Mol Genet Genomics (2010) 283:13–33
DOI 10.1007/s00438-009-0495-x

ORIGINAL PAPER

Carotenoid biosynthesis genes in rice: structural analysis, genome-wide expression profiling and phylogenetic analysis

Neetu Chaudhary · Aashima Nijhawan · Jitendra P. Khurana · Paramjit Khurana

Received: 4 July 2009 / Accepted: 12 October 2009 / Published online: 5 November 2009

Abstract Carotenoids, important lipid-soluble antioxidants in photosynthetic tissues, are known to be completely absent in rice endosperm. Many studies, involving transgenic manipulations of carotenoid biosynthesis genes, have been performed to get carotenoid-enriched rice grain. Study of genes involved in their biosynthesis can provide further information regarding the abundance/absence of carotenoids in different tissues. We have identified 16 and 34 carotenoid biosynthesis genes in rice and *Populus* genomes, respectively. A detailed analysis of the domain structure of carotenoid biosynthesis enzymes in rice, *Populus* and *Arabidopsis* has shown that highly conserved catalytic domains, along with other domains, are present in these proteins. Phylogenetic analysis of rice genes with *Arabidopsis* and other characterized carotenoid biosynthesis genes has revealed that homologous genes exist in these plants, and the duplicated gene copies probably adopt new functions. Expression of rice and *Populus* genes has been analyzed by full-length cDNA- and EST-based expression profiling. In rice, this analysis was complemented by real-time PCR, microarray and signature-based expression profiling, which reveal that carotenoid biosynthesis genes are highly expressed in light-grown tissues, have differential expression pattern during vegetative/reproductive development and are responsive to stress.

Keywords Carotenoids · Microarray · MPSS · *Populus* · Phylogenetic analysis · Rice

Communicated by Y. Van de Peer.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-009-0495-x) contains supplementary material, which is available to authorized users.

N. Chaudhary · A. Nijhawan · J. P. Khurana · P. Khurana (✉)
Interdisciplinary Centre for Plant Genomics,
Department of Plant Molecular Biology,
University of Delhi South Campus,
Benito Juarez Road, New Delhi 110021, India
e-mail: param@genomeindia.org

Introduction

Plastids are plant organelles producing diverse array of compounds that perform important functions in the plastid and cell, and are also important for agriculture and human nutrition. Plastidic isoprenoid synthesis represents a major source of such compounds and produces an important group of lipid-soluble antioxidants in photosynthetic tissues, i.e., the carotenoids. Most of the carotenoids are C_{40} tetraterpenoids derived from phytoene and comprise a large family of more than 700 structures (Britton et al. 2004). They accumulate in nearly all types of plastids, including the chloroplast and chromoplast, and are thus found in most plant organs and tissues. In seeds, roots and dark-grown tissues, they are present in amyloplasts (in starchy seeds) or elaioplasts (oil seeds), leucoplasts and etioplasts, respectively. Non-oxygenated carotenoids are referred to as carotenes, whereas their oxygenated derivatives are designated as xanthophylls. The most commonly occurring carotenes are β-carotene in chloroplasts and lycopene in chromoplasts of some fruits and flowers. The most abundant xanthophylls (lutein, violaxanthin and neoxanthin) are key components of the light harvesting complexes (Woitsch and Römer 2003; Dall'Osto et al. 2006) and are present in great abundance in photosynthetic plant tissues. In plants, carotenoids play an important role in photosystem assembly, light harvesting and photoprotection (Kulheim et al. 2002; Lokstein et al. 2002; Holt et al. 2005; Dall'Osto et al.

2007), photomorphogenesis (Park et al. 2002; Howitt and Pogson 2006), non-photochemical quenching (Pogson et al. 1998; Franco et al. 2007), lipid peroxidation (Havaux and Niyogi 1999; Davidson et al. 2002; McNulty et al. 2007), and viability and aging of the seeds (Calucci et al. 2004). In recent years, there have been many evidences for the presence of carotenoid-derived signaling compounds regulating diverse aspects of plant development, for example apical dominance and branching (Beveridge et al. 2003; McSteen and Leyser 2005), root and shoot development (Van Norman et al. 2007) and rhizosphere signaling (López-Ráez et al. 2008). Carotenoid cleavage products (enzymatic and photooxidative derivatives) are precursors for the formation of plant hormone abscisic acid (Nambara and Marion-Poll 2005) and also provide important components of food, fragrance and cosmetics (Bouvier et al. 2003a; Bouvier et al. 2003b; Simkin et al. 2004a, b).

During the past decade, primarily as a result of molecular genetic and biochemical genomics-based approaches in the model organisms *Arabidopsis* and *Synechocystis* sp. PCC6803, a near complete set of genes required for the synthesis of these compounds were identified (Dellapenna and Pogson 2006). A number of breeding and transgenic approaches have thus been used to modify the types and levels of these important compounds in agricultural crops and have provided important insight into their regulation, activity and integration (Lu and Li 2008). The availability of the complete genome sequence for four angiosperm species, *Oryza sativa* (IRGSP 2005), *Sorghum bicolor* (Paterson et al. 2009), *Populus trichocarpa* (Tuskan et al. 2006) and *Arabidopsis thaliana* (AGI 2000) provides an opportunity to conduct genomic analyses in these model species, as it has been demonstrated that all these four species have undergone at least one round of genome-wide duplication (Bowers et al. 2003; Raes et al. 2003; Tuskan et al. 2006; Paterson et al. 2009). Comparative analysis of carotenoid biosynthesis genes will provide valuable insight into the complex origin, evolution and conservation of an important metabolic pathway, as many biosynthetic genes are already known to have homologs, resulting from gene duplication and sequence divergence. The results of this analysis would provide a framework for detecting conserved sequences having common functions, and expansion of genes resulting in non-, neo- or sub-functionalization. Rice endosperm is known to be completely devoid of carotenoids and a number of transgenic approaches have been followed to improve its carotenoid content (Ye et al. 2000; Paine et al. 2005). The expression profile of carotenoid biosynthesis genes in different photosynthetic and non-photosynthetic tissues could provide important insights into their observed presence and absence, respectively. In this study, we report the identification of carotenoid biosynthesis genes in rice and *Populus*, characterization of these genes in rice and their comparison with *Arabidopsis* and other known carotenoid biosynthesis genes. We have analyzed the gene structure and organization and duplication events that might have contributed to their expansion, while phylogenetic analysis was performed to study their evolutionary relationships. We also analyzed the expression of rice genes in different tissues/developmental stages and under different abiotic stress conditions, using quantitative PCR, microarray and signature-based expression profiling.

Materials and methods

Identification of carotenoid biosynthesis genes in rice and *Populus*

To identify the complete set of carotenoid biosynthesis genes in rice (*O. sativa*), a BLAST search of all the annotated proteins in the rice genome at TIGR (release 6) was performed using HMMER program (version 2.3.2), which is based on hidden Markov models (HMM) (Durbin et al. 1998). This program utilized the HMM profiles generated by alignments of known carotenoid biosynthetic proteins, from *A. thaliana* and other plants species, as query. In all, ten HMM profiles were generated to identify proteins similar to phytoene synthase (PSY), phytoene desaturase (PDS), ζ-carotene desaturase (ZDS), carotenoid isomerase (CRTISO), lycopene *β*-cyclase (*β*LCY), lycopene ε-cyclase (εLCY), *β*-carotene hydroxylase (*β*OH), ε-carotene hydroxylase (εOH), zeaxanthin epoxidase (ZE) and violaxanthin de-epoxidase (VDE) enzymes in rice. In the results of the HMMER search, all proteins with score >500 were taken; proteins corresponding to different gene models present at the same locus were removed and the remaining ones were studied by alignment and conserved domain analysis of protein sequences.

The protein sequences of rice carotenoid biosynthesis genes were used as a query sequence to search against the *Populus* genome database (http://genome.jgi-psf.org/poptr1/Poptr1.home.html) using the BLASTP algorithm (Altschul et al. 1997). The newly identified genes were used reiteratively to search the same sequence database. All the significant hits in BLAST result were taken and analyzed by alignment with known genes and presence of conserved domains. TBLASTN searches were performed to identify the genes in the genome that may not have been annotated.

Sequence analysis

Multiple sequence alignments to identify the conserved regions were done using ClustalX (version 1.83; Thompson et al. 1997) and, for phylogenetic analysis, ClustalX

(version 1.83) and MEGA (version 4.0, Tamura et al. 2007) programs were used to construct neighbor-joining (NJ) and maximum parsimony (MP) trees, respectively, by generating 1,000 random bootstrap replicates. The phylogenetic trees were displayed using TreeView (version 1.6.6, http://www.taxonomy.zoology.gla.ac.uk/rod/treeview.html). DNA and protein sequence analysis were performed using Gene Runner program (version 3.04, http://www.genenames.com). Domain search program SMART (http://smart.embl/heidelberg.de/) was used to deduce the domain structure of putative protein sequences. Analysis of putative protein sequences was carried out using ChloroP 1.1 program (http://www.cbs.dtu.dk/services/ChloroP/) to identify the presence and length of putative chloroplast transit peptide. To obtain information on the intron/exon structures, cDNA sequences of carotenoid biosynthesis genes were aligned with their corresponding genomic sequences. Automatic annotation is sometimes known to be associated with potential errors (Rouze et al. 1999), hence the gene structures were also studied by refined analysis of intron–exon positions and by comparing gene models with the cDNA and/or EST sequences wherever the information was available.

Localization on rice chromosomes and segmental duplication

Each of the carotenoid biosynthesis genes was positioned on rice chromosome pseudomolecules available at TIGR (release 5) (http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml) and IRGSP (build 05) (http://rgp.dna.affrc.go.jp/E/IRGSP/Build5/build5.html) by BLASTN search. Segmental duplication of rice genes available at TIGR (http://www.tigr.org./tdb/e2k1/osa1/segmental_dup/index.shtml) was used to determine the presence of carotenoid biosynthesis genes on duplicated chromosomal segments, with the maximal length distance permitted between collinear gene pairs of 100 kb.

Plant material and growth conditions

Rice (*O. sativa* L. ssp. *indica* var. Pusa basmati 1) seeds were disinfected with 0.1% $HgCl_2$ solution for 1 h and thoroughly washed with sterilized RO (reverse-osmosis) water before soaking overnight in RO water, in the dark. Seedlings were grown on cotton saturated with RO water at 28 ± 1°C, either in 14-h light and 10-h dark photoperiodic cycle or in complete darkness. Developing and mature seeds were harvested from rice plants grown in the greenhouse.

RNA isolation

Total RNA from shoot and root tissues was extracted using the RNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's instructions, followed by treatment with RNase-free DNaseI (Qiagen) to remove any genomic DNA contamination. The integrity of RNA samples was monitored by electrophoresis on 1.2% agarose gel. For each RNA sample, absorption at 260 nm was measured and concentration calculated as $A_{260} \times 40$ (μg/ml) × dilution factor. To isolate RNA from mature and immature seeds, the RNA isolation method used for wheat and other carbohydrate-rich seeds was used (Singh et al. 2003), followed by treatment with DNase I to remove genomic DNA contamination.

Full-length cDNA and EST evidence search

For full-length cDNA-based expression profiling, the gene expression evidence search page (http://www.tigr.org/tdb/e2k1/osa1/locus_expression_evidence.shtml) available at TIGR rice genome annotation was used. TIGR locus IDs corresponding to the rice carotenoid biosynthesis genes were searched to find the availability of corresponding full-length cDNA sequences. For genes, where no corresponding cDNA sequence was available in the TIGR database, BLAST search was performed in NCBI database using the coding sequence as query. For EST-based expression analysis, BLAST search was performed using the cDNA sequences of rice carotenoid biosynthesis genes against the EST database (http://www.ncbi.nlm.nih.gov/dbEST/index.html) available at NCBI. MEGABLAST was used for homology search against EST databases of rice (*O. sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), *S. bicolor*, sugarcane (*Saccharum officinarum*) and barley (*Hordeum vulgare*). The searches were limited to one species at a time to detect all the existing ESTs that showed significant similarity. For each carotenoid biosynthesis gene, details of only top hits of the BLASTN search results showing score ≥500 were extracted into a separate file. An optimized value of word size of 11 and expected value of 10, without the low complexity filter, was used for analysis to record the optimum of matches between rice genes and ESTs from other monocots. For *Populus* carotenoid biosynthesis genes, BLAST search was performed using the partial/full-length coding sequence as query, in the EST database of *Populus* at NCBI (http://www.ncbi.nlm.nih.gov/dbEST/index.html) using MEGABLAST. For full-length coding sequences, results having score ≥500 were taken, whereas for partial coding sequences search results having score ≥200 were considered.

