New Phytologist

Identification of key amino acid residues determining product specificity of 2,3-oxidosqualene cyclase in *Oryza* species

Zheyong Xue[1,2]*, Zhengwei Tan[1,3,4]*, Ancheng Huang[2], Yuan Zhou[1,3], Juncong Sun[1,3], Xiaoning Wang[5], Ramesha B. Thimmappa[2], Michael J. Stephenson[2], Anne Osbourn[2] and Xiaoquan Qi[1]

[1]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; [2]Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK; [3]University of Chinese Academy of Sciences, Beijing 100049, China; [4]Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China; [5]Department of Natural Product Chemistry, Key Laboratory of Chemical Biology (Ministry of Education), School of Pharmaceutical Sciences, Shandong University, 44 West Wenhua Road, Jinan 250012, China

Authors for correspondence:
Xiaoquan Qi
Tel: +86 10 62836671
Email: xqi@ibcas.ac.cn

Anne Osbourn
Tel: +44 1603 450407
Email: anne.osbourn@jic.ac.uk

Received: *1 November 2017*
Accepted: *5 February 2018*

New Phytologist (2018) **218:** 1076–1088
doi: 10.1111/nph.15080

Key words: chemical diversity, conformation, orysatinol, oxidosqualene cyclase (OSC), parkeol, product specificity, rice, triterpene.

Summary

- Triterpene synthases, also known as 2,3-oxidosqualene cyclases (OSCs), synthesize diverse triterpene skeletons that form the basis of an array of functionally divergent steroids and triterpenoids. Tetracyclic and pentacyclic triterpene skeletons are synthesized via protosteryl and dammarenyl cations, respectively. The mechanism of conversion between two scaffolds is not well understood.
- Here, we report a promiscuous OSC from rice (*Oryza sativa*) (OsOS) that synthesizes a novel pentacyclic triterpene orysatinol as its main product. The *OsOS* gene is widely distributed in *indica* subspecies of cultivated rice and in wild rice accessions. Previously, we have characterized a different OSC, OsPS, a tetracyclic parkeol synthase found in *japonica* subspecies.
- Phylogenetic and protein structural analyses identified three key amino acid residues (#732, #365, #124) amongst 46 polymorphic sites that determine functional conversion between OsPS and OsOS, specifically, the chair–semi(chair)–chair and chair–boat–chair interconversions. The different orientation of a fourth amino acid residue Y257 was shown to be important for functional conversion
- The discovery of orysatinol unlocks a new path to triterpene diversity in nature. Our findings also reveal mechanistic insights into the cyclization of oxidosqualene into tetra- and pentacyclic skeletons, and provide a new strategy to identify key residues determining OSC specificity.

Introduction

Metabolic enzymes produce an enormous array of chemicals that provide adaptive strategies for plants in challenging terrestrial environments (Weng, 2014). Over 20 000 steroids and triterpenoids with *c.* 200 different skeletons have been identified in eukaryotic organisms (Xu *et al.*, 2004). In addition to the essential function of sterols in maintaining cell membrane fluidity and permeability (Bloch, 1965; Benveniste, 1986; Parks & Casey, 1995), a large number of structurally diverse plant triterpenoids exist that have important functions in crop defense, food quality (Osbourn *et al.*, 2011) and as drug leads (Augustin *et al.*, 2011; Moses *et al.*, 2014). 2,3-Oxidosqualene cyclases (OSCs) catalyze the first committed step in triterpene biosynthesis, namely the cyclization of the universal triterpene precursor 2,3-oxidosqualene (**1**), and therefore define sterol and triterpene skeletal diversity (Thimmappa *et al.*, 2014). Most steroids reported from nature to date are seemingly derived from tetracyclic skeletons, including cycloartenol, lanosterol and parkeol (**2**) via a chair–boat–chair (C-B-C) conformation prosteryl cation path. By contrast, the most common pentacyclic triterpenoid skeletons include lupeol (**3**) and β-amyrin (**4**), generated via a chair–chair–chair (C-C-C) conformation dammarenyl cation path (Fig. 1a). However, despite the intensive study of the OSC enzyme family, its functional diversity is yet to be fully explored.

Certain key amino acid residues have been shown to determine the functional diversity of OSCs (Chappell, 2002; Wu *et al.*, 2008). Random mutagenesis and structure-guided, site-directed mutagenesis experiments have identified a total of at least 17 amino acid residues that are essential for the initiation or cyclization process resulting in product specificity (Hart *et al.*, 1999; Herrera *et al.*, 2000; Matsuda *et al.*, 2000; Wu & Griffin, 2002; Salmon *et al.*, 2016) (Supporting Information Table S1). Eight directly interacting residues (DIRs), the catalytic D455 of the conserved DCTAE motif (Christianson, 2006), aromatic amino

*These authors contributed equally to this work.

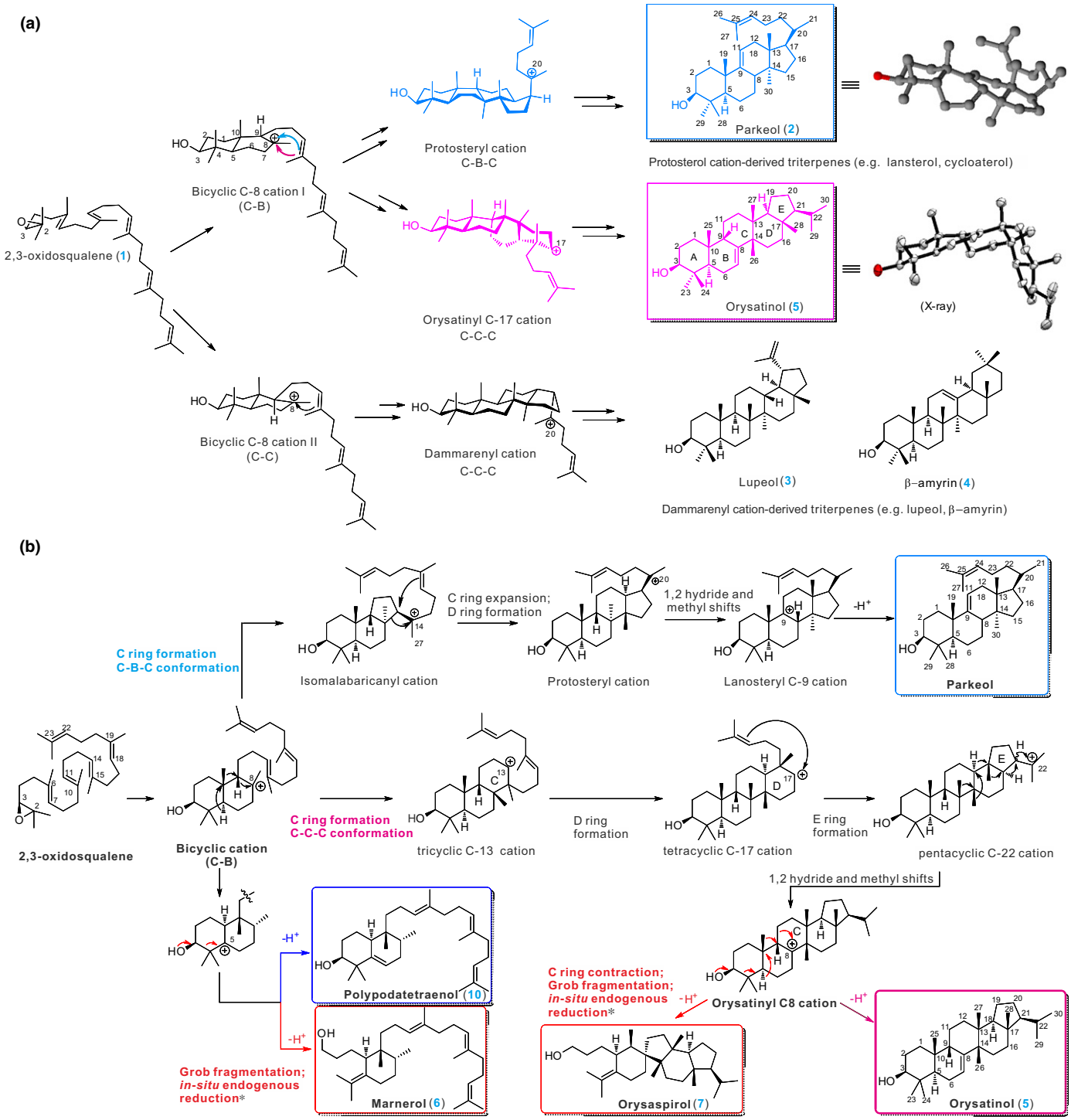

Fig. 1 Carbocation intermediates and products of 2,3-oxidosqualene cyclization by triterpene cyclases. (a) Substrate folding leads to the protosteryl C-B-C (chair–boat–chair), orysatinol C-17 C-C-C (chair–(chair)–chair) or dammarenyl C-C-C conformation in initial cyclization. Three-dimensional (3D) views of orysatinol and parkeol represented by X-ray structure (ORTEP drawing, shown at 30% probability) (Supporting Information Notes S1) and 3D-optimized structure from ACD-lab, respectively. (b) Proposed reaction scheme of cyclization. The cyclization pathway of parkeol, polypodatetraenol, marnerol and orysatinol, which initially shared the same bicyclic C-8 cation I (C-B: chair–boat) rather than biocyclic C-8 cation II (C-C: chair–chair). *Marnerol and orysaspirol are the reduction products of their aldehyde form by endogenous reductases in *Pichia pastoris*.

acids Y98, F444, Y503, W581, F696 and Y704, and the positively charged amino acid H232, are predicted to interact with intermediate cations by providing π-cation- or electric charge-based interactions (these residues are numbered here using human lanosterol synthase, HsLAS). Mutagenesis of these amino acid residues resulted in either loss of functions or the formation

of premature terminated cyclization products. Mutations at the six indirectly interacting residues (IIRs), W230, G441, F442, S454, S699 and C700, and three other residues (T381, C449 and V453), caused changes in product types and/or profiles (Table S1). Although extensive mutagenesis experiments have been conducted in many laboratories, very little is known about the key residues required for functional conversion between tetra- and pentacyclic triterpene synthases.

