UAB

Universitat Autònoma de Barcelona

UNIVERSITAT AUTÒNOMA DE BARCELONA

FACULTAT DE BIOCIÈNCIES

Molecular mechanisms of plastidial differentiation

SALVADOR TORRES MONTILLA

2021

UNIVERSITAT AUTÒNOMA DE BARCELONA

FACULTAT DE BIOCIÈNCIES

PROGRAMA DE DOCTORAT EN BIOLOGIA I
BIOTECNOLOGIA VEGETAL

Molecular mechanisms of plastidial differentiation

Memòria presentada per Salvador Torres Montilla per optar al títol de doctor per la
Universitat Autònoma de Barcelona

Dr. Manuel Rodríguez Concepción
Director de la tesi

Salvador Torres Montilla
Doctorand

Dra. Roser Tolrà Pérez
Tutora

Salvador Torres Montilla (2021)

Agradecimientos

Esta tesis doctoral ha sido un viaje personal de más de cinco años, que empezó con mi marcha de mi querida Granada hacia Barcelona, de la que me he enamorado, y que termina en Valencia. En esta aventura me comparo con mi yo que empezó y me doy cuenta de lo mucho que he crecido científicamente, pero nada comparado el desarrollo personal, condicionado por toda la gente que me ha acompañado en el camino.

En primer lugar, quiero agradecer a Manuel. Desde el momento en que me recibiste para aquella entrevista en mi máster, me has dado unas oportunidades enormes y me has enseñado mucho del mundo científico, su rigor y su funcionamiento. Pero, sobre todo, quiero agradecerte el haberme escuchado cuando más falta hacía, y el haberme apoyado cuando más lo he necesitado. Gracias también a Briardo, que me enseñaste a trabajar en el laboratorio y que a tu lado arranqué este proyecto. Gracias a Ari, a Di, a Luca, a Laura, a Edison, a Jordi, y a todas las personas que han formado parte del 2.01 en estos años y que hemos podido ser compañeros. Gracias a todos los amigos y amigas que he podido hacer en el CRAG desde que llegué. Me habéis hecho sentir muy cómodo y feliz allí, y solo me llevo buenos recuerdos. Gracias también Jordi, María, Lucio, Vero, Mari Carmen, Beltrán, Bea, Buri, y todas las personas que me habéis recibido en Valencia.

Miguelón, Vicky, Miguel y Neto, gracias de corazón. Las risas, los buenos momentos, los malos, todo lo que me llevo de estos años a vuestro lado no tiene precio. Desde el día que os conocí y me acogisteis a vuestro lado estoy en deuda con vosotros. Allá donde nos lleve la vida, esperadme que iré. Rosa, Veni, Mireia y Álex, sois increíbles. Gracias por haberme hecho sentir siempre como uno más de la familia. Vuestro apoyo, cariño y alegría me los guardo con especial cariño.

Evi, poulaki mou. Thank you so much for everything. Your love and support have always been there, your friendship is saved in my heart.

Gracias a Tomás, que me has recibido con los brazos abiertos. Las risas en casa, las tonterías en la cocina, los bailes con gatico... Gracias de corazón.

Thank you Silvia, James, Rosi and Andrea. The arrival to Milano was not convectional at all, but you made my life easy and I only have good memories from there thanks to you. Especialmente a Aure. Desde que me respondiste a aquel email sin necesidad ninguna, solamente me has ayudado. Gracias por enseñarme, gracias por

y en la que he sido tremendamente feliz. A todos y todas los que habéis formado parte de ella, gracias, gracias y gracias. Os quiero mucho.

Summary

Carotenoids are isoprenoids produced by all photosynthetic and some non-photosynthetic organisms. They have functions related to photosynthesis, photo-protection, pigmentation and signaling. Despite humans cannot synthesize them, carotenoids provide important nutritional and health-promoting properties, mainly as vitamin A precursors. Plants synthesize carotenoids in plastids from geranylgeranyl diphosphate (GGPP) synthesized by the methylerythritol 4-phosphate (MEP) pathway. Chromoplasts are plastids specialized in carotenoid production and accumulation which are usually differentiated from pre-existing chloroplasts, but the differentiation mechanism remains unknown. Chloroplast-to-chromoplast transition only occurs in some organs of some plant species, normally in parallel to many other processes such as those related to fruit ripening. The goal of this thesis work has been to characterize an artificial system of chloroplast-to-chromoplast differentiation in leaves of *Nicotiana benthamiana*, triggered by the transient expression of the bacterial *crtB* gene which encode a phytoene synthase, the first step of the carotenoid biosynthetic pathway.

To better characterize our artificial system, we initially focused on *N. benthamiana* leaves four days after the agroinfiltration with *crtB* (once a yellow phenotype was fully established), and compared them to regions of the same leaves expressing *GFP* as a control. *crtB*-expressing leaves accumulated phytoene, increased total carotenoid content, stopped their photosynthetic activity and changed the morphology of their chloroplats to chromoplast-like plastids. Strikingly, this phenotype was only induced when crtB localized in plastids, producing phytoene from GGPP derived from the MEP pathway. RNA-seq at 96 hours post-infiltration (hpi) showed global transcription profiles with similarities to the tomato fruit ripening system (where chloroplasts naturally differentiate into chromoplasts) but not to the leaf senescence process in *Arabidopsis thaliana* (where chloroplasts degenerate to become gerontoplasts). To support our results, a new genome of *N. benthamiana* was annotated, using *Arabidopsis* genes as reference and identifying homologs for different gene families.

The second part of the thesis was focused on describing gene expression events during chloroplast-to-chromoplast transition. A second RNA-seq experiment was carried out encompassing eight time-points, from 22 to 56 hpi. Analysis of these data using the newly annotated *N. benthamiana* genome and comparison of the results with RNA-seq experiments covering chloroplast-to-chromoplast differentiation in

fruits showed that this process is very heterogenous among different plant systems, sharing just a small proportion of their gene expression profiles. In *N. benthamiana*, two peaks were found in gene expression changes that correlated with drops in photosynthetic activity: one at the beginning and another at the end of the time course, with a relaxation in the middle. A first event of down-regulation affected primary metabolism and chaperone gene expression, probably caused by GGPP consumption or as a response to crtB presence. This down-regulation was followed by an up-regulation correlated to the first event of phytoene accumulation and photosynthesis decrease that affected MEP pathway and jasmonic acid (JA) related gene expression. The final peak of gene expression changes occurred simultaneously with the second and definite event of photosynthesis reduction and total carotenoid increase. JA hormone levels increased at this last event, while carotenoid biosynthesis genes were not affected during the investigated time-course. Besides contributing to a better understanding of chromoplastogenesis, these results provide *in-silico* curated annotations for future studies in *N. benthamiana*.

Resumen

Los carotenoides son isoprenoides que tanto los organismos fotosintéticos como algunos no fotosintéticos producen. Tienen funciones relacionadas con la fotosíntesis, la fotoprotección, la señalización y la pigmentación. A pesar de que los humanos no podemos sintetizarlos, los carotenoides proporcionan propiedades nutricionales muy favorables para nuestra salud, principalmente por su papel como precursores de vitamina A. Las plantas sintetizan carotenoides en los plastos a partir del geranilgeranil difosfato (GGPP) que se sintetiza a partir de la ruta del metileritritol 4-fosfato (MEP). Los cromoplastos son plastos especializados en la producción y acumulación de carotenoides que normalmente se originan por la diferenciación a partir de cloroplastos, pero el mecanismo de diferenciación no ha sido identificado. La transición cloroplasto-cromoplasto solo ocurre en algunos órganos de algunas especies, normalmente en paralelo con otros muchos procesos como todos aquellos relacionados con la maduración del fruto. El objetivo de esta tesis doctoral ha sido la caracterización de un sistema artificial de diferenciación cloroplasto-cromoplasto en hojas de *Nicotiana benthamiana*, desencadenado por la expresión transitoria del gen bacteriano *crtB*, que codifica una fitoeno sintasa, el primer paso de la ruta biosintética de carotenoides.

Para una mejor caracterización de nuestro sistema artificial, inicialmente nos enfocamos en hojas de *N. benthamiana* cuatro días después de la agroinfiltración con *crtB* (una vez que un fenotipo amarillo estaba plenamente establecido), y las comparamos con regiones de las mismas hojas expresando *GFP* como control de agroinfiltración. Las regiones que expresaban *crtB* acumulaban fitoeno, aumentaban el contenido de carotenoides totales, detenían su actividad fotosintética y cambiaban la morfología de sus plastos hacia una similar a cromoplastos. De hecho, este fenotipo solamente se inducía cuando crtB se localizaba en los plastos, produciendo fitoeno a partir del GGPP derivado de la ruta del MEP. El análisis de RNA-seq a las 96 horas post-infiltración (hpi) mostró unos perfiles de transcripción globales con similaridades con el sistema de maduración de fruto en tomate (en el que los cloroplastos se diferencian en cromoplastos de forma natural) pero no con el proceso de senescencia en hojas de *Arabidopsis thaliana* (donde los cromoplastos degeneran a gerontoplastos). Para respaldar nuestros resultados, anotamos un nuevo genoma de *N. benthamiana*, utilizando genes de *Arabidopsis* como referencia e identificando homólogos para diferentes familias génicas.

La segunda pare de la tesis estuvo enfocada en describir los eventos de expresión génica durante la transición cloroplasto-cromoplasto. Un segundo experimento de RNA-seq se llevó a cabo, abarcando ocho puntos desde las 22 a las 56 hpi. Los análisis de estos datos usando el nuevo genoma anotado de *N. benthamiana* y la comparación de los resultados con experimentos de RNA-seq enfocados en la diferenciación cloroplasto-cromoplasto en frutos mostró se trata de un proceso muy heterogéneo entre distintas especies vegetales, compartiendo solamente un pequeño porcentaje de sus perfiles de expresión génica. En *N. benthamiana*, se encontraron dos picos de cambios de expresión que correlacionan con las caídas en la actividad fotosintética: uno al principio y otro al final del proceso, con una etapa de relajación entre medias. Se observó un primer evento de represión en la expresión que afectaba a la expresión génica relacionada con el metabolismo primario y las chaperonas, probablemente causado por el consumo de GGPP o por una respuesta ante la presencia de crtB. Esta represión fue seguida de una inducción en la expresión correlacionada con la acumulación de fitoeno y el descenso de la fotosíntesis, que afectaba a la expresión de genes de la ruta del MEP y genes relacionados con el ácido jasmónico (JA). El pico final de cambio en la expresión génica ocurrió simultáneamente con el segundo y definitivo evento de reducción fotosintética y el incremento de carotenoides totales. Los niveles hormonales de JA incrementaron en este último evento, mientras que los genes de la biosíntesis de carotenoides no estuvieron afectados durante el experimento. Además de contribuir a un mejor entendimiento de la cromoplastogénesis, estos resultados proveen de anotaciones realizadas *in-silico* para futuros estudios en *N. benthamiana*.

Preface

This thesis has been written in LaTeX using the Master/Doctoral Thesis template with some modifications, and Overleaf platform. All sections, figures, tables and references in this document have a link that drives you to the content in the PDF when you click. Due to the big amount of files that were generated during this work, I have created a public repository in GitHub (`https://github.com/salvillas/PhD_TorresMontilla`), including tables with the lists of genes and GO terms from the different analyses, as well as representative scripts that I have used to analyze data and create plots in RStudio. Most of the figures were modified using Adobe Illustrator and/or Adobe Photoshop. Also, BioRender platform was used to create some of the figures (mainly the models). A palette from green (e.g., #1b9e77, #66c2a5) to orange (e.g., #d95f02, #fc8d62) colors has been used to illustrate comparisons control vs. treatment, respectively, all over this work. Websites as `https://colorbrewer2.org/` and `https://gka.github.io/palettes/#/5|d|1b9e77,8ecdb6,ffffbf|ffffbf,fc8d62,d95f02|1|1` were used to generate them, taking special attention to be color blind friendly.

Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Abbreviations

ABA	Abscisic acid
BP	Biological process
CC	Cell component
CCDs	Carotenoid cleavage dioxygenases
CCPs	Carotenoid cleavage products
CHRC	Chromoplast-specific carotenoid-associated protein
CHRD	Chromoplast protein D
DEGs	Differentially expressed genes
DMAPP	Dimethylallyl diphosphate
DNA	Deoxyribonucleic acid
dpi	Days post infiltration
ER	Endoplasmic reticulum
FBN	Fibrillin
FC	Fold change
FPKM	Fragments per kilobase per million mapped reads
FPP	Farnesyl diphosphate
GFP	Green fluorescent protein
GGPP	Geranylgeranyl diphosphate
GO	Gene ontology
hpi	Hours post infiltration
HPLC	High-performance liquid chromatography
HSP	Heat shock protein
IPP	Isopentenyl diphosphate
JA	Jasmonic acid
LOXs	Lipoxygenases
MEP	Methylerythritol 4-phosphate
MF	Molecular function
MG	Mature green
MVA	Mevalonate
OR	Orange protein / Orange stage
PAGE	Parametric analysis of gene set enrichment
PAM	Pulse-amplitude modulation
PCA	Principal component analysis
PK	Protein kinase

PLBs	Prolamellar bodies
PQC	Protein quality control
PSII	Photosystem II
ϕPSII	Effective quantum yields of photosystem II
PSY	Phytoene synthase
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
RR	Red ripe
SNP	Single nucleotide polymorphism
SP1	*SUPRESSOR OF ppi1*
SPL2	SP1-like2
TEM	Transmission electron microscopy
TF	Transcription factor
TPM	Transcript per kilobase million
TR	Transcription regulator
TIC	Translocon at the outer chloroplast envelope
TOC	Translocon at the inner chloroplast envelope

Introduction

1.1 Introduction to carotenoids

Carotenoids are isopresnoid compounds produced by all photosynthetic organisms (including plants, algae and cyanobacteria) as well as some non-photosynthetic archaea, bacteria, fungi and animals (Rodriguez-Concepcion et al., 2018). In plants, carotenoids can be considered as both primary (essential) and secondary (specialized) metabolites. Carotenoids, such as lutein, β-carotene, violaxanthin, and neoxanthin are required for photosynthesis and photoprotection (Domonkos et al., 2013; Hashimoto et al., 2016). The absence of carotenoids in plants, due to a genetic or chemical inhibition of the pathway, hence results in a non-viable albino phenotype (Pokhilko et al., 2015; Qin et al., 2007; Ruiz-Sola et al., 2016a). Also essential is their function as precursors of important phytohormones regulating plant development, growth or stress responses, such as abscisic acid (ABA) or strigolactones (Hou et al., 2016; Moreno et al., 2020). Carotenoids can also be considered specialized metabolites due to their role as communication signals between plants and their environment. Most carotenoids are pigments that provide distinctive colors to the fruits and flowers of some plants for attracting animals for seed dispersal and pollination (Llorente et al., 2017; Rodriguez-Concepcion et al., 2018). Addtionally, carotenoid cleavage can generate volatile compounds that contribute to the flavor of some flowers and fruits. These properties as natural pigments and aromas make carotenoids an important economic target of cosmetic and food industries. However, their main interest for humans is their nutritional and health-promoting properties. Dietary carotenoids are used as precursors for the production of retinoids (including vitamin A). Moreover, carotenoids-enriched diets reduce the risk of several degenerative diseases, including age-related macular degeneration, cognitive malfunctioning, cardiovascular diseases, and some types of cancer (Eggersdorfer and Wyss, 2018; Rodriguez-Concepcion et al., 2018). Due to the importance of carotenoids for human (and animal) health and their economic value for industry, a large number of biotechnological strategies have been developed to enrich plant tissues with carotenoids (Giuliano, 2017; Zheng et al., 2020). Plant carotenoid levels are the result of three interacting variables: (1) biosynthetic rate, (2) degradation rate and (3) storage capacity of the cell to accumulate and sequester them. Most biotechnological approaches to alter carotenoid contents in plants have been focused on

manipulating their biosynthesis or degradation by overexpressing or silencing sequences encoding exogenous or endogenous enzymes or their regulators (Giuliano, 2017; Nisar et al., 2015; Ruiz-Sola and Rodríguez-Concepción, 2012; Shumskaya and Wurtzel, 2013; Sun et al., 2017; Sun and Li, 2020; Zheng et al., 2020). Improving the sink capacity of plant cells to accumulate enhanced levels o carotenoids, however, has received much less attention, in part because of our limited knowledge on how to manipulate this process.

1.1.1   Chemistry of carotenoids

The main structural feature of carotenoids is the system of conjugated double bonds that forms the core of the carotenoid molecules, known as 'polyene chain' (Figure 1.1). Carotenoids with a partially saturated polyene chain (such as phytoene) do not absorb light in the visible wavelength. However, they are rarities within the carotenoid family, as desaturation steps early in the biosynthetic pathway readily convert them into downstream products with colors in the visible (yellow to red) spectrum. Carotenoids can be either acyclic or cyclic attending to the absence of presence of end rings in their structure, respectively. They are also classified as carotenes (when they only contain carbon and hydrogen) and xanthophylls (when they also contain oxygen). Carotenoids can also have *trans* (*E*) or *cis*(*Z*) isomers which can markedly differ in shape. Furthermore, some carotenoids form optical isomers due to the presence of chiral centers in their molecules (Britton, 1995; Rodriguez-Concepcion et al., 2018).

1.1.2   Biosynthetic pathway

Plants synthesize carotenoids in plastids from the isoprenoid precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), derived from the methylerythritol 4-phosphate (MEP) pathway (Phillips et al., 2008; Rodriguez-Concepcion et al., 2018; Ruiz-Sola and Rodríguez-Concepción, 2012). Three IPP and one DMAPP molecules are condensed by geranylgeranyl diphosphate (GGPP) synthase to form GGPP, which is the common precursor of the carotenoids and other photosynthesis-related plastidial isoprenoids such as chlorophylls, plastoquinone, phylloquinones, and tocopherols, but also of the plant hormones gibberellins (Ruiz-Sola et al., 2016a; Ruiz-Sola et al., 2016b; Zhou et al., 2017) (Figure 1.1). The first reaction specific of the plant carotenoid pathway is the condensation of two molecules of GGPP for the synthesis of the non-colored carotenoid 15-*cis*-phytoene, catalyzed by the enzyme phytoene synthase (PSY). This is considered a bottleneck reaction and a major control steep of the metabolic flux to carotenoids (Fraser et al., 2002; Ruiz-Sola and Rodríguez-Concepción, 2012). Different enzymes catalyze the sequential desaturation and isomerization reactions that convert 15-*cis*-phytoene into the red-colored all-*trans*-lycopene (Nisar et al., 2015; Rodriguez-Concepcion et al., 2018). Formation of β- or/and ϵ-ionone rings in the ends of the linear lycopene molecule represents a

branch point of the pathway leading to either α-carotene (β, ϵ branch) or β-carotene (β, β branch) (Nisar et al., 2015). Hydroxylation and epoxidation of the rings in these carotenes result in different types of xanthophylls, such as lutein from α-carotene or zeaxanthin, violaxanthin and neoxanthin from β-carotene (Figure 1.1).

1.1.3 Breakdown pathways

The electron-rich polyene backbone of carotenoids makes them very susceptible to oxidative breakdown. Carotenoid degradation can be mediated by carotenoid cleavage dioxygenases (CCDs) that catalyze the oxidative breakdown of carbon-carbon double bonds in different locations of the polyene backbone to produce carotenoid cleavage products (CCPs) containing carbonyl groups (aldehyde or ketone groups) in the cleaving ends. While carotenoid levels are negatively correlated with the activity of these enzymes in some plants and tissues, CCD-independent degradation can be a main contributor to carotenoid loss in other cases (Adami et al., 2013; Gonzalez-Jorge et al., 2013; Schaub et al., 2018). For example, lipoxygenases (LOXs) that oxidize polyunsaturated fatty acids can produce hyperperoxides that cause the degradation (i.e., cooxidation) of carotenoids (Chedea and Jisaka, 2013; Gao et al., 2019). Carotenoids can also be oxidized non-enzymatically by reactive oxygen species (ROS) produced after photooxidative stress (Havaux, 2014). The CCPs resulting from either CCD activity or non-enzymatic oxidation are typically referred to as apocarotenoids. Many of them have signaling roles as pigments and flavors, and some have been shown to function as hormones and defense compounds (Dhar et al., 2020; Moreno et al., 2020). The role of many other apocarotenoids, however, remains unknown.

1.1.4 Interaction with other molecules

Carotenoids can be found in plant tissues either free or associated with other molecules (proteins, fatty acids or sugars). They can also form aggregates as a result of weak and reversible bonding by hydrogen bonds, van der Waals interactions, dipole forces and the hydrophobic effects of hydrophobic molecules, their polar groups and the surrounding solven (Gruszecki, 1999; Popova and Andreeva, 2013; Simonyi et al., 2003). With only some exceptions (e.g., carotenoids harboring carboxylic groups such as bixin and azafrin), free carotenoids are hydrophobic compounds usually found in lipid-rich environments. Carotenes are typically more lipophilic than xanthophylls (Britton, 1995). The conjugated double bonds in their polyene chain causes rigidity of the carotenoid molecules. While *cis* isomers have an angular shape and are less susceptible to aggregation, those in all-*trans* configuration typically exhibit a rod-like shape (Figure 1.1). Besides solubility, these differences influence the ability of carotenoids to fit into cellular structures or to interact with enzymes and other proteins, eventually impacting their bioaccessibility (i.e., the amount of carotenoid released from the food matrix during digestion and

made available for absorption) and bioavailability (i.e., the fraction that is actually absorbed by our bodies and available for biological use).

