PLOS ONE

RESEARCH ARTICLE

A study of CCD8 genes/proteins in seven monocots and eight dicots

Ritu Batra[1], Priyanka Agarwal[1], Sandhya Tyagi[2], Dinesh Kumar Saini[1], Vikas Kumar[1], Anuj Kumar[3], Sanjay Kumar[4], Harindra Singh Balyan[1], Renu Pandey[2], Pushpendra Kumar Gupta[1]*

1 Department of Genetics and Plant Breeding, CCS University, Meerut, India, **2** Division of Plant Physiology, ICAR-Indian Agricultural Research Institute, New Delhi, India, **3** Advance Center for Computational & Applied Biotechnology, Uttarakhand Council for Biotechnology (UCB), Dehradun, India, **4** Bioinformatics Centre, Biotech Park, Lucknow, India

* pkgupta36@gmail.com

Check for updates

OPEN ACCESS

Citation: Batra R, Agarwal P, Tyagi S, Saini DK, Kumar V, Kumar A, et al. (2019) A study of CCD8 genes/proteins in seven monocots and eight dicots. PLoS ONE 14(3): e0213531. https://doi.org/10.1371/journal.pone.0213531

Editor: Guangyuan He, Huazhong University of Science and Technology, CHINA

Received: September 30, 2018

Accepted: February 23, 2019

Published: March 12, 2019

Copyright: © 2019 Batra et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

In plants, the enzyme CCD8 (carotenoid cleavage dioxygenase 8) is involved in the synthesis of an important hormone, strigolactone, and therefore, plays an important role in controlling growth and development. Using cDNA and protein sequence derived from the gene *ZmCCD8* from maize, we identified putative orthologs of the gene encoding CCD8 in six other monocots and eight dicots; the sequence similarity ranged from 52–75.9% at the gene level and 60.9–93.7% at the protein level. The average length of the gene was ~3.3 kb (range: 2.08 to 3.98 kb), although the number of introns within the genes differed (4 or 5 in dicots and 3 or 4 in monocots, except in *T. urartu* with 6 introns). Several cis-acting regulatory elements were identified in the promoters of CCD8 genes, which are known to respond to biotic and abiotic stresses. The N-terminal end (up to ~70 amino acids) of CCD8 proteins was highly variable due to insertions, deletions and mismatches. The variation in genes and proteins were particularly conspicuous in *T. urartu* and *Ae. tauschii* among the monocots and *A. thaliana* and *P. persica* among the dicots. In CCD8 proteins, 12 motifs were also identified, of which 6 were novel; 4 of these novel motifs occurred in all the 15 species. The 3D structures of proteins had the characteristic features of the related enzyme apocarotenoid oxygenase (ACO) of *Synechocystis* (a representative of cyanobacteria). The results of qRT-PCR in wheat revealed that under phosphorous (P)-starved condition (relative to expression under optimum P used as control), the expression of TaCCD8 genes increased ~37 fold in root tissue of the cultivar C306 and ~33 fold in shoot tissue of the cultivar HUW468 (the two cultivars differed in their P-use efficiency). This suggested that expression of TaCCD8 genes is genotype-dependent and tissue-specific and is regulated under different levels of P supply.

Introduction

In plants, CCD8 (carotenoid cleavage dioxygenase 8) is an important enzyme belonging to the family of CCD enzymes. They derive their name from the fact that the CCD-mediated reactions involve cleavage of carotenoids, involving use of dioxygen or molecular oxygen (O_2). The cleavage products in most cases are apocarotenoids, which carry each an aldehyde or ketone group at the cleavage site. CCDs are further classified into CCD1, CCD2, CCD4, CCD7 and CCD8 on the basis of cleavage position and/or their substrate preferences [1,2]. In different plant species, the CCD8 is also known by the following different names: MAX4 (Arabidopsis), RMS1 (pea), D10 (rice), and DAD1 (petunia) [1,2].

Among all the CCDs, CCD7 and CCD8 received major attention, since these are involved in the catalysis of two consecutive upstream steps of the biosynthesis of strigolactones (SLs), which are important plant hormones with diverse functions including their role as signaling molecules. Through a series of reactions including *cis-trans* isomerization, CCD8 is involved in the production of carlactone (Fig 1) which serves as a substrate for P_{450} enzymes, leading to the production of different forms of SLs [3–5]. Since the expression level of CCD8 directly determines the level of the synthesis of SLs, it is believed to be the critical enzyme in SL biosynthesis [6]. The CCD8 genes are also subjected to a feedback regulation [7], because the mutants exhibiting low concentration of SLs, have been shown to experience an upregulation of the expression of CCD8 in several plant systems including pea, petunia, kiwi fruit and the moss *Physcomitrella patens* [8–10]. CCD8 and SLs have also been shown to respond to abiotic stresses, particularly involving the conditions of low inorganic phosphate (Pi) [5,11].

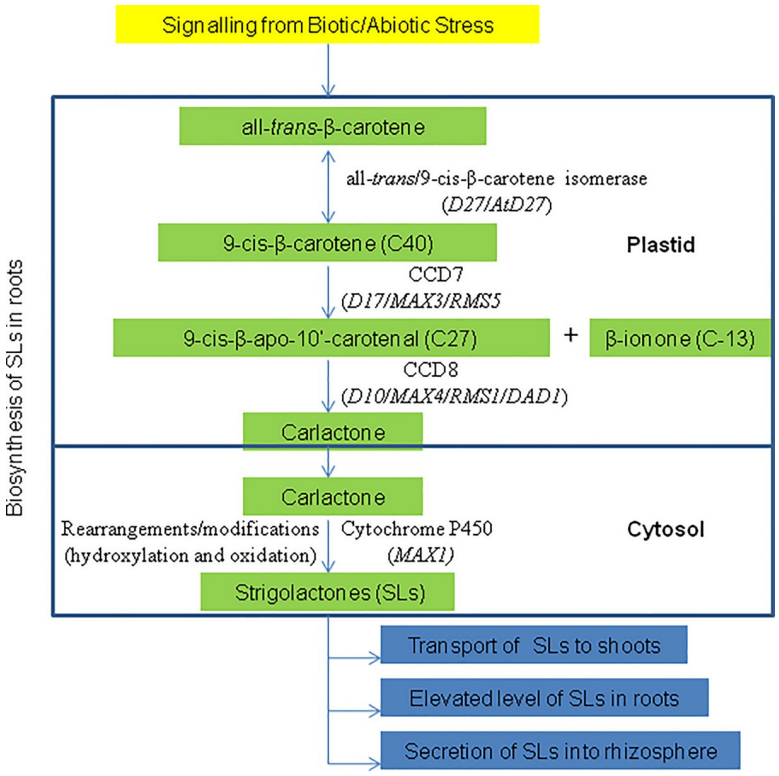

Fig 1. Pathway for the biosynthesis of strigolactone. *D*, dwarf; *AtD*, *Arabidopsis thaliana* dwarf; *MAX*, more axillary growth; *RMS*, ramosus; *DAD*, decreased apical dominance [3–5].

https://doi.org/10.1371/journal.pone.0213531.g001

The enzyme CCD8 is largely produced in the roots in a tissue-specific manner, although the gene is also known to be expressed in the shoot and inflorescence [12]. In the roots, the enzyme takes part in the synthesis of SLs, which at higher concentration modulate root system architecture (RSA). SLs are also secreted from the roots into the rhizosphere and exhibit interactions with arbuscular mycorrhizal fungi (AMF), which promote the absorption of nutrients by plants from the soil [13,14]. These are also transported from the roots to the aerial parts of the plant [1,5,15]. The CCD8 (through synthesis of SLs) also controls shoot architecture by inhibiting axillary bud growth and responds to abiotic stresses including low Pi in the soil [5,16,17]. Therefore, CCD8 knock-out mutants showing enhanced branching may be used for improving plant architecture for increased productivity[18–22].

The phylogenetic analysis of CCD8 genes from a number of monocots and dicots has also been conducted, suggesting conserved nature of these genes [23,24]. However, the earlier studies involving characterization of CCD8 genes cover only a few plant systems, including Arabidopsis (MAX4), pea (RMS1), rice (D10), petunia (DAD1), tomato, potato, and maize [6,17,19,24–27]. Therefore, the variation in the structure of these genes and the corresponding proteins in a wide spectrum of monocots and dicots still remains to be worked out.

The present report describes the structure of CCD8 genes and the corresponding proteins in 15 species, which include seven monocots (including maize) and eight dicots. *In silico* expression of CCD8 genes in seven of the above 15 species was also examined in different cells/tissues, at different developmental stages and under conditions of limited supply of phosphate. Using wet-lab experiments involving qRT-PCR, expression levels of TaCCD8 genes was also examined under different conditions of P supply in two wheat cultivars (C306 and HUW468), which are known to differ in their response to abiotic stresses. Evolutionary relationships among these CCD8 genes were inferred using CCD8 proteins. The 3D structures of the CCD8 proteins belonging to all the 15 species were studied for the first time, which will be helpful in the detailed analysis of their functions. The detailed structure of CCD8 genes resolved during the present study will help in the study of allelic variation of the CCD8 genes in monocots and dicots. This should prove useful in planning strategies to improve the plant architecture for enhanced productivity and for designing genotypes suitable for low input conditions.