Rice (*Oryza sativa*) is an important crop. Its genome contains an expanded OSC gene family with 12 members, four of which have been shown previously to synthesize a variety of triterpenes (Inagaki *et al.*, 2011; Ito *et al.*, 2011; Xue *et al.*, 2012; Sun *et al.*, 2013). *OsOSC7* from the ssp. *japonica* of cultivated rice encodes an accurate tetracyclic triterpene synthase (OsOSC7/OsPS) that cyclizes 2,3-oxidosqualene to form tetracyclic parkeol (Ito *et al.*, 2011; Xue *et al.*, 2012). Sequence variation between the *OsOSC7* genes in the *japonica* and *indica* subspecies suggests that OsOSC7 from *indica* may produce non-parkeol triterpenes. In this study, *OsOSC7* genes from a total of 34 accessions of cultivated rice and the closely related wild species, *Oryza rufipogon*, *Oryza nivara* and *Oryza meridionalis* (Kovach *et al.*, 2007; Sang & Ge, 2007), were cloned to investigate the functional diversity of the cognate gene products. Using a combined approach that relied on phylogenetic, molecular evolution, structural modeling and biochemical analyses, the key amino acid residues underlying triterpene product specificity were identified, providing catalytic insights into the functional conversion between tetra- and pentacyclic triterpenes.

Materials and Methods

Plant collection and growth

The cultivated rice (*Oryza sativa* L.) and wild rice (*Oryza rufipogon*, *Oryza nivara* and *Oryza meridionalis*) accessions were from collections of rice accessions maintained at the Institute of Botany, Chinese Academy of Sciences (IBCAS) in Beijing, and China National Rice Research Institute (CNRRI) in Hangzhou, China. The accessions were selected on the basis of the germplasm database records of phenotypic data and sampling localities to maximize genetic and geographic diversity, and were maintained by selfing.

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR)

The seeds of Zhonghua 11 (*O. sativa* ssp. *japonica,* ZH11) and Guangluai 15 (*O. sativa* ssp. *indica*, GL15) were soaked in sterile distilled water and germinated on wet filter paper at 30°C. The radicles, coleoptiles and plumules were collected from 2-d-old seedlings; 20-d-old seedlings were transplanted to the field at IBCAS. The roots, leaf sheaths, leaf blades and shoot apical meristems were collected from plants at the five-leaf stage, and roots, nodes, culms (stems), leaf sheaths, leaf blades, lemmas, paleas, stamens and pistils were collected at the flowering stage. Embryos and endosperms were collected 21 d after pollination. Tissue samples were frozen in liquid nitrogen before RNA extraction using TRI-REAGENT (Sigma catalog no. T9424). For qRT-PCR analysis, cDNA synthesis was carried out using Superscript III Reverse Transcriptase (Invitrogen) and 2 μg of DNase-treated total RNA, according to the manufacturer's instructions. Primers for the amplification of fragments of *OsOSC7j*, *OsOSC7i* and *OsACTIN1* (*LOC_Os03g50885*, as a reference gene) are listed in Table S2. qRT-PCR analysis was carried out on a Rotor-Gene 3000 thermocycler (Corbett Research, Mortlake, Australia) using the SYBR Green ER qPCR SuperMix Universal Kit (Invitrogen). The relative expression levels for each gene were normalized to the actin gene *OsACTIN1*, with three biological replicates per targeted gene.

Gene cloning and yeast transformation

cDNA derived from leaf sheath material was used to clone OsOSC7 coding sequences (CDSs). The amplified products were cloned into the pGEM-T easy vector (Promega) and sequenced from both ends. They were then cloned into the expression vector pPICZA (Invitrogen) between the *Spe*I and *Xba*I restriction sites to place the OSC open reading frame (ORF) under the control of the methanol-inducible promoter *AOX1*. *Pichia pastoris* wild-type strain X33 was transformed using electroporation according to the protocol described in the EasySelect™ Echo™-Adapted Phicia Expression kit (Catalog no. ET230-02; Invitrogen) using Bio-Rad Gene Pulser Xcell.

Purification and structural elucidation of *O. sativa* OSC (OsOS) and *O. rufipogon* polypodatetraenol synthase (OrPtS) products

Orysatinol was extracted from a 2-l *P. pastoris* culture expressing wild-type *OsOS*. The cell pellet was mixed with 0.5 l of saponification reagent (20% (w/v) KOH in 50% (v/v) ethanol) and incubated at 70°C for 2 h before extracting twice with an equal volume of hexane. The hexane extracts was loaded onto a silica gel column (zcx. II, granularity 200–300; Haiyang, Qingdao, China), 30 cm long and 2.4 cm in diameter, and eluted with hexane : ethyl acetate (6 : 1 v/v). Fractions were collected in 10-ml tubes and analyzed by thin layer chromatography (TLC) and gas chromatography-mass spectrometry (GC-MS), as described below. Fractions containing orysatinol were dried in a rotary evaporator and further purified by reverse-phase high-performance liquid chromatography (HPLC) on an Agilent 1200 series liquid chromatograph (Santa Clara, CA, USA) equipped with a semi-preparative column (Eclipse XDB-C18, 5 μm, 9.4 mm × 250 mm, Santa Clara, CA, USA) using a gradient from 95% to 100% methanol for 35 min at a flow rate of 2.5 ml min^{-1} at 40°C. The semi-preparative fractions were collected manually based on UV absorbance at 210 nm and monitored by GC-MS using the methods described below. Fractions containing the compound of interest were concentrated to dryness.

Marnerol, orysaspirol and polypodatetraenol were purified as described above for orysatinol from 40 l of *P. pastoris* culture expressing OsOS substitution strains and OrPtS, respectively. The purified compounds were analyzed by nuclear magnetic

resonance (NMR) (including ^{1}H, ^{13}C, Distortionless Enhancement by Polarization Transfer, Homonuclear chemical shift, Heteronuclear Single Quantum Coherence, Heteronuclear Multiple-Bond Correlation, Nuclear Overhauser Effect Spectroscopy experiments).

General considerations for NMR

NMR spectra were recorded in Fourier transform mode at a nominal frequency of 800 MHz for ^{1}H and 200 MHz for ^{13}C NMR, or 400 MHz for ^{1}H NMR and 100 MHz for ^{13}C NMR, using the specified deuterated solvent. Chemical shifts were recorded in parts per million (ppm) and referenced to the residual solvent peak or to an internal tetramethylsilane standard. Multiplicities are described as: s, singlet; d, doublet; dd, doublet of doublets; dt, doublet of triplets; t, triplet; q, quartet; m, multiplet; br, broad; appt, apparent; coupling constants are reported in hertz.

X-Ray crystallographic analysis of orysatinol

Single crystals suitable for X-ray analysis were obtained by recrystallization from methanol. A colorless platelet crystal having approximate dimensions of $0.10 \times 0.75 \times 0.79\ mm^3$ was used for analysis. X-Ray diffraction data were collected on a Rigaku MicroMax 002+ diffractometer (The Woodlands, TX, USA) using Cu Kα radiation at a wavelength of 1.54187 Å from a fine-focus sealed tube generator operated at 45 kV and 88 mA and in the ω–κ scan mode. The crystal-to-detector distance was 45 mm. Crystal data of **5**: $(C_{30}H_{49}O)_2 \cdot (CH_3OH)_3$, $M = 406.39$, triclinic, space group P1, $a = 7.3976(10)$ Å, $b = 7.4237(12)$ Å, $c = 28.840(9)$ Å, $\alpha = 84.84(3)°$, $\beta = 88.94(3)°$, $\gamma = 67.91(3)°$, $V = 1461.4(6)$ Å^3, $Z = 2$, $D_{calcd} = 1.077\ Mg/cm^3$, $F(000) = 548$. Program used to refine structure: SHELXL-97; refinement on F^2, full-matrix least-squares calculations. All non-hydrogen atoms were refined anisotropically, and all hydrogen atoms were placed in geometrically calculated positions and refined as riding atoms with the relative isotropic parameters. A total of 7101 reflections (6635 unique, $R_{int} = 0.1622$) was collected from 3.08° to 73.33° in θ and index ranges $8 \geq h \geq -4$, $9 \geq k \geq -9$, $35 \geq l \geq -35$. The final stage converged to $R_1 = 0.1054$ ($wR_2 = 0.2537$) for 3438 observed reflections (with $I > 2\sigma(I)$) and 617 variable parameters, and $R_1 = 0.1309$ ($wR_2 = 0.2972$) for all unique reflections and GoF = 0.985. The refined fractional atomic coordinates, bond lengths, bond angles and thermal parameters have been deposited at the Cambridge Crystallographic Data Centre (CCDC). CCDC-1546188 contains the supplementary crystallographic data for this article. These data can be obtained free of charge via http://www.ccdc.cam.ac.uk/deposit, or from the CCDC, 12 Union Road, Cambridge, CB2 1EZ, UK (fax: C44 1223 336 033; e-mail: deposit@ccdc.cam.ac.uk).

