Genome-wide identification, phylogenetic relationships, and expression analysis of the carotenoid cleavage oxygenase gene family in pepper

X.H. Zhang¹, H.Q. Liu¹, Q.W. Guo², C.F. Zheng³, C.S. Li¹, X.M. Xiang¹, D.F. Zhao¹, J. Liu⁴, J. Luo⁵, D.K. Zhao⁶, J.Q. Zheng⁶ and H.J. Wan¹

¹Quzhou Academy of Agricultural Sciences, Quzhou, Zhejiang, China
²Zhejiang Mariculture Research Institute, Wenzhou, China
³Wulanchabu Academy of Agricultural Sciences, Wulanchabu, Inner Mongolia, China
⁴Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang, China
⁵College of Agricultural, Yunnan University, Kunming, Yunnan, China
⁶Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang, China

Corresponding author: H.J. Wan
E-mail: wanhongjian@sina.com

Received April 5, 2016
Accepted May 16, 2016
Published October 5, 2016
DOI http://dx.doi.org/10.4238/gmr.15048695

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. Carotenoid cleavage oxygenases (CCOs) are a family of dioxygenases, which specifically catalyze the cleavage of conjugated...
double bonds in carotenoids and apocarotenoids in plants. In this study, genome-wide analysis of CCO genes in pepper plants was performed using bioinformatic methods. At least 11 members of the CCO gene family were identified in the pepper genome. Phylogenetic analysis showed that pepper and tomato CCO genes could be divided into two groups (CCDs and NCEDs). The CCD group included five subgroups (CCD1, CCD4, CCD7, CCD8, and CCD-like). These results indicate that there is a close genetic relationship between the two species. Sequence analysis using the online tool, Multiple Expectation Maximization for Motif Elicitation (MEME), showed that the CCO proteins comprise multiple conserved motifs, with 20 to 41 amino acids. In addition, multiple cis-acting elements in the promoter of CCO genes were identified using the online tool PlantCARE, and were found to be involved in light responsiveness, plant hormone regulation, and biotic and abiotic stresses, suggesting potential roles of these proteins under different conditions. RNA-seq analysis revealed that the CCO genes exhibit distinct patterns of expression in the roots, stems, leaves, and fruit. These findings suggest that the CCO genes have important roles in the vegetative and reproductive development of pepper plants.

**Key words:** Carotenoids; CCO genes; Bioinformatics; Pepper

**INTRODUCTION**

Carotenoids are a class of lipophilic compounds that contain C40 as their basic skeleton, and are comprised of several conjugated double bonds. To date, more than 700 types of C40 carotenoids have been identified (Britton et al., 2004). Carotenoids have important biological functions in organisms. Approximately 50 carotenoids are vitamin A precursors, which are needed for the human body to fight against cancer (Olson, 1989). Carotenoids are the main factors contributing to the yellow, orange, and red in color in some fruits (Nagal et al., 2012; Jabeen et al., 2013; Jarquín-Enríquez et al., 2013). In addition, some carotenoids participate in photosynthesis, and play an important role in light absorption and electron transport, and in the removal of triplet oxygen and superoxide anion species (Bartley and Scolnik, 1995; Tracewell et al., 2001; Woitsch and Römer, 2003). The results also showed that carotenoid cleavage products possess important biological functions. Some of the apocarotenoids derived from carotenoid cleavage are important determinants of flavor in agricultural products. In addition, some products of carotenoid-derived zeaxanthin aldehyde, which can be transformed into the plant hormone abscisic acid, can regulate stress, seed development, and other important functions (Winterhal and Schreier, 1995; Huang et al., 2009; Ilg et al., 2010; Liang et al., 2011; Heo et al., 2013; Sui et al., 2013). These cleavage products are mainly catalyzed by the carotenoid cleavage oxygenases (CCOs) (Bouvier et al., 2005; Heo et al., 2013).

