Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Functional and evolutionary analysis of *DXL1*, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in *Arabidopsis thaliana*

Lorenzo Carretero-Paulet ^{a,b,c,*}, Albert Cairó ^{a,1}, David Talavera ^{a,2}, Andreu Saura ^a, Santiago Imperial ^a, Manuel Rodríguez-Concepción ^{b,3}, Narciso Campos ^{a,b,1,3}, Albert Boronat ^{a,b,**}

^a Departament de Bioquímica i Biologia Molecular, Facultat de Biologia, Universitat de Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain

^b Department of Molecular Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Edifici CRAG, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain

^c Departamento de Biología Aplicada, Área de Genética, Universidad de Almería, La Cañada de San Urbano s/n, 04120 Almería, Spain

ARTICLE INFO

Article history:

Accepted 26 October 2012

Available online xxx

Keywords:

Gene duplication

Functional specialization

DXS

Arabidopsis thaliana

Molecular evolution

Secondary metabolism

ABSTRACT

The synthesis of 1-deoxy-D-xylulose 5-phosphate (DXP), catalyzed by the enzyme DXP synthase (DXS), represents a key regulatory step of the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway for isoprenoid biosynthesis. In plants DXS is encoded by small multigene families that can be classified into, at least, three specialized subfamilies. *Arabidopsis thaliana* contains three genes encoding proteins with similarity to DXS, including the well-known *DXS1/CLA1* gene, which clusters within subfamily I. The remaining proteins, initially named DXS2 and DXS3, have not yet been characterized. Here we report the expression and functional analysis of *A. thaliana* DXS2. Unexpectedly, the expression of DXS2 failed to rescue *Escherichia coli* and *A. thaliana* mutants defective in DXS activity. Coherently, we found that DXS activity was negligible *in vitro*, being renamed as DXL1 following recent nomenclature recommendation. DXL1 is targeted to plastids as DXS1, but shows a distinct expression pattern. The phenotypic analysis of a *DXL1* defective mutant revealed that the function of the encoded protein is not essential for growth and development. Evolutionary analyses indicated that *DXL1* emerged from *DXS1* through a recent duplication apparently specific of the Brassicaceae lineage. Divergent selective constraints would have affected a significant fraction of sites after diversification of the paralogues. Furthermore, amino acids subjected to divergent selection and likely critical for functional divergence through the acquisition of a novel, although not yet known, biochemical function, were identified. Our results provide with the first evidences of functional specialization at both the regulatory and biochemical level within the plant DXS family.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Isoprenoids represent a wide family of structurally and functionally diverse compounds. Some of them participate in essential processes for plant biology, ranging from photosynthesis (chlorophylls and carotenoids) or electron transport (plastoquinones and ubiquinones) to developmental regulation (phytohormones like brassinosteroids,

cytokinins, gibberellins and abscisic acid). However, most plant isoprenoids are synthesized as secondary metabolites influencing different aspects of plant-ecological interactions, including protection against herbivores and pathogens, attraction of pollinators and seed-dispersing animals and allelochemical modulation of ecological competition with other plant species (Croteau et al., 2000; Estevez et al., 2000; Phillips et al., 2008; Rodríguez-Concepción and Boronat,

Abbreviations: aLRT, approximate likelihood ratio tests; Ap, ampicillin; BEB, Bayes Empirical Bayes; cDNA, DNA complementary to RNA; Cm, chloramphenicol; DMAPP, dimethylallyl diphosphate; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXSP, 1-deoxy-D-xylulose 5-phosphate; ESTs, expressed sequence tags; GFP, green fluorescent protein; GUS, beta-glucuronidase; IPP, isopentenyl diphosphate; IPTG, isopropyl β-D-thiogalactopyranoside; LB, Luria-Bertani medium; LRT, likelihood ratio tests; MEP, 2-C-methyl-D-erythritol 4-phosphate; ML, maximum likelihood; MS, Murashige and Skoog medium; PCR, polymerase chain reaction; MVA, mevalonate; PP, posterior probability; RT, Reverse transcription; UTR, Untranslated Transcribed Region.

* Correspondence to: L. Carretero-Paulet, Department of Biological Sciences, SUNY-University at Buffalo (North Campus), 109 Cooke Hall, Buffalo, NY 14260. Tel. + 1 7166454984.

** Correspondence to: A. Boronat, Departament de Bioquímica i Biologia Molecular, Facultat de Biologia, Universitat de Barcelona, Avda. Diagonal 643, 08028 Barcelona, Spain. Tel.: + 34 934021194, + 34 935636600; fax: + 34 934021559, + 34 935636601.

E-mail addresses: lorenzoc@buffalo.edu (L. Carretero-Paulet), albertcairo@yahoo.com (A. Cairó), David.Talavera@manchester.ac.uk (D. Talavera), cyb_nu@hotmail.com (A. Saura), simperial@ub.edu (S. Imperial), manuel.rodriguez@cragenomics.es (M. Rodríguez-Concepción), ncampos@ub.edu (N. Campos), aboronat@ub.edu (A. Boronat).

¹ Tel.: + 34 934021194; fax: + 34 934021559.

² Current address: Faculty of Life Sciences, Michael Smith Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK.

³ Tel.: + 34 935636600; fax: + 34 935636601.

0378-1119/\$ – see front matter © 2012 Elsevier B.V. All rights reserved.

<http://dx.doi.org/10.1016/j.gene.2012.10.071>

Please cite this article as: Carretero-Paulet, L., et al., Functional and evolutionary analysis of *DXL1*, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in *Arabidopsis thaliana*, Gene (2012), <http://dx.doi.org/10.1016/j.gene.2012.10.071>

- 2002). Despite their diversity in structures and functions, all isoprenoids are derived from the common precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Until recently, it was assumed that IPP and DMAPP were solely synthesized through the mevalonic acid (MVA) pathway (Chappell, 1995). However, it is now well established that they can also be produced by the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway (Lichtenthaler, 1999; Rodriguez-Concepcion and Boronat, 2002).
- Most isoprenoids synthesized by plants are tissue-, developmental stage- and/or species-specific, or are specifically produced in an adaptative manner in response to different environmental constraints. The great biochemical flexibility required for isoprenoid biosynthesis is primarily acquired by plants through compartmentalization of IPP/DMAPP and isoprenoid production in different subcellular locations. Plastidial IPP and DMAPP are synthesized through the MEP pathway, whereas cytosolic IPP is synthesized through the MVA pathway. Mitochondrial isoprenoids are synthesized from MVA-derived IPP that is imported from the cytosol. Furthermore, a cross-talk between the different isoprenoid biosynthetic pathways acting in distinct cell compartments is well documented (Botella-Pavia et al., 2004; Hemmerlin et al., 2003; Kasahara et al., 2002; Laule et al., 2003; Rodriguez-Concepcion, 2006; Skorupinska-Tudek et al., 2008).
- Another mechanism known to provide plasticity to plant metabolic responses is provided by functional diversification of specific enzymes. Enzymes involved in secondary metabolic pathways are commonly encoded by multigene families composed of duplicated genes maintained in plant genomes during evolution through the acquisition of specialized functions (Barkman, 2003; Benderoth et al., 2006; Des Marais and Rausher, 2008; Durbin et al., 2000; Fucile et al., 2008; Li et al., 2009; Yang et al., 2002). An interesting example of functional specialization in isoprenoid metabolic pathways is provided by 1-deoxy-D-xylulose 5-phosphate synthase (DXS), the thiamine pyrophosphate (TPP) dependent enzyme that catalyzes the first step in the MEP pathway. DXS catalyzes the condensation of (hydroxyethyl)thiamin, derived from the decarboxylation of pyruvate, on the carbonyl group of D-glyceraldehyde 3-phosphate (GAP) to yield 1-deoxy-D-xylulose 5-phosphate (Lois et al., 1998). It is widely accepted that DXS plays a key regulatory role in the biosynthesis of plastidial isoprenoids (for a recent review see Cordoba et al., 2009).
- Evidences of functional diversification within DXS are provided by the occurrence of at least three functionally specialized types of DXS: type I, type II and type III (Cordoba et al., 2011; Kim et al., 2009; Walter et al., 2002). It has been proposed that type I DXS would be specifically involved in the synthesis of essential isoprenoids (such as photosynthetic pigments and phytohormones) in chloroplasts. In turn, type II DXS is thought to be located in non-photosynthetic plastids where they would be involved in the synthesis of specific isoprenoids as secondary metabolites. In addition, it has been suggested that type III DXS might participate in the synthesis of essential isoprenoids required at low levels, such as phytohormones (Cordoba et al., 2011). Functional specialization has only been verified in gene regulatory regions, outside gene regions coding for mature functional proteins, as DXS activity has been shown for members of the three subfamilies (Bouvier et al., 1998; Chahed et al., 2000; Cordoba et al., 2011; Estevez et al., 2000; Kim et al., 2009; Lois et al., 2000; Phillips et al., 2007; Walter et al., 2002).
- From the three DXS-annotated sequences encoded by the *Arabidopsis thaliana* genome, AT4G15560, AT3G21500 and AT5G11380, only AT4G15560, corresponding to the *DXS1/CLA1* gene had been characterized at the molecular and functional level (Estevez et al., 2000). Here we report the expression analysis of the *A. thaliana* *DXS2* gene (AT3G21500) and the targeting of the protein to the plastids. Unexpectedly, we have found that *DXS2* does not have DXS activity. We also present a phylogenetic, molecular evolutionary and functional divergence analysis about the origin and the evolutionary fate of the *DXS2* gene. Our results support *DXS2* arising from a recent gene duplication event within subfamily I DXS of Brassicaceae and evolving through the possible acquisition of a novel, although not yet known, biochemical function(s) through changes in selective constraints affecting specific amino acids.
- ## 2. Materials and methods
- ### 2.1. Plant material and growth conditions
- Biological material used in this work was derived from *A. thaliana* plants of ecotype Col 0, and the mutant lines *chs5* and SALK_N612762. Seeds from the mutant lines were kindly provided by the Arabidopsis Biological Resource Center at the Ohio State University. The *chs5* mutant (stock number CS 8004) had been previously characterized (Araki et al., 2000). Seeds were surface-sterilized and germinated on Petri dishes containing solid Murashige and Skoog (MS) medium as described (Carretero-Paulet et al., 2002). Unless otherwise stated growth conditions, under axenic and non-axenic conditions, were as described in a previous work (Carretero-Paulet et al., 2002). *A. thaliana* cell suspension line T87 was cultured as described (Axelos et al., 1992).
- ### 2.2. Nucleic acid purification and gene expression analysis
- RNA was extracted from green siliques, inflorescences, leaves, stems, roots, T87 cells, light-grown and etiolated *A. thaliana* seedlings using the RNeasy kit (QIAGEN, Hilden, Germany), according to the instructions provided by the supplier. The full-length *A. thaliana* *DXS2* cDNA was amplified by RT-PCR using *DXS2.S1* (see Supplementary file 1 for a list of primers used in this study) and *DXS2.S4* as primers. cDNA was generated from total RNA obtained from inflorescences and young siliques using the RETROscript kit (Ambion, Austin, TX) according to the manufacturer's instructions. The corresponding PCR product was cloned into pBLUESCRIPT SK- (obtaining plasmid pLB.DXS2) and further sequenced.
- Analysis of *DXS2* expression was performed through semiquantitative RT-PCR using the QuantumRNA kit (Ambion) using 18S ribosomal RNA as internal standard. *DXS2* transcript was specifically amplified using *DXS2.S1* and *DXS2.S6* primers, whereas 18S rRNA was amplified using the Universal primers and competitors provided in the kit mixed in a ratio of 2:8.
- ### 2.3. Complementation assays in *Escherichia coli*
- A. thaliana* *DXS2* cDNA was amplified from pLB.DXS2 in four different PCR reactions to obtain: i) a cDNA fragment coding for the full-length *DXS2* sequence (*DXS2*), ii) a cDNA fragment encoding a *DXS2* variant lacking the first 33 amino acids corresponding to the predicted plastidial transit peptide (*DXS2C*), iii) a cDNA fragment encoding a *DXS2* variant lacking the N-terminal 44 amino acids and thus only including the homology region shared with DXS prokaryotic counterparts (*DXS2CC*) and iv) a cDNA encoding a *DXS2CC* variant containing an insertion of the region extending from residues 267 to 309 of *A. thaliana* *DXS1* integrated within the corresponding positions of *DXS2CC* (*DXS2CCdom*). *DXS2*, *DXS2C* and *DXS2CC* cDNA fragments were amplified using oligonucleotides *DXS2.NheI*, *DXS2.CNheI* and *DXS2.CCNheI* as forward primers, respectively, and the universal T7 primer as reverse primer. The resulting cDNA fragments were digested with *NheI* and *EcoRI* and cloned into the expression plasmid pBADM1 (Carretero-Paulet et al., 2002), previously digested with the same restriction enzymes to generate plasmids pLBAD.DXS2, pLBAD.DXS2C and pLBAD.DXS2CC. The plasmid pLBAD.DXS2CCdom was generated following a site-directed mutagenesis strategy based on inverse PCR using pLBAD.DXS2CC as a template and primers *DXS2.domA* and *DXS2.domB*. In these plasmids, the cloned cDNAs are under the control of the pBAD promoter, which can be induced with L-arabinose.
- Complementation assays were performed by transforming the *E. coli* *dxs*-deficient strain EcAB4-2 with plasmids pLBAD.DXS2, pLBAD.DXS2C, pLBAD.DXS2CC and pLBAD.DXS2CCdom. In strain EcAB4-2 the *dxs* gene