Plant carotenoids can also be found associated to proteins (e.g., in photosynthetic complexes) or conjugated with sugar (e.g., glucose) or lipid (e.g., fatty acids) moieties. Carotenoid glycosylation is most frequent in microorganisms as a natural mechanism to increase their hydrophilicity (Wurtzel, 2019). Glycosylated carotenoids appear to play important roles in maintaining cell wall structure and stabilizing thylakoid membranes in cyanobacteria (Mohamed et al., 2005). In plants, the best-known examples are crocins and picrocrocin, water-soluble glycosylated apocarotenoids that contribute to the color and taste of the saffron (*Crocus sativus*) spice. They are produced from precursors resulting from the cleavage of zeaxanthin by a specific plastid-localized CCD enzyme (CCD2) that leave the plastid to be eventually glycosylated by identifies glycosyltransferases and then imported into vacuoles by specific ABC transporters (Demurtas et al., 2018; Demurtas et al., 2019; Diretto et al., 2019; López et al., 2021). While glycosylation increases water solubility and decreases reactivity, compartmentalization into vacuoles allows for long-term storage and higher stability (Demurtas et al., 2019). Glycosylation of xanthophyll-derived apocarotenoids also occurs in fruits and leaves of several plant species, presumably as a compensatory mechanism for increased carotenoid flux (Demurtas et al., 2018; Koschmieder et al., 2021; Lätari et al., 2015; Martí et al., 2020).
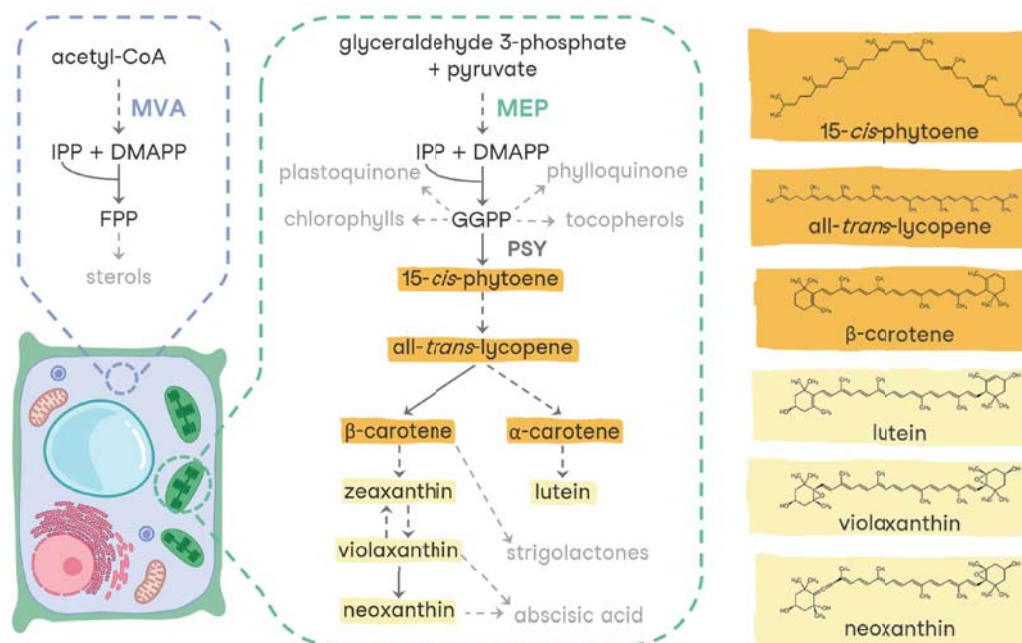
Figure 1.1: Isoprenoid biosynthesis in plants. IPP and DMAPP are produced in the cytosol from the mevalonate (MVA) pathway and in plastids from the MEP pathway. In the cytosol they are mainly used for sterols biosynthesis. In plastids, they are the precursors of many different isoprenoids, including GGPP-derived photosynthesis-related compounds such as carotenoids. Among carotenoids, carotenes and xanthophylls are boxed in orange and yellow, respectively.

The most frequent type of plant carotenoid modification is esterification with fatty acids (Britton, 1995; Rodriguez-Concepcion et al., 2018). Because the acylation reaction must necessarily be performed over a hydroxyl group, only hydroxy-xanthophylls may become acyl esters. They are ubiquitous in foods of plant origin (Mercadante et al., 2017). The ripening of many fruits is actually associated with the esterification of xanthophylls, supporting the conclusion that esterification is important to increase the carotenoid accumulation capacity of plant cells. While association with sugars makes carotenoids more water-soluble, esterification of xanthophylls renders them more lipophilic and promotes their sequestration in subplastidial structures (Van Wijk and Kessler, 2017). The increase in the liposolubility compared to free xanthophylls likely improves the integration of esterified xanthophylls into membrane structures such as plastoglobules, which eventually results in higher stability and enhanced bioavailability (Rodriguez-Concepcion et al., 2018; Watkins and Pogson, 2020).

1.2 Carotenoids in plastids

In nature, plant carotenoids are exclusively produced in plastids. While carotenoids are fairly sensitive to oxidation or isomerization by light and temperature when tissues are disrupted, they are relatively stable in their natural environment (i.e., inside plastids). In fact, the storage of carotenoids is heavily influenced by the presence of appropriate subplastidial structures that sequester them and promote their accumulation. Plastids are ubiquitous in plant cells, organs and species, but there are different types and they can fluctuate from one type to another depending on the developmental program and environmental conditions (Sadali et al., 2019; Solymosi et al., 2018). Carotenoids are only absent in proplastids (Howitt and Pogson, 2006; Li et al., 2016), which are undifferentiated plastids found in meristematic, reproductive and dedifferentiated tissues. They act as progenitors for all the other types of plastids (Solymosi et al., 2018; Sun et al., 2017).

Etioplasts are plastids that only occur in tissues that are not exposed to light (i.e., in seedlings that germinate in darkness). They develop when proplastids cannot be differentiated into chloroplasts. Etioplasts accumulate low amounts of carotenoids as well as the chlorophyll precursor protochlorophyllide in special structures called prolamellar bodies (PLBs) (Solymosi and Schoefs, 2010; Sundqvist and Dahlin, 1997). Carotenoids produced in etioplasts (such as violaxanthin and lutein) facilitate the transition to photosynthetic development when they differentiate into chloroplasts during de-etiolation (Park et al., 2002; Rodríguez-Villalón et al., 2009). Carotenoid levels increase to protect plastids from photooxidative damage as protochlorophyllide is converted into chlorophylls and PLBs into thylakoids during the etioplast-to-chloroplast differentiation process (Llorente et al., 2017; Quian-Ulloa and Stange, 2021).

Chloroplasts are photosynthetic plastids that are found in all green tissues. Chloroplasts accumulate high levels of carotenoids, with lutein and β-carotene being the most abundant ones (45% and 25-30% of the total, respectively) followed by violaxanthin and neoxanthin (10-15% each) in most plants (Ruiz-Sola and Rodríguez-Concepción, 2012). These carotenoids are often associated with proteins in different complexes of the photosynthetic apparatus (Domonkos et al., 2013; Hashimoto et al., 2016). Xanthophylls are primarily found in the light-harvesting complexes, whereas β-carotene is usually located in both photosystems and the cytochrome b6f complex. Despite their high carotenoid content, chloroplast-containing tissues are green due to the presence of chlorophylls. Carotenoids and chlorophylls are actually maintained in a tightly controlled proportion to ensure correct functioning of the photosynthetic machinery. Carotenoids function as accessory light-harvesting pigments and contribute to the assembly, stabilization, and repair of the photosynthetic apparatus (Domonkos et al., 2013; Hashimoto et al., 2016; Tóth et al., 2015). Another essential function of carotenoids in chloroplasts is photoprotection against photooxidative damage. Carotenoids prevent the accumulation of ROS under high light conditions via two complementary mechanisms: by dissipating excess energy as heat (through a mechanism called as non-photochemical quenching) and by quenching excited triplet chlorophyll and singlet oxygen (hence promoting free radical detoxification) (Domonkos et al., 2013; Hashimoto et al., 2016; Quian-Ulloa and Stange, 2021; Ramel et al., 2012; Ruban et al., 2012). Chloroplasts carotenoids can also be found in the envelope membranes, which are considered as their major production site as most carotenoid biosynthetic enzymes appear to be located there (Joyard et al., 2009; Ruiz-Sola and Rodríguez-Concepción, 2012). They are also present at low levels in plastoglobules, which are lipid bodies associated to thylakoids that may function as metabolite trafficking structures to transport carotenoids and other plastidial isoprenoids from their synthesis to their destination sites (Sun et al., 2017).

Developmental and environmental cues can promote the transformation of chloroplasts into other plastid types. For example, natural or dark-induced senescence transforms chloroplasts into gerontoplasts. Typically, gerontoplasts harbor disaggregated thylakoids and huge plastoglobules that accumulate carotenoids and their degradation products (Mulisch and Krupinska, 2013; Solymosi et al., 2018). During senescence, the chlorophylls/carotenoids ratio decreases due to chlorophyll degradation and carotenoid retention, and leaves turn to typical yellow and orange autumn colors (Egea et al., 2010).

In many fleshy fruits, chloroplasts differentiate into carotenoid-overaccumulating plastids called chromoplasts. Chromoplasts, which can also derive from other plastid types such as leucoplasts or amyloplasts, are non-photosynthetic plastid that provide color to organs such as fruits, flowers and storage roots (Egea et al., 2010; Li et al., 2016; Li and Yuan, 2013; Lopez-Juez and Pyke, 2005; Sun et al., 2017). Their structure and composition, in contrast

to chloroplasts, is very diverse. Chromoplast carotenoids can be stored in membrane systems, in plastoglobules, in lipoprotein complexes forming fibrils or as crystals. These storage structures define the type of chromoplast, which can be classified as membranous, globular, fibrillar, or crystalline (Egea et al., 2010; Li et al., 2016; Ljubesic et al., 1991; Sun et al., 2017). While the carotenoid composition often determines the type of storage structure, chromoplasts harboring different carotenoid-bearing structures and chromoplasts of different types can be found in the same organ (Egea et al., 2010; Ljubesic et al., 1996; Jeffery et al., 2012). Globular chromoplasts, found in saffron stigmas, mango, or citrus fruit, are enriched in plastoglobules (Gómez-Gómez et al., 2017; Jeffery et al., 2012; Ljubesic et al., 1996). Besides carotenoids, plastoglobules may accumulate tocopherols (Vidi et al., 2006), phylloquinone (Lohmann et al., 2006) and plastoquinone (Havaux, 2020; Zbierzak et al., 2010). Plastoglobule morphology is very variable depending on the nature of polar lipids and proteins that compose them, but also on the content of carotenoids and their esterification (Ytterberg et al., 2006). For example, tomato fruit chromoplasts show globular plastoglobules (Simkin et al., 2007) whereas fibrillar chromoplasts are characteristic of pepper fruit (Deruere et al., 1994). Membranous chromoplasts display concentrical multi-layer membranes derived from a proliferation of the inner envelope membrane, where carotenoids are accumulated. They were reported in *Narcissus* (daffodil) flowers (Hansmann et al., 1987) and the orange curd of *Or* mutants of cauliflower (Paolillo et al., 2004). Crystalline chromoplasts are abundant in tomato and carrot, where lycopene and β-carotene crystals, respectively, are encased in a bilayer of lipids (Jeffery et al., 2012; Maass et al., 2009). This heterogeneity of types and structures illustrates the huge versatility of plastids to accumulate any kind of carotenoid, making chromoplasts the best-known machinery for carotenoid production and storage in plants. Indeed, promoting the differentiation of chromoplasts appears as a useful strategy for carotenoid biofortification. Control of chromoplast differentiation has been a hot topic for decades, but the underling mechanisms have only recently began to be unveiled (Bian, 2012; Egea et al., 2010; Li et al., 2016; Li and Yuan, 2013; Llorente et al., 2020; Sadali et al., 2019).

1.3 Chloroplast-to-chromoplast differentiation

1.3.1 Non-photosynthetic tissues

The above-mentioned diversity and heterogeneity of chromoplasts in the plant kingdom makes it difficult to generalize a plan of chromoplastogenesis (Sadali et al., 2019). Chromoplasts can differentiate from a variety of pre-existing plastids: directly from proplastids (e.g., in papaya fruit), from amyloplasts (e.g., in red stigmas of saffron or carrot roots) or from chloroplasts (e.g., in tomato fruit or daffodil flowers) (Grilli-Caiola and Canini, 2004; Hansmann et al., 1987; Harris and Spurr, 1969; Kim

et al., 2010; Schweiggert et al., 2011). The chloroplast-to-chromoplast differentiation process has been best studied during fruit ripening (mostly in tomato and pepper) from microscopic, transcriptomic, proteomic and metabolic point of views. The increase on carotenoid content, the disbanding of thylakoids and the degradation of chlorophylls are among the few common processes that have been observed during fruit ripening of different species. Actually, the degradation of chlorophylls can be uncoupled from chromoplast differentiation, since the tomato mutant *green flesh* (*gf*) showed detectable levels of chlorophylls in chromoplast-containing ripe tomatoes (Cheung et al., 1993).

In order to find specific molecular features that could be associated to chromoplast differentiation, the proteins from chromoplast-containing non-photosynthetic tissues of six different crop species (tomato, papaya, watermelon, and pepper fruits, orange cauliflower and carrot root) were compared (Wang et al., 2013). This study showed the presence of proteins related to carotenoid biosynthesis, primary metabolism, amino acid metabolism or protein import machinery in most of the crops, as well as proteins related to photosynthesis, suggesting their possible relevance despite the thylakoid disruption. In these systems, however, chromoplast differentiation occurs in parallel with many other developmental processes associated to ripening that are not directly connected with the plastid conversion mechanism. That makes the study of chloroplast-to-chromoplast differentiation very complicated, since it cannot be separated from events unrelated to the chromoplast biogenesis process. For instance, studies in tomato fruit ripening have revealed the role of hormones like ethylene or ABA, changes in the primary metabolism (e.g., high accumulation of hexoses, increase in the TCA cycle or degradation of starch) or degradation of cell walls (Alexander and Grierson, 2002; Biais et al., 2014; Carrari and Fernie, 2006; Quinet et al., 2019; Zhang et al., 2009). But other systems where chromoplasts are also differentiated from chloroplasts show a different profile, suggesting that chromoplast differentiation is not linked to these features, or that it is very variable among different systems, or both (Klie et al., 2014).

Despite these issues, some molecular factors have been described to be generally associated to chromoplastogenesis process in different systems. Fibrillins are plastid proteins associated to carotenoid sequestration and plastoglobule structural development that have been found to be associated to chromoplasts (Singh and McNellis, 2011) (Figure 1.2). For instance, fibrillin 1 (FBN1) is a major component of the pepper fruit chromoplast fibrils (Deruere et al., 1994). Other structural proteins like chromoplast-specific carotenoid-associated protein (CHRC) or chromoplast protein D (CHRD) have also been associated to chloroplast-to-chromoplast differentiation (Kilambi et al., 2013; Leitner-Dagan et al., 2006a; Leitner-Dagan et al., 2006b). Transgene-mediated overexpression of some of these proteins was shown to alter carotenoid levels in tomato fruit, often through changes in chromoplast development (Li et al., 2019; Rey et al., 2000). Some proteins involved in protein import from

the translocons at the outer and inner chloroplast envelope (TIC and TOC, respectively) are also affected in plastid transitios. In particular, SP1 (*SUPRESSOR OF ppi1*) and SP1-like2 (SPL2) have shown to play a role in the chloroplast-to-chromoplast transition in tomato by accelerating the reconfiguration of the protein import machinery (Ling et al., 2021) (Figure 1.2). Members of the protein quality control (PQC) system have also found to be associated to chromoplast biogenesis (Figure 1.2). They participate in the refolding and stabilization of proteins that lose their native conformation, or degrading mature proteins that are damaged or need to be removed (D'Andrea and Rodriguez-Concepcion, 2019). Many plastidial chaperones are upregulated during tomato fruit ripening, presumably to deal with proteome changes and protein folding stress (Barsan et al., 2010). One of the small heat shock proteins, Hsp21, was associated to chloroplast-to-chromoplast differentiation (Neta-Sharir et al., 2005). Other chaperones induced during the ripening like Hsp70 and ClpB3 have been described to promote proper folding and activity of biosynthetic enzymes in the chromoplast (D'Andrea et al., 2018; Llamas et al., 2017; Pulido et al., 2016). But among chaperones, Orange (OR) is the only one described to directly promote chromoplast differentiation (Figure 1.2). OR and a version with a single nucleotide polymorphism (OR[His]) were found to induce carotenoid overaccumulation in orange cultivars of cauliflower (Li et al., 2001; Lu et al., 2006), melon (Tzuri et al., 2015) and other crops (Berman et al., 2017; Lopez et al., 2008; Park et al., 2015; Yazdani et al., 2019). OR interacts with PSY, ensuring proper activity and preventing degradation (Park et al., 2016; Welsch et al., 2018; Zhou et al., 2015), and it also affects carotenoid degradation, plastid protein import and plastid division (Sun et al., 2020). Nonetheless, the mechanism explaining how OR promotes chromoplast differentiation remains unknown.
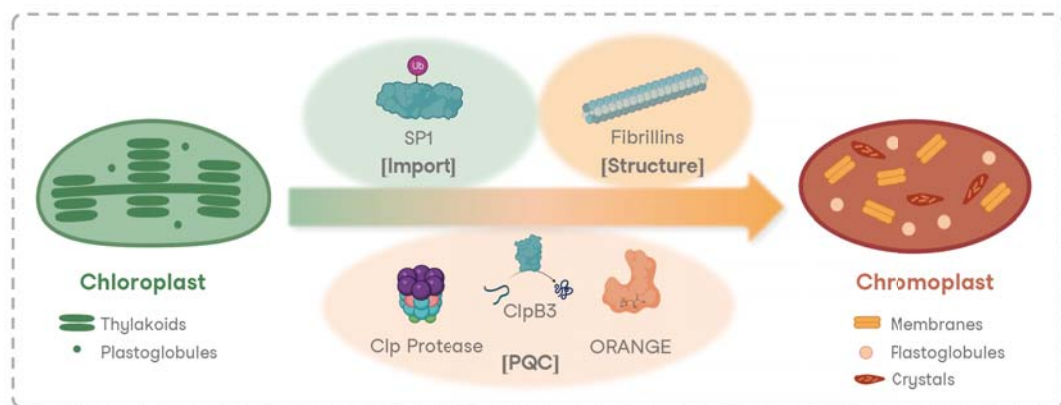
Figure 1.2: Chloroplast-to-chromoplast differentiation. Schematic representation of the differentiation of chloroplast to chromoplasts and factors associated to the process.

1.3.2 Chromoplastogenesis in leaves

Among the different organs that have been described to differentiate chromoplasts, leaves are probably the rarest, as only leaves of a handful species such as

Buxus sempervirens appear to naturally differentiate chloroplasts into chromoplasts (Hormaetxe et al., 2004; Hormaetxe et al., 2005; Koiwa et al., 1986). An interesting feature of this system is the capacity of regreening, i.e., the reversibility of the differentiation process and the recovering of the chloroplast stage (Hormaetxe et al., 2004; Hormaetxe et al., 2005; Koiwa et al., 1986). Recent results have shown that chromoplasts can also be artificially (and irreversibly) differentiated in leaves (Llorente et al., 2020). The first observation leading to this discovery was that leaves from different plant species turned their color from green to yellow after virus-mediated expression of a bacterial phytoene synthase, crtB (Majer et al., 2017). The strategy was initially focused on engineering a synthetic pathway for the extraplastidial production of carotenoids by introducing bacterial enzymes transforming cytosolic IPP and DMAPP into GGPP (crtE), GGPP into phytoene (crtB) and phytoene into lycopene (crtI). This approach worked successfully, inducing the accumulation of relatively high levels of lycopene that resulted in a red coloration of the leaves (Majer et al., 2017). The expression of crtB alone was intended as a control to confirm that single enzymes could not produce cytosolic carotenoids due to the absence of substrates. Yet, it induced an unexpected yellow phenotype as a result of an enhancement in total carotenoid levels in plastids. A discussed possibility was that crtB-produced cytosolic phytoene was activating a putative mechanism that induced chloroplasts to overproduce carotenoids. This hypothesis was the starting point of this PhD thesis, that has been focused on the characterization of the chromoplastogenesis process that crtB was inducing in the leaf.

Objectives

crtB, a bacterial phytoene synthase, was described to induce a yellow phenotype by increasing carotenoid levels when transiently expressed in leaves (Majer et al., 2017). It was hypothesized that phytoene, the first committed carotenoid of the pathway, was responsible to trigger the phenotype due to its overaccumulation, which was not previously reported under normal conditions in leaves. Chromoplast differentiation was proposed in our lab at the start of this PhD thesis by observing the changes caused by crtB after agroinfiltration of *N. benthamiana* leaves (Llorente et al., 2020). Because chloroplast-to-chromoplast transitions are only described in very few cases in leaves (Koiwa et al., 1986; Hormaetxe et al., 2004; Hormaetxe et al., 2005), we refer to the crtB-mediated process as "artificial chromoplastogenesis". The goal of this thesis has been to describe the molecular mechanisms supporting the artificial chloroplast-to-chromoplast differentiation in leaves of *N. benthamiana* after crtB expression. Specifically, we set up the two main objectives described below:

2.1   Objective 1

The first objective of this thesis was to characterize the artificial chromoplast differentiation process at biochemical, molecular, metabolic and physiological levels, with a main focus on the end-point, once the yellow phenotype was already established in *N. benthamiana* leaves after *crtB* agroinfiltration.

2.2   Objective 2

The second objective was to obtain a dynamic description of the chromoplastogenesis process by analyzing the transcriptomic changes that occur at different times after *crtB* expression in leaves of *N. benthamiana*.