Quantitative real-time PCR analysis

Transcript levels for all rice carotenoid biosynthesis genes in different RNA samples were quantified by real-time PCR analysis using the ABI Prism 7000 Sequence Detection

System and Software (PE applied Biosystems, USA) as per manufacturer's instructions. Primers were designed from 3′UTRs or other unique regions of the genes using Primer Express 2.0 (Applied Biosystems, USA; Supplemental Table S1 in Electronic Supplementary Material). To ensure that the primers amplify a unique and specific fragment of the cDNA, each primer pair was checked by homology searches against the rice genomic sequence at the TIGR Web site. First strand cDNA was synthesized by reverse transcription of 3 μg of total RNA using High Capacity cDNA Archive kit (Applied Biosystems, USA) as per manufacturer's instructions. These cDNA samples were mixed with 200 nM of each primer and SYBR Green PCR Master Mix (Applied Biosystems, USA) and real-time PCR analysis was carried out using ABI Prism 7000 sequence detection system and software (PE Applied Biosystems, USA). PCR reactions were performed using the following parameters: 2 min at 50°C, 10 min at 95°C, 40 cycles of 15 s at 95°C and 1 min at 60°C in the 96-well optical reaction plates (Applied Biosystems, USA). The identity of the amplicons and the specificity of the reaction were verified by agarose gel electrophoresis and melting curve analysis, respectively. The relative mRNA levels for each of the carotenoid biosynthetic gene in different RNA samples were normalized with respect to the internal standard *actin*, to normalize for variance in the quality of RNA and the amount of input cDNA. Real-time PCR analysis was based on three biological replicates of each sample and two technical replicates of each biological replicate. The values in the graphs are the mean of three biological replicates and the error bars show the standard deviation from mean values.

Rice microarray analysis

For microarray-based expression analysis, the data set submitted earlier by our group and available in the Gene Expression Omnibus database at the National Center for Biotechnology Information, under the series accession numbers GSE6893 and GSE6901 was used to study the expression profile of rice carotenoid biosynthesis genes in different stages of vegetative and reproductive development, and also under abiotic stress conditions. The data were imported in ArrayAssist™ (Stratagene, La Jolla, CA) microarray analysis software, normalized using GC-RMA algorithm and Log2 transformed. The average of three biological replicates was used to get each expression value. Unique Affymetrix probe-set IDs, as mentioned for each gene in TIGR, were used to extract the expression data for carotenoid biosynthesis genes. Euclidean distance metric and complete linkage rule of hierarchical clustering was used to perform cluster analysis on rows of expression values.

Rice MPSS analysis

For rice massively parallel signature sequencing (MPSS) data analysis, expression evidence for mRNA from MPSS tags was determined from rice MPSS project mapped to TIGR gene models (http://mpss.udel.edu/rice/) (Nakano et al. 2006). MPSS expression data from 22 mRNA libraries, representing 18 different tissues/organs were used in the analysis, and data for 17-base significant signatures (classes 1, 2, 5 and 7 identifying the sense strand), which uniquely identify an individual gene and show a perfect match (100% identity over 100% length of the tag), were retrieved. The normalized abundance (transcript per million, tpm) of these signatures for a given gene in any library represents a quantitative estimate of expression of that gene.

Results

Carotenoid biosynthesis genes in rice and *Populus*

BLAST searches of the rice genome using HMM profiles of carotenoid biosynthesis enzymes as query, and analysis of the putative genes with SMART, revealed the presence of 16 potential non-redundant carotenoid biosynthesis genes in rice. Each carotenoid biosynthesis gene was given a name based on the enzymatic reaction, as given in the *Arabidopsis* carotenoid biosynthesis pathway (Dellapenna and Pogson 2006), also taking into consideration their position on rice chromosomes 1–12 from top to bottom. The carotenoid biosynthesis gene name, the Institute of Genomic Research (TIGR) locus ID, gene length, open reading frame (ORF) length, protein length, chromosomal location, neighboring marker and other related information for all 16 carotenoid biosynthesis genes of rice are listed in Table 1. Similarly, genes retrieved from rice were used as a query for BLASTP search in *Populus* database, resulting in the identification of 34 genes corresponding to carotenoid biosynthesis pathway enzymes. The coding sequences of 24 genes of *Populus* carotenoid biosynthesis pathway were partial (Supplemental Table S2).

Structure of carotenoid biosynthesis genes

To study the gene structure of rice carotenoid biosynthesis genes, their exon–intron organization was analyzed, revealing that only 1 of the total 16 carotenoid biosynthesis genes was intronless; among those having introns, the number of introns varied from 4 to 15 (Supplemental Fig. S1). For carotenoid biosynthesis genes of rice, *Populus* and *Arabidopsis*, the number of exons, protein length, genomic strand and chromosomal localization were studied (Supplemental

Table 1 Carotenoid biosynthesis genes in rice

S. no.	Gene name	Locus ID[a]	Accession no.[b]	Gene length[c] (kbp)	cDNA length[d] (bp)	No. of introns[e]	ORF length[f] (bp)	Protein length[g] (aa)	Mol. wt[h] (Da)	p*I*[i]	Length of cTP[j] (in aa)	Genomic locus[k]		Neighboring marker[l]	cM position[m]
												BAC/PAC name	Chromosome no.		
1	*OsPSY1*	LOC_Os06g51290	AK070716 AY445521* AK059535	3,956	1,900	5	1,263	420	47,582.66	9.01	–	OSJNBa0069C14	6	R220S	124.4
2	*OsPSY2*	LOC_Os12g43130	AK073290	3,112	1,486	5	1,197	398	44,728.98	8.62	80	OSJNBb0016P08	12	R496	107.4
3	*OsPSY3*	LOC_Os09g38320	AK108154	3,427	1,485	5	1,335	444	49,367.24	7.38	54	OSJNBa0043I04	9	C632	90.7
4	*OsPDS*	LOC_Os03g08570	AF049356* AK058668 (partial)	5,729	2,312	14	1,737	578	64,769.51	7.96	87	OSJNBa0032G08	3	S2769	20.3
5	*OsZDS*	LOC_Os07g10490	AK065213 AK066223 AK062079	5,635	2,360	13	1,737	578	63,728.81	8.56	42	P0431A02	7	R3113	41.7
6	*OsCRTISO*	LOC_Os11g36440	EF417892* (partial) NM_001074659*	4,193	2,172	12	1,761	586	63,543.10	6.88	58	OSJNBb0077M24	11	E61360S	84.6
7	*OsβLCY*	LOC_Os02g09750	NM_001052686*	1,859	1,482	0	1,470	489	53,564.72	7.77	33	OSJNBb0031B09	2	C10215S	25.5
8	*OsεLCY*	LOC_Os01g39960	AK072015 AK066182	3,745	1,861	9	1,623	540	60,136.89	7.18	52	B1097D05	1	Y2820R	93.1
9	*OsβOH1*	LOC_Os03g03370	AK287823	2,843	1,160	5	858	285	31,456.52	9.99	61	OJ1293G11	3	C721	6.0
10	*OsβOH2*	LOC_Os04g48880	AK060559	2,069	1,457	5	930	309	33,780.95	10.12	59	OSJNBa0011J08	4	R2732	96.0
11	*OsβOH3*	LOC_Os10g38940	CT837931*	2,026	1,172	4	879	292	31,684.62	9.97	61	OSJNBa0053C23	10	R716	68.6
12	*OsεOH1*	LOC_Os02g07680	AK066680 AK062235	10,919	2,401	14	1,716	571	64,133.59	7.09	42	OSJNBa0073A21	2	C1796	18.4
13	*OsεOH2*	LOC_Os02g57290	AK068163	8,098	4,217	15	1,932	643	69,861.79	5.72	41	OJ1136_C12	2	C1445	153.5
14	*OsεOH3*	LOC_Os10g39930	AK065689	4,980	1,876	9	1,686	561	62,138.48	6.87	40	OSJNBa0001014	10	C809	72.5
15	*OsZE*	LOC_Os04g37619	AB050884*	7,181	2,345	15	1,980	659	71,798.27	8.02	–	OSJNBa0064H22	4	S12568	67.2
16	*OsVDE*	LOC_Os04g31040	AF411133* AK061554	2,564	1,887	5	1,407	468	52,906.00	6.11	–	OSJNBa0083D01	4	C12216S	41.5

[a] TIGR locus ID of corresponding gene in release 5
[b] Accession no. of full-length cDNA sequence from KOME or NCBI (marked with asterisk)
[c] Length of genomic fragment in kb in TIGR release 5
[d] Length of cDNA (open reading frame + 5′UTR + 3′UTR)
[e] No. of introns within cDNA
[f] Length of open reading frame
[g] Length (no. of amino acids) of the deduced polypeptide
[h] Molecular weight of the deduced polypeptide in Dalton
[i] Isoelectric point of the deduced polypeptide
[j] Predicted length of the chloroplastic transit peptide by ChloroP 1.1 program; – indicates absence of cTP prediction
[k] Name of the BAC/PAC clone and number of chromosome on which the gene is present
[l] Neighboring marker of the gene
[m] Approximate cM position of the neighboring marker

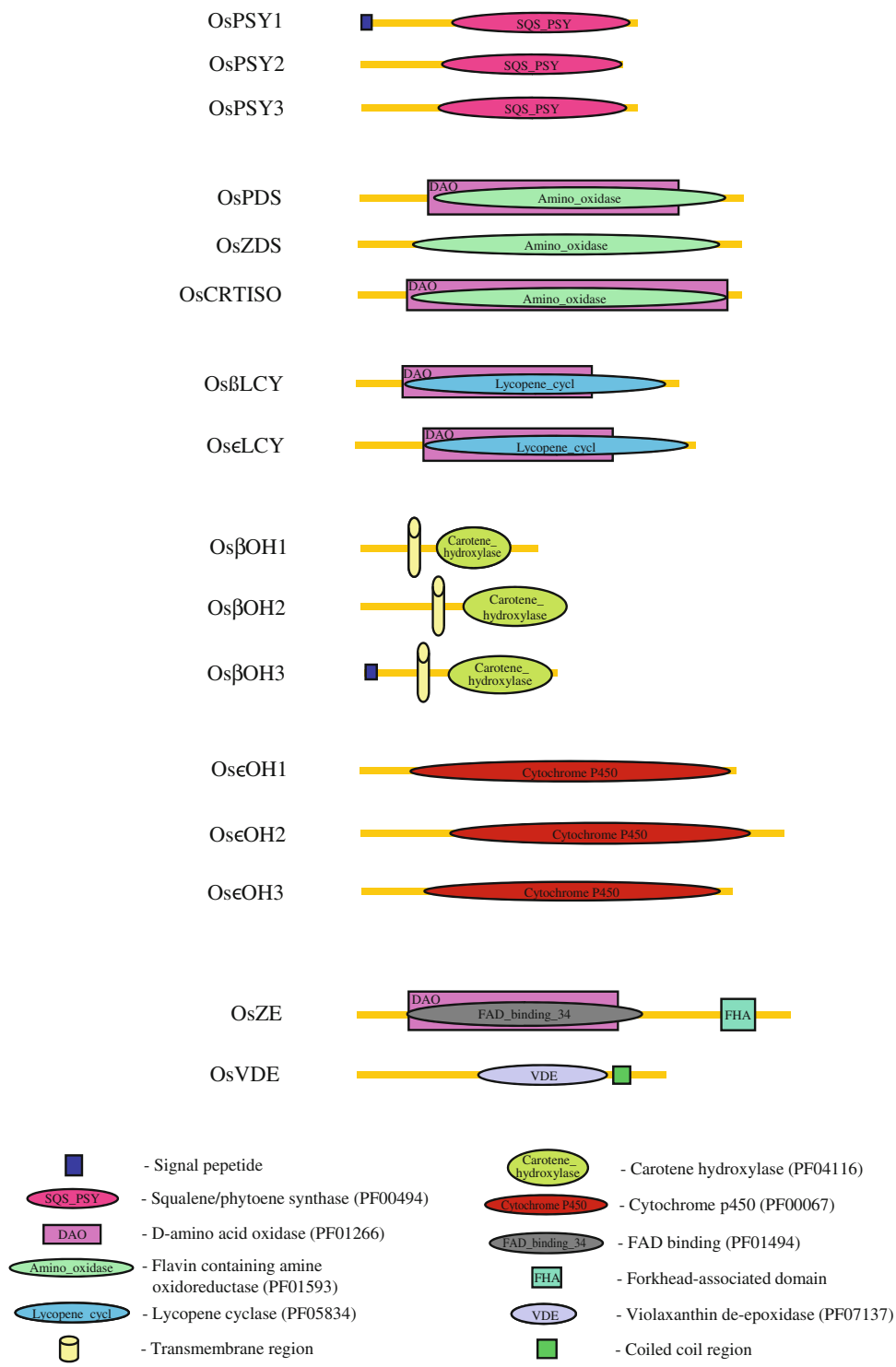


Fig. 1 Structure of proteins encoded by rice carotenoid biosynthesis genes. The protein structure is based on the presence of catalytic and other additional domains as identified by SMART. Different domains are highlighted in *different colors*

Table S2). In *Populus*, the number of introns was found to vary from 1 to 15, and out of 34 genes, 8 were intronless.