is disrupted by the insertion of the CAT gene, encoding chloramphenicol acetyltransferase. This strain contains a copy of the MVA+ operon (Campos et al., 2001) stably integrated in the chromosome, which enables synthesis of IPP from mevalonate supplied exogenously in the growth medium. The pTAC.DXS1 plasmid, expressing the full-length *A. thaliana* DXS1 cDNA under the control of the pTAC promoter (Browner et al., 1991), was used as a positive control.

2.4. Heterologous expression of DXS1 and DXS2 in *E. coli* and enzyme assays

The DXS1 cDNA sequence encoding the mature form of the enzyme (lacking the first 58 residues of the N-terminal region) was amplified by PCR (Pfu DNA polymerase) using primers DXS1.NheI and DXS1.HindIII to create a NheI site at the 5' end and a HindIII site at the 3' end. The resulting amplicon was subcloned into pGFPuv (Clontech) in the SmaI site and then subsequently cloned into the NheI/HindIII sites of pET23b(+) (Novagen) in frame with the C-terminal 6xHis tag. An equivalent strategy was used for DXS2. The sequence encoding the predicted mature form of the protein (lacking the first 44 amino acid sequence of the N-terminal region) was amplified by PCR (Pfu DNA polymerase) using primers DXS2.NheI and DXS2.XhoI to create a NheI site and a XhoI site at the 5' and 3' ends, respectively. The PCR product was first cloned into the SmaI site of pBluescript SK+, and then into the NheI/HindIII sites of pET23b(+) in frame with the C-terminal 6xHis tag. The generated constructs (pET23-DXS1 and pET23-DXS2) were confirmed by DNA sequencing.

The pET23-DXS1 and pET23-DXS2 plasmids were used to transform competent *E. coli* BL21 (DE3) pLysS cells. Bacterial cells transformed with pET23-DXS1 were grown up to an OD₆₀₀ of 0.4–0.5 at 37 °C in LB medium with Ap (100 µg/ml) and Cm (34 µg/ml). Cells transformed with pET23-DXS2 were grown at 22 °C under the same conditions. For optimal DXS1 protein expression, the bacterial culture was induced with 0.5 mM IPTG and incubated for 2 h at 37 °C. For DXS2 expression, the bacterial culture was induced with 0.01 mM IPTG and incubated for 16 h at 22 °C. Cells were harvested by centrifugation at 6000 rpm (4 °C), resuspended in 20 ml of lysis buffer (40 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mg/ml lysozyme, 10 mM imidazole, 0.5 mM EDTA) supplemented with a cocktail of protease inhibitors EDTA-free (Roche), and lysed by sonication (8 × 60 s, 40–60 W). The lysate was incubated 8:1 with 1% (w/v) protamine sulfate, and the soluble fraction was obtained by centrifugation at 15,000 rpm for 30 min at 4 °C. Each recombinant enzyme was purified by immobilized metal affinity chelating chromatography as follows: briefly, the supernatant of the previous step was passed through a 0.2 µm filter and applied on to a HiTrap Chelating FF column (GE Healthcare) equilibrated in 40 mM Tris–HCl buffer, pH 8.0 containing 4 mM 2-mercaptoethanol, 100 mM NaCl, Na₂S₂O₄ 0.04% (w/v) and 10 mM imidazole. The column was washed with the same buffer and eluted using an imidazole gradient (10–500 mM) in equilibration buffer. Fractions were monitored for protein content by measuring absorbance at 280 nm and analyzed by SDS-polyacrylamide gel electrophoresis and Coomassie brilliant blue R-250 staining. Fractions corresponding to the DXS1 and DXS2 recombinant enzymes showed a band with an apparent molecular mass of 70 kDa and 60 kDa, respectively. These fractions were pooled and assayed for protein using the Bradford microassay with bovine serum albumin as standard. Finally, glycerol was added up to a 50% (v/v) and enzyme preparations were stored at –20 °C until use.

DXS enzymatic activity was determined by a spectrophotometrical assay using purified 1-deoxy-d-xylulose 5-phosphate reductoisomerase (DXR) from *Synechocystis* sp PCC6803 as coupled enzyme (Altincicek et al., 2000). Recombinant *Synechocystis* DXR was obtained as described by Yin and Proteau (2003) with minor modifications. The standard reactions were done in 100 mM Tris–HCl buffer (pH 7.8) containing 1 mM TPP, 1 mM MgCl₂, 1 mM MnCl₂, 1 mM DTT, 0.15 mM NADPH, 1 mM pyruvate, 6 µg DXR and the assayed sample of DXS in a final volume of

0.2 ml. The reaction was initiated by adding 1 mM DL-glyceraldehyde 3-phosphate. Initial reaction velocities were determined following the oxidation of NADPH by monitoring the absorbance at 340 in a Benchmark Plus spectrophotometer (Bio-Rad). One enzyme unit is defined as the amount of DXS catalyzing the transformation of 1 mmol of pyruvate and D-glyceraldehyde 3-phosphate into DXP under the conditions of the assay.

2.5. Generation and analysis of *A. thaliana* transgenic plants

A 1.6-kb fragment of the DXS2 5'-flanking region was amplified by Pfu-based PCR using *A. thaliana* genomic DNA as template and the oligonucleotides DXS2.PROM5A and DXS2.PROM3XbaI as primers. The amplification product was cloned into plasmid pBLUESCRIPT-SK- previously digested with SmaI. The insert was recovered as an XbaI fragment and transferred to plasmid pCambia1303 previously digested with XbaI and SpeI, to substitute for the CaMV 35S promoter. The resulting plasmid, named DXS2::GUS, contains 1184 bp of the DXS2 5'-flanking untranscribed region, 43 bp of the DXS2 5'-UTR, and the first 33 bp of the DXS2 coding sequence, cloned in frame with the coding region of the *E. coli* uidA (GUS, beta-glucuronidase) + Green Fluorescent Protein (GFP) chimeric reporter genes. *A. thaliana* wt Col-0 plants were transformed with DXS2::GUS via Agrobacterium-mediated infiltration and further analyzed as described (Carretero-Paulet, 2003; Carretero-Paulet et al., 2002).

2.6. DXS1 and DXS2 expression in *chs5* transgenic plants

Transgenic *A. thaliana* lines overexpressing either DXS1 or DXS2 were generated in the *chs5* mutant genetic background via Agrobacterium-mediated infiltration as described (Carretero-Paulet, 2003). To this purpose, two constructs designated 35S::DXS1 and 35S::DXS2, in which the coding region of the DXS1 and DXS2 cDNAs were under the control of the cauliflower mosaic virus 35S (CaMV 35S) promoter, were used. The 35S::DXS1 construct was previously generated and used for transformation of *A. thaliana* plants (Botella-Pavia et al., 2004). To obtain the 35S::DXS2 construct, a cDNA fragment encoding the full-length *A. thaliana* DXS2 was excised from plasmid pLB.DXS2 as a BamHI–HincII fragment and cloned into plasmid pCambia1303 previously digested with BglII and Eco72I. All constructs were confirmed by DNA sequencing.

