Rice Research: Open Access
ISSN: 2375-4338

Umate, J Rice Res 2015, 4:1
DOI: 10.4172/2375-4338.1000160

Research Article Open Access

Comparison of Genes Encoding Enzymes of Sterol Biosynthesis from Plants to Orthologs in Yeast

Pavan Umate*

Sai Chakra, Beside R.B.V.R.R. College, Narayanguda, Hyderabad-500027, Telangana State (T.S.), India

Abstract

Sterols are ubiquitous membrane lipid components found in eukaryotic organisms. Sterols are known to serve novel functions in eukaryotes. Sterol biosynthesis pathways differ among fungi, plants and vertebrates. Lipid biosynthesis has been extensively studied in the model yeast *Saccharomyces cerevisiae*. In this study here, by the aid of computational approach, yeast orthologs of sterol pathway genes and their genomic copy number alterations in the model magnoliophytes (*Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*), in a lycophyte (*Selaginella moellendorffii*) and a bryophyte (*Physcomitrella patens*) are identified. The study demonstrates that the basic set of sterol enzymes harbored by these organisms is well conserved. Of particular interest is the wide variation in genomic copy number of sterol pathway genes. It is puzzling to identify two orthologs each of Δ8-Δ7 sterol isomerase, Δ7 sterol C-5 desaturase, sterol Δ24-reductase and sterol Δ7-reductase in the woody angiosperm *P. trichocarpa*. The identification of a surprisingly high number of genes coding for cycloartenol synthase (10 no.), sterol C-4 methyl oxidase (9 no.) and cytochrome P450 51 (7 no.) in *O. sativa* is intriguing. In addition, the report has identified two genes each coding for C-14 reductase in *O. sativa* and Δ24-reductase in *P. patens*. This analysis has brought new insights in sterol pathway genes in model plants.

Keywords: Lipids; Orthologs; Plants; Protein alignment; Yeast

Abbreviations: 3βHSD/D: 3β-hydroxysteroid-dehydrogenase/C4-decarboxylase; BRs: brassinosteroids; CAS: cycloartenol synthase; CPI: cyclopropyl sterol isomerase; cyt: cytochrome; CYP51: cytochrome P450 51; CYP710A: cytochrome P450 710A; erg: ergosterol; FAD: flavin adenine dinucleotide; GI: genbank Ids; HYD: HYDRA; JGI: Joint genome institute; LASI: lanosterol synthase; NCBI: National Center for Biotechnology Information; RGAP: Rice Genome Annotation Project; SQE: squalene epoxidase; SQS: squalene synthase; STE: sterol; SMO: sterol methyl oxidase; SMT: sterol methyltransferase; ST14R: sterol Δ14 reductase; ST24R: sterol Δ24 reductase; ST7R: sterol Δ7-reductase; TAIR: The Arabidopsis Information Resource; 8,7 SI: Δ8-Δ7 sterol isomerase

Introduction

A large family of lipid triterpenoids (a class of isoprenoids) called hopanoids and sterols are found in bacteria and eukaryotes, respectively. Synthesized via the isoprenoid biosynthesis pathway, sterols are essential components of all eukaryotic cell membranes. Sterols possess different functions and they contribute to cellular physiology in eukaryotes. They are known to maintain structural integrity of most eukaryotic cells. Sterols interact with phospholipids and regulate membrane fluidity and permeability. A number of steroid hormones like testosterone and estrogen in mammals, brassinosteroids (BRs) in plants, and ecdysteroids in arthropods are synthesized from sterols as precursor. Sterols are known to be involved in cell signaling, in transport and distribution of lipophilic molecules, and in formation of lipid rafts [1-7].

Specific sterols are produced in different organisms. The predominant sterols found in fungi and vertebrates are ergosterol and cholesterol, respectively. A variety of sterols are synthesized in plants, mainly represented by sitosterol, campesterol and stigmasterol. The sterol biosynthesis pathway has been well elucidated in fungi, higher plants and animals [1]. Sterols in higher plants and fungi are synthesized by different pathways. Identical reaction steps of conversion from isopentenyl PP (IPP) to squalene epoxide (SQE) are found in these pathways. The two pathways producing either lanosterol or cycloartenol are diverged at the step of SQE cyclization (Figure 1) [8]. Lanosterol, the first tetracyclic intermediate in animals and fungi is converted to cholesterol in vertebrates and to ergosterol in fungi; and cycloartenol, the plant-specific first tetracyclic intermediate is converted to campesterol, sitosterol and stigmasterol in land plants. These conversions proceed through a series of oxidation, reduction and demethylation reactions. The product of cyclization of SQE thus follows different routes in photosynthetic and nonphotosynthetic organisms. Accordingly, to open the cyclopropane ring of cycloartenol, the photosynthetic organisms require different set of enzymes, such as cycloartenol synthase and cycloeucalenol-obtusifoliol isomerase. Nevertheless, the sterol biosynthesis in plants, yeast and animals, share most of the enzymatic steps [9,10].

In plants, sterol biosynthesis pathway is comprised of two branches. Major membrane sterols being produced are stigmasterol and sitosterol forming one branch, and BR synthesis represents the other branch. Sterol biosynthesis has been extensively studied in yeast. Functions of some animal and plant sterol enzymes were confirmed via functional complementation of yeast mutants [11,12].

Genes involved in plant sterol (phytosterol) biosynthesis were identified and cloned based on heterologous expression or sequence similarity [9,13]. However, ergosterol-deficient mutants and complementation assays were implemented to identify genes involved in yeast ergosterol biosynthesis. Usually, the selection of nystatin-

***Corresponding author:** Pavan Umate, Beside R.B.V.R.R. College, Narayanguda, Hyderabad, Telangana State (T.S.), India, Tel: 7207805184; E-mail: pavan_umate@rediffmail.com

Received November 21, 2015; **Accepted** December 23, 2015; **Published** December 28, 2015

Citation: Umate P (2015) Comparison of Genes Encoding Enzymes of Sterol Biosynthesis from Plants to Orthologs in Yeast. J Rice Res 4: 160. doi:10.4172/2375-4338.1000160

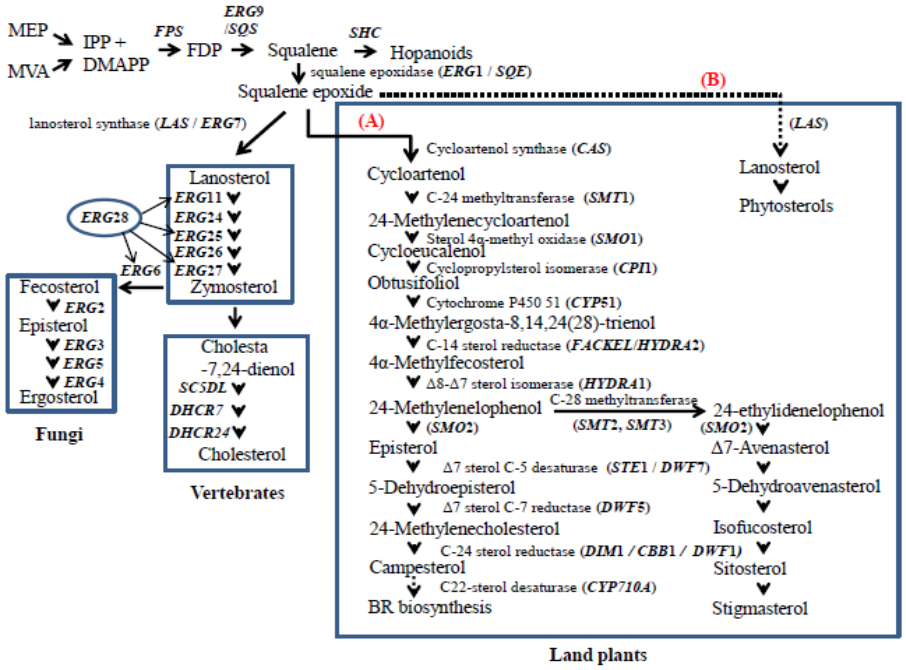

Figure 1: Scheme showing sterol biosynthesis pathways in diverse organisms. Distinct enzymes of the pathway are given beside arrows and gene names are indicated in bold and italics. The major cycloartenol (A) and the minor lanosterol (B) routes of phytosterol biosynthesis are shown.

resistant yeast mutants was one approach to identify such genes [14-16].