Domain analysis

All rice, *Populus* and *Arabidopsis* proteins were analyzed for domain structures employing SMART database and compared with each other. The domains identified in different carotenoid biosynthesis enzymes in rice are depicted in Fig. 1. The possible roles and a comparison of rice and *Populus* carotenoid biosynthesis enzymes with that of *Arabidopsis* are given in Table 2. For most rice proteins, the presence of a chloroplast transit peptide was predicted by ChloroP 1.1 program.

Table 2 Comparative analysis of domain organization in carotenoid biosynthesis enzymes of *Arabidopsis*, rice and *Populus*

Step in carotenoid biosynthesis pathway	Enzyme involved	*Arabidopsis* homologs	pfam domains present in *Arabidopsis* protein	Rice homologs	pfam domains present in rice protein	*Populus* homologs	pfam domains present in *Populus* protein
Dimerization of two molecules of GGPP (geranylgeranyl pyrophosphate) to phytoene	Phytoene synthase	AtPSY	SQS-PSY	OsPSY1	SQS-PSY and signal peptide	PtPSY1[a]	SQS-PSY
						PtPSY2[a]	SQS-PSY
				OsPSY2	SQS-PSY	PtPSY3[a]	SQS-PSY
						PtPSY4[a]	SQS-PSY
				OsPSY3	SQS-PSY	PtPSY5[a]	No domain
						PtPSY6[a]	SQS-PSY
						PtPSY7[a]	SQS-PSY
Desaturation of phytoene to yield ζ-carotene	Phytoene desaturase	AtPDS	Amine oxidase and DAO	OsPDS	Amine oxidase and DAO	PtPDS1	Amine oxidase and DAO
						PtPDS2[a]	Amine oxidase, DAO and Pyr_redox_2
Desaturation of phytoene to produce tetra-*cis*-lycopene	ζ-Carotene desaturase	AtZDS	Amine oxidase	OsZDS	Amine oxidase	PtZDS1	Amine oxidase and DAO
						PtZDS2[a]	Amine oxidase, DAO and Pyr_redox_2
Isomerization of tetra-*cis*-lycopene to all-trans-lycopene	Carotenoid isomerase	AtCRTISO	DAO	OsCRTISO	Amine oxidase and DAO	PtCRTISO1[a]	DAO and amine oxidase
						PtCRTISO2[a]	DAO and amine oxidase
Cyclization of lycopene (linear) to yield *β*-carotene (cyclic) by adding *β* rings	Lycopene *β*-cyclase	At*β*LCY	Lycopene cyclase and DAO	Os*β*LCY	Lycopene cyclase and DAO	Pt*β*LCY 1	Lycopene cyclase and DAO
						Pt*β*LCY 2	Lycopene cyclase and DAO
						Pt*β*LCY 3	Lycopene cyclase, D AO and Thi4
						Pt*β*LCY 4	Lycopene cyclase and DAO
						Pt*β*LCY 5[a]	Lycopene cyclase
						Pt*β*LCY 6[a]	Lycopene cyclase
						Pt*β*LCY 7[a]	No domain
Cyclization of lycopene (linear) to yield lutein (cyclic) by adding ε ring	Lycopene ε-cyclase	AtεLCY	Lycopene cyclase	OsεLCY	Lycopene cyclase and DAO	PtεLCY[a]	Lycopene cyclase

Table 2 continued

Step in carotenoid biosynthesis pathway	Enzyme involved	*Arabidopsis* homologs	pfam domains present in *Arabidopsis* protein	Rice homologs	pfam domains present in rice protein	*Populus* homologs	pfam domains present in *Populus* protein
Beta ring hydroxylation to produce β-carotene	Beta-carotene hydroxylase	AtβOH1	Transmem brane and FA hydroxylase	OsβOH1	Transmembrane and FA hydroxylase	PtβOH1[a]	Transmembrane and FA hydroxylase
		AtβOH2	Transmembrane and FA hydroxylase	OsβOH2	Transmembrane and FA hydroxylase	PtβOH2	Transmembrane and FA hydroxylase
		CYP97A3[b]	Cytochrome P450	OsβOH3	Transmembrane, FA hydroxylase and signal peptide	PtβOH3	Transmembrane and FA hydroxylase
						PtβOH4[a]	Transmembrane and FA hydroxylase
Epsilon ring hydroxylation to produce lutein	Epsilon-carotene hydroxylase	AtεOH1 (LUT1/ CYP97C1)	Cytochrome P450	OsεOH1	Cytochrome P450	PtεOH1[a]	Cytochrome P450
				OsεOH2	Cytochrome P450	PtεOH2[a]	Cytochrome P450
				OsεOH3	Cytochrome P450	PtεOH3	Cytochrome P450
Epoxidation of zeaxanthin to produce violaxanthin	Zeaxanthin epoxidase	AtZE	FAD binding_3 and FHA	OsZE	FAD_binding_3, FHA and DAO	PtZE1	FAD_binding_3, FHA and DAO
						PtZE2[a]	FAD_binding_3, FHA and DAO
						PtZE3[a]	FAD_binding_3 and DAO
						PtZE4[a]	No domain
De-epoxidation of violaxanthin to produce zeaxanthin again (xanthophyll cycle)	Violaxanthin de-epoxidase	AtVDE	VDE domain	OsVDE	VDE and coiled coil	PtVDE1[a]	VDE domain
						PtVDE2[a]	VDE domain

SQS-PSY squalene/phytoene synthase domain, *DAO* D-amino acid oxidase domain, *FHA* forkhead-associated domain, *Pyr_redox_2* pyridine nucleotide-disulfide oxidoreductase domain, *Thi4* domain involved in thiamine biosynthesis

[a] Partial protein sequence

[b] Tian et al. (2004)

Chromosomal distribution

To determine the chromosomal location of all the 16 carotenoid biosynthesis genes in rice, we determined the position (in base pair) of each gene on the rice chromosome pseudomolecules available at TIGR (release 5) and IRGSP (build 05). Sixteen carotenoid biosynthesis genes were found to be distributed on 10 of the 12 rice chromosomes. A diagrammatic representation of chromosomal distribution of carotenoid biosynthesis genes on the ten rice chromosomes is depicted in Fig. 2 (the exact position on rice chromosome pseudomolecules, in base pairs is given in Supplemental Table S2). Carotenoid biosynthesis genes of *Populus* were present on 12 of the 19 chromosomes. Thirteen genes were assigned to scaffolds that have not yet been assigned to any chromosome (Supplemental Table S2). To understand the mechanism underlying the evolution of carotenoid biosynthesis genes in rice, both tandem and segmental duplication events were studied. In this case, duplication was confined to only chromosomal block duplications, because none of them were found to be arranged in tandem. Only three carotenoid biosynthesis genes *OsβOH1* (LOC_Os03g03370), *OsβOH2* (LOC_Os04g48880) and *OsβOH3* (LOC_Os10g38940) were found to be present on the duplicated segmental regions of rice chromosomes, mapped by TIGR (Fig. 2; Supplemental Table S3) at a maximal length distance permitted between collinear gene pairs of 100 kb.

Phylogenetic analysis

To investigate the evolutionary relationship of rice genes, separate unrooted NJ and MP trees were generated for each enzyme, by alignment of their protein sequence with the corresponding homologs in *Arabidopsis* and other plant species (Fig. 3; Supplemental Fig. S2; Supplemental Table S4). Carotenoid biosynthesis enzymes of *Populus* were not considered for phylogenetic study as quite a few of the protein sequences were not of full length. Phylogenetic analysis of PSY in rice, *Arabidopsis* and other monocot and dicot plant species revealed that all PSYs cluster into two separate monocot- and dicot-specific clades, where most of the members show monophyletic pattern of origin (Fig. 3a). As shown in Fig. 3b, the product of *PDS*, a single copy gene in rice, was grouped with maize and daffodil PDS in monocot-specific clade, with separate clustering of *Arabidopsis* and other dicot PDS in a dicot-specific clade. Similarly, ZDS protein sequence, encoded by a single gene in rice, also formed two distinct monocot- and dicot-specific clades; OsZDS grouped with maize gene in monocot-specific clade and *Arabidopsis* and other dicot ZDS were clustered in dicot-specific clade (Fig. 3c). While studying the next enzyme of the pathway, CRTISO, it was observed that maize and *Oncidium* CRTISO clustered with rice homolog in monocot-specific group and all other dicot-specific CRTISO formed another group (Fig. 3d). For cyclases of carotenoid biosynthesis pathway, lycopene *β*- and ε-cyclase, two different and specific clades for *β*- and ε-cyclases were formed, which further subgrouped into monocot- and dicot-specific clades. Os*β*LCY grouped separately from *Arabidopsis* and *Citrus* *β*LCYs, similar to OsεLCY, which also formed a separate clade with TaεLCY (Fig. 3e). In *Arabidopsis*, two genes coding for *β*OH are known; however, in rice, three genes coding for *β*OH enzyme were found to be present, which grouped together with maize homologs in a separate monocot-specific branch (Fig. 3f). Two *Arabidopsis* *β*OHs were found in dicot-specific clade with *β*OHs from *Capsicum annuum* and other dicots. Different homologs for εOH were also found in three separate groups: OsεOH1 and OsεOH2, grouped with *Glycine max*

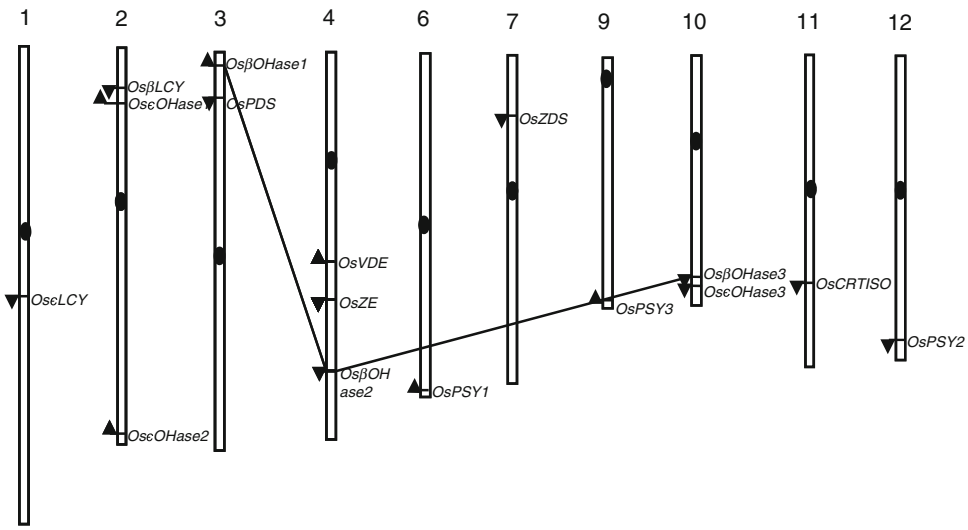


Fig. 2 Genomic distribution of carotenoid biosynthesis genes on rice chromosomes. *Black ovals* on the chromosomes (*vertical bars*) indicate the position of centromeres. The *arrows* next to gene names show the direction of transcription. The chromosome numbers are indicated at the *top* of *each bar*. The carotenoid biosynthesis genes present on duplicated chromosomal segments are connected by *lines*. The position of carotenoid biosynthesis gene on TIGR rice chromosome pseudomolecule (release 5) is given in Supplemental Table S2