Alignment, phylogenetic, network and inferred ancestral sequences analysis

Multiple alignments of OSC protein sequences were performed, and a codon matrix was produced using the Muscle alignment package in Mega6 (Tamura *et al.*, 2013). The evolutionary history was inferred using the maximum likelihood method based on the JTT matrix-based model. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed. Median-joining network analysis was carried out using DNA Alignment v.1.3.1.1, Network v.4.6.1.1 and Network Publisher v.1.3.0.0 software (Fluxus Technology, Clare, Suffolk, UK). The ancestral sequences at nodes P and S were inferred by ancestral sequence inference in Mega6 using the maximum likelihood method under the general time reversible (GTR) model. The rate and pattern were gamma distributed with invariant sites (G + I).

Molecular evolutionary analysis

The free ratio model of Codeml, implemented within the Paml4 software package (Yang, 2007), was used to estimate the lineage specificity of the non-synonymous to synonymous substitution ratio ω. A branch site analysis, which compared the nearly neutral model with Model A, was performed to test the assumption that the foreground ω value of a specific branch was > 1 at sites at which positive selection appeared to have acted within a specific lineage. The resulting likelihood ratio tests were performed at the 5% level.

Modeling and plasticity residue identification

Three-dimensional models of the OsPS and OsOS proteins were generated by modeling with HsLAS (1W6K) template (Thoma *et al.*, 2004) using Swiss-Model software (Biasini *et al.*, 2014). The three-dimensional structure of parkeol was obtained from https://pubchem.ncbi.nlm.nih.gov/. Docking searches were performed using the Lamarckian genetic algorithm, with a maximum of 25 000 000 energy evaluations and other options set as default in Autodocking tools (Morris *et al.*, 2009). Potential models were returned ranked on the basis of binding energy and the top ranked model was assumed to be the most likely. The models were graphically rendered using Chimera software (https://www.cgl.ucsf.edu/chimera/olddownload.html) (Pettersen *et al.*, 2004). Four amino acid residues were identified within a 5-Å radius of the active site.

Mutagenesis experiments

Mutagenesis was performed using the QuikChange site-directed mutagenesis method (Cat. 200519; Stratagene, Santa Clara, CA, USA). The primers used for site-directed mutagenesis are listed in Table S2 with the substitutions underlined. PCR mix was composed of 0.1 mM of deoxynucleoside triphosphates (dNTPs), 40 ng of plasmid template, 1 × Phusion HF buffer, 0.25 μM of each primer, 2 U of Phusion DNA polymerase (New England Biolabs, Ipswich, MA, USA) and double-distilled H_2O to a final volume of 20 μl. The reaction mixture was denatured at 98°C for 30 s, and then run for 20 cycles of denaturation at 98°C for 10 s, annealing at 60°C for 30 s, polymerization at 72°C for 6 min; a final extension was carried out at 72°C for 30 min. PCR products

were incubated with *Dpn*I at 37°C for 2 h to digest the parental supercoiled DNA. The plasmid DNAs were isolated by a plasmid DNA purification kit (Macherey-Nagel, Düren, Germany), according to the manufacturer's instruction. The mutations were confirmed by DNA sequencing (Eurofins Genomics, Ebersberg, Germany) and the plasmids were subsequently electroporated into the *P. pastoris* wild-type strain X33 and selected for growth on yeast extract peptone dextrose medium with sorbitol (YPDS) plates with 100 μg ml^{-1} of zeocin.

Metabolite extraction and GC-MS analysis

Transformed *P. pastoris* strains were grown at 30°C in glycerol minimal medium (MGY, 1.34% yeast nitrogen base, 1% glycerol, 4×10^{-5}% biotin) in a shaking incubator (250–300 rpm) until the culture reached the log phase (optical density at 600 nm (OD_{600}) = 2–6). Cells were collected by centrifugation and the cell pellet was resuspended to an OD_{600} = 1 in minimal methanol medium (MM, 1.34% yeast nitrogen base, 4×10^{-5}% biotin, 0.5% methanol). The resuspended cells were then incubated at 30°C for 72 h, with the addition of 0.5% methanol every 24 h. Cell pellets were collected from 4 ml of culture and saponified in 1 ml of saponification reagent (20% (w/v) KOH in 50% (v/v) ethanol) with constant shaking at 70°C for 1 h. Water (0.5 ml) was added to the resulting product and the mixture was extracted twice with 1.5 ml of *n*-hexane. Hexane extracts were combined to obtain the crude extract. Aliquots of hexane solution (100 μl) were dried under N_2 and derivatized using 50 μl of trimethylsilyl imidazole/pyridine reagent (Sigma-Aldrich 92718, UK) at 70°C for 30 min. Reaction solutions were diluted with 50 μl of hexane and analyzed using an Agilent 6890 instrument under electronic impact at 70 eV with a Zebron ZB-5 HT capillary column (0.25 mm × 30 m) (Torrance, CA, USA). The oven temperature was initially set at 170°C, raised from 170 to 290°C (6°C min^{-1}), maintained for 4 min, and then elevated from 290 to 340°C (10°C min^{-1}). The relative quantification of the compounds was carried out by comparison with an internal standard (betulin; Sigma-Aldrich B9757, UK).

Immunoblotting

To extract protein, 300 mg of cultured *P. pastoris* cells were suspended in 2 ml of protein extraction buffer (50 mM Tris-Cl, pH 7.5, 150 mM NaCl, 5 mM ethylenediaminetetraacetic acid (EDTA), 10% glycerol, 1% w/v polyvinylpolypyrrolidone (PVPP), 1% Triton-X100, 1 × complete protease inhibitor tablet (Roche)) and lysed three times using a French press at 7756.6 kPa, followed by incubation at 4°C for 1.5 h. The lysates were centrifuged (12 000 ***g***, 12 min, 4°C), and the supernatant was mixed with an equal volume of 4 × sodium dodecylsulfate (SDS) loading buffer and heated at 95°C for 10 min. A 10-μL aliquot of each sample was loaded onto a NuPAGE Novex 4–12% Bis-Tris gel (Invitrogen) and electrophoresed at 150 V for 60 min. A wet transfer cell (Bio-Rad, Hercules, CA, USA) was used to transfer proteins from the gel to the membrane (60 min, 50 V). The membranes were blocked by exposure to 5% w/v powdered skimmed milk in Tris-buffered saline with 0.05% (v/v) Tween 20 (TBST, 10 mM Tris, 100 mM NaCl) overnight at 4°C. The membranes carrying the transferred proteins were incubated with a mouse monoclonal anti-OsPS antibody (Abmart, Shanghai, China) (1 : 250 dilution in TBST and 5% w/v skimmed milk) for 1 h at room temperature, washed in TBST at room temperature (3 × 10 min), incubated with a goat anti-mouse immunoglobulin G–horseradish peroxidase (IgG-HRP)-conjugated secondary antibody (Sigma, UK) diluted 1 : 5000 in TBST and 5% w/v skimmed milk, and rinsed thoroughly. Visualization of the conjugated secondary antibody was performed by supplying SuperSignal® West Dura Extended Duration Substrate (Thermo Scientific, Waltham, MA, USA) for 5 min, followed by exposure to X-ray film.

Results

Characterization of a novel pentacyclic orysatinol synthase from *indica* subspecies

Twelve *OSC* genes were identified from *O. sativa* ssp. *indica* (cv Guangluai 15, GL15) (Fig. S1a) based on the previously annotated 12 *OSC*s from ssp. *japonica* (cv Nipponbare, Nip) (Inagaki *et al.*, 2011). Genes from *indica* and *japonica* are designated with the suffixes '*i*' and '*j*', respectively. The sequences of *OsOSC7* (*Os11g08569*) in GL15 and Nip have 35 non-synonymous and seven synonymous substitutions, giving the highest ratio of K_a/K_s (1.559) among the 12 pairs of *OsOSCs* evaluated (Table S3). Both *OsOSC7i* and *OsOSC7j* are highly expressed in the plumules at the germination stage and in the sheath at the seedling and flowering stages (Fig. S2). The co-linearity and high sequence identity of OSCs between the two subspecies (96%) indicate that *OsOSC7i* is orthologous to *OsOSC7j* (Fig. S1b,c).

OsOSC7j is a parkeol synthase (OsPS) (Ito *et al.*, 2011; Xue *et al.*, 2012). The high K_a/K_s ratio (1.559), combined with the large number of non-synonymous nucleic acid base changes (35 nucleotides), indicates that *Os*OSC7i and OsPS have undergone rapid amino acid sequence divergence, suggesting that *Os*OSC7i may have a different function. GC-MS analysis of metabolites from *P. pastoris* cells expressing OsOSC7i from GL15 revealed one major product, orysatinol (66.7%) (Fig. 1a), and at least 13 minor compounds compared with the empty vector control (Fig. S3a). The structures of three products were elucidated by spectroscopic data (Figs S3b–e, S4a,b). The major product orysatinol is a novel triterpene with an unprecedented chair–(semi-chair)–chair–chair envelope (C-sC-C-C) conformation featuring a *cis*-methyl (C-26 and C-27) orientation at the C–D ring-fused carbons (C-13 and C-14) (Fig. 1; Notes S1, S2). In addition to marnerol (**6**), which has been characterized in a previous study on *Arabidopsis thaliana* PEN5 (Xiong *et al.*, 2006) (Fig. S4a), another minor product, named orysaspirol (**7**), has an extremely unusual seco A ring and spiral B–C ring architecture (Fig. S4b).