Chapter 1: Molecular characterization of the crtB-induced phenotype in *Nicotiana benthamiana* leaves

3.1 crtB induces carotenoid accumulation only when acting in the plastid

The transient expression of the *Pantoea ananatis crtB* gene in leaves of *Nicotiana benthamiana* plants was previously reported to induce a yellow leaf phenotype and enhanced levels of endogenous carotenoids (Majer et al., 2017). In my master dissertation, I could establish that crtB only triggers the yellow phenotype when it is located in chloroplasts, synthesizing phytoene from the pool of GGPP that derives from the MEP pathway (Torres-Montilla, 2017). The unmodified bacterial crtB was found mainly in the cytosol but also in the plastids of agroinfiltrated leaf cells (Majer et al., 2017), suggesting that either it contains a cryptic plastid-targeting signal in its sequence or it uses a different mechanism to enter the plastid. An exclusively plastid-located construct of crtB harboring a plastid targeting peptide (renamed as (p)crtB) produced a yellow phenotype similar to that caused by crtB while an exclusively cytosol-located construct of crtB (renamed as (c)crtB) did not trigger the yellow phenotype (Llorente et al., 2020; Torres-Montilla, 2017). In this thesis work, we investigated how these different versions of crtB impacted carotenoid accumulation in agroinfiltrated leaves and compared it with the effect of transiently expressing PSY-encoding plant genes (Figure 3.1). The accumulation of phytoene, the first committed intermediate of the carotenoid pathway (Figure 1.1), was hardly observed in photosynthetically active chloroplasts of control *N. benthamiana* leaves agroinfiltrated with GFP, likely because downstream carotenoid enzymes consume it. The transient expression of constructs encoding crtB or (p)crtB, but not (c)crtB, resulted in a strong accumulation of phytoene and also of endogenous downstream carotenoids (Figure 3.1). By contrast, plant PSY genes from Arabidopsis thaliana and Solanum lycopersicum (AtPSY and SlPSY2 respectively) did not cause such accumulation (Figure 3.2). For this reason, the induction of the yellow phenotype was

deduced to be highly associated with the accumulation of non-colored phytoene but caused by the enhanced production of downstream colored carotenoids in the leaf plastids. Unlike that observed when using viral vectors (Majer et al., 2017), chlorophylls did not change their levels after the agroinfiltration of any crtB construct, showing the same abundance as the GFP control (Figure 3.1).
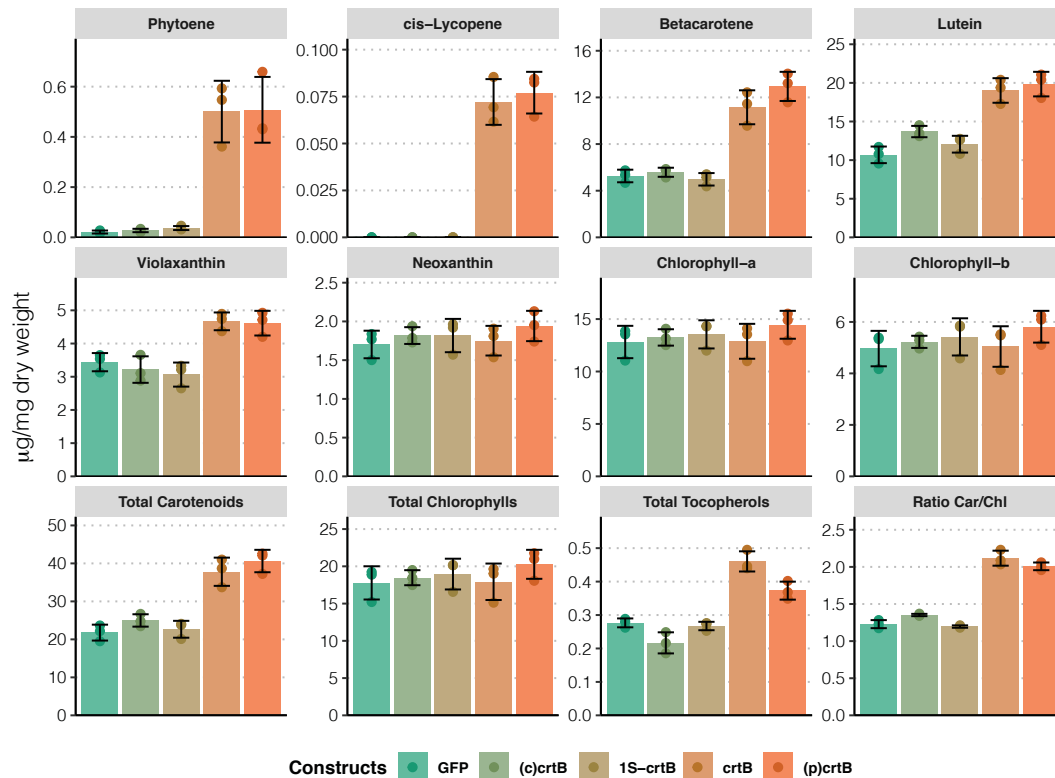
Figure 3.1: Carotenoids, chlorophylls and tocopherols in leaves agroinfiltrated with different crtB constructs. Levels of carotenoids, chlorophylls, tocopherols and ratio carotenoids/chlorophylls (Car/Chl) in leaves of *N. benthamiana* agroinfiltrated with GFP, (c)crtB, 1S-crtB, crtB and (p)crtB, identified with different colors.

One of the hypothesis about how crtB enters the plastids was through the endoplasmic reticulum (ER), according to metabolic continuity that may exist between the ER membranes and the plastids (Mehrshahi et al., 2014). To test this possibility, we used a crtB construct fused to the transmembrane domain of the *Arabidopsis* HMGR enzyme (1S domain), which anchors the protein to the cytosolic side of the ER membrane (Ferrero et al., 2015). However, this 1S-crtB construct was not able to induce the yellow phenotype in the leaves, and the HPLC profile of carotenoid and related isoprenoids (chlorophylls and tocopherols) was similar to the (c)crtB construct (Figure 3.1) (Andersen et al., 2021). Consequently, we concluded that crtB was not accessing to the plastics through the ER but it might include some cryptic transit-peptide in its sequence for the entry to the plastids, even though it was not possible to detect it with the traditional in silico predictors (TargetP and ChloroP).
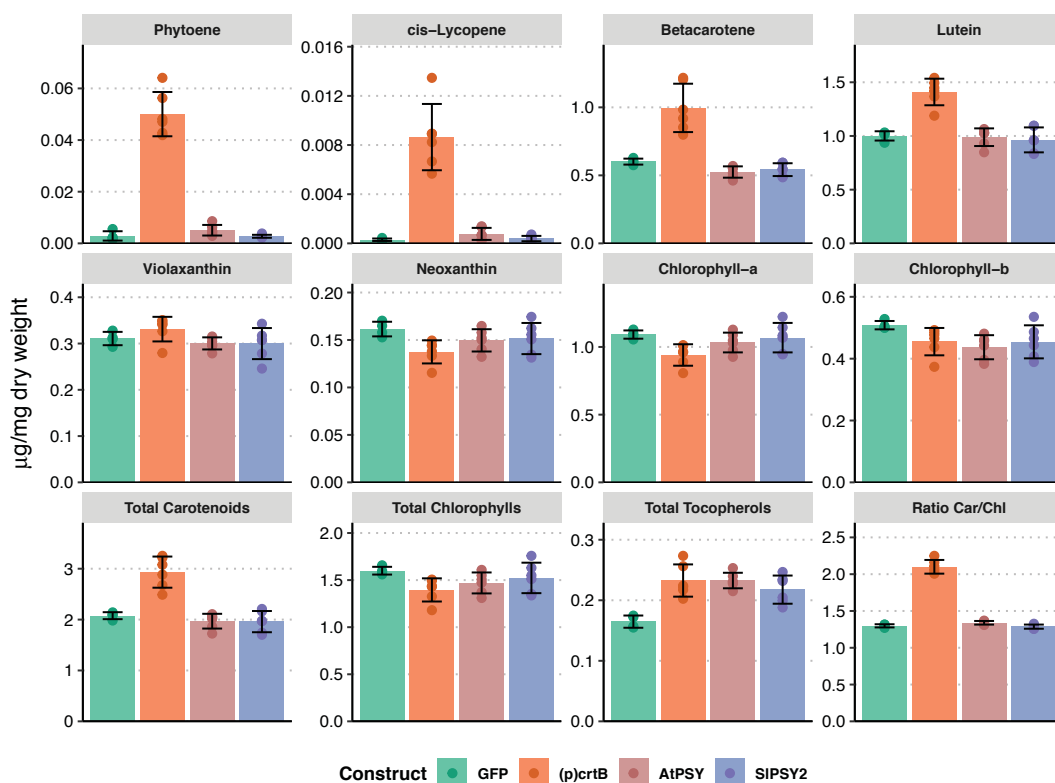
Figure 3.2: Carotenoids, chlorophylls and tocopherols in leaves agroinfiltrated with AtPSY, SlPSY2 and crtB. Levels of carotenoids, chlorophylls, tocopherols and ratio carotenoids/chlorophylls (Car/Chl) in leaves of *N. benthamiana* agroinfiltrated with AtPSY, SlPSY2 and crtB, identified with different colors.

We addressed the possibility of the malfunction of (c)crtB and 1S-crtB constructs. Although IPP and DMAPP are produced in the cytosol via the MVA pathway, most of these metabolic precursors are consumed to form farnesyl diphosphate (FPP), used for the synthesis of sterols, while the pool of GGPP (i.e., the substrate for crtB enzymes) is very limited in the cytosol. Due to this limitation of substrate, it was troublesome to associate the undetectable production of phytoene and the absent of phenotype in (c)crtB and 1S-crtB agroinfiltrated leaves to the inactivity of these proteins or to the malfunctioning of these constructs. To resolve this question, another bacterial enzyme (crtE) was co-agroinfiltrated to enhance the production of GGPP from MVA-derived IPP and DMAPP in the cytosol. The results showed that both (c)crtB and 1S-crtB were functional as the levels of phytoene were slightly but significantly increased in the presence of crtE (Figure 3.3). However, these levels were still remarkably lower than those observed in crtB and (p)crtB agroinfiltrations, which means that crtB is likely active in both compartments but it only induces the plastid differentiation process by producing high levels of phytoene in the chloroplast from MEP-derived precursors. The capacity of crtB to produce phytoene in the cytosol can be dramatically boosted by increasing the supply of IPP and DMAPP through the co-expression of a construct encoding a deregulated version of the main rate-limiting enzyme of the MVA pathway (Andersen et al., 2021). While this strategy results in

a very effective extraplastidial biofortification of leaves with cytosolic phytoene, it is unable to trigger the yellow phenotype caused by the plastid-localized versions of crtB (Andersen et al., 2021). These results further support our conclusion that crtB only induces the yellow leaf phenotype resulting from endogenous carotenoid overaccumulation when acting in plastids.
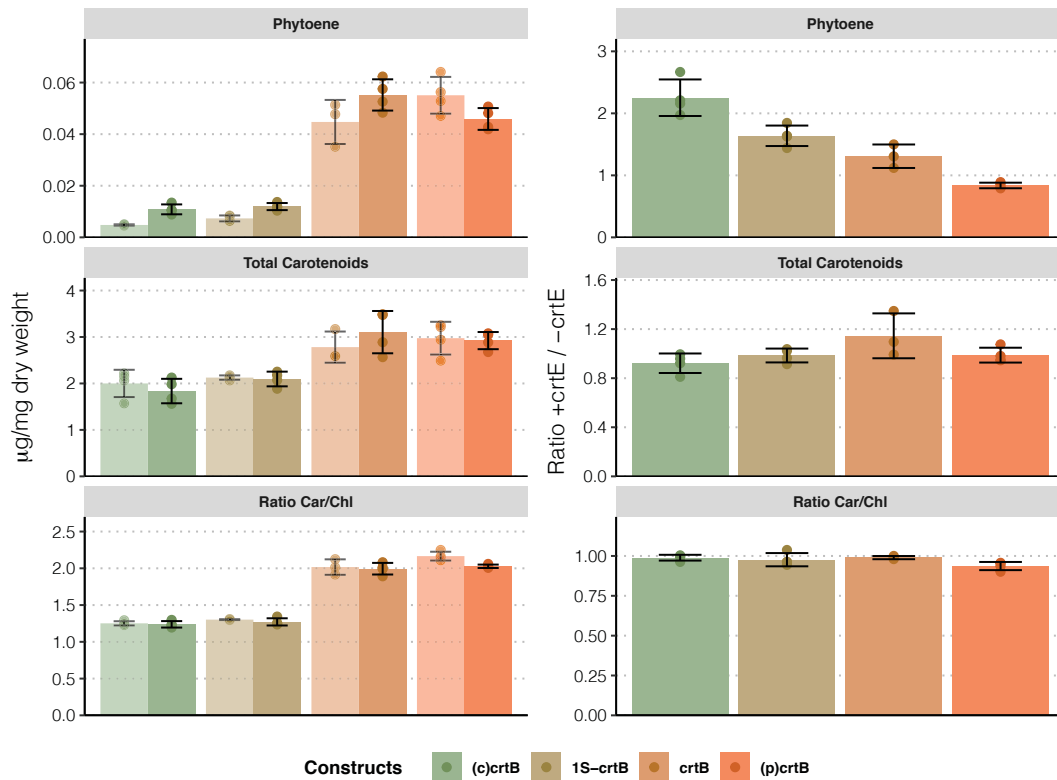
Figure 3.3: Phytoene, total carotenoids and ratio Car/Chl of co-agroinfiltrations with crtE. Levels of phytoene, total carotenoids and ratio Car/Chl in leaves of *N. benthamiana* agroinfiltrated with (c)crtB, 1S-crtB, crtB and (p)crtB, and co-infiltrated with crtE. All values are plotted in left column, showing infiltrations without crtE in a lighter color and co-infiltrations with crtE in a more intense color. Right columns shows the ratio of co-infiltrations values related to single infiltrations.

3.2 crtB transforms chloroplasts into chromoplast-like plastids

To study whether plastid ultrastructure was altered in crtB-producing yellow leaf sections, transmission electron microscopy (TEM) images were taken in *N. benthamiana* leaves agroinfiltrated with different constructs. In crtB and (p)crtB samples, TEM images showed a distinctive morphology of the plastids, which were very different from the chloroplasts observed in the GFP agroinfiltrated controls or in the (c)crtB samples (Figure 3.4B) (Llorente et al., 2020). These particular plastids from yellow sections lack organized photosynthetic thylakoids and grana, contained electron-dense (i.e., lipid-containing) membrane stacks much more tightly

appressed than grana, and showed a proliferation of small electron-dense round vesicles tentatively identified as plastoglobules, a morphology that resembled that of a chromoplast-like plastid. The ultrastructure of these plastids was very different from that of the gerontoplasts that developed in dark-incubated senescent *N. benthamiana* leaves (Figure 3.4D). Gerontoplasts still contained thylakoids and grane, and they showed huge plastoglobules vesicles that were very different from the electro-dense and relatively small vesicles detected in leaves agroinfiltrated with crtB (Figure 3.4D).
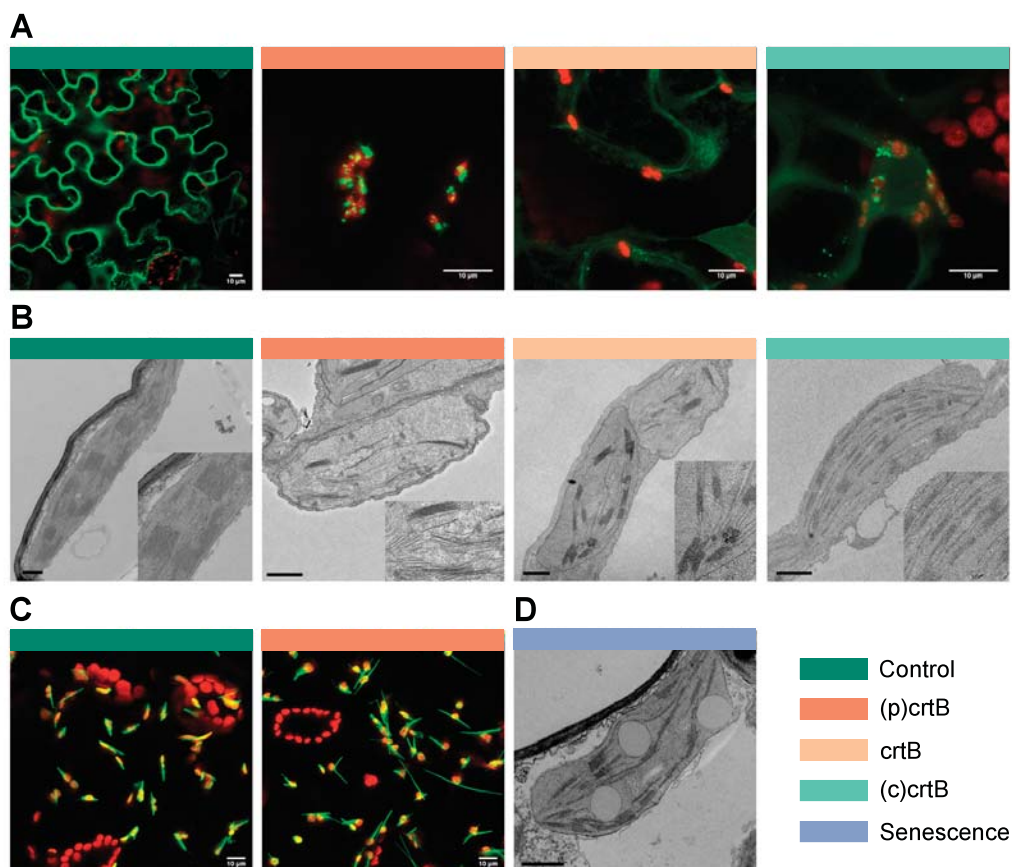
Figure 3.4: Subcelullar localization, ultrastructure and plastid morphologies observed in agroinfiltrations. (A) GFP (green) and chlorophyll (red) fluorescence of *N. benthamiana* leaves agroinfiltrated with indicated constructs at 4 days post infiltration (dpi). (B) TEM images of representative plastids and magnifications of *N. benthamiana* leaves agroinfiltrated with indicated constructs at 5 days post infiltration (dpi). (C) Plastid-located GFP (green) and chlorophyll (red) fluorescence of *N. benthamiana* leaves agroinfiltrated with indicated constructs at 4 days post infiltration (dpi). (D) Gerontoplast from a *N. benthamiana* leaf harvested from the plant and kept in the dark for 10 days (senescent). Control refers to GFP agroinfiltrated leaves in A and B, and plastid-located GFP in C.

A more dynamic study of plastid changes was carried out by staining the plastid stroma with a plastid-targeted GFP version named (p)GFP and promoting the differentiation of chromoplast-like plastids by co-expressing (p)crtB. The distribution

of the stromal GFP signal was analyzed by confocal microscopy. The most prominent structures were slim filaments that emerged from plastids, corresponding to stromules (Figure 3.4C). These structures could be observed both in control (only (p)GFP) and induced plastids (both (p)GFP and (p)crtB), but they showed a sharper and longer pattern in the (p)crtB condition. Plastid area was also studied using GFP and chlorophyll signals at two different magnifications (x0.75 and x2.25) and at different time-points from 48 to 120 hours post-infiltration (hpi). The quantification of the areas revealed that control plastids were larger at all time-points, suggesting that (p)crtB causes a reduction in plastid size (Supplementary Figure 1).
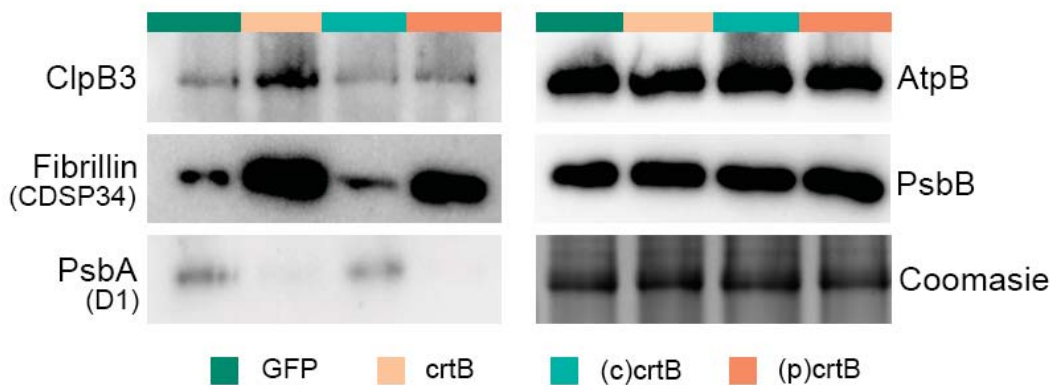
Figure 3.5: crtB induces chromoplast protein markers. Immunoblot analysis of chromoplast associated proteins (ClpB3 and fibrillin) and photosyntesis-related proteins (psbA, AtpB and psaB).

To further confirm that the plastids observed after the overproduction of phytoene by crtB were indeed chromoplasts, the levels of some protein markers (Figure 1.2) were studied by Western-blot. Fibrillins are structural proteins that have been associated to chromoplasts, mainly in fruits (Deruere et al., 1994; Langfelder and Horvath, 2008; Simkin et al., 2007; Van Wijk and Kessler, 2017). Fibrillin/CDSP34 proteins showed higher levels in crtB and (p)crtB treatments compared to GFP and (c)crtB, consistent with the induction of typical carotenoid sequestering structures of chromoplasts (Figure 3.5). ClpB3, a Hsp100 chaperone reported to be up-regulated during tomato fruit ripening (D'Andrea et al., 2018), also exhibited higher protein levels in crtB and (p)crtB agroinfiltrated areas (Figure 3.5). On the other hand, the PsbA/D1 protein, one of the core components of photosystem II (PSII) in the antenna complex, was only detected in GFP and (c)crtB treatments, suggesting the disorganization of the photosystems in crtB and (p)crtB agroinfiltrated areas (Figure 3.5). These results concur with the disaggregation of thylakoids observed in the TEM images. However, PsaB (a protein associated to the photosystem I) and atpB (a chloroplastic ATP synthase) did not show any difference between any treatments. That could mean that not every component related to photosynthesis are degraded in the plastid. In summary, it was concluded that crtB was inducing a

differentiation process when it was able to produce phytoene in the plastids, changing the morphology of these organelles and transforming them into chromoplast-like plastids. The following rounds of experiments were performed using plastid-targeted (p)crtB instead of the unmodified version of crtB to avoid any interference potentially derived from the activity of the crtB enzyme in the cytosol.

3.3 Photosynthetic loss and positional effect in the leaf

Chromoplasts are non-photosynthetic plastids that lose their photosynthetic capacity during their development. In chloroplast-to-chromoplast transition programs in nature, photosynthesis stops when the thylakoidal structures disaggregate. In the case of the chromoplast-like plastids that developed in (p)crtB samples, this disaggregation was observed in the TEM images (Figure 3.4A), and supported by the loss of some photosynthetic proteins (e.g., PsbA) was revealed by Western-blot (Figure 3.5). However, other photosynthesis-related proteins were not changed (e.g., PsaB and AtpB) (Figure 3.5) and chlorophyll levels in yellow sections were similar to those in control, chloroplast-containing samples (Figure 3.1). Therefore, we aimed at measuring photosynthetic parameters to test whether this central determinant of chloroplast identity was really stopping. For this purpose, the proportion of absorbed light that is used in photosystem II photochemistry under normal light conditions (named as effective quantum yield of photosystem II or ϕPSII) was monitored in *N. benthamiana* leaf sections agroinfiltrated with either GFP or (p)crtB constructs. In a time-course from the agroinfiltration event, ϕPSII stayed stable in GFP sections but it showed an abrupt decrease from 45 hpi in (p)crtB sections to reach values close to 0 (i.e. no photosynthesis) by the end of the time-course at 96 hpi (Figure 3.6). The disruption of photosynthesis in (p)crtB agroinfiltrated leaves is another chromoplastidial feature that appends to the distinctive plastid morphology, the enhancement of carotenoids, the yellow phenotype and the protein markers.

