Characterization of the GGPP synthase gene family in Arabidopsis thaliana

Author(s):
Beck, Gilles; Coman, Diana; Herren, Edgar; Ruiz-Sola, M. Aguila; Rodriguez-Concepcion, Manuel; Gruissem, Wilhelm; Vranova, Eva

Publication date:
2013-07

Permanent link:
https://doi.org/10.3929/ethz-b-000068554

Originally published in:
Plant Molecular Biology 82(4-5), https://doi.org/10.1007/s11103-013-0070-z

Plant Mol Biol (2013) 82:393–416
DOI 10.1007/s11103-013-0070-z

Characterization of the *GGPP* synthase gene family in *Arabidopsis thaliana*

Gilles Beck · Diana Coman · Edgar Herren · M. Águila Ruiz-Sola · Manuel Rodríguez-Concepción · Wilhelm Gruissem · Eva Vranová

Received: 1 February 2013 / Accepted: 5 May 2013 / Published online: 1 June 2013

Abstract Geranylgeranyl diphosphate (GGPP) is a key precursor of various isoprenoids that have diverse functions in plant metabolism and development. The annotation of the *Arabidopsis thaliana* genome predicts 12 genes to encode geranylgeranyl diphosphate synthases (GGPPS). In this study we analyzed GGPPS activity as well as the subcellular localization and tissue-specific expression of the entire protein family in *A. thaliana*. GGPPS2 (At2g18620), GGPPS3 (At2g18640), GGPPS6 (At3g14530), GGPPS7 (At3g14550), GGPPS8 (At3g20160), GGPPS9 (At3g29430), GGPPS10 (At3g32040) and GGPPS11 (At4g36810) showed GGPPS activity in *Escherichia coli*, similar to activities reported earlier for GGPPS1 (At1g49530) and GGPPS4 (At2g23800) (Zhu et al. in Plant Cell Physiol 38(3):357–361, 1997a; Plant Mol Biol 35(3):331–341, b). GGPPS12 (At4g38460) did not produce GGPP in *E. coli*. Based on DNA sequence analysis we propose that *GGPPS5* (*At3g14510*) is a pseudogene. GGPPS–GFP (green fluorescent protein) fusion proteins of the ten functional GGPP synthases localized to plastids, mitochondria and the endoplasmic reticulum, with the majority of the enzymes located in plastids. Gene expression analysis using quantitative real time-PCR, *GGPPS promoter-GUS* (*β-glucuronidase*) assays and publicly available microarray data revealed a differential spatio-temporal expression of *GGPPS* genes. The results suggest that plastids and mitochondria are key subcellular compartments for the synthesis of ubiquitous GGPP-derived isoprenoid species. GGPPS11 and GGPPS1 are the major isozymes responsible for their biosynthesis. All remaining paralogs, encoding six plastidial isozymes and two cytosolic isozymes, were expressed in specific tissues and/or at specific developmental stages, suggesting their role in developmentally regulated isoprenoid biosynthesis. Our results show that of the 12 predicted GGPPS encoded in the *A. thaliana* genome 10 are functional proteins that can synthesize GGPP. Their specific subcellular location and

The gene ID numbers of the GGPPS characterized in this study are: GGPPS1 (GGPPS6 in Zhu et al. 1997b; Okada et al. 2000) is At1g49530; GGPPS2 is At2g18620; GGPPS3 (GGPPS4 in Okada et al. 2000) is At2g18640; GGPPS4 (GGPPS5 in Zhu et al. 1997a; GGPPS2 in Okada et al. 2000) is At2g23800; GGPPS5 is At3g14510; GGPPS6 is At3g14530; GGPPS7 (GGPPS3 in Okada et al. 2000) is At3g14550; GGPPS8 is At3g20160; GGPPS9 is At3g29430; GGPPS10 is At3g32040; GGPPS11 (GGPPS1 in Okada et al. 2000) is At4g36810; GGPPS12 (GGR in Okada et al. 2000) is At4g38640.

Gilles Beck, Diana Coman contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-013-0070-z) contains supplementary material, which is available to authorized users.

G. Beck · D. Coman · E. Herren · W. Gruissem · E. Vranová
Department of Biology, Plant Biotechnology, ETH Zurich, Universitaetstrasse 2, 8092 Zurich, Switzerland
e-mail: beckg@ethz.ch

D. Coman
e-mail: dcoman@ethz.ch

E. Herren
e-mail: eherren@ethz.ch

W. Gruissem
e-mail: wgruissem@ethz.ch

M. Á. Ruiz-Sola · M. Rodríguez-Concepción
Department of Molecular Genetics, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, 08193 Barcelona, Spain
e-mail: aguila.ruiz@cragenomica.es

M. Rodríguez-Concepción
e-mail: manuel.rodriguez@cragenomica.es

differential expression pattern suggest subfunctionalization in providing GGPP to specific tissues, developmental stages, or metabolic pathways.

Keywords Arabidopsis · Isoprenoids · Branchpoint · Prenyl diphosphate synthase · Geranylgeranyl diphosphate synthase

Abbreviations

ABA	Abscisic acid
DMAPP	Dimethylallyl diphosphate
ER	Endoplasmic reticulum
FPP	Farnesyl diphosphate
GA	Gibberellic acid
GFP	Green fluorescent protein
GGPP	Geranylgeranyl diphosphate
GGPPS	Geranylgeranyl diphosphate synthase
GPP	Geranyl diphosphate
GUS	β-Glucuronidase
IPP	Isopentenyl diphosphate
MEP	Methylerythritol
MVA	Mevalonate

Introduction

Short chain prenyl diphosphate synthases are enzymes of the isoprenoid pathway that use isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), basic building blocks of the isoprenoid pathway synthesized by either the mevalonate (MVA) or the methylerythritol (MEP) pathway, to produce central intermediates in the isoprenoid metabolism. Short chain prenyl diphosphates are further recruited by branch point enzymes to synthesize different isoprenoid end products. Short chain prenyl diphosphate synthases are represented in plants by three enzymes—geranyl diphosphate (GPP) synthase, farnesyl diphosphate (FPP) synthase and geranylgeranyl diphosphate (GGPP) synthase. These enzymes localize to all compartments where the biosynthesis of isoprenoids takes place, e.g., cytosol, ER, mitochondria and plastids (Fig. 1), and their localization seems to be associated with the need of each compartment for a particular prenyl diphosphate. For example, *A. thaliana* FPP synthases localize to the cytosol and the mitochondria because isoprenoids such as sterols, brassinosteroids, triterpenoids and sesquiterpenoids that originate from FPP are synthesized in the cytosol/ER and mitochondria but not in plastids (Vranová et al. 2011; Fig. 1). The localization of prenyl diphosphate synthases is not only influenced by the need for the specific substrate in the given compartment but the subcellular localization of prenyl diphosphate synthases themselves can influence the outcome of biosynthesis. For example, terpene synthases can synthesize either monoterpenes or sesquiterpenes depending on the subcellular availability of GPP or FPP as a substrate (Huang et al. 2010). Therefore, a full understanding of the function of prenyl diphosphate synthases in the plant isoprenoid network requires a comprehensive analysis of their subcellular localization, their temporal and spatial expression, as well as experimental confirmation of their activity. Prenyl diphosphate synthases are highly similar at the amino acid sequence level and their activity cannot be predicted solely based on homology (Okada et al. 2000; Wang and Dixon 2009; Hsieh et al. 2011). The function of predicted *A. thaliana* FPP and GPP synthases was reported elsewhere (Delourme et al. 1994; Cunillera et al. 1996, 1997; Bouvier et al. 2000; van Schie et al. 2007; Wang and Dixon 2009; Closa et al. 2010). In this manuscript we focus on the functional characterization of the GGPP synthase (GGPPS) gene family in the model plant *A. thaliana.*

GGPP is a central precursor for the synthesis of primary and secondary isoprenoid compounds such as chlorophylls, carotenoids and derivatives including the hormones abscisic acid (ABA) and strigolactones, gibberellins, plastoquinones, ubiquinones, phylloquinones, tocopherols, diterpenoids, polyprenols, dolichols, and prenylated proteins (Fig. 1). GGPPS functions as a homodimer and catalyzes successive additions of IPP to DMAPP, GPP and FPP (Vandermoten et al. 2009). GGPPS genes have been cloned from a number of organisms such as bacteria (Ohnuma et al. 1994), yeast (Jiang et al. 1995), fungi (Sandmann et al. 1993), plants (Okada et al. 2000), mammals (Kainou et al. 1999) and insects (Hojo et al. 2007). In higher plants, GGPPS is usually encoded by gene paralogs, forming a GGPPS gene family of two to twelve members in diverse plant genomes (http://bioinformatics.psb.ugent.be/plaza/; Proost et al. 2009). In *A. thaliana*, twelve isozymes were predicted by sequence similarity (Lange and Ghassemian 2003). However, GGPPS12 (At4g38460) was shown in two independent studies to lack GGPPS activity in vitro (Okada et al. 2000; Wang and Dixon 2009) and rather has GPPS activity when coexpressed in vitro with the catalytic large subunit of the heteromeric GPPS (Wang and Dixon 2009). The information for the other GGPPS genes and isozymes (1–11) remains incomplete because studies published to date have only included specific members of the large gene family.

The GGPPS activity of six isozymes was confirmed by in vitro enzymatic assay and/or by genetic complementation of *E. coli* (Zhu et al. 1997a, b; Okada et al. 2000;

Present Address:
E. Vranová (✉)
Faculty of Science, Institute of Biology and Ecology,
P. J. Šafárik University Košice, Mánesova 23,
041 54 Košice, Slovakia
e-mail: eva.vranova@upjs.sk

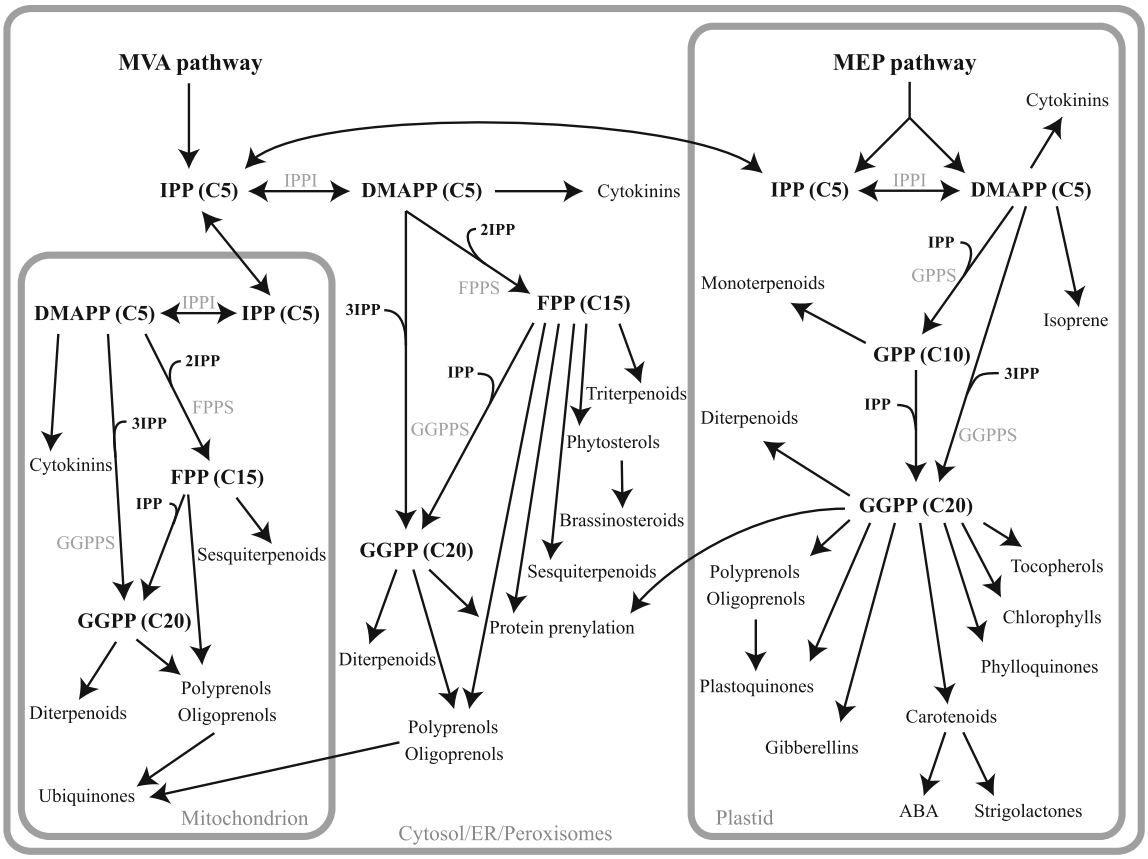


Fig. 1 Subcellular compartmentalization of isoprenoid biosynthesis in *A. thaliana*. Based on the pathway network constructed by Vranová et al. (2011). Enzymes are shown in *grey* and isoprenoids in *black*. Abbreviations are as follows: MVA: mevalonic acid; MEP: methylerythritol phosphate; IPP: isopentenyl diphosphate; DMAPP: dimethylallyl diphosphate; GPP: geranyl diphosphate; FPP: farnesyl diphosphate; GGPP: geranylgeranyl diphosphate; ABA: abscisic acid; IPPI: isopentenyl diphosphate isomerase; GPPS: geranyl diphophate synthase; FPPS: farnesyl diphosphate synthase; GGPPS: geranylgeranyl diphosphate synthase. The AGI numbers of the twelve putative GGPPS paralogs are as follows: GGPPS1, At1g49530; GGPPS2, At2g18620; GGPPS3, At2g18640; GGPPS4, At2g23800; GGPPS5, At3g14510; GGPPS6, At3g14530; GGPPS7, At3g14550; GGPPS8, At3g20160; GGPPS9, At3g29430; GGPPS10, At3g32040; GGPPS11, At4g36810; GGPPS12, At4g38640

Wang and Dixon 2009). In *A. thaliana*, all predicted GGPP synthases have a putative localization signal for translocation into different subcellular compartments, such as chloroplasts, ER and mitochondria (TargetP, http://www.cbs.dtu.dk/services/TargetP/; (Emanuelsson et al. 2000) and PSORT, http://psort.hgc.jp/form.html). For GGPPS1, 3, 4, 7 and 11, the localization was confirmed using transit peptide-GFP fusion proteins targeting individual enzymes to the plastids (GGPPS7, 11), mitochondria (GGPPS1) and ER (GGPPS3, 4) (Zhu et al. 1997b; Okada et al. 2000). In addition to different subcellular localization, the analyzed genes showed differential spatio-temporal expression both by Northern analysis and expression of the β-glucuronidase (GUS) gene under the control of GGPPS promoters (Okada et al. 2000). GGPPS11, encoding a plastidial GGPPS, was expressed throughout the plant, except of roots, while the activity of GGPPS7 promoter was restricted to the hypocotyl and the vascular tissue of roots. GGPPS3 and GGPPS4, both encoding ER-targeted proteins, were expressed in the vascular tissue, flowers, stamens and root tips (Okada et al. 2000). Considering the importance of GGPP as a key precursor for isoprenoid biosynthesis, we therefore performed a comprehensive analysis of enzyme activity, subcellular localization and tissue-specific expression for the entire *A. thaliana* GGPPS family members.