These three products can be assigned to different carbocations in the cyclization pathway following the formation of bicyclic C-8 cation I, as proposed in Fig. 1b. Marneral is known to be

formed via Grob fragmentation of the bicyclic C-8 cation (Xiong *et al.*, 2006) (Figs 1b, S4a). Reduction of marneral by endogenous reductases in *P. pastoris* gives marnerol. Orysatinol is likely to be derived from the deprotonation of H-7 of the pentacyclic C8 cation from the proposed cyclization pathway (Fig. 1b). The unprecedented conformation of orysatinol is likely to originate from the formation of the C ring via the attack of the bicyclic C-8 cation by the side-chain C-13 and C-14 double bond from below. The formation of the C-ring then ushers in an entirely new cyclization pathway that yields hitherto unseen triterpene structures, such as pentacyclic orysatinol (**5**) and tetracyclic orysaspirol (**7**). The formation of orysaspirol is likely to proceed via the C ring contraction of the pentacyclic C-8 carbocation, followed by 1,2-hydride and methyl shifts, and terminated by Grob fragmentation, forming an aldehyde form which might be reduced by endogenous reductases in *P. pastoris*, as is the case for marneral. OsOSC7i from GL15 is referred to as orysatinol synthase (OsOS) hereafter.

Metabolite analysis of the sheath extracts of GL15 by GC-MS showed that the detected orysatinol was identical to that of the authentic standard, and the two minor peaks were also minor triterpene products (**8**) and (**9**) of *P. pastoris* cells expressing OsOS (Figs 2, S4c–f). However, parkeol, which has been characterized in *japonica* cv. Zhonghua 11 (ZH11), was not detected in the sheath extract of GL15 (Fig. 2). These results suggest that OsOS has the same catalytic function both *in planta* and when heterologously expressed in *P. pastoris*.

Functional diversity of OsPS and OsOS among *Oryza* species

We were able to identify 11 distinct protein variants (V1–V11) from *OsOSC7* CDSs cloned from two subspecies, *indica* and *japonica*, and only three closely related wild species (AA genome), *O. rufipogon*, *O. nivara* and *O. meridionalis* (Kovach *et al.*, 2007; Sang & Ge, 2007) (34 accessions in total; Table S4; Notes S3). The 11 variants were divided into two separate clades. Clade I consists of 23 sequences from ssp. *indica* and three wild species, *O. rufipogon*, *O. nivara* and *O. meridionalis,* whereas clade II contains 11 sequences that are only present in *japonica* cultivars and a few accessions of *O. rufipogon* (Fig. 3a). Expression of OsOSC7 proteins from clade II in *P. pastoris* showed that V9 did not produce any triterpenes, despite being successfully expressed (Figs 3a, S5, S6). V10 from *O. rufipogon* accession NEPc (Nepal) and V11 from eight cultivars of *japonica,* including ZH11, produced a single product parkeol (Fig. S5), and therefore act as parkeol synthases (Fig. 3a). Interestingly, V8 from accession KHM0225 (*O. rufipogon*, Cambodia) produces a compound that is distinct from parkeol, orysatinol and other minor products of OsOS (Figs 3a, S5). Structural elucidation by one-dimensional (1D) and two-dimensional (2D) NMR analysis established the compound's structure as polypodatetraenol (**10**), an iridal-type bicyclic triterpene, which was characterized in the previous study of yeast $ERG7^{C703I/H}$ mutants (Fig. S7), and is referred to as polypodatetraenol synthase (OrPtS). Polypodatetraenol is likely to be synthesized via a 1,2-hydride and methyl shift, followed by deprotonation of the bicyclic C-8 cation (Chang *et al.*, 2012) (Fig. 1b). In clade I, V1, V2, V3 (OsOS) and V4–V7 derived from *indica* cultivars and those of three wild species, *O. rufipogon*, *O. nivara* and *O. meridionalis*, were shown to be promiscuous OSCs that produce orysatinol as the predominant product with the formation of marnerol, orysaspirol and several other minor, uncharacterized compounds (Figs 3a,b, S5).

OsOS probably represents the ancestral function of OsOSC7, as it was identified in all four AA genomes of *Oryza* species (*O. sativa*, *O. rufipogon*, *O. nivara* and *O. meridionalis*) from a

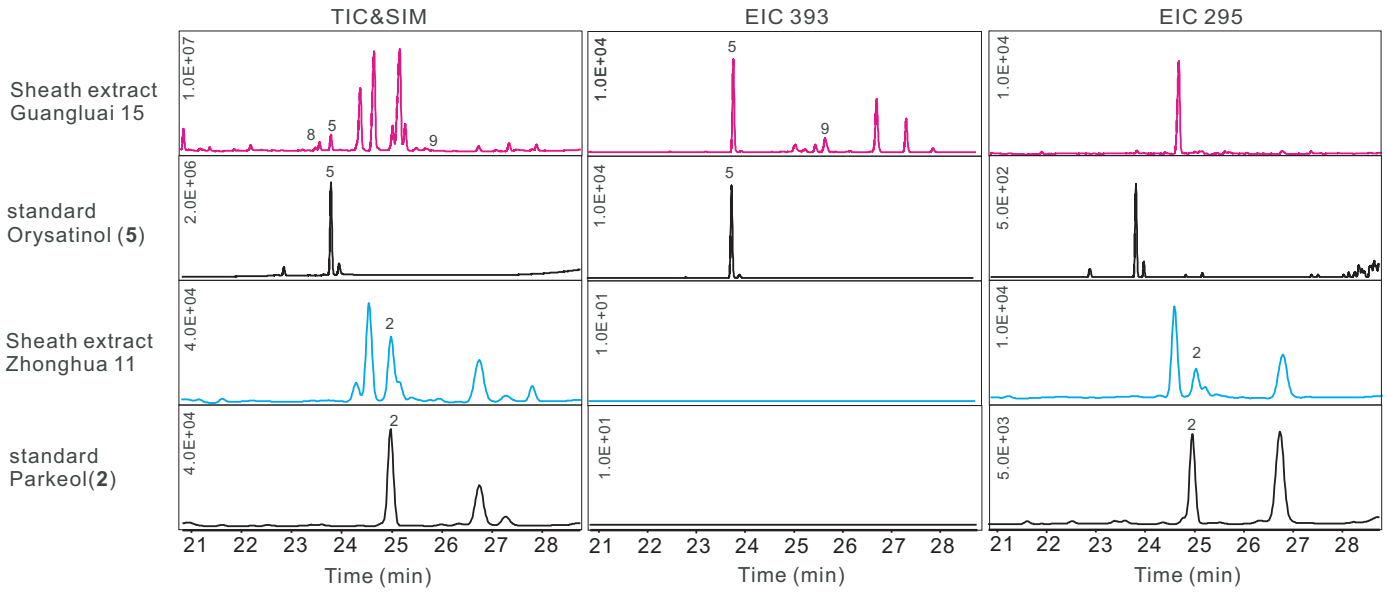

Fig. 2 Identification of parkeol and orysatinol in rice sheath. Gas chromatography-mass spectrometry (GC-MS) analysis of sheath extracts of Guangluai 15 and Zhonghua 11: **2**, parkeol; **5**, orysatinol; **8**, **9**, uncharacterized products. TIC, total ion chromatogram; EIC 393, extracted ion chromatogram at *m/z* 393; EIC 295, extracted ion chromatogram at *m/z* 295; SIM, selected ion (295, 385) model for Zhonghua 11 and parkeol.

