Universitat Autònoma de Barcelona

Exploring the diversity of geranylgeranyl diphosphate synthases in Arabidopsis thaliana and Solanum lycopersicum

M. Victoria Barja Afonso

Universitat Autònoma de Barcelona

UNIVERSITAT AUTÒNOMA DE BARCELONA

FACULTAT DE BIOCIÈNCES

Programa de Doctorat en Biologia i Biotecnologia Vegetal

Exploring the diversity of geranylgeranyl diphosphate synthases in *Arabidopsis thaliana* and *Solanum lycopersicum*

M. VICTORIA BARJA AFONSO

2019

UNIVERSITAT AUTÒNOMA DE BARCELONA

FACULTAT DE BIOCIÈNCES

Programa de Doctorat en Biologia i Biotecnologia Vegetal

PhD Thesis

Exploring the diversity of geranylgeranyl diphosphate synthases in *Arabidopsis thaliana* and *Solanum lycopersicum*

Dissertation presented by Victoria Barja for the degree of Doctor of Plant Biology and Biotechnology at Autonomous University of Barcelona. This work was developed in the Centre for Research in Agricultural Genomics

Thesis Director	Thesis tutor	PhD candidate
Dr. Manuel Rodríguez Concepción	Dr. Roser Tolrà Perez	M. Victoria Barja Afonso

Barcelona, January 2019

But doing so, controlling our own fate, requires that a large fraction of us understand and appreciate Science: How it operates. What it teaches us about the Universe, the Earth, and Life.

Kip Thorne.

Agradecimientos

Esta tesis va dedicada a todos los que, de una forma u otra, habéis hecho esto posible. Gracias a todos por hacer de este viaje una aventura espectacular.

A Manuel. Gracias por haberme dado la maravillosa oportunidad de poder formar parte de tu grupo. Desde que te vi en aquella sala Darwin supe que quería trabajar contigo, lo que no sabía es que me costaría tanto dejar de hacerlo. Gracias por tu dedicación y por tu saber hacer, pero especialmente, por todo lo que me has enseñado. Siempre serás mi referente.

Al Dr. Jules Beekwilder, por acogerme en su laboratorio y enseñarme nuevas formas de trabajar. También a Rumyana, por la disposición y ayuda constantes. Gracias, porque la Ciencia compartida siempre suma.

A aquellos que han participado en el desarrollo de esta investigación. Co-autores y colaboradores, gracias por vuestras aportaciones, dedicación y entusiasmo. Al personal de servicios y administración del CRAG por el soporte y el esfuerzo constante, gracias por hacer de la investigación un camino más fácil. También, al Ministerio de Educación, Cultura y Deporte por financiar mi tesis doctoral.

Al 2.01. Gracias a todos y cada uno de vosotros por hacer de mi trabajo el hobby más maravilloso. Gracias por remar todos codo con codo, por hacer que los logros de uno sean los logros de todos. Por todo, esta tesis también es vuestra. Nunca tendré un equipo mejor.

A todos mis amigos y compañeros del CRAG. Gracias por enseñarme que el networking está en una sonrisa, en una caña y en dormir una hora para darte aliento cuando más lo necesitas. *See you on the way. May the force be with you all.*

A mis amigos de Wageningen, por enseñarme que no es el sitio, sino la gente.

A mi familia de Barcelona. Porque cuando llegamos al río, el puente siempre lo cruzamos juntos. No tengo palabras para describir lo importantes que sois para mí. Me ha encantado cumplir mi sueño con vosotros, y que ahora seáis parte de él.

A Miguel. Gracias por venir a mi vida, pero sobre todo por quedarte. Contigo todo es mejor.

A mi gente de Pamplona. Gracias por hacer que no pase el tiempo, por darme un lugar al que volver.

A mis padres. Gracias por enseñarme el valor de la constancia y el esfuerzo, sois el mejor ejemplo que podría tener. Gracias por enseñarme que '*primero hay que ser persona*' antes que cualquier otra cosa. Y gracias por haberlo sacrificado todo para que nosotros cumplamos nuestros sueños.

A mis hermanos, Marta y Enrique. A vosotros os debo todo lo que soy. Gracias por enseñarme a compartir, a ayudar, a querer. Con vosotros aprendí por primera vez lo que era ser un equipo y que con vosotros yo ya he ganado para siempre. Perdonadme por todas las veces que no he podido estar. Nunca nadie sentirá más orgullo del que yo siento por vosotros. Os admiro y os quiero. *Siempre*.

Summary

Isoprenoids form the largest family of metabolites in nature and are especially abundant and diverse in the plant kingdom. Many plant isoprenoids with essential and specialized functions derive from geranylgeranyl diphosphate (GGPP). The biosynthesis of this isoprenoid precursor is catalyzed by the GGPP synthase (GGPPS) protein family. The GGPPS family has been most studied in *Arabidopsis thaliana*. In this model plant, five gene paralogs encode differentially localized GGPPS isoforms. Strikingly, only the disruption of the gene encoding the AtG11 isoform results in lethality as it is essential for chloroplast and embryo development. Little is known about GGPPS enzymes and their regulation in other plant species of human interest. The goals of this thesis have been (1) to unveil the molecular mechanism responsible for the duality of lethal phenotypes observed in *AtG11* loss-of-function alleles, and (2) to identify the members of the GGPPS family in *Solanum lycopersicum* (tomato) and characterize the contribution of the plastidial isoforms to the biosynthesis of GGPP-derived health-promoting nutrients such as carotenoids.

In the first part of the thesis we demonstrate that the *AtG11* gene produces transcripts of different lengths that result in two differentially targeted GGPPS enzymes. Longer transcripts encode a plastid-targeted enzyme that produces GGPP for isoprenoids involved in photosynthesis, including chlorophylls and carotenoids. Loss of function of this activity results in an albino-lethal phenotype. Shorter transcripts, instead, lack the first ATG codon but are translated from a second in-frame ATG codon. The resulting protein retains GGPPS activity but remains in the cytosol due to the loss of the N-terminal plastid-targeting peptide. This shorter isoform produces GGPP that is required for the progression of embryo development beyond the heart stage. In the second chapter, we show that the five putative GGPPS-encoding genes that are present in the tomato genome encode proteins with different subcellular localizations. Among them, three GGPPS isoforms were found to be targeted to plastids (SlG1, SlG2 and SlG3) and to be specifically associated with carotenoid biosynthesis in particular plant tissues. The *SlG1* gene was induced during root mycorrhization where carotenoid-derived metabolites are required. *SlG2* expression was mainly associated to photosynthetic processes, where carotenoids act as photoprotectants. Finally, *SlG3* was mostly activated during fruit ripening, when carotenoid pigments with high nutritional value are accumulated. The isoform-specific transcriptional profiles and the differential subcellular distribution suggest a strong subfunctionalization of these paralog genes.

The data provided in this thesis contribute to understand the complexity of the GGPPS protein family in plants. This information will be useful to design sustainable strategies to manipulate plants for optimal production of specific groups of GGPP-derived metabolites in particular tissues and subcellular compartments.

Resumen

Los isoprenoides forman la familia más amplia de metabolitos naturales y son especialmente diversos en el reino vegetal. Muchos isoprenoides vegetales con funciones esenciales y especializadas derivan del geranilgeranil difosfato (GGPP). La biosíntesis de este precursor es catalizada por la familia de proteínas GGPP sintasa (GGPPS), que ha sido estudiada principalmente en *Arabidopsis thaliana*. Esta planta modelo consta de cinco genes parálogos que codifican para GGPPSs con diferentes localizaciones subcelulares. Curiosamente, sólo la ausencia del gen que codifica para la isoforma AtG11 produce letalidad, ya que es esencial para el desarrollo embrionario y del cloroplasto. Poco se sabe acerca de las enzimas GGPPS y su regulación en especies vegetales de interés humano. Los objetivos de esta tesis han sido (1) descifrar el mecanismo molecular responsable de la dualidad de fenotipos letales observados en diferentes alelos de pérdida de función del gen *AtG11*, y (2) identificar los miembros de la familia GGPPS en tomate (*Solanum lycopersicum*) y caracterizar el papel de las isoformas plastídicas en la biosíntesis de nutrientes beneficiosos para la salud que derivan del GGPP como son los carotenoides.

En la primera parte de la tesis demostramos que el gen *AtG11* produce tránscritos de diferente longitud que resultan en dos enzimas GGPPS diferencialmente localizadas. Las transcripciones largas se traducen en una proteína de localización plastídica que produce GGPP para la síntesis de isoprenoides involucrados en la fotosíntesis, como clorofilas y carotenoides. La pérdida de función de esta isoforma resulta en un fenotipo albino letal. Los tránscritos cortos, en cambio, carecen del primer codón ATG pero se traducen a partir de un segundo codón ATG que mantiene el marco de lectura. La proteína resultante mantiene la actividad GGPPS pero permanece en el citoplasma debido a la pérdida del péptido de tránsito al plasto en la región N terminal. Esta isoforma corta produce GGPP citosólico que es necesario para la progresión del desarrollo embrionario más allá de la etapa de corazón. En el segundo capítulo mostramos que los cinco genes que codifican para posibles GGPPSs en tomate dan lugar a proteínas con diferente distribución subcelular. Entre ellos encontramos que tres isoformas se localizan en plastos (SlG1, SlG2 y SlG3) y que están asociadas a la síntesis de carotenoides de forma específica en diferentes tejidos vegetales. El gen *SlG1* se induce durante la micorrización donde se necesitan metabolitos derivados de carotenoides. La expresión de *SlG2* se asocia principalmente a procesos fotosintéticos donde los carotenoides actúan como fotoprotectores. Finalmente, encontramos que el gen *SlG3* se activa mayoritariamente durante la maduración del fruto donde se acumulan pigmentos de tipo carotenoide de alto valor nutricional. Los perfiles de expresión específicos y la localización subcelular diferencial sugieren una fuerte subfuncionalización de estos genes.

Los datos presentados en esta tesis contribuyen a entender mejor la complejidad de la familia GGPPS en plantas. Esta información será útil para diseñar estrategias genéticas para generar plantas que produzcan metabolitos derivados de GGPP de alto interés en tejidos o compartimentos celulares particulares de forma más sostenible.

Table of Contents

GENERAL INTRODUCTION	**1**
1. Isoprenoids: a large family of both essential and specialized metabolites	3
2. Biosynthesis of isoprenoids	3
2.1. Biosynthesis of universal isoprenoid precursors	3
2.2. Biosynthesis of prenyl diphosphate intermediates	5
2.3. Types of short-chain prenyltransferases	6
2.3.1. Geranyl diphosphate (GPP) synthases	6
2.3.1. Farnesyl diphosphate (FPP) synthases	7
2.3.1. Geranylgeranyl diphosphate (GGPP) synthases	8
3. *Arabidopsis thaliana* GGPPS family	9
4. GGPP synthases in other crops	11
OBJECTIVES	**15**
CHAPTER I	**17**
1. Introduction	19
2. Results	23
2.1. Different AtG11-defective alleles show a range of phenotypes from variegation to embryo lethality	23
2.2. The distinct phenotypes of AtG11 alleles correlate with differential subcellular localization and activity of the corresponding enzymes	24
2.3. Several transcription initiation sites in the AtG11 gene lead to the production of isoforms with different N-terminal ends	27
2.4. AtG11 activity is essential to produce both plastidial isoprenoids and unidentified extraplastidial isoprenoids required for embryo development.	29
3. Discussion	31
4. Materials and Methods	33
4.1. Plant material and constructs	33
4.2. GGPPS activity assays	33
4.3. Microscopy	34
4.4. 5′-RACE	34
4.5. Analysis of metabolite levels	35
5. Acknowledgements	35
6. References	35
7. Supplemental Information	39
CHAPTER II	**47**
1. Introduction	49
2. Results	52

2.1. Identification of the tomato GGPPS protein family 52
2.2. The predicted tomato GGPPS isoforms localize to different cell compartments 54
2.3. SlG1, SlG2 and SlG3 are functional GGPP-producing enzymes 56
2.4. Tissue-specific gene co-expression networks suggest functional specialization of tomato GGPPS enzymes 58
2.5. Gene expression analyses confirm an isoform-specific association with particular developmental processes involving an activation of carotenoid biosynthesis 62
2.6. A reduction of *SlG3* transcript levels decreases phytoene accumulation in red tomato fruits 64
3. Discussion 66
4. Materials and Methods 71
4.1. Plant material and growth conditions 71
4.2. Constructs 72
4.3. Phylogenetic analysis 72
4.4. Subcellular localization assays 73
4.5. Confocal microscopy 73
4.6. Stable transformation of tomato plants 73
4.7. GGPPS activity assay and product identification 74
4.8. Kinetic analyses 75
4.9. Gene Co-expression Network (GCN) analyses 75
4.10. Gene expression analysis by RT-qPCR 75
4.11. Extraction and analysis of GGPP-derived metabolites 76
5. Acknowledgements 77
6. References 77
7. Supplemental Information 85
GENERAL DISCUSSION 103
1. Subfunctionalization of plant GGPPS paralogs 105
2. The Arabidopsis GGPPS family 107
3. The tomato GGPPS family 109
4. Modulation of GGPP supply by protein-protein interactions 110
5. Open questions 111
CONCLUSIONS **113**
GENERAL REFERENCES **117**
Annex 1 i
Annex 2 ii
Annex 3 iii
Annex 4 iv

List of Figures

GENERAL INTRODUCTION

Figure 1. Distribution of isoprenoid biosynthetic pathways in plant cells 4

Figure 2. Biosynthesis of short-chain prenyl diphosphate intermediates 6

CHAPTER I

Figure 1. Isoprenoid biosynthetic pathways in Arabidopsis 20

Figure 2. AtG11 mutant alleles and associated phenotypes 24

Figure 3. In vitro GGPPS activity assays 26

Figure 4. Transcription and translation start sites in the *AtG11* gene 28

Figure 5. Levels of MVA-derived isoprenoid products in plants with reduced AtG11 activity 30

Figure S1. Arabidopsis *AtG11* gene and mutants 39

Figure S2. Sequence alignment of Arabidopsis GGPPS isoforms 40

Figure S3. Subcellular localization of AtG11-GFP and sG11-GFP fusion proteins in transgenic Arabidopsis plants 41

Figure S4. *In vitro* activity of truncated AtG11 isoforms 42

Figure S5. Embryogenesis is blocked at the heart stage in the *g11-3* mutant. 43

Figure S6. Developmental and metabolic phenotypes of the *ggpps1-1* mutant 44

CHAPTER II

Figure 1. Compartmentalization of isoprenoid biosynthesis in plant cells 51

Figure 2. Phylogenetic tree of Arabidopsis (At) and tomato (Sl) prenyltransferases 54

Figure 3. Subcellular localization of the tomato GGPPS candidates fused to GFP 55

Figure 4. In vitro GGPPS activity assays 57

Figure 5. Gene co-expression networks (GCNs) of tomato genes encoding plastidial GGPPS enzymes in vegetative and fruit tissues. 60

Figure 6. Expression profiles of genes encoding tomato PSY and GGPPS paralogs during different carotenoid biosynthetic processes 63

Figure 7. Transcript and carotenoid levels in red fruits of T1 transgenic lines 65

Figure 8. Model of the channeling of MEP-derived isoprenoid precursors towards the biosynthesis of carotenoids with specific functions in particular tomato tissues 69

Figure S1. Sequence alignment of Arabidopsis and tomato PT proteins 85

Figure S2. Dual subcellular localization of the SlG4-GFP fusion protein. 88

Figure S3. Purification of recombinant 6xHis-tagged GGPPS proteins 89

Figure S4. Transcript levels of *PSY* and plastidial *GGPPS* paralog genes in different tomato tissues 90

Figure S5. Carotenoid levels in fruits from T2 transgenic lines 91

Figure S6. Chlorophyll, tocopherol and phylloquinone levels in fruits from T2 transgenic lines 92

GENERAL DISCUSSION

Figure 1. Comparative representation of our knowledge on Arabidopsis and tomato GGPPS protein families 107

Figure 2. Transcript levels of GGPPS paralog genes in different tomato tissues 110

List of Tables

CHAPTER I

Table 1. Families of isoprenoid biosynthetic genes in Arabidopsis with paralogs that encode more than one isoform. 22

Table S1. Primers used in this work 45

Table S2. List of T-DNA insertion *ggpps* mutant lines used in this work 45

Table S3. Constructs and cloning details 45

CHAPTER II

Table 1. List of predicted GGPPS isoforms in tomato 53

Table 2. Kinetic parameters of plastidial GGPPS enzymes 58

Table S1. List of Arabidopsis (At) and tomato (Sl) prenyltransferases used for sequence and phylogenetic analyses 93

Table S2. List of plastidial isoprenoid-related genes used for the tomato GGPPS GCN analyses 94

Table S3. Individual GCNs of plastidial GGPPS paralogs (Guide genes) positively co-expressed with isoprenoid-related genes (Query genes) in different tomato tissues 98

Table S4. Primers used in this work 100

Table S5. Constructs and cloning details 101

List of Abbreviations

ABA	Abscisic acid
At	*Arabidopsis thaliana*
BRs	Brassinosteroids
CKs	Cytokinins
CLD	Chain-length determination
DMAPP	Dimethylallyl diphosphate
DXP	Deoxyxylulose 5-phosphate
DXS	DXP synthase
ER	Endoplasmic reticulum
FPP	Farnesyl diphosphate
FPPS	FPP synthase
FARM	First aspartate-rich motif
GAs	Gibberellins
GAP	Glyceraldehyde 3-phosphate
GPP	Geranyl diphosphate
GGPP	Geranylgeranyl diphosphate
GFPP	Geranylfarnesyl diphosphate
GPPS	GPP synthase
GGPPS	GGPP synthase
GFPPS	GFPP synthase
GGR	Geranylgeranyl reductase
Hl	*Humulus lupulus*
HMG-CoA	Hydroxymethylglutaryl- CoA
HMGR	HMG-CoA reductase
IPP	Isopentenyl diphosphate
IDI	IPP/DMAPP isomerase
LB	Luria-Bertani
LC-MS	Liquid-chromatography coupled to mass-spectrometry
LSU	Large subunit
MEP	Methylerythritol 4-phosphate
MS	Murashige and Skoog
MVA	Mevalonic acid
Os	*Oryza sativa*
PCR	Polymerase chain reaction
PPi	Inorganic pyrophosphate
PPP	Polyprenyl diphosphate
PPPS	PPP synthase
PSY	Phytoene synthase
PT	Prenyltransferase
RACE	Rapid amplification of cDNA ends

RT-qPCR	Real time quantitative PCR
SARM	Second aspartate-rich motif
SC-PT	Short-chain prenyltransferase
Sl	*Solanum lycopersicum*
SLs	Strigolactones
SPS	Solanesyl diphosphate synthase
SSU	Small subunit
SSUI	Type I SSU
SSUII	Type II SSU
TLC	Thin-layer chromatography

GENERAL INTRODUCTION

Plants are self-sufficient organisms that only use water, minerals and atmospheric CO_2 as raw materials, sunlight as energy source and photosynthesis as workforce to produce organic matter and O_2, the pillars that currently sustain life on Earth. Besides being the primary source of carbon and energy in all food chains, photosynthetic organisms also produce the greatest diversity of biological compounds in nature. This is the result of an evolutionary path to overcome their sessile lifestyle and survive in almost any kind of environment. The development of humankind has always been linked to the use of many of these plant products as foodstuffs but also to manufacture materials or medicines throughout the history. Nowadays, improving our understanding of how plants genuinely transform solar power into useful metabolites is one of the major challenges for humanity to improve our supply of valuable products and successfully fight climate change, overpopulation, and land and water shortage threats.

1. Isoprenoids: a large family of both essential and specialized metabolites

Isoprenoids (also known as terpenoids) comprise, by far, the largest family of plant metabolites, showing a countless variety of chemical structures (Vickers et al., 2014; Tholl, 2015). Some plant isoprenoids, considered 'primary' metabolites, are present in almost all plant species due to their irreplaceable functions in many vital metabolic processes. Among such primary or essential isoprenoids, many are involved in photosynthesis. They include chlorophylls (the main pigments involved in light harvesting and energy transfer), phylloquinones and plastoquinones (prenylquinones participating in the electron transport chain), and carotenoids and tocopherols (protectants of the photosynthetic apparatus against the excess of light). Other essential isoprenoids are ubiquinone (a prenylquinone that is fundamental in the respiratory electron transport chain), phytosterols (lipids that stabilize and confer fluidity to cell membranes), and many plant hormones, such as cytokinins (CKs), brassinosteroids (BRs), gibberellins (GAs), abscisic acid (ABA) and strigolactones (SLs). The vast majority of plant isoprenoids, however, are 'secondary' metabolites that participate in very restricted and specialized processes mostly related with plant-environment interactions. These secondary or specialized isoprenoids are typically confined to particular plant species and/or organs and their synthesis is usually activated in response to environmental challenges. They include volatiles, pigments and defense molecules, many of which are highly relevant for human needs as drugs, flavors, colorants, polymers and, more recently, as nutraceuticals or biofuels (Pulido et al., 2012; Rodríguez-Concepción and Boronat, 2015; Tholl, 2015).

2. Biosynthesis of isoprenoids

2.1 Biosynthesis of universal isoprenoid precursors

The stunning variety of isoprenoid biological functions is explained by the great diversity of chemical structures encompassed in this family of metabolites. However, despite this structural complexity, only two five-carbon (C5) universal precursors, isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP), give rise to all isoprenoids occurring in nature. The biosynthesis of these precursors in plant cells involves two independent pathways (Figure 1), the mevalonic acid (MVA) pathway in the cytosol and the methylerythritol 4-phosphate (MEP) pathway in plastids (Rodríguez-Concepción and Boronat, 2002; Hemmerlin et al., 2012; Rodríguez-Concepción and Boronat, 2015). Ubiquinone, sterols or BRs derive from MVA-derived precursors, whereas chlorophylls, phylloquinone,

plastoquinone, tocopherols, carotenoids, ABA, SLs and GAs use the plastidial pools of IPP and DMAPP produced by the MEP pathway (Figure 1). In the case of secondary metabolites, most C15 sesquiterpenes, C20 diterpenes, and C30 triterpenes derive from MVA-precursors, whereas C10 monoterpenes and C40 tetraterpenes (*i.e.* carotenoid pigments) are usually synthesized from MEP-derived IPP and DMAPP (Figure 1). Although an exchange of common precursors between cell compartments has been demonstrated, the rate under normal conditions is not high enough to overcome the absence of activity of one of the two pathways, thus explaining their coexistence in plants (Bick and Lange, 2003; Schuhr et al., 2003; Rodriguez-Concepcion, 2004; Hemmerlin et al., 2012; Vranová et al., 2013). The tight and sometimes antagonist regulation of both pathways at transcriptional, post-

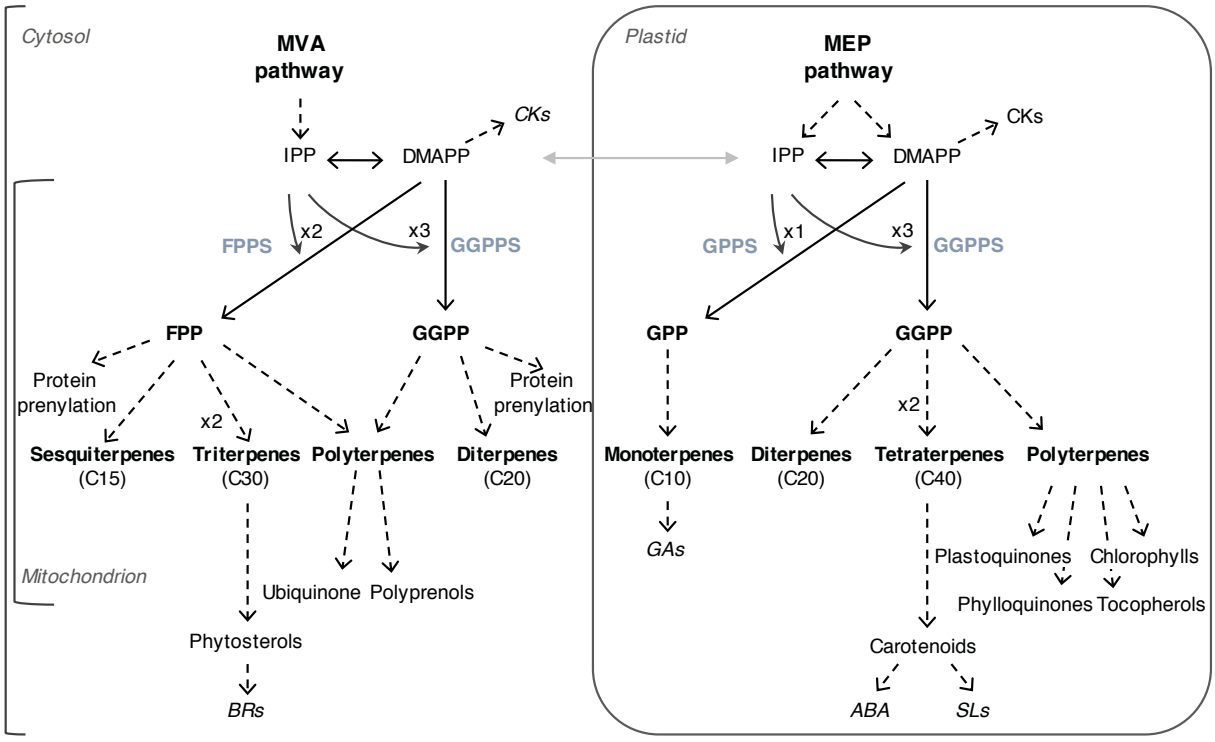


Figure 1. Distribution of isoprenoid biosynthetic pathways in plant cells. IPP and DMAPP, the universal C5 isoprenoid units, are produced by the MVA pathway in the cytosol and the MEP pathway in plastids and can be transported between subcellular compartments (gray arrow). Short-chain prenyl transferase (SC-PT) enzymes (in light blue) catalyze the condensation of one to three molecules of IPP into one molecule of DMAPP producing prenyl diphosphate intermediates of different length (*i.e.* C10 GPP, C15 FPP and C20 GGPP) in different cell compartments (including cytosol, plastids and mitochondria). While GPP synthesis is mainly plastidial, FPP and GGPP can be produced in several cell locations. Some terpene synthases can use different prenyl diphosphates as substrates to produce longer molecules (e.g. ubiquinone can be synthesized from FPP or GGPP). Short-chain prenyl diphosphates and their derived isoprenoid groups are in bold letters. Isoprenoid-derived hormones are indicated in italic letters. Solid arrows represent single enzymatic step and dashed arrows indicate multiple reactions. The brackets on the left indicate biosynthetic steps that can occur in the cytosol and/or in the mitochondrion. *See* Table of Contents for abbreviations.

transcriptional and post-translational level finely modulates the flux of precursors towards the final isoprenoid metabolites (Rodríguez-Concepción and Boronat, 2015).

2.2. Biosynthesis of prenyl diphosphate intermediates

The sequential condensation of an increasing number of IPP units into one molecule of DMAPP results in linear prenyl diphosphate intermediates, the basic backbone structures for the majority of isoprenoid groups. These intermediates, ranging from ten to thousands of carbon atoms, represent the starting point for the astonishing diversification of this metabolic family since they can then undergo a series of enzymatic modifications mainly based on self-condensation, cyclization, isomerization, conjugation and redox reactions. They can also be incorporated to proteins or non-isoprenoid metabolites (*e.g.* chlorophylls or prenylquinones) by alkylation reactions. (McGarvey and Croteau, 1995; Tholl, 2015). The synthesis of these branch-point precursors is mediated by prenyltransferase (PT) enzymes, also known as isoprenyl diphosphate synthases.

PTs catalyze the elimination of the diphosphate moiety from the allylic substrate, that remains as an allylic cation prone to be attacked by an IPP molecule. The addition of the IPP unit generates a new 1'-4 double-bond in the product (Figure 2). Depending on the stereochemical conformation of these double bonds formed during the elongation of the prenyl diphosphate, PTs are classified as *trans*- and *cis*-PTs (Ogura and Koyama, 1998; Liang et al., 2002; Vandermoten et al., 2009). Despite sharing the same substrates and carrying out similar enzymatic reactions, *trans*- and *cis*-PTs form genetically unrelated protein families with completely different protein sequences and different catalytic and substrate binding mechanisms. In general terms, *trans*-PTs are characterized for generating *trans* (E) double bonds in prenyl diphosphates with lengths of up to C50, whereas *cis*-PTs usually produce longer carbon chains with *cis* (Z) double bonds. However, exceptions to this rule include *trans*-PTs and *cis*-PTs that produce longer (Hsieh et al., 2011) or shorter products (Ambo et al., 2008; Sallaud et al., 2009; Schilmiller et al., 2009; Akhtar et al., 2013), respectively. Attending to the length of their products, *trans*-PTs can then be divided in short- (C10-C20), medium- (C25-C35), and long-chain PTs (C40 or longer prenyl diphosphates). Among them, short-chain prenyltransferases (SC-PTs) synthesize C10 geranyl diphosphate (GPP), C15 farnesyl diphosphate (FPP) and C20 geranylgeranyl diphosphate (GGPP), that represent central gears of isoprenoid metabolism operating early in the biosynthetic pathway (Vandermoten et al., 2009; Tholl, 2015). GPP represents the precursor of monoterpenes, FPP is used for the production of

sesqui- and triterpenes, and GGPP is the precursor of di- and tetraterpenes (Figure 1).

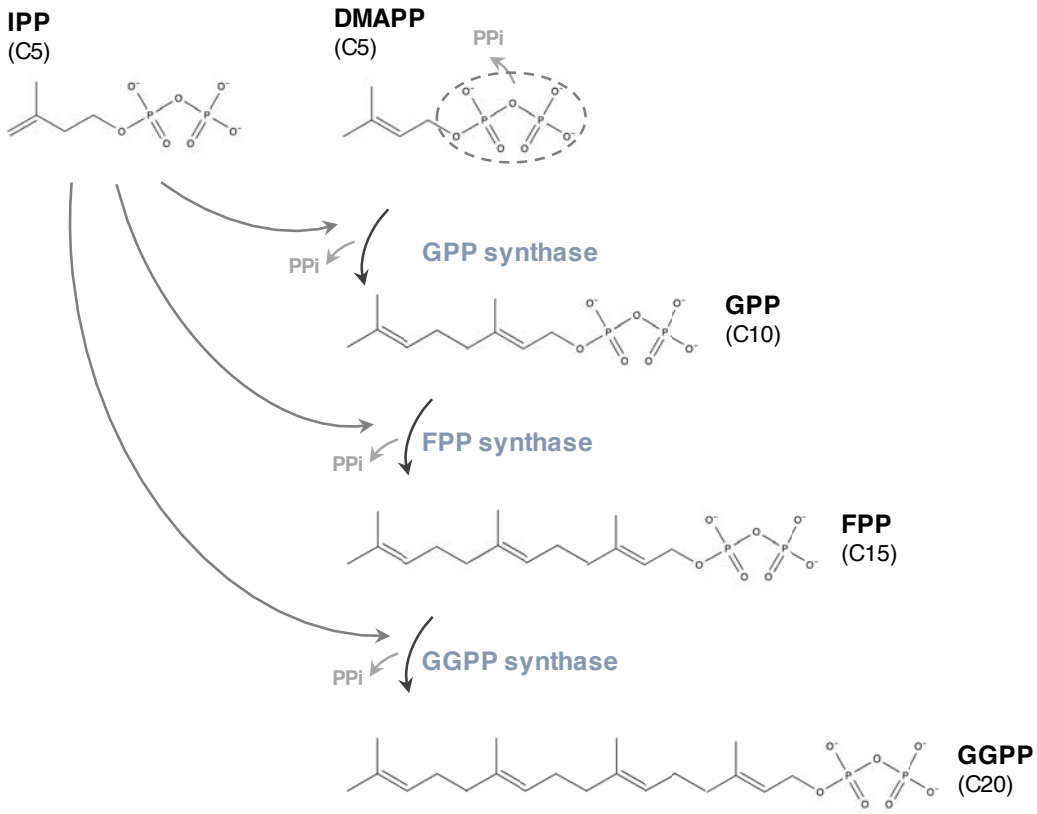


Figure 2. Biosynthesis of short-chain prenyl diphosphate intermediates. Linear prenyl diphosphate molecules result from the sequential addition of IPP units to DMAPP by SC-PT enzymes (in light blue). Each IPP condensation involves the release of one PPi molecule (in gray). Solid arrows represent one enzymatic step/molecule. *See* Table of Contents for abbreviations.

2.3. Types of short-chain prenyltransferases

2.3.1. Geranyl diphosphate (GPP) synthases

GPP synthases (GPPSs) catalyze the condensation of one DMAPP and one IPP molecules to produce GPP, the precursor of all monoterpenes (also called monoterpenoids). Plant monoterpenes are normally produced in plastids as essential oils and volatiles that are mainly involved in allelopathic, defense or pollination processes (Figure 1). Some of them are of industrial interest as fragrances (*e.g.* geraniol), flavors (*e.g.* menthol) or anti-cancer drugs (*e. g.* vinblastine and vincristine) (Rai et al., 2013). GPPS activity can be provided by heterodimeric or homodimeric enzymes in plants. Heterodimeric GPPS enzymes formed by a large and a small subunit (LSU and SSU, respectively) provide GPP in many plant species (Burke et al.,

1999; Tholl et al., 2004; Chang et al., 2010; Rai et al., 2013; Chen et al., 2015). LSUs display significant sequence homology to typical PTs, particularly to GGPP synthases (*see* Section 2.3.3). By contrast, SSUs show lower similarity to PTs and they are only functionally active when physically interact with the catalytic LSU monomer determining its product specificity (Burke et al., 2004; Wang and Dixon, 2009; Tholl, 2015). Two types of SSUs exist in plants (Wang and Dixon, 2009). GPP synthesis and gene expression of type I SSUs (SSUI) correlate well with the production of specific monoterpenes in monoterpene-rich plants (Wang et al., 2008; Dudareva et al., 2003; Wang and Dixon, 2009). The role of type II SSUs (SSUII) is more controversial. Although SSUII monomers can yield GPP when interacting with LSUs (Wang and Dixon, 2009), they have also been found to enhance the production of GGPP when interacting with GGPP-producing LSUs in some plant species (Zhou et al., 2017; Wang et al., 2018b).

Homodimeric GPPS enzymes found in some angiosperms and gymnosperms are also related to GGPP-producing PTs but have suffered different evolutionary pressure, forming two independent lineages (Hsiao et al., 2008; Schmidt and Gershenzon, 2008; Schmidt et al., 2010; Rai et al., 2013). Recently, the activity of some angiosperm homodimeric GPPS proteins (Bouvier et al., 2000; van Schie et al., 2007) has been re-evaluated, showing that these PTs can actually produce long-chain prenyl diphosphates in plastids and mitochondria (Hsieh et al., 2011; Ducluzeau et al., 2012; Jones et al., 2013). The diversity described for GPP-producing enzymes illustrates how difficult it is to predict and even experimentally determine the activity of particular members of the SC-PT family and, hence, to understand their role in plant isoprenoid metabolism.

2.3.2. Farnesyl diphosphate (FPP) synthases

FPP synthases (FPPSs) are enzymes that sequentially condense two IPP units into one molecule of DMAPP to produce FPP. FPPSs constitute the most studied SC-PT subfamily in many organisms, becoming a model for *trans*-PT mechanistic studies (Clarke et al., 1987; Tarshis et al., 1994; Poulter, 2006; Tholl, 2015). FPPSs are enzymatically active as homodimers and form small protein families in plants (Cunillera et al., 1996; Gaffe et al., 2000; Hemmerlin et al., 2003; Richter et al., 2015). FPP is required in the cytosol and mitochondria for the biosynthesis of primary isoprenoids (such as phytosterols and derived BR hormones) and specialized sesqui- and triterpenes, as well as for protein prenylation (Figure 1; (Poulter, 2006)). The model plant *Arabidopsis thaliana* contains two FPPS-encoding genes, named *FPS1* and

FPS2 (Cunillera et al., 1996). While *FPS2* encodes a cytosolic isozyme (FPPS2), the *FPS1* gene is transcriptionally regulated to encode long (FPPS1L) and short (FPPS1S) isoforms that localize to mitochondria and cytosol, respectively (Cunillera et al., 1997; Keim et al., 2012). Despite functioning in different cell compartments and showing specific implications in some particular cell/tissue processes, single mutants hardly differ from wild-type plants (Masferrer et al., 2002; Manzano et al., 2006; Closa et al., 2010; Keim et al., 2012). Only double-mutants (*i.e.* plants completely lacking FPPS activity) show a drastic phenotype of embryo lethality (Closa et al., 2010). FPPS activity is also fundamental during germination, for a proper chloroplast development and in biotic and abiotic responses (Manzano et al., 2016). This basic need of FPP is believed to be mainly associated to the biosynthesis of phytosterols, that are key components for the integrity of biological membranes. A strong impact on sterol levels affects all processes in which membranes are involved such as cell expansion and division, vacuole trafficking, fluidity, permeability or the activity of membrane-bond proteins. The biological relevance of plant sterols would explain the redundant roles of FPPS paralogs (Hartmann, 1998; He et al., 2003; Lenucci et al., 2012; Horvath and Daum, 2013; Grosjean et al., 2015; Li et al., 2015).

2.3.3. Geranylgeranyl diphosphate (GGPP) synthases

GGPP synthase (GGPPS) enzymes generate GGPP by three sequential IPP condensation steps to DMAPP, GPP and finally FPP. Together with FPP, GGPP is a major branch-point intermediate for the biosynthesis of many essential and specialized isoprenoid metabolites in different subcellular locations (Figure 1). GGPP is the precursor of most essential isoprenoids synthesized in plastids, including hormones (GAs, ABA, SLs) and photosynthesis-related isoprenoids such as carotenoids and the side chain of chlorophylls, tocopherols, plastoquinone and phylloquinone. In the cytosol and mitochondria, GGPP also contributes to protein geranylgeranylation and to the production of specialized diterpenes and polyprenols (Liang et al., 2002; Vandermoten et al., 2009; Tholl, 2015). All these GGPP-dependent pathways require differentially localized GGPPS isoforms, that normally form gene families. The most extensive revision of a plant GGPPS family so far has been done in the model plant *Arabidopsis thaliana*.

3. *Arabidopsis thaliana* GGPPS family

Although several GGPPS-encoding gene paralogs were cloned during the last decade of the 20th century (Scolnik and Bartley, 1994, 1995, 1996, Zhu et al., 1997a, 1997b), the first genome-wide list of candidates and unified nomenclature of the Arabidopsis GGPPS family was released in 2003 (Lange and Ghassemian, 2003). Initially, twelve genes (*AtGGPPS1-12*) were predicted in the genome of Arabidopsis to encode GGPPS homologs. However, the subsequent characterization of the candidates revealed that the Arabidopsis GGPPS family only contains five members. *AtGGPPS5* was described as a pseudogene and the AtGGPPS12 isoform was later identified as a type II SSU of a heterodimeric GPP synthase and renamed as AtSSUII (Wang and Dixon, 2009; Beck et al., 2013). The rest of the predicted isozymes (ten) were found to produce GGPP *in vitro* or to genetically complement the absence of GGPPS activity in *Escherichia coli* strains engineered to produce carotenoids (Zhu et al., 1997a, 1997b; Okada et al., 2000; Wang and Dixon, 2009; Beck et al., 2013). However, the use of highly-sensitive analytical methods for prenyl diphosphate identification showed, in further studies (Nagel et al., 2015; Wang et al., 2016), that six of the ten isoforms (AtGGPPS1, and AtGGPPS6-10) produced C25 geranylfarnesyl diphosphate (GFPP) instead of GGPP as major product *in vitro*. AtGGPPS8 was found to also produce even longer prenyl diphosphates. GGPP was detected as the primary product of AtGGPPS2, 3, 4, and 11 isoforms. Traces of GGPP were found in the product mixture of GFPP-producing enzymes (Nagel et al., 2015; Wang et al., 2016), being up to 25% for the AtGGPPS1 isoform in Nagel et al (2015). These remnants of GGPP can explain the *in vitro* and *in vivo* detection of GGPPS activity for these isoforms in previous studies. After these findings AtGGPPS6, 7, 9 and 10 were renamed as AtGFPP1, 2, 3 and 4 respectively, while AtGGPP8 was considered as a polyprenyl diphosphate synthase (AtPPPS2) (Wang et al., 2016). AtGGPPS2, 3, 4 and 11 retained the nomenclature, as well as AtGGPPS1, since its activity differ from one study to another and in both cases significant amounts of GGPP were detected. In summary, based on currently available evidence we can conclude that the Arabidopsis GGPPS family is formed by five members: AtGGPPS1, 2, 3, 4, and 11.

Regarding subcellular localization, GFP-fusions of individual members of the Arabidopsis GGPPS protein family were found in mitochondria (AtGGPPS1), plastids (AtGGPPS2 and 11) and endoplasmic reticulum (ER; AtGGPPS3 and 4), consistent with the cellular distribution of GGPP-consuming pathways (Okada et al., 2000; Bick and Lange, 2003; Beck et al., 2013). Genes encoding the organellar GGPPS enzymes (*AtGGPPS1*, *2* and *11*) are constitutively expressed in most organs and

tissues, with *AtGGPPS11* showing the highest mRNA levels, particularly in photosynthetic tissues. The expression of the genes encoding the ER-associated isoforms (AtGGPPS3 and 4) is, nevertheless, restricted to siliques, flowers and roots (Beck et al., 2013). A deeper study of the genes encoding the two plastidial isoforms showed that the *AtGGPPS2* paralog was mostly co-expressed with genes encoding GA biosynthetic enzymes, whereas the *AtGGPP11* gene was co-regulated with genes involved in the MEP pathway and in the production of photosynthesis-related isoprenoids (Ruiz-Sola et al., 2016a, *see* **Annex 1**). Consistent with the described gene expression pattern and co-regulation network, the AtGGPPS11 protein was found to physically interact with phytoene synthase (PSY), geranylgeranyl reductase (GGR) and solanesyl diphosphate synthase 2 (SPS2), which are GGPP-consuming enzymes involved in the synthesis of photosynthesis-related isoprenoids (Ruiz-Sola et al., 2016a, *see* **Annex 1**). PSY directs the GGPP precursor to carotenoid biosynthesis, GGR transforms GGPP into the phytyl chain of chlorophylls, tocopherols and phylloquinone, and SPS2 uses GGPP to generate the isoprenoid side moiety of plastoquinone. Recent results have confirmed that AtGGPPS11 and PSY work closely together to efficiently channel GGPP into the carotenoid pathway in Arabidopsis (Camagna et al., 2018). Finally, the interaction of AtGGPPS11 with AtSSUII (AtGGPPS12) was found to shift its product specificity from GGPP to GPP, probably to face specific requirements by diverting MEP precursors to monoterpene biosynthesis (Wang and Dixon, 2009; Chen et al., 2015). These findings evidence how the formation of specific AtGGPPS11 protein complexes can contribute to the production of particular classes of plastidial isoprenoids in Arabidopsis.

Unlike that observed for the two Arabidopsis FPPS isoforms, analysis of individual mutants defective in plastidial GGPPS enzymes revealed that AtGGPPS2 and AtGGPPS11 are far from being redundant. Despite being co-expressed with GA biosynthetic genes, the loss of AtGGPPS2 activity did not show any detrimental phenotype compared to wild-type plants indicating that AtGGPPS11 can rescue the absence of this paralog (Ruiz-Sola et al., 2016a, *see* **Annex 1**). By contrast, severe deleterious phenotypes were observed in different mutant alleles of *AtGGPPS11* gene, indicating that the AtGGPPS2 isoform cannot complement the loss of AtGGPPS11 function. Lower levels of *AtGGPPS11* transcripts in the knock-down allele *ggpps11-5* resulted in plants with a pale phenotype, growth delay and smaller mesophyll chloroplasts. These alterations were accompanied by a significant reduction in the levels of carotenoids, chlorophylls and prenylquinones, which were restored to wild-type levels when complementing the *ggpps11-5* mutant line with the

native gene (Ruiz-Sola et al., 2016a, *see* **Annex 1**). These findings confirmed that AtGGPPS11 is a hub GGPPS isozyme for the production of most photosynthesis-related plastidial isoprenoids, while AtGGPPS2 might participate in more specialized processes.

T-DNA insertion mutants causing different truncated versions of the AtGGPPS11 protein were found, however, to result in distinct lethal phenotypes. Truncation of the N-terminal region of the protein resulted in seedling-lethal albino phenotypes, similar to those of knock-out mutants of the MEP pathway, confirming the relevance of this isozyme for the biosynthesis of plastidial isoprenoids such as those involved in photosynthetic processes (Ruppel et al., 2013). Curiously, truncation of the C-terminal region of AtGGPPS11 was found to generate embryo-lethality (Ruppel et al., 2013; Ruiz-Sola et al., 2016a, *see* **Annex 1**), similar to that observed in double mutants devoid of FPPS activity (Closa et al., 2010). **The molecular mechanism underlying these differential phenotypic effects has been addressed in the first chapter of this doctoral thesis** (*see* Chapter I, **Annex 2**).

4. GGPP synthases in other crops

While Arabidopsis is tremendously useful as a model to gain molecular and mechanistic knowledge about basic plant metabolic processes, studies in medicinal and crop species then allow to understand how these processes impact plant features of specific human interest. In this framework, the analysis of GGPPS families from several plant species should contribute to better understand how GGPP-derived metabolites with interesting functions are produced.

Many specialized metabolites synthesized from GGPP in gymnosperms are involved in defense against biotic stresses, including oleoresins in *Picea abies* or taxol in *Taxus* species (Schmidt et al., 2010; Hefner et al., 1998; Liao et al., 2005). Different GGPPS enzymes involved in the production of these compounds under particular conditions have been reported (Schmidt et al., 2010; Hefner et al., 1998; Liao et al., 2005). The functional specialization of the GGPPS family in medicinal plants such as *Tripterigyum wilfordii* and *Andrographis paniculata* has also been addressed to identify the GGPPS isoform(s) involved in the production of bioactive compounds (Zhang et al., 2015; Wang et al., 2018a).

Recently, the study of GGPPS enzymes in a handful of crops has revealed important regulation mechanisms for specific GGPP allocation. It has been discovered that the GGPPS activity in hop (*Humulus lupulus*), rice (*Oryza sativa*) and pepper (*Capsicum*

annum) can be regulated by the interaction with SSU proteins (Wang and Dixon, 2009; Zhou et al., 2017; Wang et al., 2018b). Hop trichomes accumulate myrcene, a monoterpene added as a flavor during beer production. Myrcene derives from GPP, which is produced by the interaction between a GGPPS monomer acting as LSU (HlLSU) and a type I SSU protein (HlSSUI). While the *HlLSU* gene is constitutively expressed to probably supply GGPP for primary biosynthetic pathways, *HlSSUI* is mainly expressed in trichomes, likely mediating myrcene production by changing the product specificity of HlLSU from GGPP to GPP (Wang and Dixon, 2009). In rice, only one GGPPS isoform is targeted to plastids (OsGGPPS1). This enzyme can form homodimers in the stroma of plastids, but it can also be physically recruited to the thylakoid membranes by a type II SSU protein (named OsGRP). The interaction with OsGRP improves the catalytic efficiency of OsGGPPS1 to produce GGPP (Zhou et al., 2017). In thylakoids, the heterodimer further interacts with enzymes involved in chlorophyll biosynthesis, suggesting that the assembly of specific multiprotein complexes containing these enzymes might be a mechanism to efficiently control the flux of GGPP to the production of chlorophylls when the synthesis of these pigments is required (Zhou et al., 2017). In red pepper, carotenoid accumulation is responsible for the characteristic color of ripe fruit. Fruit carotenoid biosynthesis was associated to only one GGPPS isoform (CaGGPPS1), which can also interact with a type II SSU (CaSSUII). The CaSSUII protein promotes GGPP production and also the interaction with pepper PSY (Wang et al., 2018b). This result demonstrates again the ability of type II SSU enzymes to recruit GGPPS monomers for the biosynthesis of specific isoprenoid groups.

Tomato (*Solanum lycopersicum*) is one of the best plant systems to study the biosynthesis of GGPP-derived plastidial isoprenoids and its regulation. In particular, tomato carotenoids play important roles in roots (e.g. mycorrhization), photosynthetic tissues (e.g. photoprotection), and fruits (e.g. pigmentation). However, information on GGPPS number, distribution and function is scarce in this species. Only two GGPPS isoforms (SlGGPPS1 and SlGGPPS2) have been identified (Ament et al., 2006). The corresponding genes showed particular expression patterns and responded differentially to biotic stress conditions, suggesting subfunctionalization. However, their subcellular localization was not experimentally confirmed and their enzymatic activity was only indirectly inferred (Ament et al., 2006). **In the second chapter of this thesis, we address the identification of the tomato GGPPS family members, carry out the functional characterization of the**

plastidial isoforms, and provide new information about GGPP allocation for carotenoid biosynthesis in particular tomato tissues.

OBJECTIVES

This research work aims to deepen into the molecular regulation of GGPP supply in different plant cell compartments for the synthesis of essential and specialized isoprenoids. For this, the following specific objectives were established:

1. Unveil the molecular mechanism responsible for the distinct lethal phenotypes observed in different loss-of-function alleles of *Arabidopsis thaliana* GGPPS11.

2. Identify the tomato GGPPS family members, functionally characterize the plastidial isoforms, and investigate how GGPP is allocated for carotenoid biosynthesis in different tomato tissues.

CHAPTER I

A Single Arabidopsis Gene Encodes Two Differentially Targeted Geranylgeranyl Diphosphate Synthase Isoforms.

Abstract

A wide diversity of isoprenoids is produced in different plant compartments. Most groups of isoprenoids synthesized in plastids and some produced elsewhere in the plant cell derive from geranylgeranyl diphosphate (GGPP) synthesized by GGPP synthase (GGPPS) enzymes. In *Arabidopsis thaliana*, 5 genes appear to encode GGPPS isoforms localized in plastids (2), the endoplasmic reticulum (2), and mitochondria (1). However, the loss of function of the plastid-targeted GGPPS11 isoform (referred to here as AtG11) is sufficient to cause lethality. Here we show that the absence of a strong transcription initiation site in the *AtG11* gene results in the production of transcripts of different lengths. The longer transcripts encode an isoform with a functional plastid import sequence that produces GGPP for the major groups of photosynthesis-related plastidial isoprenoids. However, shorter transcripts are also produced that lack the first translation initiation codon and rely on a second in-frame ATG codon to produce an enzymatically active isoform lacking this N-terminal domain. This short enzyme localizes in the cytosol and is essential for embryo development. Our results confirm that the production of differentially targeted enzyme isoforms from the same gene is a central mechanism to control the biosynthesis of isoprenoid precursors in different plant cell compartments.

---

*The results presented in this Chapter have been published in the *Plant Physiology* research journal (**Annex 2**).

A Single Arabidopsis Gene Encodes Two Differentially Targeted Geranylgeranyl Diphosphate Synthase Isoforms.

M. Victoria BARJA[1], M. Águila RUIZ-SOLA[1], David MANZANO[1], Briardo LLORENTE[1], Bert SCHIPPER[2], Jules BEEKWILDER[2], Manuel RODRIGUEZ-CONCEPCION[1]

1. Program of Plant Metabolism and Metabolic Engineering, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, 08193 Barcelona, Spain.
2. Plant Research International, PO Box 16, 6700, AA, Wageningen, The Netherlands.

NOTE: The experiments included in this thesis Chapter were co-designed and performed by the PhD candidate under the supervision of the PhD director. The contribution of the rest of the authors to the results reported here was the following: M.A.R.-S. started the project and provided some biological materials (constructs and transgenic lines); D.M. helped with the measurement of sterol levels; B.S., and J.B. performed prenylquinone profiling experiments; B.L. provided technical assistance and discussion on the 5'-RACE experimental design. The published version of this Chapter complemented with additional experiments is attached as **Annex 2**.

1. Introduction

Plants produce tens of thousands of isoprenoid compounds, including some that are essential for respiration, photosynthesis, and regulation of growth and development. Despite their structural and functional diversity, all isoprenoids derive from the same five-carbon (C5) precursors, the double-bond isomers isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which can be interconverted by IPP/DMAPP isomerase (IDI) enzymes. Plants use two unrelated pathways to synthesize these units (Figure 1). The mevalonic acid (MVA) pathway synthesizes IPP in the cytosol, whereas the methylerythritol 4-phosphate (MEP) pathway supplies both IPP and DMAPP in the plastid (Bouvier et al., 2005; Vranova et al., 2013; Rodriguez-Concepcion and Boronat, 2015). IPP and DMAPP units can be exchanged between cell compartments to a certain level. For example, MVA-derived IPP can be imported by mitochondria for the biosynthesis of ubiquinone (Lütke-Brinkhaus et al., 1984; Disch et al., 1998). However, this limited exchange of common isoprenoid precursors is not active enough to rescue a genetic or pharmacological

blockage of one of the pathways with IPP/DMAPP produced by the non-inhibited pathway (Bouvier et al., 2005; Vranova et al., 2013; Rodriguez-Concepcion and Boronat, 2015). Addition of IPP units to DMAPP generates longer prenyl diphosphate molecules, including C10 geranyl diphosphate (GPP), C15 farnesyl diphosphate (FPP), and C20 geranylgeranyl diphosphate (GGPP), which are then used in specific downstream pathways to produce particular isoprenoids (Figure 1). FPP and GGPP pools represent nodes of the major metabolic branch points in the isoprenoid biosynthesis network (Vranova et al., 2011; Vranova et al., 2013). As prenyl diphosphates grow longer, however, their transport between cell compartments becomes increasingly restrained (Bick and Lange, 2003).

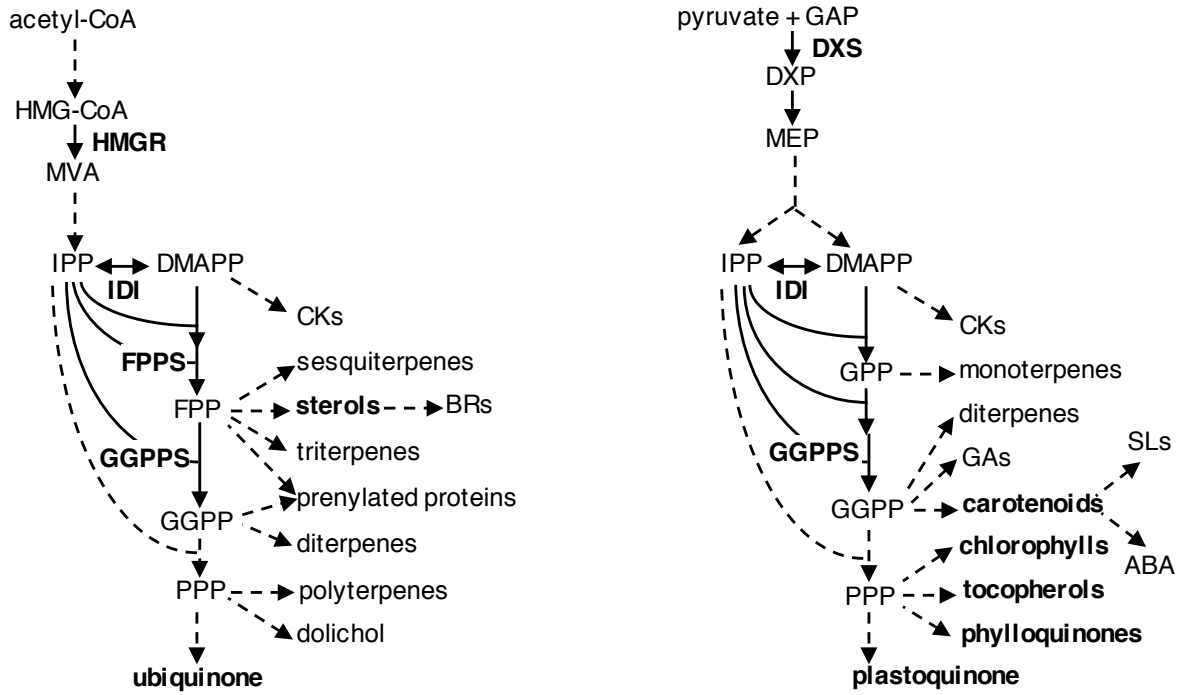


Figure 1. Isoprenoid biosynthetic pathways in Arabidopsis. Mevalonic acid (MVA) pathway and its derived metabolites are represented on the left. Methylerythritol 4-phosphate (MEP) pathway derivatives are depicted on the right. Solid arrows indicate single enzymatic steps and dashed arrows represent multiple steps. HMG-CoA, hydroxymethylglutaryl-CoA; GAP, glyceraldehyde 3-phosphate; DXP, deoxyxylulose 5-phosphate; HMGR, HMG-CoA reductase; DXS, DXP synthase; IDI, IPP/DMAPP isomerase. *See* Table of Contents for the rest of abbreviations.

The two pathways for the production of isoprenoid precursors have been extensively studied in *Arabidopsis thaliana*. All the MEP pathway enzymes are encoded by nuclear genes and imported into plastids, whereas cytosolic, endoplasmic reticulum (ER), and peroxisomal-associated locations have been found for MVA enzymes (Pulido et

al., 2012; Rodriguez-Concepcion and Boronat, 2015). The main rate-determining enzymes of the MEP and MVA pathways are deoxyxylulose 5-phosphate synthase (DXS) and hydroxymethylglutaryl CoA reductase (HMGR), respectively (Figure 1). Most plants contain small gene families encoding these two enzymes (Rodriguez-Concepcion and Boronat, 2015). While several Arabidopsis genes encode proteins with homology to DXS, only one of them produces a functional enzyme with DXS activity (Phillips et al., 2008b). In the case of HMGR, the *HMG1* gene produces long and short transcripts encoding two enzyme isoforms (HMGR1L and HMGR1S, respectively) that only differ in their N-terminal region, whereas the *HMG2* gene produces only one isoform, HMGR2 (Caelles et al., 1989; Enjuto et al., 1994; Lumbreras, 1995). The three HMGR isoforms are primarily attached to the ER and have the same topology in the membrane, with the highly divergent N-terminal region and the highly conserved catalytic domain exposed to the cytosol (Table 1).

Downstream enzymes such as IDI, FPP synthase (FPPS), and GGPP synthase (GGPPS) are also encoded by small gene families in Arabidopsis and localize to multiple subcellular compartments (Table 1). The two genes encoding IDI in Arabidopsis, *IDI1* and *IDI2*, produce long and short transcripts encoding enzyme isoforms that differ in length at their N-terminal ends (Okada et al., 2008; Phillips et al., 2008a; Sapir-Mir et al., 2008). The long IDI1 isoform is targeted to plastids, the long IDI2 isoform is transported to mitochondria, and both short isoforms are sorted to peroxisomes. The two genes encoding FPPS in Arabidopsis produce three isoforms (Cunillera, 1997; Manzano et al., 2006; Keim et al., 2012). *FPS1* encodes a long isoform targeted to mitochondria (FPP1L) and a short one lacking the N-terminal end that remains in the cytosol, whereas *FPS2* only produces a cytosolic enzyme (Table 1). Unlike IDI and FPPS, GGPPS paralogs are encoded by a high number of genes in plant genomes, with a particularly large gene family present in Arabidopsis (Lange and Ghassemian, 2003; Coman et al., 2014). From the twelve initially reported genes, however, only four have been conclusively shown to encode true GGPPS enzymes (Nagel et al., 2015; Wang et al., 2016). Two of them, *GGPPS3* and *GGPPS4*, encode proteins sorted to the ER, and the other two, *GGPPS2* and *GGPPS11*, encode plastidial isoforms (Zhu et al., 1997a; Zhu et al., 1997b; Okada et al., 2000; Wang and Dixon, 2009; Beck et al., 2013; Coman et al., 2014; Ruiz-Sola et al., 2016a, *see* **Annex 1**). The *GGPPS1* gene encodes the only mitochondrial member of the family, but the *in vivo* activity of the protein is still unclear (Zhu et al., 1997b; Okada et al., 2000; Beck et al., 2013; Nagel et al., 2015; Wang et al., 2016). To date, the production of more than one enzyme isoform from a single GGPPS-encoding gene has not been reported.

Table 1. Families of isoprenoid biosynthetic genes in Arabidopsis with paralogs that encode more than one enzyme isoform. Abbreviations: HMGR, HMG-CoA reductase; DXS, DXP synthase; IDI, IPP/DMAPP isomerase; FPPS, FPP synthase; GGPPS, GGPP synthase; ER, endoplasmic reticulum. *Isoforms reported in this work. **GGPPS activity unclear.

Enzyme	Gene	Accession	Isoform	Localization
HMGR	*HMG1*	At1g76490	HMGR1S	ER-cytosol
			HMGR1L	ER-cytosol
	HMG2	At2g17370	HMGR2	ER-cytosol
DXS	*DXS*	At4g15560	DXS	Plastids
IDI	*IDI1*	At5g16440	IDI1S	Peroxisomes
			IDI1L	Plastids
	IDI2	At3g02780	IDI2S	Peroxisomes
			IDI2L	Mitochondria
FPPS	*FPS1*	At5g47770	FPPS1S	Cytosol
			FPPS1L	Mitochondria
	FPS2	At4g17190	FPPS2L	Cytosol
GGPPS	*GGPPS1*	At1g49530	GGPPS1**	Mitochondria
	GGPPS2	At2g18620	GGPPS2	Plastids
	GGPPS3	At2g18640	GGPPS3	ER
	GGPPS4	At2g23801	GGPPS4	ER
	GGPPS11	At4g36810	GGPPS11L / AtG11*	Plastids
			GGPPS11S / sG11*	Cytosol

Despite the presence of at least two GGPPS enzymes in Arabidopsis plastids, *GGPPS11* (*At4g36810*, from herein referred to as *AtG11*) is by far the most abundant and ubiquitously expressed paralog (Beck et al., 2013; Ruiz-Sola et al., 2016a, *see* **Annex 1**). AtG11 is required for the production of all major groups of plastidial isoprenoids, including carotenoids and the side chains of chlorophylls, tocopherols, and prenylated quinones (Ruiz-Sola et al., 2016a, *see* **Annex 1**). Strikingly, several phenotypes have been described for *AtG11*-defective mutant alleles (Ruppel et al., 2013; Ruiz-Sola et al., 2016a, *see* **Annex 1**). By carrying out a comprehensive analysis of these alleles, we uncovered here the existence of two differentially targeted AtG11 enzymes, each of them indispensable for a distinct developmental process likely through the production of different types of essential isoprenoids.

2. Results

2.1. Different *AtG11*-defective alleles show a range of phenotypes from variegation to embryo lethality

To better understand the role of AtG11, the most abundant GGPPS isoform in Arabidopsis, we carefully revised the phenotypes associated to partial or complete loss of function mutants (Figure 2). The *ggpps11-1* mutant, originally named *ggps1-1* (Ruppel et al., 2013) and here referred to as *g11-1*, harbors a point mutation that changes a conserved residue (D163R) in AtG11 (Figure 2A and Figure S1). The *g11-1* allele shows a temperature-dependent variegated phenotype that resembles that of the *chs5* mutant (Araki et al., 2000), later renamed as *dxs-3* (Phillips et al., 2008), which harbors a D627N mutation in DXS. It is therefore likely that the phenotype of these mutants might be associated to the temperature sensitivity of the corresponding mutant enzymes, both of them involved in the production of photosynthesis-related isoprenoids (Figure 1).

A second partial loss-of-function allele, *ggpps11-5* (SALK_140601, *g11-5*), shows a pale phenotype and a developmental delay (Figure 5A), probably, because a T-DNA insertion upstream the predicted ATG translation start codon (Figure 2A and Figure S1) results in a decreased accumulation of *AtG11* transcripts (Ruiz-Sola et al., 2016a, *see* **Annex 1**). In this mutant, lower levels of AtG11-encoding transcripts are expected to result in an overall reduction in the accumulation of fully active, wild-type AtG11 protein (Ruiz-Sola et al., 2016a, *see* **Annex 1**). Similarly, a general inhibition of the MEP pathway with sublethal concentrations of the DXS inhibitor clomazone, also results in a pale phenotype (Pulido et al., 2013; Perello et al., 2014). Therefore, the phenotype of *g11-1* and *g11-5* plants is consistent with the reported role of AtG11 as the major isoform transforming MEP-derived precursors into GGPP for photosynthesis-related isoprenoid products. Further supporting this conclusion, a seedling-lethal albino phenotype visually identical to that of knock-out MEP pathway mutants such as *dxs-1*, also known as *cla1* (Phillips et al., 2008), was observed in the case of the *ggpps11-2* line (SALK_015098, *ggps1-2* or *g11-2*), which harbors a T-DNA insertion in the N-terminal end of the protein coding region of the *AtG11* gene (Figure 2, and Figure S1) (Ruppel et al., 2013). By contrast, T-DNA insertions interrupting the C-terminal end of the AtG11 protein in alleles *ggpps11-3* (SALK_085914, *ggps1-3* or *g11-3*) and *ggpps11-4* (SAIL_712_D06, *g11-4*) cause an arrest of embryo development (Figure 2 and Figure S1) (Ruppel et al., 2013; Ruiz-Sola et al., 2016a, *see* **Annex 1**). This embryo lethal phenotype has never been observed in MEP pathway mutants (Phillips et al., 2008b).

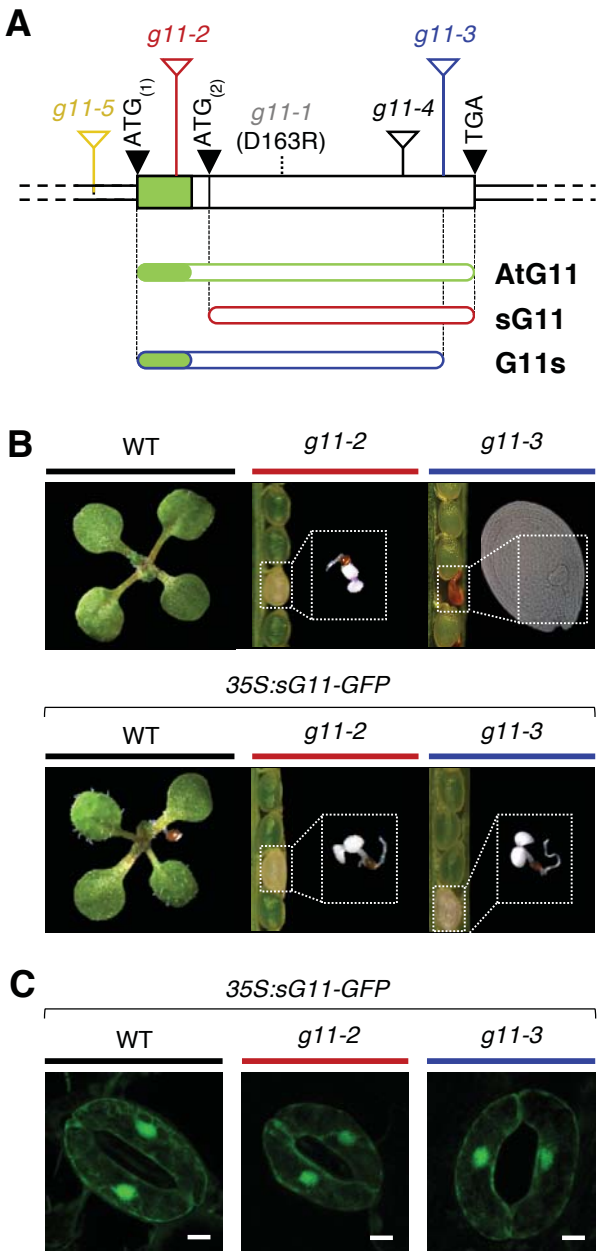


Figure 2. *AtG11* mutant alleles and associated phenotypes. (A) *AtG11* gene model according to TAIR v10 annotation. The protein-coding sequence (which lacks introns) is shown as a wider box with a green section corresponding to the predicted plastid targeting sequence. The position of translation start and stop codons is shown with black triangles. The position of T-DNA insertions is represented with empty triangles. The position of the point mutation in the *g11-1* allele is shown with a dashed line. Lower bars represent encoded proteins. **(B)** Phenotype of *AtG11*-defective mutants either producing or not producing a sG11 protein. Representative seedlings of the indicated genotypes grown under long-day conditions for 10 d are shown to the same scale. Segregating populations of seeds in siliques of plants heterozygous for the *g11-2* and *g11-3* mutations is also shown. Boxed seeds correspond to the homozygous albino mutants represented in the right. Brownish seeds did not produce seedlings due to a blockage in embryo development at the heart stage (as shown in the corresponding picture). **(C)** Cytosolic localization of the sG11-GFP protein. Pictures show GFP fluorescence from the sG11-GFP protein in stomata from 10-d-old seedlings of the indicated genotypes (White bars, 5 µm).

2.2. The distinct phenotypes of *AtG11* alleles correlate with differential subcellular localization and activity of the corresponding enzymes

To investigate the molecular basis of the puzzling phenotype differences observed between *g11-2* (albino, seedling lethal) and *g11-3* (embryo lethal) plants (Figure 2B), the position of the T-DNA in the mutant genomes had been previously validated in the lab by PCR amplification and sequencing of the insertion sites (Figure S1). Insertion of the T-DNA within the predicted plastid targeting sequence in the *g11-2*

allele (Figure 2A and Figure S1) is expected to prevent the transcription of a full-length *AtG11* cDNA harboring the first ATG codon ($ATG_{(1)}$). However, a second in-frame ATG codon ($ATG_{(2)}$) exists downstream the T-DNA insertion that could potentially act as an alternative translation initiation point to generate a shorter protein, that we named sG11 (Figure 2A and Figure S2). We speculated that this shorter protein might not be imported to plastids as it lacked the N-terminal plastid targeting domain. To test this prediction, the localization of GFP fusion proteins was analyzed in transgenic Arabidopsis lines previously generated in the lab (Ruiz-Sola et al., 2016a, *see* **Annex 1;** Ruiz-Sola et al., 2016b, *see* **Annex 2**). As shown in Figure 2C and Figure S3, green fluorescence corresponding to the sG11-GFP protein (Ruiz-Sola et al., 2016b, *see* **Annex 2**) was excluded from plastids and localized in the cytosol. By contrast, transgenic plants expressing a similar construct with the wild-type AtG11 sequence (Ruiz-Sola et al., 2016a, *see* **Annex 1**) showed a predominant association of GFP fluorescence to plastids (Figure S3).

We next evaluated whether sG11 retained the genuine enzymatic activity of the full-length enzyme (Figure 3). Recent *in vitro* activity assays followed by analysis of reaction products by LC-MS (Nagel et al., 2015) or TLC (Wang et al., 2016) confirmed that AtG11 synthesizes GGPP as the main product. They also revealed that other proteins previously believed to be true GGPPS isoforms actually belong to a novel type of prenyl diphosphate synthases that mainly produce C25 geranylfarnesyl diphosphate (GFPP) or longer products (Nagel et al., 2015; Wang et al., 2016). The sG11 protein lacks 19 residues predicted to be present in the N-terminal region of the mature AtG11 enzyme (Figure 2A and Figure S2). While this N-terminal sequence is not conserved in other GGPPS enzymes and does not include residues determining product length (Figure S2), we aimed to experimentally confirm whether its absence in sG11 had any impact on the ability of the protein to produce GGPP from IPP and DMAPP. Previous experiments in the lab showed that the sG11 protein was able to complement for the loss of GGPPS activity in *E. coli* strains engineered to produce carotenoids (Ruiz-Sola et al., 2016b, *see* **Annex 2**). As a complementary approach, we carried out *in vitro* activity assays like those described in Nagel et al. (2015) using protein extracts from *E. coli* cells overproducing sG11 or a pseudomature form of AtG11 lacking the predicted plastid-targeting sequence (Figure S4). Analysis of reaction products by LC-MS detected the production of similar amounts of GGPP in both AtG11 and sG11 extracts (Figure 3). No GPP, FPP or GFPP were detected in the assays (Figure S4), further indicating that the sG11 protein remains as an active GGPPS enzyme. Together, our results suggest that the lack of the N-terminal region

in the sG11 protein produced by *g11-2* plants prevents its targeting to plastids but does not override GGPPS activity.

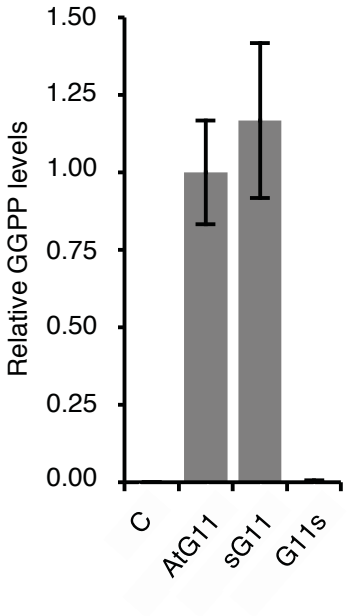


Figure 3. *In vitro* GGPPS activity assays. Protein extracts from *E. coli* cells overproducing similar amounts of the indicated proteins or transformed with an empty vector (C) were mixed with IPP and DMAPP, and the production of GGPP was quantified by LC-MS. Levels are represented relative to those in AtG11 samples. Mean and SD of n=3 extracts are shown. No GGPP was detected in C samples, and only traces were identified in G11s extracts.

In the case of the *g11-3* mutant, the T-DNA interrupts the sequence encoding the highly conserved C-terminal region of AtG11 (Figure S2), resulting in a shorter protein that we named G11s (Figure 2A). Previous sequencing of the T-DNA insertion site in the *g11-3* genome had confirmed that the last 21 residues of the wild-type AtG11 protein are replaced by a single serine residue in the G11s protein (Figure S1). To test whether the loss of the C-terminal region compromised GGPPS activity in the G11s protein, the corresponding DNA sequence was amplified from *g11-3* seedlings and cloned into plasmids for *E. coli* expression. Activity assays showed that the recombinant G11s protein is unable to produce GGPP (Figure 3). These results suggest that the blockage of embryo development at the heart stage observed in the *g11-3* mutant (Figure 2B and Figure S5) and the *g11-4* allele (Ruiz-Sola et al., 2016a, *see* **Annex 1**) might be due to a complete lack of AtG11 activity. This embryo lethal phenotype was complemented by expressing a genomic *AtG11* sequence including the promoter and the full protein-coding region (Ruiz-Sola et al., 2016a, *see* **Annex 1**). Most interestingly, embryo development was also rescued by expressing the cytosolic sG11-GFP protein in *g11-3* plants (Figure 2B and Figure 2C). Transgenic *g11-3 35S:sG11-GFP* plants, however, showed a seedling-lethal (albino) phenotype resembling the *g11-2* mutant. As expected, the cytosolic sG11-GFP protein was unable to rescue the albino phenotype of the *g11-2* mutant (Figure 2B). We therefore concluded that embryo development beyond the heart stage required the presence of

AtG11 activity in the cytosol (despite the existence of two ER-associated GGPPS enzymes in Arabidopsis, GGPPS3 and 4), whereas photosynthetic seedling development required the activity of AtG11 in the plastid (although a second plastidial enzyme with the same activity, GGPPS2, is found in this plant) (Table 1).

2.3. Several transcription initiation sites in the AtG11 gene lead to the production of isoforms with different N-terminal ends

A number of Arabidopsis genes encoding isoprenoid biosynthetic enzymes have been shown to produce transcripts of different lengths encoding isoforms with or without N-terminal transit peptides for plastids and mitochondria (Table 1). To determine whether a similar alternative transcription initiation mechanism also occurs for *AtG11*, rapid amplification of cDNA 5′ ends (5′-RACE) experiments were performed on RNA extracted from different tissues to assess the length of *AtG11* transcripts *in vivo* (Figure 4). The protocol used reverse primers optimized for 5′-RACE (Figure 4A and Table S1) to amplify gene-specific PCR products while ruling out the possibility of genomic DNA contamination (*see* Methods). Separation of PCR-amplified products by gel electrophoresis showed the presence of cDNAs of different sizes (Figure 4B), suggesting that the *AtG11* gene lacks a strictly defined transcription start site. However, amplified fragments could be grouped in two major classes: "long" (*ca*. 0.5 kb or longer) and "short" (*ca*. 0.4 kb or shorter). The 5′-RACE products amplified from siliques were cloned and sequenced. Analysis of inserts revealed that all the "long" products included the $ATG_{(1)}$ codon and hence encoded the full-length AtG11 protein. While most "short" products contained the $ATG_{(1)}$ codon with or without a few nucleotides upstream (up to 25), some lacked $ATG_{(1)}$ and so they can only produce the cytosolic sG11 protein by using the $ATG_{(2)}$ codon (Figure 4A and Figure S1). The similar pattern of "long" and "short" transcripts detected in seedlings, rosette leaves, and flowers by 5′-RACE experiments (Figure 4B) suggests that both types of transcripts are likely produced in all tissues. To verify whether the relative abundance of transcripts either lacking or not the $ATG_{(1)}$ codon was indeed similar in different tissues, 5′-RACE products from seedlings were cloned and compared with those from siliques by digestion with *Nco*I. As shown in Figure 4A, a *Nco*I target site overlaps the $ATG_{(1)}$ codon and therefore it can be used to rapidly identify clones lacking this sequence. In both seedlings and siliques, about 10% of the clones could not be cleaved by *Nco*I, confirming that transcripts exclusively encoding sG11 are low abundant but indeed detectable at similar levels in different tissues.

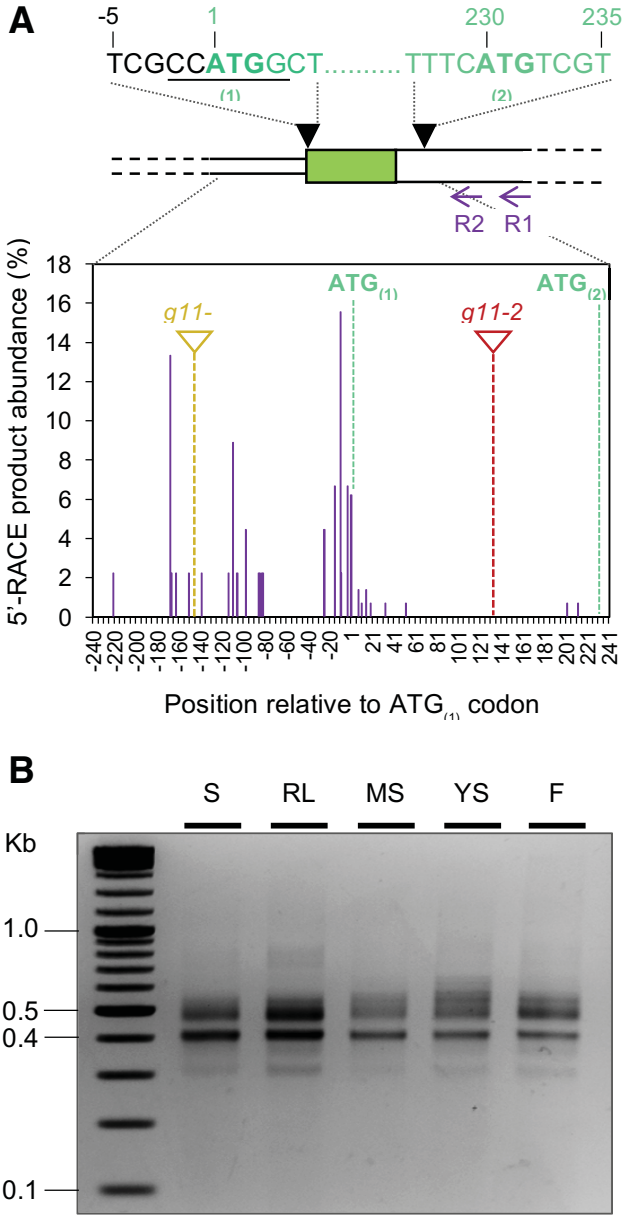


Figure 4. Transcription and translation start sites in the *AtG11* gene. (A) *AtG11* gene model showing the position of presumed translation start codons ATG(1) and ATG(2) (in bold). The NcoI site in the ATG(1) region is underlined. Numbers indicate the position relative to the first nucleotide of the ATG(1) codon. The position of primers AtG11-R-R1 and R2 used for 5′-RACE experiments is also shown. Amplified 5′-RACE products (B) were cloned and sequenced to determine the position of transcription initiation sites. The graph represents the percentage of transcripts found to start at the indicated position in siliques based on the analysis of 145 clones (purple columns). The position of ATG(1) and ATG(2) codons and T-DNA insertion sites of the indicated alleles is also shown. **(B)** Agarose gel electrophoresis of 5′-RACE products isolated from seedlings (S), rosette leaves (RL), mature siliques (MS), young siliques (YS), and flowers (F).

2.4. AtG11 activity is essential to produce both plastidial isoprenoids and unidentified extraplastidial isoprenoids required for embryo development.

We next aimed to determine the nature of the isoprenoids derived from GGPP produced by plastidial and cytosolic forms of AtG11. Previous results with the *g11-5* allele showed that the pale phenotype of the mutant (Figure 5A) was due to a decreased expression of the *AtG11* gene, which caused a reduced accumulation of the major groups of MEP-derived isoprenoids produced from GGPP in the plastid, *i.e.* carotenoids, chlorophylls, tocopherols, phylloquinones, and plastoquinone (Ruiz-Sola et al., 2016a, *see* **Annex 1**). It was also found that carotenoid levels in *g11-2* seedlings were similar to those in the *dxs-1* mutants (with a complete block of the MEP pathway; Phillips et al., 2008b) but much higher than the amounts detected in *psy-1* seedlings (which show a similar albino phenotype due to a specific blockage in the carotenoid pathway; Pokhilko et al., 2015) (Ruiz-Sola et al., 2016b, *see* **Annex 2**). We therefore conclude that the albino phenotype of *g11-2* seedlings is due to a defective production of MEP-derived GGPP to feed downstream isoprenoids in the plastid.

To identify the isoprenoid metabolite required for embryo development that is produced from sG11-derived GGPP, it would be necessary to compare the metabolite profile of *g11-2* and *g11-3* embryos in the transition from globular to heart stage (Figure 2 and Figure S5). Because this is extremely challenging, we followed an alternative approach and evaluated whether the levels of extraplastidial isoprenoids were altered in the *g11-5* mutant, which is expected to produce lower amounts not only of plastid-localized AtG11 but also of cytosolic sG11 enzymes. While MVA-derived precursors are used to produce a wide variety of isoprenoids (Figure 1), only sterols and ubiquinone appear to be required for proper embryo development (Schrick et al., 2000; Okada et al., 2004). Sterols are not synthesized from GGPP but from FPP (Figure 1). We therefore expected that reducing sG11 activity (*i.e.* cytosolic GGPP production) in *g11-5* seedlings would not cause a decreased sterol accumulation. Consistently, the levels of sterols (campesterol, stigmasterol and sitosterol) were found to be similar in wild-type and *g11-5* seedlings (Figure 5B)

In the case of ubiquinone, the predominant form in Arabidopsis (UB-9) contains a C45 solanesyl moiety synthesized by a mitochondrial polyprenyl diphosphate synthase that elongates an initial FPP or GGPP acceptor with IPP units (Hsieh et al., 2011; Ducluzeau et al., 2012). Mitochondria import MVA-derived IPP from the cytosol, as they lack their own biosynthetic pathway (Lütke-Brinkhaus et al., 1984; Disch et al., 1998). Then, specific isoforms of FPPS and GGPPS enzymes targeted to

mitochondria (in Arabidopsis, FPPS1L and GGPPS1; Table 1) are presumed to produce FPP and GGPP for ubiquinone synthesis. FPPS1L-defective *fps1* mutants only showed a limited decrease in UB-9 levels (Closa et al., 2010), suggesting that the biosynthesis of the ubiquinone solanesyl chain might predominantly rely on the supply of GGPP by GGPPS1, the only GGPPS isoform known to be targeted to mitochondria (Beck et al., 2013; Nagel et al., 2015). However, we found that the T-DNA insertion mutant *ggpps1-1* (SAIL_559_G01) contains wild-type levels of UB-9 (Figure S6). Levels of other GGPP-derived isoprenoids were also unaltered in *ggpps1-1* seedlings, which showed a wild-type phenotype in terms of plant growth and development (Figure S6). Actually, the role of GGPPS1 as a true GGPPS enzyme still remains controversial as it has not been conclusively established whether its main product is GGPP (Wang et al., 2016) or GFPP (Nagel et al., 2015). In any case, our data show that the product of mitochondrial GGPPS1 is not required for ubiquinone biosynthesis. Analysis of UB-9 contents in seedlings, flowers and young siliques of the *g11-5* mutants detected slightly reduced levels compared to wild-type samples, but the differences were not found to be statistically significant (Figure 5C). In the case of mutant siliques, however, the observed reduction in ubiquinone levels was close to statistical significance (Student's *t* test, p=0.051). These results together suggest that sG11 might be somehow involved in the production of GGPP for ubiquinone synthesis.