The enzymes involved in de novo synthesis of sterols in fungi, plants, and vertebrates have been identified and well characterized [6]. Sterol biosynthesis involves the following necessary steps: 1. epoxidation of squalene followed by cyclization (squalene monooxygenation, oxidosqualene cyclization); 2. loss of a methyl group at C-14 position (C-14 demethylation, C-14 reduction); 3. loss of two methyl groups at C-4 position (C-4 methyl oxidation, C-3 dehydrogenation/C-4 decarboxylation, C-3 ketoreduction); 4. reduction of Δ-8 double bond (Δ-8, Δ-7 isomerisation); 5. formation of double bond between C-5 and C-6 (C-5 desaturation); 6. addition of methyl groups (C-24 or C-28 methylation); 7. removal of C7-8 and C24-25 double bonds (Δ-7 and Δ-24 reduction); 8. formation of double bond between C-22 and C-23 (C-22 desaturation) [1].

The study of genes and their copy number variation in plant-specific metabolic pathways is not well understood. Moreover, the distribution of genes involved in sterol biosynthesis in *Oryza sativa*, *Populus trichocarpa*, *Physcomitrella patens* and *Selaginella moellendorffii* has not yet been studied. Complete genome sequences for representatives of model angiosperms (*Arabidopsis thaliana*, *P. trichocarpa*, *O. sativa*,), a bryophyte (*P. patens*) and a lycophyte (*S. moellendorffii*) are now available. An elaborate analysis of genes encoding sterol enzymes is now feasible among organisms representing plant diversity. Since the genes involved in ergosterol biosynthesis were elaborately studied [17,18], sequence comparisons should uncover orthologs of yeast ERGs in other organisms. The current study was aimed to identify the complete set of sterol genes and their copy number in the model angiosperms, a bryophyte and a lycophyte. The occurrence of potential orthologs of sterol enzymes in the genomes of model eukaryotes is described. This analysis might allow functional approaches to candidate genes for sterol enzymes in higher plants and other eukaryotic phyla. Here we report a systematic analysis of genes and their genomic copy number alterations in eukaryotic-specific sterol biosynthesis pathway.

Materials and Methods

Orthologs of sterol enzymes in angiosperms (*Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa*) and a bryophyte (*Physcomitrella patens*) were identified with protein sequences of *Saccharomyces cerevisiae* (yeast) using blastp (protein-protein blast) of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/mapview) database. Sequences for a lycophyte (*Selaginella moellendorffii*) were not available on the NCBI, therefore the orthologs of sterol enzymes in *S. moellendorffii* were identified by the aid of the Department of Energy Joint Genome Institute (JGI) database (Table 1).

Initial blastp searches, with *S. cerevisiae* full-length protein sequences (ERG1-7, ERG9, ERG11, ERG24-28) as query, were made against RefSeq database of the NCBI, and Genbank Ids (gi) and protein sequences were retrieved. To identify the locus and copy number of a gene, these protein sequences were blasted (blastp search) against the *Arabidopsis* information resource (TAIR) (http://www.arabidopsis.org) for *A. thaliana*, rice genome annotation project (RGAP) (http://rice.plantbiology.msu.edu) for *O. sativa* subsp. Japonica (rice) and the JGI database (http://www.jgi.doe.gov) for *P. trichocarpa*, *P. patens* and *S. moellendorffii*. Blastp search was made against specific datasets, TAIR10 proteins of *A. thaliana*, genes in MSU RGAP release 7 – protein sequences of *O. sativa*, *P. trichocarpa* Jamboree gene models (proteins) of *P. trichocarpa*, *P. patens* v1.1 filteredmodels3 (proteins) of *P. patens*, and *S. moellendorffii* v1.0 non-redundant filtered model proteins (Selmolmodels_filteredmodels3_aa) of *S. moellendorffii*. The best hit in each organism is listed in Table 2, and other probable hits in *A. thaliana*, *P. trichocarpa*, *O. sativa*, *P. patens* and *S. moellendorffii* are also enlisted (Table 2) (Supplemental Tables 1-5).

Retrieved protein sequences were aligned using ClustalX [19].

S. cerevisiae	*A. thaliana*	% I/P=aa**	*P. trichocarpa*	% I/P=aa**	*O. Sativa*	% I/P=aa**	*P. patens*	% I/P=aa**
gi|1323310 (ERG1)	gi|18406296	37/52=453	gi|222853783	38/51=431	gi|115451723	37/50=464	gi|168058650	38/53=443
gi|6323858 (ERG2)	gi|79336241	24/42=135	No hit	N/A	No hit	N/A	No hit	N/A
gi|6323085 (ERG3)	gi|15232935	31/49=235	gi|222848137	32/52=262	gi|115434372	32/54=194	gi|168035529	29/50=232
gi|6321426 (ERG4)	gi|18409727	28/48=376	gi|222833435	28/47=388	gi|115446059	28/43=384	gi|168004207	28/46=360
gi|6323657 (ERG5)	gi|15226758	28/49=500	gi|222848961	27/47=485	gi|115435230	27/49=497	gi|168007574	30/49=494
gi|6323635 (ERG6)	gi|15240691	53/68=311	gi|222849723	50/66=347	gi|115471111	52/67=310	gi|168067590	49/65=311
gi|330443583 (ERG7)	gi|15225650	40/56=694	gi|222853808	40/55=749	gi|115444137	37/53=757	gi|168011729	40/58=662
gi|6321984 (ERG9)	gi|15236168	43/60 = 369	gi|222850173	41/59=382	gi|115456049	45/61=323	gi|168046876	42/59=381
gi|6321795 (ERG11)	gi|15221075	32/49=479	gi|222846614	31/48=515	gi|115485695	32/48=452	gi|168033822	31/50=478
gi|6324049 (ERG24)	gi|18409727	36/55=372	gi|222833435	36/55=358	gi|115446059	31/48=371	gi|168004207	34/54=359
gi|398365435 (ERG25)	gi|18390767	37/56=278	gi|222845595	36/57=279	gi|115470159	36/55=267	gi|168010783	39/60=248
gi|6321437 (ERG26)	gi|18401656	37/53=354	gi|222854485	36/54=353	gi|115453453	34/53=348	gi|168051790	34/53=364
gi|6323129 (ERG27)	No hit	N/A	No hit	N/A	No hit	N/A	No hit	N/A
gi|6320883 (ERG28)	gi|18391101	32/51=85	gi|222842858	33/54=82	gi|115489726	33/51=93	gi|168006376	25/47=126

**I/P, Identities/Positives; aa = amino acids compared

Table 1: Genbank (gi) Ids and per cent (%) homology for orthologs of yeast genes of the ergosterol (ERG) pathway in model organisms.

S. No.	*A . thaliana* $	*P. trichocarpa* #	*O. Sativa***	*P. patens* #	*S. moellendorffii* #
1	AT1G58440 (Squalene epoxidase 1) (SQE1)	Poptr1_1|832433	Os03g12900	Phypa1_1|224792	Selmo1|150266
2	AT1G05440 (C-8 sterol isomerases)	No hit	No hit	No hit	No hit
	AT1G20050 (C-8 sterol isomerase) (HYD1/MAD4)	Poptr1_1|554665*	Os01g01369*	Phypa1_1|234885*	Selmo1|89288*
3	AT3G02580 (Δ 7-sterol-C-5-desaturase) (BUL1/DWF7/STE1)	Poptr1_1|831136	Os01g04260	Phypa1_1|135889	Selmo1|146325
4	AT3G19820 (Sterol C-24 reductase) (DWF1/DIM1/CBB1)	Poptr1_1|765722*	Os10g25780*	Phypa1_1|234317*	Selmo1|139528*
5	AT2G34500 (C-22 sterol desaturase) (CYP710A1)	Poptr1_1|800680	Os01g11300	Phypa1_1|204792	Selmo1|75218
6	AT5G13710 (Sterol 1) (CPH/SMT1)	Poptr1_1|202903	Os07g10600	Phypa1_1|228167	Selmo1|89663
7	AT2G07050 (Cycloartenol synthase 1) (CAS1)	Poptr1_1|832441	Os02g04710	Phypa1_1|206381	Selmo1|266790
	AT3G45130 (Lanosterol synthase 1) (LAS1)	Poptr1_1|417048*	Os02g04710*	Phypa1_1|206381*	Selmo1|269539*
8	AT4G34640 (Squalene synthase 1) (ERG9/SQS1)	Poptr1_1|833037	Os03g59040	Phypa1_1|59853	Selmo1|96714
9	AT1G11680 (Obtusifoliol 14-α demethylase) (CYP51/EMB1738)	Poptr1_1|750989	Os11g32240	Phypa1_1|134753	Selmo1|178739
10	AT3G52940 (Sterol C-14 reductase) (ELL1/FK/HYD2)	Poptr1_1|826487	Os02g26650	Phypa1_1|67627	Selmo1|99117
11	AT1G07420 (C-4 sterol methyl oxidase) (ATSMO1/SMO2-1)	Poptr1_1|752102	Os07g01150	Phypa1_1|205949	Selmo1|173304
12	AT1G47290 (3β-HSD/C4-decarboxylase) (AT3βHSD)	Poptr1_1|415947	Os03g29170	Phypa1_1|145942	Selmo1|84229
13	AT1G10030 (Homolog of yeast ergosterol28) (ERG28)	Poptr1_1|830270	Os12g43670	Phypa1_1|109549	Selmo1|94581

* Identified using *A. thaliana* protein sequence as query. $TAIR Ids; **RGAP Ids; #JGI Ids

Table 2: Locus Ids of sterol pathway genes in model organisms.