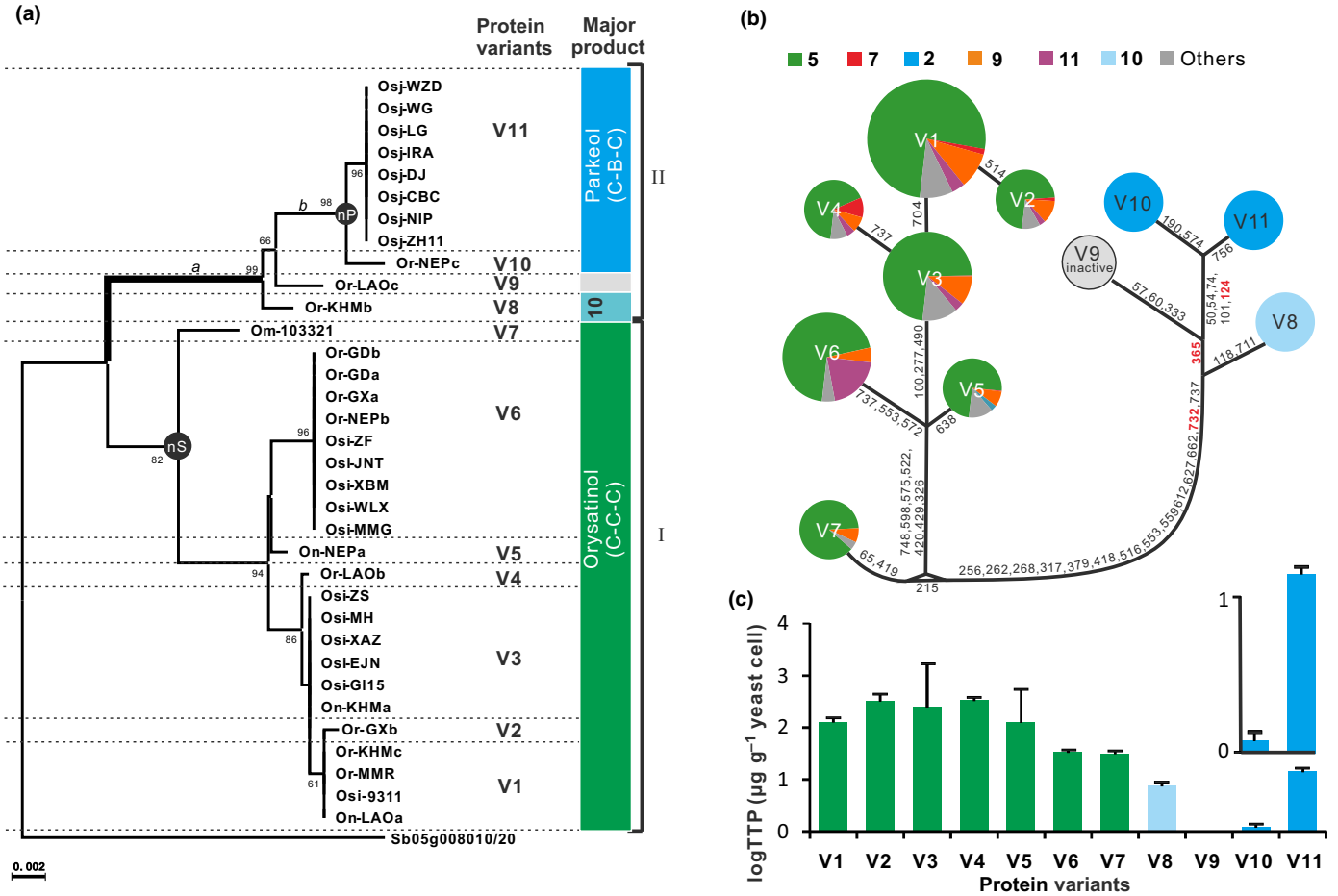

Fig. 3 Phylogenetic, network and functional analyses of OsOSC7 variants. (a) A maximum likelihood (ML) tree of 34 OsOSC7 sequences from different rice cultivars and wild relatives. 2,3-Oxidosqualene cyclase (OSC) amino acid sequences (Supporting Information Notes S3) were aligned using MUSCLE packages with default parameters as implemented in the program MEGA6. Evolutionary distances were computed using the JTT matrix-based method. Evolutionary analyses were conducted in MEGA6 with 1000 bootstrap replicates. The scale bar (bottom, left) indicates 0.002 amino acid substitutions per site. The outgroup sequences used were from *Sorghum bicolor* (http://www.plantgdb.org/SbGDB/, ID Sb05g008010 and Sb05g008020). nP and nS denote nodes P and S for inferred ancestral sequences; *a* and *b* denote branches for branch-site model analysis; bold line, branch under positive selection; C-B-C, chair–boat–chair conformation; C-C-C, chair–chair–chair conformation. (b) Median-joining network and triterpene product profile (%) of 11 distinct OsOSC7 protein variants. Circle size corresponds to the number of species identified that carry particular protein types. Lines represent genetic distance between protein sequences, typically indicating an amino acid difference. **5**, orysatinol; **7**, orysaspirol; **2**, parkeol; **10**, polypodatetraenol; **9**, **11**, uncharacterized products. The numbers on the lines indicate amino acid residues that differ among variants. (c) The quantities of total triterpene products (TTPs) produced by 11 OsOSC7 proteins determined by gas chromatography-mass spectrometry (GC-MS) analysis, using betulin as an internal standard. The amounts of the expressed OsOSC7 proteins were quantified by Western blot analysis. The activities of wild-type and all mutants are presented as means ± SE, $n = 3$ ($\log_{10}$ of micrograms of the produced TTP per gram of yeast cells; n is three different biological replicates). V1–V11 represent distinct protein variants.

wide range of geographic regions (Fig. 3a,b; Table S4). OsOS is a promiscuous enzyme making orysatinol and 12 other products (Fig. 3b), whereas OsPS only produces the specialized single product, parkeol, and was only present in cultivars of *japonica* and its closely related *O. rufipogon*. These accessions are from Cambodia, Laos and Nepal, and are close to the first domesticated area of *japonica* from its wild relative *O. rufipogon* (Huang *et al.*, 2012). OrPtS (V8) in *O. rufipogon* (accession KHM0225, Cambodia) has at least 16 amino acid changes, relative to the ancestral sequence (S) of V1–V7, which resulted in the production of the bicyclic triterpene polypodatetraenol (Fig. 3b). Further, an additional six amino acid changes were responsible for the emergence of ancestral sequence (nP), producing parkeol (C-B-C) (shown in next section) (Fig. 3a, b). Molecular evolutionary analysis of OsOSC7 proteins using PAML software (Yang, 2007) revealed that branch *a*, which leads to parkeol synthase, was under highly significant positive selection ($2\delta = 8.55$, $P < 0.01$) (Table S5). Bayes empirical Bayes analysis suggested #732 as one of the positive sites ($\omega > 1$) having the highest probability (93.6%), whereas, for the seven other sites, the probability ranges from 57.0% to 85.6% (Table S5).

Protein variants V1, V2, V3 (OsOS), V4 and V5 generated products with similar triterpene profiles and with similar orysatinol yields (125–324 μg g^{-1} yeast cells), whereas V6 and V7 produced much smaller amounts of orysatinol (31–33 μg g^{-1} yeast cells) and fewer minor products (Fig. 3b,c), even though the

protein expression levels were higher (Fig. S6). V11 (OsPS) from *japonica* produced 12 times more parkeol (14.6 $\mu g\,g^{-1}$ yeast cells) than V10 of *O. rufipogon* NEPc (Nepal) (1.21 $\mu g\,g^{-1}$ yeast cells); however, both proteins were expressed at a similar level (Figs 3c, S6). These results demonstrate that OsOSC7 variants from cultivated rice and its close relatives are highly divergent as inferred from their product profiles. Not only do they vary in catalytic activity, but also in the conformation of cations *en route* to the formation of the diverse triterpene structures.

Identification of amino acid sites determining product specificity

In total, 46 amino acid variations were identified across OsOSC7 sequences among the 34 rice cultivars or accessions analyzed. There were 21 amino acid differences between the ancestral sequences at nodes nP and nS (MEGA analysis). Expression of the synthetic ancestral gene of *nP* in *P. pastoris* gave parkeol, and that of the *nS* gene produced orysatinol (Fig. S8a), suggesting that the key residues that determine the functional divergence between these two lineages must be among these 21 amino acids (Figs 4a, b, S8b). Three-dimensional structural models of OsOS and OsPS were generated using SWISS-MODEL with the HsLAS structure as a template (Fig. S8b). There are 25 amino acid residues predicted to be within the 5-Å region of the active site. Of these, four sites (#124, #365, #553 and #732) vary between the predicted catalytic regions of OsOS and OsPS (Fig. 4b).

To verify which of these four sites are responsible for the functional specificity of parkeol synthase, *P. pastoris* expression constructs were built in such a way that each combination of residues of OsPS at these four sites was replaced by the corresponding sites from OsOS. In total, all 15 possible OsPS variants were heterologously expressed in *P. pastoris* (Table S6). GC-MS analysis of the resultant *P. pastoris* cell extracts showed that replacement of V553 by alanine had no influence on the product profiles of OsPS (Fig. S9), suggesting that this site does not play an important role in the functional specificity of OsPS. By contrast, single site mutants with substitutions of phenylalanine by leucine ($OsPS^{F124L}$ and $OsPS^{F365L}$) resulted in substantially decreased parkeol production (20.98 ± 0.32 $\mu g\,g^{-1}$ to 4.98 ± 2.40 $\mu g\,g^{-1}$ and 2.62 ± 0.32 $\mu g\,g^{-1}$, respectively) without the formation of additional triterpene products. Mutation of both phenylalanine residues ($OsPS^{F124L/F365L}$) resulted in a complete loss of activity (Figs. 4c, S10; Table S7), although it made stable proteins (Fig. S11). Interestingly, when isoleucine was changed to alanine at position #732, the mutant protein $OsPS^{I732A}$ became a multifunctional enzyme, producing both parkeol (55.93 ± 29.93 $\mu g\,g^{-1}$) and orysatinol (5.76 ± 2.71 $\mu g\,g^{-1}$), in addition to several other triterpenes (20.78% of total triterpene products (TTPs)) (Figs. 4c, S10; Table S7). The $OsPS^{F124L/I732A}$ and $OsPS^{F365L/I732A}$ double mutants yielded dramatically less parkeol, but produced orysatinol and several other triterpenes (Figs 4c, S10; Table S7), indicating that positions #124 and #365 are very important for the maintenance of the parkeol synthase function of OsPS, and position #732 plays a critical role in the catalytic specificity that distinguishes parkeol and orysatinol synthesis. Finally, the triple mutant variant, $OsPS^{F124L/F365L/I732A}$, does not produce any parkeol, and becomes an orysatinol synthase, producing orysatinol as the major product (46.32 ± 8.57 $\mu g\,g^{-1}$, *c.* 61.2% of TTP) and several other triterpenes which are similar to that of the wild-type OsOS (66.7%) (Figs 4c,e, S10; Table S7).