Materials and methods

Identification of 'putative' orthologs

Full-length cDNA and protein sequence of *Z. mays* CCD8 were available in the database and were used as a reference in tBLASTx and tBLASTn analysis. Thus, 'putative' orthologs for the genes encoding CCD8 were identified in 14 other species of higher plants including six monocots (*Triticum aestivum*, *Triticum urartu*, *Aegilops tauschii*, *Oryza sativa*, *Brachypodium distachyon* and *Sorghum bicolor*) and eight dicots (*Arabidopsis thaliana*, *Glycine max*, *Vitis vinifera*, *Solanum lycopersicum*, *Theobroma cacao*, *Populus trichocarpa*, *Prunus persica* and *Medicago truncatula*). The methods used for the identification of putative orthologs were described earlier [28–30]. Briefly, the following criteria were used for the identification of 'putative' orthologs: (*i*) high level (>60%) of sequence similarity and high query coverage along the protein length; (*ii*) presence of all domains and motifs available in the original query sequence; (*iii*) conservation of the relative size and sequence among motifs and domains of the query sequence. The orthologous sequences thus obtained were used to identify full length gene sequences from Ensembl Plants (http://plants.ensembl.org/index.html).

Analysis of gene structure, synteny and collinearity

Intron-exon junctions in the full length gene sequences were determined using the genomic and coding DNA sequences (CDS) for different species. Intron phases (phase 0,1,2) were identified following criteria used by us earlier [29]. Ka/Ks values defining the ratio of non-synonymous to synonymous substitutions were calculated using MEGA version 6.0.6 (dated April 2015) employing Juke Cantor substitution model [31].

The gene sequences were also evaluated for the presence of simple sequence repeats (SSRs) and retro-elements. For this purpose, repeatmasker version 4.0.5 (http://www.repeatmasker.org/), with default parameters was used. One kb genomic regions upstream of the translation start site (ATG) were evaluated for the presence of cis-regulatory response elements in the promoter regions, using PlantCARE database [32]. Only the response elements on the sense strand showing a matrix value of $\geq$5 were accepted, following the criteria used by us earlier [29].

Using blocks of 31 genes including CCD8 gene of maize (15 genes flanking each of the two borders of CCD8 gene), synteny and collinearity of the orthologs were studied using the online tool Genomicus [33].

Protein analyses

Multiple sequence alignment. The consensus amino acid sequence was generated through multiple sequence alignment of amino acid sequences belonging to all the orthologs through Geneious software ver 6.6.1 with default settings (http://www.geneious.com). The criteria used for this purpose are described elsewhere [34]. For protein sequence similarity, amino acids at different positions of CCD8 protein in a particular species were compared with the consensus sequence. In monocots, a similarity scale of 1–9 (a separate scale for each of the three wheat homoeologues), and in dicots, a similarity scale of 1–8 was used. A value of zero indicated complete lack of similarity with the consensus, while a value of 9 in monocots and 8 in dicots suggested conservation of amino acids in all the species. Conserved domain in the consensus protein sequence was identified through CDD analysis (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Physico-chemical properties. The ProtParam tool (http://web.expasy.org/protparam/) of ExPASy was used to compute the following physico-chemical properties: (i) amino acid composition (%), (ii) molecular weight, (iii) theoretical isoelectric point (pI), (iv) number of positively/negatively charged residues, (v) instability index, (vi) aliphatic index, and (vii) Grand Average of Hydropathy (GRAVY). Similarly, properties at the secondary level were computed using SOPMA = Self Optimized Prediction Method with Alignment (http://npsapbil.ibcp.fr/cgibinnpsaautomat.pl?page=/NPSA/npsa_sopma.html) tool of NPS (Network Protein Sequence Analysis). Motif search was performed using MEME suite (http://meme-suite.org/) using the following parameters: (*i*) maximum number of motifs being 50; (*ii*) optimum width of the motif being 50; and (*iii*) the site being 2–600. The sub-cellular localization was detrmined using ProtComp (http://www.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc).

3D structures. The 3D structures were generated employing Swiss Model using as template the 3D X-ray crystallographic structure of a mutant apocarotenoid oxygenase (ACO) of *Synechocystis* (PDB id: 5kk0.1.A). 3D structures thus generated were verified by both geometric and energetic means using the following servers: (*a*) Structure Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES), and (*b*) Swiss-Model server using structure assessment tool. SAVES employed the following software: (*i*) PROCHECK to find out the relative proportion of amino acids, which fall in favoured region, relative to other

regions [35]; (*ii*) VERIFY3D to determine the compatibility of an atomic model (3D) with its own amino acid sequence [36] and (*iii*) ERRAT to analyse the statistics of non-bonded interactions between different atom types [37].

Molecular dynamics simulation, superimposition, ligand binding site analysis and functional annotation of 3D structures

Initially, structure topology files for 3D protein model were prepared by applying Optimized Potential for Liquid Simulations (OPLS) all atom force fields [38,39]. Each protein model was solvated by adding SPC water atoms in cubic periodic box. Ions were added using Genion tool. Energy minimizations were done by applying 1000 steps of steepest descent minimization algorithm followed by 50000 steps of conjugate gradient to minimum energy. Two equilibrium minimizations (NVT and NPT) were done by applying 50000 steps leap-frog integrator for 100ps. Four steps of LINCS algorithm [40] were applied for constrain bond parameter minimizations; Partial Mesh Ewald [41] method was applied for long range electrostatics minimization under 300K temperature. For pressure coupling, the method of Parrinello-Rahman [42] was used, where a Langevin thermostat was used for temperature control. The periodic boundary conditions and SHAKE algorithm [43] were applied for each system for 10 ns of molecular dynamic simulation.

The compactness of the predicted 3D models during MD simulation analysis was examined using radius of gyration (Rg) of each protein, assuming that relatively steady value of Rg would suggest stably folded proteins; Rg value that changes over time was taken to suggest an unfolded protein.

FATCAT server [44] was used to confirm the 3D structures by superimposing the energy minimized 3D structure of CCD8 for each plant species on the 3D structure of maize CCD8. Using energy minimized 3D structures, COACH metaserver (http://zhanglab.ccmb.med.umich.edu/COACH/) and ProFunc server (http://www.ebi.ac.uk/thornton-srv/databases/profunc/) were used to predict the ligand binding sites and to perform functional annotation, respectively.

Phylogenetic analysis

Phylogenetic analysis was based on amino acid sequences of CCD8 proteins and was undertaken using MEGA version 6 [45]. Phylogenetic tree was constructed using maximum likelihood method, with a bootstrap involving 1000 iterations.

In silico expression

Using microarray data in Genevestigator platform, *in silico* expression analysis was conducted for the CCD8 orthologs of the following seven species: *Z. mays*, *T. aestivum*, *O. sativa*, *B. distachyon*, *A. thaliana*, *G. max*, and *M. truncatula*. The expression analysis was conducted in different tissues, at different developmental stages and under varying levels of phosphate.

qRT-PCR analysis of TaCCD8 genes

Plant growing conditions. Seeds of two wheat cvs., namely C306 and HUW468 (which are known to differ for their response to abiotic stresses) were surface sterlized with 0.1% $HgCl_2$ for 2 min followed by 5–6 washings with distilled water. The seedlings were raised in hydroponics under controlled conditions at National Phytotron Facility, ICAR-IARI, New Delhi following the method described earlier [46]. Five days after germination, the seedlings were transferred to plastic containers (10 L capacity) in Hoagland solution with low (5 μM)

and optimum (500 μM) phosphorous (P) concentrations. The nutrient solution was changed every third day and pH of the nutrient solution was maintained at 5.6 throughout the experiment. Fresh root and shoot tissue samples of the control (optimum P; 500 μM) and treatments were collected on 21st, 24th and 25th days in the following order: (i) on 21st day from seedlings grown in low P (LP; 5 μM); (ii) on 24th day from seedlings that were completely starved of P for 3 days after 21 days in low P (P starvation = PS; 0 μM); (iii) on 25th day from seedlings after optimum P was restored (PR; 500 μM) for one day after 24 days. All experiments were carried out in two replications.

RNA isolation, cDNA synthesis, primer design and qRT-PCR analysis. Total RNA was isolated from root and shoot tissues using a TRI reagent (Sigma) followed by RNase-free DNase I (Qiagen) treatment for removal of DNA contamination. Reverse transcription reactions were performed using 2.0 μg of total RNA and M-MuLV Reverse Transcriptase (Promega) according to the manufacturer's instructions.

A common set of primers for the three wheat TaCCD8 genes (located on chromosomes 3A, 3B and 3D) was designed using Primer3 software (forward primer: `GCAGCCTCTCGCGGCTG`; reverse primer: `TCTGTGACGGCGGCAGC`). qRT-PCR was performed with PikoReal Real-Time PCR Systems (Thermo Scientific) using PowerUp SYBR Green Master Mix (Applied Biosystems) in two biological replicates each containing three technical replicates. Following cycles of reactions were performed under the following conditions: 95˚C for 30 sec, 40 cycles involving 95˚C for 5 sec, and 60˚C for 34 sec. Constitutive expression of *TaAct2* gene of wheat was used as an endogenous control. The transcript abundance for each gene was normalized with the internal control. The $2^{-\Delta\Delta}$Ct values [fold change in gene expression under low P (LP), P starvation (PS) and P replete (PR) conditions vs. the control] were calculated as follows: $2^{-\Delta\Delta}$Ct = [(Ct LP/PS/PR test–Ct LP/PS/PR *TaAct*) − (Ct cont test–Ct cont *TaAct*)] [47]. Water in place of cDNA was used as the negative control.

