



Universitat de Lleida

Isoprenoid and flavonoid biosynthesis and regulation in higher plants

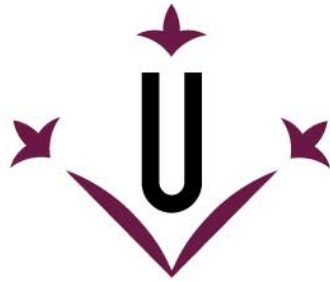
Xin Jin

<http://hdl.handle.net/10803/667579>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



Universitat de Lleida

TESI DOCTORAL

**Isoprenoid and flavonoid biosynthesis and regulation
in higher plants**

Xin Jin

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en Ciència i Tecnologia Agrària i Alimentària

Directors
Changfu Zhu
Paul Christou

Front Cover: Yuanyang Terraced Rice Fields, Yunnan County (China)
Back Cover: *Gentiana lutea* L. var. *aurantiaca* flower

Supervisors: **Changfu Zhu**

Paul Christou

Department of Plant Production and Forestry Science

School of Agricultural and Forestry Engineering

University of Lleida

September of 2019

Signatures

C. Zhu

P. Christou

DEDICATION

To my parents

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Changfu Zhu and co-supervisor Dr. Paul Christou for supporting me over the years with great enthusiasm and patience. It has been an absolute pleasure to have your guidance in all the time of research and writing of this thesis. Without your guidance and encouragement this PhD would not have been achievable.

A very special thanks to Dr. Teresa Capell for her insightful comments and selfless supporting throughout all these years. Thank you Dr. Vicente Medina, I appreciate all your kind help with the microcopy work and it is really my honour to have you as a good friend.

I gratefully acknowledge the University of Lleida for offering me the platform for my PhD study. Thanks to “Agencia de Gestión de Ayudas Universitarias y de Investigación” (AGAUR, Spain) for providing my PhD fellowship, and EUROCAROTEN for providing the fellowship for conferences and scientific visiting to London (COST Action CA15136).

I would like to thank Dr. Paul Fraser (Royal Holloway, University of London, UK) for making available UPLC and GC-MS facilities for metabolic analysis and Dr. Margit Drapal (Royal Holloway, University of London, Britain) for her kind help with UPLC and GC-MS data analysis. Dr. Lourdes Gomez-Gomez (University of Castilla-La Mancha, Spain) and Dr. Gianfranco Diretto (Italian National Agency for New Technologies, Italy) for pelargonidin glycosides identification and great contribution to the manuscript. Special thanks to Dr. Chao Bai and Dr. Judith Berman for their brilliant previous work that helped me a lot.

I would like to thank Dr. Ludovic Bassie for helping in techniques, Nuria Gabernet for organizing the official paperwork and Jaume Capell for taking care of all my plants. Also thanks to all my colleagues, Dr. Gemma Farre and Dr. Gemma Massip for helping me getting started in the lab. To Can Baysal for helping me with western blot analysis, confocal microscopy and electron microscopy. To Lucia and Erika for being incomparable lab colleagues. To all other lab colleagues from then and now: Yi,

Andrew, Daniela, Edu, Vicky, Pedro, Jose, Derry, Amaya, Sarai, Jaume and Marco, thank you for all of the assistance and encouragement.

I would like to thank my entire family, my grandma, Mom and Dad, thank you for all your love and supports. All my love and thanks for everything. Words are powerless to express my gratitude.

SUMMARY

This thesis mainly focuses on functional analysis and characterization of a number of secondary metabolite biosynthetic genes and the regulation of the corresponding secondary metabolite biosynthetic pathway in plants. The overall aims were to elucidate the transcriptional regulation of β -carotene hydroxylase 2 gene (*BCH2*) in maize, the functional analysis of rice isopentenyl diphosphate isomerases (*OsIPPI*), and determine their subcellular localization. Simultaneously, the influence of light on the metabolic pathway in rice was studied and the pelargonidin quantification and essential pelargonidin biosynthesis genes in *Gentiana lutea* L. var. *aurantiaca* were identified.

Maize and rice plants were transformed with transcription factor genes *ZmMYB* and *ZmPBF*, via transient gene expression and stable transformation respectively. The results indicated that both *ZmPBF* and *ZmGAMYB* can transactivate *ZmBCH2* expression in maize endosperm and *ZmPBF* and *ZmGAMYB* independently transactivate the *ZmBCH2* promoter in rice. The two *IPPI* paralogs (*OsIPPI1* and *OsIPPI2*) isolated previously in rice had a different expression pattern; *OsIPPI1* mRNA was more abundant than *OsIPPI2* mRNA in all tissues. Confocal fluorescence microscopy and immuno-electron microscopy were used to determine the localization of both proteins. These localized to the endoplasmic reticulum (ER) as well as peroxisomes and mitochondria, whereas only *OsIPPI2* was detected in plastids. The detection of both isoforms in the ER indicates that DMAPP can be synthesized *de novo* in this compartment. UPLC, GC-MS and qRT-PCR were used to profile the primary and secondary metabolites and gene expression in de-etiolated rice seedlings. The results revealed both primary and secondary metabolism and the corresponding genes are regulated by light, especially isoprenoids biosynthesis in rice leaves. Eleven pelargonidin derivatives were identified in the petals of *G. lutea* and the biosynthetic pathway genes were profiled, revealing DFR, ANS and 3GT mainly affect the accumulation of pelargonidin glucosides. Collectively my results provide novel insights of the regulation of carotenoid and flavonoid biosynthesis in higher plants at different levels.

RESUMEN

Esta tesis se centra principalmente en el análisis funcional y en la caracterización de los genes que codifican para algunos metabolitos secundarios y en el estudio de su regulación en las plantas. Los objetivos generales fueron (a) profundizar en el conocimiento de la regulación transcripcional del gen de la biosíntesis de los carotenoides, la β -caroteno hidroxilasa 2 (*BCH2*) en el maíz, y (b) analizar la función de las dos isopentenil difosfato isomerasas (*OsIPPI*) de arroz, determinando además su localización subcelular. Simultáneamente, se estudió cómo la luz afecta a la vía metabólica y a la producción de pelargonidina en el arroz; se identificaron también los genes esenciales de su biosíntesis en *Gentiana lutea* L. var. *aurantiaca*.

Las plantas de maíz y arroz se transformaron con los genes de los factores de transcripción *ZmMYB* y *ZmPBF*. Se analizó la expresión génica transitoria y se realizó transformación estable. Los resultados obtenidos indicaron que tanto *ZmPBF* como *ZmGAMYB* pueden transactivar la expresión de *ZmBCH2* en endospermo de maíz, y *ZmPBF* y *ZmGAMYB* transactivar independientemente el promotor de *ZmBCH2* en arroz. Los dos parálogos de *IPPI* (*OsIPPI1* y *OsIPPI2*) aislados previamente en arroz tuvieron un patrón de expresión diferente; el ARNm de *OsIPPI1* fue más abundante que el ARNm de *OsIPPI2* en todos los tejidos. Se usó la microscopía de fluorescencia confocal y microscopía inmunoelectrónica para determinar la localización de ambas proteínas. Estas se localizan en el retículo endoplásmico (RE), así como en los peroxisomas y las mitocondrias, mientras que solo se detectó *OsIPPI2* en los plastidios. La detección de ambas isoformas en el RE indica que DMAPP se puede sintetizar *de novo* en este compartimiento. Diferentes técnicas como UPLC, GC-MS y qRT-PCR también se utilizaron para perfilar los metabolitos primarios y secundarios y la expresión génica en plántulas de arroz des-etioladas. Los resultados revelaron que los genes involucrados en la en el metabolismo primario y secundario están regulados por la luz, especialmente en la biosíntesis de isoprenoides en hojas de arroz. Once derivados de pelargonidina se identificaron en los pétalos de *G. lutea* y se perfilaron los genes de la vía de biosíntesis, revelando que DFR, ANS y 3GT afectan principalmente a la acumulación de los glucósidos de pelargonidina. Todos estos resultados contribuyen al conocimiento, a diferentes niveles, de la regulación de las rutas biosintéticas de los carotenoides en plantas superiores.

RESUM

Aquesta tesi es centra principalment en l'anàlisi funcional i la caracterització dels gens que codifiquen per a alguns metabòlits secundaris i en l'estudi de la seva regulació en les plantes. Els objectius generals varen ser (a) entendre millor la regulació transcripcional del gen de la biosíntesi dels carotenoids, la β -carotè hidroxilasa 2 (*BCH2*) en el blat de moro i (b) l'anàlisi funcional de les dues isopentenil difosfat isomerasas (*OsIPPI*) d'arròs i determinar la seva localització subcel·lular. Simultàniament, es va estudiar com la llum afecta la via metabòlica i a la producció de pelargonidina en l'arròs; es van identificar també els gens essencials de la seva biosíntesi en *Gentiana lutea* L. var. *aurantiaca*.

Les plantes de blat de moro i arròs es varen transformar amb els gens dels factors de transcripció *ZmMYB* i *ZmPBF*. Es va analitzar l'expressió gènica transitòria i es va realitzar transformació estable. Els resultats obtinguts indiquen que tant *ZmPBF* com *ZmGAMYB* poden transactivar l'expressió de *ZmBCH2* a l'endosperm del blat de moro, i *ZmPBF* i *ZmGAMYB* transactiven independentment el promotor de *ZmBCH2* en arròs. Els dos paràlegs de *IPPI* (*OsIPPI1* i *OsIPPI2*) aïllats prèviament en arròs varen tenir un patró d'expressió diferent; l'ARNm de *OsIPPI1* va ser més abundant que l'ARNm de *OsIPPI2* en tots els teixits. Es va usar la microscòpia de fluorescència confocal i microscòpia immunoelectrònica per determinar la localització de les dues proteïnes. Aquestes es localitzen en el reticle endoplasmàtic (RE), així com en els peroxisomes i les mitocondries, mentre que només es va detectar *OsIPPI2* en els plastidis. La detecció d'ambdues isoformes en el RE indica que DMAPP es pot sintetitzar *de novo* en aquest compartiment. Diferents tècniques com UPLC, GC-MS i qRT-PCR també es varen utilitzar per perfilar els metabòlits primaris i secundaris i l'expressió gènica relacionada en plàntules d'arròs des-etioladas. Els resultats varen revelar que el metabolisme primari i secundari i els gens corresponents estan regulats per la llum, especialment en la biosíntesi d'isoprenoides en fulles d'arròs. Onze derivats de pelargonidina es varen identificar en els pètals de *G. lutea* i es varen perfilar els gens de la seva via de biosíntesi, revelant que DFR, ANS i 3GT afecten principalment a l'acumulació dels glucòsids de pelargonidina. Tots aquests resultats suggereixen la idea que la biosíntesi dels carotenoids en plantes superiors és regulada a diferents nivells.

Table of contents

Acknowledgments.....	I
Summary.....	III
Resumen.....	V
Resum.....	VII
Table of Contents.....	IX
Index of Figures.....	XIII
Index of Tables.....	XVI
List of Abbreviations.....	XVII
Chapter I. General Introduction.....	1
1.1 General Introduction.....	3
1.2 References.....	7
Aims and Objectives.....	13
Chapter II. <i>ZmPBF</i> and <i>ZmGAMYB</i> transcription factors independently transactivate the promoter of the maize (<i>Zea mays</i>) β-carotene hydroxylase 2 gene.....	15
2.0. Abstract.....	17
2.1. Introduction.....	18
2.2. Aims.....	20
2.3. Material and Methods.....	20
2.3.1. Plant material.....	20
2.3.3. Cloning the <i>ZmBCH2</i> promoter sequence.....	21
2.3.4. Cloning and sequencing the <i>ZmPBF</i> and <i>ZmGAMYB</i> cDNAs.....	21
2.3.5. RNA blot analysis.....	21
2.3.6. Vector construction for rice transformation.....	22
2.3.7. Rice transformation.....	23
2.3.8. Yeast one-hybrid analysis.....	24
2.3.9. Vectors for transient expression in maize endosperm and embryo tissue.....	24
2.3.10. Gene expression analysis by quantitative real-time PCR.....	25
2.3.11. Histochemical and fluorimetric GUS assays.....	26
2.3.12. Statistical analysis.....	27
2.4. Results.....	27
2.4.1. Cloning and bioinformatic analysis of the <i>ZmBCH2</i> 5'-flanking region.....	27
2.4.2. Characterization of the 5'-flanking region of <i>ZmBCH2</i> in transgenic rice....	29
2.4.3. Expression patterns of <i>ZmBCH2</i> , <i>ZmPBF</i> and <i>ZmGAMYB</i> in maize B73 leaf, endosperm and embryo tissues.....	31
2.4.4. <i>ZmPBF</i> /P-box and <i>ZmGAMYB</i> /AACA motif interactions.....	32
2.4.5. Transient expression of <i>ZmPBF</i> and/or <i>ZmGAMYB</i> and the impact on <i>ZmBCH2</i> and <i>ZmPSY1</i> expression in maize endosperm and embryo tissues.....	33
2.4.6. Overexpression of <i>ZmPBF</i> and/or <i>ZmGAMYB</i> in transgenic rice plants and the impact on <i>gusA</i> expression and GUS activity in leaf and seed.....	37
2.5. Discussion.....	42

2.6. Conclusions.....	46
2.7. References.....	47
Chapter III. The two isoforms of isopentenyl diphosphate isomerase in rice (<i>Oryza sativa</i>) are targeted to different subcellular compartments and have distinct roles in isoprenoid biosynthesis.....	53
3.0. Abstract.....	55
3.1. Introduction.....	56
3.2. Aims.....	57
3.3. Material and Methods.....	57
3.3.1. Plant material.....	57
3.3.2. RNA isolation and cDNA synthesis.....	58
3.3.3. Cloning and sequencing the <i>OsIPPI1</i> and <i>OsIPPI2</i> cDNAs.....	58
3.3.4. Construction of plasmids.....	58
3.3.5. Functional complementation analysis.....	59
3.3.6. Measurement of β -carotene levels in <i>E. coli</i>	60
3.3.7. Rice transformation.....	60
3.3.8. Protein extraction and western blot analysis.....	60
3.3.9. Gene expression analysis by quantitative real-time RT-PCR.....	61
3.3.10. DNA sequence analysis and prediction of protein subcellular localization.....	62
3.3.11. Confocal microscopy.....	62
3.3.12. Immuno-electron microscopy.....	63
3.3.13. Ultra-high-performance liquid chromatography.....	64
3.3.14. Gas chromatography–mass spectrometry.....	64
3.4. Results.....	65
3.4.1. Cloning and characterization of the <i>OsIPPI1</i> and <i>OsIPPI2</i> cDNAs.....	65
3.4.2. <i>OsIPPI2</i> encodes a functional enzyme.....	67
3.4.3. <i>OsIPPI1</i> mRNA is more abundant than <i>OsIPPI2</i> mRNA in all rice tissues.....	69
3.4.4. <i>OsIPPI1/2</i> and <i>OsHDR1/2</i> show distinct and dynamic expression profiles in 14-day etiolated leaves during de-etiolation with corresponding metabolic changes.....	70
3.4.5. Targeting of IPPI fusion proteins in transgenic rice leaves.....	72
3.5. Discussion.....	79
3.6. Conclusions.....	82
3.7. References.....	83
Chapter IV. Integrated metabolite and gene expression profiling of de-etiolated rice seedlings upon exposure to light.....	89
4.0. Abstract.....	91
4.1. Introduction.....	92
4.2. Aims.....	94
4.3. Material and Methods.....	94
4.3.1. Plant material.....	94
4.3.2. RNA isolation and cDNA synthesis.....	95
4.3.3. Gene expression analysis by quantitative real-time PCR.....	95
4.3.4. Ultra-high-performance liquid chromatography.....	96
4.3.5. Gas chromatography–mass spectrometry.....	96
4.4. Results.....	97

4.4.1 Amino acid responses to continuous WL illumination	97
4.4.2 Sugar and organic acid responses to continuous WL illumination	97
4.4.3 Fatty acid and other metabolite responses to continuous WL illumination.....	100
4.4.4 Isoprenoid responses to continuous WL illumination and isoprenoid biosynthetic gene expression profiles.....	103
4.5. Discussion.....	110
4.6. Conclusions.....	116
4.7. References.....	116
Chapter V. Differential accumulation of pelargonidin glycosides in petals at three different developmental stages of the orange-flowered gentian (<i>Gentiana lutea</i> L. var. <i>aurantiaca</i>).....	123
5.0. Abstract.....	125
5.1. Introduction.....	126
5.2. Aims.....	128
5.3 Material and Methods.....	130
5.3.1. Chemicals.....	130
5.3.2. Plant materials.....	130
5.3.3. Anthocyanin and precursor extraction and identification.....	130
5.3.4. Cloning of UDP-glucose:Flavonoid 5- <i>O</i> -glucosyltransferase (5GT) and anthocyanin 5-aromatic acyltransferase (5AT) gene fragments from petals of <i>G. lutea</i> L. var. <i>aurantiaca</i>	132
5.3.5. qRT-PCR analysis of <i>DFR</i> , <i>ANS</i> , <i>3GT</i> , <i>5GT</i> and <i>5AT</i> expression in leaf, S1, S3 and S5 petals of <i>G. lutea</i> L. var. <i>aurantiaca</i>	134
5.3.6. Bioinformatics analyses.....	135
5.4. Results.....	135
5.4.1 Identification of anthocyanins in petals of three developmental stages of <i>G. lutea</i> L. var. <i>aurantiaca</i>	135
5.4.2. Quantification of pelargonidin derivatives and precursors in petals at three developmental stages.....	140
5.4.3. Pelargonidin glucoside biosynthesis gene expression.....	144
5.4.4. Integration of transcript and metabolite data of gentian phenylpropanoid pathway.....	150
5.5. Discussion.....	153
5.6. Conclusions.....	153
5.7. References.....	154
GENERAL DISCUSSION.....	161
General discussion.....	163
References.....	165
GENERAL CONCLUSIONS.....	169
General conclusions.....	171
Outputs.....	173

Table of contents

Publications.....	175
Participation in congresses.....	175
Oral presentations.....	175
Presentations.....	175

INDEX OF FIGURES

Chapter I

Figure 1.1. Isoprenoid biosynthesis pathways in higher plants.....	5
---	---

Chapter II

Figure 2.1. Sequence of the maize (<i>Zea mays</i> , inbred line B73) b-carotene hydroxylase 2 gene (<i>ZmBCH2</i>) promoter and part of the cDNA region (GenBank KF941345).....	28
--	----

Figure 2.2. Schematic representation of <i>gusA</i> - fusion constructs and histochemical b-glucuronidase (GUS) assays of representative transgenic rice tissues expressing <i>gusA</i> driven by the constitutive CaMV35S promoter (CaMV35SPro- <i>gusA</i>).....	29
--	----

Figure 2.3. Quantitative analysis of β -glucuronidase activity in transgenic rice.....	30
---	----

Figure 2.4. Expression of endogenous <i>ZmBCH2</i> , <i>ZmPBF</i> and <i>ZmGAMYB</i> genes in different tissues of maize (<i>Zea mays</i> , inbred line B73).....	32
---	----

Figure 2.5. Identification and characterization of P-box and AACAA motifs in the maize <i>ZmBCH2</i> basal promoter region which could be recognized and bound by the <i>ZmPBF</i> and <i>ZmGAMYB</i> transcription factors, respectively.....	33
---	----

Figure 2.6. Schematic representation of expression vectors pAHC25, pAL76- <i>ZmGAMYB</i> and pAL76- <i>ZmPBF</i>	34
---	----

Figure 2.7. Expression of <i>ZmBCH2</i> , <i>ZmGAMYB</i> and <i>ZmPBF</i> in maize (<i>Zea mays</i> , inbred line B73) endosperm and embryo tissue.....	34
---	----

Figure 2.8. Sequence of the maize phytoene synthase 1 (<i>ZmPSY1</i>) promoter and part of the cDNA region.....	35
--	----

Figure 2.9. Quantitative RT-PCR analysis of maize phytoene synthase 1 gene (<i>ZmPSY1</i>) expression in endosperm and embryo determined by transient expression.....	37
--	----

Figure 2.10. Relative average expression (normalized to <i>OsACTIN</i>) in rice T3 seeds (25 d after pollination, DAP) and T2 mature flag leaves for two maize transcription factor genes (<i>ZmPBF</i> and <i>ZmGAMYB</i>) and the reporter transgenic lines representing each transgenic event (TE).....	38
--	----

Figure 2.11. Quantitative analysis of b-glucuronidase (GUS) activity in transgenic rice.....	40
---	----

Figure 2.12. Histochemical b-glucuronidase (GUS) assays of representative transgenic rice T0 mature flag leaves and T1 seeds (at 25 d after pollination).....	41
--	----

Chapter III

Figure 3.1. Multiple alignments of predicted amino acid sequences encoding isopentenyl diphosphate isomerases.....	66
---	----

Figure 3.2. Partial structures of rice isopentenyl diphosphate isomerase genes, within the bounds of the open reading frame.....	67
Figure 3.3. Color complementation and enhanced β -carotene accumulation due to the expression of <i>OsIPPI1</i> or <i>OsIPPI2</i> in <i>E. coli</i> engineered for β -carotene biosynthesis.....	68
Figure 3.4. Analysis of <i>OsIPPI1</i> and <i>OsIPPI2</i> gene expression in different rice tissues at different growth stages.....	69
Figure 3.5. Analysis of <i>OsIPPI1/2</i> and <i>OsHDR1/2</i> expression and phytosterol, chlorophyll and carotenoid levels in the etiolated leaves of 14-day-old rice plants during de-etiolation.....	71
Figure 3.6. Fusion gene construction and western blot analysis in T3 homozygous transgenic rice leaves.....	72
Figure 3.7. Confocal laser scanning microscopy of wild-type rice leaves (A-D) and transgenic plants lines expressing <i>OsIPPI1-sGFP</i> (E-H) and <i>OsIPPI2-sGFP</i> (I-L) fusion gene driven by the CaMV35S promoter.....	74
Figure 3.8. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice mitochondria by labeling sGFP.....	75
Figure 3.9. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice peroxisomes by labeling sGFP.....	76
Figure 3.10. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice endoplasmic reticulum by labeling sGFP.....	77
Figure 3.11. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice chloroplasts by labeling sGFP.....	78
 Chapter IV	
Figure 4.1. Relative amino acid contents of WT rice seedlings under different illumination time.....	98
Figure 4.2. Relative sugar and organic acid contents of WT rice seedlings under different illumination time.....	99
Figure 4.3. Relative fatty acid contents of WT rice seedlings under different illumination time...	100
Figure 4.4. Relative TCA cycle metabolite contents of WT rice seedlings under different illumination time.....	101
Figure 4.5. Relative nucleotide and cell wall contents of WT rice seedlings under different illumination time.....	102
Figure 4.6. Tocopherol composition and content in the etiolated leaves of 14-day-old rice plants during de-etiolation. in the same leaves during de-etiolation.....	103
Figure 4.7. Relative expression levels of MVA pathway and MEP pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light.....	105

Figure 4.8. Relative expression levels of carotenoids pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light.....	106
Figure 4.9. Relative expression levels of tocophyreols pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light.....	107
Figure 4.10. Relative expression levels of chlorophyll biosynthesis pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light.....	109
 Chapter V	
Figure 5.1. The gentian phenylpropanoid pathway.....	129
Figure 5.2. Typical HPLC-DAD/UV isoplot chromatograms and HPLC-PDA chromatographic profile of the anthocyanins in <i>Gentiana lutea</i> L. var. <i>aurantiaca</i> petals at the S5 stage of development.....	136
Figure 5.3. ESI-MS/MS spectra of anthocyanins isolated from petals of <i>Gentiana lutea</i> L. var. <i>aurantiaca</i>	138
Figure 5.4. Levels of pelargonidin derivatives identified in leaf and petals of <i>Gentiana lutea</i> L. var. <i>aurantiaca</i> in three different developmental stages by HPLC-ESI-MS/MS analyses.....	141
Figure 5.5. Row-directed Hierarchical clustering (HCL) visualization of anthocyanin metabolites detected in flower petals of <i>G. lutea</i> L. var. <i>aurantiaca</i>	142
Figure 5.6. Levels of anthocyanin precursors identified in petals of <i>Gentiana lutea</i> L. var. <i>aurantiaca</i> in three different developmental stages by HPLC-ESI-MS/MS analyses.....	144
Figure 5.7. Alignments of partial 5GT and 5AT cDNA sequences encoding UDP-glucose: flavonoid 5-O-glucosyltransferase (5GT; A) and anthocyanin 5-aromatic acyltransferase (5AT; B) between <i>Gentiana triflora</i> (Gt) and <i>G. lutea</i> L. var. <i>aurantiaca</i> (Gla).....	145
Figure 5.8. Alignments of the deduced amino acid sequences encoded by UDP glucose: flavonoid 5-O-glucosyltransferase (5GT) gene (A), and anthocyanin 5-aromatic acyltransferase (5AT) gene (B) from <i>Gentiana triflora</i> and <i>G. lutea</i> L. var. <i>aurantiaca</i>	148
Figure 5.9. Quantitative expression of anthocyanin genes, normalized on the ubiquitin housekeeping gene in leaf and petals of <i>Gentiana lutea</i> L. var. <i>aurantiaca</i>	149
Figure 5.10. Integration of transcript-metabolite data involved in <i>G. lutea</i> L. var. <i>aurantiaca</i> anthocyanin metabolism.....	152

INDEX OF TABLES

Chapter II

Table 2.1. Primers used for RNA blot analysis.	22
Table 2.2. Primers used for β -glucuronidase reporter gene vector construction.	23
Table 2.3. Primers used to prepare the constructs pAL76- <i>ZmPBF</i> and pAL76- <i>ZmGAMYB</i>	25
Table 2.4. Primers used for qRT-PCR analysis.	26
Table 2.5. Carotenogenic genes in maize (<i>Zea mays</i>).	36
Table 2.6. Partition of the observed variance in the relative expression of <i>gusA</i> and GUS activity in rice based on standard analysis of variance (ANOVA) and covariance (ANCOVA)....	39

Chapter III

Table 3.1. Primers used for qRT-PCR analysis in this study.....	61
Table 3.2. Prediction of OsIPPI1 and OsIPPI2 subcellular localization.	73

Chapter IV

Table 4.1. List of primers used for the qRT-PCR.....	95
---	----

Chapter V

Table 5.1. Identification of anthocyanins from <i>Gentiana lutea</i> L. var. <i>aurantiaca</i> petals by HPLC-ESI-MS/MS.	137
Table 5.2. Node strength (ns) and network strength (NS) of gentian phenylpropanoid genes and metabolites, expressed as the average of all the rs yielded by a node and the average of the ns, respectively.	151

ABBREVIATIONS

2,4-D	4-dichlorophenoxyacetic acid
3GT	flavonoid 3-O-glucosyltransferase
5-UTR	5' -untranslated region
5AT	anthocyanin 5-aromatic acyltransferase
5GT	flavonoid 5-O-glucosyltransferase
ANCOVA	analysis of covariance
ANOVA	analysis of variance
ANS	anthocyanidin synthase
ATP	adenosine triphosphate
BCH2	β -carotene hydroxylase 2
C4H	4-hydroxylase
CCD1	carotenoid cleavage dioxygenase 1
CDP-ME	4-diphosphocytidyl-2-C-methyl-D-erythritol
CDP-MEP	4-diphospho- cytidyl-2-C-methyl-D-erythritol 2-phosphate
CHI	chalcone isomerase
CHLG	chlorophyll synthase
CMK	CDP-ME kinase
DAP	days after pollination
DFR	dihydroflavonol 4-reductase
DHK	dihydrokaempferol
DHM	dihydromyricetin
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
DW	dry weight
DXP	1-deoxy-D-xylulose 5-phosphate
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
DXS	1-deoxy-D-xylulose-5-phosphate synthase
ER	endoplasmic reticulum
F3'5'H	flavonoid 3',5'-hydroxylase
F3'H	flavonoid 3'-hydroxylase
F3H	flavanone 3-hydroxylase
FPP	farnesyl diphosphate
GABA	methionine, threonine
GC-MS	gas chromatography mass spectrometry
GGPP	geranylgeranyl diphosphate
GGPPS	geranylgeranyl pyrophosphate synthase
GGPS	glucosylglycerol-phosphate synthase
GGR	geranylgeranyl reductase
GPP	geranyl diphosphate
GRP	GGPPS recruiting protein
GUS	β -glucuronidase (GUS)
gusA	β -glucuronidase reporter gene
GWAS	genome-wide association studies
HDR	HMBPP reductase
HDS	1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMBPP) synthase
HGGT	homogentisate geranylgeranyltransferase
HMBPP	4 hydroxy-3-methylbut-2-enyl diphosphate
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase

HMGS	HMG synthase
HPLC	high-performance liquid chromatography
HPLC-ESI-MS/MS	High-performance liquid chromatography-electrospray ionization-tandem mass spectrometry
Hpt	hygromycin phosphotransferase
IPP	isopentenyl diphosphate
IPPI	isopentenyl diphosphate isomerase
LB	lysogeny broth
LCYB	lycopene β -cyclase
LCYE	lycopene ε -cyclase
LIL3	light-harvesting-like protein
MCT	rice MEP cytidyltransferase
MDS	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ME-2,4cPP)
MEcDP	2-C-methyl- D-erythritol 2,4-cyclodiphosphate
MEP	2-C-methyl-D-erythritol 4-phosphate
MK	MVA kinase
mRNA	messenger RNA
MU	4 methylumbelliferone
MVA	mevalonate
MVD	MVA diphosphate decarboxylase
ORF	open reading frame
PBF	P-box binding factor
P-box	prolamin box
PCR	polymerase chain reaction
PDS	phytoene desaturase
PMK	phospho-MVA kinase
POR	protochlorophyllide oxidoreductase
PSY	phytoene synthase
qRT-PCR	quantitative real-time PCR
QTL	quantitative trait locus
RNAi	RNA interference
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
sGFP	synthetic green fluorescent protein
TBST	tris-buffered saline (TBS) and Tween 20
TCA	tricarboxylic acid cycle
TPS	transit peptide sequence
UPLC	ultra high performance liquid chromatography
UV	ultraviolet
VIGS	virus induced gene silencing
VTE1	tocopherol cyclase
VTE2	homogentisate phytyltransferase;
VTE3	MPBQ/MSBQ methyltransferase
VTE4	gamma-tocopherol methyltransferase
WT	wild type
ZDS	zeta-carotene desaturase

Chapter I. GENERAL INTRODUCTION

1. 1. General Introduction

Plants produce an amazing diversity of compounds called primary and secondary metabolites. Primary metabolites, such as amino acids, sugars, and organic acids, are mainly involved in plant growth, development, and reproduction. They are also essential for plant survival. In contrast, secondary metabolites, including isoprenoids and flavonoids, are mainly involved in the stress responses and environmental adaptations of plants. Plant secondary metabolites also have a wide range of uses in the agricultural, agrochemical pharmaceutical and cosmetic industries (Verpoorte *et al.*, 2002; Vasconsuelo *et al.*, 2007). Many biotic and abiotic factors, such as pests, plant diseases, light, temperature, and water affect the biosynthesis of secondary metabolites. The effectiveness of elicitation as a tool to enhance the production of secondary metabolites depends on complex interactions between the elicitor and the plant cells (Vasconsuelo and Boland, 2007).

Isoprenoids (also called terpenoids) are an important group of secondary metabolites, forming the largest family of natural compounds that can be found in all living organisms (Cervantes *et al.*, 2006). Isoprenoids are remarkably diverse in structure and function, especially among diverse phototrophic organisms, like mosses, liverworts, and ferns as well as higher plants (Holstein and Hohl, 2004 ;Seemann *et al.*, 2006; Zhu *et al.*, 2008). In addition, isoprenoids regulate gene expression levels, vitamin synthesis, antimicrobial agent production, mating pheromones and reproductive hormones. They also play important roles in signal transduction pathways, and in electron transport and photosynthetic machinery (Holstein and Hohl, 2004). Isoprenoids have a number of commercial uses as a solvents, pigments, fragrances, flavors, drugs, and nutraceuticals (Vranová *et al.*, 2013).

Higher plants synthesize thousands of isoprenoids that have great effects on plant growth and development, in addition to stress responses (Chappell, 1995; McGarvey and Croteau, 1995; Croteau *et al.*, 2000). In particular, isoprenoids have great effects on photosynthetic processes, including light harvesting, energy conversion, electron transfer, and excited chlorophyll triplet quenching in plants (Gershenzon and Kreis, 1999; Rodriguez-Concepcion and Boronat, 2002; Vranová *et al.*, 2013; Sapir-Mir *et al.*, 2008).

Isoprenoids are derived from two independent and highly conserved pathways, the mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP). The MVA pathway is localized in the cytosol of all eukaryotes, archaea, and some bacteria; and the MEP pathway operates in most bacteria and in the plastids of plants (Vranova *et al.*, 2013). Some bacteria and plants have the ability to use both pathways (Vranova *et al.*, 2013). All the isoprenoid products are derived from the same precursors, the five-carbon building units isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are interconverted by the enzyme isopentenyl diphosphate isomerase (IPPI) (Rodriguez-Concepcion *et al.*, 2004).

Higher plants harbor both the MVA and the MEP pathway, the former in the cytosol and the latter in plastids, thus allowing the optimization and regulation of isoprenoid biosynthesis according to the availability of fixed carbon and ATP (Rodriguez-Concepcion *et al.*, 2004; Vranova *et al.*, 2013). IPP or IPP-derived products can be exchanged between the cytoplasm and plastids in plants (Nagata *et al.*, 2002). Some isoprenoids in plants are derived only from the MVA pathway, including phytosterols (which modulate the membrane architecture and plant growth and developmental processes), sesquiterpenes, triterpenes, and the side chain of ubiquinone (which is formed from IPP synthesized in the cytosol and imported into the mitochondria). In addition to the monoterpenes, diterpenes, carotenoids, and the side chains of chlorophylls, tocopherols, and hormones (including gibberellins and abscisic acid) are all derived from the MEP pathway (Rodriguez-Concepcion *et al.*, 2004; Vranova *et al.*, 2013; Zhu *et al.*, 2013; Nogueira *et al.*, 2018).

The synthesis of isoprenoids in plants begins with the head-to-tail condensation of DMAPP and one or more IPP molecules. In the MVA pathway, IPP is synthesized from acetyl-CoA, which is the key intermediate of this metabolic route (Bloch, 1992), through seven steps, to form DMAPP. In the MEP pathway, IPP and DMAPP are carbohydrate derivatives, with their synthesis starting from pyruvate and D-glyceraldehyde phosphate via 1-deoxy-D-xylulose 5-phosphate, MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol, 4-diphospho-cytidyl-2-C-methyl-D-erythritol 2-phosphate, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate, and 4-hydroxy-2-methylbut-2-enyl 1-phosphate (HMBPP) (Seemann, 2006). DMAPP is primarily used as a chemically active substrate and is extended by adding IPP units to form short-chain

prenyl diphosphates, such as geranyl diphosphate, farnesyl diphosphate, and geranylgeranyl diphosphate. In the cytosol, DMAPP is produced from IPP by IPPI, without which the MVA pathway is blocked. In the plastids, however, HMBPP is converted into a mixture of IPP and DMAPP by HMBPP reductase (Vranova *et al.*, 2013). IPPI is the only enzyme that features in both the MVA and MEP pathways and, in this context, it can be used to regulate the levels of IPP and DMAPP in multiple subcellular compartments (Okada *et al.*, 2008; Pankratov *et al.*, 2016).

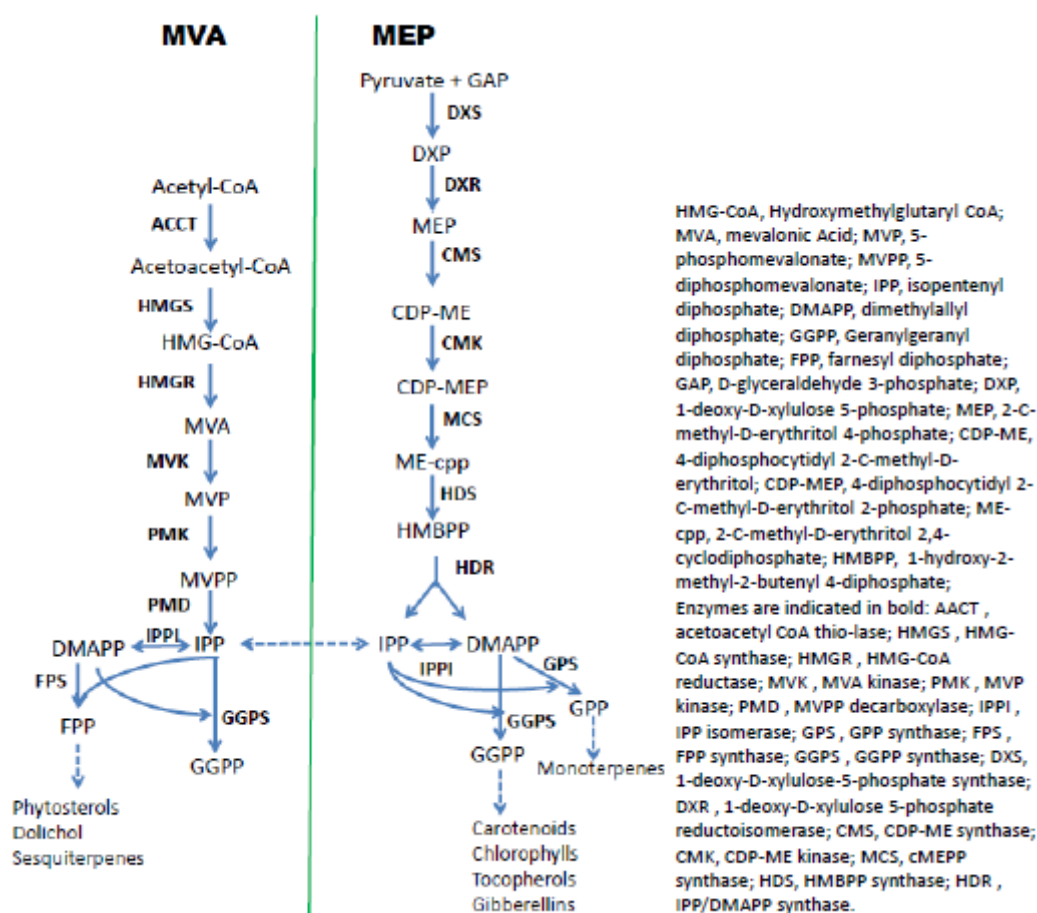


Figure 1.1 Isoprenoid biosynthesis pathways in higher plants.

Carotenoids form a very important group of isoprenoids. In humans and animals, various carotenoids derived from plant sources act as antioxidants and protect against important diseases, while other carotenoids are precursors of vitamin A and retinoid compounds, which are essential for a variety of critical functions in humans. In plants, carotenoids not only form pigments, but they also have essential roles in photosynthesis, the regulation of growth and development and anti-stress responses (Bai *et al.*, 2016).

Phytosterols, another very important group of isoprenoid compounds, regulate plasma membrane fluidity and permeability in eukaryotes, dolichols and long chain polyprenols, acting as carbohydrate carriers, the prenyl chains of quinones from electron transport chains (ubiquinones, men-aquinones, and plastoquinone), and carotenoids and phytol from chloroplasts, as well as a multitude of secondary metabolites and many medically important compounds, such as antioxidants, and anticancer and antimalarial drugs (Seemann, 2006).

Flavonoids, a group of compounds enriched in vegetables and fruits, are important in UV-protection, pollinator attraction, and seed dispersal. For human health, flavonoids are beneficial for the prevention of cardiovascular (Gaziano *et al.*, 1995), flavonoid inflammatory, and other diseases (Middleton *et al.*, 1992; Noda *et al.*, 2002). Flavonoid biosynthesis starts with the amino acid phenylalanine and the end products include anthocyanins, flavones/isoflavones (Patra *et al.*, 2013). Anthocyanidins are an important group of flavonoids. Despite the wide range of anthocyanin' structures, most of them are biosynthesized from three kinds of anthocyanidins: pelargonidin, cyanidin, and delphinidin (Nakamura *et al.*, 2010). Previous studies already revealed several essential enzymes involved in the pelargonidin biosynthesis. Dihydroflavonol 4-reductase is a key enzyme in anthocyanin biosynthesis and flower color. It catalyzes dihydrokaempferol to leucopelargonidin in petunia mutants that are deficient in F3' H and F3' 5' H, resulting in pelargonidin production and orange color (Elomaa *et al.*, 1995; Tanaka *et al.*, 1995; Nakamura *et al.*, 2010).

1.2. References

Bai C, Capell T, Berman J, Medina V, Sandmann G, Christou P, Zhu C (2016). Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor-product balance. *Plant Biotechnol. J.*, **14**:195-205.

Bloch K (1992) Sterol molecule: structure, biosynthesis, function. *Steroids*, **57**:378–383.

Cervantes-Cervantes M, Gallagher C E, Zhu C, Wurtzel T E (2006). Maize cDNAs Expressed in Endosperm Encode Functional Farnesyl Diphosphate Synthase with Geranylgeranyl Diphosphate Synthase Activity. *Plant Physiol.*, **141**:220-231.

Chappell J (1995). Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**:521-547.

Croteau R, Kutcahn TM, Lewis NG (2000). Natural products. In *Biochemistry and Molecular Biology of Plants* (B. Buchanan, W. Gruissem and R. Jones, eds). Rockville, MD: American Society of Plant Physiologists, pp. 1250– 1318.

Elomaa O, Helariutta Y, Griesbach RJ, Kotilainen M, Seppanen P, Teeri TH (1995) Transgenic inactivation in *Petunia hybrida* is influenced by the properties of the foreign gene. *Mol. Gen. Genet.*, **248**:645–649.

Gaziano JM, Manson JE, Branch LG, Colditz GA, Willett WC, Buring JEA (1995) Prospective study of consumption of carotenoids in fruits and vegetables and decreased cardiovascular mortality in the elderly. *Ann. Epidemiol.*, **5**:255- 260.

Gershenzon J, Kreis W (1999) Biosynthesis of monoterpenes, sesquiterpenes, diterpenes, sterols, cardiac glycosides and steroid saponins. In: M Wink, ed, *Biochemistry of Plant Secondary Metabolism, Annual Plant Reviews*, Vol 2. Sheffield Academic Press, Sheffield, UK, pp 222–229.

Holstein SA, Hohl RJ (2004). Isoprenoids: Remarkable diversity of form and function. *Lipids*, **39** (4):293–309.

Jiao Y, Lau OS, Deng XW (2007). Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics*, **8** (3):217–230.

Jung KH, Lee J, Dardick C, Seo YS, Cao P, Canlas P, Phetsom J, Xu X, Ouyang S, An K, Cho YJ, Lee GC, Lee Y, An G., Ronald PC (2008) Identification and functional analysis of light-responsive unique genes and gene family members in rice. *PLoS Genetics*, **4**(8):e1000164.

McGarvey DJ, Croteau R (1995). Terpenoid metabolism. *Plant Cell*, **7**:1015-1026.

Mergham O (2017). The Biochemistry and Antioxidant Properties of Carotenoids, Carotenoids, Dragan J. Cvetkovic and Goran S. Nikolic (eds.), *IntechOpen*, DOI: 10.5772/67592. Available from: <https://www.intechopen.com/books/carotenoids/the-biochemistry-and-antioxidant-properties-of-carotenoids>.

Middleton E, Kandaswani C (1992) Effects of flavonoids on immune and inflammatory cell functions. *Biochem. Pharmacol.*, **43**:1167-1179.

Nagata N, Suzuki M, Yoshida S, Muranaka T (2002). Mevalonic acid partially restores chloroplast and etioplast development in *Arabidopsis* lacking the non-mevalonate pathway. *Planta*, **216** (2):345–350.

Nakamura N, Fukuchi-Mizutani M, Fukui Y, Ishiguro K, Suzuki K, Suzuki H, Okazaki K, Shibata D, Tanaka Y (2010). Generation of pink flower varieties from blue *Torenia hybrida* by redirecting the flavonoid biosynthetic pathway from delphinidin to pelargonidin. *Plant Biotechnology*, **27**:375-383.

Noda Y, Kaneyuki T, Mori A, Packer L (2002) Antioxidant activities of pomegranate fruit extract and its anthocyanidins: delphinidin, cyanidin, and pelargonidin. *J. Agric. Food. Chem.*, **50** (1):166-71.

Nogueira M, Enfissi EM, Almeida J, Fraser PD (2018) Creating plant molecular factories for industrial and nutritional isoprenoid production. *Curr. Opin. Biotechnol.*, **49**:80–87.

Rodriguez-Concepcion M, Boronat A (2002) Elucidation of the methyl- erythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids: a metabolic milestone achieved through genomics. *Plant Physiol.*, **130**:1079–1089.

Rodriguez-Concepcion M, Forés, O., Martínez-García, J. F., González, V., Phillips, M. A., Ferrer, A., Boronat, A. (2004). Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during *Arabidopsis* seedling

development. *The Plant Cell*, **16**(1):144–156.

Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, Gidoni D, Eyal Y (2008). *Plant Physiology*, **148**:1219–1228.

Seemann M, Tse Sum Bui B, Wolff M, Miginiac-Maslow M, Rohmer M (2006). Isoprenoid biosynthesis in plant chloroplasts via the MEP pathway: Direct thylakoid/ferredoxin-dependent photoreduction of GcpE/IspG. *FEBS Letters*, **580** (6), 1547–1552.

Tanaka Y, Fukui Y, Fukuchi-Mizutani M, Holton TA, Higgins E, Kusumi T (1995) Molecular cloning and characterization of *Rosa hybrida* dihydroflavonol 4-reductase. *Plant Cell. Physiol.*, **36**:1023–1031.

Vasconsuelo A, Boland R (2007). Molecular aspects of the early stages of elicitation of secondary metabolites in plants. *Plant Science*, **172**:861-875.

Verpoorte R, Contin A, Memelink J (2002). Biotechnology for the production of plant secondary metabolites. *Phytochemistry Reviews*, **1** (1):13–25.

Vranová E, Coman D, Gruissem W (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Ann. Rev. Plant Biol.*, **64**. 10.1146/annurev-arplant-050312-120116.

Yang P, Chen H, Liang Y, Shen S (2007). Proteomic analysis of de-etiolated rice seedlings upon exposure to light. *Plant Proteomics*, **7** (14):2459-2468.

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, and Capella T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *PNAS*, **105**: 18232–18237.

Zhu C, Sanahuja G, Yuan D, Farre G, Arjo G, Berman J, Zorrilla-Lopez U, Banakar R, Bai C, Perez-Massot E, Bassie L, Capell T, Christou P (2013) Biofortification of plants with altered antioxidant content and composition: genetic engineering strategies. *Plant Biotechnol. J.*, **11**:129–141.

AIMS AND OBJECTIVES

Aims

The major aims of my research were to better understand the secondary metabolite biosynthesis and their regulation in higher plants, and how the light affects primary and secondary metabolite biosynthesis and their accumulation, establishing the expression patterns of involved genes.

Objectives

In order to reach these aims, the following specific objectives were:

1. To test whether the transcription factors ZmPBF and ZmGAMYB can transactivate the expression of *ZmBCH2* in maize and activate the minimal promoter region of *ZmBCH2* in transgenic rice plants.
2. To analyze the expression patterns and subcellular localizations of two OsIPPI and further reveal the role of *OsIPPI* genes in isoprenoid biosynthesis in rice.
3. To gain insights into the light-induced metabolites and involved genes, establishing a general metabolic profile and gene expression profile during rice de-etiolation.
4. To identify and characterize the anthocyanin molecules of the pelargonidin glycosides, and their regulation in orange-flowered *G. lutea* L. var. *aurantiaca*.

Aims and Objectives

Chapter II

ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (*Zea mays*) β -carotene hydroxylase 2 gene

2.0. Abstract

The maize (*Zea mays*) enzyme β -carotene hydroxylase 2 (*ZmBCH2*) controls key steps in the conversion of β -carotene to zeaxanthin in the endosperm. The *ZmBCH2* has an endosperm-preferred and developmentally regulated expression profile but the detailed regulatory mechanism is unknown. In order to gain insights into the regulation of *ZmBCH2*, I isolated 2036 bp of the 5'-flanking region which contains the 263-bp 5'-untranslated region (5'-UTR) including the first intron. I fused this to the β -glucuronidase reporter gene, *gusA*. I found that high-level expression of *gusA* in rice seeds requires the 5'-UTR for enhanced activation. Truncated variants of the *ZmBCH2* promoter retained their seed-preferred expression profile as long as a prolamin box and AACA motif were present. I identified candidate genes encoding the corresponding transcription factors (*ZmPBF* and *ZmGAMYB*) and confirmed that their spatiotemporal expression profiles are similar to *ZmBCH2*. Both *ZmPBF* and *ZmGAMYB* can transactivate *ZmBCH2* expression in maize endosperm. To eliminate potential confounding effects in maize, I characterized the regulation of the minimal promoter region of *ZmBCH2* in transgenic rice. This revealed that *ZmPBF* and *ZmGAMYB* independently transactivate the *ZmBCH2* promoter. The mechanism that underpins my data provides an exciting new strategy for the control of carotenoid gene expression in engineered plants.

2.1. Introduction

Carotenoids play a fundamental role in human nutrition as antioxidants and precursors of vitamin A (Fraser and Bramley, 2004). Plants are good sources of dietary carotenoids, which accumulate in various plastids such as chloroplasts, amyloplasts, and chromoplasts (Bai *et al.*, 2016; Giuliano, 2017). Genes encoding carotenogenic enzymes have therefore been used for metabolic engineering to increase the nutritional value and health-promoting properties of plants by improving the carotenoid content and composition (Farré *et al.*, 2014; Giuliano, 2017). However, the regulation of endogenous carotenogenic gene expression in higher plants is poorly understood, restricting the extent to which the impact of metabolic engineering in staple crops can be predicted (Giuliano *et al.*, 2008; Farré *et al.*, 2014).

Maize (*Zea mays* L.) kernels contain low levels of provitamin A (β -carotene, α -carotene and β -cryptoxanthin), reaching only 1–10% of the abundance of the non-provitamin A carotenoids zeaxanthin and lutein (Harjes *et al.*, 2008). The carotenoid biosynthesis pathway is well characterized in maize endosperm. Maize phytoene synthase 1 (*ZmPSY1*) controls the flux of precursors into the carotenoid pathway by catalyzing the first committed step (Zhu *et al.*, 2008). At a later branch point, lycopene ϵ -cyclase (LCYE) directs flux towards α -carotene whereas lycopene β -cyclase (LCYB) directs flux towards β -carotene, the most active form of provitamin A (Harjes *et al.*, 2008). The levels of β -carotene in maize depend not only on the synthesis steps but also on the activity of non-heme di-iron β -carotene hydroxylase 2 (*ZmBCH2*, also known as *ZmcrRB1* and *ZmHYD3*), which is preferentially expressed in the endosperm (Li *et al.*, 2010) and converts β -carotene first to β -cryptoxanthin and then to zeaxanthin (Vallabhaneni *et al.*, 2009; Li *et al.*, 2010; Yan *et al.*, 2010). *ZmBCH2* alleles associated with reduced expression correlate with higher β -carotene levels in maize endosperm (Vallabhaneni *et al.*, 2009; Yan *et al.*, 2010). Quantitative trait locus (QTL) mapping, genome-wide association studies (GWAS) and metabolic data sorting show that *ZmBCH2* controls key steps in the conversion of β -carotene to zeaxanthin in maize endosperm (Vallabhaneni *et al.*, 2009; Yan *et al.*, 2010). Furthermore, we recently demonstrated that the β -carotene content of maize endosperm increased substantially in hybrids in which *ZmBCH2* was silenced by RNA interference (RNAi), whether

ZmBCH1 (also known as *ZmcrtRB3* and *ZmHYD4*) was silenced simultaneously (Berman *et al.*, 2017).

The regulation of *ZmBCH2* expression in the endosperm is not well understood but it is important to unravel the molecular mechanisms responsible for the endosperm-preferred expression profile in order to develop rational strategies to engineer carotenoid biosynthesis in a more predictable manner. Several consensus regulatory sequences have been reported in storage protein genes, which are also expressed in an endosperm-preferred manner, including the prolamin box (P-box), GCN4 motif and AACA motif located approximately 300bp upstream of the transcriptional start site (Colot *et al.*, 1987; Mena *et al.*, 1998; Washida *et al.*, 1999; Wu *et al.*, 2000). The P-box is recognized by the P-box binding factor (PBF) which regulates many storage-protein genes in cereals (Vicente-Carbajosa *et al.*, 1997; Marzábal *et al.*, 2008). Maize PBF orthologs have been isolated from wheat (WPBF), barley (BPBF) and rice (RPBF), and all are transcriptional activators of endosperm gene expression (Mena *et al.*, 1998; Diaz *et al.*, 2002, 2005; Yamamoto *et al.*, 2006). The AACA motif is recognized by GAMYB, an R2R3-class MYB transcription factor expressed in cereal aleurone cells and endosperm in response to gibberellin (Suzuki *et al.*, 1998; Diaz *et al.*, 2002, 2005). The two AACA motifs in the rice glutelin promoter are recognized by the OsMYB5 protein *in vitro*, suggesting this is a transcriptional regulator of glutelin expression in planta (Suzuki *et al.*, 1998). Similarly, the barley transcription factor GAMYB binds AACA motifs to transactivate the native genes encoding B-hordein, the trypsin inhibitor BTI-CMe, and α -amylase in aleurone cells (Gubler *et al.*, 1995, 1999).

To investigate the regulation of *ZmBCH2* expression in detail, I determined whether the 5'-untranslated region (5'-UTR) enhanced β -glucuronidase (GUS) reporter gene (*gusA*) expression in transgenic rice and identified minimal *ZmBCH2* promoter regions responsible for endosperm-preferred expression. To specifically determine the effects of putative *cis*-regulatory elements in the minimal *ZmBCH2* promoter, such as the P-box and AACA motif present in the first 300bp, I overexpressed *ZmPBF* and/or *ZmGAMYB* in maize (transient expression assays in endosperm and embryo) and transgenic rice plants together with *gusA* under the control of the minimal *ZmBCH2* promoter, and investigated the impact of expression of *ZmPBF* and/or *ZmGAMYB* on *ZmBCH2* promoter activity in the leaf and seed.

2.2. Aims

The major aim of the work described in this chapter was to test whether the transcription factors *ZmPBF* and *ZmGAMYB* can transactivate the expression of *ZmBCH2* in maize and activate the minimal promoter region of *ZmBCH2* in transgenic rice plants. A further aim was to gain insights into the mechanism of *ZmBCH2* expression, regulated by *ZmPBF* and *ZmGAMYB* and offer new strategies for the control of carotenoid gene expression in engineered plants.

2.3. Material and Methods

2.3.1. Plant material

Maize (*Zea mays* L.) inbred line B73, wild-type rice (*Oryza sativa* L. cv. EYI105) and transgenic rice plants were grown in a greenhouse or controlled growth chamber at 28°C:20°C day:night temperature with a 10 h photoperiod and 60–90% relative humidity for the first 50 days, followed by maintenance at 21°C:18°C day:night temperature with a 16 h photoperiod thereafter (the best conditions for rice and maize cultivation in our geographical location). Plants were self-pollinated to obtain seeds. Mature leaf, seed, embryo and endosperm tissues were frozen in liquid nitrogen and stored at –80°C.

2.3.2. Nucleic acid isolation and cDNA synthesis

Genomic DNA was extracted from 5g of leaf tissue as described by Sambrook *et al.* (1989). Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, and USA) and DNA was removed with DNase I (RNase-Free DNase Set, Qiagen). Total RNA was quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and 1µg of total RNA was used as the template for first-strand cDNA synthesis with Ominiscript Reverse Transcriptase in a 20µl total reaction volume according to the manufacturer's recommendations (Qiagen).