We created all possible OsOS variants by replacement of amino acid residues at each of these three sites, alone or in combination, with the corresponding residues of OsPS, and expressed them in *P. pastoris* (Table S8). The $OsOS^{A732I}$ variant, with substitution of alanine by isoleucine at the first key site (#732), acquired parkeol synthase activity, producing 3.08 ± 0.22 $\mu g\,g^{-1}$ of parkeol (16.8% of TTP) and a small amount of orysatinol (1.59 ± 0.28 $\mu g\,g^{-1}$, 8.7% of TTP), compared with that of wild-type OsOS (112.98 ± 17.55 $\mu g\,g^{-1}$) (Figs 4d, S12; Table S9). However, replacement of leucine with phenylalanine at the second key site (#365) caused decreased orysatinol production (43.7% of the TTP) with concomitant increases in other non-parkeol triterpenes (56.3% TTP). The $OsOS^{L124F}$ and $OsOS^{L124F/L365F}$ mutants also acquired the ability to produce parkeol, whereas the synthesis of orysatinol was reduced in comparison with OsOS. The $OsOS^{L124F/A732I}$ and $OsOS^{L365F/A732I}$ double mutants and the $OsOS^{L124F/L365F/A732I}$ triple mutant were unable to produce orysatinol, but were able to synthesize parkeol (Fig. 4d). Intriguingly, parkeol synthesis was greatly enhanced in the triple mutant, producing up to 17.59 ± 0.17 $\mu g\,g^{-1}$, which is close to that of the wild-type OsPS (20.98 ± 0.32 $\mu g\,g^{-1}$) (Figs 4d,e, S12; Table S9). Our results not only elucidate the role of the three key sites (#732, #365 and #124) in using different conformation cations to synthesize diverse triterpene skeletons, such as orysatinol (C-sC-C) and parkeol (C-B-C), but also provide insights into the process by which OsPS is likely to have evolved from OsOS.

Mechanism for the function conversion between OsPS and OsOS

To understand how these four sites impact on the cyclization processes leading to parkeol and orysatinol formation, further docking of OsPS and OsOS models with the carbocation intermediates was conducted (Figs S13, S14). To our surprise, only F124 is predicted to directly interact with the C-20 positive cation (4.951 Å), potentially providing cation-π stabilization for this intermediate when OsPS is docked with the protosteryl C-20 cation (Table S10; Fig. S13). Although residues A553 and I732 are located within 3–5 Å of the carbocations of several intermediates of OsOS and OsPS models, respectively, these two non-polar and non-aromatic amino acid residues are unlikely to directly interact with any carbocation (Table S10; Figs S13, S14). Residues at site #365 in both OsPS and OsOS models are even further from the carbocations of all docked intermediates (Table S10). However, F/L365 is likely to interfere with Y257 and two other residues (Table S11). Interestingly, the orientation of Y257 in OsPS differs from that in OsOS, and the orientations of Y257 in $OsPS^{F124L/F365L/I732A}$ and $OsOS^{L124F/L365F/A732I}$ are twisted and overlapped with that of OsOS and OsPS, respectively (Fig. 5a,b).

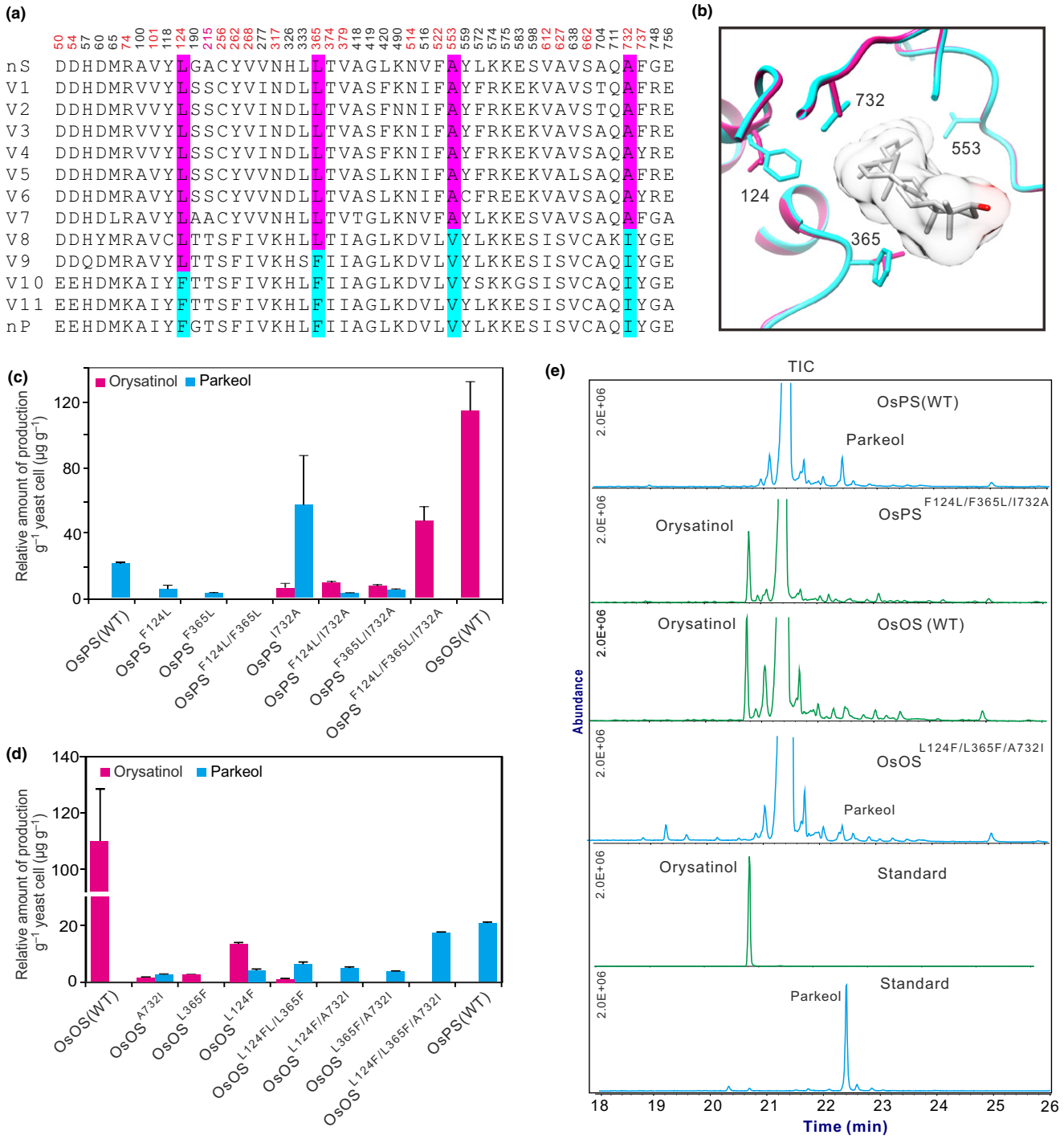

Fig. 4 Identification of amino acids for functional conversion of orysatinol and parkeol synthases in rice. (a) Sequence alignment of 11 *Os*OSC7 protein variants (V1–V11) and inferred ancestral sequences at nodes P and S. nP and nS, nodes P and S for inferred ancestral sequences. The different amino acid residues between nP and nS are shown in red font. Residues predicted to be in the active site are labeled in cyan (OsPS) and pink (OsOS). (b) Local view of the superimposed homology model of rice OsPS (cyan) and OsOS (pink) proteins docked with lanosterol. Numbers represent amino acid residue sites. (c) The quantities of orysatinol and parkeol in wild-type (WT) and mutants from the OsPS background were determined by integrating peak areas from gas chromatography-mass spectrometry (GC-MS) analysis using betulin as an internal standard. The amounts of the expressed OsPS protein were quantified by Western blot analysis. The activities of wild-type and all mutants are presented as means ± SE, $n = 3$ (micrograms of the produced triterpene product per gram of yeast cells, n is three different biological replicates). (d) The quantities of orysatinol and parkeol in WT and mutants from the OsOS background were determined by integrating peak areas from GC-MS analysis using betulin as an internal standard. The amounts of the expressed OsOS protein were quantified by Western blot analysis. The activities of wild-type and all mutants are presented as means ± SE (micrograms of the produced triterpene product per gram of yeast cells, n is three different biological replicates). (e) Total ion chromatogram (TIC) of *Pichia pastoris* expressing wild-type OsPS, OsOS and different mutant variants ($OsOS^{L124F/L365F/A732I}$ and $OsPS^{F124L/F365L/I732A}$). The retention times of parkeol and orysatinol are 22.40 and 20.80 min, respectively.