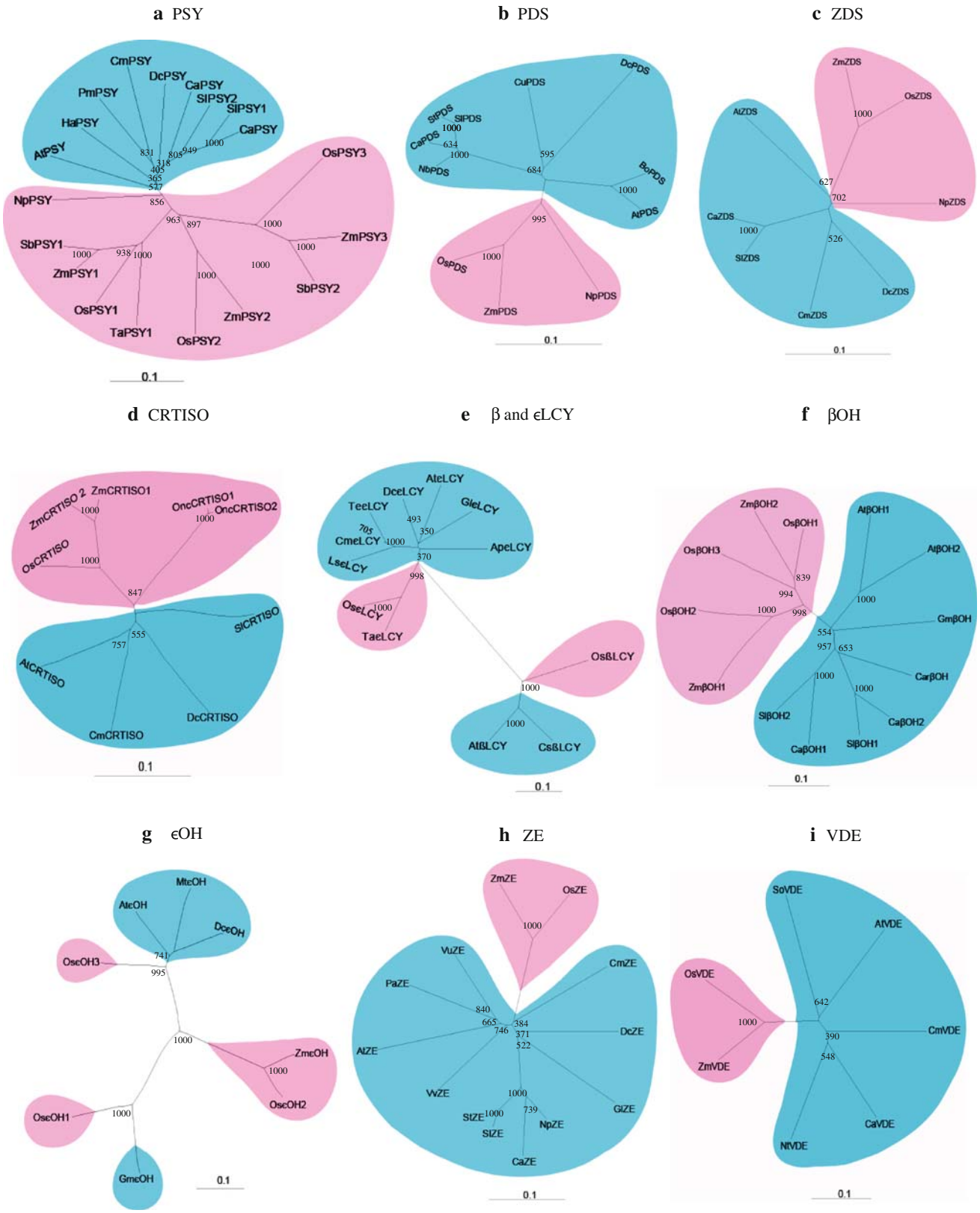


Fig. 3 Phylogenetic relationship of rice genes with *Arabidopsis* and other carotenoid biosynthesis genes. The unrooted tree was generated using ClustalX program by neighbor-joining method. Bootstrap values from 1,000 replicates are indicated at *each node*. Monocot and dicot-specific clades are indicated in *different colors*

and *Z. mays*, respectively; however, OsεOH3 was present with *Arabidopsis*, *Daucus* and *Medicago* homologs (Fig. 3g). On aligning rice ZE with other protein sequences, two separate dicot- and monocot-specific clades were formed where OsZE was clustered with ZmZE (Fig. 3h). Similarly, alignment of VDE also formed distinct dicot- and monocot-specific clades, with grouping of OsVDE with ZmVDE (Fig. 3i). For all the enzymes, the MP phylogentic trees generated by MEGA program also showed distinct clustering of monocot- and dicot-specific genes (Supplemental Fig. S2), as discussed above.

In this study, we have reported the occurrence of three and four *βOHs* in rice and *Populus*, respectively. A comparison of the genomic structures in Fig. 4 showed that both *Arabidopsis* genes, *AtβOH1* and *AtβOH2*, have seven exons and their exon sizes are very similar. The structure and exon sizes of *AtβOH2* are comparable with *OsβOH3*, where exons 2 (64 bp), 3 (72 bp) and 4 (138 bp) of *Arabidopsis* gene appear to have fused to form the second exon (276 bp) of rice gene. Accordingly, the length of 5′ and 3′UTRs of *AtβOH2* and *OsβOH3* are also quite comparable. In segmentally duplicated rice *β-carotene hydroxylases*, *OsβOH1* and *OsβOH3*, intron–exon structure and exon lengths are also very similar.

Full-length cDNA- and EST-based expression analysis

One or more full-length cDNAs were found to be available for 13 of 16 rice carotenoid biosynthesis genes in the TIGR database, and for three genes the full-length/partial cDNA sequences were present in the NCBI database (Table 1). In the NCBI database, we identified ESTs for all the rice genes, but the frequency of ESTs for individual genes varied greatly, ranging from 3 ESTs for *OsCRTISO* to 25 ESTs for *OsPSY1* (Supplemental Table S5). An analysis of EST sequences in the Gene Bank database revealed the presence of a large number of ESTs, having high sequence similarity with rice carotenoid biosynthesis genes in other monocots also (Supplemental Table S5). In *Populus*, most of the predicted genes have corresponding EST sequences in the database (Supplemental Table S6).

Quantitative real-time PCR-based expression analysis of rice carotenoid biosynthesis genes

To further study the organ-specific and light-regulated expression of rice carotenoid biosynthesis genes, real-time PCR was performed with total RNA isolated from dark-grown roots, light-grown roots, dark-grown shoots, light-grown shoots, and developing and mature seeds (Fig. 5). *OsPSY1* and *OsPSY2* have higher expression in light-grown tissues as compared to the dark-grown tissues and show almost seven to eightfold and four to fivefold higher expression, respectively, in light-grown shoot as compared to mature seed. In contrast, *OsPSY3* has almost similar expression level in light- and dark-grown root and shoot tissues. *OsPDS*, *OsZDS* and *OsCRTISO* have the highest expression in light-grown shoot, followed by developing and mature seed. *OsβCY* and *OsεLCY* also have higher expression in light-grown shoot, followed by seed and

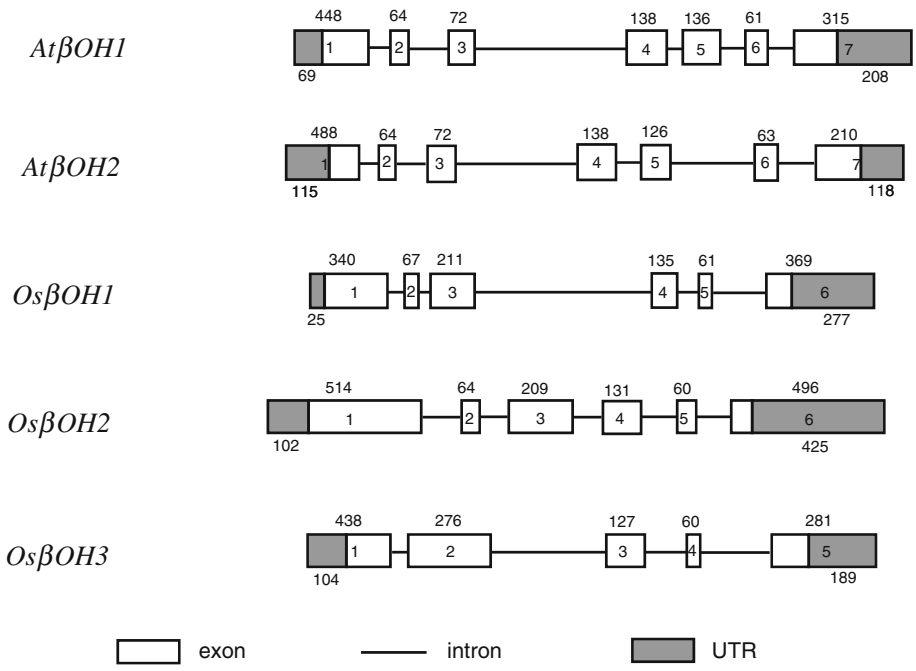


Fig. 4 Gene structure of *Arabidopsis* and rice *βOHs*. Exon–intron structures were calculated by comparing transcript sequences with genomic regions of the corresponding genes. The following full-length cDNAs were used: *AtβOH1* (AT4G25700.1), *AtβOH2* (AT5G52570.1), *OsβOH1* (AK287823), *OsβOH2* (AK060559), and *OsβOH3* (CT837931). *Numbers* within the *boxes* indicate the number of corresponding exon, *numbers* above exons represent the exon size (in base pairs) and *numbers* below the *shaded boxes* represent the length of 5′ and 3′UTRs (in base pairs)