2.3.3. Cloning the *ZmBCH2* promoter sequence

The *ZmBCH2* cDNA sequence (GenBank AY844958) was used as a blastn query against the MaizeGDB (Maize Genetics and Genomics Database) B73 RefGen_v2 (MGSC). The 5'-flanking region of *ZmBCH2* was cloned from maize B73 leaf genomic DNA by genomic PCR using forward primer 5'-GCT TAT GGT GGT ACG TGC CTG ATG ACT GAC GAT G-3' (positions -1773 to -1740, +1 assigned to the first nucleotide of the longest *ZmBCH2* cDNA) and reverse primer 5'-CTT GGC GAC GAA GCT GGT CAT CGC GGC GGC CAT-3' (positions 264 to 296). The genomic PCR products were transferred to vector PCR II TOPO (Invitrogen, Carlsbad, CA, USA) for sequencing using the Big Dye Terminator v3.1 Cycle Sequencing Kit on a 3130x1 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

2.3.4. Cloning and sequencing the *ZmPBF* and *ZmGAMYB* cDNAs

The maize *PBF* cDNA sequence (GenBank ZMU82230) was used as a query to search the B73 RefGen_v2 and NCBI GenBank databases, allowing us to retrieve a cDNA sequence containing the corresponding full-length open reading frame (ORF) representing maize inbred line B73 (GenBank JQ723698). The full-length *ZmPBF* cDNA was then isolated from a maize B73 endosperm sample at mixed developmental time points of 15, 20, 25 and 30 days after pollination (DAP). Similarly, the barley *GAMYB* cDNA sequence (GenBank X87690) was used to retrieve full-length *ZmGAMYB* cDNA (GenBank BT062692) from the same tissue. The amplified cDNAs were transferred to vector PCR II TOPO as above for sequencing.

2.3.5. RNA blot analysis

Total RNA was extracted from maize endosperm, embryo and leaf tissues and mRNA blotting was carried out as described by Li *et al.* (2010). The forward and reverse primers used to generate the probes are listed in Supporting Information Table 2.1.

Table 2.1. Primers used for RNA blot analysis.

Gene	Forward sequence (5'-3')	Reverse sequence (5'-3')
<i>ZmBCH2</i>	GCTTGTTAGCAGTCCGGTGAGTGAA	GAAAGGAAGATGGCGATAGATGTA
<i>ZmGAMYB</i>	CTCTCTCCTCCTCAAACTCTCTCTCC	CGCGCGCTGGCGGACGCGCGCTGGCGGA
<i>ZmPBF</i>	GGAAGCGTTGCTGCTATAAATAATGCAA	GATGATCCAACCAAATGATTAATTATCCA
<i>ZmPSY1</i>	TGCTGTTCTGCTGACTGGTCTCAC	CGTTGCGTGGCATGTCCATGTCCT

ZmBCH2, maize (*Zea mays*) β -carotene hydroxylase 2 gene; *ZmGAMYB*, maize GAMYB gene; *ZmPBF*, maize P-box binding factor gene; *ZmPSY1*, maize phytoene synthase 1 gene.

2.3.6. Vector construction for rice transformation

Vectors pBI101 and pBI121 (Clontech Laboratories, Mountain View, CA, USA) were digested with HindIII and EcoRI, and then the *gusA*-TNos fragment from pBI101 and the CaMV35SPro-*gusA*-TNos fragment from pBI121 were individually transferred to pUC8 to generate pUC8-*gusA*-TNos and pUC8-CaMV35S-*gusA*-TNos, respectively. The 5'-flanking region of *ZmBCH2* was transcriptionally fused to the *gusA* reporter gene in vector pUC8-*gusA*-TNos by amplifying each 5'-flanking region using primers containing additional sequences to provide appropriate restriction sites (Table 2.2). Forward primer BCH2F1 and reverse primer BCH2R1 were used to amplify the 2036-bp *ZmBCH2* 5'-flanking region containing a 1773-bp promoter fragment (abbreviated as PF in the vector names) and the 263-bp 5'-UTR. The resulting 2036bp DNA fragment was then transferred to the PCRII TOPO vector using the Invitrogen TA Cloning kit to yield intermediate vector pCR-ZmBCH2PF-5'UTR. The pCR-ZmBCH2PF-5'UTR and pUC8-*gusA*-TNos vectors were digested with HindIII and BamHI, allowing the ZmBCH2PF-5'UTR fragment to be inserted upstream of *gusA*, yielding the final construct pUC8-ZmBCH2PF-5'UTR-*gusA* (PF-5'UTR-*gusA*, -1773 to +263). The construct containing the *ZmBCH2* promoter fragment (1773 bp) upstream of the longest cDNA was similarly constructed using the specific forward primer BCH2F1 and reverse primer BCH2R2 (positions -27 to -1) (Table 2.2) to generate pUC8-ZmBCH2PF-*gusA* (PF-*gusA*, -1773 to -1). Four 5' deletions of the *ZmBCH2* promoter region were also generated by PCR, containing upstream promoter fragments and the 5'-UTR. Forward primers BCH2FD1 (positions -1271 to -1247), BCH2FD2 (positions -1264 to -1241),

BCH2FD3 (positions –518 to –494) and finally BCH2FD4 (positions –511 to –488) (Table 2.2) were combined with reverse primer BCH2R1 to generate the four stepwise deletion constructs with HindIII and BamHI restriction sites. These plasmids were named pUC8-ZmBCH2PFΔ1-5'UTR-*gusA* (PFΔ1-5'UTR-*gusA*, –1271 to +263), pUC8-ZmBCH2PFΔ2-5'UTR-*gusA* (PFΔ2-5'UTR-*gusA*, –1264 to +263), pUC8-ZmBCH2PFΔ3-5'UTR-*gusA* (PFΔ3-5'UTR-*gusA*, –518 to +263) and pUC8-ZmBCH2PFΔ4-5'UTR-*gusA* (PFΔ4-5'UTR-*gusA*, –511 to +263), respectively. The synthesized ZmBCH2PFΔ4-5'UTR fragments (–511 to +263) with a mutated P-box (mP, 5'-TGAAAAG-3' to 5'-TGgAgAG-3') or mutated AACA motif (mA, 5'-AACAAAC-3' to 5'-gAgAAcC-3') were transferred to the pUC8-*gusA*-TNos vector at the HindIII and BamHI sites to generate pUC8-ZmBCH2PFΔ4-mP-5'UTR-*gusA* (PFΔ4-mP-5'UTR-*gusA*), and pUC8-ZmBCH2PFΔ4-mA-5'UTR-*gusA* (PFΔ4-mA-5'UTR-*gusA*), respectively. The *gusA* reporter gene driven by the Cauliflower mosaic virus 35S (CaMV35S) promoter in pUC8 (pUC8-CaMV35SPro-*gusA*-TNos, CaMV35SPro-*gusA*) was also introduced into rice as a positive control plasmid. The integrity of all intermediate and final constructs was confirmed by sequencing.

Table 2.2. Primers used for β-glucuronidase reporter gene vector construction.

Primer	Oligonucleotide sequence (5'-3')
BCH2F1	<u>AAGCTT</u> GCTTATGGTGGTACGTGCCTGATGACTG
BCH2R1	<u>GGATCC</u> GCCCCTGGGCTGGTGGTATTTGTGACG
BCH2R2	<u>GGATCC</u> TGGCGGAATTGGAATGTGCTTCCCCTT
BCH2FD1	<u>AAGCTT</u> CCTTACATGTGCTAGCTCTTAGCTA
BCH2FD2	<u>AAGCTT</u> TGTGCTAGCTCTTAGCTATTGAGA
BCH2FD3	<u>AAGCTT</u> TGTAAAGAACGAACATAGAGAATAG
BCH2FD4	<u>AAGCTT</u> AACGAACATAGAGAATAGATATAG

BCH2, β-carotene hydroxylase 2; F1, Forward primer 1; R2, reverse primer 2; FD, forward deletion primer; BamHI (5'-GGATCC-3') or HindIII (5'-AAGCTT-3') sites (underlined) were introduced into the 5' end of each primer, respectively.

2.3.7. Rice transformation

Seven-day-old mature zygotic rice embryos were transferred to MS osmoticum medium [4.4g l⁻¹ Murashige-Skoog salts (Sigma-Aldrich) supplemented with 0.3g l⁻¹ casein hydrolysate, 0.5g l⁻¹ proline, 72.8g l⁻¹ mannitol and 30g l⁻¹ sucrose] 4h before transformation and then bombarded with 0.4mg gold particles coated with the transgene construct(s) and a plasmid containing hygromycin phosphotransferase (*hpt*) selectable

maker gene at a 3:1 (for one transgene), 3:3:1 (for two transgenes) or 3:3:3:1 (for three transgenes) molar ratio (Christou *et al.*, 1991). The embryos were returned to osmoticum medium for 24h before selection on MS medium (as above but without mannitol) supplemented with 30mg l⁻¹ hygromycin and 2.5mg l⁻¹ 2,4-dichlorophenoxyacetic acid (2,4-D) in the dark for 2–3 weeks. At this stage, callus appeared and this was transferred to fresh medium under selection for another 4-6 weeks. Then resistant callus was transferred to regeneration medium (MS basal medium supplemented with 2 mg l⁻¹ BAP, 0.5 mg l⁻¹ NAA, and 30mg l⁻¹ hygromycin) for minimum 4 weeks. After transgenic plantlets were transferred to rooting medium (1/2 MS medium) for 2 weeks and hardened off in soil.

2.3.8. Yeast one-hybrid analysis

The bait sequences from pUC8-ZmBCH2PFA4-5'UTR-*gusA*, pUC8-ZmBCH2PFA4-mP-5'UTR-*gusA*, and pUC8-ZmBCH2PFA4-mA-5'UTR-*gusA*, respectively, were inserted into the pAbAi vector at the HindIII and XmaI sites to form the bait vectors, which were then introduced into yeast strain Y1HGold using Yeastmaker Yeast Transformation System 2 (Clontech). To determine the minimal inhibitory concentration of aureobasidin A (AbA), the bait strains were separately screened on synthetic defined (SD) medium lacking uracil and containing different concentrations of AbA. The coding sequences of *ZmPBF* and *ZmGAMYB* were introduced into vector pGADT7 at the EcoRI and BamHI sites, respectively, to generate the prey vectors pGADT7-*ZmPBF* and pGAT7-*ZmGAMYB*. The prey vector was introduced into the bait strains and grown on SD medium lacking leucine and containing 200ng ml⁻¹ AbA for 72 h at 30°C. The integrity of all intermediate and final constructs was confirmed by sequencing.

2.3.9. Vectors for transient expression in maize endosperm and embryo tissue

The full-length ORF sequences for *ZmPBF* and *ZmGAMYB* were amplified by PCR using the primer combinations listed in Table 2.3. After sequence verification, both cDNAs were transferred to the expression vector pAL76, containing the maize *ubiquitin-1* promoter plus first intron and *nos* transcriptional terminator. The final constructs were verified by sequencing.

Table 2.3. Primers used to prepare the constructs pAL76-*ZmPBF* and pAL76-*ZmGAMYB*

Gene	Forward sequence (5'-3')	Reverse sequence (5'-3')
<i>ZmPBF</i>	<u>GGATCC</u> ATGGACATGATCTCCGGCAGCACT GCAGCA	<u>AAGCTT</u> GCATTATTTATAGCAGCAACGC TTCCACG
<i>ZmGAMYB</i>	<u>GGATCC</u> ATGTACCGGGTGAAGAGCGAGGG GGA	<u>CCCGGGT</u> CATTTGAATCCCGCCATCTG AAAAG

ZmPBF, maize (*Zea mays*) P-box binding factor gene; *ZmGAMYB*, maize *GAMYB* gene; BamHI (5'-GGATCC-3'), HindIII (5'-AAGCTT-3') or XmaI (5'-CCCGGG-3') sites (underlined) were introduced into the 5' ends of each primer, respectively.

Maize B73 kernels at 10 and 25 DAP were sterilized and the endosperm was separated from the embryo. Tangential endosperm sections were prepared to expose a large area of subaleurone endosperm cells. After dissection, the tissues were placed in sterile Petri dishes on N6 osmoticum medium (Duchefa, Haarlem, Netherlands) and were transformed by particle bombardment as previously described (Zhu *et al.*, 2008) using 0.70µg of pAHC25 (Ubi-*gusA*), 0.63µg of pAL76-*ZmPBF* (Ubi-*ZmPBF*) and/or 0.70µg of pAL76-*ZmGAMYB* (Ubi-*ZmGAMYB*). The pAHC25 plasmid carrying *gusA* driven by the maize *ubiquitin-1* promoter (Ubi-*gusA*) was used as an internal control. All samples were bombarded twice and subsequently returned to osmoticum medium for 24h at 25°C in the dark before freezing in liquid nitrogen.

2.3.10. Gene expression analysis by quantitative real-time PCR

First-strand cDNA was synthesized from 1µg total RNA using Ominiscript Reverse Transcriptase in a 20µl total reaction volume according to the manufacturer's recommendations (Qiagen). Quantitative real-time PCR (qRT-PCR) was performed on a CFX96 system (Bio-Rad Laboratories, Hercules, CA, USA) using a 25µl mixture containing 10 ng cDNA, 1x iQ SYBR Green Supermix (Bio-Rad) and 0.2µM of each primer. To calculate relative expression levels, serial dilutions (0.2–125ng) were used to produce standard curves for each gene. Triplicate PCRs in 96-well optical reaction plates were carried out with the following profile: a heating step for 3 min at 95°C was followed by 40 cycles of 95°C for 10s, 59°C for 30s, and 72°C for 20s. Amplification specificity was confirmed by melt curve analysis of the final PCR products in the temperature range 50–90°C with fluorescence acquired after each 0.5°C increment. The

fluorescence threshold value and gene expression data were calculated using CFX96 system software. Primer combinations are listed in Table 2.4.

Table 2.4. Primers used for qRT-PCR analysis.

Gene	Forward sequence (5'-3')	Reverse sequence (5'-3')
<i>ZmPBF</i>	CTTGGGCTCTCACACCTCATC	GTTGGGGTAGGCATCGTCA
<i>ZmGAMYB</i>	TGGGAACAGCCTTTATCTGCC	ATCCCCGCTTGGTCAACCT
<i>ZmBCH2</i>	GCTTGTTAGCAGTCCGGT	GAAAGGAGGATGGCGATAGAT
<i>ZmPSY1</i>	CATCTTCAAAGGGGTCGTCA	CAGGATCTGCCTGTACAACA
<i>ZmACTIN</i>	CGATTGAGCATGGCATTGTCA	CCCCTAGCGTACAACGAA
<i>OsACTIN</i>	GACTCTGGTGATGGTGTGACG	GCTTCTCCTTTATGTCTCTGAC
<i>gusA</i>	CGTGGTGATGTGGAGTATTGC	ATGGTATCGGTGTGAGCGTC

ZmPBF, maize (*Zea mays*) P-box binding factor gene; *ZmGAMYB*, maize *GAMYB* gene; *ZmBCH2*, maize β -carotene hydroxylase 2 gene; *ZmPSY1*, maize phytoene synthase 1 gene; *ZmACTIN*, maize actin gene; *OsACTIN*, rice actin gene; *gusA*, β -glucuronidase (*GUS*) reporter gene.

2.3.11. Histochemical and fluorimetric GUS assays

Histochemical GUS assays were carried out as described by Jefferson *et al.* (1987) also with minor modifications. Leaves and sectioned seeds were incubated at 37°C overnight (12 h) in the dark in 1mM X-gluc (5-bromo-4-chloro-3-indolyl- β -D-glucuronide) in 100mM sodium phosphate (pH 7.0), 10mM EDTA, 0.5mM potassium ferricyanide, 0.5mM potassium ferrocyanide, 0.3% (v:v) Triton X-100 and 20% (v:v) methanol to eliminate endogenous GUS activity (Kosugi *et al.*, 1990). After staining, tissues were destained in an ethanol series (50%, 70%, 80% and 95%) to remove chlorophyll, and then stored in 70% (v:v) ethanol, and photographed with a digital camera.

Fluorimetric GUS assays were carried out as described by Jefferson *et al.* (1987) with minor modifications. Plant tissues (100mg) were ground to powder under liquid nitrogen, dispersed in 0.5ml extraction buffer (50mM Na₂HPO₄ pH 7.0, 10mM EDTA, 0.1% (v:v) sodium dodecanoyl (methyl) aminoacetate, 10mM 2-mercaptoethanol and 0.1% (v:v) Triton X-100) and centrifuged at 15,300g for 20min at 4°C. The supernatant (250 μ l) was mixed with 250 μ l 2mM 4-methylumbelliferyl- β -D-glucuronide (MUG) on ice. Four 100- μ l aliquots were then transferred to a fresh tube, and 1.9ml of GUS stop buffer (0.2M Na₂CO₃) was immediately added to one aliquot as a control. The remaining aliquots were incubated at 37°C for 1 h before adding 1.9ml of GUS stop

buffer to each. The released fluorescent product 4-methylumbelliferone (MU) was measured on a QuantiFluor hand-held fluorimeter (Promega, Madison, WI, USA) with excitation and emission wavelengths of 365 and 455nm, respectively. The protein content was determined using the Bradford Reagent Kit (Sigma-Aldrich) with bovine serum albumin (BSA) as a standard. GUS activity was expressed in picomoles of MU per min· μ g of soluble protein. Three biological replicates and three technical replicates were performed for GUS activity analysis.

2.3.12. Statistical analysis

Significant differences in quantitative dependent variables for any two alternative treatments were calculated using Student's *t*-test, with p-values of 0.05 and 0.01 considered significant and highly significant, respectively. GUS activity and normalized *gusA* expression in rice tissues were analyzed by standard two-way factorial analysis of variance (ANOVA), using presence or absence of the transgenes encoding the two transcription factors as fixed factors. Analysis of covariance (ANCOVA) was also carried out, using as independent covariates the product of the expression of the transcription factors, rather than their presence or absence.

2.4. Results

2.4.1. Cloning and bioinformatic analysis of the *ZmBCH2* 5'-flanking region

The *ZmBCH2* cDNA sequence was used as a blastn query against sequence database B73 RefGen_v2 (MGSC) revealing that the endogenous *ZmBCH2* gene is located on chromosome 10. The 2036-bp 5'-flanking region (GenBank KF941345) was cloned from the leaf genomic DNA of maize inbred line B73 (Fig. 2.1). Because *ZmBCH2* is preferentially transcribed in the endosperm (Li *et al.*, 2010), I searched the PLACE and PlantCARE databases (Higo *et al.*, 1999; Lescot *et al.*, 2002) to identify *cis*-regulatory elements potentially responsible for this expression profile. This revealed two P-boxes (5'-TG(T/A/C)AAA(G/A)-3', positions -198 to -192 and -518 to -512), one AACA motif (5'-AACAAAC-3', position -253 to -247), one reverse and complementary P-box (5'-CTTTACA-3', position -1271 to -1265), one GCN4 motif (5'-CAAGCCA-3', position -1283 to -1277), one ACGT motif (ACGTOSGLUB1 motif, 5'-GTACGTG-3',

position –1763 to –1757) and three Skn-1 motifs (5'-GTCAT-3', positions –393 to –389, –434 to –430 and –1560 to –1556). These motifs are known to be responsible for the endosperm-specific expression of other genes in monocots (Vicente-Carbajosa *et al.*, 1997; Washida *et al.*, 1999; Wu *et al.*, 2000; Yamamoto *et al.*, 2006) and I therefore considered them likely to be involved in the endosperm-preferred expression of *ZmBCH2*.

```

-1773 GCTTATGGTG GTACGTGCCT GATGACTGAC GATGTACGAG AGAACTTCTT CCGGCAGGCG
      ACGTOSGLUB1
-1713 TGGGGCGTGC TCTCTAATGA TGAGACCATG GAAATCGTGG CATATATCGA AGTCAGCATG
-1653 ATACTGATGG TTGCAATGTA GTGGTGAAAT AATTAGATTA AAACAACAAA ATTTATGTAT
-1593 GACCAGGATC ACAAACGGAC TATGAAACAT TTTGTCAATA CAGTATAACA CACATTTTGT
      Skn-1 motif
-1533 ATATAAGTTA TCGTGGTATT ATATGTCTGC GTTGCAACGC ACGGGCACTC ACCTAGTAAA
-1473 ATTGAAAAGA GAGGGACATA GGTCTGAACT AAGGTATCGA TATGCATATA TGCACACCTA
-1413 TGTATGGGCG ATGTTTAGGT GTAAGTGAGT ACAAATTACA CGCTCACATG ATGTCCAAGC
-1353 TACCTAGTGA GGAGAGCTCT AACTACCCCT TGTGATGTCT AGTGCTGAGT CGAGTGAGAG
-1293 AGCTCAGACT CAAGCCAGAA CTCTTTACAT GTGCTAGCTC TTAGCTATTC AGATGACACA
      GCN4 motif RC-P-box
-1233 TGATACCTAG GTTGTCTCAT TCTAGAGGTG TTTTGAGAAG GGTTAACACC TCGTTTGGAT
-1173 ATAGATATTA AAGTTTGAA TTGTGAATTG CAATGATGAT TCCTAAATAA TGTGTTTGG
-1113 TTGTCTAATG AATTTAGAGT TGGAAATAAA GCTCAATTCT TCGGTAGTGG ACTATAAATA
-1053 GAAGTAGCTA GCCTCACACA ATTTAGAGCC TCGACCCTCT GTGTTGGCTG AAATTATAAT
-993 TTTGATTACA TCTATACAAC ATATTTGTAT TGTAGCTTAT TTCCAATACT AAGCGTATAT
-933 GATATTGGGA TCTAATTTAA TTCTATCACC TTCAATATCT ACGTCCAAAC ATATGTTAAG
-873 AGATCCTATA TAAGGCTGGC TCCAGCAAGG AACCCATAAG GGTTCTCTAC CCTAAATATA
-813 GAGGATCAAA TGGTCTTCTA CGATCTCCAG CAGCGTCTC TAAACGGTCC TCTAAATTTA
-753 GAGGACGCTG CTGAATTCTC TCTATATATA GTTCTCTAA ACGGTCCCTT ATCCATTGGA
-693 ATACTTTAAA TAACCGGTAT AGCAAAACTA AAATATGTAT AATACATTTG AGAGTATGAC
-633 AAATACGTAT GTACAAAAAA TAAAAATAAA AAATGTCTCT AATATAGATA TTTGAGTATA
-573 GAGGACGTTG TTGGAGAGGA AGGAGATATA AAAGATATAA TCTTTTAAAG GAGACTGTAA
      P-box
-513 AGAACGAACA TAGAGAATAG ATATAGAGAA CGTTGCTGGA GACAACCTAA GACCTTGGAC
-393 TCATAGGGTC TCCGGTCGAG TCATTATCGG GATTGATCTC TCATGTACAA CTCTGAAAAAT
      Skn-1 motif
-453 GTCATGCTGA TGCTGAACTG GACCTTATAT GAATCCGTCT TCTGCCAAAA AAAAAAAAT
      Skn-1 motif
-393 CCTGAAAAAC GTTTTACTCA TGACGTTGGC AGTGTACGCT TAAAAGACAA AACGTTTTCA
-333 CGGATCTTAC ATGTAAAAAT AACAACTGA ATCTAACGTG CGTAACATGG AAAGGGATTA
      AACA motif
-273 CTATTAAGAG TAACATGAAA AGTGATATTT GACGTGGCAC GCATCACACA GGTGCTGCG
      P-box
-153 TACTTAATTG GACAAACATG TTTGGTGT TTGAGTACTC AAGCAACTCC GTTCCGGCCT TTGGATCAAA
-93 AATCCGTCGG TAGGCGGCAG GAGCTGAGCC CTGAGTGACA AACCGTTGTG ACACGTACTT
-33 GTGAGCAAGG GGAAGCACAT TCCAATTCCG CCACCGCCAC TTCTTTCTTC CCCGCGGACG
      +1
-33 CGTCGAAAAG CGAGCCTCTG GGGAGACTCG AGGCCACTCT GCCTTCCCCT CCTATCCTGT
27 GAGCTGTAAC GCTTGGAGGT ATGTGAACTA TCCATATAAG TCCTGTGAG CTGTCTCCCG
87 CTCACTCGGT CCCTGTCCAT CCGTAGCGTT CCATTCCATT CCATCCCATC CCCGGAAAGT
      M
147 GCAAGGACGG GAGGGAGAGC GGCCGGCGCG TCACAAATAC CACCAGCCCA GGGGCCATGG
      A A A M T S F V A K
207 CCGCCGGCAT GACCAGCTTC GTCGCCAAG
    
```

Figure 2.1. Sequence of the maize (*Zea mays*, inbred line B73) b-carotene hydroxylase 2 gene (*ZmBCH2*) promoter and part of the cDNA region (GenBank KF941345). Position +1 is the first nucleotide of the longest *ZmBCH2* cDNA (GenBank AY844958). The first intron (positions 86–173) is in blue italics. The P-box, reverse and complementary P-box (RC-P-box), AACA motif, GCN4 motif, Skn-1 motif and ACGTOSGLUB1 motif are in red and underlined.

2.4.2. Characterization of the 5'-flanking region of *ZmBCH2* in transgenic rice

I prepared an expression construct in which the full-length *ZmBCH2* 5'-flanking region was fused to the reporter gene *gusA* (PF-5'UTR-*gusA*) (Fig. 2.2A). This included 1773-bp of upstream promoter sequence and the 263bp 5'-UTR containing the first intron. The PF-5'UTR-*gusA* construct was introduced into rice, and at least six independent transformants were analyzed by histochemical staining for GUS activity in the leaves and hand-cut half cross-sections of seeds. The full-length construct was sufficient to drive strong GUS activity in the seeds but not in the leaves of the transgenic rice plants (Fig. 2.2B), and the expression profile of *gusA* was similar to that of the native *ZmBCH2* in maize (Li *et al.*, 2010). These data indicated that the *ZmBCH2* 5'-flanking region in transgenic rice is regulated in a similar manner to the endogenous *ZmBCH2* promoter in maize.

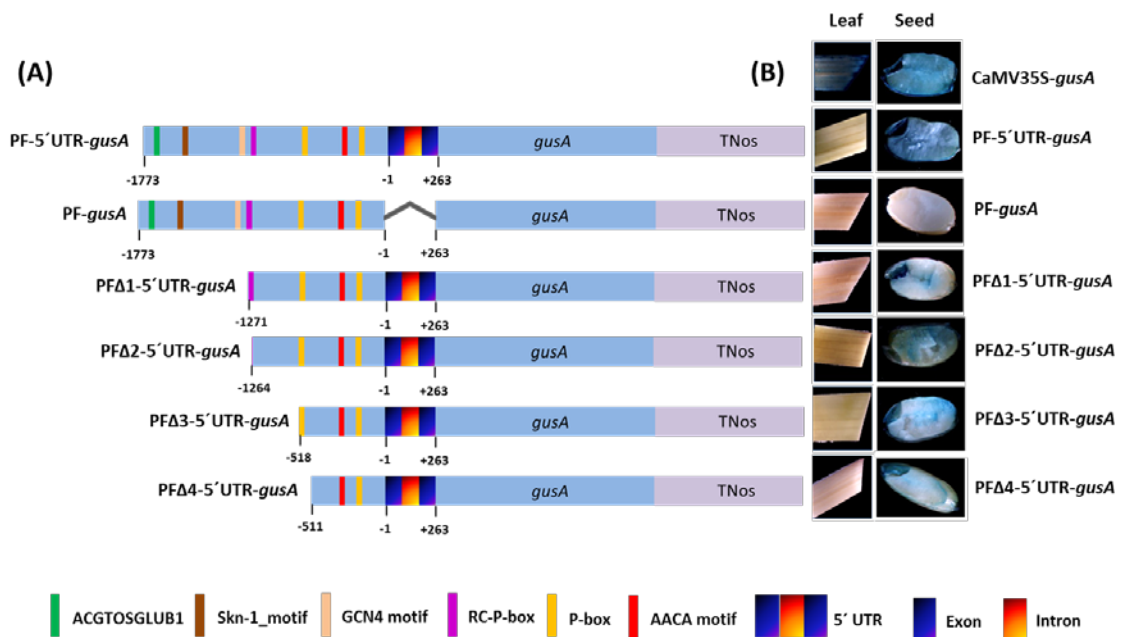


Figure 2.2. Schematic representation of *gusA*- fusion constructs and histochemical b-glucuronidase (GUS) assays of representative transgenic rice tissues expressing *gusA* driven by the constitutive CaMV35S promoter (CaMV35SPro-*gusA*), and various 5' upstream regions of maize *ZmBCH2*. (A) Schematic representation of *gusA*-fusion constructs driven by various 5' upstream regions of the *ZmBCH2* gene. (B) Histochemical GUS staining of different tissues in transgenic rice plants (T1 mature leaves and T2 seeds at 25 days after pollination (DAP)) expressing *gusA*-fusion constructs driven by various 5' upstream regions of the *ZmBCH2* gene. Rice mature leaf tissues and cross-sectioned seeds at 25 DAP were incubated in 5-bromo-4-chloro-3-indolyl-b- D-glucuronide (X-gluc) solution. Bars: 2mm.

The 5'-UTR of plant genes, particularly the first intron, is often necessary to achieve high-level gene expression (Hernandez-Garcia and Finer, 2014; Gallegos and Rose, 2015). To determine whether the 5'-UTR plays a similar role in the *ZmBCH2* gene, the 1773bp promoter region was linked to *gusA* following the removal of the 263bp 5'-UTR (PF-*gusA*). Analysis of the corresponding transgenic rice plants revealed that the 1773bp promoter fragment alone was sufficient to confer seed-specific GUS activity (Fig. 2), but the level of GUS activity determined using a fluorimetric GUS assay was 8.4 fold lower compared to the construct containing the 5'-UTR (Fig. 2.3). This confirmed that the 5'-UTR, containing the first intron, is necessary for high-level *ZmBCH2* expression.

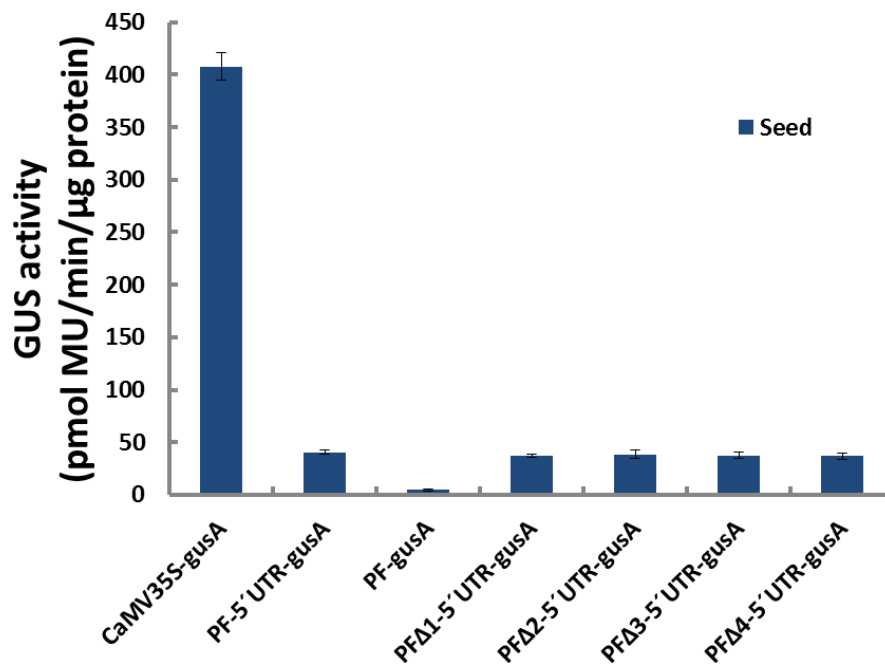


Figure 2.3. Quantitative analysis of β -glucuronidase activity in transgenic rice.

To identify the *cis*-regulatory elements responsible for the endosperm-preferred expression of *ZmBCH2*, the 5'-flanking region was progressively deleted from the 5'-end and the nested set of constructs was tested in transgenic rice (Fig. 2.2). Deletions that removed the 5' sequence beyond position -1271 removed the ACGTOSGLUB1 (ACGT motif, -1763 to -1757), a Skn-1 motif (-1560 to -1556) and a GCN4 motif at position -1283 to -1277 (Fig. 2.2A). The GUS activity was similar to that of the full-length 5'-flanking region with the 5'-UTR (Fig. 2B). Further progressive deletion to

positions –1264, –518 and –511 removed the reverse and complementary P-box (–1271 to –1265) and P-box (–518 to –512) but had little further impact on GUS activity (Fig. 2.2B). All constructs produced similar GUS profiles (Fig. 2.2B) and similar levels of GUS activity in transgenic rice seeds (Fig. 2.1) compared to the full-length construct PF-5'UTR-*gusA*. None of the constructs were active in leaves but even the shortest, with 774-bp of 5'-flanking sequence (PF Δ 4-5'UTR, –511 to +263), contained all the *cis*-regulatory elements necessary for high GUS activity in transgenic rice seeds. As shown in Figure 2.1, this 774-bp region contains a single P-box, a single AACA motif and two Skn-1 motifs, suggesting these are the key *cis*-regulatory elements responsible for seed-preferred expression, whereas the distal ACGT-motif, reverse and complementary P-box, GCN4 and Skn-1 motifs are dispensable.

2.4.3. Expression patterns of *ZmBCH2*, *ZmPBF* and *ZmGAMYB* in maize B73 leaf, endosperm and embryo

Analysis of the *ZmBCH2* 5'-flanking region in transgenic rice revealed that a 774-bp flanking region (–511 to +263) containing one AACA motif, one P-box and two Skn-1 motifs was sufficient to confer preferential expression in the endosperm. To identify the corresponding transcription factors, I used the barley *GAMYB* cDNA sequence (GenBank X87690) to search MaizeGDB and NCBI GenBank, revealing one maize ortholog, *ZmGAMYB* (GenBank BT062692). The *ZmGAMYB* gene is located on chromosome 3 and has four exons. The deduced *ZmGAMYB* amino acid sequence has 80.1% identity and 84.9% similarity with barley *GAMYB*. Similarly, by using the *ZmPBF* cDNA sequence (GenBank ZMU82230) to search the MaizeGDB, I isolated the *ZmPBF* full-length ORF cDNA from inbred line B73 (GenBank JQ723698). *ZmPBF* is located on chromosome 2 and has two exons.

RNA gel blots were then prepared from 20 μ g of total RNA and were hybridized with gene-specific probes for *ZmBCH2*, *ZmPBF* and *ZmGAMYB*. *ZmBCH2* expression levels gradually increased in the endosperm from 15 to 30 DAP (Fig. 2.4) and both *ZmPBF* and *ZmGAMYB* expression levels gradually increased in the endosperm from 15 to 25 DAP, before declining at 30 DAP. Furthermore, *ZmBCH2* and *ZmPBF* expression levels gradually increased in the embryo from 20 to 30 DAP whereas *ZmGAMYB*

expression was only detected in the embryo at 20 DAP. No transcripts for any of the genes were detected in leaves.

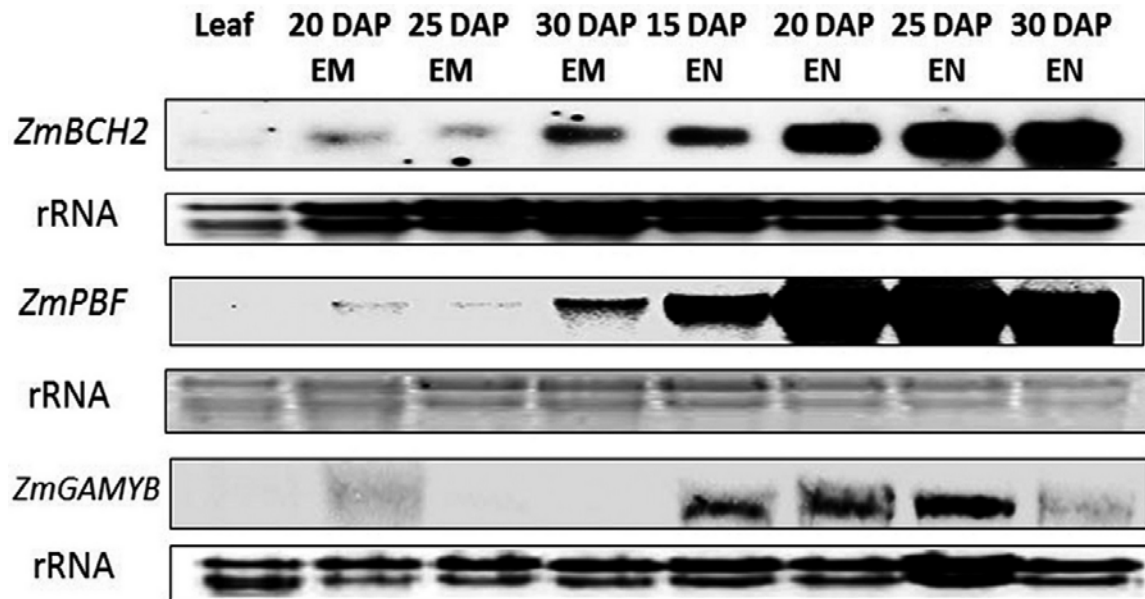


Figure 2.4. Expression of endogenous *ZmBCH2*, *ZmPBF* and *ZmGAMYB* genes in different tissues of maize (*Zea mays*, inbred line B73). We loaded 20 μ g of total RNA for each sample. DAP, days after pollination; EM, embryo; EN, endosperm.

2.4.4. *ZmPBF*/P-box and *ZmGAMYB*/AACA motif interactions

Next I used the yeast one-hybrid system to determine whether *ZmPBF* and *ZmGAMYB* interact with the P-box and AACA motifs, respectively, in the minimal *ZmBCH2* promoter. The bait plasmids pAbAi-*ZmBCH2*PF Δ 4-5'UTR (wild-type sequence), pAbAi-*ZmBCH2*PF Δ 4-mP-5'UTR (mutated P-box) and pAbAi-*ZmBCH2*PF Δ 4-mA-5'UTR (mutated AACA motif) were independently introduced into Y1HGOLD yeast cells, which can grow on medium lacking uracil but not on medium containing 200ng ml⁻¹ AbA (Aureobasidin A). I therefore chose 200ng ml⁻¹ as the minimal inhibitory concentration of AbA for the bait plasmids. The prey plasmid pGADT7-*ZmPBF* was separately introduced into the bait (pAbAi-*ZmBCH2*PF Δ 4-5'UTR), mutated P-box bait (pAbAi-*ZmBCH2*PF Δ 4-mP-5'UTR) and void bait (pAbAi) yeast strains. The transformed yeast cells with the wild-type promoter were able to grow on medium lacking leucine and containing 200ng ml⁻¹ AbA, but the transformed mutated P-box bait and void bait yeast cells could not (Fig. 2.5A), indicating that *ZmPBF* can specifically bind to the P-box element in the truncated *ZmBCH2* promoter in yeast. Likewise, I

demonstrated that *ZmGAMYB* can specifically bind to the AACA motif in the *ZmBCH2* promoter in yeast (Fig. 5B). Thus, the yeast one-hybrid system demonstrated that *ZmPBF* and *ZmGAMYB* interact with the P-box and AACA motif, respectively, in the proximal region of the truncated *ZmBCH2* promoter.

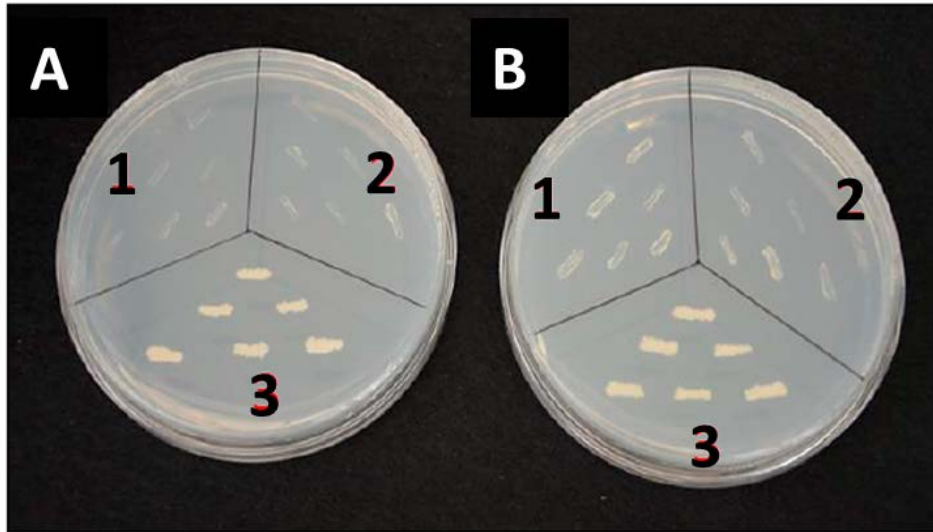


Figure 2.5. Identification and characterization of P-box and AACA motifs in the maize *ZmBCH2* basal promoter region which could be recognized and bound by the *ZmPBF* and *ZmGAMYB* transcription factors, respectively. (A) Plasmids pAbAi (A1), pAbAi-*ZmBCH2PFΔ4*-mP-5'UTR (mP, mutated P-box in 774 bp *ZmBCH2PFΔ4*-5'UTR fragment) (A2) and pAbAi-*ZmBCH2PFΔ4*-5'UTR (A3) were introduced into the yeast strain Y1HGold. These strains were retransformed with pGADT7-*ZmPBF*. The resulting strains could grow on synthetic defined medium lacking leucine (SD/-Leu). Once these yeast strains were transferred onto SD/-Leu/ aureobasidin A (AbAi) medium (200 ng ml⁻¹), only strain (A3) transformed with pAbAi-*ZmBCH2PFΔ4*-5'UTR containing an intact P-box motif could grow, whereas strains transformed with pAbAi (A1) and pAbAi-*ZmBCH2PFΔ4*-mP-5'UTR (A2) could not. (B) Plasmids pAbAi (B1), pAbAi-*ZmBCH2PFΔ4*-mA-5'UTR (mA, mutated AACA motif in 774 bp *ZmBCH2PFΔ4*-5'UTR fragment) and pAbAi-*ZmBCH2PFΔ4*-5'UTR were introduced into the yeast strain Y1HGold. These strains were retransformed with pGADT7-*ZmGAMYB*. The resulting strains could grow on SD/-Leu medium. Once these yeast strains were transferred to SD/-Leu/AbAi medium (200 ng ml⁻¹), only strain (B3) transformed with pAbAi-*ZmBCH2PFΔ4*-5'UTR containing an intact AACA motif could grow, whereas the strains transformed with pAbAi (B1) and pAbAi-*ZmBCH2PFΔ4*-mA-5'UTR (B2) could not.

2.4.5. Transient expression of *ZmPBF* and/or *ZmGAMYB* and the impact on *ZmBCH2* and *ZmPSY1* expression in maize endosperm and embryo

A transient expression assay was carried out to determine the effect of *ZmPBF* and *ZmGAMYB* on *ZmBCH2* promoter activity. Maize seeds were bombarded with the expression vectors pAHC25 (Ubi-*gusA*), pAL76-*ZmPBF* (Ubi-*ZmPBF*) and/or pAL76-

ZmGAMYB (Ubi-*ZmGAMYB*) (Fig. 2.6) and the effect on *ZmBCH2* expression was determined by qRT-PCR. In maize endosperm and embryo cells at 10 and 25 DAP, the overexpression of either *ZmPBF* or *ZmGAMYB* alone significantly increased endogenous *ZmBCH2* mRNA levels (Fig. 2.7). However, the overexpression of both transcription factors did not result in additive transactivation of the *ZmBCH2* promoter in the endosperm and embryo. These results indicate that *ZmPBF* and *ZmGAMYB* can act independently as effective stimulators of the *ZmBCH2* promoter in maize endosperm and embryo.

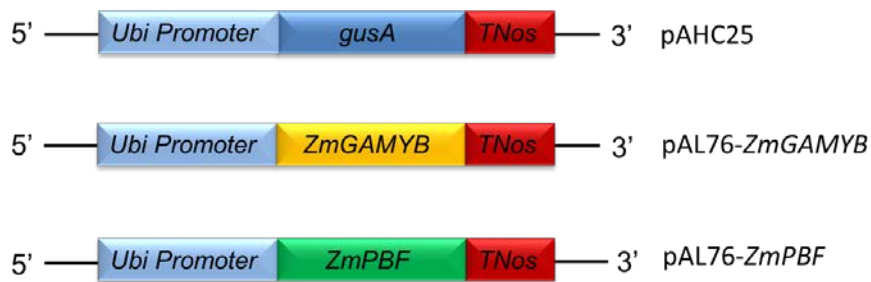


Figure 2.6. Schematic representation of expression vectors pAHC25, pAL76-*ZmGAMYB* and pAL76-*ZmPBF*.

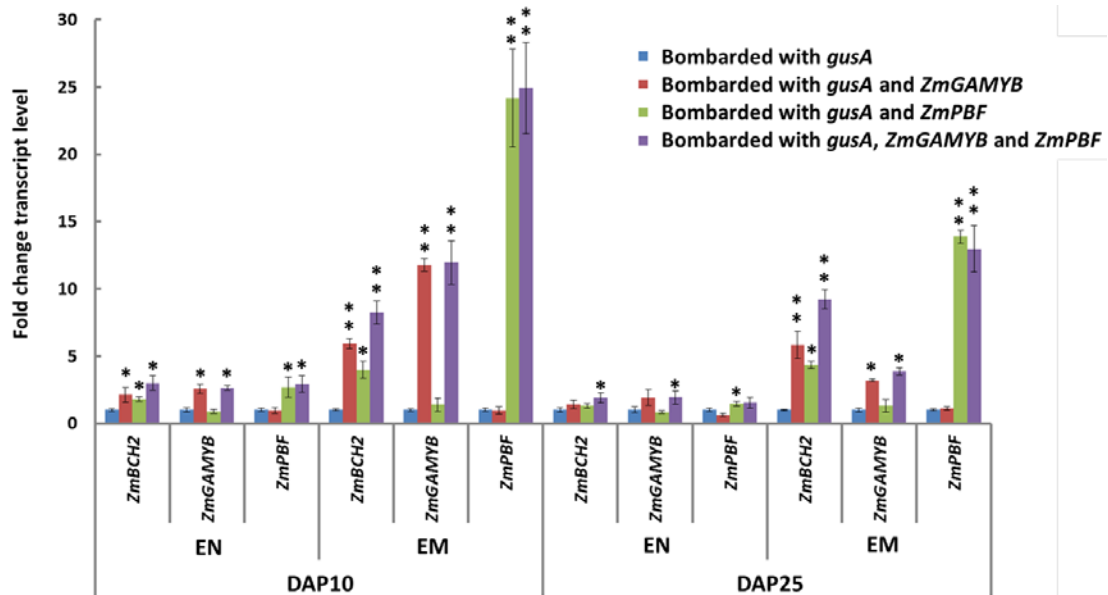


Figure 2.7. Expression of *ZmBCH2*, *ZmGAMYB* and *ZmPBF* in maize (*Zea mays*, inbred line B73) endosperm and embryo tissue as measured by transient expression analysis followed by quantitative real-time reverse transcription PCR (qRT-PCR), with mRNA levels normalized against *ZmACTIN* mRNA. Expression of *ZmPBF*, *ZmGAMYB* and *ZmBCH2* is expressed as a percentage relative to the internal control (*gusA* expression). Five biological replicates and three technical replicates were tested. Asterisks indicate statistically significant differences compared

with the control (bombarded with pAHC25) according to Student's *t*-test: *, $P < 0.05$; **, $P < 0.01$. Bars indicate \pm SD. EM, embryo; EN, endosperm.

Bioinformatic analysis of the *ZmPSYI* 5'-flanking region revealed the presence of one AACA motif (-827 to -821), two P-boxes (-1221 to -1215 and -1341 to -1335) and one reverse complementary P-box (-1600 to -1594) in the distal region (Fig. 2.8). However, P-box and AACA motifs were not found in the promoters of other carotenoid biosynthesis genes up to 2 kb upstream from the transcriptional start site (Table 2.5). RNA gel blots revealed that *ZmPSYI* is abundantly transcribed in maize endosperm from 20 DAP (Fig. S5). I therefore investigated endogenous *ZmPSYI* expression using the transient expression strategy described above. The transient expression of *ZmPBF* and/or *ZmGAMYB* caused a slight non-significant increase in *ZmPSYI* expression in maize embryo and endosperm (Fig. 2.9).

-1833	GGATCCTGGC	ACATATGGTA	TGCGCTCGGT	TCTACCCCTG	GAACCGGGTG
-1783	TGACATTGAC	CTCCTGCGCG	TGTATTTTGT	TATATAATAC	TGGGGGGAGT
-1733	ACATGATCTA	TCAAGGATAT	CTGGAAAAGC	AGAGACATCA	CACACACACG
-1683	ACACATGCAT	GCATTTTACT	AGCAGCTAGC	AGTACTCCGA	GTCCACCCCA
-1633	AGCTAGTCAA	TGACAAACCA	AACCTCCCCA	AATCTTTACA	CGTTTTATGT
				RC P-box	
-1583	TACTAGGAAC	CACATGTTTT	CTGTTATAGT	CTTGGATTTG	ATTTGACTAG
-1533	AGACTAACCG	GTGCTTTTTCT	TGCTCTGCCA	CCCAAATTGT	TCGTTCCCTC
-1483	CCCCTCGCTA	GCAGCTCACG	CCCACCAAGT	GACGACGAGA	GGGAGACGGG
-1433	ACACGCACTT	TCTGTTGCCG	TGCTCTCACC	AAACTTGCAA	CCACAATATA
-1383	TATATAAGCC	ACCGGATTGA	TAATATGGAA	ATGGGGAATA	AA TGAAAAGG
				P-box	
-1333	TGCCATGCAT	CATGCATGGA	TATGCGAATG	ATAATGCTTG	CTAGCCGAAT
-1283	GTTAGGTACG	GGACAGGACT	GAAGCTGATC	GAGCACTTGC	ATACAGTAAT
-1233	ACCTAAACCT	ACTGTAAGC	AAAATTGCCG	AGAGAAGAAG	CAGGCAAGCA
		P-box			
-1183	GCAGTGTCAA	ACAACGTGTA	GAAAGCAAAT	AAAGCGACTA	AAACAATGCT
-1133	GATCCATCCA	TCGGCATCAT	CACGTAAG	AAAAGAAAAG	GCAGAGCCGT
-1083	AAACGGTAAT	TTATTATACA	CATAATAAAT	AGCTAATAAG	CTACTCCTTC
-1033	TTGTCCTAAG	TTAAAATTAA	ATTTCGTTTA	GTTAATTAAT	AGATTCATAT
-983	AATACTTGAT	GTTGATCTAT	GTGTTTTATA	TGCGTCTAGA	TTCATCTTCA
-933	TCTATTTGAA	TATAGACATA	AAAATCAAGA	GCTAAAATAA	CTACTATTTT
-883	GGTATTTTGG	AATGGAGGGA	GTAATAGACG	ACAAGTGAGC	CTGGTGAGTT
-833	ACCTGA AACA	AACAAAGCCA	GCAGAGCCAG	AGGTCGCGGC	TATGGTAGTC
	AACA motif				
-783	TGACTGCCAT	GCATGTCACC	GCGGGTGTGG	GGGGCGCCCA	GCAGCCACGT
-733	CGGACAGGAG	CAGCCAGGGT	GAATCCGGCC	TTTTCCAGGT	GTCACCACTC
-683	AGCGTCCCTC	GAACACAGGA	GAGTCATGCG	ATGCGAGCTT	GGCGATAAGC
-633	TTATCTATCC	GCACCGCGTC	TTCTTCTCCTC	CTGGGCGACC	GGCCCTTCTT
-583	CTCTCCACGT	CTCTCCCCC	TTCTTCTCCTC	AGACCGAGCG	TACGTATGCT
-533	ACACACAGCA	ACAGCACAAAC	AGTACTAGTT	CCACCACAAG	AAGATGCCCA
-483	ATGCCAAGAA	ATAACCCATG	CTTCTTGTG	ACGATCCAGC	CGCACTAGAG
-433	ATGGCCAAAC	GGGCCGGGCC	GGGCCGGGCC	GGGCCGGGCC	CGGTGAAGCC
-383	CGGCCAAAAC	CGGCCGGGCC	CTGCTGAGCC	AGCGGGCTTA	AGTTTCTGTC
-333	CAAGCCCGGC	CCGACGCGGG	CCTAAACAGG	CCGGCCGGCC	CGTTTGTAGCA
-283	CGAAAAAACG	GGCCAAAAG	CGGGCTAAAC	GGCCGGTAA	GCACGTTTTA
-233	GTGTAATAAAA	AACGGGCTTA	ACGGGCTTAG	AGGTAAACGG	GCCGTGCCGG
-183	GCTAGCCCGC	CGTGCCTAGT	TTCTTGTCCA	AGCCCGCCCG	CTTATTCTAC
-133	CGTGCCGGGC	TCGGACCGGG	CCCAAAAAGC	GGGCTTCGTG	CCGGGCTCAC
-83	GGGCTCGTGT	CTTTTTGGCC	ATCTATGAGC	CGCACACTTA	GCATACATAC
-33	GCAAGAAGAG	GAGAGGCCGG	AGGTGCGCGT	GCTCCTTGCT	GTTCTGCTGA
				+1	
+18	CTGGTCTCAC	CATCTCATCC	CACCACCACC	ACCACCACCA	CCATCTTTAG
+68	GATAAGATAG	CAAATAT ATG			

Figure 2.8. Sequence of the maize phytoene synthase 1 (*ZmPSYI*) promoter and part of the cDNA region.

Table 2.5. Carotenogenic genes in maize (*Zea mays*).

Gene name	Chromosome number	GenBank accession number
<i>ZmPSY1</i>	6	AY773944 (genomic DNA, gDNA); AC205337 (<i>Zea mays</i> cultivar B73 chromosome 6 clone CH201-85C14); BT034021(mRNA)
<i>ZmPSY2</i>	8	AY325302 (gDNA); AC201989 (<i>Zea mays</i> cultivar B73 chromosome 8 clone CH201-498O17), BT039824 (mRNA)
<i>ZmPSY3</i>	7	DQ372936 (gDNA); EU958083 (mRNA)
<i>ZmPDS</i>	1	AF039585 (gDNA); AC211261(<i>Zea mays</i> cultivar B73 chromosome 1 clone CH201-365I14); ZMU37285 (mRNA)
<i>ZmZ-ISO</i>	10	AC202000 (<i>Zea mays</i> cultivar B73 chromosome 10 clone CH201-438P11); NM_001139248 (mRNA)
<i>ZmZDS</i>	7	AY641394 (gDNA); AC197604 (<i>Zea mays</i> cultivar B73 chromosome 7 clone CH201-21J19); AF047490 (mRNA)
<i>ZmCRTISO1</i>	4	AC205563 (<i>Zea mays</i> cultivar B73 chromosome 4 clone CH201-260P3); FJ603466 (mRNA)
<i>ZmCRTISO2</i>	2	AC183901 (<i>Zea mays</i> cultivar B73 chromosome 2 clone CH201-115P9); FJ765413 (mRNA)
<i>ZmLCYB</i>	5	AY206862 (gDNA); AC205536 (<i>Zea mays</i> cultivar B73 chromosome 5 clone CH201-106C18); AY206862 (mRNA)
<i>ZmLCYE</i>	1	AC177851 (<i>Zea mays</i> cultivar B73 chromosome 1 clone CH201-9J22); GU130217 (mRNA)
<i>ZmBCH1</i>	2	AC196442 (<i>Zea mays</i> cultivar B73 chromosome 2 clone CH201-112K16); GQ131287 (mRNA)
<i>ZmBCH2</i>	10	KF941345 (gDNA); AY844958 (mRNA)
<i>ZmCYP97A16</i>	5	AC232393 (<i>Zea mays</i> cultivar B73 chromosome 5 clone ZMMBBb-133J13); GU130216 (mRNA)
<i>ZmCYP97B2</i>	4	AC208126 (<i>Zea mays</i> cultivar B73 chromosome 4 clone ZMMBBb-605L24); HM067864 (mRNA)
<i>ZmCYP97C19</i>	1	AC177851 (<i>Zea mays</i> cultivar B73 chromosome 1 clone CH201-9J22); GU130217 (mRNA)
<i>ZmZEP1</i>	2	AC194845 (<i>Zea mays</i> cultivar B73 chromosome 2 clone CH201-463G20); XM_008669887 (mRNA)
<i>ZmZEP2</i>	10	AC206194 (<i>Zea mays</i> cultivar B73 chromosome 10 clone CH201-241O11); XM_020545718 (mRNA)
<i>ZmZEP3</i>	10	AC206194 (<i>Zea mays</i> cultivar B73 chromosome 10 clone CH201-241O11); XM_020545719 (mRNA)

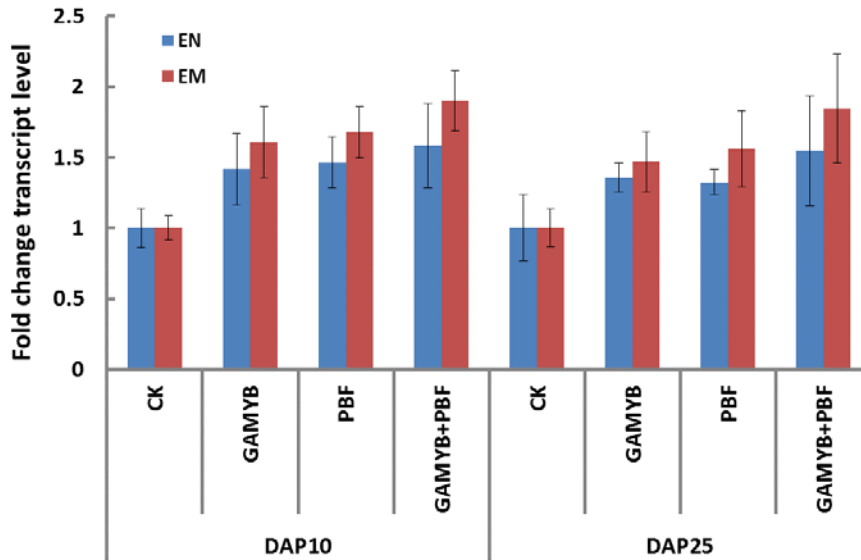


Figure 2.9. Quantitative RT-PCR analysis of maize phytoene synthase 1 gene (*ZmPSY1*) expression in endosperm and embryo determined by transient expression.