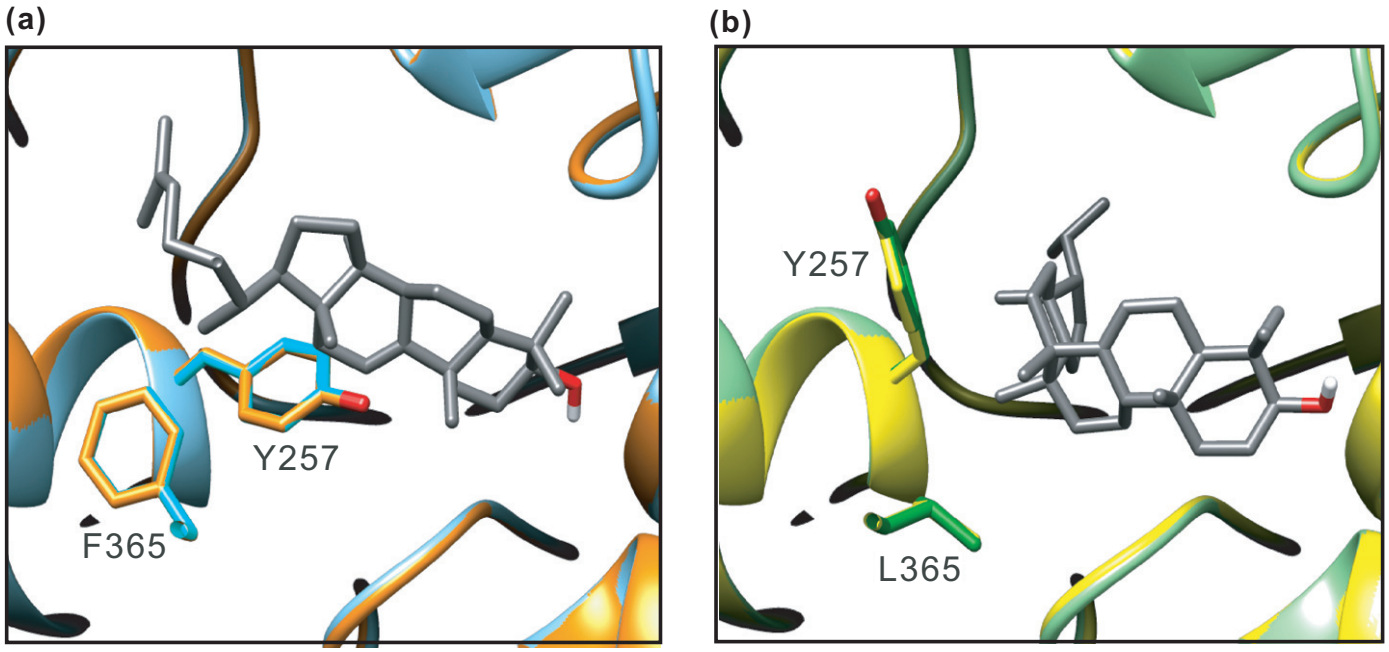

Fig. 5 Orientation and function of Y257 in parkeol and orysatinol synthases from rice. (a) OsPS and $OsOS^{L124F/L365F/A732I}$ (in orange) are docked with parkeol (in grey). The orientation of Y257 in triple mutants of $OsOS^{L124F/L365F/A732I}$ is twisted and consistent with that of OsPS (in cyan). (b) OsOS and $OsPS^{F124L/F365L/I732A}$ (in yellow) are docked with orysatinol (in grey). The orientation of Y257 in triple mutant $OsPS^{F124L/F365L/I732A}$ is twisted, consistent with that of OsOS (in green).

To analyze Y257 function in parkeol and orysatinol synthase, mutations of Y257 with leucine, phenylalanine and alanine were introduced into OsPS, OsOS and the mutant variants $OsPS^{F124L/F365L/I732A}$ and $OsOS^{L124F/L365F/A732I}$. GC-MS analysis of the *P. pastoris* cell extracts showed that Y257 is not a necessary residue for parkeol formation, as most of the OsPS and $OsOS^{L124F/L365F/A732I}$ mutants with substitutions of Y257L, Y257F and Y257A could still produce parkeol, similar to ERG^{H234X} mutants in a previous study (Wu *et al.*, 2005, 2006). By contrast, the same substitutions ($OsOS^{Y257L,F,A}$ and $OsPS^{F124L/F365L/I732A/Y257L,F,A}$) resulted in complete loss of function (Table 1), indicating that Y257 is an essential residue of orysatinol synthase. These results indicate that Y257 is more important in orysatinol formation than in parkeol formation. In addition, mutagenesis of tyrosine with leucine and alanine led to the loss of promiscuity. $OsOS^{L124F/L365F/A732I/Y257L}$ and $OsOS^{L124F/L365F/A732I/Y257A}$ could only produce parkeol (Table 1). Mutagenesis of tyrosine with another aromatic amino acid residue phenylalanine in $OsOS^{L124F/L365F/A732I/Y257F}$ still yielded the production of several minor triterpenes as side-products. This strongly indicates that the π-electrons of Y257 stabilize the intermediary cation(s) through cation-π interaction. $OsPS^{Y257F}$ is unable to produce any other triterpenes, possibly as a result of other unknown factors, such as steric bulk. These results indicate that an aromatic amino acid residue at site #257 is essential for orysatinol formation, and also important for promiscuity.

Table 1 Different function of Y257 in OsPS and OsOS from rice

Genotypes	Substitutions	Orysatinol	Parkeol	Side-products
OsPS	Wild-type		+	
OsOS		+		+
$OsPS^{F124L/F365L/I732A}$		+		+
$OsOS^{L124F/L365F/A732I}$			+	+
OsPS	Y257L		Trace	
OsOS				
$OsPS^{F124L/F365L/I732A}$				
$OsOS^{L124F/L365F/A732I}$			Trace	
OsPS	Y257F			
OsOS				
$OsPS^{F124L/F365L/I732A}$				
$OsOS^{L124F/L365F/A732I}$			+	+
OsPS	Y257A		Trace	
OsOS				
$OsPS^{F124L/F365L/I732A}$				
$OsOS^{L124F/L365F/A732I}$			Trace	

Discussion

In this study, the three key amino acid sites that determine the functional switch between orysatinol and parkeol synthesis were successfully identified through comprehensive phylogenetic analysis and structural modeling. Of relevance, a study on recently diverged rice subspecies-associated diterpene synthase orthologs led to the successful identification of a single site that is the key determinant of diterpene synthase specificity (Xu *et al.*, 2007). As the number of plant whole genome sequences increases (http://www.phytozome.net/), it is becoming possible and worthwhile to conduct functional analysis of metabolic enzyme families in multiple species from particular genera or families. The use of evolutionary information combined with protein homology modeling is an effective approach to identify new enzymes and unveil the catalytic mechanisms used by enzymes to create chemical

diversity (Weng, 2014). In this study, such an approach enabled us to identify the new triterpenes, orysatinol and orysaspirol, and allowed us to shed light on previously unknown mechanisms of OSC cyclization. Furthermore, our investigations reveal the stepwise evolution of OsPS from a multifunctional OsOS via the intermediate enzyme, polypodatetraenol synthase, by sequential substitutions of three amino acids at #732, #365 and #124.

Recently, a study of bacterial OSCs identified four residues, at positions W230, H232, Y503 and N697 (positions of HsLAS), that determine the functional interconversion of lansterol and isoarborinol synthases. These enzymes produce tetracyclic and pentacyclic triterpenes from the same protosteryl-type C-B-C conformation (Banta *et al.*, 2016). Studies (Ito *et al.*, 2013; Salmon *et al.*, 2016) on *Euphorbia tirucalli* and *Avena strigosa* β-amyrin synthase (SAD1) have identified two conserved residues at positions F696 and S699 (positions of HsLAS) which, when mutated, result in a change in product from the pentacyclic triterpenoid β-amyrin to tetracyclic cyclization products, but maintain a dammarane-type C-C-C conformation. The three key residue positions, #124, #365 and #732, in OsPS and OsOS, corresponding to L104, I335 and I702 in HsLAS, that have been identified here were not reported in the previous studies (Ito *et al.*, 2013; Banta *et al.*, 2016; Salmon *et al.*, 2016). This may be because these residues are responsible for the alternation between C-B-C and C-sC-C conformations, whereas the residues identified in the previous studies (Ito *et al.*, 2013; Banta *et al.*, 2016; Salmon *et al.*, 2016) are involved in the formation of the E ring without conformational change. More studies are required in order to gain further insights into the mechanisms of OSC cyclization and product specificity.

Cultivated rice (*O. sativa*) and its closest wild relative (*O. rufipogon*) have a broad geographical distribution, with adaptations to different ecological and agronomic conditions. We hypothesize that multifunctional enzymes, such as OsOS, have the ability to quickly create chemical diversity to enhance the survival of species in the challenging environments. Our analysis revealed a fast evolutionary process from OsOS to OsPS, which is normally caused by natural selection under biotic stress (Bergelson *et al.*, 2001). The OsPS-derived pathway may play a role in resistance to insects and/or pathogens. By contrast, parkeol, the product of OsPS, has a similar structure to cycloartenol and lanosterol (Fig. 1a), the precursors of sterol biosynthesis. The product of OsPS may thus be involved in the biosynthesis of a new steroid hormone, and hence influence plant growth and development (Bishop & Koncz, 2002; Ohyama *et al.*, 2009).

Acknowledgements

We thank Yanan Wang for NMR analysis in the Capital University of Medicine, Beijing, China. We thank Song Ge (Institution of Botany, Chinese Academy of Science (IBCAS)) for providing wild rice materials. Support for this work in X.Q.'s laboratory was provided by the National Natural Science Foundation of China (NSFC grant nos. 31370337 and 31530050), National Basic Research Program of China (973 Program grant no. 2013CB127000) and the China Scholarship Council (CSC). Research in A.O.'s laboratory was supported by the Centre of Excellence for Plant and Microbial Sciences (CEPAMS), established between the John Innes Centre and the Chinese Academy of Sciences, and funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Chinese Academy of Sciences (CAS) (grant no. CPM11; Z.X.); European Union grant KBBE-2013-7 (TriForC) (R.B.T.); the Joint Engineering and Physical Sciences Research Council/BBSRC-funded OpenPlant Synthetic Biology Research Centre grant BB/L014130/1 (M.J.S.); the BBSRC Institute Strategic Programme Grant 'Molecules from Nature' (BB/P012523/1); and the John Innes Foundation (A.O.).

