Gene 548 (2014) 223–233

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Phylogenetic analysis and evolutionary studies of plant carotenoid cleavage dioxygenase gene

R. Priya, Ramamoorthy Siva *

School of Bio Sciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 23 April 2014
Received in revised form 9 July 2014
Accepted 12 July 2014
Available online 14 July 2014

Keywords:
Carotenoids
Evolution of gene
Bayesian analyses
Sequence conservation
Functional divergence

ABSTRACT

The oxidative breakdown of carotenoid evidences the formation of apocarotenoids through carotenoid cleavage dioxygenases (*CCDs*). Numerous *CCDs* and apocarotenoids have been identified and characterized in plants. Using available sequence data, a study was performed to investigate the phylogenetic relationship among *CCD* genes and to statistically estimate the sequence conservation and functional divergence. In total, 77 genes were identified from 39 species belonging to 21 families. Our result of phylogenetic analysis indicated the existence of well-conserved subfamilies. Moreover, comparative genomic analysis showed that the gene structures of the *CCDs* were highly conserved across some different lineage species. Through functional divergence analysis, a substantial divergence was found between *CCD* subfamilies. In addition, examination of the site-specific profile revealed the critical amino acid residues accounting for functional divergence. This study mainly focused on the evolution of *CCD* genes and their functional divergence which may deliver an initial step for further experimental verifications.

1. Introduction

Carotenoids are natural tetraterpenoid pigments that have varied functions in plants and animals (Cazzonelli, 2011). In plants, they subsidize the photosynthetic machinery and prevent them against photodamage. As precursors of vitamin A and as antioxidant, carotenoids play a vital role in human nutrition (Bhakta and Siva, 2010). In addition, they also serve as yellow, orange, and red pigments in many flowers and fruits to attract animals for pollination and seed dispersal (Siva, 2007). Carotenoids are synthesized by all photosynthetic organisms and some non-photosynthetic bacteria and fungi. There are more than 700 carotenoids known that comprise into the large family of C_{40} polyenes. Carotenoids can be broadly split into two classes; carotenes that are purely hydrocarbons and contain no oxygen (like α-carotene, β-carotene, lycopene) and the xanthophylls that contain oxygen (like lutein, violaxanthin, zeaxanthin). The oxidative tailoring of carotenoids occurs in plants, animals, and microorganisms that leads to the formation of diverse bioactive products like vitamin A, plant hormone abscisic acid, numerous aroma compounds and apocarotenoid pigments (Bouvier et al., 2005).

Carotenoids can be cleaved at any of their conjugated double bonds, resulting in a diverse set of apocarotenoids (Auldridge et al., 2006a). The formation of apocarotenoids may result from nonspecific or site specific mechanisms. Three classes of models have been considered for nonspecific cleavage of carotenoid chromophores to apocarotenoids and it is difficult to understand the cleavage mechanism. The first class of enzymes is represented by lipoxygenase activity, which catalyzes the co-oxidation of β-carotene in the presence of polyunsaturated fatty acids like linoleic acid. The second class refers to xanthine oxidases, which catalyzes the co-oxidation of β-carotene in the presence of an aldehyde to yield various products like β-ionone and epoxy-β-ionone. The third class of enzymes is peroxidases, which cleave β carotene into β-ionone and several other minor derivatives. The formation of apocarotenoid can occur by nonspecific oxidation, but the biologically active forms with regulatory functions are generated via site-specific cleavage (Rodríguez-Ávila et al., 2011; Walter and Strack, 2011). Several enzymes that are capable of cleaving carotenoids at specific sites are considered to be involved in the synthesis of numerous apocarotenoids. In both plants and animals, apocarotenoids are responsible for the regulation of gene expression (Moise et al., 2005). Apocarotenoid pigments are commercially important, for example, bixin from *Bixa orellana* (Siva et al., 2010) and crocin from saffron are two different pigments used as a colorant in food, cosmetics and also used as alternative tracking dye (Siva et al., 2008). Apocarotenoids also play a key role in allelopathic interactions and plant defense mechanism (Bouvier et al., 2005).