CCOs are a class of dioxygenases, which specifically catalyze the cleavage of conjugated double bonds of carotenoids and apocarotenoids (Ilg et al., 2009; Walter et al., 2010). CCOs can be further classified as carotenoid cleavage dioxygenases (CCDs) and cis-epoxycarotenoid dioxygenases (NCEDs), based on their substrate informing an epoxy structure (Tan et al., 2003; Auldridge et al., 2006a). Recently, several studies have focused on
the identification and analysis of CCO genes of various plants (Simkin et al., 2004; Ohmiya et al., 2006; Sun et al., 2008; Adami et al., 2013; Liu et al., 2013). For example, the gene encoding the enzyme NCED (Vp14) was identified in maize (Tan et al., 1997; Woitsch and Römer, 2003). Homologs of this gene were subsequently discovered in other higher plants. In Arabidopsis, four of nine CCO genes are of the CCD type, and the remaining five have been identified as NCED genes (Schwartz et al., 2004; Auldridge et al., 2006b).

In the present study, 11 CCO genes were identified in the pepper genome. A comprehensive analysis of the CCO gene was performed, including sequence alignments, and determination of phylogenetic relationships and expression patterns. These results would aid in better understanding of the function and regulatory mechanisms of CCO genes in pepper.

MATERIAL AND METHODS

Identification of CCO genes in pepper plants

Pepper (Capsicum annuum L.) plant assembly and annotation V1.55 were downloaded from the PGP (Pepper Genome Platform) database (http://passport.pepper.snu.ac.kr/?t=PGENOME). A TBLASTp was performed using the protein coding sequence of the tomato CCO gene as the query against the pepper genome database. Subsequently, searches of candidate CCO genes in the pepper genome were repeated using BLASTp. The e-value used was 1e-5. Next, all candidate genes were evaluated for further verification using a Pfam database (http://pfam.janelia.org/), and SMART protein motif analyses (http://smart.embl-heidelberg.de/), in order to classify the CCO genes.

Alignment and phylogenetic analysis of CCO gene families

The amino acid sequences encoded by CCO genes in the pepper genome were aligned using Clustal X version 1.8, followed by manual adjustment, and were used to construct a phylogenetic tree using the Molecular Evolutionary Genetics Analysis software version 5.0 (MEGA 5.0) (Tamura et al., 2011). In addition, bootstrapping (1000 replicates) was performed to evaluate the degree of support for a particular grouping pattern in the phylogenetic tree of the CCO gene family. Missing sequence data were treated using pairwise deletions of the gaps. The branch lengths were assigned by utilizing pairwise calculations of the genetic distances.

Prediction of conserved motifs

To further investigate the diversity and structure of the CCO genes in pepper plants, their amino acid sequences were subjected to motif analyses using a Multiple Expectation Maximization for Motif Elicitation (MEME) (http://meme.sdsc.edu/meme/website/intro.html). Then, the optimal matching length of the parameter was set between 6 and 50, and all other parameters were set to the default values. Conservation of each motif among the CCO genes was determined with WebLogo version 2.8.2 (http://weblogo.berkeley.edu/) using the default settings.

Analysis of the CCO gene structure and its promoter

The exon and intron positions of pepper CCO genes were analyzed using the online
Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn), with both coding and genomic sequences. In addition, the 5′-upstream domain (1500 bp) of each CCO gene was downloaded from the PGP (http://peppergenome.snu.ac.kr/). The promoter sequences were then used to scan for cis-elements in the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Expression analysis of CCO genes

To determine the expression profiles of pepper CCO genes in different tissues and in response to different stress conditions, the table reads per kb per million reads (RPKM) of the pepper genes was searched against RNA-seq data of each tissue (http://www.nature.com/ng/journal/v46/n3/full/ng.2877.html) utilizing the locus ID given in the PGP (http://passport.pepper.snu.ac.kr/?t=PGENOME). The data obtained were analyzed and grouped based on tissue specificity.