2.7. Microbombardment assays

A cDNA fragment encoding the *A. thaliana* DXS2 putative transit peptide was amplified by PCR using oligonucleotides DXS2.NcoI-GFP and DXS2.Sall3GFP-TP. The amplification product was cloned into plasmid pGEM-T (Promega, Madison, WI) and subsequently transferred as a 126-bp NcoI–Sall fragment to pGFP-MRC plasmid previously digested with XhoI and Sall (Lois et al., 2000). In addition, the *A. thaliana* DXS1 full-length cDNA was amplified by PCR using the oligonucleotide DXS1.SallA and the universal T7 primer using as a template the EST clone H2A12T7 (Carretero-Paulet et al., 2006). The amplification product was cloned into plasmid pGEM-T and subsequently transferred as a Sall–Sall fragment to plasmid pGFP-MRC previously digested with Sall. The resulting constructs encoded the *A. thaliana* DXS2 putative transit peptide (DXS2TP, amino acids 1–42) and the whole DXS1 protein, respectively, fused to the N-terminus of the Green Fluorescent Protein (GFP) under the control of the CaMV 35S promoter. All constructs were confirmed by DNA sequencing.

Leaves of 15-day-old *A. thaliana* seedlings were microbombarded with plasmid DNA-coated tungsten particles and examined by confocal laser scanning microscopy as described (Lois et al., 2000).

2.8. Characterization of a DXS2 insertion mutant

An *A. thaliana* DXS2 knockout mutant (SALKN612762) was identified at the Salk Institute Genomic Analysis Laboratory collection (<http://signal.salk.edu>; (Alonso et al., 2003)) and obtained from the Nottingham Arabidopsis Stock Centre (<http://arabidopsis.info/>). Plants homozygous for DXS2 gene disruption were selected by PCR using the sets of primers DXS2.S5/DXS2.S10 and Lb1/DXS2.S10 (Fig. 5). For RT-PCR, total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen), DNase I treated, and reverse transcribed by the aid of Retroscript reverse transcriptase (Ambion) using random decamers. A fraction (about one-twentieth) of the first-strand cDNAs was used as a template for PCR with the gene-specific primers DXS2.S2 and DXS2.S6. Amplification of the transcript encoded by the *A. thaliana* DXS1 gene using primers DXS1.S1 and DXS1.S10 served as an internal control. PCR products were analyzed by agarose gel electrophoresis.

2.9. Sequence and phylogenetic analysis

Searches for DXS sequences were performed using different BLAST-based programs (Altschul et al., 1997) in selected databases, including PLAZA 2.5 (<http://bioinformatics.psb.ugent.be/plaza/>) (Proost et al., 2009) phytozome v8.0 (<http://www.phytozome.net/>) (Goodstein et al., 2012), Cyanidioschyzon merolae Genome Project (<http://merolae.biol.s.u-tokyo.ac.jp/>) and NCBI (<http://www.ncbi.nlm.nih.gov/>) using DNA coding and protein sequences of *A. thaliana* DXS1 (AT4G15560), DXS2 (AT3G21500) and DXS3 (AT5G11380) genes as queries. Only hits corresponding to full-length sequences were considered. Resulting hits were scanned for the presence of the INTERPRO DXS-specific domain IPR005477 and plastid transit peptides.

Phylogenetic analysis was performed on the basis of the alignment of predicted mature functional DXS protein sequences obtained using MUSCLE (Edgar, 2004). Maximum Likelihood (ML) phylogenetic reconstruction was carried out in PhyML v3.0 (Guindon and Gascuel, 2003; Guindon et al., 2010) using the LG protein evolution model (Le and Gascuel, 2008) and heterogeneity of amino acid substitution rates corrected using a gamma distribution (Γ) with eight categories, selected by ProtTest v2.4 as the best-fitting amino acid substitution model according to the Akaike information criterion (Abascal et al., 2005). Tree topology searching was optimized using the subtree pruning and regrafting option. The statistical support of the retrieved topology was assessed using the Shimodaira–Hasegawa-like approximate likelihood ratio (aLRT) test (Anisimova and Gascuel, 2006). Bayesian analysis was implemented in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001) using the JTT plus a gamma distribution (Γ) with eight categories (Jones et al., 1992). Searches were run with four Markov chains for 1 million generations and sampling every 100th tree. After stationary phase was reached (determined by the average standard deviation of split sequences approaching 0, which reflects the fact that independent tree samples became increasingly similar), the first 2500 trees were discarded as burn-in and a consensus tree was then constructed to evaluate Bayesian posterior probabilities on clades.

2.10. Molecular evolutionary analysis

Estimates of the ratio ω of non-synonymous (dN) versus synonymous (dS) nucleotide substitutions rates provide a simple and useful measurement of selective pressures driving evolution at the molecular level (simplistically, $\omega = 1$, $\omega < 1$ and $\omega > 1$ indicates neutral, purifying and positive selection, respectively). In order to examine the molecular evolutionary forces underlying putative functional specialization of *A. thaliana* DXS2, alignment of codon sequences and the corresponding ML tree was analyzed using codon substitution based models from the codeml program of the PAML v4.4 package (Yang, 2007). Nucleotide sequences were aligned in accordance with the corresponding amino acid alignment to obtain final codon alignments.

Codon alignments were subsequently edited through GBLOCKS v0.91b software to eliminate poorly aligned positions and divergent regions, such as those particularly rich in indels (Castresana, 2000). These models perform maximum likelihood estimates of ω ratios and attach a log likelihood (lnL) value to the examined tree topology and codon alignment. Two different classes of models were implemented by the codeml program: i) the branch-specific models, allowing to estimate different ω ratios among branches in the phylogeny (Yang, 1998), and ii) the branch-site models, allowing ω to vary not only in selected branches on the tree but also across codons in the sequences by defining different ω ratios site-classes (Yang and Nielsen, 2002). Among the first class, the one-ratio model 0, constraining all branches in the tree to the same ω ratio, and the two-ratio model 2, allowing for pre-specified branches in the phylogeny (foreground branches) to evolve at an independent ratio than the rest of branches, were run. Among the second class, clade model D, allowing a class of sites to be under divergent selective pressure between foreground branches and the rest of the tree (Bielawski and Yang, 2004), and model A (Zhang et al., 2005), featuring an extra class of sites with $\omega > 1$ in foreground branches to be estimated from the data.

The comparison of models was performed through Likelihood Ratio Test (LRT), which examine the significance of the difference between lnL resulting from two nested models (Goldman and Yang, 1994). This test statistic is calculated as $2\Delta\ln L$, twice the difference between the lnL under each model. If the null (simpler) model is true, LRT asymptotically follows a χ^2 distribution with a number of degrees of freedom equal to the differences in the number of parameters between models. Hence, the LRT allow examining different evolutionary hypotheses by calculating a probability for the fitting of the examined dataset to the alternative model being tested. To test for asymmetric sequence evolution, the one-ratio model 0 was compared to the two ratio model 2. (Non) neutral evolution was tested by comparing the lnL between the two-ratio branch model 2 with ω fixed at one for the selected branches and the two-ratio branch model 2 with ω estimated from the data. Furthermore, the comparison between the site-specific null discrete model 3 (which allows the ω ratio to vary among sites, but holding ω constant among branches in the tree) and the clade model D is a useful test to measure the differential selective pressures acting on a significant number of amino acids (codon sites) either under relaxed purifying selection or under positive selection (Bielawski and Yang, 2004). Finally, the comparison between the model A with ω fixed at one for the examined branch as null model and the model A was used as a conservative test to detect as opposed to relaxed purifying selection affecting a few sites in the selected branch (Zhang et al., 2005). Furthermore, model A also implements a Bayes Empirical Bayes (BEB) procedure that calculates posterior probabilities (PP) of a codon being subjected to positive selection (Yang et al., 2005).

2.11. Analysis of functional divergence

The analysis of functional divergence was performed using the DIVERGE v2.0 software (Gu and Vander Velden, 2002). DIVERGE performs the ML calculation of the theta (θ) coefficients of functional divergence type-I and type-II, based on the occurrence of altered selective constraints or radical shifts of physiochemical properties, respectively, between selected clusters in the tree after duplication (Gu, 1999, 2001, 2006). The program also estimates the posterior probabilities of amino acid sites to be responsible for the functional divergence.

2.12. Building of a 3D-structural model of *A. thaliana* DXS2

We used Modeller to build ten models of *A. thaliana* DXS2 (Marti-Renom et al., 2000), using PDB structures 2o1s and 2o1x, from *E. coli* and *Deinococcus radiodurans*, respectively, as templates (Xiang et al., 2007). Subsequently, we corrected flips using REDUCE

(Word et al., 1999) and checked the resulting models with Procheck (Laskowski et al., 1993) in order to select one model. The image corresponding to the resulting homology-based 3D-structural model has been displayed and edited through PyMOL (Delano, 2002). Accessibility of the residues in the structure was calculated using the NACCESS program (Hubbard and Thornton, 1993).

3. Results

3.1. Identification and sequence analysis of *A. thaliana* DXS2 and DXS3

After the identification of the *DXS1/CLA1* gene (AT4G15560), searches for additional DXS-related sequences in the *A. thaliana* genome led to the identification of two genes, AT3G21500 and AT5G11380, which were tentatively named DXS2 and DXS3, respectively (Carretero-Paulet, 2003; Phillips et al., 2008; Rodriguez-Concepcion and Boron, 2002). The predicted encoded mature proteins show identities of 72.8% (DXS2) and 52.9% (DXS3) with respect to DXS1. Like DXS1, both DXS2 and DXS3 have N-terminal extensions, as compared to the prokaryotic counterparts, predicted to represent plastid transit peptides (Fig. 1). *A. thaliana* DXS3 lacks many of the highly conserved residues known to be essential for DXS activity (Xiang et al., 2007) and a previous phylogenetic analysis grouped it within a clade composed of highly diverged protein sequences (Cordoba et al., 2011). However, marginal, though significant, DXS activity has been recently reported for the corresponding ortholog in maize (Cordoba et al., 2011).