Gaps in amino acid sequences were introduced to improve the alignment. Multiple parameters of gap opening 10, gap extension 0.2, delay divergent sequences 30%, Gonnet series protein weight matrix and gap separation distance of 4 were set as alignment parameters. All alignments were screened manually to identify conserved motifs. The following abbreviations are used for building all alignments, AT, *Arabidopsis thaliana*; Poptr, *Populus trichocarpa*; Os, *Oryza sativa*; Phypa, *Physcomitrella patens*; Selmol, *Selaginella moellendorffii*; and yeast, *Saccharomyces cerevisiae*; ERG, ergosterol.

Results

Genbank Ids (gi) for orthologs of yeast ERGs (ERG1-7, ERG9, ERG11, ERG24-28) in *A. thaliana*, *P. trichocarpa*, *O. sativa* and *P. patens* are given in Table 1. Four protein families for these genes are identified: (i) the cytochrome (cyt) P450 family with C-22 sterol desaturases (ERG5) and C-14 sterol demethylases (ERG11); (ii) the cytb5-dependent fatty acid hydroxylase superfamily of C-4 sterol methyl oxidases (SMOs) (ERG25) and C-5 sterol desaturases (ERG3); (iii) the highly hydrophobic reductases, which include Δ7, Δ14, and Δ24 sterol reductases; (iv) the S-adenosyl-L-methionine sterol methyltransferase (SMT) family, composed of plant-specific SMT1 and SMT2 types, and fungal C-24 sterol methyltransferase (ERG6). Genes of squalene epoxidase (SQE / ERG1), Δ8-Δ7 sterol isomerase (ERG2), lanosterol synthase (ERG7), squalene synthase (SQS/ERG9), 3β-hydroxysteroid-dehydrogenase/C4-decarboxylase (3βHSD/D) (ERG26), C-4 demethylation (ERG28), and cyclopropyl sterol isomerase (CPI) that do not fall under the above mentioned categories are also identified.

Squalene epoxidase (SQE or SQP) or ERG1

Six genes were identified to encode squalene epoxidase (SQE) in *A. thaliana* genome [20]. These are *SQE*1 (AT1G58440), *SQE*2 (AT2G22830), *SQE*3 (AT4G37760), *SQE*4 (AT5G24140), *SQE*5 (AT5G24150), and *SQE*6 (AT5G24160). Two *SQP* genes in *O. sativa* [*SQP*1 (Os03g12900) and *SQP*2 (Os03g12910)], three in *P. trichocarpa* [*SQP*1 (Poptr.832433), *SQP*2 (Poptr.788926) and *SQP*3 (Poptr.831444)], and one each in *P. patens* (Phypa.224792) and *S. moellendorffii* (Selmo1.150266) (Table 3) are identified.

There is 44-85% protein identity (63-93% similarity) in any pairwise comparison among *A. thaliana* SQEs, 87% identity (92% similarity) between *O. sativa* SQE isoforms, 79-82% identity (88-92% similarity) among *P. trichocarpa* SQE isoforms. Overall, there is only 27-37% amino acid identity (47-52% similarity) in any pairwise comparison between yeast and *A. thaliana* SQEs. Similarly, yeast SQE protein showed a 36-37% identity (50% imilarity) with *O. sativa* SQEs,

Steps	*A. thaliana*	*P. trichocarpa*	*O. sativa*	*P. patens*	*S. moellendorffii*
Squalene monooxygenation	SQE (6)	SQP (3)	SQP (2)	SQP (1)	SQP (1)
Δ-8, Δ-7 isomerization	HYD1 (1)	8,7 SI (2)	8,7 SI (1)	8,7 SI (1)	8,7 SI (1)
C-5 desaturation	STE1 (1)	STE1 (2)	STE1 (1)	STE1 (1)	STE1 (1)
Δ-24 reduction	DWF1 (1)	ST24R (2)*	ST24R (1)*	ST24R (2)*	ST24R (1)*
C-22 desaturation	CYP710A (4)	CYP710A (1)	CYP710A (4)	CYP710A (1)	CYP710A (1)
C-24 or C-28 methylation	SMT (3)	SMT (2)	SMT (3)	SMT (1)	SMT (1)
Oxydosqualene cyclization	CAS1 (1)	CAS (1)	CAS (10)	CAS (1)	CAS (1)
C-14 demethylation	CYP51 (1)	CYP51 (2)	CYP51 (7)	CYP51 (1)	CYP51 (1)
C-14 reduction	FK (1)	ST14R (1)	ST14R (2)	ST14R (1)	ST14R (1)
C-4 methyl oxidation	SMO (5)	SMO (2)	SMO (9)	SMO (2)	SMO (2)
C-3 dehydrogenation/C-4 decarboxylation	HSD (3)	HSD (1)	HSD (2)	HSD (1)	HSD (1)
C-3 ketoreduction	Not found	Not found	Not found	Not found	Not found
Endoplasmic reticulum anchoring	ERG28 (1)	ERG28 (1)	EBP28 (1)	ERG28 (1)	ERG28 (1)
Δ-7 reduction	DWF5 (1)	ST7R (2)*	ST7R (1)*	ST7R (1)*	ST7R (1)*
Cyclopropyl sterol isomerization	CPI (1)	CPI (1)*	CPI (1)*	CPI (1)*	CPI (1)*

* Identified using A. thaliana protein sequence as query

Table 3: Steps of sterol biosynthesis pathway and their corresponding genes in model organisms. Numbers in parenthesis are the copy number of a particular gene.

35-38% identity (50-51% similarity) with *P. trichocarpa*, 38% identity (52% similarity) with *P. patens,* and 37% identity (50% similarity) with *S. moellendorffii* SQE. A conserved flavin adenine dinucleotide (FAD) binding domain in SQE proteins of yeast and plants is identified [20] (Supplemental Figure 1).

Δ8-Δ7 sterol isomerase (*HYDRA*1) or ERG2

8,7 sterol isomerase (8,7 SI) is a single copy gene in yeast, *O. sativa*, *P. patens* and *S. moellendorffii*. The TAIR database search has identified two genomic loci (AT1G05440, AT1G20050) to encode 8, 7 SIs in *A. thaliana*. However, no significant similarity was found in their sequences. No probable hit was obtained in plant species (except *A. thaliana*) with yeast *ERG*2 as query sequence. Only a few amino acids of 8, 7 SI are identical across species of yeast and plants. Therefore, all plant 8, 7 SIs were identified with *A. thaliana HYD*1 as query sequence. *A. thaliana* locus AT1G20050 (annotated as *HYD*1 in the TAIR database) showed 48%-66% amino acid identity with other plant species. However, their sequence similarity with the locus AT1G05440 is poor. Thus, 8,7 SI is a single copy gene in *A. thaliana*. There are two gene copies of 8, 7 SIs in *P. trichocarpa* with 86% protein identity. Conserved amino acid residues (W, H, E, D, E and T) of 8, 7 SIs are identified in plants. Remarkably, these critical amino acids are highly conserved in plants but not in yeast 8, 7 SI (Figure 2) [21].

Δ7 sterol C-5 desaturase (*STE*1 or *DWF7*) or ERG3

Genomes of *A. thaliana*, *O. sativa*, *P. patens* and *S. moellendorffii* encode a single Δ7 sterol C-5 desaturase. Genome of *P. trichocarpa* encodes two such genes (*STE*1 and *STE*2). A blast search of *A. thaliana STE*1 showed 69% identity and 83% similarity for 272 amino acids with *O. sativa* locus Os01g04260.

Yeast ERG3 is 29-32% identical to plant Δ7 sterol C-5 desaturases. Whereas the *A. thaliana* STE1 showed high sequence identity with other plants which is 69% identical to STE1 of *O. sativa*, 79% to STE1 and 77% to STE2 of *P. trichocarpa*, 68% to STE1 of *P. patens* and 65% to STE1 of *S. moellendorffii*. The STE1 and STE2 of *P. trichocarpa* are highly conserved with 90% identity. *O. sativa* locus Os01g04260 is a fatty acid hydroxylase (FAH) in the RGAP database. However, its amino acid sequence showed a high 69% identity with *A. thaliana* STE1 and only a low 26-29% identity with *A. thaliana* SMOs. All FAHs are renamed here as sterol methyl oxidases (SMOs) (see below). Sequence identity of *A. thaliana* STE1 and *O. sativa* SMO1-9 is also poor. Thus *O. sativa* locus Os01g04260 is a Δ7 sterol C-5 desaturase. Histidine clusters were well conserved in STE1 and ERG3 [9, 12]. Conserved histidine motifs, HX3H, HX2HH, and HX3H/D are identified in STE1 proteins of *A. thaliana*, *O. sativa*, *P. trichocarpa*, *P. patens*, *S. moellendorffii* and yeast ERG3 (Figure 3).