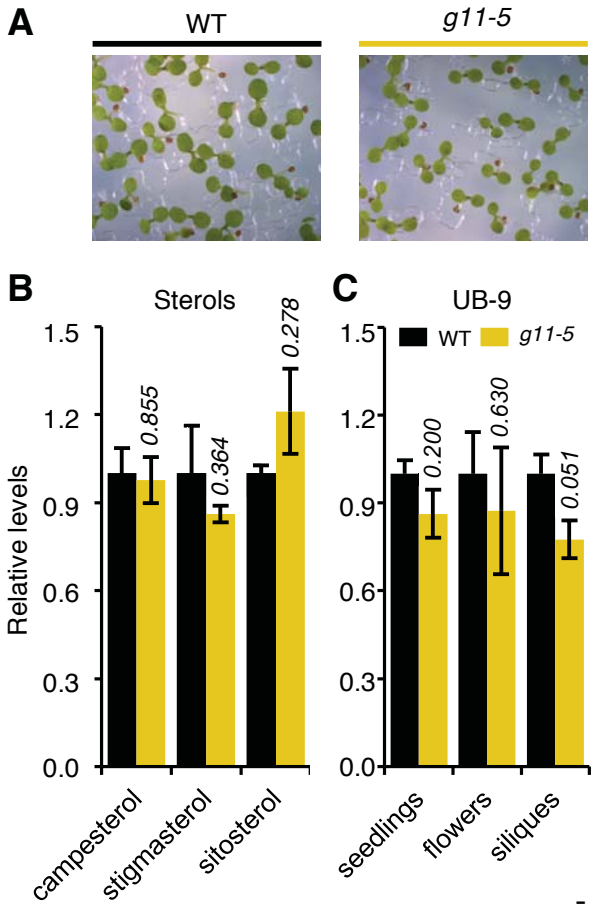


Figure 5. Levels of MVA-derived isoprenoid products in plants with reduced *AtG11* activity. (A) Seedlings of the *g11-5* mutant and the parental Columbia wild-type (WT) grown for 1 week under short-day conditions. These seedlings were used for the metabolite analyses shown in the corresponding graphs. **(B)** Levels of major sterols in WT and *AtG11*-defective *g11-5* seedlings. **(C)** Levels of the ubiquinone UB-9 in seedlings, flowers and young siliques of WT and *g11-5* plants. The graphs represent mean and SE of at least three independent samples (n=3). Italic numbers above the bars indicate P values (Student's t test) relative to WT samples.

3. Discussion

Our results support the conclusion that the Arabidopsis *AtG11* gene can produce two enzyme isoforms: one translated from the $ATG_{(1)}$ codon and carrying a plastid-targeting peptide (AtG11), and a shorter version translated from the downstream $ATG_{(2)}$ codon and lacking the plastid-targeting peptide (sG11). Mechanistically, the production of these two differentially targeted isoforms might rely on both the use of alternative transcription start sites (resulting in the synthesis of mRNAs with or without a 5'-region encoding the plastid transit peptide) or the use of alternative translation start sites ($ATG_{(1)}$ or $ATG_{(2)}$) in the long transcript. The *NetStart* algorithm for the prediction of translation start codons in Arabidopsis (http://www.cbs.dtu.dk/services/NetStart/) actually gives a similar score to $ATG_{(1)}$ (0.627) and $ATG_{(2)}$ codons (0.651), suggesting that both could be functional *in vivo*. It is remarkable that a similar situation has been reported to occur in other Arabidopsis genes encoding key isoprenoid biosynthetic enzymes, including HMGR, IDI, and FPPS. Unlike these other enzymes, however, no redundancy appears to occur in the case of Arabidopsis GGPPS enzymes, as the loss of AtG11 cannot be rescued by plastidial GGPPS2, ER-localized GGPPS3 and GGPPS4, or mitochondrial GGPPS1. Furthermore, the methionine residue encoded by the $ATG_{(2)}$ codon in AtG11 does not appear to be conserved in the rest of isoforms confirmed to synthesize GGPP (Figure S2), suggesting that *AtG11* might be the only Arabidopsis GGPPS-encoding gene producing more than one isoform.

Our analysis and those previously reported (Ruiz-Sola et al. 2016a, *see* **Annex 1**; Ruiz-Sola et al. 2016b, *see* **Annex 2**), clearly show that AtG11 activity in the plastid is indispensable for the production of plastidial isoprenoids supporting chloroplast development and photosynthesis. The low levels of carotenoids detected in *g11-2* albino seedlings (Ruiz-Sola et al. 2016b, *see* **Annex 2**) might derive from small amounts of GGPP synthesized from MEP-derived precursors by GGPPS2 (the other plastidial GGPPS enzyme found in Arabidopsis). In the case of the *dxs-1* mutant, however, the MEP pathway is completely blocked and hence it is most likely that MVA-derived IPP or DMAPP precursors are transported to the plastid and used to produce GGPP and downstream products. Alternatively, an enhanced import of cytosolic GGPP by non-differentiated plastids like those found in the albino mutants would make DXS and AtG11 (but not PSY) dispensable to produce carotenoids. No clear-cut evidence is yet available, however, of an exchange of long prenyl diphosphates such as GGPP between cell compartments (Bick and Lange, 2003). On the other hand, sG11 activity in the cytosol supplies the precursors for an

unidentified isoprenoid metabolite that is essential for embryo development. Since mitochondrial FPP and GGPP are not the main precursors for ubiquinone biosynthesis, cytosolic GGPP might be transported to mitochondria and used for the production of this respiratory metabolite. Blockage of this cytosolic source when sG11 activity is lost could actually explain why embryo development is arrested at the same heart stage in mutants defective in ubiquinone synthesis (Okada et al., 2004) and sG11 activity, *i.e. g11-3* (Figure S5) and *g11-4* (Ruiz-Sola et al., 2016a, *see* **Annex 1**). As complete loss of FPPS activity in Arabidopsis *fps1 fps2* double mutants blocks embryogenesis at the earlier globular stage (Closa et al., 2010), it can be concluded that while FPP-derived isoprenoids are needed for the transition of the embryo from globular to heart stages, progression beyond the heart stage requires ubiquinone or/and a different metabolite produced from sG11-supplied precursors. Thus, all these observations suggest two possibilities. First, sG11-derived GGPP might be critical for the biosynthesis of ubiquinone during embryogenesis. Considering that a *ca.* 50% reduction of *AtG11* transcript levels in *g11-5* seedlings only causes a 20-30% decrease in plastidial isoprenoid content (Ruiz-Sola et al., 2016a, *see* **Annex 1**), the observed 20% reduction of UB-9 levels in young siliques of the *g11-5* mutant might indeed result from partly reduced sG11 activity (and hence GGPP supply) in developing embryos or seeds. Because other GGPPS isoforms are also expressed in these tissues at different stages (Beck et al., 2013), it remains unclear why none of them is able to complement the loss of sG11 activity and hence rescue embryo development in *g11-3* and *g11-4* mutants. Alternatively, sG11 might produce GGPP for another specific class of unidentified metabolites required for embryogenesis and, perhaps, with roles in other cells and tissues during the plant life cycle, as deduced from the wide distribution of sG11-encoding transcripts (Figure 4B).

Besides synthesizing GGPP as a homodimer, the plastidial AtG11 isoform has been found to produce GPP upon heterodimerization with another plastidial GGPPS-like protein, AtSSUII (Wang and Dixon, 2009; Chen et al., 2015). Whether the cytosolic sG11 protein unveiled here also has alternative enzyme activities upon association with other proteins remains unknown. In any case, our results support the conclusion that the production of GGPP required for essential functions in different cell compartments relies on the activity of AtG11 isoforms. Other GGPPS paralogs might be maintained in the Arabidopsis genome for developmental and/or condition-specific subfunctionalization. Future experiments, including the high-resolution analysis of isoprenoid profiles in specialized tissues and organs (*e.g.* embryos) from wild-type and GGPPS-defective mutants, should provide additional

insights on the biological role of specific isoforms and their corresponding downstream GGPP-derived products, further allowing to understand the astounding complexity of the mechanisms used by plants to produce isoprenoids.

4. Materials and Methods

4.1. Plant material and constructs.

All the *Arabidopsis thaliana* lines used in this work are in the Columbia background. The T-DNA insertion alleles *g11-2*, *g11-3* and *g11-5* were already available in the lab (Pokhilko et al., 2015; Ruiz-Sola et al., 2016a, *see* **Annex 1, Table S2**). Homozygous *ggpps1-1* mutants were isolated from a segregating population of the SAIL_559_G01 line supplied by the European Arabidopsis Stock Centre (http://arabidopsis.info/). Primers for PCR-based genotyping of the mutants and sequencing of T-DNA insertion sites are indicated in Table S1. Constructs and cloning details are described in Table S3. The *35S:sG11-GFP* transgenic lines were generated as described (Ruiz-Sola et al., 2016b, *see* **Annex 2**). Seeds were surface-sterilized and germinated in Petri dishes with solid Murashige and Skoog (MS) medium without vitamins or sucrose. After stratification for 3 days at 4 °C, plates were incubated in a growth chamber at 22 °C and illuminated for 16h (long-day) or 8h (short-day) a day with fluorescent white light at a photosynthetic photon flux density of 60 μmol $m^{-2}s^{-1}$.

4.2. GGPPS activity assays

Constructs to express different truncated AtG11 versions were generated as described in Table S3. Recombinant proteins were produced in *E. coli* BL21 pGROE cells using the Overnight Express AutoInduction System 1 (Merck Millipore). After growth for 72 h at 18 °C, bacterial cells were recovered by centrifugation and pellets were resuspended in reaction buffer (25 mM HEPES pH 7.2, 10 mM $MgCl_2$, 10% v/v glycerol) supplemented with 1 mg/ml lysozyme and one tablet of complete protease inhibitor cocktail (Roche) for every 10 ml of buffer. The resuspended pellet was incubated at 4 °C for 10 min and after a brief sonication (5 pulses of 20 s at 30 W) the cell lysate was centrifuged at 19.000×*g* at 4 °C for 5 min. The supernatant was used for SDS-PAGE, protein quantification, and GGPP activity assays as described (Nagel et al., 2015). The reaction mix contained 10 μg of total protein from extracts showing similar levels of recombinant protein in 200 μl of reaction buffer, 50 μM IPP and 50 μM DMAPP. After incubation for 1 h at 30 °C, the reaction was stopped by adding

800 µl of methanol. A previously described LC-FTMS system (Henneman et al., 2008) was adapted to detect prenyl diphosphate products. A Luna C18(2) pre-column (2.0 × 4 mm) and an analytical column (2.0 × 150 mm, 100 nm, particle size 3 µm) from Phenomenex were used for chromatographic separation at 40 °C, using an Acquity UPLC (H-Class), connected to an LTQ-Orbitrap XL hybrid MS system (Waters) operating in negative electrospray ionization mode heated at 300 °C with a source voltage of 4.5 kV for full-scan LC–MS in the m/z range 100 to 1300. The injection volume was 10 µL. Compounds were separated by a linear gradient between solution A (20 mM NH_4HCO_3 with 0.1% triethylamine) and solution B (acetonitrile/H_2O, 4:1 with 0.1% triethylamine). The gradient was as follows: 0–18 min: 100% A to 80% A; 18–23 min: 80% A to 0% A; 23-25 min: 0% A; 25–30 min: 0% A to 100% A; equilibration with 100% A. Acquisition and visualization of the data were performed using *Xcalibur* software (Thermo Fischer). GPP, FPP and GGPP standards were obtained from Sigma and used for quantification.

4.3. Microscopy

Subcellular localization of the GFP fusion proteins was analyzed by direct examination of plant tissue samples with an Olympus FV 1000 Confocal Laser Scanning Microscope. Green fluorescence corresponding to the fusion proteins was detected using an argon laser for excitation with blue light at 488 nm and a 500-510 nm filter, whereas a 610-700 nm filter was used for detection of chlorophyll autofluorescence. Clearing of Arabidopsis seeds and examination of embryo developmental stages was performed as described (Ruiz-Sola et al., 2016a, *see* **Annex 1**).

4.4. 5′-RACE

For Rapid Amplification of cDNA 5′ Ends (5′-RACE) analysis, total RNA from different organs of wild-type plants was isolated using a RNA purification kit (Sigma-Aldrich) and used for first-strand cDNA synthesis with the SMARTer RACE cDNA amplification kit (Clontech). High Fidelity AccuPrime Taq DNA Polymerase (Invitrogen) was used with primers provided in the kit and the AtG11-RACE-R1 primer (Table S1) for 5′-RACE reactions, adding an initial denaturation step of 2 min at 94ºC to the recommended PCR program to activate the polymerase, and changing the elongation temperature from 72 °C to 68 °C. PCR products were cloned into the

cloning vector pCRII-TOPO (Invitrogen) for further restriction enzyme and sequencing analysis using the AtG11-RACE-R2 primer (Table S1).

4.5. Analysis of metabolite levels.

Published methods were used for the extraction, separation, and quantification of photosynthetic pigments (chlorophylls and carotenoids) (Rodríguez-Concepción et al., 2004), sterols (Closa et al., 2010), and prenylquinones, including ubiquinone (Martinis et al., 2011).

5. Acknowledgements

We are very grateful to Gaétan Glauser (University of Neuchâtel, Switzerland) for the analysis of prenylquinones. Technical support from M. Rosa Rodríguez-Goberna and members of the CRAG Services is greatly appreciated. This work was funded by grants from the Ministerio de Economía y Competitividad (BES-2009-025359, BIO2011-23680, BIO2014-59092-P, PCIN-2015-103, BIO2015-71703-REDT and FPDI-2013-018882), Ministerio de Educación, Cultura y Deporte (FPU14/05142), European Commission H2020 ERA-IB-2 program (BioProMo), and Generalitat de Catalunya (2014SGR-1434 and 2015FI_B 00007).

6. References

Araki N, Kusumi K, Masamoto K, Niwa Y and 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.

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodriguez-Concepcion M, Gruissem W and Vranova E (2013). Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. Plant Mol Biol **82:** 393-416.

Bick JA and 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:** 146-154.

Bouvier F, Rahier A and Camara B (2005). Biogenesis, molecular regulation and function of plant isoprenoids. Prog Lipid Res **44:** 357-429.

Caelles C, Ferrer A, Balcells L, Hegardt FG and Boronat A (1989). Isolation and structural characterization of a cDNA encoding *Arabidopsis thaliana* 3-hydroxy-3-methylglutaryl coenzyme A reductase. Plant Mol Biol **13:** 627-638.

Chen Q, Fan D and Wang G (2015). Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in Arabidopsis flowers. Mol Plant **8:** 1434-1437.

Closa M, Vranova E, Bortolotti C, Bigler L, Arro M, Ferrer A and 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:** 512-525.

Coman D, Altenhoff A, Zoller S, Gruissem W and Vranova E (2014). Distinct evolutionary strategies in the GGPPS family from plants. Front Plant Sci **5:** 230.

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

Disch A, Hemmerlin A, Bach TJ and Rohmer M (1998). Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. Biochem J **331:** 615-621.

Ducluzeau AL, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC and Basset GJ (2012). Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J **69:** 366-375.

Enjuto M, Balcells L, Campos N, Caelles C, Arro M and Boronat A (1994). *Arabidopsis thaliana* contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. Proc Natl Acad Sci U S A **91:** 927-931.

Henneman L, van Cruchten AG, Denis SW, Amolins MW, Placzek AT, Gibbs RA, Kulik W and Waterham HR (2008). Detection of nonsterol isoprenoids by HPLC-MS/MS. Anal Biochem **383:** 18-24.

Hsieh FL, Chang TH, Ko TP and Wang AH (2011). Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol **155:** 1079-1090.

Keim V, Manzano D, Fernandez FJ, Closa M, Andrade P, Caudepon D, Bortolotti C, Vega MC, Arro M and Ferrer A (2012). Characterization of Arabidopsis FPS isozymes and *FPS* gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds. PLoS One **7:** e49109.

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

Lumbreras V, Campos N and Boronat A (1995). The use of an alternative promoter in the *Arabidopsis thaliana HMG1* gene generates an mRNA that encodes a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase isoform with an extended N-terminal region. Plant J **8:** 541-549.

Lütke-Brinkhaus F, Liedvogel B and Kleinig H (1984). On the biosynthesis of ubiquinones in plant mitochondria. Eur J Biochem **141:** 537-541.

Manzano D, Busquets A, Closa M, Hoyerova K, Schaller H, Kaminek M, Arro M and Ferrer A (2006). Overexpression of farnesyl diphosphate synthase in Arabidopsis mitochondria triggers light-dependent lesion formation and alters cytokinin homeostasis. Plant Mol Biol **61:** 195-213.

Martinis J, Kessler F and Glauser G (2011). A novel method for prenylquinone profiling in plant tissues by ultra-high pressure liquid chromatography-mass spectrometry. Plant Methods **7:** 23.

Nagel R, Bernholz C, Vranova E, Kosuth J, Bergau N, Ludwig S, Wessjohann L, Gershenzon J, Tissier A and Schmidt A (2015). *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate, geranylfarnesyl diphosphate. Plant J **84:** 847-859.

Okada K, Kasahara H, Yamaguchi S, Kawaide H, Kamiya Y, Nojiri H and Yamane H (2008). Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in Arabidopsis. Plant Cell Physiol **49:** 604-616.

Okada K, Ohara K, Yazaki K, Nozaki K, Uchida N, Kawamukai M, Nojiri H and Yamane H (2004). The *AtPPT1* gene encoding 4-hydroxybenzoate polyprenyl diphosphate transferase in ubiquinone biosynthesis is required for embryo development in *Arabidopsis thaliana*. Plant Mol Biol **55:** 567-577.

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

Perello C, Rodriguez-Concepcion M and Pulido P (2014). Quantification of plant resistance to isoprenoid biosynthesis inhibitors. Methods Mol Biol **1153:** 273-283.

Phillips MA, D'Auria JC, Gershenzon J, Pichersky E (2008a). The *Arabidopsis thaliana* type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. Plant Cell **20:** 677-696.

Phillips MA, Leon P, Boronat A and Rodriguez-Concepcion M (2008b). The plastidial MEP pathway: unified nomenclature and resources. Trends Plant Sci **13:** 619-623.

Pokhilko A, Bou-Torrent J, Pulido P, Rodriguez-Concepcion M and Ebenhoh O (2015). Mathematical modelling of the diurnal regulation of the MEP pathway in Arabidopsis. New Phytol **206:** 1075-1085.

Pulido P, Perello C and Rodriguez-Concepcion M (2012). New insights into plant isoprenoid metabolism. Mol Plant **5:** 964-967.

Pulido P, Toledo-Ortiz G, Phillips MA, Wright LP and Rodriguez-Concepcion M (2013). Arabidopsis J-protein J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. Plant Cell **25:** 4183-4194.

Rodriguez-Concepcion M and Boronat A (2015). Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. Curr Opin Plant Biol **25:** 17-22.

Rodríguez-Concepción M, Forés O, Martínez-García JF, González V, Phillips MA, Ferrer A and Boronat A (2004). Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during Arabidopsis seedling development. Plant Cell **16:** 144-156.

Ruiz-Sola MA, Coman D, Beck G, Barja MV, Colinas M, Graf A, Welsch R, Rutimann P, Buhlmann P, Bigler L, Gruissem W, Rodriguez-Concepcion M and Vranova E (2016a). Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol **209:** 252-264. *See* **Annex 1**.

Ruiz-Sola MA, Barja MV, Manzano D, Llorente B, Schipper B, Beekwilder J and Rodriguez-Concepcion M (2016b). A single Arabidopsis gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. Plant Physiol. **172**: 1393–1402. *See* **Annex 2**.

Ruppel NJ, Kropp KN, Davis PA, Martin AE, Luesse DR and Hangarter RP (2013). Mutations in *GERANYLGERANYL DIPHOSPHATE SYNTHASE 1* affect chloroplast development in *Arabidopsis thaliana* (*Brassicaceae*). Am J Bot **100:** 2074-2084.

Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, Gidoni D and Eyal Y (2008). Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. Plant Physiol **148:** 1219-1228.

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

Vranova E, Coman D and Gruissem W (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol **64:** 665-700.

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

Wang C, Chen Q, Fan D, Li J, Wang G and Zhang P (2016). Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in *Brassicaceae* plants. Mol Plant **9:** 195-204.

Wang G and Dixon RA (2009). Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci U S A **106:** 9914-9919.

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

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

7. Supplemental Information

Figure S1. Arabidopsis *AtG11* gene and mutants. (A) The protein coding sequence according to TAIR v10 annotation is shown in uppercase (no introns are found). The first ATG codon ($ATG_{(1)}$) and the translation stop codon are marked in bold. The sequence encoding the plastid-targeting peptide is boxed in green. A second in-frame ATG codon ($ATG_{(2)}$) is boxed with a black frame. Transcription start sites identified by 5′-RACE in siliques are marked with triangles whose color represents the percentage of transcripts starting at the indicated position (white: <5%; gray: 5-10%; black: >10%). T-DNA insertions are represented with arrows pointing towards the right border. The point mutation in the *g11-1* allele is boxed in gray. **(B)** Comparison of genomic sequences and encoded proteins in wild-type (WT) and *g11-3* plants in the region harboring the T-DNA insertion in the mutant. The blue box marks the T-DNA sequence inserted and the corresponding translation.

```
AT1G49530_G1       -----------------MRPRYSLILSAMRL--------I--------------------
AT2G18620_G2       ----------MTTLNLSIFPSVKISSSASIPGFIKIQPFL---LRRKLSTV-----LSVT
AT2G18640_G3       MEAQNIFLYLLIVFLSLHFVFTTLKGRLSPANTRRLIRLLHIPIKSPVAAAIFARKDTRE
AT2G23800_G4       MEPQILFLYLSLFILSLNFFFTNLKPR--------LVRLFQPSLESRVKTALLSRKEVAA
AT4G36810_G11      MASVTLGSWIVVHHHNHHHPSSILTKSRSRSCPITLTKPISFRSKRTVSSSSSIVSSSVV
                                          :              :

AT1G49530_G1       ----------------RPSNRRLSSIASSDSEFISYMKNKAKSINKALDNSIPLCNNFVP
AT2G18620_G2       ARDEGII--------------------HNHFDFTSYMIGKANAVNEALDSAVSLRE----
AT2G18640_G3       FLDSSIKLVNEED------------DFGFSFDFKPYMISKAETINRALDEAIPLIE----
AT2G23800_G4       FLDSPIVEDEEGEEREEEEGGIVSNANFTFEFDPYMMSKAESVNKALEEAIPVGE----
AT4G36810_G11      TKEDNLRQ-----------------SEPSSFDFMSYIITKAELVNKALDSAVPLRE----
                                           :*  *:  **: :*.**:.:: : :

AT1G49530_G1       LWEPVLEVHKAMRYTLLPGGKRVRPMLCLVACELVGGQESTAMPAACAVEMIHAASLILD
AT2G18620_G2       ----PIKIHEAIRYSLLARGKRVRPVLCIAACELVGGEESVALPAACAVEMIHTMSLIHD
AT2G18640_G3       ----PLNIHKAMRYAILAGGKRVRPILCLAACELVGGEERLAIQAACAVEMIHTMSLIKD
AT2G23800_G4       ----PLKIHEAMRYAILAAGKRVRPILCLASCELVGGQENAAMPAACAVEMIHTMSLIKD
AT4G36810_G11      ----PLKIHEAMRYSLLAGGKRVRPVLCIAACELVGGEESTAMPAACAVEMIHTMSLIHD
                       :::*:*:**::*  ******:**:.:******:*  *: *********: *** *
                   FARM
AT1G49530_G1       DLPCMDDDSLRRGKPTNHKVFGEKTSILASNALRSLAVKQTLAST-SLGVTSERVLRAVQ
AT2G18620_G2       DLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALISFAFEHLATS---TAVSPARVVRAIG
AT2G18640_G3       DLPCMDNDDLRRGKPTTHKVFGESVAILSGGALLALAFEHLTEA----DVSSKKMVRAVK
AT2G23800_G4       DLPCMDNDDLRRGKPTTHKVYGEGVAILSGGALLSLAFEHMTTA----EISSERMVWAVR
AT4G36810_G11      DLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALLSFAFEHLASATSSDVVSPVRVVRAVG
                   ******:*.*******.***:** .::*:. ** ::*.::   :     ::  ::: *:

AT1G49530_G1       EMARAVGTEGLVAGQAADLAGERMSFKNEDDELRYLELMHVHKTAVLVEAAAVVGAIMGG
AT2G18620_G2       ELAKAIGSKGLVAGQVVDLTSGGMDQN--DVGLEVLEFIHVHKTAVLLEAATVLGAIVGG
AT2G18640_G3       ELAKSIGTKGLVAGQAKDLSSEGLEQN--DVGLEDLEYIHVHKTGSLLEASAVIGAVIGG
AT2G23800_G4       ELARSIGTRGLVAGQAMDISSEGLDLN--EVGLEHLEFIHVHKTAVLLETAAVLGAIIGG
AT4G36810_G11      ELAKAIGTEGLVAGQVVDISSEGLDLN--DVGLEHLEFIHLHKTAALLEASAVLGAIVGG
                   *:*:::*:.******. *::.  :. :  :  *. ** :*:***. *:*:::*:**::**
                                        SARM
AT1G49530_G1       GSDEEIERLKSYARCVGLMFQVMDDVLDETKSSEELGKTAGKDLITGKLTYPKVMGVDNA
AT2G18620_G2       GSDEEVEKLRRFARCIGLLFQVVDDILDVTKSSEELGKTAGKDLIADKLTYPKLMGLEKS
AT2G18640_G3       GTEKEIEKVRNFARCIGLLFQVVDDILDETKSSEELGKTAGKDKVAGKLTYPKVIGVEKS
AT2G23800_G4       GSDEEIESVRKFARCIGLLFQVVDDILDETKSSEELGKTAGKDQLAGKLTYPKLIGLEKS
AT4G36810_G11      GSDDEIERLRKFARCIGLLFQVVDDILDVTKSSKELGKTAGKDLIADKLTYPKIMGLEKS
                   *::.*:* :: :***:**:***:**:** ****:********* :: ******::*::::

AT1G49530_G1       REYAKRLNREAQEHLQGFDSDKVVPLLSLADYIVKRQN*
AT2G18620_G2       KDFADKLLSDAHEQLHGFDSSRVKPLLALANYIAKRQN*
AT2G18640_G3       KEFVEKLKRDAREHLQGFDSDKVKPLIALTNFIANRNH*
AT2G23800_G4       KEFVKRLTKDARQHLQGFSSEKVAPLVALTTFIANRNK*
AT4G36810_G11      REFAEKLNREARDQLLGFDSDKVAPLLALANYIAYRQN*
                   ::::.:*  :*:::* **.*.:* **::*: :*. *:.*
```

Figure S2. Sequence alignment of Arabidopsis GGPPS isoforms. Multiple alignment was performed using *Clustal Omega* with the default parameters (http://www.ebi.ac.uk/Tools/msa/clustalo/). Red and blue arrowheads mark the position of the T-DNA in the *g11-2* and *g11-3* mutants, respectively. The methionine encoded by the $ATG_{(2)}$ codon is boxed in black and the predicted plastid-targeting peptide is boxed in green. The conserved FARM (first aspartate-rich motif) and SARM (second aspartate-rich motif) signatures that form the catalytic cavity for allyl substrate binding, are boxed in gray. The fifth residue before the FARM motif, shown to be relevant to determine the chain length of the final product, is indicated with a blue frame. True GGPPS enzymes (*i.e.* those producing C20 GGPP) have a M residue in this position, whereas the presence of smaller residues (A or S) involves a preferential production of C25 GFPP (Nagel et al., 2015; Wang et al., 2016).

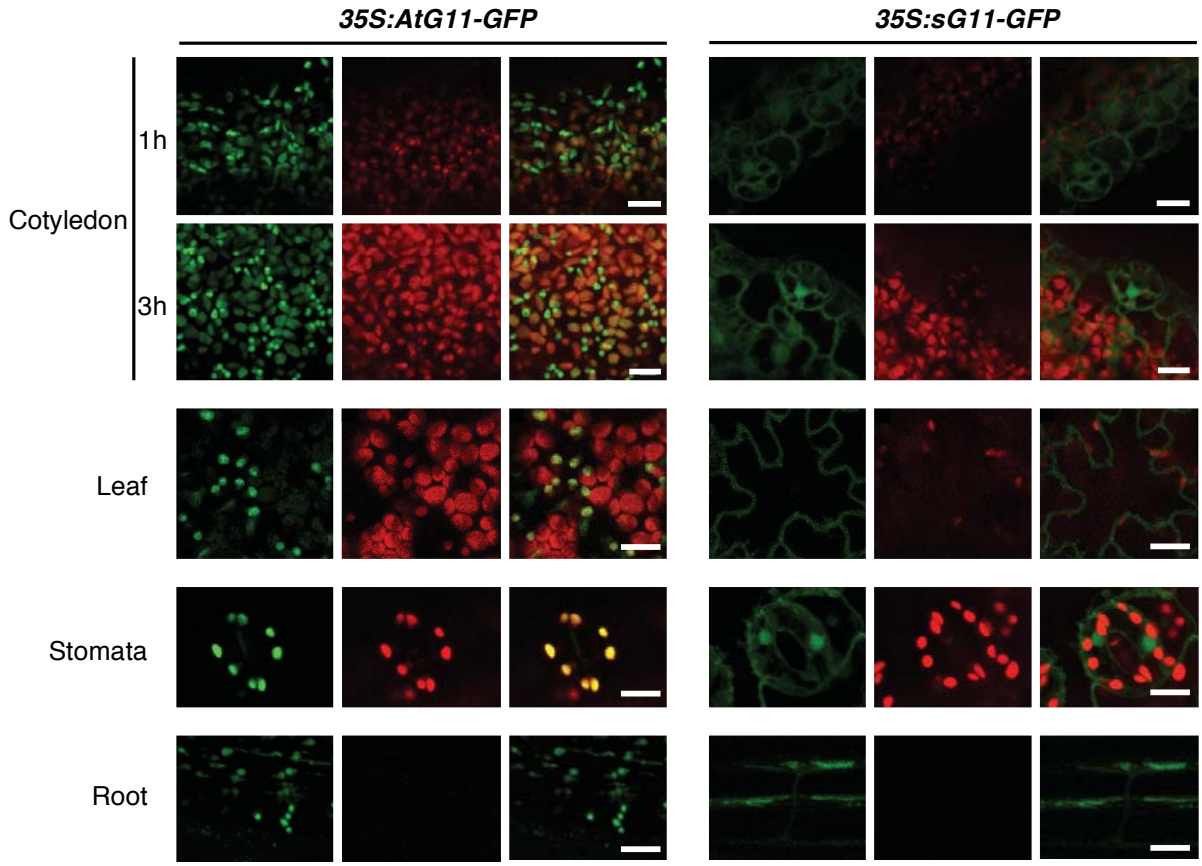


Figure S3. Subcellular localization of AtG11-GFP and sG11-GFP fusion proteins in transgenic Arabidopsis plants. Representative images from cotyledons of etiolated seedlings after 1 or 3 hours of exposure to light and from leaf epidermal cells, stomata, and roots from 10-day-old seedlings grown under long day conditions. The first column shows green fluorescence from GFP, the second shows red autofluorescence from chlorophyll, and the third shows an overlay. All confocal images were scanned using similar laser gain and offset settings. White bars represent 10 μm.

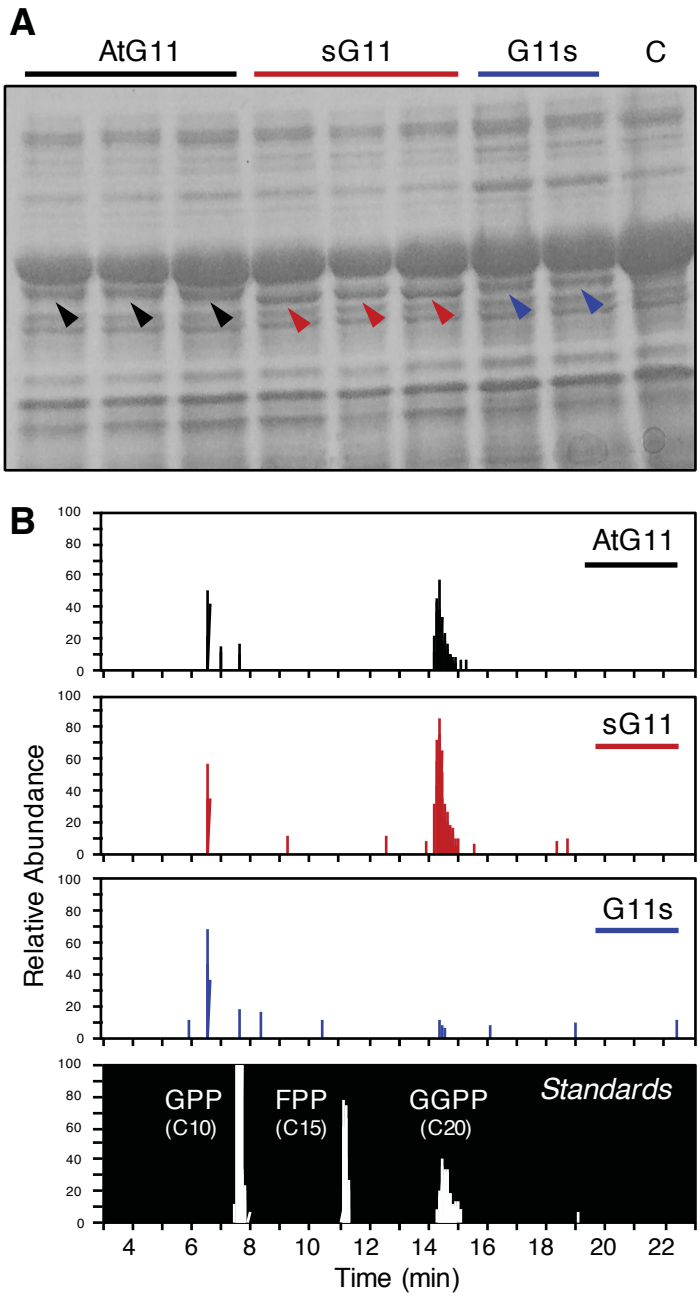


Figure S4. ***In vitro*** **activity of truncated AtG11 isoforms. (A)** SDS-PAGE of protein extracts from *E. coli* BL21 pGROE cells expressing the indicated proteins (see Figure 2A). AtG11 and G11s correspond to truncated versions lacking the predicted plastid-targeting peptide and C corresponds to empty vector control. Arrowheads mark the position of the recombinant proteins. **(B)** LC-MS chromatograms showing the results from *in vitro* enzyme activity assays using extracts like those shown in (A). IPP and DMAPP were used as substrates. Detection of prenyl diphosphates was performed using m/z 518.254 (GFPP), 449.186 (GGPP), 381.123 (FPP), and 313.061 (GPP). Retention time of available standards is also shown.

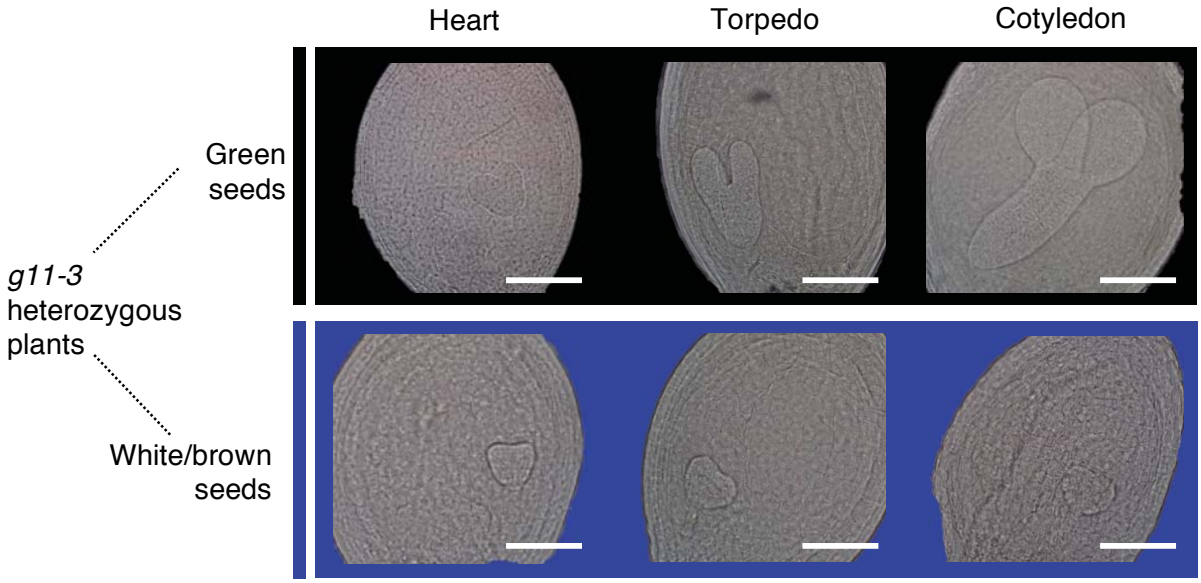


Figure S5. Embryogenesis is blocked at the heart stage in the *g11-3* mutant. Pictures show representative images of embryos in cleared seeds from heterozygous *g11-3* plants. Images in the same column correspond to seeds from the same silique that appeared either green (expected to be either azygous or heterozygous; upper row) or white/brown (expected to correspond to homozygous *g11-3* mutants; lower row, boxed in blue). Different columns correspond to siliques at different positions in the inflorescence (*i.e.* seeds at different stages of development). Embryo developmental stage in the green seed of the silique is indicated on the top. White bars, 50 µm.

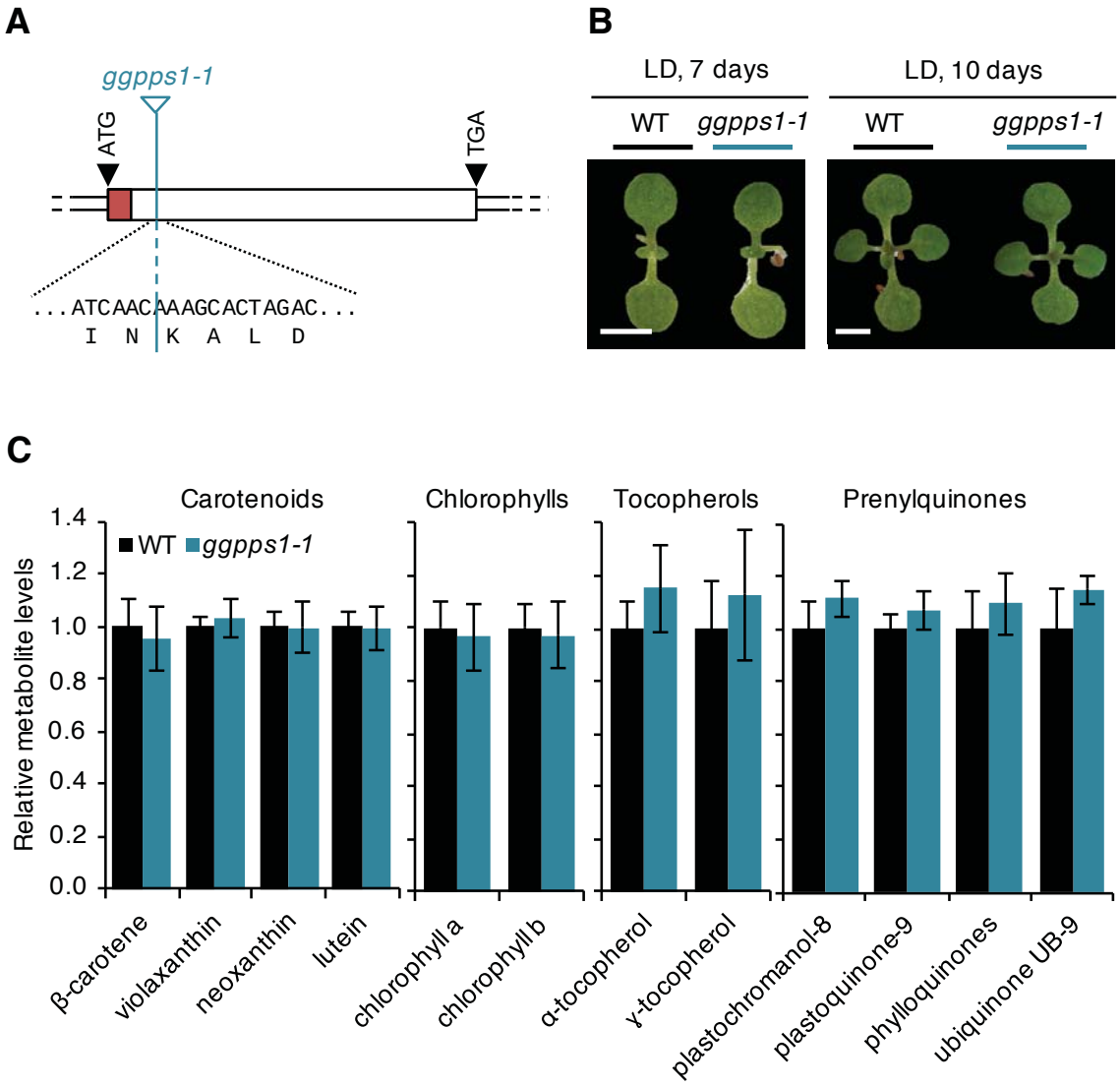


Figure S6. Developmental and metabolic phenotypes of the *ggpps1-1* mutant. (A) Schematic representation of the *GGPPS1* (At1g49530) gene according to TAIR v10 annotation. The protein coding sequence (which lacks introns) is shown as a wider box with a red section corresponding to the mitochondria-targeting peptide predicted with *TargetP* (http://www.cbs.dtu.dk/services/TargetP/). Translation start and stop codons are marked. The site of the T-DNA insertion in the *ggpps1-1* (SAIL_559_G01) mutant is also shown. The exact insertion site (shown with the encoded amino acid residues) was determined by sequencing the region after PCR amplification with primers annealing on the T-DNA and the neighbouring genomic region. **(B)** A segregating population of the SAIL_559_G01 line was analyzed by PCR to identify individuals that were azygous (WT) and homozygous (*ggpps1-1*) for the T-DNA insertion. Images show representative images of these individuals when grown together under long-day conditions (LD) for 7 and 10 days. White bars represent 2 mm. **(C)** Levels of GGPP-derived isoprenoids in WT and *ggpps1-1* seedlings. Values are shown relative to those found in WT plants and correspond to mean and SE of n=3 independent samples. No statistically significant differences (t- test, $p<0.05$) were found between WT and *ggpps1-1* samples.

Table S1. Primers used in this work.

Use	#	Name	Sequence (5'-3')
Genotyping and cloning	1	AtG11-P5F	AGAAGCTTACAAGTTGTTAAATTCG
	2	AtG11-5F	CAGATTTCAGAAATCGCCATGG
	3	AtG11-3R	ATTCCCGACAAAAGGAATCG
	4	LBb1	GCGTGGACCGCTTGCTGCAACT
	5	AtG1-LP-F	AAACTGGACCTGACCACAGC
	6	AtG1-RP-R	CCTCTGTCCCAACAGCTCTC
	7	SAIL LB3	TAGCATCTGAATTTCATAACCAATCTCGATACAC
	8	AtG11-B1-F-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTCTTCCGTTGTTACAAAAGAAG
	9	AtG11-B2-R-2	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGTTCTGTCTATAGGCAATG
	10	AtG11-B1-F-3	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGTCGTACATCATCACCAAAGC
	11	AtG11-B2-R-4	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGGACCCTAAAAGCTGATCACGAGC
5' RACE	12	AtG11-RACE-R1	TGCCACCGGCGAGAAGAGAGTAACGC
	13	AtG11-RACE-R2	TCTTGAGTGGCTCACGGAGAGGAACAGC

Table S2. List of T-DNA insertion *ggpps* mutant lines used in this work.

Gene	Gene ID	Allele	Genetic background	Original seed source ID	WT allele genotyping primers[1]	T-DNA genotyping primers[1]
AtGGPPS1	At1g49530	*ggpps1-1*	Col-0	SAIL_559_G01	5+6	6+7
AtGGPPS11	At4g36810	*g11-2*	Col-0	SALK_015098	1+3	3+4
		g11-3	Col-0	SALK_085914	2+3	3+4
		g11-5	Col-0	SALK_140601	2+3	2+4

[1]See Table S1

Table S3. Constructs and cloning details.

Use	Construct	Template[1]	Primers[2]	Sequence[3]	Plasmid backbone	Cloning method
Transgenic plants	35S:sG11-GFP	pSG11G	-	-	pCAMBIA1302	*NcoI* / Mung Bean nuclease/ T4 ligase
In vitro GGPPS activity assay	pET-AtG11	pSG11G	8+9	$AtG11_{169-1116}$	pET32-GW	Gateway
	pET-sG11	pSG11G	9+10	$AtG11_{229-1116}$	pET32-GW	Gateway
	pET-G11s	pSG11G	8+11	$AtG11_{169-1050}$	pET32-GW	Gateway

[1]Plasmid pSG11G reported in Ruiz-Sola et al. (2016a), Annex 1

[2]See Table S1

[3]Numbers refer to nucleotide positions in the protein coding sequence (positions 1-3 correspond to $ATG_{(1)}$, 2 231 to $ATG_{(2)}$, and 1114-1116 to the TGA translation stop codon)

CHAPTER II

Three Enzyme Isoforms Produce Geranylgeranyl Diphosphate for Plastidial Isoprenoid Synthesis in Different Tomato Tissues

Abstract

In plants, geranylgeranyl diphosphate (GGPP) is produced by the GGPP synthase (GGPPS) protein family and serves as a precursor for many isoprenoids, including plastid-localized carotenoids. Here we report that the tomato (*Solanum lycopersicum*) genome contains five genes encoding putative GGPPS enzymes that are differentially localized within plant cells. In particular, isoforms SlGGPPS1, 2 and 3 (or SlG1-3 in short) were found to actively produce GGPP in plastids. Different gene expression analyses revealed that each plastidial isoform is associated with a particular tissue-specific developmental process in which carotenoid biosynthesis is induced. A co-regulation was also found between plastidial GGPPS paralogs and tomato phytoene synthase (PSY) enzymes, that represent the rate-limiting step for carotenoid biosynthesis. Thus, root mycorrhization involved the activation of *SlG1* and *PSY3* genes, while *SlG2* and *PSY2* were most strongly co-regulated during photosynthetic processes. *SlG3* and *PSY1* were co-expressed in both vegetative and fruit tissues. While an additional role for SlG2 in supplying GGPP for PSY1 activity during ripening could not be discarded based on expression data, transgenic tomato plants with decreased *SlG3* transcript levels showed reduced accumulation of phytoene (the product of PSY activity) in ripe fruit, supporting the conclusion that SlG3 is the main isoform producing GGPP for carotenoid biosynthesis during ripening. Taken together, molecular and metabolic analyses indicate a functional specialization of plastidial GGPPS isoforms in tomato through a tight spatio-temporal transcriptional co-regulation with tomato PSY paralogs.

*The second Chapter of the thesis is a research article planned for publication.

Three Enzyme Isoforms Produce Geranylgeranyl Diphosphate for Plastidial Isoprenoid Synthesis in Different Tomato Tissues

M. Victoria BARJA[1], Miguel EZQUERRO[1], Elisenda FEIXES[1], Jean Luis ALCÍVAR[1], Gianfranco DIRETTO[2], Rumyana KARLOVA[3], Jules BEEKWILDER[3], Manuel RODRIGUEZ-CONCEPCION[1]

1. Program of Plant Metabolism and Metabolic Engineering, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, 08193 Barcelona, Spain.
2. Italian National Agency for New Technologies, Energy, and Sustainable Development (ENEA), Casaccia Research Centre, 00123, Rome, Italy.
3. Plant Research International, PO Box 16, 6700, AA, Wageningen, The Netherlands.

NOTE: The experiments included in this thesis Chapter were co-designed and performed by the PhD candidate under the supervision of the PhD director. The contribution of the rest of the authors to the results shown here was the following: M.E., E.F. and J.L.A provided technical support in sample collection and qPCR analysis; G.D. helped with GCN construction; R.K. and J.B provided technical advice and discussion on enzymatic activity assays.

1. Introduction

Isoprenoids are biological molecules essential in all living organisms. In particular, plants are the main source of the enormous structural and functional variety that characterizes this family of compounds (Pulido et al., 2012; Tholl, 2015). The building blocks for the biosynthesis of all isoprenoids are isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). These five-carbon (C5) universal isoprenoid units are produced in plants by the mevalonic acid (MVA) pathway in the cytosol and the methylerythritol phosphate (MEP) pathway in plastids (Figure 1, Vranová et al., 2013; Rodríguez-Concepción and Boronat, 2015). Short-chain prenyltransferases (SC-PTs) subsequently condense one or more molecules of IPP to one molecule of DMAPP giving rise to C10, C15 and C20 prenyl diphosphates known as geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), respectively. These intermediates represent the immediate precursors for downstream pathways leading to the production of the

main groups of isoprenoids (Figure 1). The SC-PT enzymes that produce each of these precursors, namely as GPP, FPP and GGPP synthases (Figure 1), contain two highly conserved domains essential for their catalytic activity referred to as FARM (first aspartate-rich motif) and SARM (second aspartate-rich motif) (Figure S1). The catalytic function of FARM ($DDx_{2-4}D$, where 'x' is any amino acid) and SARM (DDx_2D) motifs depends on their coordination with Mg^{2+} ions, required for the binding of the homoallylic substrate IPP, and one of the allylic substrates (DMAPP, GPP or FPP), allowing then the production of longer prenyl diphosphates (Marrero et al., 1992; Tarshis et al., 1994; Song and Poulter, 1994; Koyama et al., 1996; Aaron and Christianson, 2010). The length of the final product is mainly regulated by the amino acid identity of the chain-length determination (CLD) motif, located upstream to the FARM (Figure S1). In general terms, the size of the side chain of the fifth amino acid before the FARM controls the number of IPP condensations by limiting the dimensions of the elongation pocket (Ohnuma et al., 1996a, 1996b; Nagel et al., 2015; Vandermoten et al., 2009). However, other residues have been demonstrated to also regulate the product length (Tarshis et al., 1994; Wang and Ohnuma, 1999; Fernandez et al., 2000). A recent study proposed a "three-floors" model and identified residues associated with the production of C25 geranylfarnesyl diphosphate (GFPP), the precursor of rare C50 sesterpenes (Wang et al., 2016). SC-PTs are typically found in different cell compartments, consistent with the requirement of their specific prenyl diphosphate products in different subcellular locations (Figure 1). Although an important amount of information is available about SC-PTs, it is still difficult to predict their specific product, cell targeting, and biological function with absolute confidence relying solely on sequence homology criteria and hence, experimental evidence is necessary in most cases (Cunillera et al., 1996, 1997; Gaffe et al., 2000; Beck et al., 2013; Jones et al., 2013; Nagel et al., 2015; Wang et al., 2016; Zhou et al., 2017).

Carotenoids are one of the most studied groups of plant isoprenoids. They are greatly demanded as natural pigments (ranging from red to yellow) and provide benefits for human health, *e.g.* as precursors of vitamin A and other retinoids (Sandmann, 2015; Rodriguez-Concepcion et al., 2018). In plants, carotenoids have different functions. In photosynthetic tissues they contribute to light harvesting and photoprotection. They are also fundamental in growth regulation, since they are the precursors of retrograde signals and phytohormones such as abscisic acid (ABA) and strigolactones (SLs). As a secondary role, carotenoids provide distinctive colors to flowers and fruits to attract pollinators and seed dispersal animals (Nisar et al., 2015;

Yuan et al., 2015). In plants, the biosynthesis of carotenoids takes place in plastids (Ruiz-Sola and Rodríguez-Concepción, 2012; Sun et al., 2018) and they directly derive from GGPP, which is also used for the biosynthesis of gibberellins (GAs) and other essential isoprenoids involved in photosynthesis, including chlorophylls, tocopherols, and phylloquinones (Figure 1).

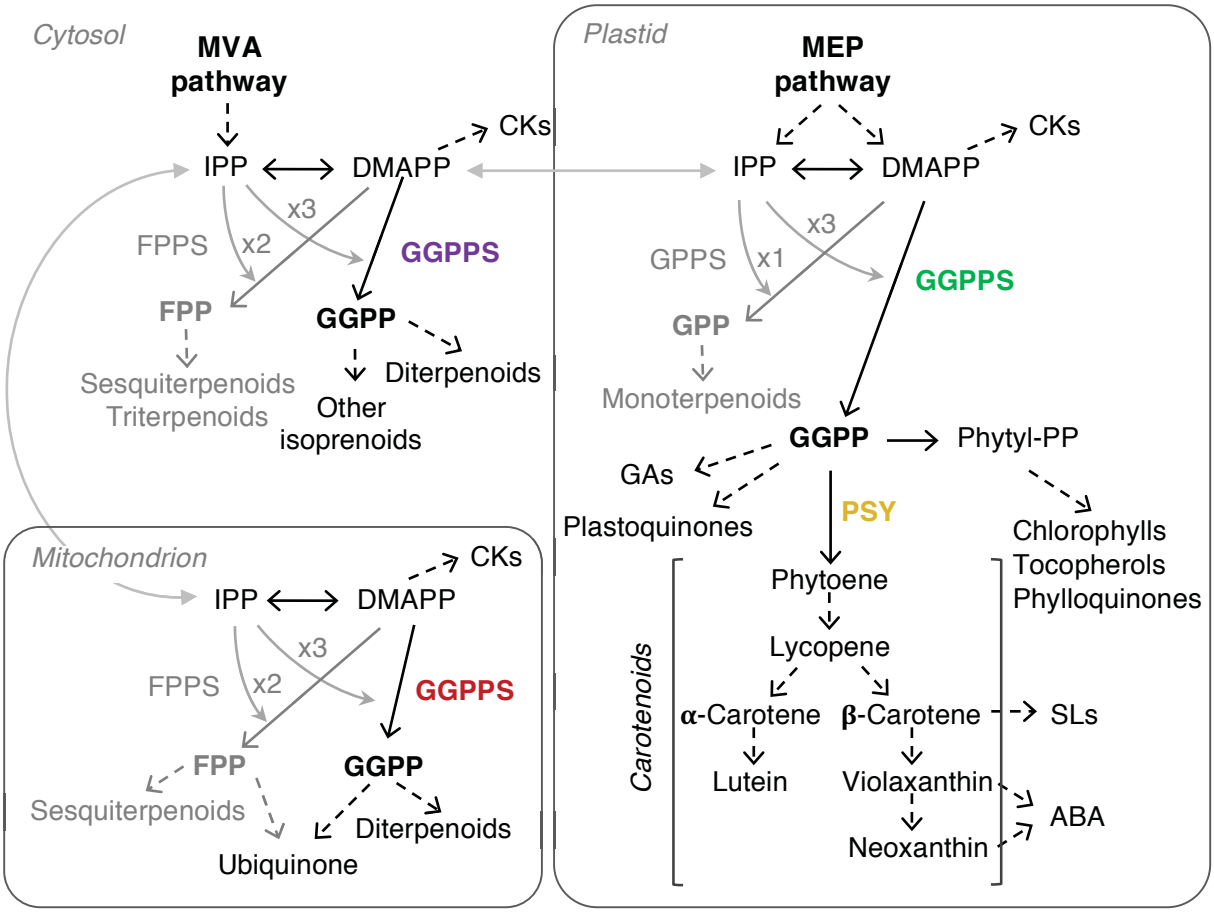


Figure 1. Compartmentalization of isoprenoid biosynthesis in plant cells. GGPP-associated pathways are shown in black and other short-chain prenyl diphosphates and their derivatives in gray. Carotenoid pathway is indicated between vertical brackets. Cytosolic GGPPS is represented in purple, mitochondrial GGPPS in red, and plastidial GGPPS in green. Phytoene synthase (PSY) is represented in yellow. Solid and dashed arrows represent single and multiple enzymatic steps, respectively. *See* Table of Contents for abbreviations.

Several GGPP synthase (GGPPS) paralogs have been retained in plants during evolution due to the acquisition of specific functions. Many studies have shown that different members of a GGPPS family provide GGPP for different metabolic branches through a network that involves gene expression co-regulation and protein-protein interactions with downstream isoprenoid biosynthetic enzymes (Beck et al., 2013; Zhang et al., 2015; Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see*

Annex 2; Zhou et al., 2017; Wang et al., 2018b). The allocation of GGPP into carotenoid biosynthesis is controlled by phytoene synthase (PSY), the first enzyme of the pathway that catalyzes the condensation of two GGPP molecules into C40 phytoene (Figure 1). It has been shown that metabolic channeling of GGPP to carotenoids is facilitated by the physical interaction of GGPPS and PSY proteins (Ruiz-Sola et al., 2016a, *see* **Annex 1**; Camagna et al., 2018; Wang et al., 2018b).

Tomato (*Solanum lycopersicum*) has become one of the best plant systems to study the biosynthesis of carotenoids and its regulation due to the variety of functions that these isoprenoids play in photosynthetic tissues (*e.g.* photoprotection) but also in roots (*e.g.* mycorrhization) and fruits (*e.g.* pigmentation). Here we report the identification and functional characterization of the plastidial GGPPS set in tomato and provide clues to understand how the supply of plastidial GGPP for the synthesis of carotenoids with different biological functions in particular tissues is regulated in this plant species.

2. Results

2.1. Identification of the tomato GGPPS protein family

Two tomato genes presumably encoding GGPPS enzymes (Solyc11g011240 and Solyc04g079960) have been previously described (Ament et al., 2006; Fraser et al., 2007; Stauder et al., 2018). To investigate the possible existence of further GGPPS-encoding genes in the tomato genome, we used previously characterized members of the *Arabidopsis thaliana* GGPPS family (Ruiz-Sola et al., 2016b, *see* **Annex 2**) as queries (Table S1). We found six putative GGPPS paralogs named as *SlGGPPS1* (Solyc11g011240), *SlGGPPS2* (Solyc04g079960), *SlGGPPS3* (Solyc02g085700), *SlGGPPS4* (Solyc02g085710), *SlGGPPS5* (Solyc02g085720) and *SlGGPPS6* (Solyc09g008920), herein referred to as *SlG1* to *SlG6*. The open reading frames of these six genes were predicted to encode proteins of 365, 363, 360, 393, 373 and 334 amino acids, respectively (Table 1).

All these putative GGPPS homologs were predicted to contain a plastid-targeting peptide except SlG5, which was predicted to localize to mitochondria according to *TargetP* and *ChloroP* algorithms (Emanuelsson et al., 2007). The identification of these proteins as putative GGPPS enzymes was initially based on sequence comparisons (Figure 2). While phylogenetic analyses showed a good separation between GGPPS candidates and other PTs, the evolutionary distinction between GGPP and GFPP

synthases was less clear. SlG1-3 were found to cluster with true Arabidopsis GGPPS isoforms targeted to plastids (GGPPS2 and GGPPS11, here referred to as AtG2 and AtG11). By contrast, SlG4 and SlG5 appeared to have followed a divergent evolutionary path and grouped separate from the cluster formed by tomato SlG1-3 and Arabidopsis GGPPS and GFPPS proteins. SlG6 was closely related to the Arabidopsis type II GPP synthase small subunit (AtSSUII) protein, which interacts with AtG11 generating a heterodimeric enzyme able to produce GPP in plastids (Wang and Dixon, 2009; Chen et al., 2015).

Table 1. List of predicted GGPPS isoforms in tomato

Enzyme	Abbreviation	Gene ID	Predicted function	Protein length (aa)	Predicted transit peptide (aa)
GGPPS1*	SlG1	Solyc11g011240	GGPP synthase	365	43
GGPPS2*	SlG2	Solyc04g079960	GGPP synthase	363	63
GGPPS3*	SlG3	Solyc02g085700	GGPP synthase	360	65
GGPPS4**	SlG4	Solyc02g085710	GGPP synthase	393	34
GGPPS5**	SlG5	Solyc02g085720	GGPP synthase	373	56
GGPPS6/ SSUII**	SlG6 / SlSSUII	Solyc09g008920	Type II small subunit	334	17

*GGPPS activity reported in this work
**Enzymatic activity not confirmed

While the first five tomato isoforms (SlG1-5) contain the two essential catalytic motifs (FARM and SARM) required for prenyltransferase activity, SlG6 lacks the SARM, again resembling AtSSUII (Figure S1). SlG1-5, but not SlG6 or AtSSUII, show a methionine (M) in the fifth position upstream to the FARM (Figure 2 and Figure S1). This specific M residue, shown to be essential for GGPP production, is not present in GFPPS enzymes (Nagel et al., 2015; Wang et al., 2016). A third difference between SlG1-5 and SlG6 is that the former contain one CxxxC domain (where 'x' is any hydrophobic residue) for protein-protein interactions (Vandermoten et al., 2009; Wang and Dixon, 2009; Zhou et al., 2017; Wang et al., 2018b), whereas two of such domains are present in SlG6 and AtSSUII proteins (Figure S1). The phylogenetic and sequence analyses together suggest the existence of five GGPPS paralogs (SlG1-5) in tomato and one type II SSU (SlG6) that might heterodimerize with one or several of the true GGPPS enzymes changing the enzyme specificity or the catalytic efficiency to control the flux of precursors to certain isoprenoid pathways.

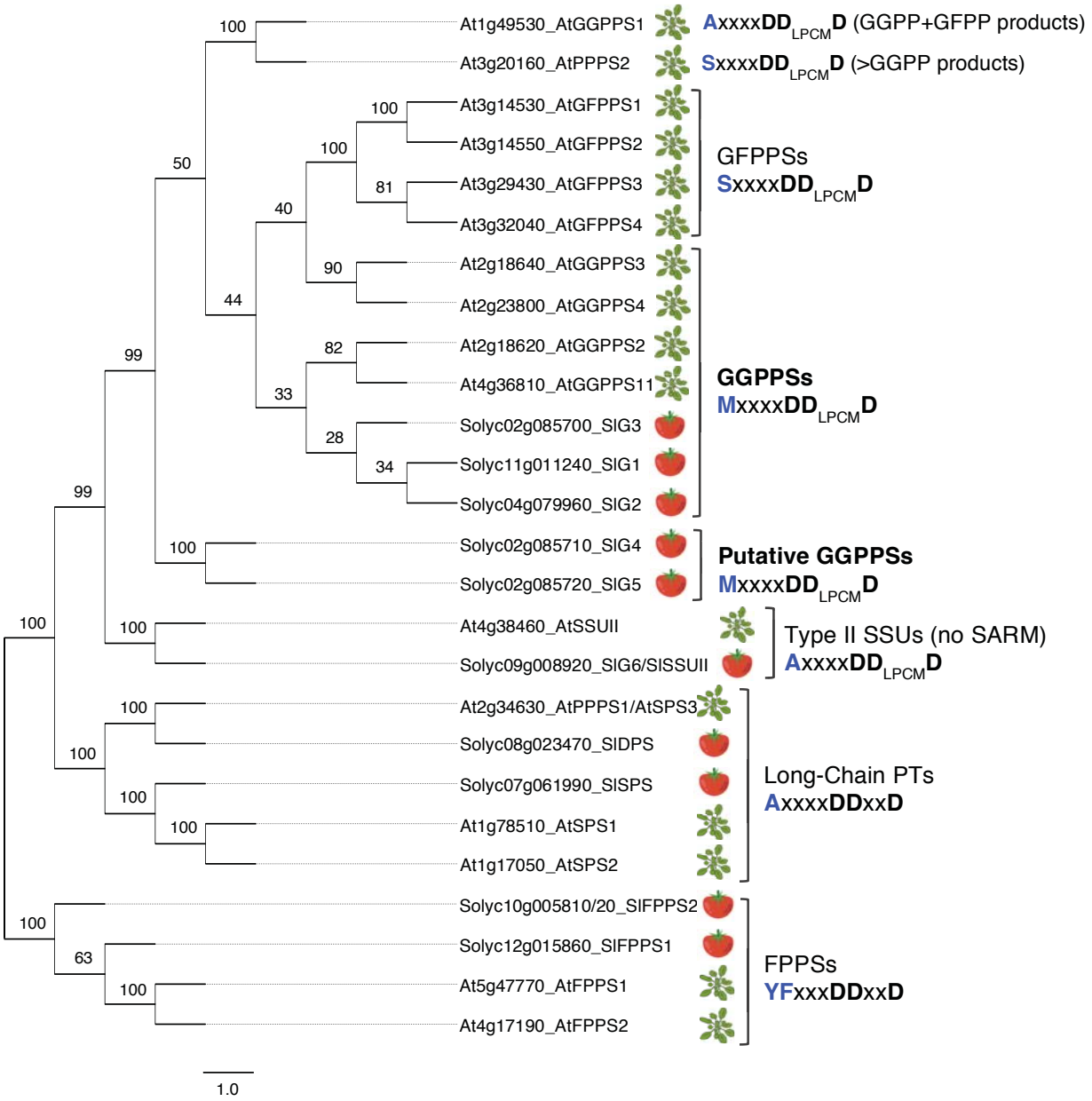


Figure 2. Phylogenetic tree of Arabidopsis (At) and tomato (Sl) prenyltransferases. Unrooted Neighbor-Joining tree constructed with Arabidopsis and tomato PT protein sequences (Table S1). The percentage of trees in which the associated sequences clustered together is indicated in each branch. The scale bar represents the mean number of substitutions per site. The chain-length determination (CLD) region (*i.e.* the FARM and the previous five residues) is indicated for each enzyme lineage. Residues determining the product length are shown in blue (positions -4 and/or -5 to the FARM) and aspartate residues are shown in black; 'x' represent any amino acid. See Table S1 for abbreviations.

2.2. The predicted tomato GGPPS isoforms localize to different cell compartments

To experimentally determine the subcellular localization of SlG1-5 enzymes, we fused the coding sequence of GFP to the 3'-end of the full-length cDNA sequences

and expressed the resulting constructs under the control of the CaMV *35S* promoter. Following agroinfiltration of *Nicotiana benthamiana* leaves, the green fluorescence emitted by each chimeric protein was analyzed by confocal microscopy at 3 dpi (days post-infiltration). The GFP fusions of SlG1-3 showed a homogeneous distribution of the GFP signal that overlapped with chlorophyll autofluorescence, confirming their location in plastids (Figure 3). SlG4-GFP mainly exhibited a cytosolic distribution of the GFP fluorescence signal, although green fluorescence was occasionally observed in the chloroplasts of some cells (Figure 3 and Figure S2). Finally, the signal corresponding to the SlG5-GFP fusion protein was detected as a punctuate pattern typical of mitochondrial proteins (Beck et al., 2013) (Figure 3).

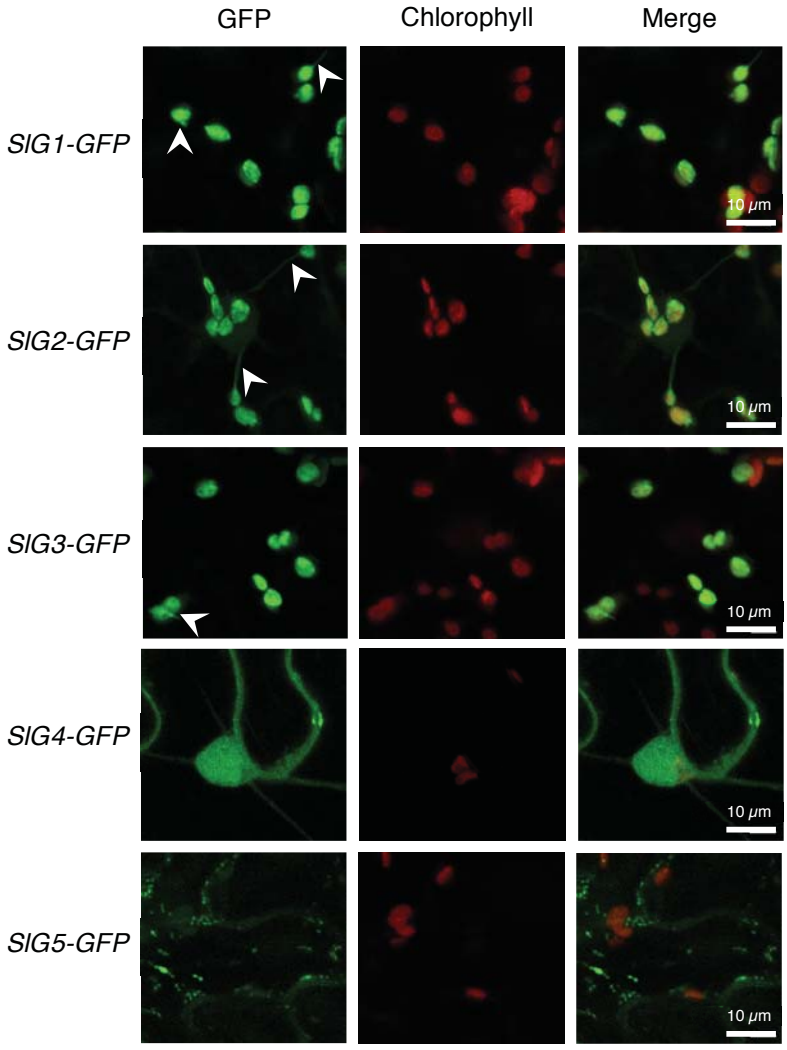


Figure 3. Subcellular localization of the tomato GGPPS candidates fused to GFP. Representative confocal microscopy images of *Nicotiana benthamiana* leaf cells transiently expressing tomato GGPPS-GFP fusion proteins. For each construct, green fluorescence from GFP (first column), red autofluorescence from chlorophyll (second column) and merged channels (third column) are shown. White arrows indicate the presence of stromules and white bars represent 10 µm.

The subcellular compartmentalization of the tested proteins is in agreement with *in silico* predictions with the only exception of SlG4-GFP, which showed a major cytosolic localization despite having a predicted plastid-targeting signal. Since the biosynthesis of carotenoids is exclusively plastidial in higher plants, a functional GGPPS enzyme involved in this pathway is expected to be mainly targeted to plastids. For this reason, we focused our research on the characterization of the plastidial isoforms SlG1-3.

2.3. SlG1, SlG2 and SlG3 are functional GGPP-producing enzymes.

The enzymatic activity of several GGPPS candidates has been assayed in *Escherichia coli* strains harboring bacterial genes for carotenoid biosynthesis but lacking GGPPS activity (Zhu et al., 1997a, 1997b; Wang and Dixon, 2009; Beck et al., 2013; Ruiz-Sola et al., 2016b, *see* **Annex 2**; Zhou et al., 2017). When transforming such *E. coli* strains with constructs encoding an active GGPPS, the biosynthesis of carotenoids takes place and the bacterial cells become colored. This method was used to conclude that SlG1 and SlG2 are GGPPS enzymes (Ament et al., 2006). However, recent studies based on liquid-chromatography coupled to mass-spectrometry (LC-MS) or thin-layer chromatography (TLC) demonstrated that this genetic complementation strategy is not enough to ascertain the enzymatic activity (*i.e.* the identity of the product) of a particular SC-PT enzyme (Nagel et al., 2015; Wang et al., 2016). To investigate whether SlG1, SlG2 and SlG3 are functional GGPPS enzymes (*i.e.* that they actually produce GGPP as major product), we carried out *in vitro* activity assays followed by the analysis of the reaction products by LC-MS. The three tomato isoforms were separately cloned into pET32-GW without their predicted plastid-targeting sequences and fused to a N-terminal 6x-histidine to facilitate purification (Figure S3A and Table S5). As positive and negative controls of the activity assays, we used constructs harboring similar versions of AtG11 and the inactive form G11s (Ruiz-Sola et al., 2016b, *see* **Annex 2**). Whole-cell protein extracts from *E. coli Rosetta* cells transformed with individual constructs and induced to produce the corresponding recombinant proteins, were directly used for activity assays in the presence of IPP and DMAPP. LC-MS analysis of the reaction products showed that SlG1, SlG2, SlG3 and AtG11 (but no G11s) produced GGPP and no other detectable prenyl diphosphate such as GPP, FPP or GFPP (Figure 4A). These results therefore confirm that the three tomato proteins are true GGPPS enzymes.

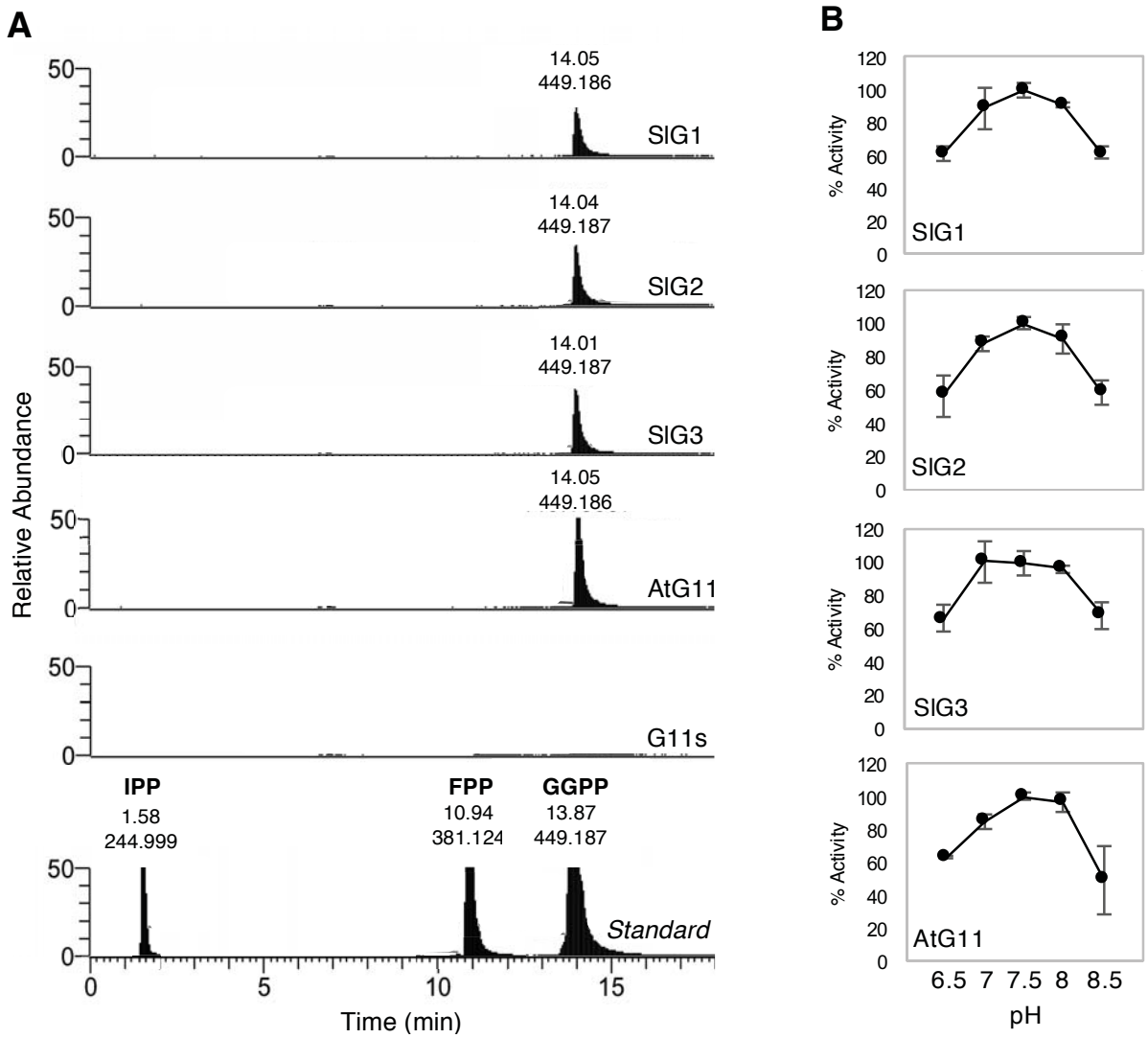


Figure 4. ***In vitro*** **GGPPS activity assays. (A)** LC-MS chromatograms. Bacterial extracts from *E. coli Rosetta* cells overproducing the indicated recombinant proteins (lacking the predicted plastid-targeting peptide and fused to an N-terminal 6x-His tag) were incubated with IPP and DMAPP and the products were analyzed by LC-MS. AtG11s- and AtG11-producing cell extracts were used as negative and positive controls, respectively. Three independent experiments were performed. Prenyl diphosphates were detected using mass-to-charge (m/z) ratios: 244.998 (IPP), 313.061 (GPP), 381.123 (FPP), 449.186 (GGPP) and 518.254 (GFPP). Retention time of available standards is also shown. **(B)** Optimal pH determination for the activity of each GGPPS assayed. Purified recombinant proteins (6x-His-GGPPS) were incubated with IPP and DMAPP under different pH conditions. Activity values are represented as the percentage of activity relative to the maximum activity obtained. Data correspond to the mean and standard deviation (SD) of three independent replicates (n=3).

In order to analyze the kinetic properties of the identified tomato GGPPS isoforms, we then purified the recombinant proteins taking advantage of their 6x-histidine tag, using again AtG11 as positive control (Figure S3). Enzymatic assays with the purified recombinant enzymes carried out as described (Barja et al., 2019, *see* **Annex 3**)

showed a similar optimal pH (around 7.5) for all the GGPPS proteins tested (Figure 4B). This result agrees with the conclusion that plastidial GGPPS activity mainly occurs in the stroma of plastids, where the pH is around 7.8-8, unlike the usual acidic pH of the thylakoid lumen (Höhner et al., 2016). Consistently, proteomic approaches (Joyard et al., 2009) and confocal microscopy (Beck et al., 2013; Perello et al., 2016) found AtG11 in the chloroplast stroma. The homogeneous pattern of GFP fluorescence found for SlG1-GFP, SlG2-GFP and SlG3-GFP within chloroplasts and in stromules (Figure 3) further supports the conclusion that all these tomato GGPPS isoforms are indeed stromal proteins.

For a deeper characterization of these enzymes, their kinetic parameters (Km and Vmax) were then calculated for their C5 isoprenyl diphosphate precursors (IPP and DMAPP) (Table 2). It is noteworthy that the apparent affinity for the substrates (Km) and Vmax values are very similar among the three tomato enzymes. They are also similar to those obtained for AtG11 here (Table 2) and elsewhere (Wang and Dixon, 2009; Camagna et al., 2018). All this information together indicates that tomato SlG1-3 and Arabidopsis AtG11 are very similar enzymes, in agreement with their close phylogenetic clustering (Figure 2).

Table 2. Kinetic parameters of plastidial GGPPS enzymes. Values correspond to the mean ± SD of three independent experimental replicates (n=3).

	DMAPP (+100 µM IPP)		**IPP** (+100 µM DMAPP)	
	Km (µM)	Vmax (nmol•min^{-1}•mg^{-1})	Km (µM)	Vmax (nmol•min^{-1}•mg^{-1})
SlG1	31.82 ± 2.92	47.47 ± 1.40	74.18 ± 7.55	59.87 ± 2.73
SlG2	49.55 ± 5.31	38.87 ± 1.53	79.75 ± 8.33	36.73 ± 1.73
SlG3	45.75 ± 6.81	26.13 ± 1.40	45.92 ± 4.86	29.13 ± 1.13
AtG11	32.86 ± 4.86	21.53 ± 1.07	38.49 ± 4.94	24.13 ± 1.07

2.4. Tissue-specific gene co-expression networks suggest functional specialization of tomato GGPPS enzymes.

It has been demonstrated that genes belonging to the same molecular process or metabolic pathway are often highly co-expressed across different spatio-temporal developmental stages, forming strong functional co-expression modules (Gachon et al., 2005; Wei et al., 2006; Heyndrickx and Vandepoele, 2012; Vranová et al., 2012;

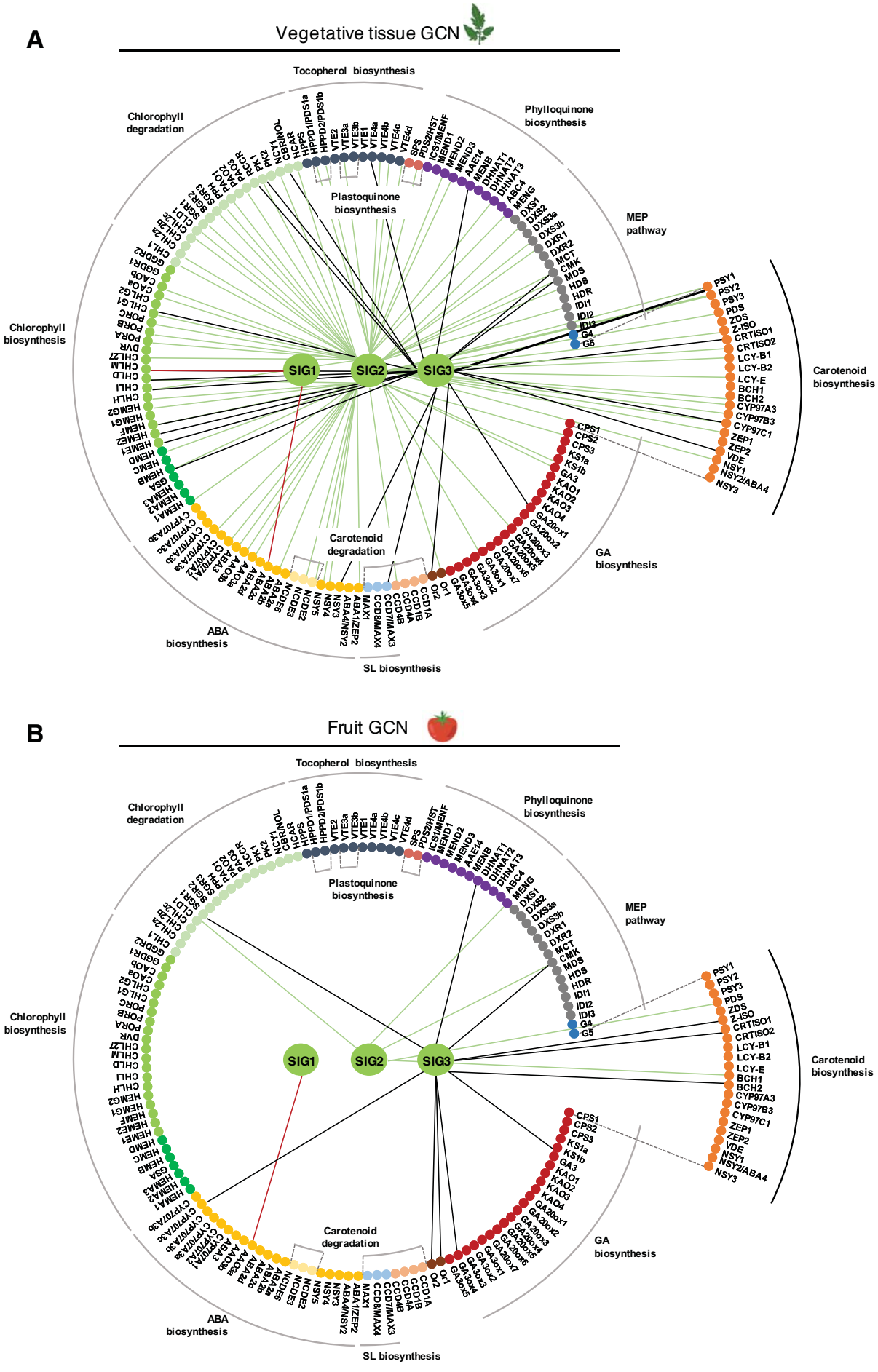
Ruiz-Sola et al., 2016a, *see* **Annex 1**). Gene co-expression network (GCN) analyses are a powerful tool to infer biological functions to non-characterized genes by applying the simple concept of "guilt-by-association", which means that a gene significantly associated with a molecular pathway at transcriptional level may share similar functions (Oliver, 2000). To gain further insight into the physiological function of SlG1, SlG2 and SlG3 we performed targeted GCN analyses with plastidial isoprenoid biosynthetic pathways in different tomato tissues. We first searched the tomato genome for genes potentially involved in plastidial GGPP-related isoprenoid pathways based on sequence homology. By using publicly available databases for plant comparative genomics (*PLAZA 4.0*, *Phytozome*), we gathered tomato genes putatively encoding the plastidial enzymes of the MEP pathway and downstream isoprenoid pathways of carotenoids, chlorophylls, tocopherols, phylloquinone, plastoquinone, gibberellins (GAs), strigolactones (SLs) and abscisic acid (ABA) (Table S2). Then, we retrieved their expression dataset in different tomato tissues from *TomExpress*, a public tomato RNA-seq database that provides a representative transcriptional landscape of vegetative and reproductive developmental processes under several environmental conditions (Zouine et al., 2017). The expression pattern of *SlG1, SlG2* and *SlG3* in particular tissues was then correlated with that of the selected isoprenoid-related genes. It was not possible to obtain GCN data for tomato roots since only two expression values are available for each gene in the *TomExpress* database. GCN analyses of vegetative (*i.e.* photosynthetic) and fruit tissue are shown in Figure 5, where edges represent a positive significant co-expression with upstream and downstream pathways of GGPPS enzymatic step (significant co-expression relationships are listed in Table S3).

The GCN analysis of vegetative tissue showed that *SlG1* is hardly co-expressed with any of the query isoprenoid biosynthetic genes, possibly indicating the absence of a primary role of the SlG1 isoform in the production of GGPP for plastidial isoprenoid biosynthesis in photosynthetic tissues. By contrast, *SlG2* and *SlG3* showed much more connectivity, being *SlG2* the paralog with the highest level of co-expression (Figure 5a). *SlG2* appeared to be significantly co-expressed with many genes from almost all selected isoprenoid pathways, suggesting a central role of this isoform in chloroplasts, where it might be producing GGPP for the synthesis of most photosynthesis-related isoprenoids. In particular, *SlG2* was co-expressed with most genes of the carotenoid pathway, showing a strong correlation with *PSY2* (Figure 5a, Table S3), the gene encoding the PSY isoform with a major role in carotenoid biosynthesis in photosynthetic tissues (Bartley and Scolnik, 1993; Giorio et al., 2008). Among the hormone pathways, *SlG2* was only highly co-expressed with ABA

biosynthetic genes. Few significant connections were found between *SlG2* and GA biosynthetic genes whereas no connection was observed with the SL pathway. In these photosynthetic tissues, the *SlG3* gene was mainly co-expressed with carotenoid and chlorophyll biosynthetic pathways but at lower level than *SlG2*. Interestingly, *SlG3* appeared to be more connected to *PSY1* than to *PSY2* (Figure 5a, Table S3), the gene that encodes the fruit-specific PSY isoform (Bartley et al., 1992; Fraser et al., 1999; Gady et al., 2012).