2.4.6. Overexpression of *ZmPBF* and/or *ZmGAMYB* in transgenic rice plants and the impact on *gusA* expression and GUS activity in leaf and seed

I transformed 7-day-old mature zygotic rice embryos with the PFA4-5'UTR-*gusA* construct with or without pAL76-*ZmPBF* and/or pAL76-*ZmGAMYB*, in which the *ZmPBF* and *ZmGAMYB* cDNAs were driven by the constitutive maize *ubiquitin-1* promoter (Fig. 2.6). In all cases, an additional construct was introduced containing the selectable marker *hpt* (encoding hygromycin phosphotransferase) also driven by the maize *ubiquitin-1* promoter. I measured transcript levels in the leaf and seed tissues of the resulting transgenic rice plants (at least 10 independent transformants for each combination of transgenes) by qRT-PCR and selected the top three in each case with the highest expression levels of *ZmPBF* and/or *ZmGAMYB* for the analysis of *gusA* expression and GUS activity in the T2 homozygous lines. Analysing the lines expressing the PFA4-5'UTR-*gusA* construct plus one of the two transcription factors, I found that *gusA* expression levels and GUS activity were consistently higher than in the control line expressing the PFA4-5'UTR-*gusA* construct alone (Figs 2.10 and 2.11), confirming that the reporter construct was transactivated by both ectopic transcription factors acting independently. In the seed, the *gusA* expression level was 2.6 (normalized against the rice *actin* gene) in the absence of any transcription factors, rising to 7.0 in

the presence of *ZmPBF* or 15.6 in the presence of *ZmGAMYB*, and to 27.3 when both transcription factors were present (Fig. 2.10). In the leaves, the corresponding values were 0.6 for no transcription factors, rising to 1.9 in the presence of *ZmPBF* or 2.8 in the presence of *ZmGAMYB*, and to 4.5 in the presence of both (Fig. 2.10). The model that incorporated the quantitative amount of transcription factor mRNA as independent covariates better explained the *gusA* expression compared to the model based on solely the presence or absence of the transgenes ($R^2 = 90.4$ vs 82.6% in the rice seed and 96.2% vs 90.2% in the leaf) (Table 2.6). *ZmGAMYB* had a stronger effect on *gusA* expression ($p = 0.0006$ and 0.0004 for seed and leaf, respectively) than *ZmPBF* ($p = 0.0298$ and 0.0070, respectively) (Table 2.6). No significant statistical interaction was detected for either the presence/absence (ANOVA) or the quantitative (ANCOVA) models, further supporting the independent additive behavior of the transcription factors in both tissues (Fig. 2.10, Table 2.6). GUS activity analysis fully supported the *gusA* expression data (Fig. 2.11, Table 2.6).

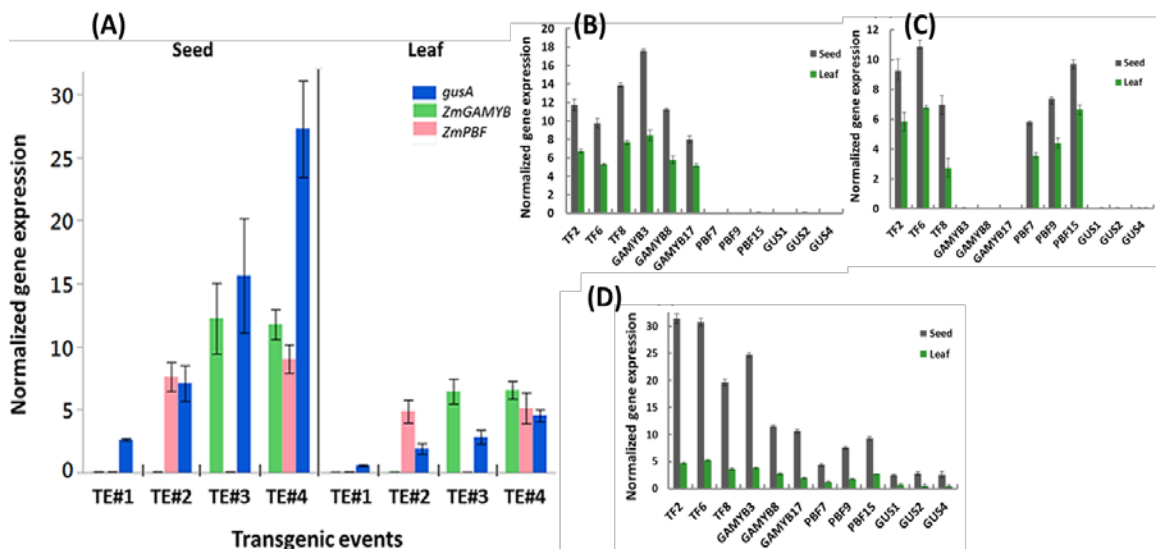


Figure 2.10. (A) Relative average expression (normalized to *OsACTIN*) in rice T3 seeds (25 d after pollination, DAP) and T2 mature flag leaves for two maize transcription factor genes (*ZmPBF* and *ZmGAMYB*) and the reporter transgenic lines representing each transgenic event (TE): TE#1 expressing gene *gusA*. Bars represent \pm SE of the mean for three independent *gusA*; TE#2 coexpressing *gusA* and *ZmPBF*; TE#3 coexpressing *gusA* and *ZmGAMYB*; and TE#4 coexpressing *gusA*, *ZmPBF* and *ZmGAMYB*; (B) Expression of *ZmGAMYB* in T3 transgenic rice; (C) Expression of *ZmPBF* in T3 transgenic rice; (D) Expression of *gusA* in T3 transgenic rice.

Table 2.6. Partition of the observed variance in the relative expression of *gusA* and GUS activity in rice based on standard analysis of variance (ANOVA) and covariance (ANCOVA). Partition of the observed variance in the relative expression of the β -glucuronidase (*GUS*) reporter gene (*gusA*) and GUS activity in rice T3 seeds and T2 mature flag leaves based on standard ANOVA (presence or absence of the transgene for two transcription factors) and on covariance analyses (ANCOVA) using as independent covariates the product of the expression of the transcription

Source	Gene Expression						GUS Activity												
	Seed			Leaf			Seed			Leaf									
	ANOVA on alleles	ANCOVA on mRNA expression	ANOVA on alleles	ANCOVA on mRNA expression	ANOVA on alleles	ANCOVA on mRNA expression	ANOVA on alleles	ANCOVA on mRNA expression	ANOVA on alleles	ANCOVA on mRNA expression	ANOVA on alleles								
	DF	Sum of Squares	F Ratio	Prob > F	Sum of Squares	F Ratio	Prob > F	Sum of Squares	F Ratio	Prob > F	Sum of Squares	F Ratio	Prob > F						
ZmGAMYB	1	826	29.5	0.0006	824	55.2	<0.0001	17.94	33.7	0.0004	18.25	129.9	<0.0001	50115	16.8	0.0034	7515	19.4	0.0023
ZmPBF	1	195	7.0	0.0298	285	19.1	0.0024	6.90	13.0	0.0070	10.17	72.4	<0.0001	18157	6.1	0.0389	3192	8.3	0.0207
ZmGAMYB*ZmPBF	1	39	1.4	0.2742	22	1.5	0.2588	0.09	0.2	0.6865	0.02	0.2	0.6904	4094	1.4	0.2751	417	1.1	0.3296
Error	8	224			119			4.26			1.12			23865			3094		
R²			0.83		0.90			0.85			0.96			0.75			0.78		

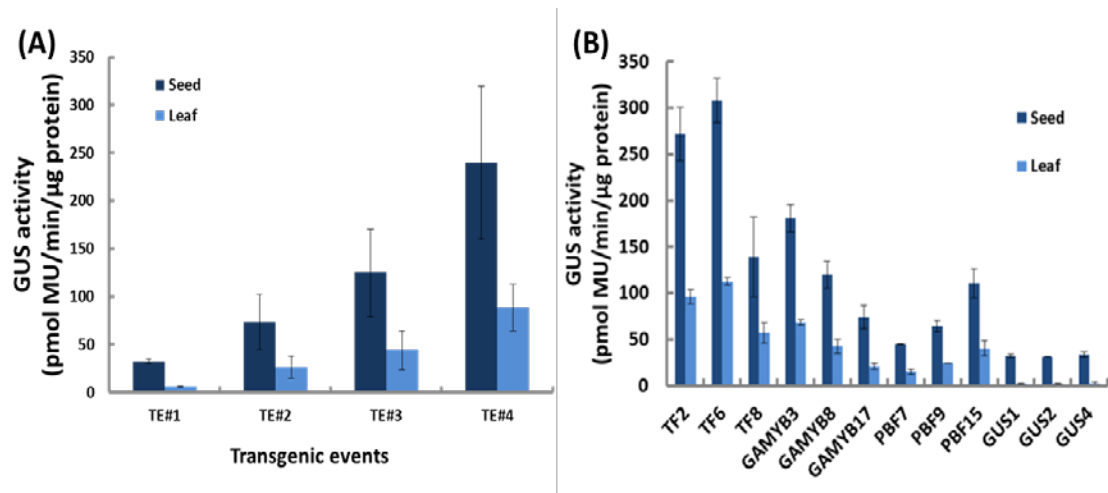


Figure 2.11. (A) Quantitative analysis of b-glucuronidase (GUS) activity in transgenic rice T3 seeds and T2 mature flag leaves. Bars represent \pm SE of the mean for three independent transgenic lines representing each transgenic event (TE): TE#1 expressing *gusA*; TE#2 coexpressing *gusA* and *ZmPBF*; TE#3 coexpressing *gusA* and *ZmGAMYB*; and TE#4 coexpressing *gusA*, *ZmPBF* and *ZmGAMYB*; (B) Quantitative analysis of b-glucuronidase activity in transgenic rice

When the experiments were repeated using PF Δ 4-mP-5'UTR-*gusA* (the mutated P-box in the truncated *ZmBCH2* promoter, 5'-TGAAAAG-3' to 5'-TGgAgAG-3') and pAL76-*ZmPBF*, or using PF Δ 4-mA-5'UTR-*gusA* (the mutated AACA motif in the truncated *ZmBCH2* promoter, 5'-AACAAAC-3' to 5'-gAgAAcC-3') and pAL76-*ZmGAMYB*, the effect of the two transcription factors was completely abolished in T0 transgenic rice leaf tissue and T1 seeds (Fig. 2.12). These data strongly suggest that the proximal P-box and AACA motif of the *ZmBCH2* promoter are the sites bound in planta by *ZmPBF* and *ZmGAMYB*, respectively, to activate the *ZmBCH2* promoter.

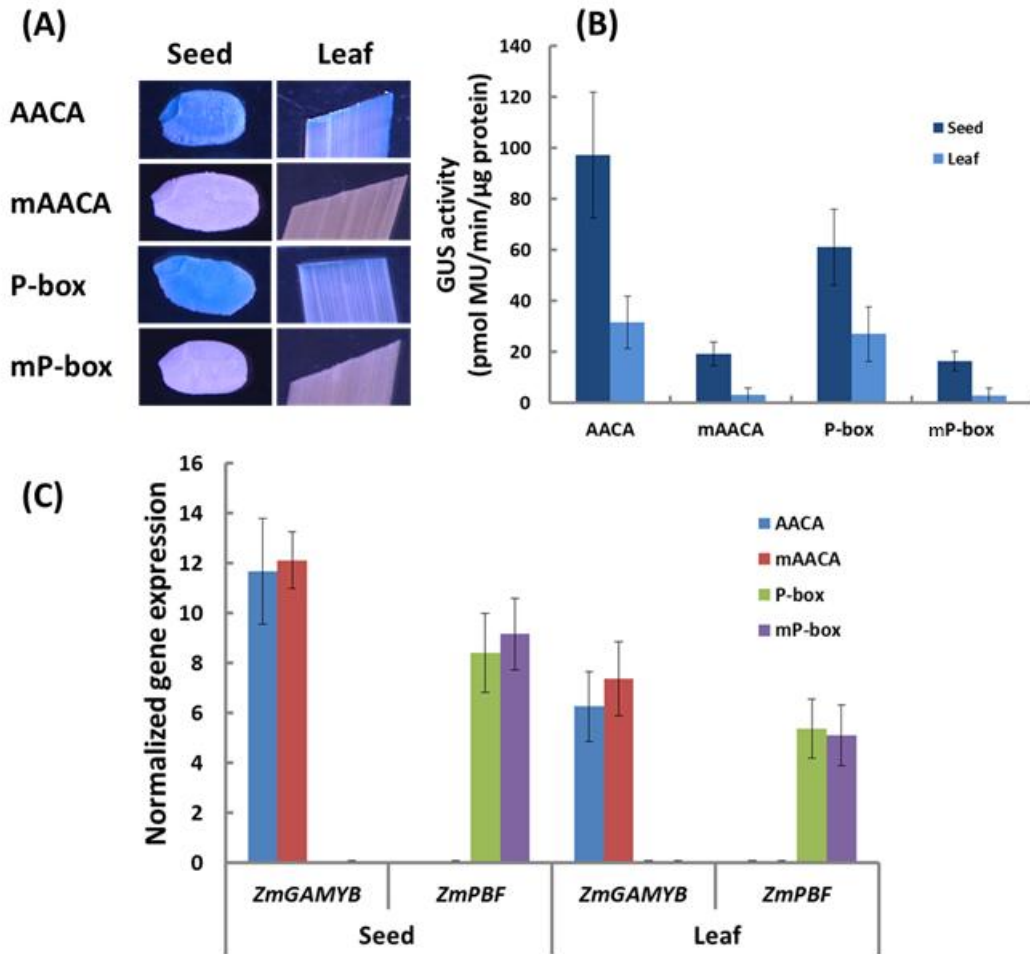


Figure 2.12. (A) Histochemical b-glucuronidase (GUS) assays of representative transgenic rice T0 mature flag leaves and T1 seeds (at 25 days after pollination, DAP); (B) quantitative analysis of GUS activity in transgenic rice T0 mature flag leaves and T1 seeds (at 25 DAP). Three biological replicates were tested for GUS coexpressing *ZmGAMYB* and *gusA* driven by the minimal promoter (PFΔ4-5'UTR-*gusA*); mAACA, transgenic rice events coexpressing *ZmPBF* and *gusA* activity. Bars represent \pm SE of the mean for three independent transgenic lines representing each transgenic event. AACA, transgenic rice events driven by the mutated AACA motif minimal promoter (PFΔ4-mA-5'UTR-*gusA*); P-box, transgenic rice events coexpressing *ZmPBF* and *gusA* driven by the minimal promoter (PFΔ4-5'UTR-*gusA*); mP-box, transgenic rice events coexpressing *ZmPBF* and *gusA* driven by the mutated P-box minimal promoter (PFΔ4-mP-5'UTR-*gusA*); MU, 4-methylumbelliferone. Bar=2mm; (C) Expression analysis of maize *ZmGAMYB* and *ZmPBF* gene in transgenic rice.

2.5. Discussion

The transcriptional regulation of carotenoid biosynthesis genes in monocots is a key intervention point to gain insight into carotenoid metabolism and in turn to devise more predictable and robust strategies for metabolic engineering. Very little is known about the regulation of these genes in cereals or indeed in other plants. A few transcription factors have been shown to influence carotenogenic genes in dicots, including *Arabidopsis* RAP2.2 and PIF1 (Welsch *et al.*, 2007; Toledo-Ortiz *et al.*, 2010), tomato RIN and BBX20 (Martel *et al.*, 2011; Xiong *et al.*, 2018), and citrus CrMYB68 and CsMADS6 (Zhu *et al.*, 2017; Lu *et al.*, 2018). RAP2.2, a member of the AP2 gene family, binds to the *Arabidopsis* *PSY* promoter and represses *PSY* and *PDS* (encoding phytoene desaturase) in root-derived callus, thus reducing the carotenoid content (Welsch *et al.*, 2007). *Arabidopsis* phytochrome-interacting factor 1 (PIF1) directly binds to the *PSY* promoter, and thus represses *PSY* expression to inhibit carotenoid accumulation (Toledo-Ortiz *et al.*, 2010). The MADS-box protein RIPENING INHIBITOR (RIN) in tomato directly induces the expression of *SIPSY1* (Martel *et al.*, 2011). SIBBX20, a B-box (BBX) zinc-finger transcription factor, can activate the expression of *SIPSY1* by directly binding to a G-box motif in its promoter, which results in the elevated levels of carotenoids in *SIBBX20* overexpression lines (Xiong *et al.*, 2018). CrMYB68, an R2R3-MYB transcription factor isolated from Green Ougan (MT), a stay-green mutant of *Citrus reticulata* cv Suavissima, can bind to the promoters of both *CrBCH2* and *CrNCED5* (*9-cis-epoxycarotenoid dioxygenase 5*) and significantly inhibits their expression (Zhu *et al.*, 2017). The MADS-box transcription factor gene *CsMADS6* was recently identified by using the promoter sequence of *CsLCYB1* as bait in a yeast one-hybrid screen for promoter-binding proteins from sweet orange (Lu *et al.*, 2018). Overexpression of *CsMADS6* in citrus callus increased the carotenoid content and induced the expression of *CsLCYB1*, *CsPSY*, *CsPDS* and *CsCCD1* (*carotenoid cleavage dioxygenase 1*) by directly binding to their promoters, highlighting the multi-target regulation of carotenoid metabolism by *CsMADS6* (Lu *et al.*, 2018). The regulation of carotenogenic genes in monocots is not well understood, even though cereals constitute a key target for metabolic engineering of the carotenoid pathway (Paine *et al.*, 2005; Zhu *et al.*, 2008; Bai *et al.*, 2016).

I focused on the regulation of *ZmBCH2* because this gene controls the accumulation of β -carotene in maize endosperm (Vallabhaneni *et al.*, 2009; Yan *et al.*, 2010), where the gene is preferentially expressed (Li *et al.*, 2010). The *cis*-regulatory elements and corresponding transcription factors which control the endosperm-preferred expression of *ZmBCH2* have not yet been identified. I hypothesized that important regulatory elements would be located not only in the upstream promoter region, but also potentially in the 5'-UTR containing the first intron, due to the common phenomenon of intron-mediated transcriptional enhancement in plants (Hernandez-Garcia and Finer, 2014; Gallegos and Rose, 2015). I therefore tested a full-length 5'-flanking region spanning from position -1773 in the *ZmBCH2* upstream promoter to position +263, which includes the entire 5'-UTR (and first intron). When linked to the reporter gene *gusA*, this construct achieved strong GUS expression in the endosperm and embryo of transgenic rice seeds (Fig. 2b), but when the 263-bp 5'-UTR was removed, the GUS activity was reduced by 8.4-fold. I therefore concluded that transcriptional enhancement mediated by the 5'-UTR is an important component of the regulation of *ZmBCH2*.

I constructed a series of 5' deletions progressively removing parts of the upstream promoter, but found that even the shortest construct (PF Δ 4-5'UTR), spanning from position -511 to +263, contained all the necessary *cis*-regulatory elements responsible for high-level *gusA* expression and strong GUS activity in transgenic rice seeds. The mechanism of intron-mediated enhancement is not well understood (Hernandez-Garcia and Finer, 2014; Gallegos and Rose, 2015) so it will be interesting to investigate how the *ZmBCH2* 5'-UTR (including the first intron) boosts the expression of *gusA* in transgenic rice by conducting loss-of-function experiments (site-directed mutagenesis) using the above minimum promoter (PF Δ 4-5'UTR).

Among the four *cis*-regulatory elements found in the minimum promoter, a single P-box and a single AACA motif were located within the first 300bp upstream of the transcription start site, often the key regulatory region in endosperm-specific genes. The P-box is a highly conserved 7-bp sequence (5'-TG(T/A/C)AAA(G/A)-3') found in the promoters of many cereal seed storage protein genes (Colot *et al.*, 1987; Vicente-Carbajosa *et al.*, 1997). As previously demonstrated using electrophoretic mobility shift assays (EMSAs), maize PBF (Vicente-Carbajosa *et al.*, 1997) and barley and wheat PBFs (Mena *et al.*, 1998) can bind to storage protein gene promoters containing the

P-box. Similarly, EMSAs have also revealed that the AACA motif is recognized by MYB-class transcription factors such as OsMYB5 (Suzuki *et al.*, 1998) and HvGAMYB (Diaz *et al.*, 2002). I confirmed that ZmPBF and ZmGAMYB can interact with the P-box and AACA motifs, respectively, in the proximal region of the truncated *ZmBCH2* promoter by using yeast one-hybrid assays. The minimal promoter fragment (774 bp) also contains two Skn-1 motifs spanning positions –393 to –389 and –434 to –430, which are involved in the endosperm-specific expression of storage protein genes in cereal grains (Washida *et al.*, 1999). The Skn-1 motif in the promoter of the rice storage protein glutelin gene *GluB-1* was found to be necessary for high expression levels in the endosperm in cooperation with other motifs (AACA, GCN4 and ACGT) by site-directed mutagenesis (Washida *et al.*, 1999). It will be interesting to investigate whether the two Skn-1 motifs are important for the regulation of *ZmBCH2* expression.

Having identified likely candidates for the transactivation of the minimal *ZmBCH2* promoter, I confirmed that the genes for both transcription factors (*ZmPBF* and *ZmGAMYB*) were preferentially transcribed in maize endosperm and regulated developmentally in a similar manner to *ZmBCH2* itself (Fig. 2.4). I then carried out transient expression experiments in maize, which revealed that the expression of *ZmPBF* and/or *ZmGAMYB* upregulated the endogenous *ZmBCH2* gene in the endosperm and embryo (Fig. 2.7). Taken together, these results confirmed that *ZmPBF*, *ZmGAMYB* and *ZmBCH2* have similar spatiotemporal expression profiles, so the transcription factors have the opportunity to regulate *ZmBCH2*, and that when present they can transactivate the *ZmBCH2* promoter, strongly supporting their role in the regulation of *ZmBCH2* in planta. The 5'-flanking region of *ZmPSY1* contains one AACA motif (–827 to –821), two P-boxes (–1221 to –1215 and –1341 to –1335) and one reverse and complementary P-box (–1600 to –1594) in the distal region (Fig. 2.8). However, the transient expression of *ZmPBF* and/or *ZmGAMYB* did not significantly upregulate endogenous *ZmPSY1* expression in maize endosperm and embryo tissues, suggesting these elements are not essential for the regulation of *ZmPSY1* promoter activity (Fig. 2.9). Other transcription factors are therefore needed for the control of *ZmPSY1* expression in maize endosperm.

I also investigated the regulatory potential of ZmPBF and ZmGAMYB in transgenic rice plants expressing *gusA* under the control of the minimal *ZmBCH2* promoter (PFΔ4-

5'UTR), because this isolates the regulatory controls from any potentially confounding factors such as the maize Opaque2 (O2) transcription factor, which interacts with *ZmPBF* in maize (Vicente-Carbajosa *et al.*, 1997). In the absence of either transcription factor, GUS activity in the transgenic rice lines expressing the PFA4-5'UTR-*gusA* construct alone was low but detectable in seeds, probably reflecting the presence of rice orthologs of these maize transcription factors, but negligible in leaves (Fig. 2b), where even rice orthologs of *ZmPBF* and *ZmGAMYB* would be absent. The expression of either transcription factor stimulated the expression of *gusA*, whereas the presence of both showed an additive effect (Fig. 2.11A). The overexpression of both *ZmPBF* and *ZmGAMYB* did not result in additive transactivation of the *ZmBCH2* promoter in the maize endosperm and embryo (Fig. 2.7), which might be due to the higher levels of endogenous *ZmBCH2* mRNA in maize endosperm and embryo tissue (Fig. 2.4). I did not detect *gusA* reporter gene expression or GUS activity in wild-type rice leaf or seeds. My results indicate that the *ZmPBF* and *ZmGAMYB* transcription factors act in a largely independent manner in both maize and rice. Similarly, the combination of rice PBF and GAMYB with the 0.2 kb *GluB-1* promoter linked to *gusA* did not reveal any significant synergistic transactivation, again suggesting little or no interaction between the two transcription factors (Yamamoto *et al.*, 2006). In contrast, barley PBF interacts with the GAMYB transcription factor to activate barley endosperm-specific genes during seed development (Diaz *et al.*, 2002) suggesting that the regulatory mechanism used by these transcription factors is species-dependent. The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9) system is an effective tool for targeted gene mutagenesis in cereals including maize (Zhu *et al.*, 2017). Knockouts of *ZmPBF* and/or *ZmGAMYB* in maize using CRISPR/Cas9 could therefore provide further insight into the regulatory mechanisms leading to the biosynthesis and accumulation of β -carotene in maize grains.

In summary, the maize gene encoding β -carotene hydroxylase 2 (*ZmBCH2*) is regulated by a combination of *cis*-regulatory elements in the upstream promoter and 5'-UTR (including the first intron). The upstream promoter contains sufficient information to restrict *ZmBCH2* expression to the seed, but the 5'-UTR is required for high-level expression, which suggests that *ZmBCH2* expression is partially dependent on 5'-UTR-mediated transcriptional enhancement. Nested deletion analysis revealed that two *cis*-regulatory elements determine the activity of the *ZmBCH2* promoter, namely the

proximal AACA motif and P-box in the -300 bp region of the promoter, which is consistent with results from other genes preferentially expressed in seeds. I isolated the candidate transcription factors (*ZmPBF* and *ZmGAMYB*) that bind these sequences based on the known regulatory mechanisms in other cereal gene promoters and confirmed that (a) the corresponding genes (*ZmPBF* and *ZmGAMYB*) are expressed in a spatiotemporally consistent profile shared with *ZmBCH2*; (b) the transcription factors physically interact with the corresponding elements in yeast one-hybrid assays; and (c) the transient expression of either or both transcription factors in maize increases the expression of endogenous *ZmBCH2*. These data show that the two transcription factors act independently. In transgenic rice plants expressing the minimal *ZmBCH2* promoter *gusA* construct (allowing us to isolate the regulatory controls), each transcription factor was able to transactivate the *gusA* expression when present alone, and when both were present the effect was additive. Statistical analysis to confirm the independent activity of the two factors revealed a stronger correlation when the quantity of each transcription factor was built into the model compared to a model based purely on presence/absence, providing a strong confirmation of our original hypothesis that the proximal P-box and AACA motifs in the *ZmBCH2* promoter are the sites bound in planta by *ZmPBF* and *ZmGAMYB*, respectively. Results indicate that transcriptional regulation by two transcription factors provides an elegant strategy for the regulation of carotenogenic gene expression in metabolically engineered plants. This sets the stage for the broader application of this approach (i.e. transcriptional control mediated by independently-acting transcription factors) in other metabolic pathways.

2.6. Conclusions

Results showed both transcription factors *ZmPBF* and *ZmGAMYB* can up-regulate the expression of *ZmBCH2* in maize endosperm and embryo at different development stages. The expression analysis of *gusA* fused with the minimal promoter of *ZmBCH2* in transgenic rice indicated high *gusA* expression as long as a prolamin box and AACA motif were present.

2.7. References

Bai C, Capell T, Berman J, Medina V, Sandmann G, Christou P, Zhu C (2016). Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor-product balance. *Plant Biotechnol. J.*, **14**:195-205.

Berman J, Zorrilla-López U, Sandmann G, Capell T, Christou P, Zhu C (2017). The silencing of carotenoid β -hydroxylases by RNA interference in different maize genetic backgrounds increases the β -carotene content of the endosperm. *Int. J. Mol. Sci.*, **18**. Pii: E2525. Doi: 10.3390/ijms18122515.

Christou P, Ford TL, Kofron M (1991). Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. *Biotechnol.*, **9**:957-962.

Colot V, Robert LS, Kavanagh TA, Bevan MW, Thompson RD (1987). Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. *EMBO J.*, **6**:3559-3564.

Diaz I, Vicente-Carbajosa J, Abraham Z, Martínez M, Isabel-La Moneda I, Carbonero P (2002). The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant J.*, **29**:453-464.

Diaz I, Martinez M, Isabel-La Moneda I, Rubio-Somoza I, Carbonero P (2005). The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J.*, **42**:652-662.

Farré G, Blancquaert D, Capell T, Van Der Straeten D, Christou P, Zhu C (2014). Engineering complex metabolic pathways in plants. *Annu. Rev. Plant Biol.*, **65**:187-223.

Fraser PD, Bramley PM (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.*, **43**:228-265.

Gallegos JE, Rose AB (2015). The enduring mystery of intron-mediated enhancement. *Plant Sci.*, **237**:8-15.

Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA (2008). Metabolic engineering of carotenoid biosynthesis in plants. *Trends Biotechnol.*, **26**:139-145.

Giuliano G (2017). Provitamin A biofortification of crop plants: a gold rush with many miners. *Curr. Opin. Biotechnol.*, **44**:169-180.

Gubler F, Kalla R, Roberts JK, Jacobsen JV (1995). Gibberellin-regulated expression of a *myb* gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. *Plant Cell*, **7**:1879-1891.

Gubler F, Raventos D, Keys M, Watts R, Mundy J, Jacobsen JV (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. *Plant J.*, **17**:1-9.

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, **319**:330-333.

Hernandez-Garcia CM, Finer JJ (2014). Identification and validation of promoters and *cis*-acting regulatory elements. *Plant Sci.*, **217-218**:109-119.

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.*, **27**:297-300.

Hwang YS, Ciceri P, Parsons RL, Moose SP, Schmidt RJ, Huang N (2004). The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. *Plant Cell Physiol.*, **45**:1509-1518.

Jefferson AR, Kavanagh TA, Bevan MW (1987). GUS fusion: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.*, **6**:3901-3907.

Kosugi S, Ohashi Y, Nakajima K, Arai Y (1990). An improved assay for β -glucuronidase in transformed cells: methanol almost completely suppresses a putative endogenous β -glucuronidase activity. *Plant Sci.*, **70**:133-140.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, **30**:325-327.

Li Q, Farré G, Naqvi S, Breitenbach J, Sanahuja G, Bai C, Sandmann G, Capell T, Christou P, Zhu C (2010). Cloning and functional characterization of the maize carotenoid isomerase and β -carotene hydroxylase genes and their regulation during endosperm maturation. *Transgenic Res.*, **19**:1053-1068.

Lu S, Zhang Y, Zhu K, Yang W, Ye J, Chai, L, Xu, Q, Deng X (2018). The *Citrus* transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. *Plant Physiol.*, **176**:2657-2676.

Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (2011). The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. *Plant Physiol.*, **157**:1568-1579.

Marzábal P, Gas E, Fontanet P, Vicente-Carbajosa J, Torrent M, Ludevid MD (2008). The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Mol Biol.*, **67**:441-454.

Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P (1998). An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. *Plant J.*, **16**:53-62.

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.*, **23**:482-487.

Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

Suzuki A, Wu CY, Washida H, Takaiwa F (1998). Rice MYB protein OSMYB5 specifically binds to the AACCA motif conserved among promoters of genes for storage protein glutelin. *Plant Cell Physiol.*, **39**:555-559.

Toledo-Ortiz G, Huq E, Rodriguez-Concepcion M (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proc. Natl. Acad. Sci. USA.*, **107**:11626-11631.

Vallabhaneni R, Gallagher CE, Licciardello N, Cuttriss AJ, Quinlan RF, Wurtzel ET (2009). Metabolite sorting of a germplasm collection reveals the *hydroxylase3* locus as a new target for maize provitamin A biofortification. *Plant Physiol.*, **151**:1635-1645.

Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997). A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc. Natl. Acad. Sci. USA*, **94**:7685-7690.

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F (1999). Identification of *cis*-regulatory elements required for endosperm expression of the rice storage protein glutelin gene *GluB-1*. *Plant Mol Biol.*, **40**:1-12.

Welsch R, Maass D, Voegel T, DellaPenna D, Beyer P (2007). Transcription factor RAP2.2 and its interacting partner SINAT2: stable elements in the carotenogenesis of *Arabidopsis* leaves. *Plant Physiol.*, **145**:1073-1085.

Wu C, Washida H, Onodera Y, Harada K, Takaiwa F (2000). Quantitative nature of the Prolamin-box, ACGT and AACCA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. *Plant J.*, **23**:415-421.

Xiong C, Luo D, Lin A, Zhang C, Shan L, He P, Li B, Zhang Q, Hua B, Yuan Z, Li H, Zhang J, Yang C, Lu Y, Ye Z, Wang T (2018). A tomato B-box protein SIBBX20 modulates carotenoid biosynthesis by directly activating *PHYTOENE SYNTHASE 1*, and is targeted for 26S proteasome-mediated degradation. *New Phytol.* doi: 10.1111/nph.15373.

Yan J, Kandianis CB, Harjes CE, Bai, L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, Fernandez MG, Zaharieva M, Babu R, Fu Y, Palacios N, Li J, Dellapenna D,

Brutnell T, Buckler ES, Warburton ML, Rocheford T (2010). Rare genetic variation at *Zea mays crtRB1* increases β -carotene in maize grain. *Nat. Genet.*, **42**:322-327.

Yamamoto MP, Onodera Y, Touno SM, Takaiwa F (2006). Synergism between RPBF Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol.*, **141**:1694-1707.

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T (2008). Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc. Natl. Acad. Sci. USA*, **105**:18232-18237.

Zhu C, Bortesi L, Baysal C, Twyman RM, Fischer R, Capell T, Schillberg S, Christou P (2017). Characteristics of genome editing mutations in cereal crops. *Trends Plant. Sci.*, **22**:38-52.

Zhu F, Luo T, Liu C, Wang Y, Yang H, Yang W, Zheng L, Xiao X, Zhang M, Xu R, Xu J, Zeng Y, Xu J, Xu Q, Guo W, Larkin RM, Deng X, Cheng Y (2017). An R2R3-MYB transcription factor represses the transformation of α - and β -branch carotenoids by negatively regulating expression of *CrBCH2* and *CrNCED5* in flavedo of *Citrus reticulata*. *New Phytol.*, **216**:178-192.

Chapter III

The two isoforms of isopentenyl diphosphate isomerase in rice (*Oryza sativa*) are targeted to different subcellular compartments and have distinct roles in isoprenoid biosynthesis

3.0. Abstract

Isoprenoids are synthesized from the precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are interconverted by the enzyme isopentenyl diphosphate isomerase (IPPI). Many plants express multiple isoforms of IPPI, the only enzyme shared by the mevalonate (MVA) and non-mevalonate (MEP) pathways, but little is known about their specific roles. Rice (*Oryza sativa*) has two IPPI paralogs (*OsIPPI1* and *OsIPPI2*). I therefore carried out a comprehensive comparison of IPPI gene expression and isoprenoid analysis in this species. I found that *OsIPPI1* mRNA was more abundant than *OsIPPI2* mRNA in all tissues, and its expression in de-etiolated leaves mirrored the accumulation of phytosterols, suggesting a key role in the synthesis of MVA pathway isoprenoids. I investigated the subcellular localization of both isoforms by constitutively expressing them as fusions with synthetic green fluorescent protein and detecting them by confocal fluorescence microscopy and immuno-electron microscopy. Both proteins localized to the endoplasmic reticulum (ER) as well as peroxisomes and mitochondria, whereas only *OsIPPI2* was detected in plastids, reflecting the presence of an N-terminal transit peptide, not present in *OsIPPI1*. Despite the plastidial location of *OsIPPI2*, the expression of *OsIPPI2* mRNA did not mirror the accumulation of chlorophylls or carotenoids, indicating that *OsIPPI2* is a dispensable component of the MEP pathway. The detection of both isoforms in the ER indicates that DMAPP can be synthesized *de novo* in this compartment. This work provides novel insight into the role of the ER in the synthesis of MVA-derived isoprenoids.

3.1. Introduction

All isoprenoids are derived from the C₅ building blocks isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are interconverted by the enzyme isopentenyl diphosphate isomerase (IPPI). Two distinct forms of the enzyme are found in nature. Type I IPPIs are Zn-containing metalloenzymes whereas type II IPPIs require flavin mononucleotide (FMN) and a reducing agent (NADPH) for activity (Thibodeaux and Liu, 2017).

IPP can be synthesized via two independent pathways: the mevalonate (MVA) pathway and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, also known as the non-mevalonate pathway (Vranova *et al.*, 2013). Archaea, many eukaryotes and some bacteria exclusively use the MVA pathway whereas most bacteria exclusively use the MEP pathway. Some bacteria as well as all higher plants, have the ability to use both pathways (Vranova *et al.*, 2013). Some isoprenoids in plants are derived solely from the MVA pathway, including phytosterols, sesquiterpenes, triterpenes and the side-chain of ubiquinone, whereas monoterpenes, diterpenes, carotenoids and the side chains of chlorophylls and tocopherols are all derived from the MEP pathway (Vranova *et al.*, 2013; Zhu *et al.*, 2013; Nogueira *et al.*, 2018). The MEP pathway in plants is compartmentalized in the plastids whereas the MVA pathway operates in the rest of the cytoplasm, thus allowing the optimization of isoprenoid biosynthesis and regulation according to the availability of fixed carbon and ATP (Vranova *et al.*, 2013).

Many plants produce different isoforms of IPPI. Two cDNAs encoding IPPIs have been isolated from Arabidopsis (*Arabidopsis thaliana*), lettuce (*Lactuca sativa*), tobacco (*Nicotiana tabacum*), tomato (*Solanum lycopersicum*) and the ornamental plant *Adonis aestivalis* (Cunningham and Gantt, 2000; Nakamura *et al.*, 2001; Pankratov *et al.*, 2016). In tobacco, one isoform features a transit peptide sequence (TPS) and is localized in the plastids, whereas the other lacks this sequence and remains in the cytosol (Nakamura *et al.*, 2001). In castor bean (*Ricinus communis*), there is biochemical evidence for the existence of mitochondrial and plastid IPPIs (Green *et al.*, 1975). Different isoforms of IPPI in Arabidopsis (Phillips *et al.*, 2008; Sapir-Mir *et al.*, 2008) and *Catharanthus roseus* (Guirimand *et al.*, 2012) are targeted to different subcellular

compartments, and this has been attributed to the expression of each *IPPI* gene as alternative transcripts.

Although only a single rice (*Oryza sativa*) *IPPI* cDNA has been reported (Cunningham and Gantt, 2000), the genome contains two genes named *IPPI1* (Os07g36190) and *IPPI2* (Os05g34180) (Jung *et al.*, 2008). Little is known about the expression, subcellular localization and function of the corresponding enzymes in the MVA and MEP pathways. Given the importance of the C₅ building blocks IPP and DMAPP for isoprenoid biosynthesis in rice, I therefore carried out a comprehensive analysis of gene expression and protein subcellular localization for both enzymes. I also investigated the effect of each gene on the biosynthesis of chlorophylls, carotenoids, and phytosterols, and investigated the expression of *OsIPPI1* and *OsIPPI2* in 14-day-old etiolated leaves during de-etiolation, which is characterized by a burst of isoprenoid biosynthesis.

3.2. Aims

The major aim of the work described in this chapter was to analyze the expression pattern and subcellular localization of *OsIPPI* the only enzyme which is shared between the MVA and MEP pathways that produce isoprenoids in higher plants. A further aim was to elucidate the role of *OsIPPI* gene in isoprenoid biosynthesis in rice.

3.3. Materials and methods

3.3.1. Plant materials

Wild-type rice (*Oryza sativa* L. cv. EY1105) and transgenic plants were grown in a controlled growth chamber with a 28/20°C day/night temperature regime, a 10h photoperiod and 60–90% relative humidity for the first 50 days, followed by maintenance at 21/18°C (day/night) with a 16-h photoperiod thereafter. Transgenic rice plants were self-pollinated to obtain seeds. To obtain 7-day- and 14-day-old rice plants, wild type rice seeds were sterilized in 3.7% NaOCl for 20min and washed with sterilized water thoroughly, then were aseptically grown on Murashige and Skoog (MS) medium containing 0.3% phytigel in pots in a controlled growth chamber at 25°C with a 10h photoperiod; the 14-day-old etiolated rice seedlings were grown in complete

darkness. Fourteen-day-old etiolated rice seedlings were illuminated with continuous white light ($100 \mu\text{mol m}^{-2} \text{s}^{-1}$) for 0.5, 1, 2, 4, 6, 8, 10, 12 or 24h. Flag leaf, endosperm, embryo, immature panicle, ovary, inner glume and outer glume tissues were collected from wild type rice grown in a controlled growth chamber; the embryo and endosperm were collected 25 days after pollination. All plant materials were rapidly frozen in liquid nitrogen and stored at -80°C .

3.3.2. RNA isolation and cDNA synthesis

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and DNA was removed using DNase I from the RNase-free DNase set (Qiagen). The RNA was quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and $1\mu\text{g}$ of total RNA was used as the template for first-strand cDNA synthesis with Ominiscript Reverse Transcriptase (Qiagen) in a $20\mu\text{l}$ total reaction volume according to the manufacturer's recommendations,

3.3.3. Cloning and sequencing the *OsIPPI1* and *OsIPPI2* cDNAs

The *OsIPPI1* and *OsIPPI2* cDNAs were cloned directly from rice leaf mRNA by RT-PCR using *OsIPPI1* forward primer 5'-ATGGCCGGCGCCGCCGCCG-3' and reverse primer 5'-TTACTTCAGCTTGTGGATGG-3', or *OsIPPI2* forward primer 5'-ATGGCGGGCGGGCGGCGCT-3' and reverse primer 5'-CTACAACCTTATGAATTGTTTTTC-3'. The primers were designed to flank the ORF of each gene based on sequences in GenBank (accession numbers AK065871 and NM_001062082). The RT-PCR products were transferred to vector pCRII TOPO (Invitrogen, Carlsbad, CA, USA) to generate pCR-*OsIPPI1* and pCR-*OsIPPI2* for sequencing.

3.3.4. Construction of plasmids

For functional analysis, two pairs of oligonucleotide primers with terminal EcoRI and BamHI restriction sites were used to amplify *OsIPPI1* and *OsIPPI2* using pCR-*OsIPPI1* and pCR-*OsIPPI2* as the templates. For *OsIPPI1*, we used forward primer 5'-GAATTCCATGGCCGGCGCCGCCGCCG-3' (EcoRI site underlined) and reverse

primer 5'-GGATCCGGACTTCAGCTTGTGGATG-3' (BamHI site underlined). For *OsIPPI2*, we used forward primer 5'-GAATTCCATGGCGGCGGCGGCGGCGCTC-3' (EcoRI site underlined) and reverse primer 5'-GGATCCGTACAACTTATGAATTGTTTTTC-3' (BamHI site underlined). The PCR products were introduced into pUC8 at the EcoRI/BamHI sites to generate vectors pUC8-*OsIPPI1* and pUC8-*OsIPPI2* for functional complementation in *E. coli* cells producing β -carotene (Kajiwara *et al.*, 1997). The integrity of all intermediate and final constructs was confirmed by sequencing as above.

For the analysis of subcellular localization study I generated two plasmids in which the rice IPPI isoforms were fused to sGFP and expressed under the control of the CaMV35S promoter and *nos* terminator. This was achieved by inserting the sGFP coding region and *nos* terminator flanked by XmaI and EcoRI sites into the same sites of pUC8-CaMV35SPro-*gusA*-TNos (Jin *et al.*, 2018) to yield the intermediate construct pUC8-CaMV35SPro-*sGFP*-TNos, which was subsequently digested with XbaI and BamHI to allow the insertion of *OsIPPI1* or *OsIPPI2* cDNAs without stop codons, yielding the final vectors pUC8-CaMV35SPro-*OsIPPI1-sGFP*-TNos and pUC8-CaMV35SPro-*OsIPPI2-sGFP*-TNos, respectively. The *OsIPPI1* cDNA without the stop codon was generated using forward primer 5'-TCTAGACATGGCCGGCGCCGCGCCGCGC-3' (XbaI site underlined) and reverse primer 5'-GGATCCGGACTTCAGCTTGTGGATG-3' (BamHI site underlined). The *OsIPPI2* cDNA without the stop codon was generated using forward primer 5'-TCTAGACATGGCGGCGGCGGCGGCGCTC-3' (XbaI site underlined) and reverse primer 5'-GGATCCGTACAACTTATGAATTGTTTTTC-3' (BamHI site underlined). The integrity of all intermediate and final constructs was confirmed by sequencing as above.

3.3.5. Functional complementation analysis

Functional complementation experiments were carried out as previously described (Kajiwara *et al.*, 1997). *E. coli* DH5 α cells were transformed with pACCAR16 Δ crtX, which contains the carotene biosynthesis genes *crtE*, *crtB*, *crtI* and *crtY* encoding GGPP synthase, phytoene synthase, phytoene desaturase and lycopene β -cyclase, respectively (Misawa *et al.*, 1995) for the generation of β -carotene. The cells were simultaneously

transformed with the pUC8 empty vector as a control or with the pUC8-OsIPPI1 or pUC8-OsIPPI2 vectors carrying the rice IPPI cDNAs, as described above.

3.3.6. Measurement of β -carotene levels in *E. coli*

E. coli cells transformed as above were plated on lysogeny broth (LB) agar plates supplemented with 100 μ g ml⁻¹ ampicillin, 37 μ g ml⁻¹ chloramphenicol and 1.0mM IPTG and incubated at 37°C overnight. Individual colonies on plates were transferred to a 50ml Falcon centrifuge tube containing 5ml of liquid LB medium with antibiotics as above and incubated at 28°C for 12h on a rotary-shaker at 180rpm. Subsequently the primary culture was transferred to a 250ml flask containing 50ml LB medium plus antibiotics. IPTG was added to a final concentration of 0.5mM when the optical density at 600nm reached 0.6, and the cultures were incubated for a further 48h under the same conditions. The cells were pelleted and freeze-dried before β -carotene was extracted from 50mg cell aliquots in 20ml methanol containing 6% KOH. The suspension was heated at 60°C for 20min and partitioned in 20ml 10% (v/v) diethyl ether in petroleum ether. The upper phase was collected and the solvent was evaporated to dryness under a stream of nitrogen gas. The pellets were redissolved with 20ml petroleum ether for quantitation of β -carotene using the appropriate extinction coefficient (Davies, 1976).

3.3.7. Rice transformation

Seven-day-old mature rice (cv. EYI105) embryos were used as explants for particle bombardment transformation with 0.4 mg gold particles coated with the transgene construct (individual pUC8-CaMV35SPro-*OsIPPI1-sGFP*-TNos, pUC8-CaMV35SPro-*OsIPPI2-sGFP*-TNos, or pUC8-CaMV35SPro-*sGFP*-TNos) and a plasmid containing hygromycin phosphotransferase (*hpt*) as a selectable maker gene at a 3:1 molar ratio as previously described (Jin *et al.*, 2018). Transgenic plantlets were regenerated and hardened off in soil as described in chapter II.

3.3.8. Protein extraction and western blot analysis

Leaf tissue (0.1–0.2g) was ground in liquid nitrogen, thawed in an equal volume of extraction buffer (20mM Tris-HCl pH 7.5, 5mM EDTA, 0.1% Tween-20, 0.1% SDS,

2mM PMSF) and vortexed for 1 h at 4°C. Cell debris was removed by centrifugation at 15,300 g for 20 min at 4°C and the protein concentration in the supernatant was determined using the Bradford assay (AppliChem, Darmstadt, Germany). The protein samples (50 µg per lane) were separated by SDS-PAGE at 200 V for 60 min followed by transfer to an Immobilon-FL transfer membrane (Merck, Darmstadt, Germany) using a semidry transfer apparatus (Bio-Rad, Hercules, CA, USA) at 20V for 45min. The membrane was immersed in 5% non-fat milk in 0.2M Tris-HCl pH 7.6, 1.37 M NaCl, 0.1% Tween-20 (TBST) for 1h at room temperature. The proteins were incubated with an anti-GFP polyclonal antibody (Sigma-Aldrich, Saint Louis, MO, USA) diluted 1:2000 in 5% non-fat milk in TBST overnight at 4°C. After three rinses in TBST for 10 min each, the membrane was incubated with a goat anti-rabbit secondary antibody conjugated to alkaline phosphatase (Sigma-Aldrich) diluted 1:5000 in 2% non-fat milk in TBST for 1h at room temperature. After three further rinses as above, the signal was detected using SIGMAFAST BCIP/NBT tablets (Sigma-Aldrich).

3.3.9. Gene expression analysis by quantitative real-time RT-PCR

First-strand cDNA was synthesized from 1µg total RNA using Ominiscript Reverse Transcriptase in a 20µl total reaction, and quantitative real-time RT-PCR (qRT-PCR) was carried out as previously described (Jin *et al.*, 2018) using the primers listed in Table 3.1. The amplified DNA fragments for each gene were confirmed by Sanger sequencing. The expression levels in different samples were normalized against rice *actin* mRNA. Five biological replicates each comprising three technical replicates were tested for each sample.

Table 3.1. Primers used for qRT-PCR analysis

Gene	Forward sequence (5'-3')	Reverse sequence (5'-3')
<i>OsIPPI1</i>	GGTGGATGAACAAGACAATG	ATCGTTGCTGGAGTAGGAGT
<i>OsIPPI2</i>	TTTAGTGGACGAACAGGACA	GCAGACCTTTGCTGAAGTAAC
<i>OsHDR1</i>	GCTGAATGAGAAGAAGCTGC	AAAGGAAGCAGTGGCAAC
<i>OsHDR2</i>	GAGTAGGCGTTGTGAATCAAAC	CCTTTCCTGAGTAGCATTGC
<i>OsActin</i>	GACTCTGGTGATGGTGTGACG	GCTTCTCCTTTATGTCTCTGAC

OsIPPI1, rice (*Oryza sativa*) isopentenyl diphosphate isomerase 1 gene; *OsIPPI2*, rice isopentenyl diphosphate isomerase 2 gene; *OsHDR1*, rice 4-hydroxy-3-methylbut-2-enyl

diphosphate reductase 1 gene; *OsHDR2*, rice 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 2 gene; *OsActin*, rice actin gene.

3.3.10. DNA sequence analysis and prediction of protein subcellular localization

The GRAMENE database (<http://www.gramene.org/>) and GenBank (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) were searched for homologous sequences using BLAST. Nucleotide and deduced amino acid sequence comparisons, and multiple sequence alignments, were carried out using Vector NTI Advance 11 (Invitrogen, Carlsbad, CA). Gene structures were predicted by aligning mRNA to genomic DNA using Spidey (<http://www.ncbi.nlm.nih.gov/spidey/>). Subcellular localization was predicted using iPSORT (<http://ipsort.hgc.jp/>; Bannai *et al.*, 2002), Predotar (<https://urgi.versailles.inra.fr/Tools/Predotar>), RSLpred (<http://crdd.osdd.net/raghava/rsllpred/> Kaundal and Raghava, 2009) and LocTree3 (<https://roslab.org/services/loctree3/>). Protein sequences were screened for chloroplast signal peptides using the ChloroP v1.1 server (<http://www.cbs.dtu.dk/services/ChloroP/>; Emanuelsson *et al.*, 1999) and TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>; Emanuelsson *et al.*, 2007). MitoFates (<http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi>; Fukasawa *et al.*, 2015), and the Peroxisomal Targeting Signal 1 (PTS1) Predictor (<http://mendel.imp.ac.at/pts1/>; Neuberger *et al.*, 2003) were used to screen for mitochondrial and peroxisomal signal peptides, respectively.

3.3.11. Confocal microscopy

Leaves from rice lines expressing pUCCaMV35SPro-*OsIPPI1-sGFP*-TNos or pUC35SPro-*OsIPPI2-sGFP*-TNos were analyzed by confocal microscopy to detect the GFP signal. Leaves from T3 homozygous transgenic lines expressing pUCCaMV35SPro-*OsIPPI1-sGFP*-TNos or pUC35SPro-*OsIPPI2-sGFP*-TNos were analyzed to reveal the subcellular localization of *OsIPPI1-sGFP* and *OsIPPI2-sGFP*. Leaves from wild-type rice and transgenic lines expressing pUC35s-*sGFP*-TNos were used as negative and positive controls, respectively. Fresh small leaf pieces (1 x 10mm) were fixed with 2% paraformaldehyde in 0.1M sodium phosphate buffer (pH 7.2) before semi-thin sections (30–40µm) were prepared using a CM3050 S Research

Cryostat (Leica Microsystems, Wetzlar, Germany). The sections were collected on poly-L-lysine glass microscope slides and 1024 x 1024 pixel images were captured using an FV1000 Laser Scanning Confocal Microscope (Olympus, Hamburg, Germany) fitted with a 100 x NA 1.40 UPlansApo (oil) objective. For GFP imaging, the excitation wavelength was 488nm provided by a multiline argon laser. Three independent transgenic rice lines for each event were analyzed.

3.3.12. Immuno-electron microscopy

Small leaf pieces (1x10mm) were fixed in 1% glutaraldehyde plus 1% paraformaldehyde in 0.1M sodium phosphate buffer (pH 7.2) for 16–24h at 4°C and washed three times (10min) with the same buffer. After fixation, samples were dehydrated in an alcohol series (30–100%) before embedding in Lowicryl K4M resin (Polysciences, Hirschberg an der Bergstrasse, Germany) in a cold-chamber at -20 to -35°C with polymerization by ultraviolet light. Ultrathin (70–90nm) sections were prepared using a Reichert-Jung ultra-cut E cryostat (Leica Microsystems). The sections were mounted on Formvar carbon-coated gold grids (200mesh) and processed for immunogold labeling using sGFP-specific monoclonal and polyclonal antibodies. The sections were incubated in blocking buffer for 15min. For the polyclonal antibody, the buffer was 200mM Tris-HCl (pH 7.4), 1% Tween-20, 0.1% gelatin and 1% bovine serum albumin (BSA). For the monoclonal antibody, the buffer was 10mM Tris-HCl (pH 7.4), 0.9% NaCl, 0.05% PEG 20,000 and 3% BSA. The grids were washed in distilled water and incubated overnight at 4°C with the primary polyclonal anti-eGFP (PA5-22688, Thermo Fisher Scientific) diluted 1:200, or the primary monoclonal anti-eGFP (11814460001, Sigma-Aldrich) diluted 1:500, with the corresponding blocking buffer as the diluent. After washing in distilled water, incubation in the blocking buffer (30min) and washing in distilled water, the grids were then incubated at room temperature for 1h with the secondary antibody (1:20 dilution in the appropriate blocking buffer). The secondary antibody was a goat-anti-rabbit IgG for the polyclonal primary antibody, or goat-anti-mouse IgG for the monoclonal antibody treatment, with both antibodies conjugated to 15nm gold particles (Electron Microscopy Sciences, Hatfield, PA, USA). Finally, the grids were contrasted with 1% uranyl acetate in water (20 min) and Reynold's lead citrate (2min) prior to observation using an EM 910

Transmission Electron Microscope (Zeiss, Oberkochen, Germany). I analyzed a minimum of two grids per antibody treatment.

The immunogold labeling of chloroplasts was also observed when the polyclonal antiserum was used in wild-type leaf samples. I therefore cross-absorbed the polyclonal antiserum (Errampalli and Fletcher, 1993) and repeated the analysis of leaves to confirm the specificity of chloroplast labeling, comparing the results to the outcome of the monoclonal antibody experiments. Three independent transgenic rice lines were analyzed for each event.