RESULTS

Genome-wide identification of pepper CCO genes

In this study, 11 CCO genes were identified in the pepper genome, and named CaCCD1, CaCCD4a, CaCCD4b, CaCCD7, CaCCD8, CaCCD-like, CaCCD-like2, CaCCD-like3, CaNCED, CaNCED2, and CaNCED3, based on homology with tomato and Arabidopsis genes (Table 1). Gene names, gene locus, chromosome locations, gene lengths, exon numbers, protein lengths, molecular weights, and isoelectric points are also shown in Table 1. CCO genes ranged in length from 954 bp (CaNCED) to 15,419 bp (CaCCD1). The number of amino acids ranged from 317 amino acids (CaNCED) to 646 amino acids (CaCCD7). The molecular weight varied from 34.96 kDa (CaCCD1) to 72.72 kDa (CaCCD7), and the isoelectric point varied from 5.40 (CaCCD-like2) to 9.39 (CaNCED).

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Description</th>
<th>Gene locus</th>
<th>Chromosome location</th>
<th>Gene length (bp)</th>
<th>Number of exons</th>
<th>Protein size (aa)</th>
<th>pI</th>
<th>MW (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaCCD1</td>
<td>Carotenoid cleavage dioxygenase 1</td>
<td>CA01g20280</td>
<td>15</td>
<td>25363</td>
<td>12</td>
<td>579</td>
<td>5.74</td>
<td>68.59</td>
</tr>
<tr>
<td>CaCCD4a</td>
<td>Carotenoid cleavage dioxygenase 4a</td>
<td>CA08g03620</td>
<td>8</td>
<td>7891</td>
<td>2</td>
<td>886</td>
<td>6.86</td>
<td>65.85</td>
</tr>
<tr>
<td>CaCCD4b</td>
<td>Carotenoid cleavage dioxygenase 4b</td>
<td>CA08g04710</td>
<td>9</td>
<td>8262</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.39</td>
</tr>
<tr>
<td>CaCCD7</td>
<td>Carotenoid cleavage dioxygenase 7</td>
<td>CA11g14280</td>
<td>4</td>
<td>3139</td>
<td>2</td>
<td>366</td>
<td>4.10</td>
<td>41.30</td>
</tr>
<tr>
<td>CaCCD8</td>
<td>Carotenoid cleavage dioxygenase 8</td>
<td>CA11g20400</td>
<td>11</td>
<td>2000</td>
<td>7</td>
<td>350</td>
<td>5.86</td>
<td>59.49</td>
</tr>
<tr>
<td>CaCCD-like2</td>
<td>cis-spectrualized degradation, putative</td>
<td>CA08g09490</td>
<td>11</td>
<td>3623</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.85</td>
</tr>
<tr>
<td>CaCCD-like3</td>
<td>cis-spectrualized degradation, putative</td>
<td>CA01g20280</td>
<td>11</td>
<td>3623</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.85</td>
</tr>
<tr>
<td>CaNCED</td>
<td>cis-spectrualized degradation, putative</td>
<td>CA01g20280</td>
<td>11</td>
<td>3623</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.85</td>
</tr>
<tr>
<td>CaNCED2</td>
<td>cis-spectrualized degradation, putative</td>
<td>CA01g20280</td>
<td>11</td>
<td>3623</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.85</td>
</tr>
<tr>
<td>CaNCED3</td>
<td>cis-spectrualized degradation, putative</td>
<td>CA01g20280</td>
<td>11</td>
<td>3623</td>
<td>12</td>
<td>576</td>
<td>5.74</td>
<td>65.85</td>
</tr>
</tbody>
</table>