When analyzing the fruit ripening GCN we observed much less connections (Figure 5b). The scarcity of nodes observed for *SlG1* and *SlG2* suggests a minor role for these GGPP-producing isoforms in the fruit. A higher degree of connectivity was detected for *SlG3*, even though it was much lower than that found in vegetative tissues. The co-expression of *SlG3* with *PSY1* observed in the GCN analysis of photosynthetic tissues was not detected in fruit. Interestingly, the *SlG3* gene was co-expressed with the two tomato genes encoding ORANGE (OR) proteins (*SlOR1* and *SlOR2*) during fruit ripening. OR proteins are chaperones that promote PSY stability and chromoplast differentiation in carotenoid-producing tissues (Zhou et al., 2015; Park et al., 2016; Chayut et al., 2017; D'Andrea et al., 2018; Welsch et al., 2018). Taken together, the GCN analyses performed in this study suggest that SlG1 might not have a relevant role for plastidial GGPP synthesis in vegetative tissues or fruits, SlG2 might be the main isozyme producing plastidial GGPP for photosynthesis-related isoprenoids, and SlG3 might have a more prominent role during tomato fruit development.

Figure 5. Gene co-expression networks (GCNs) of tomato genes encoding plastidial GGPPS enzymes in vegetative (A) and fruit (B) tissues. Genes are represented as nodes and significant positive co-expression relationships (ρ>0.55) as edges. Tomato genes encoding plastid-targeted GGPPS isoforms are depicted as central green nodes and are referred here to as *SlG1*, *SlG2* and *SlG3*. Surrounding smaller nodes represent the genes encoding enzymes from several isoprenoid pathways located upstream and downstream of GGPPS step in plastids. Red, green and black edges indicate the genes positively co-expressed with *SlG1, SlG2* and *SlG3* genes, respectively. Input gene abbreviations are listed in **Table S2** and positive co-relation ρ values are shown in **Table S3.** ►


A
Vegetative tissue GCN
Tocopherol biosynthesis
Phylloquinone biosynthesis
MEP pathway
Carotenoid biosynthesis
GA biosynthesis
SL biosynthesis
ABA biosynthesis
Carotenoid degradation
Plastoquinone biosynthesis
Chlorophyll biosynthesis
Chlorophyll degradation
SIG1
SIG2
SIG3
B
Fruit GCN
Tocopherol biosynthesis
Phylloquinone biosynthesis
MEP pathway
Carotenoid biosynthesis
GA biosynthesis
SL biosynthesis
ABA biosynthesis
Carotenoid degradation
Plastoquinone biosynthesis
Chlorophyll biosynthesis
Chlorophyll degradation
SIG1
SIG2
SIG3

2.5. Gene expression analyses confirm an isoform-specific association with particular developmental processes involving an activation of carotenoid biosynthesis.

Carotenoids in tomato contribute to mycorrhizal associations, photoprotection, and fruit pigmentation and, hence, the levels of these GGPP-derived plastidial isoprenoid metabolites increase during root mycorrhization, seedling deetiolation, and fruit ripening. In agreement with the rate-determining role of PSY for carotenoid synthesis (Fraser et al., 2002), the expression of PSY-encoding genes also increase during such developmental processes. However, not all three genes encoding tomato PSY isoforms are up-regulated in every tissue where carotenoid biosynthesis is boosted, but show a clear tissue-specific expression pattern. *PSY1* is mainly expressed in tomato fruit and its expression is boosted during ripening to produce carotenoids involved in the pigmentation of this organ (Bartley et al., 1992; Fray and Grierson, 1993; Giorio et al., 2008; Kachanovsky et al., 2012). *PSY2* is expressed in all tissues, including fruit, but transcript levels are much higher than those of *PSY1* in photosynthetic tissues, where carotenoids are involved in light harvesting and photoprotection (Bartley and Scolnik, 1993; Giorio et al., 2008). Lastly, *PSY3* is mainly expressed in roots and it is the only PSY-encoding gene induced during mycorrhization (Walter et al., 2015; Stauder et al., 2018). We experimentally confirmed these results by real-time quantitative PCR (qPCR) analysis using our own tomato samples (Figure 6). Furthermore, we found that the *PSY2* gene was more upregulated than *PSY1* during tomato seedling deetiolation (Figure 6B), consistent with the proposed major role of PSY2 isoform in the production of carotenoids for photoprotection of photosynthetic tissues. We next used the same samples to test the expression pattern of the identified *SlG1-3* genes and compare it with that of *PSY1-3* genes. According to public databases, *SlG1* is mostly expressed in roots and leaves while *SlG2* and *SlG3* are mainly expressed in leaves and fruits (Figure S4). During root mycorrhization, where the synthesis of carotenoids is induced to produce SLs and apocarotenoid molecules essential for the establishment of root symbiosis (Fester et al., 2002, 2005; Baslam et al., 2013; Stauder et al., 2018), only *SlG1* was upregulated, similarly to that observed for *PSY3* (Figure 6A). During seedling de-etiolation, when carotenoid biosynthesis is boosted to protect the emerging photosynthetic apparatus against the excess of light (Llorente et al., 2017; Pankratov et al., 2016; Rodríguez-Villalón et al., 2009), *SlG2* and *PSY2* were strongly upregulated. *SlG3* and *PSY1* were also induced with a similar profile during this process but at a much lower level (Figure 6B). During tomato fruit ripening, when a massive accumulation of

carotenoids changes the fruit color from green to orange and finally red (Klee and Giovannoni, 2011; Yuan et al., 2015), the peak of *PSY1* expression was observed at the orange stage, when carotenoids are produced more actively (Figure 6C). In the case of GGPPS-encoding genes, *SlG2* was most strongly induced at the breaker stage (earlier than that observed for *PSY1*) whereas *SlG3* peaked in orange fruits with a weaker induction (Figure 6C). Together, the qPCR-based transcriptional analyses are highly consistent with the GCN results, revealing a tissue/process-specific role for each individual GGPPS isoform and a strong correlation with particular PSY isoforms. Thus, SlG1 and PSY3 likely work together in the root, and the tandem formed by SlG2 and PSY2 might have a major function for photoprotection. More data, however, were required to conclude whether SlG3 and PSY1 might also be a functional module for carotenoid biosynthesis during fruit ripening.

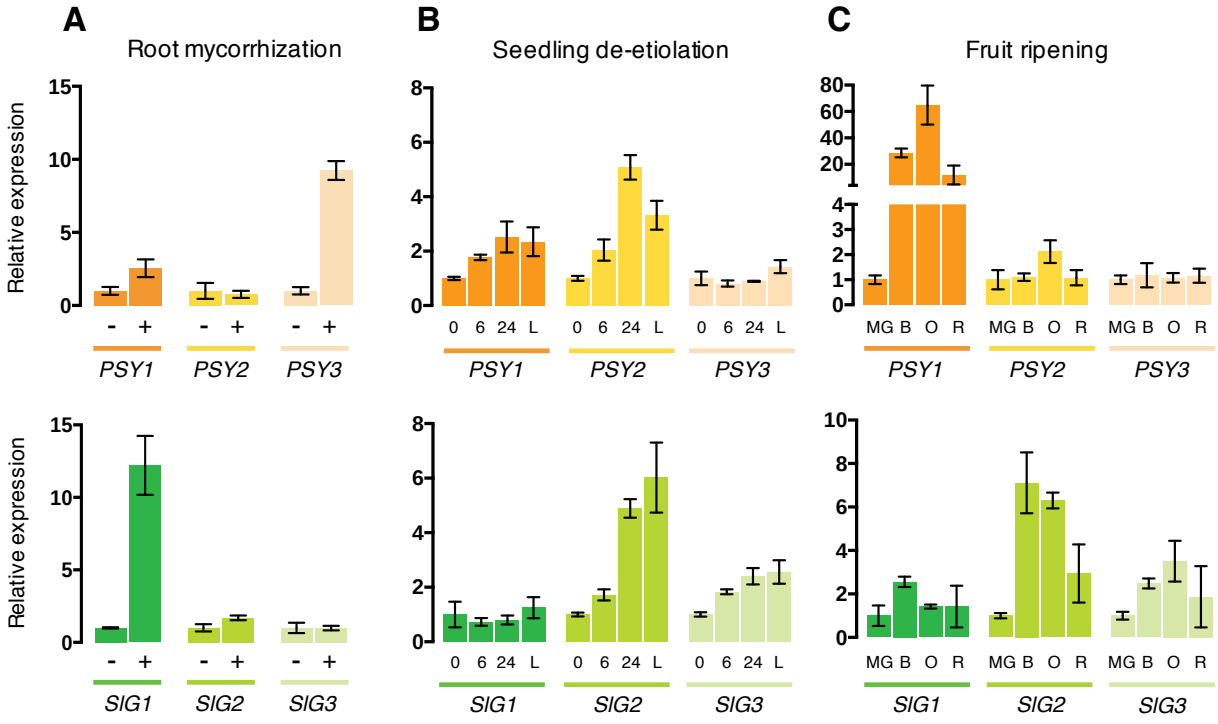


Figure 6. Expression profiles of genes encoding tomato PSY and GGPPS paralogs during different carotenoid biosynthetic processes. (A) Transcript levels in non-mycorrhized (-) and mycorrhized roots (+) at six weeks post-inoculation. **(B)** Transcript levels in 7-day-old etiolated (dark-grown) seedlings at 0, 6 and 24 h after exposure to light and in controls continuously grown in the light (L). **(C)** Transcript levels at different fruit ripening stages: MG, mature green; B, breaker; O, orange; and R, red ripe. Levels were normalized using the tomato genes *EXP* (A, B) or *ACT4* (C) and are shown relative to those in the condition with lowest carotenoid levels, *i.e.* untreated (A), non-light-exposed (B) or MG (C). Expression values represent the mean ± SD of three independent biological replicates (n=3).

2.6. A reduction of *SlG3* transcript levels decreases phytoene accumulation in red tomato fruits.

In order to determine the implication of the two tomato ripening-induced *GGPPS* genes in fruit carotenoid biosynthesis, we next generated tomato (cv. MicroTom) stable transgenic lines harboring constructs to constitutively increase the levels of SlG2 or SlG3. Control transgenic lines containing a plastid-targeted version of GFP (pGFP) were also generated. Two independent T1 (hemizygous) lines of each construct were selected for an initial characterization. Primers specific for *SlG2* and *SlG3* were used for qPCR analysis of both endogenous and transgene expression in red fruits collected from transgenic and untransformed (WT) lines. As expected, transcripts encoding SlG2 were more abundant in *35S:SlG2* fruits compared to those from WT, *35S:pGFP* and *35S:SlG3* plants (Figure 7A). Unexpectedly, however, we found that both *35S:SlG2* and *35S:SlG3* lines showed significantly reduced levels of *SlG3* transcripts in ripe fruits compared to WT and *35S:pGFP* controls (Figure 7A).

Analyses of the accumulation of carotenoids in red ripe fruits of T1 transgenic and WT plants showed that the lower accumulation of *SlG3* transcripts correlated with reduced phytoene levels in tomatoes from *35S:SlG2* and *35S:SlG3* lines (Figure 7B). Phytoene is directly produced from GGPP by PSY (Figure 1) and, therefore, its levels can better reflect GGPP availability compared to downstream carotenoids. Indeed, lycopene, beta-carotene and lutein hardly changed in all transgenic fruits compared to the WT (Figure 8). Characterization of carotenoid levels in fruits from T2 homozygous lines confirmed these results (Figure S5). A significant reduction of phytoene levels was observed in orange and red fruits from both *35S:SlG2* and *35S:SlG3* lines compared to WT and *35S:pGFP* controls, whereas no significant changes were observed for downstream carotenoids. Quantification of other GGPP-derived plastidial isoprenoids such as chlorophylls, tocopherols and phylloquinone also showed no differences between fruits from the different lines at any of the developmental stages tested (Figure S6). We therefore concluded that downregulation of SlG3 levels likely results in a decreased production of GGPP in ripening fruit, leading to lower levels of phytoene. However, the reduction is not strong enough to impact the levels of downstream carotenoids or other GGPP-derived isoprenoids.

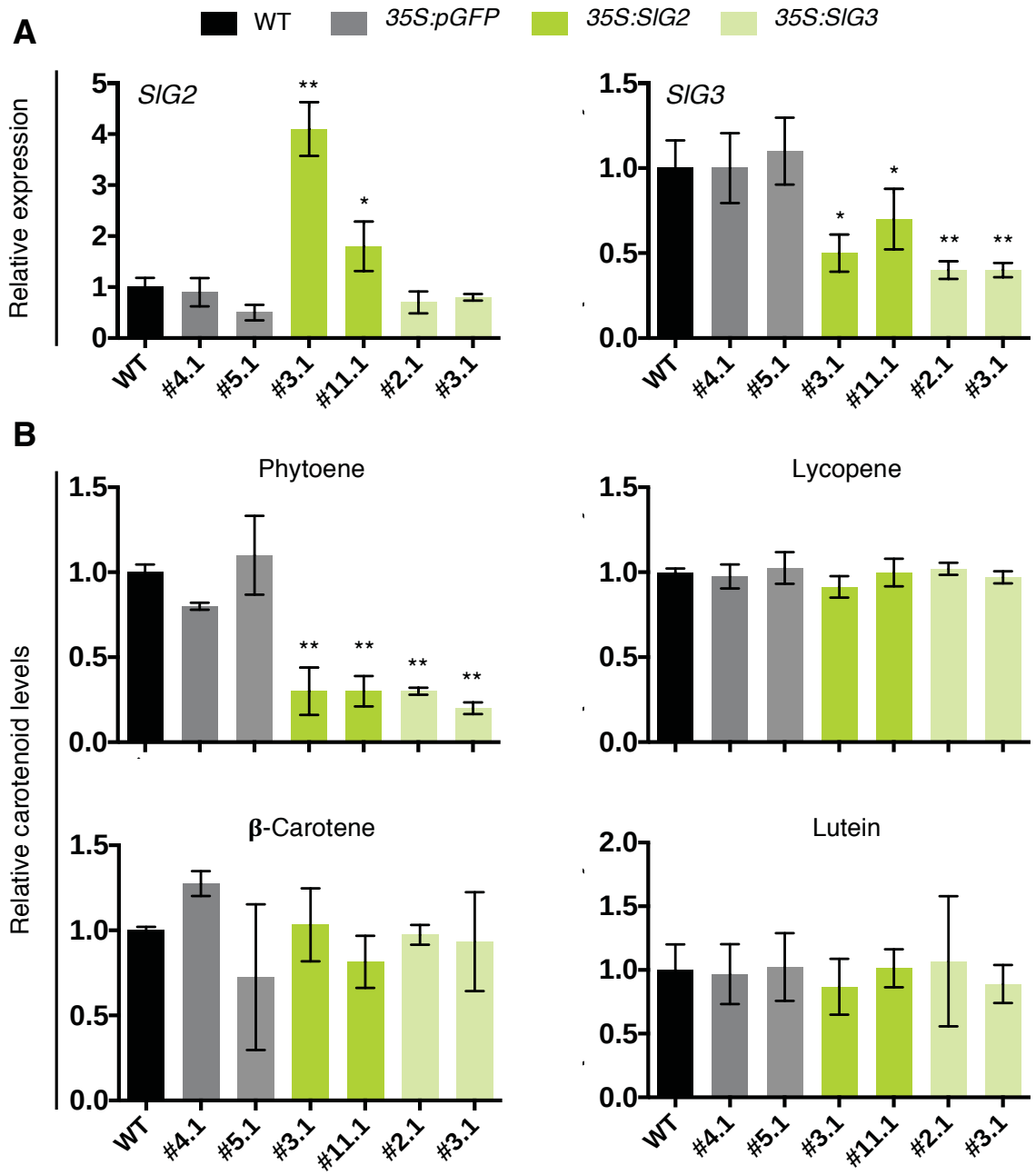


Figure 7. Transcript and carotenoid levels in red fruits of T1 transgenic lines. Ripe fruits from untransformed (WT) and T1 transgenic tomato lines carrying the indicated constructs were collected and used for **(A)** qPCR analysis of transcript levels (normalized to *ACT4*) and **(B)** HPLC analysis of the levels of the major carotenoids in ripe fruits. Values represent the mean ± SD of at least three independent biological replicates (n≥3) relative to WT levels. Asterisks indicate significant differences among means relative to WT samples (one-way ANOVA with Dunnett's multiple comparisons test, *p<0.05, **p<0.01).

3. Discussion

In plants, GGPP synthesis and allocation to the different downstream pathways is controlled by GGPPS enzyme families, whose members are regulated at transcriptional and post-transcriptional levels (Beck et al., 2013; Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see* **Annex 2**; Zhou et al., 2017; Stauder et al., 2018; Wang et al., 2018b). Several efforts have been done since the late 90's to characterize the GGPPS protein family in *A. thaliana*, establishing the fundamental basis for our knowledge of the regulation of GGPP biosynthesis in plants (Zhu et al., 1997a, 1997b; Okada et al., 2000; Beck et al., 2013; Nagel et al., 2015; Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see* **Annex 2**; Wang et al., 2016). We are also starting to understand how GGPP-related metabolites for human interest are produced in crops such as rice and other plants (Wang and Dixon, 2009; Zhang et al., 2015; Zhou et al., 2017; Wang et al., 2018a, 2018b). Strikingly, little information was available on the tomato GGPPS family despite this species is a well-established model plant. In particular, tomato fruit ripening represents a model system for the study of climacteric fruit development, which is a process associated with a boost in the production of GGPP-derived carotenoids. Here we have identified the tomato GGPPS family and characterized three plastidial GGPPS isozymes that are specifically associated with fruit ripening and other particular developmental processes causing an increased production of carotenoids, such as root mycorrhization and seedling deetiolation (Figure 8).

From the six tomato GGPPS homologs initially identified *in silico*, SlG6 was evolutionary related to Arabidopsis SSUII, also named AtSSUII or GGPPS12 (Figure 2 and Figure S1). Due to the absence of the second catalytic domain (SARM), these type of enzymes are only enzymatically active when interacting with a GGPPS enzyme, either changing its product specificity from GGPP to GPP (Wang and Dixon, 2009) or enhancing the GGPP enzymatic production in particular processes (Zhou et al., 2017; Wang et al., 2018b). The interaction between SSUII and GGPPS monomers is determined by the presence of two CxxxC motifs in the SSUII amino acidic sequence. Since all these characteristics were actually detected in SlG6 (Figure S1), we renamed this protein as SlSSUII (Solyc09g008920). Nevertheless, subcellular localization, tissue-specific gene expression and protein-protein interaction assays between SlG6 and putative GGPPS partners will be necessary to experimentally confirm whether it acts as a heterodimeric SSUII *in vivo* and then, activity assays would be required to determine the product specificity of the heterodimer formed. Heterodimer formation actually represent an important layer of post-translational regulation to tightly

control GGPP flux towards the required metabolic branches in several species, including Arabidopsis, rice, and pepper (Wang and Dixon, 2009; Zhou et al., 2017; Wang et al., 2018b), so it would not be surprising that tomato also shares this regulatory mechanism.

At the protein sequence level, the rest of the GGPPS paralogs found (SlG1-5) showed the seven domains highly conserved in SC-PTs (Koike-Takeshita et al., 1995) and all other features known to be required for GGPPS activity (Figure S1). Moreover, they were differentially located within plant cells (*i.e.* in plastids, cytosol and mitochondria). This is consistent with the need of GGPPS isozymes in several cell compartments to feed the different organelle-specific isoprenoid pathways (Bick and Lange, 2003; Beck et al., 2013; Ruiz-Sola et al., 2016b, *see* **Annex 2**; Zhou et al., 2017; Wang et al., 2018b). We confirmed that three of the five *GGPPS* homolog genes identified (*SlG1-3*) encode functional GGPPS enzymes localized in plastids when fused to GFP (Figure 3, Figure 4A). SlG5-GFP appeared to be targeted to mitochondria, whereas the GFP-fused SlG4 protein, despite showing a major preference for cytosolic location, was also observed in the chloroplasts of some *N. benthamiana* cells (Figure S2). Dual subcellular localization of many isoprenoid biosynthetic enzymes has been widely described in *Arabidopsis* (Cunillera et al., 1997; Phillips et al., 2008; Sapir-Mir et al., 2008). *AtG11*, the only GGPPS-encoding gene found to be essential in Arabidopsis, encodes two functional isoforms: a long one with an N-terminal domain enabling plastidial targeting (AtG11) and a shorter one (sG11) without such domain that localizes in the cytosol (Ruiz-Sola et al., 2016b, *see* **Annex 2**). In many cases, the production of two differentially-targeted proteins from the same gene is due to a weak transcription initiation site on the corresponding gene, giving rise to some transcripts that lack the first ATG codon of the coding sequence. Thus, translation starts from a second in-frame ATG codon producing a shorter protein lacking the N-terminal region (which usually corresponds to the organelle-specific transit peptide). Although a similar transcriptional regulation could be happening for *SlG4*, the very low frequency of plastidial localization of the constitutively overexpressed SlG4-GFP protein suggests a negligible contribution of the SlG4 enzyme to the biosynthesis of carotenoids. Further investigation would be required to ascertain whether the endogenous *SlG4* gene can actually produce transcripts of different size encoding plastidial and cytosolic versions of the enzyme.

The presence of at least three GGPPS isoforms in tomato plastids may be caused by the acquisition of specific roles during evolution either through a differential spatio-temporal gene expression, triggered by developmental and environmental signals, or through specific interactions with enzymes from distinct metabolic pathways.

Mining of public tomato gene expression databases (*Tomato eFP Browser* and *TomExpress*), GCN analyses and qPCR assays (Figure 5, Figure 6) led us to conclude that SlG1 likely has an important role in roots together with PSY3. This conclusion is supported by a recent study showing that the expression of *PSY3* and *SlG1* coordinately responds to tomato root mycorrhization and phosphate starvation (Stauder et al., 2018). The SlG1-PSY3 tandem might be deriving the flux of MEP-derived precursors towards the synthesis of carotenoid-derived molecules that are crucial for the symbiosis establishment (i.e. SL hormones and C13 α-ionol/C14 mycorradicin apocarotenoids). The possible involvement of this protein pair in the production of ABA to face abiotic stress events in roots is yet to be determined. However, the observation that *SlG1* was only co-expressed with a few genes from the ABA synthetic pathway in vegetative and fruit tissues (Figure 5) would be consistent with the contribution of SlG1 to the production of ABA under certain environmental conditions. *SlG1* expression was found to be induced in leaves under herbivore-feeding, correlating with an increase in the emissions of GGPP-derived volatiles involved in the attraction of natural enemies of the attacking insect (Ament et al., 2006; Richter et al., 2016). It is therefore possible that this isoform might be specialized in the formation of GGPP for the production of metabolites required in stress situations not only in roots, but also in other plant tissues such as leaves.

In the case of *SlG2*, GCN analysis showed a strong relationship with most photosynthesis-related isoprenoids in vegetative tissues (Figure 5A). This expression pattern is very similar to that of the *AtG11* gene that is highly co-expressed with most upstream and downstream plastidial isoprenoid pathways, consistent with its demonstrated role on the production of photosynthesis-related isoprenoids such as carotenoids, chlorophylls or prenylquinones (Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see* **Annex 2**). *SlG3* was also highly co-expressed with genes involved in chlorophyll and carotenoid metabolism in vegetative tissues (Figure 5A). However, qPCR analyses revealed that *SlG2* was much more strongly upregulated during tomato seedling de-etiolation, when an enhanced production of photosynthesis-related isoprenoids contributes to assemble a functional photosynthetic machinery (Figure 6B). The induction profile of *SlG2* was very similar to that of photosynthesis-related *PSY2*, whereas *SlG3* expression pattern during deetiolation paralleled that of ripening-associated *PSY1* (Figure 6B). These results together with the GCN analyses strongly suggest a more prominent role of SlG2 compared to SlG3 in photosynthetic tissues. They also indicate that, similar to that proposed for the SlG1-PSY3 pair, SlG2-PSY2 and SlG3-PSY1 might form modules to

direct MEP-derived precursors into the production of specific sets of carotenoids (Figure 8).

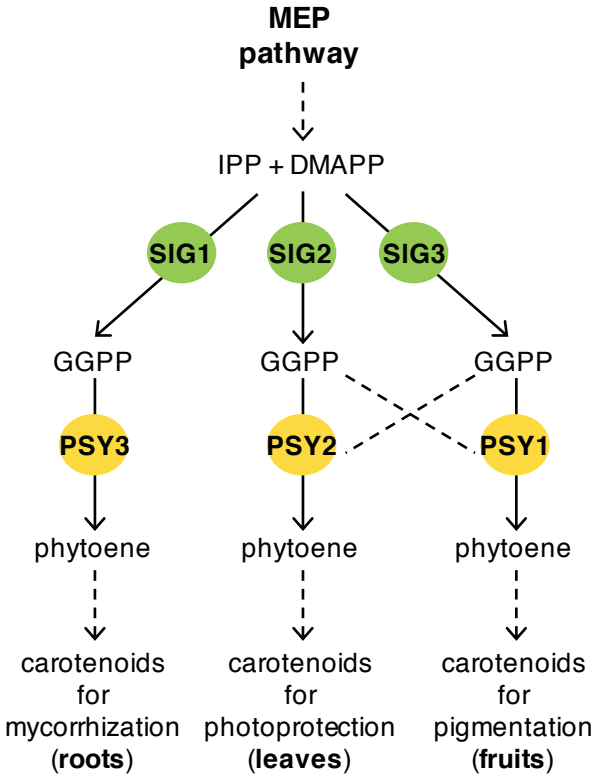


Figure 8. Model of the channeling of MEP-derived isoprenoid precursors towards the biosynthesis of carotenoids with specific functions in particular tomato tissues. Different plastidial GGPPS enzymes are transcriptionally co-regulated with specific PSY isoforms in tomato and might interact physically to provide GGPP for the synthesis of particular groups of carotenoids. Thus, SlG1 and PSY3 would cooperate in roots, while SlG2-PSY2 and/or SlG3-PSY1 pairs might function together to produce carotenoids in leaves and fruits.

The roles of SlG2 and SlG3 also seem to be interconnected during fruit ripening (Figure 8). In the fruit, *SlG3* is more highly expressed than *SlG2*, has a bigger node degree, and shows a stronger co-expression with carotenoid-related genes, including *PSY1* (Figure 5B and Figure 6C). *SlG2* was also induced during fruit ripening but peaked before *SlG3* and *PSY1* (Figure 6C). In any case, upregulation of *SlG2* and *SlG3* transcript levels during ripening was much lower than that measured for the *PSY1* gene (Figure 6C). At the breaker stage, the relative levels of *PSY1* transcripts increase around 30 times compared to its expression in mature-green fruits. By contrast, *SlG3* expression was just 2.5-fold higher, while *SlG2* was induced around 7-fold. We speculate that *SlG2* might be strongly induced at this point to help the SlG3 isoform to supply GGPP to boost the synthesis of carotenoids. When the fruit is ripe and final carotenoid levels have been reached, the expression of *PSY1* drastically decreases. The similar repression of *SlG2* and *SlG3* expression at this stage keeps supporting the involvement of both isoforms in fruit carotenoid production (Figure 6C). However, the observation that phytoene accumulation is reduced in ripe fruits from transgenic

lines with higher *SlG2* but lower *SlG3* transcript levels (Figure 7, Figure S5) strongly supports SlG3 as the main isoform producing GGPP for carotenoid biosynthesis during fruit ripening (Figure 7).

Apart from transcriptional regulation, a wide range of studies have been pointing out the existence of protein complexes regulating isoprenoid metabolism through the allocation of the precursors into the required metabolic branches. The interaction of different plant GGPPS enzymes with other proteins involved in isoprenoid biosynthesis, including PSY, was demonstrated through different protein-protein interaction approaches in the past decades (Dogbo and Camara, 1987; Camara, 1993; Maudinas et al., 1977; Fraser et al., 2000; Ruiz-Sola et al., 2016a, *see* **Annex1**; Zhou et al., 2017; Camagna et al., 2018; Wang et al., 2018b). In the light of our observations, we speculate that something similar might occur between specific GGPPS and PSY isoforms in tomato. Interaction of SlG1 and PSY3 might ensure the channelling of IPP and DMAPP into the production of carotenoids involved in mycorrhization (Figure 8). Complexes containing either SlG2 and PSY2 or SlG3 and PSY1 might cooperate to produce carotenoids in leaves and fruits, with the SlG2-PSY2 tandem preferentially involved in the biosynthesis of photosynthesis-related carotenoids and the SlG3-PSY1 pair mostly dedicated to ripening-related pigmentation (Figure 8). This relatively simple scenario, however, can be much more complex *in vivo*, based on the presence of different GGPPS-encoding genes in the same tissues and the potential of the encoded proteins to heterodimerize among them and with other related proteins (*e.g.* SlSSUII). The existence of tomato GGPPS metabolons can only be experimentally ascertained, *e.g.* by the isolation of the GGPPS-containing protein complexes in different tissues, followed by the sequencing of all protein interactors. Analysis of simple and double tomato mutants lacking specific GGPPS isoforms (*e.g.* generated using CRISPR-Cas9 technology) should confirm whether they have redundant roles or, on the contrary, their function is specific for some processes or in particular tissues. Deciphering the regulation of plastidial GGPP production and channeling will be really useful for future metabolic engineering approaches, aimed to manipulate the accumulation of carotenoids and other specific groups of GGPP-derived isoprenoids. For example, in the case of carotenoids it should allow to specifically improve the nutritional quality of tomato fruits without interfering with other vital processes such as photoprotection, mycorrhization, or tolerance to stress.

4. Materials and Methods

4.1. Plant material and growth conditions

Tomato (*Solanum lycopersicum* var. MicroTom) plants were used for most experiments. Seeds were surface-sterilized by a 15 min incubation with 40% bleach (and 2 drops of tween-20 every 50 mL of solution) and three consecutive 20 min-washes with sterilized milli-Q water. Sterile-seeds were then sown in Corning® round plates (D x H 150 mm x 25 mm, Merck) with solid Murashige and Skoog medium without vitamins or sucrose, but adding kanamycin (100 µg/mL) when required to select transgenic plants. After stratification for 3 days at 4 °C in the dark, plates were incubated in a climate-controlled growth chamber at 24 °C and illuminated for 14 h with fluorescent white light at a photosynthetic photon flux density of 140 µmol m^{-2} s^{-1}. 7-10 day seedlings were then transferred to soil and grown under standard greenhouse conditions (14 h light at 27 ± 1 °C and 10 h dark at 22 ± 1 °C).

Fruits samples were collected at 4 ripening stages according to the days post-anthesis (DPA): mature green (MG, ~30 DPA), breaker (BR, ~35 DPA), orange (OR, ~38-40 DPA) and red (R, ~45-50 DPA). The pericarp was separated from the rest of the fruit and immediately frozen in liquid nitrogen and freeze-dried during 3 days in a laboratory freeze-dryer (ALPHA 2-4 LD plus, CHRIST). Freeze-dried samples were stored at -80 °C until further analysis. For deetiolation experiments, tomato sterilized seeds were sown on sterile water-soaked cotton in transparent plastic containers. After stratification, seeds were exposed to fluorescent white light for 2-4 hours at 22 °C to induce germination. The containers were then covered with a double layer of aluminum foil and kept in darkness at 22 °C. Control samples were left under continuous light conditions. After one week, etiolated seedlings were exposed to light and samples were harvested at 0, 6 and 24 hours after illumination. Control (non-etiolated) seedlings were collected at time-point 0 hours. Samples were frozen in liquid nitrogen, subjected to an overnight freeze-drying and stored at -80 °C. At least three independent biological replicates (containers) were collected for each time point. Each biological replicate was a pool of ten tomato seedlings. *Nicotiana benthamiana* plants used for transient expression assays were grown under standard greenhouse conditions (14 h light at 26 ± 1°C and 10 h dark at 21 ± 1°C).

4.2. Constructs

Full-length cDNAs encoding SlG1-5 (Table 1) were amplified from tomato cDNA and cloned into pDONR207 entry plasmid using Gateway (GW) technology (Invitrogen, California, USA). Full length sequences were then sub-cloned into pGWB405 plasmid (Nakagawa et al., 2007) and used for transient expression in *N. benthamiana* leaves and stable transformation of tomato plants (MicroTom). Constructs in the pGWB405 vector harbor the GFP sequence in frame at the 3' end of each element cloned and the expression module is controlled by the CaMV 35S promoter. pGWB405_pGFP construct (*35S:pGFP*) was generated by N-terminal fusion of the plastid-targeting sequence of the Arabidopsis *HDS* gene (Gas et al., 2009) to GFP and it was used as an overexpression plastidial control. For protein expression in *E. coli*, SlG1-3 versions lacking the predicted transit peptide were amplified from pGWB405 constructs, cloned into the pDONR207 plasmid and then subcloned into pET32-GW plasmid fusing a 6x-Histidine tag at the N-terminal end of the cloned fragments. pET32-GW constructs containing similar Arabidopsis AtG11 versions are described in (Ruiz-Sola et al., 2016b, *see* **Annex 2**). Constructs were confirmed by restriction mapping and DNA sequence analysis. Information about primers used and cloning details are described in Supplementary Tables S4 and S5 respectively.

4.3. Phylogenetic analysis

Arabidopsis GGPPS (Beck et al., 2013; Ruiz-Sola et al., 2016b, *see* **Annex 2**) and other PT protein sequences were retrieved from *The Arabidopsis Information Resource* (https://www.arabidopsis.org/) and used as queries to search for putative tomato homologs using BLAST against tomato genome on the *Solanaceae Genomics Network* (http://solgenomics.net/) and the *National Center for Biotechnology Information* websites (www.ncbi.nlm.nih.gov/). The accession numbers of the identified homologs are listed in Table 1. The presence of transit peptides was predicted using *ChloroP* and *TargetP* algorithms as described in (Emanuelsson et al., 2007). Protein alignments of Arabidopsis and tomato PT sequences lacking the predicted transit peptides (Figure S1), were performed using *Clustal Omega* (https://www.ebi.ac.uk/Tools/msa/clustalo/) with default settings (Sievers et al., 2011; Sievers and Higgins, 2014). An unrooted tree was then constructed using *MEGA6* (Hall, 2013; Tamura et al., 2013). Evolutionary connections among PTs were predicted by the Neighbor-Joining method based on the Poisson model, where pairwise deletion was selected for gap deletions. Bootstrapping was performed on 1,000 pseudoreplicates.

4.4. Subcellular localization assays

Subcellular localization assays were performed by *Agrobacterium tumefaciens*-mediated transient expression in *N. benthamiana* leaves (Sparkes et al., 2006). *A. tumefaciens* GV3101 strains were transformed with constructs in the pGWB405 vector and grown on Luria-Bertani (LB) agar plates with the corresponding antibiotics at 28 °C for 3 days. A single PCR-confirmed colony per construct was inoculated in 5 mL antibiotic-complemented LB media and incubated overnight at 28 °C in 300 rpm continuous agitation. 500 µL of the grown culture were then inoculated in 20 mL of LB media and incubated overnight at 28 °C in 300 rpm continuous agitation. OD_{600} of each liquid culture was spectrophotometrically measured and after that, cultures were centrifuged at 4000 rpm for 10 min. Bacterial pellets were resuspended in infiltration buffer (10 mM MES pH5.5-6, 10 mM $MgSO_4$, 150 µM acetosyringone) to reach a final OD_{600} of 0.5. To prevent silencing, leaves were co-infiltrated with an Agrobacterium strain transformed with a vector expressing the HC-Pro silencing suppressor (Goytia et al., 2006). A 1:1 mixture of the two cultures was infiltrated with a syringe in the abaxial part of leaves from four to six-week old *N. benthamiana* plants. Then, plants were left in the greenhouse and leaf samples were collected three days post-infiltration (dpi) for confocal microscopy analysis.

4.5. Confocal microscopy

Subcellular localization of GFP fusion proteins was determined by analyzing agroinfiltrated leaf samples with an Olympus FV 1000 confocal laser-scanning microscope. GFP signal and chlorophyll autofluorescence were detected using an argon laser for excitation (at 488 nm) and a 500–510 nm filter for detection of GFP fluorescence and a 610–700 nm filter for detection of chlorophyll fluorescence. All images were acquired using the same confocal parameters.

4.6. Stable transformation of tomato plants

Strain GV3101 of *A. tumefaciens* was used to stably transform tomato MicroTom cotyledons with pGWB405_SlG2, SlG3 or pGFP as described in Fernandez et al (2009). *In vitro* regenerated T1 tomato transgenic lines were identified based on kanamycin resistance (100 µg/mL) and genotyping PCR, using *ACTIN4* (Solyc04g011500) as an internal control. Homozygous tomato lines containing a single T-DNA insertion were selected based on a RT-qPCR genotyping method

optimized in our lab (*see* **Annex 4)**, using tomato *LAT52* (Solyc10g007270) as an endogenous reference gene. Genotyping PCR and qPCR primers are listed in Supplementary Table S4. At least, two independent lines of each construct were used for further experiments.

4.7. GGPPS activity assay and product identification

Constructs to express different truncated GGPPS protein versions were generated in pET32-GW vector as previously described (Table S5). Competent *E. coli Rosetta* 2 (DE3) cells (Novagen, Merck KGaA, Darmstadt, Germany) were separately transformed with each construct to produce the recombinant proteins. Single colonies of each individual transformant were inoculated in 5 mL of LB medium with appropriate antibiotics and grown overnight at 37 °C. 250 µL of each pre-culture was diluted in 25 mL 2xYT medium with the required antibiotics and incubated at 37 °C and 250 rpm until an OD_{600} between 0.5 and 0.8 was reached. Then, 1 mM of IPTG was added to induce the production of the recombinant proteins and the cultures were incubated overnight at 18 °C and 250 rpm. Bacterial cells were harvested by centrifugation at 3400 rpm for 15 min and pellets were resuspended in 1 mL of assay buffer (15 mM MOPSO, 12.5% v/v glycerol, 1 mM ascorbic acid, pH 7.0, 1 mM $MgCl_2$, 2 mM DTT). About 0.2 g of zirconium/silica beads 0.1 mm (BioSpec Products) were added and bacterial lysis was carried out by two rounds of shaking for 10 seconds at speed 6.5 in a *FastPrep* machine (FP120 Bio101 Savant). Cell lysates were subsequently centrifuged during 10 min at 13,000 g and 4 °C and supernatants were collected for SDS-PAGE and GGPP activity assays. Enzymatic assays were performed in a final volume of 200 µL containing 25 µL of cell extract, 150 µM IPP and 50 µM DMAPP in assay buffer supplemented with 5 mM Na_3O_4V. The reaction mix was incubated for 2 h at 30 °C in mild agitation and stopped by adding 800 µL of 100% methanol / 0.5% formic acid. After vortexing, samples were sonicated for 15 min and centrifuged at maximum speed for 10 min. Supernatants were then evaporated in a SpeedVac concentrator (Eppendorf Concentrator plus) during 4 h and 80 µL of 100% methanol / 0.65% formic acid were added to the remnant sample. After centrifugation at maximum speed for 15 min, the supernatants were transferred glass vials. The detection of prenyl diphosphate products by LC-MS was carried out as described in Ruiz-Sola et al (2016b, see **Annex 2**). Data acquisition and visualization was performed using *Xcalibur™* software (ThermoFischer Scientific™). IPP and DMAPP substrates and FPP and GGPP standards were obtained from Echelon Biosciences Inc.

4.8. Kinetic analyses

Enzyme assays for the determination of the tomato GGPPSs kinetic properties (optimal pH and Km and Vmax) were developed as detailed in Barja et al. 2019 (*see* **Annex 3**) using 3 µg of purified SlG1, SlG2, SlG3 and AtG11 enzymes. pET32 constructs were used to produce 6x-His-tagged recombinant enzymes (Table S5) and protein purification from *E. coli Rosetta* cells was carried out using nickel-nitrilotriacetic acid (Ni-NTA) agarose (Qiagen) as described in Barja et al., 2019 (*see* **Annex 3**). Protein concentration was determined according to Bradford method (Bradford, 1976).

4.9. Gene Co-expression Network (GCN) analyses

Gene co-expression analysis was performed as previously described (Ahrazem et al., 2018). Briefly, an *ad hoc* list of tomato *GGPPS* genes was compiled and used to retrieve all the expression data available for different tomato cultivars/tissue/treatment in the *TomExpress* database (Maza et al., 2013). Subsequently, pairwise Pearson correlations between each GGPPS gene and each selected isoprenoid biosynthetic input gene (Table S2) was computed for vegetative and fruit tissues, and Fisher's Z-transformation was used to test the statistical significance of pairwise correlations. Finally, positive correlations ($\rho>0.55$) were used to draw gene co-expression networks (Table S3).

4.10. Gene expression analysis by RT-qPCR

Total RNA was isolated from freeze-dried tissue (tomato fruit pericarp or seedlings) using the *Maxwell® RSC Plant RNA Kit* with the *Maxwell® RSC Instruments* (Promega) following the manufacturer's instructions, quantified using a *NanoDrop™ 8000 spectrophotometer* (ThermoFischer Scientific™) and checked for integrity by agarose gel electrophoresis. The *Transcriptor First Strand cDNA Synthesis Kit* (Roche) was used, considering the provided recommendations, to reverse transcribe 0.5 µg of extracted RNA into 20 µL of cDNA, subsequently diluted ten-fold and stored at -20 °C for further analysis.

Relative mRNA abundance was evaluated via Real-Time Quantitative Polymerase Chain Reaction (RT-qPCR) in a reaction volume of 20 µL containing 10 µL of the LightCycler 480 SYBR Green I Master Mix (Roche), 0.3 µM of each specific forward and reverse primers and 5 µL of prepared cDNA. The RT-qPCR was carried out on a

LightCycler 480 Real-Time PCR System (Roche). Three independent biological replicates of each condition and at least two technical replicates of each biological replicate were performed, and the mean values were used for further calculations. Normalized transcript abundances were calculated as described previously (Simon, 2003) using tomato *ACT4* (Solyc04g011500) or *EXP* (Solyc07g025390) as endogenous reference genes. Primer efficiencies were calculated using serial dilutions of genomic or plasmidic DNA. Primers used and efficiencies are listed in Table S4.

Three biological replicates of cDNA samples from roots of non-mycorrhized and mycorrhized tomato plants were kindly provided by Dr. Juan Antonio López-Ráez. *Solanum lycopersicum* (cv. Reimlams Rhums) plants were inoculated with the arbuscular mycorrhizal (AM) fungus *Rhizophagus irregularis* (EEZ 58 strain) and cultivated for 6 weeks under moderate drought stress. Total RNA was extracted from roots of non-inoculated and inoculated plants and 1 µg was used to synthesize the cDNA. More details of plant material, growth conditions and cDNA synthesis can be found in Ruiz-Lozano et al. (2016). The cDNA received was then diluted 20 times before performing the RT-qPCR experiments.

4.11. Extraction and analysis of GGPP-derived metabolites

Carotenoids, chlorophylls, tocopherols and phylloquinone were extracted in 2 mL Eppendorf tubes from 15 mg of freeze-dried tomato pericarp tissue, using 1 ml of hexane/acetone/methanol 2:1:1 as extraction solvent and 15 µg of canthaxanthin (Sigma) as internal control. After vortexing for 10 secs and lysing the tissue with 4 mm glass beads for 1 min at 30 Hz in a *TissueLyser II* (QIAGEN), 100 µL of water were added. Then, 1 min of TissueLyser was carried out again and samples were centrifuged for 3 min at 3,000 rpm and 4 °C. Organic phase (upper) was kept in a new tube and the rest was re-extracted with 1 mL hexane/acetone/methanol 2:1:1, 1 min of TissueLyser and centrifuging for 5 min at maximum speed and 4 °C. The new organic phase was mixed with that previously extracted and evaporated for 1 h using a *SpeedVac system* (Eppendorf Concentrator plus). Extracted metabolites were then completely re-dissolved in 150 µL of acetone and filtered with 0.2 µm filters into amber-colored 2 mL glass vials. Separation and detection of isolated compounds was performed from 33 µL of prepared samples using an *Agilent 1200 series HPLC system* (Agilent Technologies) as described previously (Fraser et al., 2000). The HPLC equipment was coupled to a Photometric Diode Array (PDA) detector, allowing the detection of the full uv-visible absorption spectra of carotenoids and chlorophylls.

Tocopherols and phylloquinone were identified using a fluorescence detector at 330 and 420 nm, respectively. Chromatogram visualization and data analysis were performed using the *Agilent ChemStation* software.

5. Acknowledgements

We are very grateful to Juan Antonio López-Ráez (Estación Experimental del Zaidín-CSIC) for the cDNA samples of tomato mycorrhized roots. Technical support from M. Rosa Rodríguez-Goberna and members of the CRAG Services is greatly appreciated. This work was funded by grants BIO2014-59092-P, BIO2015-71703-REDT, BIO2017-90877-REDT, and BIO2017-84041-P from the Spanish Ministry of Science, Innovation, and Universities (MICINN) and grants 2014SGR-1434 and 2017SGR-710 from Generalitat de Catalunya to MRC. The financial support of the MINECO Severo Ochoa Program for Centres of Excellence in R&D 2016-2019 (SEV-2015-0533) and the Generalitat de Catalunya CERCA Program to CRAG is also acknowledged. MVB and ME were supported by PhD fellowships FPU14/05142 and BES-2017-080652, respectively.

6. References

Aaron JA and Christianson DW (2010). Trinuclear metal clusters in catalysis by terpenoid synthases. Pure Appl. Chem. **82**: 1585–1597.

Ahrazem O, Argandoña J, Fiore A, Aguado C, Luján R, Rubio-Moraga A, Marro M, Araujo-Andrade C, Loza-Alvarez P, Diretto G and Gómez-Gómez L (2018). Transcriptome analysis in tissue sectors with contrasting crocins accumulation provides novel insights into apocarotenoid biosynthesis and regulation during chromoplast biogenesis. Sci Rep. **8**: 2843.

Ament K, Van Schie CC, Bouwmeester HJ, Haring MA and Schuurink RC (2006). Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. Planta **224**: 1197–1208.

Barja MV and Rodriguez-Concepcion M (2019). A simple *in vitro* assay to measure the activity of geranylgeranyl diphosphate synthase and other short-chain prenyltransferases. Methods Mol Biol. ACCEPTED. *See* **Annex 3.**

Bartley G and Scolnik P (1993). cDNA cloning, expression during development, and genome mapping of a second phytoenen synthase. Biochemistry **268**: 25718–25721.

Bartley GE, Viitanen PV, Bacot KO and Scolnik PA (1992). A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthesis pathway. J. Biol. Chem. **267**: 5036–9.

Baslam M, Esteban R, García-Plazaola JI and Goicoechea N (2013). Effectiveness of arbuscular mycorrhizal fungi (AMF) for inducing the accumulation of major carotenoids, chlorophylls and tocopherol in green and red leaf lettuces. Appl. Microbiol. Biotechnol. **97**: 3119–3128.

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodríguez-Concepción M, Gruissem W and Vranová E (2013). Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. Plant Mol. Biol. **82**: 393–416.

Bick JA and 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**: 146–154.

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**: 248–54.

Camagna M, Grundmann A, Bär C, Koschmieder J, Beyer P and Welsch R (2018). Enzyme fusion removes competition for geranylgeranyl diphosphate in carotenogenesis. Plant Physiol.: pp.01026.2018.

Camara B (1993). Plant phytoene synthase complex: Component enzymes, Immunology and Biogenesis. Methods Enzymol. **214**: 352–365.

Chayut N, Yuan H, Ohali S, Meir A, Sa'ar U, Tzuri G, Zheng Y, Mazourek M, Gepstein S, Zhou X, Portnoy V, Lewinsohn E, Schaffer AA, Katzir N, Fei Z, Welsch R, Li L, Burger J and Tadmor Y (2017). Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. Plant Physiol. **173**: 376–389.

Chen Q, Fan D and Wang G (2015). Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in arabidopsis flowers. Mol. Plant **8**: 1434–1437.

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

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

D'Andrea L, Simon-Moya M, Llorente B, Llamas E, Marro M, Loza-Alvarez P, Li L and Rodriguez-Concepcion M (2018). Interference with Clp protease impairs carotenoid accumulation during tomato fruit ripening. J. Exp. Bot. **69**: 1557–1568.

Dogbo O and Camara B (1987). Purification of isopentenyl pyrophosphate isomerase and geranylgeranyl pyrophosphate synthase from Capsicum chromoplasts by affinity chromatography. Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab. **920**: 140–148.

Emanuelsson O, Brunak S, von Heijne G and Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. **2**: 953–971.

Fernandez AI, Viron N, Alhagdow M, Karimi M, Jones M, Amsellem Z, Sicard A, Czerednik A, Angenent G, Grierson D, May S, Seymour G, Eshed Y, Lemaire-Chamley M, Rothan C and Hilson P (2009). Flexible tools for gene expression and silencing in tomato. Plant Physiol. **151**: 1729–1740.

Fernandez SMS, Kellogg BA and Poulter CD (2000). Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity. Biochemistry **39**: 15316–15321.

Fester T, Schmidt D, Lohse S, Walter MH, Giuliano G, Bramley PM, Fraser PD, Hause B and Strack D (2002). Stimulation of carotenoid metabolism in arbuscular mycorrhizal roots. Planta **216**: 148–154.

Fester T, Wray V, Nimtz M and Strack D (2005). Is stimulation of carotenoid biosynthesis in arbuscular mycorrhizal roots a general phenomenon? Phytochemistry **66**: 1781–1786.

Fraser PD, Enfissi EMA, Halket JM, Truesdale MR, Yu D, Gerrish C and Bramley PM (2007). Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism. Plant Cell **19**: 3194–3211.

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

Fraser PD, Romer S, Shipton CA, Mills PB, Kiano JW, Misawa N, Drake RG, Schuch W and Bramley PM (2002). Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. Proc. Natl. Acad. Sci. U. S. A. **99**: 1092–7.

Fraser PD, Schuch W and Bramley PM (2000). Phytoene synthase from tomato (*Lycopersicon esculentum*) chloroplasts--partial purification and biochemical properties. Planta **211**: 361–369.

Fray RG and Grierson D (1993). Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Mol. Biol. **22**: 589–602.

Gachon CMM, Langlois-Meurinne M, Henry Y and Saindrenan P (2005). Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: Functional and evolutionary implications. Plant Mol. Biol. **58**: 229–245.

Gady ALF, Vriezen WH, Van de Wal MHBJ, Huang P, Bovy AG, Visser RGF and Bachem CWB (2012). Induced point mutations in the *phytoene synthase 1* gene cause differences in carotenoid content during tomato fruit ripening. Mol. Breed. **29**: 801–812.

Gaffe J, Bru JP, Causse M, Vidal A, Stamitti-Bert L, Carde JP and Gallusci P (2000). *LEFPS1*, a tomato farnesyl pyrophosphate gene highly expressed during early fruit development. Plant Physiol. **123**: 1351–1362.

Gas E, Flores-Pérez U, Sauret-Güeto S and Rodríguez-Concepción M (2009). Hunting for plant nitric oxide synthase provides new evidence of a central role for plastids in nitric oxide metabolism. Plant Cell **21**: 18-23.

Giorio G, Stigliani AL and D'Ambrosio C (2008). Phytoene synthase genes in tomato (*Solanum lycopersicum L.*) - New data on the structures, the deduced amino acid sequences and the expression patterns. FEBS J. **275**: 527–535.

Goytia E, Fernández-Calvo L, Martínez-García B, López-Abella D and López-Moya JJ (2006). Production of plum pox virus HC-Pro functionally active for aphid transmission in a transient-expression system. J Gen Virol. **87**: 3413–23.

Hall BG (2013). Building phylogenetic trees from molecular data with MEGA. Mol. Biol. Evol. **30**: 1229–1235.

Heyndrickx KS and Vandepoele K (2012). Systematic identification of functional plant modules through the integration of complementary data sources. Plant Physiol. **159**: 884–901.

Hirooka K, Izumi Y, An C, Nakazawa Y, Fukusaki E and Kobayashi A (2005). Functional analysis of two solanesyl diphosphate synthases from *Arabidopsis thaliana*. Biosci Biotechnol Biochem. **69**: 592-601.

Höhner R, Aboukila A, Kunz HH and Venema K (2016). Proton gradients and proton-dependent transport processes in the chloroplast. Front. Plant Sci. **7**: 1–7.

Jones MO, Perez-Fons L, Robertson FP, Bramley PM and Fraser PD (2013). Functional characterization of long-chain prenyl diphosphate synthases from tomato. Biochem. J. **449**: 729–740.

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

Jun L, Saiki R, Tatsumi K, Nakagawa T and Kawamukai M (2004). Identification and subcellular localization of two solanesyl diphosphate synthases from *Arabidopsis thaliana*. Plant Cell Physiol **45**: 1882–8.

Kachanovsky DE, Filler S, Isaacson T and Hirschberg J (2012). Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by cis-carotenoids. Proc. Natl. Acad. Sci. U. S. A. **109**: 19021–6.

Klee HJ and Giovannoni JJ (2011). Genetics and control of tomato fruit ripening and quality attributes. Annu. Rev. Genet. **45**: 41–59.

Koike-Takeshita A, Koyama T, Obata S and Ogura K (1995). Molecular cloning and nucleotide sequences of the genes for two essential proteins constituting a novel enzyme system for heptaprenyl diphosphate synthesis. J. Biol. Chem. **270**: 18396–18400.

Koyama T, Tajima M, Sano H, Doi T, Koike-Takeshita A, Obata S, Nishino T and Ogura K (1996). Identification of significant residues in the substrate binding site of *Bacillus stearothermophilus* farnesyl diphosphate synthase. Biochemistry **35**: 9533–8.

Llorente B, Martinez-Garcia JF, Stange C and Rodriguez-Concepcion M (2017). Illuminating colors: regulation of carotenoid biosynthesis and accumulation by light. Curr. Opin. Plant Biol. **37**: 49–55.

Marrero PF, Poulter CD and Edwards PA (1992). Effects of site-directed mutagenesis of the highly conserved aspartate residues in domain II of farnesyl diphosphate synthase activity. J. Biol. Chem. **267**: 21873–8.

Maudinas B, Bucholtz ML, Papastephanou C, Katiyar SS, Briedis AV and Porter JW (1977). The partial purification and properties of a phytoene synthesizing enzyme system. Arch. Biochem. Biophys. **180**: 354–62.

Maza E, Frasse P, Senin P, Bouzayen M and Zouine M (2013). Comparison of normalization methods for differential gene expression analysis in RNA-Seq experiments. Commun Integr Biol. **6**: e25849.

Nagel R, Bernholz C, Vranová E, Košuth J, Bergau N, Ludwig S, Wessjohann L, Gershenzon J, Tissier A and Schmidt A (2015). *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate, geranylfarnesyl diphosphate. Plant J. **84**: 847–859.

Nakagawa T, Suzuki T, Murata S, Nakamura S, Hino T, Maeo K, Tabata R, Kawai T, Tanaka K, Niwa Y, Watanabe Y, Nakamura K, Kimura T and Ishiguro S (2007). Improved Gateway

binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci. Biotechnol. Biochem. **71**: 2095–100.

Nisar N, Li L, Lu S, Khin NC and Pogson BJ (2015). Carotenoid metabolism in plants. Mol. Plant **8**: 68–82.

Ohnuma S, Hirooka K, Hemmi H, Ishida C, Ohto C and Nishino T (1996a). Conversion of product specificity of archaebacterial geranylgeranyl-diphosphate synthase. Identification of essential amino acid residues for chain length determination of prenyltransferase reaction. J. Biol. Chem. **271**: 18831–7.

Ohnuma SI, Narita K, Nakazawa T, Ishida C, Takeuchi Y, Ohto C and Nishino T (1996b). A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. J. Biol. Chem. **271**: 30748–30754.

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

Oliver S (2000). Guilt-by-association goes global. Nature **403**: 601–3.

Pankratov I, McQuinn R, Schwartz J, Bar E, Fei Z, Lewinsohn E, Zamir D, Giovannoni JJ and Hirschberg J (2016). Fruit carotenoid-deficient mutants in tomato reveal a function of the plastidial isopentenyl diphosphate isomerase (IDI1) in carotenoid biosynthesis. Plant J. **88**: 82–94.

Park S, Kim HS, Jung YJ, Kim SH, Ji CY, Wang Z, Jeong JC, Lee HS, Lee SY and Kwak SS (2016). Orange protein has a role in phytoene synthase stabilization in sweetpotato. Sci. Rep. **6**: 33563.

Perello C, Llamas E, Burlat V, Ortiz-Alcaide M, Phillips MA, Pulido P and Rodriguez-Concepcion M (2016). Differential subplastidial localization and turnover of enzymes involved in isoprenoid biosynthesis in chloroplasts. PLoS One **11**: e0150539.

Phillips MA, D'Auria JC, Gershenzon J, Pichersky E (2008a). The *Arabidopsis thaliana* type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. Plant Cell **20:** 677-696.

Pulido P, Perello C and Rodriguez-Concepcion M (2012). New insights into plant isoprenoid metabolism. Mol Plant **5:** 964-967.

Richter A, Schaff C, Zhang Z, Lipka AE, Tian F, Köllner TG, Schnee C, Preiß S, Irmisch S, Jander G, Boland W, Gershenzon J, Buckler ES and Degenhardt J (2016). Characterization of biosynthetic pathways for the production of the volatile homoterpenes DMNT and TMTT in *Zea mays*. Plant Cell **28**: 2651–2665.

Rodriguez-Concepcion M, Avalos J, Bonet ML, Boronat A, Gomez-Gomez L, Hornero-Mendez D, Limon MC, Meléndez-Martínez AJ, Olmedilla-Alonso B, Palou A, Ribot J, Rodrigo MJ, Zacarias L, Zhu C (2018). A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. Prog. Lipid Res. **70**: 62–93.

Rodriguez-Concepcion M and Boronat A (2015). Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. Curr Opin Plant Biol **25:** 17-22.

Rodriguez-Villalon A, Gas E and Rodriguez-Concepcion M (2009). Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings. Plant J. **60**: 424–435.

Ruiz-Lozano JM, Aroca R, Zamarreño AM, Molina S, Andreo-Jimenez B, Porcel R, Garcia-Mina, JM, Ruyter-Spira C and Lopez-Raez JA (2016). Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato. Plant, Cell Environ. **39**: 441–452.

Ruiz-Sola MA, Coman D, Beck G, Barja MV, Colinas M, Graf A, Welsch R, Rutimann P, Buhlmann P, Bigler L, Gruissem W, Rodriguez-Concepcion M and Vranova E (2016a). Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol **209:** 252-264. *See* **Annex 1**.

Ruiz-Sola MA, Barja MV, Manzano D, Llorente B, Schipper B, Beekwilder J and Rodriguez-Concepcion M (2016b). A single Arabidopsis gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. Plant Physiol. **172**: 1393–1402. *See* **Annex 2**.

Ruiz-Sola MA and Rodriguez-Concepcion M (2012). Carotenoid biosynthesis in Arabidopsis: A colorful pathway. Arab. B. **10**: e0158.

Sandmann G (2015). Carotenoids of biotechnological importance. Adv. Biochem. Eng. / Biotechnol. **148**: 449–467.

Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, Gidoni D and Eyal Y (2008). Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. Plant Physiol **148:** 1219-1228.

Sievers F and Higgins DG (2014). Clustal Omega. Curr. Protoc. Bioinforma. **48**: 3.13.1-16.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD and Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. **7**: 1–6.

Simon P (2003). Q-Gene: Processing quantitative real-time RT-PCR data. Bioinformatics **19**: 1439–1440.

Song L and Poulter CD (1994). Yeast farnesyl-diphosphate synthase: site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II. Proc. Natl. Acad. Sci. U. S. A. **91**: 3044–8.

Sparkes IA, Runions J, Kearns A and Hawes C (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat. Protoc. **1**: 2019–25.

Stauder R, Welsch R, Camagna M, Kohlen W, Balcke GU, Tissier A and Walter MH (2018). Strigolactone levels in dicot roots are determined by an ancestral symbiosis-regulated clade of the *PHYTOENE SYNTHASE* gene family. Front. Plant Sci. **9**: 1–18.

Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y and Li L (2018). Carotenoid metabolism in plants: The role of plastids. Mol. Plant **11**: 58–74.

Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30**: 2725–2729.

Tarshis LC, Sacchettini JC, Yan M and Poulter CD (1994). Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-å resolution. Biochemistry **33**: 10871–10877.

Tholl D (2015). Biosynthesis and biological functions of terpenoids in plants. In Advances in biochemical engineering/biotechnology, pp. 63–106.

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

Vranova E, Coman D and Gruissem W (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol **64:** 665-700.

Vranova E, Coman D and Gruissem W (2012). Structure and dynamics of the isoprenoid pathway network. Mol. Plant **5**: 318–333.

Walter MH, Stauder R and Tissier A (2015). Evolution of root-specific carotenoid precursor pathways for apocarotenoid signal biogenesis. Plant Sci. **233**: 1–10.

Wang C, Chen Q, Fan D, Li J, Wang G and Zhang P (2016). Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in *Brassicaceae* plants. Mol Plant **9:** 195-204.

Wang G and Dixon RA (2009). Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci U S A **106:** 9914-9919.

Wang J, Lin HX, Su P, Chen T, Guo J, Gao W and Huang LQ (2018a). Molecular cloning and functional characterization of multiple geranylgeranyl pyrophosphate synthases (ApGGPPS) from *Andrographis paniculata*. Plant Cell Rep. 38: 117-128.

Wang KC and Ohnuma S (1999). Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. Trends Biochem. Sci. **24**: 445–451.

Wang Q, Huang XQ, Cao TJ, Zhuang Z, Wang R and Lu S (2018b). Heteromeric geranylgeranyl diphosphate synthase contributes to carotenoid biosynthesis in ripening fruits of red pepper (*Capsicum annuum* var. conoides). J. Agric. Food Chem. **66:** 11691–11700.

Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C and Loraine A (2006). Transcriptional coordination of the metabolic network in arabidopsis. Plant Physiol. **142**: 762–774.

Welsch R, Zhou X, Yuan H, Álvarez D, Su T, Schlossarek D, Yang Y, Shen G, Zhang H, Rodriguez-Concepcion M Thannhauser TW and Li L (2018). Clp protease and OR directly control the proteostasis of phytoene synthase, the crucial enzyme for carotenoid biosynthesis in Arabidopsis. Mol. Plant **11**: 149–162.

Yuan H, Zhang J, Nageswaran D and Li L (2015). Carotenoid metabolism and regulation in horticultural crops. Hortic. Res. **2**.

Zhang M, Su P, Zhou YJ, Wang XJ, Zhao YJ, Liu YJ, Tong YR, Hu TY, Huang LQ and Gao W (2015). Identification of geranylgeranyl diphosphate synthase genes from *Tripterygium wilfordii*. Plant Cell Rep. **34**: 2179–88.

Zhou F, Wang CY, Gutensohn M, Jiang L, Zhang P, Zhang D, Dudareva N and Lu S (2017). A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. Proc. Natl. Acad. Sci. U. S. A. **114**: 6866–6871.

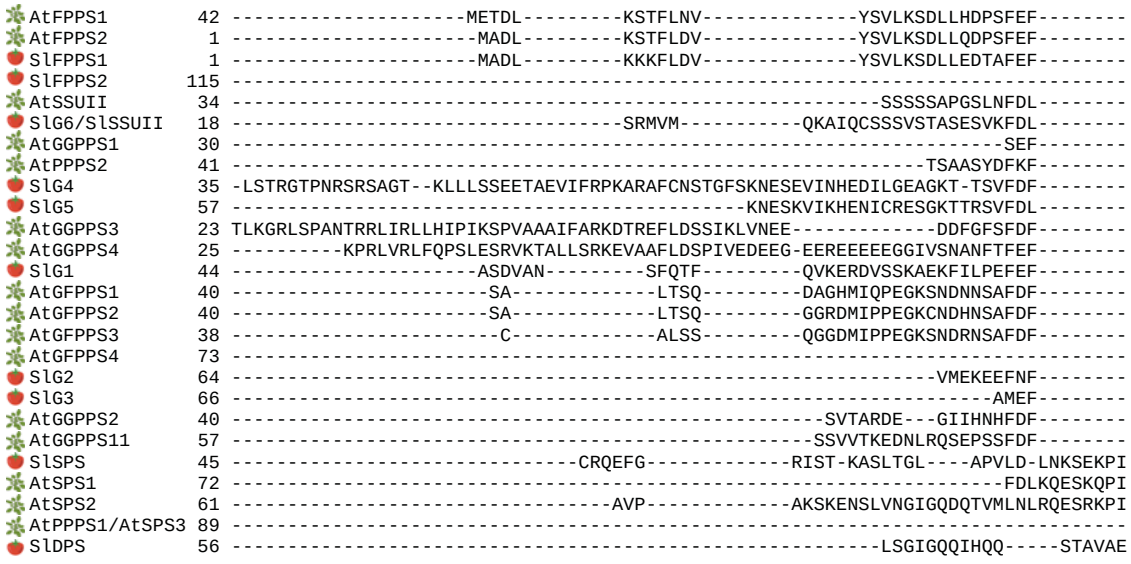
Zhou X, Welsch R, Yang Y, Álvarez D, Riediger M, Yuan H, Fish T, Liu J, Thannhauser TW and Li L (2015). Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. Proc. Natl. Acad. Sci. U. S. A. **112**: 3558–63.

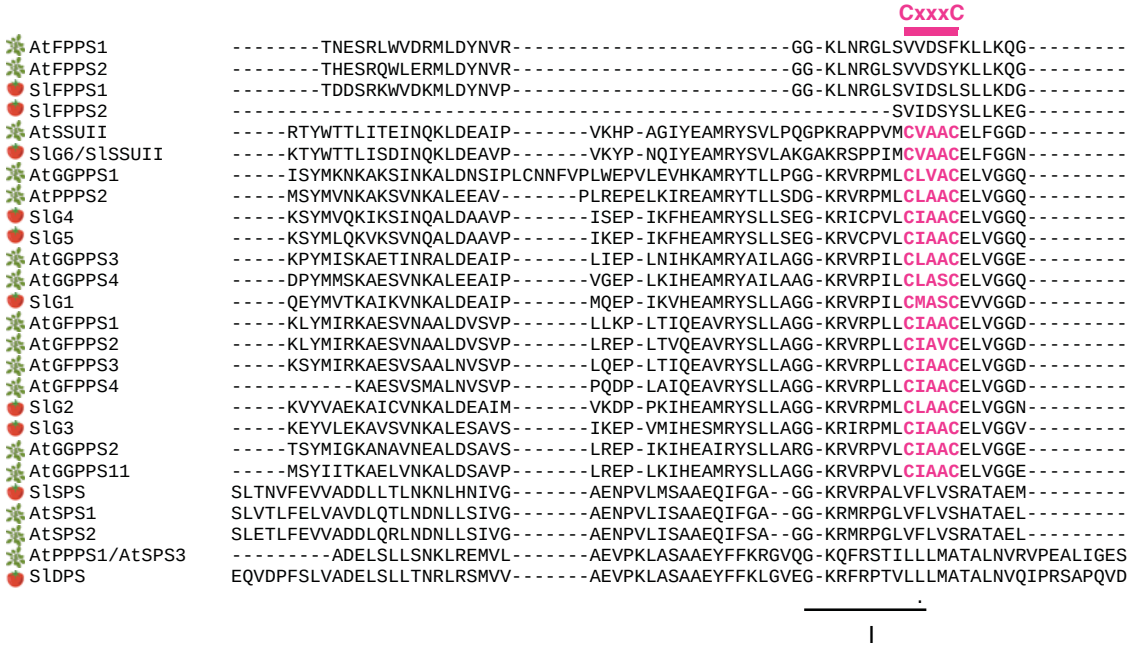
Zhu X, Suzuki K, Saito T, Okada K, Tanaka K, Nakagawa T, Matsuda H and Kawamukai M. (1997a). Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene *GGPS6* from *Arabidopsis thaliana* is localized in mitochondria. Plant Mol. Biol. **35**: 331–341.

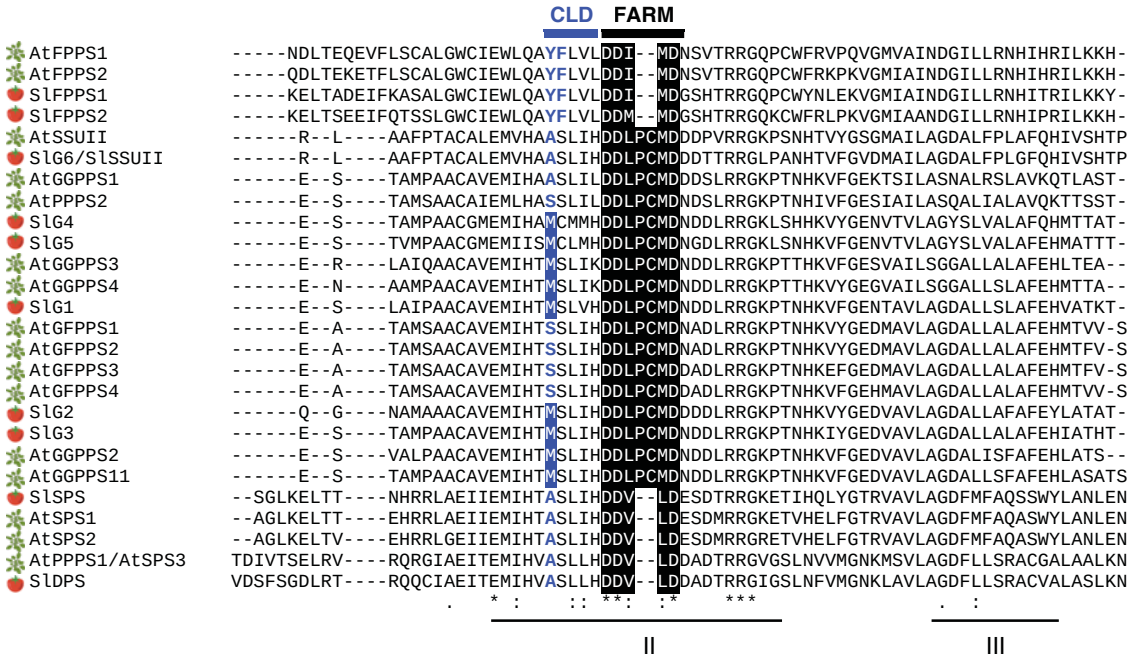
Zhu XF, Suzuki K, Okada K, Tanaka K, Nakagawa T, Kawamukai M and Matsuda K (1997b). Cloning and functional expression of a novel geranylgeranyl pyrophosphate synthase gene from *Arabidopsis thaliana* in *Escherichia coli*. Plant Cell Physiol. **38**: 357–61.

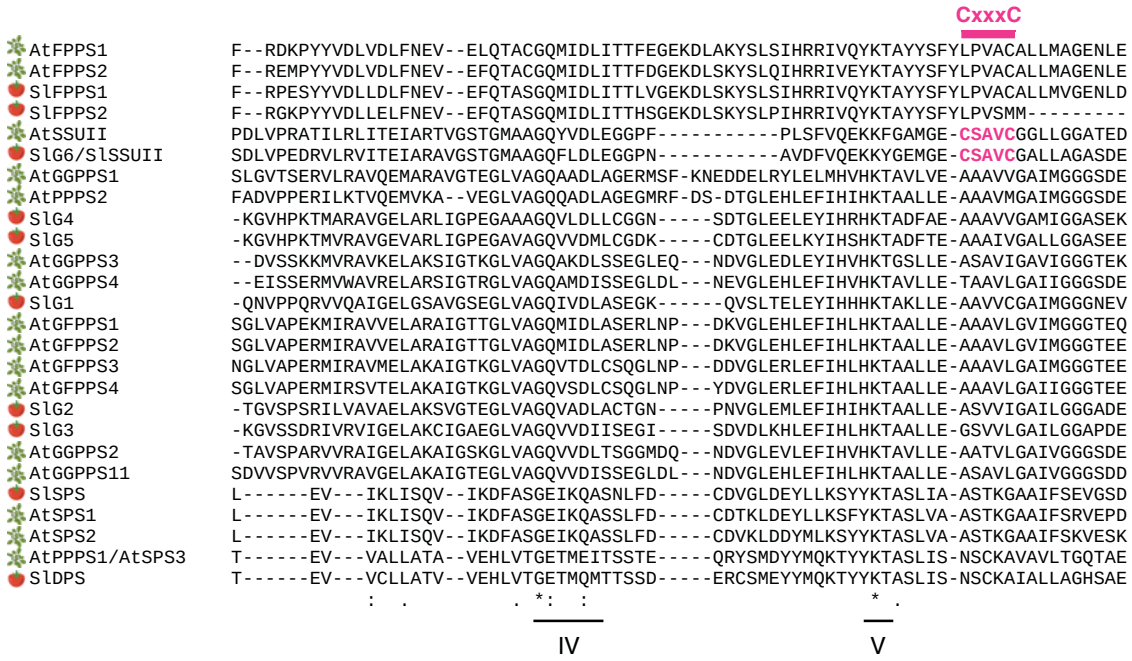
Zouine M, Maza E, Djari A, Lauvernier M, Frasse P, Smouni A, Pirrello J and Bouzayen M (2017). TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. Plant J. **92**: 727–735.

7. Supplemental Information






CLD
FARM
AtFPPS1 -----NDLTEQEVFLSCALGWCIEWLQAYFLVLDDI--MDNSVTRRGQPCWFRVPQVGMVAINDGILLRNHIHRILKKH-
AtFPPS2 -----QDLTEKETFLSCALGWCIEWLQAYFLVLDDI--MDNSVTRRGQPCWFRKPKVGMIAINDGILLRNHIHRILKKH-
SlFPPS1 -----KELTADEIFKASALGWCIEWLQAYFLVLDDI--MDGSHTRRGQPCWYNLEKVGMIAINDGILLRNHITRILKKY-
SlFPPS2 -----KELTSEEIFQTSSLGWCIEWLQAYFLVLDDM--MDGSHTRRGQKCWFRLPKVGMIAANDGILLRNHIPRILKKH-
AtSSUII ------R--L----AAFPTACALEMVHAASLIHDDLPCMDDDPVRRGKPSNHTVYGSGMAILAGDALFPLAFQHIVSHTP
SlG6/SlSSUII ------R--L----AAFPTACALEMVHAASLIHDDLPCMDDDTTRRGLPANHTVFGVDMAILAGDALFPLGFQHIVSHTP
AtGGPPS1 ------E--S----TAMPAACAVEMIHAASLILDDLPCMDDDSLRRGKPTNHKVFGEKTSILASNALRSLAVKQTLAST-
AtPPPS2 ------E--S----TAMSAACAIEMLHASSLILDDLPCMDNDSLRRGKPTNHIVFGESIAILASQALIALAVQKTTSST-
SlG4 ------E--S----TAMPAACGMEMIHAMCMMHDDLPCMDNDDLRRGKLSHHKVYGENVTVLAGYSLVALAFQHMTTAT-
SlG5 ------E--S----TVMPAACGMEMIISMCLMHDDLPCMDNGDLRRGKLSNHKVFGENVTVLAGYSLVALAFEHMATTT-
AtGGPPS3 ------E--R----LAIQAACAVEMIHTMSLIKDDLPCMDNDDLRRGKPTTHKVFGESVAILSGGALLALAFEHLTEA--
AtGGPPS4 ------E--N----AAMPAACAVEMIHTMSLIKDDLPCMDNDDLRRGKPTTHKVYGEGVAILSGGALLSLAFEHMTTA--
SlG1 ------E--S----LAIPAACAVEMIHTMSLVHDDLPCMDNDDLRRGKPTNHKVFGENTAVLAGDALLSLAFEHVATKT-
AtGFPPS1 ------E--A----TAMSAACAVEMIHTSSLIHDDLPCMDNADLRRGKPTNHKVYGEDMAVLAGDALLALAFEHMTVV-S
AtGFPPS2 ------E--A----TAMSAACAVEMIHTSSLIHDDLPCMDNADLRRGKPTNHKVYGEDMAVLAGDALLALAFEHMTFV-S
AtGFPPS3 ------E--A----TAMSAACAVEMIHTSSLIHDDLPCMDDADLRRGKPTNHKEFGEDMAVLAGDALLALAFEHMTFV-S
AtGFPPS4 ------E--A----TAMSAACAVEMIHTSSLIHDDLPCMDDADLRRGKPTNHKVFGEHMAVLAGDALLALAFEHMTVV-S
SlG2 ------Q--G----NAMAAACAVEMIHTMSLIHDDLPCMDDDLRRGKPTNHKVYGEDVAVLAGDALLAFAFEYLATAT-
SlG3 ------E--S----TAMPAACAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKIYGEDVAVLAGDALLALAFEHIATHT-
AtGGPPS2 ------E--S----VALPAACAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALISFAFEHLATS--
AtGGPPS11 ------E--S----TAMPAACAVEMIHTMSLIHDDLPCMDNDDLRRGKPTNHKVFGEDVAVLAGDALLSFAFEHLASATS
SlSPS --SGLKELTT----NHRRLAEIIEMIHTASLIHDDV--LDESDTRRGKETIHQLYGTRVAVLAGDFMFAQSSWYLANLEN
AtSPS1 --AGLKELTT----EHRRLAEIIEMIHTASLIHDDV--LDESDMRRGKETVHELFGTRVAVLAGDFMFAQASWYLANLEN
AtSPS2 --AGLKELTV----EHRRLGEIIEMIHTASLIHDDV--LDESDMRRGRETVHELFGTRVAVLAGDFMFAQASWYLANLEN
AtPPPS1/AtSPS3 TDIVTSELRV----RQRGIAEITEMIHVASLLHDDV--LDDADTRRGVGSLNVVMGNKMSVLAGDFLLSRACGALAALKN
SlDPS VDSFSGDLRT----RQQCIAEITEMIHVASLLHDDV--LDDADTRRGIGSLNFVMGNKLAVLAGDFLLSRACVALASLKN
II
III


CxxxC
AtFPPS1 F--RDKPYYVDLVDLFNEV--ELQTACGQMIDLITTFEGEKDLAKYSLSIHRRIVQYKTAYYSFYLPVACALLMAGENLE
AtFPPS2 F--REMPYYVDLVDLFNEV--EFQTACGQMIDLITTFDGEKDLSKYSLQIHRRIVEYKTAYYSFYLPVACALLMAGENLE
SlFPPS1 F--RPESYYVDLLDLFNEV--EFQTASGQMIDLITTLVGEKDLSKYSLSIHRRIVQYKTAYYSFYLPVACALLMVGENLD
SlFPPS2 F--RGKPYYVDLLELFNEV--EFQTASGQMIDLITTHSGEKDLSKYSLPIHRRIVQYKTAYYSFYLPVSMM---------
AtSSUII PDLVPRATILRLITEIARTVGSTGMAAGQYVDLEGGPF-----------PLSFVQEKKFGAMGE-CSAVCGGLLGGATED
SlG6/SlSSUII SDLVPEDRVLRVITEIARAVGSTGMAAGQFLDLEGGPN-----------AVDFVQEKKYGEMGE-CSAVCGALLAGASDE
AtGGPPS1 SLGVTSERVLRAVQEMARAVGTEGLVAGQAADLAGERMSF-KNEDDELRYLELMHVHKTAVLVE-AAAVVGAIMGGGSDE
AtPPPS2 FADVPPERILKTVQEMVKA--VEGLVAGQQADLAGEGMRF-DS-DTGLEHLEFIHIHKTAALLE-AAAVMGAIMGGGSDE
SlG4 -KGVHPKTMARAVGEARLIGPEGAAAGQVLDLLCGGN-----SDTGLEELEYIHRHKTADFAE-AAAVVGAMIGGASEK
SlG5 -KGVHPKTMVRAVGEVARLIGPEGAVAGQVVDMLCGDK-----CDTGLEELKYIHSHKTADFTE-AAAIVGALLGGASEE
AtGGPPS3 --DVSSKKMVRAVKELAKSIGTKGLVAGQAKDLSSEGLEQ---NDVGLEDLEYIHVHKTGSLLE-ASAVIGAVIGGGTEK
AtGGPPS4 --EISSERMVWAVRELARSIGTRGLVAGQAMDISSEGLDL---NEVGLEHLEFIHVHKTAVLLE-TAAVLGAIIGGGSDE
SlG1 -QNVPPQRVVQAIGELGSAVGSEGLVAGQIVDLASEGK------QVSLTELEYIHHHKTAKLLE-AAVVCGAIMGGGNEV
AtGFPPS1 SGLVAPEKMIRAVVELARAIGTTGLVAGQMIDLASERLNP---DKVGLEHLEFIHLHKTAALLE-AAAVLGVIMGGGTEQ
AtGFPPS2 SGLVAPERMIRAVVELARAIGTTGLVAGQMIDLASERLNP---DKVGLEHLEFIHLHKTAALLE-AAAVLGVIMGGGTEE
AtGFPPS3 NGLVAPERMIRAVMELAKAIGTKGLVAGQVTDLCSQGLNP---DDVGLERLEFIHLHKTAALLE-AAAVLGAIMGGGTEE
AtGFPPS4 SGLVAPERMIRSVTELAKAIGTKGLVAGQVSDLCSQGLNP---YDVGLERLEFIHLHKTAALLE-AAAVLGAIIGGGTEE
SlG2 -TGVSPSRILVAVAELAKSVGTEGLVAGQVADLACTGN-----PNVGLEMLEFIHIHKTAALLE-ASVVIGAILGGGADE
SlG3 -KGVSSDRIVRVIGELAKCIGAEGLVAGQVVDIISEGI-----SDVDLKHLEFIHLHKTAALLE-GSVVLGAILGGAPDE
AtGGPPS2 -TAVSPARVVRAIGELAKAIGSKGLVAGQVVDLTSGGMDQ---NDVGLEVLEFIHVHKTAVLLE-AATVLGAIVGGGSDE
AtGGPPS11 SDVVSPVRVVRAVGELAKAIGTEGLVAGQVVDISSEGLDL---NDVGLEHLEFIHLHKTAALLE-ASAVLGAIVGGGSDD
SlSPS L------EV---IKLISQV--IKDFASGEIKQASNLFD-----CDVGLDEYLLKSYYKTASLIA-ASTKGAAIFSEVGSD
AtSPS1 L------EV---IKLISQV--IKDFASGEIKQASSLFD-----CDTKLDEYLLKSFYKTASLVA-ASTKGAAIFSRVEPD
AtSPS2 L------EV---IKLISQV--IKDFASGEIKQASSLFD-----CDVKLDDYMLKSYYKTASLVA-ASTKGAAIFSKVESK
AtPPPS1/AtSPS3 T------EV---VALLATA--VEHLVTGETMEITSSTE-----QRYSMDYYMQKTYYKTASLIS-NSCKAVAVLTGQTAE
SlDPS T------EV---VCLLATV--VEHLVTGETMQMTTSSD-----ERCSMEYYMQKTYYKTASLIS-NSCKAIALLAGHSAE
IV
V

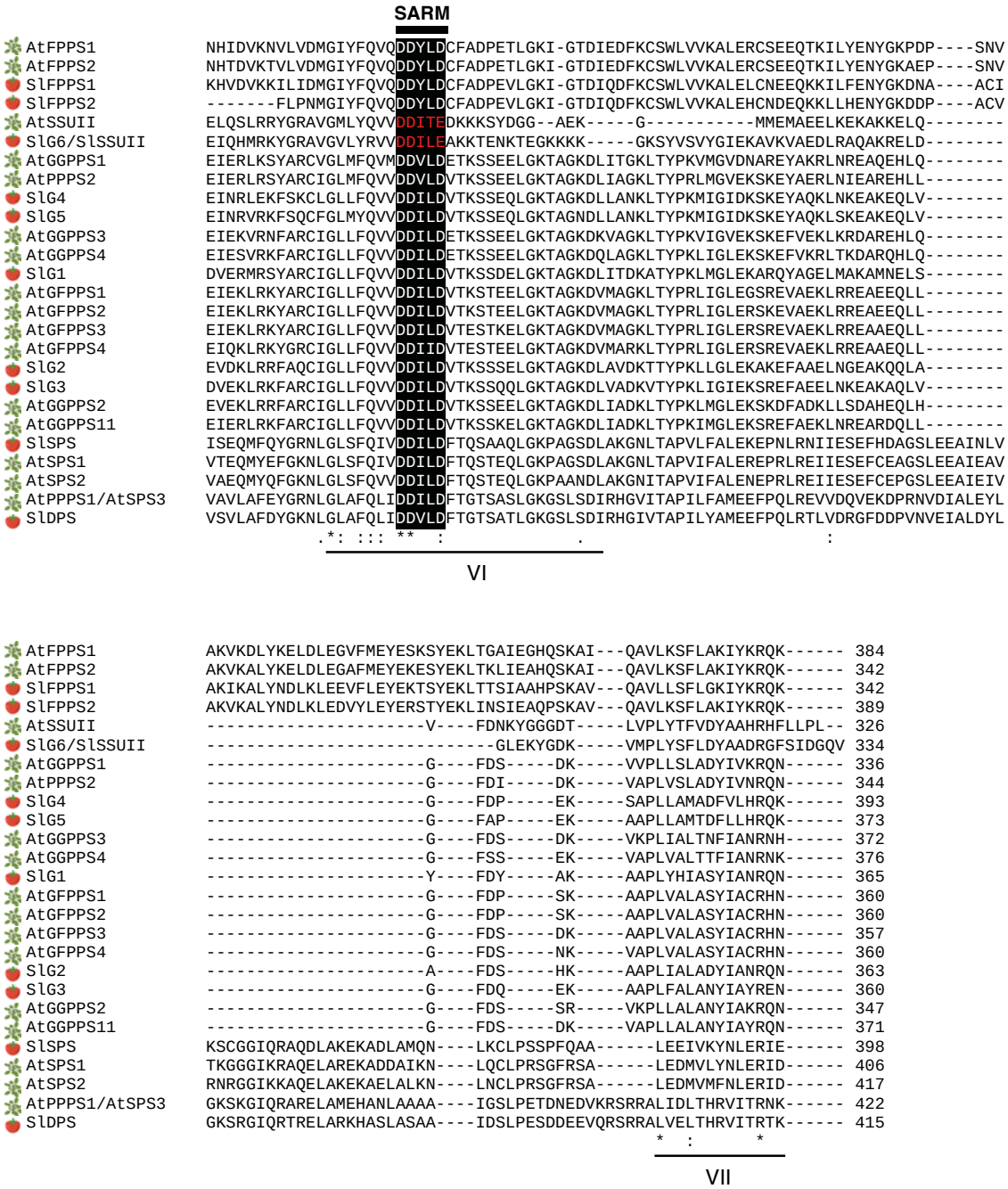


Figure S1. Sequence alignment of Arabidopsis and tomato PT proteins. Multiple sequence alignment was carried out using *Clustal Omega* (https://www.ebi.ac.uk/Tools/msa/clustalo/) with default settings**.** The predicted targeting peptides were removed for the protein alignment. Numbers at the beginning of each sequence indicate the first amino acid aligned. Numbers at the end of each sequence indicate protein length. Icons represent the species. The seven domains highly conserved in PTs are underlined. The interaction CxxxC (x = any hydrophobic residue) motifs are highlighted in pink. The catalytic motifs FARM (first aspartate-rich motif) and SARM (second-aspartate rich motif) are boxed in black. Mutated SARM signatures are indicated in red. The CLD (chain-length determination) region is marked in blue, highlighting the residues at positions –4 and /or -5 relative to the FARM which have been shown to be critical for the final length of the synthesized product. GGPPS enzymes have a methionine (M) in that position (boxed in blue).

