycopene epsilon cyclase [PLN02697] | 1 - 502                               |
| 5  | Lycopene cyclase family protein [TIGR01790] | 87 - 482                                       |
| 6  | Dehydrogenases [COG0644]          | 83 - 479                              |
| 7  | ZP6MH&RFADD0+ [COG0654]           | 84 - 481                              |
| 8  | Geranyl/geranyl reductase family [TIGR02032] | 53 - 380                              |
| 9  | FAD dependent oxidoreductase [pfam01266] | 1 - 448                              |
| 1  | Beta carotene hydroxylase [PLN02601] | 1 - 305                              |

*stands for 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases

Supplementary Figure 1: Individual amino acids% in the *P. vulgaris* lycopene β-cyclase (PvLCY-β) protein. On X axis, the amino acids are represented by a single letter code. The amino acid composition was calculated online using ProtCalc program of JustBio.
Supplementary Figure 2: Individual amino acids% in the *P. vulgaris* β-carotene hydroxylase (PvCHY-β) protein. On X axis, the amino acids are represented by a single letter code. The amino acid composition was calculated online using ProtCalc program of JustBio.

Supplementary Figure 3: Schematic diagram of carotenoid biosynthesis pathway in plants. Enzymatic conversion of the respective substrate is shown by arrows with the enzyme’s responsible for it. PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, zeta-carotene desaturase; CISO, carotene isomerase; LCY-β, lycopene beta-cyclase; LCY-ε, lycopene epsilon-cyclase; CHY-β, beta-carotene hydroxylase; CHY-ε, epsilon-carotene hydroxylase. The LCY-β and CHY-β are shown in pink and red color respectively to highlight their location in the pathway.