Results and discussion

CCD8 genes in 15 species

The details of CCD8 genes for the 15 species (including *Z. mays*) are presented in Table 1. In each species examined, a single ortholog for CCD8 was found except in case of wheat, where three orthologs were found, one each for the three sub-genomes (A, B and D sub-genomes).

Comparative gene structure

The lengths of CCD8 genes ranged from 2.08 kb (*T. urartu*) to 3.98 kb (*P. trichocarpa*). The differences in lengths were primarily due to variations in number/size of exons/introns and the size of UTRs (Fig 2 and Table 1). The average per cent similarities of cDNA and CDS sequences with those of maize were higher in monocots (cDNA: 77.82–91.48%; CDS: 80.84–94.76%) relative to those in dicots (cDNA: 56.59 to 64.65%; CDS 61.56–67.59%) (Table 1). The observed higher similarity of cDNA and CDS sequences of monocots can be attributed to maize itself being a monocot, which diverged from dicots some 200 MYA [48]. Among the monocots, the per cent similarities of the cDNA and CDS sequences were highest with sorghum because of close evolutionary proximity between sorghum and maize [49]. Among the dicots, the cDNA and CDS sequences of *V. vinifera* showed maximum similarity, presumably due to its relatively close relationship with maize.

The number of exons in individual CCD8 genes ranged from 4 to 7, which was also reflected in the number of introns (3/6; Fig 2 and S1 Table) and may be attributed to gain/loss of introns. Similar results for genes encoding AGPase enzymes were earlier reported by us [29]. Of particular interest are the CCD8 genes of *T. urartu*, *Ae. tauschii* and *P. persica*, where

Table 1. Details of CCD8 genes and corresponding cDNA and CDS in 15 selected species.

Species	Ensembl Plants Id's	Gene		cDNA		CDS	
		Length in bp (chr. no.)	Similarity (%)	Length in bp	Similarity (%)	Length in bp	Similarity (%)
Monocots							
Z. mays	Zm00001d043442_T001	3522 (3)	100	2191	100	1719	100
T. aestivum -A sub-genome	TraesCS3A02G274300	3732 (3)	65.98	2196	78.44	1683	87.19
- B sub-genome	TraesCS3B02G308000.1	3748 (3)	63.95	2210	77.82	1683	86.71
-D sub-genome	TraesCS3D02G273500.1	3718 (3)	65.42	2196	79.33	1683	86.77
T. urartu	TRIUR3_28465-T1	2088 (3)	75.81	1404	80.84	1404	80.84
Ae. tauschii	EMT16146	3172 (3)	72.06	1524	85.31	1524	87.53
O. sativa	OS01T0746400-00	3109 (1)	72	1710	87.66	1710	89.05
B. distachyon	BRADI2G49670.1	3577 (2)	68	2140	80.13	1713	86.23
S. bicolor	Sb03g034400.1	3545 (3)	75.91	1867	91.48	1740	94.76
Dicots							
A. thaliana	AT4G32810.1/Max4	3265 (4)	52.8	2026	56.59	1713	61.99
G. max	GLYMA06G09000.2	3891 (6)	52.72	2049	60.53	1692	65.88
V. vinifera	VIT_04s0008g03380.t01	2823 (4)	58.83	1782	64.65	1641	67.59
S. lycopersicum	Solyc08g066650.2.1	3075 (8)	52.36	1907	57.96	1674	61.56
T. cacao	EOY29749 (TCM_037195)	3680 (9)	52.04	2051	58	1680	63.75
P. trichocarpa	POPTR_0006s25490.1	3983 (6)	54.84	1674	63.79	1674	64.84
P. persica	EMJ23585 (PRUPE_ppa006042mg)	2893 (1)	56.11	2436	58.41	1296	67.28
M. truncatula	AES73861 (MTR_3g109610)	3439 (3)	52.31	2057	57.35	1698	63.28

https://doi.org/10.1371/journal.pone.0213531.t001

the structure of CCD8 genes deviated significantly from the general pattern. The CCD8 gene of *T. urartu* (progenitor of the A genome of wheat) surprisingly deviated from that of the CCD8 gene from A sub-genome of bread wheat; the first 396 base pairs (encoding 132 amino acids) differed in CCD8 genes of *T. urartu* and *T. aestivum* sub-genome A. The CCD8 of

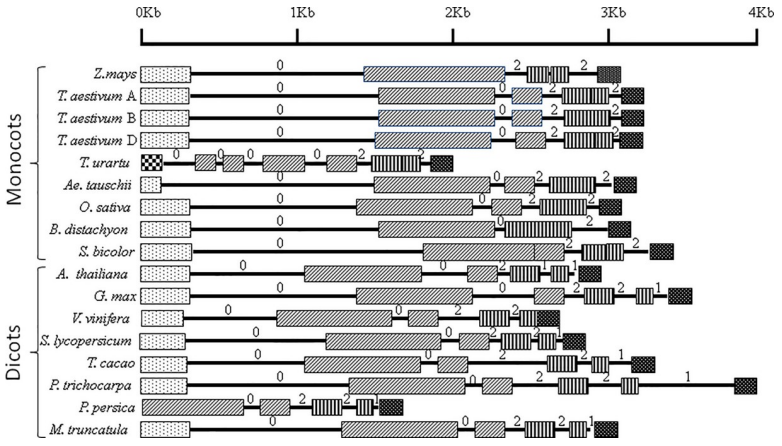

Fig 2. Representation of CCD8 structural genes (from translation start to stop sites) in seven monocots and eight dicots (for wheat, there are three genes, one each on 3A, 3B and 3D chromosomes). Solid patterned boxes indicate exons and lines connecting the exons indicate introns (a box with two patterns represents one exon that resulted due to fusion of two exons from maize *ZmCCD8*). Exons are coded based on the sequence similarity with the respective exons in the *ZmCCD8* (used as reference). Intron phases 0, 1 and 2 are marked above each intron. First exon in *T. urartu* (donor of sub-genome A of *T. aestivum*) has sequence similarity with a part of 1st intron of *T. aestivum*.

https://doi.org/10.1371/journal.pone.0213531.g002

T. *urartu* also carried a deletion of 240 bp, and the first exon of *T*. *urartu* (156 bp) resembled a part of first intron of CCD8 of *T*. *aestivum*. Apparently, these deviations in CCD8 structure of *T*. *urartu* occurred after the incorporation of A genome of *T*. *urartu* in 4x/6x wheat. Further, the deviations in *T*. *urartu* also included deletions of segments of coding sequences leading to deletions of amino acids at postions 192–194, 201–202, 207–212, 217–219 and 275–304. A deletion of 180 bp encoding 60 aa (72–131 aa) also occurred in CCD8 gene of *Ae*. *tauschii*, the donor of the D genome. In CCD8 gene of *P*. *persica* also, the first exon, first intron and a part of the second exon were absent.

Relative to the reference CCD8 gene of *Z*. *mays*, both loss and gain of introns was observed in CCD8 genes of all the other species, except *S*. *bicolor*, where the intron and exon numbers were similar to those of the *Z*. *mays*. Similar gain and loss of introns was also reported earlier in some other cases [50–51]. The intron/exon boundaries (splice sites) did not differ from the known conserved GT/AG boundaries [52]. Intron phases included all the three phases, the intron phase 1 was absent in monocots and relatively infrequent in dicots (phase 1 being 20% as against phase 0 being 40%). These results suggest that only 40% intron phases (phase 0 = 40%) are conserved, which allowed conservation of codons in the reading frame [53]. However, *A*. *thaliana*, *V*. *vinifera* and *P*. *persica* did not follow this general pattern (Fig 2). The intron insertion sites and the intron phases in monocots did not differ from those in the ancestral genome [54].

Individual exons and introns differed in size, the length of exons being 103–938 bp and that of introns being 41–1471 bp. The total length of exome ranged from 1296 bp to 1740 bp (Fig 2 and S1 Table). The average sequence similarity for exons was higher in monocots (81.67 to 94.79%) than in dicots (60.86 to 66.53%) (S2 Table).

The average Ka/Ks value for CCD8 genes was 3.3 (S3 and S4 Tables), suggesting that CCD8 genes have undergone positive selection during speciation [55].

Synteny/Collinearity analysis

Analysis of synteny conservation was undertaken using a block of 31 genes, including 15 genes flanking either side of the CCD8 gene on *Z*. *mays* chromosome 3 (S5 Table), which corresponds to wheat homoeologous group 3, rice chromosome 1, and Brachypodium chromosome 2 [56–57]. This analysis was possible for only 13 of the 15 species, since the genome sequences of *T*. *aestivum* and *M*. *truncatula* could not be utilized by Genomicus. Some degree of synteny conservation was observed among *Z*. *mays*, *S*. *bicolor*, *B*. *distachyon* and *O*. *sativa* (S1 Fig). Out of the 30 genes flanking the *ZmCCD8*, only 12 genes were syntenous in *S*. *bicolor* and *B*. *distachyon* each and 8 genes were syntenous in *O*. *sativa*. Even in these three species (*S*. *bicolor*, *B*. *distachyon* and *O*. *sativa*), the collinearity within the synteny block was rather disrupted. Some degree of synteny observed in the present study is in agreement with an earlier study, where CCD8 genes of these species were reported to occupy orthologous positions [6]. The loss of shared synteny in most of the species may be attributed to rearrangements in the genomes during the course of evolution [58]. This partial syntenic relationship was further confirmed through a study of the level of orthology between maize, wheat, rice and Brachypodium.