We established that the twelve-member gene family produces ten functional proteins that can synthesize GGPP (GGPPS1–4, 6–11). We confirmed the synthesis of GGPP in the ER, mitochondria and plastids, and demonstrated that the majority of GGPP synthases are plastid isozymes. In addition, the expression analysis of individual paralogs and spatio-temporal distributions of their transcripts showed that GGPPS11, encoding a plastidial isozyme, and GGPPS1, encoding the mitochondrial isozyme, were ubiquitously expressed throughout the whole plant in almost all tissues. Expression of the remaining paralogs, which encode six plastid and two cytosolic isozymes, was

restricted to specific tissues and/or specific developmental stages.

Materials and methods

Plant material and growth conditions

The *A. thaliana* Col0 accession was used in this study. Plants were grown either on basic Murashige-Skoog (MS) medium (Duchefa, www.duchefa.com) containing 0.8 % w/v plant agar or on soil in a climate-controlled growth chamber under long-day conditions (16 h light/8 h dark) at 22 °C.

In vivo activity assay of GGPPS proteins

The *crt* cluster responsible for the synthesis of lycopene in *Erwinia uredovora* has been cloned and widely used for complementation assays in *E. coli*. The pACCRT-BI plasmid was constructed from pACCRT-EBI (Misawa et al. 1990) by introducing a frameshift in the *Bst*XI site of the *crtE* gene encoding GGPP synthase. After digestion of pACCRT-EIB with *Bst*XI, the overhangs were filled with the Klenow fragment of DNA polymerase I to create blunt ends. After ligation and transformation with the resulting pACCRT-BI construct, positive transformants were identified by their absence of pigmentation compared to those transformed with pACCRT-EBI in which a functional crtE protein produces GGPP to further synthesize lycopene, resulting in red colonies.

pGEX-GGPPS vectors were constructed via ligation of the PCR-amplified GGPPS sequences into the *Bam*HI (SmaI for GGPPS6) and *Not*I sites of the protein expression pGEX-4T-2 vector (GE Healthcare, http://www.gehealthcare.com). PCR was performed using pENTR/D-TOPO-GGPPS-3‘ (see Subcellular Localization of GGPPS Proteins) vectors as template and primers listed in the Supplemental Table S1. The resulting constructs consisted of the fusion between the glutathione-S-transferase (GST) and truncated GGPPSs (Fig. 4b) controlled by the *tac* promoter. All constructs were checked by sequencing.

E. coli cells (Invitrogen; www.invitrogen.com) were co-transformed with both pGEX-GGPPS constructs and pACCRT-BI. Transformants containing both constructs were selected on LB plates containing both ampicillin (100 μg mL^{-1}) and chloramphenicol (25 μg mL^{-1}) antibiotics. Positive colonies were selected and grown overnight in liquid LB. The next day, 20 mL of fresh LB cultures supplemented with the appropriate antibiotics were inoculated, grown for 3 days at 20 °C and harvested. 10 mL were used for the lycopene extraction. In brief, pelleted cells were broken by vortexing for 30 s and 700 μL of acetone were added. Samples were vortexed for 30 more seconds and then incubated in the dark at 55 °C for 15 min. The tubes were centrifuged for 15 min at 4 °C at 14,000 rpm. Supernatants were collected in glass spectrophotometer cuvettes and absorbance was measured at 472 nm.

Subcellular localization of GGPPS proteins

Coding sequences of *A. thaliana* genes for GGPPS isozymes were amplified without their stop codon (the sequences of the corresponding oligonucleotides are available in the Supplemental Table S2) from total Arabidopsis cDNA and cloned into the Gateway-compatible pENTR/D-TOPO entry vector (Invitrogen, http://www.invitrogen.com) resulting in pENTR/D-TOPO-GGPPS-3′ vectors. Constructs were checked by DNA sequencing. LR reactions were then performed to clone these cDNAs into the binary pK7FWG2.0 vector (http://gateway.psb.ugent.be ; Karimi et al. 2005). The resulting constructs pGGPPS-eGFP consisted of the *GGPPS-eGFP* fusions under the control of the cauliflower mosaic virus (CaMV) *35S promoter*. Constructs were introduced into *Agrobacterium tumefaciens* (strain C58C1(pMP90); Koncz and Shell 1986) and then into wild-type Arabidopsis plants via Agrobacterium-mediated floral dip transformation (Clough and Bent 1998). Primary transformants were selected on MS medium supplemented with kanamycin (50 μg mL^{-1}). At least two stable lines per isozyme were selected for further experiments. These lines were subsequently transformed by floral dipping with mCherry organelle reporter constructs (Nelson et al. 2007) as follows: GGPPS1-GFP lines with the mitochondrial reporter construct (CD3-986); GGPPS3-GFP and GGPPS4-GFP lines with the ER reporter construct (CD3-954); GGPPS2-, 6-, 7-, 8-, 9-, 10-, 11- and 12-GFP with the plastidial (CD3-994) reporter constructs. Double transformants were selected by both kanamycin (50 μg mL^{-1}) and basta (20 μg mL^{-1}) antibiotic resistance. Leaves of young (10–15 day old) seedlings were analyzed under a Leica SP2-AOBS confocal laser-scanning microscope. eGFP was excited at 488 nm and its emission signal was collected between 500 and 550 nm. mCherry was excited at 514 nm and its emission signal was collected between 602 and 635 nm. Chlorophyll was excited at 405 nm and its emission signal was collected between 655 and 712 nm.

Relative transcript quantification

Plant material was collected from seven different organs, namely: roots, rosette leaves, cauline leaves, stems, flowers, seedlings and siliques. The root samples were collected from

18 day-old seedlings grown on MS medium. The rosette leaves, cauline leaves, stems, flowers and siliques were collected as a pool from 6 week-old plants grown in parallel in soil under long day conditions. The seedlings were grown on standard MS medium and collected 14 days after germination. The plant material was snap-frozen in liquid nitrogen immediately after collection and stored at −80 °C until use. RNA was isolated from plant tissues using the TRIZol reagent (Invitrogen, http://www.invitrogen.com) according to the manufacturer's instructions. Equal RNA amounts, quantified using a NanoDrop instrument (Thermo, http://www.nanodrop.com) were treated with RQ1 RNase-Free DNase (Promega, http://www.promega.com). cDNA was synthesized in a 20 µl reaction from 1.5 µg RNA using oligo-dT primers and RevertAid^{TM} First Strand cDNA synthesis Kit (Thermo, http://www.fermentas.com) following the manufacturer's recommendations. The absence of genomic DNA in cDNAs was verified by PCR with primers ACT-s 5′ TCCACGAGACAACCTATAAC and ACT-a 5′ GATCTTGAGAGCTTAGAAAC, spanning the second intron of *ACT2* (*At3g18780*) and visualized on 1 % agarose gel (data not shown). The relative quantification of transcripts (RT-qPCR) was performed with the Applied Biosystems 7500 Fast Real-Time PCR System (Applied Biosystems, http://www.appliedbiosystems.com) using a hydrolysis probe based assay (TaqMan, Roche, http://www.roche-applied-science.com) and the FastStart TaqMan® Probe Master Mix (Roche, http://www.roche-applied-science.com) with 30 ng of starting template, according to the manufacturer's instructions. The oligonucleotides and probes used for RT-qPCR are listed in Supplemental Table S3. The experimental design included three biological replicates and three technical replicates for each reaction being carried out.

A gene was considered to be expressed when the corresponding quantification cycle (*Cq*) value was below or equal to 35 (Karlen et al. 2007). The primer efficiencies were estimated from the raw fluorescence data (*ΔRn*) for each reaction over all PCR cycles using the LinRegPCR software (Ruijter et al. 2009). The *PavrgE* method (Karlen et al. 2007) was used to calculate an average PCR efficiency (*E*). Valid primer efficiencies ($1.75 < E < 2.25$) were included in subsequent calculations. The transcript amount of each *GGPPS* gene was normalized based on the expression levels of three reference genes (Rieu and Powers 2009). The reference genes, namely: *PP2A* (*At1g13320*), *UBC9* (*At4g27960*) and *ACT2* (*At3g18780*), were selected under the expression stability criterion across different organ types of *A. thaliana* (Czechowski et al. 2005). The expression levels of the reference genes cover the range of expression of the *GGPPS* genes (data not shown) as indicated by the RefGenes Tool from Genevestigator® (Hruz et al. 2011). The expression stability of the three reference genes in our experimental settings was tested using the geNorm software (http://medgen.ugent.be/~jvdesomp/genorm). All three reference genes received valid expression stability scores (data not shown) and were further used to estimate the normalization factor by calculating the geometric average (Vandesompele et al. 2002). To compare the differences in transcript levels of the *GGPPS* genes between different organ samples the following formula was used:

$$NQ_{sample\,x} = \frac{E_{GGPPSx}^{-Cq}}{E_{\mathrm{Ref}}^{-Cq}}$$

where $NQ_{sample\ x}$ is calculated for each *GGPPS* gene and represents the transcript quantity relative to the normalization factor in one organ type.

To compare the different expression of each *GGPPS* gene in different organs, the relative quantities were calculated. The *Cq* values were transformed to relative quantities (*RQ*) as follows: for each of the *GGPPS* and reference genes the *Cq* values corresponding to one gene in different organs were subtracted from the minimal *Cq* value of the respective gene (i.e., the maximal expression value) according to the formula:

$$\Delta Cq = \min Cq_{gene\,x} - Cq_{gene\,x}$$

$$RQ_{gene\,x} = E^{\Delta Cq}$$

where $RQ_{gene\ x}$ is the transcript amount of one gene in a certain organ sample relative to the sample with the highest expression. The $RQ_{gene\ x}$ for each *GGPPS* gene was normalized as described before and the normalized relative quantities (*NRQ*) were calculated according to the modified Pfaffl method (Pfaffl 2001) using the formula, which takes into accounts the different PCR efficiencies:

$$NRQ = \frac{RQ_{GGPPS}}{RQ_{Ref}}$$

The *NRQ* ratios were subsequently log2 transformed. The mRNA levels for each *GGPPS* gene across different organs are relative to the highest expression potential for the respective gene and are normalized to the reference genes. The data represent the mean and standard error of three biological replicates.

GGPPS promoter-GUS constructs

To clone the transcriptional regulatory elements the sequences located upstream of the ATG start codon were amplified by PCR using *A. thaliana* genomic DNA as template. In brief, the genomic DNA was extracted from seedlings using the Nucleon Phytopure system (Amersham Biosciences, http://www.amershambiosciences.com) according

to the manufacturer's instructions. To amplify the fragment of interest, the Phusion High-Fidelity DNA polymerase kit (Finnzymes, http://www.finnzymes.fi) was used. The oligonucleotides used for the amplification can be found in the Supplemental Table S4.

The amplicons were cloned into the entry vector pENTR/D-TOPO (Invitrogen, http://www.invitrogen.com) according to the manufacturer's instructions resulting in pENTR/D-TOPO-*GGPPSpro* vectors. Plant expression vectors were obtained by performing the Gateway LR reaction (Invitrogen, www.invitrogen.com) according to the manufacturer's instructions, between the entry vectors and the *GUS*-containing reporter vector pHGWFS7 (http://gateway.psb.ugent.be/; Karimi et al. 2005) resulting in *pGGPPSpro:GUS* vectors. The *Agrobacterium tumefaciens* strain C58C1 (pMP90; Koncz and Shell 1986) was transformed with the resulting binary vectors and transgenes were subsequently introduced into Arabidopsis via the floral dip method (Clough and Bent 1998). At least two stable lines per isozyme were selected for further experiments.

Visualization of GUS activity

Plant material was harvested and submerged in the X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid) solution, containing 1 mg mL^{-1} X-Gluc, 100 mM sodium phosphate buffer pH 7.0, 0.2 % v/v Triton-X and 10 mM EDTA. The samples were incubated in the dark at 37 °C until blue staining became visible (approximately 20 h). The chlorophyll was removed during serial washing with ethanol. The tissues were cleared by submerging in Histoclear solution (Brunschwig, http://www.brunschwig-ch.com) for several hours. Plants were then photographed.

Expression profile maps

GGPPS transcript levels were determined from AtGenExpress microarray experiments encompassing various organs at different developmental stages of Arabidopsis (Schmid et al. 2005). For the specific root tissues expression profiles, the microarray dataset generated by Birnbaum et al. (2003) was used. For the seed expression profiles, the microarray data generated by Le et al. (2010) were analyzed. The corresponding microarray expression data were downloaded from the Bio-Array Resource website (BAR, http://bar.utoronto.ca; Toufighi et al. 2005).

For all the experimental conditions considered in this analysis, except of Figs. 8b, 9b and 10b, the expression values were retrieved in log2 scale. Genes with an expression value below a threshold of 2.5 were treated as absent (Schmid et al. 2005).

Fig. 2 GGPPS amino acid sequence similarities. **a** Pairwise percent similarity of annotated GGPPS amino acid sequences. **b** Alignment of the amino acid sequences of putative GGPPS. The two conserved aspartate-rich motifs DD(x_{2-4})D essential for GGPPS activity are highlighted in *black*. The second aspartate-rich motif of GGPPS12 is mutated. Additional amino acids important in prenyl-substrate binding (Kellogg and Poulter 1997) are marked with a *black frame*. Some of these amino acids are not conserved in the GGPPS12 sequence. The two CxxxC motifs participating in the physical interaction between the two subunits of the heteromeric GPP synthase (Wang and Dixon 2009) are *underlined in black* and present only in the GGPPS12 sequence. The subcellular targeting sequences highlighted in *grey* were predicted by TargetP. GGPPS1, 5: mitochondrial targeting, GGPPS3, 4: ER targeting, GGPPS 2, 6, 7, 9, 10, 11: plastid targeting, GGPPS8; dual mitochondrial (aa1-40)/ER (aa1-23) targeting. The amino acid sequences of the 12 GGPP synthases were retrieved from The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) and aligned using the ClustalW2 software (Larkin et al. 2007)

Results

Characterization of the GGPPS protein family based on their amino acid sequence alignment

Using homology searches Lange and Ghassemian (2003) predicted twelve GGPPS paralogs in the *A. thaliana* genome, with GGPPS12 having the weakest similarity to the other members of the protein family (between 31 and 40 %; Fig. 2a). With the discovery of a novel class of prenyl diphosphate synthase genes, the GPP synthase small subunit (SSU)-II subfamily, GGPPS12 was reclassified as a member of this protein class (Wang and Dixon 2009). SSUs are highly similar to GGPP synthases at the amino acid level but they lack two aspartate-rich motifs DD(x_{2-4})D (where "x" is any amino acid) that are important in prenyl-substrate binding, rendering them inactive. In addition, they have two CxxxC motifs (where "x" can be any hydrophobic amino acid) that are important in physical interactions between the two subunits (Wang and Dixon 2009).