3.3.13. Ultra-high-performance liquid chromatography

Leaf samples were freeze-dried before analysis, and 10 mg of powdered tissue was extracted with 500 μ l water:methanol (50:50) and vortexed for 30 min at room temperature. I then added 500 μ l chloroform and vortexed the mixture before centrifuging at 15,300 g for 3min at room temperature. The upper phase was collected, evaporated, and re-dissolved in HPLC-grade ethyl acetate for UPLC analysis using an Acquity system (Waters, Watford, UK) with an Ethylene Bridged Hybrid (BEH C18) column (2.1 \times 100mm, 1.7mm) with a BEH C18 VanGuard pre-column (2.1 \times 50 mm, 1.7mm). Mobile phase A was methanol:water (50:50), and mobile phase B was acetonitrile:ethyl acetate (75:25) and the flow rate was 0.5ml min⁻¹. All solvents were passed through a 0.2mm filter before use. The gradient started at 30% A for 0.5 min and was then stepped to 0.1% A for 5.5min and then to 30% A for the last 2min. The column temperature was maintained at 30°C and the sample temperature at 8°C. Continuous online scanning was performed across the UV/visible range from 250 to 600nm, using a Waters extended wavelength photo-diode array detector (Nogueira *et al.*, 2013).

3.3.14. Gas chromatography–mass spectrometry

Leaf samples were freeze dried before analysis, and 10mg of powdered tissue was mixed with 400 μ l methanol and 400 μ l water and rotated for one hour. Chloroform (800 μ l) was added before vortexing the mixture thoroughly and centrifuging at 15,300 g for 5 min at room temperature. The chloroform and water phases were collected

separately and 10µg internal standards ribitol (polar) and d27-myristic acid (non-polar) before drying under nitrogen. Derivatization was performed with 30µl methoxyamine-HCl (Sigma-Aldrich, 20mg ml⁻¹ in pyridine) at 40°C for 1h, followed by incubation with 70µl N-methyl trimethylsilyl-trifluoro-acetamide (MSTFA, Macherey Nagel, Stockport, UK) for 2h at 40°C. GC-MS analysis was performed on an 7890A gas chromatograph with a 5977A MSD (Agilent Technologies, Santa Clara, CA, USA). Typically, samples were introduced using a split/splitless injector at 280°C with a 20:1 split ratio. Retention time locking to the internal standard was applied. The GC oven was held for 4 min at 70°C before ramping at 5°C min⁻¹ to 320°C. This final temperature was held for a further 10min. The interface with the MS was set at 250°C and MS performed in full scan mode using 70eV EI+ and scanned in the range 10–800 D. To identify chromatogram components found in the rice profiles, a mass spectral library was constructed from in-house standards as well as the NIST library version 2.0 (2008). Chromatograms were processed with AMIDS (version 2.69 (2010)) and a relative quantification to the internal standards was performed (Enfissi *et al.*, 2010).

3.4. Results

3.4.1. Cloning and characterization of the *OsIPPI1* and *OsIPPI2* cDNAs

The *OsIPPI1* and *OsIPPI2* cDNA sequences were cloned from leaf mRNA by RT-PCR using primers flanking the open reading frame (ORF), thus excluding any untranslated regions (UTRs). The *OsIPPI1* cDNA was 717bp in length encoding a predicted protein of 238 amino acids and an estimated mass of 27.34 kDa. The *OsIPPI2* cDNA was 882 bp in length encoding a predicted protein of 293 amino acids with an estimated mass of 32.59 kDa. The sequences showed 81.0% identity in the overlap region, but *OsIPPI2* included a 56-amino-acid N-terminal extension resembling a plastid-targeting TPS that was not present in *OsIPPI1* (Fig. 3.1). A multiple sequence alignment revealed at least 69.3% identity and 95.8% similarity (excluding TPS) among all known functional plant IPPIs (Fig.3.1). The deduced amino acid sequences of *OsIPPI1* and *OsIPPI2* featured conserved cysteine and glutamate residues at positions 139 and 207, respectively (Fig. 3.1), which correspond to essential active site residues identified in the yeast IPPI (Street *et al.*, 1994).

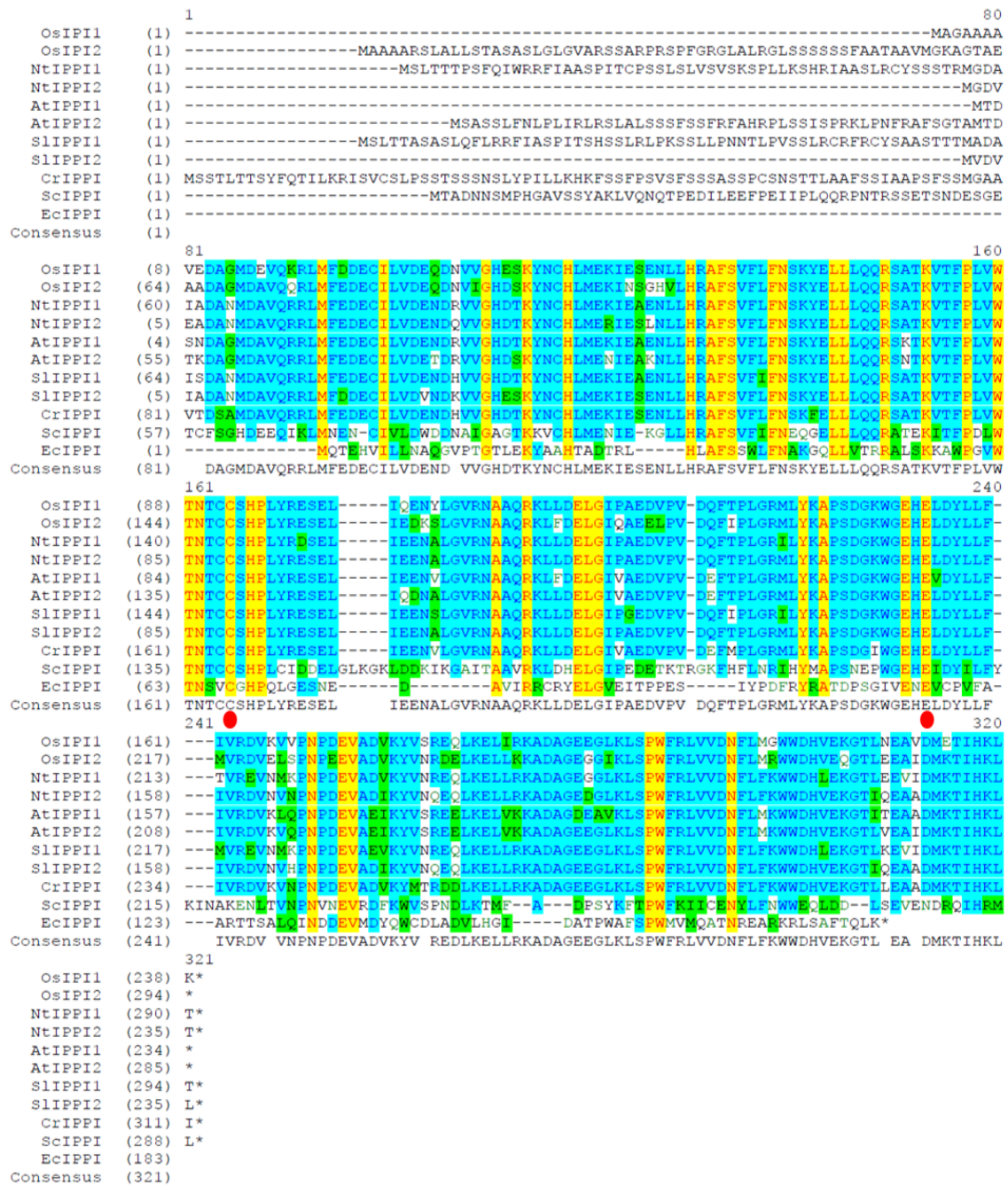


Figure 3.1. Multiple alignments of predicted amino acid sequences encoding isopentenyl diphosphate isomerases. IPPI, isopentenyl diphosphate isomerase; At, *Arabidopsis thaliana*; AtIP11, GenBank: AY065053; AtIP12, GenBank: AY143845; Cr, *Catharanthus roseus*; CrIP11, GenBank: EU135981; Ec, *Escherichia coli*; EcIP11, GenBank: NP_417365; Nt, *Nicotiana tabacum*; NtIP11, GenBank: NM_001326211; NtIP12, GenBank: NM_001325826; Os, *Oryza sativa*; OsIP11, GenBank: XM_015791312.2; OsIP12, GenBank: XM_015782225.2; *Saccharomyces cerevisiae*; ScIP11, GenBank: NP_015208; Sl, *Solanum lycopersicum*; SlIP11, gene: Solyc04g056390; SlIP12, gene: Solyc05g055760. The red spots indicate cysteine 139 and glutamic acid 207 residues in the active site.

The *OsIPPI1* and *OsIPPI2* cDNA sequences were used to screen rice genomic resources, revealing a single genomic sequence from chromosomes 7 and 5 of the Japonica Group with 100% identity, respectively, suggesting that both *OsIPPI1* and *OsIPPI2* are single-copy genes in the rice genome. The *OsIPPI1* gene has five introns and six exons, the same as *OsIPPI2* (Fig. 3.2). In both cases, exons 2–6 are similar in size whereas the translated portion of exon 1 is longer in *OsIPPI2* (242 bp) than *OsIPPI1* (74 bp) to account for the additional TPS. Both genes have similar size of intron sequence (*OsIPPI1* = 2055 bp; *OsIPPI2* = 2098 bp), but introns 1 and 5 are longer in *OsIPPI1* whereas introns 2–4 are longer in *OsIPPI2* (Fig. 3.2).

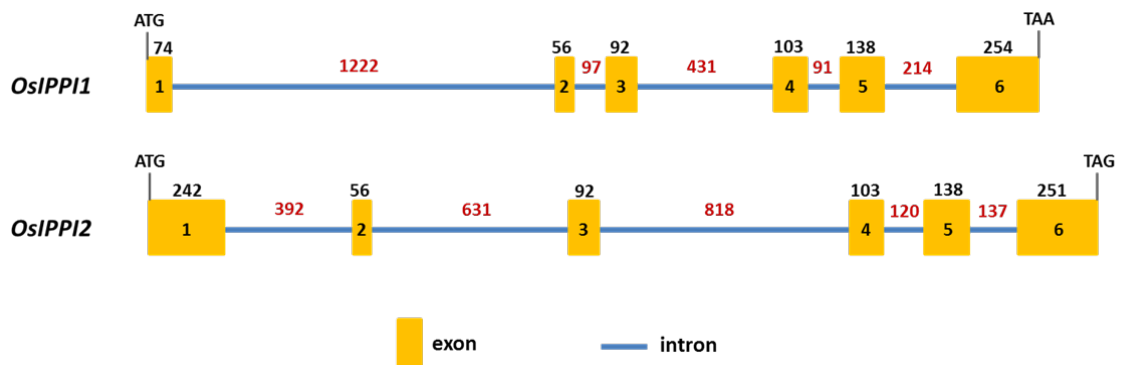


Figure 3.2. Partial structures of rice isopentenyl diphosphate isomerase genes, within the bounds of the open reading frame. Yellow boxes represent exons (or the translated parts of exons 1 and 6). Blue lines represent introns. The lengths of DNA sequences are indicated above of each exon and intron. Numbers inside the boxes indicate the number of corresponding exon. IPPI, isopentenyl diphosphate isomerase; *Os*, *Oryza sativa*; *OsIPPI1* cDNA, GenBank: XM_015791312.2; *OsIPPI1* genomic DNA, GenBank: AP014963.1; *OsIPPI2* cDNA, GenBank: XM_015782225.2; *OsIPPI2* genomic DNA, GenBank: AP014961.1).

3.4.2. *OsIPPI2* encodes a functional enzyme

Genetic color complementation assays in *Escherichia coli* are often used to determine the functionality of IPPI because the plasmid pACCAR16ΔcrtX allows the cells to produce β-carotene (Kajiwara *et al.*, 1997). I therefore introduced pACCAR16ΔcrtX into the *E. coli* DH5α strain along with either pUC8-*OsIPPI1* or pUC8-*OsIPPI2*, containing the rice *IPPI* cDNAs under the control of an inducible promoter, or the pUC8 empty vector as a control. Cells carrying pACCAR16ΔcrtX and pUC8 were able to grow on medium containing ampicillin and chloramphenicol (corresponding to the selectable markers on each plasmid) and produced pale yellow colonies representing the

accumulation of β -carotene when induced with isopropyl β -D-1-thiogalactopyranoside (IPTG) (Fig. 3.3A). Cells co-transformed with pACCAR16 Δ crtX and pUC8-OsIPPI1 formed colonies with a deeper yellow color, reflecting the additional flux provided by OsIPPI1 (Fig. 3.3A) and similar results were observed when cells were co-transformed with pACCAR16 Δ crtX and pUC8-OsIPPI2. Pigments were extracted from liquid cultures and the carotenoid levels were quantified. This revealed that co-transformation with pACCAR16 Δ crtX and the pUC8 empty vector resulted in a β -carotene yield of $340.31 \pm 39.71 \mu\text{g g}^{-1}$ DW. The addition of pUC8-OsIPPI1 resulted in a 3.2-fold increase to 1089.32 ± 117.46 addition of pUC8-OsIPPI2 resulted in a 2.8-fold increase to $950.74 \pm 90.83 \mu\text{g g}^{-1}$ DW (Fig. 3.3B). These data strongly indicate that *OsIPPI2* encodes a functional enzyme, as previously established for *E. coli* *IPPI* (Kajiwara *et al.*, 1997) and *OsIPPI1* (Cunningham and Gantt, 2000).

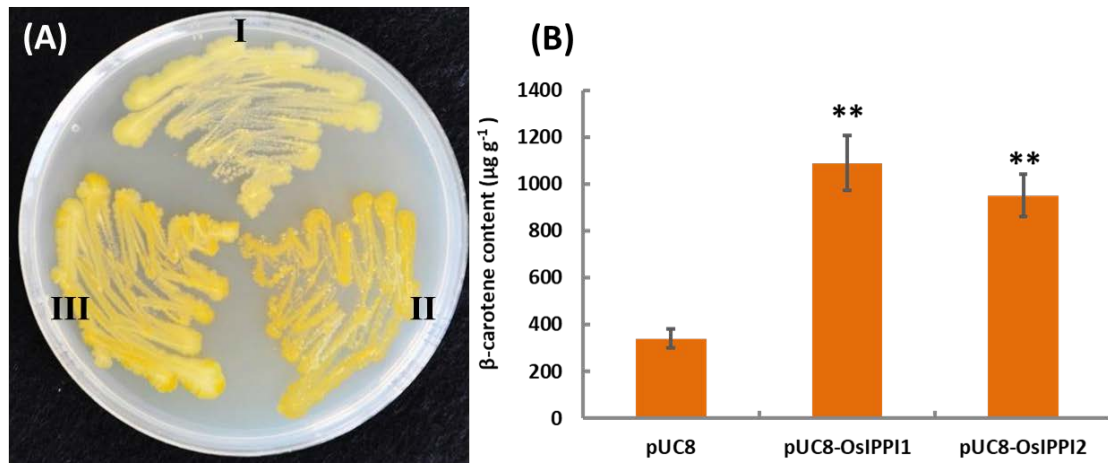


Figure 3.3. (A) Color complementation and enhanced β -carotene accumulation due to the expression of *OsIPPI1* or *OsIPPI2* in *E. coli* engineered for β -carotene biosynthesis. The plate was divided into three sections, which were inoculated separately with bacteria carrying pACCAR16 Δ crtX and pUC8 (empty vector) (I), pACCAR16 Δ crtX and pUC8-OsIPPI1 (II), or pACCAR16 Δ crtX and pUC8-OsIPPI2 (III). The bacterial clones in B and C are darker yellow due to enhanced β -carotene accumulation; (B) Production of β -carotene production by *E. coli* strains expressing rice isopentenyl diphosphate isomerase genes. Lane pUC8 = *E. coli* strain carrying pACCAR16 Δ crtX and pUC8. Lane pUC8-OsIPPI1 = *E. coli* strain carrying pACCAR16 Δ crtX and pUC8-OsIPPI1. Lane pUC8-OsIPPI2 = *E. coli* strain carrying pACCAR16 Δ crtX and pUC8-OsIPPI2. Values represent the mean of five biological replicates and error bars represent standard deviations. Double asterisks indicate highly significant differences compared with the control ($p < 0.01$, Student's *t*-test).

3.4.3. *OsIPPI1* mRNA is more abundant than *OsIPPI2* mRNA in all rice tissues

Having established that both *OsIPPI1* and *OsIPPI2* encode functional enzymes, I next determined their expression levels in different rice tissues by quantitative RT-PCR (qRT-PCR). *OsIPPI1* mRNA was much more abundant than *OsIPPI2* mRNA in all tissues (Fig. 3.4). Both *OsIPPI1* and *OsIPPI2* mRNA was most abundant in the leaf 7 days after germination (and was more than 10-fold more abundant than *OsIPPI2* mRNA in this tissue). The level of *OsIPPI1* mRNA declined sharply between 7 and 14 days after germination in the root, stem and leaves whereas the level of *OsIPPI2* mRNA declined marginally in the root and stem and increased slightly in the leaves over the same duration. Both *OsIPPI1* and *OsIPPI2* were expressed at the lowest levels in the endosperm and embryo (Fig. 3.4).

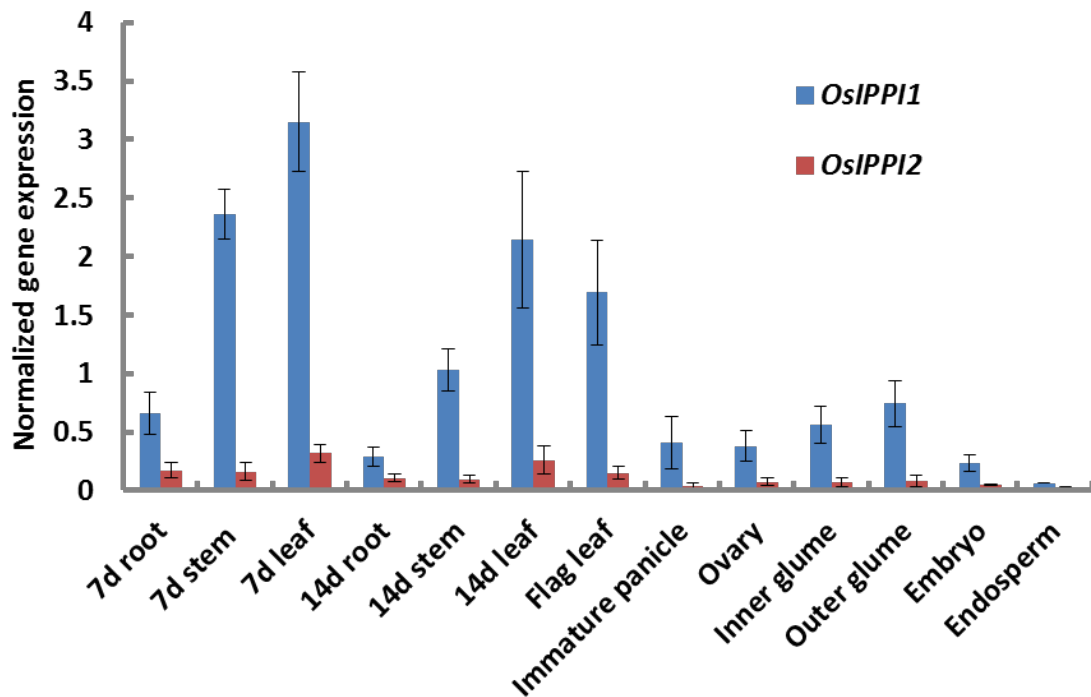


Figure 3.4. Analysis of *OsIPPI1* and *OsIPPI2* gene expression in different rice tissues at different growth stages. The expression of *OsIPPI1* and *OsIPPI2* was measured by qRT-PCR in different rice tissues and in the stem and leaves at two different stages. Expression levels were normalized against the rice *actin* gene. Each value represents the means of three different biological replicates and error bars represent standard deviations.

3.4.4 *OsIPPI1/2* and *OsHDR1/2* show distinct and dynamic expression profiles in 14-day etiolated leaves during de-etiolation with corresponding metabolic changes

The de-etiolation of etiolated leaves involves a burst of carotenoid, chlorophyll and phytosterol biosynthesis which requires the activity of IPPI and/or HDR for IPP and DMAPP synthesis. I therefore investigated the expression levels of *OsIPPI1/2* and *OsHDR1/2* by qRT-PCR during the de-etiolation of 14-day old etiolated rice leaves at 10 time points after the onset of illumination (Fig. 3.5A). As a baseline, I found that *OsIPPI1* mRNA was 6.4-fold more abundant than *OsIPPI2* mRNA in the dark control (time zero in Fig. 3.5A). Following the onset of illumination, both genes were moderately induced after 1h. Thereafter, the level of *OsIPPI2* mRNA remained stable until 12h after the onset of illumination, whereas the level of *OsIPPI1* mRNA peaked at approximately 3-fold higher than the baseline after 4h before declining thereafter until 12h after the onset of illumination. Between 12 and 24h, the expression of *OsIPPI1* increased again to more than double the baseline whereas the expression of *OsIPPI2* showed a 3-fold decline over the same period. After 24h of continuous illumination, the expression of *OsIPPI1* was more than 50-fold higher than that of *OsIPPI2* (Fig. 3.5A). The last enzyme in the plastidial MEP pathway, 3-methylbut-2-enyl diphosphate (HMBPP) reductase (HDR) is able to convert HMBPP into a mixture of IPP and DMAPP (Vranova *et al.*, 2013), I therefore carried out analyses of *OsHDR1* (Os03g52170, Jung *et al.*, 2008) and *OsHDR2* (Os03g52180, Jung *et al.*, 2008) expression during the de-etiolation of 14-day old etiolated rice leaves at the same 10 time points after the onset of illumination (Fig. 3.5A). Likewise, *OsHDR1* expression level was induced after 1h illumination. With continuous illumination, drastic rise in mRNA level of *OsHDR1* showed after 4h of illumination, which was 3.4-fold higher than the baseline. Then *OsHDR1* mRNA declined till 12h, and reached the same level as at 2h illumination. Between 12 and 24h, *OsHDR1* expression increased rapidly again to 5.7-fold higher than baseline. Nevertheless, *OsHDR2* was barely detected in rice leaves during de-etiolation.

Corresponding metabolic changes during de-etiolation were analyzed by UPLC and GC-MS at time point zero as well as after illumination 6, 12 and 24h of continuous illumination. Chlorophyll *a* was only detected in leaves after exposure to white light, but an extremely low amount of chlorophyll *b* was detected in the leaves of the dark

control ($0.75 \pm 0.10 \mu\text{g g}^{-1}$ DW). After 6–24h of continuous illumination, the levels of both chlorophyll *a* and chlorophyll *b* increased rapidly, reaching 326.67 ± 91.81 and $112.37 \pm 22.06 \mu\text{g g}^{-1}$ DW, respectively (Fig. 3.5B). The quantity of carotenoids also increased in response to illumination, with approximately 2-fold increases in abundance observed for β -carotene, lutein and violaxanthin after 24h of illumination compared to the dark control (Fig. 3.5C). The levels of phytosterols (β -sitosterol, campesterol and stigmasterol) in leaves increased by approximately 1.5-fold during the first 6h of illumination, remained steady or decline slightly until 12h and then declined to levels similar to the dark control after 24h (Fig. 3.5D).

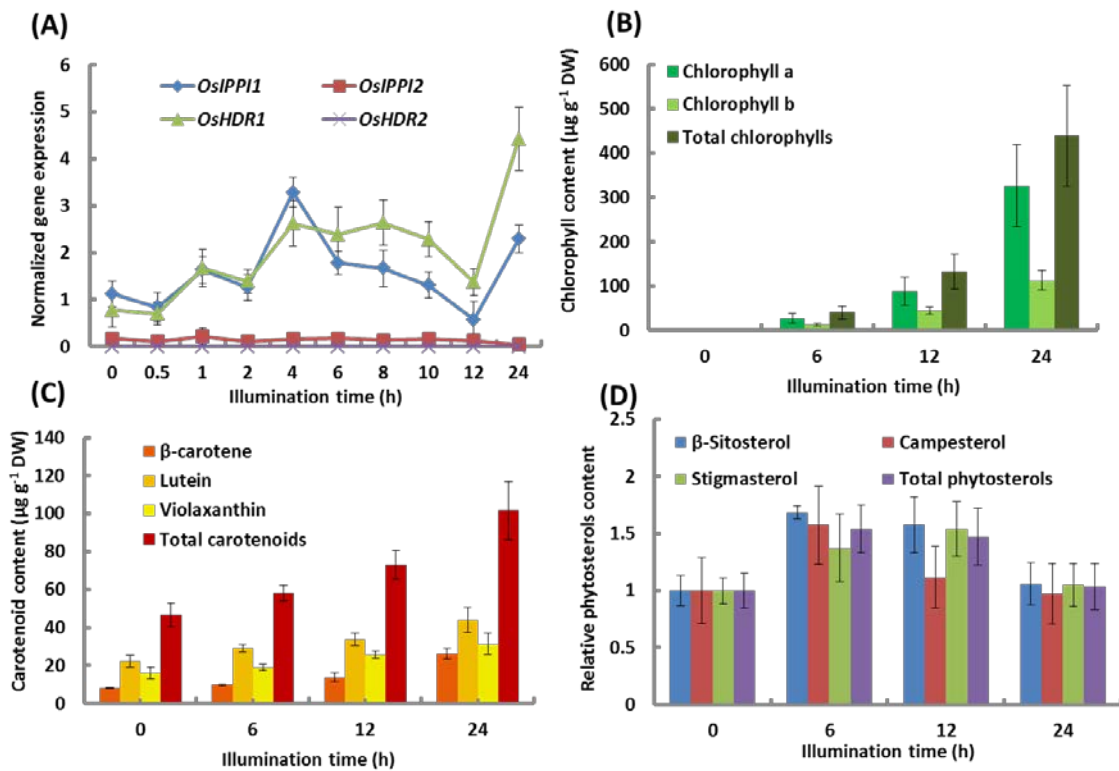


Figure 3.5. Analysis of *OsIPPI1/2* and *OsHDR1/2* expression and phytosterol, chlorophyll and carotenoid levels in the etiolated leaves of 14-day-old rice plants during de-etiolation. (A) Analysis of *OsIPPI1/2* and *OsHDR1/2* expression in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light. Expression levels were normalized to the rice *actin* gene, as determined by qRT-PCR. Each value represents the average of four experiments and the error bar indicates the standard deviation. (B) Chlorophyll composition and content in the same leaves during de-etiolation. (C) Carotenoid composition and content in the same leaves during de-etiolation. (D) Phytosterol composition and content in the same leaves during de-etiolation. Values represent the mean of five biological replicates and error bars represent standard deviations.

3.4.5. Targeting of IPPI fusion proteins in transgenic rice leaves

To investigate the subcellular localization of OsIPPI1 and OsIPPI2, I constructed chimeric genes in which the IPPI coding regions were fused to the N-terminus of a synthetic green fluorescent protein (sGFP), to facilitate tracking by confocal microscopy (Fig. 3.6A). Both fusion constructs were driven by the constitutive Cauliflower mosaic virus (CaMV) 35S promoter. I recovered 15 independent transgenic lines expressing *OsIPPI1-sGFP* and 18 expressing *OsIPPI2-sGFP*. I detected green fluorescence in the leaves of 11 *OsIPPI1-sGFP* lines and 18 *OsIPPI2-sGFP* lines. Three independent lines representing each event were then self-pollinated over two generations to yield T3 homozygous lines for extensive analysis by confocal fluorescence microscopy and immuno-electron microscopy: the IPPI1 lines OsIPPI1-sGFP-1, OsIPPI1-sGFP-10 and OsIPPI1-sGFP-12, and the IPPI2 lines OsIPPI2-sGFP-4, OsIPPI2-sGFP-6 and OsIPPI2-sGFP-16.

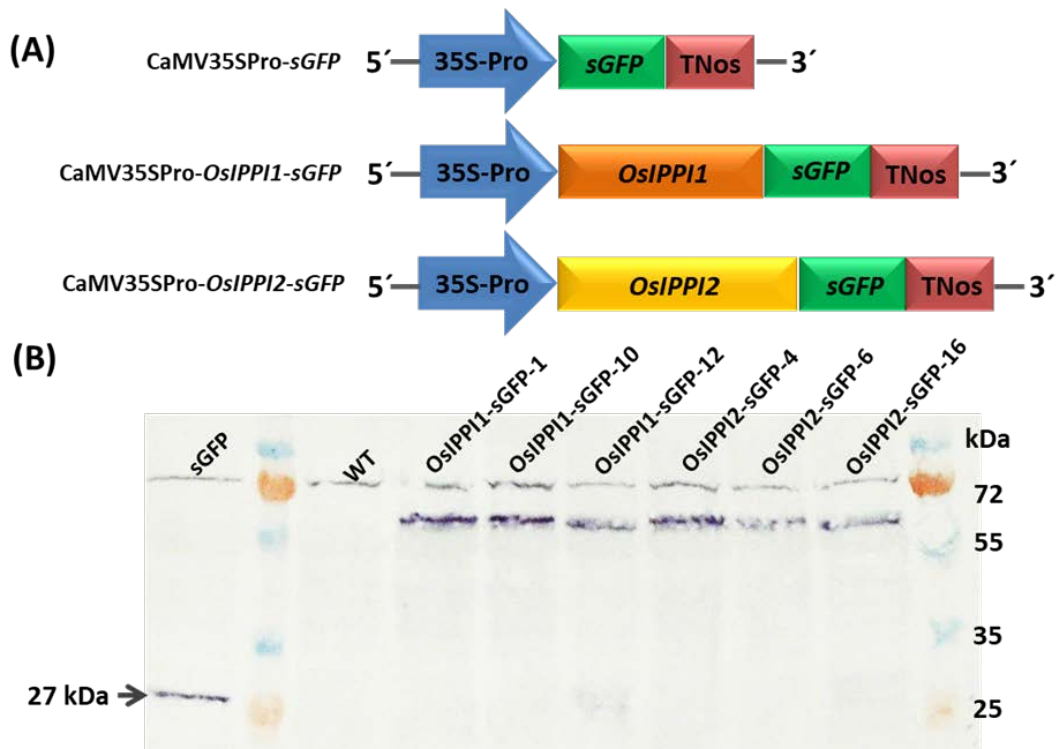


Figure 3.6. Fusion gene construction and western blot analysis in T3 homozygous transgenic rice leaves. (A) Schematic representation of the CaMV35SPro-*sGFP* construct, CaMV35SPro-*OsIPPI1-sGFP* and CaMV35SPro-*OsIPPI2-sGFP* fusion protein constructs. (B) Western blot analysis of T3 homozygous transgenic rice leaf extracts using an anti-GFP antibody. Lane sGFP = line expressing sGFP (27 kDa) under the control of the CaMV35S promoter (positive control). Lane WT = wild-type rice (negative control). Lanes OsIPPI1-sGFP-1/10/12 = three independent transgenic lines expressing the *OsIPPI1-sGFP* fusion gene under the control of the CaMV35S promoter. Lanes OsIPPI2-sGFP-4/6/16 = three independent transgenic lines expressing the *OsIPPI2-sGFP* fusion gene under the control of the CaMV35S promoter.

Analysis of T3 leaf extracts from these six lines by SDS-PAGE and western blot analysis using an anti-GFP antibody revealed protein bands with an apparent molecular mass of ~59 kDa (Fig. 3.6B). The anticipated molecular mass of the OsIPPI1-sGFP fusion protein was ~58 kDa and that of the OsIPPI2-sGFP fusion protein was ~72 kDa due to the presence of the N-terminal plastid-targeting TPS, indicating that this signal peptide was cleaved off in the transgenic plants.

The OsIPPI1 and OsIPPI2 were analyzed using various online resources to predict protein localization (Table 3.2) revealing that both isoforms contained transit peptide signals for the endoplasmic reticulum (ER), mitochondria and peroxisomes, but only OsIPPI2 featured the TPS required for import into plastids (Fig. 3.1).

Table 3.2. Prediction of OsIPPI1 and OsIPPI2 subcellular localization.

Program	OsIPPI1 Prediction	OsIPPI2 Prediction
iPSORT	No	Chloroplast Mitochondrion
Predotar	Elsewhere	Mitochondrion plastid
TargetP 1.1 Server	Other 0.923 Mitochondrion 0.103	Chloroplast 0.830 Mitochondrion 0.619
ChloroP 1.1 Server	No	Chloroplast
MitoFates	No Mitochondrial presequence	Mitochondrion
RSLpred	Mitochondrial protein	Mitochondrial protein
PTS1 Predictor	No targeted	Targeted
LocTree3	peroxisome	peroxisome

Confocal microscopy revealed the absence of green fluorescence in the leaves of wild-type plants (Fig. 3.7) but sGFP green fluorescent signals representing OsIPPI1-sGFP and OsIPPI2-sGFP, respectively, in the corresponding transgenic plants (Fig. 3.7). The signals representing OsIPPI1-sGFP and OsIPPI2-sGFP labeled structures of different sizes, and some of the OsIPPI2-sGFP signals were present as clusters, indicating distinct but overlapping patterns of subcellular localization for each isoform.

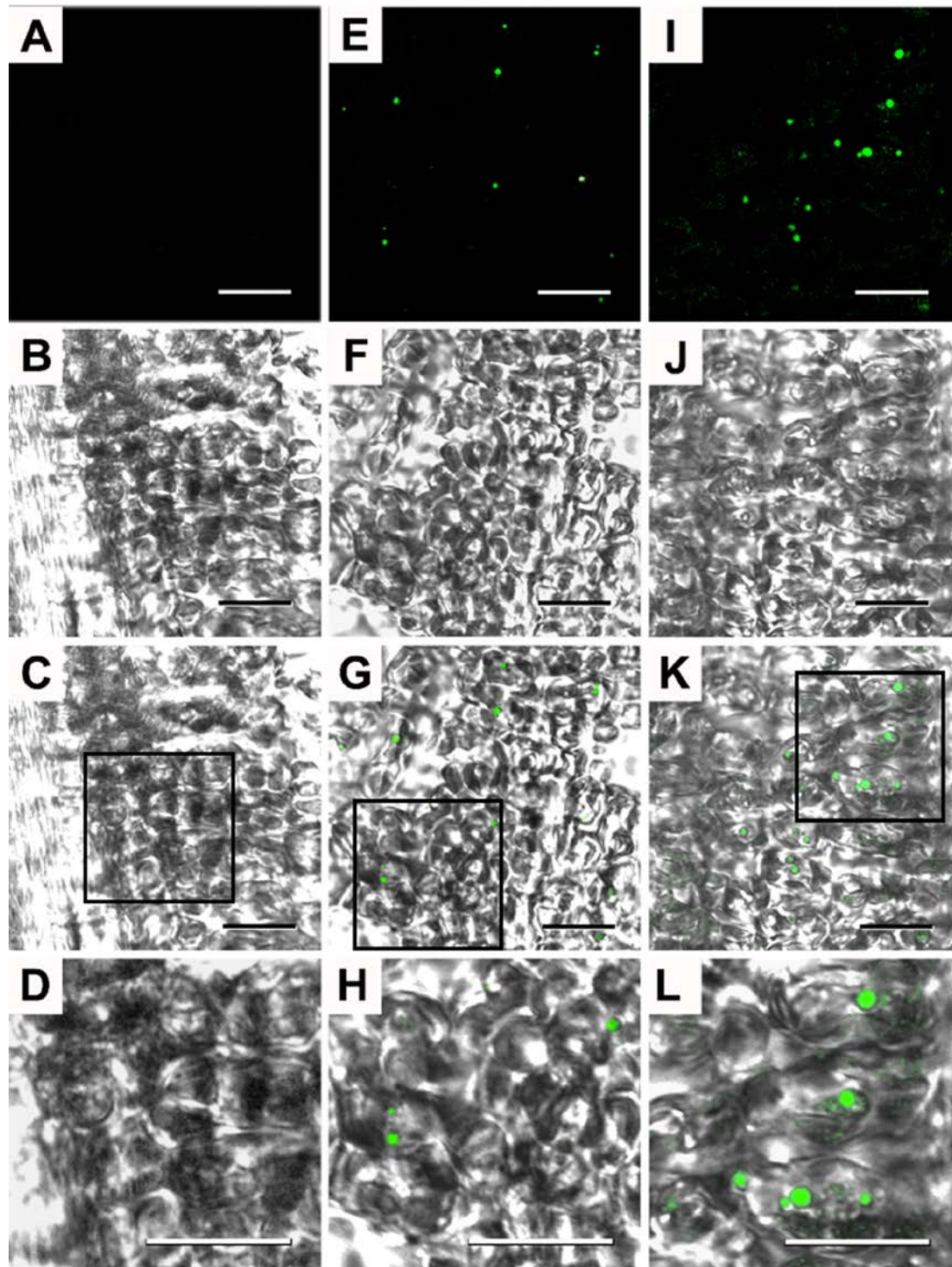


Figure 3.7. Confocal laser scanning microscopy of wild-type rice leaves (A-D) and transgenic plants lines expressing *OsIPPI1-sGFP* (E-H) and *OsIPPI2-sGFP* (I-L) fusion gene driven by the CaMV35S promoter. A, E and I: green fluorescence of sGFP. B, F and J: Bright field. C, G and K: Merged. D, H and L: Enlarged boxed areas in C, G and K. Scale bars = 20 μ m.

In order to confirm the confocal results and further explore the subcellular location of sGFP signals, I carried out detailed analysis by immuno-electron microscopy. This showed that both fusion proteins were mainly localized in the mitochondria (Fig. 3.8) and peroxisomes (Fig. 3.9), but the signal was also detected in the ER (Fig. 3.10). In the lines expressing OsIPPI2-sGFP, the fusion protein was also detected in the plastids (Fig. 3.11). The gold particles were not detected in other cellular structures, or in the leaf cells of wild-type plants. Taken together, these results indicate that OsIPPI1 and OsIPPI2 are both localized in the mitochondria, peroxisomes and ER of rice leaves, but only OsIPPI2 is imported into the chloroplasts.

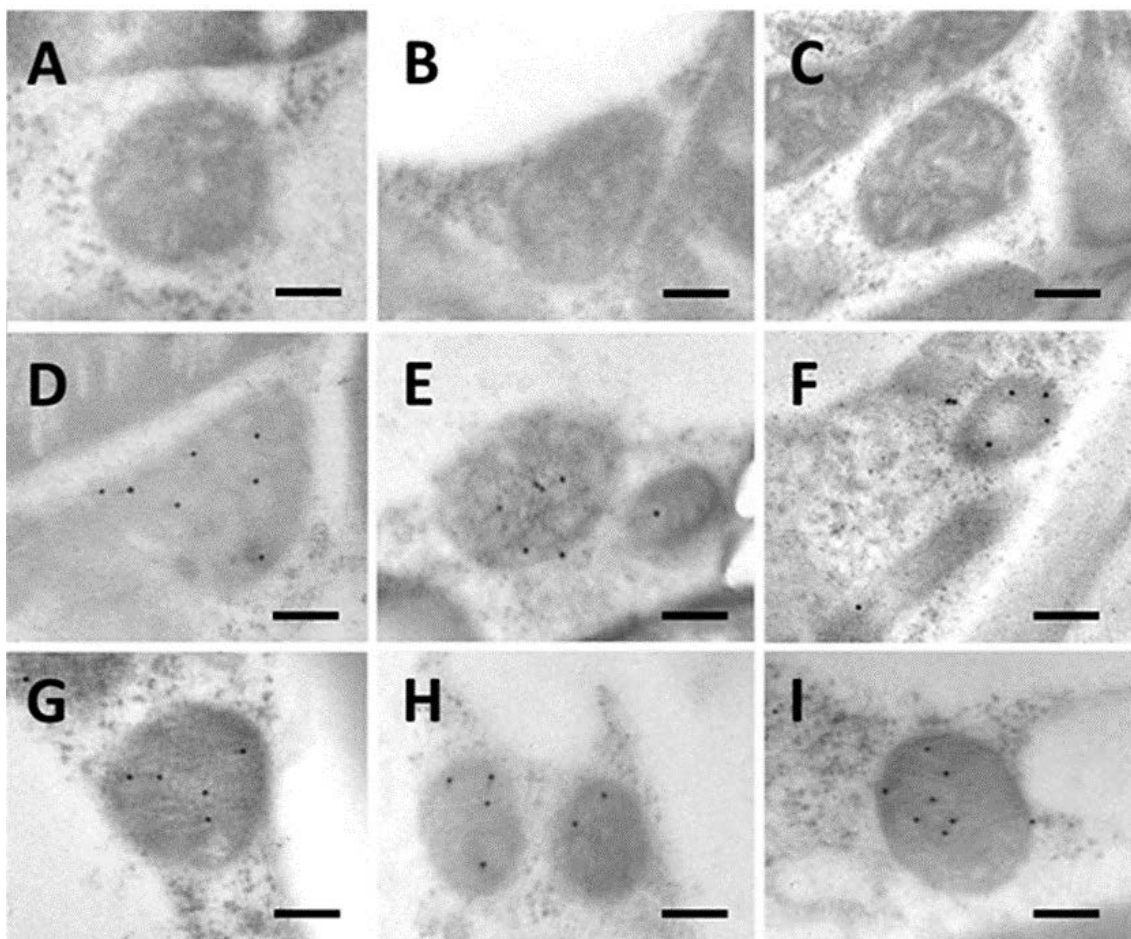


Figure 3.8. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice mitochondria by labeling sGFP. (A-C) Immunogold labeling of mitochondria in wild-type rice. (D-F) Immunogold labeling of mitochondria in *OsIPPI1-sGFP* transgenic rice. (G-I) Immunogold labeling of mitochondria in *OsIPPI2-sGFP* transgenic rice. Scale bars = 200nm; gold particle size = 15nm.

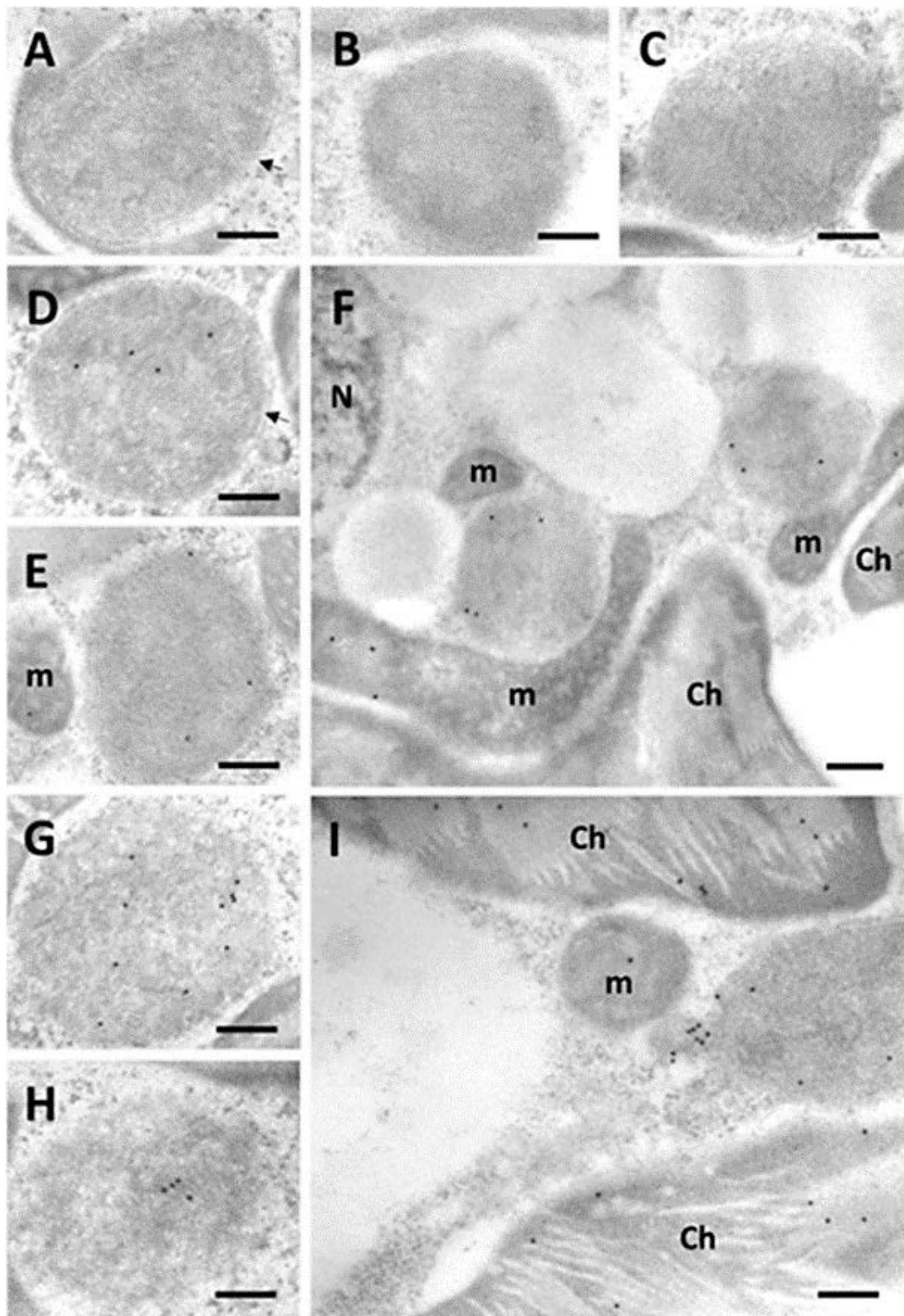


Figure 3.9. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice peroxisomes by labeling sGFP. (A-C) Immunogold labeling of peroxisome in wild-type rice. (D-F) Immunogold labeling of peroxisome in OsIPPI1-sGFP transgenic rice. (G-I) Immunogold labeling of peroxisome in OsIPPI2-sGFP transgenic rice. Ch = chloroplast; m = mitochondria, N = nucleus; arrow = peroxisome membrane. Bars = 200nm; gold particle size = 15nm.

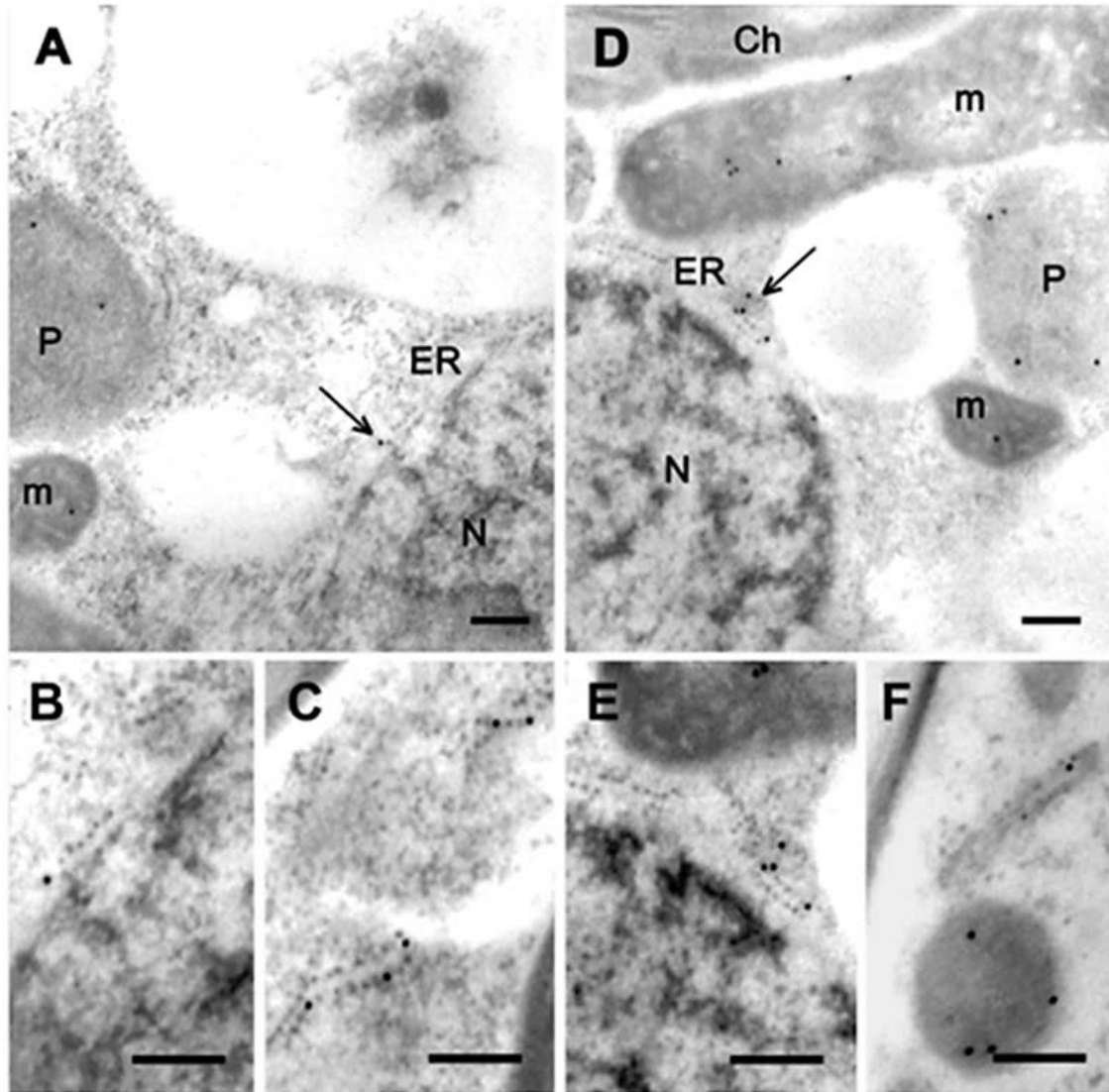


Figure 3.10. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice endoplasmic reticulum by labeling sGFP. (A-C) Immunogold labeling of endoplasmic reticulum in *OsIPPI1-sGFP* transgenic rice. (D-F) Immunogold labeling of endoplasmic reticulum in *OsIPPI2-sGFP* transgenic rice. Ch = chloroplast; m = mitochondria, N = nucleus; ER = endoplasmic reticulum; arrow = ER structure. Bars = 200nm; gold particle size = 15nm.

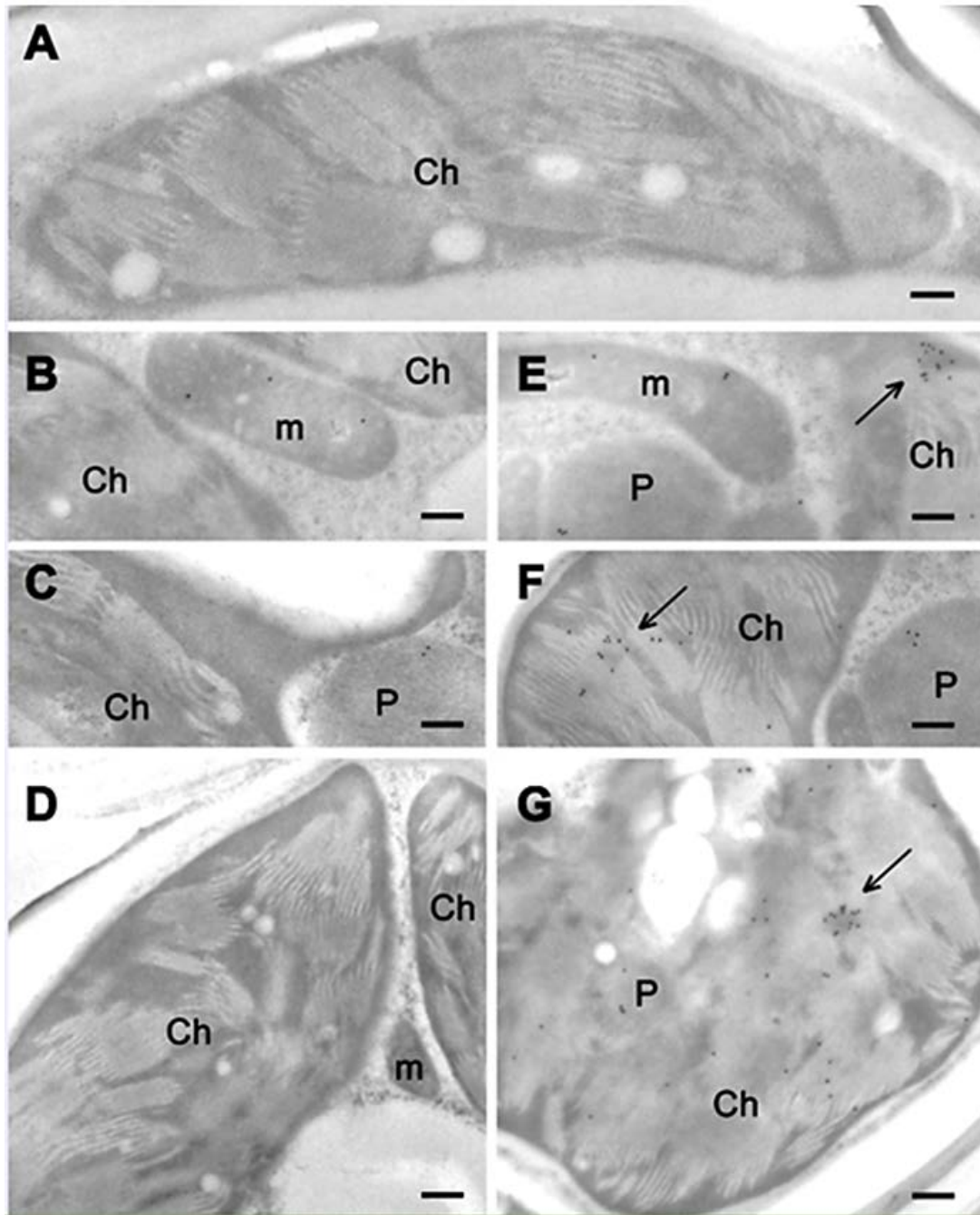


Figure 3.11. Immuno-electron microscopy showing localization of OsIPPI1-sGFP and OsIPPI2-sGFP in rice chloroplasts by labeling sGFP. (A) Immunogold labeling of chloroplasts in wild-type rice. (B-D) Immunogold labeling of chloroplasts in OsIPPI1-sGFP transgenic rice. (E-G) Immunogold labeling of chloroplasts in OsIPPI2-sGFP transgenic rice. Ch = chloroplast, m = mitochondria; P = peroxisome; arrow = gold cluster in chloroplast. Bars = 200nm; gold particle size = 15nm.

3.5. Discussion

There are two genes encoding the enzyme IPPI in rice (Jung *et al.*, 2008) but only OsIPPI1 has been functionally verified by genetic complementation (Cunningham and Gantt, 2001). The role of OsIPPI2 has not been investigated thus far and its precise function in the MVA and/or MEP pathways is unclear. Many plant genomes contain two IPPI paralogs with different expression profiles encoding proteins with different subcellular locations, as previously reported in tobacco, Arabidopsis and tomato (Nakamura *et al.*, 2001; Phillips *et al.*, 2008; Pankratov *et al.*, 2016). In tobacco, the *NtIPPI1* mRNA encodes the plastid-localized isoform and is approximately 3-fold more abundant in leaves than the *NtIPPI2* mRNA encoding the cytosolic isoform, at least under normal growth conditions (Nakamura *et al.*, 2001). The opposite profile is observed in Arabidopsis, where the *AtIPPI2* mRNA encoding the cytosolic isoform is 10-fold more abundant in leaves, 22-fold in roots, 4.9-fold in flowers, and 4.6-fold in stems compared to the *AtIPPI1* mRNA encoding the plastid-localized enzyme (Phillips *et al.*, 2008). Both mRNAs are expressed in the same range of tissues but at varying levels, e.g. *AtIPPI1* mRNA is more than 4-fold more abundant in flowers than leaves/stems and more than 1.5-fold more abundant in the roots than leaves/stems, whereas *AtIPPI2* mRNA is found at similar levels in flowers, stems and leaves but is 2.5-fold more abundant in roots (Phillips *et al.*, 2008). The characterization of *fruit carotenoid-deficient (fcd)* mutants in tomato revealed that the plastid-localized isoform SIIPPI1 is required for fruit carotenoid biosynthesis (Pankratov *et al.*, 2016). The *SIIPPI2* mRNA encoding the cytoplasmic isoform is significantly more abundant than *SIIPPI1* mRNA in leaves and green fruit tissue, but it declines during fruit development whereas *SIIPPI1* mRNA remains at constant levels until the red ripe stage, confirming that SIIPPI1 has a more important role in the ripening fruit than SIIPPI2 (Pankratov *et al.*, 2016). These studies in tobacco, Arabidopsis and tomato suggest that paralogous IPPI genes are maintained due to functional specialization and the same is likely to be true for the two paralogs in rice, but until now the specific roles of *OsIPPI1* and *OsIPPI2* have not been investigated. Similarly, with the mentioned above, I found that the two paralogs showed distinct expression profiles. Although both are nearly ubiquitously expressed, qRT-PCR analysis revealed that *OsIPPI1* is expressed at a much higher level than *OsIPPI2* in all rice tissues (Fig. 3.4). There is a significant difference in the representation of full-length *OsIPPI1* and *OsIPPI2* cDNAs in dbEST,

the database for expressed sequence tags, with 71 and 25 entries for *OsIPPI1* and *OsIPPI2*, respectively.