SSRs and retro-elements

As many as 24 SSRs were detected in different regions (exons, introns, UTRs) of CCD8 genes of 12 of the 15 species examined (no SSRs were available in CCD8 genes of *O*. *sativa*, *V*. *vinifera* and *S*. *lycopersicum*). The repeat units in SSRs ranged in size from 1–7 (mononucleotide repeats to heptanucleotide repeats), and the number of SSRs per CCD8 gene ranged from 1 to 4 (for details, see S6 Table). As known, the presence of SSRs within genes can lead to (*i*) a gain

or loss of gene function, (*ii*) affect transcription and translation, (*iii*) mRNA splicing, or (*iv*) export to cytoplasm. All these effects eventually lead to phenotypic changes [59].

Retro-elements in the form of a solitary SINE element and a solitary Helitron occurred in only one species, namely *P. trichocarpa*; no other species carried any retro-elements. The polymorphic SSRs and the polymorphic retro-elements may be utilized for developing markers for use in marker-aided breeding programmes aimed at improvement of either shoot branching or phosphate use efficiency in crops like wheat.

Promoter analysis

Promoter analysis allowed identification of some cis-acting regulatory elements (TATA, CAAT) that were common to all genes. Other cis-acting elements included light response elements (Sp1, Box 4, Box 2, MNF1, G-Box, GAG motif, GT1, TCT, I-Box, ATCT motif, 3-AF1, ACE) and response elements for tissue-specificity (Skn_1 and GCN4); multiple copies of some response elements were also common in the promoter regions of some CCD8 genes (Fig 3). These cis-elements should impart light responsive and tissue specific expression of CCD8 genes during plant development [60].

The presence of response elements for auxin (TGA, Aux-RR) and gibberellic acid (GARE, P-Box, TAGTTA) in the promoter regions of CCD8 genes in some of the species suggested that the expression of CCD8 genes and SL level is also regulated by the presence of other hormones. Recent studies demonstrated that auxin and strigolactone modulate the levels and distributions of each other, thus forming a dynamic feedback loop between the two hormones [61]. There were four response elements for biotic (CGTCA motif, TCA, TGACG element, MBSII) and the same number for abiotic stresses (ABRE, MBS, TC-rich repeats, CCAAT, HSE). This confirmed that the level of SLs is regulated by a variety of stresses including low P conditions, drought, salinity and plant-microbe interactions [62].

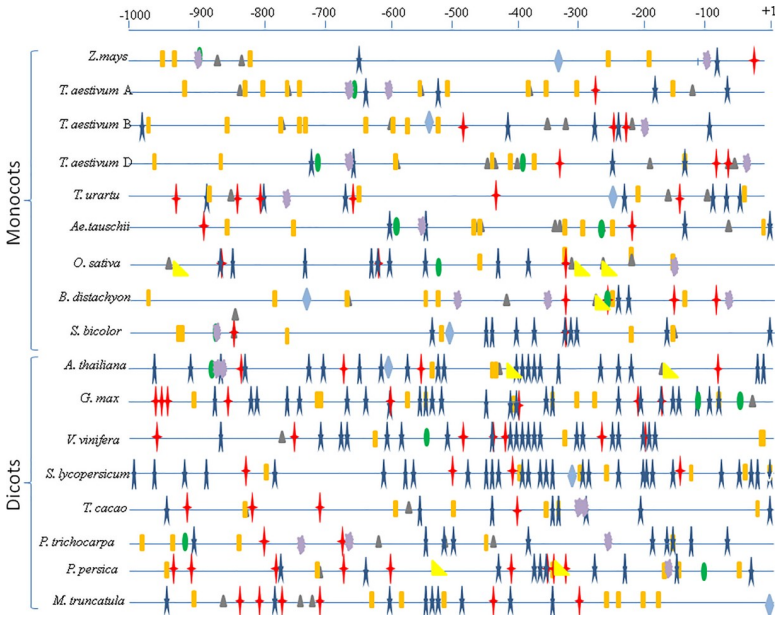

Fig 3. Regulatory elements identified in 1 kb upstream region of the translation start site (ATG) of CCD8 genes. Different symbols indicate major regulatory elements. TATA box (star), CAAT box (plus), light response elements (rectangle), abiotic stress response elements (triangle), biotic stress response elements (dispersed), endosperm-specific response elements (oval), auxin response elements (diamond) and gibberellic acid response elements (right angle triangle).

https://doi.org/10.1371/journal.pone.0213531.g003

Protein sequence and structural analyses

Primary structure. The length of predicted CCD8 proteins ranged from 431 amino acids (aa) in *P. persica* to 579 aa in *S. bicolor* (Table 2). This range matches exactly the range in the size of exomes (1296–1740 bp; 3 bp = one amino acid) recorded in the gene structure described above in this paper (Fig 2 and S1 Table). The difference of a triplet between the coding sequence and a single aa in the translated product is primarily due to the termination codon, which is a part of exome but does not code for any aa.

The variations in the length of CCD8 proteins were mainly due to deletions, insertions and mismatches, particularly in the hypervariable N-terminal region (up to ~70 aa), which corresponds to the variation in gene structure described above in this paper (comparative gene structure section). Another conspicuous deviation was observed in CCD8 protein of *P. persica*, where absence of ~125 aa residues also corresponds to the deviation in CCD8 gene of this species. Maximum number of extra aa residues were present in *A. thaliana* CCD8 protein. For instance, valine, isoleucine, asparagine, histidine, tryptophan, aspartic acid were present at positions 203–206 (isoleucine at positions 204 and 205), 274, 325 and 406, respectively (Fig 4 and S2 Fig).

A single conserved domain that occurred in the consensus sequence (116–592 aa) belongs to RPE65 super-family (RPE65 = retinal pigment epithelium-specific 65), which was also a characteristic feature of CCD8 proteins of all the species examined. Among all these species, the size of this domain was conserved ranging from 412 aa to 475 aa (Table 2). The similarity of this domain with the maize CCD8 domain was higher in monocots (82.34 to 94.75%) than in dicots (61.24 to 76.10%) (Table 2). RPE65 domain is a characteristic feature of all the

Table 2. Details of CCD8 proteins and their conserved domains in 15 selected species.

Species	Ensembl Plants Id's	Protein		Protein domain RPE65	
		Length in aa	Similarity (%)	Position; length in bp	Similarity (%)
Monocots					
Z. mays	Zm00001d043442_T001	572	100	103–564; 462	100
T. aestivum -A sub-genome	TraesCS3A02G274300	560	84.35	93–555; 463	84.27
-B sub-genome	TraesCS3B02G308000.1	560	84.71	93–555; 463	84.81
-D sub-genome	TraesCS3D02G273500.1	560	84.35	93–555; 463	84.45
T. urartu	TRIUR3_28465-T1	467	76.87	51–462; 412	82.34
Ae. tauschii	EMT16146	507	86.19	49–502; 454	86.17
O. sativa	OS01T0746400-00	569	86.99	99–561; 463	86.07
B. distachyon	BRADI2G49670.1	570	84.11	104–565; 462	84.11
S. bicolor	Sb03g034400.1	579	93.71	110–571; 462	94.75
Dicots					
A. thaliana	AT4G32810.1/Max4	570	60.87	92–566; 475	61.24
G. max	GLYMA06G09000.2	563	69.08	97–559; 463	69.1
V. vinifera	VIT_04s0008g03380.t01	546	70.66	80–542; 463	72.08
S. lycopersicum	Solyc08g066650.2.1	557	69.02	90–553; 464	70.43
T. cacao	EOY29749 (TCM_037195)	559	66.85	92–555; 464	67.96
P. trichocarpa	POPTR_0006s25490.1	557	71.25	91–553; 463	71.72
P. persica	EMJ23585 (PRUPE_ppa006042mg)	431	76.1	1–427; 427	76.1
M. truncatula	AES73861 (MTR_3g109610)	565	68.17	98–561; 464	68.91

aa, amino acid

https://doi.org/10.1371/journal.pone.0213531.t002

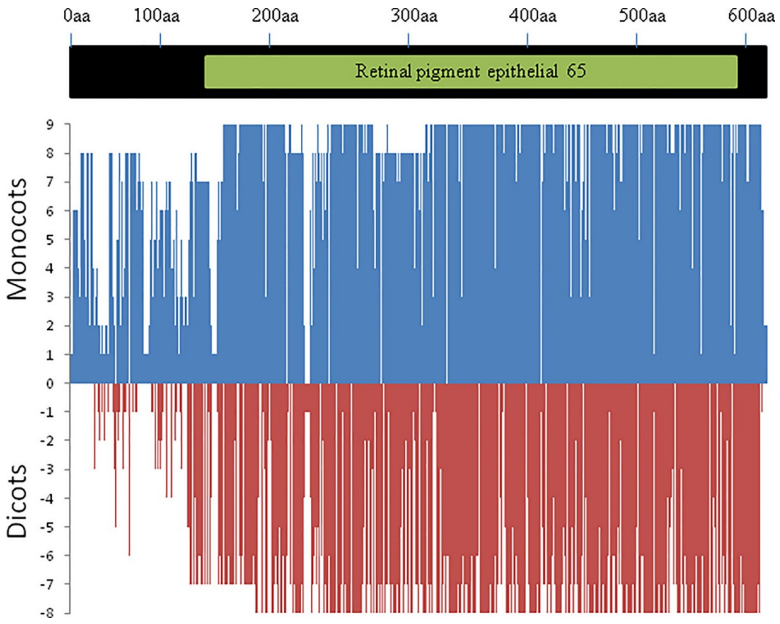

Fig 4. Amino acid sequence similarity of CCD8 proteins among seven monocots and eight dicots with respect to consensus sequence (for wheat, there are three CCD8 proteins derived one each from 3A, 3B and 3D chromosomes). Position 0 (on y-axis) indicates amino acid (aa) consensus sequence. Presence of similar aa residues against consensus is plotted on a scale of 1 to 9 in monocots and -1 to -8 in dicots. Consensus protein on top is indicated by solid-bar; a small-bar within solid-bar indicates the position of the conserved domain.

https://doi.org/10.1371/journal.pone.0213531.g004

enzymes (including CCD8) that are involved in the biosynthesis of apocarotenoids, which are the intermediate products in the synthesis of strigolactones catalysed by CCD8. Therefore, the role of the domain RPE65 in CCD8 is rather general in nature, and not specific; this is known to be a common feature in many enzymes that are involved in a variety of biosynthetic pathways.