We compared the amino acid sequences of functionally non-characterized GGPP synthases (GGPPS5, 8, 9, 10) with their active counterparts and with the GGPPS12 (Fig. 2b). All GGPP synthases except GGPPS12 have two DD(x_{2-4})D motifs and lack the second CxxxC motif, suggesting that they are functional proteins. In addition, other amino acids flanking the DD(x_{2-4})D motifs are important for prenyl-substrate binding (Kellogg and Poulter 1997). These amino acids are also conserved in all GGPP synthases except GGPPS12 (Fig. 2b). We used TargetP (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson et al. 2000) and PSORT (http://psort.hgc.jp/form.html) to identify putative subcellular localization domains in the GGPP synthases whose localization has not been reported. Based on the *in silico* predictions, GGPPS2, 6, 9 and 10 localize to plastids,

(a)

% amino acid similarity	GGPPS1	GGPPS2	GGPPS3	GGPPS4	GGPPS5	GGPPS6	GGPPS7	GGPPS8	GGPPS9	GGPPS10	GGPPS11	GGPPS12
GGPPS1		54	52	53	59	52	53	69	52	51	58	35
GGPPS2			59	58	68	61	59	56	60	59	70	39
GGPPS3				70	67	56	55	57	57	55	58	38
GGPPS4					66	57	57	59	55	56	58	34
GGPPS5						88	90	65	89	86	75	40
GGPPS6							94	57	86	87	63	38
GGPPS7								58	87	84	64	38
GGPPS8									57	56	63	31
GGPPS9										91	63	37
GGPPS10											63	38
GGPPS11												40
GGPPS12												

(b)

GGPPS8 either to mitochondria or to the ER, and GGPPS5 to mitochondria (Fig. 2b).

Inspection of the multiple sequence alignment revealed a shorter protein sequence for GGPPS5 compared to its closest paralogs GGPPS6 and GGPPS7 (Fig. 2b). Whereas GGPPS5 is 284 amino acids long, GGPPS6 and GGPPS7 are both 360 amino acids long and contain a plastid targeting sequence at the N terminus. The DNA sequences currently existing in the TAIR10 database for *GGPPS5*, *GGPPS6* and *GGPPS7* were examined in more detail (Fig. 3a). Compared to *GGPPS6* and *GGPPS7*, *GGPPS5* has an insertion of four nucleotides [GATC] that causes a frame shift if translation is initiated from the [ATG] homologous to the GGPPS6 and GGPPS7 translation initiation codons, resulting in a truncated protein sequence (Fig. 3b). Therefore another [ATG] codon was selected by TAIR as the translation initiation start. This [ATG] codon is present in all homologous sequences (e.g., *GGPPS6* and *GGPPS7*), indicating that it is not specific for *GGPPS5*. This [ATG] is also not preceded by typical translation initiation sequences such as the Kozak consensus sequence (gcc)gccRcc**AUG**G (R is a purine; Kozak 1997) or the *A. thaliana*-specific consensus sequence for AUG context aa(A/G)(A/C)a**AUG**Gcg (Rangan et al. 2008). Additionally, no cDNA is present in GenBank for *AT3G14510* (*GGPPS5*; http://gbrowse.arabidopsis.org/cgibin/gbrowse/arabidopsis/?name=AT3G14510) that would support the proposed gene model. We therefore consider *GGPPS5* a pseudogene and did not include it in our further analysis.

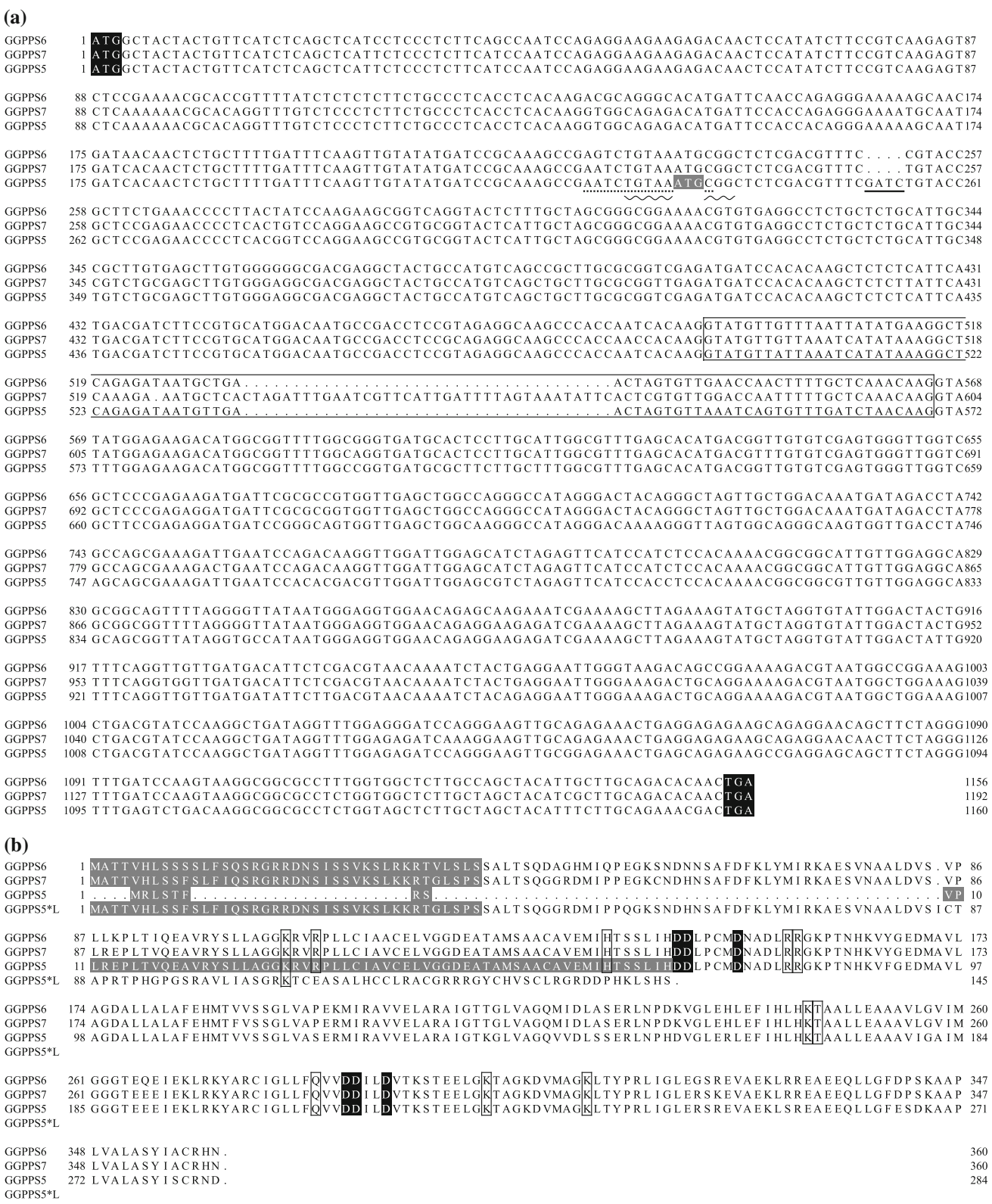
All predicted GGPP synthases encode functional enzymes

Based on DNA sequence analysis, we predict that the *A. thaliana* genome encodes ten functional GGPP synthases. Both in vitro enzymatic assays and genetic complementation of *E. coli* are methods that are widely used to characterize functionality of GGPPS. Crude extracts or purified proteins from *E. coli* cultures expressing heterologous GGPP synthases were used for enzymatic reactions with [^{14}C]IPP and an allylic substrate (DMAPP or FPP). Under these conditions GGPPS1 (Zhu et al. 1997b), GGPPS2 (Wang and Dixon 2009), GGPPS3 (Okada et al. 2000), GGPPS4 (Zhu et al. 1997a), GGPPS7 (Okada et al. 2000) and GGPPS11 (Okada et al. 2000; Wang and Dixon 2009) synthesized GGPP but GGPPS6 only a compound longer than C_{20} (Wang and Dixon 2009). GGPPS12 did not produce any prenyl diphosphate (Wang and Dixon 2009). The GGPPS activity of GGPPS1 and GGPPS4 was tested and confirmed by genetic complementation of *E. coli* (Zhu et al. 1997a, b).

Spectrum and length of prenyl-PP synthesized by prenyl transferases in vitro can be greatly affected by the type and concentration of the used substrate and by the ex vivo experimental conditions (Nishino and Rudney 1977; Ohnuma et al. 1993; Pan et al. 2002; Hsieh et al. 2011). We therefore used genetic complementation of *E. coli* to examine the activity of the uncharacterized *A. thaliana* GGPP synthases. All enzymes tested previously only in in vitro conditions (GGPPS1, 2, 3, 4, 6, 7, 11) were analyzed as well. The *crtE* gene encoding GGPPS from the *Erwinia uredovora* lycopene biosynthesis gene cluster present in the pACCRT-EBI vector (Misawa et al. 1990) was mutated to generate plasmid pACCRT-BI. After *E. coli* cells are transformed with pACCRT-BI, they only synthesize minor amounts of lycopene due to residual GGPP levels in the bacteria (Vallon et al. 2008). Lycopene production is increased after co-transformation with another plasmid harboring a protein with GGPPS activity (Fig. 4a).

All putative GGPP synthases except GGPPS1 and GGPPS4 were cloned into pGEX-4T-2 vectors after removal of the signal peptides from the N-termini to improve the solubility of the recombinant proteins (Fig. 4b). As expected, *E. coli* cells carrying the pACCRT-BI vector and the empty vector pGEX-4T-2 produced very little pigment (Fig. 4c). In contrast, *E. coli* co-transformed with the pACCRT-BI vector containing the genes for the individual *A. thaliana* putative GGPP synthases formed red-colored colonies except cells transformed with the pGEX-4T-2 vector carrying *GGPPS12*. Pigments were extracted from bacterial liquid cultures and quantified by measuring the absorbance of the extracts at 472 nm (Fig. 4c). Vectors

Fig. 3 Gene and protein models for GGPPS5. **a** Alignment of the nucleotide sequences of *GGPPS5*, *6* and *7*. The [ATG] start and [TGA] stop codons homologous to the three genes are highlighted in *black*. The intron position according to the TAIR gene models (TAIR10, www.arabidopsis.org) is shown in the *black frame*. The additional four nucleotides [GATC] that cause the translation frame shift of *GGPPS5* are *underlined in black*. The alternative [ATG] start codon, selected in the TAIR annotation and enabling in-frame translation of a functional protein is highlighted in *grey*. The Kozak consensus sequence (gcc)gccRcc**AUG**G (R is a purine; (Kozak 1997)) or *A. thaliana*–specific consensus sequence for the AUG context aa(A/G)(A/C)a**AUG**Gcg (Rangan et al. 2008) are *underlined* by a *dotted black line* or *waved black line*, respectively. **b** Alignment of the amino acid sequences of GGPPS5, 6 and 7. TAIR gene models for *GGPPS5, GGPPS6* and *GGPPS7* were retrieved from the database and after intron removal and *in silico* translation (http://expasy.org), the sequences were aligned using ClustalW2 (Larkin et al. 2007). The *GGPPS5* gene model labeled *GGPPS5**L, whose [ATG] start codon is homologous to that of *GGPPS6* and *GGPPS7*, was included in the multiple sequence alignment analysis. GGPPS5 is 284 amino acids long while GGPPS5*L is 145 amino acids long and has a premature stop codon. Both aspartate-rich motifs DD(x_{2-4})D that are essential for the substrate binding (*black boxes*) and most of the additional amino acid residues involved in substrate binding (*black frame*) were absent in GGPPS5*L. Protein targeting domains are highlighted in *grey*



containing *GGPPS2*, *GGPPS6*, *GGPPS7*, *GGPPS8*, *GGPPS10* and *GGPPS11* highly increased lycopene production in *E. coli* cells harboring pACCRT-BI, establishing that the genes encode active GGPP synthases. The absorbance of extracts from *E. coli* cells expressing *GGPPS3* and *GGPPS9* was not as strong but still significantly ($p < 0.05$) higher than the vector control,

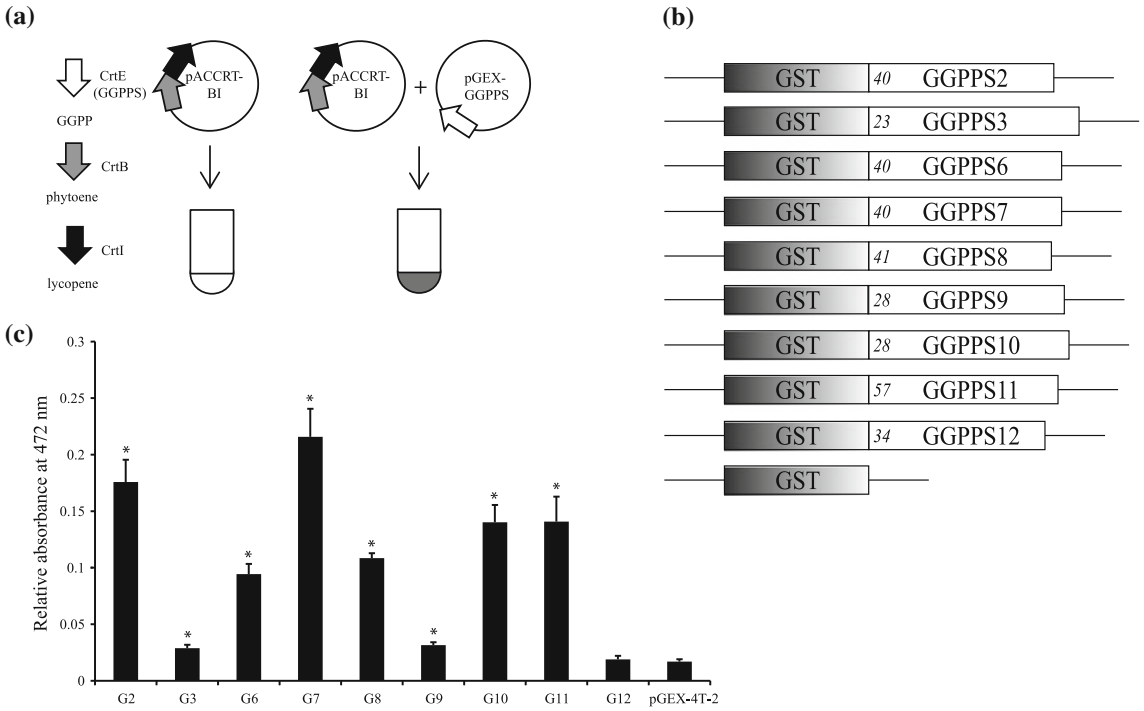


Fig. 4 Activity assay of GGPP synthases by genetic complementation of *E. coli*. The functionality of the GGPPS proteins was determined by genetic complementation of *E. coli* expressing the *crt* lycopene biosynthetic proteins lacking an active GGPP synthase. **a** Biosynthesis of lycopene by the *crt* gene cluster of *Erwinia uredovora*. The pACCRT-BI vector lacks a functional *crtE* gene (*GGPPS*). When expressed in *E. coli* in presence of a functional GGPPS, the construct is complemented and lycopene is produced. **b** Expression cassettes used to express *A. thaliana* GGPPS proteins in *E. coli*. The truncated proteins are fused to the C-terminus of the gluthatione-S-transferase (GST) and their expression is controlled by the *tac* promoter. The *numbers* indicate the position of the amino acid relative to the first methionine in the respective GGPPS protein. **c** Lycopene content of *E. coli* cells co-transformed with the pACCRT-BI and pGEX-GGPPSs or control pGEX-4T-2 vectors. Absorbance of the extracts of *E. coli* clones was measured at 472 nm. Values shown are the means ± SE of four to fifteen independent transformations. *Significantly different from the wild type ($p < 0.05$, one-tailed *t* test assuming equal variances)

suggesting that they can also produce GGPP. Only *GGPPS12* was found to lack GGPP synthase activity, confirming in vitro activity assays (Okada et al. 2000; Wang and Dixon 2009). Together, *A. thaliana* GGPPS1–4 and 6–11 are functional GGPP synthases and GGPPS12 most likely functions as a GPPS small subunit (Wang and Dixon 2009).