Phylogenetic relationships of pepper CCO genes

Phylogenetic analysis showed that pepper and tomato CCO genes could be divided into two groups (CCD and NCED groups). The former contained five sub-groups (CCD1, CCD4, CCD7, CCD8, and CCD-like), and the latter contained three sub-groups (NCED, NCED2, and NCED3) (Figure 1). Among these, SICCD1a, SICCD4b, SICCD7, and SICCD8 in tomato were orthologous to CaCCD1a, CaCCD4b, CaCCD7, and CaCCD8 in pepper, respectively. Furthermore, SINCED, SINCED2, and SINCED3 in tomato were orthologous to
CaNCED, CaNCED2, and CaNCED3 in pepper, respectively. However, there were two copies of CCD1 in the tomato genome (SlCCD1a and SlCCD1b) (Wei et al., 2016), while only CCD1 (CaCCD1a) was observed in the pepper genome. In contrast, only one member of the CCD-like (SlCCD-like) gene was found in tomato (Wei et al., 2016), and three homologous genes (SlCCD-like, SlCCD-like2, SlCCD-like3) in pepper were very similar to the CCD-like genes in tomato. These results indicated that some gene loss or gain events may have occurred during the course of evolution between the pepper and tomato plants.

Genome-wide analysis of the CCO gene family has been performed in several plant species, including Arabidopsis, tomato, rice, maize, and sorghum, which were found to contain nine, ten, five, six, and nine members, respectively (Auldridge et al., 2006b; Rubio et al., 2008). Therefore, the CCO gene family in plants might encode a small family of proteins with only a few gene members.

Figure 1. Phylogenetic tree of carotenoid cleavage oxygenase (CCO) proteins from pepper and tomato was constructed with the MEGA5.0 software using neighbor-joining method.

Multiple-sequence alignment and conserved motif analysis of the CCO proteins

To determine whether there was high-sequence homology between the CCO proteins from pepper and tomato, the amino acid sequences of all CCO genes in the pepper and tomato were aligned using the ClustalX software program. Poor conservation was observed in these proteins (Figure S1). This result is consistent with the findings in other plant species (Auldridge et al., 2006a,b; Vallabhaneni et al., 2010; Walter et al., 2010; Lashbrooke et al., 2013; Wei et al., 2016).

Using the MEME online software, the 10 conserved motifs (Motif1-Motif10) were identified in all CCO proteins from pepper plants (Table 2 and Figure 2). Seven genes (CaCCD1,
CaCCD4b, CaCCD-like, CaCCD-like2, CaCCD-like3, CaNCED2, and CaNCED3) contained all of the 10 conserved motifs. CaCCD4a contained eight conserved motifs, but not Motif5 and Motif9. CaCCD7 contained seven conserved motifs, but not Motif2, Motif5, and Motif8, and CaNCED contained five conserved motifs. CaCCD8 only contained two conserved motifs (Motif5 and Motif6).

In addition, the length of these conserved motifs ranged from 20 to 41 amino acid residues (Motif8 and Motif6, respectively). Motif7 was comprised of 30 amino acid residues. Motif9, Motif10 and Motif5 contained 29 amino acid residues. Motif1 was comprised of 28 amino acid residues. Both Motif2 and Motif3 contained 27 amino acid residues, and Motif4 contained 21 amino acid residues (Table 2).

**Table 2. Motifs of CCO proteins in pepper.**

<table>
<thead>
<tr>
<th>Motif</th>
<th>Width</th>
<th>Best possible match</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28</td>
<td>GTGVANTLEHYGGRYYAMAEDDMPYEI</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>RYGDENSISWFVPPCCFLHNWANWE</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>DEDDOGWAYTHENTWQSQVYHDAK</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>EMAYKLPREVPVYGFQDFSM</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>CPNGVYVRNGAPLFGPLAGHFWPDGQSM</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>MDRCMCHIDAFAIERYIHPDQLFPCQPMRGGQVIYD</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>GRWNEDQWNQSMTAIHPIPDVTGELFAMG</td>
</tr>
<tr>
<td>8</td>
<td>20</td>
<td>HDPRQCGQSNAPVPEQEP</td>
</tr>
<tr>
<td>9</td>
<td>29</td>
<td>KWNASYCCRYVQTDWFQNGKGRPSGFPK</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>WQWNMFMINEAVIGRKNKYYAQIANP</td>
</tr>
</tbody>
</table>