IPPI is the only enzyme common to both the MVA and MEP pathways and it is essential for the maintenance of appropriate levels of IPP and DMAPP in different subcellular compartments (Okada *et al.*, 2008; Pankratov *et al.*, 2016). In the plastids, DMAPP can be synthesized via the isomerization of IPP (by IPPI) or via the conversion of HMBPP into a mixture of IPP and DMAPP (by HDR, in the last step of the MEP pathway), which means IPPI is dispensable (Vranova *et al.*, 2013). In contrast, cytosol lacks the MEP pathway and IPPI is absolutely required for the synthesis of DMAPP, hence the mevalonate pathway is blocked without it. Localization experiments revealed that *OsIPPI1* and *OsIPPI2* are found in the mitochondria, peroxisome and the ER, whereas only *OsIPPI2* is imported into the plastids. Given the stronger expression of *OsIPPI1* compared to *OsIPPI2*, this indicates that *OsIPPI1* is predominantly responsible for the synthesis of MVA pathway-derived isoprenoids such as phytosterols whereas *OsIPPI2* is responsible for the synthesis of MEP pathway-derived isoprenoids such as chlorophylls and carotenoids. I therefore investigated the correlation between *OsIPPI1/2* gene expression during the illumination of etiolated leaves (Fig.3.5A) and the accumulation of chlorophylls (Fig. 3.5B), carotenoids (Fig. 3.5C) and phytosterols (Fig. 3.5D). The phytosterols (β -sitosterol, campesterol and stigmasterol) accumulated during the first 6h of illumination and then declined, mirroring the expression profile of *OsIPPI1* mRNA, with a slight delay (Fig. 3.5A, B), as might be expected, given that the peak in mRNA levels would take some time to translate into a peak in enzyme levels, followed by the peak in the conversion of IPP/DMAPP into downstream products. In contrast, the levels of chlorophylls and carotenoids continued to increase during the 24h illumination even though the levels of *OsIPPI2* mRNA remained at steady low levels and then declined in the last 12h (Fig. 3.5A, C, D). These results suggest that *OsIPPI2* is not required for the increase in flux observed in the MEP pathway, in agreement with reports that the enzyme is dispensable given the existence of a second route for the synthesis of IPP/DMAPP using HDR (Rodriguez-Concepcion, 2010). Indeed, I found that the accumulation of chlorophylls and carotenoids closely mirrored the expression of *OsHDR1* (Fig. 3.5A) which suggests that the flux towards these MEP-derived products is almost wholly determined by the availability of HDR activity. In 2004, Jonathan *et al* using VIGS (Virus induced gene silencing) silenced the *HDR* or *IPPI* gene successfully

in *N. benthamiana* and caused the dramatic decline of total chlorophyll and carotenoid levels.

The complex metabolic networks in plants are highly compartmentalized because the corresponding enzymes are targeted to different intracellular sites, where different steps in the pathway are catalyzed (Heinig *et al.*, 2013). In many cases, different isoforms of the same enzyme are targeted to different compartments, as observed for the two IPPI isoforms in Arabidopsis (Sapir-Mir *et al.*, 2008), a single IPPI in *C. roseus* (Guirimand *et al.*, 2012) and now rice (this study). In both Arabidopsis and *C. roseus*, long and short IPPI transcripts from each gene are synthesized giving rise to a longer polypeptide with an N-terminal signal for import into the mitochondria and plastids, as well as a shorter polypeptide which is imported into the peroxisomes (Sapir-Mir *et al.*, 2008; Guirimand *et al.*, 2012). To investigate the localization of the two IPPI isoforms in rice, I expressed fusion proteins combining each isoform with sGFP under the control of the constitutive CaMV35S promoter. The analysis of leaf extracts by SDS-PAGE and western blot revealed protein bands of ~59 kDa, corresponding to the anticipated size of the fusion protein following the cleavage of the N-terminal TPS from OsIPPI2 (Fig.18B). It seems that rice IPPI genes are transcribed solely into long mRNAs, one including a TPS and one not, allowing the specific targeting of OsIPPI2 to the plastids. In contrast, the Arabidopsis and *C. roseus* genes produce alternative transcripts containing different initiation codons, resulting in variants with and without a TPS for each IPPI gene (Sapir-Mir *et al.*, 2008; Guirimand *et al.*, 2012). Different plant species can therefore use different genetic mechanisms to produce alternatively-targeted isoforms of IPPI.

The synthesis of DMAPP in Arabidopsis and *C. roseus* takes place across various cellular locations, including the chloroplasts, peroxisomes and mitochondria (Sapir-Mir *et al.*, 2008; Guirimand *et al.*, 2012; Heinig *et al.*, 2013), with some isoprenoids (such as phytosterols) primarily synthesized in the ER (Clastre *et al.*, 2011; Pulido *et al.*, 2012; Vranova *et al.*, 2012, 2013). Using confocal fluorescence microscopy and immuno-electron microscopy, I confirmed that both OsIPPI1 and OsIPPI2 are also localized in the ER in rice (Fig. 3.10). It is therefore possible that phytosterols are also synthesized mainly in the ER in rice, particularly given the evidence based on feeding experiments that show phytosterols are derived solely from IPP/DMAPP generated via the MVA pathway in Arabidopsis (Phillips *et al.*, 2008). Other MVA pathway enzymes

are also localized in the ER, including acetyl-CoA acetyltransferase (AACT), which catalyzes the first step in the pathway, and is mainly localized to the ER and its associated vacuoles in the sun spurge *Euphorbia helioscopia* (Wang *et al.*, 2017). The MVA pathway starts with the sequential condensation of three molecules of acetyl-CoA to generate 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA), which is in turn converted to MVA by HMG-CoA reductase (HMGR), the rate-limiting step in the MVA pathway (Vranova *et al.*, 2013). Arabidopsis HMGR was found to be localized to the ER as well as unidentified spherical structures (Campos and Boronat, 1995; Leivar *et al.*, 2005). Enzymes responsible for downstream reactions, such as FPP synthase and squalene synthase have also been localized to the ER in plants and humans (Martin *et al.*, 2007; Busquets *et al.*, 2008; Stamellos *et al.*, 1993). My finding that OsIPPI1 and OsIPPI2 are also localized in the ER indicated that this compartment is sufficient for the autonomous synthesis of DMAPP from IPP.

3.6. Conclusions

The present study represents a comprehensive characterization of the IPPI gene family in rice; revealing for the first time that *OsIPPI2* encodes a functional enzyme which, along with the previously characterized *OsIPPI1*, is expressed in all rice organs and tissues. The two isoforms show distinct expression profiles; with *OsIPPI1* being expressed at much higher levels than *OsIPPI2*. The expression profile of *OsIPPI1* in de-etiolated leaves, then illuminated for 24h, mirrored the accumulation of phytosterols, suggesting a key role in the synthesis of these MVA pathway products. In contrast, *OsIPPI2* expression showed no correlation with the accumulation of chlorophylls and carotenoids, and further investigation revealed that these MEP pathway products were instead related to the expression of *OsHDR1*. OsIPPI1 was localized in the mitochondria, peroxisomes and the ER, suggesting it is primarily responsible for the synthesis of MVA pathway-derived isoprenoids. In contrast, OsIPPI2 was localized not only in these compartments but also in the plastids (reflecting the presence of an N-terminal targeting signal) but was dispensable for the synthesis of MEP pathway-derived isoprenoids. The detection of both isoforms in the ER indicates that the autonomous synthesis of DAMPP from IPP occurs in the ER, providing insights into the role of this compartment in the synthesis of DMAPP, which is required for all MVA pathway-derived isoprenoids.

3.7. References

- Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics*, **18**:298–305.
- Busquets A, Keim V, Closa M, del Arco A, Boronat A, Arro M, Ferrer A (2008). *Arabidopsis thaliana* contains a single gene encoding squalene synthase. *Plant Mol. Biol.*, **67**:25–36.
- Campos N, Boronat A (1995). Targeting and topology in the membrane of plant 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Cell*, **7**:2163–2174.
- Capell T, Christou P (2004). Progress in plant metabolic engineering. *Curr. Opin. Biotechnol.*, **15**:148–154.
- Clastre M, Papon N, Courdavault V, Giglioli-Guivarc'h N, St-Pierre B, Simkin AJ (2011). Subcellular evidence for the involvement of peroxisomes in plant isoprenoid biosynthesis. *Plant Signal Behav.*, **6**:2044–2046.
- Cunningham FX Jr, Gantt E (2000). Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in *Escherichia coli*. *Plant Cell Physiol.*, **41**:119–123.
- Davies BH (1976). Carotenoids. *Chemistry and Biochemistry of Plant Pigments*. Academic, London, pp. 38–366.
- Emanuelsson O, Nielsen H, von Heijne G (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, **8**:978–984.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.*, **2**:953–971.
- Enfissi EM, Barneche F, Ahmed I, Lichtlé C, Gerrish C, McQuinn RP, Giovannoni JJ, Lopez-Juez E, Bowler C, Bramley PM, Fraser PD (2010). Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell*, **22**:1190-1215.

Errampalli D, Fletcher J. 1993. Production of monospecific polyclonal antibodies against aster yellows mycoplasma-like organism-associated antigen. *Phytopathol.*, **83**:1279–1282.

Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell Proteomics*, **14**:1113–1126.

Green TR, Dennis DT, West CA (1975). Compartmentation of isopentenyl pyrophosphate isomerase and prenyl transferase in developing castor bean endosperm. *Biochem. Biophys. Res. Commun.*, **64**:976–982.

Guirimand G, Guihur A, Phillips MA, Oudin A, Glevarec G, Melin C, Papon N, Clastre M, St-Pierre B, Rodriguez-Concepcion M, Burlat V, Courdavault V (2012). A single gene encodes isopentenyl diphosphate isomerase isoforms targeted to plastids, mitochondria and peroxisomes in *Catharanthus roseus*. *Plant Mol. Biol.*, **79**:443–459.

Heinig U, Gutensohn M, Dudareva N, Aharoni A (2013). The challenges of cellular compartmentalization in plant metabolic engineering. *Curr. Opin. Biotechnol.*, **24**:239–246.

Jin X, Bai C, Bassie L, Nogareda C, Romagosa I, Twyman RM, Christou P, Zhu C (2018). ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (*Zea mays*) β -carotene hydroxylase 2 gene. *New Phytol.*, **222**(2):793–804.

Jung KH, Lee J, Dardick C, Seo YS, Cao P, Canlas P, Phetsom J, Xu X, Ouyang S, An K, Cho YJ, Lee GC, Lee Y, An G, Ronald PC (2008). Identification and functional analysis of light-responsive unique genes and gene family members in rice. *PLoS Genet.*, **4**:e1000164.

Kajiwara S, Fraser PD, Kondo K, Misawa N (1997). Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochem. J.*, **324**:421–426.

Kaundal R, Raghava GPS (2009). RSLpred: an integrative system for predicting subcellular localization of rice proteins combining compositional and evolutionary information. *Proteomics*, **9**:2324–2342.

Leivar P, González VM, Castel S, Trelease RN, López-Iglesias C, Arró M, Boronat A, Campos N, Ferrer A, Fernández-Busquets X (2005). Subcellular localization of Arabidopsis 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Plant Physiol.*, **137**:57–69.

Martin D, Piulachs MD, Cunillera N, Ferrer A, Belles X (2007). Mitochondrial targeting of farnesyl diphosphate synthase is a widespread phenomenon in eukaryotes. *Biochim. Biophys. Acta*, **1773**:419–426.

Misawa N, Satomi Y, Kondo K, Yokoyama A, Kajiwara S, Saito T, Ohtani T, Miki W (1995). Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol.*, **177**:6575–6584.

Nakamura A, Shimada H, Masuda T, Ohta H, Takamiya K (2001). Two distinct isopentenyl diphosphate isomerases in cytosol and plastid are differentially induced by environmental stresses in tobacco. *FEBS Lett.*, **506**:61–64.

Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003). Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. *J. Mol. Biol.*, **328**:567–579.

Nogueira M, Enfissi EM, Almeida J, Fraser PD (2018). Creating plant molecular factories for industrial and nutritional isoprenoid production. *Curr. Opin. Biotechnol.* **49**, 80–87.

Nogueira M, Mora L, Enfissi EM, Bramley PM, Fraser PD (2013). Subchromoplast sequestration of carotenoids affects regulatory mechanisms in tomato lines expressing different carotenoid gene combinations. *Plant Cell*, **25**:4560-4579.

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.

Pankratov I, McQuinn R, Schwartz J, Bar E, Fei Z, Lewinsohn E, Zamir D, Giovannoni JJ, 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.

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.

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

Rodriguez-Concepcion M (2010). Supply of precursors for carotenoid biosynthesis in plants. *Arch. Biochem. Biophys.*, **504**:118–122.

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.

Stamellos KD, Shackelford JE, Shechter I, Jiang G, Conrad D, Keller GA, Krisans SK (1993). Subcellular localization of squalene synthase in rat hepatic cells. *J. Biol. Chem.* **268**:12818–12824.

Street IP, Coffman HR, Baker JA, Poulter CD (1994). Identification of Cys139 and Glu207 as catalytically important groups in the active site of isopentenyl diphosphate:dimethylallyl diphosphate isomerase. *Biochem.*, **33**:4212–4217.

Thibodeaux CJ, Liu HW (2017). The type II isopentenyl diphosphate:dimethylallyl diphosphate isomerase (IDI-2): A model for acid/base chemistry in flavoenzyme catalysis. *Arch. Biochem. Biophys.*, **632**:47–58.

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

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

Wang M, Wang D, Zhang Q, Chai J, Peng Y, Cai X (2017). Identification and cytochemical immunolocalization of acetyl-CoA acetyltransferase involved in the

terpenoid mevalonate pathway in *Euphorbia helioscopia* laticifers. *Bot. Stud.*, **58**:62.
doi: 10.1186/s40529-017-0217-3.

Zhu C, Sanahuja G, Yuan D, Farre G, Arjo G, Berman J, Zorrilla-Lopez U, Banakar R, Bai C, Perez-Massot E, Bassie L, Capell T, Christou P (2013). Biofortification of plants with altered antioxidant content and composition: genetic engineering strategies. *Plant Biotechnol J.*, **11**:129–141.

Chapter IV

Integrated metabolite and gene expression profiling of de-etiolated rice seedlings upon exposure to light

4.0. Abstract

This study aimed to assess whether white light influences metabolite levels and the expression of related genes in rice (*Oryza sativa*) leaves. The differences in metabolite profiles between etiolated and de-etiolated rice leaves at distinct illumination times were measured by gas chromatography–mass spectrometry and ultra-high performance liquid chromatography.

We observed significant changes in primary and in secondary metabolites in rice seedlings during de-etiolation. The levels of several primary metabolites, including some amino acids, organic acids and major sugars, decreased in de-etiolated rice leaves compared with in etiolated rice leaves. There were also significant differences in some secondary metabolites between dark-grown and de-etiolated rice seedlings. The levels of chlorophylls, carotenoids and tocopherols were significantly increased, whereas the levels of phytosterols increased slightly after illumination. A gene expression analysis revealed that the expression levels of corresponding genes mirrored the levels of metabolite changes in de-etiolated rice leaves. These findings provide a comprehensive picture of the primary and secondary metabolite responses to de-etiolation and serve as a base for the further characterization of light-induced genes.

4.1. Introduction

Abiotic environmental stresses, such as water deficiency, variable light conditions and temperature changes, can occur at multiple stages of plant development and affect the growth and productivity of plants (Chen *et al.*, 2011; Nakashima *et al.*, 2012). Unlike animals, plants cannot avoid environmental stresses by easily moving. Plant survival depends on the ability to accurately sense and respond to the extracellular environment at the physiological, biochemical and molecular levels (Yang *et al.*, 2007; Nakashima *et al.*, 2012). Light is an important regulator of many developmental processes in higher plants. In addition to acting as an energy source, it is also involved in regulating different growth and developmental stages throughout plant life cycle, such as seed germination, seedling photomorphogenesis, phototropism, gravitropism, chloroplast movement, shade avoidance, circadian rhythms and flower induction. (Jiao *et al.*, 2007; Canlas *et al.*, 2008). Many plants have evolved sophisticated photosensory systems that enable them to respond appropriately (Yang *et al.*, 2007). The effects of light on plant growth and development are particularly evident in the morphology during the de-etiolation process (Yang *et al.*, 2007). Etiolated *Arabidopsis* seedlings exhibit a typical phenotype, such as a pronounced apical hook, elongated hypocotyls and undifferentiated chloroplast precursors (Chory *et al.*, 1996; Clouse, 2001). When etiolated seedlings are exposed to light, the cotyledons expanded, the apical hook opened and the hypocotyls elongated (Li *et al.*, 2011). A similar phenotype was also observed in etiolated monocotyledonous seedlings. In etiolated rice (*Oryza sativa*), elongated hypocotyls and stem and rolled and enclosed young leaves were observed. When etiolated rice seedlings are exposed to light, dramatic morphological changes have been observed, the leaves expanded and turned green gradually, and apical hooks straightened and opened (Yang *et al.*, 2007).

All these morphological changes during de-etiolation are accompanied by a series of developmental processes, including chloroplast development, pigment synthesis and accumulation, and thylakoid-related photosystem assembly (Sánchez-Retuerta *et al.*, 2018; Yang *et al.*, 2007). Phytochromes, cytochromes and phytohormones play important roles in the regulation of the de-etiolation process, individually or cooperatively (Alabadi *et al.*, 2004; Lin, 2002; Li *et al.*, Terzaghi *et al.*, 1995). Because chlorophyll biosynthesis is light-dependent, etiolated seedlings contain non-

photosynthetic etioplasts lacking chlorophyll, while the chlorophyll precursor protochlorophyllide IX complexed with NADPH-dependent protochlorophyllide IX oxidoreductase (POR) accumulates (Papenbrock *et al.*, 2001). When etiolated seedlings are exposed to light, protochlorophyllide IX is rapidly converted to chlorophyllide and then to chlorophylls, resulting in the further formation of thylakoid membranes in the developing chloroplasts (Fankhauser *et al.*, 1997; Ghassemian *et al.*, 2006). Functional photosystems also require a variety of pigments, chlorophyll *a/b* and carotenoids that can be found in photosynthetic reaction centers in varying ratios (Ghassemian *et al.*, 2006). Chlorophyll *a/b* and carotenoids play important roles in the stabilization and assembly of newly synthesized photosynthetic complexes (Ghassemian *et al.*, 2006; Hooper *et al.*, 2001).

The study of Ghassemian *et al.* (2006) demonstrated that the amounts of chlorophyll *a/b*, β -carotene, lutein and α -tocopherol significantly increased during the de-etiolation process in response to red light in *Arabidopsis* seedlings. In other studies, the metabolic flux of the 2-C-methyl-derythritol 4-phosphate (MEP) pathway was determined in spinach leaves, and the flux increased with light intensity. The light may interfere with the cross-talk between the mevalonic acid (MVA) and MEP pathways (Mongelard *et al.*, 2011; Hemmerlin *et al.*, 2003). A comparison of metabolite profiles in wild-type (WT) seedlings grown under dark conditions and exposed to far-red light (FR)/white light (WL) for 6 and 24 h, respectively, was established in *Arabidopsis* by Jumtee *et al.* (2008). A number of metabolites, including some amino acids, organic acids and major sugars, were found in smaller amounts in seedlings grown under FR/WL compared with in the dark control.

All of these changes depend on the differential expression of a large number of genes, many of which are controlled by light (Jiao *et al.*, 2007; Chaudhary *et al.*, 2009). The transcription levels of MEP pathway enzymes were light induced in different plants, like sunflower, *Arabidopsis* and beech (Schuh *et al.*, 1997; Hsieh and Goodman, 2005; Wiberley *et al.*, 2009). A proteomic analysis of de-etiolated rice seedlings identified 52 proteins that were differentially displayed after exposure to light. Among these changed proteins, 33 were down-regulated and 19 were up-regulated (Yang *et al.*, 2007). In soybean, 27 genes were identified as responding to FR using a microarray analysis (Li *et al.*, 2011). In rice NSF45K, an oligo-microarray analysis was carried out to compare

the light- and dark-grown leaf tissues of four different cultivars. In total, 365 genes were more highly induced in the light-grown plants compared with the dark control, and the expression levels of the chlorophyll biosynthetic pathway genes were positively correlated with light illumination (Jung *et al.*, 2008). Kim *et al.* (2005) reported that rice *DXS* family genes had different responses to UV-irradiation and light-illumination. After UV-irradiation, *dxs3*'s expression increased up to 9-fold compared with that of the dark-grown control. After light-illumination, the *dxs1* expression level increased twofold compared with the dark control. Carotenoid biosynthetic genes, such as *OsPSY1*, *OsPSY2*, *OsPDS*, *OsZDS* and *OsCRTISO*, were expressed higher in light-grown rice compared with in dark-grown rice. In contrast, *OsPSY3* had stable expression with light illumination (Chaudhary *et al.*, 2009).

In my study I focused on identifying novel information regarding light regulation on primary and secondary metabolite profiles. To identify the genes essential to photomorphogenesis, we carried out metabolite profiles between etiolated and de-etiolated rice under distinct illumination times and analyzed the expression levels of selected pathway genes.

4.2. Aims

The aims of the experiments describes in this chapter were to gain a general metabolic profile and the involved gene expression profile of de-etiolated rice seedlings to gain insight into the light-induced metabolites and genes.

4.3. Material and Methods

4.3.1. Plant material

To obtain 14-day-old etiolated rice plants, wild type rice (*Oryza sativa* L. cv. EYI105) seeds were sterilized in 3.7% NaOCl for 20min and washed with sterilized water thoroughly, then were aseptically grown on Murashige and Skoog (MS) medium containing 0.3% phytagel in pots (diameter 10cm, height 20cm) in a controlled growth chamber at 25°C in complete darkness. Fourteen-day-old etiolated rice seedlings were illuminated with continuous white light ($100\mu\text{mol m}^{-2} \text{s}^{-1}$) for 0.5, 1, 2, 4, 6, 8, 10, 12 or

24h. Leaves of de-etiolated rice plants were rapidly frozen in liquid nitrogen after collected and stored at -80°C .

4.3.2 RNA isolation and cDNA synthesis

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and 1 μg of total RNA was used as the template for first-strand cDNA synthesis with Ominiscript Reverse Transcriptase (Qiagen) as described previously Jin *et al.* (2018).

4.3.3 Gene expression analysis by quantitative real-time PCR

First-strand cDNA was synthesized from 1 μg total RNA using Ominiscript Reverse Transcriptase in a 20 μl total reaction, and quantitative real-time RT-PCR (qRT-PCR) was carried out as previously described (Jin *et al.*, 2018) using the primers listed in Table 4.1. The amplified DNA fragments for each gene were confirmed by sequencing. The expression levels in different samples were normalized against rice *actin* mRNA. Three biological replicates each comprising three technical replicates were tested for each sample.

Table 4.1. List of primers used for the qRT-PCR.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
<i>OsDXS1</i>	CTCAAGGGAGGGAAGAACAA	ACACCTGCTTGTTGCTGTTG
<i>OsDXS2</i>	TGTTGTGGAGCTCGCTATTG	TCCTCCCACCTAGATCCCTT
<i>OsDXS3</i>	ACCTCCTCGGGAAGAAGAAG	GAGGGACACCTGCTTGTTGT
<i>OsDXR</i>	TCCACCCACAATCTATCA	GGTGACCTCTGAGCAATA
<i>OsMCT</i>	GACTTGAGGTCAGTATG	AGCAAGTCATCAGGAGTC
<i>OsCMK</i>	GTATTTGCAGCATGCGTCGT	ACCTGGTGCTTTCTACTCACAG
<i>OsMDS</i>	TCGTCGGTGTTGATGAGG	TTGTCTCCTTGAATGGGC
<i>OsHDS</i>	GGGTTGCATTGCAATGG	CTGAATCAAGGCGTCAGT
<i>OsHDR1</i>	GCTGAATGAGAAGAAGGTGC	AAAGGAAGCAGTGGCAAC
<i>OsHDR2</i>	GAGTAGGCGTTGTGAATCAAAC	CCTTTCCTGAGTAGCATTGC
<i>OsIPPI1</i>	GGTGGATGAACAAGACAATG	ATCGTTGCTGGAGTAGGAGT
<i>OsIPPI2</i>	TTAGTGGACGAACAGGACA	GCAGACCTTTGCTGAAGTAAC
<i>OsLIL3</i>	CGTGCTGCCATGATTGGCTTCTTCATG	GTCGTCTCATCGATGAGCTTCTTCACTG
<i>OsCHLG</i>	CTGACTTCTTTGTATAGCATAGCTGG	CTGCAACGGATAATTGAGTTATGTCA
<i>OsPORA</i>	CGGCCTCGTCTGAGTTTATTAT	CCTCTCTACTGAAAGCTGAAA
<i>OsPORB</i>	GTGAGTGAGAGTGATGTGCTATT	CTCCTCCATCGATCTTCTTGG
<i>OsCAO</i>	GGGAGACCTGGATGTGTTATG	ATGGTAAGGGCATTGGATTCT
<i>OsPSY1</i>	GAGGCGGGAAGAGGAGATAAAA	GGAAGGCAGATGCTGTCT GATT
<i>OsPSY2</i>	TTCTTTTGATTTGCGACGATTTTC	TGGCAGGCTTATGGCATA CA
<i>OsPDS</i>	GATCCAAACCGTTCAATGCTGG	TGCTACTCCGTTCAACCCATT
<i>OsZDS</i>	GCTGCCCTTTTGCCTTGA	CACAATAACGAAGAACAC ACCATTG

Table 4.1. (cont.)

Gene	Forward primer (5' - 3')	Reverse primer (5' - 3')
<i>OsLYCB</i>	CGTCCAGTACGACAAGCCGTA	AAGGGCATGGCGTAGAGGAACG
<i>OsLYCE</i>	GTATGGCAGGGTTCACAGGGAC	GCCAGCGTCATAGCATCGTCTC
<i>OsBCH2</i>	TCGAGAACGTGCCCTACTTCC	ACCCACCTCCTCCAACCTCCTT
<i>OsHGGT</i>	ATGGGGATTTCTTGAGGCTTTATC	GGCTGGGCTTATTGACCTTATC
<i>OsGGR</i>	GCACCCGAGGCCTAAGAGGG	GAAGTAGATCCCTCGCCGGAG
<i>OsVTE1</i>	ATAACTTTTGGGGAAGTAGGCATG	AAACTCGACTAGAAAATTCCTGAG
<i>OsVTE2</i>	TCACTGGTTTGTGGAGGCAG-	AAGAGTTGGCTTGTAACCTTATC
<i>OsVTE5</i>	GCGGGGTCTTCGAGCAGAAAC	GCATAGCTTCTGACCTGAAAAG
<i>OsVTE4</i>	GTCACTGTCTCTTGAGGATATAAG	TCTTATCGTCTTCCACCCACTTC
<i>OsHMGS1</i>	GCCTACGCCTTCTCCCAAT	ATGCCACGTCCTTCTCTC
<i>OsHMGS2</i>	GGGATGGACGCTACGGTCTT	TAGCAGCAGCACCACTGTT
<i>OsHMGS3</i>	TTGTTGCCTCCTGGGACGTT	GATCTCCTCGTCGGCCTTCC
<i>OsHMGR1</i>	CTCAAGGGAGGGAAGAACAA	ACACCTGCTTGTGTCGTTG
<i>OsHMGR2</i>	TGTTGTGGAGCTCGCTATTG	TCCTCCACCTAGATCCCTT
<i>OsHMGR3</i>	ACCTCCTCGGGAAGAAGAAC	GAGGGACACCTGCTTGTTGT
<i>OsMK</i>	CACTAGTGGATCGCGACCGT	TGGCCTGCACAATGCAAAC
<i>OsPMK</i>	ACTTGGGTTCATCAGCTGCCA	GATCTCGTCCAGCTGCGTTG
<i>OsMVD</i>	CCGTCACGACAGCATCAGC	ATCTCCTTGCCGTTGAGCCA

OsDXS1, rice (*Oryza sativa*) 1-deoxy-D-xylulose-5-phosphate synthase 1 gene; *OsDXS2*, rice 1-deoxy-D-xylulose-5-phosphate synthase 2 gene; *OsDXS3*, rice 1-deoxy-D-xylulose-5-phosphate synthase 3 gene; *OsDXR*, rice 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene; *OsMCT*, rice MEP cytidyltransferase gene; *OsCMK*, rice CDP-ME kinase gene; *OsMDS*, rice 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) synthase gene; *OsHDS*, rice 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMBPP) synthase gene; *OsHDR1*, rice HMBPP reductase 1 gene; *OsHDR2*, rice HMBPP reductase 2 gene; *OsIPPI1*, rice isopentenyl diphosphate isomerase 1 gene; *OsIPPI2*, rice isopentenyl diphosphate isomerase 2 gene; *OsLIL3*, rice light-harvesting-like protein 3 gene; *OsCHLG*, rice chlorophyll synthase gene; *OsPORA* *OsPORB*, rice protochlorophyllide oxidoreductase gene; *OsCAO* *OsPSY1*, rice phytoene synthase 1 gene; *OsPSY2*, rice phytoene synthase 2 gene; *OsPDS*, rice phytoene desaturase gene; *OsZDS*, rice zeta-carotene desaturase gene; *OsLYCB*, rice lycopene beta-cyclase gene; *OsLYCE*, rice lycopene epsilon cyclase gene; *OsBCH2*, rice beta-carotene hydroxylase 2 gene; *OsHGGT*, rice homogentisate geranylgeranyltransferase gene; *OsVTE1*, rice tocopherol cyclase gene; *OsVTE2*, rice homogentisate phytyltransferase gene; *OsVTE3*, rice MPBQ/MSBQ methyltransferase gene; *OsVTE4*, rice gamma-tocopherol methyltransferase gene; *OsHMGS1*, rice HMG synthase 1 gene; *OsHMGS2*, rice HMG synthase 2 gene; *OsHMGS3*, rice HMG synthase 3 gene; *OsHMGR1*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 1 gene; *OsHMGR2*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 2 gene; *OsHMGR3*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 3 gene; *OsMK*, rice MVA kinase gene; *OsPMK*, rice phospho-MVA kinase gene; *OsMVD*, rice MVA diphosphate decarboxylase gene.

4.3.4 Ultra-high-performance liquid chromatography

UPLC analysis as described in chapter 3.3.13

4.3.5. Gas chromatography–mass spectrometry

GS-MS analysis as described in chapter 3.3.14

4.4. Results

4.4.1 Amino acid responses to continuous WL illumination

The effects of WL illumination on amino acid levels in etiolated seedlings were examined by GC–MS analysis. All the amino acid contents in leaves declined after 24 h continuous WL illumination (Fig. 4.1). No significant changes in amino acid levels were found in the first 6 h of illumination. The leucine, phenylalanine, lysine, proline, valine and serine contents declined sharply after 12 h of illumination, and the amounts of these amino acids were found to be significantly lower compared with the amounts in dark controls. The glycine, gamma-aminobutyric acid, methionine, threonine, pyroglutamic acid, β -alanine and glutamine contents decreased significantly as the illumination time increased to 24h. Furthermore, alanine and asparagine contents also decreased compared with their levels in the dark control, but no significant differences were detected (Fig. 4.1).

4.4.2 Sugar and organic acid responses to continuous WL illumination

In addition to amino acids, other primary metabolites, like sugars and organic acids, were also readily detected by GC–MS. The relative sugar and organic acid levels were profiled and exhibited in Figs. 4.2 and 4.3, respectively. The sucrose and galactose levels were significantly lower after 24h of illumination compared with the dark control. While the levels of other sugars, such as glucose, inositol and xylose, showed no significant changes in rice during the de-etiolation process. In contrast, the relative levels of fructose and ribitol increased compared with the dark control. Ribitol was barely detected in the leaves of the dark control, while in the 6- and 12h illuminated seedlings, the ribitol contents sharply increased along with the illumination time to 24h (Fig. 4.2).

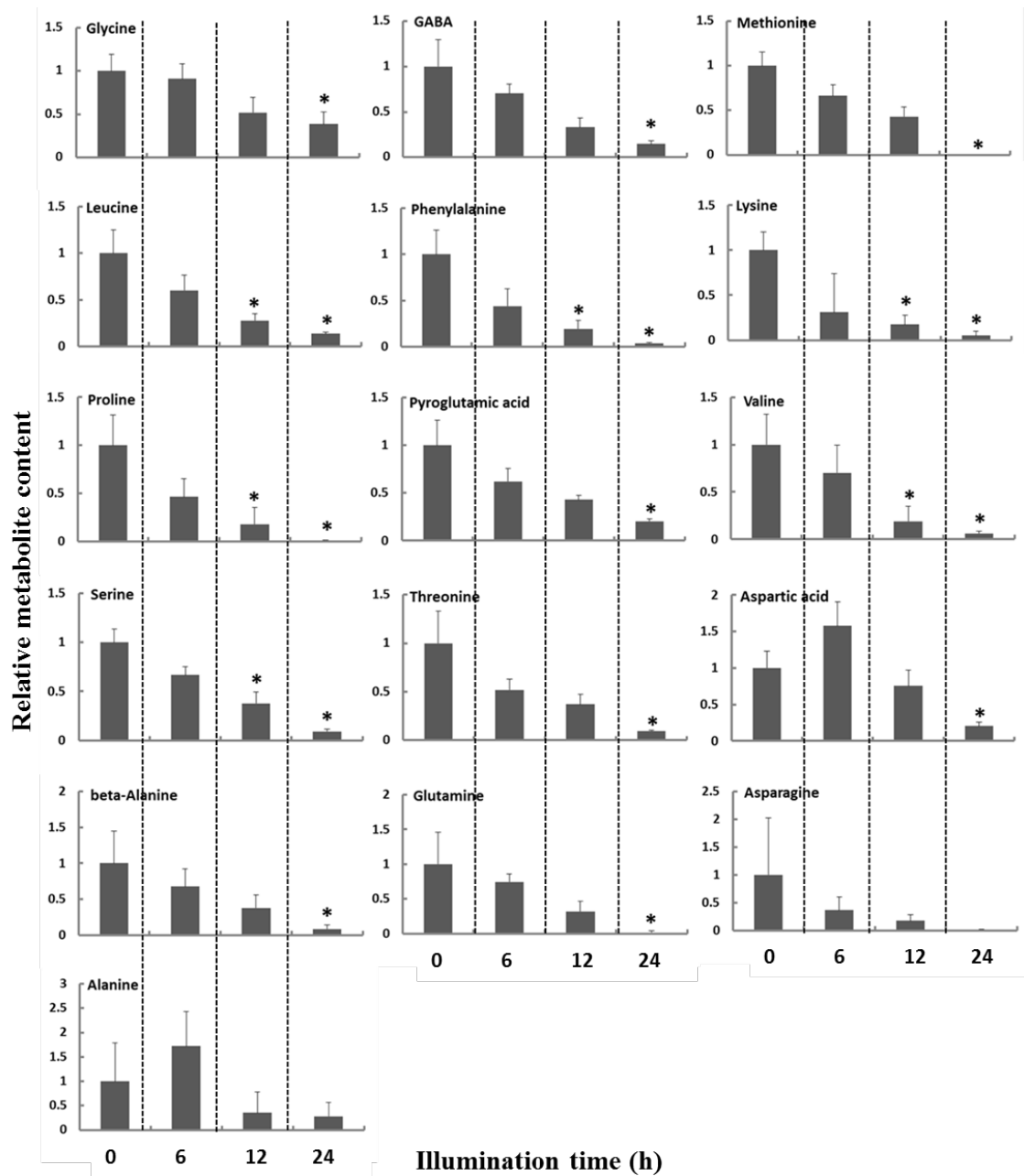


Figure 4.1. Relative amino acid contents of WT rice seedlings under different illumination time. Etiolated seedlings were subjected to white light irradiation. Metabolite levels in light-irradiated seedlings were normalized with those in the dark control seedlings. Values presented are the mean \pm SD of determinations on five replicates. An asterisk indicates values to be significantly different from value of dark control by means of Student's *t*-tests ($P < 0.05$). GABA, c-aminobutyric acid.

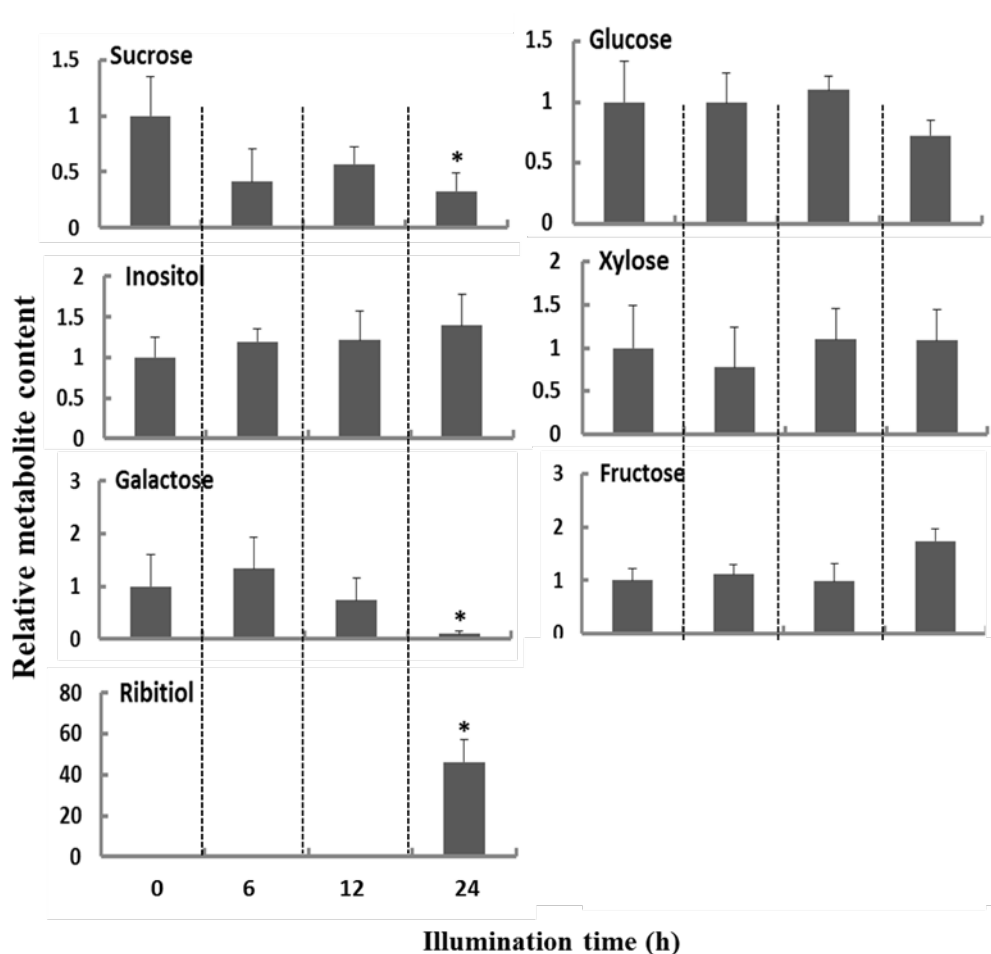


Figure 4.2. Relative sugar contents of WT rice seedlings under different illumination time. Etiolated seedlings were subjected to white light irradiation. Metabolite levels in light-irradiated seedlings were normalized with those in the dark control seedlings. Values presented are the mean \pm SD of determinations on five replicates. An asterisk indicates values to be significantly different from value of dark control by means of Student's *t*-tests ($P < 0.05$).

The dynamic changes in organic acids are shown in Fig. 4.3. The relative levels of oxalic acid were not significantly different among the varied illumination times. In addition, five more organic acids related to the TCA cycle, namely aconitic, citric, fumaric, itaconic and malic acids, were detected (Fig. 4.3). The levels of aconitic, citric and itaconic acids in rice leaves showed significant decreases after 24h of illumination. In addition, there were no significant changes in the relative levels of fumaric and malic acids. Aconitic acid was barely detected after a 24h exposure to WL.

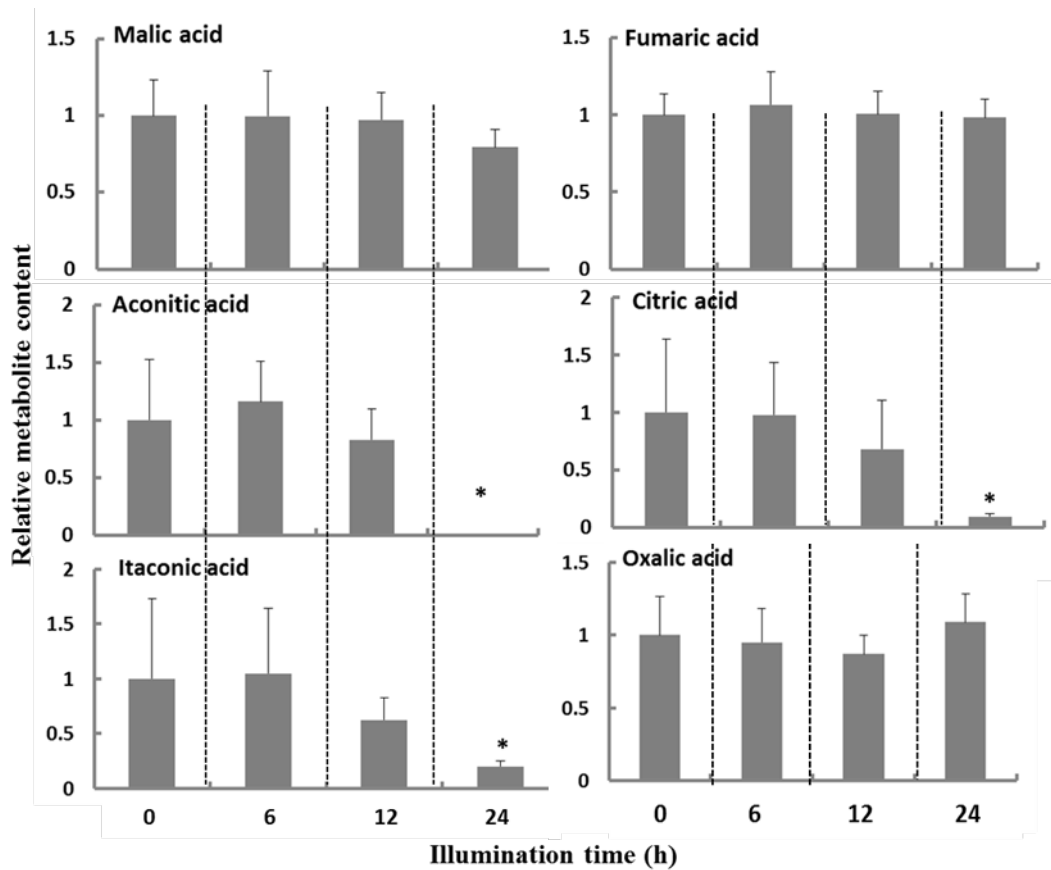


Figure 4.3. Relative organic acid contents of WT rice seedlings under different illumination time. Etiolated seedlings were subjected to white light irradiation. Metabolite levels in light-irradiated seedlings were normalized with those in the dark control seedlings. Values presented are the mean \pm SD of determinations on five replicates. An asterisk indicates values to be significantly different from value of dark control by means of Student's *t*-tests ($P < 0.05$)

4.4.3. Fatty acid and other metabolite responses to continuous WL illumination

Fatty acids were identified and profiled in Fig. 4.4. The levels of detected fatty acids were stable during the rice de-etiolation process, with no significant changes found compared with the dark control. The levels of cell wall compounds and nucleotides were also stable, without any significant changes (Fig. 4.5).

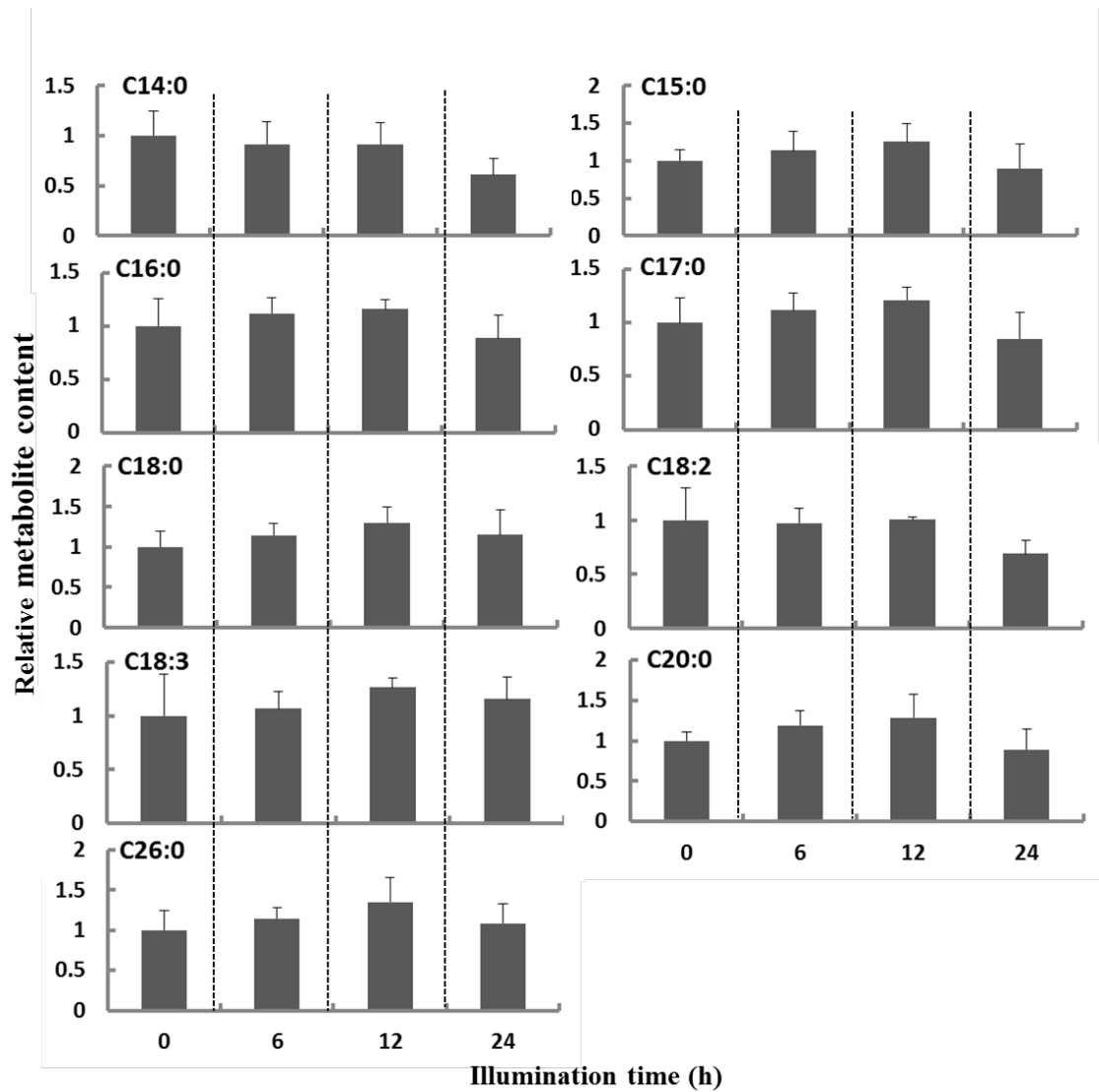


Figure 4.4. Relative fatty acid contents of WT rice seedlings under different illumination time. Etiolated seedlings were subjected to white light irradiation. Metabolite levels in light-irradiated seedlings were normalized with those in the dark control seedlings. Values presented are the mean \pm SD of determinations on five replicates.

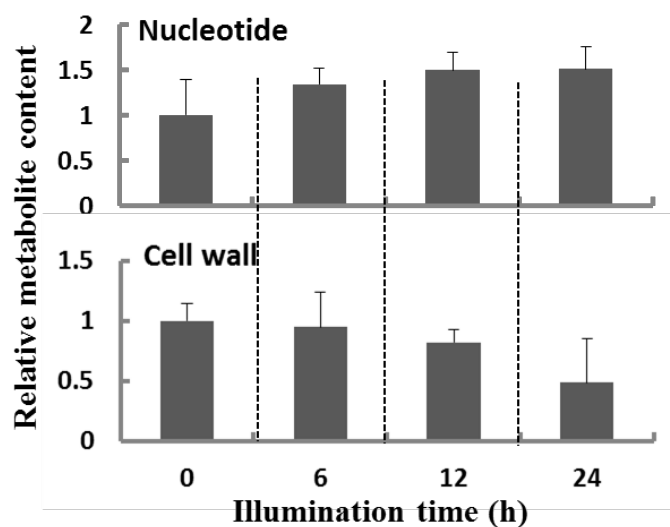


Figure 4.5. Relative cell wall and nucleotide contents of WT rice seedlings under different illumination time. Etiolated seedlings were subjected to white light irradiation. Metabolite levels in light-irradiated seedlings were normalized with those in the dark control seedlings. Values presented are the mean \pm SD of determinations on five replicates.

4.4.4. Isoprenoid responses to continuous WL illumination and isoprenoid biosynthetic gene expression profiles

The isoprenoids like chlorophylls, carotenoids, tocopherols and phytosterols were also detected and measured using UPLC or GC-MS. The levels of chlorophylls and carotenoids, and relative levels of phytosterols are shown and described in Chapter 3 Fig. 3.5. The quantity of α -tocopherol increased in response to illumination. In the first 6h after illumination, α -tocopherol increased 2.5-fold (from $35.54 \pm 16.73 \mu\text{g g}^{-1}$ dry weight to $88.83 \pm 15.87 \mu\text{g g}^{-1}$ dry weight) and then increased slightly up to \sim 3.5-fold compared with the dark control after 24h illumination. While α -tocopherol hydroquinone amounts showed an increasing trend, no significant differences were detected during rice seedling de-etiolation (Fig. 4.6).

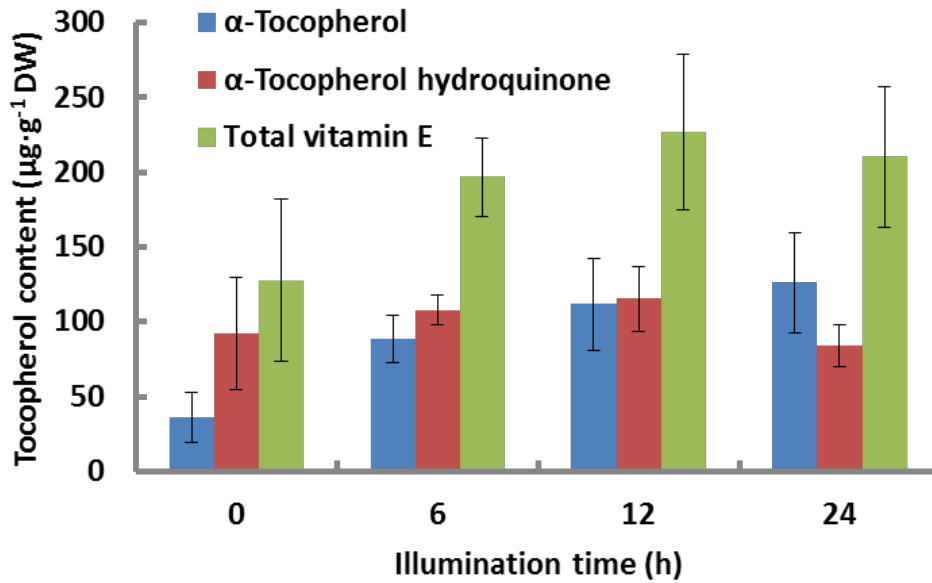


Figure 4.6. Tocopherol composition and content in the etiolated leaves of 14-day-old rice plants during de-etiolation. in the same leaves during de-etiolation.

The chlorophyll, carotenoid, phytosterol and tocopherol levels were different after distinct illumination times, and the expression levels of the genes coding for enzymes involved in isoprenoid biosynthesis were studied further. The MVA and MEP pathway genes' expression levels were investigated using qRT-PCR, and the dynamic changes that occurred along with increased illumination time are shown in Fig. 4.8. Three *OsDXS* genes, *OsDXS1–3*, from the MEP pathway had different responses to light illumination. The expression level of *OsDXS1* decreased slightly in the first 0.5h of illumination. Then, *OsDXS1* started to increase, and the expression level reached a peak after 6h illumination. The peak was fourfold greater compared with that in the dark control. After peaking, the *OsDXS1* expression level started to decline, and finally, the expression level at 24h of illumination was 2.3-fold greater than that in the dark control. The *OsDXS2* expression level was stable during the first 1h of illumination and decreased slightly after 2h of illumination, after which no significant changes were observed until 24h of illumination. *OsDXS3* had a very low expression level in rice leaves compared with the other two *OsDXS* genes, and there were no significant changes in the expression level in rice leaves during de-etiolation (Fig. 4.7). The other genes from the MEP pathway, such as *OsDXR*, *OsMCT*, *OsCMK*, *OsMDS* and *OsHDS*, had positive responses to WL illumination. The expression levels of these genes were stable in the first 0.5h of illumination and then started to increase. The expression level

of *OsDXR* increased rapidly and reached a peak after 4h of illumination (7.6-fold higher than in the dark control). It then stayed at a high level until 10h of illumination, from 10 to 12h of illumination, the gene expression level decreased dramatically (the expression level was a half that at the peak) and from 12 to 24h WL illumination the expression level of *OsDXR* showed a tendency to increase again. *OsMCT*, *OsCMK*, *OsMDS* and *OsHDS* exhibited very similar responses to continuous WL illumination at the early stage. The gene expression levels increased rapidly in the first 6h of illumination and were then stable up to 12h of illumination. Afterwards, the expression levels of *OsMCT* and *OsCMK* decreased rapidly from 12 to 24h, whereas the expression levels of *OsMDS* and *OsHDS* started to decrease after 10h of illumination (Fig. 4.7). The dynamic changes in the *OsHDR* gene family's mRNA levels have been described by Jin *et al.* (2019) previously.

Regarding the levels of chlorophylls, carotenoids, phytosterols and tocopherols were found to be different under distinct illumination times, the expression of the genes coding for enzymes, which involved in isoprenoids biosynthesis was further studied. The MVA and MEP pathway genes expression levels were investigated by qRT-PCR and the dynamic changes with increased illumination time are shown in Fig. 4.7. Three *OsDXS* genes from MEP pathway, namely *OsDXS1*, *OsDXS2* and *OsDXS3*, had different responses to light irradiation. The expression levels of *OsDXS1* decreased slightly in the first 0.5h irradiation, then *OsDXS1* started to increase and the expression level reached the peak after 6h irradiation, which was 4-fold higher compared to dark control, after that the expression level of *OsDXS1* started to decline and finally the expression level at 24h of illumination was 2.3-fold higher than dark control. While the expression level of *OsDXS2* was stable during the first 1h illumination, then decreased slightly at 2h of irradiation, and till 24h of irradiation no significant changes were observed. *OsDXS3* had very low expression in rice leaves compared with the other two *OsDXS* genes, and there were no significant changes in the expression level in rice leaves during de-etiolation (Fig. 4.7). The dynamic changes in the *OsHDR* gene family's mRNA levels have been described in chapter 3.