GGPP synthases localize to the cytosol, mitochondria and plastids

The subcellular localization of GGPPS1, GGPPS3, GGPPS4, GGPPS7 and GGPPS11 has already been determined using transit peptide-GFP fusion proteins in tobacco BY-2 cells (GGPPS1; Zhu et al. 1997b and *A. thaliana* (GGPPS1, 3, 4, 7, 11; Okada et al. 2000). GGPPS1 was shown to localize to the mitochondria, GGPPS3 and GGPPS4 to the ER, and GGPPS7 and GGPPS11 to plastids. The localization of GGPPS7 and GGPPS11 to plastids was also demonstrated by in vitro import into pea chloroplasts (Okada et al. 2000), and GGPPS11 was detected in the stroma in chloroplast proteomics studies (Joyard et al. 2009). All remaining functional GGPP synthases were predicted to have an N-terminal transit peptide for plastid localization, except of GGPPS8 that was predicted to localize to the ER or mitochondria (Fig. 2b). We fused GFP to the C-terminus of the full-length GGPPS proteins to determine the subcellular localization of the entire protein family. Figure 5a, b show that the GGPPS1-GFP signal exhibited a punctuate pattern of a size, shape and distribution that overlapped with the mCherry mitochondrial marker but not with chlorophyll fluorescence. GGPPS1 is therefore a mitochondrial isozyme, supporting earlier results obtained using a transit peptide-GFP fusion protein (Zhu et al. 1997b; Okada et al. 2000). Both GGPPS3-GFP and GGPPS4-GFP signals formed a filamentous structure that overlapped with the mCherry ER marker. GGPPS3 and GGPPS4 therefore localize to the ER, consistent with transit peptide-GFP fusions (Okada et al. 2000). The GFP signals of GGPPS2-, 6-, 7-, 8-, 9-, 10- and 11-GFP fusion proteins were localized to plastids based on both chlorophyll autofluorescence and overlap with the mCherry plastid marker, confirming plastidial localization of all remaining

proteins. These results are consistent with in vivo data (GGPPS7, 11; Okada et al. 2000) and *in silico* data except for GGPPS8, which was predicted by TargetP and PSORT to localize to either mitochondria or ER.

Expression of GGPP synthases in *A. thaliana* organs and seedlings monitored by RT-qPCR

To gain insight into the differential expression of the *GGPPS* gene family, we first used quantitative real time PCR (RT-qPCR), which allows the specific differential amplification of transcripts from highly similar genes. We analyzed the expression of individual paralogs in seedlings and plant organs such as roots, rosette and cauline leaves, stems, flowers and siliques. Expression levels lower or equal to a cycle threshold value (*Cq*) of 35 were regarded as significant (Karlen et al. 2007). The analysis revealed the distinct distribution and accumulation of the *GGPPS* transcripts in the different organs and in seedlings. *GGPPS11* had the highest expression level compared to all other paralogs in all organs and in seedlings (Fig. 6a), representing more than 90 % of the total *GGPPS* transcripts in all organs except siliques, flowers and roots (Supplemental Table S5). Only *GGPPS1*, encoding the mitochondrial GGPPS, and *GGPPS2*, encoding a plastid GGPPS, are also ubiquitously expressed in all organs and in seedlings, although at much lower levels (Fig. 6). Although *GGPPS2* is expressed in all organs, expression was significantly more pronounced in siliques and roots (Fig. 6b). Expression of *GGPPS3*, *4*, *6*, *7*, *8*, *9*, *10* is confined to specific organs and developmental stages. In seedlings, all *GGPPS* genes except *GGPPS3*, *4* and *9* were expressed. In plants, expression of the remaining *GGPPS* genes was confined mainly to roots (all except *GGPPS4*), siliques (all except *GGPPS10*) and flowers (*GGPPS4*, *6* and *7*). Together, the *GGPPS* paralogs in *A. thaliana* have significantly different quantitative and tissue-specific expression patterns.

Tissue-specific expression of *GGPPS* paralogs detected by *promoter-GUS* fusions

To obtain more detailed information on the tissue-specific expression of the *GGPPS* genes, we introduced *GGPPS promoter-GUS* fusion constructs into *A. thaliana* and analyzed the transformants using histochemical staining. Since *GGPPS1*, *GGPPS3*, *GGPPS4*, *GGPPS7* and *GGPPS11* were already characterized (Okada et al. 2000), we focused on the remaining *GGPPS* genes. We also included a *GGPPS11-GUS* promoter fusion because of the discrepancy between the reported GUS data that indicated little promoter activity in roots (Okada et al. 2000) and our gene expression data shown in Fig. 6b. To clone *GGPPS* promoters, intergenic regions of a maximum of 1.88 kb were amplified (Fig. 7a). The *GGPPS11* promoter contained also part of the upstream gene sequence and was 407 bp longer than the construct used by Okada et al. (2000). Plants expressing *GGPPS6*, *GGPPS8*, *GGPPS10*, and *GGPPS11 promoter-GUS* fusions showed discernable blue staining, while plants expressing *GGPPS2* and *GGPPS9 promoter-GUS* fusions did not give any signal even when stained for more than 48 h. The *GGPPS11* promoter expressed *GUS* ubiquitously in all seedling tissues and during all subsequent phases of plant development, including the roots (Fig. 7b). Thus, the distal 407 bp 5′ sequence of the promoter region present in our construct is required for expression of *GGPPS11* in the root. *GGPPS6*, *8* and *10* promoters directed *GUS* expression specifically in the roots of both seedlings and adult plants (Fig. 7b). The promoter region of *GGPPS10* restricted transcription of *GUS* in the root tip. The *GGPPS6* promoter was also active in the meristematic zone of the root tip, particularly in the columella and the lateral root cap. *GGPPS8* expression was detected specifically in the outer cell layers located above the mitotically active area of the root. Although expression of *GGPPS6* in flowers, siliques and seedlings, *GGPPS9* in roots and siliques and *GGPPS2* in all organs and seedlings was detected by RT-qPCR, albeit transcript levels were low (Fig. 6 and Supplemental Table S5), no GUS staining was visible in these tissues. A possible explanation for the discrepancy between the expression detected by RT-qPCR and *promoter-GUS* fusions can be either the lower sensitivity of GUS detection assay compared to RT-qPCR or the lack of regulatory sequences in the selected promoter regions.

Subfunctionalization of *GGPPS* paralogs revealed by microarray data

RT-qPCR analysis of gene expression showed that all or most of the *GGPPS* paralogs are expressed in roots and reproductive organs (siliques and flowers), respectively (Fig. 6b and Supplemental Table S5). Since the tissue-specific expression detected by RT-qPCR did not fully coincide with the *GGPPS promoter-GUS* fusion data, we used available microarray data sets for roots (Birnbaum et al. 2003) and seeds (Le et al. 2010), and extracted microarray data for flower tissues at developmental stages 9, 10/11, 12 and 15 (Smyth et al. 1990) from the Arabidopsis Development Baseline dataset (Schmid et al. 2005). Before the analysis we confirmed that the microarray data were similar to the RT-qPCR data using comparable organ samples and that probesets were specific to capture paralog-specific expression data. As shown on Supplemental Table S5, RT-qPCR and microarrays were comparable, although the sensitivity of RT-qPCR was higher as

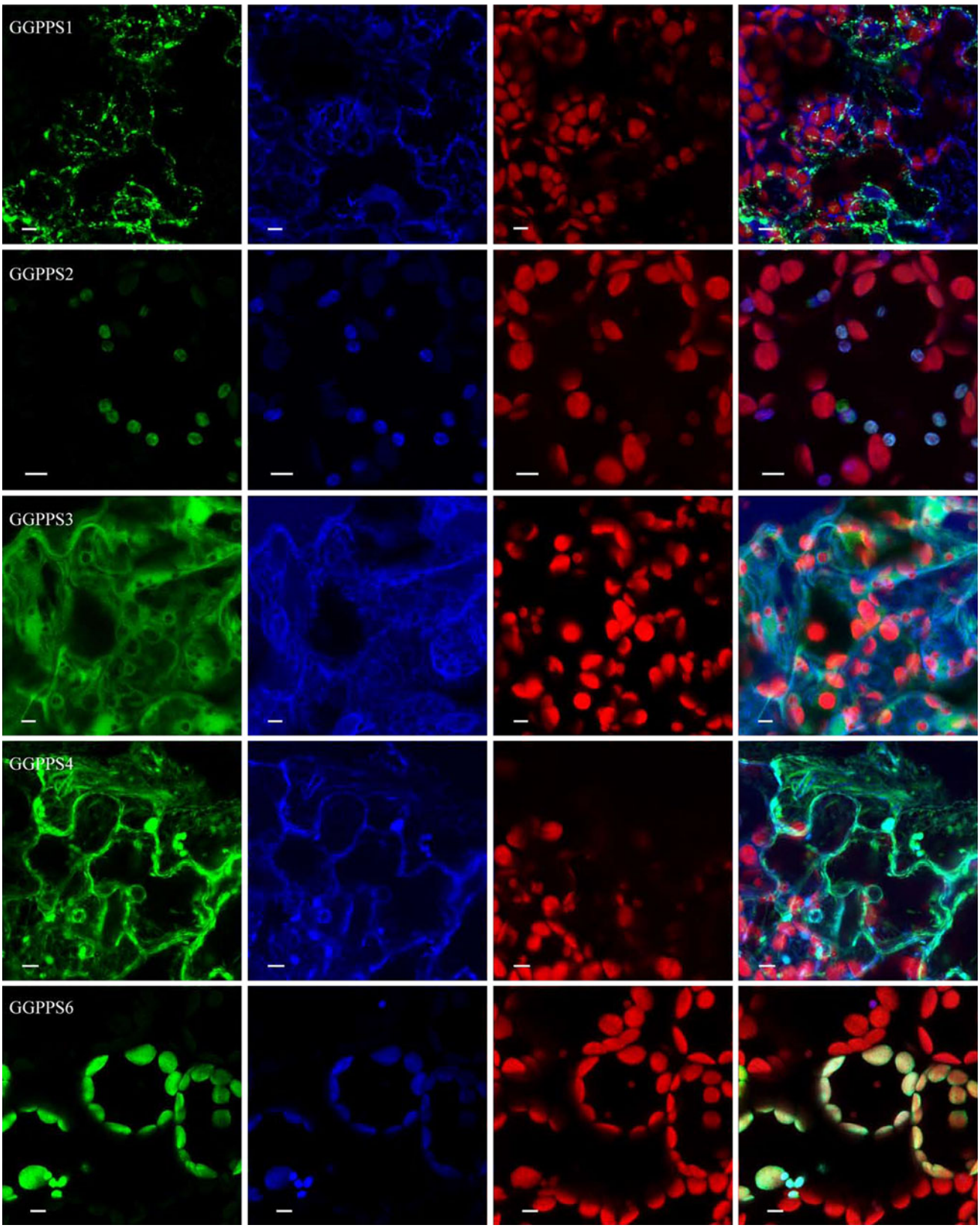


Fig. 5 Subcellular localization of the GGPPS-GFP fusion proteins. Confocal microscopy of *A. thaliana* leaves expressing GGPPS-eGFP together with the red fluorescent protein (RFP; mCherry) markers for mitochondria (GGPPS1), plastids (GGPPS2, 6–11) and ER (GGPPS3, 4). The first column shows GFP in *green*, the second shows mCherry (RFP) in *blue*, the third shows fluorescence of the chlorophyll in *red*, and the fourth shows the overlay of the three channels. *Bars* represent 5 μm