Physico-chemical analysis. The molecular weight of CCD8 proteins ranged from 47.66 to 63.96 kD. The total number of negatively charged aa residues exceeded the total number of positively charged aa residues. The isoelectric point was within the acidic range (5.75–6.99) indicating that the proteins encoded by CCD8 genes of all the selected species are sparingly soluble in aqueous medium [63]. The aliphatic index (relative volume occupied by aliphatic/hydrophobic aa: glycine, alanine, valine, isoleucine, and leucine) was relatively higher in dicots (78.06–83.29) than in monocots (75.13–78.73). It has been reported that higher aliphatic index is correlated with stability of proteins at wide range of temperatures [64]. Based on values of aliphatic index, the results of the present study suggested relatively higher stability of the proteins encoded by the dicots. The value for Grand Average Hydropathy (GRAVY, which represents the hydrobhobic and hydrophilic nature of aa) was higher in monocot proteins (range: -0.252 to -0.349) relative to dicot proteins (range: -0.313 to -0.366 except *P. persica*) (S7 Table). The CCD8 proteins of all monocots were found to be unstable except those associated with the B sub-genome of *T. aestivum* and *Ae. tauschii*. In dicots, the CCD8 proteins were stable except in *G. max*. Overall, the results indicated that the proteins encoded by the monocots had a relatively low level of stability, which might make it difficult to obtain these proteins in pure crystalline state for a study of their crystalline structure [65].

Secondary structure and motif search. A comparison of the secondary structures of CCD8 proteins in 15 species suggested that random coils dominated in the secondary structure followed by helices except for *Z. mays* and *T. urartu* among monocots and *G. max*, *T.*

cacao, *P. persica* and *M. truncatula* among dicots (S8 Table and S3 Fig). The random coils are known to form irregular structured regions permitting polypeptide chain to fold in a unique way. These results support that the proteins encoded by CCD8 genes tend to attain a globular structure which is known to be highly stable structure.

As many as 12 motifs were identified using MEME; these motifs ranged from 15 aa to 50 aa in most species except *T. urartu* (motifs 10,11 and 12 absent), *Ae. tauschii* (motif 12 absent), *A. thaliana* (motif 11 absent) and *P. persica* (motifs 9 and 12 absent) (for details, see S9 Table). Only six of these 12 motifs (motifs 1, 2, 4, 5, 9 and 10 related to carotenoid oxygenase family) were earlier characterized using Interpro Scan search (IPR004294); all other motifs were novel, which may be validated in future studies to identify their role in providing active sites for binding of regulatory proteins.

3D structures. The percent identity of 3D structures of the CCD8 proteins with the chosen template (PDB id: 5kk0.1.A) was low (23.90–32.80), which may be attributed to the non-availability of X-ray crystallographic structure of CCD8 protein in case of plants. The 3D structures had a high level of confidence, with Global Model Quality Estimation GMQE) score ranging within the acceptable range of 0.40 to 0.60; the higher value indicated higher reliability [66] (S10 Table).

The use of Ramachandran plot for evaluation of the accuracy of protein structures is widely known and has been emphasized in several recent studies [67]. In the present study, most aa residues were present in the favoured region of Ramachandran plots relative to those falling in the 'allowed' and 'disallowed' regions. Several features including the high values of quality factors through ERRAT and 3D-1D score by VERIFY3D and the values of *Dfire energy* (lower values indicate better quality structures) indicated that the modelled structures were of good quality. These can, therefore, be used for further analysis (S10 Table).

Simulation analysis of 3D models

Rg values ranged from 1.6 nm to 2.4 nm, except for the protein model belonging to *B. distachyon*, which had Rg value >4.1nm, suggesting that protein models were generally stable. When a two dimensional plot was drawn taking into consideration Rg values with respect to the simulation time, no changes were noticed in the Rg values, suggesting their stability over time also (S4 Fig). Overall, the data suggested stable folded state during 10 nsec simulation time. Based on root mean square deviation (RMSD) data also, protein models were found to be constant and stable (range: 0.3–1.3Å). Only in *T. urartu* and *Ae. tauschii*, the protein models showed higher RMSD values, although these values were still in the acceptable range (0.5–2.0Å) (S5 Fig).

Superimposition and alignment of 3D structures

The predicted values of different parameters were obtained through superimposition of the 3D protein structures (with minimized energy) for each species over the corresponding 3D protein structure of maize (Fig 5 and S11 Table). The aligned protein structures showed a fairly high level of similarity, which ranged from 50.98 to 85.19%. Physico-chemical properties also showed high level of similarity ranging from 44.12 to 78.43%. The values for RMSD were 1.72 to 3.26 Å, suggesting that the average distance of all pairs of residues in two structures were high, perhaps because a local error can arise in a big RMSD value, although the global topology is correct [68].

Sub-cellular localization, functions and ligand binding sites

CCD8 proteins of all species were largely located in the plastids, as also shown in previous studies in maize, sorghum, rice and Arabidopsis [1,22]. It is also reported that the CCDs are involved in the remodelling of carotenoids and therefore often targeted to plastids [69].

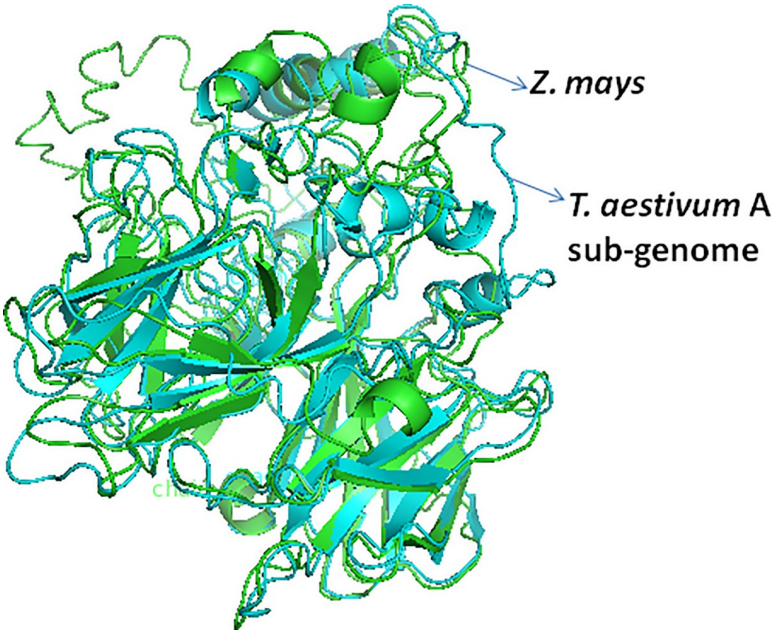

Fig 5. Representative figure showing superimposed structure of the predicted 3D structure of *T. aestivum* CCD8 protein belonging to A sub-genome over 3D structure of *Z. mays* CCD8 protein.

https://doi.org/10.1371/journal.pone.0213531.g005

Functional annotation (using gene ontology analysis) suggested that CCD8 proteins are involved in oxidation-reduction process. They have their characteristic role in oxidoreductase activity, dioxygenase activity and in metal-ion binding [1,69] (S12 Table).