Carotenoids are metabolized to apocarotenoids through the pathway catalyzed by carotenoid cleavage oxygenases (*CCOs*). Two types of carotenoid oxygenases have been identified, namely 9-cis epoxycarotenoid dioxygenases (*NCEDs*) and carotenoid cleavage dioxygenases (*CCDs*). *NCED* enzymes cleave the 11, 12 (11′, 12′) double bond of

Abbreviations: CCO, carotenoid cleavage oxygenase; CCD, carotenoid cleavage dioxygenase; NCED, 9-cis carotenoid cleavage dioxygenase; MEME, Multiple EM for Motif Elicitation.

* Corresponding author.
E-mail address: rsiva77in@rediffmail.com (R. Siva).

http://dx.doi.org/10.1016/j.gene.2014.07.037
0378-1119/

9-cis-violaxanthin or 9-cis-neoxanthin, catalyzing the first step in abscisic acid (ABA) biosynthesis. The *CCDs* catalyze the oxidative cleavage of double bonds in various positions in a variety of carotenoids. The first carotenoid cleaving enzyme (*Vp14*) was isolated from the ABA deficient viviparous maize mutant. Vp14 is an *NCED* and catalyzes the rate-limiting step in ABA biosynthesis, the cleavage of the 9-*cis*-isomer of neoxanthin or violaxanthin (Schwartz et al., 2004). Based on the sequence homology to *Vp14*, nine *CCDs* have been identified in the *Arabidopsis thaliana* genome.

In *Arabidopsis*, the *CCD* enzyme family consists of nine members that include five *NCEDs* and four *CCDs* viz., *CCD1*, *CCD4*, *CCD7* and *CCD8*. *CCDs* are distantly related to the *NCEDs*, and their substrate specificities and activities differ from those of the *NCEDs*. However, we have taken only *CCD* and their subclass genes to study the evolution of these genes. Owing to the relative high amino acid sequence similarity found among *CCD* proteins, few studies have investigated their relationships (Vallabhaneni et al., 2010). Phylogenetic study of protein family is an important tool to determine conserved or divergent regions, which can possibly lead to functional predictions. In this study, we interpreted the evolutionary relationship of the *CCD* protein family by a comprehensive phylogenetic approach and coefficients of type-I functional divergence, which might pave for functional analysis of these genes in future research.

2. Materials and methods

2.1. Data collection of CCD genes in plants

This study used amino acid sequences of all known *CCDs* that were identified by using the key term "carotenoid cleavage dioxygenase" in Protein Knowledgebase (UniProtKB) available at http://www.uniprot.org/. A collection of 77 sequence entries was identified from 39 plant species belonging to 21 different families comprising 31 (*CCD1*), 15 (*CCD4*), 14 (*CCD7*) and 17 (*CCD8*) genes. The different plant families included in this study were Actinidiaceae, Amaranthaceae, Apiaceae, Asteraceae, Brassicaceae, Bixaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Funariaceae, Iridaceae, Lauraceae, Oleaceae, Orchidaceae, Orobanchaceae, Poaceae, Rosaceae, Rubiaceae, Rutaceae, Solanaceae and Vitaceae. The dataset is constructed with great attention and comprehensive details of *CCD* gene sequences with their accession number were given in Table 1 (Agustí et al., 2007; Auldridge et al., 2006b; Baldermann et al., 2012; Bouvier et al., 2003; Cao et al., 2005; Carmen et al., 2009; Drummond et al., 2009; Guan et al., 2012; Huang et al., 2009a, 2009b; Ibdah et al., 2006; Ilg et al., 2009; Just et al., 2007; Kato et al., 2006; Ledger et al., 2010; Mathieu et al., 2006; Ohmiya et al., 2006; Proust et al., 2011; Rubio et al., 2008; Schwartz et al., 2001; Simkin et al., 2004a, 2004b, 2008; Snowden et al., 2005; Sun et al., 2008; Vogel et al., 2010).