Author contributions

Z.X., X.Q., R.B.T. and A.O. designed the research. Z.T. and Y.Z. performed the functional analysis of OsOSC7 orthologs in *Pichia*. X.W. and M.J.S. carried out the structural analysis. Z.X. and J.S. performed the evolutionary analysis mutagenesis experiments and homology modeling. Z.X., A.H. and R.B.T. analyzed the data. Z.X., R.B.T., A.H., A.O. and X.Q. wrote the paper. Z.X. and Z.T. contributed equally to this work.

ORCID

Zheyong Xue http://orcid.org/0000-0002-7203-9794
Xiaoquan Qi http://orcid.org/0000-0001-7175-115X

References

Augustin JM, Kuzina V, Andersen SB, Bak S. 2011. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**: 435–457.

Banta AB, Wei JH, Gill CC, Giner JL, Welander PV. 2016. Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proceedings of the National Academy of Sciences, USA* **114**: 245–250.

Benveniste P. 1986. Sterol biosynthesis. *Annual Review of Plant Physiology* **37**: 275–308.

Bergelson J, Kreitman M, Stahl EA, Tian D. 2001. Evolutionary dynamics of plant R-genes. *Science* **292**: 2281–2285.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L *et al.* 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**: W252–W258.

Bishop GJ, Koncz C. 2002. Brassinosteroids and plant steroid hormone signaling. *Plant Cell* **14**: S97–S110.

Bloch K. 1965. The biological synthesis of cholesterol. *Vitamins and Hormones* **15**: 119–150.

Chang CH, Chen YC, Tseng SW, Liu YT, Wen HY, Li WH, Huang CY, Ko CY, Wang TT, Wu TK. 2012. The cysteine 703 to isoleucine or histidine mutation of the oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae* generates an iridal-type triterpenoid. *Biochimie* **94**: 2376–2381.

Chappell J. 2002. The genetics and molecular genetics of terpene and sterol origami. *Current Opinion in Plant Biology* **5**: 151–157.

Christianson DW. 2006. Structural biology and chemistry of the terpenoid cyclases. *Chemical Reviews* **106**: 3412–3442.

Hart EA, Ling H, Darr LB, Wilson WK, Pang JA, Matsuda SPT. 1999. Directed evolution to investigate steric control of enzymatic oxidosqualene cyclization. An isoleucine-to-valine mutation in cycloartenol synthase allows

lanosterol and parkeol biosynthesis. *Journal of the American Chemical Society* **121**: 9887–9888.

Herrera JBR, Wilson WK, Matsuda SPT. 2000. A tyrosine-to-threonine mutation converts cycloartenol synthase to an oxidosqualene cyclase that forms lanosterol as its major product. *Journal of the American Chemical Society* **122**: 6765–6766.

Huang XH, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W *et al.* 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**: 497–501.

Inagaki YS, Etherington G, Geisler K, Field B, Dokarry M, Ikeda K, Mutsukado Y, Dicks J, Osbourn A. 2011. Investigation of the potential for triterpene synthesis in rice through genome mining and metabolic engineering. *New Phytologist* **191**: 432–448.

Ito R, Hashimoto I, Masukawa Y, Hoshino T. 2013. Effect of cation-π interactions and steric bulk on the catalytic action of oxidosqualene cyclase: a case study of phe 728 of β-amyrin synthase from *Euphorbia tirucalli* L. *Chemistry – A European Journal* **19**: 17150–17158.

Ito R, Mori K, Hashimoto I, Nakano C, Sato T, Hoshino T. 2011. Triterpene cyclases from *Oryza sativa* L.: cycloartenol, parkeol and achilleol B synthases. *Organic Letters* **13**: 2678–2681.

Kovach MJ, Sweeney MT, McCouch SR. 2007. New insights into the history of rice domestication. *Trends in Genetics* **23**: 578–587.

Matsuda SPT, Darr LB, Hart EA, Herrera JBR, McCann KE, Meyer MM, Pang J, Schepmann HG. 2000. Steric bulk at cycloartenol synthase position 481 influences cyclization and deprotonation. *Organic Letters* **2**: 2261–2263.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**: 2785–2791.

Moses T, Papadopoulou KK, Osbourn A. 2014. Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Critical Reviews in Biochemistry and Molecular Biology* **49**: 439–462.

Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T. 2009. Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **106**: 725–730.

Osbourn A, Goss RJM, Field RA. 2011. The saponins – polar isoprenoids with important and diverse biological activities. *Natural Products Reports* **28**: 1261–1268.

Parks LW, Casey WM. 1995. Physiological implications of sterol biosynthesis in Yeast. *Annual Review of Microbiology* **49**: 95–116.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**: 1605–1612.

Salmon M, Thimmappaa RB, Minto RE, Melton RE, Hughes RK, O'Maille PE, Hemmings AM, Osbourn A. 2016. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. *Proceedings of the National Academy of Sciences, USA* **113**: E4407–E4414.

Sang T, Ge S. 2007. The puzzle of rice domestication. *Journal of Integrative Plant Biology* **49**: 760–768.

Sun JC, Xu X, Xue ZY, Snyder JH, Qi XQ. 2013. Functional analysis of a rice oxidosqualene cyclase through total gene synthesis. *Molecular Plant* **6**: 1726–1729.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. 2014. Triterpene biosynthesis in plants. *Annual Review of Plant Biology* **65**: 225–257.

Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, Dehmlow H, Hennig M, Stihle M, Ruf A. 2004. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **432**: 118–122.

Weng JK. 2014. The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist* **201**: 1141–1149.

Wu TK, Chang CH, Liu YT, Wang TT. 2008. *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase: a chemistry–biology interdisciplinary study of the protein's structure-function-reaction mechanism relationships. *Chemical Record* **8**: 302–325.

Wu TK, Griffin JH. 2002. Conversion of a plant oxidosqualene-cycloartenol synthase to an oxidosqualene-lanosterol cyclase by random mutagenesis. *Biochemistry* **41**: 8238–8244.

Wu TK, Liu YT, Chang CH. 2005. Histidine residue at position 234 of oxidosqualene-lanosterol cyclase from Saccharomyces cerevisiae simultaneously influences cyclization, rearrangement, and deprotonation reactions. *ChemBioChem* **6**: 1177–1181.

Wu TK, Liu YT, Chang CH, Yu MT, Wang HJ. 2006. Site-saturated mutagenesis of histidine 234 of *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase demonstrates dual functions in cyclization and rearrangement reactions. *Journal of the American Chemical Society* **128**: 6414–6419.

Xiong QB, Wilson WK, Matsuda SPT. 2006. An *Arabidopsis* oxidosqualene cyclase catalyzes iridal skeleton formation by Grob fragmentation. *Angewandte Chemie. International Edition* **45**: 1285–1288.

Xu MM, Wilderman PR, Peters RJ. 2007. Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proceedings of the National Academy of Sciences, USA* **104**: 7397–7401.

Xu R, Fazio GC, Matsuda SPT. 2004. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**: 261–291.

Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, ÓMáille P, Osbourn A, Qi X. 2012. Divergent evolution of oxidosqualene cyclases in plants. *New Phytologist* **193**: 1022–1038.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Identification of orthologous gene *OsOSC7i* from *Oryza sativa* L. ssp. *indica*.

Fig. S2 Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis of *OsOSC7j* and *OsOSC7i*.

Fig. S3 Functional analysis of OsOS in *Pichia pastoris*.

Fig. S4 Structural identification of side-products.

Fig. S5 Functional analysis of OsOSC7 variants in *Pichia pastoris*.

Fig. S6 Expression level of OsOSC7 variants in *Pichia pastoris*.

Fig. S7 Structural identification of V8/OrPtS variant product.

Fig. S8 Functional analysis and key residue identification of ancestral sequences of nP and nS.

Fig. S9 Molecular docking of intermediates of parkeol with the OsPS model.

Fig. S10 Molecular docking of intermediates of orysatinol with the OsOS model.

Fig. S11 The role of substitution at position #553.

Fig. S12 GC-MS spectra of extracts from *Pichia pastoris* expressing OsPS mutants.

Fig. S13 Expression level of the OsPS mutants in *Pichia pastoris*

Fig. S14 Gas chromatography-mass spectrometry (GC-MS) spectrum of extracts of *Pichia pastoris* expressing OsOS mutants.

Table S1 List of essential 2,3-oxidosqualene cyclase (OSC) or squalene-hopene cyclase residues for cyclization of triterpenes

Table S2 List of primer sequences

Table S3 K_a/K_s analysis of 2,3-oxidosqualene cyclase (OSC) gene family in two subspecies *japonica* and *indica*

Table S4 List of rice accessions for cloning of *OsOSC7* genes

Table S5 Summary of statistics for detection of positive selection for evolution of the parkeol synthase

Table S6 OsPS mutants with substitutions by OsOS amino acids at four divergent sites

Table S7 Products ($\mu g\, g^{-1}$) of OsPS mutants with single, double and triple substitutions at sites #124, #365 and #732 by OsOS amino acids

Table S8 OsOS mutants with substitutions by OsPS amino acids at three divergent sites

Table S9 Products ($\mu g\, g^{-1}$) of OsOS mutants with single, double and triple substitutions at sites #124, #365 and #732 by OsPS amino acids

Table S10 Distance between intermediate carbocations and their adjacent residues

Table S11 Residues interfered by four key sites

Notes S1 Crystal X-ray diffraction experiment for orysatinol (**5**).

Notes S2 Nuclear magnetic resonance (NMR) spectra for orysatinol (**5**), marnerol (**6**), orysaspirol (**7**) and polypodatetraenol (**10**).

Notes S3 Sequence alignments of OsOSC7 variants.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.