During the execution of these experiments, we realized that the position where the agroinfiltration was performed in the leaf is not trivial, since the phenotype was observed to change depending on this position. For this reason, levels of carotenoids were studied by HPLC for GFP and (p)crtB agroinfiltrations in different leaf sections. For every condition, samples were collected from areas close to the petiole (named as proximal) and from other areas nearby the apex (named as distal). HPLC analysis of yellow (i.e., (p)crtB) sections showed that distal regions tended to accumulate higher levels of carotenoids and chlorophylls despite they showed a less intense yellow color (Figure 3.7). The higher levels of chlorophylls might attenuate the yellow phenotype caused by the accumulation of carotenoids despite they also increase their levels. When different leaves from the same plant were compared, leaf age was inversely correlated with the intensity of the yellow color, i.e., younger leaves showed the most intense phenotype (Figure 3.7). The following rounds of experiments were performed trying to apply the agroinfiltration treatments in the
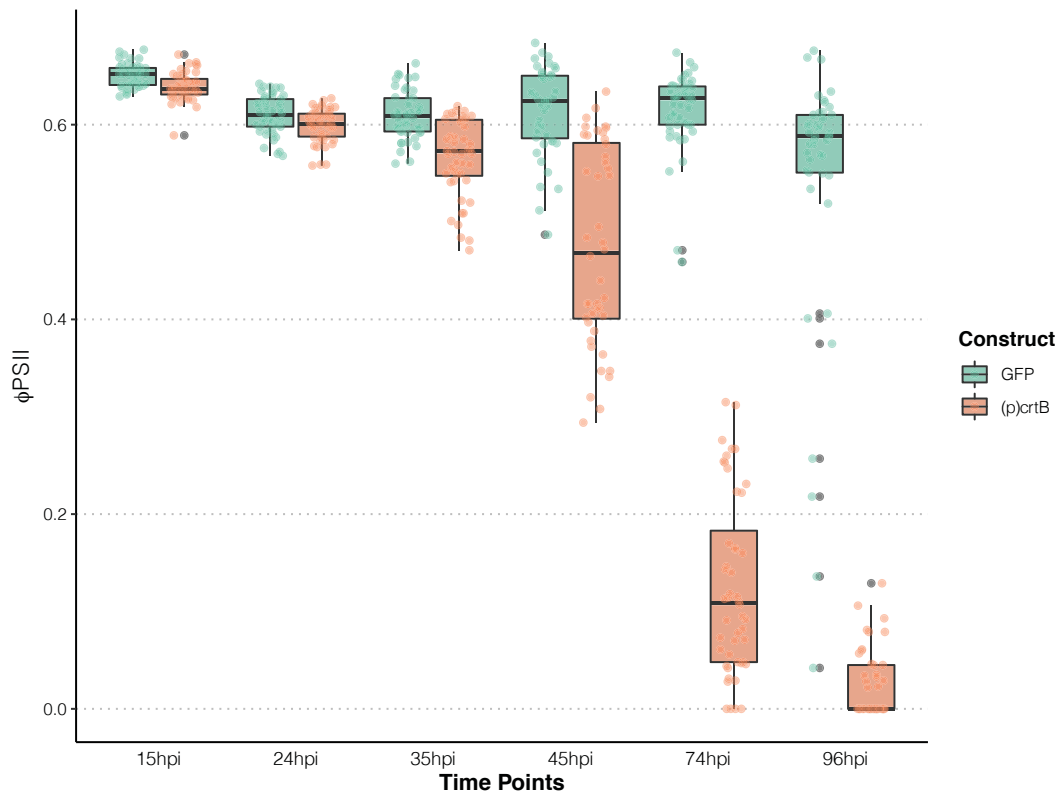
Figure 3.6: Photosynthesis stops working in crtB samples. Photosynthetic efficiency measured by ϕPSII values in GFP and (p)crtB agroinfiltrated leaves of *N. benthamiana* at different hpi.

third leaf and the proximal position in all plants to ensure homogeneity. Also, both hemispheres (halves) of the same leaf were used, one for (p)crtB and the other one for the GFP control, so they could be fully comparable.

3.4 Leaf areas developing artificial chromoplasts show dramatic changes in nuclear gene expression

To analyze the changes in gene expression associated with the (p)crtB-mediated differentiation of chloroplasts into chromoplasts, RNA-seq analysis was performed comparing RNA extracted from (p)crtB vs GFP agroinfiltrated areas from the same leaf at 96 hpi (i.e., when the yellow phenotype was very clear, the ultrastructure of plastid was already changed, photosynthesis was abolished, and total carotenoid levels were ~2 times increased compared to GFP controls).

3.4.1 First analysis: Sequentia Biotech and AgriGO

The first approach for the analysis of the RNA-seq data was performed with the Artificial Intelligence RNA-seq (AIR) software from Sequentia Biotech SL, due to the inability to perform these analyses by myself at the beginning of my PhD. The first step for RNA-seq analysis is the trimming of the reads for each sample, in order to

Figure 3.7: Positional effect on agroinfiltration.(A) Total carotenoids, chlorophylls and ratio in *N. benthamiana* agroinfiltrated in different positions of the leaf with (p)crtB. (B) Phenotype of the 3rd leaf of *N. benthamiana* agroinfiltrated in proximal and distal positions. (C) Phenotype of the 5th leaf of *N. benthamiana* agroinfiltrated in proximal and distal positions.

remove low quality sequences. The number of reads passing the quality control step of the AIR software was very low (50%), somehow questioning the reliability of the sequencing procedure (Table 3.1). The second step was the mapping of the trimmed reads to a reference genome. The genome that was used for the mapping was the last published genome for *N. benthamiana* at that time: Niben genome v.1.0.1 (Niben101), published in the Sol Genomics Network webpage (Fernandez-Pozo et al., 2015). The mapping was fine, with less than 10% of unmapped sequences and almost 80% of unique mapped reads (Table 3.1). Mapped reads were analyzed using the DESeq2 package (Love et al., 2014), setting the GFP agroinfiltrated sample as reference. The number of Differentially Expressed Genes (DEGs) in (p)crtB compared to GFP samples was high, with 3,183 up-regulated genes and 1,803 down-regulated genes using a fold change (FC) filter of FC $\geq$ 1.5 and FC $\leq$ 2/3 (i.e., $\log_2$FC $\geq$ 0.585 and $\log_2$FC $\leq$ -0.585), respectively. Hence, the total number of DEGs was 4,983, illustrating a large

impact on nuclear expression.

Table 3.1: Summary of first RNA-seq analysis. Number of raw (untrimmed) and mapped to Niben101 genome sequences per sample, and percentages of trimming and mapping.

Sample	Untrimmed sequences	% Trimming	% Mapping	Mapped sequences
GFP_96h_1	27,491,827	51.7	92.6	13,131,492
GFP_96h_2	27,644,786	45.9	93.3	11,850,312
GFP_96h_3	25,011,083	49.5	91.5	11,331,649
(p)crtB_96h_1	21,700,011	50.2	91.7	9,991,674
(p)crtB_96h_2	24,967,626	52.2	94.0	12,240,540
(p)crtB_96h_3	25,226,228	50.8	92.2	11,822,598

In order to translate lists of genes into a more comprehensible information, we performed enrichment analysis of Gene Ontologies (GOs). GOs are divided in three different categories: biological process (BP), molecular function (MF) and cellular component (CC). These terms are universal for all species, and genes are associated to GO terms based on their sequence homology once a genome is annotated (functional annotation). The GO enrichment analysis was initially performed using AgriGO v2.0 software (Tian et al., 2017). Because the Niben101 genome was not included in AgriGO database by the time we were doing the analysis for the first time, our list of genes was first adapted to the previous genome version (Niben044), losing about 1/3 of DEGs. These new lists of DEGs were used to perform Parametric Analysis of Gene Set Enrichment (PAGE) analysis, using both the gene IDs and their FC values. The output includes a list of enriched GO terms together with a p-value indicating the significance of the enrichment and a z-score, a number that represents the behavior of the genes that were included in a specific GO term. A GO term with a positive z-score indicates that the genes included are mostly up-regulated, while a negative z-score indicates the predominance of down-regulation.

The next step was to compare the resulting GO terms with those enriched during plastidial differentiation processes in other species. As a reference system for chloroplast-to-chromoplast differentiation, we compared tomato fruit at the Red Ripe and Orange stages (RR and OR, corresponding to mature and differentiating chromoplasts, respectively) with those at the Mature green stage (MG, with only chloroplasts) (Shinozaki et al., 2018). Also, we included in the analysis the transcriptome associated to gerontoplast development in *Arabidopsis*, comparing leaves of 30 days (senescent) against 16 days (non-senescent) (Woo et al., 2016). The aim of this comparison was to check if the profile of gene expression changes in our system was more similar to the chloroplast-to-chromoplast or the chloroplast-to-gerontoplast differentiation process. The fragments per kilobase per million mapped reads (FPKM) lists supplied in the papers reporting the tomato ripening and *Arabidopsis* senescence data were used to calculate the FC values manually applying the same filter as to our data ($\log_2 FC \geq 0.585$ and $\log_2 FC \leq -0.585$) to identify
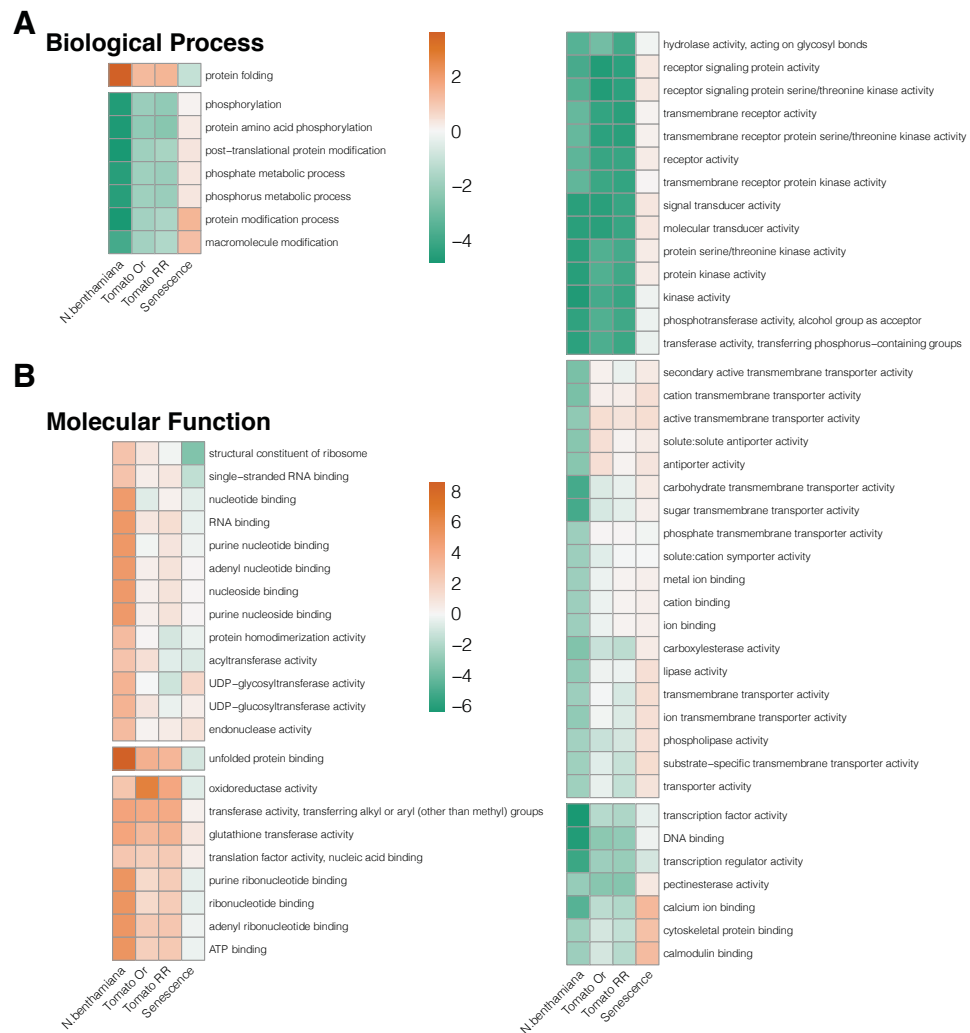
Figure 3.8: GO terms comparison between *N. benthamiana*, tomato and *Arabidopsis*. (A) Biological processes heatmap comparing (p)crtB vs GFP in *N. benthamiana*, OR and RR vs MG in tomato fruit ripening, and senescent vs non-senescent leaves in *Arabidopsis*. (B) Molecular processes heatmap comparing the four systems. Color scale shows z-score values in orange gradient for positive and green gradient for negative.

DEGs. Then, the resulting gene IDs and their FC values were used for PAGE analysis in AgriGO v2.0. After the comparison of the GO terms from the three different systems, the z-score of overlapping terms (i.e., those arising in the three experiments and the four comparisons) were plotted on heatmaps (Figure 3.8). The overlapped BP list was quite short (8 terms) and showed a clear similarity between tomato fruit ripening and our system (that we refer to as "chromoplastogenic leaves"), while senescence had an opposite behavior. "Protein folding" was a predominantly up-regulated BP in chromoplastogenic leaves and ripening fruit but it was down-regulated in senescing leaves. By contrast, phosphorylation-related and protein modification processes were down-regulated in the two first systems, while up-regulated in senescing leaves. The list of MF terms was longer but it could be

grouped in blocks. Some of these blocks showed the same clear pattern of chromoplastogenic leaves resembling ripening fruit but differing from senescing leaves, while others showed a more diffused positive or negative association. "Unfolded protein binding", transferase activities and nucleotide-related functions were up-regulated in chromoplastogenic *N. benthamiana* leaves and ripening tomato fruit, but down-regulated in senescing *Arabidopsis* leaves. The inverse behavior was observed in kinase activities, receptor-related and transferase-related functions. According to this approach, the differentiation process triggered by (p)crtB shares more resemblance to the chromoplast differentiation in fruits than to gerontoplast biogenesis in leaf senescence, hence confirming our previous conclusion.

3.4.2   Niben261: new genome and annotation

A 3-month research stay at Dr. Aureliano Bombarely's lab (University of Milan, Italy) was carried out to gain a higher quality analysis of the data. Dr. Bombarely had generated a new version of the *N. benthamiana* genome with a better quality (Niben261), assigning the different genes to chromosomes and including mitochondrial and plastidial DNA. For the functional annotation of the genome, during the stay we run three BLAST searches against TAIR (*Arabidopsis*), Swissprot and Trembl databases, respectively. The three outputs were integrated into a single file with Automated Assignment of Human Readable Descriptions (AHRD) program (`https://github.com/groupschoof/AHRD`), including GO annotations. As a result, GO terms from the three different categories were linked to genes of the Niben261 genome, based on the association of each sequence to the databases used as reference. It is important to note here, however, that GO terms are only a simplified but incomplete way to understand transcriptomic data. For instance, only eight of the twenty-four genes of the carotenogenic pathway in tomato are associated to the "Carotenoid Biosynthesis Process" term in this species. Consequently, there are two thirds of probabilities that this term would not show up to be enriched even if the genes of the pathway are included in the input DEG list for the analysis.

Model species like *Arabidopsis* or tomato have been extensively studied, and most of their genes have been validated by experimental studies, making the analysis of specific gene families more easy and accurate. Working with *N. benthamiana* implies more difficulties, as it has been scarcely studied and most of the genes in the genome have not been characterized yet, forcing us to rely only on homology searches and GOs. To address this problem and improve the analysis of our data, a high-resolution phylogenetic reconstruction was performed to identify different *N. benthamiana* gene families of interest that, in principle, might be involved in the artificial chromoplastogenesis process. Among others, we searched for the full complement of enzymes involved in the biosynthesis of carotenoids, chlorophylls,

hormones and isoprenoid precursors, as well as chaperones, fibrillins and phytochromes. The well-annotated *Arabidopsis* gene lists were used as a reliable reference, using their corresponding protein (amino acid) sequences in BLASTP searches run against *N. benthamiana*, *S. lycopersicum* and *C. annuum* genomes to obtain a list of candidates for each species. This procedure was substituted for the identification of specific protein domains using HMMER 3.3.2 software (`http://hmmer.org/`) when such domains were known to define the family. The amino acid FASTA sequences of the three candidate lists, together with the protein query list from *Arabidopsis*, were aligned using Guidance software (Sela et al., 2015). After that, IQtree software (Nguyen et al., 2015) found the optimal model that fits for each analysis, and then it was re-run to generate a phylogenetic tree. The quality of the analysis was confirmed by checking the proper location of tomato proteins in the tree, as they are relatively well annotated. Due to genome duplication events, it was expected to find one or more tomato genes for each one in *Arabidopsis*, and 1:1:2 proportions of tomato, pepper and *N. benthamiana*, respectively. The *N. benthamiana* genes were named following the tomato nomenclature.

To finish the functional annotation of the Niben261 genome, an additional analysis was performed to identify transcription factors (TFs), transcriptional regulators (TRs) and protein kinases (PKs), using the program iTAK (Zheng et al., 2016). As a result, 2,677 TFs were found, divided in 68 families; 703 TRs classified in 25 families; and 1,635 PKs with 123 families.

3.4.3 New analysis using Niben261

The 96 hpi RNA-seq data analysis was repeated in order to include all the new information from Niben261. This time, instead of using AIR software from Sequentia, the analyses were carried out by our own using an array of software programs and packages for data manipulation, calculation and graphical display integrated in Linux and R. For the trimming step, the `fastq-mcf` command was used to remove low quality and too short reads. This time, more than 90% of the reads were preserved, compared to the ~50% of the first (AIR) analysis (Table 3.2). The mapping was performed with the Niben261 genome using STAR program (Dobin et al., 2013), obtaining results that were similar to the previous ones in AIR (~80% of uniquely mapped reads and ~10% of unmapped reads) (Table 3.2). However, since the number of clean trimmed reads was much higher in this new analysis, the number of mapped reads was also higher this time (~11 millions reads in the first approach vs ~22 millions). The mapped reads were counted for gene expression using HTSeq program (Anders et al., 2015), showing a similar distribution for all samples (Figure 3.9A). Principal component analysis (PCA) analysis separated two well-defined groups: GFP samples in one side and (p)crtB samples in the other, with the first component explaining more than the 94% of the differences (Figure 3.9B). DESeq2 was

used again for differential expression analysis, applying the same filters as the previous approach to identify DEGs (p-value ≤ 0.05, $\log_2$FC ≥ 0.585 for up-regulation and $\log_2$FC ≤ -0.585 for down-regulation), obtaining 3,571 and 2,269 up- and down-regulated genes, respectively.

Table 3.2: Summary of new RNA-seq analysis. Number of raw (untrimmed) and mapped to Niben101 genome sequences per sample, and percentaje of trimming and mapping.

Sample	Untrimmed sequences	% Trimming	% Mapping	Mapped sequences
GFP_96h_1	27,491,827	99.5	91.2	24,828,366
GFP_96h_2	27,644,786	98.9	90.9	24,912,885
GFP_96h_3	25,011,083	99.0	89.2	22,030,418
(p)crtB_96h_1	21,700,011	97.5	87.8	18,487,861
(p)crtB_96h_2	24,967,626	99.0	92.6	22,804,649
(p)crtB_96h_3	25,226,228	98.4	89.4	22,120,410
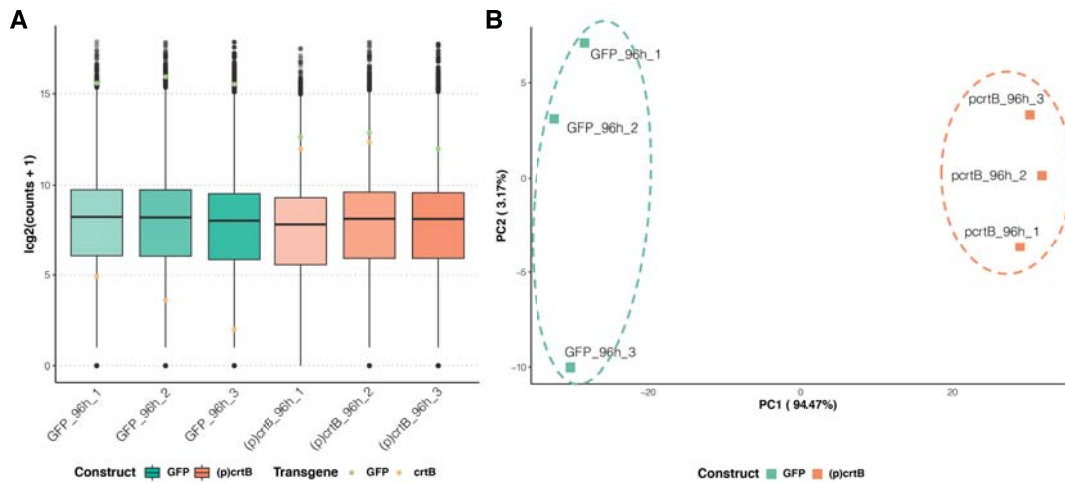
Figure 3.9: Preliminar analysis of GFP and (p)crtB samples at 96 hpi. (A) Distribution of $\log_2$ of counts. GFP and (p) crtB transgenes were highlighted in color. (B) Distribution of GFP and (p)crtB samples in a PCA.