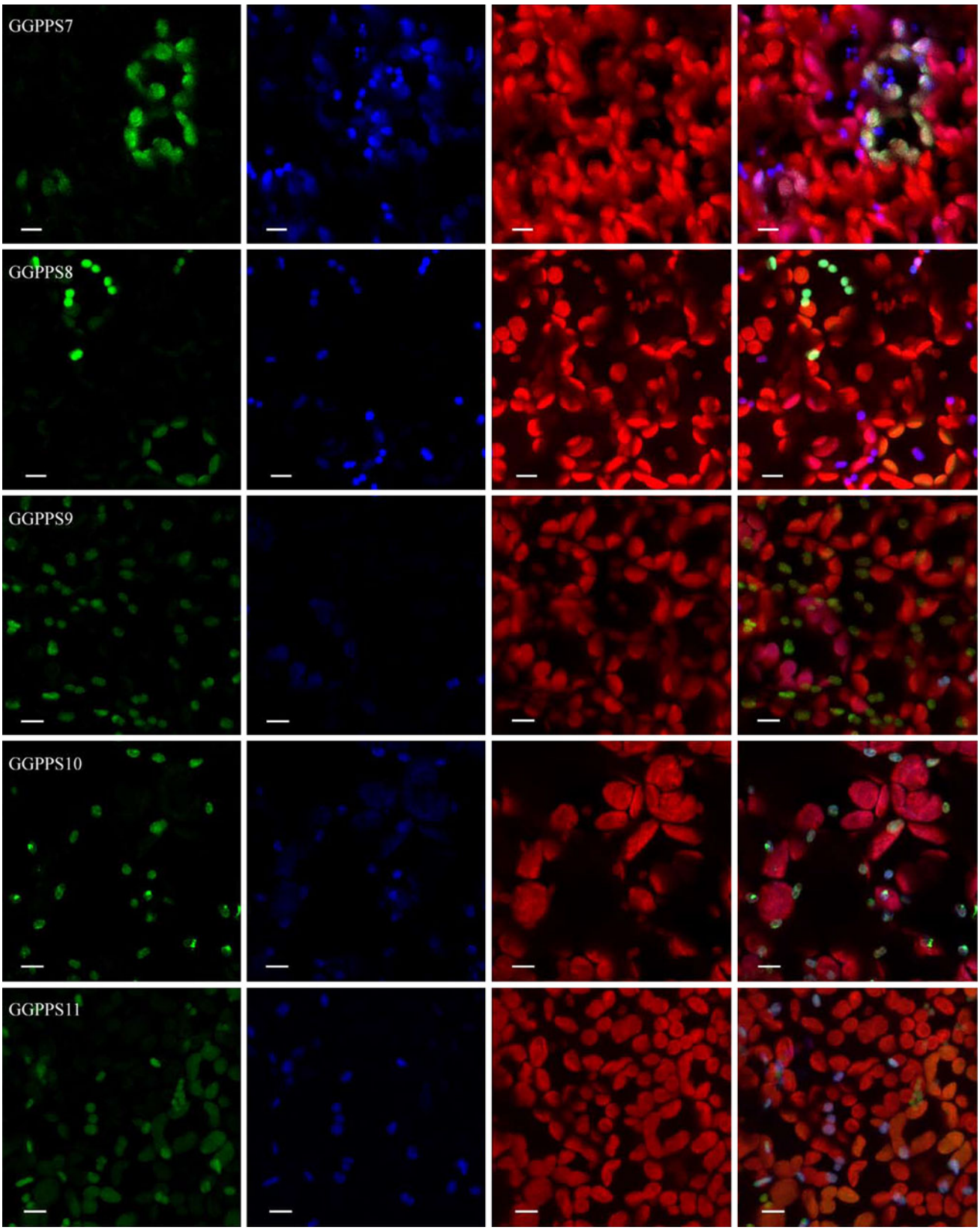


Fig. 5 continued

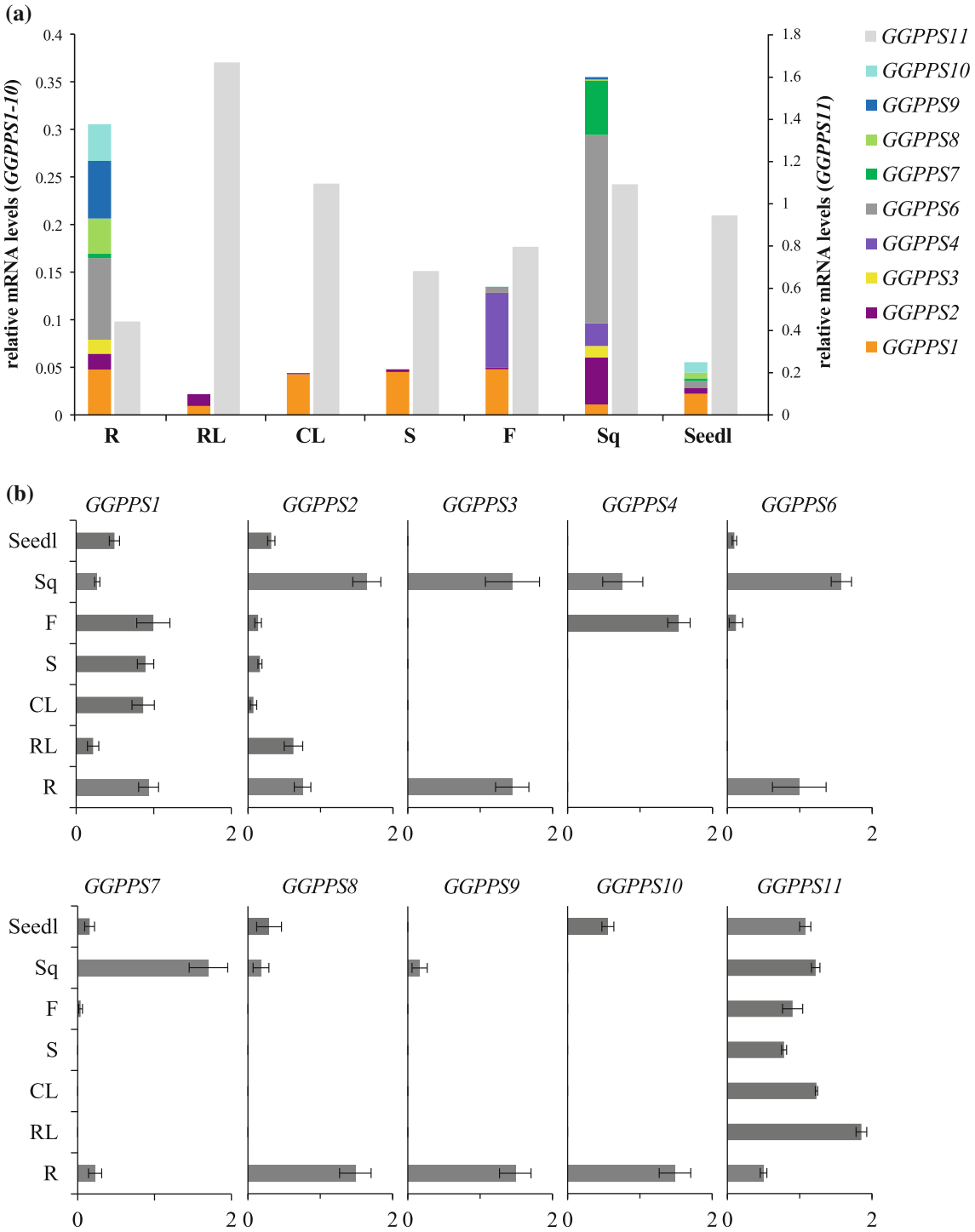


Fig. 6 Relative expression levels of the *GGPPS* genes. Seven plant organs from *A. thaliana* plants grown in standard long-day conditions were analyzed. The root (R) samples were collected from 18 day-old seedlings grown on standard MS medium. The rosette leaves (RL), cauline leaves (CL), stems (S), flowers (F) and siliques (Sq) were collected from 6 week-old plants grown in soil and the seedlings (Seedl) were collected from 14 day-old plants grown on standard MS medium. The expression levels of all *GGPPS* paralogs were determined by quantitative Real-Time PCR (RT-qPCR). **a** Transcript abundance of the *GGPPS* genes in the different plant organs. *GGPPS11* (*grey*) has the highest expression levels over all analyzed organs and is shown on the secondary y-axis (*right*). The scale units (*NQ* values, see section "Materials and methods") represent the fold changes in *GGPPS* mRNA levels relative to the reference genes. **b** Organ-specific expression of individual *GGPPS* genes relative to the organ with the maximum expression level. The scale units (*NRQ* values, see section "Materials and methods") are the arbitrary units and show abundance of a specific *GGPPS* transcript in each of the seven organs relative to the organ with the maximum expression level (**b**). For both (**a, b**), the average of three biological replicates is shown in log2 scale with the corresponding standard error (**b**)

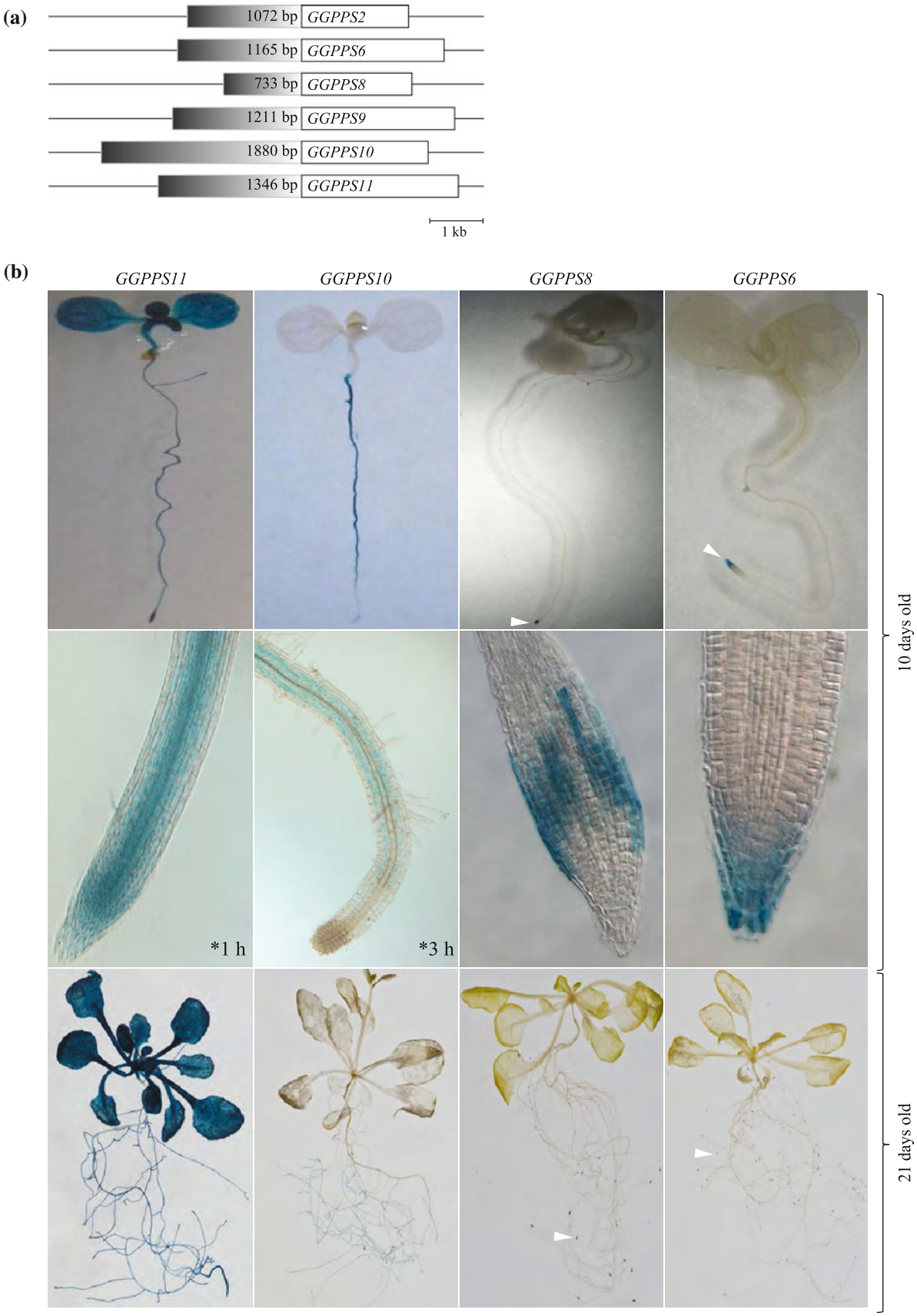
previously reported for other genes expressed at low levels (Czechowski et al. 2004). Additional signals observed in microarrays in some organs (Supplemental Table S5) can likely be attributed to the differences in sample material used by the two platforms (see section "Materials and methods" and Supplemental Table S5). Except for *GGPPS6* and *GGPPS7*, specific probesets for all paralogs are present on the ATH1 microarray and their specificity is supported by unique expression pattern for each of the paralogs (Supplemental Table S5). There is only one probeset that hybridizes to both *GGPPS6* and *GGPPS7*, and data obtained from microarrays for this probeset reflect expression of both paralogs (*GGPPS6/7*).

GGPPS1 and *GGPPS11* are expressed in all flower tissues, (Fig. 8a and Supplemental Table S6), further substantiating their ubiquitous expression pattern (Fig. 6b). Expression of *GGPPS2*, another ubiquitously expressed paralog, was not detected on the microarrays in any of the flower tissues (Fig. 8a and Supplemental Table S6), although its expression was detected in flowers by RT-qPCR (Fig. 6b). The reason for this discrepancy might be the lower sensitivity of microarrays to detect gene expression (Supplemental Table S5). Two additional paralogs, *GGPPS4* and *GGPPS6/7*, are expressed in flowers (Fig. 8a and Supplemental Table S6). *GGPPS4* is expressed at all flower developmental stages, mainly in stamens, whereas *GGPPS6/7* are expressed specifically in carpels at the later stages of flower development. The *GGPPS11* paralog is the most abundantly expressed *GGPPS* in all flower tissues, except for *GGPPS4* that is the most highly expressed paralog in stamens at flower developmental stage 12. The expression of *GGPPS11* in flowers and *GGPPS4* in anthers was also observed using *promoter-GUS* fusions of the respective genes (Okada et al. 2000).

While all paralogs are expressed in developing seeds, *GGPPS11* is expressed most highly at all developmental stages and in all tissues (Fig. 9a and Supplemental Table S6). *GGPPS11* is expressed at the highest level in the chalazal endosperm at pre-globular, globular and heart stages (Fig. 9b and Supplemental Table S6). All remaining paralogs are expressed in several tissues but at significantly lower levels and expression of each paralog peaks at different developmental stages. *GGPPS1* is expressed most highly in the suspensor at the globular stage, *GGPPS2* in the seed coat at the linear cotyledon stage, *GGPPS3* in the peripheral endosperm at mature green stage, *GGPPS4* in the seed coat at mature green stage, *GGPPS6/7* in the embryo proper at the heart stage, *GGPPS8* in the chalazal endosperm at pre-globular stage, *GGPPS9* in the chalazal endosperm at the heart stage and *GGPPS10* in the chalazal endosperm at the pre-globular stage (Fig. 9 and Supplemental Table S6).

All *GGPPS* paralogs are also expressed in roots (Fig. 10 and Supplemental Table S6). While *GGPPS11* is expressed ubiquitously at all developmental stages and in all root tissues except the procambium, its peak expression level is similar to that of *GGPPS1–4*, *8–10* (Fig. 10a and Supplemental Table S6). The other paralogs have diverse developmental and/or tissue-specific expression patterns. Similarly to *GGPPS11*, the gene for the mitochondrial *GGPPS1* is expressed ubiquitously in all tissues but its expression is the strongest in the elongation zone (Fig. 10b). *GGPPS2* is expressed predominantly in the endodermis, in the cortex and, to a lesser extent, in vascular tissues. *GGPPS3* and *GGPPS4*, which encode ER-localized GGPP synthases, are expressed in the procambium and in the epidermis in root hair cells. Although *GGPPS3* and *GGPPS4* are expressed in the same tissues, *GGPPS3* is confined to the elongation zone, while *GGPPS4* is expressed in the meristematic and maturation zones. *GGPPS3* expression in vascular tissues is also consistent with the *promoter-GUS* expression data, while expression of *GGPPS4* was observed in the root tip using a *promoter-GUS* fusion construct (Okada et al. 2000). *GGPPS6/7* is mainly expressed in columella, lateral root cap, cortex, epidermis and vascular tissues, but not the procambium. *GGPPS6/7* is the paralog that shows the strongest expression in root tissues. Expression in the columella and lateral root cap is likely resulting from *GGPPS6*, while *GGPPS7* is expressed in the vasculature as suggested by *GGPPS6* and *GGPPS7 promoter-GUS* expression patterns (Okada et al. 2000; Fig. 7b). *GGPPS8* is mainly expressed in the epidermis, especially in the elongation zone and in the lateral root cap, consistent with the *GGPPS8 promoter-GUS* expression pattern (Fig. 7b). *GGPPS9* is weakly expressed, mainly in the epidermis and phloem while *GGPPS10* is expressed mainly in the procambium (Fig. 10), although its expression was observed in the ground tissue using *GGPPS10 promoter-GUS* fusion (Fig. 7b). Together, the expression of most of the *GGPPS* paralogs in *A. thaliana* has become significantly constrained by tissue- and cell-specific transcriptional regulation.