**Figure 2.** Distribution of conserved motifs of CCO proteins in pepper.

**Promoter analyses of CCO genes and exon-intron structure**

To investigate the transcriptional activity of pepper CCO genes, the promoter regions of each CCO gene (approximately 1500-bp DNA upstream sequences) were obtained from the pepper genome sequences, and the putative cis-regulatory elements were further analyzed using the PlantCARE online tool (Table S1). A total of 88 cis-regulatory elements were identified, which could be divided into six categories according to their putative functions. The
Characterization of pepper CCO gene family

first category included those involved in light responsiveness, such as ACE and G-Box. The second category contained eight types of *cis*-acting elements, which were associated with plant hormone responses, including: the TGA-element (auxin-responsive element); TCA-element (salicylic acid responsiveness); abscisic acid-responsive element (ABRE); CGTCA-motif, TGACG-motif (MeJA responsiveness); ethylene-responsive element (ERE); GARE-motif; and TGA-box. The third category responded to biotic and abiotic stresses, including: the TC-rich repeats (involved in defense and stress responsiveness); heat stress responsiveness (HSE); ARE (cis-acting-regulatory element essential for the anaerobic induction); Box-W1 (fungal elicitor-responsive element); MBS (MYB binding site involved in drought responsiveness); LTR (involved in low-temperature responsiveness); and WUN-motif (wound-responsive element). The fourth category was related to plant growth and development and included the following components: MSA-like, circadian, O2-site, CAT-box, CCGTCC-box, Skn-1_motif, GCN4_motif, AACA_motif, and MBSI. The fifth category involved elements associated with the activities of certain biological macromolecules, including 12 types of *cis*-elements. The final category included 19 types of *cis*-elements; however, their specific functions remain unclear. Therefore, it could be speculated that pepper CCO genes are involved in plant growth and development, and may also be induced in response to light and hormones, as well as biotic and abiotic stresses.

In addition, the exon-intron patterns of all pepper and tomato CCO genes were predicted using the GSDS online tool (Figure 3). The number of introns in pepper CCO genes ranged from 0 to 11. Four genes, *CaCCD1*, *CaCCD-like*, *CaCCD-like2*, and *CaCCD-like3*, contained 11 introns. No introns were observed in the *CaNCED*, *CaNCED2*, and *CaNCED3* genes. A similar pattern was found in tomato plants.

![Figure 3. Intron-exon structure of CCO genes in pepper and tomato.](image-url)
Expression analysis of $CCO$ genes in different pepper tissues using RNA-seq

To determine the transcript levels of $CCO$ genes in different pepper tissues, an RNA-seq transcriptome was selected for use in this study (Kim et al., 2014). A total of 24 samples were collected, including roots, stems, leaves, pericarp (PC), and placenta (PL) from the pungent cultivar CM334 and the non-pungent cultivar ECW30R plants, at 6-day post-anthesis (DPA), 16 DPA, 25 DPA, mature green (MG), breaker (B), 5-day post-breaker (B5), and 10-day post-breaker (B10). Figure 4 shows that most of the $CCO$ genes were expressed in tissues selected from the pungent cultivar CM334, with the exception of Ca$CCD1$ and Ca$CCD7$. Of the expressed genes, Ca$CCD4$ was expressed at high levels in the roots and at different stages of fruit development. High expression of Ca$CCD8$ was observed in the vegetative organs. Two genes, Ca$CCD$-like and Ca$NCED$, were expressed in all tissues. The remaining two genes, Ca$CCD4b$ and Ca$NCED2$, showed tissue-specific expression.