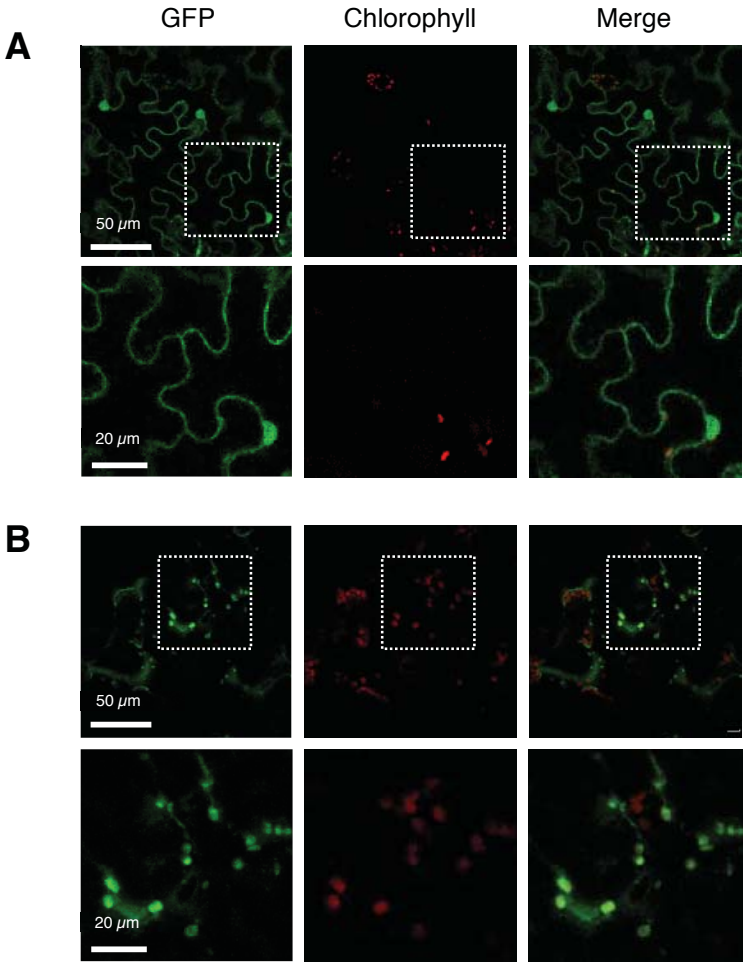


Figure S2. Dual subcellular localization of the SlG4-GFP fusion protein. Confocal microscopy images of *N. benthamiana* leaf cells transiently expressing the SlG4 isoform fused to GFP are shown. The first column shows green fluorescence from GFP, the second shows red autofluorescence from chlorophyll and the third represents the overlay of the two channels. Section **(A)** shows the most commonly observed localization of the SlG4-GFP protein (in the cytosol), whereas section **(B)** shows plastidial localization of the protein in some cells. White bars indicate the magnification of the images.

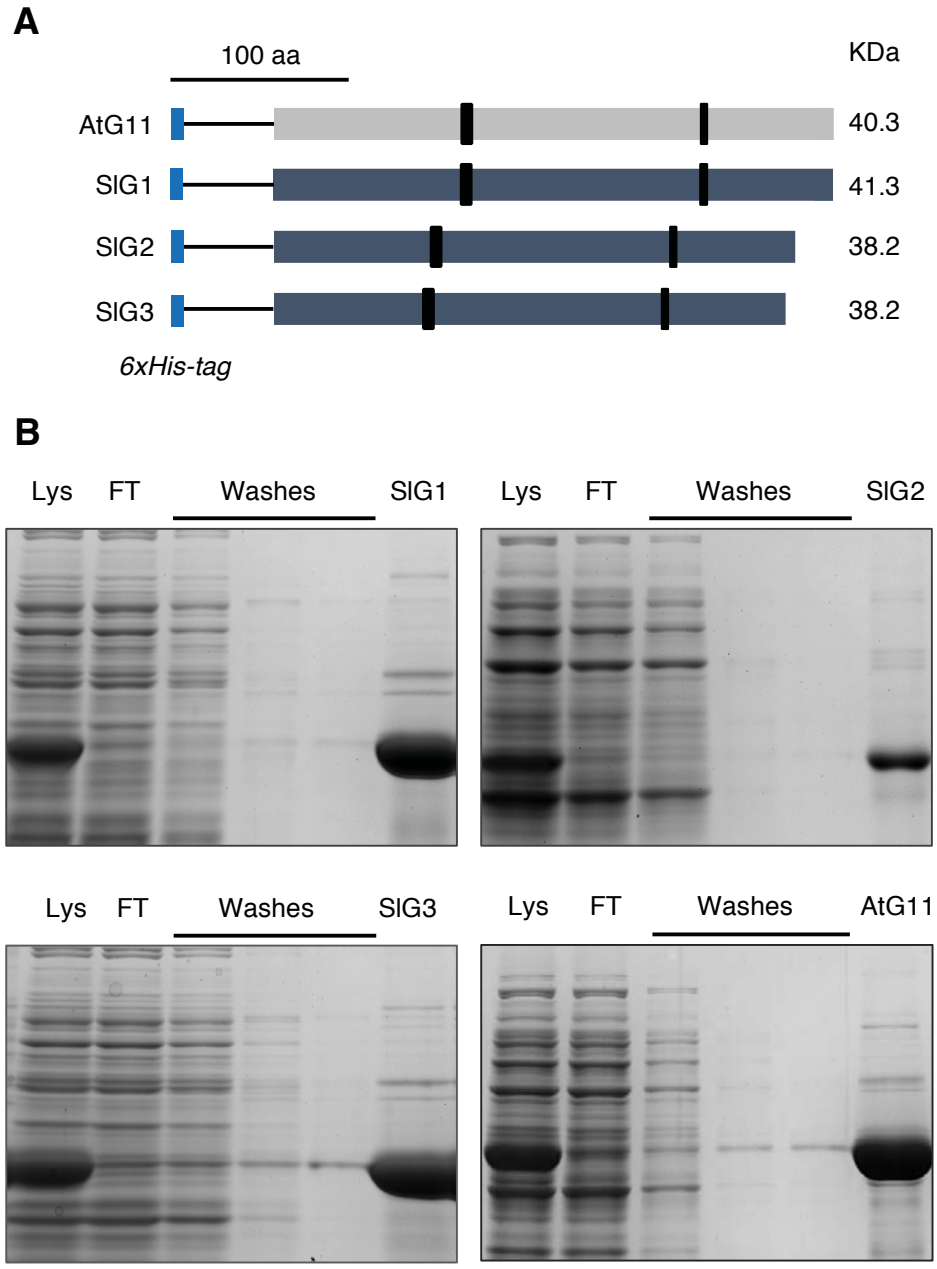


Figure S3. Purification of recombinant 6xHis-tagged GGPPS proteins. (A) Schematic representation of the purified GGPPS enzyme versions (lacking the predicted plastid-targeting peptide and fused to a 6xHis-tag in the N-terminal end). **(B)** Coomassie-Blue staining gels showing the purification of the indicated truncated proteins. The enzymes were purified as described (Barja et al., 2019, *see* **Annex 3**) from soluble lysates (Lys) of *E. coli Rosetta* cells overproducing the recombinant protein. Lysates were separately incubated with Ni-NTA beads and the staining of the flow-through (FT) shows that the recombinant protein was retained in the Ni-NTA column. After several washes with 20 mM imidazole to remove non-specific proteins attached to the column, His-tagged enzymes were eluted using 150 mM imidazole. Purified proteins were then desalted, quantified and stored with glycerol 40% at < -20 ºC.

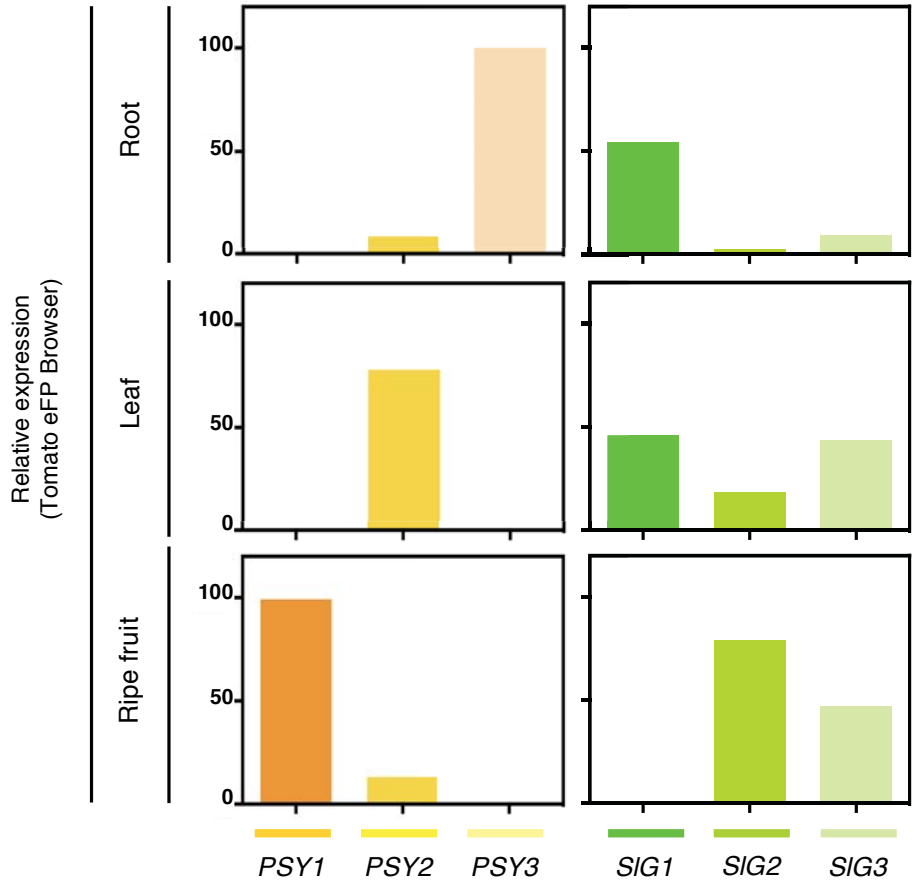


Figure S4. Transcript levels of *PSY* and plastidial *GGPPS* paralog genes in different tomato tissues. Expression data were retrieved from the *Tomato eFP Browser* database (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Graphs show the transcript levels of *PSY1-3* and *SlG1-3 in* roots, leaves and red ripe fruits. Results are represented as the percentage of expression of each gene per tissue. Thus, 100% of the expression of each gene represents the sum of the expression in each selected tissue. *PSY1* gene, for example, is only expressed in fruits (100% of expression) while around the 75% of *SlG2* expression is detected in fruits and the 25% in leaves. *See* Table 1 for accessions.

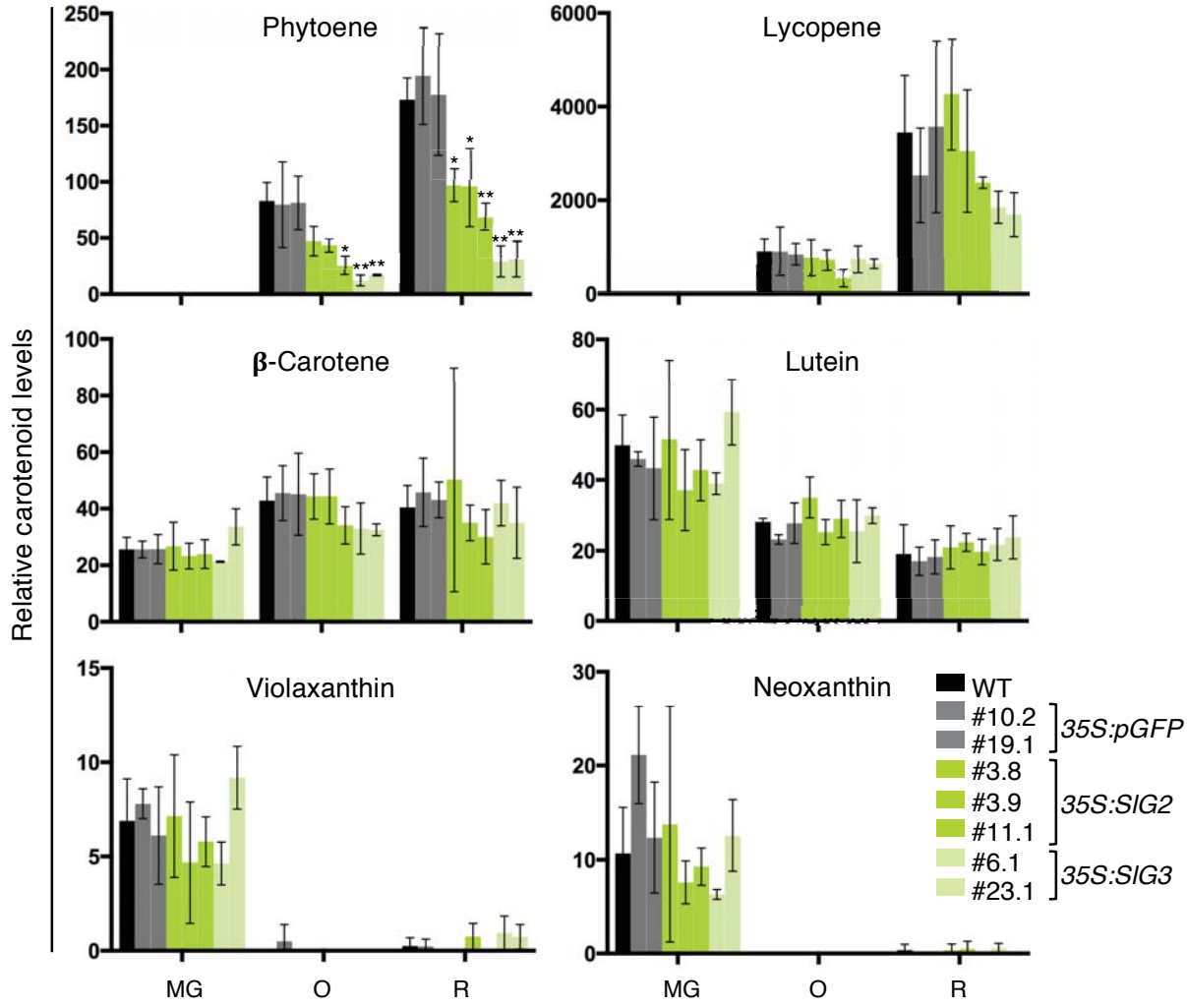


Figure S5. Carotenoid levels in fruits from T2 transgenic lines. Levels of individual carotenoids were measured in mature green (MG), orange (O) and red (R) fruit pericarp samples of WT, *35S:pGFP*, *35S:SlG2* and *35S:SlG3* lines. At least two independent transgenic lines per construct were used for the HPLC analyses. The graphs represent the mean ± SD of three independent biological replicates (n=3). Values represent percentage relative to total carotenoid contents of WT MG fruits. Asterisks indicate significant differences among means relative to WT samples in each ripening stage (one-way ANOVA with Dunnett's multiple comparisons test, *p<0.05, **p<0.01).

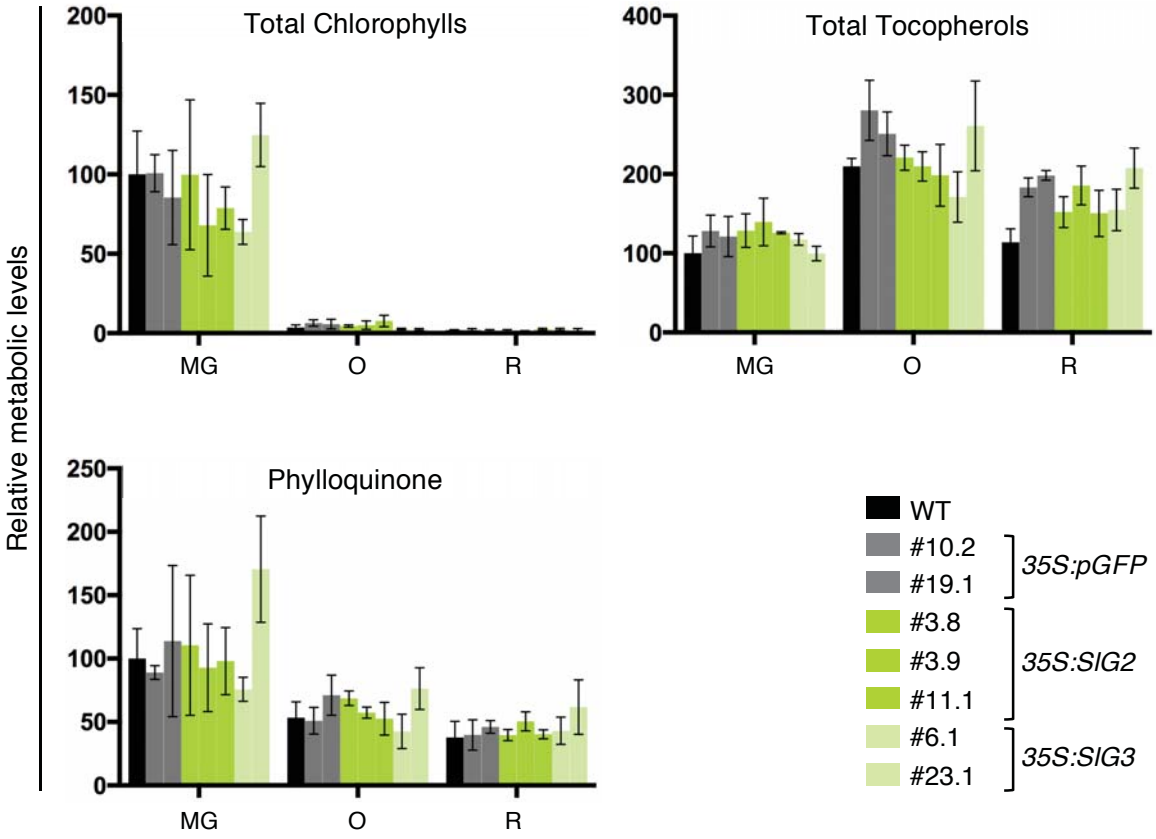


Figure S6. Chlorophyll, tocopherol and phylloquinone levels in fruits from T2 transgenic lines. Levels of total chlorophylls, total tocopherols and phylloquinone were analyzed in the pericarp of mature green (MG), orange (O) and red fruits of WT, *35S:pGFP*, *35S:SlG2* and *35S:SlG3* lines. The same samples as those used for carotenoid analysis in Figure S5 were used. The graphs represent the mean ± SD of three independent biological replicates (n=3). Values represent percentage of total contents relative to WT levels in MG fruits. No significant differences among means relative to WT samples in each ripening stage were found (one-way ANOVA with Dunnett's multiple comparisons test).

Table S1. List of Arabidopsis (At) and tomato (Sl) prenyltranferases used for sequence and phylogenetic analyses. Abbreviations: PT (prenyltransferase); FPP(S), farnesyl diphosphate (synthase); GGPP(S), geranylgeranyl diphosphate (synthase); GFPP(S), geranylfarnesyl diphosphate (synthase); SPS; solanesyl diphosphate synthase; PPP, polyprenyl diphosphate synthase; DPS, decaprenyl diphosphate syntahse; SSU, small subunit; C, cytosol; ER, endoplasmic reticulum; M, mitochondria; P, plastid

PT	Gene ID	Protein length (aa)	Predicted TP (aa)	Subcellular localization	Reference
FPP synthases					
AtFPPS1	At5g47770	384	41	M+C	Cunillera et al., 1996; 1997
AtFPPS2	At4g17190	342	-	C	Cunillera et al., 1996; Keim et al., 2012
SlFPPS1	Solyc12g015860	342	-	C**	Gaffe et al., 2000
SlFPPS2	Solyc10g005810/20	389	114	M**	Gaffe et al., 2000
GGPP synthases					
AtGGPPS1	At1g49530	336	29	M	Beck et al., 2013; Nagel et al., 2015; Wang et al., 2016
AtGGPPS2	At2g18620	347	39	P	Beck et al., 2013; Nagel et al., 2015; Wang et al., 2016
AtGGPPS3	At2g18640	372	22	C-ER	Beck et al., 2013; Wang et al., 2016
AtGGPPS4	At2g23800	376	24	C-ER	Beck et al., 2013; Wang et al., 2016
AtGGPPS11	At4g36810	371	56	P+C	Beck et al., 2013; Nagel et al., 2015; Ruiz-Sola et al. 2016b (**Annex 2**)
SlG1*	Solyc11g011240	365	43	P	Ament et al., 2006; This study
SlG2*	Solyc04g079960	363	63	P	Ament et al., 2006; This study
SlG3*	Solyc02g085700	360	65	P	This study
SlG4**	Solyc02g085710	393	34	C-ER**	This study
SlG5**	Solyc02g085720	373	56	M**	This study
GFPP synthases					
AtGFPPS1	At3g14530	360	39	P	Beck et al., 2013; Nagel et al., 2015; Wang et al. 2016
AtGFPPS2	At3g14550	360	39	P	Beck et al., 2013; Nagel et al., 2015; Wang et al. 2016
AtGFPPS3	At3g29430	357	37	P	Beck et al., 2013; Nagel et al., 2015; Wang et al. 2016
AtGFPPS4	At3g32040	360	72	P	Beck et al., 2013; Nagel et al., 2015; Wang et al. 2016
Long-chain PTs					
AtSPS1	At1g78510	406	71	C-ER	Jun et al., 2004; Hirooka et al., 2005
AtSPS2	At1g17050	417	60	P	Jun et al., 2004; Hirooka et al., 2005
AtPPPS1/AtSPS3	At2g34630	422	88	P	Ducluzeau et al., 2012
AtPPPS2	At3g20160	344	40	P	Beck et al., 2013; Wang et al., 2016
SlSPS	Solyc07g061990	398	44	P	Jones et al., 2013
SlDPS	Solyc08g023470	415	55	M	Jones et al., 2013
Type II SSUs					
AtSSUII	At4g38460	326	33	P	Wang and Dixon, 2009; Chen et al., 2015
SlG6 / SlSSUII**	Solyc09g008920	334	17	P**	This study

*Activity reported in this work

**Prediction. Needs experimental confirmation

Table S2. List of plastidial isoprenoid-related genes used for the tomato GGPPS GCN analyses. Arabidopsis genes were used as queries to search for tomato homologs in *PLAZA 4.0* (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_dicots/) and *Phytozome* (https://phytozome.jgi.doe.gov/pz/portal.html#). Genes are organized by pathways.

Input gene	Gene ID	Description
MEP pathway		
DXS1	Solyc01g067890	1-deoxy-D-xylulose-5-phosphate (DXP) synthase 1
DXS2	Solyc11g010850	1-deoxy-D-xylulose-5-phosphate (DXP) synthase 2
DXS3a	Solyc01g028900	1-deoxy-D-xylulose-5-phosphate (DXP) synthase 3a
DXS3b	Solyc08g066950	1-deoxy-D-xylulose-5-phosphate (DXP) synthase 3b
DXR1	Solyc03g114340	DXP reductoisomerase 1
DXR2	Solyc06g060860	DXP reductoisomerase 2
MCT	Solyc01g102820	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase
CMK	Solyc01g009010	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase
MDS	Solyc08g081570	2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase
HDS	Solyc11g069380	4-hydroxy-3-methylbut-2-enyl diphosphate synthase
HDR	Solyc01g109300	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
IDI1	Solyc08g075390	Isopentenyl diphosphate isomerase 1
IDI2	Solyc05g055760	Isopentenyl diphosphate isomerase 2
IDI3	Solyc04g056390	Isopentenyl diphosphate isomerase 3
GGPP biosynthesis		
G1 (SlG1)*	Solyc11g011240	GGPP synthase 1
G2 (SlG2)*	Solyc04g079960	GGPP synthase 2
G3 (SlG3)*	Solyc02g085700	GGPP synthase 3
*G4 (SlG4**)*	Solyc02g085710	Putative GGPP synthase 4
*G5 (SlG5**)*	Solyc02g085720	Putative GGPP synthase 5
Carotenoid biosynthesis		
PSY1	Solyc03g031860	Phytoene synthase 1
PSY2	Solyc02g081330	Phytoene synthase 2
PSY3	Solyc01g005940	Phytoene synthase 3
PDS	Solyc03g123760	Phytoene desaturase
ZDS	Solyc01g097810	Zeta-carotene desaturase
Z-ISO	Solyc12g098710	15-cis-zeta-carotene isomerase
CRTISO1	Solyc10g081650	Carotenoid isomerase 1
CRTISO2	Solyc05g010180	Carotenoid isomerase 2
LCY-B1	Solyc04g040190	Lycopene beta-cyclase 1
LCY-B2	Solyc10g079480	Lycopene beta-cyclase 2
LCY-E	Solyc12g008980	Lycopene epsilon-cyclase
BCH1	Solyc06g036260	Carotene beta-hydroxylase 1
BCH2	Solyc03g007960	Carotene beta-hydroxylase 2
CYP97A3	Solyc04g051190	Carotene hydroxylase (cytochrome P450)
CYP97B3	Solyc05g016330	Carotene hydroxylase (cytochrome P450)
CYP97C1	Solyc10g083790	Carotene hydroxylase (cytochrome P450)
ZEP1	Solyc06g060880	Zeaxanthin epoxidase 1
ZEP2	Solyc02g090890	Zeaxanthin epoxidase 2
VDE	Solyc04g050930	Violaxanthin de-epoxidase
NSY1	Solyc02g089050	Neoxanthin synthase 1
NSY2/ABA4	Solyc02g063170	Neoxanthin synthase 2
NSY3	Solyc03g034240	Neoxanthin synthase 3
Gibberellin (GA) biosynthesis		
CPS1	Solyc06g084240	Ent-copalyl diphosphate synthase 1
CPS2	Solyc08g005710	Ent-copalyl diphosphate synthase 2
CPS3	Solyc09g065230	Ent-copalyl diphosphate synthase 3
KS1a	Solyc07g066670	Ent-kaurene synthase 1a
KS1b	Solyc08g005720	Ent-kaurene synthase 1b
GA3	Solyc04g083160	Ent-kaurene oxidase

KAO1	Solyc01g080900	Ent-kaurenoate oxidase 1
KAO2	Solyc08g007050	Ent-kaurenoate oxidase 2
KAO3	Solyc12g006460	Ent-kaurenoate oxidase 3
KAO4	Solyc10g007860	Ent-kaurenoate oxidase 4
GA20ox1	Solyc03g006880	Gibberellin 20-oxidase 1
GA20ox2	Solyc09g009110	Gibberellin 20-oxidase 2
GA20ox3	Solyc11g072310	Gibberellin 20-oxidase 3
GA20ox4	Solyc06g035530	Gibberellin 20-oxidase 4
GA20ox5	Solyc01g093980	Gibberellin 20-oxidase 5
GA20ox6	Solyc06g050110	Gibberellin 20-oxidase 6
GA20ox7	Solyc11g013360	Gibberellin 20-oxidase 7
GA3ox1	Solyc06g066820	Gibberellin 3-oxidase 1
GA3ox2	Solyc03g119910	Gibberellin 3-oxidase 2
GA3ox3	Solyc00g007180	Gibberellin 3-oxidase 3
GA3ox4	Solyc01g058250	Gibberellin 3-oxidase 4
GA3ox5	Solyc05g052740	Gibberellin 3-oxidase 5
Orange proteins		
Or1	Solyc03g093830	Orange 1
Or2	Solyc09g010110	Orange 2
Strigolactone (SL) biosynthesis		
MAX3 (CCD7)	Solyc01g090660	More axillary branches 3 (carotenoid cleavage dioxygenase 7)
MAX4 (CCD8)	Solyc08g066650	More axillary branches 4 (carotenoid cleavage dioxygenase 8)
MAX1	Solyc08g062950	More axillary branches 1
Abscisic acid (ABA) biosynthesis		
ABA1/ZEP	Solyc02g090890	Zeaxanthin epoxidase 2
ABA4/NSY2	Solyc02g063170	Neoxanthin synthase 2
NSY4	Solyc02g086050	Neoxanthin synthase 4
NSY3	Solyc03g034240	Neoxanthin synthase 3
NSY5	Solyc06g074240	Neoxanthin synthase 5
NCED2	Solyc08g016720	9-cis-epoxycarotenoid dioxygenase 2
NCED3	Solyc07g056570	9-cis-epoxycarotenoid dioxygenase 3
NCED6	Solyc05g053530	9-cis-epoxycarotenoid dioxygenase 6
ABA2a	Solyc04g071940	Xanthoxin dehydrogenase
ABA2b	Solyc04g071960	Xanthoxin dehydrogenase
ABA2c	Solyc10g085380	Xanthoxin dehydrogenase
ABA2d	Solyc11g018600	Xanthoxin dehydrogenase
AAO3a	Solyc11g065920	Abscisic aldehyde oxidase 3a
AAO3b	Solyc11g065930	Abscisic aldehyde oxidase 3b
ABA3	Solyc07g066480	Molybdenum cofactor sulfurare
CYP707A1	Solyc04g078900	ABA 8'-hydorxylase (ABA8ox)
CYP707A2	Solyc08g075320	ABA 8'-hydorxylase (ABA8ox)
CYP707A3a	Solyc01g108210	ABA 8'-hydorxylase (ABA8ox)
CYP707A3b	Solyc08g005610	ABA 8'-hydorxylase (ABA8ox)
CYP707A3c	Solyc04g071150	ABA 8'-hydorxylase (ABA8ox)
CYP707A3d	Solyc04g080650	ABA 8'-hydorxylase (ABA8ox)
Carotenoid degradation		
CCD1A	Solyc01g087250	Carotenoid cleavage dioxygenase 1A
CCD1B	Solyc01g087260	Carotenoid cleavage dioxygenase 1B
CCD4A	Solyc08g075480	Carotenoid cleavage dioxygenase 4A
CCD4B	Solyc08g075490	Carotenoid cleavage dioxygenase 4B
MAX3 (CCD7)	Solyc01g090660	More axillary branches 3 (carotenoid cleavage dioxygenase 7)
MAX4 (CCD8)	Solyc08g066650	More axillary branches 4 (carotenoid cleavage dioxygenase 8)
NCDE2	Solyc08g016720	9-cis-epoxycarotenoid dioxygenase 2
NCDE3	Solyc07g056570	9-cis-epoxycarotenoid dioxygenase 3
NCDE6	Solyc05g053530	9-cis-epoxycarotenoid dioxygenase 6
CCDX	Solyc08g066720	Carotenoid cleavage dioxygenase
Chlorophyll biosynthesis		
HEMA1	Solyc04g076870	Glutamyl tRNA reductase 1

HEMA2	Solyc01g106390	Glutamyl tRNA reductase 2
HEMA3	Solyc01g089840	Glutamyl tRNA reductase 3
GSA	Solyc04g009200	Glutamate-1-semialdehyde 2, 1 aminomutase
HEMB	Solyc08g069030	5-aminolevulinate dehydratase
HEMC	Solyc07g066470	Porphobilinogen deaminase
HEMD	Solyc04g079320	Uroporphyrinogen III synthase
HEME1	Solyc10g007320	Uroporphyrinogen III decarboxylase 1
HEME2	Solyc06g048730	Uroporphyrinogen III decarboxylase 2
HEMF	Solyc10g005110	Coproporphyrinogen III oxidase
HEMG1	Solyc01g079090	Protoporphyrinogen IX oxidase 1
HEMG2	Solyc03g005080	Protoporphyrinogen IX oxidase 2
CHLH	Solyc04g015750	Mg chelatase H subunit
CHLI	Solyc10g008740	Mg chelatase I subunit
CHLD	Solyc04g015490	Mg chelatase D subunit
CHLM	Solyc03g118240	Mg protoporphyrin IX methyltransferase
CHL27-CRD	Solyc10g077040	Mg protoporphyrin IX monomethylester cyclase
DVR	Solyc01g067290	Divinyl reductase
PORA	Solyc12g013710	Protochlorophyllide reductase A
PORB	Solyc07g054210	Protochlorophyllide reductase B
PORC	Solyc10g006900	Protochlorophyllide reductase C
CHLG1	Solyc05g024190	Chlorophyll synthase 1
CHLG2	Solyc09g014760	Chlorophyll synthase 2
CAO1	Solyc06g060310	Chlorophyll a oxigenase 1
CAO2	Solyc11g012850	Chlorophyll a oxigenase 2
GGDR1	Solyc01g088310	Geranylgeranyl reductase 1
GGDR2	Solyc03g115980	Geranylgeranyl reductase 2
Chlorophyll degradation		
CHL1	Solyc09g065620	Chlorophyllase 1
CHL2a	Solyc06g053980	Chlorophyllase 2a
CHL2b	Solyc09g082600	Chlorophyllase 2b
CHL2c	Solyc12g005300	Chlorophyllase 2c
CLD1	Solyc02g070490	Chlorophyll dephytyllase
SGR1	Solyc08g080090	Stay-green 1, Non-yellowing 1
SGR2	Solyc12g056480	Stay-green 2, Non-yellowing 2
SGR3	Solyc04g063240	Stay-green 3, Non-yellowing 3
PPH	Solyc01g088090	Pheophytine pheophorbide hydrolase
PAO1	Solyc11g066440	Pheophorbide a oxygenase 1
PAO2	Solyc04g040160	Pheophorbide a oxygenase 2
PAO3	Solyc12g096550	Pheophorbide a oxygenase 3
RCCR	Solyc03g044470	Red chlorophyll catabolite reductase
PK1	Solyc03g071720	Phytol kinase 1
PK2	Solyc09g018510	Phytol kinase 2
NYC1	Solyc07g024000	Chlorophyll b reductase
CBR/NOL	Solyc05g032660	Chlorophyll b reductase
HCAR	Solyc09g091100	Chlorophyll a reductase
Tocopherol biosynthesis		
GGDR1	Solyc01g088310	Geranylgeranyl reductase 1
GGDR2	Solyc03g115980	Geranylgeranyl reductase 2
HPPD1/PDS1a	Solyc05g041200	4-hydroxyphenyl pyruvate dioxygenase 1
HPPD2/PDS1b	Solyc07g045050	4-hydroxyphenyl pyruvate dioxygenase 2
VTE2	Solyc07g017770	Homogentisate phytyltransferase
VTE3a	Solyc03g005230	MPBQ/MSBQ methyltransferase
VTE3b	Solyc09g065730	MPBQ/MSBQ methyltransferase
VTE1	Solyc08g068570	Tocopherol cyclase
VTE4a	Solyc08g076360	Delta/gamma-tocopherol methyltransferase
VTE4b	Solyc04g063230	Delta/gamma-tocopherol methyltransferase
VTE4c	Solyc08g077240	Delta/gamma-tocopherol methyltransferase
VTE4d	Solyc03g116150	Delta/gamma-tocopherol methyltransferase
Phylloquinone biosynthesis		
ICS1/MENF	Solyc06g071030	Isochorismate synthase 1

PHYLLO/MEND1	Solyc04g005190	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase 1
PHYLLO/MEND2	Solyc04g005200	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase 2
PHYLLO/MEND3	Solyc04g005180	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase 3
AAE14	Solyc02g069920	O-succinylbenzoyl-CoA ligase
DHNS/MENB	Solyc05g005180	1,4-dihydroxy-2-naphthoyl-CoA synthase
DHNAT1	Solyc02g078410	1,4-dihydroxy-2-naphthoyl-CoA thioesterase 1
DHNAT2	Solyc03g006440	1,4-dihydroxy-2-naphthoyl-CoA thioesterase 2
DHNAT3	Solyc03g006450	1,4-dihydroxy-2-naphthoyl-CoA thioesterase 3
ABC4	Solyc01g105460	DHNA phytyl transferase
MENG	Solyc12g019010	2-phytyl-1,4-naphthoquinone methyltransferase
Plastoquinone biosynthesis		
HPPD1/PDS1a	Solyc05g041200	4-hydroxyphenyl pyruvate dioxygenase 1
HPPD2/PDS1b	Solyc07g045050	4-hydroxyphenyl pyruvate dioxygenase 2
SPS1	Solyc07g061990	Solanesyl diphosphate synthase 1
PDS2/HST	Solyc03g051810	Homogentisate solanesyl transferase
VTE3a	Solyc03g005230	MSBQ/MPBQ methyltransferase
VTE3b	Solyc09g065730	MSBQ/MPBQ methyltransferase

*Activity reported in this work

**Prediction. Needs experimental confirmation

Table S3. Individual GCNs of plastidial GGPPS paralogs (Guide genes) positively co-expressed with isoprenoid-related genes (Query genes) in different tomato tissues. Significant pairwise Pearson (ρ) correlations (ρ>0.55) between guide and query genes are shown. See Table S2 for gene description.

Guide gene	Query gene	Gene ID	Pearson (ρ)	Guide gene	Query gene	Gene ID	Pearson (ρ)
Vegetative tissue GCN				*Vegetative tissue GCN*			
G1 (SlG1)	ABA2c	Solyc10g085380	0.64		SGR1	Solyc08g080090	0.96
	CHLM	Solyc03g118240	0.67		CCD4B	Solyc08g075490	0.96
					PAO3	Solyc12g096550	0.96
G2 (SlG2)	HEMF	Solyc10g005110	0.60		PAO1	Solyc11g066440	0.96
	GA3	Solyc04g083160	0.60		CLD1	Solyc02g070490	0.97
	CHLH	Solyc04g015750	0.60		Or1	Solyc03g093830	0.97
	BCH2	Solyc03g007960	0.60		CYP97A3	Solyc04g051190	0.97
	CHLD	Solyc04g015490	0.65		SPS1	Solyc07g061990	0.97
	GA3ox1	Solyc06g066820	0.65		NYC1	Solyc07g024000	0.97
	HEMB	Solyc08g069030	0.66		LCY-E	Solyc12g008980	0.98
	NSY1	Solyc02g089050	0.68		DXS3b	Solyc08g066950	0.98
	PDS2/HST	Solyc03g051810	0.69		CYP97C1	Solyc10g083790	0.98
	SGR3	Solyc04g063240	0.69		CYP97B3	Solyc05g016330	0.98
	DXR1	Solyc03g114340	0.69		Z-ISO	Solyc12g098710	0.98
	CMK	Solyc01g009010	0.70		PSY2	Solyc02g081330	0.98
	ABA1/ZEP	Solyc02g090890	0.71		HCAR	Solyc09g091100	0.98
	AAO3b	Solyc11g065930	0.72		CRTISO2	Solyc05g010180	0.99
	KS1b	Solyc08g005720	0.74				
	CHLI	Solyc10g008740	0.74	*G3 (SlG3)*	RCCR	Solyc03g044470	0.60
	VTE4d	Solyc03g116150	0.75		DHNS/MENB	Solyc05g005180	0.60
	CYP707A3b	Solyc08g005610	0.75		HEME2	Solyc06g048730	0.60
	MDS	Solyc08g081570	0.75		TPS27	Solyc02g079910	0.60
	PHYLLO/MEND1	Solyc04g005190	0.75		GGDS8	Solyc07g064660	0.60
	PK1	Solyc03g071720	0.76		PSY1	Solyc03g031860	0.60
	CHLG1	Solyc05g024190	0.76		NSY3	Solyc02g086050	0.61
	FDS2	Solyc12g015860	0.77		CYP97C1	Solyc10g083790	0.64
	HEMA1	Solyc04g076870	0.77		CRTISO1	Solyc10g081650	0.65
	PSY1	Solyc03g031860	0.79		VTE4a	Solyc08g076360	0.66
	GA20ox4	Solyc06g035530	0.80		GSA	Solyc04g009200	0.67
	CHL27-CRD	Solyc10g077040	0.81		HEME1	Solyc10g007320	0.67
	CPS2	Solyc08g005710	0.82		MDS	Solyc08g081570	0.70
	CAOa	Solyc06g060310	0.82		CMK	Solyc01g009010	0.73
	FDS3	Solyc10g005810	0.84		CHLG1	Solyc05g024190	0.75
	PHYLLO/MEND3	Solyc04g005180	0.84		VDE	Solyc04g050930	0.76
	NSY5	Solyc06g074240	0.85		PK1	Solyc03g071720	0.77
	CYP707A1	Solyc04g078900	0.86		ACS12	Solyc03g007070	0.77
	CHL1	Solyc09g065620	0.86		CHLD	Solyc04g015490	0.79
	DHNAT3	Solyc03g006450	0.87		Or2	Solyc09g010110	0.81
	HPPD1/PDS1a	Solyc05g041200	0.88		CHLM	Solyc03g118240	0.84
	IDI1	Solyc08g075390	0.88		GA20ox1	Solyc03g006880	0.87
	HEMD	Solyc04g079320	0.89		MAX3 (CCD7)	Solyc01g090660	0.88
	PORA	Solyc12g013710	0.89		CHLI	Solyc10g008740	0.88
	ZEP1	Solyc06g060880	0.89		HEMF	Solyc10g005110	0.89
	LCY-B1	Solyc04g040190	0.90		CBR/NOL	Solyc05g032660	0.97
	NSY4	Solyc03g034240	0.91				
	DXS2	Solyc11g010850	0.92				
				Fruit GCN			
	AAO3a	Solyc11g065920	0.92	*G1 (SlG1)*	AAO3b	Solyc11g065930	0.59
	PPH	Solyc01g088090	0.92				
	CHL2c	Solyc12g005300	0.92	*G2 (SlG2)*	CMK	Solyc01g009010	0.56

VTE4b	Solyc04g063230	0.92		CLD1	Solyc02g070490	0.57
ABA2b	Solyc04g071960	0.93		MENG	Solyc12g019010	0.60
HEMG2	Solyc03g005080	0.94		PDS	Solyc03g123760	0.71
MENG	Solyc12g019010	0.94		BCH1	Solyc06g036260	1.00
VTE3a	Solyc03g005230	0.94				
PDS	Solyc03g123760	0.94	*G3 (SlG3)*	DHNAT1	Solyc02g078410	0.55
PORB	Solyc07g054210	0.94		CRTISO1	Solyc10g081650	0.56
CHL2a	Solyc06g053980	0.94		SGR1	Solyc08g080090	0.57
VTE4c	Solyc08g077240	0.95		Z-ISO	Solyc12g098710	0.61
CAOb	Solyc11g012850	0.95		Or1	Solyc03g093830	0.64
HDS	Solyc11g069380	0.95		Or2	Solyc09g010110	0.67
NCED2	Solyc08g016720	0.95		KS1b	Solyc08g005720	0.68
PORC	Solyc10g006900	0.95		BCH2	Solyc03g007960	0.70
ABA3	Solyc07g066480	0.95		GA3ox4	Solyc01g058250	0.71
HPPD2/PDS1b	Solyc07g045050	0.96		CYP707A3c	Solyc04g071150	0.74
NCED3	Solyc07g056570	0.96		CMK	Solyc01g009010	0.76

Table S4. Primers used in this work

Use	#	Name	Sequence (5'-3')[1]
Cloning	1	SlG1-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGGCATTTTTAGCTACCATTTCTG
	2	SlG1-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGATTCTGTCGATTTGCAATATAACTAGC
	3	SlG2-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGAGATCTATGAACCTTGTTGATTC
	4	SlG2-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGATTTTGACGATTAGCAATGTAATCTG
	5	SlG3-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGAGTCTTTCAACAACAATTACAAC
	6	SlG3-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGATTCTCTCTGTAAGCAATATAATTTG
	7	SlG4-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGGCCCATACTAAGTCAAATAGG
	8	SlG4-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTTTGCCGATGAAGAACGAAATC
	9	SlG5-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGTCTATGCGAAAAGGTGTAATCC
	10	SlG5-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTTTGCCGATGAAGAAGAAAATC
	11	SlG1(-tp)-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCTAGCGATGTTGCGAACTC
	12	SlG1stop-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAATTCTGTCGATTTGCAATATAAC
	13	SlG2(-tp)-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTTATGGAAAAAGAAGAATTTAATTTC
	14	SlG2stop-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTAATTTTGACGATTAGCAATG
	15	SlG3(-tp)-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCAATGGAGTTTAAAGAATACG
	16	SlG3stop-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTAATTCTCTCTGTAAGCAATATAATTTG
	17	AtHDS-attB1-F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGACTGGAGTATTGCCAGCTC
	18	AtHDS-attB2-R	GGGGACCACTTTGTACAAGAAAGCTGGGTGATTCCGGATAACCGAAACTCTTCTC
RT-qPCR	19	SlPSY1-qPCR-F	ACAGGCAGGTCTATCCGATG
	20	SlPSY1-qPCR-R	ACGCCTTTCTCTGCCTCATC
	21	SlPSY2-qPCR-F	CAGGGCTCTCCGATGAAGAC
	22	SlPSY2-qPCR-R	CACCGGCCATCTACTAGCAG
	23	SlPSY3-qPCR-F	TTGGATGCAATAGAGGAGAATG
	24	SlPSY3-qPCR-R	ATTGAATGGCTAAACTAGGCAAAG
	25	SlG1-qPCR-F	GGCCTTTGAACATGTGGCTACC
	26	SlG1-qPCR-R	ACTCGCCAAGTCCACAATTTGC
	27	SlG2-qPCR-F	AAAGTCATCGTCGGAGCTCG
	28	SlG2-qPCR-R	GTTTAGCTTCGCCGTTGAGC
	29	SlG3-qPCR-F	AGGAGGTGCACCAGATGAAG
	30	SlG3-qPCR-R	TCAGCAACCAAGTCCTTCCC
	31	SlEXP-qPCR-F	GCTAAGAACGCTGGACCTAA
	32	SlEXP-qPCR-R	TGGGTGTGCCTTTCTGAATG
	33	SlACT-qPCR-F	CCTTCCACATGCCATTCTCC
	34	SlACT-qPCR-R	CCACGCTCGGTCAGGATCT
Genotyping PCR	35	SlG1-qPCR-F	GGCCTTTGAACATGTGGCTACC
	36	SlG2-qPCR-F	AAAGTCATCGTCGGAGCTCG
	37	SlG3-qPCR-F	AGGAGGTGCACCAGATGAAG
	38	eGFP-qPCR-2 R	TCTCGTTGGGGTCTTTGCTC
	39	ACT_genot_F	GTGAAAAGATGACCCAGATTATG
	40	ACT_genot_R	CACGCTCGGTCAGGATCTTCATC
Genotyping qPCR[2]	41	Lat1-F	AGACCACGAGAACGATATTTGC
	42	Lat2.1-R	GCCTTTTCATATCCAGACACAC
	43	npt1-5'-F	GACAGGTCGGTCTTGACAAAAAG
	44	npt1-3'-R	GAACAAGATGGATTGCACGC

[1]Gateway recombination sites in bold
[2]Described in Annex 4

Table S5. Constructs and cloning details.

Use	Construct	Template	Primers[2]	Sequence cloned[3]	Plasmid backbone
Subcellular localization assays/ Transgenic plants	*35S:SlG1-GFP*	Tomato root cDNA	1 + 2	$SlG1_{1-1095}$	pGWB405
	35S:SlG2-GFP	Tomato flower cDNA	3 + 4	$SlG2_{1-1089}$	pGWB405
	35S:SlG3-GFP	Tomato flower cDNA	5 + 6	$SlG3_{1-1080}$	pGWB405
	35S:SlG4-GFP	Tomato flower cDNA	7 + 8	$SlG4_{1-1179}$	pGWB605
	35S:SlG5-GFP	Tomato flower cDNA	9 + 10	$SlG5_{1-1119}$	pGWB605
	35S:pGFP	Arabidopsis cDNA	17 + 18	$AtHDS_{1-147}$	pGWB405
In vitro GGPPS activity assays	*6xHis-SlG1*	35S:SlG1-GFP	11 + 12	$SlG1_{130-1098}$	pET32-GW
	6xHis-SlG2	35S:SlG2-GFP	13 + 14	$SlG2_{190-1092}$	pET32-GW
	6xHis-SlG3	35S:SlG3-GFP	15 + 16	$SlG3_{196-1083}$	pET32-GW
	pET-AtG11[1]	-	-	$AtG11_{169-1116}$	pET32-GW
	pET-G11s[1]	-	-	$AtG11_{169-1050}$	pET32-GW

[1]Constructs reported in Ruiz-Sola et al. (2016b), **Annex2**

[2]See Table S1

[3]Numbers indicate the first and last nucleotide positions cloned from the coding sequence of the indicated gene.

GENERAL DISCUSSION

1. Subfunctionalization of plant GGPPS paralogs

In plants, GGPP is a central diversification point for the synthesis of isoprenoids, some of which are essential to sustain life on earth. Unlike other enzymes catalyzing earlier isoprenoid biosynthetic steps, encoded by one or two genes (Goldstein and Brown, 1990; Phillips et al., 2008; Closa et al., 2010; Vranová et al., 2013), GGPPS activity relies on a higher number of genes (Coman et al., 2014; Ruiz-Sola et al., 2016b, *see* **Annex 2**). The presence of several gene paralogs (*i.e.* intraspecific homolog genes) in a genome results from duplications of genes, genomic fragments or whole-genomes. Unlike ortholog genes (homolog genes in different species) that display similar biological roles, paralog genes normally endure in the genome of a plant species if they acquire advantageous specialized functions that improve the performance of the organism against different environmental conditions (Studer and Robinson-Rechavi, 2009). If this is not the case, a paralog will suffer deleterious mutations becoming an inactive pseudogene ("pseudogenization"), something that likely occurred in the case of the Arabidopsis *GGPPS5* gene (Beck et al., 2013). There are different ways for a paralog to specialize: it can either evolve towards a complete new function, leaving the original role to another paralog ("neofunctionalization"), or it can retain the ancestral function but specialize in specific cell compartments, tissues or processes ("subfunctionalization") (Lynch and Conery, 2000; Lynch and Force, 2000; Innan and Kondrashov, 2010).

While a single *GGPPS* copy gene is present in green algae, several *GGPPS* paralogs exist in most plants, likely arising during the diversification of land plants. Several molecular changes in important GGPPS protein motifs conferred new and more specialized functions to some paralogs, giving rise to SSU, PPPS or GFPPS enzymes in many plant species (Wang and Dixon, 2009; Coman et al., 2014; Nagel et al., 2015). While GGPP-producing enzymes contain three highly conserved motifs (FARM, SARM and the CxxxC upstream to the FARM), SSU proteins acquired a second CxxxC motif upstream to the SARM but accumulated inactivating mutations in the catalytic domains. Thus, type I SSU proteins lost both FARM and SARM and type II SSU proteins lost only the SARM (*see* Chapter II, Fig. S1). These changes allowed the proteins to heterodimerize with GGPPS monomers through the CxxxC motifs and either produce much more specialized metabolites such as GPP-derived monoterpenes or enhance GGPP synthesis for particular isoprenoid branches. On the other hand, other GGPPS paralogs that retained the canonical catalytic domains experienced mutations in the chain-length determination (CLD) region that encoded smaller amino acids, hence allowing the production of longer prenyl diphosphates in

the elongation cavity. The change of the methionine (M) present in the fifth position upstream to the FARM, for a serine (S) or an alanine (A), enabled the protein to produce C25 GFPP or longer prenyl diphosphates (PPPs), respectively, generating new groups of PTs (*see* Chapter II, Fig.2 and Fig. S1; Hsieh et al., 2011; Nagel et al., 2015; Wang et al., 2016). Other paralogs not only gained an A in the CLD region but lost the first CxxxC, producing even longer prenyl diphosphates such as solanesyl diphosphate (SPP), required for the synthesis of respiratory or photosynthetic electron transfer molecules, including ubiquinone and plastoquinone (Ducluzeau et al., 2012; Jones et al., 2013). All these molecular modifications illustrate the strong neofunctionalization of the GGPPS-like homologs during the evolution of land plants, which likely contributed to enabling a quick adaptation to the new environmental challenges. The analysis of these sequence features allowed us to predict that Arabidopsis sG11 retained GGPPs activity and that tomato *SlG1-SlG5* genes encoded GGPPS enzymes, predictions that were later confirmed. Although similar sequence-based reasoning led us to conclude that SlG6 is a SSUII protein, this still requires experimental approaches.

Among the paralogs that show GGPPS activity, different layers of regulation led to a subfunctionalization that allowed their retention in the genome of a plant species. The very first layer of subfunctionalization is the transcriptional regulation of different paralog genes, which were found to show specific expression patterns strictly controlled by different developmental or environmental stimuli in several plant species (Ament et al., 2006; Beck et al., 2013; Zhang et al., 2015; Wang et al., 2018b, 2018a). Once they are transcriptionally activated, another layer of subfunctionalization appears, which corresponds to the subcellular localization of the synthesized GGPPS isoform. The acquisition of organelle-targeting signals in the coding sequence of a gene paralog can provide specialized functions in particular cell compartments, again facilitating its functional retention in the genome. Both, Arabidopsis and tomato show these two subfunctionalization mechanisms for the GGPPS gene family (Figure 1). A third layer of subfunctionalization corresponds to interactions of GGPPS isoforms with specific protein partners to modulate enzyme activity and divert the flux of precursors towards particular GGPP-consuming downstream pathways (Figure 1).

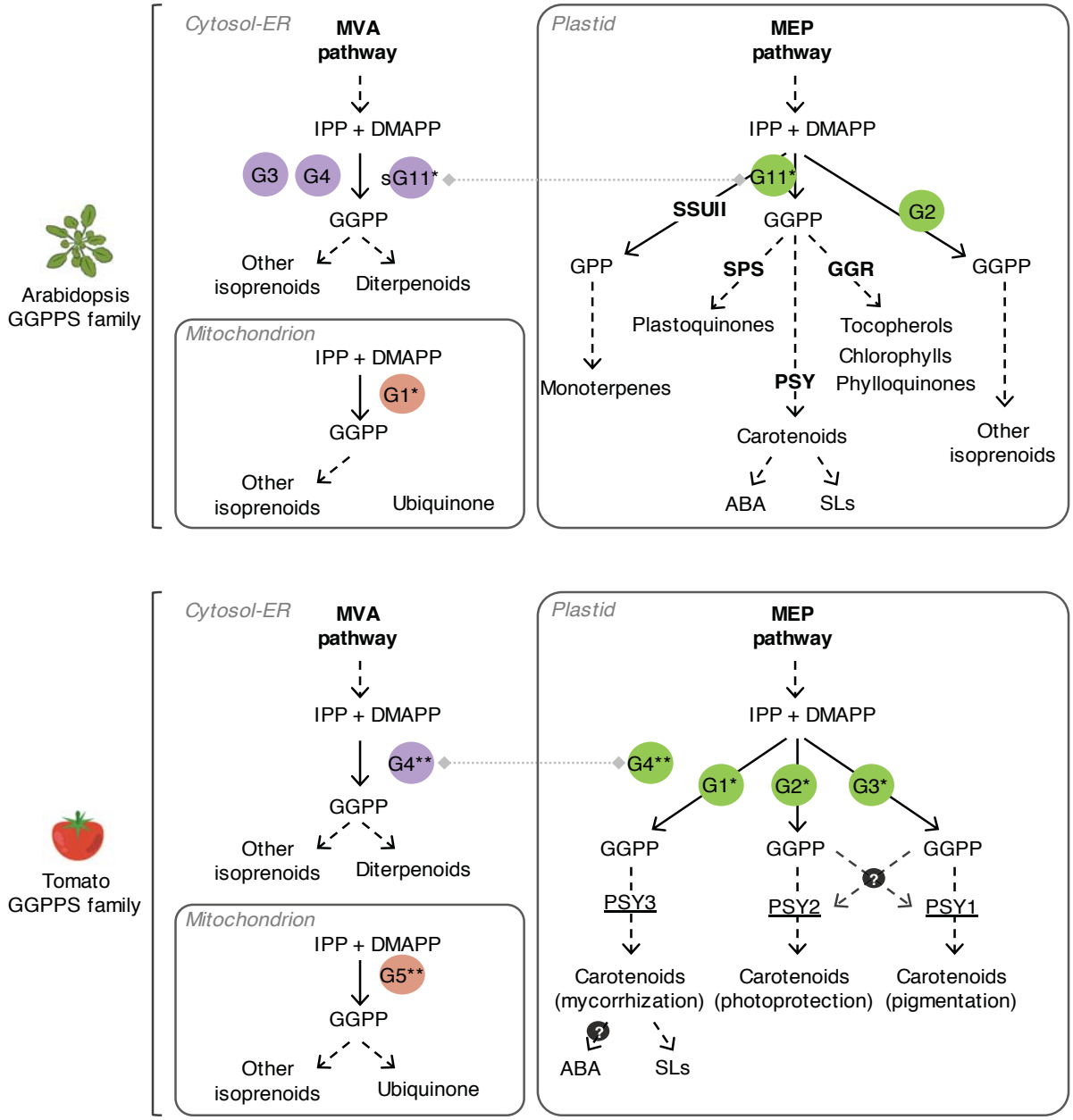


Figure 1. Comparative representation of our current knowledge on Arabidopsis and tomato GGPPS protein families. In this Figure, GGPPS isoforms lack the prefix of the species (indicated with images) for simplicity. Single asterisks indicate the isoforms analyzed in this study and double asterisks indicate results that require confirmation. Cytosolic-ER, mitochondrial and plastidial enzymes are represented in purple, red and green circles, respectively. Solid arrows represent single enzymatic steps, dashed arrows indicate multiple enzymatic steps and dotted grey lines highlight differentially localized GGPPS isoforms encoded by the same gene. Enzymes marked in bold indicate physical interaction with the upstream GGPPS isoform. Underlined enzymes represent positive co-expression and putative protein-protein interaction with the upstream GGPPS isoform. Question marks indicate hypothetical associations.

2. The Arabidopsis GGPPS family

Regarding Arabidopsis GGPPS-encoding genes, it was shown that whereas the paralogs encoding mitochondrial (AtGGPPS1) and plastidial (AtGGPPS2 and AtG11)

enzymes were ubiquitously expressed, the genes encoding the ER-targeted isoforms (AtGGPPS3 and AtGGPPS4) were restricted to specific tissues from specific organs (Beck et al., 2013). Evolutionary analyses of the Arabidopsis *GGPPS* gene sequences and expression patterns revealed that *AtGGPPS2* and *AtG11* are the oldest paralogs, indicating that the firstly needed GGPPS activity was related to plastidial isoprenoid metabolism (Coman et al., 2014). The analysis also showed that the highest levels of expression of *AtG11* would reflect that it is the paralog that retained most of the ancestral function. Later emerged genes (*AtGGPPS3* and *4*) and finally *AtGGPPS1* show other cellular locations and lower transcript levels and/or specialized expression patterns, indicating that newly acquired genes must be specialized once the essential original function is covered (Coman et al., 2014). Analysis of *AtGGPPS2* and *AtG11* mutants showed that only the AtG11 enzyme is essential for plant development (providing GGPP for primary isoprenoids involved in photosynthesis and embryo development) and suggested that *AtGGPPS2* might have a specific role in the secondary metabolism despite being constitutively expressed (Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see* **Annex 2**). The essential function of AtG11 in photosynthesis completely fits with the evolutionary analyses, explaining why this isoform is ubiquitously expressed in Arabidopsis and why the plastidial GGPP pool was initially required before gaining more paralogs and functions. However, we reported here that the essential function of the *AtG11* gene during embryo development gene relies on another layer of transcriptional regulation found for the first time for a *GGPPS* gene. We discovered that the *AtG11* gene produces transcripts of different lengths that encode active GGPPS enzymes with essential functions in different cell compartments. Longer transcripts are translated into a plastid-targeted AtG11 protein that supplies GGPP for photosynthesis-related isoprenoids, while shorter transcripts lacking the first ATG codon encode a version that lacks the N-terminal signaling peptide and remains in the cytosol to produce GGPP for embryo development (Ruiz-Sola et al., 2016a, *see* **Annex 1**; Ruiz-Sola et al., 2016b, *see* **Annex 2**).

It is possible that, during the evolution of the Arabidopsis *GGPPS* gene family, the initial role of AtG11 was to supply GGPP in plastids but it acquired a new one in the cytosol by using a second ATG codon downstream the one used to encode the full-length AtG11 protein. The use of an alternative translation start site might had been selected before the specialization of *AtGGPPS2* and before the appearance of the ER-isoforms. These results together let us conclude that Arabidopsis gene family is formed by five transcriptionally regulated genes that encode six differentially

localized protein isoforms with particular functions, illustrating again the complexity that characterizes plant metabolism (Figure 1).

3. The tomato GGPPS family

In the case of the tomato GGPPS family, of which there was very limited information at the beginning of this doctoral thesis, we found similar mechanisms regulating GGPP biosynthesis through the specialization of different paralogs. As previously mentioned we identified five GGPPS-like homologs which, accordingly to gene expression databases (*e.g. TomExpress*), show particular expression patterns. *SlG1* was mostly expressed in roots and leaves, while *SlG2* and *SlG3* showed many similarities at transcriptional level. They were constitutively expressed, showing the highest levels of transcripts among tomato GGPPS paralogs. Both genes were strongly expressed in leaves and were induced during fruit ripening. *SlG4* was expressed in flowers, while *SlG5* had a peak of expression in roots and was repressed during fruit ripening (Figure 2). In terms of organelle-targeting we found again plastidial (SlG1-3), cytosolic (SlG4) and mitochondrial (SlG5) isoforms, and even dual targeting of SlG4, that was also detected in chloroplasts (Figure 1). Our phylogenetic analyses have shown that plastid-targeted tomato GGPPS proteins are closely related to the plastidial isoforms from Arabidopsis. By contrast, SlG4 and 5 show quite a big evolutionary divergence from their paralogs indicating a different selective pressure and again clues of subfunctionalization (*see* Chapter II, Fig. 2).

Gene expression analyses of the plastidial paralogs suggest that SlG2 might be the tomato equivalent of AtG11, performing most of the original GGPPS function in photosynthetic tissues (*see* Chapter II, Fig. 5). Our results also suggest that SlG1 and SlG3 might have acquired more specialized functions in roots and fruits, respectively. In agreement with this conclusion, the transcriptional pattern of the plastidial GGPPS-encoding genes was finely-regulated and specifically activated in particular spatio-temporal conditions to enhance the biosynthesis of carotenoids with very specific functions (*i.e.* photoprotection, myccorhization, or pigmentation), hence explaining why tomato maintains three plastidial isoforms. The results further illustrate the plasticity of the GGPPS gene family to rapidly acquire new roles during evolution and support the adaptation of the organism to particular environmental conditions.

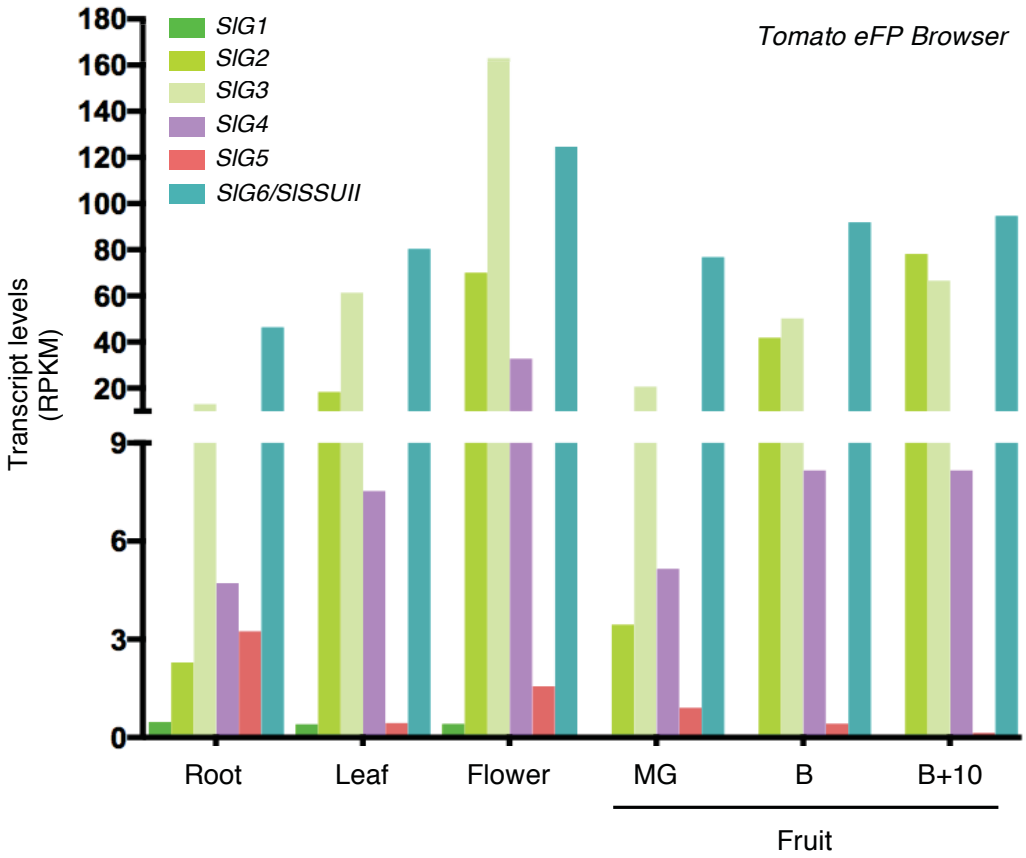


Figure 2. Transcript levels of *GGPPS* paralog genes in different tomato tissues. RNAseq data were retrieved from the *Tomato eFP browser* database (http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Graphs show the transcript levels of *SlG1-5* and *SlSSUII* in root, leaf, flower and fruit (MG, mature green; B, breaker; B+10 days). Expression data are represented as RPKM (Reads per Kilobase of transcript per Million mapped reads). For accessions: *see* Chapter II, Table 1.

4. Modulation of GGPP supply by protein-protein interactions

Several lines of evidence demonstrated that GGPPS proteins can interact with other enzymes catalyzing upstream and downstream biosynthetic steps in the plastids of different plant species (Dogbo and Camara, 1987; Camara, 1993; Maudinas et al., 1977; Fraser et al., 2000; Ruiz-Sola et al., 2016a, *see* **Annex 1**; Zhou et al., 2017; Camagna et al., 2018; Wang et al., 2018b). In the case of Arabidopsis, AtG11 has been found to interact with PSY, GGR and SPS2 enzymes (Ruiz-Sola et al., 2016a, *see* **Annex 1**; Camagna et al., 2018), presumably to directly provide GGPP for the production of specific groups of photosynthesis-related plastidial isoprenoids. This is consistent with its house-keeping role during photosynthetic-associated development. GGPPS-PSY containing protein complexes were also isolated from tomato chloroplasts and fruit chromoplasts (Maudinas et al., 1977; Fraser et al., 2000).

The transcriptional co-regulation of different tomato GGPPS and PSY paralogs reported in this work suggests that specific protein pairs could be formed to produce carotenoids depending on the tissue and the developmental process (*see* Chapter II, Fig. 6). Thus, SlG1-PSY3 would produce carotenoids for mycorrhization in roots, SlG2-PSY2 would be most relevant for photoprotection in photosynthetic tissues, and the production of carotenoids for fruit pigmentation during ripening would mainly rely on the SlG3-PSY1 tandem. It is most likely that these GGPPS paralogs could also interact with other proteins besides PSY.

Interaction of GGPPS enzymes with some proteins has been shown to facilitate the assembly of specific multienzyme complexes and modulate the enzymatic activity of different GGPPS paralogs to face particular demands. This is the case of SSU proteins, GGPPS-descending enzymes that evolved, together with specific GGPPS paralogs, to post-translationally regulate the allocation of isoprenoid precursors. This mechanism was identified in Arabidopsis, hop, rice and pepper (Wang and Dixon, 2009; Zhou et al., 2017; Wang et al., 2018b), but nothing has been reported in tomato yet. Here, we propose the existence of a SSUII homolog in tomato that contains all the structural characteristics to putatively interact with GGPPS monomers (*see* Chapter II). Further studies will be required to identify the specific GGPPS paralogs that by subfunctionalization acquired the ability to interact with SlSSUII, perhaps to produce GPP or GGPP in specialized metabolons.

5. Open questions

Experiments are currently ongoing in our lab to (1) determine the specific roles of the tomato plastid-localized GGPPS enzymes (by generating and analyzing loss-of-function mutants by CRISPR-Cas9), (2) confirm their isoform-specific interaction with PSY paralogs (by pull-down experiments in *E. coli* and *N. benthamiana* cells expressing tagged proteins), and (3) identify new protein partners (by co-immunoprecipitation of protein complexes from different tissues of tomato plants stably transformed with constructs to produce GFP-tagged GGPPS isoforms). Isolating GGPPS-containing metabolons *in vivo*, however, is a huge challenge since they are often assembled under really specific conditions or in very restricted tissues (*e.g.* in hop trichomes or in thylakoids of rice chloroplasts).

While the existence of several GGPPS-encoding genes has been demonstrated in many plant species, the biological roles of paralogs with restricted, confined, or specialized expression profiles remains virtually unknown. Moreover, so far, only

plastidial GGPPS paralogs have been deeply characterized beyond gene expression analyses, probably because of their higher expression levels and their involvement in primary processes, where an impact is more easily detectable (e.g. *AtG11* knock-down and knock-out mutants or rice *OsGGPPS1* overexpression or downregulation display visual phenotypes). In this thesis, we discovered the existence and essential function of the Arabidopsis sG11 cytosolic protein for embryo development but we could not identify the GGPP-derived isoprenoid responsible for that phenotype, since metabolic measurements in embryos at specific developmental stages is extremely tricky. Future technological advances such as single-cell metabolomics should provide the tools to address this question in the future.

Evolutionary mechanisms like those reported here for GGPPS enzymes have been also described for many gene paralog families involved in different steps of isoprenoid metabolism (Vranová et al., 2012, 2013; Rodríguez-Concepción and Boronat, 2015). These mechanisms contribute to build an intricate network of regulation of plant isoprenoid metabolism, that ultimately leads to an immense amount and diversity of functional structures. Understanding this regulatory network of isoprenoid biosynthesis in different plant species will definitely help us to improve plant traits such as crop yield, nutritional quality, stress resistance or the production of useful compounds in a more sustainable manner.

CONCLUSIONS

1. The only essential GGPPS gene in Arabidopsis (*AtG11*) encodes two differentially targeted GGPP-producing enzymes. The presence of alternative transcription initiation sites results in transcripts of different sizes, some of which lack the first ATG codon but use a second in-frame ATG codon to produce a shorter protein.
2. The long AtG11 isoform is targeted to plastids and its loss of function causes an albino-lethal phenotype consistent with this isoform providing GGPP for the production of photosynthesis-related plastidial isoprenoids.
3. The short AtG11 isoform remains in the cytosol because it lacks the N-terminal plastid-targeting signal. The GGPP produced by this cytosolic isoform is required for embryo development, perhaps because it is used for the production of ubiquinone or/and another isoprenoid in the cytosol or the mitochondria.
4. The tomato genome contains five genes encoding putative GGPPS enzymes and one gene encoding a type II SSU protein that might influence the activity of the GGPPS isoforms by forming heterodimers.
5. The tomato GGPPS isoforms are mainly located in plastids (SlG1, SlG2 and SlG3), cytosol (SlG4) and mitochondria (SlG5) when fused to GFP.
6. The three plastidial isoforms only produce GGPP *in vitro* with similar kinetic parameters compared to AtG11.
7. GCN analyses suggest functional association of SlG2 to plastidial isoprenoid biosynthesis in photosynthetic tissues and SlG3 to carotenoid biosynthesis in fruits.
8. Analyses of gene expression during carotenoid biosynthesis in different tissues showed similar co-expression profiles of particular genes encoding GGPPS and PSY isoforms. *SlG1* is co-induced with *PSY3* during mycorrhization in roots, *SlG2* and *PSY2* are similarly upregulated in de-etiolating seedlings, and *SlG3* and *PSY1* show parallel induction profiles during fruit ripening.
9. Reduced *SlG3* transcript levels correlate with decreased phytoene levels in fruit, suggesting that the SlG3 isoform diverts GGPP to PSY1 for carotenoid biosynthesis during ripening.
10. Particular GGPPS and PSY isoforms might physically interact in tomato to ensure the channelling of IPP and DMAPP into the production of carotenoids in a tissue-specific manner. Interaction of SlG1 and PSY3 would be most relevant in roots, whereas complexes containing either SlG2 and PSY2 or SlG3 and PSY1 might cooperate to produce carotenoids in leaves and fruits.

GENERAL REFERENCES

Akhtar TA, Matsuba Y, Schauvinhold I, Yu G, Lees HA, Klein SE and Pichersky E (2013). The tomato cis-prenyltransferase gene family. Plant J. **73**: 640–652.

Ambo T, Noike M, Kurokawa H and Koyama T (2008). Cloning and functional analysis of novel short-chain cis-prenyltransferases. Biochem. Biophys. Res. Commun. **375**: 536–540.

Ament K, Van Schie CC, Bouwmeester HJ, Haring MA and Schuurink RC (2006). Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. Planta **224**: 1197–1208.

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodriguez-Concepcion M, Gruissem W and Vranova E (2013). Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. Plant Mol Biol **82:** 393-416.

Bick JA and 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:** 146-154.

Bouvier F, Suire C, D'Harlingue A, Backhaus RA and Camara B (2000). Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J. **24**: 241–52.

Burke C, Klettke K and Croteau R (2004). Heteromeric geranyl diphosphate synthase from mint: construction of a functional fusion protein and inhibition by bisphosphonate substrate analogs. Arch. Biochem. Biophys. **422**: 52–60.

Burke CC, Wildung MR and Croteau R (1999). Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proc. Natl. Acad. Sci. **96**: 13062–13067.

Camagna M, Grundmann A, Bär C, Koschmieder J, Beyer P and Welsch R (2018). Enzyme fusion removes competition for geranylgeranyl diphosphate in carotenogenesis. Plant Physiol.: pp.01026.2018.

Camara B (1993). Plant phytoene synthase complex: Component enzymes, Immunology and Biogenesis. Methods Enzymol. **214**: 352–365.

Chang TH, Hsieh FL, Ko TP, Teng KH, Liang PH and Wan AHJ (2010). Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (*Mentha piperita*) reveals intersubunit regulation. Plant Cell **22**: 454–467.

Chen Q, Fan D and Wang G (2015). Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in Arabidopsis flowers. Mol Plant **8:** 1434-1437.

Clarke CF, Tanaka RD, Svenson K, Wamsley M, Fogelman AM and Edwards PA (1987). Molecular cloning and sequence of a cholesterol-repressible enzyme related to prenyltransferase in the isoprene biosynthetic pathway. Mol. Cell. Biol. **7**: 3138–46.

Closa M, Vranova E, Bortolotti C, Bigler L, Arro M, Ferrer A and 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:** 512-525.

Coman D, Altenhoff A, Zoller S, Gruissem W and Vranova E (2014). Distinct evolutionary strategies in the GGPPS family from plants. Front Plant Sci **5:** 230.

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

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

Dogbo O and Camara B (1987). Purification of isopentenyl pyrophosphate isomerase and geranylgeranyl pyrophosphate synthase from Capsicum chromoplasts by affinity chromatography. Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab. **920**: 140–148.

Ducluzeau AL, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC and Basset GJ (2012). Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J **69:** 366-375.

Dudareva N, Martin D, Kish CM, Kolosova N, Gorenstein N, Fäldt J, Miller B and Bohlmann J (2003). (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell **15**: 1227–1241.

Fraser PD, Schuch W and Bramley PM (2000). Phytoene synthase from tomato (*Lycopersicon esculentum*) chloroplasts--partial purification and biochemical properties. Planta **211**: 361–369.

Gaffe J, Bru JP, Causse M, Vidal A, Stamitti-Bert L, Carde JP and Gallusci P (2000). *LEFPS1,* a tomato farnesyl pyrophosphate gene highly expressed during early fruit development. Plant Physiol. **123**: 1351–1362.

Goldstein JL and Brown MS (1990). Regulation of the mevalonate pathway. Nature **343**: 425–30.

Grosjean K, Mongrand S, Beney L, Simon-Plas F and Gerbeau-Pissot P (2015). Differential effect of plant lipids on membrane organization: specificities of phytosphingolipids and phytosterols. J. Biol. Chem. **290**: 5810–25.

Hartmann M (1998). Plant sterols and the membrane environment. Trends Plant Sci. **3**: 170–175.

He JX, Fujioka S, Li TC, Kang SG, Seto H, Takatsuto S, Yoshida S and Jang JC (2003). Sterols regulate development and gene expression in Arabidopsis. Plant Physiol. **131**: 1258–69.

Hefner J, Ketchum REB and Croteau R (1998). Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from *Taxus canadensis* and assessment of the role of this prenyltransferase in cells induced for Taxol production. Arch. Biochem. Biophys. **360**: 62–74.

Hemmerlin A, Harwood JL and Bach TJ (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? Prog. Lipid Res. **51**: 95–148.

Hemmerlin A, Rivera SB, Erickson HK and Poulter CD (2003). Enzymes encoded by the farnesyl diphosphate synthase gene family in the Big Sagebrush *Artemisia tridentata ssp. spiciformis*. J. Biol. Chem. **278**: 32132–40.

Horvath SE and Daum G (2013). Lipids of mitochondria. Prog. Lipid Res. **52**: 590–614.

Hsiao YY, Jeng MF, Tsai WC, Chuang YC, Li CY, Wu TS, Kuoh CS, Chen WH and Chen HH (2008). A novel homodimeric geranyl diphosphate synthase from the orchid *Phalaenopsis bellina* lacking a DD(X)2-4D motif. Plant J. **55**: 719–33.

Hsieh FL, Chang TH, Ko TP and Wang AH (2011). Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol **155:** 1079-1090.

Innan H and Kondrashov F (2010). The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. **11**: 97–108.

Jones MO, Perez-Fons L, Robertson FP, Bramley PM and Fraser PD (2013). Functional characterization of long-chain prenyl diphosphate synthases from tomato. Biochem. J. **449**: 729–740.

Keim V, Manzano D, Fernandez FJ, Closa M, Andrade P, Caudepon D, Bortolotti C, Vega MC, Arro M and Ferrer A (2012). Characterization of Arabidopsis FPS isozymes and *FPS* gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds. PLoS One **7:** e49109.

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

Lenucci MS, Serrone L, De Caroli M, Fraser PD, Bramley PM, Piro G and Dalessandro G (2012). Isoprenoid, lipid, and protein contents in intact plastids isolated from mesocarp cells of traditional and high-pigment tomato cultivars at different ripening stages. J. Agric. Food Chem. **60**: 1764–75.

Li R, Sun R, Hicks GR and Raikhel NV (2015). Arabidopsis ribosomal proteins control vacuole trafficking and developmental programs through the regulation of lipid metabolism. Proc. Natl. Acad. Sci. U. S. A. **112**: E89-98.

Liang PH, Ko TP and Wang AHJ (2002). Structure, mechanism and function of prenyltransferases. Eur. J. Biochem. **269**: 3339–3354.

Liao Z, Gong Y, Kai G, Zuo K, Chen M, Tan Q, Wei Y, Guo L, Tan F, Sun X and Tang K (2005). [An intron-free methyl jasmonate inducible geranylgeranyl diphosphate synthase gene from *Taxus media* and its functional identification in yeast]. Mol. Biol. (Mosk). **39**: 14–20.

Lynch M and Conery JS (2000). The evolutionary fate and consequences of duplicate genes. Science **290**: 1151–5.

Lynch M and Force A (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics **154**: 459–73.

Manzano D, Andrade P, Caudepón D, Altabella T, Arró M and Ferrer A (2016). Suppressing farnesyl diphosphate synthase alters chloroplast development and triggers sterol-dependent induction of jasmonate- and fe-related responses. Plant Physiol. **172**: 93–117.

Manzano D, Busquets A, Closa M, Hoyerova K, Schaller H, Kaminek M, Arro M and Ferrer A (2006). Overexpression of farnesyl diphosphate synthase in Arabidopsis mitochondria triggers light-dependent lesion formation and alters cytokinin homeostasis. Plant Mol Biol **61:** 195-213.

Masferrer A, Arró M, Manzano D, Schaller H, Fernandez-Busquets X, Moncaleán P, Fernández B, Cunillera N, Boronat A and Ferrer A (2002). Overexpression of *Arabidopsis thaliana* farnesyl diphosphate synthase (FPS1S) in transgenic Arabidopsis induces a cell death/senescence-like response and reduced cytokinin levels. Plant J. **30**: 123–132.

Maudinas B, Bucholtz ML, Papastephanou C, Katiyar SS, Briedis AV and Porter JW (1977). The partial purification and properties of a phytoene synthesizing enzyme system. Arch. Biochem. Biophys. **180**: 354–62.

McGarvey DJ and Croteau R (1995). Terpenoid Metabolism. Plant Cell **7**: 1015–1026.

Nagel R, Bernholz C, Vranova E, Kosuth J, Bergau N, Ludwig S, Wessjohann L, Gershenzon J, Tissier A and Schmidt A (2015). *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate, geranylfarnesyl diphosphate. Plant J **84:** 847-859.

Ogura K and Koyama T (1998). Enzymatic aspects of isoprenoid chain elongation. Chem. Rev. **98**: 1263–1276.

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

Phillips MA, D'Auria JC, Gershenzon J, Pichersky E (2008). The *Arabidopsis thaliana* type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. Plant Cell **20:** 677-696.

Poulter CD (2006). Farnesyl diphosphate synthase. A paradigm for understanding structure and function relationships in E-polyprenyl diphosphate synthases. Phytochem. Rev. **5**: 17–26.

Pulido P, Perello C and Rodriguez-Concepcion M (2012). New insights into plant isoprenoid metabolism. Mol Plant **5:** 964-967.

Rai A, Smita SS, Singh AK, Shanker K and Nagegowda DA (2013). Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. Mol. Plant **6**: 1531–1549.

Richter A, Seidl-Adams I, Köllner TG, Schaff C, Tumlinson JH and Degenhardt J (2015). A small, differentially regulated family of farnesyl diphosphate synthases in maize (*Zea mays*) provides farnesyl diphosphate for the biosynthesis of herbivore-induced sesquiterpenes. Planta **241**: 1351–61.

Rodriguez-Concepcion M (2004). Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during arabidopsis seedling development. Plant Cell Online **16**: 144–156.

Rodriguez-Concepcion M and Boronat A (2015). Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. Curr Opin Plant Biol **25:** 17-22.

Rodriguez-Concepcion M and 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.

Ruiz-Sola MA, Coman D, Beck G, Barja MV, Colinas M, Graf A, Welsch R, Rutimann P, Buhlmann P, Bigler L, Gruissem W, Rodriguez-Concepcion M and Vranova E (2016a). Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol **209:** 252-264. *See* **Annex 1**.

Ruiz-Sola MA, Barja MV, Manzano D, Llorente B, Schipper B, Beekwilder J and Rodriguez-Concepcion M (2016b). A single Arabidopsis gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. Plant Physiol. **172**: 1393–1402. *See* **Annex 2**.

Ruppel NJ, Kropp KN, Davis PA, Martin AE, Luesse DR and Hangarter RP (2013). Mutations in *GERANYLGERANYL DIPHOSPHATE SYNTHASE 1* affect chloroplast development in *Arabidopsis thaliana* (*Brassicaceae*). Am J Bot **100:** 2074-2084.

Sallaud C, Rontein D, Onillon S, Jabes F, Duffe P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, Causse M and Tissier A (2009). A novel pathway for sesquiterpene biosynthesis from z,z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. Plant Cell **21**: 301–317.

van Schie CCN, Ament K, Schmidt A, Lange T, Haring MA and Schuurink RC (2007). Geranyl diphosphate synthase is required for biosynthesis of gibberellins. Plant J. **52**: 752–62.

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL and Pichersky E (2009). Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proc. Natl. Acad. Sci. **106**: 10865–10870.

Schmidt A and Gershenzon J (2008). Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). Phytochemistry **69**: 49–57.

Schmidt A, Wachtler B, Temp U, Krekling T, Seguin A and Gershenzon J (2010). A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. Plant Physiol. **152**: 639–655.

Schuhr CA, Radykewicz T, Sagner S, Latzez C, Zenk MH, Arigoni D, Bacher A, Rohdich F and Eisenreich W (2003). Quantitative assessment of crosstalk between the two isoprenoid biosynthesis pathways in plants by NMR spectroscopy. Phytochem. Rev. **2**: 3–16.

Scolnik PA and Bartley GE (1995). Nucleotide sequence of a putative geranylgeranyl pyrophosphate synthase (GenBank L40577) from Arabidopsis (PGR95-018). Plant Physiol. **108**: 1343.

Scolnik PA and Bartley GE (1994). Nucleotide sequence of an Arabidopsis cDNA for geranylgeranyl pyrophosphate synthase. Plant Physiol. **104**: 1469–70.

Scolnik PA and Bartley GE (1996). Two more members of Arabidopsis geranylgeranyl pyrophosphate synthase family (PGR 96–014). Plant Physiol. **110**: 1435.