Fig. 7 Histochemical localization of GUS activity in transgenic Arabidopsis plants expressing *GGPPS6, 8, 10* and *11 promoter-GUS* fusion constructs. **a** *GGPPS promoter-GUS* transcriptional fusion constructs were generated for five uncharacterized paralogs (*GGPPS2, 6, 8, 9* and *10)*. In the case of *GGPPS11,* 119 bp of the upstream gene sequence (*At4g36800*) was cloned along with the intergenic sequence. **b** GUS activity patterns resulting from expression of *GGPPS6, 8, 10* and *11 promoter-GUS* fusion constructs in 10 and 21 day-old transformed T_2 Arabidopsis plants. No staining was observed in plants transformed with *GGPPS2* and *GGPPS9 promoter-GUS* fusion constructs. The magnification of stained areas in roots of 10 day-old plants is shown. All staining for GUS expression assays was for 24 h (see section "Materials and methods") except for the 10 day-old roots expressing the *GGPPS10* and *GGPPS11 promoter-GUS* fusion constructs, which were stained for 3 h and 1 h respectively. White arrowheads indicate highly localized GUS activity


(a)
1072 bp GGPPS2
1165 bp GGPPS6
733 bp GGPPS8
1211 bp GGPPS9
1880 bp GGPPS10
1346 bp GGPPS11
1 kb
(b)
GGPPS11
GGPPS10
GGPPS8
GGPPS6
*1 h
*3 h
10 days old
21 days old

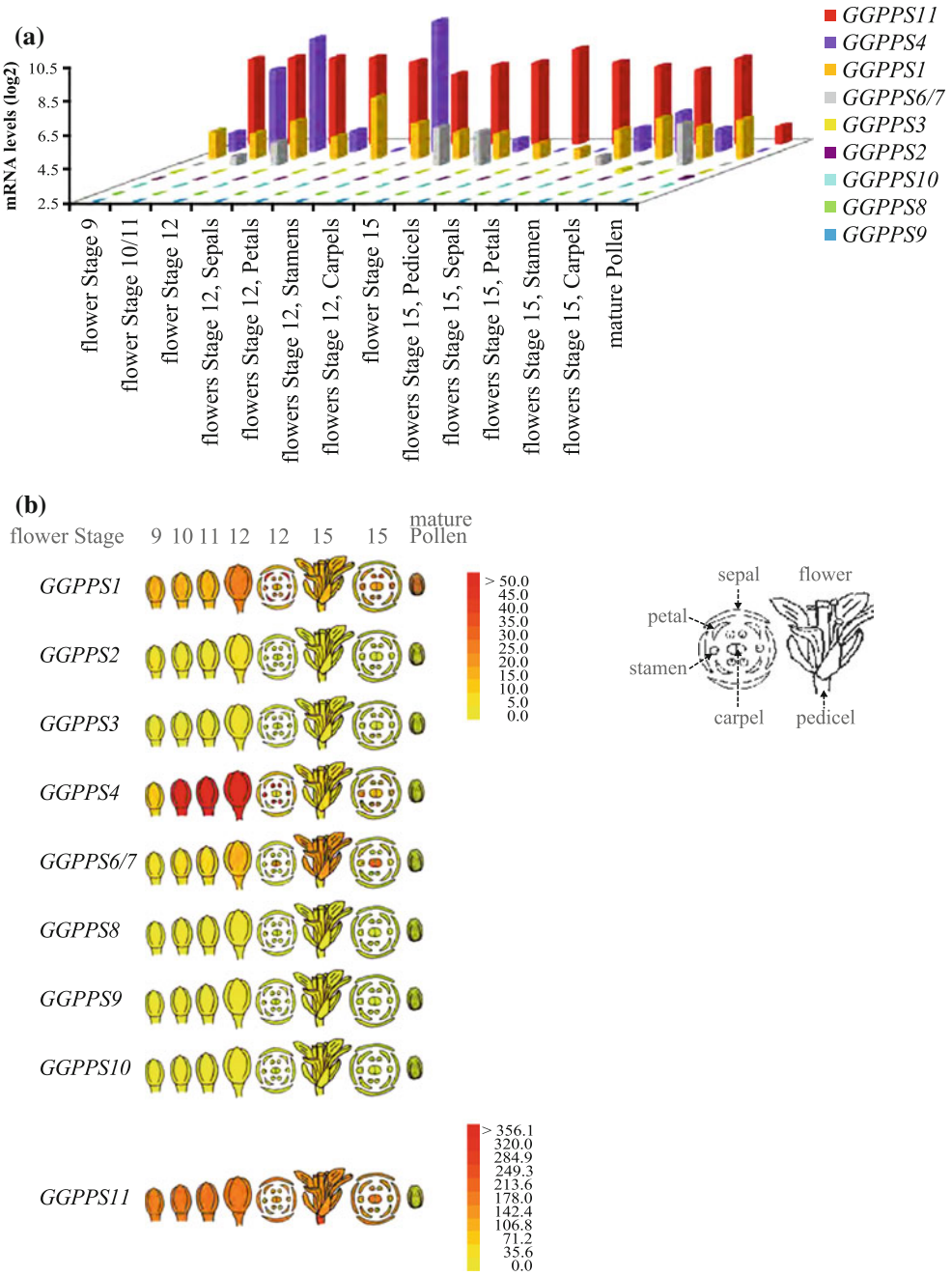


Fig. 8 Tissue-specific expression of the *GGPPS* genes during flower development. For all *GGPPS* genes except *GGPPS6* and *GGPPS7* (referred to here as *GGPPS6/7*) specific probe sets are present on ATH1 microarrays. Data were extracted from the AtGenExpress developmental series (Schmid et al. 2005), specifically selecting the flower developmental stages (Supplemental Table S6). The data were retrieved from the Bio-Array Resource website (BAR, http://bar.utoronto.ca; Toufighi et al. 2005). **a** Transcript intensity (log2 scale) and distribution of the *GGPPS* expression during flower development. The different flower organs and stages of development are shown. Signal intensities below 2.5 in log2 scale were regarded as not detectable (see section "Materials and methods"). **b** Expression of *GGPPS* genes in various flower organs during five stages of flower development and in pollen grains. Signal intensities in linear scale are shown as a heatmap as follows: undetected transcripts in *yellow*, low and medium transcript levels in *orange* and high transcript levels in *red*. *GGPPS11* has generally much higher transcript levels than other *GGPPS* genes and is therefore shown separately with the corresponding signal intensity scale

Discussion

The production of GGPP is the key branch point in the biosynthesis of major isoprenoid compounds (Fig. 1). Because of the substantial branching of the pathway at the point of GGPP synthesis and of the diverse roles of the different end products, it was proposed that the subcellular and tissue-specific expression of GGPPS isoforms allocates GGPP precursors and controls metabolic flux to distinct isoprenoid biosynthetic pathways (Tholl and Lee 2011). To date, only a partial molecular characterization of the *GGPPS* gene family in *A. thaliana* has been reported (Zhu et al. 1997a, b; Okada et al. 2000; Wang and Dixon 2009). To fully understand the role of each individual GGPPS isozyme in *A. thaliana*, it was therefore essential to establish a complete characterization of this enzyme class.

Here we have characterized the activity, subcellular localization and tissue-specific expression of the entire protein family. In *A. thaliana*, twelve GGPP synthase isozymes were predicted *in silico* to be responsible for the synthesis of GGPP (Lange and Ghassemian 2003). GGPPS12, however, has no GGPPS activity as was shown in vitro (Okada et al. 2000; Wang and Dixon 2009) and in vivo in *E. coli* (Fig. 4c). *GGPPS5* has a frameshift

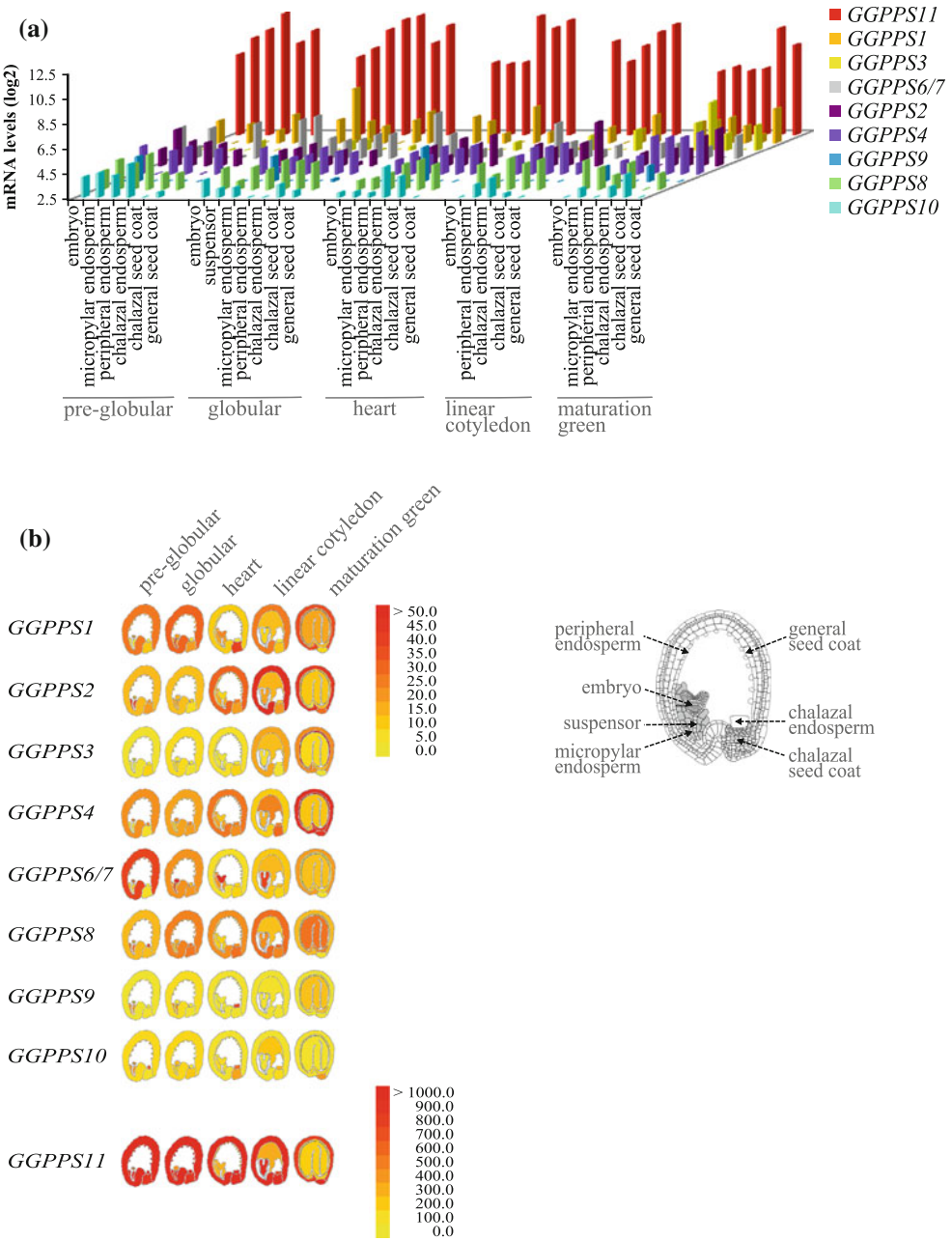


Fig. 9 Tissue-specific expression of the *GGPPS* genes during seed development. For all *GGPPS* genes except *GGPPS6* and *GGPPS7* (referred to here as *GGPPS6/7*) specific probesets are present on ATH1 microarrays. The *GGPPS* data (Supplemental Table S6) were extracted from microarray experiments reported by Le et al. (2010) and retrieved from the Bio-Array Resource website (BAR, http://bar.utoronto.ca; Toufighi et al. 2005). **a** Transcript intensity (log2 scale) and expression of the *GGPPS* genes during seed development. The different seed tissues and stages of development are shown. Signal intensities below 2.5 in log2 scale were regarded as not detectable (see section "Materials and methods"). **b** *GGPPS* expression in various seed tissues during five stages of seed development (pre-globular, globular, heart, linear cotyledon and maturation green). Signal intensities in linear scale are shown as a heatmap as follows: undetected transcripts in *yellow*, *low* and medium transcript levels in *orange* and high transcript levels in red. *GGPPS11* has much higher transcript levels than other *GGPPSs* and is therefore shown separately with the corresponding signal intensity scale. The schematic seed image was adapted based on the drawing of Meryl Hashimoto retrieved from http://www.seedgenenetwork.net/arabidopsis

mutation that would result in the synthesis of a truncated protein when translated from the ATG start codon in the appropriate context (Fig. 3) and therefore is likely a pseudogene. This reduces the number of candidates for functional GGPPS isozymes in *A. thaliana* to ten genes. We have established the functionality of all of them in vivo except GGPPS1 and GGPPS4 in *E. coli* strains engineered to synthesize lycopene but lacking GGPPS activity (Fig. 4). The activity of GGPPS1 and GGPPS4 had been shown previously in a similar heterologous in vivo system (Zhu et al. 1997a, b). Data from in vitro enzymatic activity assays (Zhu et al. 1997a, b; Okada et al. 2000; Wang and Dixon 2009) are also consistent with the predicted functionality based on the amino acid sequence alignment (Fig. 2b). Only for GGPPS6 we have obtained contrasting data. To determine GGPPS6 functionality, Wang and Dixon (2009) used DMAPP and [^{14}C]IPP as substrates in an in vitro enzymatic assay and reported a polyprenyl diphosphate product with a chain length of more than 20 carbons. At the amino acid level GGPPS6 is highly similar to GGPPS7, which can synthesize GGPP both in vitro and in vivo (Okada et al. 2000 and this study). Moreover, close inspection of the GGPPS6 amino acid sequence showed that conserved residues around the elongation cavity that

are present in *A. thaliana* polyprenyl diphosphate synthases (Hsieh et al. 2011) are not present in GGPPS6 (data not shown). Therefore we suggest that GGPPS6 is also a functional GGPPS. Nevertheless, we can not exclude at this point that GGPPS6 synthesizes multiple end products, among which both GGPP and polyprenyl diphosphate are present.