Cytochrome P450 710A (CYP710A) or ERG5

Four genes, *CYP*710*A*1, *CYP*710*A*2, *CYP*710*A*3, and *CYP*710*A*4, which encode members of the CYP710A subfamily in *A. thaliana* were found [22]. Four genes of *CYP*710A in *O. sativa* , and one each in *P. trichocarpa*, *P. patens* and *S. moellendorffii* are identified (Supplemental Tables 1-5).

Protein sequence comparisons between plant CYP710As and fungal ERG5 revealed an identity in the range of 25-30%. *A. thaliana* CYP710A1 and CYP710A2 share a high 82% protein identity and CYP710A3 and CYP710A4 are 94% identical. Sequence similarity of *A. thaliana* CYP710A1, CYP710A2 with each of CYP710A3 and CYP710A4 is in the range of 74-78%. Any pairwise sequence comparison of *A. thaliana* CYP710A isoforms with other plant CYP710As showed high identity in the range of 51-68%. High level sequence conservation in the range of 79-86% can be seen among *O. sativa* CYP710A isoforms. The characteristic domain F(L/M)FA(A/S) QDAS(T/S)S, and conserved Cys residue in the C-terminal part of plant and fungal CYP710A proteins is identified (Supplemental Figure 2) [22,23].

Sterol methyltransferase (SMT) or ERG6

There are three SMT genes, *SMT*1 (AT5G13710), *SMT*2 (AT1G20330) and *SMT*3 (AT1G76090) in *A. thaliana* [24]. Three *SMT*s in *O. sativa* (*SMT1* - Os03g04340; *SMT*2 - Os03g59290 and *SMT*3 - Os07g10600), two in *P. trichocarpa* (*SMT*1 - Poptr.202903; *SMT*2 - Poptr.829664), and one each in *P. patens* (*SMT*1 - Phypa.228167) and *S. moellendorffii* (*SMT*1 - Selmo1.89663) are identified (Table 3, Supplemental Tables 1-5).

Protein sequence of *A. thaliana* SMT1 (with 53% identity), *O. sativa* SMT3 (with 52% identity) and *P. trichocarpa* SMT1 (with 50% identity) are most similar to yeast ERG6. A high amino acid sequence identity of 83% and 73% is found between SMT2 and SMT3 in *A. thaliana* and *O. sativa*, respectively. Protein sequence of SMT1 is 39-41%

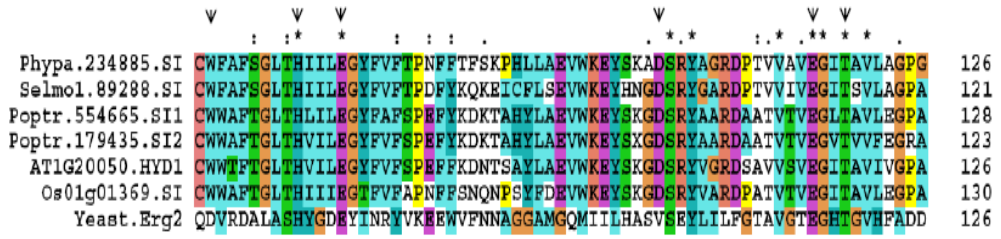

Figure 2: Protein sequence alignment of sterol Δ8-Δ7 isomerases. Conserved amino acids are indicated on top of alignment with inverted arrows [21]. In all figures showing alignments, gene Ids and protein names are identified on left and numbers of amino acids are given on right. The asterisks denote identities and the colons and periods denote similarities. The colon (double dot) denotes conservative and the period (single dot) denotes semi-conservative amino acid changes. SI, sterol Δ8-Δ7 isomerase; HYD1, HYDRA1.

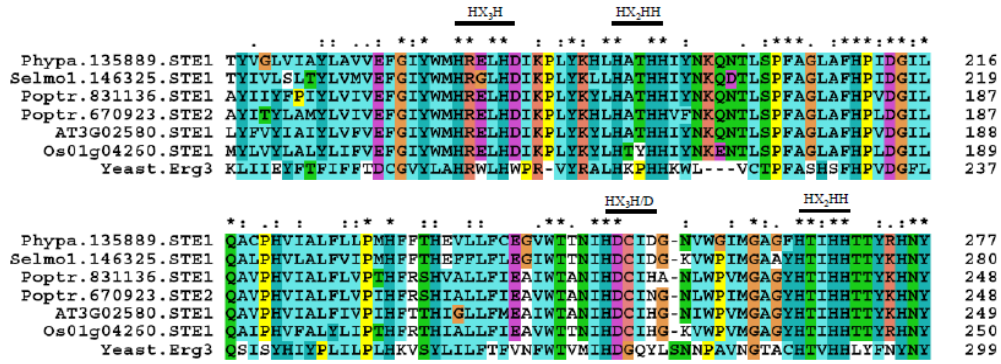

Figure 3: Aligned protein sequences of Δ7 sterol C-5 desaturases. Conserved histidine motifs (HX3H, HX2HH, and HX3H/D) are shown on top of alignment. STE, STEROL.

identical to SMT2 and SMT3 of *A. thaliana* and *O. sativa*. Proteins SMT1 and SMT2 of *P. trichocarpa* showed high sequence conservation (93% identity).

Five protein motifs in SMT sequences are identified. Three motifs SAM I [(LD(V/A)GCG(I/V)GGP], SAM II (NSFDAVYA), and SAM III (V(L/M)KPGQCFAAY), identified in most methyltransferase [25, 26] are also present in SMT protein sequences of *A. thaliana*, *O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii* (Supplemental Figure 3). The first motif is well conserved, second and third motifs are less conserved. Conserved domains SMT I (YE(F/Y/W)GWGXS(F/Y)HF) and SMT II (IEA(T/S)CHAP) are also identified in these sequences.

Lanosterol synthase (LAS) and Cycloartenol synthase (CAS) or ERG7

A. thaliana genome encodes a single *CAS*1 gene (AT2G07050). Interestingly, a preliminary search of *O. sativa* genome has identified thirteen *CAS* genes. A single copy of *CAS* gene is found in the genomes of *P. trichocarpa* (Poptr.832441), *P. patens* (Phypa.206381) and *S. moellendorffii* (Selmo1.266790). There are also three *LAS* genes (Poptr.717351, Poptr.417048 and Poptr.786184) in *P. trichocarpa* and one in *S. moellendorffii* (Selmo1.269539). Any *LAS* gene could not be identified in the genomes of *O. sativa* and *P. patens*.

Protein sequences of *A. thaliana* LAS and CAS are well conserved with 65% identity and 79% similarity. Any pairwise sequence comparison among *O. sativa* CAS isoforms showed %identity in the range of 44-91% and %similarity in the range of 60-94%. A similar pairwise comparison of *P. trichocarpa* CAS and isoforms of LAS showed %identity in the range of 70-77% and %similarity in the range of 82-87%. Any pairwise sequence comparison among LAS isoforms showed high level protein identity (66-87%) and similarity (79-91%) in *P. trichocarpa*. Amino acid sequences of CAS and LAS are highly conserved with 98% identity and 98% similarity in *S. moellendorffii*.

Any pairwise sequence comparison between yeast ERG7 and plants CAS or LAS showed %identity in the range of 33-41% and %similarity in the range of 52-59%. Blastp search with *A. thaliana* LAS1 as query has always identified *CAS* genes in *O. sativa* and *P. patens*. However, a similar search has identified *LAS* genes in *P. trichocarpa* and *S. moellendorffii*.

A conserved aspartate (D) residue in the active site region of yeast ERG7 and plant LAS and CAS sequences is identified [1]. Only ten of thirteen CAS sequences of *O. sativa* contain a conserved 'D' residue, which however was not found in sequences of loci Os02g04760, Os11g18310 and Os11g18340. Thus, ten copies of *CAS* were validated in *O. sativa* genome (Figure 4).

Squalene synthase (SQS) or ERG9

Genomes of *A. thaliana* (loci AT4G34640 and AT4G34650) and *O. sativa* (loci Os03g59040 and Os07g10130) encode two genes of *SQS*. Two *SQS* genes in the genome of *P. trichocarpa* (Poptr.833037, Poptr.818128), and one each in *P. patens* (Phypa.59853) and *S. moellendorffii* (Selmo1.96714) are identified (Supplemental Tables 2, 4 and 5). SQS is a single copy gene in yeast (*ERG*9) and human [27,28].