2.2. Sequence alignment and trimming

To construct the phylogenetic tree of *CCD* gene family, the protein sequences of this gene present in various plants were retrieved from the database to infer the relationships among them. Initially the protein sequences were aligned using clustal W (http://www.ebi.ac.uk/tools/msa/clustalW2/) with default settings. Next, the sequence was aligned for trimming by using a trimAI tool. TrimAI is a tool for the automated removal of spurious sequence or poorly aligned regions from a multiple sequence alignment to increase the quality of phylogenetic analysis. TrimAI is available at http://phylemon2.bioinfo.cipf.es/cgibin/trimAI.cgi.

2.3. Phylogenetic analysis

The phylogenetic reconstructions were performed, first by using the neighbor-joining (NJ) method with 1000 bootstrap values through PHYLIP v.3.69. Further the maximum likelihood (ML) analysis was also performed using ProtTest v2.4 to verify the reliability of the NJ tree. To estimate the most appropriate model of amino acid substitution for tree building analyses the Akaike Information criterion (AIC) was implemented in ProtTest v2.4. Next, a rooted maximum likelihood tree was constructed using the PhyML v3.0 online program, according to the best-fit model predicted by ProtTest v2.4. Finally, the phylogenetic trees were displayed using MEGA v5.05 (Tamura et al., 2007). In MEGA v5.05, we also measured the bootstrap values deriving from 1000 randomized and replicated datasets and the following parameters were selected: model used was p-distance; and gap/missing data was pairwise deletion. The tree with the highest log likelihood and the percentage of trees in which the associated taxa clustered together were calculated. The tree result suggests that the *CCD* phylogeny is robust in different tree reconstruction methods. Consequently, only the NJ method is presented in this study.

2.4. Analysis of exon and poly A site in CCD gene family

To study the promising mechanisms of structural evolution of the *CCD* paralogous, the region of exons and poly A site was predicted in the *CCD* gene family. The gene sequences used for this analysis were retrieved from the GenBank (www.genbank.org) and the Gramene database (www.gramene.org). Sequences were carefully scrutinized and corrected for annotation errors before use. The position of exons and poly A site was confirmed through FGENESH. The FGENESH (Find GENES HMM) utilizes a statistical Hidden Markov Model to predict the position of exons and poly site, acquired from soft berry package.

2.5. Conserved region and motif identification

Program clustal W was employed to detect the conserved regions in protein sequences of *CCD* family and the motif regions were identified by MEME algorithm (http://meme.nbcr.net). In sets of biological sequences MEME (Multiple EM for Motif Elicitation) is one of the most widely used tools for searching for novel signals. MEME searches for statistically significant motifs in the input sequence set.

2.6. Positive selection assessment

Using a Bayesian inference approach, the site-specific positive and purifying selection was calculated for *CCD* genes with "The Selecton Server" (Doron-Faigenboim et al., 2005; Stern et al., 2007). Several evolutionary models implemented in this server, such as MEC (Mechanistic Empirical Combination Model), M5 (gamma), M7 (beta), M8a ($\omega_s = 1$) and M8 ($\omega_s \geq 1$), each of which adopts different biological assumptions and enables contrasting hypotheses through testing which model better fits the dataset. M8 permits for positive selection operating on the protein. A proportion p_0 of the sites are drawn from a beta distribution which is experiencing purifying selection and a proportion p_1 ($=1 - p_0$) of the sites are drawn from an additional category ωs (≥ 1) which is experiencing either neutral or positive selection. M8a model is similar to the M8 model, it does not allow for positive selection by setting $\omega_s = 1$ but it allows only the neutral and purifying selection. M7 model is also similar to M8, it allows mainly for purifying selection in the protein because it assumes only the beta distribution with no additional category. M5 model differs from other M models, in assuming Ka/Ks among sites as gamma distributed and thus may allow for purifying, neutral, and positive selection (Yang et al., 2000). The MEC model is entirely different from other four M models; it allows immediate substitutions between pairs of codons that differ at 2 or 3 codon positions and thus permits different replacement probabilities between amino acids. The advantage of the MEC model than the other models is that it considers different amino-acid replacement probabilities. In this model, Ka is computed differently and a position with drastic replacements will obtain a higher Ka value than a position with more moderate replacements. All these above said