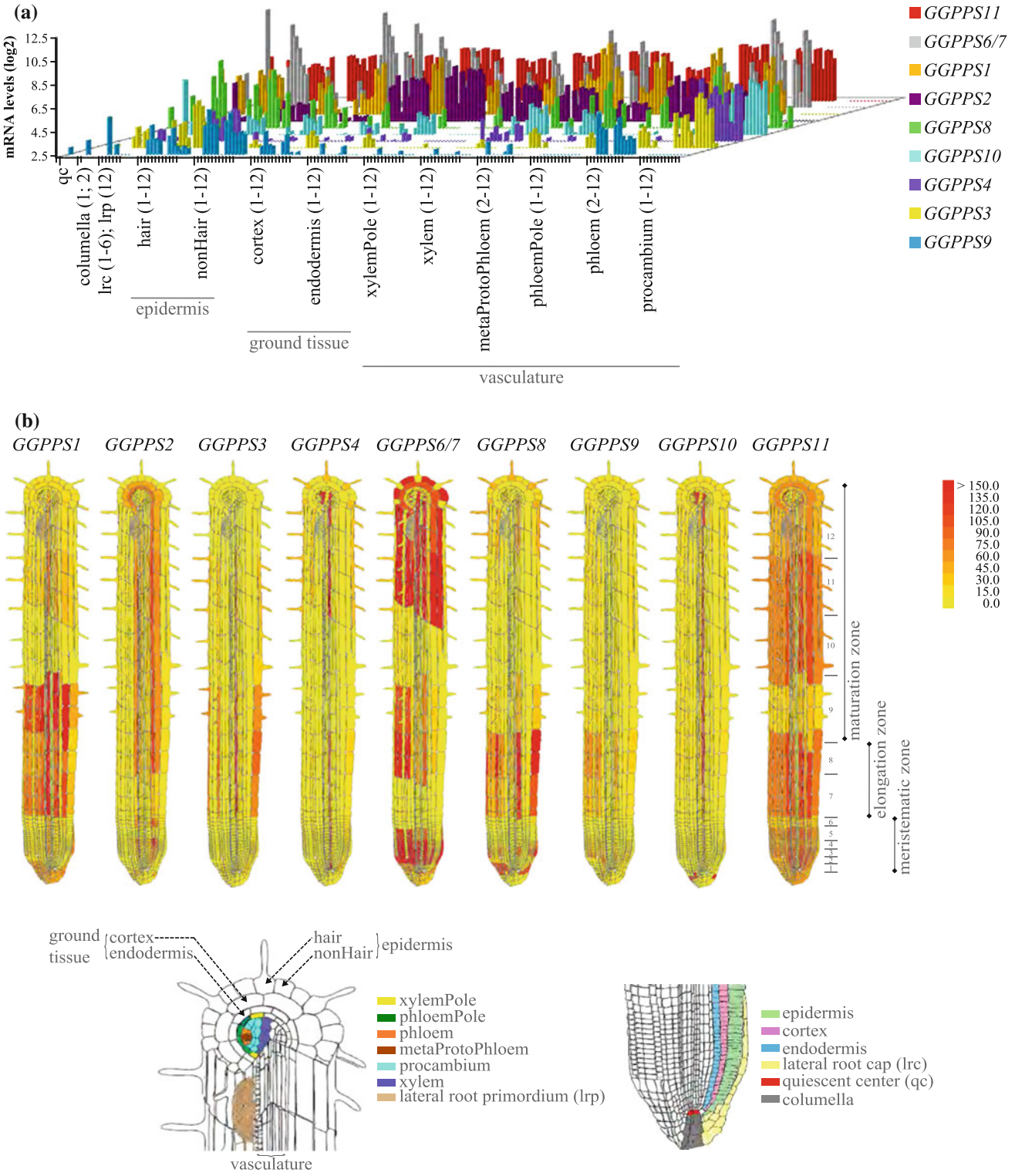
The presence of the high number of gene paralogs in the *GGPPS* family in the *A. thaliana* genome is interesting because gene paralogs can be stably maintained when they differ in their functions such as differential spatial and temporal gene expression, better performance in certain conditions, gene dosage function, or association with distinct metabolic fluxes. In addition, the role in compensating knockout mutations—often referred to as genetic network robustness—is attributed to many duplicated genes (Zhang 2003; Kuepfer et al. 2005). At least one GGPPS is associated with the cytosol, mitochondria and plastids (Fig. 5a, b), which may reflect the need for autonomous synthesis of the GGPP precursor in the subcellular compartments. Prenyl diphosphates with carbon chain length C_5–C_{15} can translocate through the plastid membrane (Bick and Lange 2003; Flügge and Gao 2005) but higher carbon chain length prenyl diphosphates, such as GGPP (C_{20}) are not translocated with appreciable efficiency (Bick and Lange 2003) and no GGPP transporter has been identified. Thus, while IPP, DMAPP, GPP or FPP can be translocated between compartments, the synthesis of GGPP is likely an organelle-autonomous process.

To establish organ and tissue-specific expression patterns of individual paralogs we used three different approaches: RT-qPCR, *GGPPS promoter-GUS* constructs and available microarray data (Figs. 6, 7, 8, 9, 10 and Supplemental Tables S5 and S6). The combined results together with the data reported by Okada et al. (2000) are summarized in Fig. 11. Each approach has its constraints that can influence the results of gene expression. For example, RT-qPCR and microarrays have different levels of sensitivity, results from analysis of *promoter-GUS* constructs can be influenced by the selection of regulatory elements, and experiments in different laboratories can be affected by individual growth and sampling conditions (Massonnet et al. 2010). Combining results from different approaches will therefore increase confidence in the expression data. As shown in Figs. 6, 7, 8, 9 and 10 and Supplemental Tables S5 and S6, and summarized in Fig. 11, the plastid paralog *GGPPS11* and the mitochondrial paralog *GGPPS1* are ubiquitously expressed in all organs and most of the tissues but root procambium (Fig. 10 and Supplemental Table S6). Surprisingly, none of the genes for cytosolic/ER GGPP synthases (*GGPPS3*, *4*) is ubiquitously expressed (Fig. 11). This suggests that the synthesis of plastid and mitochondrial GGPP is essential for isoprenoid biosynthesis in most of the plant tissues, in contrast to cytosolic GGPP. The expression of genes encoding the two cytosolic isozymes (*GGPPS3*, *4*) and the six remaining plastid isozymes (*GGPPS2*, *6*, *7*, *8*, *9*, *10*) is developmentally regulated and confined to specific tissues (Figs. 6, 7, 8, 9, 10).

Fig. 10 Tissue-specific expression of *GGPPS* genes in root tissues. ▶ For all *GGPPS* genes except *GGPPS6* and *GGPPS7* (referred to here as *GGPPS6/7*) specific probesets are present on ATH1 microarrays. The *GGPPS* data (Supplemental Table S6) were extracted from microarray experiments reported by Birnbaum et al. (2003) and retrieved from the Bio-Array Resource website (BAR, http://bar.utoronto.ca; Toufighi et al. 2005). **a** Transcript intensity (log2 scale) and expression of the *GGPPS* genes in root tissues. The root tissues are further classified in quiescent center-qc, columella, lateral root cap-lrc, lateral root primordium-lrp, epidermis (hair and nonHair), ground tissues (cortex and endodermis) and vasculature (xylemPole, xylem, metaProtoPhloem, phloemPole, phloem and procambium). The signal values of *GGPPS6/7* are much higher than those of other *GGPPSs* in several tissues (e.g., signal = 10.7 in columella, Supplemental Table S6). Signal intensities below 2.5 in log2 scale were regarded as not detectable (see section "Materials and methods"). **b** *GGPPS* expression pattern in root tissues of young Arabidopsis plants. Signal intensities in linear scale are shown as a heatmap as follows: undetected transcripts in *yellow*, low and medium transcript levels in *orange* and high transcript levels in *red*

GGPPS11 is ubiquitously expressed and produces higher mRNA levels than the other paralogs in all organs and most of the tissues except in stamens at the flower developmental stage 12 (Fig. 8a and Supplemental Table S6) and in several root tissues (Fig. 10 and Supplemental Table S6). The expression of *GGPPS11* is also reflected at the protein level because it was as the only GGPPS isozyme identified in genome-wide mass spectrometry studies with the highest number of spectral counts (http://suba.plantenergy.uwa.edu.au; http://www.grenoble.prabi.fr/at_chloro/). GGPPS11 is a plastid protein localized to the stroma (Joyard et al. 2009), and it is the only one of the seven plastid GGPPS isozymes that is highly expressed in photosynthetic tissues (Fig. 6a). Therefore we suggest that GGPPS11 has an essential function in the synthesis of photosynthesis-related isoprenoid compounds such as chlorophylls, carotenoids, plastoquinones, phylloquinones and tocopherols. In addition, the strong expression of *GGPPS11* in non-photosynthetic tissues (Figs. 6, 7, 8, 9, 10) implies that GGPPS11 might also be involved in the synthesis of other plastid isoprenoid compounds such as diterpenoids and hormones (Fig. 1).

GGPPS1 is also ubiquitously expressed, although at lower levels than *GGPPS11* (Figs. 6, 7, 8, 9, 10). However, expression of the *GGPPS1 promoter-GUS* fusion did not result in any visible GUS staining (Okada et al. 2000) and therefore the tissues in which *GGPPS1* is expressed are not known. Based on the rather ubiquitous expression profile of *GGPPS1* in all plant organs and in flower, seed and root tissues



(Figs. 6, 7, 8, 9, 10) it is likely, however, that GGPPS1 supplies GGPP for the synthesis of the essential isoprenoid end products. One of them can be ubiquinone-9. The polyprenyl moiety of the ubiquinone-9 is synthesized via the isoprenoid pathway and recently, the trans-prenyl diphosphate synthase that synthesizes solanesyl diphosphate in mitochondria and can use GGPP as a substrate has been identified (Ducluzeau et al. 2011; Hsieh et al. 2011). Another possible candidate for a

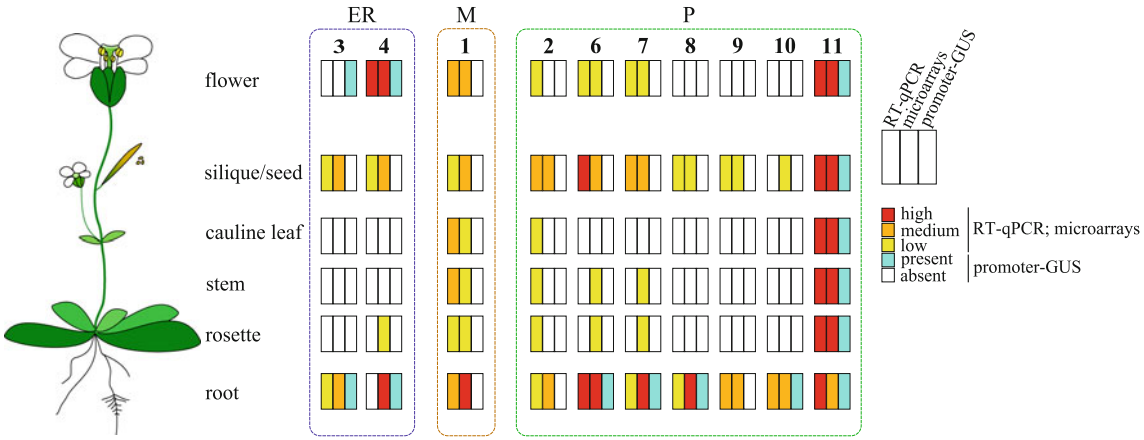


Fig. 11 Summary of Arabidopsis *GGPPS* developmental and subcellular expression patterns. Summary of expression analyses using three independent methods (RT-qPCR, *promoter-GUS* fusion constructs and available microarray data). *GGPPS* genes are represented by numbers (e.g., GGPPS1 = 1). For the RT-qPCR analysis, the *NQ* values (see section "Materials and methods") for each *GGPPS* gene were used to define three equally sized intensity classes. For the microarray data the maximum expression (log2 scale) value from different samples representing the same organ was used to define three equally sized intensity classes. Intensity classes are shown as a heatmap as follows: high expression in *red* (RT-qPCR $NQ = 0.073$-0.1 and microarray signal >7.22), medium expression in *orange* (RT-qPCR $NQ = 0.037$–0.073 and microarray signal = 5.64–7.22), low expression in *yellow* (RT-qPCR $NQ = 0.000712$–0.037 and microarray signal = 2.58–5.64) and undetected expression in *white* (see section "Materials and methods"). For the expression of *promoter-GUS* fusion constructs, data obtained in this study were combined with data reported by Okada et al. (2000). Discernable blue staining is shown in *blue color* and absence of staining in *white*. The subcellular localization is based on the experimental data obtained in this study. ER: endoplasmic reticulum, M: mitochondrium, P: plastid

mitochondrial isoprenoid that requires GGPP for its synthesis is a polyisoprenoid attached to the heme moiety, which is part of the cytochrome C complex required for respiration (Caughey et al. 1975; Cunillera et al. 1996).

The third ubiquitously expressed *GGPPS* paralog based on our RT-qPCR analysis encodes the plastid-localized GGPPS2 (Fig. 11). Expression of this isozyme, similarly to *GGPPS1*, could not be detected using a *GGPPS2 promoter-GUS* fusion construct. Based on microarray data, however, *GGPPS2* is specifically expressed in the endodermis, the cortex and to the lesser extent in the vasculature. It is therefore likely that expression of *GGPPS2*, although ubiquitous at the organ level based on RT-qPCR, is highly specific and constrained to certain tissues. Most of the genes implicated in regulating GA levels in roots are expressed more highly in the endodermis than in the surrounding tissues (Dugardeyn et al. 2008) and mutants impaired in GA biosynthesis or GA perception in endodermis have impaired root growth (Ubeda-Tomas et al. 2008, 2009). GGPPS2 can thus be a specific isozyme contributing GGPP for GA biosynthesis in the endodermis. Gene paralogs encoding plastid isozymes GGPPS6, 7, 8, 9 and 10 are expressed predominantly in specific root tissues (Figs. 7, 10), in developing seeds (Figs. 6, 9), and *GGPPS6/7* also in flowers and specifically in carpels (Fig. 8). *GGPPS6/7* might also be expressed in stem and leaf tissues, because in microarray experiments one or both of these genes were significantly expressed in both organs (Fig. 11 and Suppelmental Table S5).