In contrast, *A. thaliana* DXS2 shows a higher level of amino acid sequence similarity with DXS1 and contains the conserved amino acid residues proposed to have a role in the catalytic site of the enzyme (Xiang et al., 2007) (Fig. 1). This suggested that DXS2 could actually represent a new DXS isozyme in *A. thaliana*. However, DXS1 and DXS2 proteins differ in the N-terminal sequence corresponding to the predicted plastid transit peptide and on two deletions affecting an internal region of 32 amino acids, which is included in a disorganized external loop

located close to the dimer interface in the prokaryotic DXS (positions 197 to 241 in *E. coli* DXS; (Xiang et al., 2007)) and the C-terminal end of the DXS2 protein (Fig. 1). Although the structural or functional role of these deleted sequences is currently unknown, the corresponding regions are significantly conserved among DXS (Supplementary Fig. 2).

3.2. Functional analysis of *A. thaliana* DXS2

To evaluate, and eventually confirm, the catalytic activity of DXS2, we first undertook a complementation assay using an *E. coli* strain defective in DXS activity. A cDNA encoding the full-length DXS2 protein was first amplified by RT-PCR using RNA obtained from different organs and developmental stages of *A. thaliana* plants. Shorter versions encoding DXS2 variants lacking either the first N-terminal 33 amino acids (DXS2C) or the first 44 N-terminal amino acids (DXS2CC) were also generated. DXS2C does not contain the predicted plastid transit peptide and DXS2CC corresponds to a protein that extends over the region conserved in prokaryotic counterparts (Fig. 1). As a positive control, the cDNA encoding *A. thaliana* mature DXS1 (lacking the plastid transit peptide) was used. The lethal phenotype caused by the loss of DXS activity in the *E. coli* mutant strain was rescued with the plasmid encoding DXS1 but not with those encoding DXS2, DXS2C and DXS2CC (Fig. 2A).

One of the most apparent differences between *A. thaliana* DXS1 and DXS2 at the sequence level consists of a deletion affecting the internal part of the DXS2 protein (between N255 and C256, Fig. 1). To test whether the deletion of this sequence was responsible for the lack of DXS activity, we constructed a plasmid encoding a derivative of DXS2CC in which the corresponding region from *A. thaliana* DXS1 was integrated in DXS2CC. Expression of the resulting chimeric DXS2 protein (DXS2CCdom) neither complemented the *E. coli* dxs mutant (Fig. 2A).

The lack of complementation of the DXS defective strain suggested that *A. thaliana* DXS2 either does not have DXS activity or that its activity was not enough to support full rate isoprenoid biosynthesis

Fig. 1. Alignment of *E. coli* DXS and *A. thaliana* DXS1, DXS2 and DXS3. Amino acid residues identical in all four proteins are highlighted in black. Amino acid residues identical in three out of the four proteins are shown in gray. DXS functional residues involved in GAP-binding (black stars) and TPP-binding (white stars) as determined from crystallographic analysis of bacterial DXS (Xiang et al., 2007) are pointed out.

in *E. coli*. To check this possibility, mature DXS1 and DXS2 fused to a 6xHis tag at their C-terminus were expressed in *E. coli* and the recombinant proteins purified by immobilized metal affinity chromatography. Enzymatic DXS assays showed that the purified recombinant DXS1 had a specific activity of 2.6 $\mu\text{mol}/\text{min}/\text{mg}$, a value which was in agreement with previous reports using purified recombinant DXS from other organisms (Bailey et al., 2002; Hahn et al., 2001). In the case of recombinant DXS2 the enzyme activity was in the range of 0.05 $\mu\text{mol}/\text{min}/\text{mg}$ of protein and very close to background assay levels. It could be argued however that *A. thaliana* DXS2 was unable to rescue the DXS defective *E. coli* mutant or has detectable enzyme activity due to specific folding or stability requirements that cannot be met in bacteria. Therefore, we next tried a similar complementation assay but using a DXS defective *A. thaliana* mutant. Complete loss of DXS1 activity is lethal in *A. thaliana* but viable partial loss-of-function mutants are also available (Crowell et al., 2003). For example, a single nucleotide change in the *DXS1* gene of the *A. thaliana* *chs5* mutant leads to an amino acid substitution that results in a temperature-sensitive DXS1 enzyme (Araki et al., 2000). As a consequence, the *chs5* mutant develops albino leaves at restrictive temperatures (15 °C), due to a block in the production of chlorophyll and carotenoid precursors by the MEP pathway, but almost normal green leaves at permissive temperatures (22 °C) (Fig. 2B). When mutant *chs5* plants were transformed with constructs designed to constitutively overproduce the full length *A. thaliana* DXS1 or DXS2 isoforms, only *chs5::DXS1* lines, transformed with the 35S-DXS1 construct, showed a clear rescue of the albino phenotype when grown at 15 °C (Fig. 2B).

3.3. Intracellular targeting of *A. thaliana* DXS2

In spite of its predicted transit peptide, DXS2 could fail to rescue the *A. thaliana* *chs5* mutant because it is not localized in plastids as DXS1 does. MEP pathway enzymes are nuclear encoded and targeted to the plastids (Bouvier et al., 1998; Carretero-Paulet, 2003; Carretero-Paulet et al., 2002; Hsieh et al., 2008; Kim et al., 2009; Lange et al., 1998; Lois et al., 2000; Oudin et al., 2007; Rodriguez-Concepcion et al., 2001). To test whether DXS2 was a plastidial protein, the sequence containing the putative transit peptide of the DXS2 protein (DXS2TP, residues 1–42) was fused to the N terminus of an optimized version of GFP and the recombinant DXS2TP–GFP protein transiently expressed in *A. thaliana* leaf cells. Leaves from 15-day-old *A. thaliana* plants were microbombarded with a construct expressing DXS2TP–GFP and the subcellular localization of the green fluorescence corresponding to the recombinant protein was observed by confocal laser scanning microscopy (Fig. 3). Green fluorescence from DXS2TP–GFP is co-localized with chlorophyll autofluorescence, thus indicating that, like DXS1, DXS2 is also targeted to plastids.

3.4. Expression analysis of *A. thaliana* DXS2

To get some clues about the possible function of DXS2 in *A. thaliana*, we analyzed the expression pattern of the *DXS2* gene. DXS2 appears to be expressed at lower levels than DXS1, as suggested by the low number of ESTs found in the available databases. Consistently, *DXS2* transcripts could not be detected by northern-blot analysis (unpublished results). Using a semi-quantitative RT-PCR approach, *DXS2* transcripts were detected at highest levels in siliques and inflorescences, with very low levels detected in leaves, roots, stems, whole seedlings and T87 cultured cells (Fig. 4A). Unlike that reported for genes encoding MEP pathway enzymes in *A. thaliana* (Botella-Pavia et al., 2004; Carretero-Paulet et al., 2002; Cordoba et al., 2011; Mandel et al., 1996; Rodriguez-Concepcion, 2006) no changes in *DXS2* transcript levels were detected between light-grown and dark-grown seedlings (Fig. 4A). A more detailed analysis of the spatial and temporal pattern of *DXS2* expression was obtained using transgenic plants carrying a translational fusion between 1.2 kb of the *DXS2* 5' flanking region fused to the entire coding sequence of the *GUS* reporter gene.

Fig. 2. Functional analysis of *A. thaliana* DXS2. (A) Complementation of an *E. coli* *dxs* defective mutant with *A. thaliana* DXS1 and DXS2. The *E. coli* *dxs::CAT* mutant EcAB4-2 (Campos et al., 2001) was transformed with a control expression plasmid without insert (Control), a pTAC derivative coding for *A. thaliana* DXS1 (DXS1), three pBADM1 derivatives coding for three versions of *A. thaliana* DXS2: DXS2 (amino acid residues 1–629), DXS2C (amino acid residues 34–629), and DXS2CC (amino acid residues 45–629), and a chimeric DXS protein in which the amino acid region corresponding to positions 267–309 of *A. thaliana* DXS1 has been integrated within the corresponding positions of DXS2CC (DXS2CCdom). (B) Complementation of *A. thaliana* *chs5* mutant with *A. thaliana* DXS1 and DXS2. Images correspond to 3-week old plants grown under controlled conditions at 15–16 °C or 22–24 °C. Plants were grown on Petri dishes with MS medium. *chs5::DXS1* and *chs5::DXS2* correspond to transgenic *A. thaliana* *chs5* mutant plants transformed with a construct containing either the full-length cDNA coding for *A. thaliana* DXS1 or DXS2, respectively, under the control of the strong and constitutive promoter *CaMV* 35S.

The histochemical pattern of GUS activity was consistent with the data obtained from RT-PCR analysis. Expression levels were relatively high during reproductive development, as revealed by the intense GUS activity in young and mature inflorescences as well as in developing siliques

Fig. 3. Targeting and subcellular localization of *A. thaliana* DXS1 and DXS2. Leaves of 15-day-old *A. thaliana* seedlings were microbombarded with a construct encoding GFP (A) and either *A. thaliana* DXS1 (B) or the putative DXS2 transit peptide (C) fused to the N terminus of GFP, respectively. Cells expressing the fusion protein were studied by laser confocal microscopy.

(Figs. 4F–I). In the rest of the tissues, organs and developmental stages GUS staining could only be observed after long incubation periods (more than 24 h), with the only exception of mature roots where no GUS staining was observed (Figs. 4B–E). The expression pattern observed for DXS2 is in agreement with that reported in the *A. thaliana* eFP Browser database (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>).

3.5. Phenotypic analysis of an *A. thaliana* DXS2 defective mutant

To further investigate the biological function of DXS2 we analyzed the phenotype of a defective mutant (*dxs2-1*) containing a T-DNA insertion in exon 4 of the DXS2 gene (Fig. 5A). As estimated by RT-PCR, homozygous mutants do not show the expression of the DXS2 mRNA (Fig. 5B). The phenotype of the *dxs2-1* mutant was indistinguishable from wild type plants during the different stages of vegetative and reproductive growth analyzed (Fig. 5C). These results show that, in contrast to DXS1, the function of DXS2 is not essential for *A. thaliana* growth and development.