SQS isoforms share a high level of sequence conservation with 79% identity (89% similarity for 410 amino acids) in *A. thaliana*, 77% identity (87% similarity for 408 amino acids) in *O. sativa*, and 90% identity (94% similarity for 413 amino acids) in *P. trichocarpa*. Yeast ERG9 sequence has 43% and 40% identity with *A. thaliana* SQS1 and

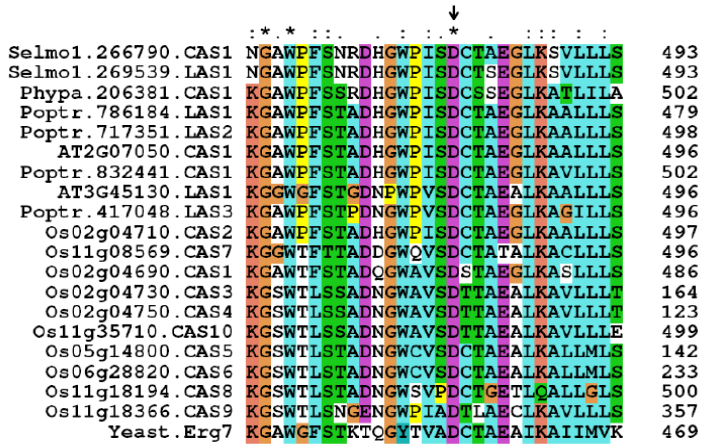

Figure 4: Aligned protein sequences of lansterol synthase (LAS) and cycloartenol synthase (CAS) from different model organisms. The conserved aspartate (D) is shown under inverted arrow.

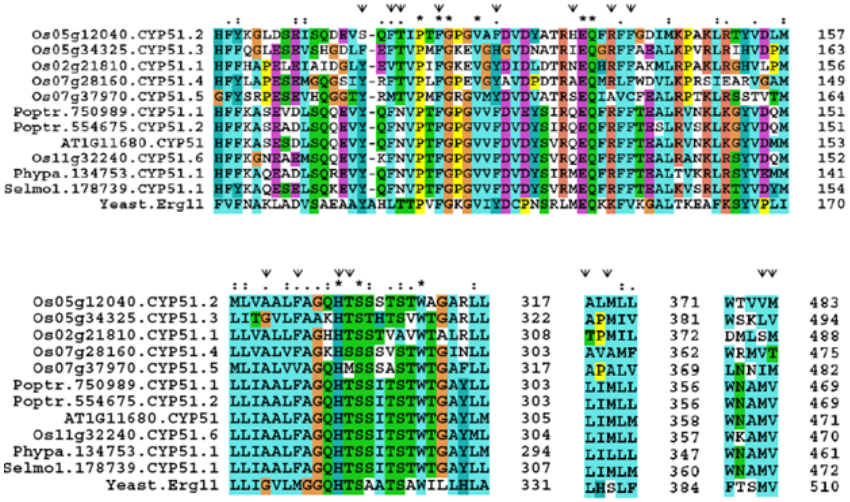

Figure 5: Protein sequence alignment of cytochrome P450 51 (CYP51). Several conserved residues scattered in CYP51 sequences of different organisms are indicated with inverted arrows on top of alignment [21].

SQS2, 45% and 42% identity with *O. sativa* SQS1 and SQS2, 41% and 42% identity with *P. trichocarpa* SQS1 and SQS2, respectively. SQS of *P. patens* and *S. moellendorffii* showed 42% and 38% identity, respectively with yeast ERG9 protein sequence.

Five domains (I to V) in SQS proteins of *A. thaliana*, *O. sativa*, *P. patens* and *S. moellendorffii* conserved with yeast SQS are identified. Domains II, III, IV and V are well conserved within SQS of yeast and plants, while domain I is less conserved. Sequence conservation of SQS proteins is especially low at N- and C-termini. Highly conserved aspartate-rich-motifs [DT(V/I)EDD and DY(L/N)ED] that occurs in domain II and domain IV of SQS proteins are also identified in SQS sequences of *A. thaliana*, *O. sativa*, *P. patens* and *S. moellendorffii*. A highly conserved sequence YC(H/Y)Y(V/A)AG(L/I)VG called the SQS motif [27] in domain III of SQS is identified [29-32] (Supplemental Figure 4).

Cytochrome P450 51 (CYP51) or ERG11

A single gene encodes CYP51 protein in yeast, *A. thaliana*, *P. patens* and *S. moellendorffii*. Interestingly, there are seven copies of *CYP*51 in *O. sativa* and two in *P. trichocarpa*. Low sequence conservation between yeast ERG11 and plant CYP51 protein sequences (27-32% identity) is found. CYP51 protein sequences of plants showed 42-84% identity when compared with *A. thaliana* CYP51. A pairwise sequence comparison of *O. sativa* CYP51 isoforms showed identity in the range of 41-65%. *P. trichocarpa* CYP51.1 and CYP51.2 isoforms showed high sequence identity (of 96%). Alignment of CYP51 protein sequences showed several conserved residues in diverse organisms (Figure 5) [33]. Only residues H and M are conserved in the sequence of locus Os07g37980 of *O. sativa* (Supplemental Table 6).

Sterol C-4 methyl oxidase (SMO) or ERG25

Five *SMO* genes in *A. thaliana*, nine in *O. sativa*, and two each in *P. trichocarpa, P. patens* and *S. moellendorffii* are identified (Table 3, Supplemental Tables 1-5). *O. sativa SMO* genes are annotated as FAHs in the RGAP database. All these FAHs are renamed here as *SMO*1-9. The locus Os01g04260, although annotated as FAH, showed high sequence similarity with *A. thaliana STE*1, a Δ7 sterol C-5 desaturase.

A high amino acid sequence identity (of 87%) is detected in *A. thaliana* SMO1 and SMO2 isoforms. A high sequence conservation is found between SMO1-1 and SMO1-2 (88% identity), SMO1-1 and SMO1-3 (76% identity) and SMO1-2 and SMO1-3 (76% identity) isoforms of *A. thaliana*. Any pairwise sequence comparison of SMO isoforms of *O. sativa* showed high sequence identity between SMO1 and SMO7 (66%), SMO2 and SMO4 (90%), SMO2 and SMO6 (74%), SMO3 and SMO9 (76%), SMO4 and SMO6 (89%) and SMO5 and SMO8 (85%). A relatively low sequence identity can be seen in other SMO isoforms of *O. sativa*. A high level of sequence conservation in SMO isoforms of *P. trichocarpa* (95% identity), *P. patens* (97% identity) and *S. moellendorffii* (82% identity) is identified. Yeast ERG25 share a relatively low 25-39% identity with SMO isoforms of *A. thaliana, O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii.*

A pairwise comparison of *A. thaliana* SMO isoforms with *O. sativa* SMO1 and SMO7 shared protein identity in the range of 53-73%. Their sequence identity with other *O. sativa* SMOs is poor. *A. thaliana* SMO2-1 and SMO2-2 revealed high sequence identity (73-84%) with SMO1 and SMO2 of *P. trichocarpa*, *P. patens* and *S. moellendorffii*. Other *A. thaliana* SMOs (SMO1-1, SMO1-2 and SMO1-3) disclosed a poor sequence similarity with SMO1 and SMO2 of *P. trichocarpa*, *P. patens* and *S. moellendorffii*.

Protein sequences of all SMOs of yeast and plants possess the characteristic histidine-rich motifs (HX3H, HX2HH, HX3H/D and HX4H) [34]. Motifs HX4H and HX2HH identified in SMO2, SMO4 and SMO6 of *O. sativa* did not align with rest of SMOs and appeared slightly displaced for the alignment parameters used. A similar situation is observed for motif HX3H in SMO3 and SMO9 of *O. sativa* (Figure 6).

3β-hydroxysteroid-dehydrogenase/C4-decarboxylase (3βHSD/D) or ERG26

Three genes to encode 3βHSDs are identified in *A. thaliana*. There are two 3βHSDs in *O. sativa* and one each in *P. trichocarpa*, *P. patens* and *S. moellendorffii.*

Amino acid sequences of yeast ERG26 and plant 3βHSDs share a low identity in the range of 30-37%. A pairwise protein sequence comparison of *A. thaliana* 3βHSD/D1 showed identities of 74%, 55% and 48% with 3βHSDs of *P. trichocarpa*, *S. moellendorffii* and *P. patens*, respectively. Protein sequence of *O. sativa* 3βHSD/D1 is 65% and 3βHSD/D2 is 35% identical to 3βHSD/D1 of *A. thaliana*. Sequence identity of *A. thaliana* 3βHSD/D2 with other plant 3βHSDs is almost similar to its 3βHSD/D1. However, *A. thaliana* 3βHSD/D showed a low sequence identity in the range of 37-46% with other plant 3βHSDs. Amino acid sequences of At3βHSD/D1 and At3βHSD/D2 share 78% identity, whereas At3βHSD/D1 and At3βHSD/D disclosed 40% identity, and At3βHSD/D2 and At3βHSD/D revealed 42% sequence identity. 3βHSDs of *O. sativa* showed a low 36% identity.