Studer RA and Robinson-Rechavi M (2009). How confident can we be that orthologs are similar, but paralogs differ? Trends Genet. **25**: 210–6.

Tarshis LC, Sacchettini JC, Yan M and Poulter CD (1994). Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-å resolution. Biochemistry **33**: 10871–10877.

Tholl D (2015). Biosynthesis and biological functions of terpenoids in plants. In Advances in biochemical engineering/biotechnology, pp. 63–106.

Tholl D, Kish CM, Orlova I, Sherman D, Gershenzon J, Pichersky E and Dudareva N (2004). Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. Plant Cell **16**: 977–92.

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

Vickers CE, Bongers M, Liu Q, Delatte T and Bouwmeester H (2014). Metabolic engineering of volatile isoprenoids in plants and microbes. Plant. Cell Environ. **37**: 1753–1775.

Vranova E, Coman D and Gruissem W (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol **64:** 665-700.

Vranova E, Coman D and Gruissem W (2012). Structure and dynamics of the isoprenoid pathway network. Mol. Plant **5**: 318–333.

Wang C, Chen Q, Fan D, Li J, Wang G and Zhang P (2016). Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in *Brassicaceae* plants. Mol Plant **9:** 195-204.

Wang G and Dixon RA (2009). Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci U S A **106:** 9914-9919.

Wang G, Tian L, Aziz N, Broun P, Dai X, He J, King A, Zhao PX and Dixon RA (2008). Terpene biosynthesis in glandular trichomes of hop. Plant Physiol. **148**: 1254–66.

Wang J, Lin HX, Su P, Chen T, Guo J, Gao W and Huang LQ (2018a). Molecular cloning and functional characterization of multiple geranylgeranyl pyrophosphate synthases (ApGGPPS) from *Andrographis paniculata*. Plant Cell Rep. 38: 117-128.

Wang Q, Huang XQ, Cao TJ, Zhuang Z, Wang R and Lu S (2018b). Heteromeric geranylgeranyl diphosphate synthase contributes to carotenoid biosynthesis in ripening fruits of red pepper (*Capsicum annuum* var. conoides). J. Agric. Food Chem. **66:** 11691–11700.

Zhang M, Su P, Zhou YJ, Wang XJ, Zhao YJ, Liu YJ, Tong YR, Hu TY, Huang LQ and Gao W (2015). Identification of geranylgeranyl diphosphate synthase genes from *Tripterygium wilfordii*. Plant Cell Rep. **34**: 2179–88.

Zhou F, Wang CY, Gutensohn M, Jiang L, Zhang P, Zhang D, Dudareva N and Lu S (2017). A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. Proc. Natl. Acad. Sci. U. S. A. **114**: 6866–6871.

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

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

ANNEX 1

PUBLICATION

Ruiz-Sola MÁ, Coman D, Beck G, Barja MV, Colinas M, Graf A, Welsch R, Rütimann P, Bühlmann P, Bigler L, Gruissem W, Rodríguez-Concepción M and Vranová E (2016a). Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol. **209**: 252–264.

NOTE: The role of the PhD candidate in this work was to perform the experiments included in Figure S3e and Figure S4, which were requested by the journal editor after the initial round of reviews.

New Phytologist

Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids

M. Águila Ruiz-Sola[1]*, Diana Coman[2]*, Gilles Beck[2], M. Victoria Barja[1], Maite Colinas[2], Alexander Graf[2], Ralf Welsch[3], Philipp Rütimann[4], Peter Bühlmann[4], Laurent Bigler[5], Wilhelm Gruissem[2], Manuel Rodríguez-Concepción[1] and Eva Vranová[2,6]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, Barcelona 08193, Spain; [2]Department of Biology, Plant Biotechnology, ETH Zurich, Universitätstrasse 2, Zurich 8092, Switzerland; [3]Faculty of Biology II, University of Freiburg, Freiburg 79104, Germany; [4]Department of Mathematics, Seminar for Statistics, ETH Zurich, Rämistrasse 101, Zurich 8092, Switzerland; [5]Department of Chemistry, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland; [6]Present address: Pavol Jozef Šafárik University in Košice, Faculty of Science, Institute of Biology and Ecology, Mánesova 23, Košice 04154, Slovakia

Authors for correspondence:
Eva Vranová
Tel: +421 55 234 2319
Email: eva.vranova@upjs.sk

Manuel Rodríguez-Concepción
Tel: +34 935 636 600 ext. 3222
Email: manuel.rodriguez@cragenomica.es

Received: *5 June 2015*
Accepted: *30 June 2015*

New Phytologist (2016) **209:** 252–264
doi: 10.1111/nph.13580

Key words: *Arabidopsis thaliana*, gene paralogues, geranylgeranyl diphosphate synthase, GGPPS11.

Summary

- Most plastid isoprenoids, including photosynthesis-related metabolites such as carotenoids and the side chain of chlorophylls, tocopherols (vitamin E), phylloquinones (vitamin K), and plastoquinones, derive from geranylgeranyl diphosphate (GGPP) synthesized by GGPP synthase (GGPPS) enzymes. Seven out of 10 functional GGPPS isozymes in *Arabidopsis thaliana* reside in plastids. We aimed to address the function of different GGPPS paralogues for plastid isoprenoid biosynthesis.
- We constructed a gene co-expression network (GCN) using *GGPPS* paralogues as guide genes and genes from the upstream and downstream pathways as query genes. Furthermore, knock-out and/or knock-down *ggpps* mutants were generated and their growth and metabolic phenotypes were analyzed. Also, interacting protein partners of GGPPS11 were searched for.
- Our data showed that *GGPPS11*, encoding the only plastid isozyme essential for plant development, functions as a hub gene among *GGPPS* paralogues and is required for the production of all major groups of plastid isoprenoids. Furthermore, we showed that the GGPPS11 protein physically interacts with enzymes that use GGPP for the production of carotenoids, chlorophylls, tocopherols, phylloquinone, and plastoquinone.
- GGPPS11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. Both gene co-expression and protein–protein interaction likely contribute to the channeling of GGPP by GGPPS11.

Introduction

Isoprenoids are the most functionally and structurally diverse group of plant metabolites reported to date. They are produced in all living organisms, but their abundance and variety in plants is unparalleled (Croteau *et al.*, 2000; Bouvier *et al.*, 2005; Pulido *et al.*, 2012). From the 10s of 1000s of plant isoprenoid compounds, only a few can be considered as 'primary' metabolites, that is, those that are essential for plant function and are therefore common to all plant species. These include molecules involved in respiration, photosynthesis, and regulation of growth and development (Fig. 1). The others are specialized metabolites whose biosynthesis is usually restricted to specific plant families or even to particular species. They typically function in protecting plants against herbivores and pathogens, in attracting pollinators and seed-dispersing animals, and as allelochemicals that influence competition among plant species. Large numbers of specialized isoprenoid metabolites have a commercial value as flavours, pigments, polymers, or drugs (Gershenzon & Dudareva, 2007; Kirby & Keasling, 2009).

Despite their structural and functional diversity, all isoprenoids are derived from the same five-carbon (C5) precursors, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), also called isoprene units (Fig. 1). The addition of IPP units to DMAPP generates prenyl diphosphate molecules of increasing size such as geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15) and geranylgeranyl diphosphate (GGPP, C20). These are the starting points for the production of the huge variety of isoprenoids found in plants. Consistent with the compartmentalization of most metabolic

*These authors contributed equally to this work.

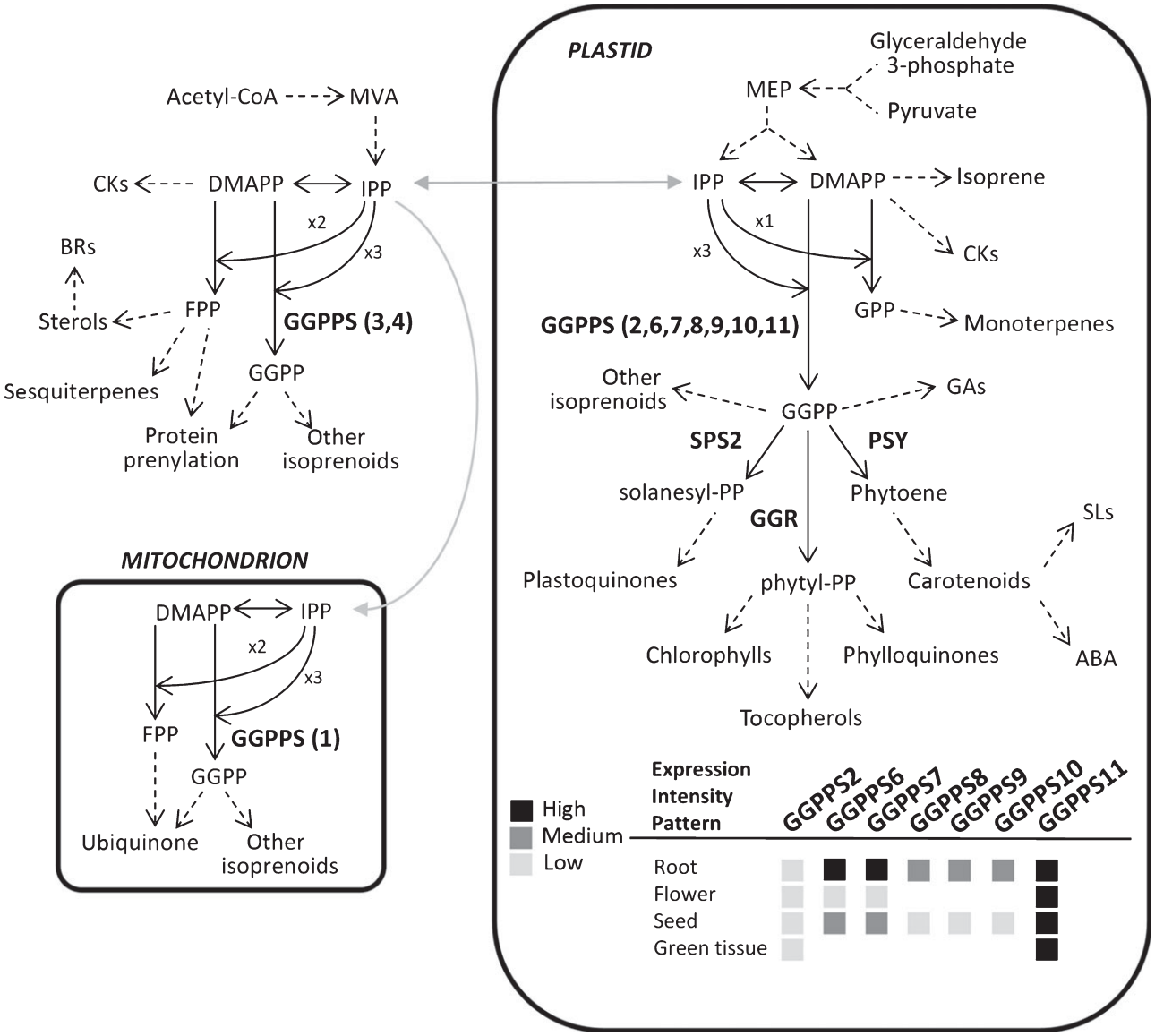


Fig. 1 Isoprenoid biosynthesis in plant cells. The universal C5 isoprenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) can be synthesized by the MVA pathway in the cytosol or the methylerythritol 4-phosphate (MEP) pathway in plastids, and then transported among cell compartments. Addition of IPP molecules to DMAPP produces prenyl diphosphates of increasing chain length, such as geranyl diphosphate (GPP, C10), farnesyl diphosphate (FPP, C15) and geranylgeranyl diphosphate (GGPP, C20). These are the starting points for the production of particular groups of isoprenoids, including the plant hormones brassinosteroids (BRs), cytokinins (CKs), gibberellins (GAs), abscisic acid (ABA) and strigolactones (SLs). GGPP is synthesized in different compartments by GGPP synthase (GGPPS) enzymes (the numbers refer to the *Arabidopsis thaliana* isoforms). Some of the enzymes that channel GGPP to specific isoprenoid pathways are indicated: SPS2, solanesyl diphosphate synthase 2; GGR, geranylgeranyl reductase; PSY, phytoene synthase. Dashed arrows indicate multiple steps. Grey arrows represent transport of isoprenoid precursors between cell compartments. Plastid GGPPS paralogues expression pattern (i.e. root, flower, seed and other plant organs) and expression intensity (i.e. black squares, high; dark grey squares, medium; light grey squares, low) are shown in the pictogram.

pathways in plants (Lunn, 2007), different steps of plant isoprenoid biosynthesis can take place in different plant tissues and subcellular compartments. All land plants use two different pathways for the production of the same universal isoprene units (Fig. 1). The mevalonic acid (MVA) pathway synthesizes cytosolic IPP for the production of sterols, brassinosteroids, sesquiterpenes, and prenyl moieties used for protein modification. MVA-derived IPP is also transported to mitochondria for the biosynthesis of ubiquinone (Disch *et al.*, 1998). Plastid IPP and DMAPP precursors are synthesized by the methylerythritol 4-phosphate (MEP) pathway (Fig. 1).

The main groups of plastid isoprenoids, including photosynthesis-related metabolites such as carotenoids and the side chain of chlorophylls, tocopherols (vitamin E), phylloquinones (vitamin K), and plastoquinones, are derived from GGPP synthesized by GGPPS enzymes (Fig. 1). Twelve paralogous *GGPPS* genes (*GGPPS1–GGPPS12*) exist in the genome of the model plant *Arabidopsis thaliana* (Lange & Ghassemian, 2003). However,

GGPPS12 (At4g38460) does not show GGPPS activity and *GGPPS5* (At3g14510) is likely a pseudogene (Okada *et al.*, 2000; Wang & Dixon, 2009; Beck *et al.*, 2013; Coman *et al.*, 2014). The remaining 10 GGPPS isozymes produce GGPP *in vitro* and/or *in vivo* and localize to different subcellular compartments (Zhu *et al.*, 1997a,b; Okada *et al.*, 2000; Wang & Dixon, 2009; Beck *et al.*, 2013). GGPPS2 (At2g18620), GGPPS6 (At3g14530), GGPPS7 (At3g14550), GGPPS8 (At3g20160), GGPPS9 (At3g29430), GGPPS10 (At3g32040) and GGPPS11 (At4g36810) are plastid enzymes. Among these plastid isoforms, only *GGPPS2* and *GGPPS11* are ubiquitously expressed, but *GGPPS11* produces much higher mRNA levels than any of the other paralogues in most organs, especially in photosynthetic tissues (Fig. 1; Beck *et al.*, 2013). Expression of genes encoding the GGPPS6, 7, 8, 9, and 10 isoforms was confined to specific organs and developmental stages, with higher levels in roots, developing seeds, and flowers (in particular, GGPPS6 and 7). Together, the *Arabidopsis GGPPS* paralogues show significantly different quantitative and tissue-specific expression patterns. Based on their expression patterns, we hypothesized that GGPPS11 might be responsible for the housekeeping production of GGPP in chloroplasts, whereas other plastid-localized GGPPS isozymes might have specific and/or minor roles in the tissues where they are expressed (Beck *et al.*, 2013).

To address the relevance of the different *GGPPS* paralogues for plastid isoprenoid biosynthesis in *Arabidopsis*, we constructed a gene co-expression network (GCN) and linked individual enzymes to GGPP-consuming downstream pathways and to the upstream pathways. Furthermore, knock-out and/or knock-down mutants for each plastid *GGPPS* gene were generated and their growth and metabolic phenotypes were analyzed. Our results show that *GGPPS11* is a hub gene and GGPPS11 is required for the production of all major groups of plastid isoprenoids and is the only plastid isozyme essential for plant development. Consistent with the connection at the gene expression level unveiled by the GCN, we also found that GGPPS11 can physically interact with enzymes that use GGPP for the production of photosynthesis-related isoprenoids, suggesting a mechanism for the channeling of GGPP to specific downstream pathways.

Materials and Methods

Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh plants were grown either on Murashige-Skoog (MS) medium (Duchefa) containing 0.8% w/v plant agar or on soil in a climate-controlled growth chamber at 22°C under long-day (16 h : 8 h, light : dark) or short-day (8 h : 16 h, light : dark) conditions. For methyl viologen treatments, seeds were surface-sterilized and sown on a sterile mesh of filter paper or synthetic fabric (Sefar Nitex 03-100/44) on top of solid MS medium in square culture dishes. Following stratification for 3 d at 4°C in the dark, plates were incubated vertically at 22°C under long-day photoperiod for 5 d. Then, the mesh with the plants was transferred to new plates containing solid MS medium either supplemented or not with 5 μM methyl viologen and grown under long-day conditions for 5 additional days. Seeds from *Arabidopsis* insertion lines belonging to SALK, SAIL, FLAG, and SM collections (Supporting Information Table S1; Tissier *et al.*, 1999; Samson *et al.*, 2002; Sessions *et al.*, 2002; Alonso *et al.*, 2003) were obtained from Torrey Mesa Research Institute (http://www.syngenta.com/), from the European *Arabidopsis* Stock Center (http://arabidopsis.info/), and from the INRA (http://urgv.evry.inra.fr/FLAGdb). The original genetic background of these lines is indicated in Table S1. The loss-of-function T-DNA lines FLAG134B10 (*ggpps2*), SAIL1148_A03 (*ggpps6*), and FLAG470E09 (*ggpps8*) were back-crossed four times to the Columbia (Col-0) accession. Then, heterozygous plants were allowed to self-pollinate and the homozygous lines were identified in the progeny by segregation analysis and PCR-based genotyping (see Table S2 for primers). A *GGPPS9*-specific 95-bp fragment in the 3′-UTR was amplified (see Table S2 for PCR primers) and cloned into the pHellsgate8 vector (Helliwell & Waterhouse, 2003) to generate lines defective in *GGPPS9* by RNA interference (RNAi; see Table S3 for cloning details). Constructs containing the *GGPPS11* promoter and/or coding region were also constructed as described in Table S3 and then used for *Agrobacterium*-mediated transformation of *Arabidopsis* wild-type (Col-0) and GGPPS11-defective mutant plants.

Homozygous transgenic lines containing a single T-DNA insertion were selected based on the segregation of the corresponding resistance marker. In the case of T-DNA insertion mutants, the chi-squared goodness of fit test was performed with 1 degree of freedom and 95% interval of confidence to verify the Mendelian segregation of the associated resistance (Table S1). In the case of the *ggpps11-5* mutant, where silencing of the kanamycin resistance gene was observed, PCR-based genotyping results were scored. In the case of the null *ggpps11-3* and *ggpps11-4* mutants, the number of green vs brown seeds was scored. Homozygous mutant lines were confirmed by PCR-based genotyping (Fig. S1a; Table S2). In most cases, reverse transcription RT-PCR was used to confirm the absence of transcripts in mutant lines using cDNA isolated from roots as template and primers spanning the T-DNA insertion site (Fig. S1a; Table S2). When necessary to amplify low abundance transcripts, nested PCR was carried out using 1 μl of the RT-PCR product as template (Fig. S1a; Table S2). The transcript down-regulation in *GGPPS9* RNAi lines (Fig. S1b) was assessed by TaqMan® RT-qPCR (Applied Biosystems, Foster City, CA, USA) using cDNA isolated from roots as template, and UPL6 (Universal Probe Library, F. Hoffmann-La Roche, Basel, Switzerland). The transcript down-regulation in *ggpps11-5* plants was assessed as described (Rodriguez-Villalon *et al.*, 2009). Primer sequences are listed in Table S2. The ImageJ software (http://imagej.nih.gov/ij/) was used for measurements of plant morphological traits (e.g. cotyledon length).

GCN analysis

The GCN was generated as described in Coman *et al.* (2014). In brief, the AtIPD database (Vranová *et al.*, 2011) was used to assemble a list comprising the guide genes (*GGPPS2*, *GGPPS6–GGPPS11*) and the query genes (genes encoding the MVA and the

MEP pathway enzymes and enzymes from downstream biosynthetic pathways). The Arabidopsis Developmental Baseline dataset generated within the AtGenExpress Consortium was used (Schmid *et al.*, 2005). Next, the pairwise Pearson correlation between each guide and query genes was computed and Fisher's *Z*-transformation was employed to test if the pairwise correlations were significant. The family wise error was controlled using Holm–Bonferroni correction. The *GGPPS* GCN was built based on the corrected *P*-values from the significance test as estimator of significant positive co-expression between pairs of genes (i.e. *P*-values ≤ 0.05 indicate significantly co-expressed genes). The GGPPS GCN was displayed as undirected graph with nodes representing genes and edges representing significant correlation between pair of genes (i.e. *P*-value ≤ 0.05) using Cytoscape (Shannon *et al.*, 2003).

Microscopy

Clearing of *Arabidopsis* ovules was performed as described (Stangeland & Salehian, 2002) with some modifications: Hoyer's solution contained 66.7 g chloral hydrate and 8.3 ml glycerol dissolved in 25 ml water. Siliques were dissected with hypodermic needles or ovules were taken out of siliques and cleared for several hours or overnight. Embryo development was studied microscopically with a Zeiss Axioplan 2 microscope equipped with differential interference contrast optics. Pictures were taken with the connected Axiocam Hrc (Carl Zeiss, Jena, Germany).

For determination of chloroplast size and abundance, wild-type and *ggpps11-5* lines were grown for 7 d under long-day conditions. Then, whole seedlings were embedded in 5% (w/v) agarose blocks and 200 µm cross-sections of their cotyledons were obtained with a Vibratome Series 1000 Sectioning System. Sections were observed with a Olympus FV 1000 Confocal Laser Scanning Microscope (Tokyo, Japan). Chloroplasts were identified based on chlorophyll autofluorescence. Pictures of chloroplasts from different stomata and mesophyll regions were taken and used to count chloroplasts with the ImageJ software. To estimate chloroplast area, only the largest chloroplast of each guard cell and the five largest chloroplasts in the pictures of the mesophyll region were used for ImageJ calculations.

Extraction and analysis of metabolites

Seedlings were grown for 10 d on MS medium supplemented with 1% of sucrose and frozen in liquid nitrogen before analysis. Frozen plant material from seven biological replicates was ground and 100 mg samples were resuspended in 1 ml of extraction mixture 1 (methanol/chloroform/water 2.5 : 1 : 1) containing two internal standards at 1 µg ml^{-1} concentration, corticosterone and C17-choline. After mixing and sonication, samples were centrifuged for 3 min at 20 000 ***g*** at 4°C. The supernatants were transferred to fresh tubes and 200 µl water and 400 µl chloroform were added. The tubes were briefly vortexed and then centrifuged again. Approximately 500 µl of the organic lower phase was transferred to a fresh tube. The pellet resulting from the first extraction was re-extracted with 300 µl of extraction mixture 2 (2-propanol/hexane/water 5 : 2 : 2.5). Samples were vortexed, sonicated, and centrifuged and the resulting supernatants were pooled to the organic phase. Extracts were dried by a speed-vacuum centrifugation at 20°C for *c.* 2 h. The dry pellets were then resuspended in 200 µl of a 7 : 3 chloroform/2-propanol solution. After vortexing, sonication, and centrifugation, 140 µl were transferred to a glass vial for UPLC-MS analysis in a system consisting of a Waters Acquity ultra high performance liquid chromatograph (UPLC) and a Bruker quadrupole time-of-flight (QTOF) high-resolution mass spectrometer equipped with electrospray ionization source. An Acquity BEH C8 2.1 × 100 mm, 1.7 µm column from Waters was used in the study. The two solvents used for gradient elution were A (H_2O +1% ammonium acetate + 0.1% acetic acid) and B (acetonitrile/isopropanol (7 : 3) + 1% ammonium acetate + 0.1% acetic acid). A further metabolic analysis of *ggpps11-5* plants was performed by extracting and quantifying photosynthetic pigments (chlorophylls and carotenoids) and prenylquinones as described (Rodríguez-Concepción *et al.*, 2004; Martinis *et al.*, 2011).

Yeast two-hybrid assays

The split-ubiquitin system was used as described (Obrdlik *et al.*, 2004). ORFs were truncated by the length of putative transit peptides predicted by ChloroP (Emanuelsson *et al.*, 1999) which was 55 aa for G11, 57 aa for phytoene synthase (PSY), 43 aa for GGR and 59 aa for SPS2, as described in Table S3. The cDNAs were cloned into pNXgate for Nub-GGPS11 or pMetYCgate for Cub fusions and transformed into the yeast strain THY.AP4 (Nub) or THY.AP5 (Cub), respectively. Separate strains carrying Nub and Cub fusions were mated and the resulting diploid cells were cultured in synthetic complete medium lacking leucin and tryptophane. Interaction growth tests were performed on synthetic minimal agar, supplemented with 150 µM methionine for PSY-Cub combinations to reduce a weak background activation of the reporter gene. For β-galactosidase assays and phytoene extraction, yeasts were cultivated overnight in synthetic complete medium supplemented with adenine and histidine at 28°C. β-Galactosidase activity was determined in triplicates as described (Obrdlik *et al.*, 2004) and expressed relative to cell density measured at 600 nm. Phytoene was quantified as described (Welsch *et al.*, 2010).

Bimolecular fluorescence complementation (BiFC)

Constructs for BiFC experiments were generated in pSPYNE(R) 173 and pSPYCE(MR) vectors (Waadt *et al.*, 2008) as described in Table S3. Onion (*Allium cepa*) epidermal peels were microbombarded with DNA-coated 1 µM gold microcarriers using a Biolistic PDS-1000/He system (Bio-Rad) and incubated at 22°C in the dark for 24 h before observation with a Leica (Leica Microsystems CMS GmbH, Mannheim, Germany) TCS 4D Confocal Laser Scanning Microscope.

Gene ID numbers

AGI locus identifiers 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; 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 is At4g38460.

Results

GGPPS11 is a hub gene in the modular *GGPPS* gene co-expression network

Gene co-expression networks (GCNs) may be indicative for modular organization and functional relationships between genes (Oliver, 2000). Genes associated with the same metabolic pathway are highly co-expressed across various spatial, temporal and/or environmental conditions compared to genes from different pathways (Gachon *et al.*, 2005; Wei *et al.*, 2006; Heyndrickx & Vandepoele, 2012). Here we have used a targeted GCN analysis to predict the association of individual plastid GGPPS isozymes with pathways in the isoprenoid metabolism.

As input expression dataset for the targeted GCN analysis we used the Developmental Baseline microarray compendium, which contains gene expression information for various stages of *Arabidopsis* development from embryogenesis to senescence as well as for various organs and tissues (Schmid *et al.*, 2005). As guide genes we used the *GGPPS* genes that encode plastid isozymes (*GGPPS2* and *GGPPS6* to *GGPPS11*). As query genes, we used the genes encoding the enzymes from the plastid MEP pathway and the cytosolic MVA pathway and the enzymes from downstream plastid biosynthetic pathways for the production of carotenoids, chlorophylls, plastoquinone, phylloquinone, tocopherols, gibberellins, abscisic acid (ABA) and strigolactones (SLs; Table S4). To build the targeted GCN, the correlation of transcript profiles between guide and query genes was assessed and tested for statistical significance (Coman *et al.*, 2014).

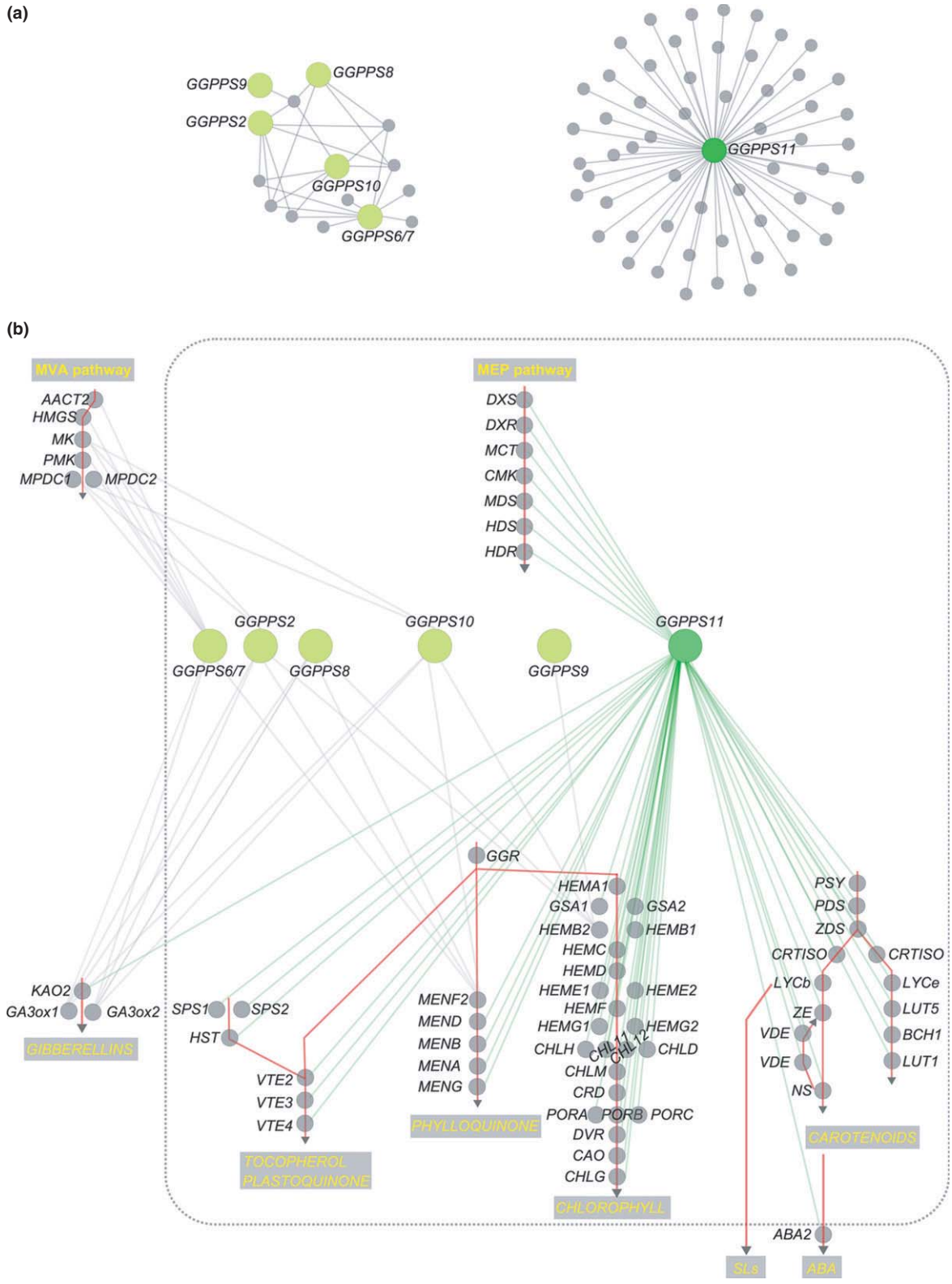
The *GGPPS* GCN has in total 73 nodes and 83 edges representing significant co-expression (P-value ≤ 0.05; Table S5) between the *GGPPS* genes encoding plastid isozymes and genes from the MEP pathway, the MVA pathway and biosynthetic pathways downstream of GGPPS (Fig. 2a). The *GGPPS* GCN shows a modular organization. The major component in the network with the highest connectivity is *GGPPS11*, which accounts for 68.6% of total edges in the GCN. The remaining GGPPS paralogues on average account for only 6.2% of total edges (Fig. 2a). The high node degree of *GGPPS11* indicates the central relevance of this gene for GGPP synthesis in plastids. The small node degree of *GGPPS2* and *GGPPS6* to *GGPPS10* suggests a minor role for these paralogues in *Arabidopsis* plastid GGPP synthesis, in agreement with gene expression profiles (Beck *et al.*, 2013).

We next mapped GCN onto the isoprenoid metabolic pathway network (Vranová *et al.*, 2011) to determine the involvement of *GGPPS11* and other *GGPPS* paralogues in different pathway branches. *GGPPS11*, which is the most ubiquitously expressed plastid *GGPPS* paralogue and the only one highly expressed in green tissues (Fig. 1; Beck *et al.*, 2013), is significantly co-expressed with all genes encoding MEP pathway enzymes and almost all genes from pathways synthesizing photosynthesis-related isoprenoids (Fig. 2b; Table S6). This is in agreement with similar network studies of the *GGPPS* genes that used different data sets and calculation methods (Wille *et al.*, 2004; Gilbert *et al.*, 2009; Meier *et al.*, 2011; Lonzano & Świrszcz, 2012; Yang *et al.*, 2014). Only few significant connections exist between *GGPPS11* and genes from the hormone biosynthetic pathways (gibberellins, ABA and strigolactones).

The other plastid *GGPPS* genes are significantly co-expressed with only a few genes from downstream plastid pathways. *GGPPS2, GGPPS6/GGPPS7, GGPPS8* and *GGPPS10* have similar co-expression pattern and they are co-expressed with genes encoding gibberellin 3-oxidase 1 (*GA3ox1, At1g15550*) and gibberellin 3-oxidase 2 (*GA3ox2, At1g80340*) from the GA biosynthetic pathways and with isochorismate synthase 2 (*ICS2, MENF2, At1g18870*) from the phylloquinone biosynthetic pathway. *GGPPS2* and *GGPPS8* to *GGPPS10* are co-expressed with the 5-aminolevulinate dehydratase (*HEMB2, At1g443180*) from the chlorophyll biosynthetic pathway (Fig. 2b). Interestingly, *GGPPS2, GGPPS6/GGPPS7* and *GGPPS10* are co-expressed with genes encoding enzymes in the cytosolic MVA pathway (Fig. 2) but not with genes encoding enzymes in the MEP pathway. Cross-compartment transport of prenyl diphosphate intermediates has been reported (Bick & Lange, 2003; Flügge & Gao, 2005) and therefore GGPPS2, GGPPS6/GGPPS7 and GGPPS10 might use precursors derived from the cytosolic MVA pathway. This would be consistent with the expression patterns of the *GGPPS2*, *GGPPS6/GGPPS7* and *GGPPS10* genes, which appear to be confined to nongreen tissues (Beck *et al.*, 2013). Yet, GGPPS11 might be the main isozyme that uses IPP and DMAPP precursors produced via the MEP pathway for GGPP biosynthesis and the main one contributing GGPP for the synthesis of photosynthesis-related isoprenoids (Fig. 2; Table S6).

Mutants in the *GGPPS* genes confirm essentiality of GGPPS11

To test if the function of GGPPS11 is essential compared with other plastid GGPPS, we established a collection of full loss-of-function mutants for plastid *GGPPS* genes (Figs 3, S1; Table S1). Homozygous lines with a single T-DNA inserted in the coding region could be obtained for all genes except *GGPPS9* and *GGPPS11* (see the Materials and Methods section and Fig. S1a). The absence of detectable full-length transcripts in *ggpps2*, *ggpps6*, *ggpps7*, *ggpps8* and *ggpps10* mutants was confirmed by standard reverse transcription PCR (RT-PCR), indicating that all are null mutants (see the Materials and Methods section and Fig. S1a). The two available T-DNA insertion mutants for *GGPPS11* were *ggpps11-3* (SALK_085914) and *ggpps11-4* (SAIL_712_D06) according to the annotation by Ruppel *et al.* (2013). Because no T-DNA insertion line for *GGPPS9* could be identified in public repositories, we generated RNAi lines for this paralogue (see the Materials and Methods section and Fig. S1b).



A 95 nucleotide tag located in the 3′UTR region of *GGPPS9* was selected to ensure specific targeting of this paralogue. Two independent homozygous lines with the lowest transcript levels were selected for further analysis, the *ggpps9-1* and *ggpps9-6* RNAi lines with 21% and 16% of transcript wild-type levels, respectively (Fig. S1b).

All lines used in this work were either originally in the Col-0 background or back-crossed into Col-0 to ensure comparability

Fig. 2 The *Arabidopsis GGPPS* gene co-expression network. Genes are shown as nodes and statistically significant positive co-expression relationships (*P*-value ≤ 0.05) are shown as edges. The *GGPPS* genes encoding plastid isozymes are indicated in green (dark green, *GGPPS11*; *GGPPS2*; light green, *GGPPS6-10*). *GGPPS6* and *GGPPS7* are ambiguously amplified by the same probe set on the Affymetrix™ (Santa Clara, CA, USA). microarray and are referred here to as '*GGPPS6/7*'. The grey nodes represent the genes encoding enzymes from the methylerythritol 4-phosphate (MEP) and the mevalonic acid (MVA) pathways and from biosynthetic pathways located downstream of geranylgeranyl diphosphate synthase (GGPPS). Only significant co-expression edges between *GGPPSs* and genes from isoprenoid pathway are shown (i.e. co-expression between *GGPPs* are not depicted here). (a) The modular co-expression network of the *GGPPSs* encoding plastid isozymes during *Arabidopsis* development is shown. *GGPPS11* has the highest connectivity and is a hub in the gene co-expression network envisioning an essential role owed to its high nonoverlapping connectivity. The remaining paralogues form overlapping modules. (b) The co-expression network of the *GGPPSs* encoding plastid isozymes, mapped onto the metabolic pathway of isoprenoid biosynthesis is shown. The plastid compartment is delimited by the grey dotted line. The MEP and the MVA pathways upstream of GGPPS and the downstream biosynthetic pathways that use GGPP as precursor are schematically represented. Multiple nodes aligned horizontally represent isozymes and the direction of the biosynthetic process is indicated by red arrows. GGPPS11 has the highest connectivity (green edges). Gene abbreviations are included in Supporting Information Table S6.

between the mutant lines (see the Materials and Methods section and Table S1). All *ggpps* mutants had segregation ratios consistent with single locus recessive nuclear mutations (Table S1). Mutant lines for GGPPS2, GGPPS6, GGPPS7, GGPPS8, GGPPS9, or GGPPS10 did not show any developmental defects compared to wild-type plants. Homozygous *ggpps11-3* and *ggpps11-4* lines could not be identified in the F2 progeny. Even screening of a larger population of 87 F2 *ggpps11-4* seedlings grown on antibiotic selection media using PCR genotyping revealed no homozygous mutant plants. Similar results were reported by Ruppel *et al.* (2013) from genotyping a segregating *ggpps11-3* population. In siliques of heterozygous *ggpps11-3* and *ggpps11-4* plants, 19% ($n = 2286$) and 22% ($n = 2021$), respectively, of developing seeds were white, and at a later stage brownish and shrunken (Fig. 3b), suggesting defective homozygous embryo and/or seed development. To discriminate between these two possibilities, we initially examined embryo development in siliques from heterozygous plants. Approx. 20% of the embryos in these siliques were found to be arrested at the heart stage (Fig. 3c). The arrested embryos could not develop further and seeds with those aborted embryos eventually collapsed and dried (Fig. 3b). The embryo lethal phenotype of the *ggpps11-4* allele was complemented by expressing a genomic fragment with the *GGPPS11* gene (*G11* minigene), including the promoter and protein coding region (Fig. S2). We therefore conclude that the embryo lethal phenotype of these plants was specifically caused by a loss of GGPPS11 activity and hence that GGPPS11 activity is essential during embryo development.

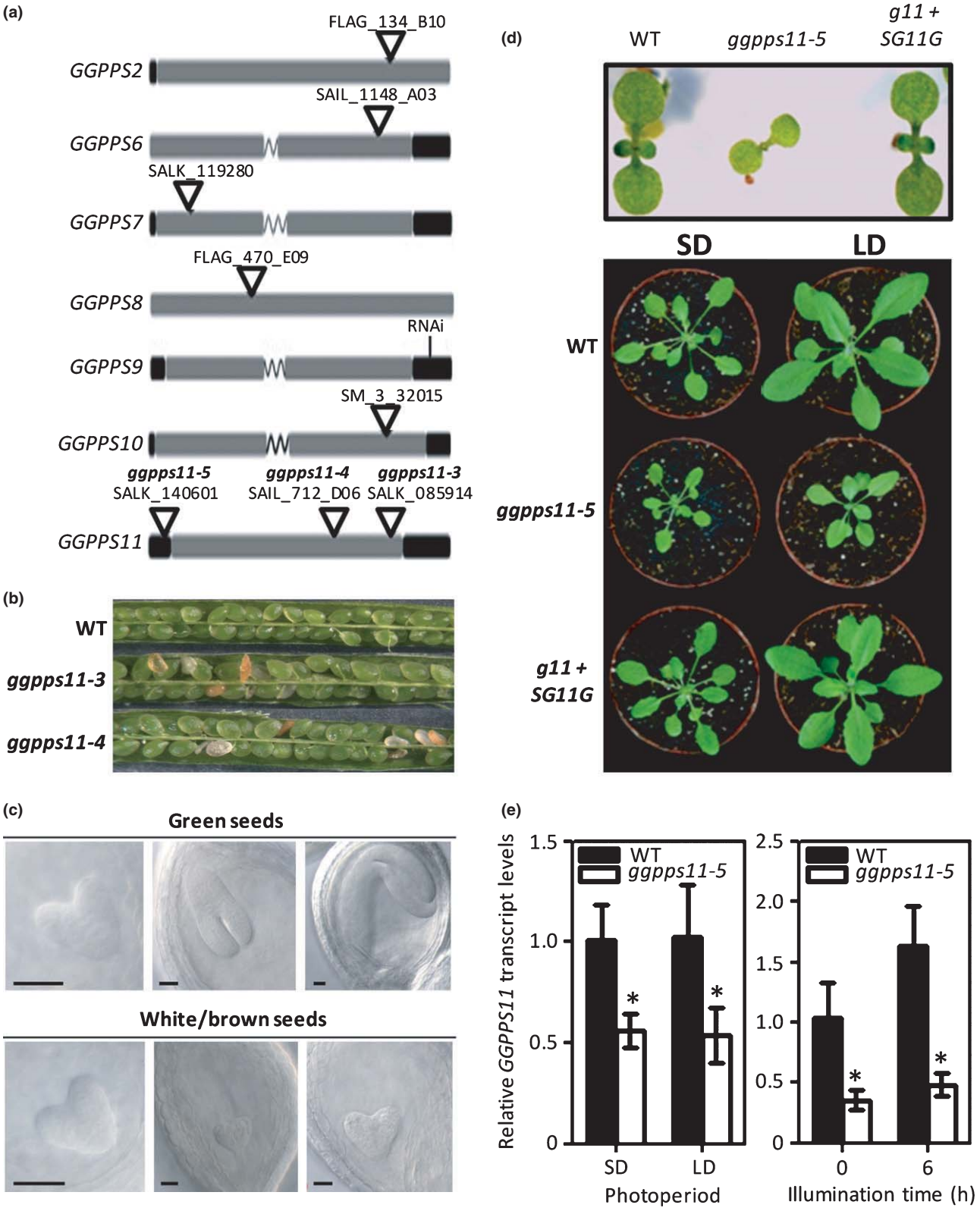
Reduced GGPPS11 levels in the *ggpps11-5* allele result in pale plants with smaller mesophyll chloroplasts

We next searched the SALK collection of T-DNA insertion lines (Alonso *et al.*, 2003) for potential knock-down mutants to understand the contribution of GGPPS11 to the production of plastid isoprenoids. Line SALK_140601 was selected and confirmed to contain a T-DNA 148 bp upstream of the first ATG codon in the *GGPPS11* gene by sequencing the insertion site in the genome of homozygous plants. The new insertion allele was named *ggpps11-5* (Fig. 3a). Unlike the variegated phenotype of the *ggpps11-1* mutant, a likely knock-down allele with a point mutation in the coding region of the *GGPPS11* gene (Ruppel *et al.*, 2013), *ggpps11-5* plants were paler than the wild-type (Figs 3d, S2, S3). But similar to *ggpps11-1* plants, *ggpps11-5* seedlings were smaller than the wild-type, had shorter roots, and showed a developmental delay when growing in long or short days (Figs 3d, S2, S3). The number of chloroplasts appeared to be similar in wild-type and *ggpps11-5* plants (Fig. S4). Also similarly, guard cell chloroplasts in the mutant had a wild-type size. However, chloroplasts in mesophyll cells were found to be smaller in *ggpps11-5* plants (Fig. S4). Consistently, the white sectors of *ggpps11-1* leaves showed an improper development of mesophyll chloroplasts but normal-appearing guard cell chloroplasts (Ruppel *et al.*, 2013).

Reverse transcription quantitative PCR (RT-qPCR) analysis of *GGPPS11* transcript levels in wild-type and *ggpps11-5* seedlings showed substantially reduced levels in mutant plants before and after de-etiolation (Fig. 3e). The visual phenotype of the *ggpps11-5* mutant was fully complemented by expressing either the *G11* minigene (Fig. S2) or the *GGPPS11* coding region fused to the green fluorescent protein (GFP) reporter under the transcriptional control of the constitutive *35S* promoter (*g11 + SG11G* lines; Figs 3d, S3), demonstrating that it was specifically caused by a defective *GGPPS11* expression.

Metabolite analysis confirms the essential role of GGPPS11 in plastid isoprenoid metabolism

We next analyzed the level of photosynthesis-related isoprenoid metabolites in wild-type, *ggpps11-5*, and complemented (*g11 + SG11G*) lines using high pressure liquid chromatography (HPLC) and UPLC-MS (see Materials and Methods). Levels of chlorophylls (*a* and *b*), carotenoids (β-carotene, violaxanthin, neoxanthin, and lutein) and prenylquinones such as tocopherols (α-tocopherol and γ-tocopherol), plastoquinones (plastochromanol-8 and plastoquinone-9) and phylloquinone were significantly reduced in *ggpps11-5* seedlings, whereas complemented lines showed a metabolite profile similar to that of wild-type controls (Figs 4, S5). It is likely that the observed reduction of chloroplast size (Fig. S4) and photosynthesis-related metabolites (Figs 4, S5) in *ggpps11-5* plants could have an impact on photosynthetic rate, as previously reported for another GGPPS11-defective allele (Ruppel *et al.*, 2013). We additionally evaluated whether the metabolic, physiological, and developmental changes associated with a reduction in GGPPS11 activity in the *ggpps11-5* mutant had an impact on the protection against oxidative



stress. We addressed this question by comparing the response of wild-type and mutant plants to treatment with methyl viologen (also known as paraquat, a widely used inductor of oxidative stress in the chloroplast). The response to oxidative stress caused by methyl viologen was similar in wild-type and *ggpps11-5* plants based on the reduction of root growth and shoot fresh weight (Fig. S3). Together, metabolite and phenotypic analysis of the *ggpps11-5* mutant suggests that partial loss of GGPPS11 function

Fig. 3 Collection of *ggpps* mutants. (a) Schematic representation of mutations disrupting the plastid geranylgeranyl diphosphate synthase (GGPPS) genes. *ggpps2* (FLAG_134_B10), *ggpps6* (SAIL_1148_A03), *ggpps7* (SALK_119280), *ggpps8* (FLAG_470_E09), *ggpps9-1* and *ggpps9-6* (RNAi), *ggpps10* (SM_3_32015), *ggpps11-3* (SALK_085914), *ggpps11-4* (SAIL_712_D06) and *ggpps11-5* (SALK_140601). The genomic location of the T-DNA insertion lines and of the gene specific tag selected as target for RNAi mediated silencing, are shown and marked by an arrowhead. Gene models are according to TAIR v10 *Arabidopsis thaliana* genome annotation and are shown in 5′-3′ orientation. Coding regions are shown in grey and untranslated regions are shown in black. (b) Representative images of wild-type (WT) siliques and the segregating population of seeds in siliques of heterozygous *ggpps11-3* and *ggpps11-4* full loss-of-function mutant plants. White and brownish seeds are observed in both mutant backgrounds. (c) Representative images of embryos in the segregating population of seeds found in siliques of heterozygous *ggpps11-4* plants at different stages of development. Images in the same column correspond to green or white/brown seeds of the same silique. Similar results were obtained for the *ggpps11-3* line. Bars, 20 μm. (d) Partial loss-of-function *ggpps11-5* mutant plants show pigmentation and growth defects. Representative images of WT, mutant *ggpps11-5*, and complemented *g11 + SG11G* seedlings grown for 7 d under long-day conditions (LD, upper panel) and for 4 wk under short-day (SD) or 3 wk under LD (lower panel) conditions. (e) Reverse transcription quantitative (RT-q)PCR analysis of *GGPPS11* transcript levels of WT and *ggpps11-5* mutant plants. The graph on the right represents *GGPPS11* transcript levels of 4-d-old etiolated WT and *ggpps11-5* seedlings immediately before and after illumination with white light for 6 h. The graph on the left represents *GGPPS11* transcript levels in WT and *ggpps11-5* seedlings grown for 11 d under SD or 7 d under LD photoperiod. The *UBC/UBC21/PEX4* (At5g25760; Czechowski *et al.*, 2005) gene was used for normalization. Data correspond to mean and standard deviation of $n = 4$ independent samples. Asterisks mark statistically significant differences ($P < 0.05$) relative to WT samples.

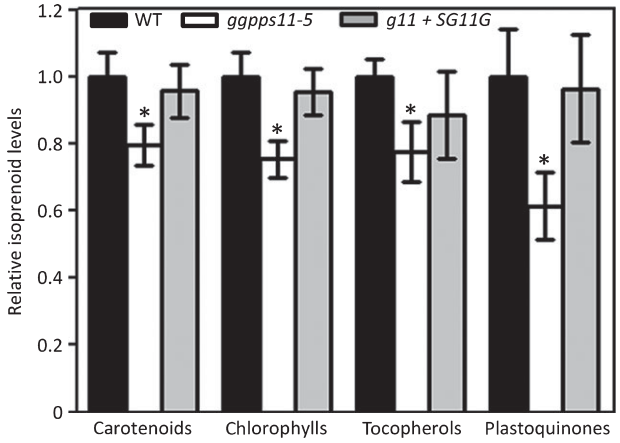


Fig. 4 Plastid isoprenoid levels are reduced in *ggpps11-5* plants. *Arabidopsis thaliana* wild-type (WT), mutant *ggpps11-5*, and complemented *g11 + SG11G* seedlings grown for 7 d under short-day conditions were used to quantify the levels of the indicated groups of plastid isoprenoids. Data of individual metabolites are shown in Fig. S5. Values are shown relative to those found in WT plants and correspond to mean and standard deviation of $n = 8$ (carotenoids and chlorophylls) or $n = 4$ (tocopherols and plastoquinones) independent samples. Asterisks mark statistically significant differences ($P < 0.05$) relative to WT samples. Absolute levels in WT plants are as follows: carotenoids, 137.9 ± 7.0 μg g^{-1} FW; chlorophylls, 177.8 ± 8.9 μg g^{-1} FW; tocopherols, 6.1 ± 0.2 μg g^{-1} FW; plastoquinones, 9.3 ± 0.9 μg g^{-1} FW.

reduces the supply of GGPP for the production of photosynthesis-related metabolites, which in turn would negatively impact chloroplast development and photosynthetic efficiency and eventually cause the growth defects observed in the *ggpps11-5* mutant.

In principle, reduced production of gibberellins (Fig. 1) could also contribute to reduced growth. However, experiments with the *ggpps11-1* mutant allele, which has a stronger growth and pigmentation phenotype compared to *ggpps11-5*, indicated that further reduction in GGPPS11 activity did not significantly affect gibberellin-controlled processes such as seed germination, hypocotyl elongation, or flowering (Ruppel *et al.*, 2013). Consistent with the conclusion that GGPPS11 does not significantly contribute to the biosynthesis of these hormones, mutant *ggpps11-5* seedlings showed wild-type levels of transcripts from most gibberellin producing and inactivating pathway genes used as markers for GA contents (Curaba *et al.*, 2004; Eriksson *et al.*, 2006; Fig. S6).

We also tested whether *ggpps2, ggpps6, ggpps7, ggpps8, ggpps9* and *ggpps10* mutants were compromised in the synthesis of photosynthesis-related isoprenoid metabolites. Profiling of chlorophylls (*a* and *b*), carotenoids (β-carotene, neoxanthin) and prenylquinones (phylloquinone) using UPLC-MS (see the Materials and Methods section) showed that none of the mutants except *ggpps8* (in which a small decrease in chlorophyll *b* level was detected) had significantly decreased levels of these isoprenoid metabolites (Fig. S5). These results are consistent with the GCN prediction that GGPPS11 is the main plastid GGPPS isozyme that produces a common pool of GGPP substrate for the biosynthesis of photosynthesis-related isoprenoids (Fig. 1).

GGPPS11 interacts with plastid enzymes that use GGPP as substrate

GGPPS enzymes have been found in multienzyme complexes containing phytoene synthase (PSY) and other isoprenoid biosynthetic enzymes in chromoplasts (Maudinas *et al.*, 1977; Dogbo *et al.*, 1988; Camara, 1993; Fraser *et al.*, 2000). Based on our GCN analysis and metabolite profiling data, we reasoned that GGPPS11 might interact with PSY but also with other enzymes that use GGPP as a substrate for the production of downstream isoprenoids in *Arabidopsis* chloroplasts (Fig. 1). To test this possibility, we used *g11 + SG11G* plants for immunoprecipitation assays. Transgenic lines expressing a GFP-tagged version of deoxyxylulose 5-phosphate, the first enzyme of the MEP pathway (Pulido *et al.*, 2013), were used as controls for immunoprecipitation experiments with a commercial anti-GFP serum. Analysis of co-immunoprecipitated proteins by mass spectrometry detected the presence of GGR (At1g74470) and SPS2 (At1g78510) only in *g11 + SG11G* samples. GGR produces the phytyl moiety of chlorophylls, tocopherols and phylloquinone (Keller *et al.*, 1998) and, SPS2 specifically converts GGPP into solanesyl diphosphate for the production of photoactive plastoquinone (Block *et al.*, 2013; see Fig. 1). However, PSY was not detected among the co-immunoprecipitated proteins. Furthermore, the

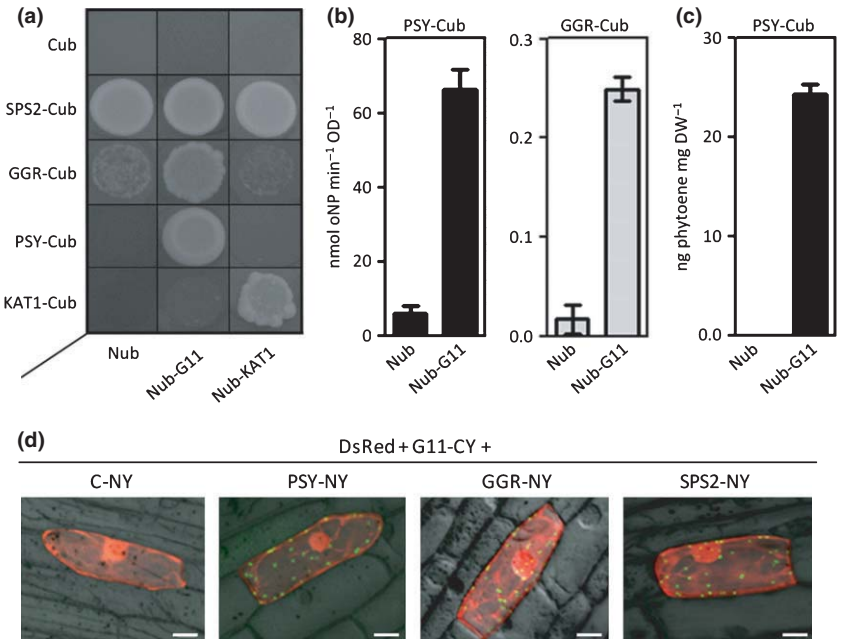


Fig. 5 Geranylgeranyl diphosphate synthase 11 (GGPPS11) interacts with phytoene synthase (PSY), GGR, and SPS2 *in vivo*. (a) Growth of yeast strains co-expressing the indicated proteins on synthetic minimal medium. Empty vectors expressing Nub (N-terminal ubiquitin moiety) or Cub (C-terminal ubiquitin moiety) alone were used as negative controls, and constructs with the KAT1 protein, known to homodimerize, were used as a positive control. Growth indicates interaction. (b) Quantification of interaction by β-galactosidase activity from yeast strains co-expressing the indicated proteins. (c) Levels of phytoene (produced by the PSY enzyme from GGPP) in yeast cells co-expressing the indicated proteins. Values in (b, c) correspond to mean and standard deviation of $n = 3$ independent samples. (d) Bimolecular fluorescence complementation (BiFC) experiments in onion (*Allium cepa*) epidermal tissue. Cells were co-bombarded with a construct to transiently express the full-length GGPPS11 protein with its C-terminus fused to the C-terminal domain of YFP (G11-CY) together with plasmids expressing full-length PSY, GGR or SPS2 proteins with their C-termini fused to the N-terminal half of YFP (PSY-NY, GGR-NY and SPS2-NY, respectively) or a negative control (C-NY). An extra plasmid encoding a cytosolic DsRed marker protein was also included to mark transformed cells. Panels show merged images obtained by observation of DsRed fluorescence (red), yellow fluorescent protein (YFP) fluorescence (green, indicative of positive interaction), and bright field. Green fluorescent dots correspond to plastids (leucoplasts). Bars, 20 μm.

immunoprecipitation results were not consistently reproducible, perhaps due to transient or unstable nature of the interactions or to their dependence on specific growth conditions or cell types.

As an alternative and more robust approach to evaluate the interaction between GGPPS11 and downstream enzymes, we used yeast two-hybrid (Y2H) analysis. Because PSY enzymes have been found associated to membranes in all plastid types (Dogbo *et al.*, 1988; Bonk *et al.*, 1997; Fraser *et al.*, 2000; Welsch *et al.*, 2000; Li *et al.*, 2008), we used a split-ubiquitin membrane-based Y2H system (Obrdlik *et al.*, 2004). A clear interaction between *Arabidopsis* GGPPS11 and PSY enzymes was detected by growth in selective medium (Fig. 5a) and β-galactosidase activity (Fig. 5b) of yeast strains co-expressing GGPPS11 fused to the N-terminal ubiquitin moiety (Nub-G11) and PSY fused to the C-terminal ubiquitin moiety (PSY-Cub). The fusion proteins maintained their enzymatic activity, as demonstrated by the accumulation of phytoene (the product of PSY) in yeast strains expressing both Nub-G11 and PSY-Cub (Fig. 5c), while this was not observed in yeasts co-expressing Nub and PSY-Cub. Y2H experiment further confirmed the interaction of GGPPS11 with GGR, another membrane-associated enzyme (Peltier *et al.*, 2004; Joyard *et al.*, 2009; Tanaka *et al.*, 2010). However, interaction with SPS2 could not be evaluated with this experimental system because the SPS2-Cub fusion alone showed a very strong activation of the reporter genes (Fig. 5).

To verify that the observed interaction of GGPPS11 with GGPP-consuming enzymes occurs under appropriate physiological conditions (i.e. inside plastids in living plant cells), we carried out bimolecular fluorescence complementation (BiFC) assays (Ohad *et al.*, 2007). Onion (*Allium cepa*) epidermal peels were microbombarded with particles coated with constructs expressing GGPPS11 fused to the C-terminal domain of the yellow fluorescent protein (YFP) (G11-CY) and the interacting enzymes fused to the N-terminal half of YFP (PSY-NY, GGR-NY, and SPS2-NY). As shown in Fig. 5(d), BiFC experiments confirmed the interaction of *Arabidopsis* GGPPS11 with PSY, GGR, and SPS2 in the plastids of plant cells.

Discussion

Using GCN (Fig. 2), mutant analyses (Figs 3, S2–S4) and metabolite profiling (Figs 4, S5) we demonstrated the central role of the GGPPS11 isozyme for the synthesis of photosynthesis-related isoprenoids and hence for chloroplast and plant development. Furthermore, immunoprecipitation, Y2H, and BiFC assays (Fig. 5) provided evidence that GGPPS11 interacts with enzymes that use GGPP as substrate.

The remaining individual plastid GGPPS enzymes appear to be dispensable for the synthesis of major plastid isoprenoids and for normal plant growth and development. So far we can

only speculate on their function. Multiple paralogues are generally maintained in the genome if they confer a selective advantage, such as a better performance under certain conditions, gene dosage or reducing metabolic cross talk by association with different metabolic fluxes (Force *et al.*, 1999). Expression of the *GGPPS2* and *GGPPS6* to *GGPPS10* paralogues is restricted to specific tissues and developmental stages mainly in flowers, seeds and in roots (Beck *et al.*, 2013) and correlates, in general, with the developmentally regulated synthesis of hormones (Bennett *et al.*, 2006; Nambara & Marion-Poll, 2006; Floss & Walter, 2009; Ruyter-Spira *et al.*, 2013; Seto & Yamaguchi, 2014). Expression of genes encoding GA biosynthetic enzymes do correlate with GGPPS-encoding genes in the GCN. Nevertheless, *ggpps* mutants do not show any apparent change from the normal phenotype that would imply GA, ABA or strigolactone hormone deficiencies. Since these plastid GGPPS isozymes seem to be redundant in their function, combinations of multiple mutants will have to be generated in order to reveal their function.

Our data provide evidence that GGPPS11 can interact with the three enzymes that channel GGPP to the production of major photosynthesis-related isoprenoids: PSY to carotenoids, GGR to chlorophylls, tocopherols, phylloquinones and SPS2 to plastoquinones (Fig. 1). The formation of multienzyme complexes containing GGPPS11 and particular GGPP-consuming enzymes could ensure an efficient allocation of substrates into particular pathways, a mechanism to control metabolic flux to specific isoprenoid biosynthetic pathways in addition to the co-regulation of gene expression. Our data are consistent with the existence of multienzyme complexes containing GGPPS and PSY in chromoplasts that are specialized in the production and accumulation of high amounts of carotenoids (Maudinas *et al.*, 1977; Dogbo *et al.*, 1988; Camara, 1993; Fraser *et al.*, 2000; Ruiz-Sola & Rodríguez-Concepción, 2012). It is therefore possible that complexes containing GGPPS11 and GGR might be found in tocopherol-rich seeds, whereas complexes containing PSY, GGR, or/and SPS2 might form, albeit transiently, in the chloroplasts of photosynthetic tissues. The formation of a particular complex might be determined by the abundance of the corresponding GGPP-consuming enzyme in a given plastid, cell, tissue, organ, or developmental stage. It is possible that competition among different enzymes for GGPPS11 binding is also part of a regulatory mechanism to balance the production of different types of isoprenoid products. These possibilities require further experimental analysis.

In summary, our data confirm the central role of GGPPS11 for the synthesis of plastid isoprenoids. The essential nature of this particular GGPPS isozyme was demonstrated by *in silico* analysis and confirmed genetically based on the severe developmental phenotype of the *ggpps11* null mutants and metabolically using the knock-down *ggpps11-5* allele. Furthermore, the formation of multienzyme complexes containing GGPPS11 and particular GGPP-consuming enzymes could be an additional mechanism besides the gene co-expression to control metabolic flux to specific isoprenoid biosynthetic pathways.

Acknowledgements

We are very grateful to G. Glauser (University of Neuchâtel, Switzerland) for the analysis of prenylquinones. We also thank V. Burlat and V. Courdavault for providing the BiFC vectors, M. Rothbart and B. Grimm for the GGR clone, P. Obrdlik for the Y2H plasmids, F. Wüst for his support with the Y2H setup and M. Nater for her phenotyping work. Technical support from M.R. Rodríguez-Goberna and members of the CRAG Services is greatly appreciated. We would finally like to thank M. Meret from the Max Planck Institute of Molecular Plant Physiology in Golm (Germany) for support in the metabolomics sample preparation procedure. This work was supported by grants from CYTED (Ibercarot-112RT0445), MINECO (BIO2011-23680), AGAUR (2014SGR-1434), ETH Zurich (TH-51 06-1), the EU FP7 contract 245143 (TiMet) and SK grant VEGA 1/0417/13. M.A.R-S. and M.V.B. were supported by MINECO FPI and AGAUR FI fellowships.

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al.* 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**: 653–657.

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodríguez-Concepción M, Gruissem W, Vranová E. 2013. Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. *Plant Molecular Biology* **82**: 393–416.

Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O. 2006. The Arabidopsis MAX pathway controls shoot branching by regulating auxin transport. *Current Biology* **16**: 553–563.

Bick JA, Lange BM. 2003. Metabolic cross talk between cytosolic and plastid pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of Biochemistry and Biophysics* **415**: 146–154.

Block A, Fristedt R, Rogers S, Kumar J, Barnes B, Barnes J, Elowsky CG, Wamboldt Y, Mackenzie SA, Redding K *et al.* 2013. Functional modeling identifies paralogous solanesyl-diphosphate synthases that assemble the side chain of plastoquinone-9 in plastids. *Journal of Biological Chemistry* **288**: 27594–27606.

Bonk M, Hoffmann B, Von Lintig J, Schledz M, Al-Babili S, Hobeika E, Kleinig H, Beyer P. 1997. Chloroplast import of four carotenoid biosynthetic enzymes *in vitro* reveals differential fates prior to membrane binding and oligomeric assembly. *European Journal of Biochemistry* **247**: 942–950.

Bouvier F, Rahier A, Camara B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. *Progress in Lipid Research* **44**: 357–429.

Camara B. 1993. Plant phytoene synthase complex: component enzymes, immunology, and biogenesis. *Methods in Enzymology* **214**: 352–365.

Coman D, Rütimann P, Gruissem W. 2014. A flexible protocol for targeted gene co-expression network analysis. *Methods in Molecular Biology* **1153**: 285–299.

Croteau R, Kutchan T, Lewis N. 2000. Natural products (secondary metabolites). In: Buchanan B, Gruissem W, Jones R, eds. *Biochemistry and molecular biology of plants*. Rockville, MD, USA: American Society of Plant Biologists, 1250–1268.

Curaba J, Moritz T, Blervaque R, Parcy F, Raz V, Herzog M, Vachon G. 2004. *AtGA3ox2*, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by *LEAFY COTYLEDON2* and *FUSCA3* in *Arabidopsis*. *Plant Physiology* **136**: 3660–3669.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology* **139**: 5–17.

Disch A, Hemmerlin A, Bach TJ, Rohmer M. 1998. Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. *Biochemistry Journal* **331**: 615–621.

Dogbo O, Laferriere A, D'Harlingue A, Camara B. 1988. Carotenoid biosynthesis: isolation and characterization of a bifunctional enzyme catalyzing the synthesis of phytoene. *Proceedings of the National Academy of Sciences, USA* **85**: 7054–7058.

Emanuelsson O, Nielsen H, von Heijne G. 1999. ChloroP, a neural network based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978–984.

Eriksson S, Bohlenius H, Moritz T, Nilsson O. 2006. GA4 is the active gibberellin in the regulation of LEAFY transcription and *Arabidopsis* floral initiation. *Plant Cell* **18**: 2172–2181.

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

Flügge UI, Gao W. 2005. Transport of isoprenoid intermediates across chloroplast envelope membranes. *Plant Biology* **7**: 91–97.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.

Fraser PD, Schuch W, Bramley PM. 2000. Phytoene synthase from tomato (*Lycopersicon esculentum*) chloroplasts – partial purification and biochemical properties. *Planta* **211**: 361–369.

Gachon CMM, Langlois-Meurinne M, Henry Y, Saindrenan P. 2005. Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Molecular Biology* **58**: 229–245.

Gershenzon J, Dudareva N. 2007. The function of terpene natural products in the natural world. *Nature Chemical Biology* **3**: 408–414.

Gilbert HN, van der Laan MJ, Dudoit S. 2009. *Joint multiple testing procedures for graphical model selection with applications to biological networks*. [WWW document] URL http://www.bepress.com/ucbbiostat/paper245/ [accessed 23 April 2015].

Helliwell C, Waterhouse P. 2003. Constructs and methods for high-throughput gene silencing in plants. *Methods* **30**: 289–295.

Heyndrickx KS, Vandepoele K. 2012. Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiology* **159**: 884–901.

Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin J, Rolland N. 2009. Chloroplast proteomics and the compartmentation of plastid isoprenoid biosynthetic pathways. *Molecular Plant* **2**: 1154–1180.

Keller Y, Bouvier F, d'Harlingue A, Camara B. 1998. Metabolic compartmentation of plastid prenyl lipid biosynthesis – evidence for the involvement of a multifunctional geranylgeranyl reductase. *European Journal of Biochemistry* **251**: 413–417.

Kirby J, Keasling JD. 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annual Review of Plant Biology* **60**: 335–355.

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

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

Lonzano AC, Świrszcz G. 2012. Multi-level Lasso for sparse multi-task regression. In: Langford J, Pineau J, eds. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. Edinburgh, UK: ACM, 361–368.

Lunn JE. 2007. Compartmentation in plant metabolism. *Journal of Experimental Botany* **58**: 35–47.

Martinis J, Kessler F, Glauser G. 2011. A novel method for prenylquinone profiling in plant tissues by ultra-high pressure liquid chromatography-mass spectrometry. *Plant Methods* **7**: 23.

Maudinas B, Bucholtz ML, Papastephanou C, Katiyar SS, Briedis AV, Porter JW. 1977. The partial purification and properties of a phytoene synthesizing enzyme system. *Archives of Biochemistry and Biophysics* **180**: 354–362.

Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET. 2011. A transcriptional analysis of carotenoid, chlorophyll and plastid isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Systems Biology* **5**: 77.

Nambara E, Marion-Poll A. 2006. Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology* **56**: 165–185.

Obrdlik P, El-Bakkoury M, Hamacher T, Cappellaro C, Vilarino C, Fleischer C, Ellerbrok H, Kamuzinzi R, Ledent V, Blaudez D *et al.* 2004. K^+ channel interactions detected by a genetic system optimized for systematic studies of membrane protein interactions. *Proceedings of the National Academy of Sciences, USA* **101**: 12242–12247.

Ohad N, Shichrur K, Yalovsky S. 2007. The analysis of protein–protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiology* **145**: 1090–1099.

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 Physiology* **122**: 1045–1056.

Oliver S. 2000. Guilt-by-association goes global. *Nature* **403**: 601–603.

Peltier JB, Ytterberg AJ, Sun Q, van Wijk KJ. 2004. New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast, and versatile fractionation strategy. *Journal of Biological Chemistry* **279**: 49367–49383.

Pulido P, Perello C, Rodríguez-Concepción M. 2012. New insights into plant isoprenoid metabolism. *Molecular Plant* **5**: 964–967.

Pulido P, Toledo-Ortiz G, Phillips MA, Wright LP, Rodríguez-Concepción M. 2013. Arabidopsis J-protein J20 delivers the first enzyme of the plastid isoprenoid pathway to protein quality control. *Plant Cell* **25**: 4183–4194.

Rodríguez-Concepción M, Forés O, Martínez-García JF, González V, Phillips MA, Ferrer A, Boronat A. 2004. Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during *Arabidopsis* seedling development. *Plant Cell* **16**: 144–156.

Rodriguez-Villalon A, Gas E, Rodríguez-Concepción M. 2009. Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown *Arabidopsis* seedlings. *Plant Journal* **60**: 424–435.

Ruiz-Sola MA, Rodríguez-Concepción M. 2012. Carotenoid biosynthesis in *Arabidopsis*: a colorful pathway. *The Arabidopsis Book/American Society of Plant Biologists* **10**: e0158.

Ruppel NJ, Kropp KN, Davis PA, Martin AE, Luesse DR. 2013. Mutations in *GERANYLGERANYL DIPHOSPHATE SYNTHASE 1* affect chloroplast development in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany* **100**: 2074–2084.

Ruyter-Spira C, Al-Babili S, van der Krol S, Bouwmeester H. 2013. The biology of strigolactones. *Trends in Plant Science* **18**: 72–83.

Samson F, Brunaud V, Balzergue S, Dubreucq B, Lepiniec L, Pelletier G, Caboche M, Lecharny A. 2002. FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Research* **30**: 94–97.

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. *Nature Genetics* **37**: 501–506.

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C *et al.* 2002. A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* **14**: 2985–2994.

Seto Y, Yamaguchi S. 2014. Strigolactone biosynthesis and perception. *Current Opinion in Plant Biology* **21**: 1–6.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498–2504.

Stangeland B, Salehian Z. 2002. An improved clearing method for GUS assay in Arabidopsis endosperm and seeds. *Plant Molecular Biology Reporter* **20**: 107–114.

Tanaka R, Rothbart M, Oka S, Takabayashi A, Takahashi K, Shibata M, Myouga F, Motohashi R, Shinozaki K, Grimm B *et al.* 2010. LIL3, a light-harvestinglike protein, plays an essential role in chlorophyll and tocopherol biosynthesis. *Proceedings of the National Academy of Sciences, USA* **107**: 16721–16725.

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD. 1999. Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. *Plant Cell* **11**: 1841–1852.

Vranová E, Hirsch-Hoffmann M, Gruissem W. 2011. AtIPD: a curated database of *Arabidopsis* isoprenoid pathway models and genes for isoprenoid network analysis. *Plant Physiology* **156**: 1655–1660.

Waadt R, Schmidt LK, Lohse M, Hashimoto K, Bock R, Kudla J. 2008. Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes *in planta*. *Plant Journal* **56**: 505–516.

Wang GD, Dixon RA. 2009. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *Proceedings of the National Academy of Sciences, USA* **106**: 9914–9919.

Wei HR, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, Somerville C, Loraine A. 2006. Transcriptional coordination of the metabolic network in *Arabidopsis*. *Plant Physiology* **142**: 762–774.

Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltran J, Chavarriaga P, Ceballos H, Tohme J, Beyer P. 2010. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. *Plant Cell* **22**: 3348–3356.

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

Wille A, Zimmermann P, Vranová E, Furholz A, Laule O, Bleuler S, Hennig L, Prelic A, von Rohr P, Thiele L *et al.* 2004. Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biology* **5**: R92.

Yang E, Lozano AC, Ravikumar P. 2014. Elementary estimators for high-dimensional linear regression. In: Jebara T, Xing EP, eds. *Proceedings of the 31st International Conference on Machine Learning (ICML-14), vol. 32.* Beijing, China: JMLR:W&CP, 388–396.

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 and Cell Physiology* **38**: 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 Molecular Biology* **35**: 331–341.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Mutants of the plastid GGPPS2, 6–10 isoforms.

Fig. S2 Complementation of GGPPS11-defective mutants with the *G11* minigene.

Fig. S3 Phenotype of seedlings and adult plants with altered GGPPS11 levels grown under different photoperiods.

Fig. S4 Chloroplast size in stomata and mesophyll cells.

Fig. S5 Levels of individual plastid isoprenoids.

Fig. S6 Level of transcripts encoding gibberellin pathway genes.

Table S1 List of *ggpps* mutant lines used in this work and their segregation ratios

Table S2 Primers used in this work

Table S3 Constructs and cloning details

Table S4 List of GCN input genes

Table S5 The GGPPS GCN representing significant co-expression relationships

Table S6 Gene abbreviations used in Fig. 2

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

SUPPORTING INFORMATION FIGURES S1-S6 AND TABLES S1-S6

SUPPORTING INFORMATION FIGURES

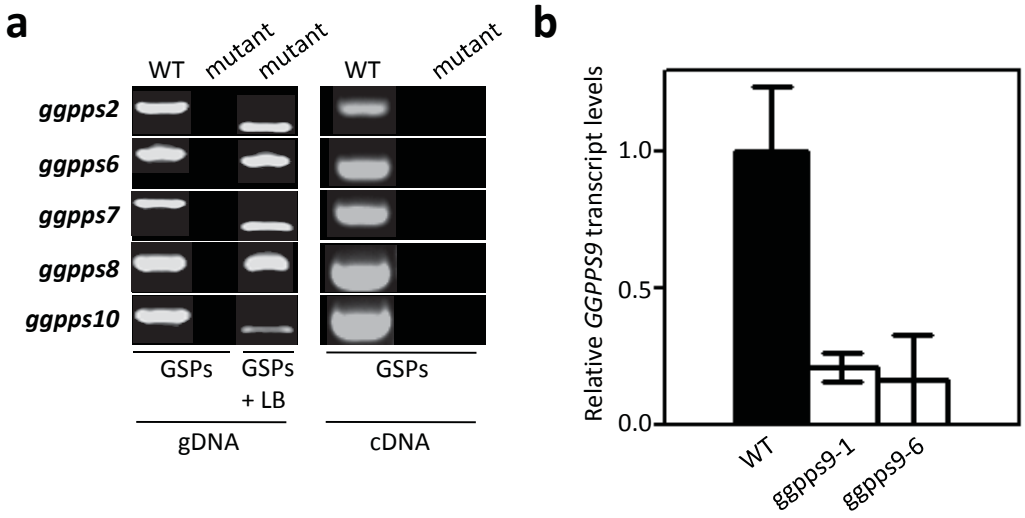


Figure S1. Mutants of the plastid GGPPS2,6-10 isoforms. (a) PCR genotyping of T-DNA insertion lines disrupting *GGPPS2*, *GGPPS6*, *GGPPS7*, *GGPPS8* and *GGPPS10* genes (left panel) and RT-PCR with corresponding cDNAs (right panel). For genotyping, the state of the transgene allele as heterozygous or homozygous was analysed using combinations of gene specific primers (GSPs) and a primer annealing to the left border of the T-DNA (LB). As control, the same GSPs were used with genomic DNA isolated from wild type Col-0 plants (WT). Homozygous mutant lines for *ggpps2*, *ggpps6*, *ggpps7*, *ggpps8* and *ggpps10* are shown. In the RT-PCR, full length cDNA spanning the T-DNA insertion site for each mutant could not be amplified, whereas in the control Col-0 (WT) the respective transcripts were present. (b) Transcript down-regulation in *ggpps9-1* and *ggpps9-6* stable RNAi lines. The actin (*ACT2; At3g18780*) gene was used for normalization. Data correspond to mean and standard deviation of n=5 independent samples. *GGPPS9* transcript levels are down-regulated to 21% in *ggpps9-1* and to 16% in *ggpps9-6* compared to the WT levels.

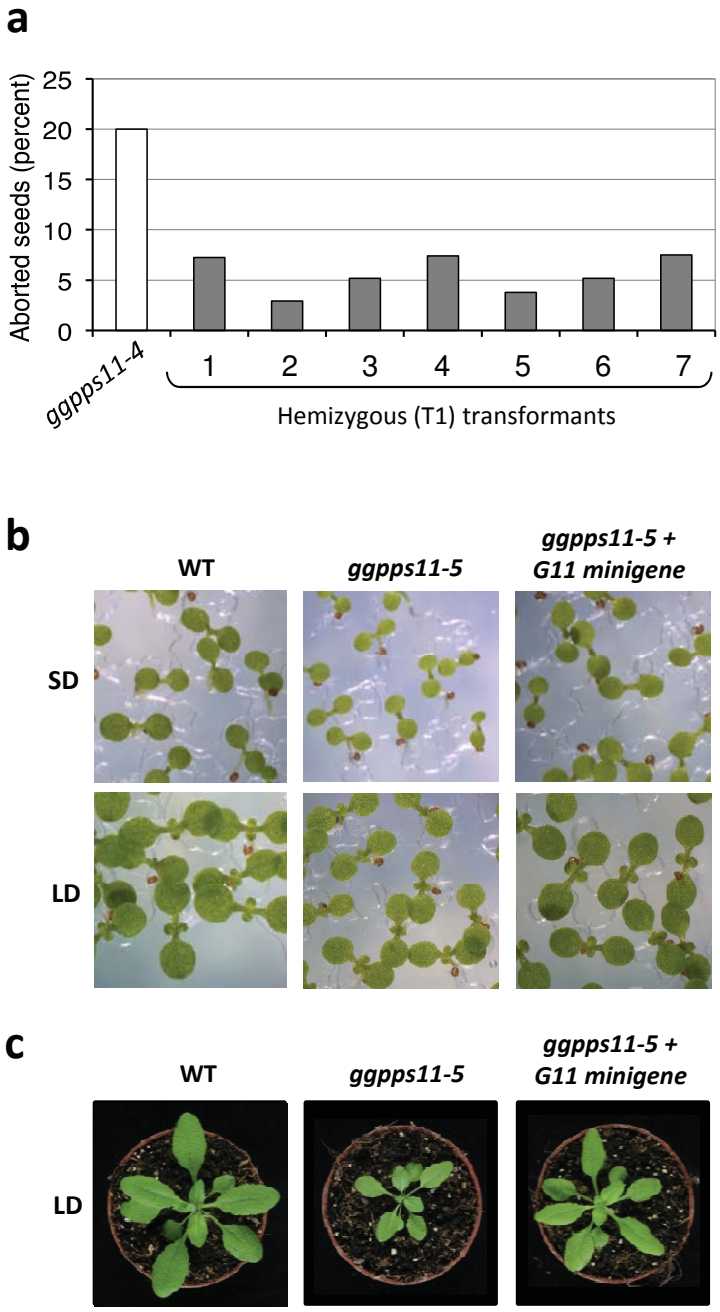


Figure S2. Complementation of GGPPS11-defective mutants with the *G11* minigene. The promoter and coding region of the *GGPPS11* gene (*G11* minigene) was amplified from the genome of Col-0 plants, cloned in a binary vector, and used to transform heterozygous *ggpps11-4* (a) and homozygous *ggpps11-5* (b,c) lines. (a) Rescue of the embryo lethal phenotype of the *ggpps11-4* allele with the G11 minigene. The graph shows the percentage of aborted seeds in siliques from heterozygous *ggpps11-4* lines either untransformed or transformed with the *G11* minigene (T1 population, in which a ~6.25 % of aborted seeds was expected assuming complementation). Between 420 and 856 seeds were counted from each plant. (b) Representative images of seedlings from wild type (WT), mutant *ggpps11-5*, and complemented *ggpps11-5* lines harbouring the *G11* minigene. Plants were grown for 7 days under short day (SD) or long day (LD) photoperiod. (c) Representative individuals of the indicated lines grown for 3 weeks under LD.

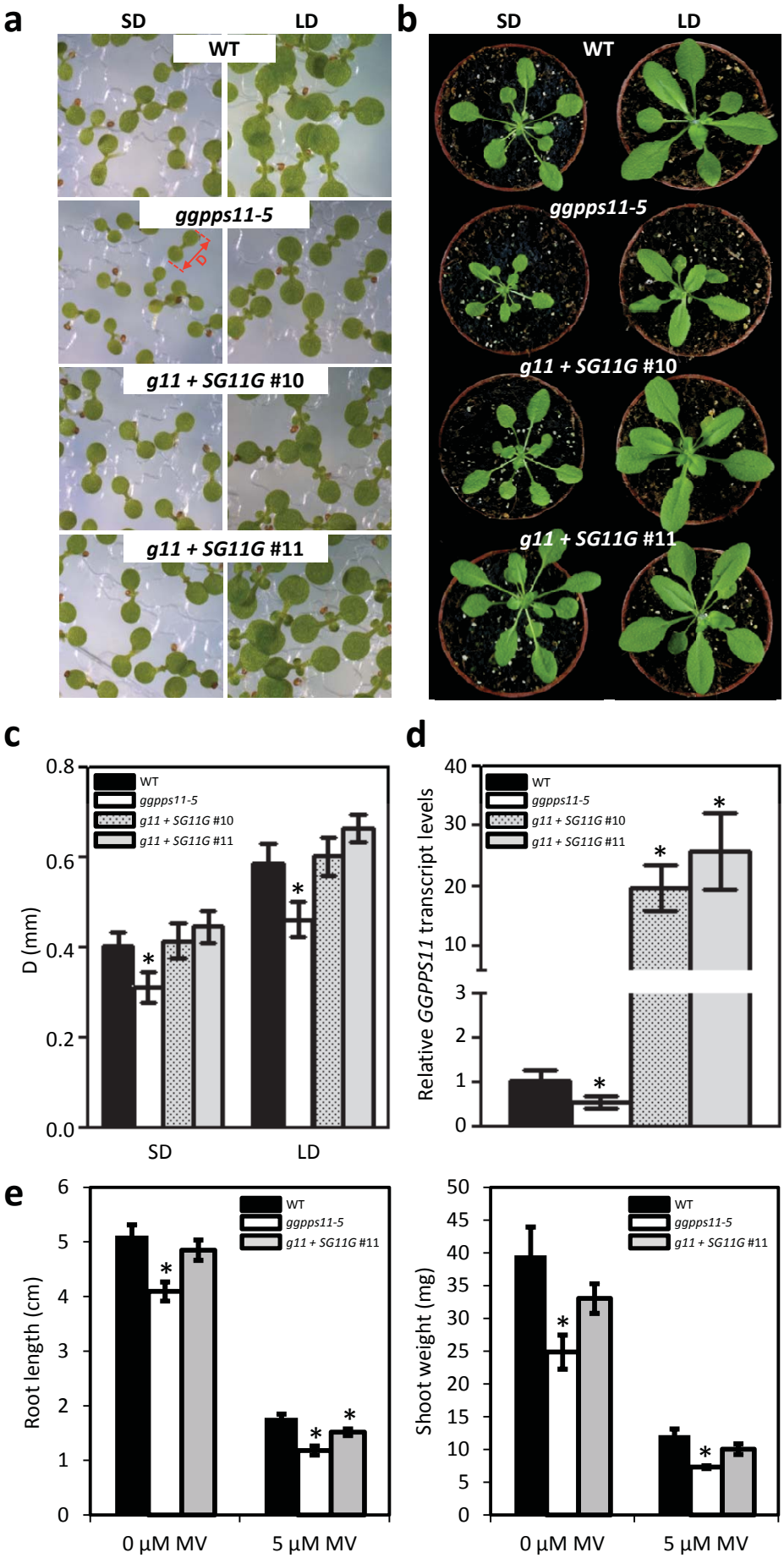

a
SD
LD
WT
ggpps11-5
D
g11 + SG11G #10
g11 + SG11G #11
b
SD
LD
WT
ggpps11-5
g11 + SG11G #10
g11 + SG11G #11
c
D (mm)
0.0
0.2
0.4
0.6
0.8
WT
ggpps11-5
g11 + SG11G #10
g11 + SG11G #11
SD
LD
d
Relative GGPPS11 transcript levels
0
1
2
3
10
20
30
40
WT
ggpps11-5
g11 + SG11G #10
g11 + SG11G #11
e
Root length (cm)
0
1
2
3
4
5
6
WT
ggpps11-5
g11 + SG11G #11
0 µM MV
5 µM MV
Shoot weight (mg)
0
5
10
15
20
25
30
35
40
45
50
WT
ggpps11-5
g11 + SG11G #11
0 µM MV
5 µM MV

Figure S3. Phenotype of seedlings and adult plants with altered GGPPS11 levels grown under different photoperiods. (a) Representative images of wild type (WT), mutant *ggpps11-5* and complemented seedlings of the indicated lines grown for 7 days under short day (SD) or long day (LD) photoperiod. (b) Representative individuals of the indicated lines grown for 4 weeks under SD or 3 weeks under LD. (c) Estimation of seedling size (D, distance between the tips of cotyledons; see red arrow on (a)) for the indicated lines grown for 7 days under SD or LD conditions. Values correspond to mean and standard deviation of at least n=25 seedlings. Asterisks mark statistically significant differences ($p<0.05$) relative to WT samples. (d) GGPPS11 transcript levels in WT, *ggpps11-5* and complemented seedlings of the indicated lines grown for 7 days under LD photoperiod. The *UBC/UBC21/PEX4* (At5g25760) gene was used for normalization. Data correspond to mean and standard deviation of n=4 independent samples. Asterisks mark statistically significant differences ($p<0.05$) relative to WT samples. (e) Root length and shoot weight of WT, *ggpps11-5* and complemented seedlings grown for 5 days under LD and then transferred to plates supplemented (+) or not (-) with 5 µM methyl viologen (MV). After 5 more days under LD, the length of the roots was monitored before removing them to quantify the fresh weight of the remaining shoot tissue. Data correspond to mean and standard deviation of at least n=45 individuals in 5 independent samples. Asterisks mark statistically significant differences ($p<0.05$) relative to WT samples.