3.6. Phylogenetic analysis of *A. thaliana* DXS2

To examine the evolutionary origin and relationships of *A. thaliana* DXS2 within the plant DXS family, we performed a ML phylogenetic analysis based on the alignment of 79 DXS protein sequences (Supplemental Fig. 2). These included the mature functional region of 74 DXS proteins from 40 species representative of the main plant evolutionary lineages, including the whole DXS family of 20 species with fully sequenced genomes, plus five bacterial representatives (Fig. 6 and Table 1). Bayesian analysis was also performed, resulting in trees with almost identical topologies (Supplemental Fig. 3). DXS from green algae and land plants clustered into independent clades, suggesting that the evolutionary expansion and diversification of the family occurred after transition to land. Previously defined subfamilies I, II and III grouped into well supported clades (Cordoba et al., 2011; Walter et al., 2002) (Fig. 6). *A. thaliana* DXS1 and DXS2 clustered within subfamily I while DXS3 clustered with a set of nine highly diverged sequences from monocots and dicots into subfamily III (Fig. 6). Interestingly, no *A. thaliana* or any Brassicaceae species representative was found within subfamily II.

The phylogenetic analysis also revealed lineage-specific gene duplication (and losses) within subfamilies I and II of DXS. This was the case of *A. thaliana* DXS2, closely clustering with its orthologs from *Arabidopsis lyrata* and three additional Brassicaceae species (Fig. 6). The sister-clade relationship between Brassicaceae DXS2 and DXS1 suggests the origin of the former through a duplication event predating the emergence of the Brassicaceae lineage. Coherently, *A. thaliana* DXS1 and DXS2 genes map in genomic regions of chromosomes 4 and 3 (<http://wolfe.gen.tcd.ie/athal/dup>) resulting from a whole genome duplication event estimated to have occurred in the last 25–40 Ma (Blanc et al., 2003), matching the time of origin and diversification of the Brassicaceae family (Couvreur et al., 2010). Furthermore, the corresponding duplicated genomic blocks showed a

Fig. 4. Analysis of *A. thaliana* DXS2 gene expression. (A) Semi-quantitative RT-PCR analysis using DXS2 gene-specific primers and rRNA 18S specific primer as internal standard. The source of the RNAs is indicated. (B–I) Histochemical analysis of GUS activity in transgenic DXS2::GUS plants. (B), germinating seed (24 h) in MS medium; (C), 7-day-old dark-grown etiolated seedlings; (D), 3-week-old light grown plants; (E), cauline leaf with emerging axillary inflorescence; (F), young inflorescences; (G), flower; (H), green silique; and (I), mature silique.

Fig. 5. Phenotypic analysis of the *A. thaliana* *dxs2-1* mutant. (A) Schematic representation of the *DXS2* gene and the position of T-DNA insertion (T-DNA is not drawn to scale); (B) genotyping (Genomic DNA) and expression (RNA) analysis of *DXS2* in the *dxs2-1* mutant using PCR and RT-PCR, respectively. Levels of *DXS2* (upper panel) and *DXS1* (lower panel) mRNAs are shown. The relative position of the *DXS2* primers used is shown in panel A; and (C) phenotype of the *dxs2-1* mutant at different stages of vegetative and reproductive growth.

high level of synteny and gene collinearity over a region of more than 1 mega base pairs, according to the analysis performed using CoGe's tool Gevo (this analysis may be regenerated by using the following link, <http://genomevolution.org/r/5dy5>).

3.7. Molecular evolutionary analysis of *A. thaliana* *DXS2*

Under the assumption that estimation of ω ratios (of nonsynonymous versus synonymous substitutions) specifically acting on Brassicaceae *DXS2* sequences would give us further insights into the evolutionary and functional fate of *A. thaliana* *DXS2* after duplication, we performed a molecular evolutionary analysis based on codon-substitution models under a likelihood framework. Our analysis was restricted to the alignment of the subset of 20 DXS codon sequences belonging to dicot species from DXS subfamily I (Table 1), which showed appropriate sequence divergence (Anisimova et al., 2001; Hernandez-Hernandez et al., 2007). The inferred ML protein phylogeny was used as starting tree and branches within the Brassicaceae *DXS2* clade were selected as foreground branches.

Firstly, to test for asymmetric evolution following gene duplication, we used the one-ratio model as null model in Likelihood Ratio Tests (LRT) comparisons with the two-ratio model (Table 2). The LRT revealed the two-ratio model, which gave estimates of $\omega_1 = 0.1384$ for the Brassicaceae *DXS2* branches significantly higher than the background ratio for the rest of branches $\omega_0 = 0.0408$, provided a significantly better fit to the data. Furthermore, to discard increased ω on Brassicaceae *DXS2* branches following gene duplication was due to neutral evolution, a LRT was constructed comparing the two-ratio model with ω_1 fixed as one as null model and the two-ratio model with ω_1 estimated from the data (Table 2). The analysis clearly rejected neutral evolution acting on *DXS2* branches.

It is reasonable to expect divergent selective constraints or positive selection acting on specific sites after gene duplication of Brassicaceae *DXS2*. To examine these hypotheses, the branch-site specific models were implemented. LRT comparing discrete model 3 as null model

with model D was performed (Table 2), supporting divergence in selective constraints acting on a substantial fraction of sites (Bielawski and Yang, 2004). Parameter estimates under model D suggested ~27% of sites evolving under strong purifying selection in dicot subfamily I *DXS* sequences ($\omega_{01} = 0.05$) and much weaker purifying selection in the Brassicaceae *DXS2* clade ($\omega_{02} = 0.41$). Moreover, to evaluate whether divergent selective pressures could also have been due to positive selection, we used the branch-site test of positive selection, which compares the null model A with model A (Zhang et al., 2005). The resulting LRT statistic did not favor the latter, arguing against positive selection (Table 2). However, the BEB analysis, which has been shown to be extremely conservative in inferring positively selected sites (Yang et al., 2005; Zhang et al., 2005), identified 16 amino acid residues putatively subjected to positive selection with high posterior probabilities (Table 2).

3.8. Analysis of functional divergence

To further examine the role of specific amino acid substitutions of Brassicaceae *DXS2* proteins into functional and evolutionary fate after gene duplication, two different statistical approaches were followed as implemented in the DIVERGE v2.0 software (Gu and Vander Velden, 2002). The first one allows detecting amino acids subjected to different evolutionary rates between clades under exam, indicative of divergent functional selective constraints (type-I functional divergence) (42, 43). The second one relies on site-specific shifts of amino acid property (type-II functional divergence) (44).

From 0 to 1, the functional divergence increases from very weak to very strong. Given the ML tree topology of 20 dicot *DXS* sequences from subfamily I, the ML estimate of the theta coefficients of type-I functional divergence between the Brassicaceae *DXS2* clade and the rest of sequences was significant (Table 3), implying that functional selective constraints at some sites have shifted significantly after gene duplication of Brassicaceae *DXS1* and *DXS2* clades. In contrast, the corresponding tests did not support type-II functional divergence (Table 3).

Fig. 6. Maximum likelihood phylogeny of DXS. ML tree depicting the phylogenetic relationships among 74 DXS protein sequences from 40 plant species (including 20 with whole sequenced genomes) and 5 bacterial DXS proteins (see Table 1 and Supplementary file 1). Statistical support on clades is indicated by aLRT values, shown next to the corresponding nodes. The tree is drawn to scale, with branch lengths proportional to evolutionary distances between nodes. The scale bar indicates the number of amino acid substitutions per site. Clades clustering plant DXS from subfamilies I, II and III and green algae DXS are indicated with bars. Clades grouping dicot DXS from subfamily I and Brassicaceae *DXS2*, used as background and foreground branches in the molecular evolutionary and functional divergence analyses, respectively, are highlighted within boxes.

Table 1

Summary of sequences used in this study.