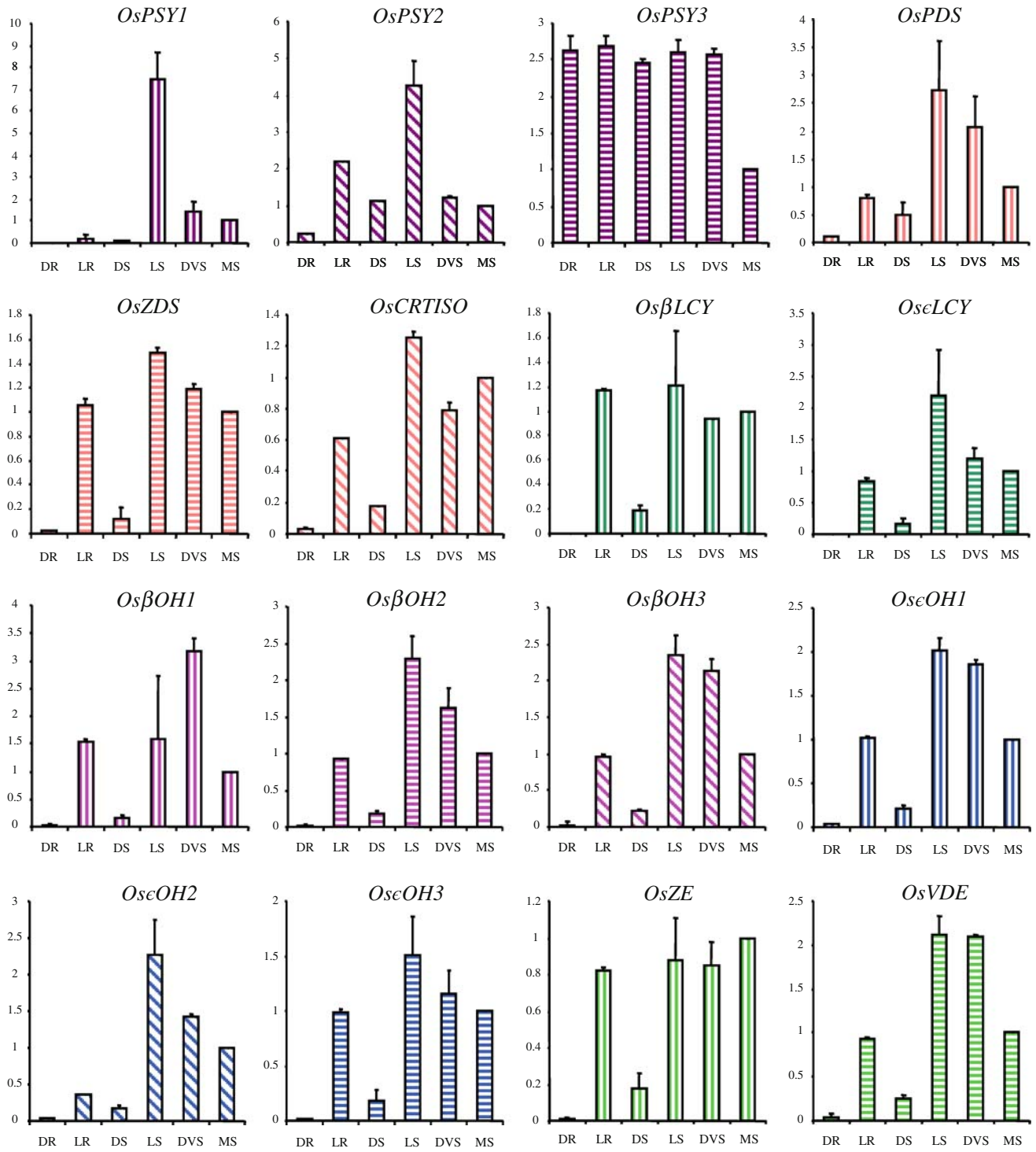


Fig. 5 Real-time PCR expression profile of rice carotenoid biosynthesis genes. For individual genes, the relative mRNA levels were normalized with respect to housekeeping gene *Actin*, in different tissues (*DR* dark-grown roots, *LR* light-grown roots, *DS* dark-grown shoots, *LS* light-grown shoots, *DVS* developing seed, *MS* mature seed)

light-grown root. All the three *βOHs* show high level of expression in light-grown shoot and root, and in developing and mature seeds, whereas the expression is low in dark-grown tissues (Fig. 5). The three *β*-hydroxylase genes, *OsβOH1*, *OsβOH2* and *OsβOH3*, have maximum expression in developing seed, light-grown shoot and light-grown shoot, respectively. The expression pattern of three εOH encoding genes of the pathway, *OsεOH1*, *OsεOH2* and

OsεOH3, is quite similar where light-grown shoot has slightly higher expression, followed by developing seed, mature seed, light-grown root and dark-grown shoot. *OsZE* has the highest expression in mature and developing seed, followed by light-grown shoot, light-grown root and dark-grown shoot (Fig. 5). The expression of *OsVDE* is highest in light-grown shoot, followed by developing and mature seed, light-grown root and dark-grown shoot.

Microarray-based expression analysis of rice carotenoid biosynthesis genes

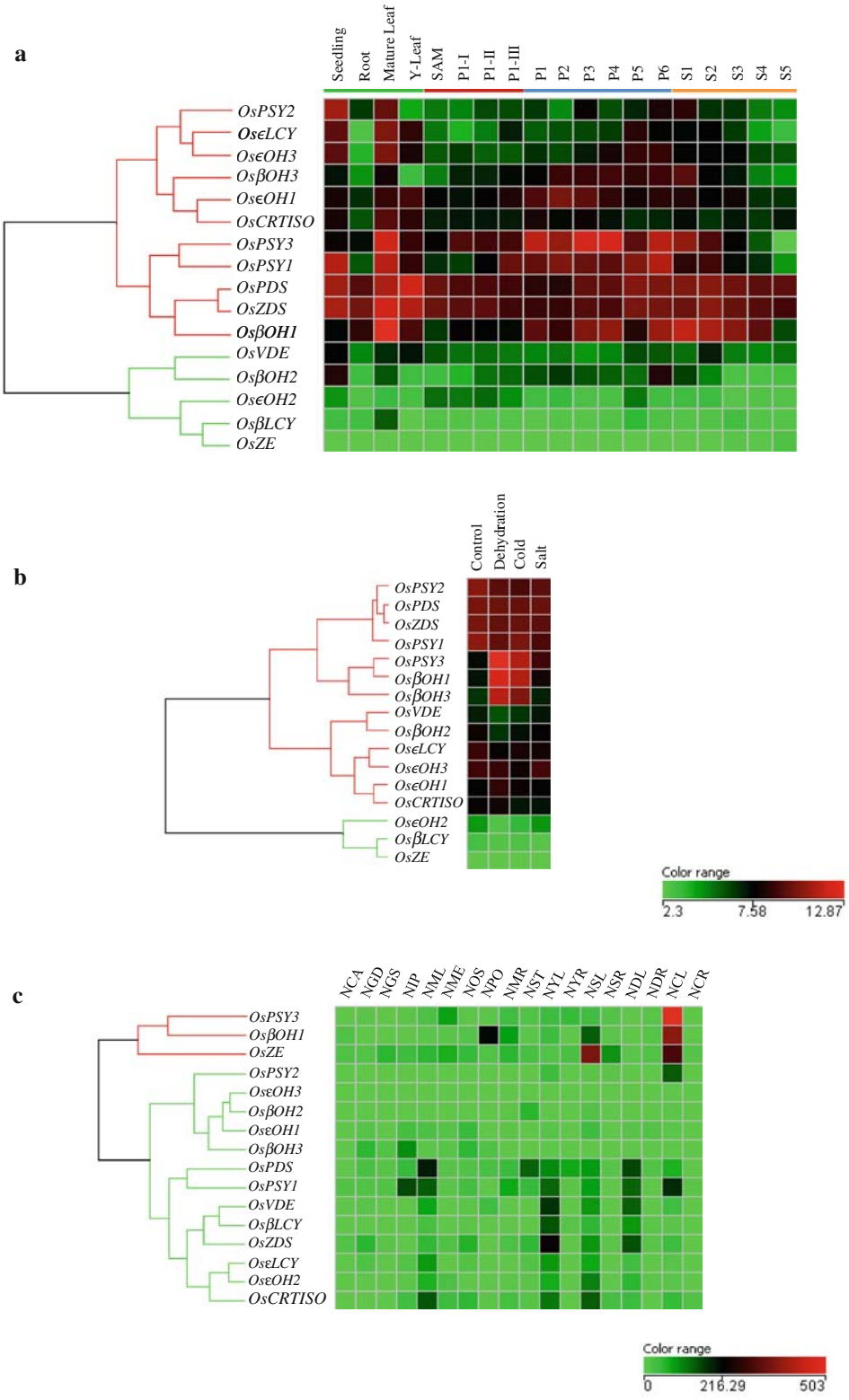
To study the transcript accumulation of rice carotenoid biosynthesis genes in different stages of vegetative and reproductive development, data set available in the public domain under the series accession number GSE6892 was utilized. On the basis of expression in different tissues/organs and developmental stages, a cluster dendrogram was made and all rice carotenoid biosynthesis genes were grouped into two main clades: the first group of genes had very low to moderate level of expression, and the second group had moderate to high level of expression in all vegetative and reproductive stages studied (Fig. 6a). In group 1 genes, all except *OsVDE* had almost negligible expression in late seed developmental stages (S3–S5) and variable expression pattern in other vegetative and reproductive stages; however, most of the genes had comparatively higher expression in photosynthetic tissues (seedling, mature leaf and Y leaf). In the second group, some genes were highly expressed in all the tissues/stages studied. *OsPSY1* and *OsPSY3* were highly expressed in photosynthetic tissues (seedling, mature leaf and Y leaf) and during panicle development (P1-III, P1-6), and showed a decrease in transcript abundance as seed maturation progressed (S3–S5). *OsβOH1* was also highly expressed in most of the stages and showed a decline in transcript level during seed maturation (S5); however, *OsPDS* and *OsZDS* were highly expressed in all the tissues and developmental stages analyzed, including late stages of seed development. The expression level of *OsεOH1* and *OsCRTISO* was reduced during late stages of panicle (P5–P6) and seed development (S1–S5), as compared to mature and Y leaf. *OsβOH3* also showed gradual reduction in expression during seed development (S2–S5), while the transcript level was highest during panicle (P2–P6) and early stage of seed development (S1) and not in photosynthetic tissues (seedling, mature leaf and Y leaf). *OsPSY2*, *OsεLCY* and *OsεOH3* showed highest expression in photosynthetic tissues (seedling and mature leaf), whereas the expression was highly reduced during late seed developmental stages (S3–S5).

To study the effect of abiotic stress on transcript profile of carotenoid biosynthesis genes in rice seedlings, expression analysis was performed under three conditions, viz. dehydration, cold and salt, using the data set available in the public domain under the series accession number GSE6901. Figure 6b clearly shows that the conditions of abiotic stress have almost no effect on the expression of *OsPSY1*, *OsPSY2*, *OsPDS*, *OsZDS*, *OsCRTISO*, *OsβLCY*, *OsεOH1* and *OsZE*. Expression of some genes, i.e., *OsVDE*, *OsβOH2*, *OsεLCY*, *OsεOH3* and *OsεOH2* is slightly reduced with respect to the control; however, transcript level of three carotenoid biosynthesis genes, *OsPSY3*, *OsβOH1* and *OsβOH3*, was found to be highly increased as compared to control. Expression of *OsβOH1* and *OsβOH3* was specifically increased in rice seedlings exposed to dehydration and cold stress. A comparison of expression profiles for duplicated carotenoid biosynthesis genes revealed that for *OsPSY* and *OsβOH*, expression of the duplicated gene copy has significant variation (Fig. 7). In case of duplicated *β-carotene hydroxylases,* transcript level of *OsβOH2* and *OsβOH3* is quite low, but they reveal a similar kind of expression pattern; however, expression pattern as well as transcript level of *OsβOH1* is strikingly different.

The strong induction of *β*OH encoding gene by abiotic stresses prompted us to check their promoter sequence (1 kb upstream the transcript start site) by searching against the plant CARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The promoter sequence of *OsβOH1* has *cis*-acting elements involved in abscisic acid and salicylic acid responsiveness (ABRE; 5 hits, TCA; 1 hit). The promoter sequence of *OsβOH3* also shows the presence of putative stress-response related *cis*-acting elements involved in abscisic acid responsiveness (ABRE; 5 hits, motif IIb; 1 hit) and low temperature responsiveness (LTR; 1 hit).

MPSS-based expression analysis of rice carotenoid biosynthesis genes

Expression analysis of rice carotenoid biosynthesis genes using MPSS data for 17-base signatures from 22 mRNA libraries showed that all the genes had MPSS signature in at least one of the libraries, indicating the presence of their transcript. On the basis of expression in different libraries, a cluster dendrogram was made where all rice carotenoid biosynthesis genes were grouped into two main clades. The first group has three genes, *OsPSY3*, *OsβOH1* and *OsZE*, having high transcript abundance in leaves under salt or cold stress conditions, and the second group includes rest 13 genes having low to moderate level of expression in all vegetative and reproductive stages studied. However, the conditions of abiotic stress were found to have almost negligible effect on the expression of these genes (Fig. 6c).