CCD8 proteins are already known to have Fe^{++} binding sites. During the present study, 10–14 clusters of ligands were predicted to bind with the Fe^{++} binding catalytic centre in CCD8 proteins. The binding sites were predicted to have four conserved histidine residues. This is a characteristic feature of apocarotenoid oxygenase enzyme (ACO) of *Synechocystis* [69] and found in majority of species examined (Fig 6). The conserved His residues/ligand binding sites

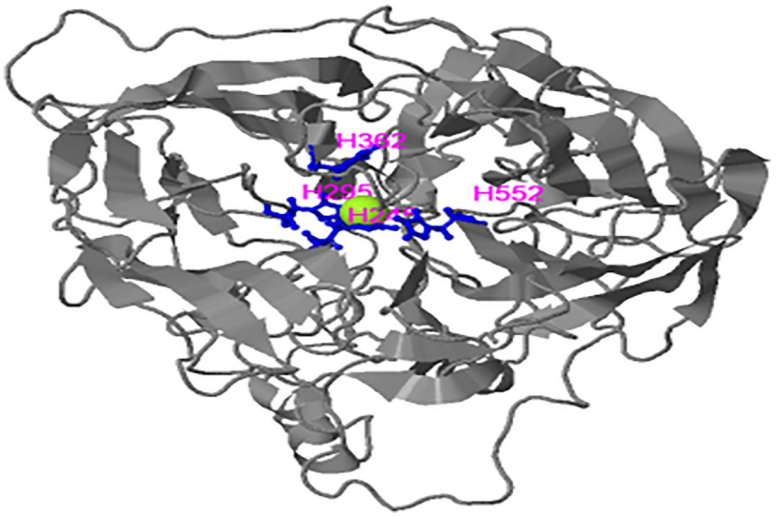

Fig 6. 3D structure of CCD8 protein of *T. aestivum* belonging to A sub-genome. The four conserved histidine residues and their positions are shown in middle. Solid sphere represents iron catalytic center.

https://doi.org/10.1371/journal.pone.0213531.g006

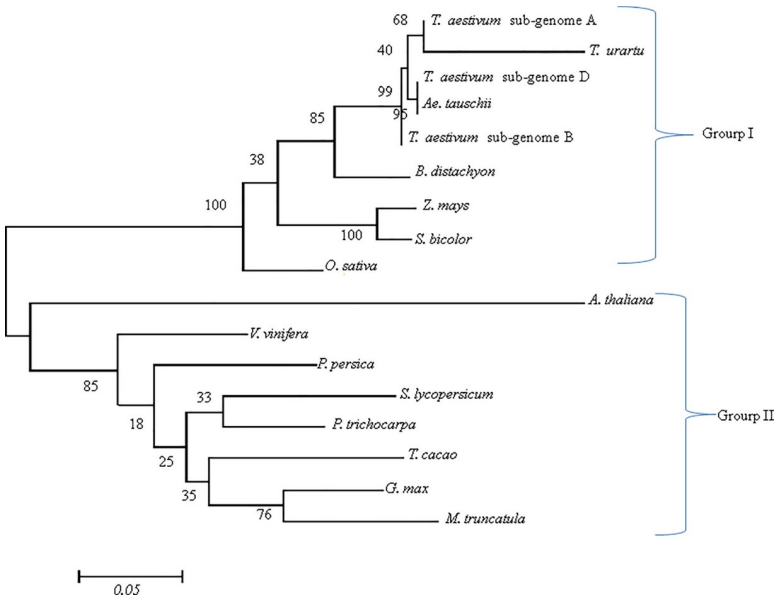

Fig 7. Phylogenetic tree obtained by maximum-likelihood method using amino acid sequences of CCD8 proteins of seven monocots and eight dicots depicting the relationship among monocots and dicots (for wheat there are 3 proteins for three homoeologues). The branch length represents magnitude of genetic change.

https://doi.org/10.1371/journal.pone.0213531.g007

were confined to the characteristic domain of CCD8 proteins. However, in *T. urartu*, 4th conserved histidine residue, and in *M. truncatula*, 2nd conserved histidine residue were absent, thus leaving only three histidine residues at the binding sites.

Phylogenetic analysis

The phylogenetic tree (based on aa sequences) had two major groups, Group I with all the monocots, and Group II with all the dicots. Among the monocots, as expected the sub-genome A of *T. aestivum* was close to *T. urartu* and *T. aestivum* sub-genome D was close to *Ae. tauschii*, whereas *Z. mays* was close to *S. bicolor* and among dicots, *S. lycopersicum* was close to *P. trichocarpa*, and *G. max* was close to *M. truncatula* (Fig 7). Phylogenetic analysis supports the conclusion that CCD8 genes in monocots and dicots diverged early in the evolutionary history. This was also inferred in earlier studies where monocots and dicots also made two separate clusters [23,24].

In silico expression analysis

The *in silico* expression analysis on the basis of microarray data was available for only seven species (*Z. mays*, *T. aestivum*, *O. sativa*, *B. distachyon*, *A. thaliana*, *G. max* and *M. truncatula*), which suggested that the level of expression was generally high in the vegetative tissues relative to reproductive tissues (S6 Fig). Abundance of CCD8 in terms of SLs in all the vegetative tissues and at all the developmental stages (S7 Fig) may be attributed to the transport of SLs to aerial parts of the plants after their synthesis in roots. These are in turn involved in important functions like leaf senescence, shoot branching/tillering etc., which suggests tissue and development stage specific expression of CCD8. Future studies may provide further clues to the role of SLs in the development during reproductive phase.

Under phosphate limited conditions, the expression was upregulated in *M. truncatula* only (fold change: >2 and P value-0.5), whereas, in *T. aestivum* and *O. sativa*, no differential

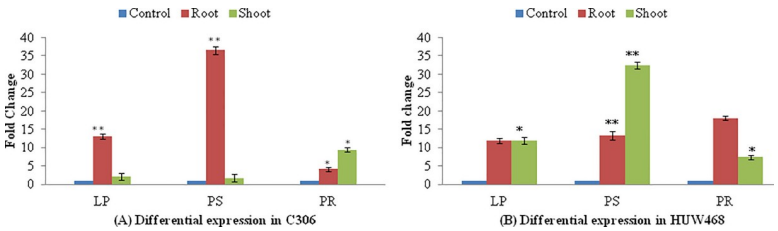

Fig 8. Combined relative expression level of three TaCCD8 genes (located on chromosomes 3A, 3B and 3D) in root and shoot tissues of wheat seedlings of cv. C306 (A) and cv. HUW 468 (B). Transcript levels were normalized with respect to expression of *TaAct* mRNA. Each data point represents mean ± SE (n = 3). Significance of combined expression level of three TaCCD8 genes (located on chromosomes 3A, 3B and 3D) in different treatments vs. the control was tested using *t*-test; *, ** represents significance at $p < 0.05$ and $p < 0.01$, respectively. C, control; LP, low P; PS, P starvation and PR, P restoration.

https://doi.org/10.1371/journal.pone.0213531.g008

expression was found in controlled vs. experimental conditions (S8 Fig). The data on P stress was not available for rest of the species. No expression data in reponse to biotic or abiotic stresses was available except in maize where synthesis of CCD8/SLs was found to be upregulated due to drought stress [6]. Similar results were also reported in *A. thaliana* where the role of SLs in positive regulation was examined under drought and salt stress using mutant and microarray analysis [70]. Therefore, a future study of the role of available regulatory sequences in response to biotic and abiotic stresses may prove rewarding, as suggested by some preliminary results obtained in the present study (see next).

qRT-PCR expression analysis of TaCCD8 genes

Two wheat cvs. (C306 and HUW468) were used for wet-lab experiments: of these two wheat cvs., C306 is known to be relatively tolerant to abiotic stresses like drought, and has a low nitrogen use efficiency (NUE). Significant differences in expression patterns of TaCCD8 genes were observed under three different P regimes (Fig 8). Following results were observed: (i) In root tissue of cv. C306, ~13 fold increase in expression of TaCCD8 genes was observed under low P, whereas ~37 fold increase in expression was observed under P starved condition. In root tissue of cv. HUW468, the expression of the gene was >10 fold higher at all the three P regimes but the maximum expression of the gene was observed after P restoration. (ii) In shoot tissue, maximum expression of TaCCD8 in cv. C306 was observed after P restoration (~9.5 fold), and that in cv. HUW468 increased ~12 fold under low P and ~33 fold under P starvation.

The expression pattern of TaCCD8 genes in root and shoot tissues seems to be in agreement with the results reported in previous studies [6]. The expression of CCD8 gene in Arabidopsis, pea, petunia, tomato, tobacco and potato was reported to be predominant in root tissue; in rice and chrysanthemum, the expression of this gene was reported to be high in stem tissue [21]. Since the combined expression of TaCCD8 genes was higher in roots of cv. C306 and shoots of cv. HUW468 under P starved conditios, it may be concluded that the expression of these genes is not only tissue specific, but also genotype-dependent [21,71]. Increased expression of TaCCD8 genes was observed under limited supply of phosphate. This is consistent with the role of CCD8 genes in nutrient uptake through SLs biosynthesis. Keeping this in view, future studies on expression using wheat mutants for CCD8 genes in root and shoot tissues under limited supply of phosphate may prove rewarding.