A conserved Yx3K catalytic motif, which is characteristic of the SDR superfamily, is identified in protein sequences of 3βHSDs. This motif is not found in AT2G43420.HSD and Os09g34090.HSD2.

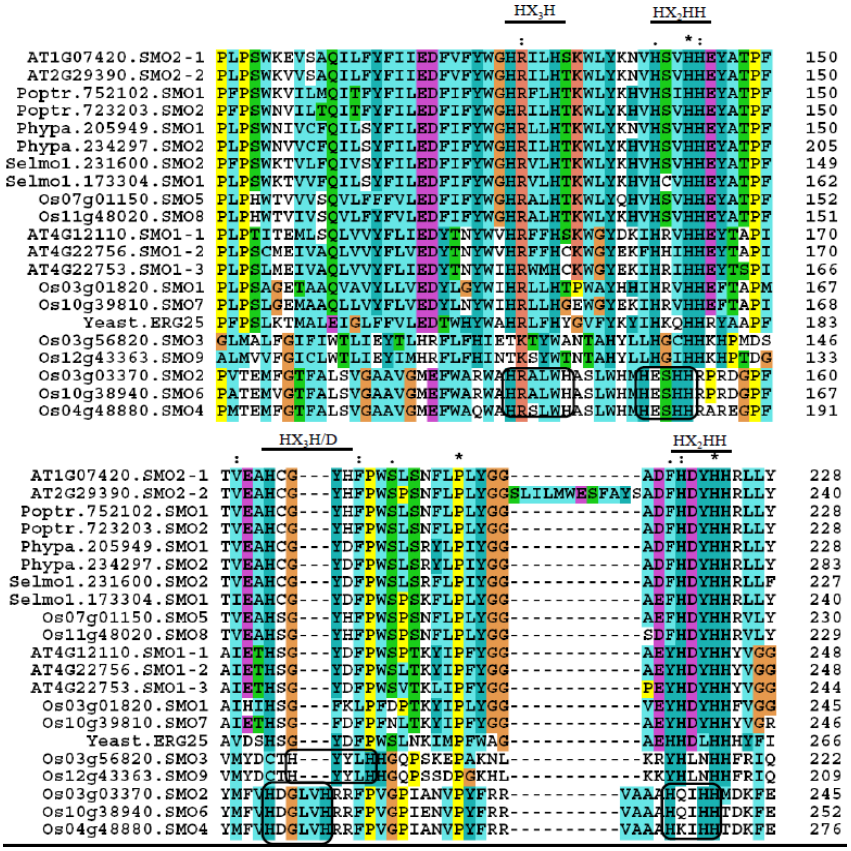

Figure 6: Aligned protein sequences of sterol C-4 methyl oxidases (SMOs). Conserved histidine-rich motifs (HX3H, HX2HH, HX3H/D and HX4H) are shown. Sequence of SMO2, SMO3, SMO4, SMO6 and SMO9 within the black box indicates slightly displaced motifs.

Glycine and aspartic residues (TGGxGxxAx18D) are conserved in 3βHSDs of plants and yeast (Supplemental Figure 5) [35-37].

- **C-3 ketoreduction or ERG27**

Blastp search did not identify an ortholog of yeast ERG27 in plant organisms.

- **C-4 demethylation or ERG28**

*ERG*28 is a single copy gene in yeast and plants. Protein sequence of ERG28 is well conserved (62-83% identity) in plants. Low sequence conservation can be seen in yeast and plant ERG28 proteins (25-33% identity). Protein alignment has identified a well conserved motif RTFG(V/T)WT in ERG28 sequences of yeast and plants (Supplemental Figure 6).

- **Cyclopropyl sterol isomerase (CPI)**

CPI is a land plant-specific enzyme that converts pentacyclic cyclopropyl sterols to conventional tetracyclic sterols [18]. *CPI*1 is a single copy gene in *A. thaliana* [18] as well as in *O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii*. *A. thaliana* CPI1 protein shared identity in the range of 58-81% with other plant CPIs. Four conserved motifs (SKRWGE, VGNYFWTHYF, YTFPS and LFYAIYF(I/F)VSFPMF) are identified in all plant CIP proteins (Supplemental Figure 7).

Δ7 (*DWF5*), Δ14 (*FACKEL*1 or *HYDRA*2) and Δ24(28) (*DWF1*) sterol reductases

(i) Sterol Δ7-reductase (ST7R)

ST7R is a single copy gene in *A. thaliana* (AT1G50430), *O. sativa* (Os02g26650), *P. patens* (Phypa.104832) and *S. moellendorffii* (Selmo1.185064), but two copies of ST7R occur in *P. trichocarpa* (Poptr.833904 and Poptr.765023). *O. sativa* locus Os02g26650 is a Δ7-sterol reductase, which is annotated as a Δ14-sterol reductase in the RGAP database. The protein sequence of locus Os02g26650 showed 83% identity with Dwf5 of *A. thaliana*. ST7R activity is not found in yeast [38].

A pairwise comparison uncovered that protein identity of ST7Rs of *P. trichocarpa* (ST7R1 and ST7R2), *P. patens* and *S. moellendorffii* with *Dwf*5 of *A. thaliana* is 87% and 84%, 66%, and 74%, respectively. Proteins of *P. trichocarpa* ST7R1 and ST7R2 showed a high 92% identity. Highly conserved domains I and II are identified in ST7R proteins of *A. thaliana*, *O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii* (Figure 7).

(ii) Sterol Δ14-reductase (ST14R) or ERG24

FACKEL (*FK*) is a single copy gene in *A. thaliana* encoded by the locus AT3G52940. A preliminary search has identified three genes to encode ST14R in *O. sativa* genome. ST14R is a single copy gene in *P. trichocarpa* (Poptr.826487), *P. patens* (Phypa.67627) and *S. moellendorffii* (Selmo1.99117).

Protein sequence of yeast ERG24 showed 34-36% identity with ST14R sequences of *A. thaliana, O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii*. ST14R1 and ST14R2 of *O. sativa* share a significant 96% identity and 98% similarity (for 370 amino acids). An earlier identified signature motif LLxSGxWGxxRH [38], and other conserved motifs DWWxGxQLNP and GFMLxFGD are identified in ST14R proteins of *A. thaliana*, *O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii*. However, only two of three ST14R proteins of *O. sativa* shared these motifs, which were not found in the sequence of locus Os02g26650. This locus has been annotated as Δ14-sterol reductase in the RGPA database. However, it shares only 27% identity, whereas ST14R1 and ST14R2 are 70-71% identical to the gene *FK* of *A. thaliana*. Two gene copies of *ST14R* (*ST14R1* and *ST14R2*) are validated in the *O. sativa* genome (Supplemental Figure 8).

(iii) Sterol Δ24-reductase (ST24R) or ERG4

ST24R is a single copy gene in *A. thaliana* (AT3G19820), *O. sativa* (Os10g25780) and *S. moellendorffii* (Selmo1.139528). Two genes of ST24R in *P. trichocarpa* (Poptr.765722 and Poptr.822314) and *P. patens* (Phypa.234317 and Phypa.85741) are identified.

Overall amino acid sequence similarity of ST24Rs from yeast, *A. thaliana*, *O. sativa*, *P. trichocarpa, P. patens* and *S. moellendorffii* is relatively poor. A very limited similarity can be seen scattered over small stretches in these protein sequences. Blast search with yeast ERG4 as query has revealed a high similarity with ST14Rs and not with ST24Rs of *O. sativa*, *P. trichocarpa, P. patens* and *S. moellendorffii*. However, the *DWF1* of *A. thaliana* showed a high 73-84% identity with ST24Rs of *O. sativa*, *P. trichocarpa, P. patens* and *S. moellendorffii*. The locus Os10g25780 of *O. sativa* which is currently annotated as FAD-linked oxidoreductase showed 80% identity and 89% positives (for 559 amino acids) with *DWF1* of *A. thaliana*. Based on this high sequence similarity, the locus Os10g25780 can be considered as ST24R in *O. sativa*. *ST24R* is a single copy gene in *O. sativa*. High sequence conservation for ST24Rs of *P. trichocarpa* (97% identity) and *P. patens* (88% identity) is revealed. A significant similarity to FAD-binding domain $(GxGxxG(x)_{15}E)$ is identified in ST24R proteins of *A. thaliana*, *O. sativa*, *P. trichocarpa*, *P. patens* and *S. moellendorffii*. This domain is absent in yeast ERG4 [39] (Figure 8).