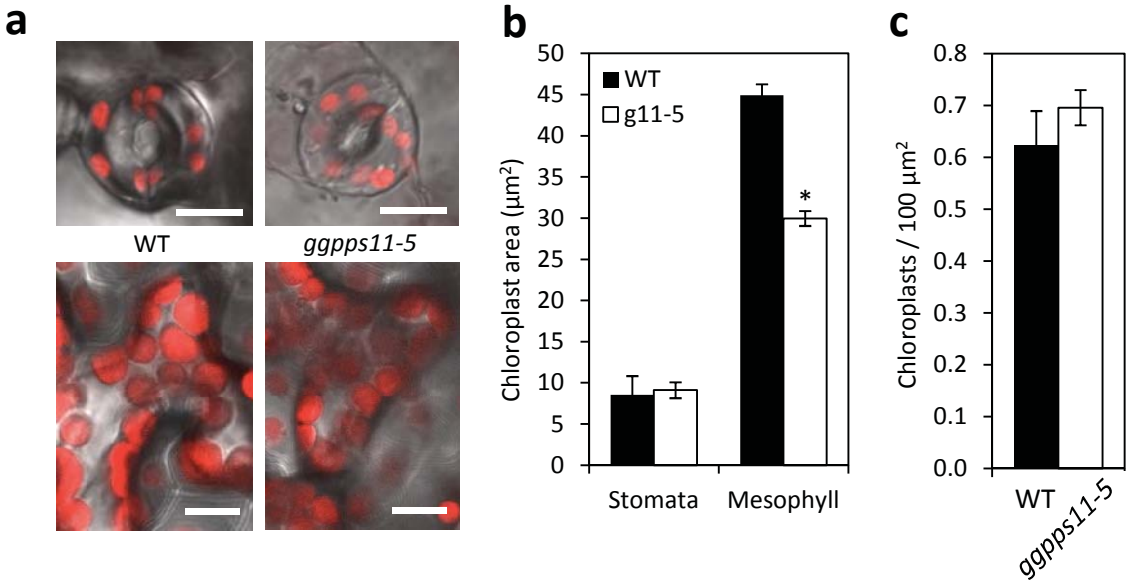


Figure S4. Chloroplast size in stomata and mesophyll cells. (a) Representative images of chloroplasts from WT and *ggpps11-5* stomata (upper panels) and mesophyll cells (lower panels). Bar, 10 µm. (b) Chloroplast area. Values correspond to mean and standard error of n=9 (guard cells) or n=45 (mesophyll cells) chloroplasts from 3 different sections of 3 cotyledons from independent plants. Asterisk marks statistically significant difference ($p<0.01$) relative to the WT. (c) Chloroplast concentration in the mesophyll. Values correspond to mean and standard error of n=6 pictures from 3 different sections of the cotyledons from 2 independent plants. No statistically significant differences between WT and mutant samples were found.

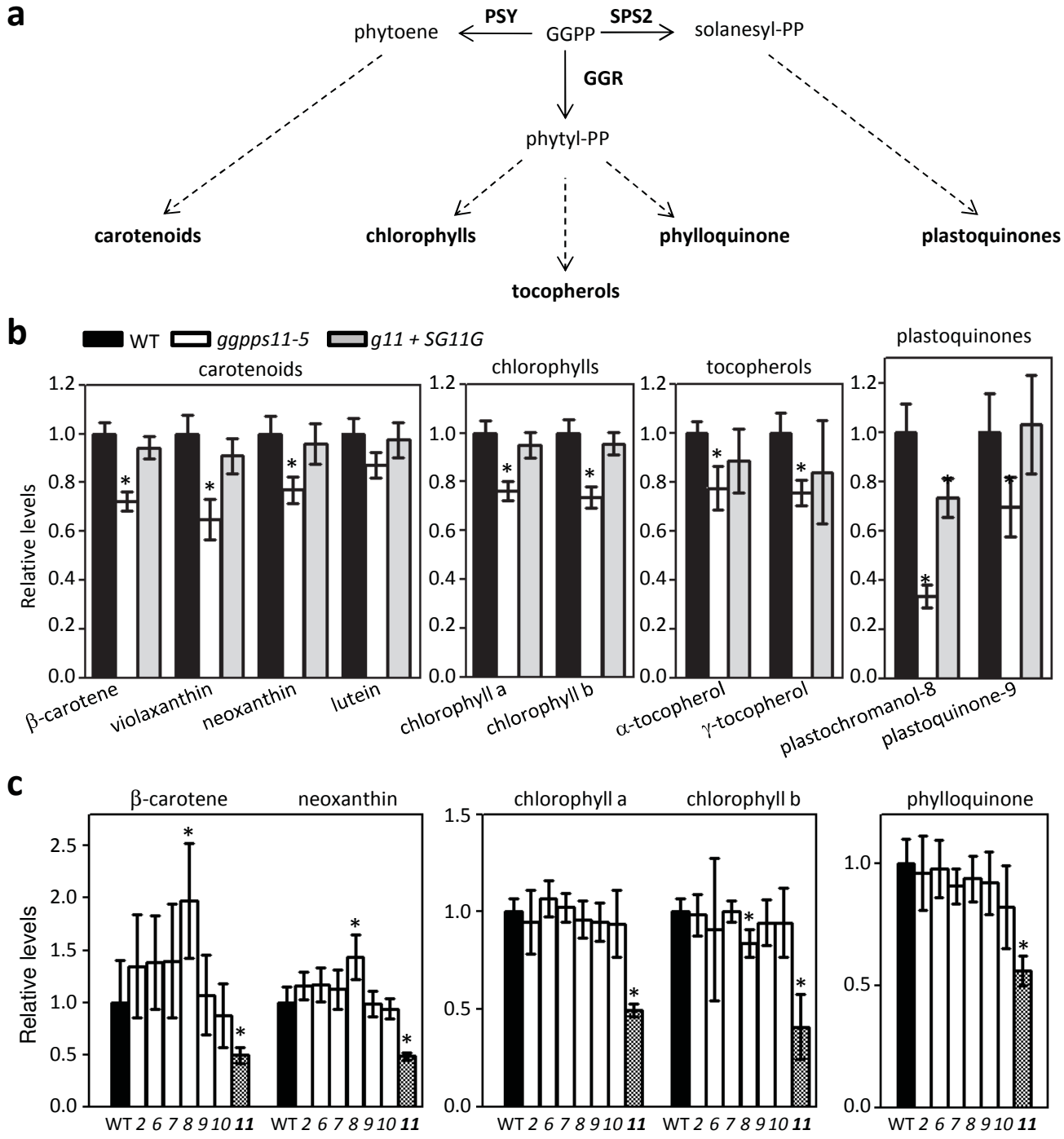


Figure S5. Levels of individual plastid isoprenoids. (a) Schematic representation of GGPP-consuming pathways. (b) Levels of individual plastid isoprenoids in wild type (WT), mutant *ggpps11-5*, and complemented *g11+SG11G* seedlings grown for 7 days under short day. Values are shown relative to those found in WT plants and correspond to mean and standard deviation of n=8 (carotenoids and chlorophylls) or n=4 (tocopherols and plastoquinones) independent samples. Asterisks mark statistically significant differences ($p<0.05$) relative to WT samples. (c) Levels of individual plastid isoprenoids in WT and mutant *ggpps2*, *ggpps6*, *ggpps7*, *ggpps8*, *ggpps9*, *ggpps10* and *ggpps11-5* seedlings grown for 10 days under long day photoperiod. Data are represented relative to those of WT plants and correspond to mean and standard deviation of n=7 independent samples. Asterisks mark statistically significant differences ($p<0.05$) relative to WT samples.

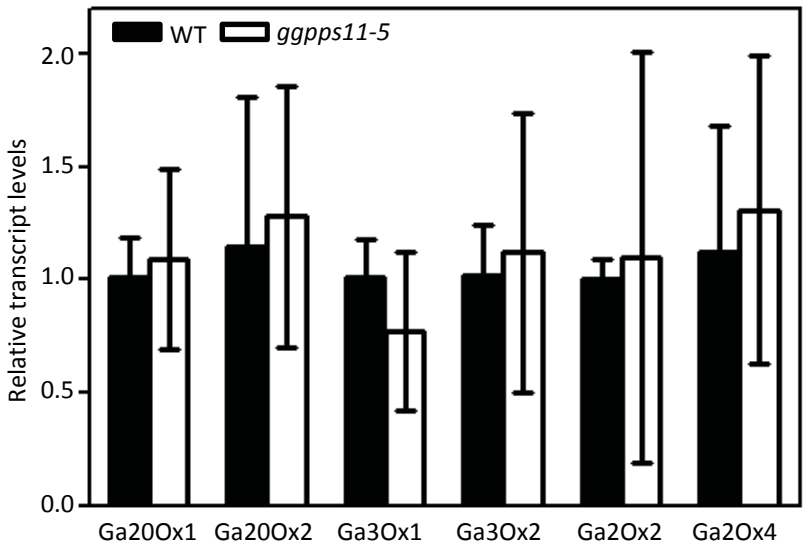


Figure S6. Level of transcripts encoding gibberellin pathway genes. Transcripts of different Arabidopsis genes encoding enzymes involved in the production of biologically active gibberellins (GA20ox and GA3ox) or in their inactivation (GA2ox) were quantified by RT-qPCR in samples from wild type (WT) and mutant *ggpps11-5* seedlings grown for 11 days under short day. The *UBC/UBC21/PEX4* (At5g25760) gene was used for normalization. Data correspond to mean and standard deviation of n=4 independent samples. No statistically significant differences ($p<0.05$) were found between WT and mutant seedlings.

SUPPORTING INFORMATION TABLES

Table S1. List of *ggpps* mutant lines used in this work and their segregation ratios.

Gene	AGI	Allele	Genetic Background	Original Genetic Background	Mutagen	Original Seed Source ID	Resistance	Segregation (resistant /sensitive)	χ^2	Mendelian monogenic recessive segregation ($\chi^2_{0.05}$ = 3.84)
GGPPS2	At2g18620	*ggpps2*	Col0	Ws	T-DNA	FLAG_134_B10	Kanamycin	3,18	0,35	yes
GGPPS6	At3g14530	*ggpps6*	Col0	Col3qrt1	T-DNA	SAIL_1148_A03	PPT(Basta)	2,57	0,24	yes
GGPPS7	At3g14550	*ggpps7*	Col0	Col0	T-DNA	SALK_119280	Kanamycin	2,75	0,04	yes
GGPPS8	At3g20160	*ggpps8*	Col0	Ws	T-DNA	FLAG_470_E09	Kanamycin	3,56	1,79	yes
GGPPS9	At3g29430	*ggpps9-1*	Col0	Col0	RNAi	in house	Kanamycin	3,01	0,9	yes
		ggpps9-6	Col0	Col0	RNAi	in house	Kanamycin	4,88	0,06	yes
GGPPS10	At3g32040	*ggpps10*	Col0	Col0	T-DNA	SM_3_32015	PPT(Basta)	4,26	0,16	yes
GGPPS11	At4g36810	*ggpps11-3*	Col0	Col0	T-DNA	SALK_085914	Kanamycin	3.24 (*1)	3,37E-11	yes
		ggpps11-4	Col3qrt1	Col3qrt1	T-DNA	SAIL_712_D06	PPT(Basta)	3.55 (*1)	2,78407E-14	yes
		ggpps11-5	Col0	Col0	T-DNA	SALK_140601	Kanamycin	1.8 (*2)	0,3	yes

(*1) Segregation analysis based on the number of green vs. brown seeds

(*2) Segregation analysis based on the PCR-based genotyping

Table S2. Primes used in this work.

Use	No.	Name	Sequence (5'-3')
Cloning	1	G11-5F	GATTTCAGAAATCGCCATGG
	2	G11-3R	ATTCCCGACAAAAGGAATCG
	3	G11-SpeI-3R	AACACACTAGTGTTCTGTCTATAGG
	4	G11 prom-5F	CCGCTCGAGGAAACTCACACCCACACAC
	5	G11 prom-3R	GACTAGTGGCGATTTCTGAAATCTG
	6	G11-SpeI-5F	GAAACTAGTATGGCTTCAGTGACTCTAG
	7	G11-SmaI-3R	CACCCCGGGGTTCTGTCTATAGG
	8	PSY-SpeI-5F	AACACTAGTATGTCTTCTTCTGTAGCAGTGTTATGG
	9	PSY-SmaI-3R	TCGCCCGGGTATCGATAGTCT
	10	GGR-SpeI-5F	CAAACTATGATGGCGACGACGGTTACACTCAAATC
	11	GGR-SmaI-3R	TGTCCCGGGAACACTAAGCTTCTCAATCTCTCTC
	12	SPS2-SpeI-5F	GTCACTATGATGATGATGTCATGTCG
	13	SPS2-SmaI-3R	TTGCCCGGGATCAATCCTTTCAAGATTAAACATTAC
	14	G11-B1	ACAAGTTTGTACAAAAAAGCAGGCTCTCCAACCACCATGTCTTCTTCCGTTGTTACAAAAG
	15	G11-B2	TCCGCCACCACCAACCACTTTGTACAAGAAAGCTGGGTAGTTCTGTCTATAGGCAATGTAATT
	16	PSY-B1	ACAAGTTTGTACAAAAAAGCAGGCTCTCCAACCACCATGTCTTTTGTAAGGAACCGAAGTAG
	17	PSY-B2	TCCGCCACCACCAACCACTTTGTACAAGAAAGCTGGGTATATCGATAGTCTTGAACTTGAAG
	18	GGR-B1	ACAAGTTTGTACAAAAAAGCAGGCTCTCCAACCACCATGGCAGCCAGAGCCACTCCCAAAC
	19	GGR-B2	TCCGCCACCACCAACCACTTTGTACAAGAAAGCTGGGTAAACACTAAGCTTCTCAATCTCTC
	20	SPS2-B1	ACAAGTTTGTACAAAAAAGCAGGCTCTCCAACCACCATGAGAGCTGTTCCGGCTAAATCC
	21	SPS2-B2	TCCGCCACCACCAACCACTTTGTACAAGAAAGCTGGGTAATCAATCCTTTCAAGATTAAAC
	22	G9-a	CCGCTCGAGGGTCCATTAGATTGATCC
	23	G9-b	CGCGAATTCCGTAATACTAAAATGCAAGC
	24	G9-c	GCTCTAGAGGTCCATTAGATTGATCC
	25	G9-d	CGCAAGCTTCGTAATACTAAAATGCAAGC
	26	M13-R	GTCATAGCTGTTTCCTG
Genotyping	27	G2-F	GCACTTAGCAACGTCAACGGCTG
	28	G2-R	CCAGCAAAAGCCATTGTCGGAG
	29	G6-F	CTAGGTGTATTGGACTACTGTTTC
	30	G6-R	GCACACTTTCAACATGAGCTGGC
	31	G7-F	AAGTTCTATTAGAGACGGTG
	32	G7-R	GACATGGCAGTAGCCTCGTC
	33	G8-F	CACCATGGAAAATCGAGAAGTTTTTGTT
	34	G8-R	TCAGTTTTGTCTGTTGACAATGTAATC
	35	G10-F	GATGATCCGTTCTGTAACTGAGC
	36	G10-R	ACGTTATAATGGACACAGAGAG
	37	Salk-LBb1	GCGTGGACCGCTTGCTGCAACT
	38	Sail-LB	TAGCATCTGAATTTCATAACCAATCTCGATACAC
	39	SM-LB	TACGATAACGGTCGGTACGG
	40	Flag-LB	TCCAGGGCGTGTGCCAGGTGC

Use	No.	Name	Sequence (5'-3')
RT-PCR	41	G2-F	ATGACTACACTCAATCTATCAATT
	42	G2-R	CTAATTCTGTCTTTTGGCAATGTAATT
	43	G6-F	CATGAGAGAGATGATCACACAGC
	44	G6-R	CCTCACCTCACAAGACGCAGGGC
	45	G7-F	AAGTTCTATTAGAGACGGTG
	46	G7-R	GAGATGATCACACGGCATAG
	47	G8-F	ATGGAAAATCGAGAAGTTTTTGTT
	48	G8-R	TCAGTTTTGTCTGTTGACAATGTAATC
	49	G10-F	TGTGCTCTCACCTCACAAGG
	50	G10-R	TCAGTTGTGTCTGCAAGCAATGTAGC
	51	G2-n-F	GCACTTAGCAACGTCAACGGCTG
	52	G2-n-R	CTAATTCTGTCTTTTGGCAATGTAATT
	53	G6-n-F	TCAGTTGTGTCTGCAAGCAATGTAGCT
	54	G6-n-R	AGGCAGCGGCAGTTTTAG
	55	G7-n-F	AAGTTCTATTAGAGACGGTG
	56	G7-n-R	CGCCTCCCACAAGCTCGCAGACG
	57	G10-n-F	TGTGCTCTCACCTCACAAGG
	58	G10-n-R	TCAGTTGTGTCTGCAAGCAATGTAGC
RT-qPCR	59	G9-F	TTAGGAGCCATAATGGGAGGT
	60	G9-R	TCGCATATTTTCTTAGCTTTTCG
	61	G11-F	CCTCTTTCGATTTCATGTCGT
	62	G11-R	ATTAGTCAACAAAGCTTTAGATTCAGC
	63	GA20ox1-F	CTTCCATCAACGTTCTCGAGC
	64	GA20ox1-R	GGTTTTGAAGGTCGATGAGAGG
	65	GA20ox2-F	AGAAACCTTCCATTGACATTCCA
	66	GA20ox2-R	AGAGATCGATGAACGGGACG
	67	GA3ox1-F	CATCCCATTCACCTCCCACACTCTCACATAC
	68	GA3ox1-R	AGGAGAAGGAGCAGCGGAGAAGAGGAG
	69	GA3ox2-F	GACTTGCTCCACATTTTAACCAACGGAATCTTC
	70	GA3ox2-R	CCACAGGTAAGCCATTGAGAACCGAGATC
	71	GA2ox2-F	GCAGGAGGCTATTGGCTTCTTCG
	72	GA2ox2-R	CTGAGGATTAGCATTGAGGAGGAGATAC
	73	GA2ox4-F	AGTGTAAGGCATAGAGCATTGAC
	74	GA2ox4-R	TACAACCGTGGCTGATTCATC
	75	UBC-F	CTGCTTGGACGCTTCAGTCTG
	76	UBC-R	CGCAGTTAAGAGGACTGTCCG

Table S3. Constructs and cloning details.

Use	Construct	Template	Primers (see Table S5)	Insert or PCR product	Plasmid backbone	Cloning method
Initial cloning	pCR-G11-p+cds	Genomic DNA	4+2	$GGPPS11_{(-1345)-1160}$ promoter+CDS	pCR2.1 TOPO	TOPO TA cloning
	pCR-G11-p	Genomic DNA	4+5	$GGPPS11_{(-1345)-(-1)}$ promoter	pCR2.1 TOPO	TOPO TA cloning
	pCR-G11-cds+	Seedling cDNA	1+3	$GGPPS11_{1-1160}$ cDNA/no stop	pCR2.1 TOPO	TOPO TA cloning
	pCR-PSY-cds+	Seedling cDNA	8+9	PSY_{1-1311} cDNA/no stop	pCR2.1 TOPO	TOPO TA cloning
	pCR-GGR-cds+	Seedling cDNA	10+11	GGR_{1-1401} cDNA/no stop	pCR2.1 TOPO	TOPO TA cloning
	pCR-SPS2-cds+	Seedling cDNA	12+13	$SPP2_{1-1251}$ cDNA/no stop	pCR2.1 TOPO	TOPO TA cloning
Transgenic plants	pG9-GST	Seedling cDNA	22+23+24+25	$GGPPS9_{1140-1235}$ cDNA	pHellsgate 8	*Xba*I / *Xho*I / *Eco*RI / *Hind*III
	G11 minigene	pTOPO-G11-p+cds	26+2	$GGPPS11_{(-949)-1160}$ promoter+CDS	pCAMBIA1302	*Bam*HI / *Pml* I
	pSG11G	pTOPO-G11-cds+	-	$GGPPS11_{1-1160}$ cDNA/no stop	pCAMBIA1302	*Nco*I / *Spe*I
Y2H-SUS	pNub-G11	pCR-G11-cds	14+15	$GGPPS11_{213-1160}$ cDNA/no stop	pNXgate	*in vivo* recombination
	pPSY-Cub	pCR-PSY-cds+	16+17	$PSY_{210-1311}$ cDNA/no stop	pMetYCgate	*in vivo* recombination
	pGGR-Cub	pCR-GGR-cds+	18+19	$GGR_{129-1401}$ cDNA/no stop	pMetYCgate	*in vivo* recombination
	pSPS2-Cub	pCR-SPS2-cds+	20+21	$SPS2_{177-1251}$ cDNA/no stop	pMetYCgate	*in vivo* recombination
BiFC	pG11-CY	pTOPO-G11-cds	6+7	$GGPPS11_{1-1160}$ CDS/no stop	pSYCE	*Spe*I / *Sma*I
	pG11-NY	pTOPO-G11-cds	6+7	$GGPPS11_{1-1160}$ CDS/no stop	pSYNE	*Spe*I / *Sma*I
	pPSY-NY	pTOPO-PSY-cds	-	PSY_{1-1311} cDNA/no stop	pSYNE	*Spe*I / *Sma*I
	pPSY-CY	pTOPO-PSY-cds	-	PSY_{1-1311} cDNA/no stop	pSYCE	*Spe*I / *Sma*I
	pGGR-NY	pTOPO-GGR-cds	-	GGR_{1-1401} cDNA/no stop	pSYNE	*Spe*I / *Sma*I
	pGGR-CY	pTOPO-GGR-cds	-	GGR_{1-1401} cDNA/no stop	pSYCE	*Spe*I / *Sma*I
	pSPS2-NY	pTOPO-SPS2-cds	-	$SPS2_{1-1251}$ cDNA/no stop	pSYNE	*Spe*I / *Sma*I
	pSPS2-CY	pTOPO-SPS2-cds	-	$SPS2_{1-1251}$ cDNA/no stop	pSYCE	*Spe*I / *Sma*I

Table S4. List of GCN input genes. Genes are grouped by pathways according to AtIPD (*). Genes which are part of several pathways are included in each pathway group (e.g. At1g74470). AGI, Arabidopsis Genome Initiative gene code.

Pathway	Affymetrix™ Probeset ID	AGI
MEP PATHWAY		
	245281_at	At4g15560
	247401_at	At5g62790
	267220_at	At2g02500
	266863_at	At2g26930
	260324_at	At1g63970
	247637_at	At5g60600
	253235_at	At4g34350
MVA PATHWAY		
	248779_at	At5g47720
	248690_at	At5g48230
	254845_at	At4g11820
	259983_at	At1g76490
	264856_at	At2g17370
	246778_at	At5g27450
	246286_at	At1g31910
	266414_at	At2g38700
	251881_at	At3g54250
GGPP BIOSYNTHESIS		
	265924_at	At2g18620
	258121_s_at	At3g14530
	258121_s_at	At3g14550
	257117_at	At3g20160
	256738_at	At3g29430
	256684_at	At3g32040
	246198_at	At4g36810
GIBBERELLINS BIOSYNTHESIS		
	255461_at	At4g02780
	262891_at	At1g79460
	246864_at	At5g25900
	264586_at	At1g05160
	266335_at	At2g32440
	259453_at	At1g44090
	259714_at	At1g60980
	254065_at	At4g25420
	250611_at	At5g07200
	248371_at	At5g51810
	261768_at	At1g15550
	260299_at	At1g80330
	260300_at	At1g80340
	254397_at	At4g21690
PLASTOQUINONE BIOSYNTHESIS		
	263122_at	At1g78510
	262526_at	At1g17050
	248207_at	At5g53970
	262635_at	At1g06570
	258755_at	At3g11945
	251118_at	At3g63410
TOCOPHEROL BIOSYNTHESIS		
	248207_at	At5g53970
	262635_at	At1g06570
	266938_at	At2g18950
	251118_at	At3g63410
	253394_at	At4g32770
	262875_at	At1g64970
	260236_at	At1g74470
PHYLLOQUINONE BIOSYNTHESIS		
	261428_at	At1g18870
	262177_at	At1g74710
	259643_at	At1g68890
	252293_at	At3g48990
	264920_at	At1g60550
	245484_at	At4g16210
	264963_at	At1g60600
	263044_at	At1g23360
	260236_at	At1g74470

Pathway	Affymetrix™ Probeset ID	AGI
CHLOROPHYLL BIOSYNTHESIS		
	264660_at	At1g09940
	256020_at	At1g58290
	264085_at	At2g31250
	252318_at	At3g48730
	247392_at	At5g63570
	245245_at	At1g44318
	260370_at	At1g69740
	246033_at	At5g08280
	245042_at	At2g26540
	255826_at	At2g40490
	257219_at	At3g14930
	264820_at	At1g03475
	255402_at	At4g03205
	255537_at	At4g01690
	250221_at	At5g14220
	250243_at	At5g13630
	254623_at	At4g18480
	248920_at	At5g45930
	261695_at	At1g08520
	254105_at	At4g25080
	251664_at	At3g56940
	250006_at	At5g18660
	264839_at	At1g03630
	253871_at	At4g27440
	248197_at	At5g54190
	246308_at	At3g51820
	245242_at	At1g44446
	260236_at	At1g74470
CAROTENOID BIOSYNTHESIS		
	250095_at	At5g17230
	245284_at	At4g14210
	259092_at	At3g04870
	260821_at	At1g06820
	246411_at	At1g57770
	247936_at	At5g57030
	259140_at	At3g10230
	246268_at	At1g31800
	251969_at	At3g53130
	254020_at	At4g25700
	248311_at	At5g52570
	247025_at	At5g67030
	264799_at	At1g08550
	263873_at	At2g21860
	255857_at	At1g67080
STRIGOLACTONE BIOSYNTHESIS		
	266129_at	At2g44990
	253398_at	At4g32810
	267380_at	At2g26170
ABA BIOSYNTHESIS		
	259669_at	At1g52340
	263570_at	At2g27150
	256190_at	At1g30100
	255857_at	At1g67080
	260797_at	At1g78390
	257280_at	At3g14440
	257242_at	At3g24220
	254668_at	At4g18350
	247025_at	At5g67030

(*) **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:** 1655-1660.

Table S5. The GGPPS GCN representing significant co-expression relationships. The GGPPS paralogs ("Guide Gene") significantly co-expressed with genes encoding enzymes in the MVA pathway, MEP pathway and downstream pathways using GGPP as precursor ("Query Gene").

Guide Gene	Query Gene
GGPPS11	At1g78510
	At1g74470
	At4g27440
	At5g54190
	At3g51820
	At1g44446
	At2g32440
	At1g68890
	At1g60550
	At1g60600
	At1g23360
	At4g15560
	At2g26930
	At4g34350
	At1g64970
	At3g11945
	At3g63410
	At2g18950
	At5g62790
	At2g02500
	At1g63970
	At5g60600
	At1g52340
	At1g17050
	At5g17230
	At4g14210
	At3g04870
	At1g06820
	At1g57770
	At5g57030
	At3g10230
	At1g31800
	At3g53130
	At4g25700
	At5g67030
	At1g08550
	At2g21860
	At1g67080
	At1g58290
	At3g48730
	At5g63570
	At1g69740
	At5g08280
	At2g26540
	At2g40490
	At3g14930
	At1g03475
	At4g01690
	At5g14220
	At5g13630
	At4g18480
	At5g45930
	At1g08520
	At4g25080
	At3g56940
	At5g18660
	At1g03630

Guide Gene	Query Gene
GGPPS2	At1g44318
	At1g15550
	At1g80340
	At1g18870
	At5g27450
	At2g38700
GGPPS6/7	At1g15550
	At1g80340
	At1g18870
	At5g27450
	At2g38700
	At5g48230
	At4g11820
	At1g31910
	At3g54250
GGPPS8	At1g44318
	At1g80340
	At1g18870
	At1g15550
GGPPS9	At1g44318
GGPPS10	At5g27450
	At2g38700
	At1g44318
	At1g15550
	At1g80340
	At1g18870

Table S6. Gene abbreviations used in Figure 2.

Gene	AGI	Description
AACT2	At5g48230	acetoacetyl-CoA thiolase 2
HMGS	At4g11820	3-hydroxy-3-methylglutaryl-CoA synthase
MK	At5g27450	mevalonate kinase
PMK	At1g31910	phosphomevalonate kinase
MPDC1	At2g38700	mevalonate diphosphate decarboxylase 1
MPDC2	At3g54250	mevalonate diphosphate decarboxylase 2
DXS	At4g15560	1-deoxy-D-xylulose-5-phosphate synthase
DXR	At5g62790	DXP reductoisomerase
MCT	At2g02500	2-*C*-methyl-D-erythritol 4-phosphate cytidyltransferase
CMK	At2g26930	4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol kinase
MDS	At1g63970	2-*C*-methyl-D-erythritol-2,4-cyclodiphosphate synthase
HDS	At5g60600	4-hydroxy-3-methylbut-2-enyl diphosphate synthase
HDR	At4g34350	HMBPP reductase
KAO2	At2g32440	*ent*-kaurenoic acid oxidase 2
GA3ox1	At1g15550	gibberellin 3-oxidase 1
GA3ox2	At1g80330	gibberellin 3-oxidase 2;
GGR	At1g74470	geranylgeranyl reductase
SPS1	At1g78510	solanesyl-diphosphate synthase 1
SPS2	At1g17050	solanesyl-diphosphate synthase 2
HST	At3g11945	homogentisate solanesyltransferase
VTE2	At2g18950	homogentisate phytyltransferase
VTE3	At3g63410	MSBQ/MPBQ methyltransferase
VTE4	At1g64970	delta/gamma-tocopherol methyltransferase
MENF2	At1g18870	isochorismate synthase 2
MEND	At1g68890	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (PHYLLO)
MENB	At1g60550	1,4-dihydroxy-2-naphtoyl-CoA synthase
MENA	At1g60600	1,4-dihydroxy-2-naphthoate phytyltransferase
MENG	At1g23360	demethylphylloquinone methyltransferase
HEMA1	At1g58290	glutamyl-tRNA reductase 1
GSA1	At5g63570	glutamate-1-semialdehyde 2,1-aminomutase
GSA2	At3g48730	glutamate-1-semialdehyde 2,1-aminomutase
HEMB1	At1g69740	5-aminolevulinate dehydratase
HEMB2	At1g44318	5-aminolevulinate dehydratase
HEMC	At5g08280	porphobilinogen deaminase
HEMD	At2g26540	uroporphyrinogen III-synthase
HEME1	At3g14930	uroporphyrinogen III decarboxylase
HEME2	At2g40490	uroporphyrinogen III decarboxylase
HEMF	At1g03475	coproporphyrinogen III oxidase
HEMG1	At4g01690	protoporphyrinogen IX oxidase
HEMG2	At5g14220	protoporphyrinogen IX oxidase
CHLH	At5g13630	magnesium chelatase H subunit
CHL11	At4g18480	magnasium chelatase I subunit
CHL12	At5g45930	magnasium chelatase I subunit
CHLD	At1g08520	magnesium chelatase D subunit
CHLM	At4g25080	Mg-protoporphyrin IX methyltransferase
CRD	At3g56940	Mg-protoporphyrin IX monomethylester cyclase
PORA	At5g54190	protochlorophyllide reductase A
PORB	At4g27440	protochlorophyllide reductase B
PORC	At1g03630	protochlorophyllide reductase C
DVR	At5g18660	divinyl reductase
CAO	At1g44446	chlorophyll a oxygenase
CHLG	At3g51820	chlorophyll synthetase
ABA2	At1g52340	short-chain alcohol dehydrogenase activity
PSY	At5g17230	phytoene synthase
PDS	At4g14210	phytoene desaturase
ZDS	At3g04870	zeta-carotene desaturase
CRTISO1	At1g06820	carotenoid isomerase 1
CRTISO2	At1g57770	carotenoid isomerase 2
LYCb	AT3g10230	lycopene beta-cyclase
LYCe	At5g57030	lycopene epsilon-cyclase
ZE	At5g67030	zeaxanthin epoxidase
VDE	At1g08550	violaxanthin de-epoxidase
VDEr	At2g21860	violaxanthin de-epoxidase related gene
NS	At1g67080	neoxanthin synthase
LUT5	At1g31800	carotene hydroxylase CYP97A3
BCH1	At4g25700	carotene hydroxylase
LUT1	At3g53130	carotene hydroxylase CYP97C1

ANNEX 2

PUBLICATION

Ruiz-Sola MÁ*, Barja MV*, Manzano D, Llorente B, Schipper B, Beekwilder J and Rodríguez-Concepción M (2016b). A single Arabidopsis gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. Plant Physiol. **172**: 1393–1402.

*These authors contributed equally to the work

NOTE: Only the work specifically performed by the PhD candidate in this article is included in the first Chapter of the thesis.

Research Report

A Single Arabidopsis Gene Encodes Two Differentially Targeted Geranylgeranyl Diphosphate Synthase Isoforms[1][OPEN]

M. Águila Ruiz-Sola[2], M. Victoria Barja[2], David Manzano, Briardo Llorente, Bert Schipper, Jules Beekwilder, and Manuel Rodriguez-Concepcion*

Program of Plant Metabolism and Metabolic Engineering, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB Bellaterra, 08193 Barcelona, Spain (M.A.R.-S., M.V.B., D.M., B.L., M.R.-C.); and Plant Research International, 6700 AA Wageningen, The Netherlands (B.S., J.B.)

ORCID IDs: 0000-0002-2281-6700 (M.A.R.-S.); 0000-0002-3846-4885 (M.V.B.); 0000-0002-5483-7223 (D.M.); 0000-0002-3727-1395 (B.L.); 0000-0002-1280-2305 (M.R.-C.).

A wide diversity of isoprenoids is produced in different plant compartments. Most groups of isoprenoids synthesized in plastids, and some produced elsewhere in the plant cell derive from geranylgeranyl diphosphate (GGPP) synthesized by GGPP synthase (GGPPS) enzymes. In Arabidopsis (*Arabidopsis thaliana*), five genes appear to encode GGPPS isoforms localized in plastids (two), the endoplasmic reticulum (two), and mitochondria (one). However, the loss of function of the plastid-targeted GGPPS11 isoform (referred to as G11) is sufficient to cause lethality. Here, we show that the absence of a strong transcription initiation site in the *G11* gene results in the production of transcripts of different lengths. The longer transcripts encode an isoform with a functional plastid import sequence that produces GGPP for the major groups of photosynthesis-related plastidial isoprenoids. However, shorter transcripts are also produced that lack the first translation initiation codon and rely on a second in-frame ATG codon to produce an enzymatically active isoform lacking this N-terminal domain. This short enzyme localizes in the cytosol and is essential for embryo development. Our results confirm that the production of differentially targeted enzyme isoforms from the same gene is a central mechanism to control the biosynthesis of isoprenoid precursors in different plant cell compartments.

Plants produce tens of thousands of isoprenoid compounds, including some that are essential for respiration, photosynthesis, and regulation of growth and development. Despite their structural and functional diversity, all isoprenoids derive from the same five-carbon precursors, the double-bond isomers isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which can be interconverted by IPP/DMAPP isomerase (IDI) enzymes. Plants use two unrelated pathways to synthesize these units (Fig. 1). The mevalonic acid (MVA) pathway synthesizes IPP in the cytosol, whereas the methylerythritol 4-phosphate (MEP) pathway supplies both IPP and DMAPP in the plastid (Bouvier et al., 2005; Vranová et al., 2013; Rodriguez-Concepción and Boronat, 2015). IPP and DMAPP units can be exchanged between cell compartments to a certain level. For example, MVA-derived IPP can be imported by mitochondria for the biosynthesis of ubiquinone (Lütke-Brinkhaus et al., 1984; Disch et al., 1998). However, this limited exchange of common isoprenoid precursors is not active enough to rescue a genetic or pharmacological blockage of one of the pathways with IPP/DMAPP produced by the noninhibited pathway (Bouvier et al., 2005; Vranová et al., 2013; Rodriguez-Concepción and Boronat, 2015). Addition of IPP units to DMAPP generates longer prenyl diphosphate molecules, including C10 geranyl diphosphate (GPP), C15 farnesyl diphosphate (FPP), and C20 geranylgeranyl diphosphate (GGPP), which are then used in specific downstream pathways to produce particular isoprenoids (Fig. 1). FPP and GGPP pools represent nodes of the major metabolic branch points in the isoprenoid biosynthesis network (Vranová et al., 2011; Vranová et al., 2013). As prenyl

[1] This work was funded by grants from the Ministerio de Economía y Competitividad (BES-2009-025359, BIO2011-23680, BIO2014-59092-P, PCIN-2015-103, BIO2015-71703-REDT, and FPDI-2013-018882), Ministerio de Educación, Cultura y Deporte (FPU14/05142), European Commission H2020 ERA-IB-2 program (BioProMo), and Generalitat de Catalunya (2014SGR-1434 and 2015FI_B 00007).

[2] These authors contributed equally to the article.

* Address correspondence to manuel.rodriguez@cragenomica.es.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Manuel Rodriguez-Concepcion (manuel.rodriguez@cragenomica.es).

M.A.R.-S. and M.R.-C. conceived the project and the original research plan; M.A.R.-S., M.V.B., and M.R.-C. designed the experiments; M.A.R.-S. and M.V.B. performed most of the experiments; D.M., M.V.B., B.S., and J.B. performed metabolite profiling experiments; B.L. provided technical assistance and discussion on experimental design to M.A.R.-S. and M.V.B.; M.A.R.-S., M.V.B., D.M., B.L., B.S., J.B., and M.R.-C. analyzed the data; M.R.-C. wrote the article with contributions of all the authors.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.01392

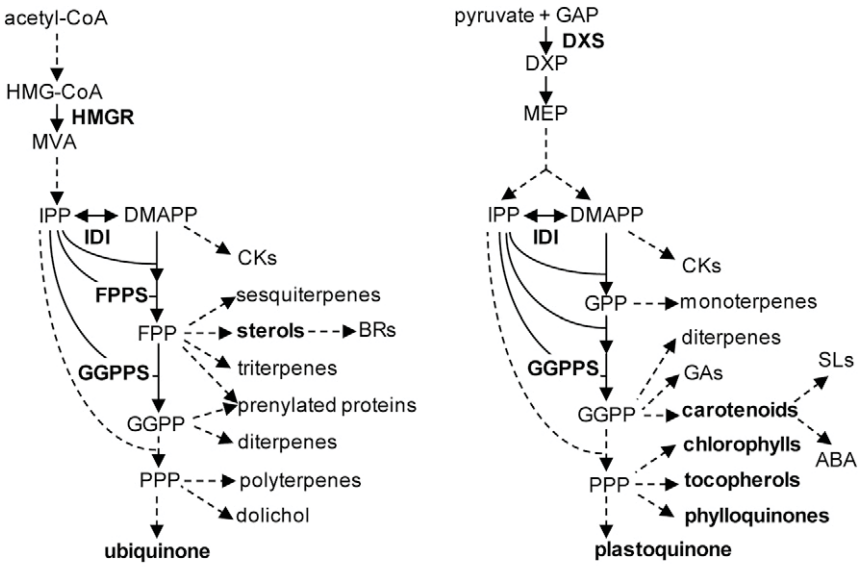


Figure 1. Isoprenoid biosynthetic pathways and enzymes in Arabidopsis. Solid arrows indicate single enzymatic steps, and dashed arrows represent multiple steps. Mevalonic acid (MVA) pathway: HMG-CoA, hydroxymethylglutaryl-CoA. Methylerythritol 4-phosphate (MEP) pathway: GAP, glyceraldehyde 3-phosphate; DXP, deoxyxylulose 5-phosphate. Prenyl diphosphates: IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; PPP, polyprenyl diphosphate. Hormones: CKs, cytokinins; BRs, brassinosteroids; GAs, gibberellins; SLs, strigolactones; ABA, abscisic acid. Enzymes: HMGR, HMG-CoA reductase; DXS, DXP synthase; IDI, IPP/DMAPP isomerase; FPPS, FPP synthase; GGPPS, GGPP synthase. *GGPPS activity unclear. **Isoforms reported in this work.

Enzyme	Gene	Accession	Isoform	Localization
HMGR	*HMG1*	At1g76490	HMGR1S	ER-cytosol
			HMGR1L	ER-cytosol
	HMG2	At2g17370	HMGR2	ER-cytosol
DXS	*DXS*	At4g15560	DXS	Plastids
IDI	*IDI1*	At5g16440	IDI1S	Peroxisomes
			IDI1L	Plastids
	IDI2	At3g02780	IDI2S	Peroxisomes
			IDI2L	Mitochondria
FPPS	*FPS1*	At5g47770	FPPS1S	Cytosol
			FPPS1L	Mitochondria
	FPS2	At4g17190	FPPS2L	Cytosol
GGPPS	*GGPPS1*	At1g49530	GGPPS1*	Mitochondria
	GGPPS2	At2g18620	GGPPS2	Plastids
	GGPPS3	At2g18640	GGPPS3	ER
	GGPPS4	At2g23801	GGPPS4	ER
	GGPPS11	At4g36810	GGPPS11 / G11**	Plastids
			GGPPS11S / sG11**	Cytosol

diphosphates grow longer, however, their transport between cell compartments becomes increasingly restrained (Bick and Lange, 2003).

The two pathways for the production of isoprenoid precursors have been extensively studied in Arabidopsis (*Arabidopsis thaliana*; Fig. 1). All the MEP pathway enzymes are encoded by nuclear genes and imported into plastids, whereas cytosolic, endoplasmic reticulum (ER), and peroxisomal-associated locations have been found for MVA enzymes (Pulido et al., 2012; Rodriguez-Concepción and Boronat, 2015). The main rate-determining enzymes of the MEP and MVA pathways are deoxyxylulose 5-phosphate synthase (DXS) and hydroxymethylglutaryl-CoA reductase (HMGR), respectively (Fig. 1). Most plants contain small gene families encoding these two enzymes (Rodriguez-Concepción and Boronat, 2015). While several Arabidopsis genes encode proteins with homology to DXS, only one of them produces a functional enzyme with DXS activity (Phillips et al., 2008a). In the case of HMGR, the *HMG1* gene produces long and short transcripts encoding two enzyme isoforms (HMGR1L and HMGR1S, respectively) that only differ in their N-terminal region, whereas the *HMG2* gene produces only one isoform, HMGR2 (Caelles et al., 1989; Enjuto et al., 1994; Lumbreras et al., 1995). The three HMGR isoforms are primarily attached to the ER and have the same topology in the membrane, with the highly divergent N-terminal region and the highly conserved catalytic domain exposed to the cytosol.

Downstream enzymes such as IDI, FPP synthase (FPPS), and GGPP synthase (GGPPS) are also encoded by small gene families in Arabidopsis and localize to multiple subcellular compartments (Fig. 1). The two

genes encoding IDI in Arabidopsis, *IDI1* and *IDI2*, produce long and short transcripts encoding enzyme isoforms that differ in length at their N-terminal ends (Okada et al., 2008; Phillips et al., 2008b; Sapir-Mir et al., 2008). The long IDI1 isoform is targeted to plastids, the long IDI2 isoform is transported to mitochondria, and both short isoforms are sorted to peroxisomes. The two genes encoding FPPS in Arabidopsis produce three isoforms (Cunillera et al., 1997; Manzano et al., 2006; Keim et al., 2012). *FPS1* encodes a long isoform targeted to mitochondria (FPP1L) and a short one lacking the N-terminal end that remains in the cytosol, whereas *FPS2* only produces a cytosolic enzyme (Fig. 1). Unlike IDI and FPPS, GGPPS paralogs are encoded by a high number of genes in plant genomes, with a particularly large gene family present in Arabidopsis (Lange and Ghassemian, 2003; Coman et al., 2014). From the 12 initially reported genes, however, only four have been conclusively shown to encode true GGPPS enzymes (Nagel et al., 2015; Wang et al., 2016). Two of them, *GGPPS3* and *GGPPS4*, encode proteins sorted to the ER, and the other two, *GGPPS2* and *GGPPS11*, encode plastidial isoforms (Okada et al., 2000; Beck et al., 2013; Coman et al., 2014; Ruiz-Sola et al., 2016). The *GGPPS1* gene encodes the only mitochondrial member of the family, but the in vivo activity of the protein is still unclear (Zhu et al., 1997; Okada et al., 2000; Beck et al., 2013; Nagel et al., 2015; Wang et al., 2016). To date, the production of more than one enzyme isoform from a single GGPPS-encoding gene has not been reported.

Despite the presence of at least two GGPPS enzymes in Arabidopsis plastids, GGPPS11 (At4g36810, from herein referred to as G11) is by far the most abundant and ubiquitously expressed isoform (Beck et al., 2013; Ruiz-Sola et al., 2016). G11 is required for the production of all major groups of plastidial isoprenoids, including carotenoids and the side chains of chlorophylls, tocopherols, and prenylated quinones (Ruiz-Sola et al., 2016). Strikingly, several phenotypes have been described for G11-defective mutant alleles (Ruppel et al., 2013; Ruiz-Sola et al., 2016). By carrying out a comprehensive analysis of these alleles, we uncovered here the existence of two differentially targeted G11 enzymes, each of them indispensable for a distinct developmental process likely through the production of a different group of essential isoprenoids.

RESULTS AND DISCUSSION

Different G11-Defective Alleles Show a Range of Phenotypes from Variegation to Embryo Lethality

To better understand the role of G11, the most abundant GGPPS isoform in Arabidopsis, we carefully revised the phenotypes associated to partial or complete loss-of-function mutants (Fig. 2). The *ggpps11-1* mutant, originally named *ggps1-1* (Ruppel et al., 2013) and here referred to as *g11-1*, harbors a point mutation that

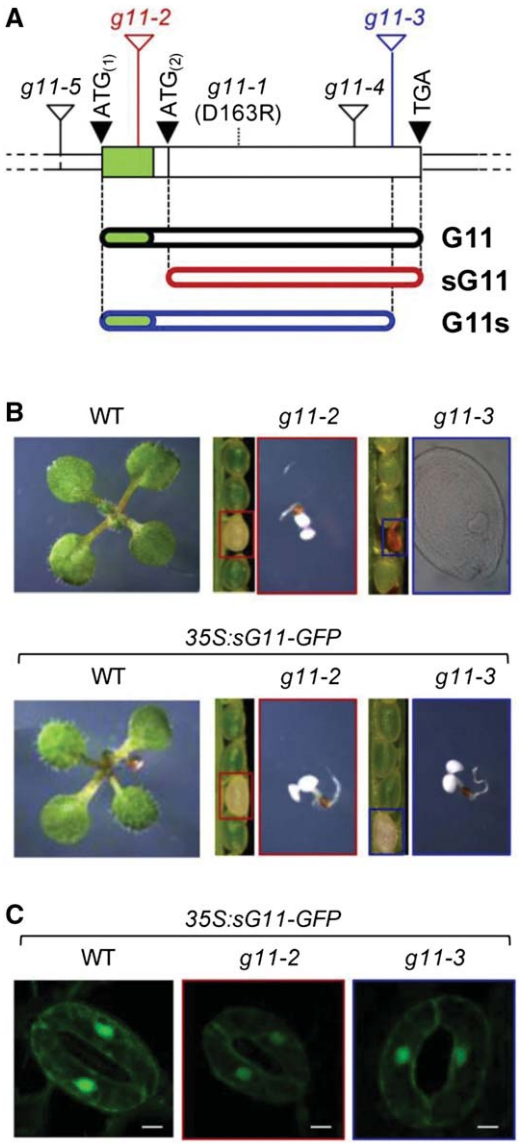


Figure 2. G11 mutant alleles and associated phenotypes. A, *G11* gene model according to TAIR v10 annotation. The protein-coding sequence (which lacks introns) is shown as a wider box with a green section corresponding to the predicted plastid targeting sequence. The position of translation start and stop codons is shown with black triangles. The position of T-DNA insertions is represented with white triangles. The position of the point mutation in the *g11-1* allele is shown with a dashed line. Lower bars represent encoded proteins. B, Phenotype of G11-defective mutants either producing or not producing a sG11 protein. Representative seedlings of the indicated genotypes grown under long-day conditions for 10 d are shown to the same scale. Segregating populations of seeds in siliques of plants heterozygous for the *g11-2* and *g11-3* mutations is also shown. Boxed seeds correspond to the homozygous albino mutants represented in the right. Brownish seeds did not produce seedlings due to a blockage in embryo development at the heart stage (as shown in the corresponding picture). C, Cytosolic localization of the sG11-GFP protein. Pictures show GFP fluorescence from the sG11-GFP protein in stomata from 10-d-old seedlings of the indicated genotypes (bars, 5 μm).

changes a conserved residue (D163R) in G11 (Fig. 2A; Supplemental Fig. S1). The *g11-1* allele shows a temperature-dependent variegated phenotype that resembles that of the *chs5* mutant (Araki et al., 2000), later renamed *dxs-3* (Phillips et al., 2008a), which harbors a D627N mutation in DXS. It is therefore likely that the phenotype of these mutants might be associated to the temperature sensitivity of the corresponding DXS or G11 mutant enzymes, both of them involved in the production of photosynthesis-related isoprenoids (Fig. 1).

A second partial loss-of-function allele, *ggpps11-5* (SALK_140601, *g11-5*), shows a pale phenotype and a developmental delay, likely because a T-DNA insertion upstream of the predicted ATG translation start codon (Fig. 2A; Supplemental Fig. S1) results in a decreased accumulation of *G11* transcripts (Ruiz-Sola et al., 2016). In this mutant, lower levels of G11-encoding transcripts are expected to result in an overall reduction in the accumulation of fully active, wild-type G11 protein (Ruiz-Sola et al., 2016). Similarly, a general inhibition of the MEP pathway with sublethal concentrations of the DXS inhibitor clomazone also results in a pale phenotype (Pulido et al., 2013; Perelló et al., 2014). Therefore, the phenotype of *g11-1* and *g11-5* plants is consistent with the reported role of G11 as the major isoform transforming MEP-derived precursors into GGPP for photosynthesis-related isoprenoid products. Further supporting this conclusion, a seedling-lethal albino phenotype visually identical to that of knockout MEP pathway mutants such as *dxs-1*, also known as *cla1* (Phillips et al., 2008a), was observed in the case of the *ggpps11-2* line (SALK_015098, *ggps1-2* or *g11-2*), which harbors a T-DNA insertion in the N-terminal end of the protein coding region of the *G11* gene (Fig. 2; Supplemental Fig. S1; Ruppel et al., 2013). By contrast, T-DNA insertions interrupting the C-terminal end of the G11 protein in alleles *ggpps11-3* (SALK_085914, *ggps1-3* or *g11-3*) and *ggpps11-4* (SAIL_712_D06, *g11-4*) cause an arrest of embryo development (Fig. 2; Supplemental Fig. S1; Ruppel et al., 2013; Ruiz-Sola et al., 2016). This embryo lethal phenotype has never been observed in MEP pathway mutants (Phillips et al., 2008a).

The Distinct Phenotypes of *G11* Alleles Correlate with Differential Subcellular Localization and Activity of the Corresponding Enzymes

To investigate the molecular basis of the puzzling phenotype differences observed between *g11-2* (albino, seedling lethal) and *g11-3* (embryo lethal) plants (Fig. 2B), we first validated the position of the T-DNA in the mutant genomes by PCR amplification and sequencing of the insertion sites (Supplemental Fig. S1). Insertion of the T-DNA within the predicted plastid targeting sequence in the *g11-2* allele (Fig. 2A; Supplemental Fig. S1) is expected to prevent the transcription of a full-length *G11* cDNA harboring the first ATG codon ($ATG_{(1)}$). However, a second in-frame ATG codon ($ATG_{(2)}$) exists downstream of the T-DNA insertion that could potentially act as an alternative translation initiation point to generate a shorter protein, which we named sG11 (Fig. 2A; Supplemental Fig. S2). We speculated that this shorter protein might not be imported into plastids as it lacked the N-terminal plastid-targeting domain. To test this prediction, a DNA sequence encoding sG11 was generated by removing the $ATG_{(1)}$ codon. The generated sequence was then fused to the N terminus of the GFP reporter in the pCAMBIA1302 vector to obtain the *35S:sG11-GFP* construct. As shown in Figure 2C and Supplemental Figure S3, green fluorescence corresponding to the sG11-GFP protein was excluded from plastids and localized in the cytosol of cells from Arabidopsis plants transformed with the construct. By contrast, transgenic plants expressing a similar construct with the wild-type G11 sequence (Ruiz-Sola et al., 2016) showed a predominant association of GFP fluorescence to plastids (Supplemental Fig. S3).

We next evaluated whether sG11 retained the genuine enzymatic activity of the full-length enzyme (Fig. 3). Recent in vitro activity assays followed by analysis of reaction products by liquid chromatography-mass spectrometry (LC-MS; Nagel et al., 2015) or thin-layer

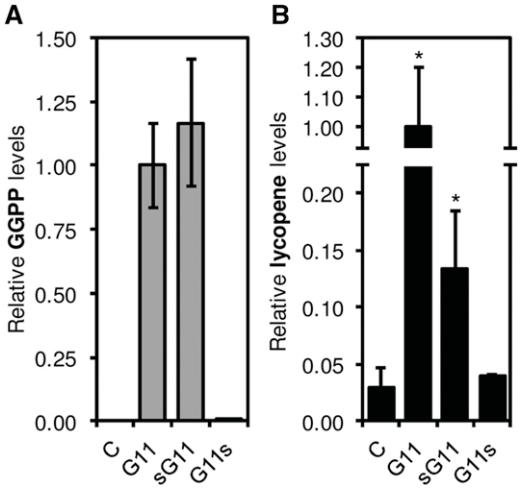


Figure 3. GGPPS activity assays. A, In vitro assays. Protein extracts from *E. coli* cells overproducing similar amounts of the indicated proteins or transformed with an empty vector (C) were mixed with IPP and DMAPP, and the production of GGPP was quantified by LC-MS. Levels are represented relative to those in G11 samples. Mean and SD of $n = 3$ extracts are shown. No GGPP was detected in C samples, and only traces were identified in G11s extracts. B, In vivo assays. Bacterial cells were cotransformed with pACCRT-BI (which lacks a functional GGPPS-encoding gene to produce lycopene) together with a plasmid expressing the indicated G11 isoform or an empty vector (C). Positive transformants were used to measure lycopene levels by normalizing A_{472} to bacterial growth (OD 600 nm). Levels are represented relative to those in G11 samples. Data correspond to the mean and SD of $n = 10$ independent transformants. Asterisks mark statistically significant differences from the C control ($P < 0.05$, one-tailed Student's *t* test assuming equal variances).

chromatography (Wang et al., 2016) confirmed that G11 synthesizes GGPP as the main product. They also revealed that other proteins previously believed to be true GGPPS isoforms actually belong to a novel type of prenyldiphosphate synthases that mainly produce C25 geranylfarnesyl diphosphate (GFPP) or even longer products (Nagel et al., 2015; Wang et al., 2016). The sG11 protein lacks 19 residues predicted to be present in the N-terminal region of the mature G11 enzyme (Fig. 2A; Supplemental Fig. S2). While this N-terminal sequence is not conserved in other GGPPS enzymes and does not include residues determining product length (Supplemental Fig. S2), we aimed to experimentally confirm whether its absence in sG11 had any impact on the ability of the protein to produce GGPP from IPP and DMAPP. In vitro activity assays like those described in Nagel et al. (2015) were carried out using protein extracts from *Escherichia coli* cells overproducing sG11 or a pseudomature form of G11 lacking the predicted plastid-targeting sequence (Supplemental Fig. S4). Analysis of reaction products by LC-MS detected the production of similar amounts of GGPP in both G11 and sG11 extracts (Fig. 3A). No GPP, FPP, or GFPP were detected in the assays (Supplemental Fig. S4), further indicating that the sG11 protein remains an active GGPPS enzyme. To next determine whether sG11 could also produce GGPP in vivo, we used a heterologous system based on carotenoid (lycopene) production in *E. coli* strains carrying the pACCRT-BI vector (Beck et al., 2013). Plasmids encoding the full-length G11 or sG11 proteins were used together with pACCRT-BI to cotransform *E. coli* cells. While cells cotransformed with pACCRT-BI and an empty plasmid control synthesized minor amounts of lycopene due to the presence of only trace levels of GGPP in the bacteria (Vallon et al., 2008), those harboring G11 and sG11 constructs showed significantly increased levels of the carotenoid (Fig. 3B), supporting the conclusion that sG11 synthesizes GGPP in vivo. Together, our results suggest that the lack of the N-terminal region in the sG11 protein produced by *g11-2* plants prevents its targeting to plastids but does not override GGPPS activity.

In the case of the *g11-3* mutant, the T-DNA interrupts the sequence encoding the highly conserved C-terminal region of G11 (Supplemental Fig. S2), resulting in a shorter protein that we named G11s (Fig. 2A). Sequencing of the T-DNA insertion site in the *g11-3* genome confirmed that the last 21 residues of the wild-type G11 protein are replaced by a single Ser residue in the G11s protein (Supplemental Fig. S1). To test whether loss of the C-terminal region compromised GGPPS activity in the G11s protein, the corresponding DNA sequence was amplified from *g11-3* seedlings and cloned into plasmids for *E. coli* expression. Both in vitro (Fig. 3A) and in vivo (Fig. 3B) activity assays showed that the recombinant G11s protein is unable to produce GGPP. These results suggest that the blockage of embryo development at the heart stage observed in the *g11-3* mutant (Fig. 2B; Supplemental Fig. S5) and the *g11-4* allele (Ruiz-Sola et al., 2016) might be due to a complete lack of G11 activity. This embryo-lethal phenotype was complemented by expressing a genomic *G11* sequence including the promoter and the full protein-coding region (Ruiz-Sola et al., 2016). Most interestingly, embryo development was also rescued by expressing the cytosolic sG11-GFP protein in *g11-3* plants (Fig. 2, B and C). Transgenic *g11-3 35S:sG11-GFP* plants, however, showed a seedling-lethal (albino) phenotype resembling the *g11-2* mutant. As expected, the cytosolic sG11-GFP protein was unable to rescue the albino phenotype of the *g11-2* mutant (Fig. 2B). We therefore concluded that embryo development beyond the heart stage required the presence of G11 activity in the cytosol (despite two ER-associated GGPPS enzymes exist in Arabidopsis, GGPPS3 and 4), whereas photosynthetic seedling development required the activity of G11 in the plastid (despite a second plastidial enzyme with the same activity, GGPPS2, is found in this plant).

Several Transcription Initiation Sites in the G11 Gene Lead to the Production of Isoforms with Different N-Terminal Ends

A number of Arabidopsis genes encoding isoprenoid biosynthetic enzymes have been shown to produce transcripts of different lengths encoding isoforms with or without N-terminal transit peptides for plastids and mitochondria (Fig. 1). To determine whether a similar alternative transcription initiation mechanism also occurs for *G11*, rapid amplification of cDNA 5′ ends (5′-RACE) experiments were performed on RNA extracted from different tissues to assess the length of *G11* transcripts in vivo (Fig. 4). The protocol used reverse primers optimized for 5′-RACE (see Fig. 4A; Supplemental Table S1) to amplify gene-specific PCR products while ruling out the possibility of genomic DNA contamination (see "Materials and Methods"). Separation of PCR-amplified products by gel electrophoresis showed the presence of cDNAs of different sizes (Fig. 4B), suggesting that the *G11* gene lacks a strictly defined transcription start site. However, amplified fragments could be grouped in two major classes: "long" (approximately 0.5 kb or longer) and "short" (approximately 0.4 kb or shorter). The 5′-RACE products amplified from siliques were cloned and sequenced. Analysis of inserts revealed that all the "long" products included the $ATG_{(1)}$ codon and hence encoded the full-length G11 protein. While most "short" products contained the $ATG_{(1)}$ codon with or without a few nucleotides upstream (up to 25), some lacked $ATG_{(1)}$ and so they can only produce the cytosolic sG11 protein by using the $ATG_{(2)}$ codon (Fig. 4A; Supplemental Fig. S1). The similar pattern of "long" and "short" transcripts detected in seedlings, rosette leaves, and flowers by 5′-RACE experiments (Fig. 4B) suggests that both types of transcripts are likely produced in all tissues. To verify whether the relative abundance of transcripts either lacking or not the $ATG_{(1)}$ codon was indeed

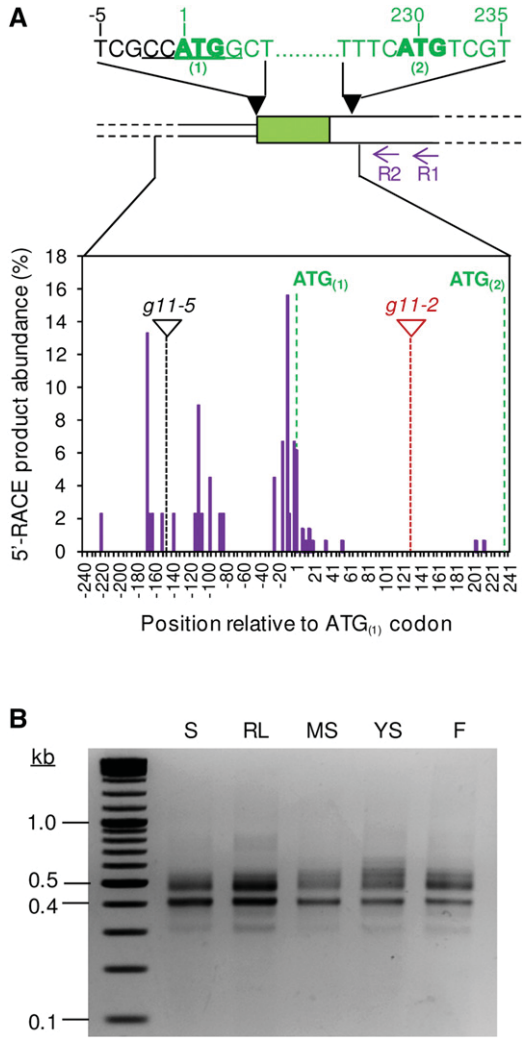


Figure 4. Transcription and translation start sites in the *G11* gene. A, *G11* gene model showing the position of presumed translation start codons $ATG_{(1)}$ and $ATG_{(2)}$ (in bold). The *Nco*I site in the $ATG_{(1)}$ region is underlined. Numbers indicate the position relative to the first nucleotide of the $ATG_{(1)}$ codon. The position of primers G11-R-R1 and R2 used for 5′-RACE experiments is also shown. Amplified 5′-RACE products (B) were cloned and sequenced to determine the position of transcription initiation sites. The graph represents the percentage of transcripts found to start at the indicated position in siliques based on the analysis of 145 clones (purple columns). The position of $ATG_{(1)}$ and $ATG_{(2)}$ codons and T-DNA insertion sites of the indicated alleles is also shown. B, Agarose gel electrophoresis of 5′-RACE products isolated from seedlings (S), rosette leaves (RL), mature siliques (MS), young siliques (YS), and flowers (F).

similar in different tissues, 5′-RACE products from seedlings were cloned and compared with those from siliques by digestion with *Nco*I. As shown in Figure 4A, a *Nco*I target site overlaps the $ATG_{(1)}$ codon, and therefore it can be used to rapidly identify clones lacking this sequence. In both seedlings and siliques, about 10% of the clones could not be cleaved by *Nco*I, confirming that transcripts exclusively encoding sG11 are low abundant but indeed detectable at similar levels in different tissues.

G11 Activity Is Essential to Produce Both Plastidial Isoprenoids and Unidentified Extraplastidial Isoprenoids Required for Embryo Development

We next aimed to determine the nature of the isoprenoids derived from GGPP produced by plastidial and cytosolic forms of G11. Our previous results with the *g11-5* allele showed that the pale phenotype of the mutant was due to a decreased expression of the *G11* gene, which caused a reduced accumulation of the major groups of MEP-derived isoprenoids produced from GGPP in the plastid, i.e. carotenoids, chlorophylls, tocopherols, phylloquinones, and plastoquinone (Ruiz-Sola et al., 2016). To verify whether the albino phenotype of *g11-2* seedlings was the result of a blockage in the production of plastidial GGPP-derived isoprenoids, we analyzed the levels of carotenoids in the *g11-2* mutant (Fig. 5). As controls, we used *dxs-1*, a MEP pathway null mutant with a complete loss of DXS activity (Phillips et al., 2008a), and *psy-1*, a knockout mutant in which a specific blockage in the carotenoid pathway results in a similar albino phenotype (Pokhilko et al., 2015; Fig. 5A). Analysis of *g11-2* seedlings showed low but detectable levels of carotenoids such as lutein, β-carotene, and β,β-xanthophylls (Fig. 5B). The levels of all these metabolites in *g11-2* seedlings were similar to those in the *dxs-1* mutants but much higher than the amounts detected in *psy-1* seedlings. The results suggest that loss of G11 or DXS activity do not completely block the production of carotenoids in seedlings (as the loss of PSY activity does). The carotenoids detected in *g11-2* seedlings might derive from small amounts of GGPP synthesized from MEP-derived precursors by GGPPS2 (the other plastidial GGPPS enzyme found in Arabidopsis). In the case of the *dxs-1* mutant, however, the MEP pathway is completely blocked, and hence it is most likely that MVA-derived IPP or DMAPP precursors are transported to the plastid and used to produce GGPP and downstream products. Alternatively, an enhanced import of cytosolic GGPP by nondifferentiated plastids like those found in the albino mutants would make DXS and G11 (but not PSY) dispensable to produce carotenoids. In any case, such a residual production of plastidial isoprenoids is clearly insufficient to support photosynthetic development. We therefore conclude that the albino phenotype of *g11-2* seedlings is due to a defective production of GGPP and downstream isoprenoids in the plastid.

To identify the isoprenoid metabolite required for embryo development that is produced from sG11-derived GGPP, it would be necessary to compare the metabolite profile of *g11-2* and *g11-3* embryos in the transition from globular to heart stage (Fig. 2; Supplemental Fig. S5). Because this is extremely challenging, we followed an alternative approach and evaluated whether the levels of extraplastidial

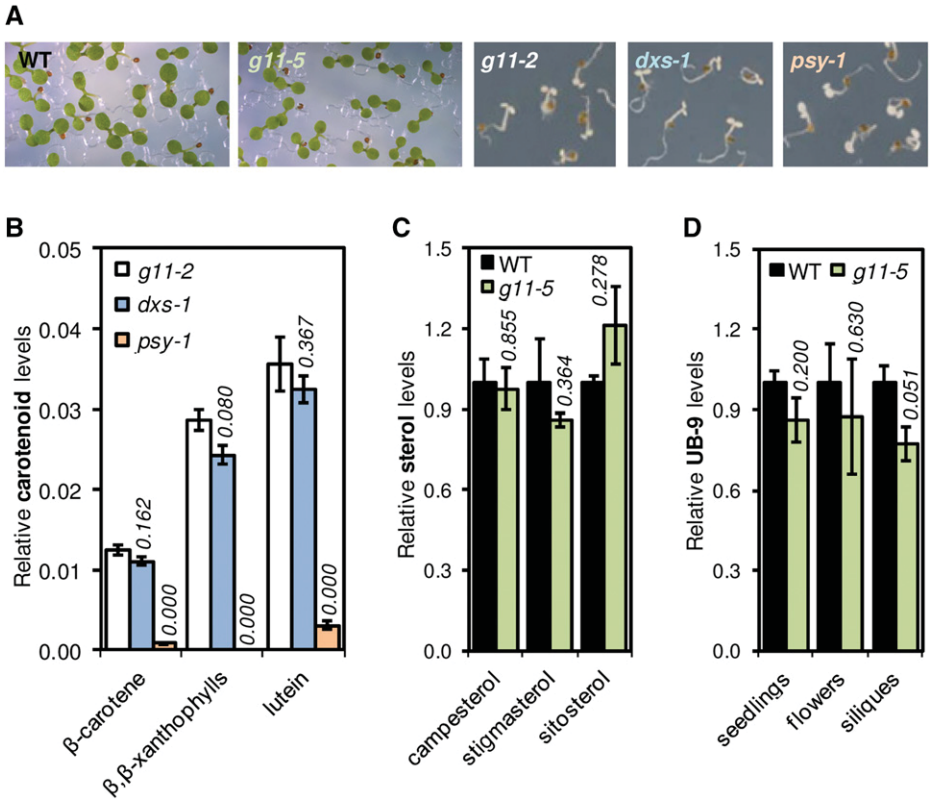


Figure 5. Levels of MEP-derived and MVA-derived isoprenoid products in plants with reduced G11 activity. A, Seedlings of the indicated mutants and the parental Columbia wild-type (WT) grown for 1 week under short-day conditions. These seedlings were used for the metabolite analyses shown in the corresponding graphs. B, Levels of the indicated carotenoids in albino mutants relative to those in wild-type seedlings. C, Levels of major sterols in wild-type and G11-defective *g11-5* seedlings. D, Levels of the ubiquinone UB-9 in seedlings, flowers, and young siliques of wild-type and *g11 -5* plants. The graphs represent mean and SE of at least $n = 3$ independent samples. Italic numbers above the bars indicate *P* values (Student's *t* test) relative to *g11-2* (B) or wild-type (C and D) samples.

isoprenoids were altered in the *g11-5* mutant, which is expected to produce lower amounts not only of plastid-localized G11 but also of cytosolic sG11 enzymes. While MVA-derived precursors are used to produce a wide variety of isoprenoids (Fig. 1), only sterols and ubiquinone appear to be required for proper embryo development (Schrick et al., 2000; Okada et al., 2004). Sterols are not synthesized from GGPP but from FPP (Fig. 1). We therefore expected that reducing sG11 activity (i.e. cytosolic GGPP production) in *g11-5* seedlings would not cause a decreased sterol accumulation. Consistently, the levels of sterols (campesterol, stigmasterol, and sitosterol) were found to be similar in wild-type and *g11-5* seedlings (Fig. 5C).

In the case of ubiquinone, the predominant form in Arabidopsis (UB-9) contains a C45 solanesyl moiety synthesized by a mitochondrial polyprenyl diphosphate synthase that elongates an initial FPP or GGPP acceptor with IPP units (Hsieh et al., 2011; Ducluzeau et al., 2012). Mitochondria import MVA-derived IPP from the cytosol, as they lack their own biosynthetic pathway (Lütke-Brinkhaus et al., 1984; Disch et al., 1998). Then, specific isoforms of FPPS and GGPPS enzymes targeted to mitochondria (in Arabidopsis, FPPS1L, and GGPPS1; Fig. 1) are presumed to produce FPP and GGPP for ubiquinone synthesis. FPPS1L-defective *fps1* mutants only showed a limited decrease in UB-9 levels (Closa et al., 2010), suggesting that the biosynthesis of the ubiquinone solanesyl chain might predominantly rely on the supply of GGPP by GGPPS1, the only GGPPS isoform known to be targeted to mitochondria (Beck et al., 2013; Nagel et al., 2015). However, we found that the T-DNA insertion mutant *ggpps1-1* (SAIL_559_G01) contains wild-type levels of UB-9 (Supplemental Fig. S6). Levels of other GGPP-derived isoprenoids were also unaltered in *ggpps1-1* seedlings, which showed a wild-type phenotype in terms of plant growth and development (Supplemental Fig. S6). Actually, the role of GGPPS1 as a true GGPPS enzyme still remains controversial as it has not been conclusively established whether its main product is GGPP (Wang et al., 2016) or GFPP (Nagel et al., 2015).

In the absence of a mitochondrial source of GGPP, cytosolic GGPP might be transported to this organelle and used for ubiquinone synthesis. Blockage of this cytosolic source when sG11 activity is lost could actually explain why embryo development is arrested at the same heart stage in mutants defective in ubiquinone synthesis (Okada et al., 2004) and sG11 activity, i.e. *g11-3* (Supplemental Fig. S5) and *g11-4* (Ruiz-Sola et al., 2016). Analysis of UB-9 contents in seedlings, flowers, and young siliques of the *g11-5* mutants detected slightly reduced levels compared to wild-type samples, but the differences were not found to be statistically significant (Fig. 5D). In the case of mutant siliques, however, the observed reduction in ubiquinone levels was close to statistical significance (Student's *t* test; $P = 0.051$). Together, these results suggest two possibilities. First, sG11-derived GGPP might be critical for the biosynthesis of ubiquinone during embryogenesis,

allowing embryo development to progress beyond the heart stage. Considering that a ca. 50% reduction of *G11* transcript levels in *g11-5* seedlings only causes a 20%–30% decrease in plastidial isoprenoid content (Ruiz-Sola et al., 2016), the observed 20% reduction of UB-9 levels in young siliques of the *g11-5* mutant might indeed result from partly reduced sG11 activity (and hence GGPP supply) in developing embryos or seeds. Because other GGPPS isoforms are also expressed in these tissues at different stages (Beck et al., 2013), it remains unclear why none of them is able to complement the loss of sG11 activity and hence rescue embryo development in *g11-3* and *g11-4* mutants. Alternatively, sG11 might produce GGPP for another specific class of unidentified metabolites required for embryogenesis and, perhaps, with roles in other cells and tissues during the plant life cycle, as deduced from the wide distribution of sG11-encoding transcripts (Fig. 4B).

CONCLUSION

Our results support the conclusion that the Arabidopsis *G11* gene can produce two enzyme isoforms: one translated from the $ATG_{(1)}$ codon and carrying a plastid-targeting peptide (G11) and a shorter version translated from the downstream $ATG_{(2)}$ codon and lacking the plastid-targeting peptide (sG11). Mechanistically, the production of these two differentially targeted isoforms might rely on both the use of alternative transcription start sites (resulting in the synthesis of mRNAs with or without a 5′-region encoding the plastid transit peptide) or the use of alternative translation start sites ($ATG_{(1)}$ or $ATG_{(2)}$) in the long transcript. The NetStart algorithm for the prediction of translation start codons in Arabidopsis (http://www.cbs.dtu.dk/services/NetStart/) actually gives a similar score to $ATG_{(1)}$ (0.627) and $ATG_{(2)}$ codon (0.651), suggesting that both could be functional in vivo. It is remarkable that a similar situation has been reported to occur in other Arabidopsis genes encoding key isoprenoid biosynthetic enzymes, including HMGR, IDI, and FPPS. Unlike these other enzymes, however, no redundancy appears to occur in the case of Arabidopsis GGPPS enzymes, as the loss of G11 cannot be rescued by plastidial GGPPS2, ER-localized GGPPS3 and GGPPS4, or mitochondrial GGPPS1. Furthermore, the Met residue encoded by the $ATG_{(2)}$ codon in G11 does not appear to be conserved in the rest of isoforms confirmed to synthesize GGPP (Supplemental Fig. S2), suggesting that *G11* might be the only Arabidopsis GGPPS-encoding gene producing more than one isoform.

Whereas G11 activity in the plastid is indispensable for the production of plastidial isoprenoids supporting chloroplast development and photosynthesis, sG11 activity in the cytosol supplies the precursors for an unidentified isoprenoid metabolite that is essential for embryo development. Interestingly, loss of sG11 activity in *g11-3* and *g11-4* mutants prevents embryo development to proceed beyond the heart stage (Supplemental Fig. S5; Ruiz-Sola et al., 2016), similar to that observed in ubiquinone-defective mutants (Okada et al., 2004), whereas complete loss of FPPS activity in Arabidopsis *fps1 fps2* double mutants blocks embryogenesis at the earlier globular stage (Closa et al., 2010). These results suggest that while FPP-derived isoprenoids are needed for the transition of the embryo from globular to heart stages, progression beyond the heart stage requires ubiquinone or/and a different metabolite produced from sG11-supplied precursors. Besides synthesizing GGPP as a homodimer, the plastidial G11 isoform has been found to produce GPP upon heterodimerization with another plastidial GGPPS-like protein (Wang and Dixon, 2009; Chen et al., 2015). Whether the cytosolic sG11 protein unveiled here also has alternative enzyme activities upon association with other proteins remains unknown. In any case, our results support the conclusion that the production of GGPP required for essential functions in different cell compartments relies on the activity of G11 isoforms. Other GGPPS paralogs might be maintained in the Arabidopsis genome for developmental and/or condition-specific subfunctionalization. Future experiments, including the high-resolution analysis of isoprenoid profiles in specialized tissues and organs (e.g. embryos) from wild-type and GGPPS-defective mutants, should provide additional insights on the biological role of specific isoforms and their corresponding downstream GGPP-derived products, further allowing to understand the astounding complexity of the mechanisms used by plants to produce isoprenoids.

MATERIALS AND METHODS

Plant Material and Constructs

All the Arabidopsis (*Arabidopsis thaliana*) lines used in this work are in the Columbia background. The T-DNA insertion alleles *g11-3*, *g11-4*, *g11-5*, *dxs-1*, and *psy-1* were already available in the lab (Pokhilko et al., 2015; Ruiz-Sola et al., 2016). Homozygous *ggpps1-1* mutants were isolated from a segregating population of the SAIL_559_G01 line supplied by the European Arabidopsis Stock Centre (http://arabidopsis.info/). Primers for PCR-based genotyping of the mutants and sequencing of T-DNA insertion sites are indicated in Supplemental Table S1. The *35S:sG11-GFP* construct was generated by deleting the $ATG_{(1)}$ codon of the *G11* sequence in the pSG11G plasmid (Ruiz-Sola et al., 2016). As shown in Supplemental Table S2, the pSG11G plasmid was digested with *Nco*I (which has a target sequence overlapping the $ATG_{(1)}$ codon; Fig. 4A) and subsequently treated with Mung Bean nuclease (New England Biolabs) to remove single-stranded extensions and generate blunt ends that were eventually ligated using T4 ligase enzyme (Roche). After *Agrobacterium*-mediated plant transformation, homozygous lines containing a single T-DNA insertion of the *35S:sG11-GFP* construct were selected based on the segregation of the resistance marker (hygromycin). Seeds were surface-sterilized and germinated in petri dishes with solid Murashige and Skoog medium without vitamins or Suc. After stratification for 3 d at 4°C, plates were incubated in a growth chamber at 22°C and illuminated for 16 h (long-day) or 8 h (short-day) a day with fluorescent white light at a photosynthetic photon flux density of 60 μmol m^{-2} s^{-1}.

GGPPS Activity Assays

Constructs to express different truncated G11 versions were generated as described (Supplemental Table S2). Recombinant proteins were produced in *Escherichia coli* BL21 pGROE cells using the Overnight Express AutoInduction System 1 (Merck Millipore). After growth for 72 h at 18°C, bacterial cells were

recovered by centrifugation, and pellets were resuspended in reaction buffer (25 mM HEPES, pH 7.2, 10 mM $MgCl_2$, 10% v/v glycerol) supplemented with 1 mg/mL lysozyme and one tablet of complete protease inhibitor cocktail (Roche) for every 10 mL of buffer. The resuspended pellet was incubated at 4°C for 10 min, and after a brief sonication (five pulses of 20 s at 30 W), the cell lysate was centrifuged at 19,000*g* at 4°C for 5 min. The supernatant was used for SDS-PAGE, protein quantification, and GGPP activity assays as described (Nagel et al., 2015). The reaction mix contained 10 μg of total protein from extracts showing similar levels of recombinant protein in 200 μL of reaction buffer, 50 μM IPP, and 50 μM DMAPP. After incubation for 1 h at 30°C, the reaction was stopped by adding 800 μL of methanol. A previously described LC-FTMS system (Henneman et al., 2008) was adapted to detect prenyl diphosphate products. A Luna C18(2) precolumn (2.0 × 4 mm) and an analytical column (2.0 × 150 mm, 100 nm, particle size 3 μm) from Phenomenex were used for chromatographic separation at 40°C, using an Acquity UPLC (H-Class), connected to an LTQ-Orbitrap XL hybrid mass spectrometry system (Waters) operating in negative electrospray ionization mode heated at 300°C with a source voltage of 4.5 kV for full-scan LC-MS in the *m/z* range 100 to 1300. The injection volume was 10 μL. Compounds were separated by a linear gradient between solution A (20 mM NH_4HCO_3 with 0.1% triethylamine) and solution B (acetonitrile/H_2O, 4:1 with 0.1% triethylamine). The gradient was as follows: 0 to 18 min, 100% A to 80% A; 18 to 23 min, 80% A to 0% A; 23 to 25 min, 0% A; 25 to 30 min, 0% A to 100% A; equilibration with 100% A. Acquisition and visualization of the data were performed using Xcalibur software (Thermo Fischer). GPP, FPP, and GGPP standards were obtained from Sigma and used for quantification.

For in vivo activity assays, *E. coli* TOP10 cells were cotransformed with both pACCRT-BI (Beck et al., 2013) and plasmids harboring the corresponding G11 versions generated as described (Supplemental Table S2). Transformants containing both plasmids were selected on LB medium containing both ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL). Positive colonies were selected and grown overnight at 37°C in liquid LB. Fresh liquid LB medium supplemented with the appropriate antibiotics was then inoculated with the overnight culture and incubated for 7 more days at 30°C. Aliquots of 10 mL of grown culture were harvested and used for lycopene extraction and quantification as described (Beck et al., 2013).

Microscopy

Subcellular localization of the GFP fusion proteins was analyzed by direct examination of plant tissue samples with an Olympus FV 1000 Confocal Laser Scanning Microscope. Green fluorescence corresponding to the fusion proteins was detected using an argon laser for excitation with blue light at 488 nm and a 500 to 510-nm filter, whereas a 610 to 700-nm filter was used for detection of chlorophyll autofluorescence. Clearing of Arabidopsis seeds and examination of embryo developmental stages was performed as described (Ruiz-Sola et al., 2016).

5′-RACE

For 5′-RACE analysis, total RNA from different organs of wild-type plants was isolated using an RNA purification kit (Sigma-Aldrich) and used for first-strand cDNA synthesis with the SMARTer RACE cDNA amplification kit (Clontech). High-Fidelity AccuPrime Taq DNA Polymerase (Invitrogen) was used with primers provided in the kit and the G11-R-R1 primer (Supplemental Table S1) for 5′-RACE reactions, adding an initial denaturation step of 2 min at 94°C to the recommended PCR program to activate the polymerase, and changing the elongation temperature from 72°C to 68°C. PCR products were cloned into the cloning vector pCRII-TOPO (Invitrogen) for further restriction enzyme and sequencing analysis using the G11-R-R2 primer (Supplemental Table S1).

Analysis of Metabolite Levels

Published methods were used for the extraction, separation, and quantification of photosynthetic pigments (chlorophylls and carotenoids; Rodríguez-Concepción et al., 2004), sterols (Closa et al., 2010), and prenylquinones, including ubiquinone (Martinis et al., 2011).

Accession Numbers

Accessions are deposited in the TAIR database (www.arabidopsis.org): At1g49530, *GGPPS1*; At4g36810, *GGPPS11* (here referred to as *G11*). See also Figure 1.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Arabidopsis *G11* gene and mutants.

Supplemental Figure S2. Sequence alignment of Arabidopsis GGPPS isoforms.

Supplemental Figure S3. Subcellular localization of G11-GFP and sG11-GFP fusion proteins in transgenic Arabidopsis plants.

Supplemental Figure S4. In vitro activity of truncated G11 isoforms.

Supplemental Figure S5. Embryogenesis is blocked at the heart stage in the *g11-3* mutant.

Supplemental Figure S6. Developmental and metabolic phenotypes of the *ggpps1-1* mutant.

Supplemental Table S1. Primers used in this work.

Supplemental Table S2. Constructs and cloning details.

ACKNOWLEDGMENTS

We are very grateful to Gaétan Glauser (University of Neuchâtel, Switzerland) for the analysis of prenylquinones. Technical support from M. Rosa Rodríguez-Goberna and members of the CRAG Services is greatly appreciated.

Received September 7, 2016; accepted September 30, 2016; published October 5, 2016.