Name ^a	Gene identifier	Species	Lineage
Aaestivalis	117553507	<i>Adonis aestivalis</i>	Dicots
AannuaDXS1	8000868	<i>Artemisia annua</i>	Dicots – asterids
Alyrata1	AL7G28180	<i>A. lyrata</i> ^a	Dicots – rosids
Alyrata2	AL3G23620	<i>A. lyrata</i> ^a	Dicots – rosids
Alyrata3	AL6G11170	<i>A. lyrata</i> ^a	Dicots – rosids
AmajusDXPS	56785668	<i>Antirrhinum majus</i>	Dicots – asterids
Apaniculata	30144404	<i>Andrographis paniculata</i>	Dicots – asterids
AthalianaDXS1	AT4G15560	<i>Arabidopsis thaliana</i> ^a	Dicots – rosids
AthalianaDXS2 (AtDXL1)	AT3G21500	<i>A. thaliana</i> ^a	Dicots – rosids
AthalianaDXS3	AT5G11380	<i>A. thaliana</i> ^a	Dicots – rosids
Aviabilis	75910729	<i>Anabaena variabilis</i> ^a	Bacteria
Bdistachyon1	BD2G25550	<i>Brachypodium distachyon</i> ^a	Monocots
Bdistachyon2	BD1G54890	<i>B. distachyon</i> ^a	Monocots
Brapa1	Bra033495	<i>Brassica rapa</i> ^a	Dicots – rosids
Brapa2	Bra001832	<i>B. rapa</i> ^a	Dicots – rosids
Brapa3	Bra008967	<i>B. rapa</i> ^a	Dicots – rosids
Brapa4	Bra012779	<i>B. rapa</i> ^a	Dicots – rosids
CannuumTKT2	3559816	<i>Capsicum annuum</i>	Dicots – asterids
Cmerolae	CMF089C	<i>Cyanidioschyzon merolae</i> ^a	Red algae
Cmorifolium	87299429	<i>Chrysanthemum x morifolium</i>	Dicots – asterids
Cpapaya1	CP00016G00910	<i>Carica papaya</i> ^a	Dicots – rosids
Cpapaya2	CP00055G00080	<i>C. papaya</i> ^a	Dicots – rosids
Cpapaya3	CP00091G00340	<i>C. papaya</i> ^a	Dicots – rosids
Creinhardtii	CR07G09120	<i>Chlamydomonas reinhardtii</i> ^a	Green algae
CroesusDXPS	11691594	<i>Catharanthus roseus</i>	Dicots – asterids
Crubella1	Carubv10004244m.g	<i>Capsella rubella</i> ^a	Dicots – rosids
Crubella2	Carubv10016105m.g	<i>C. rubella</i> ^a	Dicots – rosids
Crubella3	Carubv10000557m.g	<i>C. rubella</i> ^a	Dicots – rosids
DradioduransDXS	1800297	<i>Deinococcus radiodurans</i> ^a	Bacteria
EcolidXS	6059543	<i>Escherichia coli</i> ^a	Bacteria
EguineensisDXS1	46488696	<i>Elaeis guineensis</i>	Monocots
Gbiloba1	46361691	<i>Ginkgo biloba</i>	Gymnosperms
Gbiloba2	40849969	<i>G. biloba</i>	Gymnosperms
HbrasiliensisDXS1	46410857	<i>Hevea brasiliensis</i>	Dicots – rosids
HbrasiliensisDXS2	94421689	<i>H. brasiliensis</i>	Dicots – rosids
McitrifoliaDXS2	16930804	<i>Morinda citrifolia</i>	Dicots – asterids
MpiperitaDXPS	3114573	<i>Mentha x piperita</i>	Dicots – asterids
MRCC299	MRCC299_01G06320	<i>Micromonas sp. RCC299</i> ^a	Green algae
MtruncatulaDXS1	21322713	<i>Medicago truncatula</i> ^a	Dicots – rosids
MtruncatulaDXS2	21322715	<i>M. truncatula</i> ^a	Dicots – rosids
NpseudonarcissusDXS2	10178920	<i>Narcissus pseudonarcissus</i>	Monocots
Olucimarinus	OL02G03930	<i>Ostreococcus lucimarinus</i> ^a	Green algae
Osativajaponica1	LOC_Os05g33840	<i>Oryza sativa ssp. japonica</i> ^a	Monocots
Osativajaponica2	LOC_Os07g09190	<i>O. sativa ssp. japonica</i> ^a	Monocots
Osativajaponica3	LOC_Os06g05100	<i>O. sativa ssp. japonica</i> ^a	Monocots
Otauri	OT02G03880	<i>Ostreococcus tauri</i> ^a	Green algae
PabiesDXS1	153799533	<i>Picea abies</i>	Gymnosperms
PabiesDXS2A	153799535	<i>P. abies</i>	Gymnosperms
PabiesDXS2B	153799537	<i>P. abies</i>	Gymnosperms
PdensifloraDXS1	183206796	<i>Pinus densiflora</i>	Gymnosperms
PdensifloraDXS2	184071334	<i>P. densiflora</i>	Gymnosperms
PmontanaDXS	35187002	<i>Pueraria montana</i>	Dicots – rosids
Ppatens1	PP00022G01710	<i>Physcomitrella patens</i> ^a	Bryophyta
Ppatens2	PP00011G01360	<i>P. patens</i> ^a	Bryophyta
Ppatens3	PP00042G00320	<i>P. patens</i> ^a	Bryophyta
Ppatens4	PP00139G00580	<i>P. patens</i> ^a	Bryophyta
Ptrichocarpa1	PT08G19910	<i>Populus trichocarpa</i> ^a	Dicots – rosids
Ptrichocarpa2	PT06G17230	<i>P. trichocarpa</i> ^a	Dicots – rosids
Ptrichocarpa3	PT06G24870	<i>P. trichocarpa</i> ^a	Dicots – rosids
Ptrichocarpa4	PT07G08920	<i>P. trichocarpa</i> ^a	Dicots – rosids
RcapsulatusDXSA	141323	<i>Rhodobacter capsulatus</i> ^a	Bacteria
RcapsulatusDXSB	^b	<i>R. capsulatus</i> ^a	Bacteria
ShabrochaitesDXS2	51234083	<i>Solanum habrochaites</i>	Dicots – asterids
Sitalica1	Si021346m.g	<i>S. italica</i> ^a	Monocots
Sitalica2	Si029067m.g	<i>S. italica</i> ^a	Monocots
Sitalica3	Si005941m.g	<i>S. italica</i> ^a	Monocots
SlycopersiconDXS1	5059160	<i>Solanum lycopersicon</i>	Dicots – asterids
Smoellendorffii1	SM00038G00390	<i>Selaginella moellendorffii</i> ^a	Lycophyta
Smoellendorffii2	SM00006G04950	<i>S. moellendorffii</i> ^a	Lycophyta
SrebaudianaDXS2	22797427	<i>Stevia rebaudiana</i>	Dicots – asterids
TerectaDXS2	9971824	<i>Tagetes erecta</i>	Dicots – asterids
Thalophila1	Thhalv10024542m.g	<i>Thellungiella halophila</i> ^a	Dicots – asterids
Thalophila2	Thhalv10020205m.g	<i>T. halophila</i> ^a	Dicots – asterids
Thalophila3	Thhalv10012843m.g	<i>T. halophila</i> ^a	Dicots – asterids
Tmedia	46361693	<i>Taxus x media</i>	Gymnosperms

Table 1 (continued)

Name ^a	Gene identifier	Species	Lineage
Vcarteri	VC00067G00340	<i>Volvox carteri</i> ^a	Green algae
ZmaysDXS1	ZM06G24060	<i>Zea mays</i> ^a	Monocots
ZmaysDXS2	ZM07G03860	<i>Z. mays</i> ^a	Monocots
ZmaysDXS3	ZM09G05000	<i>Z. mays</i> ^a	Monocots

Sequences examined in the molecular evolutionary and functional divergence analyses are shown in bold.

^a Species with whole sequenced genomes.

^b Sequence reported in Hahn et al. (2001).

Table 2

Parameter estimates, likelihood values and LRTs of codon-substitution evolutionary models.

Model	p ^a	Parameter estimates (frequency, <i>f</i> and ω values)				lnL	BEB PS sites ^b
One ratio model 0 ($\omega_0 = \omega_1$)	40	$\omega_0 = \omega_1 = 0.0534$				−10,964.31	NA
Two-ratio model 2 (ω_0, ω_1)	41	$\omega_0 = 0.0408; \omega_1 = 0.1384$				−10,925.15	NA
Two-ratio model 2 ($\omega_0, \omega_1 = 1$)	40	$\omega_0 = 0.0395; \omega_1 = 1$				−11,073.35	NA
Model 3 (discrete)	44	NA				−10,771.94	NA
Clade model D ($K = 3$)	45	Site class	0	1	2	−10,726.63	NA
		<i>f</i> :	0.62923	0.10186	0.26891		
		ω_0 :	0.00604	0.30210	0.04997		
		ω_1 :	0.00604	0.30210	0.40898		
Model Anull ($\omega = 1$)	42	NA				−10,809.52	NA
Model A ($0 < \omega < 1$)	43	Site class	0	1	2a	2b	86*, 127, 165*, 227*, 259*, 284, 300*, 311*, 354, 357, 462, 534, 558, 567*, 631, 636*
		<i>f</i> :	0.79760	0.03540	0.15991	0.00710	
		ω_0 :	0.03463	1.00000	0.03463	1.00000	
		ω_1 :	0.03463	1.00000	1.00000	1.00000	

LRT for asymmetric sequence evolution: one-ratio model 0 ($\omega_0 = \omega_1$) vs two-ratio model 2 (ω_0, ω_1) $2\Delta\ln L = 78.31$; $df = 1$; $p = 8.8085E - 19$.

LRT for (non)neutral evolution: two-ratio model 2 ($\omega_0, \omega_1 = 1$) vs two-ratio model 2 (ω_0, ω_1) $2\Delta\ln L = 296.39$; $df = 1$; $p = 2.0164E - 66$.

LRT for divergent selection: model 3 (discrete) vs clade model D ($K = 3$) $2\Delta\ln L = 90.63$; $df = 1$; $p = 1.7296E - 21$.

LRT for positive selection: model anull ($\omega_2 = 1$) vs model A ($0 < \omega_0 < 1$) $2\Delta\ln L = 0$; $df = 1$; $p = 1$.

All branches within the Brassicaceae DXS2 clade were used as foreground branches (ω_1). ^a, number of parameters in the ω distribution; ^b, amino acids detected in the BEB analysis

as fixed by positive selection with posterior probabilities >95% (>99%, indicated with asterisks) are shown. In bold type, amino acids also identified in the analysis of functional divergence. Numbering refers to *A. thaliana* DXS2. df = degrees of freedom; p = probability values.

Moreover, 34 specific amino acid residues were predicted as responsible for type-I functional divergence with posterior probabilities >0.9 (Table 3). Interestingly, seven of these positions were also identified by the BEB analysis as putatively fixed by positive selection. Many of these sites are highly conserved in DXS sequences, but variable in Brassicaceae DXS2 or vice versa, suggesting a change in their functional roles (Supplementary Fig. 2).

4. Discussion

Gene duplication provides with a unique mechanism for the evolution of divergent gene functions. In most cases, it is expected that one duplicated gene retains the ancestral function and the other one evolves neutrally, free from selective constraints, rapidly becoming a pseudogene that will be inactivated or even deleted from the genome (Lynch and Conery, 2000; Ohno, 1970; Zhang, 2003). However, in a higher than expected number of cases, after a duplication event, both duplicated genes are maintained in the genome. This occurs when the conservation of the ancestral function by both duplicated

genes is selectively advantageous. Alternatively, duplicated genes evolve through the acquisition of new or specialized functions at the expression or protein level (Conant and Wolfe, 2008; Moore and Purugganan, 2005).

Medicago truncatula, *Picea abies* and *A. thaliana* provide interesting examples of functional specialization following gene duplication within the plant DXS family (Fig. 6). *M. truncatula* DXS1 and DXS2, belonging to subfamilies I and II of DXS, respectively, show differential regulation and plastid location. While the former is preferentially expressed in many developing plant tissues except roots, the latter is expressed at low levels in most tissues, but is strongly stimulated in roots upon colonization by mycorrhizal fungi (Walter et al., 2002). Additional DXS from subfamilies I and II have been characterized, supporting the involvement of the former in the biosynthesis of essential isoprenoids (such as photosynthetic pigments and phytohormones) and the involvement of the latter in the synthesis of specific isoprenoids as secondary metabolites. *P. abies* DXS2A and DXS2B, belonging to subfamily II, showed distinct expression patterns suggestive of their involvement in isoprenoid inducible synthesis through separate jasmonate and salicylate-mediated defense responses signaling pathways (Phillips et al., 2007). In both cases, DXS enzymatic activity has been reported for both paralogs. *A. thaliana* DXS2 arose from a whole duplication event estimated to have occurred at the time of emergence of the Brassicaceae family (Couvreur et al., 2010). From the DXS1–DXS2 pair of *A. thaliana* subfamily I paralog genes, only DXS1 has been characterized up to date (Carretero-Paulet et al., 2002, 2006; Estevez et al., 2000, 2001; Mandel et al., 1996).