Supporting information

S1 Fig. Representative figure showing synteny and collinearity of a block of 31 genes of *Zea mays* (15 genes on either side of *ZmCCD8* gene) with respective genes of *S. bicolor*, *B.*

***distachyon* and *O. sativa*.** The numbers given above the genes correspond to the gene number given in S5 Table.
(TIF)

S2 Fig. Multiple sequence alignment of CCD8 proteins of all the selected 15 species and their comparison with the CCD8 consensus protein sequence. The figure shows insertions, deletions and mismatches in the CCD8 protein of the individual species with respect to the consensus sequence.
(JPG)

S3 Fig. Representative figure showing secondary structure of protein sequence encoded by CCD8 gene of *Z. mays*.
(TIF)

S4 Fig. Radius of gyration plot for each of the 17 simulated CCD8 proteins belonging to seven monocots and eight dicots (for wheat there are 3 proteins for three homoeologues).
(TIF)

S5 Fig. RMSD plot of CCD8 protein models of seven monocots and eight dicots representing protein backbone atoms using 1 to 10 nsec trajectory data (for wheat there are 3 proteins for three homoeologues).
(TIF)

S6 Fig. Representative figure showing *in silico* expression of TaCCD8 in different tissues of wheat.
(TIF)

S7 Fig. Representative figure showing *in silico* expression of TaCCD8 at different development stages of wheat plant. (1) germination, (2) seedling growth, (3) tillering, (4) stem elongation, (5) booting, (6) inflorescence emergence, (7) anthesis, (8) milk stage, (9) dough development stage, and (10) ripening.
(TIF)

S8 Fig. *In silico* expression of CCD8 genes belonging to *Oryza sativa*, *Triticum aestivum* and *Medicago truncatula* at different dose of phosphorous.
(TIF)

S1 Table. Details of the positions of exons (upper row) and introns (lower row) of CCD8 structural gene sequences (in bp) in 15 different species. The position of first exon is marked from translation start site.
(DOCX)

S2 Table. Per cent similarity of exons in CCD8 genes of 14 different species with respect to *Z. mays*.
(DOCX)

S3 Table. Values of non-synonymous substitutions (Ka; upper row) and synonymous substitutions (Ks; lower row) in CCD8 genes of seven monocots (including homoeologues on group 3 chromosomes of wheat).
(DOCX)

S4 Table. Values of non-synonymous substitutions (Ka; upper row) and synonymous substitutions (Ks; lower row) in CCD8 genes of eight dicots.
(DOCX)

S5 Table. Details of 31 genes (including the CCD8 gene of *Z. mays* at position 16) used for synteny and collinearity analysis.
(DOCX)

S6 Table. Simple sequence repeats (SSRs) and retro-elements identified in CCD8 genes belonging to 12 species.
(DOCX)

S7 Table. Primary protein sequence analyses of CCD8 proteins of 15 selected species.
(DOCX)

S8 Table. Analysis of secondary structure of CCD8 proteins of selected 15 species.
(DOCX)

S9 Table. Predicted values of different parameters for the motifs identified in CCD8 proteins using MEME suite.
(DOCX)

S10 Table. Details of 3D structures of CCD8 proteins (using Swiss-model) and their quality assessment parameters in selected 15 species obtained using SAVES and structure assessment tool of Swiss-Model.
(DOCX)

S11 Table. Predicted values of different parameters obtained after superimposition of 3D protein structures of CCD8 of selected 14 species over 3D protein structure of CCD8 of maize.
(DOCX)

S12 Table. Predicted scores of different parameters related to cellular component, biological processes and biochemical functions of CCD8 proteins of all the selected 15 species obtained through their functional analysis.
(DOCX)

Acknowledgments

This work was initiated during a workshop held at CCS University, Meerut and sponsored by the Department of Science and Technology, Government of India, under its CP-STIO (Collaborative Project with Scientists and Technologists of Indian Origin Abroad) Program. The study was carried out in the Bioinformatics Infrastructure Facility laboratory, Department of Genetics & Plant Breeding, CCS University, Meerut.

Author Contributions

Conceptualization: Ritu Batra, Harindra Singh Balyan, Renu Pandey, Pushpendra Kumar Gupta.

Data curation: Ritu Batra, Priyanka Agarwal, Sandhya Tyagi, Dinesh Kumar Saini, Vikas Kumar, Anuj Kumar, Sanjay Kumar.

Formal analysis: Ritu Batra, Priyanka Agarwal, Sandhya Tyagi.

Methodology: Ritu Batra, Renu Pandey.

Resources: Harindra Singh Balyan, Renu Pandey, Pushpendra Kumar Gupta.

Supervision: Harindra Singh Balyan, Renu Pandey, Pushpendra Kumar Gupta.

Writing – original draft: Ritu Batra, Sanjay Kumar.

Writing – review & editing: Harindra Singh Balyan, Renu Pandey, Pushpendra Kumar Gupta.

References

1. Auldridge ME, Block A, Vogel JT, Dabney-Smith C, Mila I, Bouzayen M, et al. Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. Plant J. 2006; 14: 982–993. https://doi.org/10.1111/j.1365-313X.2006.02666.x
2. Schwartz SH, Qin X, Loewen MC. The biochemical characterization of two carotenoid cleavage enzymes from Arabidopsis indicates that a carotenoid-derived compound inhibits lateral branching. J Biol Chem. 2004; 279: 46940–46945. https://doi.org/10.1074/jbc.M409004200 PMID: 15342640
3. Czarnecki O, Yang J, Weston DJ, Tuskan GA, Chen J. A dual role of strigolactones in phosphate acquisition and utilization in plants. Int J Mol Sci. 2013; 14(4): 7681–7701. https://doi.org/10.3390/ijms14047681 PMID: 23612324
4. Seto Y, Yamaguchi S. Strigolactone biosynthesis and perception. Curr Opin Plant Biol. 2014; 21: 1–6.
5. Saeed W, Naseem S, Ali Z. Strigolactones biosynthesis and their role in abiotic stress resilience in plants: A critical review. Front Plant Sci. 2017; 8: 1487. https://doi.org/10.3389/fpls.2017.01487 PMID: 28894457
6. Guan JC, Koch KE, Suzuki M, Wu S, Latshaw S, Petruff T, et al. Diverse roles of strigolactone signaling in maize architecture and the uncoupling of a branching-specific subnetwork. Plant Physiol. 2012; 160: 1303–1317. https://doi.org/10.1104/pp.112.204503 PMID: 22961131
7. Mashiguchi K, Sasaki E, Shimada Y, Nagae M, Ueno K, Nakano M, et al. Feedback-regulation of strigolactone biosynthetic genes and strigolactone-regulated genes in Arabidopsis. Biosci Biotechnol Biochem. 2009; 73: 2460–2465. https://doi.org/10.1271/bbb.90443 PMID: 19897913
8. Foo E, Morris SE, Parmenter K, Young N, Wang H, Jones A, et al. Feedback regulation of xylem cytokinin content is conserved in pea and Arabidopsis. Plant Physiol. 2007; 143: 1418–1428. https://doi.org/10.1104/pp.106.093708 PMID: 17277096
9. Proust H, Hoffmann B, Xie X, Yoneyama K, Schaefer DG, Yoneyama K, et al. Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*. Development. 2011; 138: 1531–1539. https://doi.org/10.1242/dev.058495 PMID: 21367820
10. Ledger SE, Janssen BJ, Karunairetnam S, Wang T, Snowden KC. Modified CAROTENOID CLEAVAGE DIOXYGENASE8 expression correlates with altered branching in kiwifruit (Actinidia chinensis). New Phytol. 2010; 188: 803–813.
11. Kohlen W, Charnikhova T, Liu Q, Bours R, Domagalska MA, Beguerie S, et al. Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in non-arbuscular mycorrhizal host Arabidopsis. Plant Physiol. 2010; 155: 974–987. https://doi.org/10.1104/pp.110.164640 PMID: 21119045
12. Al-Babili S, Bouwmeester HJ. Strigolactones, a novel carotenoid-derived plant hormone. Annu Rev Plant Biol. 2015; 66: 161–86. https://doi.org/10.1146/annurev-arplant-043014-114759 PMID: 25621512
13. Yokota T, Sakai H, Okuno K, Yoneyama K, Takeuchi Y. Alectrol and orobanchol, germination stimulants for *Orobanche* minor, from its host red clover. Phytochemistry. 1998; 49: 1967–1973.
14. Besserer A, Puech-Pagès V, Kiefer P, Gomez-Roldan V, Jauneau A, Roy S, et al. Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. PLoS Biol. 2006; 4: 1239–1247.
15. López-Ráez JA, Charnikhova T, Gómez-Roldán V, Matusova R, Kohlen W, De Vos R, et al. Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. New Phytol. 2008; 178: 863–874. https://doi.org/10.1111/j.1469-8137.2008.02406.x PMID: 18346111
16. Umehara M, Hanada A, Yoshida S, Akiyama K, Arite T, Takeda-Kamiya N, et al. Inhibition of shoot branching by new terpenoid plant hormones. Nature. 2008; 455: 195–200.
17. Snowden KC, Simkin AJ, Janssen BJ, Templeton KR, Loucas HM, Simons JL, et al. The decreased apical dominance1/Petunia hybrida *CAROTENOID CLEAVAGE DIOXYGENASE8* gene affects branch production and plays a role in leaf senescence, root growth, and flower development. Plant Cell. 2005; 17: 746–59. https://doi.org/10.1105/tpc.104.027714 PMID: 15705953
18. Sorefan K, Booker J, Haurogné K, Haurogne K, Goussot M, Bainbridge K, et al. MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea. Genes Dev. 2003; 17: 1469–1474. https://doi.org/10.1101/gad.256603 PMID: 12815068

19. Johnson X, Brcich T, Dun EA, Goussot M, Haurogné K, Beveridge CA, et al. Branching genes are conserved across species. Genes controlling a novel signal in pea are coregulated by other long-distance signals. Plant Physiol. 2006; 142: 1014–1026. https://doi.org/10.1104/pp.106.087676 PMID: 16980559

20. Arite T, Iwata H, Ohshima K, Maekawa M, Nakajima M, Kojima M. DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. Plant J. 2007; 51(6): 1019–1029. https://doi.org/10.1111/j.1365-313X.2007.03210.x PMID: 17655651

21. Gao J, Zhang T, Xu B, Jia L, Xiao B, Liu H, et al. CRISPR/Cas9 mediated mutagenesis of Carotenoid Cleavage Dioxygenase 8 (CCD8) in tobacco affects shoot and root architecture. Int J Mol Sci. 2018; 19 (4): 1062.