LITERATURE CITED

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

Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodríguez-Concepción M, Gruissem W, Vranová E (2013) Characterization of the GGPP synthase gene family in Arabidopsis thaliana. Plant Mol Biol **82:** 393–416

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:** 146–154

Bouvier F, Rahier A, Camara B (2005) Biogenesis, molecular regulation and function of plant isoprenoids. Prog Lipid Res **44:** 357–429

Caelles C, Ferrer A, Balcells L, Hegardt FG, Boronat A (1989) Isolation and structural characterization of a cDNA encoding Arabidopsis thaliana 3-hydroxy-3-methylglutaryl coenzyme A reductase. Plant Mol Biol **13:** 627–638

Chen Q, Fan D, Wang G (2015) Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in Arabidopsis flowers. Mol Plant **8:** 1434–1437

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:** 512–525

Coman D, Altenhoff A, Zoller S, Gruissem W, Vranová E (2014) Distinct evolutionary strategies in the GGPPS family from plants. Front Plant Sci **5:** 230

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

Disch A, Hemmerlin A, Bach TJ, Rohmer M (1998) Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. Biochem J **331:** 615–621

Ducluzeau AL, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC, Basset GJ (2012) Gene network reconstruction identifies the authentic trans-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in Arabidopsis. Plant J **69:** 366–375

Enjuto M, Balcells L, Campos N, Caelles C, Arró M, Boronat A (1994) Arabidopsis thaliana contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. Proc Natl Acad Sci USA **91:** 927–931

Henneman L, van Cruchten AG, Denis SW, Amolins MW, Placzek AT, Gibbs RA, Kulik W, Waterham HR (2008) Detection of nonsterol isoprenoids by HPLC-MS/MS. Anal Biochem **383:** 18–24

Hsieh FL, Chang TH, Ko TP, Wang AH (2011) Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol **155:** 1079–1090

Keim V, Manzano D, Fernández FJ, Closa M, Andrade P, Caudepón D, Bortolotti C, Vega MC, Arró M, Ferrer A (2012) Characterization of Arabidopsis FPS isozymes and FPS gene expression analysis provide insight into the biosynthesis of isoprenoid precursors in seeds. PLoS One **7:** e49109

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

Lumbreras V, Campos N, Boronat A (1995) The use of an alternative promoter in the *Arabidopsis thaliana* HMG1 gene generates an mRNA that encodes a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase isoform with an extended N-terminal region. Plant J **8:** 541–549

Lütke-Brinkhaus F, Liedvogel B, Kleinig H (1984) On the biosynthesis of ubiquinones in plant mitochondria. Eur J Biochem **141:** 537–541

Manzano D, Busquets A, Closa M, Hoyerová K, Schaller H, Kamínek M, Arró M, Ferrer A (2006) Overexpression of farnesyl diphosphate synthase in Arabidopsis mitochondria triggers light-dependent lesion formation and alters cytokinin homeostasis. Plant Mol Biol **61:** 195–213

Martinis J, Kessler F, Glauser G (2011) A novel method for prenylquinone profiling in plant tissues by ultra-high pressure liquid chromatography-mass spectrometry. Plant Methods **7:** 23–34

Nagel R, Bernholz C, Vranová E, Košuth J, Bergau N, Ludwig S, Wessjohann L, Gershenzon J, Tissier A, Schmidt A (2015) Arabidopsis thaliana isoprenyl diphosphate synthases produce the C25 intermediate geranylfarnesyl diphosphate. Plant J **84:** 847–859

Okada K, Kasahara H, Yamaguchi S, Kawaide H, Kamiya Y, Nojiri H, Yamane H (2008) Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in Arabidopsis. Plant Cell Physiol **49:** 604–616

Okada K, Ohara K, Yazaki K, Nozaki K, Uchida N, Kawamukai M, Nojiri H, Yamane H (2004) The AtPPT1 gene encoding 4-hydroxybenzoate polyprenyl diphosphate transferase in ubiquinone biosynthesis is required for embryo development in Arabidopsis thaliana. Plant Mol Biol **55:** 567–577

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:** 1045–1056

Perelló C, Rodríguez-Concepción M, Pulido P (2014) Quantification of plant resistance to isoprenoid biosynthesis inhibitors. Methods Mol Biol **1153:** 273–283

Phillips MA, D'Auria JC, Gershenzon J, Pichersky E (2008b) The Arabidopsis thaliana type I Isopentenyl Diphosphate Isomerases are targeted to multiple subcellular compartments and have overlapping functions in isoprenoid biosynthesis. Plant Cell **20:** 677–696

Phillips MA, León P, Boronat A, Rodríguez-Concepción M (2008a) The plastidial MEP pathway: unified nomenclature and resources. Trends Plant Sci **13:** 619–623

Pokhilko A, Bou-Torrent J, Pulido P, Rodríguez-Concepción M, Ebenhöh O (2015) Mathematical modelling of the diurnal regulation of the MEP pathway in Arabidopsis. New Phytol **206:** 1075–1085

Pulido P, Perello C, Rodriguez-Concepcion M (2012) New insights into plant isoprenoid metabolism. Mol Plant **5:** 964–967

Pulido P, Toledo-Ortiz G, Phillips MA, Wright LP, Rodríguez-Concepción M (2013) Arabidopsis J-protein J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. Plant Cell **25:** 4183–4194

Rodríguez-Concepción M, Boronat A (2015) Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. Curr Opin Plant Biol **25:** 17–22

Rodríguez-Concepción M, Forés O, Martinez-García JF, González V, Phillips MA, Ferrer A, Boronat A (2004) Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during Arabidopsis seedling development. Plant Cell **16:** 144–156

Ruiz-Sola MA, Coman D, Beck G, Barja MV, Colinas M, Graf A, Welsch R, Rütimann P, Bühlmann P, Bigler L, et al (2016) Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol **209:** 252–264

Ruppel NJ, Kropp KN, Davis PA, Martin AE, Luesse DR, Hangarter RP (2013) Mutations in GERANYLGERANYL DIPHOSPHATE SYNTHASE 1 affect chloroplast development in Arabidopsis thaliana (Brassicaceae). Am J Bot **100:** 2074–2084

Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, Gidoni D, Eyal Y (2008) Peroxisomal localization of Arabidopsis isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes. Plant Physiol **148:** 1219–1228

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

Vallon T, Ghanegaonkar S, Vielhauer O, Müller 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:** 175–182

Vranová E, Coman D, Gruissem W (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol **64:** 665–700

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:** 1655–1660

Wang C, Chen Q, Fan D, Li J, Wang G, Zhang P (2016) Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. Mol Plant **9:** 195–204

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

SUPPLEMENTAL FIGURES

Figure S1. Arabidopsis *G11* gene and mutants. (A) The protein coding sequence according to TAIR v10 annotation is shown in uppercase (no introns are found). The first ATG codon ($ATG_{(1)}$) and the translation stop codon are marked in bold. The sequence encoding the plastid-targeting peptide is boxed in green. A second in-frame ATG codon ($ATG_{(2)}$) is boxed in white. Transcription start sites identified by 5'-RACE in siliques are marked with triangles whose color represents the percentage of transcripts starting at the indicated position (white: <5%; gray: 5-10%; black: >10%). T-DNA insertions are represented with arrows pointing towards the right border. The point mutation in the *g11-1* allele is boxed in gray. (B) Comparison of genomic sequences and encoded proteins in wild-type (WT) and *g11-3* plants in the region harboring the T-DNA insertion in the mutant. The blue box marks the T-DNA sequence inserted and the corresponding translation.

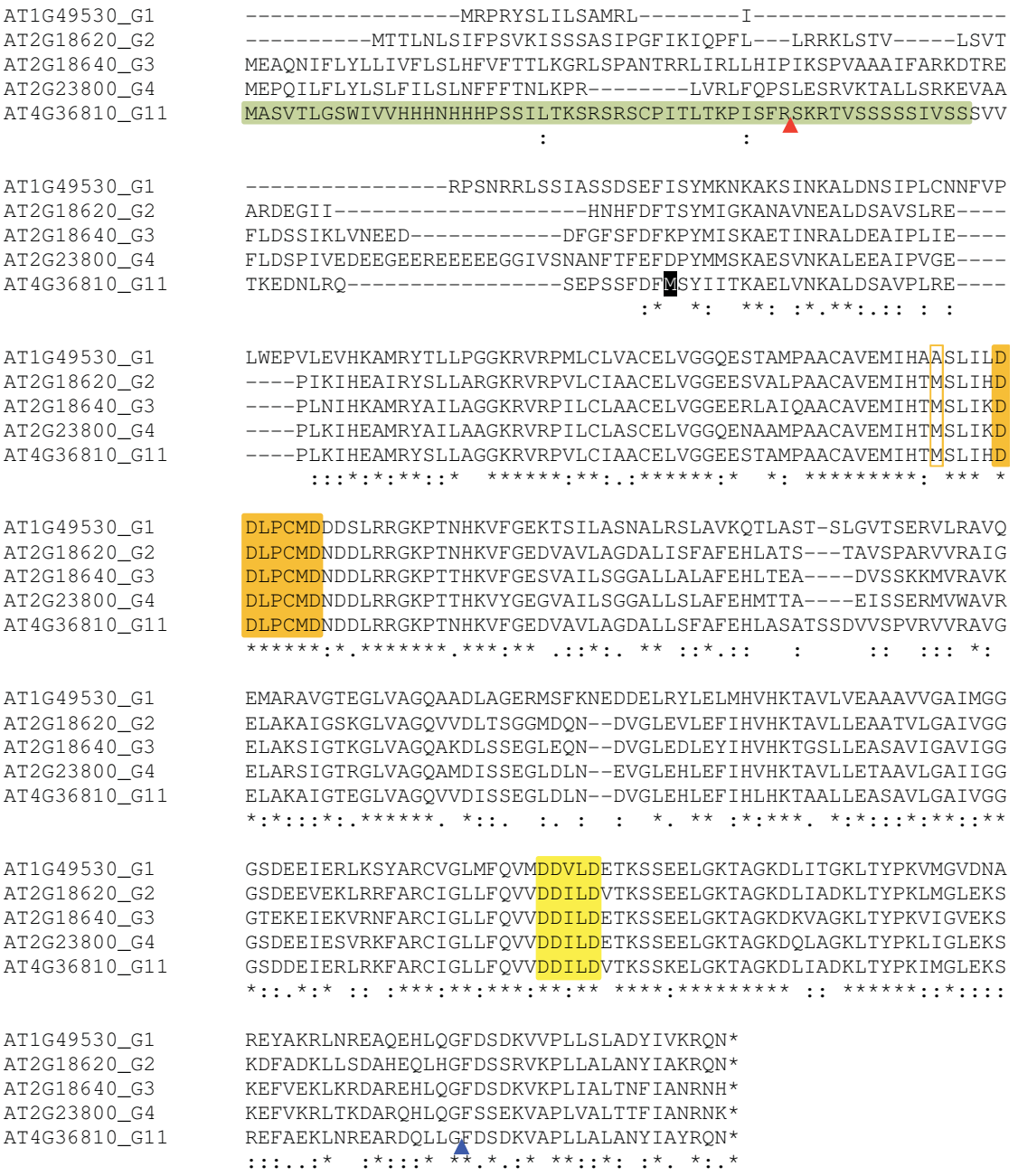


Figure S2. Sequence aligment of Arabidopsis GGPPS isoforms. Multiple aligment was performed using Clustal Omega with the default parameters (http://www.ebi.ac.uk/Tools/msa/clustalo/). Red and blue arrowheads mark the position of the T-DNA in the *g11-2* and *g11-3* mutants, respectively. The methionine encoded by the $ATG_{(2)}$ codon is boxed in black and the predicted plastid-targeting peptide is boxed in green. The conserved FARM (first aspartate-rich motif) and SARM (second aspartate-rich motif) signatures that form the catalytic cavity for allyl substrate binding, are boxed in orange and yellow, respectively. The fifth residue before the FARM motif, shown to be relevant to determine the chain length of the final product, is also indicated with an orange frame. True GGPPS enzymes (i.e. those producing C20 GGPP) have a M residue in this position, whereas the presence of smaller residues (A or S) involves a preferential production of C25 GFPP (Nagel et al., 2015; Wang et al., 2016).

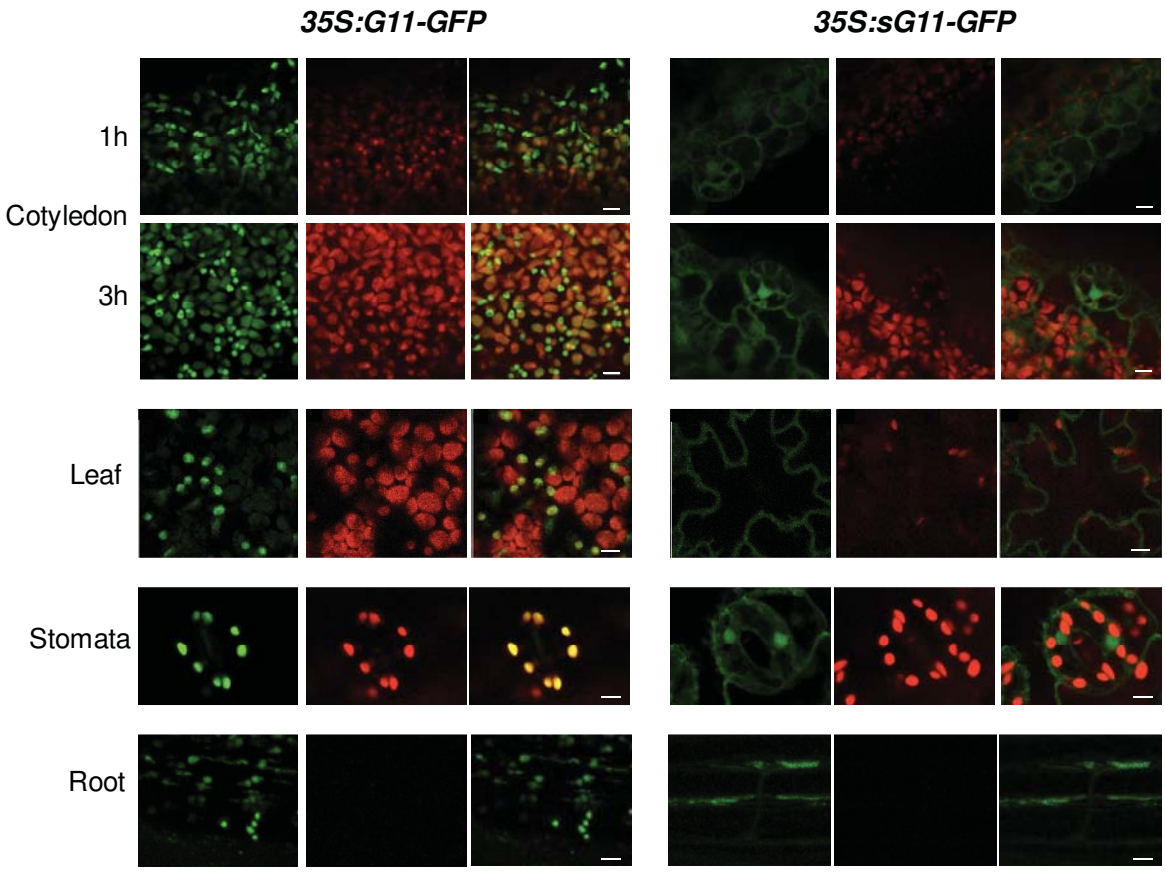


Figure S3. Subcellular localization of G11-GFP and sG11-GFP fusion proteins in transgenic Arabidopsis plants. Representative images from cotyledons of etiolated seedlings after 1 or 3 hours of exposure to light and from leaf epidermal and mesophyll cells, stomata, and roots from 10-day-old seedlings grown under long day conditions are shown. The first column shows green fluorescence from GFP, the second shows red autofluorescence from chlorophyll, and the third shows an overlay. All confocal images were scanned using similar laser gain and offset settings.

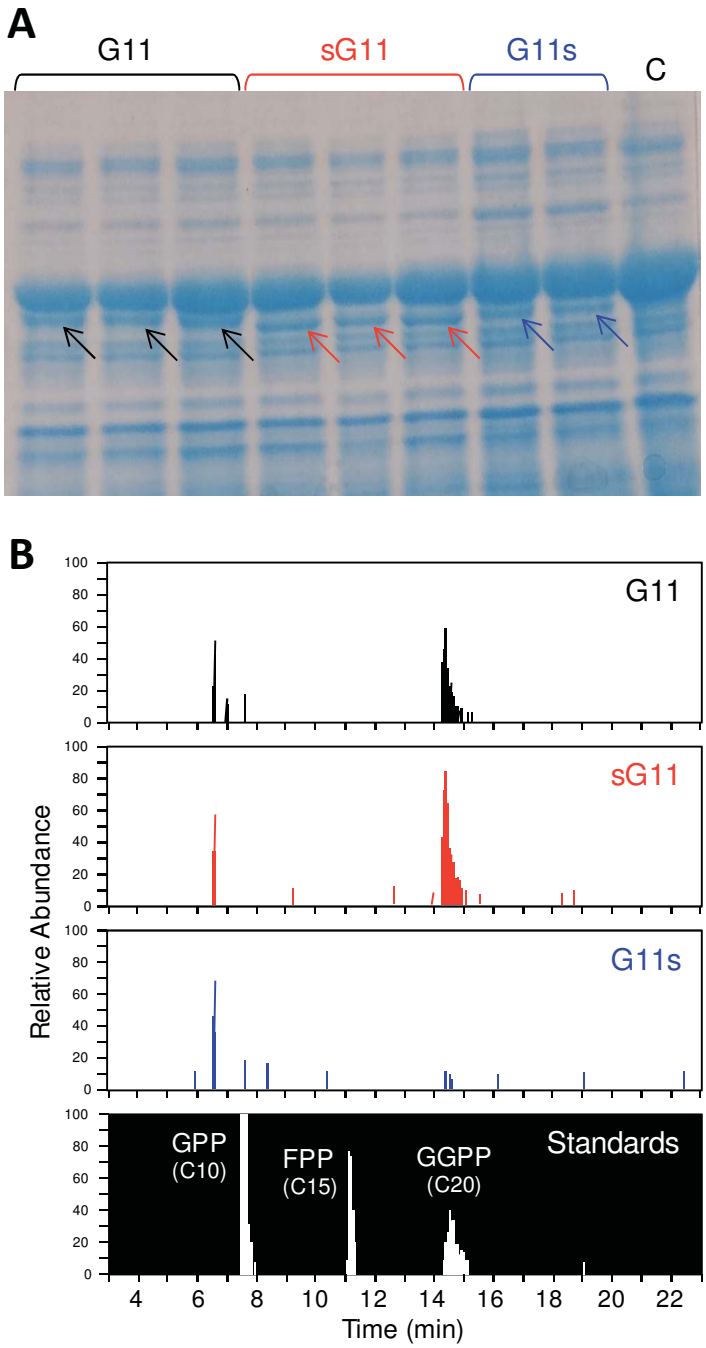


Figure S4. *In vitro* activity of truncated G11 isoforms. (A) SDS-PAGE of protein extracts from *Escherichia coli* BL21 pGROE cells expressing the indicated proteins (see Figure 2A; G11 and G11s correspond to truncated versions lacking the predicted plastid-targeting peptide). C, empty vector control. Arrows mark the position of the recombinant proteins. (B) LC-MS chromatograms showing the results from *in vitro* enzyme activity assays using extracts like those shown in (A). IPP and DMAPP were used as substrates. Detection of prenyl diphosphates was performed using *m/z* 518.254 (GFPP), 449.186 (GGPP), 381.123 (FPP), and 313.061 (GPP). Retention time of available standards is also shown.

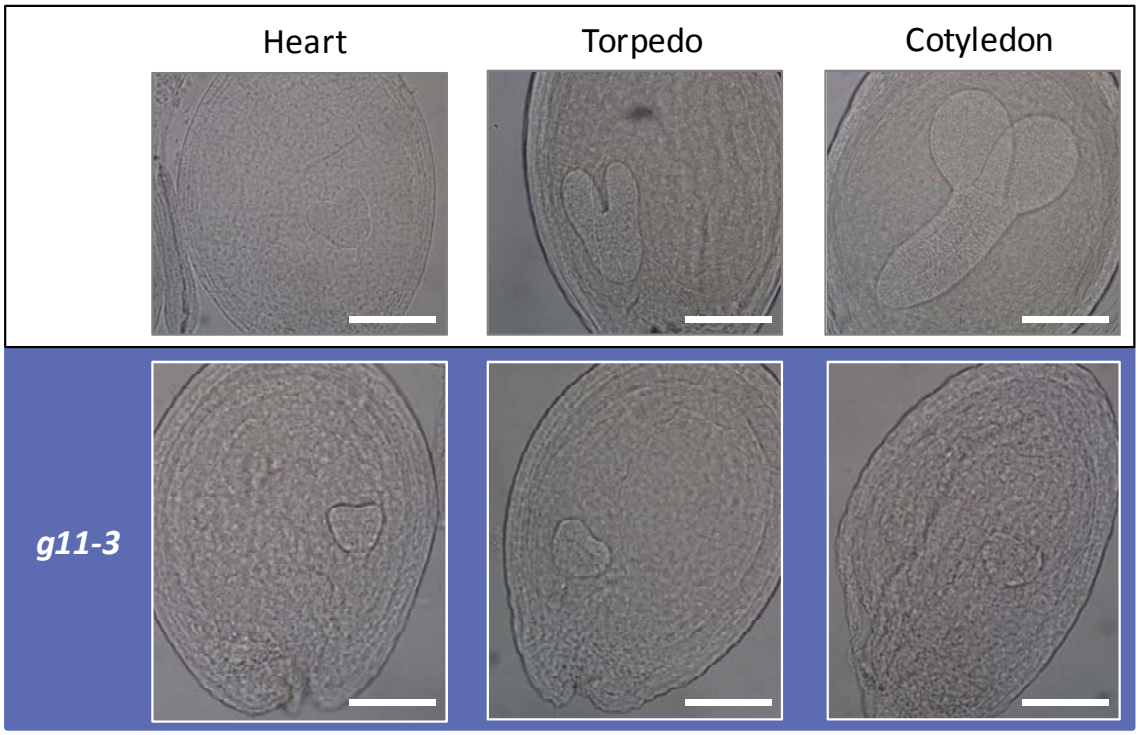


Figure S5. Embryogenesis is blocked at the heart stage in the *g11-3* mutant. Pictures show representative images of embryos in cleared seeds from heterozygous *g11-3* plants. Images in the same column correspond to seeds from the same silique that appeared either green (expected to be either azygous or heterozygous; upper row) or white/brown (expected to correspond to homozygous *g11-3* mutants; lower row, boxed in blue). Different columns correspond to siliques at different positions in the inflorescence (i.e. seeds at different stages of development). Embryo developmental stage in the green seed of the silique is indicated on the top. Bars, 50 µm.

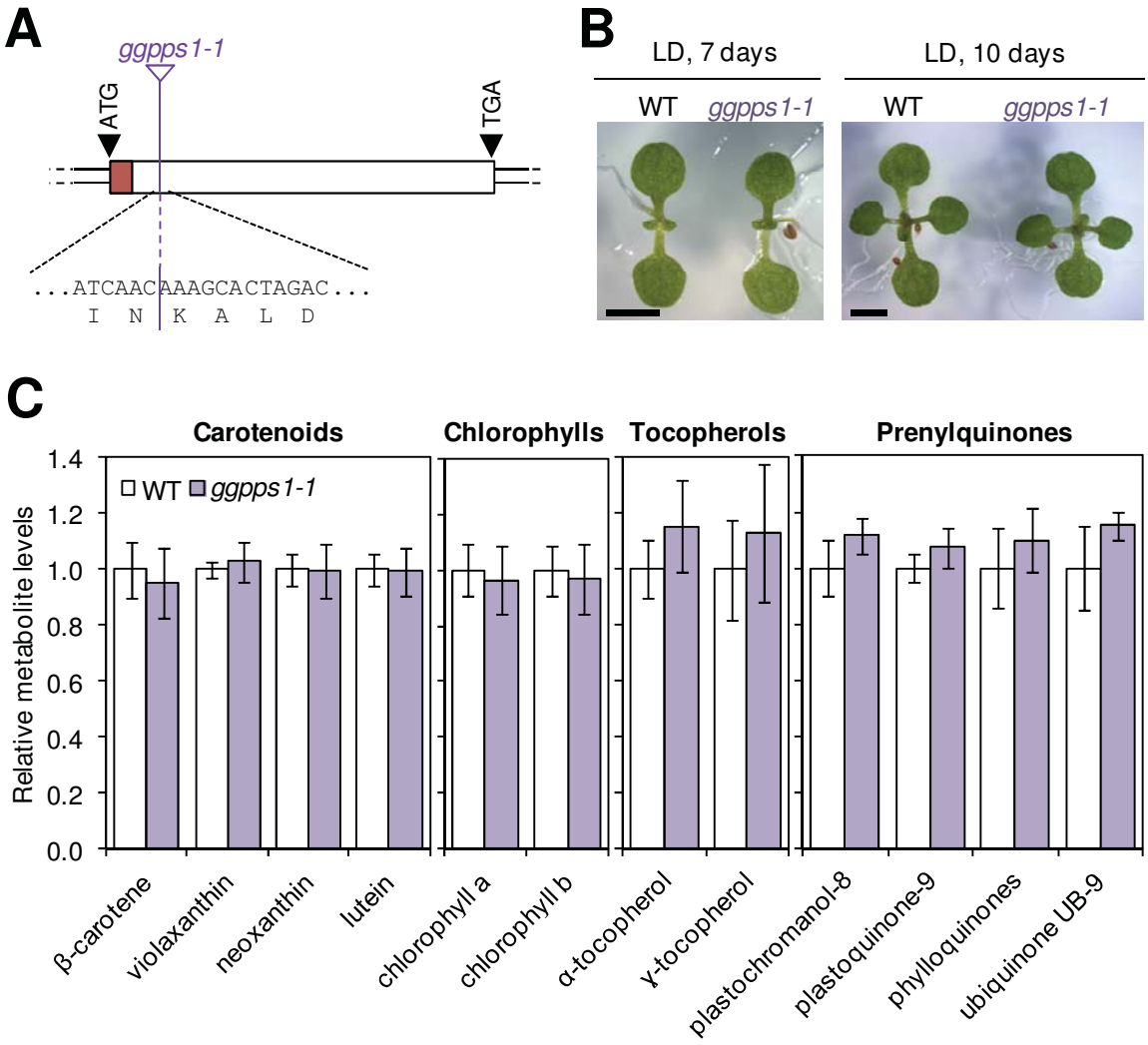


Figure S6. Developmental and metabolic phenotypes of the *ggpps1-1* mutant. (A) Schematic representation of the *GGPPS1* (At1g49530) gene according to TAIR v10 annotation. The protein coding sequence (which lacks introns) is shown as a wider box with a purple section corresponding to the mitochondria-targeting peptide predicted with TargetP. Translation start and stop codons are marked. The site of the T-DNA insertion in the *ggpps1-1* (SAIL_559_G01) mutant is also shown. The exact insertion site (shown with the encoded amino acid residues) was determined by sequencing the region after PCR amplification with primers annealing on the T-DNA and the neighbouring genomic region. (B) A segregating population of the SAIL_559_G01 line was analyzed by PCR to identify individuals that were azygous (WT) and homozygous (*ggpps1-1*) for the T-DNA insertion. Images show representative images of these individuals when grown together under long-day conditions (LD) for 7 and 10 days. Bars, 2 mm. (C) Levels of GGPP-derived isoprenoids in WT and *ggpps1-1* seedlings. Values are shown relative to those found in WT plants and correspond to mean and standard error of n=3 independent samples. No statistically significant differences (*t*-test, $p<0.05$) were found between WT and *ggpps1-1* samples.

SUPPLEMENTAL TABLES

Table S1. Primers used in this work.

Use	#	Name	Sequence (5'-3')
Genotyping and cloning	1	G11-P5F	AGAAGCTTACAAGTTGTTAAATTCG
	2	G11-5F	CAGATTTCAGAAATCGCCATGG
	3	G11-3R	ATTCCCGACAAAAGGAATCG
	4	LBb1	GCGTGGACCGCTTGCTGCAACT
	5	G1-LP-F	AAACTGGACCTGACCACAGC
	6	G1-RP-R	CCTCTGTCCCAACAGCTCTC
	7	SAIL LB3	TAGCATCTGAATTTCATAACCAATCTCGATACAC
	8	G11-B1-F-1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGTCTTCCGTTGTTACAAAAGAAG
	9	G11-B2-R-2	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGTTCTGTCTATAGGCAATG
	10	G11-B1-F-3	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGATGTCGTACATCATCACCAAAGC
	11	G11-B2-R-4	GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAGGACCCTAAAAGCTGATCACGAGC
5'-RACE	12	G11-R-R1	TGCCACCGGCGAGAAGAGAGTAACGC
	13	G11-R-R2	TCTTGAGTGGCTCACGGAGAGGAACAGC

Table S2. Constructs and cloning details.

Use	Construct	Template (1)	Primers (2)	Sequence (3)	Plasmid backbone	Cloning method
Transgenic plants	35S:sG11-GFP	pSG11G	-	-	pCAMBIA1302	*Nco*I / Mung Bean nuclease / T4 ligase
In vitro GGPPS activity assay	pET-G11	pSG11G	8+9	$G11_{169-1116}$	pET32-GW	Gateway
	pET-sG11	pSG11G	9+10	$G11_{229-1116}$	pET32-GW	Gateway
	pET-G11s	pSG11G	8+11	$G11_{169-1050}$	pET32-GW	Gateway
In vivo GGPPS activity assay	pCR-G11	Genomic DNA (WT)	2+3	$G11_{1-1116}$	pCR2.1 TOPO	TOPO TA cloning
	pCR-sG11	Genomic DNA (*g11-2*)	3+4	$G11_{130-1116}$	pCR2.1 TOPO	TOPO TA cloning
	pCR-G11s	Genomic DNA (*g11-3*)	2+4	$G11_{1-1049}$	pCR2.1 TOPO	TOPO TA cloning

(1) Plasmid pSG11G reported in Ruiz-Sola et al. (2016)

(2) See Table S1.

(3) Numbers refer to nucleotide positions in the protein coding sequence (positions 1-3 correspond to $ATG_{(1)}$, 229-231 to $ATG_{(2)}$, and 1114-1116 to the TGA translation stop codon).

ANNEX 3

MANUSCRIPT

Barja MV and Rodríguez-Concepción M (2019). A simple in vitro assay to measure the activity of geranylgeranyl diphosphate synthase and other short-chain prenyltransferases.

NOTE: The PhD candidate designed and performed all the work described here and wrote the manuscript with the supervision of the PhD director. This article has been accepted for publication in the *Plant and Food Carotenoids* issue of the journal *Methods in Molecular Biology*.

A simple *in vitro* assay to measure the activity of geranylgeranyl diphosphate synthase and other short-chain prenyltransferases

M. Victoria Barja and Manuel Rodriguez-Concepcion

Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Barcelona, Spain.

i. Running head

Determination of GGPP synthase activity

ii. Abstract

Most carotenoids are C40 metabolites produced from C20 geranylgeranyl diphosphate (GGPP). The enzymes that produce this precursor, GGPP synthases (GGPPS), are members of the short-chain prenyltransferase (SC-PT) family. SC-PTs are enzymes that catalyze the sequential head-to-tail addition of one or more C5 molecules of isopentenyl diphosphate (IPP) to dimethylallyl diphosphate (DMAPP) with the concomitant release of pyrophosphate (PPi). SC-PTs produce linear isoprenyl diphosphates of up to C20 (GGPP) that serve as precursors for many groups of isoprenoids with a wide range of essential biological functions in Eucarya, Bacteria and Archaea. Enzymatic analysis of SC-PT activity normally requires complex, laborious and expensive methods such as radioactivity-based assays or liquid chromatography-mass spectrometry (LC-MS). Here we describe a fast and inexpensive spectrophotometric protocol for determining the kinetic parameters of SC-PTs in purified enzyme preparations, using an adapted assay for PPi quantification. We developed the method using the *Arabidopsis thaliana* GGPPS11 enzyme, which produces geranylgeranyl diphosphate for the synthesis of carotenoids in the chloroplast.

iii. Key Words

Short-chain prenyltransferases, isoprenyl diphosphate synthases, enzymatic activity, kinetic parameters, EnzCheck kit, geranylgeranyl diphosphate synthase, GGPP, isoprenoids, carotenoids.

1. Introduction

Isoprenoids (also known as terpenoids) are a widely diverse group of natural metabolites, and many of them are essential in all-living organisms. Particularly,

plant isoprenoids show an enormous structural and functional diversity. Plants produce a myriad of specialized isoprenoids involved in their interaction with the environment, but also a number of essential isoprenoids that participate in photosynthesis (chlorophylls, carotenoids, tocopherols, prenylquinones), respiration (ubiquinone) or growth regulation (cytokinins, brassinosteroids, gibberellins, abscisic acid, strigolactones) [1]. All isoprenoids derive from the C5 units isopentenyl diphosphate (IPP) and its double-bond isomer dimethylallyl diphosphate (DMAPP) (Figure 1). Condensation of one or several IPP molecules to one DMAPP unit catalyzed by isoprenyl diphosphate synthases (IDSs), also called prenyl transferases (PTs), produces linear isoprenyl diphosphates of different chain length that represent the first intermediates of nearly all isoprenoid groups. IDS enzymes can be classified as *trans*- or *cis*-PTs depending on the stereochemistry of the double bonds formed in the synthesized product, forming two evolutionary distinct groups of proteins [2,3]. With some exceptions [4–8], most *trans*-PTs generate isoprenyl diphosphates of up to C50, whereas *cis*-PTs synthesize much longer molecules.

Among the *trans*-PTs, short-chain prenyltransferases (SC-PTs) catalyse the sequential head-to-tail addition of up to three molecules of IPP to one molecule of DMAPP producing C10 geranyl diphosphate (GPP), C15 farnesyl diphosphate (FPP) or C20 geranylgeranyl diphosphate (GGPP) [2,3] (Figure 1). This enzymatic reaction requires two highly conserved domains, referred to as First Aspartate-Rich Motif (FARM, $DDx_{2-4}D$) and Second Aspartate-Rich Motif (SARM, DDx_2D), that are involved in substrate binding [9–14] and catalysis using Mg^{2+} as cofactor [15]. The sequential addition of IPP molecules to the isoprenyl diphosphate product releases a pyrophosphate (PPi) molecule per IPP unit added. This process takes place in the elongation pocket of the SC-PT, a hydrophobic cavity that controls the hydrocarbon tail length by the size of the side chain of some amino acid residues located on the N-terminal side of the FARM. In most cases, the fifth amino acid upstream to this motif is the responsible of the isoprenyl product chain length [16–22]. However, other residues could also play a role controlling the number of IPP condensations during product elongation [3].

Despite all the studies highlighting the role of the elongation pocket on final product length, predictions are difficult and highly sensitive analytical methods are still required to determine the actual product of uncharacterized SC-PTs. Radioactive enzymatic assays followed by the hydrolysis of the products require specialized equipment for the measurements such as radio-gas chromatography (radio-GC), radio-high-performance liquid chromatography (radio-HPLC), thin

layer chromatography (TLC), or liquid scintillation counting (LSC) [22–28]. These methods, however, often fail to conclusively demonstrate the chain length (i.e. identity) of the product. Moreover, they are extremely time-consuming when used to calculate the enzymatic kinetics of the tested SC-PT. Nonradioactive methods based on liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) have been developed to accurately determine isoprenyl diphosphate metabolites in a much more precise and faster way [29–31]. However, they are still complex and expensive for the enzymatic characterization of SC-PTs (i.e. for the calculation of their optimal pH or kinetic parameters such as Vmax and Km).

Here we describe an easy and inexpensive spectrophotometric protocol to enzymatically characterize previously identified SC-PTs using purified enzyme preparations. The described method is an adaptation of the commercial *EnzCheck Pyrophosphate Assay Kit (E-6645)*, that allows the detection of the PPi released by a biochemical reaction. The PPi produced in the reaction mix is converted into two molecules of inorganic phosphate (Pi) by an inorganic pyrophosphatase. The Pi then reacts with the substrate 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) by a purine nucleoside phosphorylase (PNP) producing ribose 1-phosphate and 2-amino-6-mercapto-7-methyl-purine, which can be measured at 360 nm, hence allowing the quantification of the PPi released during the enzymatic reaction.

In this article, the method is used to measure the activity of the *Arabidopsis thaliana* GGPPS11 protein (At4g36810), a GGPP synthase (EC 2.5.1.29) that supplies GGPP for the production of carotenoids and other groups of plastidial isoprenoids [32]. This enzyme has been well characterized in previous studies [21,31,33], which served as a reference to validate the results obtained with the reported assay.

2. Materials

1. Reaction buffer (20X): 1 M Tris-HCL, 20 mM $MgCl_2$, pH 7.5 (*see* **Note 1**).
2. Solid Tris Base. Tris(hydroxymethyl)aminomethane.
3. Hydrogen chloride (HCl) 37%
4. Solid magnesium dichloride ($MgCl_2$).
5. 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG): 1 mM in milli-Q

water (store at -20 ºC) (*see* **Note 2**).

6. Purine nucleoside phosphorylase (PNP): 100 U/mL in milli-Q water (store at 4 ºC) (*see* **Note 3**).
7. Inorganic pyrophosphatase: 3 U/mL in 1X Reaction buffer (store at 4 ºC) (*see* **Note 4**).
8. PPi standard: 500 μM in milli-Q water (*see* **Note 5**).
9. IPP and DMAPP (Echelon Biosciences Inc.): 1 mM in milli-Q water (store at -80 ºC) (*see* **Note 6**).
10. Eppendorf Safe-Lock Tubes™ of 1.5 and 2 mL capacity.
11. Sterile polystyrene 96 well-plates (costar®) with low evaporation lid and flat bottom.
12. SpectraMax M3 multi-mode microplate reader (Molecular Devices).
13. GraphPad Prism Software

3. Methods

3.1. Standard GGPPS Activity Assay

All 200 μL enzyme reactions should be carried out in 96-well plates (Figure 2). The empty wells can be reused for subsequent experiments. Always perform three technical replicates per condition and at least two independent experiments.

1. Prepare the reaction mixture (Table 1) by adding 10 μL of 20x reaction buffer (*see* **Note 7**), 40 μL of 1 mM MESG substrate, 2 μL of 100 U/mL PNP and 2 μL of the 3 U/mL inorganic pyrophosphatase (*see* **Note 8**). Add enough water to reach a final total volume of 200 μL minus the volumes of the enzyme, IPP and DMAPP that will be added later (Table 1).
2. Add 5 μg of the purified enzyme (*see* **Note 9** and **Note 10**) to the reaction mixture and mix well by up and down pipetting.
3. Pre-incubate at room temperature for 10 min (*see* **Note 11**).
4. Add the enzyme substrates IPP and DMAPP as shown in Table 1 for a

standard reaction (*see* **Note 12**) and mix well by up and down pipetting. It is very important to always perform a control assay in parallel in which solvent buffer will be added instead of the experimental substrates (*no-substrate control*, see **Note 13**, Figure 2).

5. Immediately introduce the plate in the spectrophotometer, shake it and start measuring the absorbance at 360 nm as a function of time during 1 h at 25ºC. Reading the absorbance every 2 min should be enough to build the activity curve (*see* **Note 14**).
6. When analyzing the data, remember subtracting the *no-substrate control* values from the corresponding experimental samples (*see* **Note 15**).
7. Calculate the enzyme activity obtained for each condition as the increase of PPi production per minute and μg of enzyme (μM PPi min^{-1} μg^{-1}) (see **Note 16**). Use them to build a pH curve (*see* **Note 17**) or to obtain the kinetic parameters of the enzyme (Vmax and Km) from the Michaelis-Menten curve using the GraphPad Prism software (*see* **Note 18**).

Table 1. Reagents and volumes to prepare a standard 200 μL single GGPPS enzyme reaction

Reagent	**Stock concentration**	**Assay concentration**	**Vol. added**
[a]**20X Reaction buffer**	1 M Tris-HCL 20 mM $MgCl_2$ pH 7.5	50 mM Tris-HCL 1 mM $MgCl_2$ pH 7.5	[a]10 μL
MESG	1 mM	0.2 mM	40 μL
PNP	100 U/mL	1 U/mL	2 μL
Inorganic pyrophosphatase	3 U/mL	0.03 U/mL	2 μL
[b]**IPP**	1 mM	90 μM	18 μL
[b]**DMAPP**	1 mM	30 μM	9 μL
[c]**Enzyme**			[c]3-5 μg
Water			Up to 200 μL

[a]Volume of Reaction buffer used depends on the stock solution concentration (*see* **Note 7**)

[b]Volume of IPP and DMAPP will change when performing the assays for the determination of the kinetic parameters (*see* **Note 12**)

[c]Volume of enzyme will depend on the concentration of the purified enzyme

3.2. Standard Curve for PPi Quantification

The linear range for the quantification of PPi using the EnzChek Kit is from 1 μM to 75 μM.

1. Follow the standard reaction (Table 1) adding increasing amounts of the 500 μM PPi standard solution (*see* **Note 19**) and omitting the volumes of the substrates (IPP and DMAPP) and the experimental enzyme.
2. Prepare a *no-PPi control* without PPi (0 μM point, *see* **Note 19**). Also, include a *no-pyrophosphatase control* as a blank with no PPi and no inorganic pyrophasphatase (*see* **Note 11**).
3. After mixing all the reagents, incubate the plate at 25 ºC for 60 min (*see* **Note 20**).
4. Measure the absorbance at 360 nm.
5. Subtract the *no-pyrophosphatase control* absorbance value from each experimental reaction and build the PPi standard curve plotting the absorbance at 360 nm as a function of PPi concentration.

4. Notes

1. The 20X Reaction buffer provided in the *EnzChek* Kit contains 2 mM of sodium azide, that acts as a preservative. The Kit allows to perform the activity assay over a pH range of 6.5 to 8.5, but in this case the buffer must be prepared with the pH of interest (here, sodium azide can be omitted). To generate a pH curve, we recommend to prepare a 20x stock of Reaction buffer at pH 9.5. Then, distribute it in 5 ml aliquots (as many as pH values to be tested). Adjust the pH of every aliquot to the desired value with 37% HCl and then add milli-Q water up to 10 mL to get 10X aliquots of each pH. Remember that the volume of these 10X aliquots added per activity reaction must be doubled (as the recipe in Table 1 contains a 20X Reaction buffer).
2. The Kit includes 6.3 mg (20 μmoles) of MESG. Add 20 mL of milli-Q water directly to the container to prepare a 1 mM stock. Immediately, aliquot the homogenized MESG solution and store the aliquots at -20 ºC. As each individual reaction requires 40 μL of MESG substrate, and normally triplicates will be performed per condition, we recommend to prepare both

200 μL and 500 μL aliquots. Immediately before use, thaw the required aliquot of MESG at 37 ºC (no more than 5 min), vortex strongly and place on ice. If more than one aliquot is needed, we recommend to mix all of them together and use the same solution for all reactions. As described in the manufacturer's instructions, MESG solution is stable on ice at least for 4h at pH 7.5. Be aware that the half-life of this substrate may change in different conditions. It is not recommended to freeze and reuse MESG leftovers.

3. The Kit provides two vials of 50 U of freeze-dried PNP. To prepare a 100 U/mL stock solution add 0.5 mL of milli-Q water to one of the vials. This solution can be stored at 4 ºC for at least one month. The non-reconstituted PNP vial may be stored at -20 ºC.
4. The Kit contains one vial of 6 U of lyophilized inorganic pyrophosphatase. Add 200 μL of milli-Q water to the vial to prepare a stock of 30 U/mL and store it at 4 ºC. Before performing the activity assay, prepare a 3 U/mL aliquot by diluting 10-fold the 30 U/mL stock into 1X Reaction buffer (previously diluted from 20X Reaction buffer). Never dilute the inorganic pyrophosphatase into 20X reaction buffer. Leftover 3 U/mL solution can be stored at 4ºC for one week.
5. The Kit provides 500 μL of 50 mM $Na_4P_2O_7$ (with 2 mM sodium azide) as a source of PPi. Prepare a 500 μM working solution of PPi standard by diluting 100-fold a portion of the given stock in milli-Q water. Preparing 500 μL of working solution should be enough for a regular standard curve (see **Note 19**).
6. IPP and DMAPP are provided by Echelon Biosciences Inc. as Tris-ammonium salts. Prepare 4 mM stock solutions in milli-Q water, dispense in 50 μL aliquots and store at -80 ºC. Prior to use, dilute an aliquot of these substrates in water to prepare 1 mM working solutions. Leftover 1mM solutions can be frozen again and stored at -80 ºC.
7. This volume can change when testing different pH reaction buffers. If 10X buffers are used instead of 20X, 20 μL should be added.
8. Depending on the number of reactions, a mix of the first four reagents can be prepared.
9. The amount of purified enzyme added to the reaction mixture should be determined empirically. It is important that the release of PPi by the enzyme tested does not surpass the activity of the inorganic pyrophophatase. In our case, 3 μg of the *Arabidopsis* GGPPS11 enzyme also worked well in the assays.
10. *Arabidopsis* GGPPS11 was purified using the pET-G11 construct [31], which

harbors a version of the enzyme lacking the plastid targeting peptide and fused to an N-terminal 6x-histidine tag. The recombinant protein was produced in the *E. coli* Rossetta 2 (DE) strain (Novagen, Merck KGaA, Darmstadt, Germany). The cells were grown at 37ºC in 100 mL of 2xYT medium supplemented with the corresponding antibiotics. When they reached an OD_{600} of 0.5-0.8, 1 mM IPTG was added. After the induction, the culture was incubated over night at 18 ºC and bacterial cells were then recovered by centrifugation. The pellet was resuspended in 10 mL of buffer A (100 mM Tris-HCl pH7.5, 100 mM NaCl, 10% glicerol) supplemented with 1 mM DTT, 1 mg/mL lysozyme and one tablet of complete protease inhibitor cocktail (Roche), and incubated in ice for 20 min. After a brief sonication (five pulses of 30 s, 17%), the cell lysate was centrifuged at 12,500 rpm at 4°C for 20 min. Then 2 μg/ml DNase I and 10 μg/ml RNase A were added directly to the tube and gently mixed. The mixture was incubated 20 min in ice and then centrifuged again at 17,500 rpm at 4 ºC for 50 min. The supernatant was filtered using a 0.2 μm filter and incubated for 2 h with 2 mL of nickel-nitrilotriacetic acid (Ni-NTA) beads (Qiagen) previously equilibrated with washing buffer (buffer A + 20 mM imidazole). After the incubation, the mixture was placed into a filter column where the flowthrough is discarded and the Ni-NTA beads are stacked with the recombinant protein attached. To remove the unspecific proteins retained in the column five washes of 1 mL of washing buffer were performed, and finally the protein was recovered in one tube by five elutions with 150 μL of elution buffer (buffer A + 150 mM imidazole). The eluted sample was desalted by Thermo Scientific™ Zeba™ Desalting Columns using buffer A as exchange buffer and the purified protein was quantified, aliquoted with glycerol 40% and finally stored at -20 ºC.

11. Reagents and enzymes may be contaminated with Pi which will interfere with the measurements. The reagents should be tested in a standard reaction (Table 1) with no inorganic pyrophosphatase (*no-pyrophosphatase control*). If present, contaminating Pi should be subtracted out from the experimental reaction measurements. In addition to the reagents, the purified enzyme may also show Pi contamination that could prevent the obtainment of conclusive data. To reduce the contaminating Pi, the enzyme sample can be pre-incubated for 10 min before the addition of the substrates (e.g. IPP and DMAPP). In this pre-incubation step, the reaction catalysed by the PNP can act as a "Pi mop" lowering the contaminating Pi to submicromolar levels. It

is recommended to always carry out this step to avoid any possible Pi interference.

12. The concentration of IPP and DMAPP was empirically calculated to observe an increase of enzymatic activity throughout the time using 3 μg of enzyme. We add 3-fold more IPP than DMAPP since every GGPP molecules is formed from 3 molecules of IPP and only one of DMAPP. The concentration and proportion of the substrates may be determined empirically for other SC-PTs. For the generation of **pH curves** we used the amount of substrates specified in Table 1.

    For **kinetic parameters determination assays**, the concentration of the substrates will change and so will their required volumes. We calculated the basic kinetic parameters (Vmax and Km) for each substrate separatedly, designing the plate to test different concentrations of one of the substrates fixing the other one in 100 μM (Figure 2). The range of concentrations used to calculate the kinetics of the enzyme for each substrate was: 5, 10, 20, 50, 75, 100 and 200 μM.
13. The *no-substrate control* must be done whenever a reagent of the *Standard Reaction* change. For example, when testing the activity of the experimental enzyme in the presence of different pH media, a *no-substrate control* must be included for each pH (column 4 of the Figure 2, light grey). However, for kinetic parameters determination assays, only one *no-substrate control* is needed (well A4 of the Figure 2) since only the volume of the substrates changes.
14. The assay may require to set up the measuring time points to stablish the linearity region of PPi production *versus* time.
15. The *no-substrate control* absorbance value should not increase during the one-hour assay, as an indication that there is not too much Pi or PPi contamination that could be transformed after the 10 min "Pi-mop" pre-incubation. When the absorbance of the *no-substrate control* is stable, the mean of the values obtained in each time point can be calculated and subtracted from the values of the experimental reactions.
16. Use the linear equation from a PPi standard curve (*see* **Section 3.2**) to quantify the PPi produced in each time point and, after plotting it, select the linear range of PPi production to calculate to calculate the enzyme activity (μM PPi min^{-1} μg^{-1}) for each condition.
17. To build the pH curve represent the enzyme activities *versus* the pH. Relative values can be calculated after giving a 100% to the optimal performance.

18. Other softwares may be used to build the Michaelis-Menten curve and calculate the enzyme kinetics. Using Prism, we created an XY table selecting "Enzyme kinetics – Michaelis-Menten" as sample data and added the triplicates of the enzyme activities obtained for each substrate concentration. To obtain the regression curve and the kinetic parameters from the generated graph we performed a "Nonlinear regression" analysis selecting "Enzyme Kinetics – Substrate vs Velocity" and "Michaelis-Menten equation". As a result of the analysis, the regression curve is superimposed on the graph and a table with the values of Vmax and Km together with the statistical parameters is retrieved.
19. A standard curve with 9 different PPi concentrations (0, 5, 10, 20, 30, 40, 50, 60, 75 μM) is enough to obtain the linear regression equation with a square of the Pearson correlation coefficient (R^2) > 0.99. Prepare at least three replicates per PPi concentration to build the standard curve.
20. The manufacturer's instructions indicate an incubation of 30-60 min, enough time to transform all the PPi in the reaction mixture into 2-amino-6-mercapto-7-methylpurine. We incubated the *Standard Curve* 60 min adjusting the time to the *Standard Reaction* for GGPPS activity measurement. Around an hour is needed to (1) synthesize the GGPP and release the PPi molecules and then to (2) consume the PPi by the kit enzymes.

5. References

1. Rodríguez-Concepción, M.; Boronat, A. Breaking new ground in the regulation of the early steps of plant isoprenoid biosynthesis. *Curr. Opin. Plant Biol.* **2015**, *25*, 17–22, doi:10.1016/j.pbi.2015.04.001.
2. Liang, P. H.; Ko, T. P.; Wang, A. H. J. Structure, mechanism and function of prenyltransferases. *Eur. J. Biochem.* **2002**, *269*, 3339–3354, doi:10.1046/j.1432-1033.2002.03014.x.
3. Vandermoten, S.; Haubruge, É.; Cusson, M. New insights into short-chain prenyltransferases: Structural features, evolutionary history and potential for selective inhibition. *Cell. Mol. Life Sci.* **2009**, *66*, 3685–3695, doi:10.1007/s00018-009-0100-9.
4. Ambo, T.; Noike, M.; Kurokawa, H.; Koyama, T. Cloning and functional analysis of novel short-chain cis-prenyltransferases. *Biochem. Biophys. Res. Commun.* **2008**, *375*, 536–540, doi:10.1016/j.bbrc.2008.08.057.
5. Sallaud, C.; Rontein, D.; Onillon, S.; Jabes, F.; Duffe, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A. A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato Solanum habrochaites. *Plant Cell Online* **2009**, *21*, 301–317, doi:10.1105/tpc.107.057885.
6. Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl

diphosphate. *Proc. Natl. Acad. Sci.* **2009**, *106*, 10865–10870, doi:10.1073/pnas.0904113106.
7. Hsieh, F.-L.; Chang, T.-H.; Ko, T.-P.; Wang, A. H.-J. Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. *Plant Physiol.* **2011**, *155*, 1079–90, doi:10.1104/pp.110.168799.
8. Akhtar, T. A.; Matsuba, Y.; Schauvinhold, I.; Yu, G.; Lees, H. A.; Klein, S. E.; Pichersky, E. The tomato cis-prenyltransferase gene family. *Plant J.* **2013**, *73*, 640–652, doi:10.1111/tpj.12063.
9. Marrero, P. F.; Poulter, C. D.; Edwards, P. A. Effects of site-directed mutagenesis of the highly conserved aspartate residues in domain II of farnesyl diphosphate synthase activity. *J. Biol. Chem.* **1992**, *267*, 21873–8.
10. Joly, A.; Edwards, P. A. Effect of site-directed mutagenesis of conserved aspartate and arginine residues upon farnesyl diphosphate synthase activity. *J. Biol. Chem.* **1993**, *268*, 26983–9.
11. Tarshis, L. C.; Yan, M.; Poulter, C. D.; Sacchettini, J. C. Crystal Structure of Recombinant Farnesyl Diphosphate Synthase at 2.6-.ANG. Resolution. *Biochemistry* **1994**, *33*, 10871–10877, doi:10.1021/bi00202a004.
12. Song, L.; Poulter, C. D. Yeast farnesyl-diphosphate synthase: site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 3044–8.
13. Koyama, T.; Tajima, M.; Sano, H.; Doi, T.; Koike-Takeshita, A.; Obata, S.; Nishino, T.; Ogura, K. Identification of significant residues in the substrate binding site of Bacillus stearothermophilus farnesyl diphosphate synthase. *Biochemistry* **1996**, *35*, 9533–8, doi:10.1021/bi960137v.
14. Koyama, T.; Gotoh, Y.; Nishino, T. Intersubunit location of the active site of farnesyl diphosphate synthase: reconstruction of active enzymes by hybrid-type heteromeric dimers of site-directed mutants. *Biochemistry* **2000**, *39*, 463–9.
15. Aaron, J. A.; Christianson, D. W. Trinuclear Metal Clusters in Catalysis by Terpenoid Synthases. *Pure Appl. Chem.* **2010**, *82*, 1585–1597, doi:10.1351/PAC-CON-09-09-37.
16. Ohnuma, S.; Hirooka, K.; Hemmi, H.; Ishida, C.; Ohto, C.; Nishino, T. Conversion of Product Specificity of Archaebacterial Geranylgeranyl-diphosphate Synthase. *J. Biol. Chem.* **1996**, *271*, 18831–18837, doi:10.1074/jbc.271.31.18831.
17. Ohnuma, S. i; Narita, K.; Nakazawa, T.; Ishida, C.; Takeuchi, Y.; Ohto, C.; Nishino, T. A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. *J. Biol. Chem.* **1996**, *271*, 30748–54.
18. Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D. Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl. Acad. Sci.* **1996**, *93*, 15018–15023, doi:10.1073/pnas.93.26.15018.
19. Wang, K.; Ohnuma, S. Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends Biochem. Sci.* **1999**, *24*, 445–51.
20. Stanley Fernandez, S. M.; Kellogg, B. A.; Poulter, C. D. Farnesyl diphosphate synthase. Altering the catalytic site to select for geranyl diphosphate activity. *Biochemistry* **2000**, *39*, 15316–21.
21. Nagel, R.; Bernholz, C.; Vranová, E.; Košuth, J.; Bergau, N.; Ludwig, S.; Wessjohann, L.; Gershenzon, J.; Tissier, A.; Schmidt, A. Arabidopsis thaliana isoprenyl diphosphate synthases produce the C 25 intermediate, geranylfarnesyl diphosphate. *Plant J.* **2015**, *84*, 847–859, doi:10.1111/tpj.13064.
22. Wang, C.; Chen, Q.; Fan, D.; Li, J.; Wang, G.; Zhang, P. Structural Analyses of Short-Chain Prenyltransferases Identify an Evolutionarily Conserved GFPPS Clade in

Brassicaceae Plants. *Mol. Plant* **2016**, *9*, 195–204, doi:10.1016/j.molp.2015.10.010.
23. Cunillera, N.; Arró, M.; Delourme, D.; Karst, F.; Boronat, A.; Ferrer, A. Arabidopsis thaliana Contains two differentially expressed farnesyl-diphosphate synthase genes. *J. Biol. Chem.* **1996**, *271*, 7774–7780, doi:10.1074/jbc.271.13.7774.
24. Masferrer, A.; Arró, M.; Manzano, D.; Schaller, H.; Fernández-Busquets, X.; Moncaleán, P.; Fernández, B.; Cunillera, N.; Boronat, A.; Ferrer, A. Overexpression of Arabidopsis thaliana farnesyl diphosphate synthase (FPS1S) in transgenic Arabidopsis induces a cell death/senescence-like response and reduced cytokinin levels. *Plant J.* **2002**, *30*, 123–32.
25. Chang, T.-H.; Hsieh, F.-L.; Ko, T.-P.; Teng, K.-H.; Liang, P.-H.; Wang, A. H.-J. Structure of a Heterotetrameric Geranyl Pyrophosphate Synthase from Mint (Mentha piperita) Reveals Intersubunit Regulation. *Plant Cell* **2010**, *22*, 454–467, doi:10.1105/tpc.109.071738.
26. Kim, O. T.; Bang, K. H.; Jung, S. J.; Kim, Y. C.; Hyun, D. Y.; Kim, S. H.; Cha, S. W. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. *Biol. Plant.* **2010**, *54*, 47–53, doi:10.1007/s10535-010-0007-1.
27. Schmidt, A.; Wächtler, B.; Temp, U.; Krekling, T.; Séguin, A.; Gershenzon, J. A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in Picea abies. *Plant Physiol.* **2010**, *152*, 639–55, doi:10.1104/pp.109.144691.
28. Arró, M.; Manzano, D.; Ferrer, A. Farnesyl Diphosphate Synthase Assay. *Methods Mol. Biol.* **2014**, *1153*, 41–53, doi:10.1007/978-1-4939-0606-2_4.
29. Henneman, L.; van Cruchten, A. G.; Denis, S. W.; Amolins, M. W.; Placzek, A. T.; Gibbs, R. A.; Kulik, W.; Waterham, H. R. Detection of nonsterol isoprenoids by HPLC-MS/MS. *Anal. Biochem.* **2008**, *383*, 18–24, doi:10.1016/j.ab.2008.08.023.
30. Nagel, R.; Gershenzon, J.; Schmidt, A. Nonradioactive assay for detecting isoprenyl diphosphate synthase activity in crude plant extracts using liquid chromatography coupled with tandem mass spectrometry. *Anal. Biochem.* **2012**, *422*, 33–38, doi:10.1016/j.ab.2011.12.037.
31. Ruiz-Sola, M. Á.; Barja, M. V.; Manzano, D.; Llorente, B.; Schipper, B.; Beekwilder, J.; Rodriguez-Concepcion, M. A single gene encodes two differentially targeted geranylgeranyl diphosphate synthase isoforms. *Plant Physiol.* **2016**, *172*, 1393–1402, doi:10.1104/pp.16.01392.
32. Ruiz-Sola, M. Á.; Coman, D.; Beck, G.; Barja, M. V.; Colinas, M.; Graf, A.; Welsch, R.; Rütimann, P.; Bühlmann, P.; Bigler, L.; Gruissem, W.; Rodríguez-Concepción, M.; Vranová, E. Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. *New Phytol.* **2016**, *209*, 252–264, doi:10.1111/nph.13580.
33. Wang, G.; Dixon, R. a Heterodimeric geranyl(geranyl)diphosphate synthase from hop (Humulus lupulus) and the evolution of monoterpene biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 9914–9919, doi:10.1073/pnas.0904069106.

6. Figures

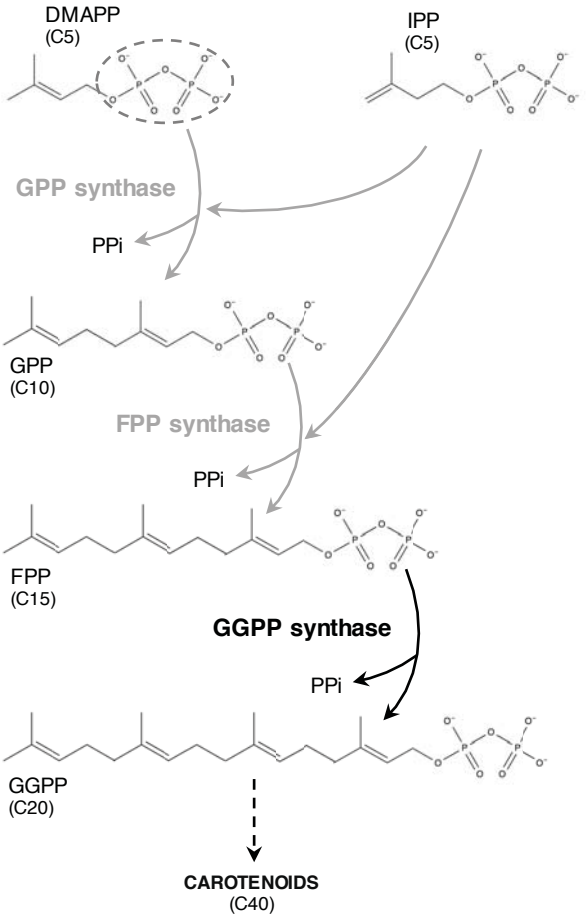


Figure 1. Biosynthesis of short-chain prenyl diphosphates by short-chain prenyltransferases. Each IPP condensation involves the elongation of the prenyl diphosphate molecule with the subsequent release of one PPi molecule (dashed circle). Solid arrows represent one enzymatic step and dashed arrows indicate multiple enzymatic steps. Abbreviations: DMAPP, dimethylallyl diphosphate; IPP, isopentenyl diphosphate; PPi, inorganic pyrophospate; GPP, geranyl diphosphate; FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GPPS, geranyl diphosphate synthase; FPPS, farnesyl diphosphate synthase; GGPP, geranylgeranyl diphosphate synthase.

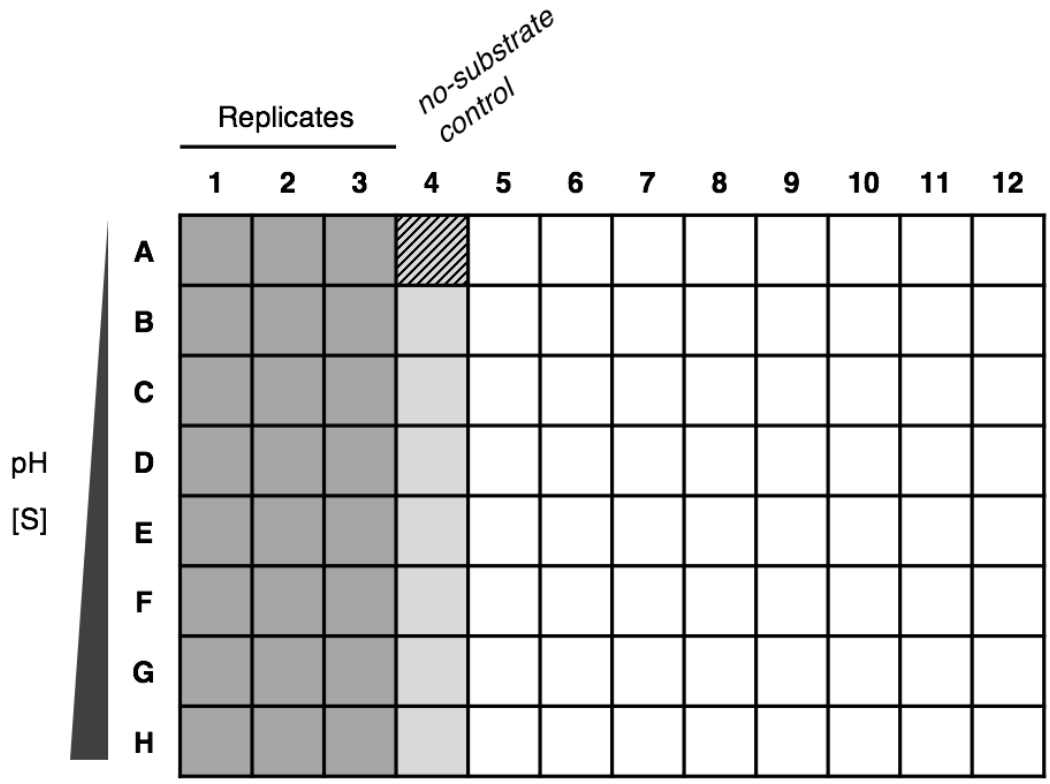


Figure 2. 96-well plate design. Black triangle indicate increasing pH or substrate concentration from top to bottom. Dark grey squares represent standard enzymatic reactions in triplicates per condition. Light grey squares (column 4) indicate *no-substrate control*. Perform a *no-substrate control* whenever a reagent is changed among conditions. If only the concentration of a reagent changes but the reagent solution is the same, only one *no-substrate control* is needed (dashed square).

ANNEX 4

SUPERVISED BSc THESIS

Alcívar JL (2018). Exploiting qPCR for the molecular characterization of transgenic plants.

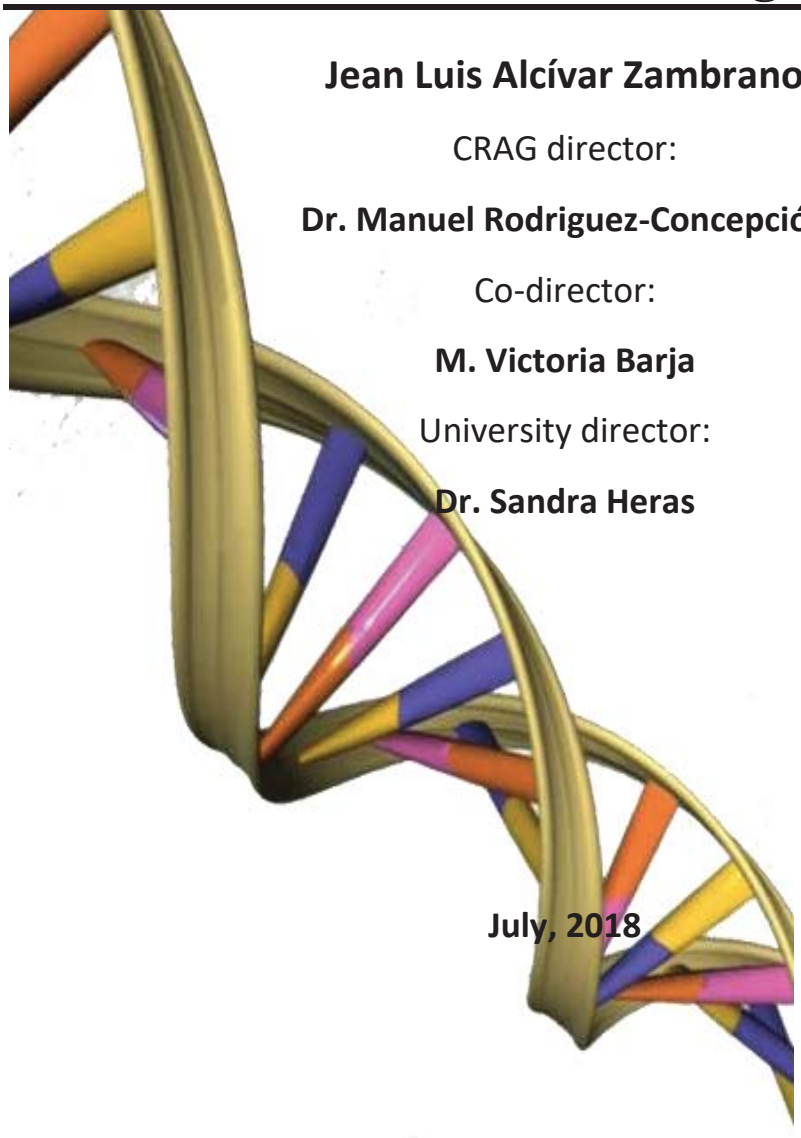
NOTE: This is the BSc thesis (TFG, Treball Final de Grau) presented by Jean Luis Alcívar to obtain his degree in Biotechnology by the Universitat de Girona. The PhD candidate designed the experiments, provided the biological materials, supervised the experimental work of the BSc candidate in the lab, and together with the PhD director co-supervised the analysis of the data and the writing of the TFG document shown here.

Universitat de Girona

CRAG CENTRE DE RECERCA EN AGRIGENÒMICA

Final degree project

Exploiting qPCR for the molecular characterization of transgenic plants



Jean Luis Alcívar Zambrano

CRAG director:

Dr. Manuel Rodriguez-Concepción

Co-director:

M. Victoria Barja

University director:

Dr. Sandra Heras

July, 2018

Abstract

Since the Green Revolution, plant research has been increasingly growing, fostering new methodologies of genetic engineering. One of the best-known procedures in plant biotechnology is the use of *Agrobacterium* strains to generate transgenic organisms. These soil bacteria are capable of transferring and integrating a DNA segment called T-DNA into the genome of an infected plant cell. Notwithstanding the integration is stable, *Agrobacterium* can insert more than one copy of T-DNA per genome. To select for positive insertion events, the engineered T-DNA normally carries a marker gene providing resistance to an herbicide that allows growth of primary transgenic individuals (T1 generation) on herbicide-supplemented medium. Self-crossing of T1 plants generates T2 individuals with a segregating presence of the resistance marker gene that is normally used to deduce T-DNA number (one or more per genome). The analysis of the resistance segregation in the following T3 generation then serves to estimate zygosity (azygosis, hemizygosis, or homozygosis). Hence, it normally takes three generations to get a homozygous line by performing segregation experiments. Selecting the most appropriate line also requires to analyze the expression of the transgene in the selected line.

All this work can be laborious apart from being a lingering process. Therefore, we aimed to develop a quantitative PCR (qPCR) method to estimate T-DNA copy number in the genome of transgenic plants. As a model, we used tomato (*Solanum lycopersicum*), a diploid (2n) species with a huge economic interest. Before the start of this project, tomato plants had been transformed with constructs to overproduce enzymes involved in the biosynthesis of carotenoids, which are health-promoting natural pigments that naturally accumulate in tomato ripe fruits. Our method assigned a value of 0.5 when only 1 T-DNA copy was present in the 2n genome of transformed lines, a value of 1 when 2 copies were present, 1.5 to 3 copies, etc. Identifying T1 plants with a value of 0.5 rapidly allowed to select transgenic lines with a single T-DNA insertion. Homozygous individuals could then be identified in the next (T2) generation as those showing a value of 1. Our method hence reduces the number of T1 lines to propagate and allows to start working with homozygous individuals in the T2 generation. The last part of the project used qPCR to estimate transgene expression levels in selected tomato lines.

Resum

Des de la Revolució Verda, la investigació en plantes ha anat creixent, promovent noves metodologies d'enginyeria genètica. Una de les més conegudes a la biotecnologia vegetal és l'ús de soques d'*Agrobacterium*, el quals són capaços de transferir i integrar un fragments de DNA anomenat T-DNA al genoma de la cèl·lula vegetal infectada. Tot i ésser una integració estable, pot inserir més d'una copia de T-DNA per genoma. Per seleccionar aquests esdeveniments de transformació, normalment el T-DNA porta un gen de resistència a un herbicida que permet el creixement d'individus transgènics primaris (generació T1) en un medi selectiu. L'autocreuament de plantes T1 genera individus T2 que presenten una segregació al marcador de resistència normalment emprat per deduir el número de T-DNA (un o més per genoma). L'anàlisi en la segregació de la resistència a la següent línia T3 serveix per estimar la zigositat (azigosis, hemizigosi, o homozigosis). Conseqüentment, es triguen tres generacions per aconseguir una línia homozigòtica mitjançant experiments de segregació. A més, seleccionar la línia més adient requereix l'anàlisi en l'expressió del transgen a la línia homozigòtica identificada.

Tota aquesta feina pot ésser molt laboriosa apart d'un procés perllongat. Per aquest motiu, el nostre objectiu va ser desenvolupar un mètode basat en PCR quantitativa (qPCR) per estimar el número de còpies del T-DNA al genoma de plantes transgèniques. El model que vam utilitzar va ser *Solanum lycopersicum*, una espècie diploide de gran interès econòmic. Aquestes plantes de tomàquet van ser transformades amb una construcció per la sobreproducció d'enzims involucrats en la biosíntesi de carotenoides, pigments naturals i beneficiosos per la salut. El nostre mètode assignava un valor de 0.5 per aquell T-DNA present en 1 còpia en genomes diploides de les línies transformants, un valor d'1 per dues còpies presents, 1.5 per 3 còpies, etc. La identificació de plantes T1 amb valors de 0.5 ens va permetre seleccionar les línies transgèniques amb una sola inserció de T-DNA. De manera que els individus homozigòtics de la següent generació (T2) es van poder seleccionar gràcies a que mostraven un valor 1. Així mateix, el nostre mètode permet reduir el nombre de línies T1 per reproduir i permet començar a treballar amb individus homozigòtics a la generació T2. En l´última part del projecte es va utilitzar la qPCR per estimar els nivells d'expressió de les línies seleccionades.

Resumen

Desde la Revolución Verde, la investigación en plantas ha ido creciendo, promoviendo nuevas metodologías de ingeniería genética. Una de las más conocidas en biotecnología vegetal es el uso de cepas de *Agrobacterium,* bacterias capaces de transferir e integrar fragmentos de DNA llamados T-DNA en el genoma de la célula vegetal infectada. Aunque la integración es estable, *Agrobacterium* puede insertar más de una copia de T-DNA por genoma. Para la selección de los eventos de transformación positivos, normalmente el T-DNA contiene un gen de resistencia a un herbicida que permite el crecimiento de individuos transgénicos primarios (generación T1) en medio selectivo. El autocruzamiento de las plantas T1 genera los individuos T2 que presentan una segregación del gen de resistencia que normalmente se usa para deducir el número de T-DNAs (uno o más por genoma). El análisis de la segregación de la resistencia en la siguiente generación T3 sirve para estimar la cigosidad (acigosis, hemicigosis, o homocigosis). Consecuentemente, se tardan tres generaciones en conseguir una línea homocigótica mediante experimentos de segregación. Finalmente, se analizan los niveles de expresión del transgén en la línea homocigota identificada.

Todo este trabajo puede ser laborioso además de prolongado. Por ende, nuestro objetivo era desarrollar un método de PCR cuantitativa (qPCR) para estimar el número de copias de T-DNA en plantas transgénicas. Como modelo, usamos el tomate (*Solanum lycopersicum)*, una especia diploide (2n) de gran interés económico. Las plantas de tomate habían sido transformadas con construcciones para sobreproducir unas enzimas involucradas en la biosíntesis de carotenoides, pigmentos naturales con beneficios para la salud. Nuestro método consistía en asignar un valor 0.5 cuando solo había 1 copia de T-DNA presente en un genoma 2n de líneas transformantes, un valor de 1 cuando había 2 copias, 1.5 para 3 copias, etc. La identificación de individuos con 0.5 en plantas T1 nos permitió seleccionar líneas transgénicas con una sola inserción de T-DNA. De esta manera, los individuos homocigóticos se pudieron identificar en la siguiente generación T2 como aquellos que mostraban un valor de 1. Así mismo, nuestro método reduce el número de líneas T1 a propagar y permite empezar a trabajar con individuos homocigóticos en la generación T2. En la última parte del proyecto se usó la qPCR para estimar los niveles de expresión en las líneas de transgénicas seleccionadas.

Contents

Introduction 1

Agrobacterium: from tumor to tool 1

Transformation by means of *Agrobacterium tumefaciens* 2

The classical method for transgenic plant selection 3

Beyond qPCR technology 4

Carotenoids: a practical case 6

Objectives **7**

Materials and methods **7**

Plant material and growth conditions 7

Sample collection and treatment 7

Total genomic DNA extraction from leaf tissue 7

Quantification of nucleic acids 8

Genotyping PCR 8

Genotyping qPCR 9

Transcript levels analysis 10

Method validation experiment 11

Sustainability and ethical criteria 12

Results 12

Determination of T-DNA number copy in transgenic plants by qPCR 12

T1 generation 12

T2 generation 15

T3 generation 16

Validation of the method 18

Determination of transgene expression levels in selected transgenic plants 20

T1 generation 20

T3 generation 22

Discussion 24

Determination of T-DNA number copy 24

Validation of the method 26

Expression levels 26

Conclusions 27

Acknowledgment 28

Bibliography 28

Introduction

Agrobacterium: from tumor to tool

Agrobacterium was reported for the first time in 1897 by Fridiano Cavara, who described galls formed at the base of grapevines (Kado, 2014). This Gram-negative soil bacterium has the conspicuous ability of introducing foreign genes into plant cells. Thanks to this property, today, is one of the most powerful molecular tools for genetic engineering in plants, giving us the opportunity not only to introduce exogenous genes but also to generate loss-of-function mutants.

All members of the *Agrobacterium* genus have been described to cause cortical hypertrophy, *Agrobacterium rhizogenes, Agrobacterium tumefaciens* and *Agrobacterium rubi* can induce an abnormal root growth. The only species that doesn't cause tumor-like growth is *Agrobacterium radiobacter* (Sawada et al., 1993). This genus has a wide range of tumorigenic host plant species to infect, both dicotyledonous plants and gymnosperms (Gelvin, 2012). Despite the existence of several species in the wild, the most studied has been *A. tumefaciens,* becoming the most used for genetic engineering in plants.

The detailed infection mechanism of *A. tumefaciens* is not fully understood yet, but the main aspects of this process are fairly well-known. This bacterium possesses a large plasmid called Ti (tumor inducing)-plasmid that contains two important regions: *vir* (virulence) genes and T(transferred)-DNA. The *vir* genes comprise six groups of operons (*virA*, *virB*, *virC*, *virD*, *virE*, *virG)* that are essential for the transfer of T-DNA to plant cells. The T-DNA, i.e. the DNA region that is transferred from the bacteria to the eukaryotic plant cell, harbors gene sequences that encode enzymes for the production of opines (amino acid derivatives used by *A. tumefacies* as carbon and energy source) and hormones (such as auxins and cytokines) responsible for the uncontrolled growth that generates the typical

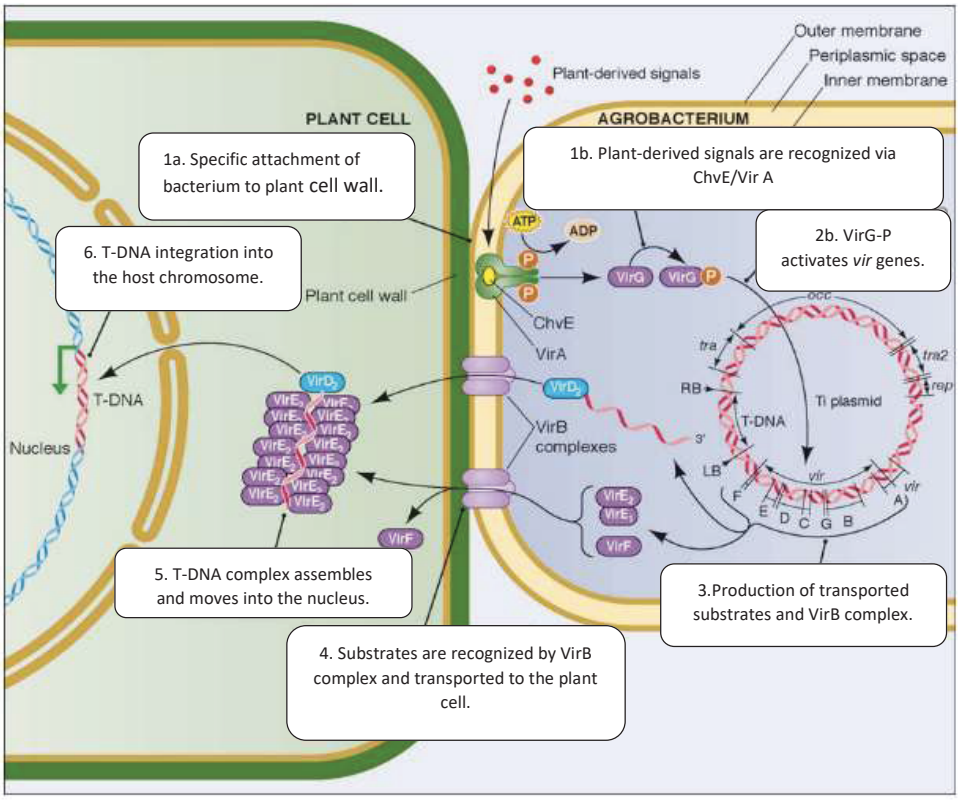


Figure 1. General model of *Agrobacterium*-mediated transformation. Adapted from McCullen & Binns, 2006.

crown gall tumors associated with the bacterial infection. The T-DNA is flanked by short sequence repeats on each end. Transfer starts at the right border (RB) and ends at the left border (LB), but sometimes the insertion is not complete and the resulting genome insertion lacks the LB sequence (**Figure 1**) (Nester, 2015).

The infection mechanism starts when a plant cell is wounded. As a consequence, three signals are released: phenolic compounds, a variety of monosaccharides coming from plant cell walls and acidic conditions required for several steps in the induction process. These signals induce the expression of the *vir* genes which will process and transport the T-DNA from *Agrobacterium* to the wounded cell cytoplasm. The T-DNA will finally enter into the nucleus as a single-strand molecule, and by illegitimate recombination it will be integrated randomly in the genome of the infected plant cell (Gelvin, 2000; Nester, 2015).

This process, schematically represented in Figure 1, can occur more than once. As a result, a mutant plant cell may end up having one or several copies of the T-DNA inserted in different locations of its genome (Nester, 2015). In some cases, several copies of the T-DNAs are inserted in tandem in the same position of the genome.

Transformation by means of *Agrobacterium tumefaciens*

The current way of using *A. tumefaciens* for biotechnological purposes is by replacing the genes located in the T-DNA region (responsible for the tumor formation) with the gene of interest (GOI) combined with a selectable marker. Plant transformation is normally performed using two separate plasmids as a binary system. One of them (known as the binary plasmid) harbors the engineered T-DNA and the other one (known as helper plasmid) contains the *vir* genes (Hoekema et al., 1983).

Standard procedures for *Agrobacterium*-mediated transformation at plant biotechnology labs worldwide are basically two. For the model plant *Arabidopsis thaliana* the most common method is floral dip. This is a fast an efficient method based on dipping developing flowers into a solution carrying the *Agrobacterium* strain of interest (Zhang et al., 2006). In this case, only female gametes are transformed. After pollination, the resulting seeds are used to select for primary (T1) transformants based on the herbicide resistance provided by the marker gene in the T-DNA. For other species such as tomato (*Solanum lycopersicum*), the transformation procedure involves tissue culture. Plant tissue (e.g. cut cotyledons) is incubated in a solution containing *Agrobacterium* and then placed into *callus* induction medium supplemented with selection herbicide to only allow the growth of cells that have incorporated the T-DNA in their genome. The resulting transgenic (T1) *calli* are subsequently transferred to media containing hormone mixtures that allow the development of shoots and roots. The generated T1 plantlets are eventually transferred to soil for normal growth in the greenhouse (Dan et al., 2006).

Since *Agrobacterium* inserts the T-DNA in an illegitimate (random) location in the genome, it is important to work with more than just one line. Differences in transgene expression levels are expected depending on the transcriptional activity of the genomic region where the T-DNA is inserted. Also, insertion of the T-DNA disrupting or affecting an endogenous gene can have unanticipated effects for the plant, jeopardizing the viability of the

cell or giving a phenotype that could mask the function of the GOI. Importantly, transgenic lines with a single T-DNA insertion are preferred because (1) the described insertional effects are less likely, (2) their genetic analysis is easier, and (3) the generation of homozygous lines is faster.