Results presented in this study reveal specialization at the level of expression pattern for the DXS2 gene (Fig. 4) with respect to DXS1. Thus, whereas DXS1 expression is particularly high in green developing tissues and is strongly and rapidly up-regulated by light (Estevez et al., 2000), coherently with its proposed role in plastidial isoprenoid biosynthesis, DXS2 expression is not up-regulated by light, and shows

Table 3

Analysis of functional divergence.

Functional divergence	Coefficient $\theta \pm SE$ (p)	Critical amino acid sites
Type I	$\theta_1 = 0.936 \pm 0.121$ ($p = 1.5177E - 10$)	127, 165, 227, 256, 259, 275, 277, 293, 298, 300, 304, 308, 348, 352, 353, 354, 357, 376, 445, 446, 479, 481, 494, 498*, 503*, 533*, 543, 564*, 565, 591, 595, 601, 621, 639*
Type II	$\theta_2 = 0.014 \pm 0.026$	

p = posterior probability values. Critical amino acid sites detected as related to functional divergence with $p > 90\%$ (>95%, indicated with asterisks) are listed. In bold type, amino acids also identified in the BEB analysis of positive selection. Numbering refers to *A. thaliana* DXS2 protein sequence.

weaker expression levels, with the only exception of reproductive organs (Fig. 4). In any case, the putative function of DXS2 does not seem to be essential for growth and development since a DXS2 defective mutant shows no visible phenotype (Fig. 5). Interestingly, complementation assays using DXS-defective *E. coli* or *A. thaliana* mutants suggested that, in spite of the high level of sequence similarity with DXS1, *A. thaliana* DXS2 might not have DXS activity (Fig. 2). In vitro enzyme assays have revealed that DXS2 protein shows negligible levels of DXS activity (estimated as less than 2% of DXS1 activity). The DXS activity provided by DXS2 in the in vivo complementation assays is clearly not enough to sustain active rates of isoprenoid biosynthesis. Although it cannot be ruled out that *A. thaliana* DXS2 could have DXS activity under particular conditions or in the presence of particular interacting partners, our results strongly suggest that DXS2 is not playing a role in isoprenoid biosynthesis and has evolved (or is evolving) to a new function.

We observed accelerated evolutionary rates along the DXS2 branches. Though this observation might be compatible with *A. thaliana* DXS2 evolving as a pseudogene, our evolutionary analyses clearly rejected neutral evolution for *A. thaliana* DXS2 and is conclusive about most of the DXS2 coding sequences evolving under functional selective constraints.

Our results support divergent selective pressures acting on a substantial fraction of sites following gene duplication. 27% of sites would be evolving under relaxed purifying selection in Brassicaceae DXS2 sequences (Table 2). Altered functional selective constraints after gene duplication could be indicative of functional divergence, as further supported the DIVERGE analysis (Table 3). In addition, 34 of the sites' potentially key in functional divergence of DXS2 were identified (Table 3). However, the role of positive selection in divergent selective pressures at specific sites is uncertain. The corresponding LRT does not provide better fit to the model accounting for positive selection, though the BEB analysis detected 16 amino acid changes putatively fixed through positive selection (Table 2). Two features might be hiding the signal of positive selection in the corresponding LRT: i) selection tests might not be sensitive to detect episodic positive selection acting at few sites; and, ii) some positively selected sites might have been deleted during codon alignment correction and edition.

Among them, we found seven of the positions predicted to be keys for functional divergence (Tables 2 and 3).

Considering the high level of similarity between DXS2 and the rest of plant DXS proteins (Supplementary Fig. 2) and the conservation of the amino acid residues proposed to be involved in the catalytic site of prokaryotic DXS (Fig. 1) (Xiang et al., 2007), it is likely that both the loss of DXS activity and the eventual gain of a new catalytic function could be related with changes affecting specific amino acid residues in DXS2. Functional divergence and putative positively selected sites were mapped over the corresponding *A. thaliana* DXS2 protein model. Most of them are located on the protein surface i.e. showed more than 5% accessibility to solvent (unpublished results; Figs. 7A and B). As surface residues can be involved in the recognition and binding of other molecules (Jones and Thornton, 1997; Yeh et al., 2003), such changes might be indicative of conformational changes and subsequent functional shifts in the ancestral DXS activity. Thus, some of these substitutions could affect the access of the ligand to the binding pocket or the hydrogen-bonding. Considering the possibility that specific residues could also affect the interaction of DXS2 with GAP and/or TPP we checked for particular changes mapping close to residues reported to play a role in the DXS active site. We found two residues detected as both related to functional divergence and identified through the BEB analysis to be fixed by positive selection mapping very close (± 4 residues, one helix turn) to residues involved in the binding of GAP (K165 and Q357). Q357 was located in a region of five residues extending from positions 352 to 357, which included four sites predicted to be involved in functional divergence, two of them also predicted to be fixed by positive selection. We also found an additional positive selection site (M462) in the proximity of the GAP binding site. Some other sites are also highly conserved in DXS sequences, but variable in Brassicaceae DXS2, or vice versa, maybe suggesting functional shifts (Supplementary Fig. 2).

Therefore, it is likely that these changes could modify the spatial and/or the electrostatic environment of the catalytic site affecting the interactions with the substrate. This observation would be in agreement with the proposal that DXS2 may have evolved to a novel biochemical function through changes affecting substrate specificity

Fig. 7. 3D-structural model of *A. thaliana* DXS2. (A) Front and (B) back molecular surfaces of DXS2. (C) Cartoon representation of the TPP-binding pocket. (D) Cartoon representation of the GAP-binding domain. Residues predicted as responsible for functional divergence are shown in green, positions putatively fixed by positive selection in red, and positions both predicted as related to functional divergence and fixed by PS in orange. GAP-binding residues are colored cyan and TPP-binding residues are displayed in marine. The images were produced with PyMol (Delano, 2002).

while retaining the same TPP-dependent reaction mechanism. Moreover, it is likely that particular amino acid changes in DXS2 located close to the GAP and TPP binding site residues could also be related with the loss of DXS activity. This hypothesis is supported by results derived from the molecular characterization of *E. coli* mutants defective in DXS activity (Sauret-Gueto et al., 2003). None of the identified mutants harbored mutations directly affecting the residues currently known to play a role in the active site of DXS. Instead, most of the mutations affected residues located close (+/− 1–2 positions) to the GAP and TPP binding site residues described in *E. coli* DXS (Xiang et al., 2007). Moreover, in spite that the structural and/or functional role of the internal and C-terminal sequences deleted in DXS2 (Fig. 1) is currently unknown, their involvement in the loss of DXS activity and putative gain of a new biochemical function cannot be ruled out.

The loss of DXS activity in *A. thaliana* DXS2 raises a nomenclature issue concerning the naming of the protein and the corresponding gene. Thus, following the recommendations of Phillips et al. (2008), and until the biological function of DXS2 is known, we propose to rename DXS2 as DXL1 (for DXS-like 1) and *DXL1* as the corresponding gene.

Enzyme families involved in plant secondary metabolism pathways provide interesting examples of the role of divergent selective constraints, relaxed purifying selection and positive selection in promoting functional specialization following gene duplication featured by distinct substrate specificity and/or reaction enzymatic products. Well documented examples include chalcone synthases and dihydroflavonol-4-reductases (DFR), involved in flavonoid biosynthesis (Des Marais and Rausher, 2008), shikimate kinases, participating in the shikimate pathway (Fucile et al., 2008) and methylthioalkylmalate synthases, controlling an early step in the biosynthesis of glucosinolates (Benderoth et al., 2006). Results presented here provide the first evidences supporting functional specialization at both the expression and biochemical level within the plant DXS family and reveal the occurrence of lineage-specific gene duplications. The plant DXS family emerges as an interesting model to examine the molecular evolutionary basis of plant secondary metabolism diversification.

Authors' contributions

LCP conducted the expression analysis, complementation assays and subcellular location studies. AC characterized the *A. thaliana* DXS2 mutant. LCP performed the sequence, phylogenetic, molecular evolutionary and functional divergence analyses. DT carried out the building of a 3D-structural model. AS and SI performed the heterologous expression and purification of recombinant DXS1 and DXS2 and enzyme assays. LCP, MRC, NC and AB conceived of the study, participated in its design and coordination and helped to draft and refine the manuscript. LCP and AB wrote and prepared the manuscript. All authors read and approved the final manuscript.

Uncited reference

Davis et al., 2007

Acknowledgments

We thank Drs. Françoise Paulet-Dubois and David L. Robertson for the critical reading of the manuscript. We also thank all our laboratory members for stimulating discussions and suggestions. LCP and AC were co-financed by the Spanish Ministry of Education and Science and the European Social Fund ("Juan de la Cierva", and "Formación del Personal Investigador" programs, respectively). This work was supported by grants from the Spanish Ministry of Education and Science and the Spanish Ministry of Science and Innovation (BIO2006-03704 and BIO2009-09523 to A.B., BFU2011-24208 to N.C., including FEDER funds) and the Generalitat de Catalunya (grant 2009SGR-0026). This work

has been carried out within the University of Almería, the University of Barcelona, the Centre Consolider-Ingenio for Research in Agrigenomics (CSD2007-00036) funded by the Spanish Ministry of Science and Innovation and the University of Manchester.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2012.10.071>.