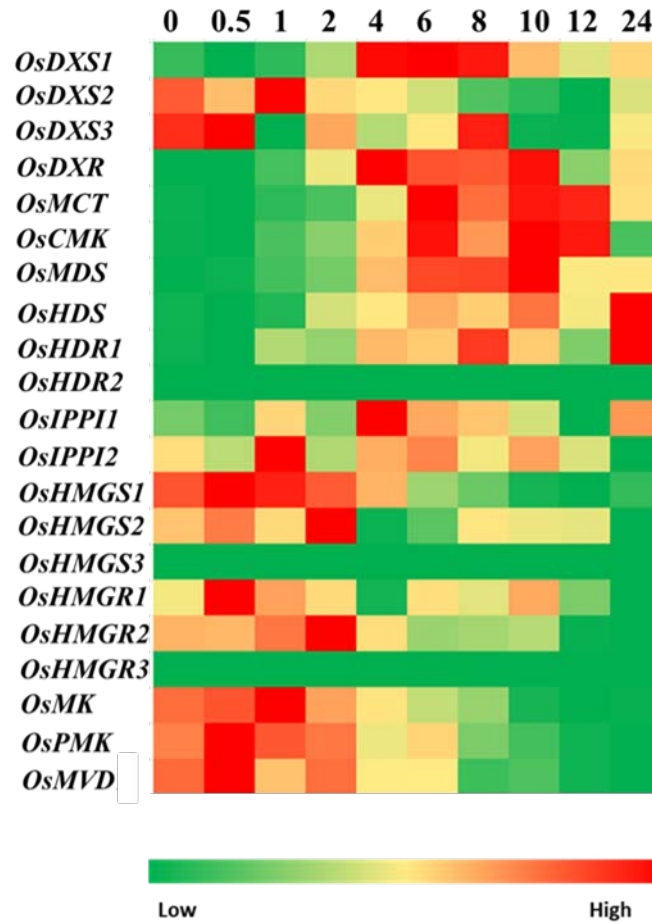


Figure 4.7. Relative expression levels of MVA pathway and MEP pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light

OsDXS1, rice 1-deoxy-D-xylulose-5-phosphate synthase 1 gene; *OsDXS2*, rice 1-deoxy-D-xylulose-5-phosphate synthase 2 gene; *OsDXS3*, rice 1-deoxy-D-xylulose-5-phosphate synthase 3 gene; *OsDXR*, rice 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene; *OsMCT*, rice MEP cytidyltransferase gene; *OsCMK*, rice CDP-ME kinase gene; *OsMDS*, rice 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cPP) synthase gene; *OsHDS*, rice 1-hydroxy-2-methyl-2-butenyl 4-diphosphate (HMBPP) synthase gene; *OsHDR1*, rice HMBPP reductase 1 gene; *OsHDR2*, rice HMBPP reductase 2 gene; *OsIPPI1*, rice isopentenyl diphosphate isomerase 1 gene; *OsIPPI2*, rice isopentenyl diphosphate isomerase 2 gene; *OsHMGS1*, rice HMG synthase 1 gene; *OsHMGS2*, rice HMG synthase 2 gene; *OsHMGS3*, rice HMG synthase 3 gene; *OsHMGR1*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 1 gene; *OsHMGR2*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 2 gene; *OsHMGR3*, rice 3-hydroxy-3-methylglutaryl-CoA reductase 3 gene; *OsMK*, rice MVA kinase gene; *OsPMK*, rice phospho-MVA kinase 1 gene; *OsMVD1*, rice MVA diphosphate decarboxylase gene;

In the MAV pathway, the *OsHMGR* gene family showed distinct reactions to continuous WL illumination. There were no significant changes in the expression of *OsHMGR1* until after 12h of illumination, while from 12 to 24h of illumination the expression level dropped by 50% compared with the dark control. *OsHMGR2* showed a

twofold increase at this stage. Then, the expression of *OsHMGR2* started to decrease sharply at 4h of illumination, and its mRNA level in the dark control was 7.9-fold greater than after 24h of illumination (Fig. 4.7). The expression of *OsHMGR3* showed a slight increase in the first 0.5h of illumination and then started to decrease until after 24h of illumination. The expression level of *OsHMGR3* in the dark control was ~18-fold greater than in the leaves of seedlings exposed to light for 24h. The expression changes in the *OsIPPI* gene family, involved in both MVA and MEP pathways, are described in Figure 3.5

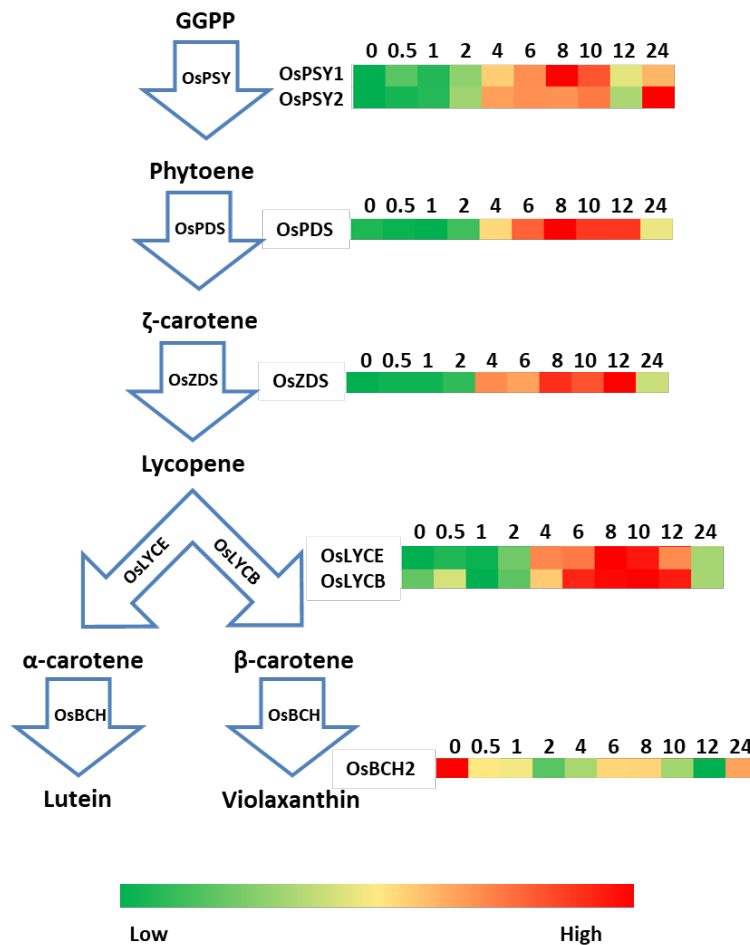


Figure 4.8. Relative expression level of carotenoids pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light. *OsPSY1*, rice phytoene synthase 1 gene; *OsPSY2*, rice phytoene synthase 2 gene; *OsPDS*, rice phytoene desaturase gene; *OsZDS*, rice zeta-carotene desaturase gene; *OsLYCE*, rice lycopene epsilon cyclase gene; *OsLYCB*, rice lycopene beta-cyclase gene; *OsBCH2*, rice beta-carotene hydroxylase 2 gene.

The expression levels of selected genes related to the carotenoid biosynthetic pathway are shown in Fig. 4.8. *OsPSY1* and *OsPSY2*'s expression levels started to increase immediately after a 10h exposure to continuous WL illumination (~5- and 25-fold, respectively, greater than the expression level in the dark control) and decreased sharply after 10h. There were no significant changes in the expression level of *OsPSY1* from 12 to 24h of illumination. In contrast, the expression level of *OsSPY2* increased rapidly again after 12h, and the expression level was the highest at 24h. The expression levels of *OsPDS*, *OsZDS*, *OsLYCB* and *OsLYCE* increased in the first 12h of illumination (3.8-, 8.5-, 3.7- and 2.2-fold greater, respectively, compared with the dark control). They later showed decreasing trends from 12 to 24h of illumination. Unlike the other genes, *OsBCH2*'s expression level decreased after exposure to WL, and the expression level after 24h of illumination decreased to ~50% of that of the dark control (Fig. 4.8).

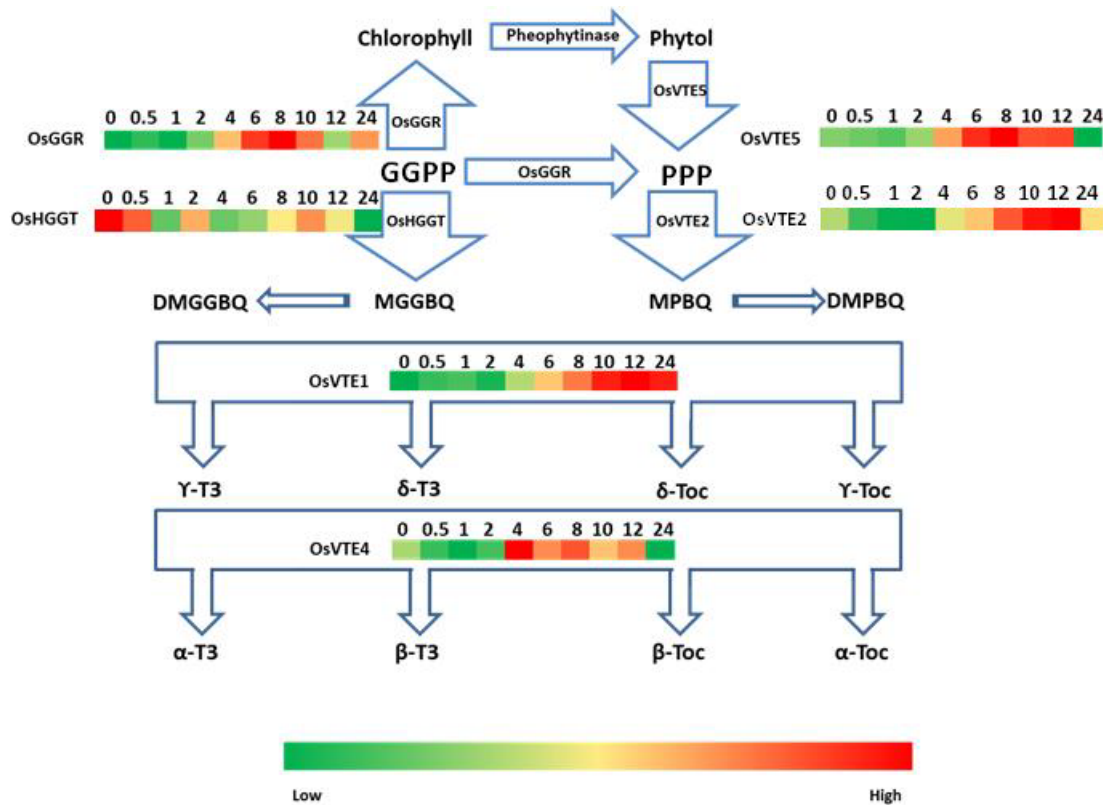


Figure 4.9. Relative expression level of tocopherol pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light. *OsGGR*, rice geranylgeranyl reductase gene; *OsHGGT*, rice homogentisate geranylgeranyltransferase gene; *OsVTE2*, rice homogentisate phytyltransferase gene; *OsVTE1*, rice tocopherol cyclase gene; *OsVTE4*, rice gamma-tocopherol methyltransferase gene; *OsVTE5*, rice phytol kinase gene

The expression levels of selected tocopherol biosynthetic pathway genes in de-etiolated rice are shown in Fig. 4.9. The expression levels of *OsGGR*, *OsVTE5*, *homogentisate phytyltransferase (VTE2)*, *tocopherol cyclase (VTE1)* and *γ-tocopherol methyltransferase (VTE4)* were all up-regulated by WL. The expression levels of *OsGGR*, *OsVTE5* and *OsVTE2* increased rapidly during the first 6 h of illumination and remained stable up to 10 h of illumination. Afterward, the expression levels started to decrease. Those of *OsGGR* and *OsVTE2* increased again after 12h of illumination whereas *OsVTE5* steadily decreased, showing its lowest expression level at 24h of illumination. The expression level of *OsVTE1* kept increasing after exposure to light, and the highest expression levels was detected at 12h of illumination. There were no significant differences in expression levels among 10-, 12- and 24h illuminated samples. The expression level of *OsVTE4* increased rapidly in the first 4h of illumination and kept decreasing until after 24h of illumination. In contrast, *OsHGGT* was down-regulated by WL. After exposure to light, the expression level decreased to 3-fold greater than the expression in the dark control after 24h of illumination.

Changes in the expression levels of selected chlorophyll biosynthetic pathway genes during rice de-etiolation are shown in Fig. 4.10. The expression level of *OsCHLIL3* increased sharply in the first 4h and then remained stable. The expression levels of rice *protochlorophyllide oxidoreductase (POR) A* and *PORB* showed different reactions to WL illumination. The *OsPORB* mRNA level increased significantly in the first 8h, then decreased rapidly until 12h of illumination. No significant difference in the expression level was detected between 12 and 24h of illumination, and the expression level at 24h was approximately 3-fold greater than in the dark control. In contrast, the expression of *OsPORA*, started to decline when it was exposed to WL. This result indicated that *OsPORA* was down-regulated by light. The rice *chlorophyll synthase* gene showed an increasing trend in the first 8h of illumination and remained stable from 8 to 12h. It then decreased immediately to the same level as in the dark control. The expression level of *OsCAO* reached a peak after 6h of illumination, then decreased till 12h, and slightly increased again from 12 to 24h.

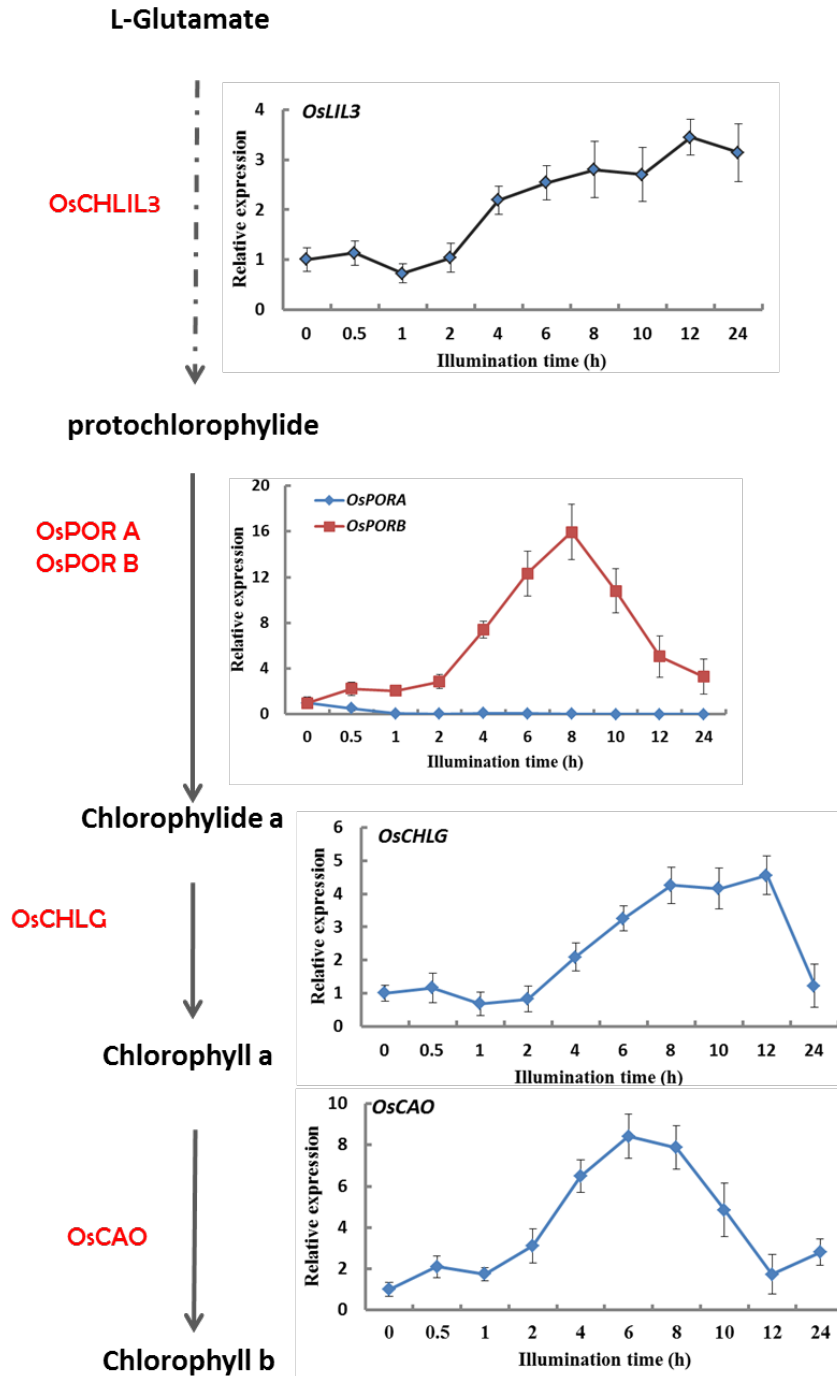


Figure 4.10. Relative expression levels of chlorophyll biosynthesis pathway genes in etiolated leaves during de-etiolation at various times after the onset of irradiation with white light. *OsLIL3*, rice light-harvesting-like protein 3 gene; *OsCHLG*, rice chlorophyll synthase gene; *OsPORA*, rice protochlorophyllide oxidoreductase A gene; *OsPORB*, rice protochlorophyllide oxidoreductase B gene; *OsCAO*, rice chlorophyll a oxygenase gene.

4.5. Discussion

In this study, UPLC and GC–MS were used as reliable methods for identifying and quantifying the metabolites of rice seedlings during de-etiolation. According to amino acid profiles, we observed a continuous reduction in the amino acid contents of de-etiolated rice leaves from 6 to 24h of illumination. During the first 6h of WL illumination, we did not observe any significant differences in the relative levels of the detected amino acids.

In *Arabidopsis*, the study of Jumtee *et al.* (2008) showed that alanine, phenylalanine, leucine, isoleucine and valine levels significantly decreased after 24h WL illumination, similar to our results. Interestingly, aspartic acid and glutamine contents were significantly increased after 6h. We also observed increases in the contents of aspartic acid and alanine during the first 6h of illumination but the differences were not significant. They also investigated changes in the amino acid contents of *Arabidopsis* after exposure to FR. However, no significant changes were detected during 6h of illumination, after which, the contents of β -alanine, methionine, phenylalanine, proline, glycine, leucine, lysine, valine, serine and threonine decreased significantly after 24h of FR illumination (Jumtee *et al.*, 2008). The dynamic changes in the contents of these amino acids were very similar to the amino acid level changes during rice de-etiolation. As essential precursors of protein synthesis, the consumption of large amounts of amino acids is required during the rice photomorphogenesis. In addition to protein synthesis, some amino acids (phenylalanine, tyrosine and tryptophan) also act as precursors to secondary metabolites (such as pigments, alkaloids, hormones and cell wall components), which are important for plant growth and development, as well as environmental responses (Vered and Gad, 2010; Maeda and Dudareva, 2012; Michard *et al.*, 2011). These processes could also increase the demand for amino acids in de-etiolated rice leaves. The amino acid composition and biosynthetic pathways are highly regulated by light (Stitt *et al.*, 2002; Ma *et al.*, 2001; Jumtee *et al.*, 2008). The determinant of the amino acid accumulation level in de-etiolated rice was the balance between biosynthesis and catabolism. We detected a significant decrease in the proline content after exposed to WL, and similar changes have also been reported in *Arabidopsis* (Jumtee *et al.*, 2008). The gene encoding pyrroline-5-carboxylate synthetase, which converts glutamate to glutamate-semialdehyde during proline

biosynthesis is induced by light (Abrahám *et al.*, 2003; Szabados *et al.*, 2009). Additionally, the expression of proline dehydrogenase, which is repressed by light in proline catabolism, was reported (Sato *et al.*, 2002). The up-regulated biosynthesis and down-regulated catabolism should cause an increase in the proline content; however, the proline content decreased significantly in both de-etiolated Arabidopsis and rice. This could be a consequence of the activation of protein biosynthesis and regulation of plant development. Because proline is involved in the production of polyamines, protection of protein integrity under stress, enhancement of enzyme activities, and regulation of plant development, high levels can be required (Szabados *et al.*, 2009; Ashrafa *et al.*, 2007). Amino acids, like valine, serine, glycine and phenylalanine, are mainly involved in protein biosynthesis, and reductions in these amino acid levels indicate large amounts of protein production during plant de-etiolation (Jumtee *et al.*, 2008).

Sugar is not only a carbon and energy source for plants, it also regulates plant growth and development and has an essential role in defense responses (Ciereszko *et al.*, 2018). Carbon dioxide is fixed through the Calvin cycle, and hexose phosphates are produced through gluconeogenesis. Then, these sugar products are converted into starch and stored in plants. All these pathways require the energy provide by photosynthesis, indicating the indispensable role of light in sugar biosynthesis (Osanai *et al.*, 2007; Liam *et al.*, 2013). The sugar biosynthetic genes are up-regulated by light (Ma *et al.*, 2001; Jumtee *et al.*, 2008). In WT Arabidopsis, most of the sugars were consumed after 24h of FR illumination, whereas only the ribose and fucose contents were reduced under WL illumination. However, in our results, only sucrose and galactose levels showed significant decreases in rice leaves after 24h of WL illumination, while the others remained unchanged or increased slightly. The ribitol level showed a sharp increase after 24h of WL illumination. The up-regulation of sugar biosynthetic genes may stabilize or increase the accumulation of sugars in rice. Light might be required for sugar consumption, such as in the conversion of sugar into fatty acids during photomorphogenesis in plants. This could be why the sugar content decreases during de-etiolation (Ghassemian *et al.*, 2006).

Jumtee *et al.* (2008) found relatively steady pools of organic acids in Arabidopsis compared with those of amino acids and sugars. The levels of organic acids in WT seedlings showed no changes during FR illumination, whereas the glycolate content

decreased significantly during WL illumination, which could be caused by the up-regulation of the *glycolate oxidase* gene by light (Barak *et al.*, 2001; Jumtee *et al.*, 2008). The genes associated with the TCA cycle were reportedly up-regulated by light (Ma *et al.*, 2001; Nunes-Nesi *et al.*, 2013). The expression of succinate dehydrogenase, which converts succinate to fumaric acid, in *Arabidopsis* was found to increase first and then decrease after transitioning from dark to light (Popov *et al.*, 2010). This may explain why the fumaric acid level remained stable during rice de-etiolation.

Plant fatty acids are synthesized mainly in chloroplasts, and this synthesis increases in the light and decreases in the dark in *Arabidopsis* (Sasaki *et al.*, 1997; Rawsthorne, 2002). In my results, fatty acid levels remained stable during rice de-etiolation. This could occur because of the equivalence between the biosynthetic and catabolic rates during rice de-etiolation. The biosynthesis increased because the involved genes were up-regulated, while plant photomorphogenesis required increasing amounts of fatty acids for cellular membrane synthesis in all the cells.

The chlorophyll, carotenoid and tocopherol levels increased after transitioning from dark to light in various plants (Zhou *et al.*, 2017; Vranová *et al.*, 2013; Sapir-Mir *et al.*, 2008). In rice leaves, this included chlorophyll *a/b*, carotenoid, tocopherol and phytosterol levels. The study of Ghassemian *et al.* (2006) demonstrated that the amounts significantly increased during the de-etiolation process in response to FR in *Arabidopsis* seedlings (Ghassemian *et al.*, 2006).

Carotenoids, the side chains of chlorophylls and tocopherols are all derived from the MEP pathway, while phytosterols are derived from the MVA pathway (Vranová *et al.*, 2013; Cordoba *et al.*, 2009; Ghassemian *et al.*, 2006; Rodriguez-Concepcion *et al.*, 2006). The analysis in *Arabidopsis* revealed that the genes involved in the MVA and MEP pathways behaved differently in response to light stimulation. The expression levels of MVA pathway genes are inhibited by light, whereas the expression levels of MEP pathway genes are upregulated by light (Vranová *et al.*, 2013).

The MEP pathway has seven enzymatic steps, starting with the condensation of pyruvate and glyceraldehyde-3-phosphate to form 1-deoxy-d-xylulose 5-phosphate (DXP), which is catalyzed by 1-deoxy-d-xylulose 5-phosphate synthase (DXS)

(Sprenger *et al.*, 1997; Lois *et al.*, 1998). Next, DXP reductoisomerase (DXR) catalyzes the isomerization and reduction of DXP to MEP. MEP is then converted into isopentenyl diphosphate (IPP) and dimethylallyl pyrophosphate through five enzymatic reaction (MCT, CMK, MDS, HDS and HDR) steps. *A. thaliana* has three DXS-like genes and only one encodes a functional DXS enzyme. The other two proteins do not have DXS activity (Estevez *et al.*, 2000; Vranová *et al.*, 2013). The metabolic flux of the MEP pathway in spinach leaves also showed that its flux increases with light intensity (Mongelard *et al.*, 2011). In rice, three isomerases of DXS have been reported (OsDXS1–3). Kim *et al.* (2005) reported that, in the case of light-illumination, only the relative expression of *dxs1* increased, based on an unilluminated control. In our results, *OsDXS1* increased fourfold after 4h of illumination, whereas the expression levels of *OsDXS2* and *OsDXS3* decreased after exposure to light. Thus, the *DXS* genes played distinct roles in the control of the MEP pathway. The appearance of isoprenoids in the de-etiolated rice leaves was coincident with the time of the positive response of *OsDXS1* to light, suggesting that *OsDXS1* might be involved in chlorophyll biosynthesis. The study of Huang *et al.* (2017) reported the levels of photosynthetic pigments in an *OsMDS* mutant, which displayed a yellow–green leaf phenotype, and the chlorophyll and carotenoid levels were significantly lower than those in WT rice. *OsMDS* is essential for the biosynthesis of chlorophylls and carotenoids (Huang *et al.*, 2017). The rest of the rice MEP pathway genes, which encode functional enzymes had positive correlations with light illumination that mirrored the MEP pathway-derived isoprenoid accumulation.

Phytosterols are derived from MVA pathway. In our study, the level of β -sitosterol increased significantly in the first 6h of light illumination, and then decreased after 12h of illumination to the same level as the dark control. After 24h of light illumination, there were no differences between the levels of campesterol and stigmasterol compared with the dark control. The MVA pathway is catalyzed by acetyl-coenzyme A(CoA) C-acetyltransferase (AACT), then acetoacetyl-CoA is converted to 3-hydroxy-3-methylglutaryl-CoA by HMG synthase. 3-Hydroxy-3-methylglutaryl-CoA is converted to MVA in two reduction steps, which are catalyzed by the enzyme 3-hydroxy-3-methylglutaryl-CoA reductase. Then, MVA is phosphorylated to MVA 5-diphosphate through two steps, catalyzed by MVA kinase and phospho-MVA kinase. The last step

of IPP biosynthesis is a decarboxylation of MVA 5-diphosphate, which is catalyzed by MVA diphosphate decarboxylase (Vranová *et al.*, 2013).

Because the biosynthesis of chlorophyll, carotenoid and tocopherol (Vitamin E) are all derived from the products of the MEP pathway, the genes involved in their biosynthesis are regarded as downstream of the MEP pathway and their expression patterns are predicted to be similar, having positive responses to light (Burlat *et al.*, 2004; Aharoni *et al.*, 2005; Jung *et al.*, 2008). Dimethylallyl pyrophosphate in the MEP pathway is converted into geranyl diphosphate and geranylgeranyl diphosphate (GGPP), and in the MVA pathway it is converted into farnesyl diphosphate and GGPP (Vranová *et al.*, 2013). GGPP represents an essential metabolic node, as a precursor for vital metabolic branches, including chlorophyll, carotenoid and tocopherol biosynthesis, which are produced by plastidic GGPP synthase (GGPPS) in plants (Zhou *et al.*, 2017). In rice, only one functionally active GGPPS (*OsGGPPS1*) has been found in chloroplasts (Zhou *et al.*, 2017). Zhou *et al.* (2017) overexpressing *OsGGPPS1* in rice, and the transgenic rice plants showed increased growth and enhanced chlorophyll, but not carotenoid, contents. The expression analysis results showed that *OsGGPPS1* was up-regulated by light, with the expression level being increased more than 10-fold after 8h of light illumination compared with in the dark control. Yang *et al.* (2007) carried out a proteomic analysis of rice seedlings during de-etiolation, and they detected two proteins related to light reactions in chlorophyll biosynthesis, indicating that the biosynthesis of chlorophyll was correlated with the transition from dark to light (Yang *et al.*, 2007). The other genes encoding enzymes involved in chlorophyll biosynthesis, such as rice light-harvesting-like protein 3, chlorophyll synthase, *OsPORB* and *OsCAO*, were up-regulated by light, except *OsPORA*.

Phytoene synthase (*PSY*) is the rate-limiting enzyme of carotenogenesis. In Arabidopsis seedlings, the expression level of *PSY* increased under the illumination of both continuous FR and red light (Von Lintig *et al.*, 1997). In addition, the results of global microarray expression correlation analysis revealed that *PSY* is highly co-expressed with many photosynthesis-related genes, including many isoprenoid-related biosynthetic pathway genes (Meier *et al.*, 2011). In white mustard, an up-regulation of *PSY* expression was observed with light exposure, while *PDS* and *GGPS* expression levels remained constant (Welsch *et al.*, 2000). In green algae, the expression levels of

PSY, *PDS*, *BCH* and *LYCE* were rapidly up-regulated in response to light (Steinbrenner *et al.*, 2003; Bohne *et al.*, 2002). The expression levels of *OsPSY1/2*, *OsPDS*, *OsZDS*, *OsLYCB* and *OsLYCE* increased after exposed to WL, whereas *OsBCH2*'s expression level decreased compared with that of the dark control. Chaudhary *et al.* (2010) reported that *OsPSY1* and *OsPSY2* have greater expression levels in light-grown tissues compared with in dark-grown tissues. In contrast, *OsPSY3* has similar expression levels in light- and dark-grown root and shoot tissues (Chaudhary *et al.*, 2010). *OsPDS*, *OsZDS*, *OsCRTISO*, *Os β CY* and *Os ϵ LCY* also have greater expression levels in light-grown tissues. The three β -hydroxylase genes, *Os β OHI-3*, also have maximum expression levels in light-grown shoots. The expression patterns of the three ϵ OH-encoding genes of the pathway, *Os ϵ OHI-3*, are quite similar, and the expression levels in light-grown shoots were slightly greater than in dark-grown shoots. *OsZE* and *OsVDE* also have greater expression levels in light-grown shoots (Chaudhary *et al.*, 2010).

In plants, tocopherols are synthesized at the inner chloroplast envelope (Mene-Saffrane and DellaPenna, 2010). The formation of 2-methyl-6-phytyl-benzoquinone (MPBQ) from the condensation of homogentisic acid and a prenyl side chain (phytyl diphosphate) is the first committed step in the pathway and is catalyzed by VTE2. Studies in *Arabidopsis* and *Synechocystis* showed that *AtVTE2* and *SynVTE2* activities are limiting in tocopherol synthesis (Collakova and DellaPenna, 2003a, 2003b). Total tocopherol levels increase in plants when VTE2 is overexpressed compared with WT controls (Li *et al.*, 2012; Maeda *et al.*, 2006; Schelbert *et al.*, 2009). Our results indicated that *OsVTE2* was up-regulated by light. MPBQ can be methylated to form 2,3-dimethyl-6-phytylbenzoquinone by MPBQ methyltransferase. In rice, the greatest expression levels of *MPBQ methyltransferase* were detected at 12h of illumination (twofold greater than in the dark control). MPBQ and 2,3-dimethyl-6-phytylbenzoquinone can then be cyclized via VTE1 to form δ - and γ -tocopherol, respectively. *OsVTE1* was also up-regulated by WL in rice. The last enzyme of the pathway, VTE4, methylates δ - and γ -tocopherol to form β - and α -tocopherol, respectively (Li *et al.*, 2012). We found that *OsVTE4* increased rapidly during the early stage of rice de-etiolation. In rice, the tocopherol biosynthetic genes were up-regulated by light, which mirrored the increasing tocopherol content during de-etiolation.

4.6. Conclusions

The results showed the primary metabolites like amino acids, some sugars and TCA acids showed decline in the rice seedling de-etiolation process, while other metabolites like fatty acids and organic acid showed stable metabolites pool during de-etiolation. The expression profiles of genes involved in isoprenoid biosynthesis in de-etiolated rice leaves mirrored the accumulation of the corresponding metabolites. Furthermore isoprenoid levels (metabolites) and mRNA levels of the corresponding genes were positively correlated and with illumination time.

4.7. References

- Aharoni A, Jongsma MA, Bouwmeester HJ (2005). Volatile science? Metabolic engineering of terpenoids in plants. *Trends Plant Sci.*, **10**(12):594-602.
- Abrahám E, Rigó G, Székely G, Nagy R, Koncz C, Szabados L (2003). Light-dependent induction of proline biosynthesis by abscisic acid and salt stress is inhibited by brassinosteroid in Arabidopsis. *Plant Mol. Biol.*, **51**:363–372.
- Alabadí D, Gil J, Blázquez MA, García-Martínez JL (2004). Gibberellins repress photomorphogenesis in darkness. *Plant Physiol.*, **134**:1050–1057.
- Ashraf M, Foolad MR (2007). Roles of Glycine Betaine and Proline in Improving Plant Abiotic Stress Resistance. *Environmental and Experimental Botany*, **59**(2):206-216.
- Barak S, Nejidat A, Heimer Y, Volokita M (2001). Transcriptional and posttranscriptional regulation of the glycolate oxidase gene in tobacco seedlings. *Plant Mol. Biol.*, **45**:399–407.
- Burlat V, Oudin A, Courtois M, Rideau M, St-Pierre B (2004). Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.*, **38**(1):131-41.

Ciereszko I (2018). Regulatory roles of sugars in plant growth and development. *Acta Soc. Bot. Pol.*, **87**(2):3583.

Chaudhary N, Nijhawan A, Khurana JP, Khurana P (2010) Carotenoid biosynthesis genes in rice: structural analysis, genome-wide expression proWling and phylogenetic analysis. *Mol. Genet. Genomics*, **283**:13–33.

Chen I-C, Hill JK, Ohlemüller,R, Roy DB, Thomas CD (2011) Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, **333**:1024-1026.

Chory, J., Chatterjee, M., Cook, R. K., Elich, T, Fankhauser C, Li J, Nagpal P, Neff M, Pepper A, Poole D, Reed J, Vitart V (1996). From seed germination to flowering, light controls plant development via the pigment phytochrome. *Proc. Natl. Acad. Sci. USA*, **93**:12066–12071.

Clouse SD (2001). Integration of light and brassinosteroid signals in etiolates seedling growrh. *Trends Plant Sci.*, **6**(10):443–445.

Collakova E, DellaPenna D (2003a) Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in Arabidopsis. *Plant Physiol.*, **131**: 632–642.

Collakova E, DellaPenna D (2003b). The role of homogentisate phytyl- transferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. *Plant Physiol.*, **133**:930–940.

Cordoba E, Salmi M, León P (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.*, **60**:2933–43.

Estevez JM, Cantero A, Romero C, Kawaide H, Jimenez LF, Jiménez LF, Kuzuyama T, Seto H, Kamiya Y, León P (2000). Analysis of the expression of CLA1, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol- 4-phosphate pathway in Arabidopsis. *Plant Physiol.*, **124**:95–104.

Fankhauser C (2001). The phytochromes, a family of red/far-red absorbing photoreceptors. *J. Biol. Chem.*, **276**:11453–11456.

Fankhauser C, Chory J (1997). Light control of plant development. *Annu. Rev. Cell Dev. Biol.*, **13**:203–229.

Ghassemian M, Lutes J, Tepperman JM, Chang HS, Zhu T, Wang X, Quail PH, Lange BM (2006). Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis. *Arch. Biochem. Biophys.*, **448**:45–59.

Hemmerlin A., Hoeffler J-F, Meyer O, Tritsch D, Kagan I, Grosdemange-Billiard C, Rohmer M, Bach TJ (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.*, **278**: 26666-26676.

Hsieh MH, Goodman HM (2005). The Arabidopsis IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. *Plant Physiol.*, **138**:641–653.

Hooper JK, Eggink LL (2001). A potential role of chlorophylls b and c in assembly of light-harvesting complexes *FEBS Lett.*, **489**:1–3.

Huang R, Wang Y, Wang P, Li C, Xiao F, Chen N, Li N, Li C, Sun C, Li L, Chen R, Xu Z, Zhu J, Deng X (2018) A single nucleotide mutation of IspF gene involved in the MEP pathway for isoprenoid biosynthesis causes yellow-green leaf phenotype in rice, *Plant Mol. Biol.*, **96**(1-2):5-16.

Jiao Y, Lau OS, Deng XW (2007) Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genetics*, **8**(3):217-230.

Jumtee K, Bamba T, Okazawa A, Fukusaki E, Kobayashi A (2008). Integrated metabolite and gene expression profiling revealing phytochrome A regulation of polyamine biosynthesis of Arabidopsis thaliana. *J. Exp. Botany*, **59**(6):1187–1200.

Jung KH, Lee J, Dardick C, Seo YS, Cao P, Canlas P, Phetsom J, Xu X, Ouyang S, An K, Cho YJ, Lee GC, Lee Y, An G, Ronald PC (2008). Identification and functional analysis of light-responsive unique genes and gene family members in rice. *PLoS Genetics*, **4**(8):e1000164.

Kim B-R, Kim S-U, Chang Y-J (2005). Differential expression of three 1-deoxy-D-xylulose-5-phosphate synthase genes in rice. *Biotechnology Letters*, **27**:997–1001.

Li J, Nagpal P, Vitart V, McMorris TC, Chory J (1996). A role for brassinosteroids in light-dependent development of Arabidopsis. *Science*, **272**:398–401.

Li Y, Swaminathan K, Hudson ME (2011). Rapid, Organ-Specific Transcriptional Responses to Light Regulate Photomorphogenic Development in Dicot Seedlings. *Plant Physiol.*, **156**:2124–2140.

Li Z, Keasling JD, Niyogi KK (2012). Overlapping Photoprotective Function of Vitamin E and Carotenoids in Chlamydomonas. *Plant Physiol.*, **158**:313–323.

Lin C (2002). Blue light receptors and signal transduction. *Plant Cell*, **14**:S207–S225.

Liam EO, Matthew JP, Wingler A (2013). How Do Sugars Regulate Plant Growth and development? New insight into the role of trehalose-6-phosphate. *Mol. Plant*, **6**(2):261–274.

Lois LM, Campos N, Putra SR, Danielsen K, Rohmer M, Boronat A (1998). Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. *Proc Natl Acad Sci USA*, **95**: 2105-2110.

Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H, Deng XW. 2001. Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *The Plant Cell*, **13**:2589–2607.

Maeda H, Dudareva N (2012). The Shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.*, **63**:73-105.

Maeda H, Song W, Sage TL, DellaPenna D (2006). Tocopherols play a crucial role in low-temperature adaptation and Phloem loading in Arabidopsis. *The Plant Cell*, **18**:2710–2732.

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

Mene-Saffrane L, DellaPenna D (2010). Biosynthesis, regulation and functions of tocochromanols in plants. *Plant Physiol Biochem.*, **48**:301–309.

Michard E, Lima PT, Borges F, Silva AC, Maria Portes MT, Carvalho JE, Gilliam M, Liu L-H, Obermeyer G, Feijó JA (2011). Glutamate receptor-like genes form Ca²⁺ channels in pollen tubes and are regulated by pistil D-serine. *Science*, **332**:434-437.

Mongelard G, Seemann M, Boisson AM, Rohmer M, Bligny R, Rivasseau C (2011) Measurement of carbon flux through the MEP pathway for isoprenoid synthesis by ³¹P-NMR spectroscopy after specific inhibition of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate reductase. Effect of light and temperature. *Plant, Cell and Environment*, **34**:1241–1247.

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012). NAC transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta*, **1819**:97–103.

Nunes-Nesi A, Araújo WL, Obata T, Fernie AR (2013) Regulation of the mitochondrial tricarboxylic acid cycle. *Current Opinion in Plant Biology*, **16**:335–343.

Osanai T, Azuma M, Tanaka K (2007). Sugar catabolism regulated by light- and nitrogen-status in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photochem. Photobiol. Sci.*, **6**:508–514.

Papenbrock J, Grimm B (2001). Regulatory network of tetrapyrrole biosynthesis—Studies of intracellular signalling involved in metabolic and developmental control of plastids. *Planta*, **213**:667–681.

Popov VN, Eprintsev AT, Fedorin DN, Igamberdiev AU (2010) Succinate dehydrogenase in *Arabidopsis thaliana* is regulated by light via phytochrome A. *FEBS Lett.*, **584**:199-202.

Rawsthorne S (2002). Carbon flux and fatty acid synthesis in plants. *Progress in Lipid Research*, **41**:182–196.

Rodriguez-Concepcion M. (2006). Early steps in isoprenoid biosynthesis: multilevel regulation of the supply of common precursors in plant cells. *Phytochem. Rev.*, **5**:1–15.

Sánchez-Retuerta C, Suárez-López P, Henriques R (2018). Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs. *Front Plant Sci.*, **9**:962.

Sasaki Y, Kozaki A, Hatano M (1997) Link between light and fatty acid synthesis: Thioredoxin-linked reductive activation of plastidic acetyl-CoA carboxylase. *Proc. Natl. Acad. Sci.* **94**:11096–11101.

Satoh R, Nakashima K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2002) ACTCAT, a novel cis-acting element for proline and hypoosmolarity-responsive expression of the ProDH gene encoding proline dehydrogenase in Arabidopsis. *Plant Physiol.*, **130**:709–719.

Schelbert S, Aubry S, Burla B, Agne B, Kessler F, Krupinska K, Hörtensteiner S (2009). Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis. *The Plant Cell*, **21**:785–787.

Schuh G, Heiden AC, Hoffmann T, Kahl J, Rockel P, Rudolph J, Wildt J (1997) Emission of volatile organic compounds from sunflower and beech: dependence on temperature and light intensity. *J. Atmos. Chem.*, **27**:291–318.

Sprenger GA, Schorken U, Wiegert T, Grolle S, de Graaf AA, Taylor SV, Begley TP, Bringer-Meyer S, Sahm H (1997). Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-d-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proc. Natl. Acad. Sci. USA*, **94**:12857–12862.

Szabados L, Savoure A (2009) Proline: a multifunctional amino acid, *Trends in Plant Science*, **15**(2):89-97.

Terzaghi WB, Cashmore AR (1995). Light-regulated transcription. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **46**:445–474.

von Lintig J, Welsch R, Bonk M, Giuliano G, Batschauer A, Kleinig H (1997). Light-dependent regulation of carotenoid biosynthesis occurs at the level of phytoene synthase expression and is mediated by phytochrome in *Sinapis alba* and *Arabidopsis thaliana* seedlings. *Plant J.*, **12**(3):625-34.

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

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.

Wiberley AE, Donohue AR, Westphal MM, Sharkey TD (2009) Regulation of isoprene emission from poplar leaves throughout a day. *Plant Cell Environ.*, **32**:939–947.

Yang P, Chen H, Liang Y, Shen S (2007). Proteomic analysis of de-etiolated rice seedlings upon exposure to light. *Proteomics*, **7**:2459–2468.

Zhou F, Wang C-Y, Gutensohn M, Jiang L, Zhang P, Zhang D, Dudarevaf N, Lua S (2017). A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *PNAS*, **114**: 6866–6871.

Chapter V

Differential accumulation of pelargonidin glycosides in petals at three different developmental stages of the orange-flowered gentian (*Gentiana lutea* L. var. *aurantiaca*)

5.0. Abstract

Corolla color in *Gentiana lutea* L. exhibits a yellow/orange variation. We previously demonstrated that the orange petal color of *G. lutea* L. var. *aurantiaca* is predominantly caused by newly synthesized pelargonidin glycosides that confer a reddish hue to the yellow background color, derived from the carotenoids. However, the anthocyanin molecules of these pelargonidin glycosides are not yet fully identified and characterized. Here, I investigated the regulation, content and type of anthocyanins determining the petal coloration of the orange-flowered *G. lutea* L. var. *aurantiaca*. Eleven pelargonidin derivatives were identified in the petals of *G. lutea* L. var. *aurantiaca* for the first time, but quantitative and qualitative differences were observed at each developmental stage. The highest levels of these pelargonidin derivatives were reached at the fully open flower stage (S5) where all anthocyanins were detected. In contrast, not all the anthocyanins were detected at the budlet stage (S1) and mature bud stage (S3) and those corresponded to more complex pelargonidin derivatives. The major pelargonidin derivatives found at all the stages were pelargonidin 3-*O*-glucoside, pelargonidin 3,5-*O*-diglucoside and pelargonidin 3-*O*-rutinoside. Furthermore, the expression of DFR (dihydroflavonol4-reductase), ANS (anthocyanidin synthase), 3GT (UDP-glucose:flavonoid3-*O*-glucosyltransferase), 5GT (UDP-glucose:flavonoid5-*O*-glucosyltransferase) and 5AT (anthocyanin 5- aromatic acyltransferase) genes was analyzed in the petals of three developmental stages, showing that the expression level of DFR, ANS and 3GT parallels the accumulation of the pelargonidin glucosides. Overall, this study enhances the knowledge of the biochemical basis of flower coloration in *Gentiana* species, and lays a foundation for breeding of flower color and genetic variation studies in *Gentiana* varieties.

5.1. Introduction

Polymorphism in flower color has been associated with the preferences of pollinators for certain colors, primarily due to the ability of pollinators to perceive and distinguish among different hues (Harborne, 1962; Fenster *et al.*, 2004). The final color of a flower is determined by the presence of different pigments, mainly carotenoids, flavonoids or betalains (Tanaka *et al.*, 2008). Anthocyanins are water-soluble pigments derived from the flavonoid branch of the phenylpropanoid pathway (Tanaka *et al.*, 2008; Shi and Xie, 2014; Zhang *et al.*, 2014). The genes encoding for enzymes of the anthocyanin pathway are grouped in two classes (Fig. 5.1): early biosynthetic genes, which are common to other flavonoids, encoding enzymes including chalcone synthase (CHS) which is responsible for the formation of naringenin chalcone from 4-coumaroyl CoA and malonyl CoA substrates, chalcone isomerase (CHI) which generates flavones, flavanone 3-hydroxylase (F3H) which produces dihydrokaempferol, flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) which generate dihydroquercetin and dihydromyricetin respectively; and late biosynthetic genes, specific to the anthocyanin pathway, encoding enzymes including dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS) (Tanaka *et al.*, 2008; Shi and Xie, 2014; Zhang *et al.*, 2014). Differences in the hydroxylation of the B-ring of the anthocyanin skeleton confer specific variants in the color range of anthocyanidins: pelargonidins, being orange to red; cyanidins, red to red-purple (by the activity of F3'H); and delphinidins, being red-purple to blue (by the activity of F3'5'H), depending on many other factors (Holton, 1995; Tanaka *et al.*, 2008; Zhang *et al.*, 2014). The resulting anthocyanidin structures are inherently unstable. Therefore, these pigments accumulate exclusively as glycosylated forms, where C3 is linked through oxygen to a sugar residue, most frequently glucose. UDP-glucose: flavonoid 3-O-glucosyltransferase (3GT) catalyzes the 3-O-glucosylation of anthocyanidins as well as other flavonoids (Tanaka *et al.*, 2008). The generated glycosylated anthocyanins at the 3-position serve as substrates for further modifications such as glycosylation at the 5-position (Fig 5.1), by the action of UDP-glucose: flavonoid 5-O-glucosyltransferase (5GT) (Kamsteeg *et al.*, 1978; Nakatsuka *et al.*, 2008). The glycosylation reactions provide sugar residues for decoration with acyl groups by the activity of anthocyanin acyltransferase, which can incorporate aromatic acids such as p-coumaric, caffeic, ferulic, sinapic, gallic or p-hydroxybenzoic acids, and/or aliphatic acids including malonic, malic, acetic, succinic

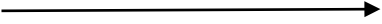
or oxalic acids. These acyl substituents are commonly bound to the C3 sugar, esterified to the 6-OH or much less frequently to the 4-OH group of the sugars (Giusti and Wrolstad, 2003). Other modifications also occur, such as acylation of the glucose moiety at the 5-position of anthocyanin catalyzed by anthocyanin 5-aromatic acyltransferase (5AT) (Fujiwara *et al.*, 1997). Both glycosyltransferases and acyltransferases contribute to the extensive range of natural anthocyanins identified to date, which differ in their side chain decorations, affecting color and increasing pigment stability (Shi and Xie, 2014). The genus *Gentiana* includes over 400 species (Kohlein, 1991), with a huge range of flower color including orange, pink, red, magenta, purple, and blue to blue-black. This color variation is mainly due to the accumulation of anthocyanins and/or carotenoids. Most gentian species of commercial interest in the flower market include those with an intense blue coloration, due to the accumulation of major blue gentiodelphin, a polyacylated delphinidin-type anthocyanin, and minor red gentiocyanin, a cyanidin-derived anthocyanin in petals of *G. triflora* (Fukuchi-Mizutani *et al.*, 2003; Yoshida *et al.*, 2009) and those with pink coloration, associated to the exclusive accumulation of red gentiocyanins in petals of *G. scabra* (Hosokawa *et al.*, 1995a, b; Nakatsuka *et al.*, 2005). Both *G. triflora* and *G. scabra* do not accumulate carotenoids in the petals (Nakatsuka *et al.*, 2006). *G. leucomelaena* manifests dramatic flower color polymorphism, with both blue- and white-flowered individuals both within a population and on an individual plant (Mu *et al.*, 2017). Flower color of *G. lutea* varies among populations, ranging from yellow, orange to almost red shades (Renobales, 2003). The yellow petals of *G. lutea* L. var. *lutea* are due to the exclusive accumulation of abundant amounts of β -carotene and xanthophylls, while the orange color of the flowers of *G. lutea* L. var. *aurantiaca* (M. Laínz) is caused by newly synthesized pelargonidin glycosides that confer a reddish hue to the yellow background color, derived from the carotenoids (Berman *et al.*, 2016). However, the anthocyanin molecules of these pelargonidin glycosides in petals of orange-flowered *G. lutea* L. var. *aurantiaca* are not yet fully identified and characterized. To date, the most used methods for separation and quantitation of anthocyanins are high performance liquid chromatography (HPLC) with UV-Vis or diode array detectors (DAD) (Harborne and Sherratt, 1957; Harborne, 1962; Harborne and Williams, 2001; Hong and Wrolstad, 1990). However, reverse-phase HPLC coupled to mass spectrometry (MS) offers the most powerful approach for identification of individual anthocyanins (Wu and Prior, 2005; Sun *et al.*, 2012). In this study, we identified and characterized the pelargonidin

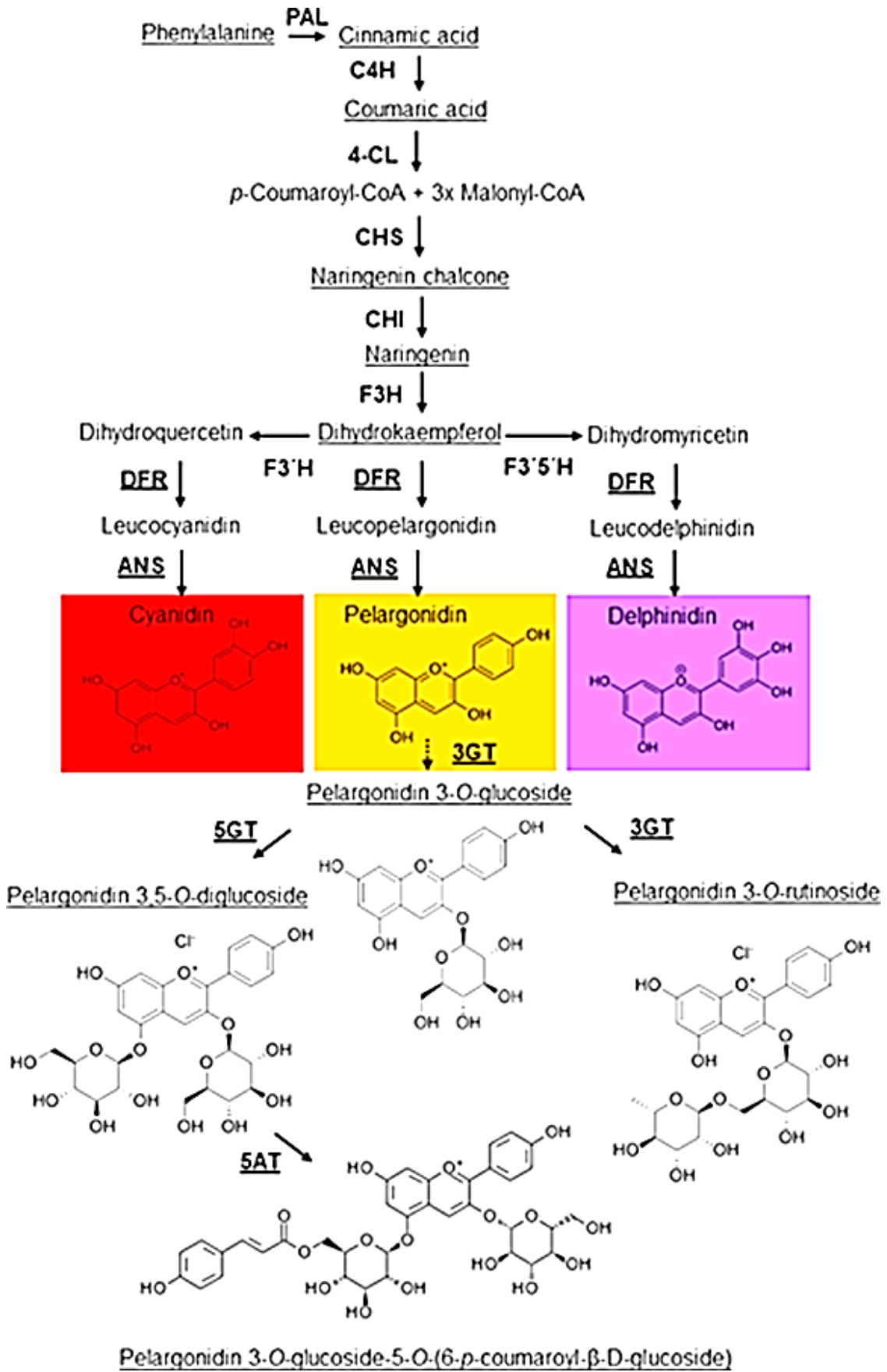
derivatives present in the petals of *G. lutea* L. var. *aurantiaca*, and their evolution during the development of the petals. We also isolated the key late anthocyanin biosynthetic gene (*5GT* and *5AT*) fragments (cDNAs), and analyzed the expression levels of DFR, ANS, 3GT, 5GT and 5AT to determine the molecular mechanism responsible for the biosynthesis and accumulation of the pelargonidin glycosides in petal of *G. lutea* L. var. *aurantiaca* during flower development.

5.2. Aims

The aims of the experiments described in this chapter were to identify and characterize the anthocyanin molecules of these pelargonidin glycosides and investigate the regulation, content and type of anthocyanins determining the petal coloration of the orange-flowered *G. lutea* L. var. *aurantiaca*. Characterize and quantify anthocyanins from the petals of *G. lutea* by HPLC-ESI-MS/MS

Figure 5.1. The gentian phenylpropanoid pathway. Anthocyanidins (pelargonidin, cyanidin, and delphinidin) are modified by the addition of sugars and other moieties to form anthocyanins in a species-specific manner. Pink-flowered gentians contain gentiocyannin in their petals, and blue-flowered gentians mainly gentiodelphin, whereas orange-flowered gentian accumulates exclusively pelargonidin glycosides. Abbreviations: ANS, anthocyanidin synthase; 5AT, anthocyanin 5-aromatic acyltransferase; CHI, chalcone isomerase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; DFR, dihydroflavonol 4-reductase; DHK, Dihydrokaempferol; DHM, Dihydromyricetin; DHQ, Dihydroquercetin; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; 3GT, UDP-glucose:flavonoid 3-O-glucosyltransferase; 5GT, UDP-glucose:flavonoid 5-O-glucosyltransferase; PAL, phenylalanine ammonia-lyase. Underlined names indicate enzymes/metabolites whose encoding gene expression/accumulation.





5.3 Material and Methods

5.3.1. Chemicals

All the chemicals and solvents were liquid chromatography–mass spectrometry (LC/MS) grade (LiChrosolv, Merck). Water was reverse osmosis Milli-Q water (18.2 MO) (Millipore, USA). The standard pelargonidin was purchased from Sigma (St. Louis, MO, USA). Pelargonidin 3-*O*-glucoside, pelargonidin 3,5-*O*-diglucoside, naringenin, cinnamic acid, coumaric acid, and dihydrokaempferol were purchased from TransMIT (Marburg, Germany). Formononetin was purchased from Indofine Chemicals (indofinechemical.com).