In general, the temporal and tissue-specific expression of plastid GGPP synthases is correlated with the developmentally regulated synthesis of hormones such as ABA, GA or strigolactones (Nambara and Marion-Poll 2005; Bennett et al. 2006; Yamaguchi 2008; Ruyter-Spira et al. 2011). Additionally, plastid GGPP synthases likely provide GGPP for the synthesis of apocarotenoids or diterpenoids involved in flower scent, fruit flavour, plant-plant and plant-pathogen interactions (Floss and Walter 2009). Apocarotenoid biosynthesis is not well understood, but terpene synthases and enzymes regulating GA and ABA homeostasis are encoded by gene families expressed in different plant organs and tissues (Tan et al. 2003; Lefebvre et al. 2006; Mitchum et al. 2006; Dugardeyn et al. 2008; Hu et al. 2008). Plastid GGPP synthases can thus be part of regulons synthesizing isoprenoid hormones, diterpenoids or apocarotenoids required at different developmental stages or in response to certain environmental cues.

GGPPS genes encoding the ER-localized isozymes *GGPPS3* and *GGPPS4* are expressed in different tissues of developing seeds, mainly in the endosperm and the seed coat of mature green seeds (Fig. 9), in root procambium and root hairs (Okada et al. 2000; Fig. 10) and *GGPPS4* is also strongly expressed in anthers (Okada et al. 2000; Fig. 8). Expression of *GGPPS3* was also detected by Okada et al. (2000) in flowers using a *GGPPS3 promoter-GUS* construct, but this expression could not be confirmed by microarray and RT-qPCR data (Figs. 6, 8). Cytosolic

GGPP would be required as substrate in the synthesis of cytosolic oligoprenols (C_{25}–C_{45}) such as the ubiquinone side chain, dolichols or diterpenoids, in the synthesis of polyisoprenoids (C_{50}–C_n), and in protein prenylation (Vranová et al. 2011; Fig. 1). However, the developmental roles of GGPPS3 and GGPPS4 in flux distribution and the origin of cytosolic GGPP for the synthesis of cytosolic isoprenoid end products in organs and tissues in which expression of the cytosolic isozymes could not be detected remain to be clarified.

Final considerations

Our work represents a comprehensive characterization of the GGPPS family in *A. thaliana*. Our results show that ten of the twelve gene family members predicted by sequence similarity (Lange and Ghassemian 2003) encode functional enzymes that can synthesize GGPP in *E. coli* (Fig. 4c). *GGPPS5* (*At3g14510*) is likely a pseudogene (Fig. 3) and GGPPS12 (At4g38460) is not a functional GGPPS (Fig. 4c). The functional isozymes are targeted to different subcellular compartments, as shown in Fig. 5a, b. GGPPS1 is mitochondrial, GGPPS3 and GGPPS4 are in the ER and GGPPS2, 6, 7, 8, 9, 10 and 11 are localized in the plastids. The patterns of expression of the GGPPS paralogs differ substantially. Only two of them (*GGPPS1*, *GGPPS11*) are expressed constitutively in all organs and almost all tissues (Figs. 6, 7, 8, 9, 10). Based on the abundance and spatio-temporal expression pattern of *GGPPS11* transcripts, GGPPS11 has likely a house-keeping function (Fig. 6a). Expression of the remaining isozymes is restricted to specific tissues and developmental stages mainly in flowers (Figs. 6, 8), seeds (Figs. 6, 9) and in roots (Figs. 6, 7, 10). These results are summarized in Fig. 11.

The contribution of each isozyme to isoprenoid synthesis is still unclear, although their differential expression and localization suggest the enzymes are associated with specific developmentally-regulated isoprenoid biosynthesis pathways that use GGPP as substrate (see section "Discussion"). The hypothesis proposed here is currently being evaluated using *ggpps* single mutants.

Acknowledgments This work was supported by a grant from ETH Zurich (TH-51 06-1) and the EU FP7 contract 245143 (TiMet). The Spanish Ministerio de Ciencia e Innovacion (www.micinn.es) provided grants BIO2008-00432 and BIO2011-23680 to MRC and a doctoral FPI fellowship to ARS. We thank Biswapriya Biswavas Misra and Christian Barucker for their contribution to the work on subcellular localization. We thank Dr. Axel Schmidt for useful discussions on enzymatic activity assays.

Conflict of interest The authors declare that they have no competing interests.

References

Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O (2006) The Arabidopsis MAX pathway controls shoot branching by regulating auxin transport. Curr Biol 16(6):553–563

Bick JA, Lange BM (2003) Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Arch Biochem Biophys 415(2):146–154

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A gene expression map of the Arabidopsis root. Science 302(5652):1956–1960

Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B (2000) Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J 24(2):241–252

Caughey WS, Smythe GA, O'Keeffe DH, Maskasky JE, Smith MI (1975) Heme A of cytochrome c oxicase. Structure and properties: comparisons with hemes B, C, and S and derivatives. J Biol Chem 250(19):7602–7622

Closa M, Vranová E, Bortolotti C, Bigler L, Arró M, Ferrer A, Gruissem W (2010) The *Arabidopsis thaliana* FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest. Plant J 63(3):512–525

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J 16(6):735–743

Cunillera N, Arró M, Delourme D, Karst F, Boronat A, Ferrer A (1996) *Arabidopsis thaliana* contains two differentially expressed farnesyl-diphosphate synthase genes. J Biol Chem 271(13):7774–7780

Cunillera N, Boronat A, Ferrer A (1997) The *Arabidopsis thaliana* FPS1 gene generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate synthase isoform. J Biol Chem 272(24):15381–15388

Czechowski T, Bari RP, Stitt M, Scheible W-R, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J 38(2):366–379

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol 139(1):5–17

Delourme D, Lacroute F, Karst F (1994) Cloning of an *Arabidopsis thaliana* cDNA coding for farnesyl diphosphate synthase by functional complementation in yeast. Plant Mol Biol 26(6):1867–1873

Ducluzeau A-L, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC, Basset GJC (2011) Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J 69(2):366–375

Dugardeyn J, Vandenbussche F, Van Der Straeten D (2008) To grow or not to grow: what can we learn on ethylene-gibberellin cross-talk by in silico gene expression analysis? J Exp Bot 59(1):1–16

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300(4):1005–1016

Floss DS, Walter M (2009) Role of carotenoid cleavage dioxygenase 1(CCD1) in apocarotenoid biogenesis revisited. Plant Signaling Behav 4(3):172–175

Flügge U-I, Gao W (2005) Transport of isoprenoid intermediates across chloroplast envelope membranes. Plant Biol 1:91–97

Hojo M, Matsumoto T, Miura T (2007) Cloning and expression of a geranylgeranyl diphosphate synthase gene: insights into the

synthesis of termite defence secretion. Insect Mol Biol 16(1):121–131

Hruz T, Wyss M, Docquier M, Pfaffl M, Masanetz S, Borghi L, Verbrugghe P, Kalaydjieva L, Bleuler S, Laule O, Descombes P, Gruissem W, Zimmermann P (2011) RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 12(1):156

Hsieh F-L, Chang T-H, Ko T-P, Wang AH-J (2011) Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol 155(3):1079–1090

Hu J, Mitchum MG, Barnaby N, Ayele BT, Ogawa M, Nam E, Lai W-C, Hanada A, Alonso JM, Ecker JR, Swain SM, Yamaguchi S, Kamiya Y, Sun T-p (2008) Potential sites of bioactive gibberellin production during reproductive growth in Arabidopsis. Plant Cell 20(2):320–336

Huang M, Abel C, Sohrabi R, Petri J, Haupt I, Cosimano J, Gershenzon J, Tholl D (2010) Variation of herbivore-induced volatile terpenes among Arabidopsis ecotypes depends on allelic differences and subcellular targeting of two terpene synthases, TPS02 and TPS03. Plant Physiol 153(3):1293–1310

Jiang Y, Proteau P, Poulter D, Ferro-Novick S (1995) BTS1 encodes a geranylgeranyl diphosphate synthase in *Saccharomyces cerevisiae*. J Biol Chem 270(37):21793–21799

Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin Jrm, Rolland N (2009) Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways. Mol Plant 2(6):1154–1180

Kainou T, Kawamura K, Tanaka K, Matsuda H, Kawamukai M (1999) Identification of the GGPS1 genes encoding geranylgeranyl diphosphate synthases from mouse and human. Biochim Biophys Acta Mol Cell Biol Lipids 1437(3):333–340

Karimi M, De DeMeyer B, Hilson P (2005) Modular cloning in plant cells. Trends Plant Sci 10(3):103–105

Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N (2007) Statistical significance of quantitative PCR. BMC Bioinformatics 8(1):131

Kellogg BA, Poulter CD (1997) Chain elongation in the isoprenoid biosynthetic pathway. Curr Opin Chem Biol 1(4):570–578

Koncz C, Shell C (1986) The promoter of T1-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by novel type of Agrobacterium binary vector. Mol Gen Genet 204: 383–396

Kozak M (1997) Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J 16(9): 2482–2492

Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes *in Saccharomyces cerevisiae*. Genome Res 15(10):1421–1430

Lange BM, Ghassemian M (2003) Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. Plant Mol Biol 51(6):925–948

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci USA 107(18):8063–8070

Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, Nambara E, Marion-Poll A (2006) Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. Plant J 45(3):309–319

Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisec J, Beemster GT, Meyer RC, Messerli G, Gronlund JT, Perkovic J, Wigmore E, May S, Bevan MW, Meyer C, Rubio-Diaz S, Weigel D, Micol JL, Buchanan-Wollaston V, Fiorani F, Walsh S, Rinn B, Gruissem W, Hilson P, Hennig L, Willmitzer L, Granier C (2010) Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three Arabidopsis accessions cultivated in ten laboratories. Plant Physiol 152(4):2142–2157

Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K (1990) Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J Bacteriol 172(12):6704–6712

Mitchum MG, Yamaguchi S, Hanada A, Kuwahara A, Yoshioka Y, Kato T, Tabata S, Kamiya Y, Sun T-p (2006) Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. Plant J 45(5):804–818

Nambara E, Marion-Poll A (2005) Abscisic acid biosynthesis and catabolism. Annu Rev Plant Biol 56(1):165–185

Nelson BK, Cai X, Nebenfuhr A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 51(6):1126–1136

Nishino T, Rudney H (1977) Effects of detergents on the properties of 4-hydroxybenzoate. Polyprenyl transferase and the specificity of the polyprenyl pyrophosphate synthetic system in mitochondria. Biochemistry 16(4):605–609

Ohnuma S, Koyama T, Ogura K (1993) Alteration of the product specificities of prenyltransferases by metal ions. Biochem Bioph Res Co 192(2):407–412

Ohnuma S, Suzuki M, Nishino T (1994) Archaebacterial ether-linked lipid biosynthetic gene. Expression cloning, sequencing, and characterization of geranylgeranyl-diphosphate synthase. J Biol Chem 269(20):14792–14797

Okada K, Saito T, Nakagawa T, Kawamukai M, Kamiya Y (2000) Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. Plant Physiol 122(4):1045–1056

Pan J–J, Kuo T-H, Chen Y-K, Yang L-W, Po-Huang L (2002) Insight into the activation mechanism of *Escherichia coli* octaprenyl pyrophosphate synthase derived from pre-steady-state kinetic analysis. Biochim Biophys Acta Protein Struct Mol Enzymol 1594(1):64–73

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29(9):e45

Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Plant Cell 21(12):3718–3731

Rangan L, Vogel C, Srivastava A (2008) Analysis of context sequence surrounding translation initiation site from complete genome of model plants. Mol Biotechnol 39(3):207–213

Rieu I, Powers SJ (2009) Real-time quantitative RT-PCR: design, calculations, and statistics. Plant Cell 21(4):1031–1033

Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, Moorman AFM (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37(6):e45

Ruyter-Spira C, Kohlen W, Charnikhova T, van Zeijl A, van Bezouwen L, de Ruijter N, Cardoso C, Lopez-Raez JA, Matusova R, Bours R, Verstappen F, Bouwmeester H (2011) Physiological effects of the synthetic strigolactone analog GR24 on root system architecture in Arabidopsis: another belowground role for strigolactones? Plant Physiol 155(2):721–734

Sandmann G, Misawa N, Wiedemann M, Vittorioso P, Carattoli A, Morelli G, Macino G (1993) Functional identification of al-3 from *Neurospora crassa* as the gene for geranylgeranyl pyrophosphate synthase by complementation with crt genes, in vitro characterization of the gene product and mutant analysis. J Photoch Photobio B 18(2–3):245–251

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37(5): 501–506

Smyth D, Bowman J, Meyerowitz E (1990) Early flower development in Arabidopsis. Plant Cell 2:755–767

Tan B-C, Joseph LM, Deng W-T, Liu L, Li Q-B, Cline K, McCarty DR (2003) Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. Plant J 35(1):44–56

Tholl D, Lee S (2011) Terpene Specialized Metabolism in *Arabidopsis thaliana*. The Arabidopsis Book:e0143

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The botany array resource: e-Northerns, expression angling, and promoter analyses. Plant J 43(1):153–163

Ubeda-Tomas S, Swarup R, Coates J, Swarup K, Laplaze L, Beemster GTS, Hedden P, Bhalerao R, Bennett MJ (2008) Root growth in Arabidopsis requires gibberellin/DELLA signalling in the endodermis. Nat Cell Biol 10(5):625–628

Ubeda-Tomás S, Federici F, Casimiro I, Beemster GTS, Bhalerao R, Swarup R, Doerner P, Haseloff J, Bennett MJ (2009) Gibberellin signaling in the endodermis controls arabidopsis root meristem size. Curr Biol 19(14):1194–1199

Vallon T, Ghanegaonkar S, Vielhauer O, Muller A, Albermann C, Sprenger G, Reuss M, Lemuth K (2008) Quantitative analysis of isoprenoid diphosphate intermediates in recombinant and wild-type *Escherichia coli* strains. Appl Microbiol Biotechnol 81(1): 175–182

van Schie CC, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC (2007) Geranyl diphosphate synthase is required for biosynthesis of gibberellins. Plant J 52(4):752–762

Vandermoten S, Haubruge E, Cusson M (2009) New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. Cell Mol Life Sci 66(23):3685–3695

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7):research0034.0031–research0034.0011

Vranová E, Hirsch-Hoffmann M, Gruissem W (2011) AtIPD: a curated database of arabidopsis isoprenoid pathway models and genes for isoprenoid network analysis. Plant Physiol 156(4): 1655–1660

Wang G, Dixon RA (2009) Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci USA 106(24): 9914–9919

Yamaguchi S (2008) Gibberellin metabolism and its regulation. Annu Rev Plant Biol 59(1):225–251

Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18(6):292–298

Zhu X, Suzuki K, Okada K, Tanaka K, Nakagawa T, Kawamukai M, Matsuda K (1997a) Cloning and functional expression of a novel geranylgeranyl pyrophosphate synthase gene from Arabidopsis thaliana in *Escherichia coli*. Plant Cell Physiol 38(3):357–361

Zhu X, Suzuki K, Saito T, Okada K, Tanaka K, Nakagawa T, Matsuda H, Kawamukai M (1997b) Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from *Arabidopsis thaliana* is localized in mitochondria. Plant Mol Biol 35(3):331–341