In order to obtain human-readable information of these genes and to be able to compare them with other systems, the GO terms strategy was selected again, but with a different approach this time. Instead of using AgriGO software, topGO package was applied in R (Alexa and Rahnenfuhrer, 2021). The use of this package allowed us to implement the new Niben261 genome annotation that we created, obtaining more updated and accurate results. The number of enriched GO terms was notably higher in the three categories (BP, MF and CC). This pipeline was also used to analyze raw data from tomato fruit ripening and *Arabidopsis* leaf senescence, obtaining the corresponding list of DEGs and enriched GO terms. The lists of GO terms from the four systems (chromoplastogenic *N. benthamiana* leaves at 96 hpi, orange and red ripe stages in tomato and leaf senescence in *Arabidopsis*) were compared in a Venn diagram, showing only a 4% of overlap among GO terms in all datasets (Figure 3.10). When comparing only chromoplastogenic systems, the overlap was even lower, slightly higher than 2% (Figure 3.10A). However, among the 46 overlapped processes between the four systems, 42 (91%) of them were up- or down-regulated

both in *N. benthamiana* and tomato systems, while *Arabidopsis* showed an opposite profile in 19 (41%) BPs (Figure 3.10B).
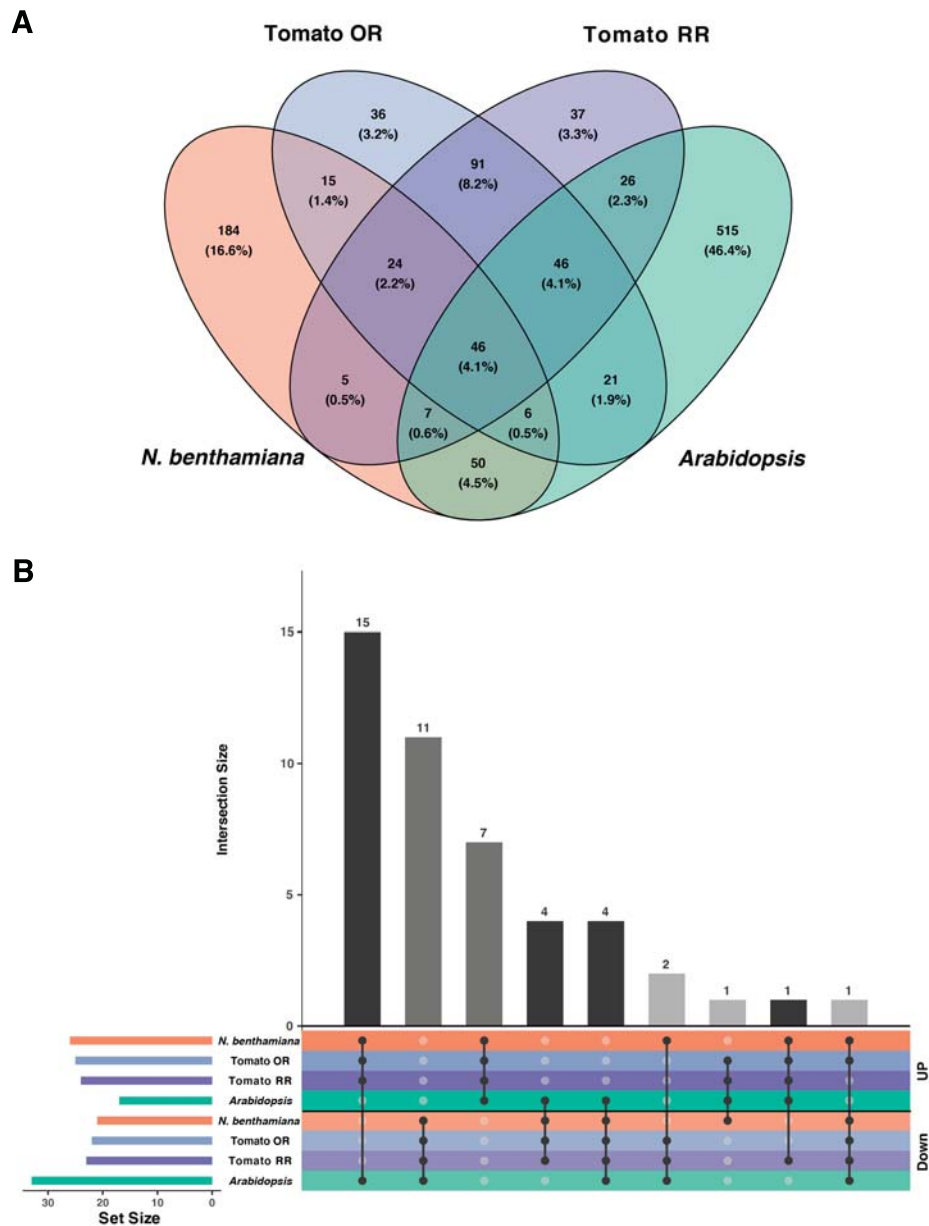
Figure 3.10: Low overlapping between systems. (A) Venn diagram of BPs up- and down-regulated in *N. benthamiana* comparison of (p)crtB vs. GFP, tomato fruit ripening comparisons of OR vs. MG and RR vs MG and *Arabidopsis* comparison of senescent vs non-senescent leaves. (B) UpSet diagram of the 46 overlapped BPs between the four systems. Each system present same colors as in the Venn diagram, but are divided in two: up-regulated (above) and down-regulated (below). Processes with the same profile in *N. benthamiana* and tomato, but opposite in *Arabidopsis* are represented in black bars; processes with the same profile in the four systems are represented in grey bars; processes with opposite profile in *N. benthamiana* and tomato are represented in light grey bars.

For a deeper study of gene expression at 96 hpi, the expression profiles of the different gene families that we manually annotated in the Niben261 genome were compared to the homologous in tomato and *Arabidopsis* experiments. Several genes

of the carotenoid biosynthesis pathway showed a significant induction at 96 hpi in *N. benthamiana* leaves (Figure 3.11). However, those encoding PSY showed no change in their expression, which makes sense since there was an external enzyme producing phytoene (crtB) and hence there was no physiological requirement to induce the expression of the endogenous phytoene-producing genes. In tomato fruit ripening, genes that encode enzymes producing phytoene and lycopene (PSY, PDS, ZDS, ZISO and CRTISO) were up-regulated, while those encoding enzymes transforming lycopene into downstream carotenoids were mostly down-regulated (Figure 3.11), resulting in the characteristic accumulation of lycopene that gives the red color to ripe tomatoes. During leaf senescence in *Arabidopsis*, most of the carotenoid pathway genes were very down-regulated (Figure 3.11). Carotenoid accumulation was only stimulated in *N. benthamiana* leaves and tomato fruit, but in a different manner. The entrance of the pathway was fully activated with high levels of phytoene synthases (crtB in agroinfiltrated leaves and PSY1 in tomato, producing and accumulating phytoene). However, the pathway was stopped at the lycopene level in tomato since the system is focused on the accumulation of this carotenoid, while the synthesis of other downstream carotenoids (mainly lutein and β-carotene) takes place in chromoplastogenic *N. benthamiana* leaves. On the other hand, senescence is not a process where carotenoid production is stimulated. The gene expression profile therefore parallel the metabolic (carotenoid) profile in these processes.

Regarding the supply of precursors for carotenoids and other isoprenoids, the complete gene sets of the two biosynthetic pathways (MEP and MVA) were identified. Only a few isolated genes from the plastidial MEP pathway were significantly up-regulated in *N. benthamiana* leaves, while most showed a strong up-regulation in the tomato fruit (Supplementary Figure 2A). By contrast, all MEP pathway genes, except DXS2, were down-regulated during *Arabidopsis* leaf senescence. The genes from the cytosolic MVA pathway showed a more heterogenous profile in tomato, with some isoforms up-regulated and others down-regulated, while the response was milder in *N. benthamiana* and *Arabidopsis* (Supplementary Figure 2B). The genes for downstream isomerases and GGPP synthases did not show differences in their expression compared to the control in *N. benthamiana*, while most of them were significantly induced or repressed in tomato fruit ripening (Supplementary Figure 2C). The expression of these genes was also affected in leaf senescence, also showing significant inductions and repressions. In summary, the artificial chromoplast differentiation in leaves does not imply the intense changes on the expression of the isoprenoid precursor biosynthetic genes that was observed in the tomato fruit ripening, but neither showed the general repression exhibited in a senescent process.

The expression of genes for fibrillins, reported to participate in carotenoid sequestration and considered as chromoplast markers, was also addressed. Except both *FBN2* and one *FBN3*, all expressed fibrillin genes showed up-regulated expression in tomato fruit ripening (Figure 3.12). The expression of this family was not so
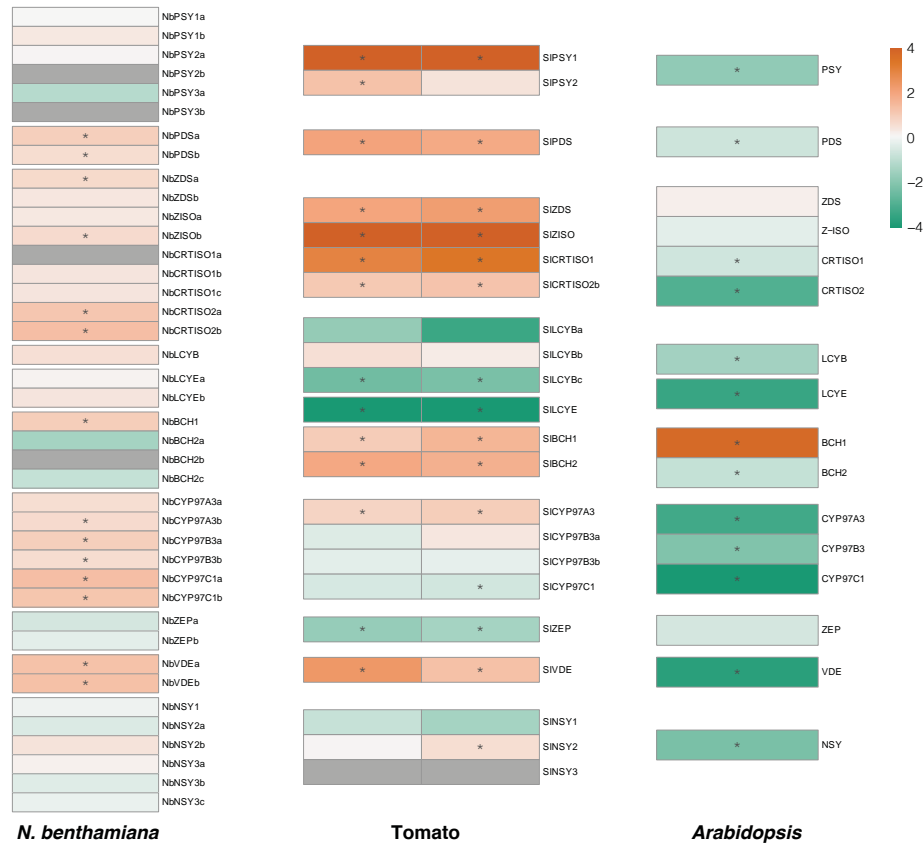
Figure 3.11: Carotenoids gene expression in *N. benthamiana*, tomato and *Arabidopsis*. Gene expression of carotenoid biosynthetic pathways in $\log_2$FC, in *N. benthamiana* comparison of (p)crtB vs GFP, tomato fruit ripening comparisons of OR vs. MG and RR vs. MG and *Arabidopsis* comparison of senescent vs non-senescent leaves.

affected in chromoplastogenic *N. benthamiana* leaves, but isoforms of *FBN1*, *FBN3*, *FBN7* and *FBN8* also showed an induction in their expressions. In opposition, fibrillin expression was mostly down-regulated in senescing *Arabidopsis* leaves, showing also more similarities between the chloroplast-to-chromoplasts systems than the chloroplast-to-gerontoplast transition.

Despite the unchanged levels of chlorophylls in chromoplastogenic *N. benthamiana* leaves, the interruption of photosynthesis and the loss of thylakoidal structures, genes for the biosynthesis of chlorophylls were up-regulated (Supplementary Figure 3A). Surprisingly, most of these genes were also up-regulated in tomato, where chlorophylls levels endure a pronounced decrease across ripening. On the other hand, this family of genes was mostly down-regulated in senescing *Arabidopsis* leaves. Genes from the chlorophyll cycle did not show any change on their expression compared to the control in *N. benthamiana*, while some of the homologs in tomato and *Arabidopsis* do show an up-regulation (Supplementary Figure 3B). Specific chlorophyll degradation genes were up-regulated in both *N. benthamiana* and tomato, but not the same homologs (Supplementary Figure 3C). In our system,
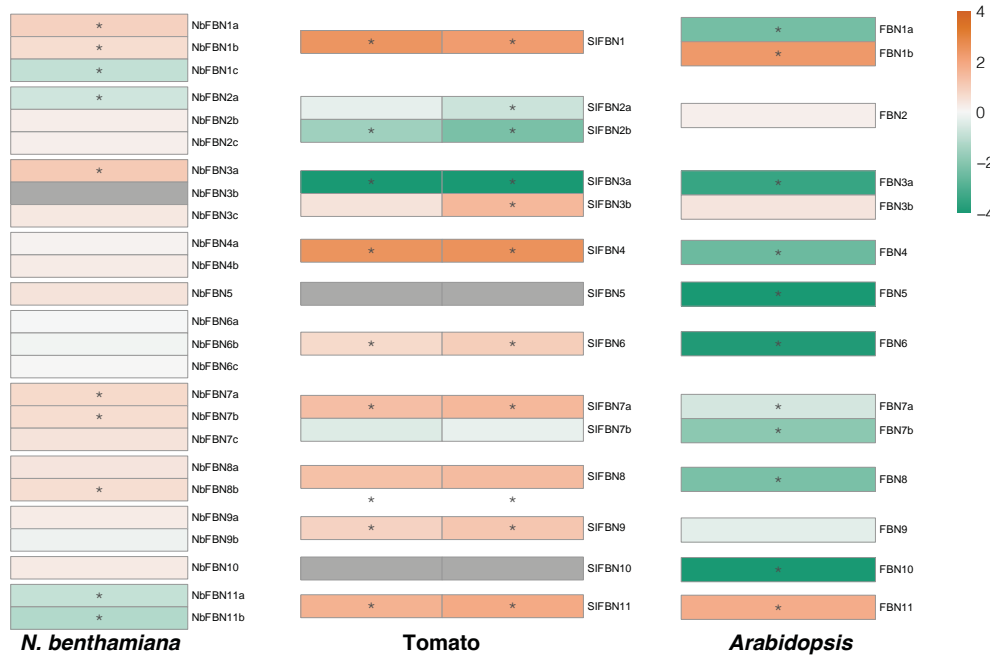
Figure 3.12: FBNs gene family expression in *N. benthamiana*, tomato and *Arabidopsis*. Gene expression of *FBNs* in $\log_2$FC, in *N. benthamiana* comparison of (p)crtB vs GFP, tomato fruit ripening comparisons of OR vs. MG and RR vs. MG and *Arabidopsis* comparison of senescent vs non-senescent leaves.

the up-regulation was on *CHL2a* and *CHL2b*, that encode the hydroxymethyl chlorophyll a reductase, that synthesize chlorophyllide a, but the homologs in tomato were down-regulated. Tomato fruit ripening activates the expression of *PPH2*, encoding a pheophytinase that catalyzes the degradation of pheophytin into pheophorbide a, while the homologs in *N. benthamiana* did not show any significant change on their expression (Supplementary Figure 3C).

A general induction was observed on the expression of different chaperones in *N. benthamiana*. We were especially interested on the HSP100 family of chaperones, since we already studied the protein levels of ClpB3, associated to chromoplast differentiation and reported to promote the correct folding and enzymatic activity of DXS. Most genes from this family showed an up-regulation in the three systems, but ClpB3 isoforms were only induced in *N. benthamiana* and tomato, while it was down-regulated in *Arabidopsis* (Figure 3.13).

3.5 crtB transgenic lines in *Arabidopsis*

All previous experiments were performed using transient expression of *(p)crtB* in *N. benthamiana* leaves. Transient expression has some disadvantages, like the interference of *Agrobacterium* infection, the small region affected on the plant, the heterogeneity of the samples or the temporality of the results. Most importantly, it does not work well in *Arabidopsis*, the model plant for genetic and transcriptomic studies

Figure 3.13: HSP100 gene family expression in *N. benthamiana*, tomato and *Arabidopsis*. Gene expression of **HSP100** in log$_2$FC, in *N. benthamiana* comparison of (p)crtB vs GFP, tomato fruit ripening comparisons of OR vs. MG and RR vs. MG and *Arabidopsis* comparison of senescent vs non-senescent leaves.

that will be essential to characterize the crtB-mediated artificial chromoplastogenesis process. To overcome these problems, different stable lines of *Arabidopsis* were designed to express different constructs of crtB under the control of constitutive or inducible promoters.

3.5.1 Constitutive expression

The first attempts to check whether crtB could also induce chromoplastogenesis in *Arabidopsis* were done using the *35S:(p)crtB* construct, and we also used the *35S:(c)crtB* construct as a control Table 3.3. The phenotype of *35S:(c)crtB* lines was undistinguishable from that of untransformed plants, with no changes in the green color or the carotenoid content. However, *35S:(p)crtB* lines showed a slightly yellowish phenotype on the leaves that could be reflected on the carotenoid profile, showing an accumulation of phytoene and increased levels of total carotenoids (mainly β-carotene and lutein) that were not observed in *35S:(c)crtB* lines (Figure 3.14). These results confirmed that *Arabidopsis* could be used as a model plant to study crtB-induced chromoplastogenesis.

3.5.2 Inducible expression

The constitutive expression of *(p)crtB* has the inconvenience of showing only the final effect of the protein in the plant. And because chromoplastogenesis impairs photosynthesis, transgenic lines with strong activity of the *crtB* transgen are expected to be lethal or to show a deleterious phenotype. To be able to study the sequence of events from normal chloroplasts to the establishment and differentiation
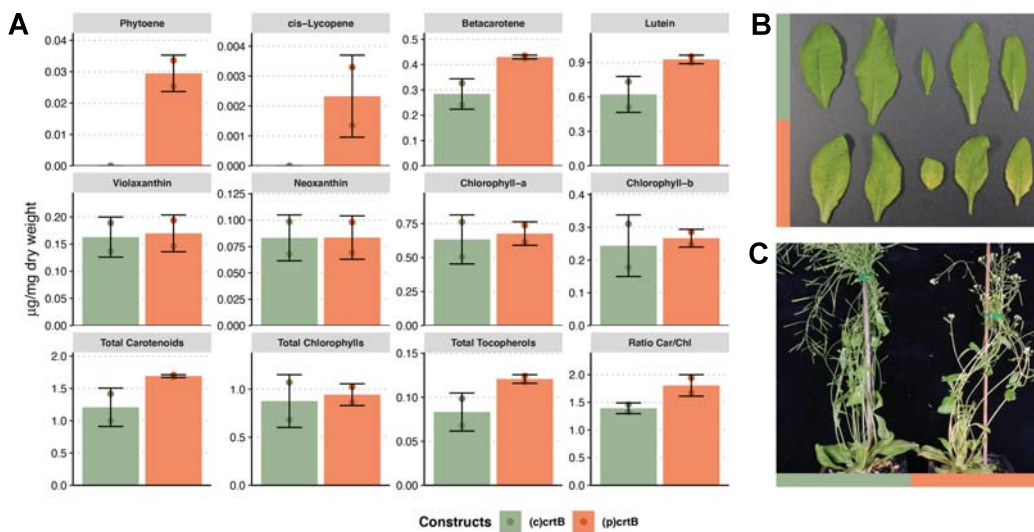
Figure 3.14: Constitutive expression of crtB in *Arabidopsis*. (A) Levels of carotenoids, chlorophylls and tocopherols in transgenic lines of *Arabidopsis* plants expressing *35S:(c)crtB* and *35S:(p)crtB*. (B) Representative leaves 38 days after germination. (C) Representative plants 38 days after germination.

of chromoplasts, inducible lines were designed in different backgrounds of *Arabidopsis*. Two different strategies to induce the expression of *(p)crtB* were considered: heat-shock induction (*pHsp70:(p)crtB*) and β-estradiol induction (*XVE:(p)crtB*) (Table 3.3). The heat-shock inducible *pHsp70:(p)crtB* lines produced phytoene when exposed to heat but did not develop a yellow phenotype in response to any heat treatment, probably because this strategy was found to only induce the expression of the transgene for a few minutes, while *(p)crtB* likely requires to be expressed for hours or days to produce the phenotype (Figure 3.15A). For β-estradiol induction, not only wild-type *Arabidopsis* plants were transformed with the *XVE:(p)crtB* construct, but also a PSY-defective line to test if crtB could complement the lethal (albino) phenotype of this mutants (Pokhilko et al., 2015) (Table 3.3). Complementation of this mutant was never observed despite β-estradiol was applied to the germination medium at different concentrations and times. However, when wild-type plants containing the same XVE:(p)crtB transgene were sowed in the presence of β-estradiol, dwarf yellow seedling germinated in all strategies, but they could not grow and generate mature plants, probably due to the lack of photosynthesis (Figure 3.15C). For a later induction, different strategies were attempted. Plants were transferred from non-supplemented medium to hydroponic cultures with β-estradiol (Figure 3.15B) or to plates with β-estradiol (Figure 3.15D). In other cases, the estrogen was applied directly to the leaves of soil-grown plants with a brush or sprayed all over the plants. However, most of the approaches did not show any phenotype, and in the few cases when leaves turned yellow, they did not show differences in carotenoid levels (Figure 3.15). These negative results prompted us to abandon the original idea of using *Arabidopsis* to study the time course of chloroplast-to-chromoplast transition at the transcriptomic level. Instead, we turned to agroinfiltrated *N. benthamiana* leaves as

the model system to address this question. The results are present in the following Chapter.
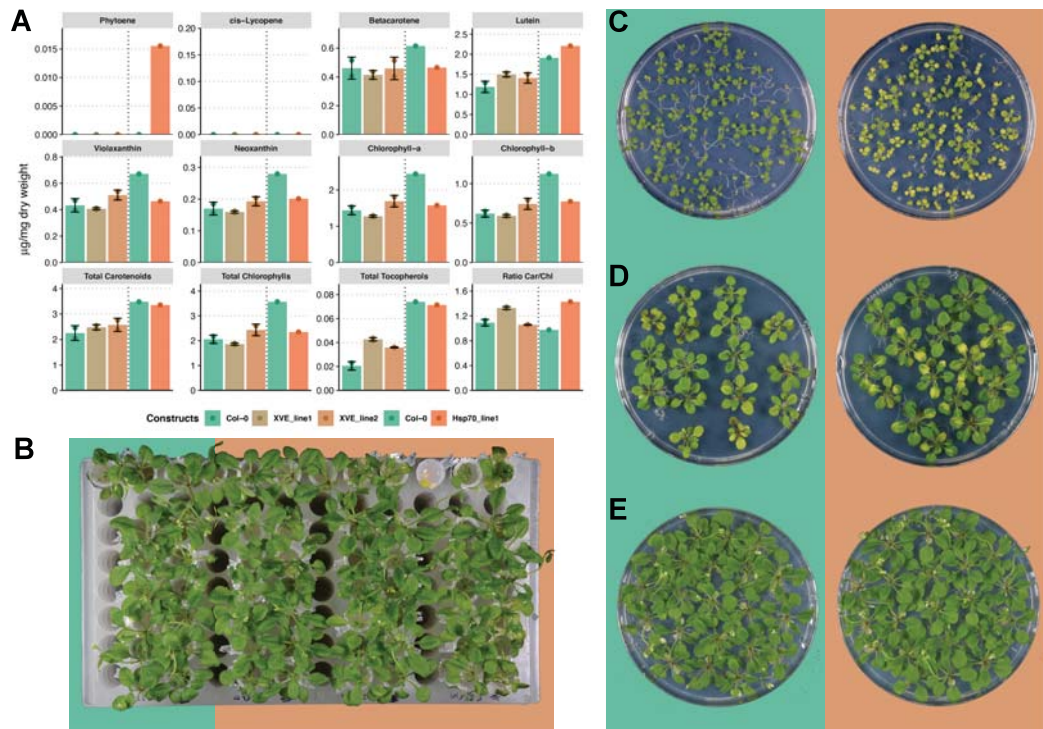
Figure 3.15: Inducible expression of crtB in *Arabidopsis*. (A) Levels of carotenoids, chlorophylls and tocopherols in transgenic lines of *Arabidopsis* plants expressing *XVE:(p)crtB* and *pHsp70:(p)crtB*. (B) *Arabidopsis* plants transferred to hydroponic cultures with β-estradiol. (C) *Arabidopsis* plants sowed in plates with β-estradiol. (D-E) *Arabidopsis* plants transferred to plates with β-estradiol. Wild-type are marked in green color, and different lines expressing *XVE:(p)crtB* are marked in brown.

Table 3.3: Transgenic lines.

Construct	Background	Ecotype	Inducible	Resistance
35S:(p)crtB	wild-type	*Col-0*	No	Basta
35S:(c)crtB	wild-type	*Col-0*	No	Basta
XVE:(p)crtB	wild-type	*Col-0*	β-estradiol	Hygromycin
XVE:(p)crtB	*psy*	*Col-0*	β-estradiol	Hygromycin
pHsp70:(p)crtB	wild-type	*Col-0*	heat-shock	Hygromycin

Chapter 2: Dynamic analysis of transcriptomic changes during chloroplast to chromoplast transition in *Nicotiana benthamiana* leaves

4.1 High-resolution transcriptional analysis by RNA-seq in *N. benthamiana* leaves

After comparing gene expression profiles of *N. benthamiana* leaves agroinfiltrated with GFP or (p)crtB at 96 hpi (i.e., harboring either chloroplasts or differentiated chromoplasts), we moved to analyze the timeline of gene expression events that take place as chloroplasts lose their photosynthetic identity and become competent to be transformed into chromoplasts, using our *N. benthamiana* leaf system.