Discussion

During an extensive search of the rice genome and domain analysis of the putative candidate genes, carotenoid biosynthesis genes in rice were found to have 16 potential non-redundant members. A similar bioinformatics-based approach revealed the presence of 34 genes coding for carotenoid biosynthesis enzymes in *Populus*. However, in *Arabidopsis*, 11 genes coding for carotenoid biosynthesis pathway enzymes have been characterized (DellaPenna and

◀ **Fig. 6** Expression profiles of rice carotenoid biosynthesis genes based on rice microarray and mRNA MPSS data. Microarray-based expression profile of rice carotenoid biosynthesis genes **a** during different vegetative and reproductive developmental stages, and **b** under abiotic stress conditions. The average log signal values of rice carotenoid biosynthesis genes in various tissues/organs/developmental stages (7-day-old light-grown seedling, roots from 7-day-old light-grown seedlings, mature leaf, Y leaf, different stages of panicle: up to 0.5 mm, SAM; 0.5–2.0 mm, P1-I; 2.1–5.0 mm, P1-II; 5.1–10.0 mm, P1-III; 0–3 cm, P1; 3–5 cm, P2; 5–10 cm, P3; 10–15 cm, P4; 15–22 cm, P5; 22–30 cm, P6, and seed: 0–2 dap, S1; 3–4 dap, S2; 5–10 dap, S3; 11–20 dap, S4; 21–29 dap, S5), and abiotic stress treatments (control; 7-day-old light-grown seedling, salt; 200 mM NaCl solution for 3 h, desiccation; dried for 3 h between folds of tissue paper at 28 ± 1°C, cold; 4 ± 1°C for 3 h) are presented by *cluster display*. The *color scale* (representing log signal values) is displayed at the *bottom*. **c** Rice mRNA MPSS data-based expression profile of rice carotenoid biosynthesis genes. The number of significant signatures are shown as transcript per million (tpm). The *color scale* (representing tpm) is shown at the *bottom*. The description of the mRNA libraries is: *NCA* 35-day-old callus, *NGD* 10 days germinating seedling grown in dark, *NGS* 3 days germinating seed, *NIP* 90 days immature panicle, *NML* 60 days mature leaf (representing an average of four replicates; A, B, C and D), *NME* 60 days meristematic tissue, *NOS* ovary and mature stigma, *NPO* mature pollen, *NMR* 60 days mature root (representing an average of two replicates; A and B), *NST* 60 days stem, *NYL* 14 days young leaves, *NYR* 14 days young roots, *NSL* 14 days young leaves stressed in 250 mM NaCl for 24 h, *NSR* 14 days young roots stressed in 250 mM NaCl for 24 h, *NDL* 14 days young leaves stressed in drought for 5 days, *NDR* 14 days young roots stressed in drought for 5 days, *NCL* 14 days young leaves stressed at 4°C for 24 h, *NCR* 14 days young roots stressed at 4°C for 24 h. On the *left side* of all three expression maps cluster dendrogram is shown

Pogson 2006). In *Populus*, such a high number of genes could be due to incomplete finishing of chromosomal BACs/PACs or alleles from unassembled haplotypes, which are potential artifacts from shotgun assembly of this highly heterozygous genome. However, whole-genome (salicoid) duplication event accounting for the presence of paralogous sequences and recent evolution of this genus (Tuskan et al. 2006) might have also contributed to the evolution of carotenoid biosynthesis genes. Post-duplication, the gene copies could have diverged functionally and acquired specific functions, as evidenced by significant deviation in the frequency of ESTs (Supplemental Table S6; Segerman et al. 2007).

The relationship among different members of a gene family can be analyzed by gene structure, i.e., the number and distribution of introns (Li et al. 2006). A cross comparison among rice, *Populus* and *Arabidopsis* genes revealed more or less similar number of exons in homologous genes, although the length of introns was quite variable (Supplemental Table S2). Interestingly, all the seven gene copies coding for *β*LCY in *Populus* were found to be intronless, similar to their *Arabidopsis* and rice counterparts. These results indicate that in the common ancestor of each gene there would be an ancestral exon–intron structure (or intronless), and after the divergence of genes (e.g., by speciation), some genes changed and others retained the original pattern. The enzymes involved in carotenoid biosynthesis were found to possess specific domains depending on the type of reaction they catalyze in the pathway (Fig. 1). All the homologous genes in rice, *Populus* and *Arabidopsis*, were found to have a highly conserved catalytic domain (Table 2); some additional domains (from the same superfamily) were also present in rice and/or *Populus* homologs. Overlapping domains, involved in catalyzing similar kind of reactions, and domains unrelated to carotenoid biosynthesis, which can have possible roles in diverse metabolic and regulatory activities, were also present. While studying the chromosomal location of rice carotenoid biosynthesis genes, their distribution pattern revealed no evident clustering, and chromosomal block duplications were found to be responsible for the evolution of carotenoid biosynthesis genes in rice.

While studying the evolutionary relationship of rice carotenoid biosynthesis genes with *Arabidopsis* and other plant species, most of the rice enzymes grouped with other monocot homologs with the support of good bootstrap scores (Fig. 3; Supplemental Fig. S2). Duplication of PSY, catalyzing the first step of the pathway, was expected to have taken place only before the members of the grass family diverged, since three *PSY* copies were conserved in the Poaceae family (Li et al. 2008a). Welsch et al. (2008) have shown that *OsPSY1* is the ancestral gene, as it has the highest structural resemblance with *AtPSY*, and *OsPSY2* and *OsPSY3* is the sister pair resulting from the second duplication event. This phylogenetic relationship is clearly evident in Fig. 3a, as three distinct subclades with a good bootstrap support are formed within the monocot-specific clade. The PSYs of rice, maize and *Sorghum* are not clustered with tomato PSYs, which indicates that the duplication of *PSY* in the Solanaceae and Poaceae family has taken place independently. OsPDS, OsZDS, OsCRTISO, Os*β*LCY and OsεLCY were also found to be grouped in separate monocot-specific clades suggesting a close evolutionary relationship among monocot homologs (Fig. 3b–e). The three *βOHs* of rice were also present in a separate monocot-specific clade with further subgrouping into smaller clades, which indicates that they are proteins encoded by homologous genes (Fig. 3f). It is not clear whether the duplication of *βOHs* is unique for rice or they have been duplicated in the whole Poaceae family (like *PSY* duplication) or in a large group of monocots. MP phylogenetic trees, generated by MEGA program, also illustrated a similar pattern of grouping where rice genes were close to monocot homologs with good bootstrap values (Supplemental Fig. S2).

Lineage-specific expansion of gene families has played an important role in the evolution of bacterial, archaeal and eukaryotic genomes (Jordan et al. 2001; Lespinet et al.

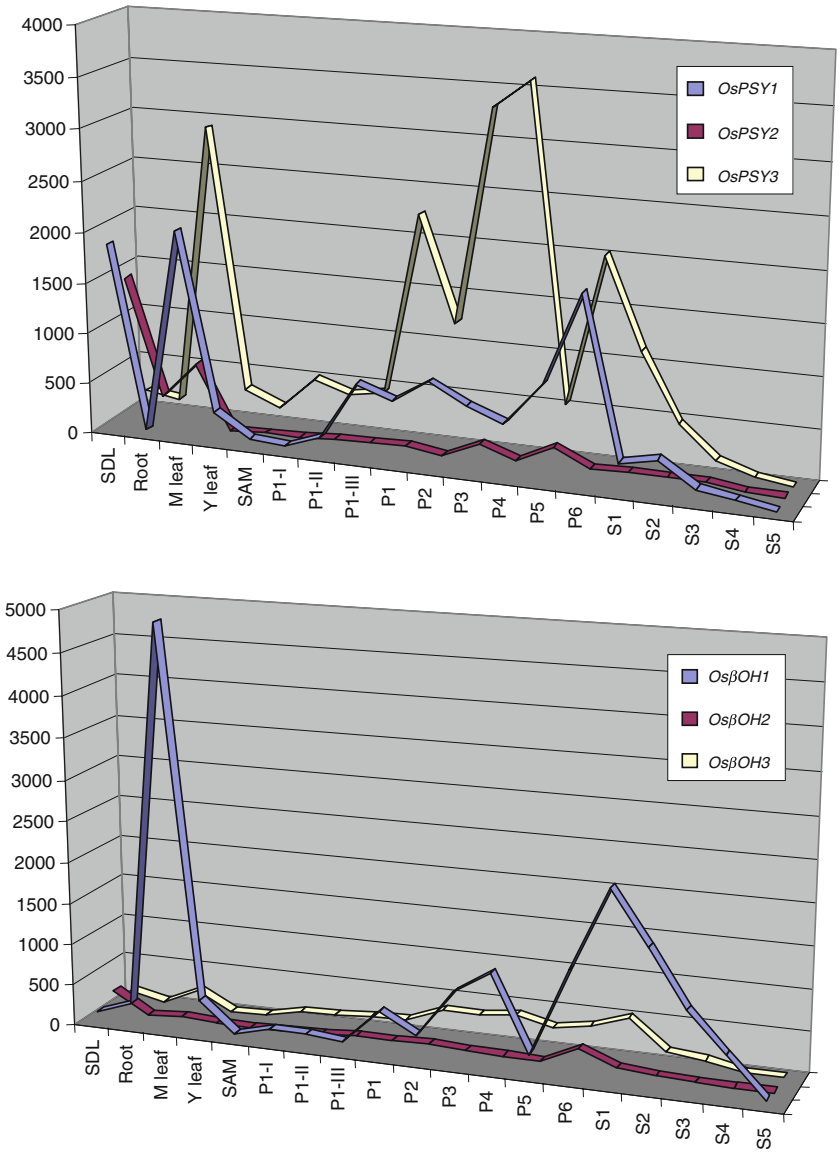


Fig. 7 Expression profile of duplicated carotenoid biosynthesis genes in rice genome. *X*-axis represents the developmental stages, *Y*-axis represents the raw expression values obtained using microarrays

2002). It has been observed that in rice, some carotenoid biosynthesis genes are present as multiple copy genes, while in *Arabidopsis* the copy number is less. It appears that the higher number of *PSY* copies provides an ability to fine tune and control the flux in carotenoid pathway where each homolog has a different tissue-specific expression, plays a different role in plant growth and development, and is differentially regulated. In tomato, two *PSY* are expressed in different types of plastids (Fraser et al. 1999); rice and maize *phytoene synthase* have different tissue-specific expression, play different roles and are differentially regulated by light or abiotic stress (Welsch et al. 2008; Li et al. 2008b). Another enzyme of the pathway, encoded by multiple gene copies in rice and *Populus*, is βOH, which is a non-heme diiron monooxygenase and catalyzes *β*-hydroxylation of carotene. In *Arabidopsis*, two βOH closely related at protein levels are present. These genes have different expression levels, as they are not entirely equivalent and likely to have specialized functions (Tian et al. 2004). Similarly, two closely related, but functionally distinct *βOHs*, are maintained in tomato and pepper (Dellapenna and Pogson 2006). During the course of evolution, the process of gene duplication and sequence divergence might have resulted in homologs having slightly different structure and new properties. These small changes in the sequences may lead to generation of new domains and altered protein functions, thus playing an important role in evolution, since additional domains are known to modulate the function of the conserved catalytic domain (Bashton and Chothia 2007).

Another hydroxylase of the pathway, εOH, which belongs to CYP97 family of P450 enzymes and catalyzes ε-hydroxylation of carotene, is encoded by a single copy gene in *Arabidopsis* (Tian et al. 2004). In rice, three related

genes encoding for CYP97 family of P450 enzymes (Clans A, B and C) were found to be present in the genome (Fig. 3g). In an earlier study, the substrate specificity and activity of two of rice P450 CYP97 enzymes (Clans A and C) was studied by functional complementation in *E. coli* and revealed that Clan A enzyme, CYP97A4, is a βOH with only minor ε-hydroxylation activity; however, Clan C enzyme CYP97C2 is an εOH (Quinlan et al. 2007). Till now, there is no functionally demonstrated carotene hydroxylase representative for members of CYP97 Clan B, which is expected to be involved in the synthesis of some common plant metabolite (Quinlan et al. 2007). We have considered CYP97 Clan B enzyme for our analysis, as further studies regarding functional validation are needed to confirm the absence of carotene hydroxylation activity. Although these two types of carotene hydroxylases use different mechanisms and have evolved independently, still β-ring hydroxylation can be performed either by non-heme diiron β-ring hydroxylases or CYP97A4 (Tian et al. 2004), and βOH can also show some activity toward ε-rings (Sun et al. 1996). It appears to be a case of convergent evolution, where enzymes having different protein structure are observed to have plasticity in their function.

To analyze the expression profile of these genes, the frequency of cDNAs or ESTs present in different databases can serve as an important resource (Adams et al. 1995). For all carotenoid biosynthesis genes of rice and *Populus*, one or more full-length cDNAs/ESTs were found to be present in the TIGR/NCBI database (Table 1; Supplemental Table S5; Supplemental Table S6), which confirm their existence and expression. ESTs, showing significant homology with rice carotenoid biosynthesis genes, were also found in other monocots, indicating existence of their homologs in other species (Supplemental Table S5).