Discussion

Sterol biosynthesis, diversity, and nature of sterols have been

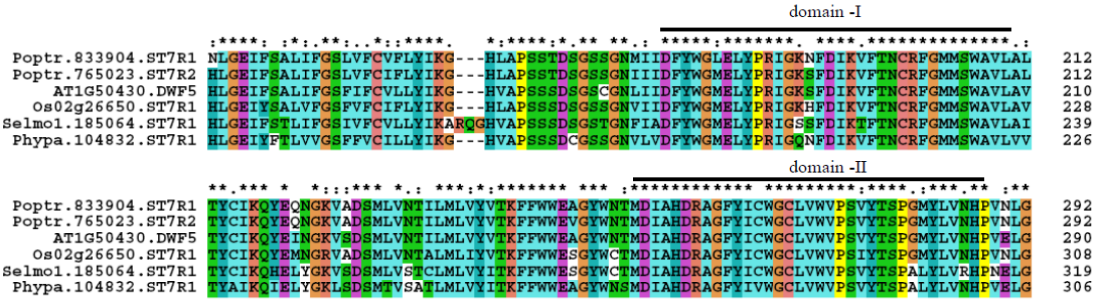

Figure 7: Alignment of protein sequences of sterol Δ7-reductase (ST7R). Highly conserved domain I and II are shown.

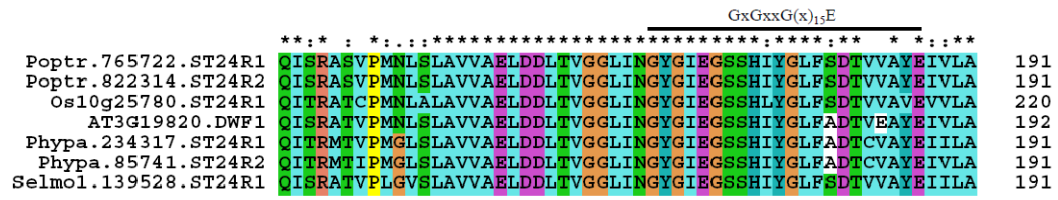

Figure 8: Aligned protein sequences of Δ24 sterol reductase (ST24R). The domain (GxGxxG(x)15E) is shown.

studied in fungi, higher plants and animals [1]. Many of the reactions of yeast ergosterol synthesis are identical with those of cholesterol pathway in vertebrates [4]. However, sterol biosynthesis requires four additional genes in plants. These are cycloartenol synthase [10], sterol Δ7-reductase [38], a second sterol methyltransferase [40] and cyclopropyl sterol isomerase [18]. The present investigation has identified these plant-specific sterol genes in model magnoliophytes (*A. thaliana, P. trichocarpa*, *O. sativa*), a bryophyte (*P. patens*) and a lycophyte (*S. moellendorffii*).

Commonly, homologous gene products are known to perform same enzymatic steps of sterol biosynthesis in fungi, land plants and vertebrates, with the exception that the non-homologous ERG2 in fungi and HYD1 in land plants perform Δ-8, Δ-7 isomerization step, and the non-homologous ERG4 in fungi and DWF1 in land plants perform Δ-24 reduction [1]. The current study also supports the above conclusion.

Squalene is the first intermediate unique to sterol biosynthesis. Eleven, twelve and fourteen enzymes are involved in the conversion of squalene to cholesterol (in vertebrates), ergosterol (in fungi), and stigmasterol (in land plants), respectively [41]. The basic set of sterol enzymes harbored by the woody angiosperm *P. trichocarp*, a bryophyte *P. patens*, and a lycophyte *S. moellendorffii* is identical to that of *A. thaliana* and *O. sativa*.

The ergosterol pathway in fungi like yeast is only partially similar to that of phytosterols: the first cyclization product of SQE in yeast (and animals capable of synthesizing sterols), depending on the activity of LAS is lanosterol, whereas in the phytosterol pathway it is by far predominantly cycloartenol, synthesized by CAS. Although the biochemical steps of sterol metabolism are well elucidated [24,42,43], however, why plants synthesize sitosterol via the major cycloartenol route remains unclear? Nevertheless, the existence of both LAS and CAS in *A. thaliana* suggests functional redundancy of the first step in plant sterol biosynthesis [44,45]. Recently, a CAS1-specific functional sterol pathway was engineered in yeast, and bulk dependence on CAS1-mediated sterol biosynthesis was shown in tobacco [46]. It might be noteworthy that in green algae with a completely sequenced genome like *Chlamydomonas reinhardtii* a CAS does exist, but the final sterol that accumulates is ergosterol, like in fungi [**8**]. In this report, both *CAS* and *LAS* genes were identified in *P. trichocarpa* and *S. moellendorffii*, however their role in sterol metabolism needs further investigation. Also, the functional aspects of a large pool of *CAS* genes in *O. sativa* deserve further investigations. The sequence identity that is distributed over a range within and between species might also point to other functions of such gene isoforms. Further, sterol properties in plants from varied ecological habitats also might differ.

The physiological significance of cyclopropyl sterol intermediates in plants is not fully understood [44]. Furthermore, an enzyme C-3 ketoreductase present in fungi and vertebrates is absent in land plants. A gene for this enzyme was proposed based on phylogenomics approach in land plants. The candidate gene was identified as succinate-semialdehyde dehydrogenase [NAD(P)+]. In this report, no sequence similarity in yeast ERG27 and succinate-semialdehyde dehydrogenases of *A. thaliana* or *O. sativa* was found. Yet an unknown gene encoding the elusive C-3 ketoreductase might exist in land plants [1].

Protein-protein interactions among sterol enzymes are less studied [47]. Feedback regulation of enzymes by the end product ergosterol has been studied in yeast [48,49] but not in plants. It is still unclear if multiple copies of a gene lead to redundant or unique function [20]. Hence the role of individual gene isoforms in plant sterol biosynthesis remains unclear.

Complementation of yeast mutants with *A. thaliana* orthologs has resulted in much of the information about sterol pathway in *A. thaliana* [50]. A similar approach is needed to decipher the functions of individual sterol pathway genes in other model plants like *P. trichocarpa*, *O. sativa*, *P. patens* and *S. moellendorffii*. Mutational and transgenic studies will also provide new insights into the roles of these genes in sterol pathway, and sterols in plant development.

Genomic copy number alteration is known to contribute substantially to phenotypic variation and population diversity. This is also relevant to complex phenotypes and has functional consequences, such as differential expression of genes [51]. Frequently, isogenes are differentially expressed in 'space and time' (plant organs and different tissues, during development, etc). Therefore, further investigation of expression of sterol genes in diverse plants is required.

In conclusion, genes for sterol enzymes are found in diverse plant phyla. The protein sequence similarity of sterol enzymes in yeast and plants is poor. Highly conserved regions can be found among sterol proteins in various plants. Genomic copy number variation is enriched among sterol pathway genes. The functions for several sterol pathway genes in plants remain elusive. The current finding provide the most comprehensive and systematic cataloging of multiple isoforms of key genes of plant sterol biosynthesis.

References

1. Desmond E, Gribaldo S (2009) Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. Genome Biol Evol 1: 364-81.
2. Rohmer M (2008) From molecular fossils of bacterial hopanoids to the formation of isoprene units: discovery and elucidation of the methylerythritol phosphate pathway. Lipids 43: 1095-107.
3. Goldstein JL, Brown MS (1990) Regulation of the mevalonate pathway. Nature 343: 425-30.
4. Mo C, Bard M (2005) Erg28p is a key protein in the yeast sterol biosynthetic enzyme complex. J Lipid Res 46: 1991-8.
5. Devarenne TP, Ghosh A, Chappell J (2002) Regulation of squalene synthase, a key enzyme of sterol biosynthesis in tobacco. Plant Physiol 129: 1095-106.
6. Tomazic ML, Najle SR, Nusblat AD, Uttaro AD, Nudel CB (2011) A novel sterol

desaturase-like protein promoting dealkylation of phytosterols in *Tetrahymena thermophila*. Eukaryot Cell 10: 423-34.

7. Laloi M, Perret AM, Chatre L, Melser S, Cantrel C, et al. (2007) Insights into the role of specific lipids in the formation and delivery of lipid microdomains to the plasma membrane of plant cells. Plant Physiol 143:461-72.

8. Brumfield KM, Moroney JV, Moore TS, Simms TA, Donze D (2010) Functional characterization of the *Chlamydomonas reinhardtii* ERG3 ortholog, a gene involved in the biosynthesis of ergosterol. PLoS One 5: e8659.