4.1.1 Sampling, quality and clustering

N. benthamiana plants were agroinfiltrated with *(p)crtB* in one hemisphere of the leaf, using the other for the agroinfiltration of *GFP* as control. Samples were collected every 3 hours from 22 hpi until 46 hpi. Two additional time-points were collected at 56 and 68 hpi. Three replicates of (p)crtB sections and the corresponding GFP controls of the same leaf were collected for each of the eleven time-points. Carotenoid levels and photosynthetic efficiency (i.e., ϕPSII) were measured in every sample, showing an increasing and decreasing pattern, respectively, across the timeline in (p)crtB samples (Figure 4.1). Phytoene levels were not reliably detected until the second time-point (25 hpi), gradually increasing until 56 hpi, and showing a decrease at the last time-point (68 hpi). Downstream carotenoids, however, did not show a prominent increase until 40 hpi (Figure 4.1). While ϕPSII did not change in GFP controls, (p)crtB samples showed a slight decrease from the third time-point (28 hpi) until 34 dpi, followed by a lag period with no changes until 41 hpi and a prominent event of reduction at 46 hpi, arriving close to 0 from 56 hpi up

to the end of the experiment (68 hpi) (Figure 4.1). It is possible that the initial accumulation of phytoene causes the first drop in ϕPSII, whereas the accumulation of downstream carotenoids eventually leads to the second and more drastic reduction in ϕPSII (Supplementary Figure 4A). When (p)crtB/GFP ratios of total carotenoids and ϕPSII were plotted together, samples were linearly distributed according to time points, from left to right (ϕPSII) and down to up (carotenoid levels) (Supplementary Figure 4B). First and last time-points were clearly defined in terms of values clustering together, while samples values were more mixed across different time-points in the middle of the process.
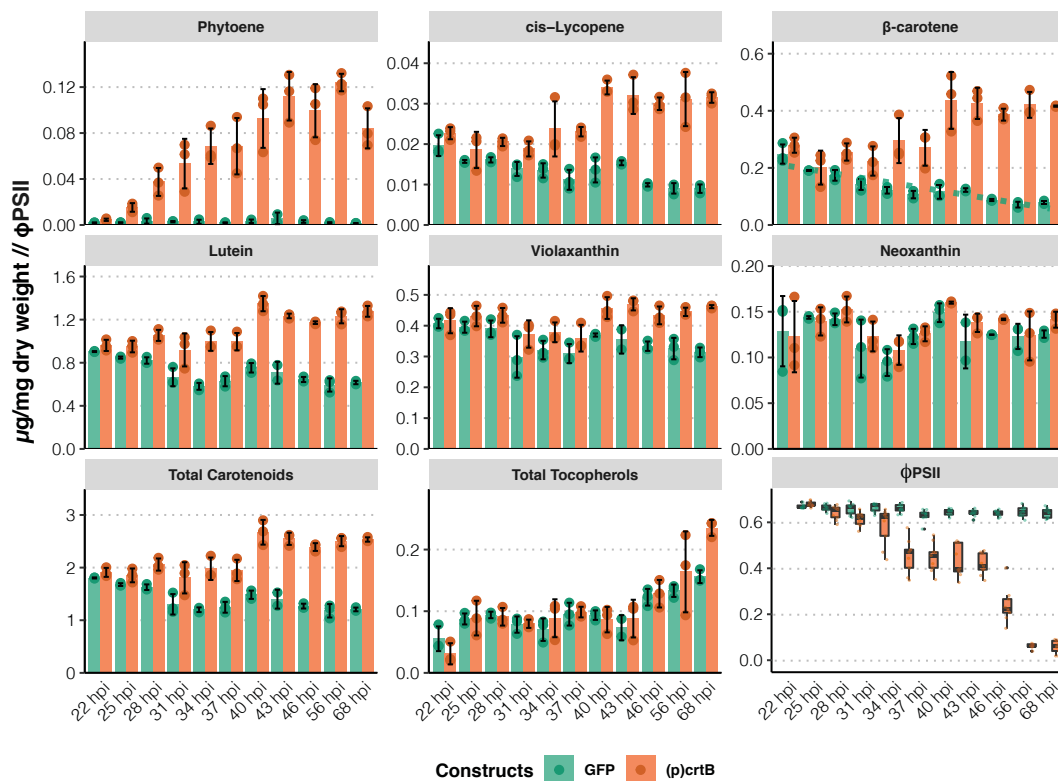
Figure 4.1: Carotenoids, tocopherols and ϕPSII in *N. benthamiana* leaves agroinfiltrated during time course. Levels of carotenoids, tocopherols and ϕPSII in leaves of *N. benthamiana* agroinfiltrated with GFP and (p)crtB, identified with different colors.

Eight of the eleven time-points were selected for the RNA-seq analysis based on carotenoid and ϕPSII data. These time points corresponded to events (i) before phytoene accumulation and ϕPSII changes (22 hpi), and occurring during (ii) beginning of phytoene accumulation and low slope drop of ϕPSII (25, and 28 hpi), (iii) lag period (34, 37 and 40 hpi), and (iv) high slope drop in ϕPSII and total carotenoids increase (46 and 56 hpi) (Supplementary Figure 4). After the sequencing, the last pipeline with Niben261 genome applied to the 96 hpi experiments was used (subsection 3.4.3). After the mapping and the counting of the transcripts, all samples showed a similar profile of their counts (Figure 4.2A). The *GFP* and *(p)crtB* transgenes were highlighted to check their abundance. All samples showed high levels

of *GFP* transcripts, which was expected since the *(p)crtB* construct includes a *GFP* sequence that is transcribed but not translated. The first time-point at 22 hpi showed lower levels of transgenes counts compared to the other samples, while these levels were quite similar among the other time-points. According to this, the transgenes did not reach the highest levels of expression until the second time-point (25 hpi), staying at these levels until the end of the time-course (Figure 4.2A). Detectable *(p)crtB* transcripts were observed in some *GFP*-agroinfiltrated samples, that could imply small levels of contamination.
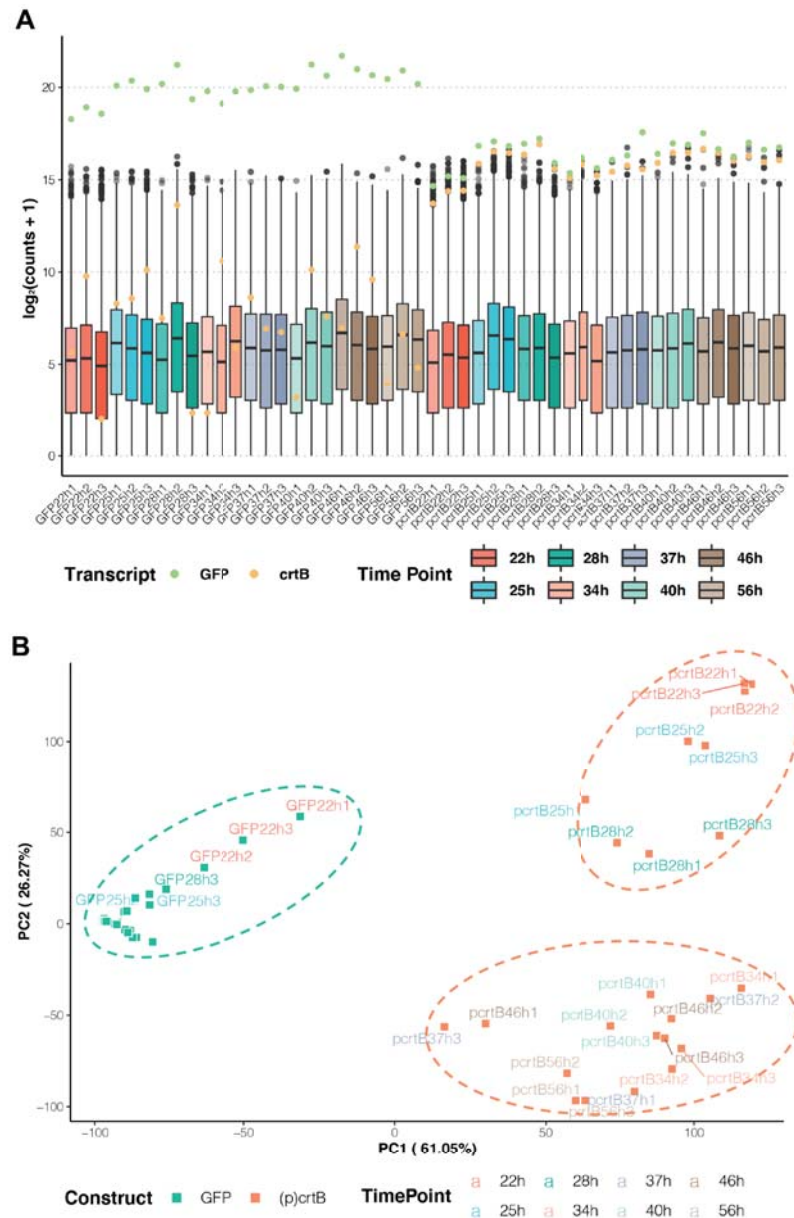
Figure 4.2: RNA-seq sequencing time course samples. (A) Distribution of $\log_2$ of counts in all time points. *GFP* and *(p)crtB* transgenes were highlighted in color. (B) Distribution of GFP and (p)crtB samples in a PCA. GFP samples are rounded in green, while (p)crtB samples are rounded in orange.

PCA plots were created to see the distribution of the samples. By only introducing the transcriptomic data to the PCA, PC1 ($\sim$61% of the variation) clearly showed the distribution of the samples according to the construct that was agroinfiltrated: GFP samples were all plotted on the left forming a compact and well-defined group, and (p)crtB samples were all plotted on the right but showing a higher separation among them as expected (Figure 4.2B). These results suggest that the possible contamination of GFP samples with (p)crtB constructs did not cause a relevant effect to the samples. PC2 ($\sim$26% of the variation) showed the differences related to the time component, showing the series from 22 to 56 hpi from the top to the bottom of the plot (Figure 4.2B). The (p)crtB samples could be divided in two different groups: early samples from events (i) and (ii) (22, 25 and 28 hpi) and late samples from events (iii) and (iv) (from 34 to 56 hpi). On the other side, GFP samples did not show that kind of variations related to the time component, being grouped in a more stable group (Figure 4.2B). When phenotypic data (carotenoid levels and ϕPSII values) was included into the PCA after a block normalization to adjust each type of data to their size, the plot showed similar results. GFP and (p)crtB samples were still well-separated according to the PC1 ($\sim$72% of the variation), but this time PC2 ($\sim$10% of the variation) showed more diversity in GFP samples and more separation between (p)crtB replicates than the previous PCA plot (Supplementary Figure 5).

A different approach was also performed to cluster the samples, called distance matrix heatmap. This cluster groups the samples according to the transcriptomic data, based on their correlation. Samples were grouped mainly based on the time variation, where event (i) (22 hpi) samples of GFP and (p)crtB were grouped in an independent cluster (Figure 4.3). Event (ii) samples (25 and 28 hpi) were grouped in another cluster, also separated by the construct variant. Event (iii) 34 and 37 hpi samples were organized in an independent cluster, although two replicates of (p)crtB were included with event (iv) 46 and 56 hpi samples. 40 hpi samples were separated in an independent group, while 46 and 56 hpi samples were organized by the construct variant (Figure 4.3). Taking all the information from the different cluster methods together, the samples were behaving as expected, as they could be easily clustered according to the construct that was agroinfiltrated (GFP or (p)crtB) and the time-point when they were collected.

4.1.2 Differential Expression Analysis and Gene Ontology Enrichment

Accordingly with the quality and clustering results, samples from each time-point were treated independently for the differential expression analysis by DESeq2 (Love et al., 2014), using GFP agroinfiltrated samples as reference. After applying
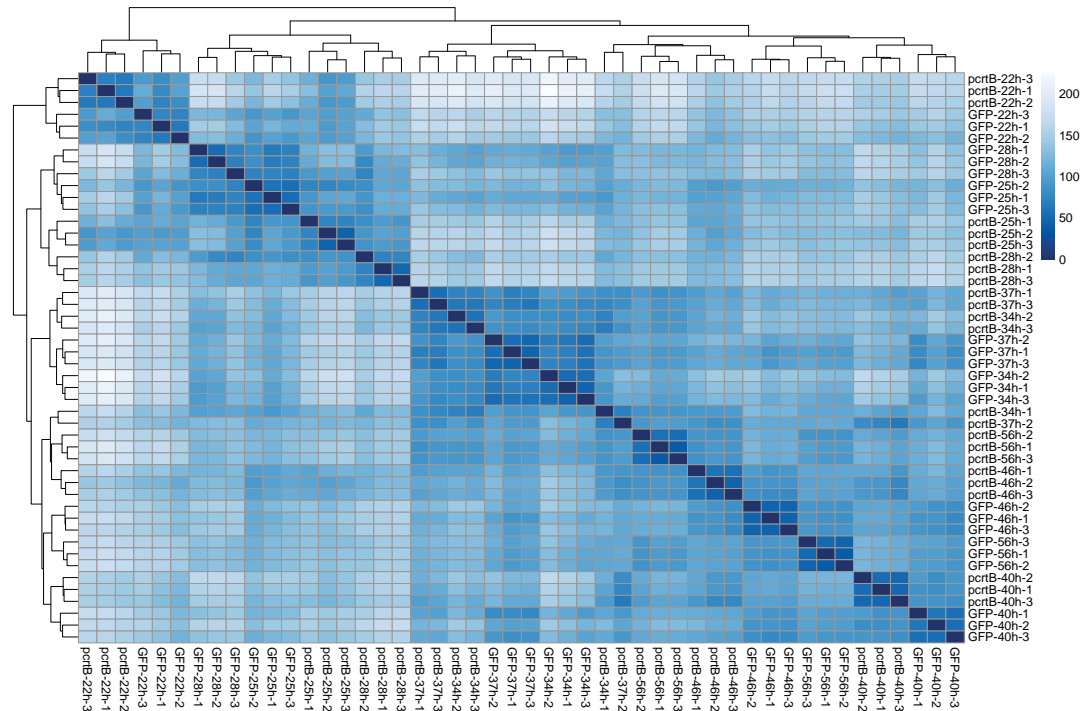
Figure 4.3: Distance matrix heatmap. Clustering of samples from time course in a distance matrix heatmap. Samples were clustered according to the distance (the higher the distance, the less relationship).

the same filtering as in the 96 hpi analysis (p-value $\leq$ 0.05, $\log_2$FC $\geq$ 0.585 for up-regulation and $\log_2$FC $\leq$ -0.585 for down-regulation) (subsection 3.4.3), the number of up- and down-regulated DEGs at the different time-points were plotted (Figure 4.4). There were two different hotspots of induction and repression of gene expression in the time-course, but not completely synchronized. The first big increase on gene induction was observed during event (ii), with a maximum at 28 hpi ($\sim$2.000 DEGs) (Figure 4.4A), followed by a decrease during event (iii) with a minimum at 37 hpi ($\sim$400 DEGs). The second hotspot of gene induction corresponded to event (iv), reaching the maximum at the last time-point ($\sim$2.500 DEGs). Down-regulation of genes was already notable at the first time-point, i.e., during event (i), that showed more than 1.500 DEGs (Figure 4.4B). Event (ii) 25 and 28 hpi time-points showed more than 2.000 down-regulated DEGs, but the peak was at 25 hpi. The valley of event (iii) was followed by a similar figure of more down-regulated genes during event (iv), as the induced genes, with a maximum at 56 hpi ($\sim$2.500 genes). The gene expression dynamics therefore appears to be strongly linked to the profile observed in the photosynthetic efficiency (ϕPSII), which in turn correlates with phytoene production and downstream carotenoid accumulation (Figure 4.1, Figure 4.4).

To check if the profile of changes in gene expression was gradual, Venn diagrams were created with all sets for up- and down-regulation. The resulting combinations were plotted according to their abundance in particular plots performed with the UpSetR package from R (Gehlenborg, 2019). Up-regulated genes showed a
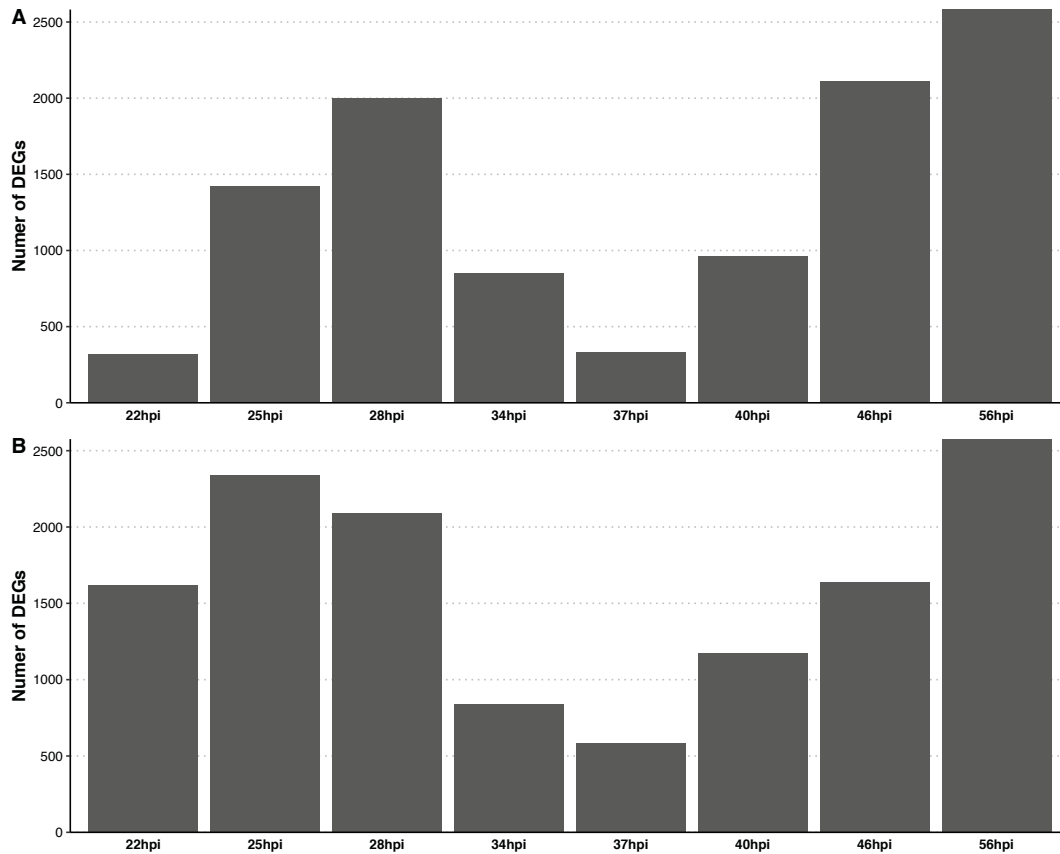
Figure 4.4: Distribution of DEGs in time course. Number of up- (A) and down-regulated (B) DEGs in each time point of the time course after DESeq2 analysis.

very heterogeneous distribution with low numbers of DEGs induced in more than one time-point. For instance, the three most abundant groups of induced DEGs were those that were only differentially expressed in 56, 28 and 46 hpi, respectively (Figure 4.5A). However, event (iv) 46 and 56 hpi, and event (ii) 25 and 28 hpi shared relatively high numbers of induced DEGs (651 and 333, respectively), but the numbers of DEGs that were differentially induced in three or more time-points were very low (e.g., only 27 genes were induced in all time-points from 25 to 56 hpi). Down-regulation was very heterogenous as well, but there were more genes differentially repressed in different time-points than the induced ones (Figure 4.5B). 145 genes were differentially down-regulated in all time-points, 82 from 25 to 56 hpi and 75 from 34 to 56 hpi.

Gene Ontology enrichment was performed with topGO (Alexa and Rahnen-fuhrer, 2021) to see the biological processes, molecular functions and cellular components representing the DEGs of each time-point. The distribution of the GOs among the different time points was also analized using UpSetR package, generating the same type of plots as before. In general, the distribution of GOs was similar to the one observed with the DEGs, showing a predominance of GO terms that represent DEGs from only one of the time-points and a relatively high number of terms shared by the two timepoints that represent events (ii) and (iv). For Biological Processes
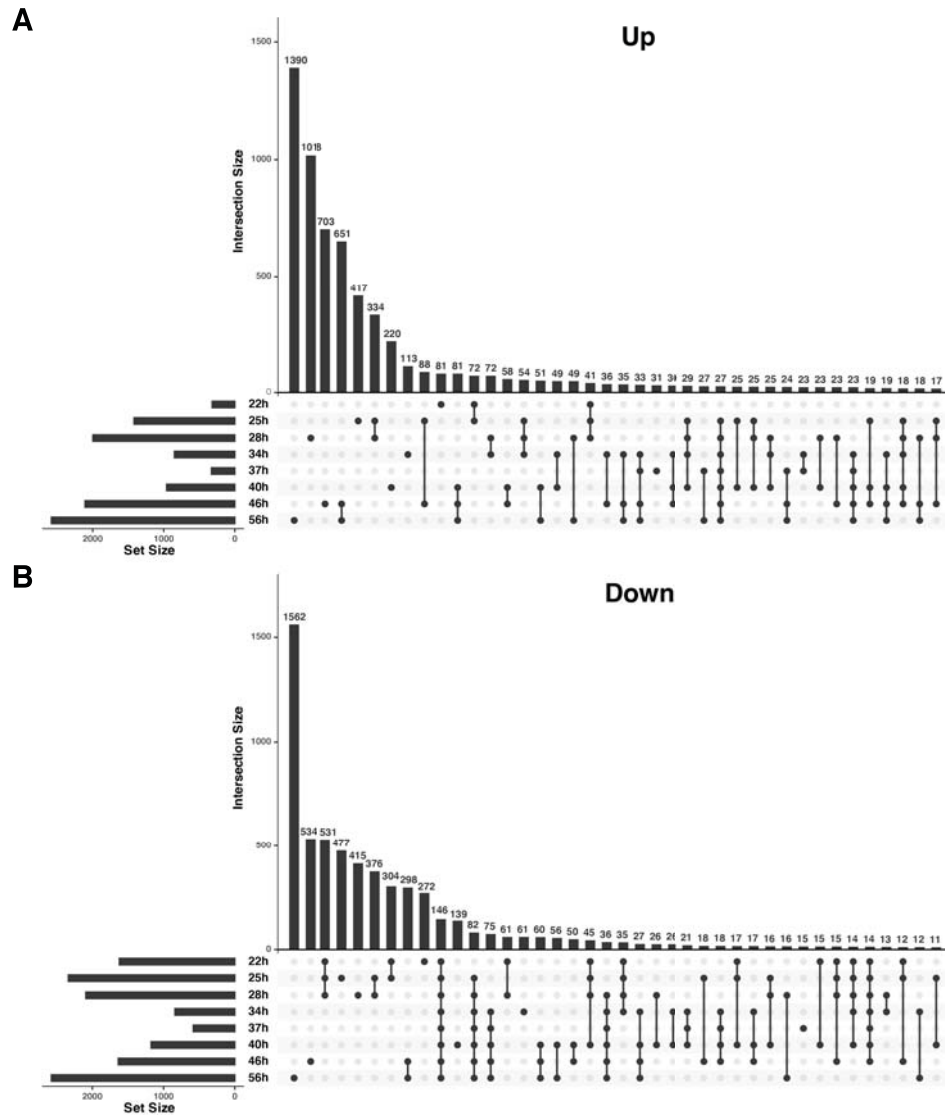
Figure 4.5: UpSet diagram of DEGs in time course. Distribution of up- (A) and down-regulated (B) DEGs of the different time points. Groups were distributed according to their abundance (from left to right). Number of DEGs in each group are represented in bar plots on the top; time points including in each group are highlighted below; number of DEGs in each time points are represented in horizontal bar plots on the left.