Given the role of carotenoids in photoprotection and photomorphogenesis (Dellapenna and Pogson 2006), their biosynthesis genes are expected to be highly expressed in light-grown tissues. Since the first gene of carotenoid biosynthesis pathway, *PSY* is known to be light-regulated (von Lintig et al. 1997; Welsch et al. 2000), expression of other genes was also studied in light/dark-grown tissues, revealing differential expression of these genes. Higher expression of *OsPSY1* and *OsPSY3* in light-grown tissues and similar transcript level of *OsPSY2* under light/dark conditions (Fig. 5) are in accordance with the earlier published literature where *OsPSY3*, unlike its paralogs, was not found to be light regulated (Welsch et al. 2008). Expression of *OsPDS*, *OsZDS* and *OsCRTISO* was also found to be highest in light-grown shoot, suggesting control of expression by light. *OsPDS* can play an important regulatory role in rice carotenoid biosynthesis, as regulation of *PDS* level is shown to be an important component in the regulation of ABA level during maize seed development (Hable et al. 1998). Transcript levels of *OsβCY* and *OsεLCY* were also found to be higher in light-grown shoot, suggesting their possible regulation by light. In maize, variation at *εLCY* locus affected the flux down α- or β-carotene branch of the pathway, thereby creating natural variation for carotenoid composition and vitamin A precursors (Harjes et al. 2008). Three *β*- and *εOHs* of rice were expressed in almost all tissues, with low transcript level in dark-grown tissues. Similar to the observed expression pattern, two *βOHs* in *Arabidopsis* were expressed in all tissues, and in leaves their expression was induced by excess light (Tian and Dellapenna 2001; Rossel et al. 2002). *OsZE* showed the highest transcript level in seed tissue similar to *Arabidopsis*, where expression of *ZE* was shown to be enhanced during the maturation phase of seed development, when ABA accumulation was maximal (Audran et al. 2001). Increased expression of *OsZE* in seed, as compared to other tissues, can be related to ABA accumulation as it is the last step of carotenoid biosynthesis pathway and provides direct supply of precursor for ABA synthesis. *OsVDE* had the highest transcript level in light-grown shoot followed by seed tissue. In higher plants, VDE forms part of the xanthophyll cycle (Niyogi 1999), thus explaining higher expression of *OsVDE* in light-grown shoot. In *Citrus* fruits, transcript level for many carotenoid biosynthesis genes were found to increase during fruit ripening and maturation, which led to a change from β, ε-carotenoids to β, β-carotenoids resulting in a shift in peel color (Kato et al. 2004). Similarly, enhanced expression of carotenoid biosynthesis genes in seed tissue can account for increased ABA levels in seed, as ABA is required for seed maturation and dormancy and carotenoid biosynthesis pathway provides precursors for ABA biosynthesis (Fang and Chu 2008).

Earlier studies have shown that carotenoids or carotenoid intermediates are abundant in photosynthetic tissues, but absent in rice endosperm (Burkhardt et al. 1997). Gallagher et al. (2004) demonstrated that both *OsPSY1* and *OsPSY2* are transcribed and encode for functional enzymes in the endosperm; however, endosperm carotenoid levels correlate only with *OsPSY1* levels. Thus, transcript accumulation and possible roles of other genes in rice carotenogenesis were analyzed revealing that almost all genes were expressed in green photosynthetic tissues and at different stages of panicle/seed development, although there was variation in transcript abundance (Fig. 6a). However, localization of encoded enzymes in different cell compartments can also affect their function in vivo, as earlier studies have shown the existence of carotenoid biosynthesis in different cell compartments: chloroplast in leaf and amyloplast in seed, both having a different plastid and enzyme structure (Gallagher et al. 2004). Even in chloroplast, envelope and thylakoid membrane fractions have different carotenoid profiles (Jeffery et al. 1974; Siefermann-Harms et al. 1978).

The observed accumulation of carotenoids in any tissue is the net result of its biosynthesis, turnover and stable storage. Earlier reports have shown that carotenoids are synthesized in nearly all types of plastids, but the highest accumulation is observed in chloroplasts and chromoplasts (Howitt and Pogson 2006). In chromoplasts, carotenoid-lipoprotein sequestering substructures were found to be present for effective sequestration and retaining of large amounts of carotenoids (Vishnevetsky et al. 1999). In carrot and cauliflower *or* mutants, high levels of carotenoid accumulation were found to associate with large amounts of carotenoid sequestering structures (Li et al. 2001; Lopez et al. 2008). These observations suggest that carotenoid sequestering structures may prevent the carotenoid end products from overloading plastid membranes, the site of carotenoid biosynthesis, and avoid a negative feedback of the biosynthesis pathway by the end products. Since lipids are not known to be present in starchy endosperm of rice, it can result in the absence of carotenoid sequestering structures in elaioplasts and consequently carotenoid-free rice endosperm.

The transcript profile of duplicated genes can indicate whether they are homologs created by gene duplication and perform different functions (neofunctionalization), or together perform part of the original function (subfunctionalization; Shan et al. 2007). The three *PSYs* in rice are already known to differ from each other with respect to tissue-specific expression and regulation (Welsch et al. 2008), which is also evident in the present expression analysis (Fig. 7a). Similarly, in case of beta-carotene hydroxylases, *OsβOH2* and *OsβOH3* have a similar and low level of expression, whereas expression of *OsβOH1* is quite high and shows a lot of variability (Fig. 7b).

In a natural environment, plants frequently encounter adverse growth conditions and thereby develop various adaptive mechanisms to deal with different kinds of abiotic stresses, and the one mediated via phytohormone abscisic acid is widely studied (Tuteja 2007). Under abiotic stress conditions, production of ABA is increased, which in turn promotes the expression of many genes involved in stress signaling and finally enables the plant to tolerate and survive stress by maintaining homeostasis. Carotenoid biosynthesis is known to provide precursors (9-*cis*-epoxycarotenoids) for ABA biosynthesis and may be a potentially limiting factor affecting flux of these compounds to this pathway (Nambara and Marion-Poll 2005). This observation is consistent with an earlier report by Welsch et al. (2008), where they have discussed increased expression of *OsPSY3* in roots, after salt treatment and drought. In the present study, however, transcript abundance of *OsPSY3* in seedlings increased significantly under dehydration and cold stress, whereas salt stress caused only a slight increase in its expression. Increased expression of carotenoid biosynthesis genes during stress conditions provides the plant an ability to combat deleterious stress effects; however, activity of their biosynthesis enzymes might represent a bottleneck for increased production. Probably to overcome this limitation, their expression is up-regulated by abiotic stress conditions. Increased transcript abundance of *OsPSY3*, *OsβOH1* and *OsβOH3* during dehydration and cold stress could also be involved in providing increased supply of precursors for stress-induced ABA accumulation, which may be translocated to other plant tissues having limiting precursors for ABA biosynthesis. During osmotic stress conditions, ABA levels were found to increase in maize roots and a part of this ABA was translocated to other plant tissues (Rivier et al. 1983). In tomato leaves, the expression of ABA biosynthesis gene (*NCED*) is known to increase under abiotic stress conditions (Thompson et al. 2000a); however, the expression of upstream carotenoid biosynthesis enzymes is not affected since the precursors (epoxycarotenoids) are abundantly present (Thompson et al. 2000b). Unlike leaves, drought stress-induced accumulation of ABA in tomato roots was associated with increased transcript levels of *NCED*, as well as *ZE* and *βOH* (Thompson et al. 2000b, 2007), which shows that ABA precursors are limited in roots. The presence of ABRE (ABA-response elements) in the promoter region of *OsβOH1* and *OsβOH3* indicates their possible activation by ABA. Initially, expression of these genes is increased in response to abiotic stress, which provides more carotenoid precursors for ABA biosynthesis. The increased amount of ABA would possibly increase the expression of these genes further, resulting in ABA-enhanced ABA biosynthesis. The presence of SA-response elements was also observed in the promoter sequence of *OsβOH1*. In *Arabidopsis*, salicylic acid is also known to play a role in plant response to adverse environmental conditions, such as it potentiates the abiotic stress response during salt and osmotic stress conditions by increasing reactive oxygen species generation in the photosynthetic tissue (Borsani et al. 2001).

For obtaining a comprehensive coverage of gene expression, measurements by using a combination of transcription profiling technologies is recommended (Oudes et al. 2005). Further quantitative assessment of transcript abundance by a different transcription profiling technique, MPSS, a digital method for analyzing gene expression, was utilized as it generates 100s of 1,000s of molecules per reaction, thereby facilitating detection of significant differential expression even for genes having very low expression levels (Nobuta et al. 2007). Significant signatures (which uniquely identify any individual gene) were found to be available for all the genes in at least 1 of the 22 libraries (Fig. 6c). MPSS data from abiotic stress treatments showed that the transcript level of *OsPSY3* and *OsβOH1* increased in cold-stressed

young leaves (NCL), as also observed in microarray-based expression analysis (Fig. 6b); however, contrary to microarray analysis, the expression of these genes was unaffected in drought-stressed leaves (NDL). Similarly, *OsβOH3* also did not show any change in transcript abundance while analyzing MPSS data from cold-stressed (NCL) or drought-stressed (NDL) leaves. During microarray-based expression analysis, transcript level of *OsZE* was found to be unaffected by stress conditions, whereas MPSS expression data showed enhanced expression of this gene in leaves under salt and cold stress conditions. Some earlier studies have also shown that microarray and MPSS technologies did not correlate well on a quantitative basis for transcript abundance measurements, thereby suggesting the MPSS platform to be more variable than microarrays for RNA abundance measurement (Chen et al. 2007). Moreover, data from two microarray technologies (Affymetrix and Agilent) were found to be more consistent with each other than with MPSS (Chen et al. 2007).

To evaluate the utility of carotenoid biosynthesis genes for functional analysis, we gathered additional information regarding the phenotype of insertion lines of rice *Tos17* retrotransposon insertion mutants (Miyao et al. 2003) using the BLAST program with the aid of *Tos17* mutant panel database (http://tos.nias.offrc.go.jp/). A total of 13 insertion mutants, corresponding to five genes of the carotenoid biosynthesis pathway, revealed variant phenotypes from being super dwarf to viviparous (Supplemental Table S7), suggesting an important role of these compounds in regulating plant growth, development and different metabolic processes by synthesizing antioxidants and signaling molecules.

In conclusion, using an in silico approach, 16 and 34 genes involved in carotenoid biosynthesis were found to be present in rice and *Populus* genome, respectively. Sequence comparison with *Arabidopsis* and other known carotenoid biosynthesis genes showed that the encoded proteins had a highly conserved structure. Genomic organization and phylogenetic analysis revealed the potential mechanisms responsible for the evolution of genes within and among species. Full-length cDNA- and EST-based expression analysis showed that all the rice and *Populus* carotenoid biosynthesis genes were expressed. Most of the rice genes appear to be light-regulated, a majority of genes are differentially expressed in different tissues/organs, vegetative and reproductive development stages, and expression of some genes is enhanced under abiotic stress conditions. During rice seed development, almost all the genes are expressed in the mature seed stage which indicates that some other regulatory mechanism (at post-transcriptional/translational level or during localization in different subcellular compartments) affects enzyme synthesis/activity resulting in a carotenoid-free endosperm.

Acknowledgments This work was financially supported by the Department of Biotechnology, Government of India, the University Grants Commission and the Council of Scientific and Industrial Research, New Delhi (research fellowship to N. C. and A. N.). The TIGR, KOME, TAIR, NCBI and *Populus* database resources are gratefully acknowledged for the detailed sequence information.