9. Choe S, Noguchi T, Fujioka S, Takatsuto S, Tissier CP, et al. (1999) The *Arabidopsis* dwf7/ste1 mutant is defective in the Δ^7 Sterol C-5 desaturation step leading to brassinosteroid biosynthesis. Plant Cell 11: 207-21.

10. Corey EJ, Matsuda SPT, Bartel B (1993) Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. Proc Natl Acad Sci USA 90: 11628-32.

11. Souter M, Topping J, Pullen M, Friml J, Palme K, et al. (2002) *hydra* mutants of *Arabidopsis* are defective in sterol profiles and auxin and ethylene signaling. Plant Cell 14: 1017-31.

12. Gachotte D, Husselstein T, Bard M, Lacroute F, Benveniste P (1996) Isolation and characterization of an *Arabidopsis thaliana* cDNA encoding a Δ^7-sterol-C-5-desaturase by functional complementation of a defective yeast mutant. Plant J 9: 391-8.

13. Bach TJ, Benveniste P (1997) Cloning of cDNAs or genes encoding enzymes of sterol biosynthesis from plants and other eukaryotes: heterologous expression and complementation analysis of mutations for functional characterization. Prog Lipid Res 36: 197-226.

14. Woods RA (1971) Nystatin-resistant mutants of yeast: alterations in sterol content. J Bacteriol 108: 69-73.

15. Bard M (1972) Biochemical and genetic aspects of nystatin resistance in *Saccharomyces cerevisiae*. J Bacteriol 111: 649-57.

16. Bard M, Woods RA, Bartón DH, Corrie JE, Widdowson DA (1977) Sterol mutants of *Saccharomyces cerevisiae*: chromatographic analyses. Lipids 12: 645-54.

17. Gachotte D, Barbuch R, Gaylor J, Nickels E, Bard M (1998) Characterization of the *Saccharomyces cerevisiae* ERG26 gene encoding the C-3 sterol dehydrogenase (C-4 decarboxylase) involved in sterol biosynthesis. Proc Natl Acad Sci, USA 95: 13794-9.

18. Lovato MA, Hart EA, Segura MJR, Giner J-L, Matsuda SPT (2000) Functional cloning of an *Arabidopsis thaliana* cDNA encoding cycloeucalenol cycloisomerase. J Biol Chem 275: 13394-7.

19. 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: 4876-82.

20. Rasbery JM, Shan H, LeClair RJ, Norman M, Matsuda SPT, et al. (2007) *Arabidopsis thaliana s*qualene epoxidase 1 is essential for root and seed development. J Biol Chem 282: 17002-13.

21. Rahier A, Pierre S, Riveill G, Karst F (2008) Identification of essential amino acid residues in a sterol 8,7-isomerase from *Zea mays* reveals functional homology and diversity with the isomerases of animal and fungal origin. Biochem J 414: 247-59.

22. Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, et al. (2006) Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in *Arabidopsis* and tomato. Plant Cell 18: 1008-22.

23. Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol 54:629-67.

24. Diener AC, Li H, Zhou W-x, Whoriskey WJ, Nes WD, et al. (2000) STEROL METHYLTRANSFERASE 1 controls the level of cholesterol in plants. Plant Cell 12: 853-70.

25. Bouvier-Nave P, Husselstein T, Desprez T, Benveniste P (1997) Identification of cDNAs encoding sterol methyl-transferases involved in the second methylation step of plant sterol biosynthesis. Eur J Biochem. 246: 518-29.

26. Kagan RM, Clarke S (1994) Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem. Biophys 310: 417-27.

27. Manavalan LP, Chen X, Clarke J, Salmeron J, Nguyen HT (2012) RNAi-mediated disruption of squalene synthase improves drought tolerance and yield in rice. J Exp Bot 63: 163-75.

28. Hata S, Sanmiya K, Kouchi H, Matsuoka M, Yamamoto N (1997) cDNA cloning of squalene synthase genes from mono- and dicotyledonous plants, and expression of the gene in rice. Plant Cell Physiol 38: 1409-13.

29. Huang Z, Jiang K, Pi Y, Hou R, Liao Z, et al. (2007) Molecular cloning and characterization of the yew gene encoding squalene synthase from *Taxus cuspidate*. J Biochem Mol Biol 40: 625-35.

30. Devarenne TP, Shin DH, Back K, Yin S, Chappell J (1998) Molecular characterization of tobacco squalene synthase and regulation in response to fungal elicitor. Arch Biochem Biophys 349: 205-15.

31. Robinson GW, Tsay YH, Kienzle BK, Smith-Monroy CA, Bishop RW (1993) Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation. Mol Cell Biol 13: 2706-17.

32. Kribii R, Arro M, Del Arco A, Gonzalez V, Balcells L, et al. (1997) Cloning and characterization of the *Arabidopsis thaliana* SQS1 gene encoding squalene synthase - involvement of the C-terminal region of the enzyme in the channeling of squalene through the sterol pathway. Eur J Biochem 249: 61-9.

33. Podust LM, Poulos TL, Waterman MR (2001) Crystal structure of cytochrome P450 14α-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. Proc Natl Acad Sci, USA 98: 3068-73.

34. Darnet S, Rahier A (2004) Plant sterol biosynthesis: identification of two distinct families of sterol 4α-methyl oxidases. Biochem J 378: 889-98.

35. Rahier A, Darnet S, Bouvier F, Camara B, Bard M (2006) Molecular and enzymatic characterizations of novel bifunctional 3β-hydroxysteroid-dehydrogenases/C4-decarboxylases from *Arabidopsis thaliana*. J Biol Chem 281: 27264-77.

36. Kallberg Y, Oppermann U, Jörnvall H, Persson B (2002) Short-chain dehydrogenases/reductases (SDRs). Eur J Biochem 269: 4409-17.

37. Rahier A, Bergdoll M, Geneviève Génot, Bouvier F, Camara B (2009) Homology modeling and site-directed mutagenesis reveal catalytic key amino acids of 3β-hydroxysteroid-dehydrogenase/C4-decarboxylase from *Arabidopsis*. Plant Physiol 149: 1872-86.

38. Lecain E, Chenivesse X, Spagnoli R, Pompon D (1996) Cloning by metabolic interference in yeast and enzymatic characterization of *Arabidopsis thaliana* sterol delta 7-reductase. J Biol Chem 271: 10866-73.

39. Dym O, Eisenberg D (2001) Sequence-structure analysis of FAD-containing proteins. Protein Sci 10: 1712-28.

40. Bouvier-Nave P, Husselstein T, Benveniste P (1998) Two families of sterol methyltransferases are involved in the first and the second methylation steps of plant sterol biosynthesis. Eur J Biochem. 256: 88-96.

41. Marijanovic Z, Laubner D, Möller G, Gege C, Hunsen B, et al. (2003) Closing the gap: identification of human 3-ketosteroid reductase, the last unknown enzyme of mammalian cholesterol biosynthesis. Mol Endocrinol 17: 1715-25.

42. Jang JC, Fujioka S, Tasaka M, Seto H, Takatsuto S, et al. (2000) A critical role of sterols in embryonic patterning and meristem programming revealed by the fackel mutants of *Arabidopsis thaliana*. Genes Dev 14: 1485-97.

43. Schrick K, Mayer U, Horrichs A, Kuhnt C, Bellini C, et al. (2000) FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. Genes Dev 14: 1471-1484.

44. Babiychuk E, Bouvier-Nave´ P, Compagnon V, Suzuki M, Muranaka T, et al. (2008) Allelic mutant series reveal distinct functions for *Arabidopsis* cycloartenol synthase 1 in cell viability and plastid biogenesis. Proc Natl Acad Sci, USA 105: 3163-3168.

45. Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T (2009) Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in *Arabidopsis*. Proc Natl Acad Sci, USA 106: 725-730.

46. Gas-Pascual E, Berna A, Bach TJ, Schaller H (2014) Plant oxidosqualene metabolism: cycloartenol synthase-dependent sterol biosynthesis in *Nicotiana benthamiana*. PLoS One 9: e109156.

47. Mo C, Valachovic M, Randall SK, Nickels JT, Bard M (2002) Protein-protein interactions among C-4 demethylation enzymes involved in yeast sterol biosynthesis. Proc Natl Acad Sci, USA 99: 9739-9744.

48. He J-X, Fujioka S, Li T-C, Kang SG, Seto H, et al. (2003) Sterols regulate

development and gene expression in *Arabidopsis*. Plant Physiol 131: 1258-1269.

49. Smith SJ, Crowley JH, Parks LW (1996) Transcriptional regulation by ergosterol in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol 16: 5427-5432.

50. Benveniste P (2004) Biosynthesis and accumulation of sterols. Annu Rev Plant Biol 55: 429-57.

51. Locke ME, Milojevic M, Eitutis ST, Patel N, Wishart AE, et al. (2015) Genomic copy number variation in *Mus musculu*. BMC Genomics 16: 497.