(BPs) that represented up-regulated DEGs, there was a massive dominance of exclusive terms (248 out of 340 terms represented a single time-point, 71%) (Supplementary Figure 6A). However, among those few GOs that represented many time-points were interesting processes like "cell wall biogenesis" and "xyloglucan metabolic process" from 22 to 56 hpi, "isoprenoid biosynthetic process", "negative regulation of endopeptidase activity" and "response to wounding" from 25 to 56 hpi, and different terms related to photosynthesis, terpenoid and sterol metabolism, jasmonic acid and cell wall representing 4 or more time-points (Venn_TimeCourse_BP_Up.txt). Only 27, 10 and 16 terms were exclusively included in events (i), (ii) and (iv), respectively. These terms were very diverse, but specific ones related to plastids were included, including "chloroplast-nucleus signaling pathway" in event (i) or

"1-deoxy-D-xylulose 5-phosphate biosynthetic process" in event (ii). Most of the BPs that represented down-regulated genes were also exclusive, but less than the induced ones (193 out of 411, 47%) (Supplementary Figure 6B). Relatively many processes (24) were representing the first stages of the process (from 22 to 28 hpi), as well as the full set from 22 to 56 hpi. This last group of processes was enriched in protein folding and response to abiotic stress (Venn_TimeCourse_BP_Down.txt). 19, 16 and 13 BPs were specifically included in events (i), (ii) and (iv), respectively, but they were very diverse. The GOs showed a similar profile for Molecular Functions (MFs), very exclusive but more for up-regulation than down-regulation (Supplementary Figure 6C-D). Some functions related to carotenoid pathway were in the few cases of representing induced genes from many time-points (Venn_TimePoints_Cornell_MF_Up.txt), while also the chaperone activities and protein folding-related functions were predominant among the most shared MFs from down-regulated genes (Venn_TimePoints_Cornell_MF_Down.txt). Chloroplast, membranes and cell wall were the most affected Cellular Components (CCs) for up-regulated DEGs at the different time-points (Venn_TimePoints_Cornell_CC_Up.txt), while the endoplasmic reticulum, different regions and components of the cytosol and the mitochondria were more affected for the down-regulation (Venn_TimePoints_Cornell_CC_Down.txt).

4.1.3   Incorporation of end-point results at 96 hpi

In the first chapter of this thesis, RNA-seq data from GFP and (p)crtB samples at 96 hpi were analyzed and compared to those from plant systems in which chloroplasts were transformed into chromoplasts or gerontoplasts (subsection 3.4.3). Phenotypically, 96 hpi samples corresponded to chromoplasts based on lack of photosynthesis and high carotenoid contents, a condition that was already evident at the end of event (iv), in 56 hpi samples. After adding the 96 hpi (i.e. end-point) data to the time-course and clustering all samples in a PCA plot, unexpected results were observed. While all samples were separated in two well-defined groups according to the construct that was agroinfiltrated (GFP or (p)crtB) when the experiments were analyzed separately (Figure 3.9A and Figure 4.2B), three different groups were found when both time-course and end-point analyses were clustered together (Figure 4.6A). GFP and (p)crtB samples from the time-course experiment were still clustering separately in two different groups, but samples from 96 hpi were grouped together in an independent third group (Figure 4.6A). Only (p)crtB samples from the 22 hpi time-point were included in this third group.

Despite this issue, DEGs from 96 hpi were compared to the other time-points to see if there were similarities after the normalization of the data. Most DEGs at 96 hpi were very exclusive of this stage and there was no close relationship with any of the other time-points (Supplementary Figure 7). Even in the more specific comparison
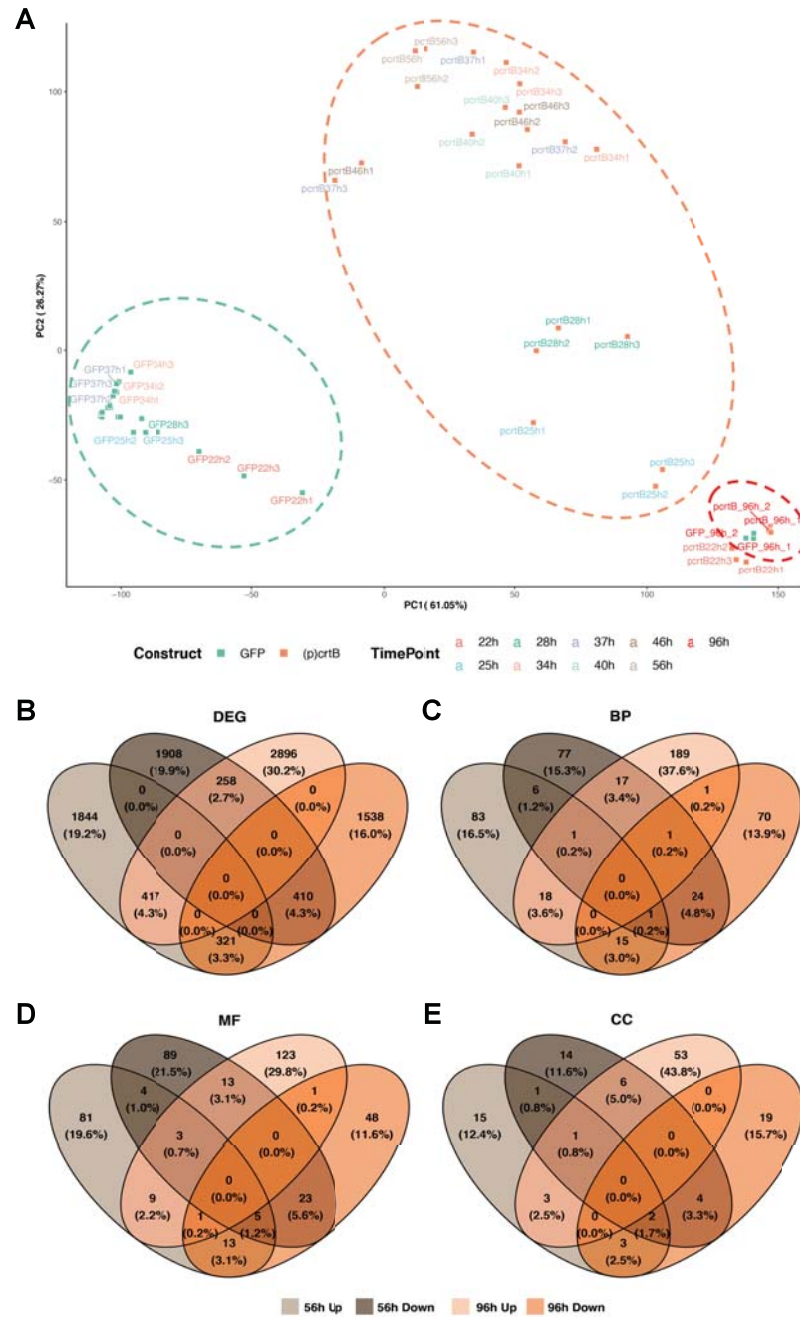
Figure 4.6: Comparison Time course vs end-point analyses. (A) PCA plot including transcriptomic data from all samples in the time course and the end-point experiments. Green circle surrounds GFP samples from time course; orange circle surrounds (p)crtB samples from time course; and red circle surrounds all samples from end-point. Venn diagrams comparing up- and down-regulated DEGs (B) from 56 and 96 hpi, and the BPs (C), MFs (D) and CCs (E) representing them.

with 56 hpi, overlapped DEGs were residual and there were almost the same correlations of Up-Up and Down-Down than Up-Down or Down-Up (56 and 96 hpi, respectively) (Figure 4.6B). When GOs from both time-points were compared, similar results were found, with less than 10% overlapped BPs, MFs or CCs (Figure 4.6C-E).

Both 56 and 96 hpi experiments were therefore showing different frames of the process from the transcriptomic point of view, in spite of their phenotypical similarities.

The results suggest that once the process reaches the final phenotypic transformation of chloroplasts into chromoplasts (as deduced from the yellow color of the tissue, higher levels of carotenoid content, and dissipation of photosynthesis), the system does not stay static and gene expression continues changing. An alternative explanation would be the existence of experimental differences between time-course and end-point experiments, as samples were agroinfiltrated and collected in different seasons of the year (with 30 months between the two experiments), different plants were used, and different services sequenced the RNA. To distinguish between the two possible explanations for the poor similarity of the transcriptomes from 56 and 96 hpi samples, the expression of some marker genes was tested by qPCR in a third independent experiment. These marker genes (*Chap60*, *Hsp18* and *Hsp3*) were selected based on their opposite abundance in (p)crtB vs. GFP samples at 56 and 96 hpi (Figure 4.7). In the new experiment, *N. benthamiana* leaves were agroinfiltrated with *GFP* and *(p)crtB* constructs and samples were collected at 4 time-points: 24, 48, 72 and 96 hpi. In the RNA-seq experiments, these marker genes were expressed at lower levels in (p)crtB vs. GFP samples at the time-course event (iv), i.e., 46 and 56 hpi, but at higher levels in (p)crtB vs. GFP samples at the end-point 96 hpi. Similar results were obtained by qPCR in the third experiment (Figure 4.7), supporting the conclusion that the observed changes in gene expression between the first two experiments were not due to experimental variations. In conclusion, 96 hpi represents the consequences of having chromoplasts instead of chloroplasts in leaves for several days, while the time course from 22 to 56 hpi would inform us about the consecutive steps in the process of deconstruction of chloroplasts and their initial transformation into chromoplasts in leaves.

4.1.4 Similarities and differences with other ripenning systems

In order to test if the synthetic conversion of chloroplasts into chromoplasts in leaves was similar to the natural chloroplast-to-chromoplast differentiation in fruit ripening, enriched GOs from the different time-points in our system were compared to those representing different stages of tomato and pepper fruit ripening. To draw more robust conclusions, two different datasets of tomato fruit ripening were included. One of them was already used in the first part of the thesis (section 3.4) (Shinozaki et al., 2018), while the new one derived from the original work published by (Yazdani et al., 2019). In the second study, tomato lines overexpressing the chromoplastogenic chaperone OR[His] were generated and RNA-seq analyses were carried out at different fruit ripening stages. Chromoplast differentiation was ahead in this line compared to the wild-type, so OR[His] fruits were already orange at the mature green (MG) stage while they were still green in the wild-type plants (Yazdani et al., 2019). For our analysis, we decided to compare the transgenic fruits vs. the wild type
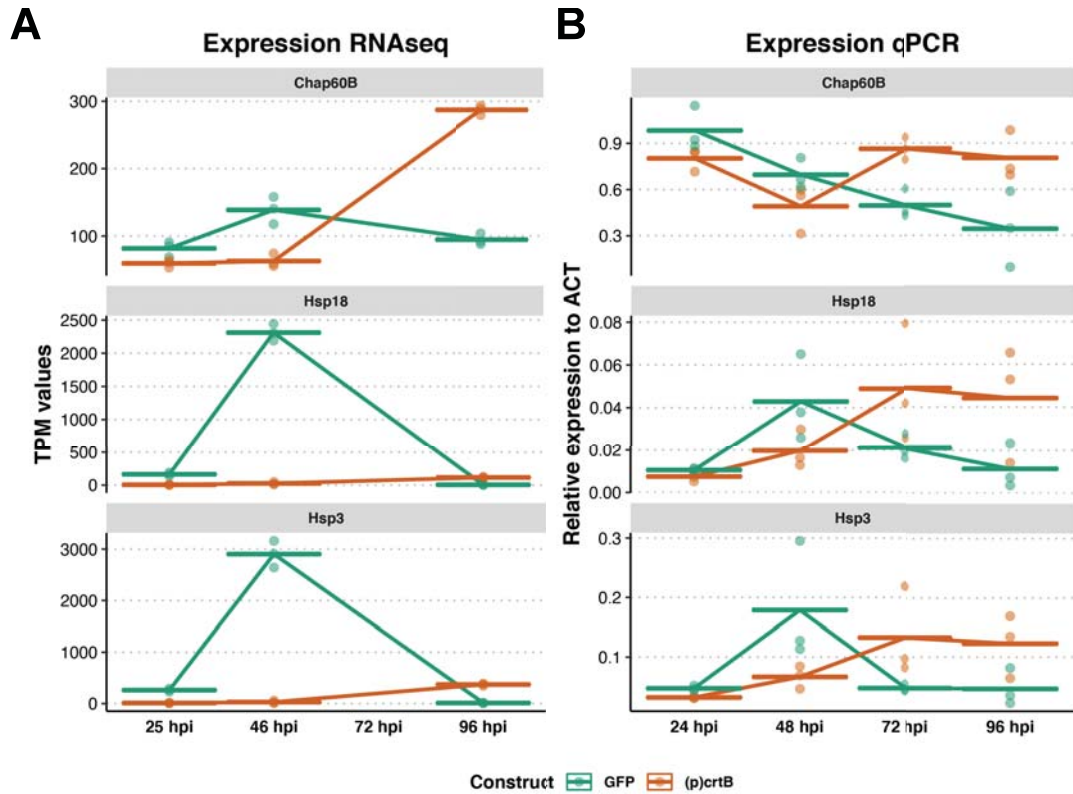
Figure 4.7: Gene expression of markers genes. Transcript levels of three marker genes measured by RNA in Transcript Per Kilobase Million (TPM) (A) and by qPCR (B), relative to *Actine* (ACT) levels.

at the same ripening stage (MG and Breaker) in order to add another artificial system of chromoplast differentiation. The same experiment of *Arabidopsis* leaf senescence used for the analysis of 96 hpi samples was also included as a control of chloroplast dismantling not leading chromoplast differentiation. In summary, we included (a) three different systems of natural chromoplast differentiation in fruit (two in tomato and one in pepper), comparing different ripening stages vs. MG (Breaker, Orange and Red in tomato; Breaker and Red in pepper), (b) two different systems of synthetic chromoplast differentiation (one in tomato fruit and other in leaves of *N. benthamiana*), comparing the transgenic samples vs. the control at the same time point (OR[His] vs. wild-type in tomato; (p)crtB vs. GFP in *N. benthamiana*), and (c) one extra system of leaf senescence in *Arabidopsis*, comparing old senescent leaves (containing gerontoplasts) vs. mature photosynthetic leaves (containing chloroplasts). The actual data from all these experiments (González-Gordo et al., 2019; Shinozaki et al., 2018; Yazdani et al., 2019; Woo et al., 2016) were retrieved from the SRA database and analyzed following the same pipeline used for our own time course data.

The GO enrichment analyses resulted in long and heterogenous list of GOs in all systems. For simplification, the medium (0.7) filter of REVIGO software (Supek et al., 2011) was applied to all of them to reduce the redundancy. We were only focused on BPs since this is the most meaningful category for our purpose of understanding the biology of the chromoplastogenesis process, and we divided the terms in up-

and down-regulated, regarding the DEGs that they were representing. Fruit ripening systems showed a similar number of BPs representing the up-regulated DEGs in their different stages, while *N. benthamiana* showed the above-mentioned profile of two peaks, with the largest number of BPs at 96 hpi (Figure 4.8A). On the other hand, down-regulated BPs had a different profile among the systems. Tomato fruit ripening was again very homogeneous, while pepper fruit ripening and the synthetic differentiation of chromoplasts in OR^{His} tomato showed an increment through the process (Figure 4.8B). A two peaks profile was defining the process in *N. benthamiana* again, but the largest number of terms was at the beginning (25 hpi). *Arabidopsis* showed the highest number (2-fold) of BPs affected, both up- and down-regulated (>200 terms each).

In order to compare the presence of specific GO terms in the different systems, some categories were created regarding the sets that each term was representing. "Exclusive" refers to those that were only represented in a single set; "Tomato-Ripening" was given to those terms that were present in more than one set exclusively from the tomato fruit ripening systems; "Pepper-Ripening" was similar to the previous one, but only from the pepper system; terms from "Ripening" category were those terms that represented more than one set of samples in the tomato and pepper fruit ripening and not in any artificial chromoplast differentiation system (tomato *OR^{His}* fruit or *N. benthamiana* agroinfiltrated leaves); "Nicotiana-independent" were terms that were representing different sets of samples from any of the systems except for *N. benthamiana*; "OR-Dependent" represented terms from both sets of the artificial chromoplast differentiation in *OR^{His}* tomato and not from any other system; "Ripening-independet" was given to BPs included in the two artificial systems and leaf senescence but not in any of the fruit ripening sets; "Nicotiana" terms were representing more than one set from the *N. benthamiana* system alone; "Fruit-independent" was given to terms that were representing only in *N. benthamiana* and *Arabidopsis*, the two leaf systems; and "Spread" were those terms that could not be included in any of the categories described before, since these terms were spread across all systems. After labeling all terms from the different systems according to this categorization, the results were plotted to see the distribution of categories in each set, both for up- and down-regulated processes (Figure 4.8). The distribution of the different categories among the systems was very heterogeneous. Only tomato fruit ripening showed a very similar profile in all samples, with a low proportion of "exclusive" terms. This homogeneity among the two different experiments of ripening in tomato fruit indicates a big consistence of the approach and that the experimental noise was not significant. On the other hand, fruit ripening in pepper showed a notable disparity compared to tomato. The number of "exclusive" terms was high (from 30 to 50%) and the number of terms shared with tomato was low. Hence, it can be seen that fruit ripening works differently in different species, even from the same family, so the chromoplastogenesis-derived processes represent a small proportion of similarities. Even more disparate was the *OR^{His}* transgenic
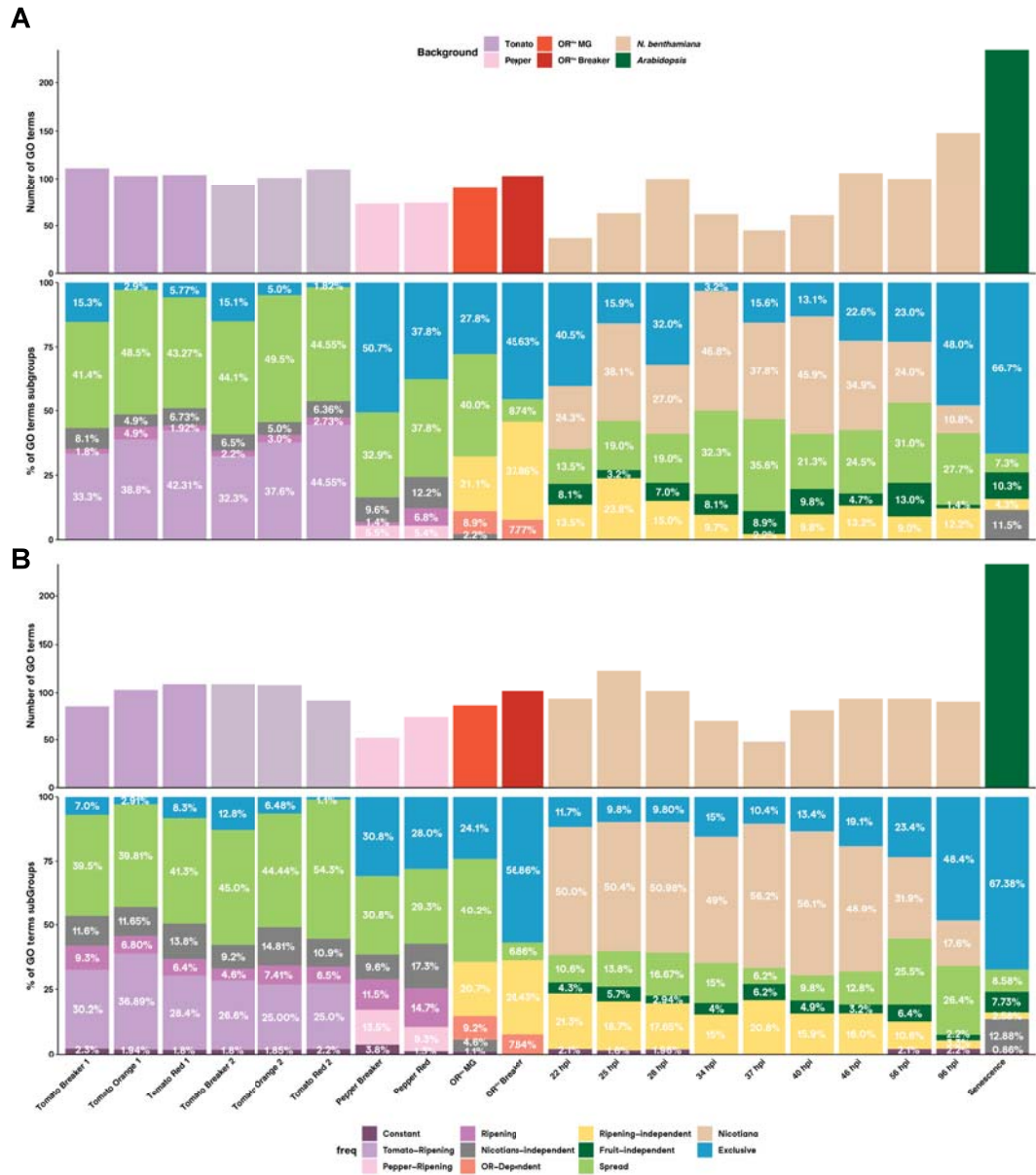
Figure 4.8: Distribution GO terms in different plastid transition experiments. Absolute and relative abundance of up- (A) and down-regulated (B) BPs in different sets of experiments. Absolute abundances are colored according to the different experiments, while relative abundances are colored according to the different categories of distribution of GO terms.