22. Mashiguchi K, Sasaki E, Shimada Y, Nagae M, Ueno K, Nakano M, et al. Feedback-regulation of strigolactone biosynthetic genes and strigolactone-regulated genes in Arabidopsis. Biosci Biotechnol Biochem. 2009; 73: 2460–2465. https://doi.org/10.1271/bbb.90443 PMID: 19897913

23. Vallabhaneni R, Bradbury LM, Wurtzel ET. The carotenoid dioxygenase gene family in maize, sorghum, and rice. Arch Biochem Biophys. 2010; 504: 104–111.

24. Wei Y, Wan H, Wu Z, Wang R, Ruan M, Ye Q, et al. A comprehensive analysis of carotenoid cleavage dioxygensae genes in *Solanum lycopersicum*. Plant Mol Biol Rep. 2016; 34: 512–523.

25. Morris SE, Turnbull CG, Murfet IC, Beveridge CA. Mutational analysis of branching in pea. evidence that Rms1 and Rms5 regulate the same novel signal. Plant Physiol. 2001; 126: 1205–1213. PMID: 11457970

26. Napoli C. Highly branched phenotype of the petunia dad1-1 mutant is reversed by grafting Plant Physiol. 1996; 111: 27–37. PMID: 12226274

27. Pasare SA, Ducreux LJM, Morris WL, Campbell R, Sharma SK, Roumeliotis E, et al. The role of the potato (Solanum tuberosum) CCD8 gene in stolon and tuber development. New Phytol. 2013; 198: 1108–1120. https://doi.org/10.1111/nph.12217 PMID: 23496288

28. Dhaliwal AK, Mohan A, Gill KS. Comparative analysis of ABCB1 reveals novel structural and functional conservation between monocots and dicots. Front Plant Sci. 2014; 5: 657. https://doi.org/10.3389/fpls.2014.00657 PMID: 25505477

29. Batra R, Saripalli G, Mohan A, Gupta S, Gill KS, Varadwaj PK, et al. Comparative analysis of AGPase genes and encoded proteins in eight monocots and three dicots with emphasis on wheat. Front Plant Sci. 2017; 8: 1–16. https://doi.org/10.3389/fpls.2017.00001

30. Kumar A, Batra R, Gahlaut V, Gautam T, Kumar S, Sharma M, et al. Genome-wide identification and characterization of gene family for RWP-RK transcription factors in wheat (Triticum aestivum L.). PLoS One. 2018; 13(12): e0208409. https://doi.org/10.1371/journal.pone.0208409 PMID: 30540790

31. Jukes TH, Cantor CR. Evolution of protein molecules. Mamm Protein Metab. 1969; 21–132.

32. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res. 2002; 30: 325–327. PMID: 11752327

33. Muffato M, Louis A, Poisnel CE, Crollius HR. Genomicus: A database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics. 2010; 26: 1119–1121. https://doi.org/10.1093/bioinformatics/btq079 PMID: 20185404

34. Navarro C, Moore J, Ott A, Baumert E, Mohan A, Gill KS, et al. Evolutionary, comparative and functional analyses of the Brassinosteroid receptor gene, bri1, in wheat and its relation to other plant genomes. PLoS One. 2015; 10: 1–16.

35. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr. 1993; 26: 283–291.

36. Eisenberg D, Luthy R. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzy Mol. 1997; 277: 396–404.

37. Colovos C, Yeates T. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993; 2(9): 1511–1519. https://doi.org/10.1002/pro.5560020916 PMID: 8401235

38. Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved peptide and protein torsional energetics with the OPLS-AA force field. J Chem Theory Comput. 2015; 11: 3499–3509. https://doi.org/10.1021/acs.jctc.5b00356 PMID: 26190950

39. Gajula MNVP, Kumar A, Ijaq J. Protocol for molecular dynamics simulations of proteins. Bio-protocol. 2016; 6: e2051.

40. Ryckaert JP, Ciccotti G, Berendsen H. Numerical integration of the Cartesian Equations of motion of a system with constraints: Molecular dynamics of n-Alkanes. J Comput Phys. 1977; 23: 327–341.

41. Kolafa J, Perram JW. Cutoff errors in the {E}wald summation formulae for point charge systems. Mol Simul. 1992; 9: 351–368.

42. Parrinello M, Rahman A. Crystal structure and pair potentials: A {Molecular-Dynamics} study. Phys Rev Lett. 1980; 45: 1196–1199.
43. Andersen HC. Rattle: A "Velocity" version of the shake algorithmfor molecular dynamics calculations. J Comput Phys. 1983; 52: 24–34.
44. Ye Y, Godzik A. FATCAT: A web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 2004; 32: 582–585.
45. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28: 2731–2739. https://doi.org/10.1093/molbev/msr121 PMID: 21546353
46. Pandey R, Lal MK, Vengavasi K. Differential response of hexaploid and tetraploid wheat to interactive effects of elevated [CO2] and low phosphorus. Plant Cell Rep. 2018; 37(9): 1231–1244. https://doi.org/10.1007/s00299-018-2307-4 PMID: 29868985
47. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc. 2008; 3: 1101–1108. PMID: 18546601
48. Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci U S A. 1989; 86: 6201–6205. PMID: 2762323
49. Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, et al. Acc homoeoloci and the evolution of wheat genomes. Proc Natl Acad Sci U S A. 2016; 105: 9691–9696. https://doi.org/10.1073/pnas.1611405113
50. Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet. 2006; 7: 211–21. https://doi.org/10.1038/nrg1807 PMID: 16485020
51. Roy SW, Penny D. Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. Mol Biol Evol. 2007; 24(1): 171–81. https://doi.org/10.1093/molbev/msl159 PMID: 17065597
52. Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV. Analysis of evolution of exon-intron structure of eukaryotic genes. Brief Bioinform. 2005; 6: 774–778.
53. Long M, Deutsch M. Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. Mol Biol Evol. 1999; 16: 1528–1534. https://doi.org/10.1093/oxfordjournals.molbev.a026065 PMID: 10555284
54. Trapp SC, Croteau RB. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics. 2001; 158: 811–832.
55. Huo N, Vogel JP, Lazo GR, You FM, Ma Y, McMahon S, et al. Structural characterization of Brachypodium genome and its syntenic relationship with rice and wheat. Plant Mol Biol. 2009; 70: 47–61.
56. Ahn S, Tanksley SD. Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci U S A. 1993; 90: 7980–7984. PMID: 8103599
57. Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, et al. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. The Plant Cell. 2008; 20: 11–24. https://doi.org/10.1105/tpc.107.056309 PMID: 18178768
58. Bennetzen JL, Chen M. Grass genomic synteny illuminates plant genome function and evolution. Rice. 2008; 1: 109–118.
59. Fankhauser C, Chory J. Light control of plant development. Annu Rev Cell Dev Biol. 1997; 13: 203–229. https://doi.org/10.1146/annurev.cellbio.13.1.203 PMID: 9442873
60. Morishima A. Identification of preferred binding sites of a light-inducible DNA-binding factor (MNF1) within 5'-upstream sequence of C4-type phosphoenolpyruvate carboxylase gene in maize. Plant Mol Biol. 1998; 38: 633–646. PMID: 9747808
61. Hayward A, Stirnberg P, Beveridge C, Leyser O. Interactions between auxin and strigolactone in shoot branching control. Plant Physiol. 2009; 151: 400–412. https://doi.org/10.1104/pp.109.137646 PMID: 19641034
62. Andreo-Jimenez B, Ruyter-Spira C, Bouwmeester HJ, Lopez-Raez JA. Ecological relevance of strigolactones in nutrient uptake and other abiotic stresses, and in plant-microbe interactions below-ground. Plant and Soil. 2015; 394(1–2): 1–19.
63. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engg. 1990; 4: 155–161.
64. Gupta SK, Rai AK, Kanwar SS, Sharma TR. Comparative analysis of zinc finger proteins involved in plant disease resistance. PLoS One. 2012; 7: e42578. https://doi.org/10.1371/journal.pone.0042578 PMID: 22916136

65. Dawar C, Jain S, Kumar S. Insight into the 3D structure of ADP-glucose pyrophosphorylase from rice (Oryza sativa L.). J Mol Model. 2013; 19: 3351–3367. https://doi.org/10.1007/s00894-013-1851-7 PMID: 23674369

66. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014; 42: W252–W258. https://doi.org/10.1093/nar/gku340 PMID: 24782522

67. Kumar A, Kumar S, Kumar U, Suravajhala P, Gajula MNVP. Functional and structural insights into novel DREB1A transcription factors in common wheat (Triticum aestivum L.): A molecular modeling approach. Comp Biol Chem. 2016; 64: 217–216.

68. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins Struct Funct Genet. 2004; 57: 702–710. https://doi.org/10.1002/prot.20264 PMID: 15476259

69. Kloer DP, Ruch S, Al-Babili S, Beyer P, Schulz GE. The structure of a retinal-forming carotenoid oxygenase. Science. 2005; 308: 267–9. https://doi.org/10.1126/science.1108965 PMID: 15821095

70. Ha C Van, Leyva-González MA, Osakabe Y, Tran UT, Nishiyama R, Watanabe Y, et al. Positive regulatory role of strigolactone in plant responses to drought and salt stress. Proc Natl Acad Sci U S A. 2014; 111: 851–6. https://doi.org/10.1073/pnas.1322135111 PMID: 24379380

71. López-Ráez JA, Charnikhova T, Gómez-Roldán V, Matusova R, Kohlen W, De Vos R, et al. Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. New Phytol. 2008; 178: 863–874. https://doi.org/10.1111/j.1469-8137.2008.02406.x PMID: 18346111