References

- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105.
- Alonso, J.M., et al., 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.
- Altincicek, B., Hintz, M., Sanderbrand, S., Wiesner, J., Beck, E., Jomaa, H., 2000. Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: an approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* 190, 329–333.
- Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552.
- Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18, 1585–1592.
- Araki, N., Kusumi, K., Masamoto, K., Niwa, Y., Iba, K., 2000. Temperature-sensitive *Arabidopsis* mutant defective in 1-deoxy-d-xylulose 5-phosphate synthase within the plastid non-mevalonate pathway of isoprenoid biosynthesis. *Physiol. Plant.* 108, 19–24.
- Axelos, M., Curie, C., Mazzolini, L., Bardet, C., Lescure, B., 1992. A protocol for transient gene expression in *Arabidopsis thaliana* protoplasts. *Plant Physiol. Biochem.* 30 (1), 123–128.
- Bailey, A.M., Mahapatra, S., Brennan, P.J., Crick, D.C., 2002. Identification, cloning, purification, and enzymatic characterization of *Mycobacterium tuberculosis* 1-deoxy-D-xylulose 5-phosphate synthase. *Glycobiology* 12, 813–820.
- Barkman, T.J., 2003. Evidence for positive selection on the floral scent gene isoeugenol-O-methyltransferase. *Mol. Biol. Evol.* 20, 168–172.
- Benderoth, M., Textor, S., Windsor, A.J., Mitchell-Olds, T., Gershenzon, J., Kroymann, J., 2006. Positive selection driving diversification in plant secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9118–9123.
- Bielawski, J.P., Yang, Z., 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.* 59, 121–132.
- Blanc, G., Hokamp, K., Wolfe, K.H., 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* 13, 137–144.
- Botella-Pavia, P., Besumbes, O., Phillips, M.A., Carretero-Paulet, L., Boronat, A., Rodriguez-Concepcion, M., 2004. Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. *Plant J.* 40, 188–199.
- Bouvier, F., d'Harlingue, A., Suire, C., Backhaus, R.A., Camara, B., 1998. Dedicated roles of plastid transketolases during the early onset of isoprenoid biogenesis in pepper fruits. *Plant Physiol.* 117, 1423–1431.
- Browner, M.F., Rasor, P., Tugendreich, S., Fletterick, R.J., 1991. Temperature-sensitive production of rabbit muscle glycogen phosphorylase in *Escherichia coli*. *Protein Eng.* 4, 351–357.
- Campos, N., Rodriguez-Concepcion, M., Sauret-Gueto, S., Gallego, F., Lois, L.M., Boronat, A., 2001. *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *Biochem. J.* 353, 59–67.
- Carretero-Paulet, L., 2003. Molecular characterization of 1-deoxy-D-xylulose 5-phosphate synthase and 1-deoxy-D-xylulose 5-phosphate reductoisomerase genes of *Arabidopsis thaliana*. *Biochemistry and Molecular Biology*. Universidad de Barcelona, Barcelona, p. 270.
- Carretero-Paulet, L., et al., 2002. Expression and molecular analysis of the *Arabidopsis* DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. *Plant Physiol.* 129, 1581–1591.
- Carretero-Paulet, L., et al., 2006. Enhanced flux through the methylerythritol 4-phosphate pathway in *Arabidopsis* plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Mol. Biol.* 62, 683–695.
- Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.
- Chahed, K., et al., 2000. 1-Deoxy-D-xylulose 5-phosphate synthase from periwinkle: cDNA identification and induced gene expression in terpenoid indole alkaloid-producing cells. *Plant Physiol. Biochem.* 38, 559–566.
- Chappell, J., 1995. Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46, 521–547.
- Conant, G.C., Wolfe, K.H., 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950.
- Cordoba, E., Salmi, M., Leon, P., 2009. Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60, 2933–2943.

- Cordoba, E., et al., 2011. Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. *J. Exp. Bot.* 62, 2023–2038.
- Couvreur, T.L., Franzke, A., Al-Shehbaz, I.A., Bakker, F.T., Koch, M.A., Mummenhoff, K., 2010. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* 27, 55–71.
- Croteau, R., Kitchan, T.M., Lewis, N.G., 2000. Secondary metabolites. In: Buchanan, B., W.G., Jones, R. (Eds.), *Biochemistry & Molecular Biology of Plants: American Society of Plant Physiologists*, pp. 1250–1318.
- Crowell, D.N., Packard, C.E., Pierson, C.A., Giner, J.L., Downes, B.P., Chary, S.N., 2003. Identification of an allele of CLA1 associated with variegation in *Arabidopsis thaliana*. *Physiol. Plant.* 118, 29–37.
- Davis, I.W., et al., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383.
- Delano, W., 2002. PyMOL. DeLano Scientific, San Carlos, CA.
- Des Marais, D.L., Rausher, M.D., 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454, 762–765.
- Durbin, M.L., McCaig, B., Clegg, M.T., 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol. Biol.* 42, 79–92.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Estevez, J.M., et al., 2000. Analysis of the expression of CLA1, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol-4-phosphate pathway in *Arabidopsis*. *Plant Physiol.* 124, 95–104.
- Estevez, J.M., Cantero, A., Reindl, A., Reichler, S., Leon, P., 2001. 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* 276, 22901–22909.
- Fucile, G., Falconer, S., Christendat, D., 2008. Evolutionary diversification of plant shikimate kinase gene duplicates. *PLoS Genet.* 4, e1000292.
- Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11, 725–736.
- Goodstein, D.M., et al., 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186.
- Gu, X., 1999. Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.* 16, 1664–1674.
- Gu, X., 2001. Maximum-likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* 18, 453–464.
- Gu, X., 2006. A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol. Biol. Evol.* 23, 1937–1945.
- Gu, X., Vander Velden, K., 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18, 500–501.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Hahn, F.M., Eubanks, L.M., Testa, C.A., Blagg, B.S., Baker, J.A., Poulter, C.D., 2001. 1-Deoxy-D-xylulose 5-phosphate synthase, the gene product of open reading frame (ORF) 2816 and ORF 2895 in *Rhodobacter capsulatus*. *J. Bacteriol.* 183, 1–11.
- Hemmerlin, A., et al., 2003. Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.* 278, 26666–26676.
- Hernandez-Hernandez, T., Martinez-Castilla, L.P., Alvarez-Buylla, E.R., 2007. Functional diversification of B MADS-box homeotic regulators of flower development: Adaptive evolution in protein-protein interaction domains after major gene duplication events. *Mol. Biol. Evol.* 24, 465–481.
- Hsieh, M.H., Chang, C.Y., Hsu, S.J., Chen, J.J., 2008. Chloroplast localization of methylerythritol 4-phosphate pathway enzymes and regulation of mitochondrial genes in *ispD* and *ispE* albino mutants in *Arabidopsis*. *Plant Mol. Biol.* 66, 663–673.
- Hubbard, S., Thornton, J., 1993. NACCESS. Department of Biochemistry and Molecular Biology. University College London, London.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Jones, S., Thornton, J.M., 1997. Analysis of protein–protein interaction sites using surface patches. *J. Mol. Biol.* 272, 121–132.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- Kasahara, H., Hanada, A., Kuzuyama, T., Takagi, M., Kamiya, Y., Yamaguchi, S., 2002. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. *J. Biol. Chem.* 277, 45188–45194.
- Kim, Y.B., et al., 2009. Regulation of resin acid synthesis in *Pinus densiflora* by differential transcription of genes encoding multiple 1-deoxy-D-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase genes. *Tree Physiol.* 29, 737–749.
- Lange, B.M., Wildung, M.R., McCaskill, D., Croteau, R., 1998. A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2100–2104.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.
- Laule, O., et al., 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6866–6871.
- Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320.
- Li, W., Liu, B., Yu, L., Feng, D., Wang, H., Wang, J., 2009. Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytyldienoate acid reductase gene family in plants. *BMC Evol. Biol.* 9, 90.
- Lichtenthaler, H.K., 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 47–65.
- Lois, L.M., Campos, N., Putra, S.R., Danielsen, K., Rohmer, M., Boronat, A., 1998. Cloning and characterization of a gene from *Escherichia coli* encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2105–2110.
- Lois, L.M., Rodriguez-Concepcion, M., Gallego, F., Campos, N., Boronat, A., 2000. Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase. *Plant J.* 22, 503–513.
- Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155.
- Mandel, M.A., Feldmann, K.A., Herrera-Estrella, L., Rocha-Sosa, M., Leon, P., 1996. CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. *Plant J.* 9, 649–658.
- Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A., 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325.
- Moore, R.C., Purugganan, M.D., 2005. The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128.
- Ohno, S., 1970. *Evolution by Gene Duplication*. Springer-Verlag.
- Oudin, A., et al., 2007. Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Mol. Biol.* 65, 13–30.
- Phillips, M.A., et al., 2007. Functional identification and differential expression of 1-deoxy-D-xylulose 5-phosphate synthase in induced terpeneoid resin formation of Norway spruce (*Picea abies*). *Plant Mol. Biol.* 65, 243–257.
- Phillips, M.A., Leon, P., Boronat, A., Rodriguez-Concepcion, M., 2008. The plastidial MEP pathway: unified nomenclature and resources. *Trends Plant Sci.* 13, 619–623.
- Proost, S., et al., 2009. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* 21, 3718–3731.
- Rodriguez-Concepcion, M., 2006. Early steps in isoprenoid biosynthesis: multilevel regulation of the supply of common precursors in plant cells. *Phytochem. Rev.* 5, 1–15.
- Rodriguez-Concepcion, M., Boronat, A., 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130, 1079–1089.
- Rodriguez-Concepcion, M., et al., 2001. 1-Deoxy-D-xylulose 5-phosphate reductoisomerase and plastid isoprenoid biosynthesis during tomato fruit ripening. *Plant J.* 27, 213–222.
- Sauret-Gueto, S., Ramos-Valdivia, A., Ibanez, E., Boronat, A., Rodriguez-Concepcion, M., 2003. Identification of lethal mutations in *Escherichia coli* genes encoding enzymes of the methylerythritol phosphate pathway. *Biochem. Biophys. Res. Commun.* 307, 408–415.
- Skorupinska-Tudek, K., et al., 2008. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of dolichols in plants. *J. Biol. Chem.* 283, 21024–21035.
- Walter, M.H., Hans, J., Strack, D., 2002. Two distantly related genes encoding 1-deoxy-d-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant J.* 31, 243–254.
- Word, J.M., Lovell, S.C., Richardson, J.S., Richardson, D.C., 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285, 1735–1747.
- Xiang, S., Usunow, G., Lange, G., Busch, M., Tong, L., 2007. Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *J. Biol. Chem.* 282, 2676–2682.
- Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917.
- Yang, Z., Yang, J., Huang, J., Gu, H., Zhong, Y., 2002. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). *Mol. Biol. Evol.* 19, 1752–1759.
- Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118.
- Yeh, B.K., et al., 2003. Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors. *Proc. Natl. Acad. Sci. U. S. A.* 100, 2266–2271.
- Yin, X., Proteau, P.J., 2003. Characterization of native and histidine-tagged deoxyxylulose 5-phosphate reductoisomerase from the cyanobacterium *Synechocystis* sp. PCC6803. *Biochim. Biophys. Acta* 1652, 75–81.
- Zhang, J., 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298.
- Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479.