References

Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O et al (1995) Initial assessment of human genome diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377:3–17

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Audran C, Liotenberg S, Gonneau M, North H, Frey A, Tap-Waksman K, Vartanian N, Marion-Poll A (2001) Localization and expression of *zeaxanthin epoxidase* mRNA in *Arabidopsis* in response to drought stress and during seed development. Aust J Plant Physiol 28:1161–1173

Bashton M, Chothia C (2007) The generation of new protein functions by the combination of domains. Structure 15:85–99

Beveridge CA, Gresshoff PM, Rameau C, Turnbull CGN (2003) Additional signaling compounds are required to orchestrate plant development. J Plant Growth Regul 22:15–24

Borsani O, Valpuesta V, Botella MA (2001) Evidence for a role of salicylic acid in the oxidative damage generated by NaCl and osmotic stress in *Arabidopsis* seedlings. Plant Physiol 126:1024–1030

Bouvier F, Dogbo O, Camara B (2003a) Biosynthesis of the food and cosmetic plant pigment bixin (annatto). Science 300:2089–2091

Bouvier F, Suire C, Mutterer J, Camara B (2003b) Oxidative remodeling of chromoplast carotenoids: identification of the carotenoid dioxygenase *CsCCD* and *CsZCD* genes involved in *Crocus* secondary metabolite biogenesis. Plant Cell 15:47–62

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Britton G, Jensen SL, Pfander H (2004) Carotenoids handbook. Birkhäuser Verlag, Basel

Burkhardt P, Beyer P, Wunn J, Kloti A, Armstrong G, Schledz M, von Lintig J, Potrykus I (1997) Transgenic rice (*Oryza sativa*) endosperm expressing daffodil (*Narcissus pseudonarcissus*) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. Plant J 11:1071–1078

Calucci L, Capocchi A, Galleschi L, Ghiringhelli S, Pinzino C, Saviozzi F, Zandomeneghi M (2004) Antioxidants, free radicals, storage proteins, puroindolines, and proteolytic activities in bread wheat (*Tritium aestivum*) seeds during accelerated aging. J Agric Food Chem 52:4274–4281

Chen J, Agrawal V, Rattray M, West MAL, St Clair DA, Michelmore RW, Coughlan SJ, Meyers BC (2007) A comparison of microarray and MPSS technology platforms for expression analysis of *Arabidopsis*. BMC Genomics 8:414

Dall'Osto L, Lico C, Alric J, Giuliano G, Havaux M, Bassi R (2006) Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection in vivo under strong light. BMC Plant Biol 6:32

Dall'Osto L, Fiore A, Cazzaniga S, Giuliano G, Bassi R (2007) Different roles of alpha- and beta-branch xanthophylls in photosystem assembly and photoprotection. J Biol Chem 282:35056–35068

Davidson PA, Hunter CN, Horton P (2002) Overexpression of *beta-carotene hydroxylase* enhances stress tolerance in *Arabidopsis*. Nature 418:203–206

Dellapenna D, Pogson BJ (2006) Vitamin synthesis in plants: tocopherols and carotenoids. Annu Rev Plant Biol 57:711–738

Durbin R, Eddy SR, Krogh A, Mitchison GJ (1998) Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge University Press, London

Fang J, Chu C (2008) Abscisic acid and the pre-harvest sprouting in cereals. Plant Signal Behav 3:1046–1048

Franco AC, Matsubara S, Orthen B (2007) Photoinhibition, carotenoid composition and the co-regulation of photochemical and non-photochemical quenching in neotropical savanna trees. Tree Physiol 27:717–725

Fraser PD, Kiano JW, Truesdale MR, Schuch W, Bramley PM (1999) Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. Plant Mol Biol 40:687–698

Gallagher CE, Matthews PD, Li F, Wurtzel ET (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of grasses. Plant Physiol 135:1776–1783

Hable WE, Oishi KK, Schumaker KS (1998) Viviparous-5 encodes phytoene desaturase, an enzyme essential for abscisic acid (ABA) accumulation and seed development in maize. Mol Gen Genet 257:167–176

Harjes C, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG et al (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319:330–333

Havaux M, Niyogi KK (1999) The violaxanthin cycle protects plants from photooxidative damage by more than one mechanism. Proc Natl Acad Sci USA 96:8762–8767

Holt NE, Zigmantas D, Valkunas L, Li XP, Niyogi KK, Fleming GR (2005) Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science 307:433–436

Howitt CA, Pogson BJ (2006) Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ 29:435–445

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jeffery SW, Douce R, Benson AA (1974) Carotenoid transformations in the chloroplast envelope. Proc Natl Acad Sci USA 71:807–810

Jordan IK, Makarova KS, Spouge JL, Wolf YI, Koonin EV (2001) Lineage-specific gene expansions in bacterial and archaeal genomes. Genome Res 11:555–565

Kato M, Ikoma Y, Matsumoto H, Sugiura M, Hyodo H, Yano M (2004) Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. Plant Physiol 134:824–837

Kulheim C, Agren J, Jansson S (2002) Rapid regulation of light harvesting and plant fitness in the field. Science 308:267–269

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res 12:1048–1059

Li L, Paolillo DJ, Parthasarathy MV, DiMuzio EM, Garvin DF (2001) A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). Plant J 26:59–67

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J, Zhang D (2006) Genome-wide analysis of basic/helix–loop–helix transcription factor family in rice and *Arabidopsis*. Plant Physiol 141:1167–1184

Li F, Vallabhaneni R, Wurtzel ET (2008a) PSY3, a new member of the phytoene synthase gene family conserved in the Poaceae and regulator of abiotic stress-induced carotenogenesis. Plant Physiol 146:1333–1345

Li F, Vallabhaneni R, Yu J, Rocheford T, Wurtzel ET (2008b) The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis and thermal stress tolerance. Plant Physiol 147:1334–1346

Lokstein H, Tian L, Polle JEW, Dellapenna D (2002) Xanthophyll biosynthetic mutants of *Arabidopsis thaliana*: altered nonphotochemical quenching of chlorophyll fluorescence is due to changes in photosystem II antenna size and stability. Biochim Biophys Acta 1553:309–319

Lopez AB, Van Eck J, Conlin BJ, Paolillo DJ, O'Neill J, Li L (2008) Effect of the cauliflower o*r* transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. J Exp Bot 59:213–223

López-Ráez JA, Charnikhova T, Gómez-Roldán V, Matusova R, Kohlen W, De Vos R, Verstappen F, Puech-Pages V, Bécard G, Mulder P, Bouwmeester H (2008) Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. New Phytol 178:863–874

Lu S, Li L (2008) Carotenoid metabolism: biosynthesis, regulation and beyond. J Integr Plant Biol 50:778–785

McNulty HP, Byun J, Lockwood SF, Jacob RF, Mason RP (2007) Differential effects of carotenoids on lipid peroxidation due to membrane interactions: X-ray diffraction analysis. BBA Biomembranes 1768:167–174

McSteen P, Leyser O (2005) Shoot branching. Annu Rev Plant Biol 56:353–374

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771–1780

Nakano M, Nobuta K, Vemaraju K, Tej SS, Skogen JW, Meyers BC (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34:D731–D735

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56:165–185

Niyogi KK (1999) Photoprotection revisited: genetic and molecular approaches. Annu Rev Plant Biol 56:165–185

Nobuta K, Venu RC, Lu C, Beló A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ, Wang G, Meyers BC (2007) An expression atlas of rice mRNAs and small RNAs. Nat Biotechnol 25:473–477

Oudes AJ, Roach JC, Walashek LS, Eichner LJ, True LD, Vessella RL, Liu AY (2005) Application of Affymetrix array and massively parallel signature sequencing for identification of genes involved in prostate cancer progression. BMC Cancer 5:86

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) A new version of golden rice with increased pro-vitamin A content. Nat Biotechnol 23:482–487

Park H, Kreunen SS, Cuttriss AJ, Dellapenna D, Pogson BJ (2002) Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. Plant Cell 14:321–332

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pogson BJ, Niyogi KK, Bjorkman O, Dellapenna D (1998) Altered xanthophylls compositions adversely affect chlorophyll accumulation and nonphotochemical quenching in *Arabidopsis* mutants. Proc Natl Acad Sci USA 95:13324–13329

Quinlan RF, Jaradat TT, Wurtzel ET (2007) *Escherichia coli* as a platform for functional expression of plant P450 carotene hydroxylases. Arch Biochem Biophys 458:146–157

Raes J, Rohde A, Christensen JH, van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiol 133:1051–1071

Rivier L, Léonard JF, Cottier JP (1983) Rapid effect of osmotic stress on the content and exodiffusion of abscisic acid in *Zea mays* roots. Plant Sci Lett 83:133–137

Rossel JB, Wilson IW, Pogson BJ (2002) Global changes in gene expression in response to high light in *Arabidopsis*. Plant Physiol 130:1109–1120

Rouze P, Pavy N, Rombauts S (1999) Genome annotation: which tools do we have for it? Curr Opin Plant Biol 2:90–95

Segerman B, Jansson S, Karlsson J (2007) Characterization of genes with narrow expression patterns in *Populus*. Tree Genet Genomes 3:351–362

Shan H, Zhang N, Liu C, Xu G, Zhang J, Chen Z et al (2007) Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes. Mol Phylogenet Evol 44:26–41

Siefermann-Harms D, Joyard J, Douce R (1978) Light-induced changes of the carotenoid levels in chloroplast envelopes. Plant Physiol 61:530–533

Simkin AJ, Schwartz SH, Auldridge M, Taylor MG, Jee HJ (2004a) The tomato carotenoid cleavage dioxygenase I gene contributes to the formation of the flavor volatiles beta-ionone, pseudo ionone, and geranylacetone. Plant J 40:882–892

Simkin AJ, Underwood BA, Auldridge M, Loucas HM, Shibuya K, Schmelz E et al (2004b) Circadian regulation of the *PhCCD1* carotenoid cleavage dioxygenase controls emission of β-ionone, a fragrance volatile of petunia flowers. Plant Physiol 136:3504–3514

Singh G, Kumar S, Singh P (2003) A quick method to isolate RNA from wheat and other carbohydrate-rich seeds. PMB Report 21:93a-f

Sun ZR, Gantt E, Cunningham FX (1996) Cloning and functional analysis of the beta-carotene hydroxylase of *Arabidopsis thaliana*. J Biol Chem 271:24349–24352

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTALX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Thompson AJ, Jackson AC, Parker RA, Morpeth DR, Burbidge A, Taylor IB (2000a) Abscisic acid biosynthesis in tomato: regulation of zeaxanthin epoxidase and 9-*cis*-epoxycarotenoid dioxygenase mRNAs by light/dark cycles, water stress and abscisic acid. Plant Mol Biol 42:833–845

Thompson AJ, Jackson AC, Symonds RC, Mulholland BJ, Dadswell AR, Blake PS, Burbidge A, Taylor IB (2000b) Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes overproduction of abscisic acid. Plant J 23:363–374

Thompson AJ, Mulholland BJ, Jackson AC, McKee JM, Hilton HW, Symonds RC, Sonneveld T, Burbidge A, Stevenson P, Taylor IB (2007) Regulation and manipulation of ABA biosynthesis in roots. Plant Cell Environ 30:67–78

Tian L, DellaPenna D (2001) Characterization of a second *carotenoid β-hydroxylase* gene from *Arabidopsis* and its relationship to the *LUT1* locus. Plant Mol Biol 47:379–388

Tian L, Musetti V, Kim J, Magallanes-Lundback M, Dellapenna D (2004) The *Arabidopsis LUT1* locus encodes a member of the cytochrome P450 family that is required for carotenoid epsilon-ring hydroxylation activity. Proc Natl Acad Sci USA 101:402–407

Tuskan GA et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Tuteja N (2007) Abscisic acid and abiotic stress signaling. Plant Signal Behav 2:135–138

van Norman, Jaimie M, Sieburth, Leslie E (2007) Dissecting the biosynthetic pathway for the *bypass1* root-derived signal. Plant J 49:619–628

Vishnevetsky M, Ovadis M, Zuker A, Vainstein A (1999) Molecular mechanisms underlying carotenogenesis in the chromoplast: multilevel regulation of carotenoid-associated genes. Plant J 20:423–431

von Lintig J, Welsch R, Bonk M, Giuliano G, Batschauer A, Kleinig H (1997) Light-stimulated carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. Plant J 12:625–634

Welsch R, Beyer P, Hugueney P, Kleinig H, von Lintig J (2000) Regulation and activation of phytoene synthase, a key enzyme in carotenoid biosynthesis, during photomorphogenesis. Planta 211:846–854

Welsch R, Wüst F, Bär C, Al-Babili S, Beyer P (2008) A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. Plant Physiol 147:367–380

Woitsch S, Römer S (2003) Expression of xanthophyll biosynthetic genes during light-dependent chloroplast differentiation. Plant Physiol 132:1508–1517

Ye XD, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287:303–305