The classical method for transgenic plant selection

As it is explained above, T1 plants come from *callus* (or transformed seeds in the case of *Arabidopsis*). The fact that T1 plants grow in selective medium means that they carry the transgene, more specifically in hemizygosis (as it is almost impossible that 2 copies are inserted in the same location of the maternal and the paternal chromosome generating homozygous plants). When T1 plants are in adult stage they are self-pollinated giving rise to many seeds. The classical method to select transgenic lines with a single T-DNA copy consists on sowing seeds from T1 lines in a selective medium, referring them as T2 plants (**Figure 2**). In case of a single T-DNA insertion in a T1 plant, the transgene is expected to segregate in a Mendelian way in the T2 offspring (25% azygous plants –without the transgene-, 50% hemizygous and 25% homozygous for the transgene). Therefore, it is also expected that 75% (3:1 proportion) of the T2 seedlings will grow in selective medium, since they harbor the T-DNA insertion with the GOI and the selectable marker (conferring the herbicide resistance). The remaining 25% plants will die due to the lack of the transgene under selective conditions (Figure 2). Usually, T2 plants are also used to assess transgene expression so only those lines with an active transgene will be used for subsequent steps. Thus, several herbicide-resistant T2 plants from lines expressing the transgene from a single T-DNA

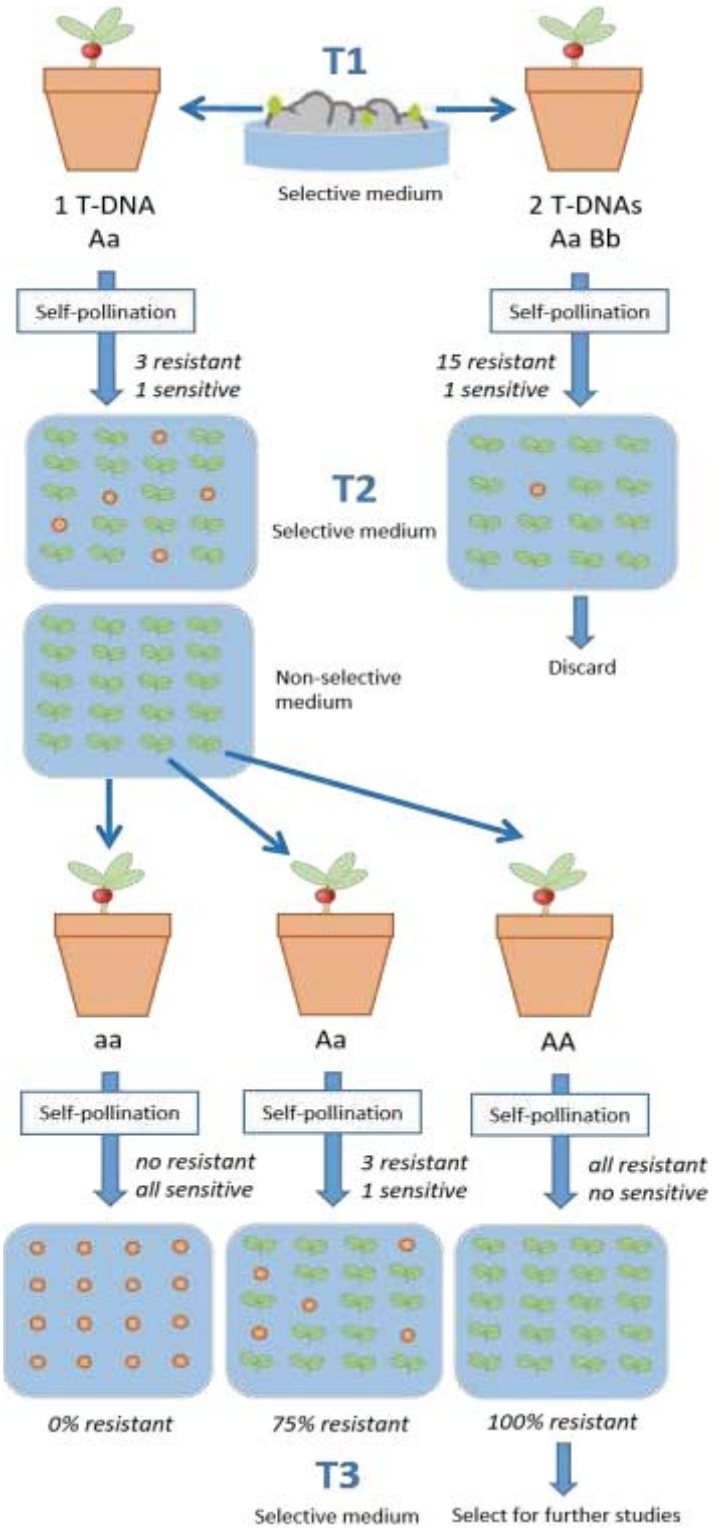


Figure 2. Conventional method of selection for homozygous transgenic plants.

locus (based on the segregation of the resistance marker) are typically transferred to soil and grown to obtain the T3 generation (Figure 2). The seeds of each T2 resistant plant are collected and subsequently sown again in selective medium, where 2 situations can occur: (1) if the T2 resistant plant is hemizygous, its T3 offspring will segregate again in a 3:1 (resistant:sensitive) proportion; (2) if the T2 plant is homozygous (expected in 1/3 of the resistant T2 plants), all of its offspring will grow in selective medium. Since we want the transgene in homozygosis, T3 plants derived from the homozygous T2 plant will be selected and propagated for further experiments.

Selection of transgenic homozygous lines using this method, based on analyzing the segregation of the herbicide resistance provided by the marker gene in the T-DNA, works reasonably well for Arabidopsis but it involves a huge amount of work and plant growth facility (e.g. greenhouse) space for other plant species, including tomato. First, because this selection process must be done for each T1 plant and for many T2 resistant plants coming from each T1 line. Second, because herbicide selection is difficult to implement at large scale in plant species with seedlings that do not develop well on plates. And third, because as a consequence of the number of lines and the life cycle of the plant species, obtaining the final homozygous lines can take several months or even years in case of plant with long reproductive cycles. Due to these inconveniences, faster and more precise methods must be used or developed to reduce the time and costs needed to identify the final transgenic lines of interest. Taking this into account, this project arose to set up a new method in the lab for the selection of tomato transgenic lines with a single T-DNA insertion. This method is based on a technology routinely used in the lab for the analysis of transgene expression: quantitative Polymerase Chain Reaction (qPCR).

Beyond classic qPCR

The qPCR (quantitative PCR) technology is based on PCR (polymerase chain reaction) that allows to determine the exact amount of amplified DNA by labelling it with a fluorescent dye. For this project SYBER Green I Dye was used, this dye is capable of intercalating between dsDNA formed during the PCR reaction leading to an increase of fluorescence signal compared to the unbound dye. After elongation, all the replicon is amplified and the maximum amount of SYBR Green is intercalated, and finally the fluorescence is measured at 530 nm (Bustin, 2000).

The signal coming from the intercalated SYBER Green allows to monitor the amount of dsDNA amplified during every PCR cycle. When the signal reaches a predetermined fluorescent threshold, it is detected by the thermal cycler. The cycle number when this occurs is called C_t (cycle of threshold) or C_p (crossing point). In this way, the lower abundant is the target the more cycles will require to be detected and the higher C_t value will be obtained (Bustin, 2000).

Besides template abundance (C_t), qPCR also informs about the specific melting curve of each amplicon. Each DNA fragment has a characteristic melting temperature (Tm) and it depends on the size and nucleotide composition

of the fragment. The melting curves allow to identify the desired amplified products and to distinguish them from unspecific amplifications or primer dimer events. These melting peaks are analogous to the bands on an electrophoresis gel (Nolan et al., 2006).

Real-time qPCR is commonly used to quantify gene expression, procedure known as RT-qPCR (reverse transcription quantitative polymerase chain reaction). This procedure consists of a reversed transcription of mRNA into cDNA, followed by cDNA amplification by PCR, and finally the detection and quantification of amplified cDNA in real time. (Livak & Schmittgen, 2001).

Although qPCR is most frequently used for quantification of gene expression levels (i.e. mRNA or transcript abundance), in this project we propose another method of using qPCR, as a main objective of genotyping transgenic plants.

First, we extracted gDNA and performed PCR in order to check the absence or the presence of at least one T-DNA. Since PCR doesn't specify the number of copies of T-DNA, we next performed genotyping qPCR on plants confirmed to contain the T-DNA. For genotyping qPCR, we used *LAT52* as the reference endogenous gene since it was reported to be specific for tomato plant (Yang et al., 2005). *LAT52* is a single copy gene in homozygosis, which is what we need to perform our method. The gene associated with the T-DNA selected for qPCR was *nptII* but we could have also targeted the T-DNA gene encoding GFP, since this is another gene that is not present in the plant and therefore will not amplify in non-transgenic tomato lines.

In order to determine the T-DNA copy number from each line we used the formula as follows:

$$\frac{E1^{Ct(RG)}}{E2^{Ct(GOI)}}$$

where

***E1= Primer efficiency of reference gene**

***E2= Primer efficiency of T-DNA gene**

Ct(RG)= Ct value of reference gene

Ct(GOI)= Ct value of T-DNA gene

*E-method was used for primer efficiency whose arithmetic formula is given by: $E=10^{-1/slope}$.

Taking into account the ploidy of the organism we are working with, in our case a diploid plant (2n), the reference gene (*LAT52*) is present in 2 copies in the genome or 1 per each set of chromosomes. Hence, we can obtain different values that represent different T-DNA copy number; **0.5 value for 1 copy, 1 value for 2 copies, 1.5 value for three copies and so on** (Ingham et al., 2001).

Since the insertion of T-DNA is illegitimate, the likelihood that different T-DNA got integrated in homozygosis is remote. Thus, when more than one T-DNA is integrated in the T1 plant genome we assume that each of them will be in hemizygosis (i.e. present in just one set of chromosomes) in that generation.

Carotenoids: a practical case

The research group where this project was carried out is focused on studying the biosynthesis of carotenoids and its regulation. Carotenoids are isoprenoid molecules that function as natural pigments in the range from yellow to red. Carotenoids can be synthesized by all photosynthetic organisms and some non-photosynthetic bacteria and fungi. These metabolites have a great impact on human health. They protect against chronic and degenerative diseases and they also have anticancer and anti-inflammatory effects. However, their most important function is that they are essential precursors of retinoids (including vitamin A). Thanks to their properties, carotenoids are highly demanded as health promoting nutrients and food and feed additives. (Jaswir et al., 2012). In order to reduce their chemical synthesis in the industry and sustainably produce them using natural platforms, such as plants, we first need to decipher the regulation of the biosynthetic pathway.

Carotenoids have important functions in plants. Besides acting as pigments in many flowers and fruits, they are essential for photoprotection of green tissues (against the excess of light) and for growth regulation (being the precursors of important phytohormones such as abscisic acid or strigolactones) (Ruiz-Sola & Rodríguez-Concepción, 2012).

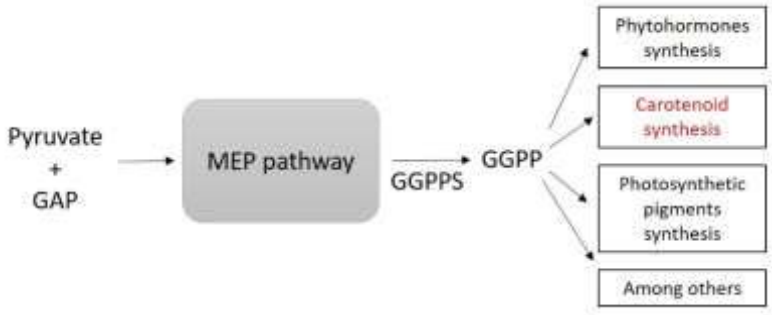


Figure 3. A global view of GGPPS role in plants.

GAP, glyceraldehid 3-phosphate; MEP, 2-C-methyl-D-erythriol 4-phosphate; GGPP, geranylgeranyl diphosphate; GGPPS, GGPP synthase.

Tomato fruits accumulate high amounts of carotenoids such as lycopene (the red pigment of ripe tomatoes) and beta-carotene (the main precursor of vitamin A). Because of this, tomato has become a model species for the study of carotenoid biosynthesis. Carotenoids are synthesized in plastids from geranylgeranyl diphosphate (GGPP). GGPP is also the precursor for many other isoprenoid metabolites in different cell compartments and it is produced by a family of enzymes called GGPP synthases (Rodriguez-Concepcion et al., 2018) (**Figure 3**). Previous work in the lab demonstrated that tomato has 3 GGPP synthases in plastids (G1, G2 and G3), where carotenoids are produced. To better elucidate the physiological role of these enzymes for carotenoid production, transgenic plants were generated to express versions of G1, G2 and G3 fused to the Green Fluorescent Protein (GFP) under the control of a constitutive and strong promoter (*35S::G1-GFP, 35S::G2-GFP and 35S::G3-GFP*). Tomato lines overexpressing the GFP reporter fused to an N-terminal plastid targeting peptide (*35S::pGFP*) were also generated to serve as a control for possible interferences of the GFP in plastids.

To analyze whether increasing the levels of G1, G2 or G3 has an impact on the carotenoid content and the nutritional quality of transgenic tomato fruits, we first needed to select individual lines with a single T-DNA insertion per haploid (2n) genome and take them to homozygosis. In this context, we decided to develop the current project to reduce the time and costs needed for this aim, moving on to the molecular characterization of transgenic plants and leaving behind the classical selection methods.

Objectives

The generation of transgenic plants is an incredibly useful tool for metabolic engineering and molecular biology research. In order to make the genetic characterization of transgenic lines more precise, faster and cheaper, this project aims to accomplish the following objectives.

- Develop a genomic qPCR technique for the determination of the T-DNA copy number in the genome of transgenic tomato plants.
- Determine the transgene expression level in selected tomato lines.

Material and Methods

Plant material and growth conditions

Tomato (*Solanum lycopersicum* var. MicroTom) plants were grown under standard greenhouse long day (LD) conditions, (16 h light at 24-28 ºC, 8 h dark at 22-24 ºC, and at 50-70% of relative humidity). WT and transgenic tomato plants overexpressing *SlG1* (*p35S::SlG1-GFP*), *SlG2* (*p35S::SlG2-GFP*), *SlG3* (*p35S::SlG3-GFP*) and *pGFP* (*p35S::AtpHDS-GFP*)were used in this project. All the constructions also harbored the *ntp II* for kanamycin resistance.

Sample collection and treatment

Leaf was the tomato tissue chosen for the molecular characterization of the transgenic plants. The leaf samples must be collected in the same life stage and localized in the upper part of the plant, so that obtain homogeneous data further on. Leaves were cut and placed into 2 mL Eppendorf tubes. Immediately, the samples were frozen using liquid N_2 and lyophilized using the Freeze Dryer ALPHA 2-4 LD PLUS (Martin Christ®) during at least 24h. After lyophilizing, the samples were stored at -80ºC until next treatment.

Total genomic DNA extraction from leaf tissue

Once the samples were lyophilized, the DNA extraction was performed using CTAB protocol. The lyophilized leaf tissue is grinded to a fine powder with the TissueLysser II (Qiagen®), during 1 minute at 30 s^{-1}. Then, 600 µl of ice-cold extraction buffer (50 mM Tris·HCl pH 8.0 and 20 mM EDTA pH 8.0) is added to the tissue and it is vortexed for several seconds until there is no powder on the lid. 80 µl of 10% SDS were added, it was vortexed for 4 minutes and the mix was incubated at room temperature for 15 minutes. Then, the mix is vortex for 2 min and

then 180 µl of 3M NaAc pH 5.2 were added and it is incubated in ice for 30 minutes. Afterwards, the tissue-debris is spun down at 10K rpm for 15 minutes at 4 °C. Then, the supernatant is transferred to a new tube of 2 ml (700 or 650 µl). An equal volume of isopropanol is added and is mixed by inverting the Eppendorf at least 3 times, and incubated in ice for 0.5 h. (or more at -20 °C). After that, gDNA was spun down at 10K rpm for 10 minutes, the supernatant was poured into waste using a pippete. The pellet is then resuspended in 375 µl of 10 mM Tris·HCl pH 8.0, vortexed for several seconds. An equal volume of CTAB buffer (2% CTAB, 2M NaCl, 0.2 M Tris-HCl pH 8.0, 0,05M EDTA) was added and incubated for 15 min at 65 °C. In the fume hood, an equal volume of chloroform was added, it is mixed by inverting the Eppendorf, and spun for 5 min at max speed (13K rpm). Afterwards, the aquatic phase is rescued and the DNA is precipitated with equal volume of isopropanol, it is incubated at least 1.5 h. at -20 °C or overnight at same temperature. Then mixture is spun down at 10K rpm for 10 minutes, and the supernatant is carefully removed. Afterwards the Eppenderf is left drying inverted on top of paper with the lid open, and incubated at 37 °C until is completely dried. Finally, the pellet is suspended with 100 µl of H_2Odd.

Quantification of nucleic acids

The quantity and quality (**Table 1**) of isolated DNA and RNA was measured with the NanoDrop® ND-8000 spectrophotometer (Thermo Scientific™), using the same solvent as blanking reagent.

Table 1. The ideal values for each ratio and its meaning in terms of sample purity.

Ratio	Determines	Ideal value
260/280	Presence of proteins, phenol or other contaminants	1.8 (DNA), 2.0 (RNA)
260/230	Presence of EDTA, phenol or other organic solvents.	2.0-2.2 (both)

Genotyping PCR

Genotyping PCR (Polymerase Chain Reaction) was performed after genomic DNA extraction in order ensure whether the transgene was integrated. In addition to the specific transgene, *ACTIN4* was also amplified for all samples as an internal control of the extraction. The primer pairs used for genotyping are showed in **Table 2**.

Table 2. Sequence of primers used for genotyping PCR.

Amplified product	Primer pair	Sequence (5'→3')
G1-GFP	SlG1-qPCR-F	GGCCTTTGAACATGTGGCTACC
	eGFP_R_qPCR2	TCTCGTTGGGGTCTTTGCTC
G2-GFP	SlG2-qPCR-F	AAAGTCATCGTCGGAGCTCG
	eGFP_R_qPCR2	TCTCGTTGGGGTCTTTGCTC
G3-GFP	SlG3-qPCR-F	AGGAGGTGCACCAGATGAAG
	eGFP_R_qPCR2	TCTCGTTGGGGTCTTTGCTC
p-GFP	UnivF-attB1short F	GGGGACAAGTTTGTACAAAAAAGCAGGCT
	eGFP_R_qPCR2	TCTCGTTGGGGTCTTTGCTC
Actin	Actin-genot-F	GTGAAAAGATGACCCAGATTATG
	Actin-genot-R	CACGCTCGGTCAGGATCTTCATC

The PCR was performed using the NZY Taq 2× Green Master Mix, which contains: Taq DNA Polymerase, dNTP's, reaction buffer and $MgCl_2$. The reaction mix is specified in **Table 3** and the thermal cycler program is described in **Table 4**.

Table 3. Volumes used in genotyping PCR.

Reagent	Volume
NZYTaq 2× Green Master Mix	5 µL
Forward primer	0.5 µL
Reverse primer	0.5 µL
Nuclease free-water	3 µL
Template DNA	1 µL
Total volume	10

Table 4. PCR program performed for genotyping PCR. Tm and annealing temperature are specific for each primer.

Step	Temperature	Time	Cycles
Initial Denaturation	95ºC	5 min.	1
Denaturation	95ºC	1 min.	30
Annealing	T_m	30 sec.	
Elongation	72ºC	1Kb/min.	
Final elongation	72 ºC	5 min.	1
Cooling	16 ºC	∞	-

Genotyping qPCR

The number of copies and the zygosity of the T-DNA inserted in the transgenic lines was evaluated using total genomic DNA as template. The abundance of the transgene was determined by quantitative PCR on a LightCycler® 480 II real-time PCR system (Roche) using LightCycler® 480 SYBR Green I Master Mix 2x (Roche) which contains: FastStart Taq DNA polymerase, reaction buffer, dNTPs (with dUTP instead of dTTP), SYBER Green I dye and $MgCl_2$. The other reagents added to the mix are specified in **Table 5**, and the qPCR conditions in **Table 6**. Three technical replicates of each biological replicate were performed. Normalized number of copies were calculated as it is described as described in introduction , measuring the abundance of the kanamycin resistance gene (*ntpII*) present in the T-DNA and using *LAT52* (Gene ID: 101261755) as a endogenous reference gene (Yang et al., 2005). Further details of these primers are shown in **Table 7**. In this case, the efficiency of both primer pairs was considered as 2, value considered to be the best efficiency for a primer pair.

Table 5. Reagents used for mix in qPCR.

Reagent	Volume
Syber Green	10 µL
F primer	0.6 µL
R primer	0.6 µL
Nuclease free-water	3,8 µL
Template cDNA	5 µL
Total volume	20

Table 6. qPCR program.

Step	Temperature	Ramp Rate	Time	Cycles
Taq Activation	95ºC	4.4ºC/s	10 min.	1
Denaturation	95ºC	4.4ºC/s	10 sec.	40
Annealing and Elongation	60ºC	2.2ºC/s	30 sec.	
Cooling	40ºC	2.2ºC/s	30 sec.	1

Table 7. Primers utilized for genotyping qPCR.

Amplified product	Primer Pair	Sequence (5'→3')
ntpII	npt1-5'-F	GACAGGTCGGTCTTGACAAAAAG
	npt1-3'-R	GAACAAGATGGATTGCACGC
LAT51	Lat1-F	AGACCACGAGAACGATATTTGC
	Lat2.1-R	GCCTTTTCATATCCAGACACAC

Transcript levels analysis

Two step RT-qPCR (quantitative reverse transcription) protocol was used (RNA was first transformed into cDNA, and then cDNA was used as a template for qPCR). RNA extraction was performed using Maxwell® RSC Plant RNA purification kit (Promega) following the manufacturer's instructions. Isolated RNA was quantified by a NanoDrop® ND-8000 spectrophotometer (Thermo Scientific™), and its integrity was checked through 1% agarose gel electrophoresis. Afterwards, the synthesis of cDNA was performed using a First-Strand cDNA Synthesis Kit (Roche). The conditions of cDNA synthesis and the reaction volumes utilized are in **Table 8** and **Table 9** respectively. The resulting cDNA was then diluted 1:10 and 5 µL of it was used for qPCR.

Table 8. Program performed for cDNA synthesis.

Step	Temperature	Time
Initial Denaturation	65ºC	5 min.
Annealing	55 ºC	30 min.
Elongation	70ºC	30 sec.
Cooling	16 ºC	∞

Table 9. Volumes of the reagents used for cDNA synthesis.

Reagent	Volume
dNTP	2 µL
Buffer 5x	4 µL
RNase INH	0.5 µL
RT enzyme	0.5 µL
OdT	1 µL
Water	Up to 20 µL
Template RNA	500 ng
Total volume	20 µL

Relative mRNA abundance was evaluated via quantitative PCR using the same real-time PCR system as described in the "Genotyping qPCR" section, including the same conditions and volumes explained before in **Tables 5** and **6**. Primers and their efficiency are given in **Table 10**. At least two technical replicates of each biological replicate were performed. Normalized transcript abundances were calculated as described in Simon (2003) using tomato *ACT4* (Solyc04g011500) as endogenous reference gene.

Table 10. Primer pair used for qPCR and its efficiency so that calculate expression levels.

Amplified product	Primer Pair	Sequence (5'→3')	Efficiency
GFP	eGFP_F_qPCR2	CACTACCAGCAGAACACCCC	1.95
	eGFP_R_qPCR2	TCTCGTTGGGGTCTTTGCTC	
ACT	ACT_genot_F	GTGAAAAGATGACCCAGATTATG	1.91
	ACT_genot_R	CACGCTCGGTCAGGATCTTCATC	

Segregation experiments

Three different lines were selected:

- p-GFP T1 line #32.1 (*nptII*/*LAT52* ratio of 0.4): a transgenic line that is supposed to contain one T-DNA insertion per diploid (2n) genome and hence hemizygous for the *ntpII* gene that confers resistance to kanamycin.
- WT line: wild-type tomato line without the *nptII* gene
- ntpII line: A transgenic line that contains the *ntpII* gene in homozygosis.

The experiment was performed with control (MS) medium without kanamycin, and MS medium with two different concentrations of kanamycin (100 µg/mL and 150 µg/mL) to test the resistance of the selected lines.

Tomato seeds were surface-sterilized with bleach 40%. The seeds were placed into a 15 mL Falcon tube (20-30 seeds/Falcon tube). Afterwards, in a laminar air flow 10 mL of sterilization solution (40 % bleach, 2 drops of

Tween per 50 ml of solution) was added to the Falcon tube and the seeds were incubated 15 minutes in agitation. Then, seed sterilization solution was discarded and three washes of 10 mL of milli-Q water were performed during 20 minutes per wash in agitation.

The sterile seeds were sown on sterile Murashige and Skoog (MS) medium (salts (2,2g/L), MES (250mg/L), Bacto-Agar (8g/L) and no sucrose, with a final pH of 5.7) supplemented with kanamycin when required. After sowing, plates were kept at 4ºC in dark during 3 days (stratification process) so as to synchronize the timing of germination. Finally, the plates were placed in *in vitro* chambers with LD conditions (16h of light/8h of dark) and controlled temperature (24 ºC) and relative humidity (60%) during 10 days.

Pictures used to differentiate sensitive and resistant seedlings were taken with Camera Nikon D700 with micro nikkor 105 mm. lens.

Sustainability and ethical criteria

The classical method to identify one T-DNA insertion homozygous lines is a process that normally takes at least 3 generations. As far as we are concerned, for plants with relatively long reproductive cycles each generation involves a huge amount of time, but foremost a vast amount of resources like water, energy, materials and such. Bearing in mind that each generation of Microtom tomatoes requires about 3 months to reach their adult stage and bear viable seeds, at least 9 months as a minimum would be needed to obtain the T3 generation, hence wasting water and energy. By using genotyping qPCR, the use of resources and materials would decrease up to 30%, since we just need two generations to select the proper transgenic candidates.

Regarding the waste generated during this project, since we were working with transgenic plants and other hazardous materials it was indispensable to treat them correctly. Thereby, all waste generated was deposited in its corresponding bin according to the current law on waste treatment in CRAG. In addition, when new incorporations arrive to CRAG we are all required to attend to a “Waste management and lab safety” course to learn about good laboratory practices.

Results

Determination of T-DNA copy number in transgenic tomato plants by qPCR

G1-GFP, G3-GFP and p-GFP (control) plants were molecularly analyzed in the T1 generation, whereas G2-GFP lines were analyzed both in T2 and T3 generations since this project started when G2-GFP T2 plants were already in adult stage.

- T1 Generation

After *in vitro* transformation with the G1-GFP construct we obtained 24 kanamycin resistant T1 plants. After genomic DNA extraction from leaves we measured the presence of the transgene by regular PCR and we

observed that most of the lines had integrated the transgene (**Figure 4**). A few of them (#15.1, #6.2 and #11.1 lines, with a circle in **Figure 4**) seemed not to be actually transformed since the transgene was not detected, whereas the endogenous *ACT4* gene was perfectly amplified (indicative of the presence of sufficient gDNA in the sample). As mentioned before, this method is useful to identify transgenic plants but it does not inform us about the number of T-DNA insertions. To determine the transgene copy number, we next measured *nptII* and *LAT52* genes by qPCR in each line.

Those lines that did not show any amplification in the PCR, in fact showed a low number of *ntpII* copies by qPCR. However, there were some lines (#11.1, #12.1 and #19.1) that also show a low *ntpII*/*LAT52* ratio of transgene copies despite having shown to contain the transgene, resulting in inconclusive results (**Figure 4**). Therefore, these lines were not used for further analyses. On the other hand, the rest of the plants that seemed to be transformed can be grouped in three clusters. First of all, those that presented quite clear *ntpII*/*LAT52* value of 0.5, meaning 1 copy in hemizygosis (#5.1, #14.1, #14.2, #18.1 and #24.1 lines). Then, those plants that showed clearly more than one copy (#9.1, 3 copies; #16.1, 6 copies; #17.1, 2 copies and #20.3, 2 copies). And finally those that showed an intermediate ratio at around 0.5 which are the rest, such as #8.2 and #25.1 lines (Figure 4). We considered one-insertion transgenic lines when the ratio was between 0.3 and 0.7, something that of course must be checked in the next generation.

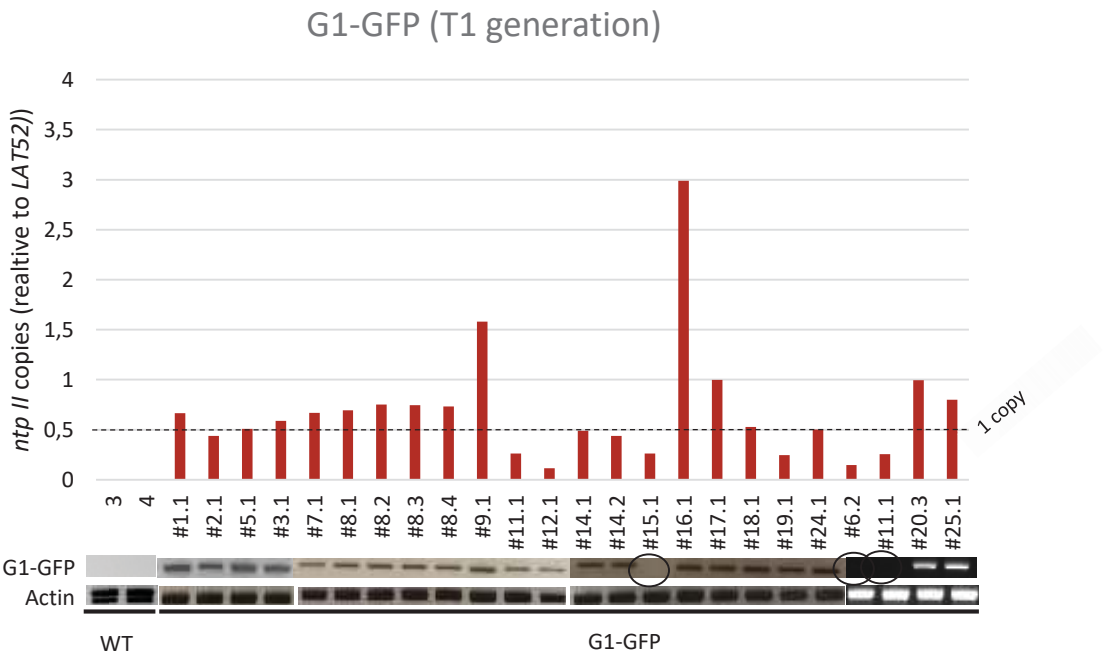


Figure 4. Number of *ntp II* copies relative to *LAT52* for G1-GFP plants in T1 generation. Results from genotyping PCR are also shown below each line. Black circles correspond to transgenic plants that did not show amplification. The dotted line in 0.5 value suggests 1 *ntp II* copy.

Regarding the G3-GFP tomato transformation, we obtained 13 kanamycin resistant plants. All the T1 lines showed a positive amplification by PCR (**Figure 5**). Some of them showed 0.5 ratio, therefore 1 copy (#4.1, #23.1

and #17.1 lines). Then, those that presented more than 1 copy (#2.1, #3.1, and #4.2), and finally the rest of the lines that seem to be ambiguous or confusing about the number of copies (**Figure 5**). However, again were considered as one-insertion lines when the ratio was from 0.3 to 0.7.

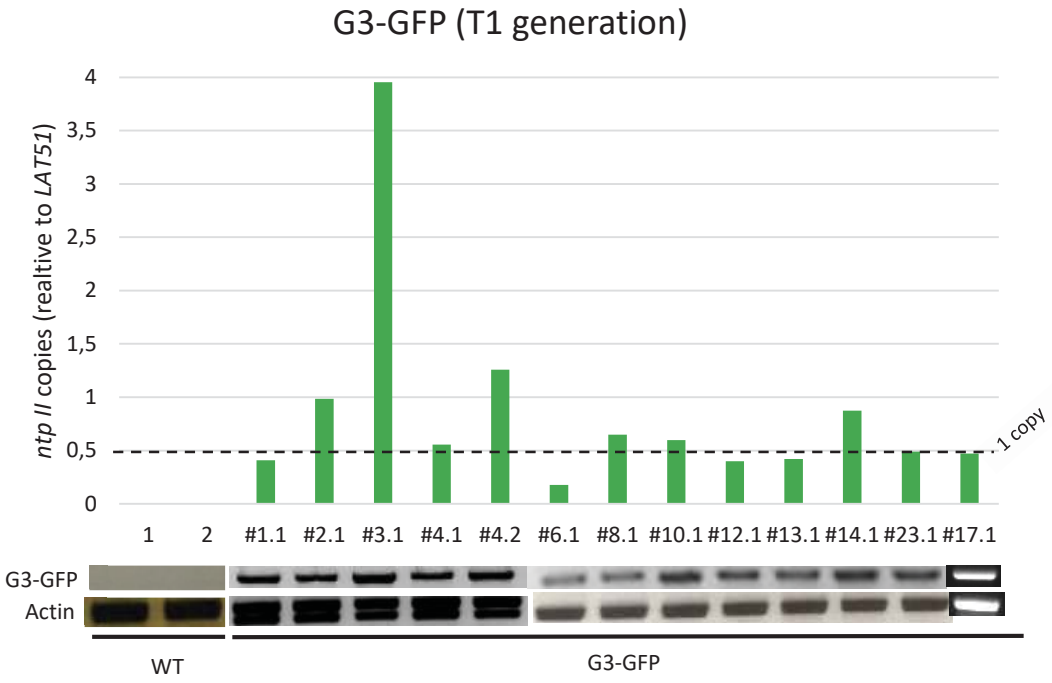


Figure 5. Number of *ntp II* copies relative to *LAT52* for G3-GFP lines in T1 generation. Results from genotyping PCR are also shown below each line. The dotted line in 0.5 value suggest 1 ntp II copy.

In the case of plants transformed with p-GFP, we had 25 kanamycin resistant lines. Three of them were negative for regular genotyping PCR (#1.2, #22.1 and #36.1 lines) indicating that the transformation may have not occurred (**Figure 6**). When performing the genotyping qPCR we observed that the ratio was close to 0 in these plants, except in line #22.1.

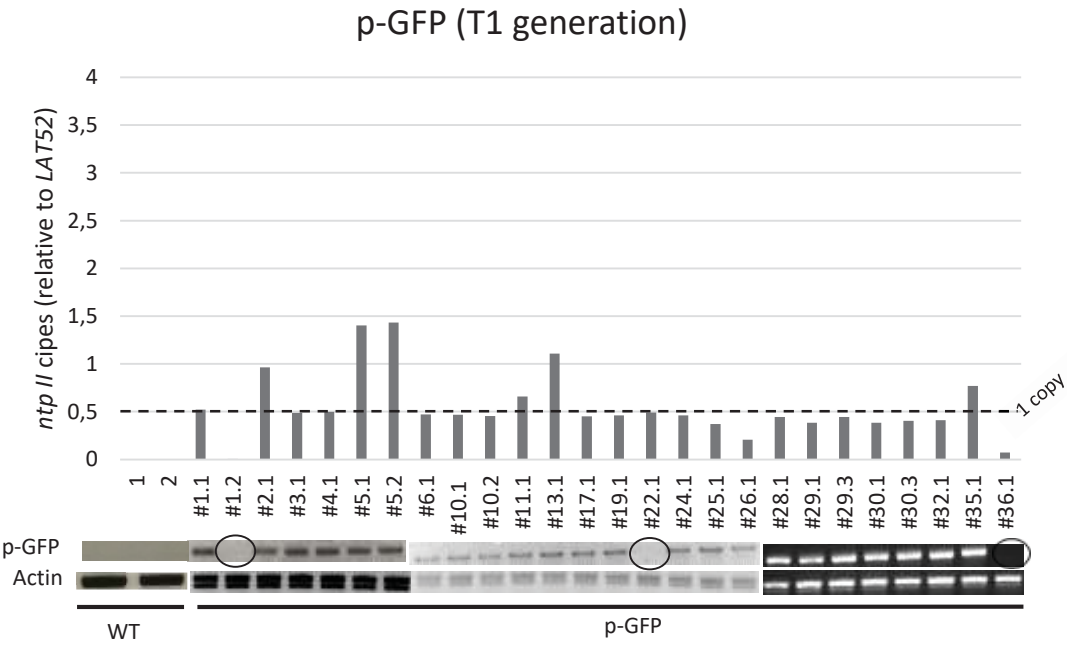


Figure 6. Number of *ntp II* copies relative to *LAT52* for p-GFP lines in T1 generation. Black circles correspond to transgenic plants that didn't show amplification. Results from genotyping PCR are also shown below each line. The dotted line in 0.5 value suggest 1 *ntp II* copy.

The rest of the lines showed a clear band for the transgene after the PCR, showing most of them a ratio close to 0.5 in the genotyping qPCR. There were some lines that showed more than one T-DNA copy or more than 0.5 ratio (#2.1, #5.1, #5.2 and #13.1), and two lines that are unclear whether they have one (ratio 0.5) or two (ratio 1) copies (#11.1 and 35.1 lines respectively) (**Figure 6**).

- T2 generation

Regarding the G2-GFP transgenic lines, it is important to highlight here that they correspond to the first tomato transformation performed. When the T1 generation was first obtained (10 kanamycin resistant plants), the genotyping qPCR technique was not available in the lab. Because of this, the T1 plants were selected according to the levels of transgene expression. The expression data analyzed showed that only two T1 lines had high levels of G2-GFP expression: #3.1 and #11.1. These were the only lines propagated to the next generations. Thus, only the T2 and T3 generations of these two lines were studied in this project.

T-DNA copy number in different T2 plants of the G2-GFP #3.1 line showed a complex profile (**Figure 7**). Some plants seemed not to have the transgene (plants #3.1-4 and -14) as they showed a 0 ratio and did not show amplification by PCR. Other plants showed ratios of 0.5 (#3.1-12) or lower (#3.1-2 and -5), 1 (#3.1-1, -8, -10, and -13) and more than 1 (rest of the plants). It therefore seemed that the G2-GFP #3.1 line was a high copy number line since its offspring showed a segregation with wide range of different values (**Figure 7**).

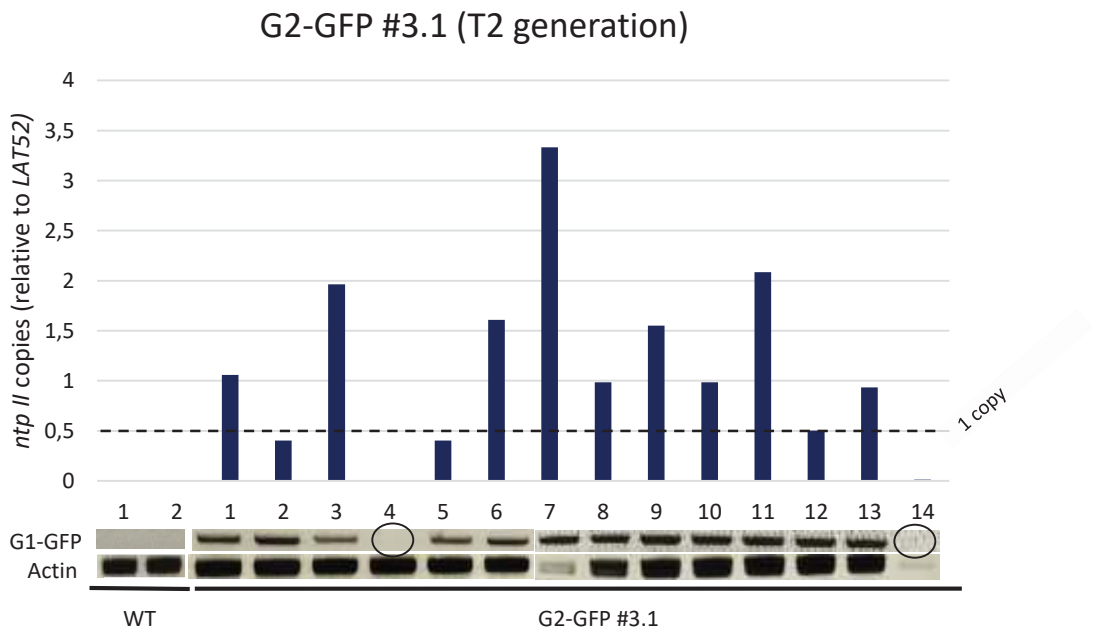


Figure 7. Number of *ntp II* copies relative to *LAT52* for G2-GFP #3.1 plants in T2 generation. Results from genotyping PCR are also shown below each line. Black circles correspond to transgenic plants that did not show amplification. The dotted line in 0.5 value suggests 1 *ntp II* copy.

Line G2-GFP #11.1 T2 segregation, by contrast, was consistent with a single T-DNA insertion. In this case we can see that plant #11.1-7 did not seem to have any copy. The other plants of the T2 generation showed values around 0.35 (#11.1-2, -3, and -5) and 0.8 (#11.1-4 and -6), but no values of 0.5 or 1 (Figure 8). We interpret this result as 1 T-DNA copy per 2n genome in the first group (hemizygotes) and 2 copies per 2n genome in plants #3.1-4 and -6 (homozygotes). However, it was also possible that line G2-GFP #11.1 had 2 T-DNA insertions and hence plants #3.1-4 and -6 would be double hemizygotes (**Figure 8**).

For T3 generation we selected 3.1-1 line (ratio 1) to propagate in order to know if its 2 copies of T-DNA were in hemizygous or homozygosis; and #11.1-2 line (**Figure 8)**, in order to check whether it actually has one copy and whether Mendelian segregation (1:2:1) is accomplished, reinforcing that it is possible to detect this by qPCR.

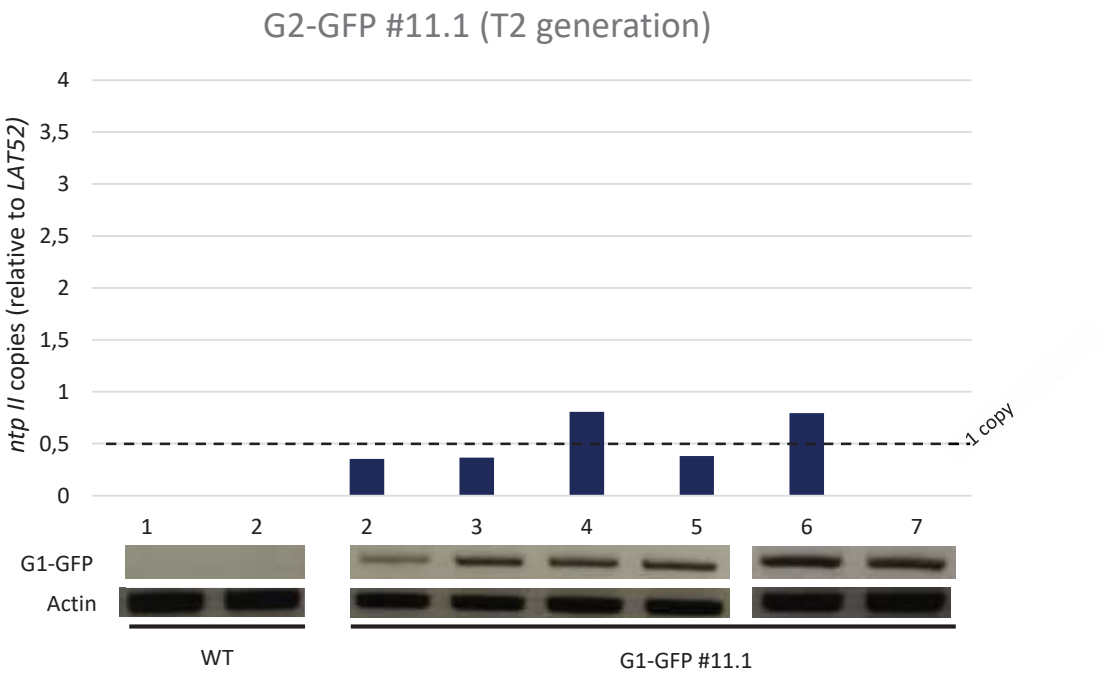


Figure 8. Number of *ntp II* copies relative to *LAT52* for G2-GFP #11.1 plants in T2 generation. Results from genotyping PCR are also shown below each line. The dotted line in 0.5 value suggests 1 *ntp II* copy.

- T3 Generation

To obtain individual plants of lines G2-GFP #3.1 and #11.1 with a single T-DNA insertion site in their genome, we analyzed the T3 offspring of plants #3.1-1 and #11.1-2. When we analyzed the T3 generation of G2-GFP #3.1-1 by PCR we found that all plants presented the transgene, whereas WT plants, as expected, didn't show a band for the transgene. By genotyping qPCR we observed that the T3 offspring showed a wide range of T-DNA copies per plant, which means that the T2 mother of these plants didn't have one insertion in homozygosis but more than one copy in hemizygosis, which in fact are segregating among the T3 individuals. However, we did obtain three T3 plants (#3.1-1-5, -8 and -9) with a ratio of 0.5, meaning that, after segregation, they only contain in their

genome one of the copies (in hemizygosis). We can now keep these lines as one-insertion transgenic plants (**Figure 9**).

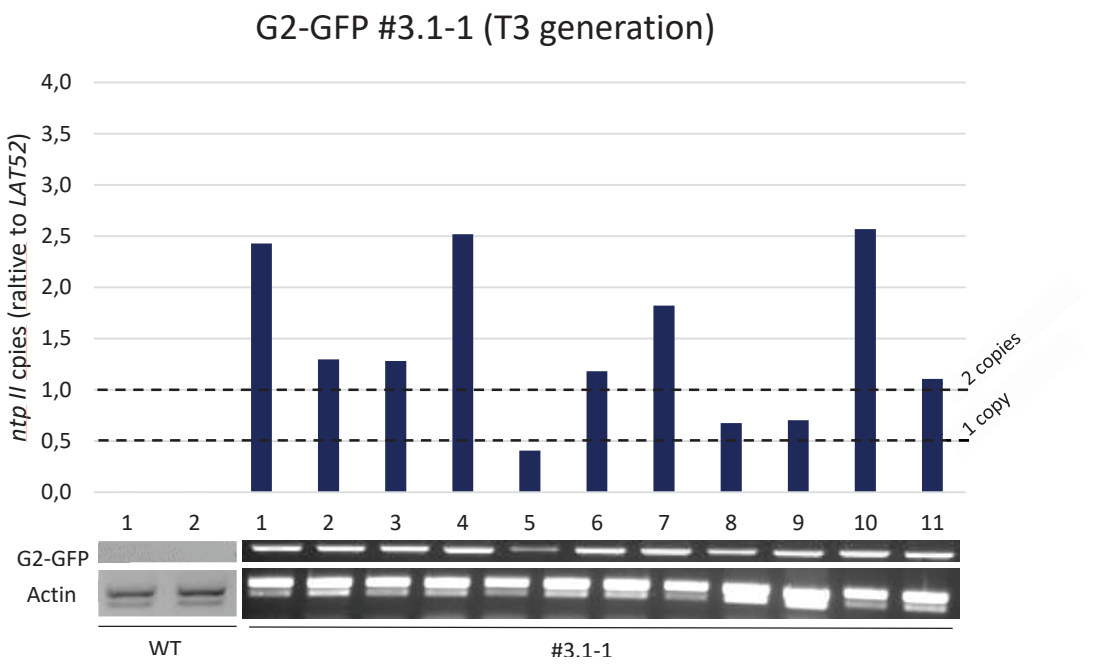


Figure 9. Number of *ntp II* copies relative to *LAT52* for G2-GFP #3.1-.1 plants in T3 generation. Results from genotyping PCR are also shown below each line. The dotted line in 0.5 and in 1 value suggest 1 and 2 *ntp II* copies respectively.

The T3 offspring of plant G2-GFP #11.1-2 only showed ratios close to 1 (plants #11.1-2-1, -2, -6 and -7), and 0.5 (the rest) (**Figure 10**). Taking into account that the T2 mother plant had a ratio of 0.5, the data obtained in the T3 generation strongly suggests that we have a single T-DNA insertion in this line. Among the T3 offspring we would have hemizygous plants (ratio 0.5) and homozygous plants for the transgene (ratio 1). We were also expecting to find azygous plants among the T3 generation, since a Mendelian segregation should be taking place. However, we did not obtain any plant lacking the transgene, which could be due to the low number of T3 plants tested.

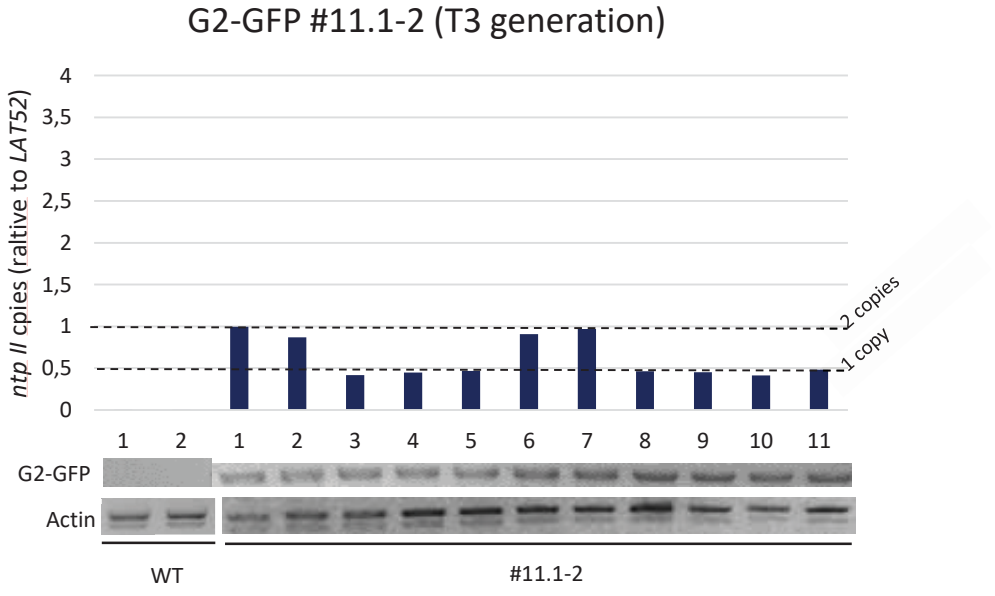


Figure 10. Number of *ntp II* copies relative to *LAT52* for G2-GFP #11.1-2 plants in T3 generation. Results from genotyping PCR are also shown below each line. The dotted line in 0.5 and in 1 value suggest 1 and 2 *ntp II* copies respectively.

- Validation of the genotyping qPCR technique using the classical method for transgenic plants selection

In order to confirm whether our qPCR method is actually informing us about the right T-DNA copy number, we performed a segregation experiment in kanamycin supplemented media. For this experiment we select three different lines described in Material and Methods.

As we can observe in **Table 11**, the percentage of germination in control medium (MS) for WT (without the kanamycin resistance gene), *ntp II* line (with the kanamycin resistance gene in homozygosis) and p-GFP #32.1 (ratio 0.5) are 60%, 87% and 77% respectively. When germination rate is taken into account, no WT seeds grew in kanamycin, as expected. In the case of *ntp II* line, 92% of them grow in 100 µg/mL kanamycin +MS, and 107% of them grow in MS + kanamycin (relative to the germination in rate in MS), confirming that the resistance gene is not segregating. Finally, p-GFP #32.1 line showed a 78% of seeds that grew in MS + Kanamycin (100 µg/ml), whereas the 61% of the seeds grew in MS + Kanamycin (150 µg/ml). These kanamycin resistant rates are close to a Mendelian segregation for one T-DNA insertion (75% of resistance), validating that a line with a ratio of 0.5 actually harbors one copy of the transgene in its (2n) genome.

Table 11. Red row corresponds to the germination rate for each line in different mediums. Orange rows correspond to the percentage of plants that survived taking into account the germination rate for each line and medium. (n= number of seeds sown per line).

	WT (n=15)	*ntp II* line (n=15)	p-GFP #32.1 (n=30)
MS	(9/15) 60%	(13/15) 87%	(23/30) 77%
MS + Kanamycin (100 µg/ml)	0%	(12/13) 92%	(18/23) 78%
MS + Kanamycin (150 µg/ml)	0%	(14/13) 108%	(14/23) 61%

As shown in **Figure 11**, sensitive plants were not able to develop a proper root and most of them showed pale or albino leaves. We bear in mind these two characteristics when deciding which plants were resistant or sensitive.

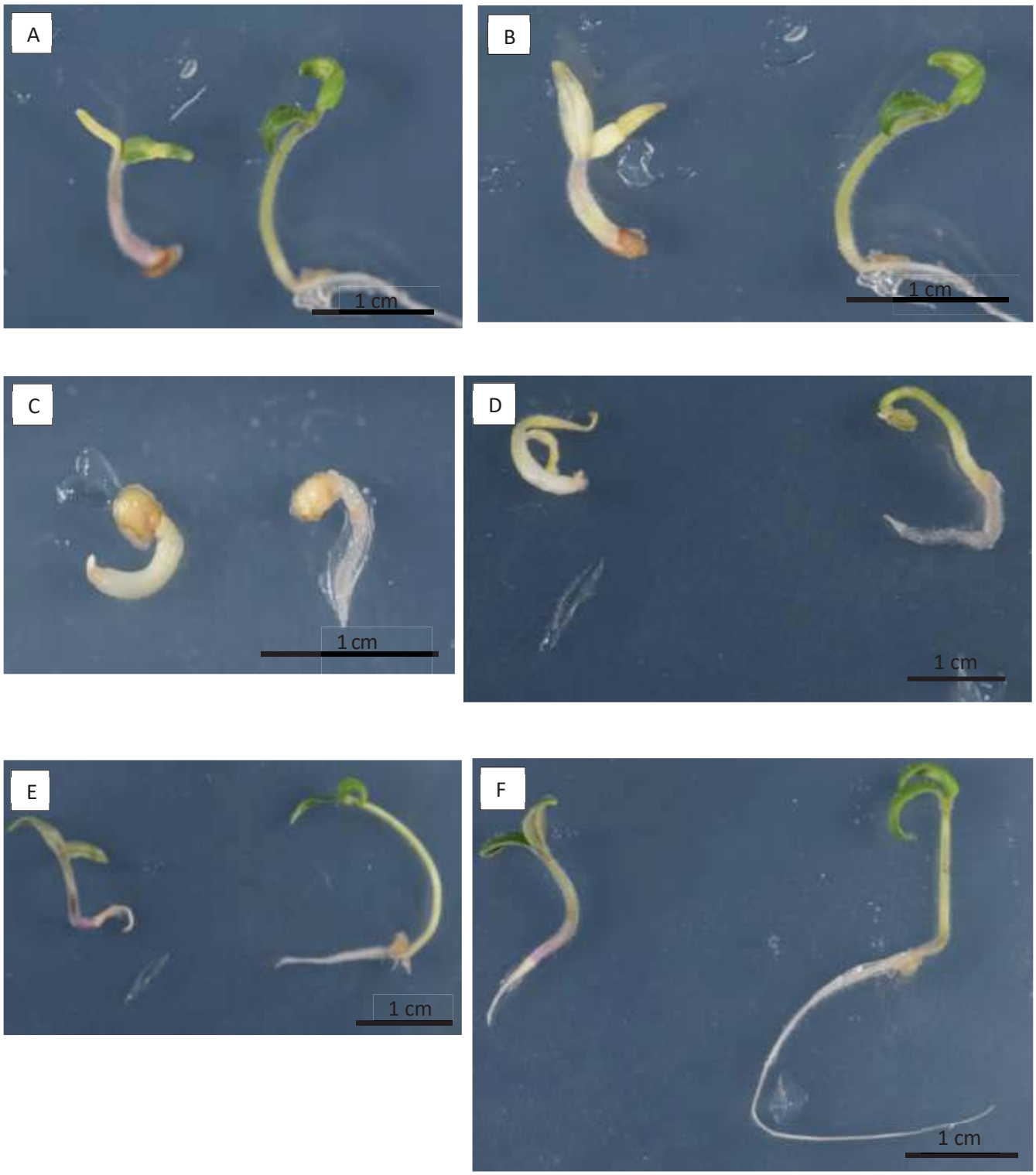


Figure 11. Kanamycin **sensitive or resistant tomato seedlings**. For each picture, seedlings located on the right were considered resistant whereas the ones located on the left were considered sensitive to kanamycin. **(A-B)** The seedlings located on the left show two examples of albino phenotype as a consequence of kanamycin, these type of seedlings are not viable. **(C-F)** They show different stages of development from the less developed to the most developed (10 days' seedlings) despite stratification process. In all of them we can notice that the differentiation area of the root is well-developed for the plants on the right, on the contrary this area is not well-developed or not even developed for the plants on the left side.

Determination of transgene expression levels in selected transgenic plants

After validating the genotyping qPCR method, we preferentially selected transgenic plants harboring only one T-DNA insertion (ratio 0.5) for further molecular characterization. Sometimes, although the T-DNA is inserted in the plant genome, the transgene is not actually expressing due to some events more detailed in the discussion. Because of this, it is important to measure the transcript levels of the transgene in each selected transgenic line.

- T1 Generation

Among the G1-GFP T1 generation, we selected 9 plants showing a ratio close to 0.5 (meaning one T-DNA insertion) to further analyze the expression level of the transgene. Firstly, we can observe in **Figure 12** that WT lines did not express *GFP*. In contrast, the selected G1-GFP lines showed *GFP* expression, although the levels varied among plants. The line whose expression is the highest is #3.1, while #24.1 line showed the lowest expression.

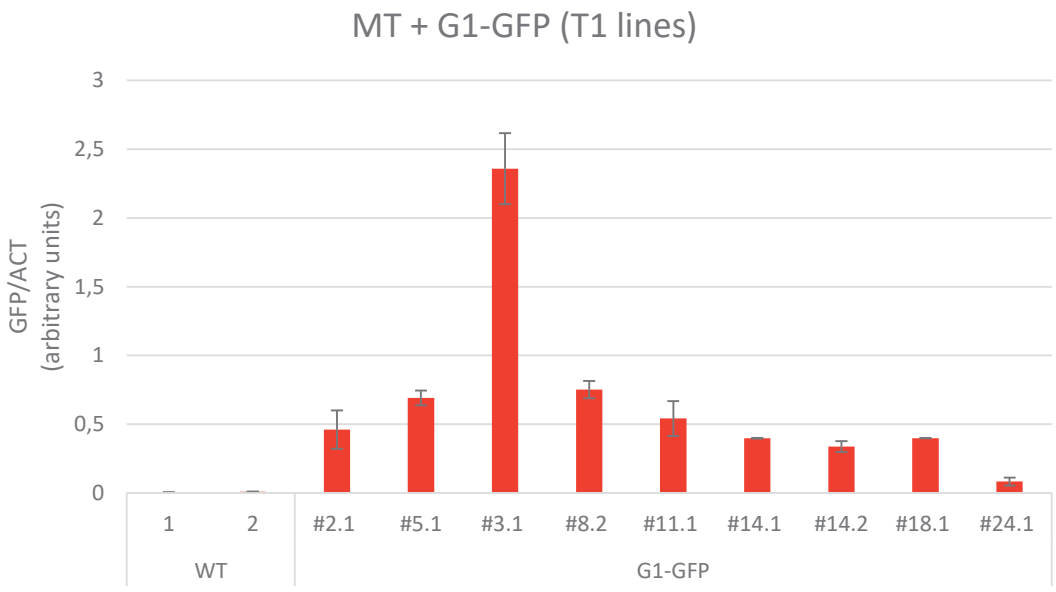


Figure 12. *GFP* expression levels of G1-GFP and WT plants in T1 generation expressed in arbitrary units. WT plants were used as a control.

Regarding the G3-GFP T1 transgenic lines, we selected 8 plants, again, with a ratio around 0.5 to analyze the transcript levels of the transgene. As we can see in **Figure 13**, the control plants (WT) didn't show *GFP* expression, while the transgenic lines showed a wide variety of *GFP* expression values. Plants #6.1, #8.1, #23.1, and #17.1 showed the greatest expression among all lines. However, #12.1 showed a very low expression level.

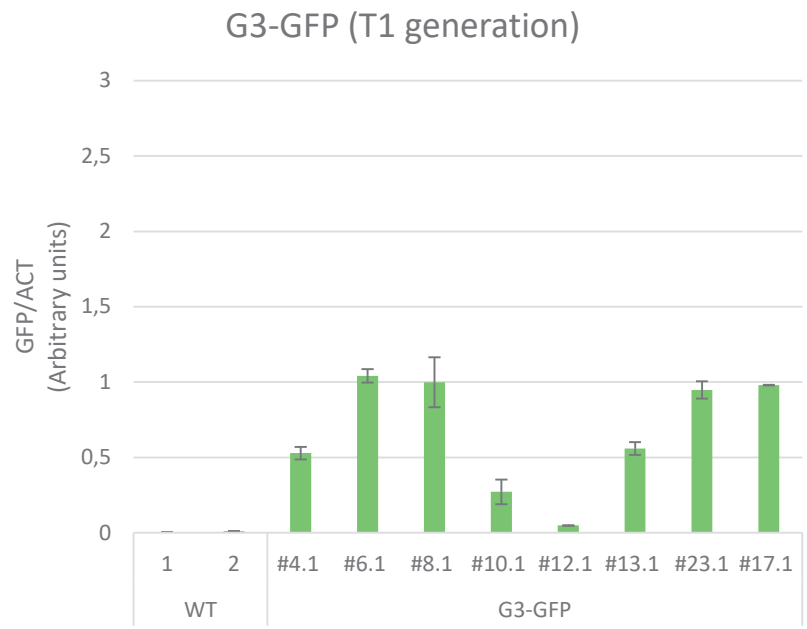


Figure 13. *GFP* expression levels of G3-GFP and WT plants in T1 generation expressed in arbitrary units. WT plants were used as a control.

From p-GFP T1 generation we selected 11 independent one-insertion plants. Among them, #6.1 line showed huge *GFP* expression levels **Figure 14**. These *GFP* transcript levels were also much higher than the previous transgenic lines described in **Figures 12** and **13**. In order to better compare *GFP* expression of the rest of the plants, the data from #6.1 line was removed in **Figure 14**.

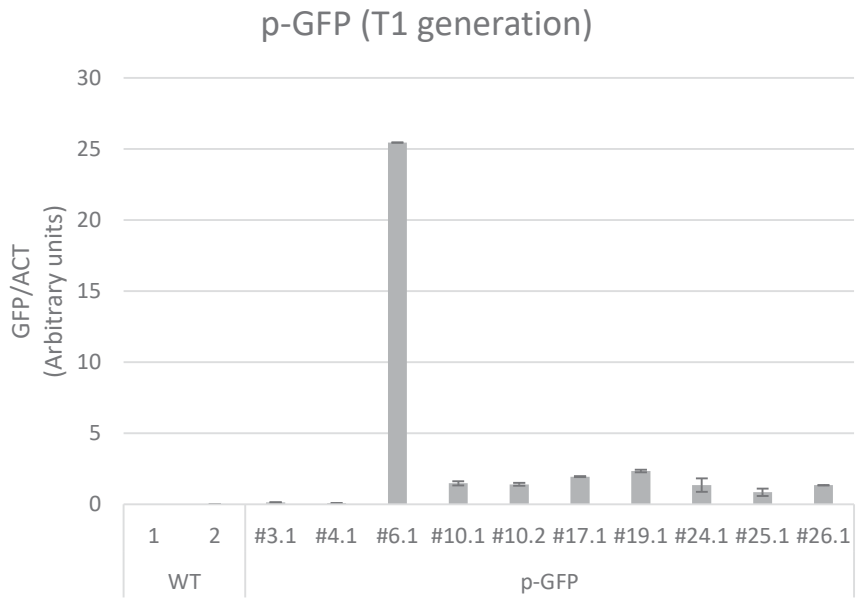


Figure 14. *GFP* expression levels of p-GFP and WT plants in T1 generation expressed in arbitrary units. WT plants were used a control.

Once the line #6.1 line was removed, we can observe in **Figure 15** that WT lines didn't show any *GFP* expression, as expected. Apart from #6.1 line, #17.1 and #19.1 plants showed the highest *GFP* expression levels, which were

more similar to those obtained in G1-GFP and G3-GFP transgenic lines. We also found some lines with very low levels of *GFP* expression (#3.1 and #4.1)

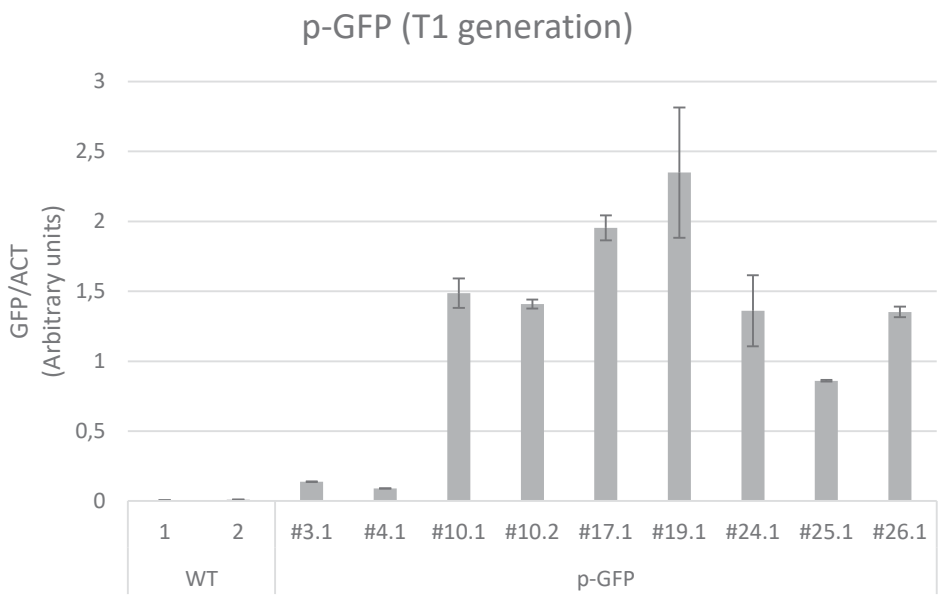


Figure 15. *GFP* expression levels of p-GFP and WT plants in T1 generation expressed in arbitrary units where #6.1 line was removed. WT plants were used as a control.

From all these results, we were able to select 3 independent T1 transgenic lines for each transgene to propagate to the T2 generation based on (a) the existence of a single T-DNA insertion in their genome, and (b) their capacity to express the transgene. We chose G1-GFP #5.1, #11.1 and # 14.2 lines, G3-GFP #6.1, #8.1 and #23.1 lines, and pGFP #10.1, #10.2 and #19.1 lines. Sowing their seeds, we will be able to identify homozygous plants directly in the T2 generation by genotyping qPCR and to use them for future studies regarding the biosynthesis of carotenoids in plants.

- T3 generation

As for the G2-GFP lines we analyzed the expression of the T3 offspring, regardless of the number of copies. Control plants (WT) didn't show any *GFP* expression (**Figure 16**). Plant #3.1-1-9 showed the highest expression despite having a low copy number ratio (0.7), confirming that transgene expression levels do not depend on the number of T-DNA insertions but mainly on the genome area where the integration occurs (**Figure 16**).

For the T2 line #11.1-2 we only selected three T3 candidates, two homozygous and one hemizygous plants (**Figure 17).** The three of them showed similar levels of expression.

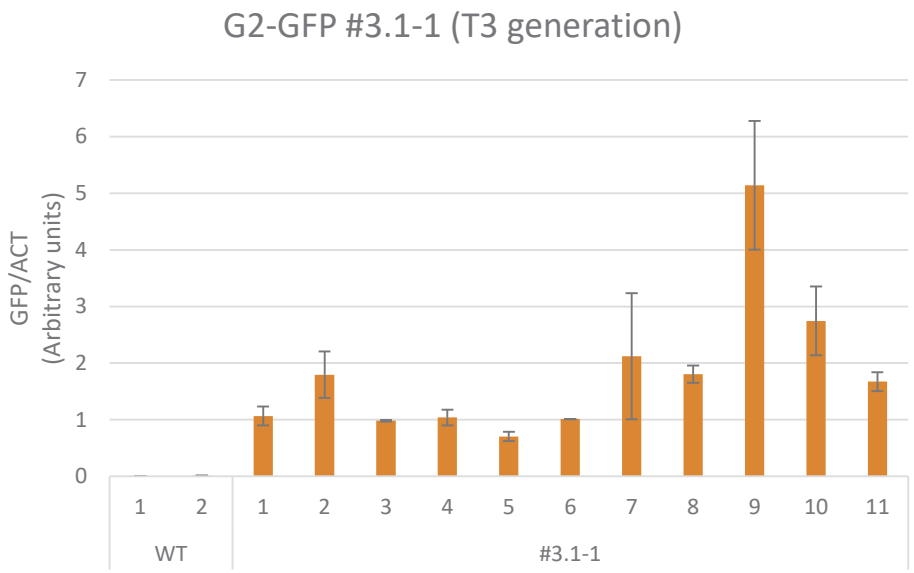


Figure 16. *GFP* expression levels of G2-GFP # 3.1-1 and WT plants in T3 generation expressed in arbitrary units. WT plants were used as a control.

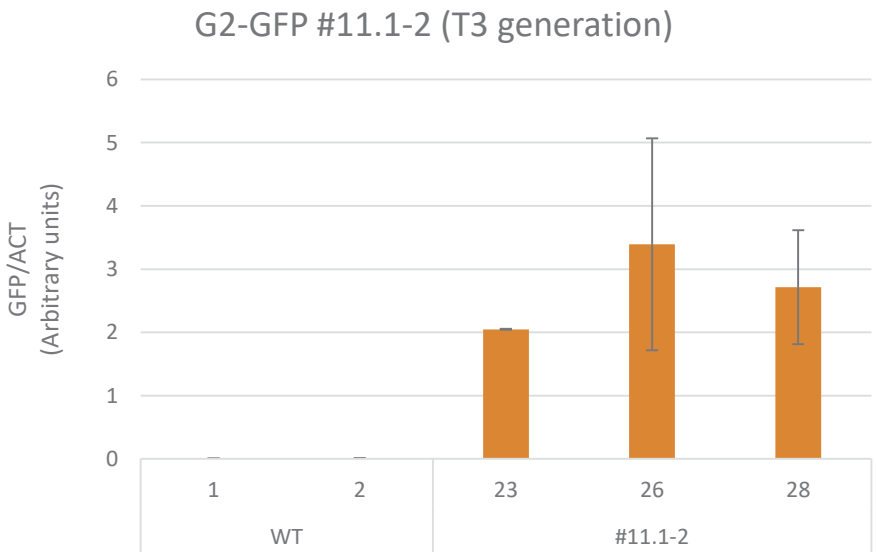


Figure 17. *GFP* expression levels of G2-GFP and WT plants in T1 generation expressed in arbitrary units. WT plants were used as a control.

From the G2-GFP #3.1-1 transgenic plants we chose #3.1-1-8 and -9, hoping that they harbor only 1 of the insertions, something that we will check in the T4 generation. Among #11.1-2 offspring we decided to sow seeds coming from #11.1-2-2 and -7 homozygous plants.

Discussion

Determination of T-DNA copy number

The results of this project have proved that the T-DNA copy number can be determined by means of genotyping qPCR. However, this method can have a few drawbacks, for instance, when values are obtained that are not clear-cut. In some of the genotyped lines there were some values that showed a ratio much lower than 0.5 such as G1-GFP #11.1 T1 (Figure 4), G3-GFP #6.1 T1 (Figure 5), p-GFP #26.1 T1 (Figure 6), among others. Taking into account that the T-DNA is integrated since they were PCR-positive and the ratio didn't exceed 0.5 led to the suggestion that those plants could just have 1 T-DNA copy per 2n genome (i.e., they are likely single hemizygotes). Furthermore, there were other values between 0.5 and 1, such as G3-GFP #4.2 T1 (Figure 5), or p-GFP #35.1 T1 (Figure 6). In these cases, we were not able to ensure if the plant had one or more T-DNA insertions. Therefore, we could either discard them for future experiments or propagate them to confirm the copy number in the next generation.

Sometimes we obtained some contradictory results. We found that some lines showed an absence of the transgene by PCR and a positive amplification by genotyping qPCR (G1-GFP #6.2, #11.1 and #15.1 T1 plants (Figure 5); and pGFP #22.1 and #36.1 T1 (Figure 6). For regular PCR, we detected the presence of the transgene using a primer pair that specifically detects the GOI and the GFP: the forward primer anneals in the GOI (G1, G2 or G3) while the reverse primer is specific for the GFP tag. Thus, we ensure that the integration of our transgene in the plant has taken place. However, for the qPCR genotyping experiment we used a primer pair to detect the kanamycin resistant gene (*nptII*), which is also integrated as part of the T-DNA. Assuming that qPCR has a low error-prone as is a very sensitive technique, and controls (WT plants) were negative in qPCR and PCR, led us to the hypothesis that T-DNA was somehow wrongly or partially inserted. In this way, only a fragment of the T-DNA (harboring the *nptII* gene) might have been integrated in the genome, whereas the rest was lost, thus being undetectable by PCR with the primers we used. Some studies have suggested that both ssDNA (single-strand DNA) and dsDNA (double-strand) T-DNA integration can occur in plants (Chilton & Que, 2003). However, the way how it is actually integrated in the plant genome remains unknown (Gelvin, 2017). Therefore, it is not easy to set up a possible scenario of how T-DNA could be wrongly or partially inserted, but is still a possibility we cannot discard. To simply overcome this contradictory results, we could use a qPCR primer pair to amplify *GFP* instead of *nptII*, which would give us closer results to the genotyping PCR. If the transgene or the GFP are lost during the T-DNA integration we shouldn't detect it neither by PCR nor by qPCR (or RT-qPCR).

Agrobacterium mediated T-DNA integration is illegitimate or random and, as a consequence, the probability that the T-DNA was inserted two times in the same location on both chromosomes is very scarce. Hence, when transformation occurs, either 1 or more copies are considered to be in hemizygosis in the very first transgenic generation (T1). Using genotyping qPCR not only allows us to select the best T1 lines to propagate (according to their copy number), but also to know whether the T-DNA of the T2 plants is in homozygosis or hemizygosis. By

contrast, the classical method requires to analyze the segregation of the T3 generation, which is a much longer process, especially in tomato. When the copy number is determined in the next generations, instead of measuring it in the T1, it is much more difficult to obtain conclusions (Figure 7). A T2 plant with a ratio of 0.5 clearly informs that it contains 1 T-DNA insertion in hemizygosis (#3.1-2, #3.1-5 and #3.1-12 lines) (Figure 7). However, sister plants with ratios higher than 1 only inform that the mother had multiple T-DNA insertions. In this case, for instance, a plant with a ratio of 1 (*e.g*. #3.1-1) informs that it has 2 copies. However, it is impossible to know if the plant has one insertion in homozygosis or two independent insertions in hemizygosis. This should be determined propagating its offspring (T3 generation). The best situation we can find in the offspring of an unknown copy-number T1 line, is that all the daughters show ratios of 0, 0.5 and 1, preferably in Mendelian proportions. This would mean that the mother only had one T-DNA insertion in its genome, and now, in its offspring homozygous plants can be identified (those showing a ratio of 1). We obtained something similar to this situation in the T2 generation of the G2-GFP #11.1 line (Figure 8). We obtained 1 azygous plant (ratio 0), 3 hemizygous plants (ratio 0.5) and 2 homozygous plants (ratio 1), clearly suggesting that we have one-insertion transgenic line.

We were also interested in checking the distribution of the T-DNA in the T3 offspring of the T2 plant G2-GFP #3.1-1 (*nptII*/*LAT52* ratio of 1) (Figure 7). Since we don't know the ratio of the parental #3.1 line, different possibilities could have occurred. The simplest one would have been that the two copies from #3.1-1 line were in homozygosis (meaning one T-DNA insertion, AA) resulting in an entire homozygous offspring, which did not occur. The other possibility, is that the 2 copies were in hemizygosis (AaBb). According to the Punnet Table A2, we would obtain:

- 1/16 (AABB) → Ratio=2 (4 copies, two T-DNA insertions in homozygosis)
- 4/16 (AABb + AaBB) → Ratio=1.5 (3 copies, one insertion in homozygosis and the other in hemizygosis)
- 6/16 (AaBb + AAbb + aaBB) → Ratio=1 (2 copies, many possibilities)
- 4/16 (Aabb + aaBb) → Ratio= 0.5 (1 copy in hemizygosis)
- 1/16 (aabb) → Ratio=0 (0 copies, azygous)

Table 12. Punnet table showing the Mendelian segregation for two different insertions. Capital letters mean the presence of T-DNA, where the different letters (A or B) mean different T-DNA loci.

gametes	AB	Ab	aB	Ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	Aabb

Proportions obtained from the Punnet table are hardly consistent with the values showed in Figure 8 (G2-GFP #3.1-1 T3 generation). The use of this method for multiple-insertion lines can hinder the determination of the real copy number. Foremost the small quantity of samples is not representative to know whether the obtained proportions accomplish the Mendelian segregation proposed in Table 2. In addition, in Figure 8, some plants show 2.5 values, meaning 5 copies, a number of copies that is not consistent with a parent with 2 copies in hemizygosis, but it does for 3 or more T-DNA insertions. This could have happened, since other sisters of the parental line #3.1-1 showed ratios up to 3 (meaning up to 6 insertions). Copy number of 3 (in hemizygosis) for parent #3.1-1 could be proposed, but its ratio was far from 1.5 value (ratio=1.06).

Despite the inconclusive result, we reinforce the importance of genotyping the very first generation by qPCR. This will facilitate the determination of the copy number of the T1 generation and the T-DNA zygosity of the T2 generation.

When we analyzed the offspring of the G2-GFP #11.1-2 line (*nptII*/*LAT52* ratio of 0.5) we didn't obtain any azygous plant but we found that all the plants showed ratios close to 0.5 (7, 64%) and 1 (4, 36%), confirming again the presence of only one T-DNA insertion in the T1 #11.1 line (Figure 9). The absence of azygous plants could be due to the low number of individuals tested; however, these proportions remain somehow a Mendelian segregation (25% azygous, 50% hemizygous and 25% homozygous).

- Validation of the method

Before performing the experiment, we assumed that p-GFP #32.1 line with 0.4 ratio has 1 T-DNA insertion since its ratio is close to 0.5. First, even though WT seedlings seemed to have a low germination rate in MS medium, none of them grew in the presence of kanamycin, which means that the medium was actually selective. Furthermore, *ntpII* lines grew in selective medium as expected for plants resistant to kanamycin. The p-GFP line showed a germination rate of 77% (23/30) in MS medium. Assuming that this line has a similar germination rate in the selective medium, we can consider that approximately 78% (18/23) and 61%(14/23) of the viable offspring is actually resistant to kanamycin, strongly suggesting that the transgene is segregating according to Mendelian proportions for only one insertion (only the hemizygous - 50% - and homozygous - 25% - plants should be resistant to kanamycin). Due to the 3:1 ratio suggested by the viable offspring from p-GFP line under selective medium (78% and 61%), we can conclude that p-GFP with 0.4 ratio has 1 copy of T-DNA in hemizygosis (Table 11). This experiment, eventually validated the use of qPCR for the determination of the copy number in transgenic plants.

Expression level

As it is explained in the Results section, the transgenic lines analyzed showed a different level of *GFP* expression. We found that the expression level does not correlate either directly or indirectly with the T-DNA copy number.

The actual variable that determines T-DNA expression level depends on chromosome packaging, the nucleotide sequence where it is inserted, methylations, posttranscriptional regulations and silencing (Ziemienowicz et al., 2008). Apart from natural processes or regulations from the plant itself, the fact of inserting more genes already harbored in the plant can trigger RNA silencing and interfere with the expression pattern (Schubert et al., 2004), which could be occurring in those lines with really poor levels of transgene expression. To verify if this is happening we should further analyze the expression levels of the endogenous genes (*G1*, *G2* and *G3*) in the transgenic lines comparing the values with the those obtained from wild-type and pGFP plants. These analyses would confirm the overexpression of these genes in some transgenic plants together with the *GFP* expression measurement, but also they would show if some of the lines are silencing even the endogenous tomato gene, especially in those that showed such a low *GFP* transcript levels.

Taking into account the T-DNA copy number and the transgene expression level of the tested tomato transgenic plants, we were able to select the best ones to propagate and eventually obtain the homozygous lines to study the biological processes in which the overexpressed genes are involved. As specified in Results, we selected G1-GFP #5.1, #11.1 and # 14.2 lines (Figure 11); G3-GFP #6.1, #8.1 and #23.1 lines (Figure 12); and pGFP #10.1, #10.2 and #19.1 (Figure 14) lines from the T1 generation of transgenic plants. From the T3 G2-GFP generation we chose #3.1-1-8, #3.1-1-9 (Figure 15), #11.1-2-2 and #11.1-2-7 (Figure 16). Although some single T-DNA insertion plants showed higher levels of *GFP* expression (e.g. G1-GFP #3.1 or pGFP#19.1) we could not propagate them because their fruits were seedless, something that can happen because of the *in vitro* plant regeneration or the T-DNA genomic insertion area. Moreover, we decided to select independent transgenic plants overexpressing similar levels of the gene of interest, in order to compare further physiological and molecular experiments. This will allow us to better determine the biological function of the three genes selected (*G1*, *G2* and *G3*) in tomato carotenoid biosynthesis, using the pGFP lines as control.

Conclusions

The results obtained during the development of this project allowed to deduce the following conclusions:

- Genotyping qPCR it is a method that works to assess T-DNA copy number in transgenic tomato plants. Results obtained with this method for the p-GFP #32.1 were in agreement with the results obtained by classical marker segregation assays.
- If the genotyping qPCR analysis is applied from the T1 generation, a homozygous line with a single T-DNA insertion per genome could be obtained in the next generation (T2).
- Analysis of transgene expression by RT-qPCR demonstrate that there is no direct or inverse correlation between T-DNA copy number and transgene expression levels.

Acknowledgment

First of all, I would like to thank Manuel Rodriguez-Concepción for opening his lab doors for me and for the good treat during my stay, not to mention everything I learnt from him. I also want to thank all my lab mates for making me feel at home as well as their kind and selfless help whenever I needed it.

I would like to mention "Sci-hub" because they have allowed me, as well as other students, to support this project in a scientific and proven way breaking down barriers in order to make this whole science world wider and accessible.

I want to thank very very much to Victoria Barja for her patience, attention, assistance, kindness, support, among many others. In addition, not only for having transmitted such rich knowledge in me, but also for teach and strengthen me important values and attitudes in the research world that I will never forget, being one of the best teachers I have ever had.

Last but not least, I would like to acknowledge to my family and my friends for their support and help.

Bibliography

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, *25*(2), 169–193. doi: https://doi.org/10.1677/jme.0.0250169

Chilton, M.-D., & Que, Q. (2003). Targeted Integration of T-DNA into the Tobacco Genome at Double-Stranded Breaks: New Insights on the Mechanism of T-DNA Integration. *Plant Physiology*, *133*(3), 956–965. doi: https://doi.org/10.1104/pp.103.026104

Dan, Y., Yan, H., Munyikwa, T., Dong, J., Zhang, Y., & Armstrong, C. L. (2006). MicroTom — a high-throughput model transformation system for functional genomics. *Plant Cell Reports*, *25*(5), 432–441. doi: https://doi.org/10.1007/s00299-005-0084-3

Gelvin, S. B. (2000). *Agrobacterium* and plant genes involved in T-DNA transfer and integration. *Annual Review of Plant Physiology and Plant Molecular Biology*, *51*(1), 223–256. doi: https://doi.org/10.1146/annurev.arplant.51.1.223

Gelvin, S. B. (2012). Traversing the Cell: *Agrobacterium* T-DNA's Journey to the Host Genome. *Frontiers in Plant Science*, *3*(March), 1–11. doi: https://doi.org/10.3389/fpls.2012.00052

Gelvin, S. B. (2017). Integration of *Agrobacterium* T-DNA into the Plant Genome. *Annual Review of Genetics*, *51*(1), annurev-genet-120215-035320. doi: https://doi.org/10.1146/annurev-genet-120215-035320

Hoekema, A., Hirsch, P. R., Hooykaas, P. J. J., & Schilperoort, R. A. (1983). A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. *Nature*, *303*, 179. doi: http://dx.doi.org/10.1038/303179a0

Ingham, D. J., Beer, S., Money, S., & Hansen, G. (2001). Quantitative Real-Time PCR Assay for Determining Transgene Copy Number in. *BioTechniques*, *31*(1), 132–140. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11464506

Jaswir, I., Noviendri, D., Fitri, R., & Octavianti, F. (2011). Carotenoids: Sources, medicinal properties and their application in food and nutraceutical industry. *Journal of Medicinal Plants Research*, *5*(33), 7119–7131. doi: https://doi.org/10.5897/JMPRX11.011

Kado, C. I. (2014). Historical account on gaining insights on the mechanism of crown gall tumorigenesis induced by *Agrobacterium tumefaciens*. *Frontiers in Microbiology*, *5*(340), 1–15. doi: https://doi.org/10.3389/fmicb.2014.00340

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, *25*(4), 402–408. doi: https://doi.org/10.1006/meth.2001.1262

McCullen, C. A., & Binns, A. N. (2006). *Agrobacterium tumefaciens* and Plant Cell Interactions and Activities Required for Interkingdom Macromolecular Transfer. *Annual Review of Cell and Developmental Biology*, doi: *22*(1), 101–127. doi: https://doi.org/10.1146/annurev.cellbio.22.011105.102022

Nester, E. W. (2015). *Agrobacterium*: nature's genetic engineer. *Frontiers in Plant Science*, *5*(730), 1–16. doi: https://doi.org/10.3389/fpls.2014.00730

Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, *1*(3), 1559–1582. doi: https://doi.org/10.1038/nprot.2006.236

Rodriguez-Concepcion, M., Avalos, J., Bonet, M. L., Boronat, A., Gomez-Gomez, L., Hornero-Mendez, D., Limon M. C., Meléndez-Martínez, A. J., Olmedilla-Alonso, B., Palou, A., Ribot, J., Rodrigo, M. J., Zacarias, & L., Zhu, C. (2018). A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. *Progress in Lipid Research*, *70*, 62–93. doi: https://doi.org/10.1016/J.PLIPRES.2018.04.004

Ruiz-sola, M. Á., & Rodríguez-Concepción, M. (2012). Carotenoid Biosynthesis in Arabidopsis : A Colorful Pathway. *The Arabidopsis Book / American Society of Plant Biologists*, *10*(e0158), 1–28. doi: https://doi.org/10.1199/tab.0158

Sawada, H., Ieki, H., Oyaizu, H., & Matsumoto, S. (1993). Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *International Journal of Systematic Bacteriology*, *43*(4), 694–702. doi: https://doi.org/10.1099/00207713-43-4-694

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., & Schmidt, R. (2004). Silencing in *Arabidopsis* T-DNA Transformants: The Predominant Role of a Gene-Specific RNA Sensing Mechanism versus Position Effects. *The Plant Cell Online*, *16*(10), 2561–2572. doi: https://doi.org/10.1105/tpc.104.024547

Simon, P. (2003). Q-Gene: Processing quantitative real-time RT-PCR data. *Bioinformatics*, *19*(11), 1439–1440. doi: https://doi.org/10.1093/bioinformatics/btg157

Yang, L., Pan, A., Jia, J., Ding, J., Chen, J., Cheng, H., Zang, C., & Zhang, D. (2005). Validation of a tomato-specific gene, *LAT52*, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic tomatoes. *Journal of Agricultural and Food Chemistry*, *53*(2), 183–190. doi: https://doi.org/10.1021/jf0493730Ingha

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., & Chua, N.-H. (2006). Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nature Protocols*, *1*(2), 641–646. doi: http://dx.doi.org/10.1038/nprot.2006.97

Ziemienowicz, A., Tzfira, T., & Hohn, B. (2008). Mechanisms of T-DNA integration BT - Agrobacterium: From Biology to Biotechnology. In T. Tzfira & V. Citovsky (Eds.) (pp. 395–440). New York, NY: Springer New York. doi: https://doi.org/10.1007/978-0-387-72290-0_11