5.3.2. Plant material

Gentiana lutea L. var. *aurantiaca* flowers were collected from Torrestio, Leon (Spain) in the Eurosiberian phytogeographic region (43° 03' N; 6° 04' W; 1600 m above sea level). All samples were collected in August of 2013. After harvest, all samples were frozen in liquid nitrogen and stored at -80°C until use. Petal tissues were collected at three stages of floral development, freeze-dried and stored at -80°C prior to pigment analysis. Each developmental stage was defined in terms of flower bud length as previously described (Berman *et al.*, 2016): stage 1 (S1), budlet stage, petals <1.5 cm long with hard bud; stage 3 (S3), mature bud stage, petals 2.5–3.5 cm long with petals slightly loosened; and stage 5 (S5), fully open flower stage.

5.3.3. Anthocyanin and precursor extraction and identification

All extractions were performed in triplicate. Freeze-dried petals were ground into a powder and homogeneously ground using 500µL of 85:15 methanol:1 N HCl, and left in the dark at 4°C for 24h. The extracted liquid was filtered with 0.22µm membrane filter and used for HPLC-DAD analyses as described previously (Berman *et al.*, 2016). Anthocyanins were extracted from 10 mg of homogeneously ground petal tissue, using 500µL of 85:15 methanol:1 N HCl with 2µg L⁻¹ of formononetin (Sigma-Aldrich, USA) as internal standard for quantitative analysis. After shaking for 12h and centrifugation for 10min at 15,300 g at 25°C, 0.25mL of supernatant was transferred to

new Eppendorf vials and dried. Samples were then resuspended in 0.1mL 75% methanol (plus 0.1% formic acid) and centrifuged (10min at 15,300g at 25°C). Finally, 70µL of supernatant was transferred to HPLC vials for MS (mass spectrometry) analysis, and 5µL of extracts were injected to the HPLC-DAD/MS. HPLC analysis was performed using a C18 Luna column (Phenomenex, Aschaffenburg, Germany) (150 × 2.0mm; 3µm). 10µL of each extract were injected at a flow of 0.2mL min⁻¹. The mobile phases were: 0.1% formic acid (A) in water and acetonitrile plus 0.1% formic acid (B). The method used for separation was: 95% A: 5% B for 1min, and a linear gradient to 25% A: 75% B over 40min, plus an additional 2min, before returning to the initial conditions for 18min. Detection was performed continuously from 200 to 600nm with an online Accela Surveyor photodiode array detector (PDA; Thermo Fisher Scientific). Mass spectrometry analysis was performed by Diretto (Italian National Agency for New Technologies) using a quadrupole-Orbitrap Q-exactive system (ThermoFisher scientific, USA), as previously described (Rambla *et al.*, 2016) with slight modifications. Metabolite ionization was carried out with a heated electrospray ionization (HESI) source operating in positive ion mode, and nitrogen was used as sheath and auxiliary gas (45 and 15 units, respectively). Mass spectrometer parameters were as follows: capillary and vaporizer temperatures 30°C and 270°C, respectively, discharge current 4.0 KV, probe heater temperature at 370°C, S-lens RF level at 50 V. The acquisition was carried out in the 110/1600 m/z scan range, with the following parameters: resolution 70,000, microscan 1, AGC target 1e6, and maximum injection time 50. A first full scan MS with data-dependent MS/MS fragmentation was used to identify the anthocyanins in the *G. lutea* L. var *aurantiaca* flower extracts. Subsequently, a single ion monitoring (SIM) with targeted MS/MS fragmentation was used to identify anthocyanins for which Data-dependent MS/MS fragmentation was not successful, and to further validate the tentative identifications. Data were analyzed using the Xcalibur 3.1 software (ThermoFisher scientific, USA). Metabolites were identified as M⁺ adducts, based on their accurate masses (m/z) and MS fragmentation, using both in house database and public sources (e.g. KEGG, MetaCyc, ChemSpider, PubChem, Metlin, Phenol-Explorer). Relative abundances of the metabolites studied were calculated using the Xcalibur 3.1 software (ThermoFisher scientific, USA). The two most abundant fragments per compound were used to determine the relative abundances, and data were normalized through dividing each peak area by the value of

the internal standard peak area (Maloney *et al.*, 2014). Data are presented as means and standard deviation of at least three independent biological replicates.

For precursor analyses 15mg of freeze-dried, homogenized *G. lutea* L. var. *aurantiaca* petals were extracted with 0.75mL cold 75% (v/v) methanol, with 0.1% (v/v) formic acid and spiked with 5µg mL⁻¹ formononetin. Samples were shaken for 40min at 20Hz using a Mixer Mill 300 (Qiagen), before centrifugation for 15min at 15,300 g. Finally, 0.6mL of supernatants was placed in the HPLC tubes. HPLC-DAD/MS analyses were carried out also by Diretto using a Q-exactive quadrupole Orbitrap mass spectrometer (ThermoFisher Scientific), operating in positive/negative heated electrospray ionization (HESI), and coupled to an Ultimate HPLC-DAD system (Thermo Fisher Scientific, Waltham, MA). HPLC conditions were as in the section 4.4.ESI-MS ionization was carried out as follows: sheath and aux gas flow rate set at 40 and 25 units, respectively; vaporizer and capillary temperature were used at 250 and 30°C, discharge current was set at 4.5µA and Slens RF level at 50. Acquisition was achieved as previously reported. Metabolite identification was carried out using a single ion monitoring (SIM) scan mode, with targeted MS/MS fragmentation, and by comparing chromatographic and spectral properties with authentic standards. Data were analyzed using Xcalibur 3.1 software (ThermoFisher scientific, USA). Metabolites were quantified in a relative way by normalization on the internal standard amounts. Data are presented as means and standard deviation of at least 3 independent biological replicates.

5.3.4. Cloning of UDP-glucose: Flavonoid 5-O-glucosyltransferase (5GT) and anthocyanin 5-aromatic acyltransferase (5AT) gene fragments from petals of *G. lutea* L. var. *aurantiaca*

Three grams of leaf, petals at S1, S3 and S5 stages, respectively, or mixed petals from S1, S3 and S5 stages (1g from each stage) were ground into a fine powder using a mortar and pestle in liquid nitrogen. From these samples, 100mg samples were used to isolate total RNA for each extraction. Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) and DNA was removed with DNase I (RNase-free DNase Set, Qiagen). The integrity of RNA was confirmed by gel electrophoresis in a denaturing 1.2% (w/v) agarose gel containing formaldehyde. Total

RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Vernon Hills, Illinois, USA), and 1µg total RNA was used as template for first strand cDNA synthesis with QuantiTect Reverse Transcription Kit (Qiagen) in a 20µL total reaction volume, following the manufacturer's recommendations. A 2µL aliquot of the first strand cDNA product with isolated total RNAs from mixed petals of S1, S3 and S5 stages of *G. lutea* L. var. *aurantiaca* was used in a 50 µL RT-PCR reaction containing 1x Green GoTaq Reaction Buffer (Promega, Madison, WI, USA), 0.2 mM of each dNTP, 0.5 µM of each primer, and 1.25 U GoTaq DNA Polymerase (Promega, Madison, WI, USA). Reverse transcription polymerase chain reaction (RT-PCR) was performed using primer sequences designed on *5GT* and *5AT* sequences from *G. triflora*, to amplify partial protein coding sequences of cDNAs for *5GT* and *5AT* from *G. lutea* L. var. *aurantiaca*. The primer sequences for the *5GT* were 5' CTGCTTCCTTGGGCTGCCGATGTGGCT-3' (forward primer) and 5'-TGCGGATGGAATTCAACGCTGGAGAGC-3' (reverse primer). The primer sequences for the *5AT* were 5'-CCGCTCGTAGCCGTGCAAGTAACCGTTTTTTCCT-3' (forward primer) and 5'-GATTTTGGATGGGGAAAGCCTGCAAATTTGA-3' (reverse primer). The PCR conditions for *5GT* and *5AT* genes involved heating the reaction to 95°C for 3 min, followed by 35 cycles at 95°C for 45 s, 60°C for 45 s and 72°C for 90 s, and a final extension step at 72°C for 10min. The resulting expected size products were purified from a 1.0% w/v agarose gel using the GeneClean II Kit (BIO101 Systems, Solon, OH, USA) and cloned in pGEM-T easy vector (Promega, Madison, WI, USA) for sequencing with the Big Dye Terminator v3.1 Cycle Sequencing Kit on a 3130x1 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Four and six independent clones provided the identical sequences for isolated *5GT* and *5AT* gene fragments, respectively, from 20 and 26 independent sequenced clones. The nucleotide sequences were translated to the respective deduced amino acid sequences using

Vector NTI software, which was also used for alignments of cDNA and deduced amino acid sequences, respectively. Both cloned *5GT* and *5AT* showed sequence similarity to previously characterized anthocyanin structural genes and BLAST searches against the National Center for Biotechnology Information (NCBI) nucleotide and protein databases were used to confirm homology with previously characterized *5GT* and *5AT* genes from higher plants (Fujiwara *et al.*, 1997).

5.3.5. qRT-PCR analysis of *DFR*, *ANS*, *3GT*, *5GT* and *5AT* expression in leaf, S1, S3 and S5 petals of *G. lutea* L. var. *aurantiaca*

From the above *5GT* and *5AT* partial cDNA sequences, gene-specific primers were designed for these two genes, in order to perform expression analysis by quantitative real-time RT-PCR (qRT-PCR). The primers for the *5GT* were 5'-GCCTTGCAAGCAATCCCAA-3' (forward primer) and 5'-CAACCATCCATGGCAGTCCT-3' (reverse primer). The primers for the *5AT* were 5'-AGCTGTTGGGGATGCCATTG-3' (forward primer) and 5'-CGATCCACTAATCCCGAGCAA-3' (reverse primer). Gene-specific oligonucleotides for *DFR*, *ANS*, *3GT* and *UBQ* (ubiquitin gene) were the same as reported previously (Berman *et al.*, 2016). Specific primers that amplified a single fragment of no more than 200 bp were designed in regions of the coding sequences isolated from orange-flowered *G. lutea* L. var. *aurantiaca* petals. qRT-PCRs were carried out using a BioRad CFX96 system. Each 25 μ L reaction volume comprised 5ng of cDNA, 1x iQ SYBR green supermix (BioRad, Hercules, CA, USA), and 5 μ M of forward and reverse primers. Relative expression levels were calculated on the basis of serial dilutions of cDNA (100–0.16ng), which were used to generate standard curves for each gene. Triplicate amplifications (biological replicates) were performed in 96-well optical reaction plates by first heating to 95°C for 5min, followed by 40 cycles at 95°C for 30s, 58°C for 30s and 72°C for 30s. Amplification specificity was confirmed by product melt curve analysis over the temperature range 50–90°C with fluorescence acquired after every 0.5°C increase. The fluorescence threshold value and gene expression data were calculated with BioRad CFX96 software. Data were calculated from three biological replicates with at least three technical replicates for each biological replicate and with error bars representing the standard deviation. Amplification efficiencies were compared by plotting the Δ Ct values of different primer combinations of serial dilutions against the log of starting template concentrations using the CFX96 software. The Ct values were adjusted to the standard curves and normalized against the levels of ubiquitin (*UBQ*) housekeeping gene (Berman *et al.*, 2016).

5.3.6. Bioinformatics analyses

Correlation analyses of transcript and metabolite data were performed as previously described (Ahrazem *et al.*, 2018; Gomez-Gomez *et al.*, 2017). Briefly, Pearson correlation coefficients for each data pair (gene-gene, metabolite-metabolite, gene-metabolite) were compiled in Excel and subjected to significance analysis by SPSS software; only correlations showing correlations with P values ≤ 0.05 were further processed. Subsequently, correlation a matrix was generated using R version 2.10.0, and colored with the Morpheus software (<https://software.broadinstitute.org/morpheus/>). Correlation network analysis was performed using R to generate a network correlation input file, which was finally imported in Cytoscape version 2.6.2 (<http://www.cytoscape.org/>), and visualize using a Prefuse force-directed layout. Node strength (ns) and network strength (NS) were calculated as reported previously (Diretto *et al.*, 2010; Sulli *et al.*, 2017), and represent the average of all the ps (Pearson correlation coefficients) yielded by each node, and the average of all the node strengths in the network, respectively.

5.4. Results

5.4.1 Identification of anthocyanins in petals of three developmental stages of *G. lutea* L. var. *aurantiaca*

HPLC reverse phase C18 columns have been extensively used for the separation of anthocyanins in different plant species (Kim *et al.*, 2012; Koshioka *et al.*, 2015; Shim *et al.*, 2015). The acidic methanol extracts of the petals of *G. lutea* L. var. *aurantiaca* showed mixtures of several peaks (Fig.5.2). The anthocyanins corresponding to the peaks (500nm) observed in the HPLC-DAD chromatogram were characterized and quantified by HPLC-ESI-MS/MS analyses.

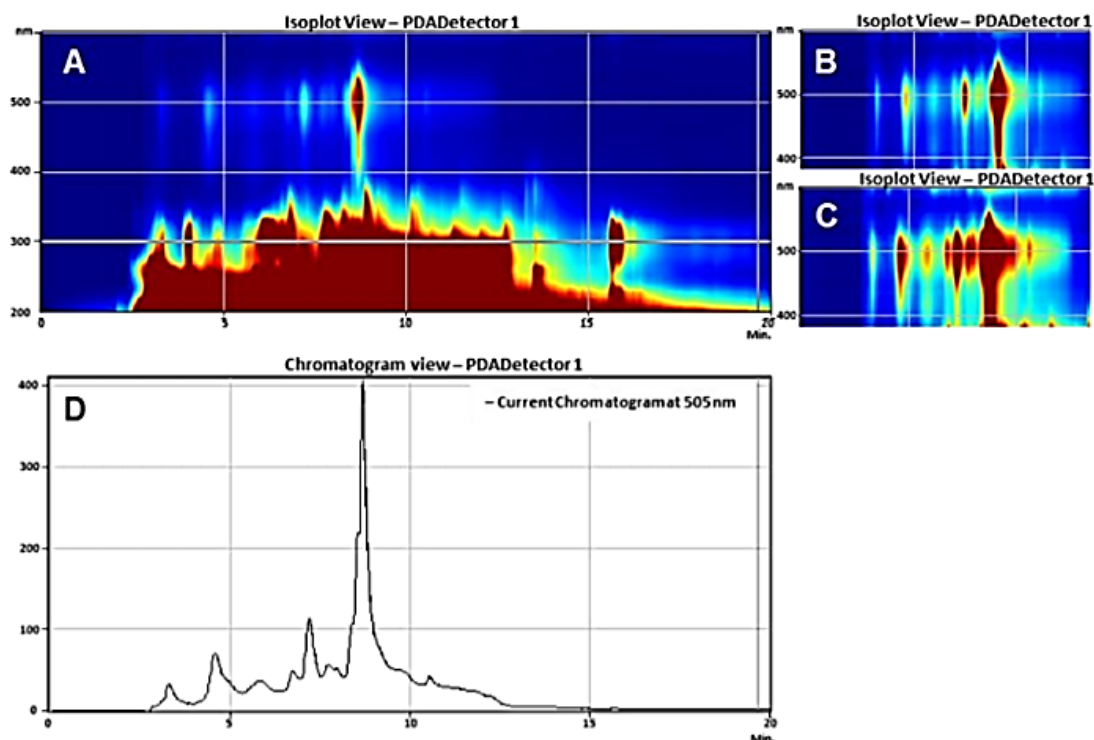


Figure 5.2. Typical HPLC-DAD/UV isoplot chromatograms and HPLC-PDA chromatographic profile of the anthocyanins in *Gentiana lutea* L. var. *aurantiaca* petals at the S5 stage of development. A) HPLC-DAD/UV isoplot from 200–600 nm. Anthocyanins are detected in the 500 nm range. B) and C) correspond to magnifications in the 500 nm range to show up the anthocyanins present in lower amounts. D) HPLCDAD chromatogram at 505 nm, showing the profile of the anthocyanins present in the S5 extract.

All the detected peaks showed a maximum absorbance between 495–505 nm (Fig. 5.2), characteristic of pelargonidin derivatives (Yildirim *et al.*, 2016). The glycosidic substitution pattern of anthocyanins can be inferred by the presence of absorption in the 400–460 nm region, and the sugar substitution in the 3-OH is deduced by the ratio of E_{440}/E_{max} about two times greater than those for sugar substitution in both 3- and 5-OH positions (Harborne, 1958). The E_{440}/E_{max} ratios obtained for peak 2 was 25% while peak 4 showed a E_{440}/E_{max} of 52%, providing evidence that pigment 2 was a 3,5-glycoside and the latter was a 3-glycoside. Identification and peak assignment of anthocyanins was based on the analysis of the mass fragmentation pattern, followed by the comparison of their retention times and mass spectral data with those of standards, if available, and literature data. All the glycosides were tentatively identified using the molecular weight of the aglycone pelargonidin, m/z 271.06008 (Fig 5.3). In total, eleven major pelargonidin derivatives were identified in the petals at different developmental stages (Table 5.1).

Table 5.1. Identification of anthocyanins from *Gentiana lutea* L. var. *aurantiaca* petals by HPLC-ESI-MS/MS.

Peak number	Rt (min)	UV-maxima (nm)	Experimental [M] ⁺ (m/z)	Experimental MS/MS (m/z)	Formula	Assignment	Petal Stage
1	6.88	274, 426, 498	741.22180	579.16992/433.11237/ 271.05966	C ₃₃ H ₄₁ O ₁₉	Pelargonidin 3-O-rutinoside-5-O-β-D-glucoside	S3, S5
2	8.61	282, 497	595.16516	433.11273/271.05972	C ₂₇ H ₃₁ O ₁₅	Pelargonidin 3,5-O-diglucoside	S1, S3, S5
3	11.15	498	579.17078	433.11307/271.05981/ 146.05790	C ₂₇ H ₃₁ O ₁₄	Pelargonidin 3-O-rutinoside	S1, S3, S5
4	12.53	498	433.11203	271.05969	C ₂₁ H ₂₁ O ₁₀	Pelargonidin 3-O-glucoside	S1, S3, S5
5	12.73	282, 496	1094.29358	932.23883/771.21368/ 433.11273/271.06000	C ₅₂ H ₅₅ O ₂₆	Pelargonidin 3-O-[2-O-(6-(E)-feruloyl-β-D-glucopyranosyl)-6-O-(E)-caffeoyl-β-D-glucopyranoside]-5-O-(β-D-glucopyranoside)	S3, S5
6	13.07	447, 496	1079.30284	932.23871/771.21368/ 433.11273/271.05972	C ₅₂ H ₅₅ O ₂₅	Pelargonidin 3-O-[2-O-(6-(E)-feruloyl-β-D-glucopyranosyl)-6-O-(E)-p-coumaroyl-β-D-glucopyranoside]-5-O-(β-D-glucopyranoside)	S3, S5
7	13.87	275, 366, 500	843.22095	681.16675/595.16333/ 519.11456/433.11392/ 271.06033	C ₃₉ H ₃₉ O ₂₁	Pelargonidin 3-O-(6-O-caffeoyl-D-glucoside)-5-O-(6-O-malonyl-β-D-glucoside)	S3, S5
8	14.28	492	827.21936	579.17004/519.11230/ 271.05972	C ₃₆ H ₄₃ O ₂₂	Pelargonidin 3-O-(6-p-coumaroyl-D-glucoside)-5-(4-O-malonyl-β-D-glucoside)	S3, S5
9	14.94	447, 496	681.16565	519.11298/433.11243/ 271.05963	C ₃₀ H ₃₃ O ₁₈	Pelargonidin 3-O-(6-O-malonyl-β-D-glucoside)-5-β-D-glucoside	S3, S5
10	15.22	282, 496	517.09755	433.11307/271.06073	C ₂₄ H ₂₁ O ₁₃	Pelargonidin 3-O-(6-O-malonyl-β-D-glucoside)	S1, S3, S5
11	15.66	447, 496	579.15012	433.11307/271.05957	C ₃₀ H ₂₇ O ₁₂	Pelargonidin 3-O-(6-p-coumaroyl)glucoside	S1, S3, S5

The anthocyanins identified were mainly 3-O-glycoside conjugates and their derivatives. Peak 1 had m/z of 741.22180 and three mass fragment ions at m/z 579.16992, due to the elimination of one hexose molecule; at m/z 433.11237, indicating the loss of one molecule of rhamnose, and 271.05966, corresponding to pelargonidin (Fig 5.3G).

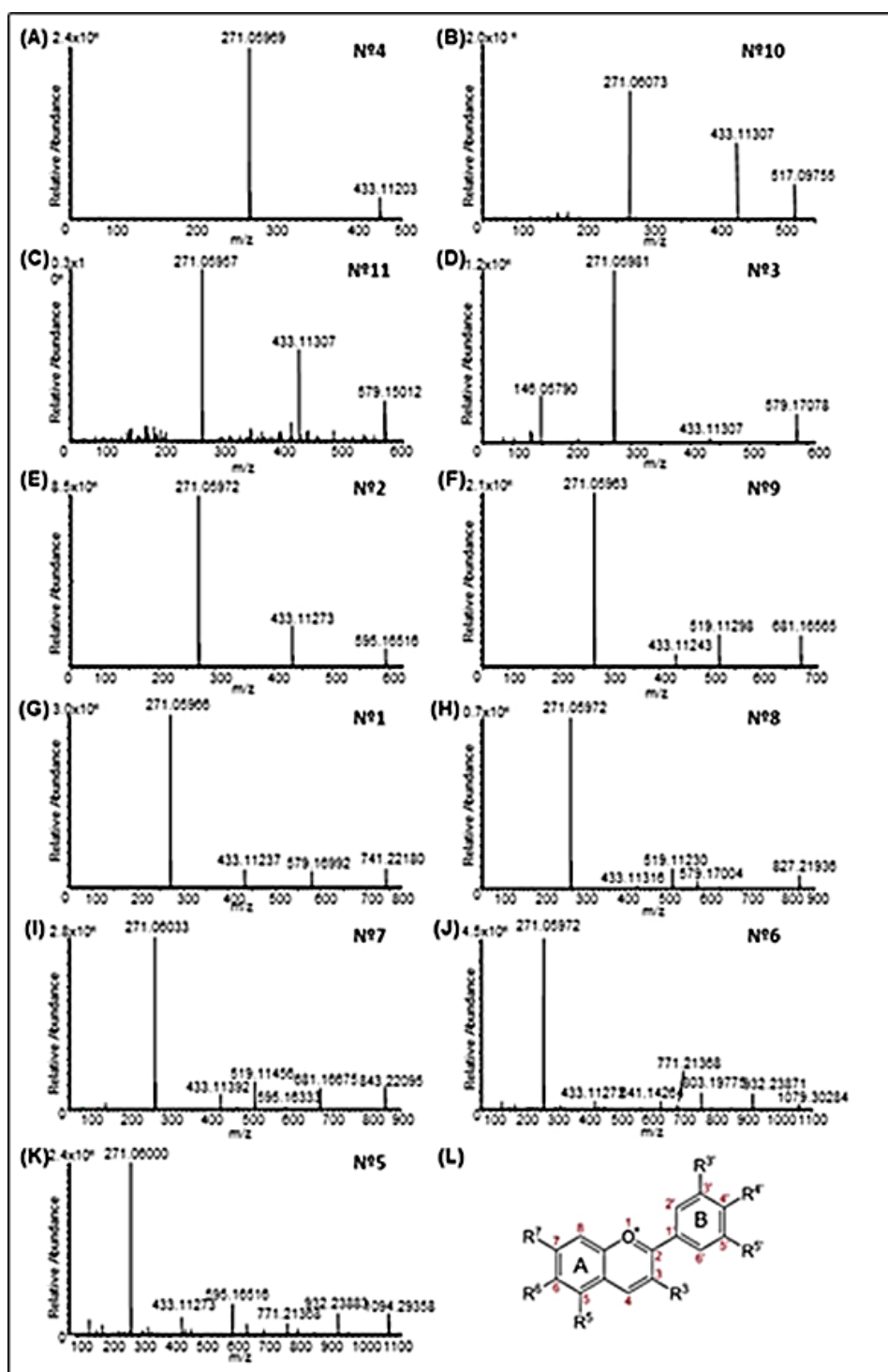


Figure 5.3. ESI-MS/MS spectra of anthocyanins isolated from petals of *Gentiana lutea* L. var. *aurantiaca*. The number from 1 to 11 corresponded with the anthocyanins characterized and listed in Table 5.1. The image in L corresponded to the basic structure of the anthocyanin molecule.

Therefore, peak 1 was assigned to pelargonidin 3-*O*-rutinoside-5-*O*- β -D-glucoside (C₃₃H₄₁O₁₉) (Rodriguez-Saona *et al.*, 1998). Peak 2 (Fig 5.3E) showed a molecular

ion at m/z 595.16516, and two fragment ions of 433.11273 and 271.05972, respectively. Peak 2 was identified as pelargonidin 3,5-*O*-diglucoside (C₂₇H₃₁O₁₅) (Barnes and Schug, 2011). Peak 3 had a molecular ion at m/z 579.17078, and three mass fragment ions, one at m/z 433.11307, due to the elimination of one molecule of rhamnose; one at m/z 271.05981, indicating the loss of rutinoside; and one at 146.05790 for rhamnose. Thus, peak 3 was assigned to pelargonidin 3-*O*-rutinoside (C₂₇H₃₁O₁₄; Fig 5.3D). Peak 4 displayed at m/z 433.11203 and a main mass fragment ion at m/z 271.05969, which corresponded to the pelargonidin molecule, due to the elimination of one molecule of glucose (Fig 5.3A). Therefore, the peak was identified as pelargonidin 3-*O*-glucoside (C₂₁H₂₁O₁₀; Fig 5.3A). Peak 5 was characterized by a molecular ion m/z 1094.29358, and four fragment ions: m/z 932.23883, which indicated loss of one hexose; m/z 771.21368, indicative of the loss of caffeoylglucoside; m/z 433.11273, corresponding to pelargonidin and one glucose molecule; and m/z 271.06000, the pelargonidin aglycone. For this reason, peak 5 was assigned as pelargonidin 3-*O*-[2-*O*-(6-(*E*)-feruloyl-β-D-glucopyranosyl)-6-*O*-(*E*)-caffeoyl-β-D-glucopyranoside]-5-*O*-(β-D-glucopyranoside) (C₅₂H₅₅O₂₆; Fig 5.3K). Peak 6 had a molecular ion m/z 1079.30284, and four fragment ions: m/z 932.23871 and m/z 771.21368 indicated the loss of coumaroyl glucoside; m/z 771.21368 and m/z 433.11273 indicated the loss of feruloyl glucoside, which resulted in pelargonidin glucoside; and m/z 271.05972, corresponding to pelargonidin (Fig 5.3J). Thus, peak 6 was tentatively assigned to pelargonidin 3-*O*-[2-*O*-(6-(*E*)-feruloyl-β-D-glucopyranosyl)-6-*O*-(*E*)-*p*-coumaroyl-β-D-glucopyranoside]-5-*O*-(β-D-glucopyranoside) (C₅₂H₅₅O₂₅). Peak 7 showed a m/z 843.22095 and five mass fragment ions at m/z 681.16675, due to the loss of a caffeic acid; m/z 595.16333 [M-248.05321 (malonylhexose)]; m/z 519.11456 [M-324.08451 (caffeoylhexose)]; m/z 433.11392, corresponding to pelargonidin with one glucose; and, finally, m/z 271.06033, the pelargonidin aglycone. This peak (Fig 5.3I) was tentatively assigned to pelargonidin 3-*O*-(6-*O*-caffeoyl-D-glucoside)-5-*O*-(6-*O*-malonyl-β-D-glucoside) (C₃₉H₃₉O₂₁) (Hosokawa *et al.*, 1995). Peak 8 displayed m/z 827.21936, and three mass fragment ions at 579.17004 and 519.11230, which corresponded to the loss of malonylglucoside and coumaryl glucoside, respectively; and 271.05972, which represented pelargonidin. Peak 8 was assigned to pelargonidin 3-*O*-(6-*p*-coumaroyl-D-glucoside)-5-(4-*O*-malonyl-β-D-glucoside) (C₃₆H₄₃O₂₂; Fig 5.3H). Peak 9 was identified as pelargonidin 3-*O*-(6-*O*-malonyl-β-D-glucoside)-5-β-D-glucoside (C₃₀H₃₃O₁₈) on the basis of full MS at m/z 681.16565 which, upon the loss of a glucose molecule, showed a fragment of

m/z 519.11298; and the loss of a malonyl glucoside yielded a fragment at m/z 433.11243, corresponding to the pelargonidin unit and a glucose molecule (Fig 5.3F). Peak 10 and peak 11, finally, showed a molecular ions at m/z of 517.09755 and 579.15012, respectively, and were tentatively assigned to pelargonidin 3-*O*-(6-*O*-malonyl- β -D-glucoside) (C₂₄H₂₁O₁₃; Fig 5.3B) and pelargonidin 3-*O*-(6-*p*-coumaroyl)glucoside (C₃₀H₂₇O₁₂; Fig 5.3C), respectively.

5.4.2. Quantification of pelargonidin derivatives and precursors in petals at three developmental stages

The anthocyanins determined by HPLC-ESI-MS/MS analyses in the petals of *G. lutea* L. var. *aurantiaca* at three developmental stages showed strong variation in abundance (Fig 5.4). Pelargonidin 3-*O*-glucoside and pelargonidin 3,5-*O*-diglucoside were found to be, by far, the predominant anthocyanins in all the samples analyzed. Interestingly, none of the pelargonidin glycosides were detectable in leaves of *G. lutea* L. var. *aurantiaca* (Fig 5.4). Petals at stage one (S1) were green, and in this stage pelargonidin 3-*O*-glucoside (62.5%) and pelargonidin 3,5-*O*-diglucoside (29.1%) were the main anthocyanins detected (Fig 5.4 and Fig 5.5), followed by pelargonidin 3-*O*-(6-*O*-malonyl- β -D-glucoside) (6.8%). Other pelargonidin derivatives identified, albeit at very low amounts, were pelargonidin 3-*O*-rutinoside and pelargonidin 3-*O*-(6-*p*-coumaroyl)glucoside. Stage 3 (S3) displayed the presence of new pelargonidin derivatives not detected at S1, with four main derivatives: pelargonidin 3,5-*O*-diglucoside (33.4%), pelargonidin 3-*O*-glucoside (32.5%), pelargonidin-3-*O*-(6-*p*-coumaryl-D-glucoside)-5-(4-*O*-malonyl- β -D-glucoside) (10.3%) and pelargonidin 3-*O*-rutinoside (9.6%) (Fig 5.4 and Fig 5.5). Stage 5 (S5), corresponding to the flower at anthesis, showed the highest anthocyanin content. In this tissue, the major derivatives of pelargonidin corresponded to pelargonidin 3-*O*-glucoside (34.5%), followed by pelargonidin 3,5-*O*-diglucoside (25.1%), pelargonidin 3-*O*-rutinoside (19.10%) and pelargonidin 3-*O*-(6-*O*-malonyl- β -D-glucoside) (5.5%) (Fig 5.4). The formation of pelargonidin 3-*O*-glucoside, pelargonidin 5-*O*-glucoside, pelargonidin 3,5-*O*-diglucoside and pelargonidin 3-*O*-rutinoside are the most likely first steps of glycosylation in gentian petals (Fig 5.1), and are further modified by acylation reactions, leading to more diverse and complex pelargonidin molecules (Sasaki *et al.*, 2014).

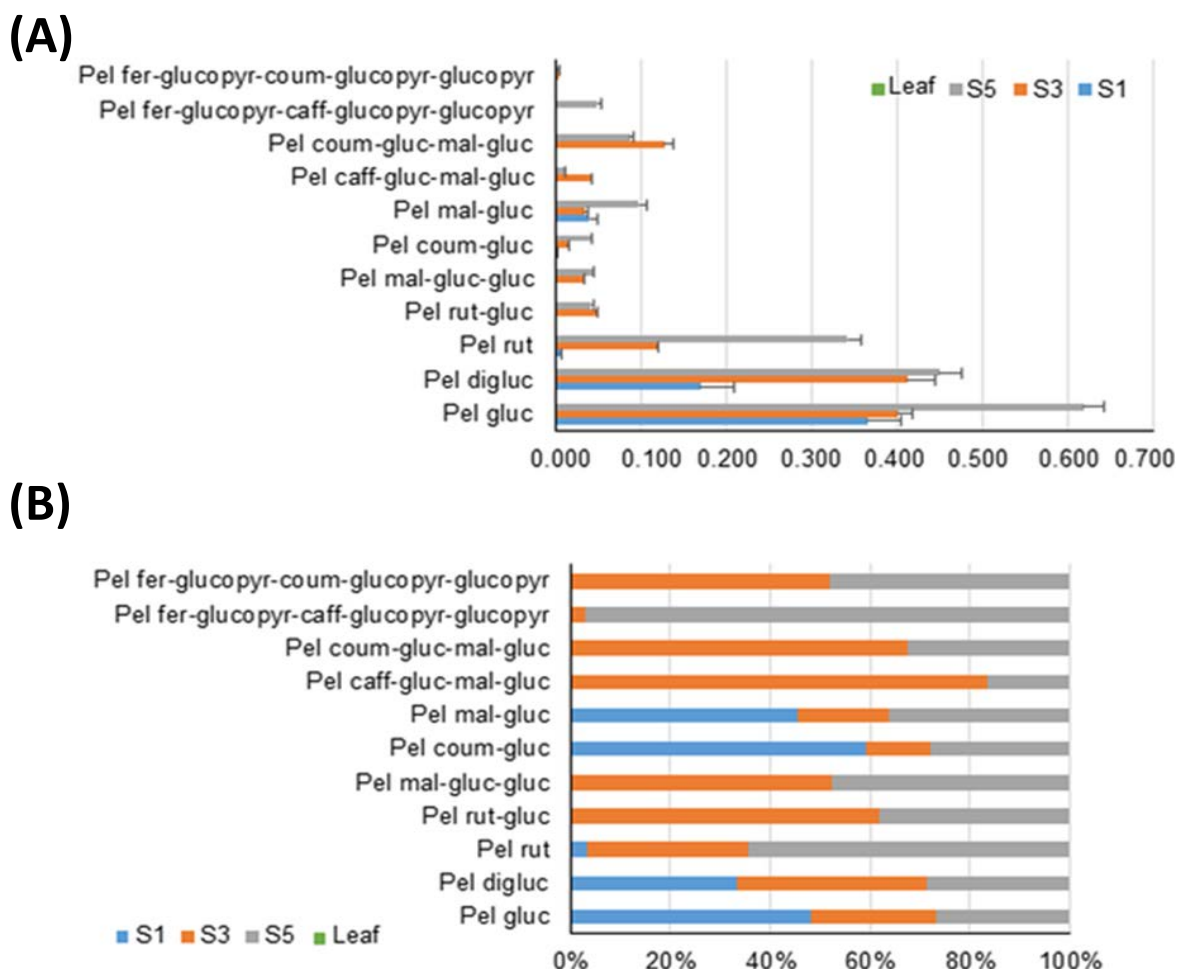


Figure 5.4. Levels of pelargonidin derivatives identified in leaf and petals of *Gentiana lutea* L. var. *aurantiaca* in three different developmental stages by HPLC-ESI-MS/MS analyses. (A) Data are expressed as average and standard deviation which represent, for each anthocyanin metabolite, the fold over the internal standard (IS), have been obtained by using, at least, 3 independent biological replicates. For more details, see Materials and Methods. (B) Percent relative abundance of each anthocyanin among the three developmental stages (S1, S3, S5) under study. Abbreviations: Pel coum-gluc, Pelargonidin 3-*O*-(6-*p*-coumaroyl)glucoside; Pel digluc, Pelargonidin 3,5-*O*-diglucoside; Pel caff-gluc-mal-gluc, Pelargonidin 3-*O*-(6-*O*-caffeoyl-*D*-glucoside)-5-*O*-(6-*O*-malonyl- β -*D*-glucoside); Pel fer-glucopyr-caff-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-*E*)-feruloyl- β -*D*-glucopyranosyl]-6-*O*-(*E*)-caffeoyl- β -*D*-glucopyranoside]-5-*O*-(β -*D*-glucopyranoside); Pel fer-glucopyr-coum-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-*E*)-feruloyl- β -*D*-glucopyranosyl]-6-*O*-(*E*)-*p*-coumaroyl- β -*D*-glucopyranoside]-5-*O*-(β -*D*-glucopyranoside); Pel gluc, Pelargonidin 3-*O*-glucoside; Pel mal-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -*D*-glucoside); Pel rut, Pelargonidin 3-*O*-rutinoside; Pel rut-gluc, Pelargonidin 3-*O*-rutinoside-5-*O*- β -*D*-glucoside; Pel mal-gluc-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -*D*-glucoside)-5- β -*D*-glucoside; Pel coum-gluc-mal-gluc, Pelargonidin 3-*O*-(6-*O*-coumaroyl-*D*-glucoside)-5-(4-*O*-malonyl- β -*D*-glucoside).

Notably, the anthocyanin molecules modified with aromatic acyl groups in their 3- and/or 5-positions appear to provide a reddish-purple color (without the interaction with

co-pigments and/or metal ions) (Sasaki *et al.*, 2014; Sasaki and Nakayama, 2015). Overall, the biochemical analyses showed that there was a significant increase in pelargonidin glycosides as the petals developed, and such accumulation parallels the increasing orange coloration of the petals during development. In most plants, total anthocyanin content increases with flower development and reaches its peak just before the buds open (Justesen *et al.*, 1997; Lou *et al.*, 2017), with a rapid accumulation of anthocyanins in petals at the later stages of their development (Lai *et al.*, 2012; Yamagishi *et al.*, 2010). In order to better understand the relationships among the different anthocyanins detected, we performed a row-directed hierarchical clustering analysis (HCL) (Fig 5.5).

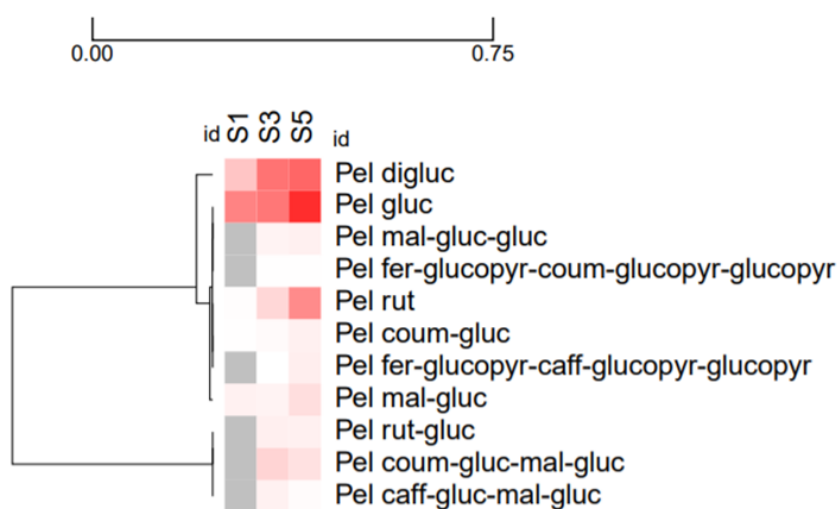


Figure 5.5. Row-directed Hierarchical clustering (HCL) visualization of anthocyanin metabolites detected in flower petals of *G. lutea* L. var. *aurantiaca*. Data represent, for each metabolite, the fold over the internal standard (IS) intensity. Abbreviations: Pel coumgluc, Pelargonidin 3-*O*-(6-*p*-coumaroyl) glucoside; Pel digluc, Pelargonidin 3,5-Odiglucoside; Pel caff-gluc-mal-gluc, Pelargonidin 3-*O*-(6-*O*-caffeoyl-*D*-glucoside)-5-*O*-(6-*O*-malonyl- β -*D*-glucoside); Pel fer-glucopyr-caff-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-(*E*)-feruloyl- β -*D*-glucopyranosyl)-6-*O*-(*E*)-caffeoyl- β -*D*-glucopyranoside]-5-*O*-(β -*D*glucopyranoside); Pel fer-glucopyr-coum-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-(*E*)-feruloyl- β -*D*-glucopyranosyl)-6-*O*-(*E*)-*p*-coumaroyl- β -*D*-glucopyranoside]-5-*O*-(β -*D*glucopyranoside); Pel gluc, Pelargonidin 3-*O*-glucoside; Pel mal-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -*D*-glucoside); Pel rut, Pelargonidin 3-*O*-rutinoside; Pel rut-gluc, Pelargonidin 3-*O*-rutinoside-5-*O*- β -*D*-glucoside; Pel mal-gluc-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -*D*-glucoside)-5- β -*D*-glucoside; Pel coum-gluc-mal-gluc, Pelargonidin 3-*O*-(6-*p*-coumaroyl-*D*-glucoside)-5-(4-*O*-malonyl- β -*D*-glucoside).

This analysis confirmed the presence of a main cluster placed in the top region and composed of pelargonidin 3-*O*-rutinoside and pelargonidin 3-*O*-glucoside, and some of their derivatives, such as pelargonidin 3-*O*-(6-*p*-coumaroyl) glucoside, pelargonidin3-

O-(6-*p*-coumaroyl-D-glucoside)-5-(4-*O*-malonyl- β -D-glucoside), and pelargonidin-3-*O*-(6-*O*-malonyl- β -D-glucoside), thus partially confirming the hypothesis that anthocyanins with simpler compositions as pelargonidin 3-*O*-glucoside serve as precursors for more complex derivatives.

We further analyzed the levels of the flavonoid precursors in the anthocyanin pathway of the same developmental stages (Fig 5.6). Phenylalanine acts as the universal precursor of the phenylpropanoid pathway, leading to the production of coumarins, hydroxycinnamic acids, lignins, flavonoids, isoflavonoids, stilbenes, and a wide variety of other phenolic compounds (Shi and Xie, 2014; Zhang *et al.*, 2014). The levels of phenylalanine decreased from S1 to S3, and reached minimum levels at S5 (Fig. 5.6). Simultaneously, cinnamic acid levels peaked at stages S3 and S5, a pattern opposite to that of coumaric acid (Fig. 5.6), which peaked at S1; finally late anthocyanin precursors such as naringenin chalcone and naringenin gradually increased from S1 to S5, while dihydrokaempferol levels were greatly reduced during flower development. The divergent trends of accumulation between a first metabolic group composed by cinnamic acid, naringenin chalcone and naringenin (increased levels from S1 to S5), and a second including coumaric acid and dihydrokaempferol (higher levels at S1, and lower level at S3 and S5), may suggest the presence of a series of rate-limiting steps in the gentian phenylpropanoid pathway, for example cinnamic acid 4-hydroxylase (C4H) catalyzing aromatic ring-4 hydroxylation of cinnamic acid into *p*-coumaric acid, and/or flavonoid 3-hydroxylase (F3H) converting naringenin to yield dihydrokaempferol, which constrain pelargonidin accumulation in the petals during their development. In agreement with this hypothesis, PAL (phenylalanine ammonia-lyase) and C4H are part of a feedback loop controlling phenylpropanoid biosynthesis (Blount *et al.*, 2000), which acts at the entry point of the pathway; while F3H plays a role in regulating anthocyanin production in *Antirrhinum* (Martin and Gerats, 1993).

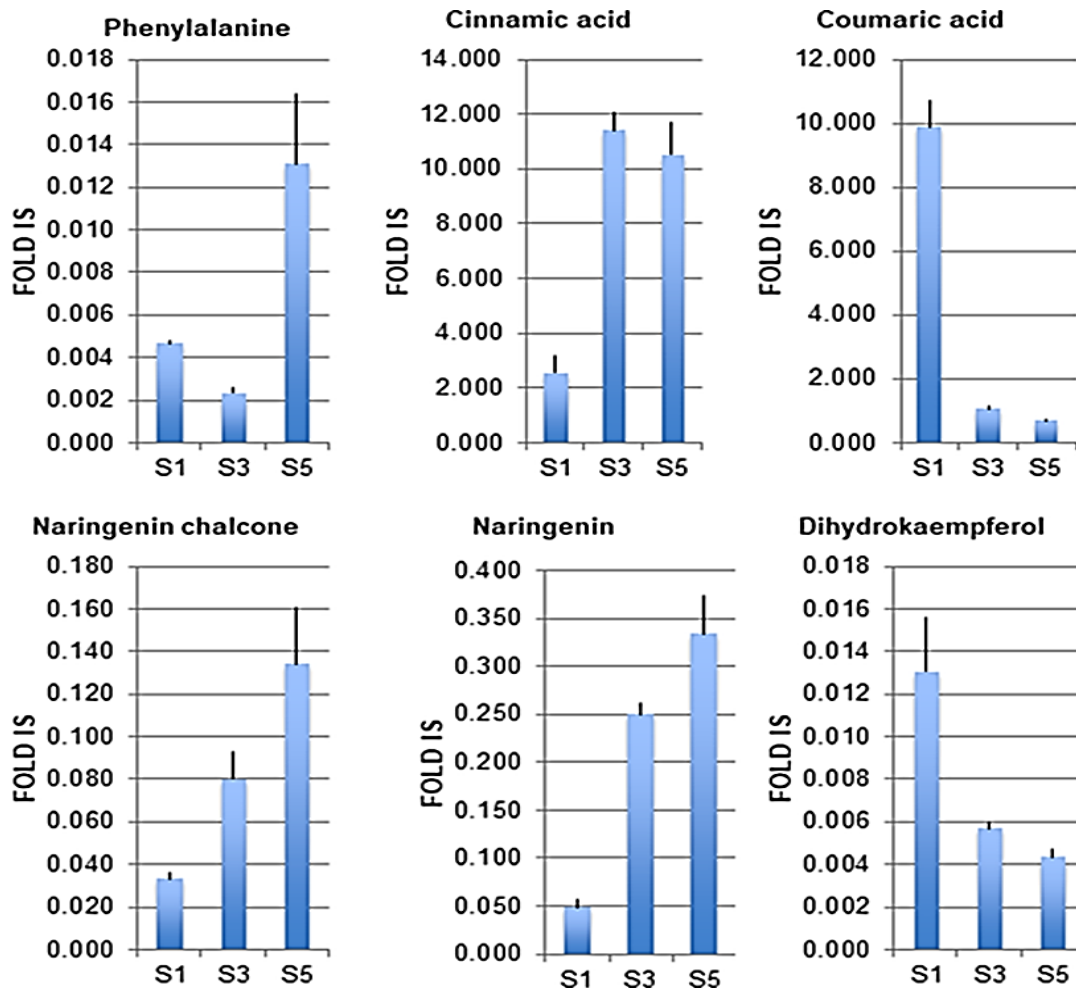


Figure 5.6. Levels of anthocyanin precursors identified in petals of *Gentiana lutea* L. var. *aurantiaca* in three different developmental stages by HPLC-ESI-MS/MS analyses. Data are expressed as average and standard deviation which represent, for each metabolite, the fold over the internal standard (IS), have been obtained by using, at least, three independent biological replicates.

5.4.3. Expression of pelargonidin glucoside biosynthetic genes

As anthocyanin genes are mainly regulated at the level of transcription (Nakatsuka *et al.*, 2005), I analyzed the expression levels of five late anthocyanin biosynthetic genes (*DFR*, *ANS*, *3GT*, *5GT* and *5AT*) which are involved in pelargonidin glycoside biosynthesis (Fig 5.1) in mature leaves and petals of *G. lutea* L. var. *aurantiaca* at S1, S3 and S5. *DFR*, *ANS* and *3GT* genes had been isolated previously by our group (Berman *et al.*, 2016). In this work, I cloned *5GT* and *5AT* gene fragments from mixed petals of stages S1, S3 and S5 by RT-PCR. Each PCR yielded a single product of the expected size. Cloning, sequencing and alignments of these cDNA fragments and their deduced encoding proteins (Fig. 5.7 and Fig. 5.8) demonstrated the isolated *5GT* and

5AT partial cDNAs had high identity (more than 88.0%, eliminating primer sequences) with those from *G. triflora* (Nakatsuka *et al.*, 2008). Quantitative RT-PCR (qRT-PCR) of anthocyanin genes of interest are shown in Fig. 5.9. The petals of *G. lutea* L. var. *aurantiaca* contained substantially higher levels of *DFR*, *ANS*, *3GT*, *5GT* and *5AT* mRNA, while the genes did not express in leaves. Higher relative levels of *DFR*, *ANS* and *3GT* mRNAs at stages S3 and S5 than at stage S1 were found, while similar levels of *5AT* mRNA at stages S1, S3 and S5 were observed (Fig. 5.9). These findings are in agreement with previous studies, which showed relatively high expression levels of acyltransferase genes in the initial stages of petal development (Fujiwara *et al.*, 1998). Notably, the levels of *5GT* mRNA at stage S1 were higher than at stages S3 and S5. A similar observation has been reported in carnation (Matsuba *et al.*, 2010) and peony (Zhao *et al.*, 2014), where *5GT* is also expressed independently of the other genes for anthocyanin biosynthesis (Yabuya *et al.*, 2002).

Figure 5.7. (next two pages) Alignments of partial 5GT and 5AT cDNA sequences encoding UDP-glucose: flavonoid 5-*O*-glucosyltransferase (5GT; A) and anthocyanin 5-aromatic acyltransferase (5AT; B) between *Gentiana triflora* (Gt) and *G. lutea* L. var. *aurantiaca* (Gla). The underlined cDNA sequences indicate the primers used to isolate cDNA fragments from petals of *G. lutea* L. var. *aurantiaca*. Gaps are inserted with a dash (-) in one of the sequences. Abbreviations: Gt, *Gentiana triflora*; Gla, *G. lutea* L. var. *aurantiaca*; 5GT, UDP-glucose: flavonoid 5-*O*-glucosyltransferase gene; 5AT, anthocyanin 5-aromatic acyltransferase gene. GenBank accession numbers: Gt5GT, AB363839; Gt5AT, AB010708. The partial cDNA sequences of 5GT and 5AT genes from *G. lutea* L. var. *aurantiaca* have been isolated by the authors in this study.