![Figure 4. Expression of $CCO$ genes in different tissues of the pungent pepper, CM334, and non-pungent pepper, ECW. PC and B indicate pericarp and breaker stage with days post-anthesis, respectively.](image)

The expression profiles of $CCO$ genes were compared in placental tissue from the pungent cultivar CM334 with placental tissue from the non-pungent cultivar ECW (Figure 5). These results showed that three genes ($CaCCD1$, $CaCCD4b$, and $CaNCED2$) were not expressed in any of the tissues studies. In contrast, two genes, $CaCCD$-like and $CaNCED$, were expressed in all of the tissues sampled. $CaCCD8$ was expressed at similar levels in the early stages of placental development in CM334 and ECW. Additionally, expression of $CaCCD7$ was observed in the ECW-PL-B plant, but not in the CM334-PL-B plant. Similarly, $CaCCD1$ was expressed at high levels in the ECW-PL-6DPA and ECW-PL-13DPA cultivars, but not in CM334-PL-6DPA and CM334-PL-16DPA.
DISCUSSION

The CCOs are a class of carotenoid cleavage oxygenases, which specifically catalyze the cleavage of conjugated double bonds in carotenoids and pro-carotenoids (Ilg et al., 2009; Walter et al., 2010). These proteins catalyze the cleavage of carotenoids, and help to determine the flavor quality of agricultural products, as well as the catalytic dehydration of the carotenoids, in order to produce a yellow aldehyde, and adjust the plant responses to stress (Espasandin et al., 2014). Therefore, it is beneficial to explore the biological characteristics of CCO genes.

In the present study, 11 members in the CCO gene family were identified in the pepper genome, while 10 members of the CCO gene family were identified in the tomato genome, which is suggestive of a gene duplication or deletion event following the differentiation of pepper and tomato plants. Three CCD-like genes (CCD-like, CaCCD-like3, and CaCCD-like2) in the CCD-like sub-group were identified in the pepper genome. However, only one CaCCD-like gene was identified in the tomato genome. The results of the present study suggest that duplication of CCD-like genes in the pepper genome occurred after the differentiation of pepper and tomato plants. However, only one CCD1 gene (CaCCD1) was observed in the pepper genome, and two CCD1 genes (SlCCD1a and SlCCD1b) were observed in the CCD1 sub-group in the tomato genome. These results indicate that deletion events involving the CCD1 gene may have occurred in the pepper genome.

In the present study, preferential expression of the CaCCD4a genes in roots and fruit was observed. This has previously been reported in other plant species (Sun et al., 2008; Ahrazem et al., 2010; Brandi et al., 2011; Adami et al., 2013). The CaCCD4b gene was expressed in a tissue-specific manner in the roots, which suggests a possible function in the growth and development of the roots. Additionally, the highest expression of CaCCD8 was
observed in the leaves, roots, and stems, followed by the early stages of fruit development. Recently, similar patterns of expression have been observed for orthologous genes of *CaCCD8* in other plant species, which indicates that the putative functions of *CCD8* genes are conserved in different plant species (Zhang et al., 2009; Ledger et al., 2010; Pasare et al., 2013). In conclusion, this comprehensive analysis of the *CCO* gene family lays the foundation for investigation of their potential roles in the future.

**Conflicts of interest**

The authors declare no conflict of interest.

**ACKNOWLEDGMENTS**

Research supported by the New Cultivar Breeding Program of Zhejiang Province (grant #2012C12903-1-11), the State Key Laboratory Breeding Base for Zhejiang Sustainable Pest and Disease Control (#2010DS700124-KF1516), the Young Talent Training Program of Zhejiang Academy of Agricultural Sciences (#2015R23R08E09), the Scientific Special of Northern Jiangsu Province (#BN2015096), and the Zhejiang Provincial Natural Science Foundation (#LQ13C030002).

**REFERENCES**


Characterization of pepper CCO gene family


Kim S, Park M, Yeom SI, Kim YM, et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46: 270-278. [dx.doi.org/10.1038/ng.2877]


**Supplementary material**

**Table S1.** Information on 88 different cis-regulatory elements identified in 11 *CaCCO* genes, including their numbers and the function of the respective elements.

**Figure S1.** Multiple-sequence alignments of SICCO and CaCCO proteins.