A

		551		600
Gt5GT	(551)	GAATCCTGTATACTACT	CTGCTTCCTTGGGCTGCCGATGTGGCTCGTGAA	
Gla5GT	(1)	-----	CTGCTTCCTTGGGCTGCCGATGTGGCTCGTGAA	
Consensus	(551)		CTGCTTCCTTGGGCTGCCGATGTGGCTCGTGAA	
		601		650
Gt5GT	(601)	TTTCGGTTACCCCTCTGTGCTTCTGTGGACACA	CCAGTCACAACCTTTCCT	
Gla5GT	(34)	TTTCGGATACCATCGGTACTTCTGTGGAGCCAG	CCAGTCACAACCTTGGT	
Consensus	(601)	TTTCGC TACC TC GT CTTCTGTGGAC CA	CCAGTCACAACCTT T	
		651		700
Gt5GT	(651)	GACATTCCTACTACTATTTCACTGGCTATGAAGAT	GCTATCAA CAAGGTAC	
Gla5GT	(84)	GACATTCCTACTACTATTTCACTGGCTATGAAGAC	GCATCAA GGAAGTAT	
Consensus	(651)	GACATTCCTACTACTATTTCA TGGCTATGAAGA GC	ATCAA A GTA	
		701		750
Gt5GT	(701)	GCAATCAACAAGGTACAGAAGATGATTC	CACTATTCAGCTTCC TAGACTT	
Gla5GT	(134)	GCAAT-----GGTACAG-----	ATTCACGATTCAGCTTCCAAGACTT	
Consensus	(701)	GCAAT GGTACAG	ATTC AC ATTCAGCTTCC AGACTT	
		751		800
Gt5GT	(751)	CCCTGCTGTCTAGTCGTGATTTACATTCTTTCATGCT	ACCATCTAATCC	
Gla5GT	(172)	CCCTGCTGTCTAGCCGGGATCTTCATTCTTTCATGCT	TCCGTC AATCC	
Consensus	(751)	CCCTGCTGTCTAG CG GAT T CATTCTTTCATGCT	CC TC AATCC	
		801		850
Gt5GT	(801)	GTTTAAAGGTGCATCAACACGGTTCAAAGAGCACCT	GGAAGCACTTGATG	
Gla5GT	(222)	GTTTAAAGGTGCAC TCAAGACGGTTCAAAGAGCACCT	GGAAGCGCTTGACG	
Consensus	(801)	GTTTAAAGGTGC TCAA ACGTTCAAAGAGCACCT	GGAAGC TTGA G	
		851		900
Gt5GT	(851)	CAGAA GAAAC TCCTCAA CCAATTC	TTGTTAACAGTTACGATGCGTTAGAG	
Gla5GT	(272)	CAGAA CAAAT TCCGACAA---	TTATTGTTAACAGTTACGACGCTTTAGAG	
Consensus	(851)	CAGAA AAA TCC CAA TT	TTGTTAACAGTTACGA GC TTAGAG	
		901		950
Gt5GT	(901)	GAAGAGGCCTTGCA GCGATGAT	CCCAAAGTACAAAACAATGGGGATCGG	
Gla5GT	(319)	GAAGAGGCCTTGCAAGCAAT---	CCCAAAGTACAAAACAATGGGGATGG	
Consensus	(901)	GAAGAGGCCTTGCA GC AT	CCCAAAGTACAAAACAATGGGGAT GG	
		951		1000
Gt5GT	(951)	ACCATTGATTCCTCGTCCGTTTTCGAC	ACCAAGGAAACAACAATGTGAGG	
Gla5GT	(366)	ACCATTGATTCCTCATCCATTTTGAT	ACCAAGGAAACA---TGTGAGG	
Consensus	(951)	ACCA TGATTCCTC TCC TTTT GA	ACCA GGAAACA T TGAGG	
		1001		1050
Gt5GT	(1001)	TGGTTTCTCTGTGTTCCAGATTGCG	GCAAAAAGTCA AAGGACGACTGCCAG	
Gla5GT	(413)	T---TTCTCTGTGTTGCATATTTGTT	GCAAAAAGTCA AAGGAC-----	
Consensus	(1001)	T TTCTCTGTGTT CA ATTTG	GCAAAAAGTC AAGGAC	
		1051		1100
Gt5GT	(1051)	TGGCATGGATGGTTGAACTCAAAA	CCAGAGGGGTCAGTGATTTATGTATC	
Gla5GT	(451)	TGCCATGGATGGTTGAACTCAAAA	CCCAAGGGTCAGTGATTTATGTATC	
Consensus	(1051)	TG CATGGATGGTTGAACTCAAAA C	A GGGTCAGTGATTTATGTATC	
		1101		1150
Gt5GT	(1101)	ATTCGGAAGTCATGTC AAGCAATCTAAA	GTC A AACC GGAGGAGATAGCAA	
Gla5GT	(501)	ATTTGGAAGTCATGT AAGCAATCTAAA	TCC A AACC GGAGGAGATAGCAA	
Consensus	(1101)	ATT GGAAGTCATGT AAGCA TCTAAA C	CAAACC GGAGGAGATAGCAA	
		1151		1200
Gt5GT	(1151)	AAGGGCTCCTCGCGAGTGGCCATCCATTCT	TGTGGGTGATTACATCGAAC	
Gla5GT	(551)	AAGGGCTCCTCGCGAGTGGCCATCCATTCT	TGTGGGTGATTACATCA AAT	
Consensus	(1151)	AAGGGCTCCTCGCGAGTGGCCATCCATTCT	TGTGGGTGATTACATC AA	
		1201		1250
Gt5GT	(1201)	GAAGAAGAGG-----AGGGT	GATCAAAATAATGGAACA A AATTTGGTGGAA	
Gla5GT	(601)	GAGGAGGAGGCCGAAAAGGAAAAGT	ACTAATGGAACGAAATTTGCTGAA	
Consensus	(1201)	GA GA GAGG A GG A G A	TAATGGAAC AAATTTG T G A	
		1251		1300
Gt5GT	(1245)	GGAAATTCAAGAAAAAGGGATGATGA	TAGTACCATGGTGTGCTCAGTTTC	
Gla5GT	(651)	GGAAATTCAAGAAAAAGGGATGA---	TAGTACCATGGTGTGCTCAGATTTC	
Consensus	(1251)	GGAAATTCAAGAAAAAGGGATGA	TAGTACCATGGTGTGCTCAG TTC	
		1301		1350
Gt5GT	(1295)	AGGTGCTAAAGCATCCGCTCT	GTGGGATGCTTCATGACACATTGCCGATGG	
Gla5GT	(698)	AGGTGCTAAAGCATCAC	TGGGATGTTTCATGACACATTGCCGATGG	
Consensus	(1301)	AGGTGCTAAAGCATC TC	GTGGGATG TTCATGACACATTGCCGATGG	
		1351		1400
Gt5GT	(1345)	AATTC AACGCTGGAGAGCATAGCT	TGTGGGTGCCTATGATTGGTTTTC	
Gla5GT	(748)	AATTC AACGCTGGAGAGC-----		
Consensus	(1351)	AATTC AACGCTGGAGAGC		

B

		451		500
Gt5AT	(451)	AAGTGATC	CCGCTCGTAGCCGTGCAAGTAACCGTTTTTCCTAACCGTGGC	
Gla5AT	(1)	-----	CCGCTCGTAGCCGTGCAAGTAACCGTTTTTCCTAACCA	TGGC
Consensus	(451)		CCGCTCGTAGCCGTGCAAGTAACCGTTTTTCCTAAC	TGGC
		501		550
Gt5AT	(501)	ATAGCCGTGGCTCTGACGG	CATCAT	TCAATTGCAGATGCTAAAAGTTT
Gla5AT	(43)	ATAGCCGTGGCTCTGACG	ACGCATCA	TCAATTGCAGATGGAAGAAGTGC
Consensus	(501)	ATAGCCGTGGCTCTGACG	C CATCA	TCAATTGCAGATG A AAGT
		551		600
Gt5AT	(551)	TGTAATGTTTCAT	CAATGCTTGGGCTATATTAACAAATTTGGGAAAGACG	
Gla5AT	(93)	TGTAAGTTCAT	GAATGCTTGGGCTATATTAACAAATTTGGGAAAGAG	
Consensus	(551)	TGTAAGTTCAT	AATGCTTGGGCTATATTAACAAATTTGG	AAAGA G
		601		650
Gt5AT	(601)	CGGACTTGTGT	CCGCGAATCTTCTTCCATCTTT	GATAGATCGATAATC
Gla5AT	(143)	CAGACTTATTTGT	GTAGAAATCTTCTTCCATCTTT	GATAGATCCATAATC
Consensus	(601)	C GACTT TGT	AATCTTCTTCCATCTTT	GATAGATC ATAATC
		651		700
Gt5AT	(651)	AAAGATCT	GTATGGCCTAGAGGAAACATTTTGGAACGAAATGCAAGATGT	
Gla5AT	(193)	AAAGATCC	GTATGGCCTAGAGGAAATATTTTGGAACGAAATGCAAGATAT	
Consensus	(651)	AAAGATC	GTATGGCCTAGAGGAAA	ATTTTGGAACGAAATGCAAGAT T
		701		750
Gt5AT	(701)	TCTTGAAATGTTTCTCTAGATT	TGGAA	GCAAACCCCTCGATTCAACAAGG
Gla5AT	(243)	TCTCGAATGTTTCTCTAGATT	TGGAA	CAAACCCCTCGATTCAACAAGG
Consensus	(701)	TCT GAA	TGTTCTCTAGATT	TGGAA CAAACCCCTCGATTCAACAAGG
		751		800
Gt5AT	(751)	TACGAGCTACATATGT	CCTCCCTTG	TGAAATCCAGAAGCTAAAGAAC
Gla5AT	(293)	TACGAGCCACATATGT	ACTATCCCTTG	TGAAATCGAGGCTAAAGAAC
Consensus	(751)	TACGAGC	ACATATGT CT	TCCCTTG TGAAATC AGA GCTAAAGAAC
		801		850
Gt5AT	(801)	AAAGTACTGAATCTCAGAGGAT	CCGAACCGACAATACGTGTAACGACGTT	
Gla5AT	(343)	AAAGTACTGAATCTCAGAGGAT	GTGAACCAACAATACGCGTAACACATT	
Consensus	(801)	AAAGTACTGAATCTCAGAGGAT	GAACC	ACAATACG GTAAC AC TT
		851		900
Gt5AT	(851)	CACAATGACG	TGTGGATACGTATGGACATGCATGGT	CAATCAAAAGATG
Gla5AT	(393)	CACAGTGACAT	TGTGGATACATATGGACATGCATGGT	TAAATCAAAAGAG
Consensus	(851)	CACA	TGAC TGTGGATAC	TATGGACATGCATGGT AAATCAAAAGA G
		901		950
Gt5AT	(901)	ACGTCGT	ATCAGAGGAATCATCGAACGACG	AAAATGAGCTCGAGTACTTC
Gla5AT	(443)	GCA	---ATCAGAGGAATCATCAAACGACAAAATGAGCTCGAGTACTTC	
Consensus	(901)	C T	ATCAGAGGAATCATC AACGAC	AAAATGAGCTCGAGTACTTC
		951		1000
Gt5AT	(951)	AGTTTACAGCG	GATTGCCGAGGACTTCTGACGCCCC	CGTGCCGCCTAA
Gla5AT	(490)	AGTTTACAGCT	CAATGCCGAGGCTTCTGACGCCCC	CGTGCCGCCTAA
		1001		1050
Gt5AT	(1001)	CTACTTTGG	CAACTGTCTTGCGT	CATGGCTTGCAAAAGCAACACATAAAG
Gla5AT	(540)	CTACTTTGGA	AACTGTCTTGCGCGGTGCTTGCAAAAGCAACACATAAAG	
Consensus	(1001)	CTACTTTGG	AACGTCTTGCG C	TGC TTGCAAAAGCAACACATAAAG
		1051		1100
Gt5AT	(1051)	AGTTAGTT	GGGATAAAGGGCTTCTTGTTCAGTTGCAGCT	ATTGGAGAA
Gla5AT	(590)	AGTTAAT	CGGAAATAAAGGGTTCCTTGTTCAGTTGCAGCT	ATTGGGGAT
Consensus	(1051)	AGTTA	T GG	ATAAAGGG TTCTTGTTCAGTTGCAGCT TTGG GA
		1101		1150
Gt5AT	(1101)	GCCATTGAAAAGAGG	TGCACAAAGG	C GTTCTTGAGATGCAAA
Gla5AT	(640)	GCCATTGAAAAGAGG	TGCAGAACAAAGAGG	TGTCTTGAGATGCAAA
Consensus	(1101)	GCCATTGAAAAGAGG	TGCA AAC AA	AAGG GTTCTTGAGATGCAAA
		1151		1200
Gt5AT	(1151)	AACTTGGTTATC	GGAATCTAA	TGGAATCCCTTCAAAAAGATTCTCGGGA
Gla5AT	(690)	AACTTGGTTATC	GGAATCTAAAGGAATCCCTTCA	GAAAGATTGCTCGGGA
Consensus	(1151)	AACTTGGTTATC	GAATCTAA	GGAATCCCTTCA AAAGATT CTCGGGA
		1201		1250
Gt5AT	(1201)	TTACCGGATCGCCTAAGTTCGATT	TCGTATGGTGTAGATTTTGGATGGGGA	
Gla5AT	(740)	TTAGTGGATCGCCTAAGTTCGATT	TCGTATGGTGTAGATTTTGGATGGGGA	
Consensus	(1201)	TTA	GGATCGCCTAAGTTCGATTTCGTATGGTGTAGATTTTGGATGGGGA	
		1251		1300
Gt5AT	(1251)	AAGCCTGCAAAATTGA	CATTACCTCTGTGATTATGCAGAATTGATTTA	
Gla5AT	(790)	AAGCCTGCAAAATTGA	-----	
Consensus	(1251)	AAGCCTGCAAAATTGA		

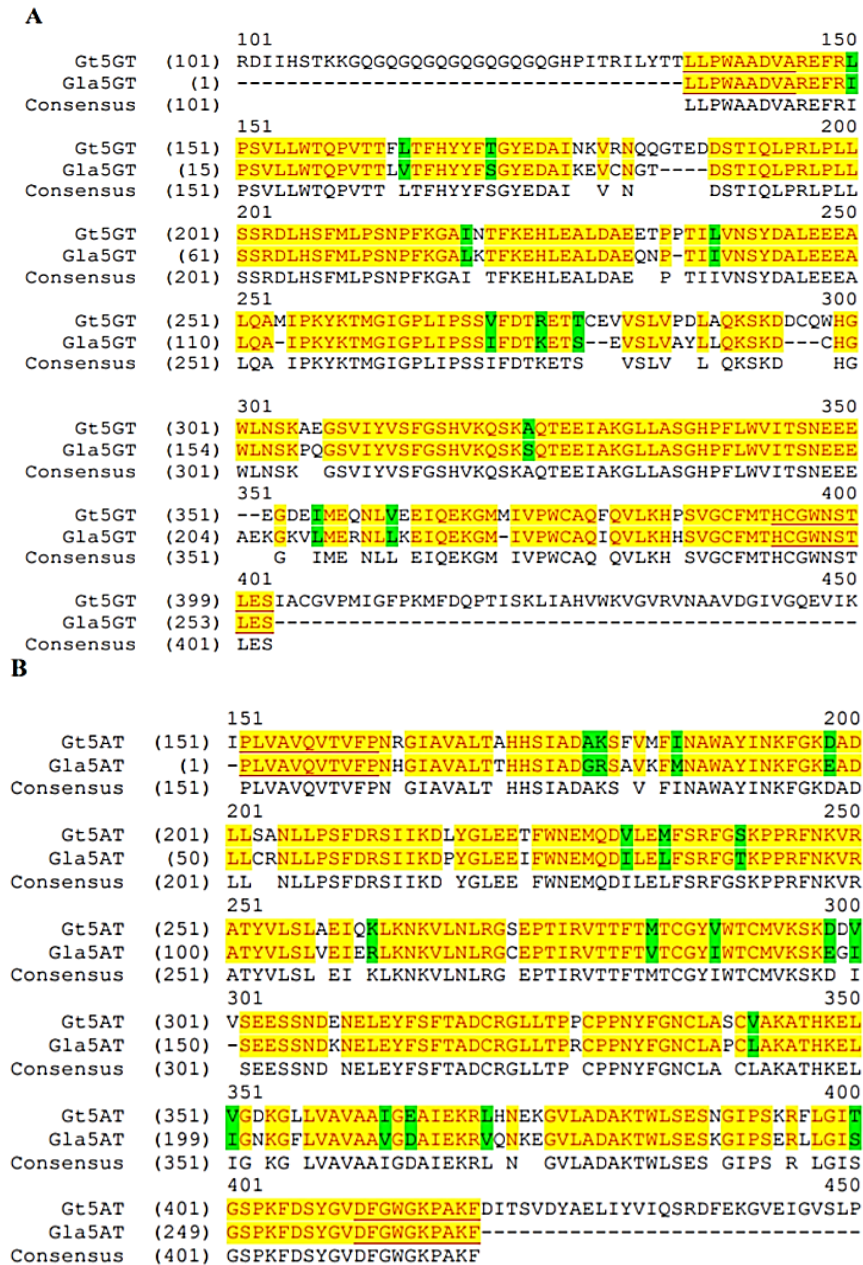


Figure 5.8. Alignments of the deduced amino acid sequences encoded by UDP glucose: flavonoid 5-*O*-glucosyltransferase (5GT) gene (A), and anthocyanin 5-aromatic acyltransferase (5AT) gene (B) from *Gentiana triflora* and *G. lutea* L. var. *aurantiaca*. Gaps are inserted with a dash (-) in one of the sequences. The underlined amino acid sequences from *G. lutea* L. var. *aurantiaca* were deduced from primers. Abbreviations: Gt, *Gentiana triflora*; Gla, *G. lutea* L. var. *aurantiaca*; 5GT, UDPglucose: flavonoid 5-*O*-glucosyltransferase; 5AT, anthocyanin 5-aromatic acyltransferase. GenBank accession numbers: Gt5GT, AB363839; Gt5AT, AB010708. The partial amino acid sequences of 5GT and 5AT of *G. lutea* L. var. *aurantiaca* were deduced from partial cDNA sequences cloned by the authors in this study.

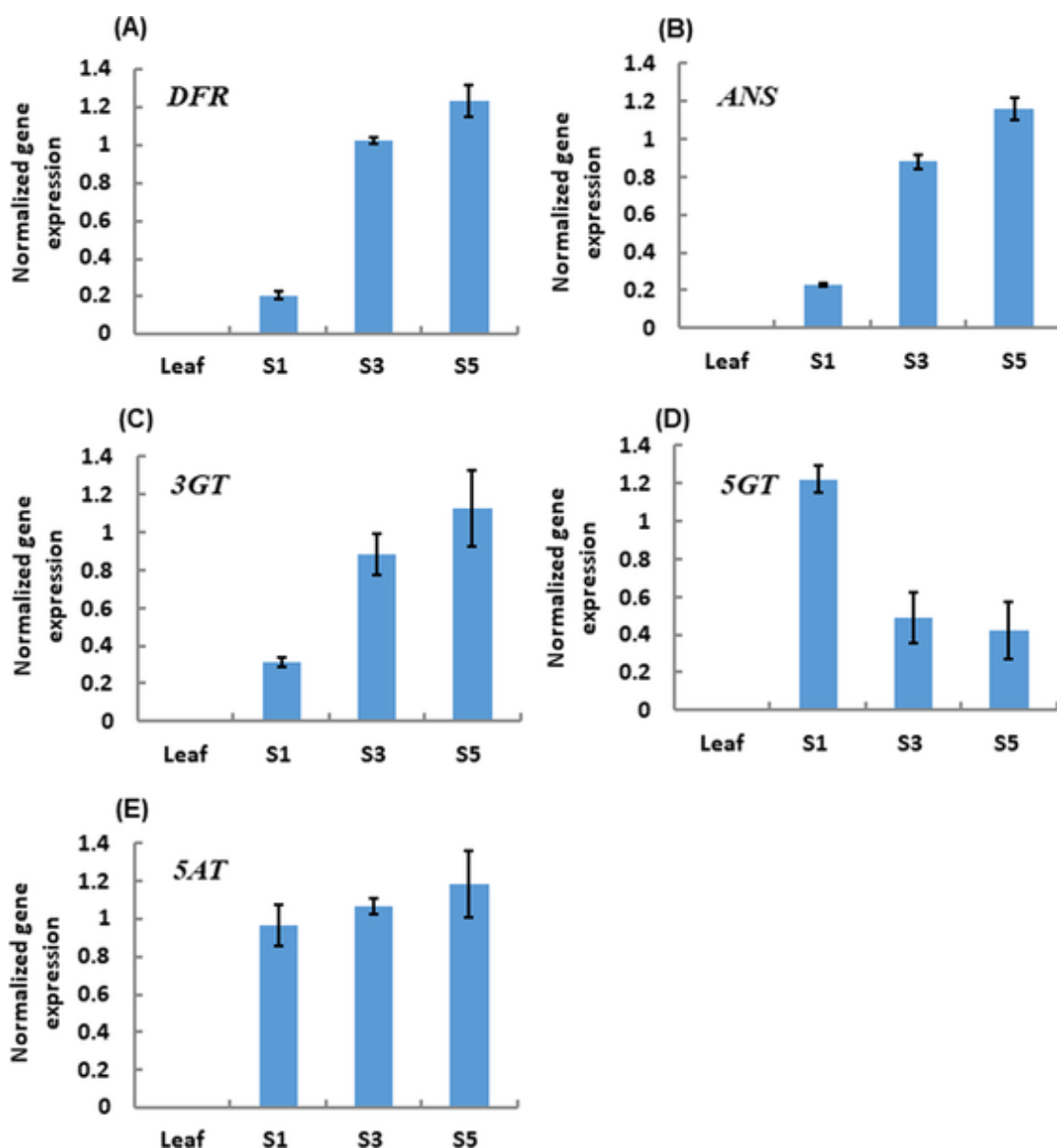


Figure 5.9. Quantitative expression of anthocyanin genes, normalized on the ubiquitin housekeeping gene in leaf and petals of *Gentiana lutea* L. var. *aurantiaca*. qRT-PCT data were calculated from three biological replicates with at least three technical replicates for each biological replicate and with error bars representing the standard deviation. Abbreviations: 5AT, anthocyanin 5-aromatic acyltransferase gene; ANS, anthocyanidin synthase gene; DFR, dihydroflavonol 4-reductase gene; 3GT, UDP-glucose: flavonoid 3-*O*-glucosyltransferase gene; 5GT, UDP-glucose:flavonoid 5-*O*-glucosyltransferase gene.

5.4.4. Integration of transcript and metabolite data of gentian phenylpropanoid pathway

Pearson correlation coefficients (ρ s) were calculated for each data pair (transcript-transcript, transcript-metabolite, metabolite-metabolite), followed by correlation matrix and network visualizations (Figs. 5.9 and 5.10, respectively). Overall, most of the correlations observed were of positive sign and very strong ($\rho \geq 0.65$), indicating that most transcript and metabolite levels exhibited similar trends in their expression and accumulation. Thus, at statistical level, this finding suggests the presence of a high extent of coordination in the synthesis and accumulation of pelargonidin anthocyanins along flower petal development. However, a notable exception was found within *5GT* transcript (Fig 5.10). The presence of a divergent altitude for this element may be associated with specific controls at the gene expression or metabolite levels (rate limiting steps or negative feedback, as mentioned earlier). In addition, force-directed correlation network was exploited to highlight the most relevant elements in the dataset (“hubs” of the network) mainly responsible for anthocyanin production and accumulation (Fig 5.10). Interestingly, most of the pelargonidin anthocyanins placed, in a right- and top-side region in the network. Since, in a correlation network, the topology is generated by the correlations of the input file themselves, and nodes which do not display high number of significant correlations towards the rest of the nodes in the dataset normally localize in distal regions of the network, this evidence provides clues about a non-active role in the control of the pathway. Conversely most of the genes were located in the central area of the network. Among them, *DFR*, *ANS* and *3GT* displayed the highest number of very significant correlations, as evidenced by their topology and by the node strength (ns) values (e.g. the average of all the $|\rho|$ yielded by each node) (Table 5.2).

Table 5.2. Node strength (ns) and network strength (NS) of gentian phenylpropanoid genes and metabolites, expressed as the average of all the $|\rho_s|$ yielded by a node and the average of the ns, respectively.

ID	Kind	Node strength (ns)
ANS	gene	0.84
3GT	gene	0.84
Pelargonidin 3- <i>O</i> -(6- <i>O</i> -caffeoyl- β -D-glucoside)-5- <i>O</i> -(6- <i>O</i> -malonyl- β -D-glucoside)	anthocyanin	0.84
DFR	gene	0.84
Dihydrokaempferol	precursor	0.83
Pelargonidin 3,5- <i>O</i> -diglucoside	anthocyanin	0.83
5AT	gene	0.82
Pelargonidin 3- <i>O</i> -[2- <i>O</i> -(6- <i>E</i>)-feruloyl- β -D-glucopyranosyl]-6- <i>O</i> -(<i>E</i>)- <i>p</i> -coumaroyl- β -D-glucopyranoside]-5- <i>O</i> -(β -D-glucopyranoside)	anthocyanin	0.82
5GT	gene	0.81
Pelargonidin 3- <i>O</i> -rutinoside	anthocyanin	0.78
Pelargonidin 3- <i>O</i> -rutinoside-5- <i>O</i> - β -D-glucoside	anthocyanin	0.76
Pelargonidin 3- <i>O</i> -(6- <i>p</i> -coumaroyl)glucoside	anthocyanin	0.76
Pelargonidin 3- <i>O</i> -glucoside	anthocyanin	0.71
Pelargonidin 3- <i>O</i> -(6- <i>O</i> -malonyl- β -D-glucoside)-5-glucoside	anthocyanin	0.67
Pelargonidin 3- <i>O</i> -[2- <i>O</i> -(6- <i>E</i>)-feruloyl- β -D-glucopyranosyl]-6- <i>O</i> -(<i>E</i>)-caffeoyl- β -D-glucopyranoside]-5- <i>O</i> -(β -D-glucopyranoside)	anthocyanin	0.66
Pelargonidin 3- <i>O</i> -(6- <i>O</i> -malonyl- β -D-glucoside)	anthocyanin	0.61
Pelargonidin 3- <i>O</i> -(6- <i>p</i> -coumaroyl- β -D-glucoside)-5-(4- <i>O</i> -malonylglucoside)	anthocyanin	0.45
	Network strength (NS)	0.76

Finally, the determination of such high network strength (NS; equal to 0.76) obtained by calculating all the $|\rho_s|$ in the network, confirms the existence of strong and harmonic relationships both at the transcript and metabolite levels, underlying the generation of all the anthocyanin set in the petals of the orange gentian.

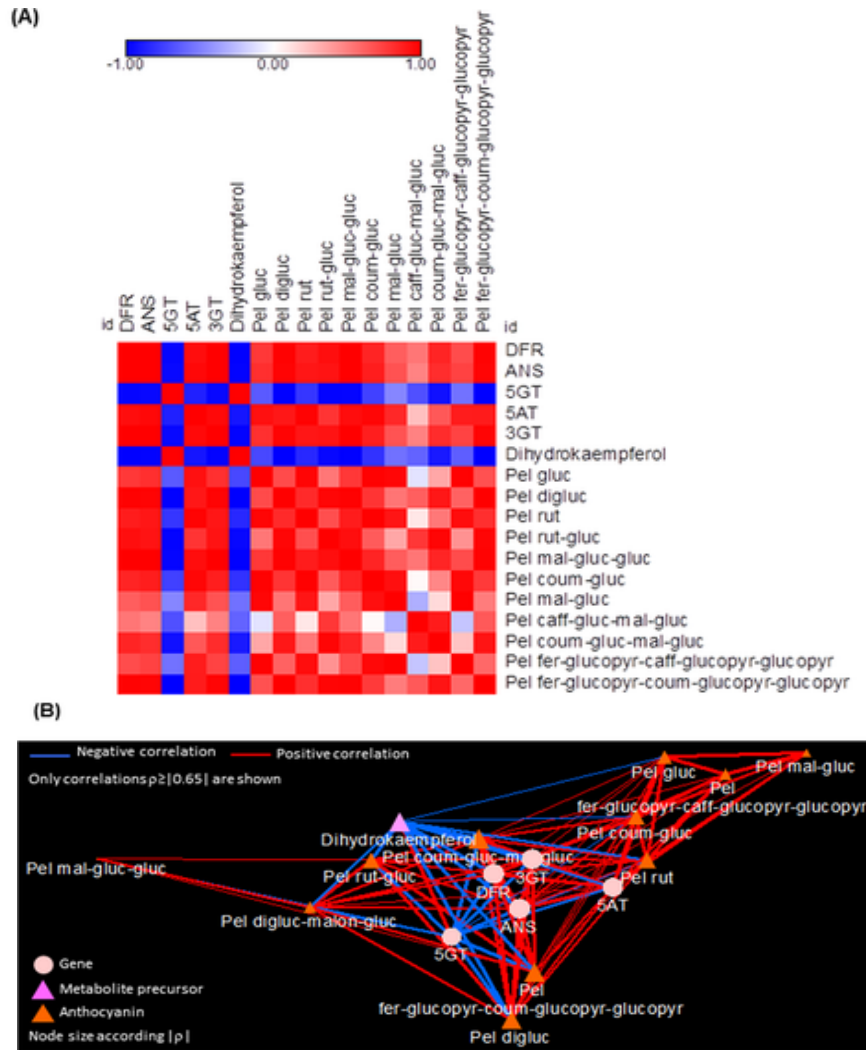


Figure 5.10. Integration of transcript-metabolite data involved in *G. lutea* L. var. *aurantiaca* anthocyanin metabolism. (A) Pearson coefficient-based correlation matrix. Legend on the right corresponds to the names of the anthocyanin transcripts and metabolites. Red and blue shaded boxes represent different levels of positive and negative correlations, respectively; white boxes represent no correlation. (B) Anthocyanin transcript/metabolite correlation network using a prefuse force-directed layout (only $\rho > 0.65$ are shown). Transcripts, anthocyanins and anthocyanin precursors are represented, respectively, as pink rounds, and orange and violet triangles. Blue and red edges refer to negative and positive correlations, respectively. Node size is according to the node strength (ns, representing the average of all the ρ s of each node). Lines joining the nodes indicate positive (red) and negative (blue) correlations, of width proportional to each corresponding $|\rho|$.

Abbreviations: ANS, anthocyanidin synthase; 5AT, anthocyanin 5-aromatic acyltransferase; DFR, dihydroflavonol 4-reductase; 3GT, UDP-glucose:flavonoid 3-*O*-glucosyltransferase gene; 5GT, UDP-glucose: flavonoid 5-*O*-glucosyltransferase; Pel coum-gluc, Pelargonidin 3-*O*-(6-*p*-coumaroyl)glucoside; Pel digluc, Pelargonidin 3,5-*O*-diglucoside; Pel caff- gluc-mal-gluc, Pelargonidin 3-*O*-(6-*O*-caffeoyl- β -D-glucoside)-5-*O*-(6-*O*-malonyl- β -D-glucoside); Pel fer-glucopyr-caff-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-*E*)-feruloyl- β -D-glucopyranosyl]-6-*O*-(*E*)-caffeoyl- β -D-glucopyranoside]-5-*O*-(β -D-glucopyranoside); Pel fer-glucopyr- coum-glucopyr-glucopyr, Pelargonidin 3-*O*-[2-*O*-(6-*E*)-feruloyl- β -D-glucopyranosyl]-6-*O*-(*E*)-*p*-coumaroyl- β -D-glucopyranoside]-5-*O*-(β -D- glucopyranoside); Pel gluc, Pelargonidin 3-*O*-glucoside; Pel mal-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -D-glucoside); Pel rut, Pelargonidin 3-*O*- rutinoside; Pel rut-gluc, Pelargonidin 3-*O*-rutinoside-5-*O*- β -D-glucoside; Pel mal-gluc-gluc, Pelargonidin 3-*O*-(6-*O*-malonyl- β -D-glucoside)-5- β -D- glucoside; Pel coum-gluc-mal-gluc, Pelargonidin 3-*O*-(6-*p*-coumaroyl-D-glucoside)-5-(4-*O*-malonyl- β -D-glucoside).

5.5. Discussion

Pelargonidin derivatives have been found in different plant species including orange cultivars of *Pelargonium* where they can account for up to 98% of total anthocyanins (Mitchell *et al.*, 1998), *Dahlia variabilis* (Ohno *et al.*, 2011), magenta flowers of *Verbena* (Toki *et al.*, 1995), red-purple flowers of *Matthiola incana* (Saito *et al.*, 1996a) and *Ipomoea purpurea* (Saito *et al.*, 1996b), red flowers of *Hyacinthus orientalis* (Hosokawa *et al.*, 1995), maroon flowers of *Pharbitis nil* (Lu *et al.*, 1992), and white and pink flowers of *Muscari* spp. (Lou *et al.*, 2017). All identified anthocyanin molecules reported to increase anthocyanin stability in aqueous solution and may likely alter their light absorption properties (Yonekura-Sakakibara *et al.*, 2009; Shi and Xie, 2014). All metabolites identified contained at least one sugar group. The hydroxyl groups at C3 and C5 positions of pelargonidin have been reported to be the two most common targets of glucosylation (Yonekura-Sakakibara *et al.*, 2009). In the present study, my results showed that pelargonidin 3-*O*-glucoside, pelargonidin 3,5-*O*-diglucoside and pelargonidin 3-*O*-rutinoside are the main pelargonidin derivatives present in *G. lutea* L. var *aurantiaca* petals. However, in addition to these molecules, other more complex pelargonidin derivatives were tentatively identified. I analyzed the expression of late *DFR*, *ANS*, *3GT*, *5GT* and *5AT* genes specifically in anthocyanin biosynthesis in the petals of *G. lutea* L. var *aurantiaca* to determine the molecular mechanisms responsible for the accumulation of pelargonidin glycosides during flower development. The accumulation of the pelargonidin glucosides parallels the expression levels of the genes *DFR*, *ANS* and *3GT*, suggesting a major role in the accumulation of these anthocyanins in *G. lutea* L. var *aurantiaca* flowers, with *DFR* most probably acting as the limiting step in this process (Berman *et al.*, 2016). The resulting data will enhance our understanding of the structural diversity of anthocyanin diversity in *Gentiana* species, and lays a foundation for breeding for flower color and genetic variation studies in *Gentiana* varieties.

5.6. Conclusions

Eleven pelargonidin derivatives were identified in the petals of *G. lutea* L. var. *aurantiaca*, and different amounts were observed at each developmental stage. The major pelargonidin derivatives at all the stages were pelargonidin 3-*O*-glucoside,

pelargonidin 3,5-*O*-diglucoside and pelargonidin 3-*O*-rutinoside. Gene expression analysis revealed *DFR*, *ANS* and *3GT* parallel the accumulation of pelargonidin glucosides.

5.7. References

Ahrazem O, Argandona J, Fiore A, Aguado C, Lujan R, Rubio-Moraga A, Marro M, Araujo-Andrade C, Loza-Alvarez P, Diretto G, 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.

Barnes JS, Schug KA (2011). Structural characterization of cyanidin-3,5-diglucoside and pelargonidin-3,5-diglucoside anthocyanins: Multi-dimensional fragmentation pathways using high performance liquid chromatography-electrospray ionization-ion trap-time of flight mass spectrometry. *Int. J. Mass Spectrom.*, **308**:71–80.

Berman J, Sheng Y, Gomez-Gomez L, Veiga T, Ni X, Farre G, Capell T, Guitián J, Guitián P, Sandmann G, Christou P, Zhu C (2016). Red anthocyanins and yellow carotenoids form the color of orange-flower gentian (*Gentiana lutea* L. var. *aurantiaca*). *PLoS One*, **11**:e0162410.

Blount JW, Korth KL, Masoud SA, Rasmussen S, Lamb C, Dixon RA (2000). Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiol.*, **122**:107–116.

Diretto G, Al-Babili S, Tavazza R, Scossa F, Papacchioli V, Migliore M, Beyer P, Giovanni G (2010). Transcriptional-metabolic networks in beta-carotene-enriched potato tubers: the long and winding road to the Golden phenotype. *Plant Physiol.*, **154**:899–912.

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD (2004). Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.*, **35**:375–403.

- Fukuchi-Mizutani M, Okuhara H, Fukui Y, Nakao M, Katsumoto Y, Yonekura-Sakakibara K, Kusumi T, Hase T, Tanaka Y (2003). Biochemical and molecular characterization of a novel UDP-glucose:anthocyanin 3'-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from gentian. *Plant Physiol.*, **132**:1652–1663.
- Fujiwara H, Tanaka Y, Fukui Y, Nakao M, Ashikari T, Kusumi T (1997). Anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. Purification, characterization and its role in anthocyanin biosynthesis. *Eur. J. Biochem.*, **249**:45–51.
- Fujiwara H, Tanaka Y, Yonekura-Sakakibara K, Fukuchi-Mizutani M, Nakao M, Fukui Y, Yamaguchi M, Ashikari T, Kusumi T (1998). cDNA cloning, gene expression and subcellular localization of anthocyanin 5-aromatic acyltransferase from *Gentiana triflora*. *Plant J.*, **16**:421–431.
- Giusti MM, Wrolstad RE (2003). Acylated anthocyanins from edible sources and their applications in food systems. *Biochem. Eng. J.*, **14**:217–225.
- Gomez-Gomez L, Parra-Vega V, Rivas-Sendra A, Segui-Simarro JM, Molina RV, Pallotti C, Rubio-Moraga A, Direccion G, Prieto A, Ahrazem O (2017). Unraveling massive crocins transport and accumulation through proteome and microscopy tools during the development of saffron stigma. *Int. J. Mol. Sci.*, **18**: pii:E76.
- Harborne JB (1958a). Spectral methods of characterizing anthocyanins. *Biochem J.*, **70**:22–28.
- Harborne JB (1958b). The chromatographic identification of anthocyanin pigments. *J. Chromatogr.*, **1**:210–224.
- Harborne JB. Anthocyanins and their sugar components (1962). *Fortschr. Chem. Org. Naturst.*, **20**:165–199.
- Harborne JB, Sherratt HS (1957). Variations in the glycosidic pattern of anthocyanins. II. *Experientia*, **13**:486–487.
- Harborne JB, Williams CA (2001). Anthocyanins and other flavonoids. *Nat. Prod. Rep.*, **18**:310–333.

Holton TA (1995). Modification of flower colour via manipulation of P450 gene expression in transgenic plants. *Drug Metabol. Drug Interact.*, **12**: 359–368.

Hong V, Wrolstad RE (1990). Use of HPLC separation photodiode array detection for characterization of anthocyanins. *J. Agric. Food Chem.*, **38**:707–715.

Hosokawa K, Fukushi E, Kawabata J, Fuji C, Ito T, Yamamura S (1995a). Three acylated cyanidin glucosides in pink flowers of *Gentiana*. *Phytochem.*, **40**:941–944.

Hosokawa K, Fukunaga Y, Fukushi E, Kawabata J (1995b). Seven acylated anthocyanins in the blue flowers of *Hyacinthus orientalis*. *Phytochem.*, **38**:1293–1298.

Justesen H, Andersen AS, Brandt K (1997). Accumulation of anthocyanins and flavones during bud and flower development in *Campanula isophylla* Moretti. *Ann Bot.*, **79**:355–360.

Kamsteeg J, van Brederode J, van Nigtevecht G (1978). Identification, properties, and genetic control of UDPglucose: cyanidin-3-rhamnosyl-(1 leads to 6)-glucoside-5-O-glucosyltransferase isolated from petals of the red campion (*Silene dioica*). *Biochem Genet.*, **16**:1059–1071.

Kim HW, Kim JB, Cho SM, Chung MN, Lee YM, Chu SM, Chu M, Che JH, Kim SN, Kim SY, Cho YS, Kim JH, Park HJ, Lee DJ (2012). Anthocyanin changes in the Korean purple-fleshed sweet potato, Shinzami, as affected by steaming and baking. *Food Chem.*, **130**:966–972.

Kohlein F (1991). *Gentians*. Portland, Oregon, USA: Timber Press Inc.; pp. 25–27.

Koshioka M, Umegaki N, Boontiang K, Pornchuti W, Thammasiri K, Yamaguchi S, Tatsuzawa F, Nakayama M, Tateishi A, Kubota S. (2015). Anthocyanins in the bracts of *Curcuma* species and relationship of the species based on anthocyanin composition. *Nat. Prod. Commun.*, **10**:453–456.

Lai YS, Shimoyamada Y, Nakayama M, Yamagishi M (2012). Pigment accumulation and transcription of LhMYB12 and anthocyanin biosynthesis genes during flower development in the Asiatic hybrid lily (*Lilium* spp.). *Plant Sci.*, **193–194**:136–147.

Lou Q, Wang L, Liu H, Liu Y (2017). Anthocyanin profiles in flowers of *Grape hyacinth*. *Molecules*, **22**.pii: E688.

Lu TS, Saito N, Yokoi M, Shigihara A, Honda T (1992). Acylated pelargonidin glycosides in the red-purple flowers of *Pharbitis nil*. *Phytochem.*, **31**:289–295.

Maloney GS, DiNapoli KT, Muday GK (2014). The anthocyanin reduced tomato mutant demonstrates the role of flavonols in tomato lateral root and root hair development. *Plant Physiol.*, **166**:614–631.

Martin C, Gerats T (1993). Control of pigment biosynthesis genes during petal development. *Plant Cell*, **5**:1253–1264.

Matsuba Y, Sasaki N, Tera M, Okamura M, Abe Y, Okamoto E, Nakamura H, Funabashi H, Takatsu M, Saito M, Matsuoka H, Nagasawa K, Ozeki Y (2010). A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent glucosyltransferase in the petals of carnation and delphinium. *Plant Cell*, **22**:3374–3389.

Mitchell KA, Markhama KR, Boase MR (1998). Pigment chemistry and colour of *Pelargonium* flowers. *Phytochem.*, **47**:355–361.

Mu J, Yang Y, Luo Y, Su R, Niklas KJ (2017). Pollinator preference and pollen viability mediated by flower color synergistically determine seed set in an Alpine annual herb. *Ecol. Evol.*, **7**:2947–2955.

Nakatsuka T, Nishihara M, Mishiba K, Yamamura S (2005). Temporal expression of flavonoid biosynthesis-related genes regulates flower pigmentation in gentian plants. *Plant Sci.*, **168**:1309–1318.

Nakatsuka T, Nishihara M, Mishiba K, Hirano H, Yamamura S (2006). Two different transposable elements inserted in flavonoid 3',5'-hydroxylase gene contribute to pink flower coloration in *Gentiana scabra*. *Mol. Genet. Genomics*, **275**:231–241.

Nakatsuka T, Nishihara M, Mishiba K, Yamamura S (2005). Two different mutations are involved in the formation of white-flowered gentian plants. *Plant Sci.*, **169**:949–958.

Nakatsuka T, Sato K, Takahashi H, Yamamura S, Nishihara M (2008). Cloning and characterization of the UDP-glucose:anthocyanin 5-O-glucosyltransferase gene from blue-flowered gentian. *J. Exp. Bot.*, **59**:1241–1252.

Ohno S, Hosokawa M, Hoshino A, Kitamura Y, Morita Y, Park KI, Nakashima A, Deguchi A, Tatsuzawa F, Doi M, Iida S, Yazawa S. (2011). A bHLH transcription factor, DvIVS, is involved in regulation of anthocyanin synthesis in dahlia (*Dahlia variabilis*). *J. Exp. Bot.*, **62**:5105–5116.

Rambla JL, Trapero-Mozos A, Diretto G, Rubio-Moraga A, Granell A, Gomez-Gomez L, Ahrazem O (2016). Gene-Metabolite Networks of Volatile metabolism in airen and tempranillo grape cultivars revealed a distinct mechanism of aroma bouquet production. *Front. Plant Sci.*, **7**:1619.

Renobales G (2003). Notas acerca del tratamiento de las gentianaceae para flora ibérica. *Anales del Jardín Botánico de Madrid*. **60**:461–469.

Shi MZ, Xie DY (2014). Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat. Biotechnol.*, **8**:47–60.

Rodriguez-Saona LE, Giusti MM, Wrolstad RE (1998). Anthocyanin pigment composition of red-fleshed potatoes. *J. Food Sci.*, **63**:458–465.

Saito N, Tatsuzawa F, Hongo A, Win KW, Yokoi M, Shigihara A, Honda T (1996). Acylated pelargonidin 3-sambubioside-5-glucosides in *Matthiola incana*. *Phytochem.*, **41**:1613–1620.

Saito N, Tatsuzawa F, Yokoi M, Kasahara K, Iida S, Shigihara A, Honda T (1996). Acylated pelargonidin glycosides in red-purple flowers of *Ipomoea purpurea*. *Phytochem.*, **43**:1365–1370.

Sasaki N, Nishizaki Y, Ozeki Y, Miyahara T (2014). The role of acyl-glucose in anthocyanin modifications. *Molecules*, **19**:18747–18766.

Sasaki N, Nakayama T (2015). Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant Cell Physiol.* **56**:28–40.

Shim YS, Yoon WJ, Kim DM, Watanabe M, Park HC, Jang HW, Lee J, Ha J (2015). The simple determination method for anthocyanidin aglycones in fruits using ultra-high-performance liquid chromatography. *J. Chromatogr. Sci.*, **53**:1646–1653.

- Sulli M, Mandolino G, Sturaro M, Onofri C, Diretto G, Parisi B, Giovanni G (2017). Molecular and biochemical characterization of a potato collection with contrasting tuber carotenoid content. *PLoS One*, **12**:e0184143.
- Sun J, Lin LZ, Chen P (2012). Study of the mass spectrometric behaviors of anthocyanins in negative ionization mode and its applications for characterization of anthocyanins and non-anthocyanin polyphenols. *Rapid Commun Mass Spectrom.*, **26**:1123-1133.
- Tanaka Y, Sasaki N, Ohmiya A (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.*, **54**:733–749.
- Toki K, Saito N, Terahara N, Honda T (1995). Pelargonidin 3-glucoside-5-acetylglucoside in *Verbena* flowers. *Phytochem.*, **40**:939–940.
- Wu X, Prior RL (2005). Systematic identification and characterization of anthocyanins by HPLC-ESI-MS/MS in common foods in the United States: fruits and berries. *J. Agric. Food Chem.*, **53**:2589–2599.
- Yabuya T, Yamaguchi M, Imayama T, Katoh K, Ino I (2002). Anthocyanin 5-O-glucosyltransferase in flowers of *Iris ensata*. *Plant Sci.*, **162**:779–784.
- Yamagishi M, Shimoyamada Y, Nakatsuka T, Masuda K (2010). Two *R2R3-MYB* genes, homologs of *Petunia AN2*, regulate anthocyanin biosyntheses in flower Tepals, tepal spots and leaves of asiatic hybrid lily. *Plant Cell. Physiol.*, **51**:463–474.
- Yildirim S, Kadioglu A, Saglam A, Yasar A, Sellitepe HE (2016). Fast determination of anthocyanins and free pelargonidin in fruits, fruit juices, and fruit wines by high-performance liquid chromatography using a core-shell column. *J. Sep. Sci.* **39**:3927–3935.
- Yonekura-Sakakibara K, Nakayama T, Yamazaki M, Saito K (2009). Modification and stabilization of anthocyanins. In: Gould K, Davies K, Winefield C, editors. *Anthocyanins: Biosynthesis, functions and applications*. New York; Springer. pp. 169–185.
- Yoshida K, Mori M, Kondo T (2009). Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat. Prod. Rep.*, **26**:884–915.

Zhang Y, Butelli E, Martin C (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant. Biol.*, **19**:81–90.

Zhao D, Jiang Y, Ning C, Meng J, Lin S, Ding W, Tao J (2014). Transcriptome sequencing of a chimaera reveals coordinated expression of anthocyanin biosynthetic genes mediating yellow formation in herbaceous peony (*Paeonia lactiflora* Pall.). *BMC Genomics*, **15**:689.

Chapter VI
GENERAL DISCUSSION

6.1. General discussion

The biosynthesis and accumulation of secondary metabolites have always been a prime focus of researches because of their beneficial effects on human health and also industrial usage, and they are strictly controlled in a spatial and temporal manner and also influenced by biotic and abiotic factors (Patra *et al.*, 2013). A number of genes related to the biosynthesis of secondary metabolites have been identified and used to obtain engineered plants (Patra *et al.*, 2013). For example, the carotenoid biosynthetic pathway was engineered by overexpression of several genes that controlling flux into this pathway (Patra *et al.*, 2013; Giuliano, 2014, 2017), overexpression of structural carotenoid biosynthetic pathway genes showed that there is a correlation between transcripts and carotenoid content or composition (Ye *et al.*, 2000; Giuliano, 2017). Zhu *et al.* (2008) expressed *Zmpsyl* in white maize, led to a 53-fold increase of total carotenoids, and suggested that PSY1 is the key enzyme limiting carotenoid accumulation in maize endosperm of white maize. The total carotenoid content and composition in rice callus showed huge differences after expressing *ZmPSY1*, *PaCRTI*, *AtDXS* or all genes simultaneously in rice (Bai *et al.*, 2014). It is very well known, that both tissue specific-gene expression and gene activity were determined by gene transcription (Latchman, 1997). Gene transcription can be regulated at different levels, it occurs at post-translation level in some cases, in most cases happens at transcription level (Latchman, 1997). And transcription factors have been demonstrated to regulate the transcription of the gene either positively or negatively at the transcriptional level in many pathways (Latchman, 1997; Alves *et al.*, 2013). In my research I focused on the regulation of *ZmBCH2*, the gene controls the accumulation of β -carotene in maize endosperm, and confirmed that the genes for both transcription factors (*ZmPBF* and *ZmGAMYB*) are preferentially transcribed in maize endosperm and regulated developmentally in a similar manner to *ZmBCH2* itself. The transient expression experiments in maize and analysis of transgenic rice, which expressed *ZmPBF* and/or *ZmGAMYB* upregulated the *ZmBCH2* gene expression. In Arabidopsis, the genes related to flavonoids biosynthesis (such as *CHS*, *CHI* and *F3H*), are positively regulated by three functionally MYB TFs (MYB11, MYB12 and MYB111) (Patra *et al.*, 2013). In tobacco bHLH, R2R3 MYB TFs, NtAN1 and NtAN2 transcription factors were found to regulate the anthocyanin accumulation respectively in flowers (Pattanaik *et al.*, 2010; Schwinn *et al.*, 2006). WRKY transcription factor family are also has been

demonstrated to regulate secondary metabolite biosynthesis genes in different types of higher plants (Kato *et al.*, 2007; Ma *et al.*, 2009; Li *et al.*, 2013; Mao *et al.*, 2011; Suttipanta *et al.*, 2011). In *C. roseus*, overexpression of octadecanoid-responsive Catharanthus AP2-domain transcription factors (ORCAs) showed two-fold increase of alkaloid production in transgenic plants (van der Fits, 2000). In tobacco, MAPK phosphorylates transcription factors were found to upregulate the expression of genes which play a critical role in the biosynthesis of secondary metabolites such as phytoalexins (Yang *et al.*, 2001). Another study demonstrated that, phytochrome interacting transcription factors (PIFs) that bind to G-boxes and down regulate the expression of light- induced genes (Leivar *et al.*, 2009). This transcription factors also have been shown to have an important role in regulating the transcription of PSY and other carotenoid and chlorophyll biosynthesis genes during light-induced de-etiolation (Toledo-Ortiz *et al.*, 2010).

Light has been confirmed to be an important abiotic factor which regulates the secondary metabolites biosynthesis and the expression of related genes at different levels. In higher plants, light regulates photoreceptors and their activity and the regulation happens mainly at the post-translational level (Jiao *et al.*, 2007). In Arabidopsis, various regulators downstream of photoreceptors have been identified, including transcription factors, kinases, phosphatases and degradation-pathway proteins (Chen *et al.*, 2004). In both Arabidopsis and rice, at least 20% of the genome showed differential expression between seedlings under light-grow and that are under dark growth (Ma *et al.*, 2001; Tepperman *et al.*, 2001; Jiao *et al.*, 2005). Anthocyanin biosynthesis has been confirmed regulated by light, in Arabidopsis, MYBL2 has been proposed to play a key regulatory role in the induction of anthocyanin accumulation by high light (Teng *et al.*, 2005; Loreti *et al.*, 2008). Rodríguez-Concepción *et al.* (2004) reported that in Arabidopsis environmental light conditions are sensed and transduced by distinct signaling pathways that coordinate the metabolism of isoprenoids. Also, I found the dynamic changes of metabolites of rice leaves during de-etiolation mirrored the expression levels of the MVA and/or MEP pathway and the related genes, these results suggested the secondary metabolites biosynthesis were affected by light via regulated the expression of involved genes.

I also identified pelargonidin derivatives from the petals of *G. lutea* L. var *aurantiaca* and profiled the expression levels of the *DFR*, *ANS*, *3GT*, *5GT* and *5AT* genes, results showed DFR most probably acting as the limiting step in the accumulation of these anthocyanins in orange-flower gentian. In *Torenia hybrida*, additional expression of rose dihydroflavonol 4-reductase (DFR) gene elevated the level of pelargonidin and yielded darker pink petals. Expression of pelargonidin DFR gene instead of the rose DFR gene increased the level of pelargonidin and darkened the petal color. Introducing the two genetic constructs containing a DFR gene into a violet torenia cultivar, which had more anthocyanins and a darker color than the blue one, further increased the level of pelargonidin and pink color intensity. Downregulation of the *F3'5'H* and *F3'H* genes is insufficient and that addition of a suitable the proper *DFR* gene is necessary to generate a metabolic flux toward pelargonidin (Nakamura *et al.*, 2010).

6.2. References

Alves M S, Dadalto S P, Gonçalves A B, De Souza G B, Barros V A, Fietto L G (2013) Plant bZIP Transcription Factors Responsive to Pathogens: A Review. *Int. J. Mol. Sci.*, **14**: 7815-7828.

Bai C, Rivera S M., Medina V, Alves R, Vilaprinyo E, Sorribas A, Canela R, Capell T, Sandmann G, Christou P and Zhu C (2014) An *in vitro* system for the rapid functional characterization of genes involved in carotenoid biosynthesis and accumulation. *The Plant Journal*, **77**: 464–475.

Chen M, Chory J and Fankhauser C (2004) Light signal transduction in higher plants. *Annu. Rev. Genet.* **38**, 87–117.

Giuliano G (2014) Plant carotenoids: genomics meets multi-gene engineering. *Curr Opin Plant Biol*, **19**: 111–117.

Giuliano G (2017) Provitamin A biofortification of crop plants: a gold rush with many miners. *Curr Opin Biotechnol*, **44**: 169–180

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES (2008) Natural

genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, **319**: 330-333.

Jiao Y, Lau OS, Deng X W (2007) Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics*, **8**:217-230.

Jiao Y, Ma L, Strickland E, Deng X W (2005) Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. *Plant Cell*, **17**: 3239–3256.

Kato N, Dubouzet E, Kokabu Y, Yoshida S, Taniguchi Y, Dubouzet JG, Yazaki K, Sato F (2007). Identification of a WRKY protein as a transcriptional regulator of benzyloquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol.*, **48**:8–18.

Latchman D (1997). Transcription Factors: An Overview. *Int. J Biochem. Cell Biol.*, **29**: 1305-1312.

Leivar P, Tepperman JM, Monte E, Calderon RH, Liu TL, Quail PH (2009) Definition of early transcriptional circuitry involved in light-induced reversal of PIF- imposed repression of photomorphogenesis in young Arabidopsis seedlings. *Plant Cell*, **21**:3535-3553.

Li S, Zhang P, Zhang M, Fu C, Yu L (2013). Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in *Taxus chinensis*. *Plant Biol.*, **15**:19–26.

Loreti E, Povero G, Novi G, Solfanelli C, Alpi A, Perata P (2008). Gibberellins, jasmonate and abscisic acid modulate the sucrose-induced expression of anthocyanin biosyn- thetic genes in Arabidopsis. *New Phytol.*, **179**:1004–1016.

Ma D, Pu G, Lei C, Ma L, Wang H, Guo Y, Chen J, Du Z, Wang H, Li G, Ye H, Liu B (2009). Isolation and characterization of AaWRKY1, an *Artemisia annua* transcription fac- tor that regulates the amorpha-4,11-diene synthase gene, a key gene of artemisinin biosynthesis. *Plant Cell Physiol.*, **50**:2146–2161.

Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H, Deng XW (2001) Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell*, **13**: 2589–2607.

Mao G, Meng X, Liu Y, Zheng Z, Chen Z, Zhang S (2011). Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin bio- synthesis in Arabidopsis. *Plant Cell*, **23**:1639–1653.

Nakamura N, Fukuchi-Mizutani M, Fukui Y, Ishiguro K, Suzuki K, Suzukic H, Okazaki K, Daisuke S, Tanaka Y (2010) Generation of pink flower varieties from blue *Torenia hybrida* by redirection of its flavonoid biosynthetic pathway from delphinidin to pelargonidin. *Plant Biotechnol.*, **27**:375–383.

Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Ling Y (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochimica et Biophysica Acta*, **1829**: 1236–1247.

Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochimica et Biophysica Acta*, **1829**(11):1236–1247.

Pattanaik S, Kong Q, Zaitlin D, Werkman JR, Xie CH, Patra B, Yuan L (2010). Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco, *Planta*, **231**:1061–1076.

Rodríguez-Concepción M, Forés O, Martínez-García JF, Víctor González, 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**(1):144–156.

Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell*, **18**:831–851.

Suttipanta N, Pattanaik S, Kulshrestha M, Patra B, Singh SK, Yuan L (2011) The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.*, **157**: 2081–2093.

Teng S, Keurentjes J, Bentsink L, Koornneef M, Smekens S (2005). Sucrose-specific induction of anthocyanin biosynthesis in *Arabidopsis* requires the MYB75/PAP1 gene. *Plant Physiol.*, **139**:1840–1852.

Tepperman J M, Zhu T, Chang H S, Wang X, Quail P H (2001) Multiple transcription-factor genes are early targets of phytochrome A signaling. *Proc. Natl Acad. Sci. USA*, **98**: 9437–9442.

Toledo-Ortiz G, Huq E, Rodriguez-Concepcion M (2010) Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proc. Natl. Acad. Sci. USA*, **107**:11626-11631.

Vallabhaneni R, Wurtzel ET (2009) Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. *Plant Physiol*, **150**: 562–572.

Van der Fits L (2000) Transcriptional regulation of stress-induced plant secondary metabolism. PhD thesis, University of Leiden, The Netherlands.

Yang K, Liu Y, Zhang S (2001) Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco, *Proc. Natl. Acad. Sci. U.S.A.*, **16**:741–746.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**: 303–305.

Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, and Capella T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *PNAS*, **105**: 18232–18237.

GENERAL CONCLUSIONS

General conclusions

1. Maize β -carotene hydroxylase 2 (ZmBCH2) is regulated by two *cis*-regulatory elements (AACAA motif and P-box) in the upstream promoter and the 5'-untranslated region (5'-UTR).
2. The ZmPBF and ZmGAMYB transcription factors act independently *in planta* to bind P-box and AACAA motifs, respectively, and transactivate the *ZmBCH2* promoter.
3. Both rice *IPPI* paralogs encode functional enzymes, which show distinct expression profiles.
4. The expression profile of *OsIPPI1* in de-etiolated leaves mirrors the accumulation of phytosterols, suggesting a key role in the synthesis of these MVA pathway products.
5. *OsIPPI1* is localized in mitochondria, peroxisomes and the ER, whereas, *OsIPPI2* is localized not only in these compartments but also in the plastids, suggesting distinct functions of these two enzymes.
6. The expression profiles of genes involved in isoprenoid biosynthesis in de-etiolated rice leaves mirrored the accumulation of the corresponding metabolites. Furthermore, isoprenoid levels (metabolites) and mRNA levels of the corresponding genes were positively correlated and increased with illumination time.
7. Eleven pelargonidin derivatives were identified in the petals of *G. lutea* L. var. *aurantiaca*. These derivatives accumulated at different levels at each developmental stage.
8. Gene expression analysis showed that *DFR*, *ANS* and *3GT* parallel the accumulation of pelargonidin glucosides, suggesting a major role in the accumulation of these anthocyanins in *G. lutea* L. var. *aurantiaca* flowers, with *DFR* most probably acting as the limiting step in the anthocyanin biosynthetic pathway.

OUTPUTS

Outputs

- Publications

Zhu C, **Jin X**, Baysal C, Gao L, Medina V, Drapal M, Ni X, Sheng Y, Shi L, Capell T, Fraser P, Christou P (2019) The subcellular localization of two isopentenyl diphosphate isomerases in rice suggests a role for the endoplasmic reticulum in isoprenoid biosynthesis (submitted to *The Plant Journal*)

Diretto G, **Jin X**, Capell C, Zhu C, Gomez-Gomez L (2019). Differential accumulation of pelargonidin glycosides in petals at three different developmental stages of the orange-flowered gentian (*Gentiana lutea* L. var. *aurantiaca*). *PLoSOne*, **14**(2):e0212062. DOI: 10.1371/ journal.pone.0212062

Jin X, Bai C, Bassie L, Nogareda C, Romagosa I, Twyman RM, Christo P, Zhu C (2018). ZmPBF and ZmGAMYB transcription factors independently transactivate the promoter of the maize (*Zea mays*) β -carotene hydroxylase 2 gene. *New Phytol.*, **222**(2):793-804.

Bortesi L, Zhu C, Zischewski J, Perez L, Bassié L, Nadi R, Forni G, Lade SB, Soto E, **Jin X**, Medina V, Villorbina G, Muñoz P, Farré G, Fischer R, Twyman RM, Capell T, Christou P, Schillberg S (2016). Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol. J.*, **14**(12):2203-2216.

- Participation in congresses

Oral presentations

Jin X, Bai C, Berman J, Farre G, Capell T, Christou P, Sandmann G, Zhu C. (2017). Regulation of β -carotene biosynthesis and accumulation in maize endosperm. EUROCAROTEN *Workshop on Sustainable Production of Carotenoids - From Biosynthesis to Biotechnology*, 16-17 October, Trogir (Croatia).

Jin X, Bai C, Berman J, Farre G, Capell T, Christou P, Sandmann G, Zhu C. (2017). Regulation of maize (*Zea mays* L.) β -carotene hydroxylase 2 (*ZmBCH2*) gene expression “*Reunión Nacional sobre CAROTENOIDES en Microorganismos, Plantas, Alimentación y Salud*”, November, Valencia (Spain)

- Conferences

Jin X (2018). Regulation of β -carotene biosynthesis and accumulation in maize endosperm. Departmental Seminar. University of Lleida.