vs. wild-type comparisons in tomato, specially at the Breaker stage. Although same conditions were used in one of the tomato fruit ripening experiments, performed by the same researchers at the same time, the overexpression of OR^{His} causes a big impact on the fruit that did not correspond to an acceleration of the fruit ripening program. However, the most distinct system was the leaf senescence in *Arabidopsis*, showing 2/3 of "exclusive" terms and less than 10% of "Spread" terms. *N. benthamiana* samples contained a considerable proportion of "Nicotiana" and "Spread" terms, while the "Exclusive" category was relatively low. Also "Fruit-independent" terms were not abundant, showing a bigger resemble with the chromoplast differentiation systems than with *Arabidopsis*, despite they were the only two systems in

leaves. Only the 96 hpi time point was considerable different, representing a more independent event. In summary, there is a low level of similarities between different systems showing chloroplast-to-chromoplast differentiation, but there is still a correlation between them that did not correspond to the non-chromoplast system.
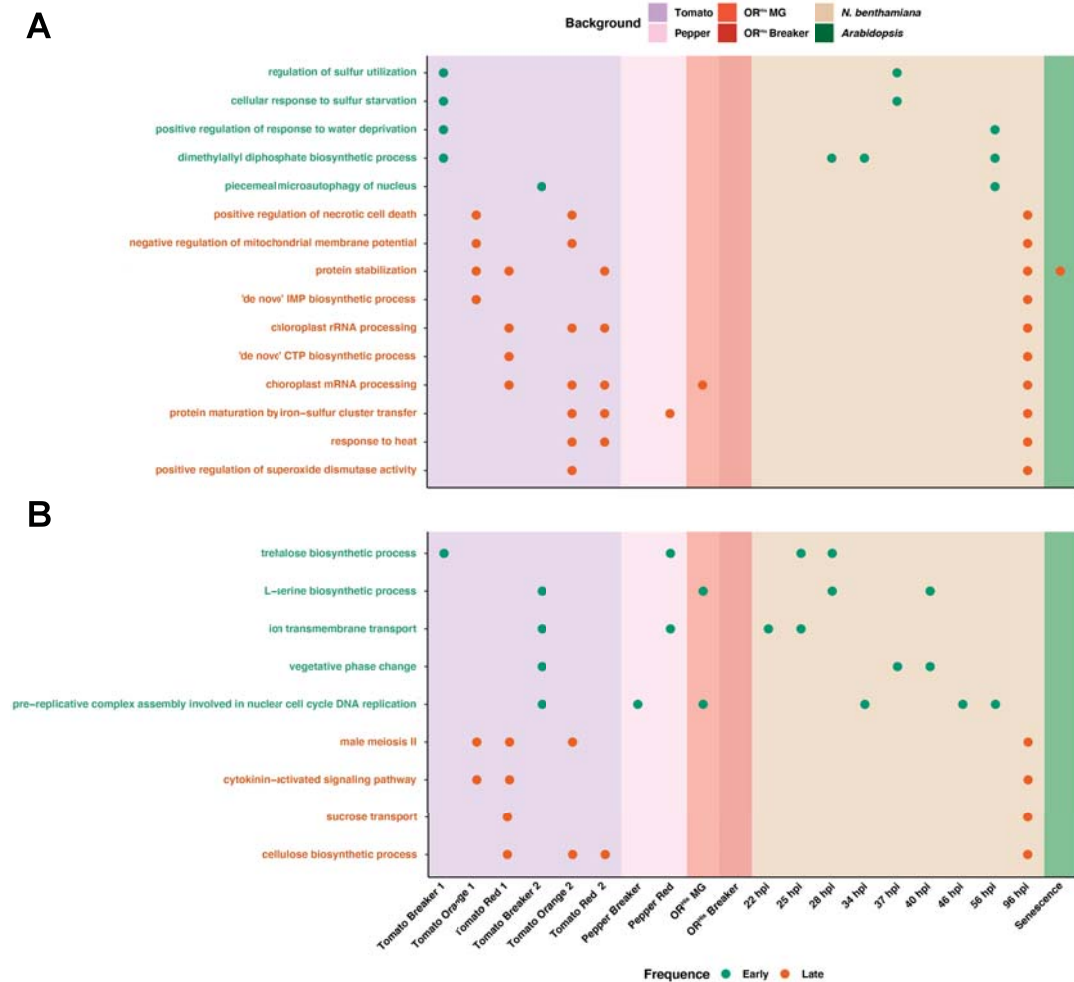
Figure 4.9: Selection of "Spread" GO terms. Subset of terms from Supplementary Figure 8 and Supplementary Figure 9, distributed in up- (A) and down-regulated (B) BPs in the different sets of samples, differenciated in "early" (green) and "late" (orange). Background was colored according to data sets (Figure 4.8).

Terms that were representing the four chromoplastogenic systems were very scarce, encompassing only "rhythmic process" for up-regulation, and "microtubule-based process", "auxin-activated signaling pathway", "sterol biosynthetic process" and "pre-replicative complex assembly involved in nuclear cell type DNA replication" for down-regulation (Supplementary Figure 8 and Supplementary Figure 9). All the terms from this category were re-categorized depending on the time-points of *N. benthamiana* that were represented: if only terms from time course samples were represented, they were labeled as "early" processes; if they represented only end-point at 96 hpi, the category was "late"; and if they represented time-points dispersed throughout both experiments, they were named as "disperse". There was

a high number of terms that did not show a clear association with other systems. Tomato was the most robust system, showing a big homogeneity and abundance of terms shared with *N. benthamiana*. In order to simplify and facilitate the interpretation of information, a subset of terms was performed according to the correspondence with tomato ripening systems: "early" processes that were also representing a Breaker stage of tomato, and "late" processes that were also representing any Orange or Red stages of tomato (Figure 4.9). The resulting lists of BPs was not very informative, since most of them were not representing both experiments of tomato fruit ripening, nor pepper or transgenic OR^{His} tomato systems. In conclusion, the diversity of processes among different systems showed the heterogeneity of chromoplastogenesis, without a well-defined transverse associated pattern to explain it. For that reason, we decided to focus on our most comprehensive approach, the 22 to 56 hpi time course, in order to obtain the most informative description of our system and to simplify its complexity.

4.1.5 Weighted Gene Co-Expression Network Analysis: all genes

The ensemble of time-points from 22 to 56 hpi included all events from the increase of phytoene production and accumulation after *(p)crtB* expression to the chromoplast establishment in a yellow leaf, with higher carotenoid levels and without photosynthesis. As it was shown before, the chromoplastogenesis process continued after the last time-point (56 hpi), but we decided to not include the additional samples of 96 hpi in order to focus on the critical steps that determine the dismantling of chloroplasts and the initial build-up of chromoplasts (subsection 4.1.3, subsection 4.1.4). The application of GO terms that represented the DEGs from the different time-points was not efficient due to the low level of overlapped processes and the difficulties to find any dynamics of events from the beginning to the end of the process. A new approach based on the clustering of genes according to their co-expression profiles was applied to the data. This method, named Weighted Gene Co-expression Analysis (WGCNA) (Langfelder and Horvath, 2008), is included in a package from R that takes normalized expression values from all samples as input, creates groups (or clusters) of genes based on their co-expression, and associates them to the phenotypic data that was also provided (e.g. sampling time, type of construct or carotenoid levels). A total of twenty-two clusters were generated, each one identified with a color, and correlated with the phenotypic data (Supplementary Figure 10). Average expression of all genes from a cluster was plotted for each sample, in order to see the profiles of each cluster with respect to time and construct variables (Figure 4.10). Only six of the modules showed a clearly differential expression profile in control (GFP) vs. (p)crtB samples throughout time: "cyan" cluster showed higher levels of expression in (p)crtB samples in the first time-points; "lightcyan" contained genes more expressed in (p)crtB samples at the end of the time course; "lightyellow" was a very interesting cluster with higher expression levels in (p)crtB samples from the 28 hpi until the end; "magenta" cluster were genes showing

a peak of expression at 28 hpi only in (p)crtB; "pink" cluster were genes induced in GFP samples from 34 hpi forward, but not in (p)crtB; and "tan" was a cluster with down-regulated genes in (p)crtB samples at the two last time-points. The content of DEGs of these clusters was contrasted and showed a correlation with the dynamics observed before (Figure 4.10). Hubgenes at each cluster were defined as those genes with a high Module Membership (≥ 0.8) that better represented the profile of the cluster. Hubgenes were then correlated to the expression value of the (p)crtB transgene to find those that were more related to the treatment (GeneSignificance ≥ 0.2). Some of these hubgenes were related to isoprenoid biosynthesis in some clusters, like "magenta" or "lightyellow" Tables HubGenes.
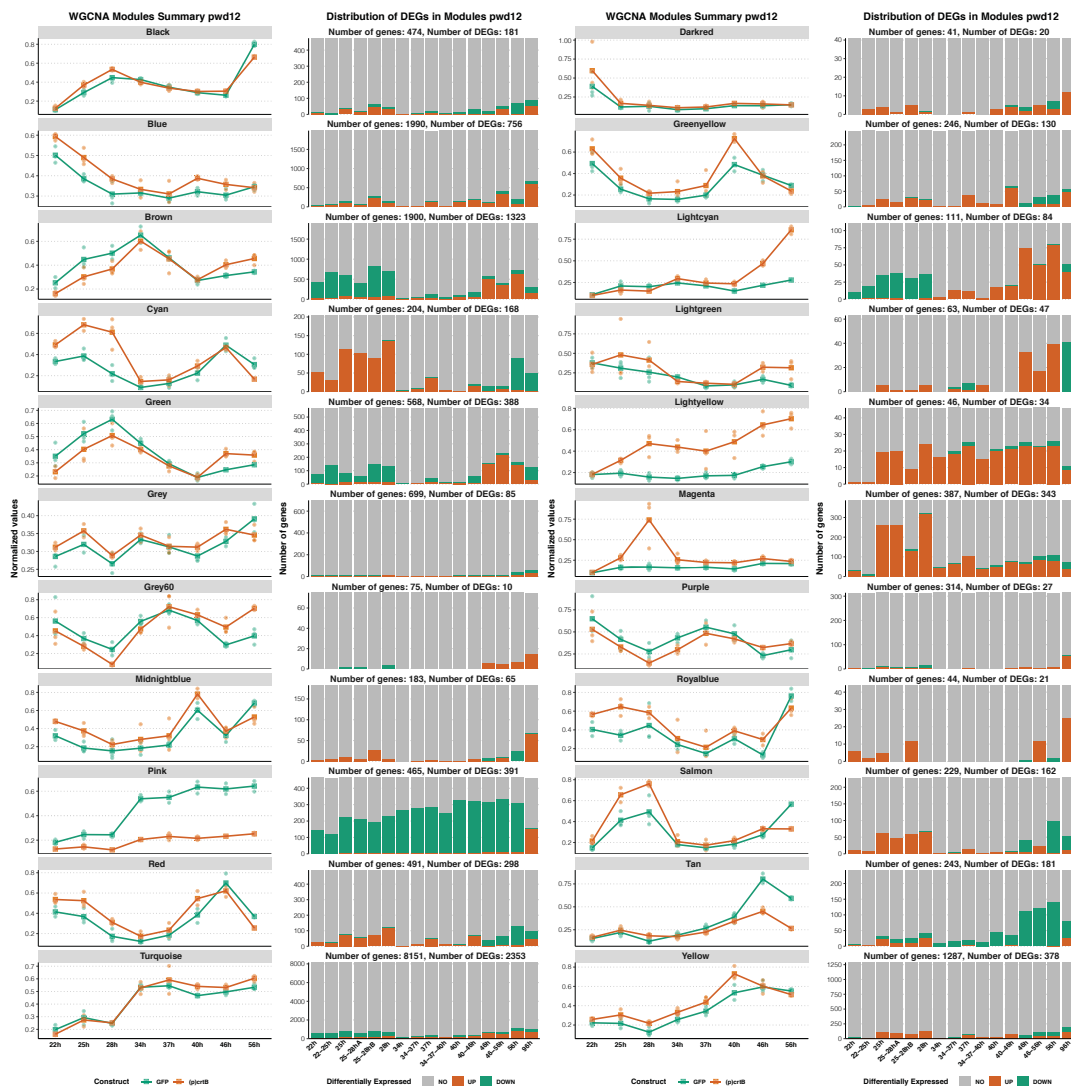
Figure 4.10: Modules from WGCNA analysis with all genes. Average expression of genes included in each sample from time course RNA-seq analysis is showed in green for GFP and orange for (p)crtB, for each module from WGCNA analysis. Bar plots shows the abundance of DEGs in each comparison of RNA-seq, classified in up-regulated (orange), down-regulated (green) or not differentially expressed (grey).

To study the biological processes that were representing each cluster, GO enrichment was performed with topGO package (Alexa and Rahnenfuhrer, 2021). The "cyan" module showed an enrichment in GO terms related to cell wall, illustrating that (p)crtB impacts biology of this structure at the beginning of the process (first three time-points) (Supplementary Figure 11). The "magenta" module (induction at 28 hpi) was enriched in many metabolic processes, with a predominance of the biosynthetic ones (Supplementary Figure 12). Other biosynthetic processes like MEP-pathway or terpenoids, together with biotic stress related processes, were representing the "lightyellow" module (Supplementary Figure 13). The induction of processes at the end of the time course in (p)crtB ("lightcyan" module) was mainly defined by response to biotic defense processes (Supplementary Figure 14), while a greater diversity of processes was down-regulated at this stage (Supplementary Figure 15). Protein folding related processes were the most prevalent GO terms enriched in the "pink" module, as well as some abiotic stress and energy production related processes, indicating that these were down-regulated events during chromoplast differentiation (Supplementary Figure 16).

4.1.6 Weight Gene Co-Expression Network Analysis: DESeq2

All previous information was collected focusing on modules that showed a visible association with the 'construct' variable (i.e., those that showed clear differences in (p)crtB vs GFP). However, most of them were a consequence of the 'time' variable, showing the dynamic of gene expression throughout the time-points, with no differences between constructs. Even in the modules of interest, not all genes showed significant differences. Therefore, we decided to do a new WGCNA analysis focusing only on genes that were differentially expressed in (p)crtB vs GFP at least in one time-point. Instead of using all expressed genes as input, a prior analysis was done with DESeq2, using all samples in a special design in order to obtain DEGs according to the 'construct' variable towards the 'time'. The resulting genes were significantly and differentially expressed in (p)crtB samples compared to GFP at some time-point. Then, this list was used as input for WGCNA to cluster these DEGs in new modules, according to their expression profiles. As a result, twenty-one new modules were generated, identified with colors again and correlated to phenotype data (Supplementary Figure 17). They were also plotted to show the average expression of all genes in each sample, to see the profiles of the expression with respect to the 'time' and 'construct' variables (Figure 4.11). There were some correspondence between the previous modules and the new ones, like cyan vs. greenyellow, magenta vs. pink, lightyellow vs. midnightblue or pink-purple (previous vs. new, respectively). However, in this second analysis all modules were meaningful as they only contained DEGs (Figure 4.11). The top up- and down-regulated hubgenes (positive and negative, respectively) were listed for the different modules, showing some enrichments like protease inhibitors in midnightblue module, or heat-shock proteins in purple (Tables of HubGenes approaches). To organize these new modules into a

Figure 4.11: Modules from WGCNA analysis with DEGs. Average expression of genes included in each sample from time course RNA-seq analysis is showed in green for GFP and orange for (p)crtB, for each module from WGCNA analysis. Bar plots shows the abundance of DEGs in each comparison of RNA-seq, classified in up-regulated (orange), down-regulated (green) or not differentially expressed (grey).

logic story-telling mode, the most meaningful ones were ordered according to their profile of expression FC in (p)crtB vs. GFP in a heatmap, and gathered in eight functional groups (Figure 4.12):

- **Group 1**: Contained **magenta** and **purple** modules, that were characterized by the down-regulation of their genes all over the time course.

- **Group 2**: Corresponded to the **yellow** module, which showed a slight up-regulation at the beginning, followed by a neutral expression in the middle and a final down-regulation at the end of the time course.

- **Group 3**: Contained **black**, **green** and **red** modules, that showed a profile similar to the previous group, but with a more marked initial up-regulation, and

a down-regulation restricted to the last time-point.

- **Group 4**: Restricted to the **greenyellow** module, in which the up-regulation was more intense at the first time-points, specially at the 28 hpi, and the down-regulation was also restricted to the last time-point.

- **Group 5**: **Pink** module, that was similar to Group 4 but it showed a more remarkable up-regulation at the 28 hpi time point.

- **Group 6**: A small group that only contained the **midnightblue** module, consisting of up-regulated genes from the 28 hpi forward.

- **Group 7**: Another very small group, only with the **lightyellow** module. These genes showed a down-regulation at the beginning of the time course, followed by an up-regulation at the end.

- **Group 8**: The largest group, composed by **brown**, **turquoise** and **blue** modules. The expression of the genes was characterized by a slight down-regulation at the beginning, a neutral expression in the middle and an up-regulation by the end of the process (opposite to groups 2 and 3).

Enriched GO terms representing each group were obtained, and a significant diversity of them was found (Supplementary Figure 18, Supplementary Figure 19, Supplementary Figure 20, Supplementary Figure 21, Supplementary Figure 22, Supplementary Figure 23, Supplementary Figure 24, Supplementary Figure 25). Venn diagrams of the different GOs were created in order to check if there were many overlapped terms (Tables GitHub). Most of the terms were exclusive representing only one group, but some metabolic and stress related processes were overlapped in two or three different groups (Supplementary Figure 26).

4.2 The role of specific functional families of genes

Previous approaches let us to obtain a general overview of the differentiation of chromoplasts from pre-existing chloroplast in leaves of *N. benthamiana*, comparing it to other systems. Considering these results, different functional groups were identified to be particularly involved in the process. We annotated the corresponding genes in the new *N. benthamiana* genome, in order to study their pattern of expression in the samples from 22 to 56 hpi. The expression data were plotted in a heatmap, using all the information we have previously generated: all possible groups of samples for the FC expression (attending to the time-point that they were collected or how they were grouping in the PCA analysis, mixing samples from different time-points), including asterisks for those that were significantly expressed in that group, adding two columns to indicate the module where each gene was clustered in both WGCNA approaches (first approach using all expressed genes and the second when
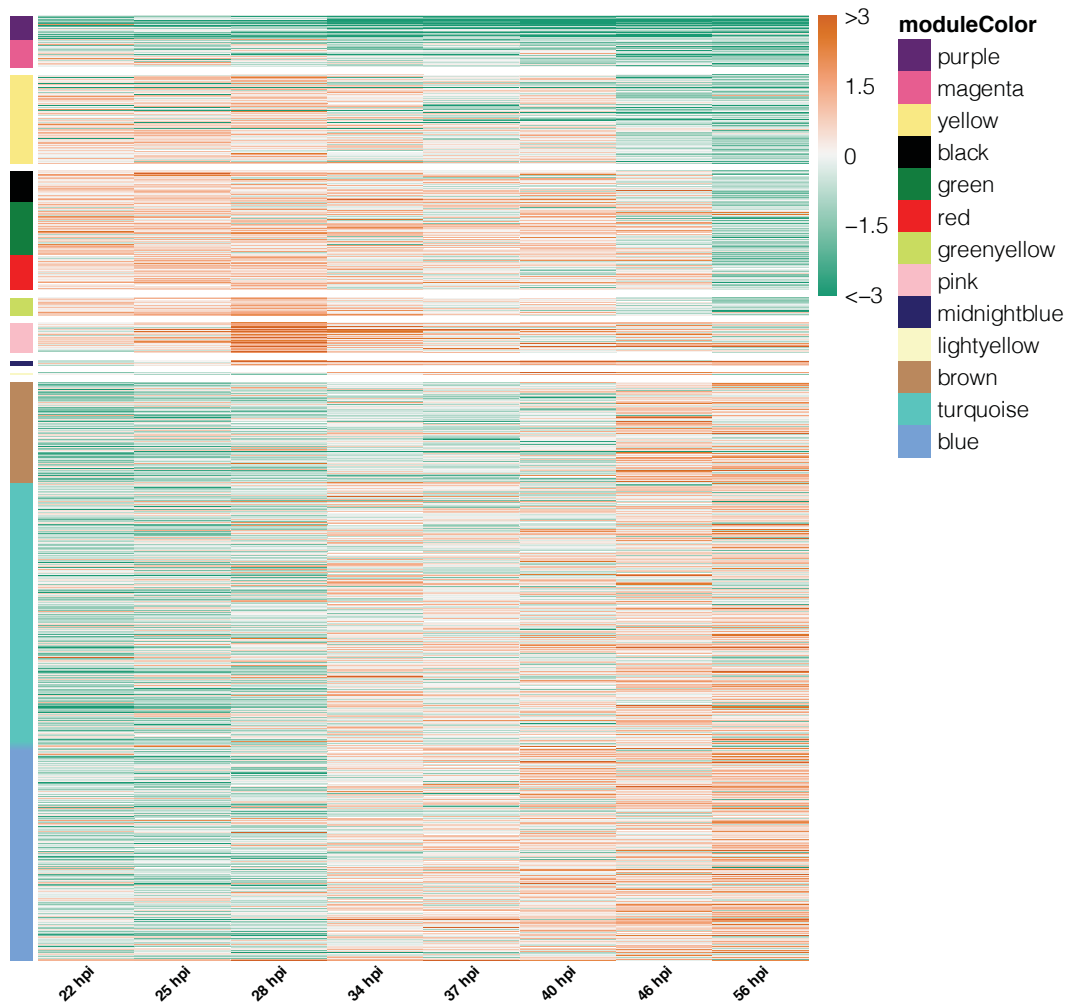
Figure 4.12: Heatmap of modules from second WGCNA analysis. Row normalized expression of genes included in the different modules from WGCNA analysis.

only DESeq2 significantly expressed DEGs were used) and the subcellular location of the encoding protein, predicted with TargetP 2.0 (Armenteros et al., 2019).

4.2.1 Carotenoids

Despite the (p)crtB-dependent phytoene overproduction and accumulation and the 2-fold increase of total carotenoids by the end of the time course (Figure 4.1), the expression of most of the endogenous carotenoid biosynthetic genes was not affected in any point of the time-course experiment (Figure 4.13). Upstream lycopene there was no affection at all, while some of the downstream genes showed changes on their expression by the end of the time course, most of them downregulated. However, they were not clustered in the same module, and most of them were not included in the DESeq2 modules, showing that they did not show a common expression profile and their expression was not significantly different between (p)crtB and

GFP samples throughout time. This is a clear example of the limitations of gene expression changes as readouts of actual biological changes (e.g., an active carotenoid synthesis occurs without major changes in the expression of biosynthetic genes).

Figure 4.13: Carotenoid biosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of carotenoid biosynthesis genes. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction).

4.2.2 MEP pathway

Genes encoding MEP pathway enzymes involved in the production of carotenoid precursors in plastids, showed a significant pattern of induction at the 25 and 28 hpi (Figure 4.14A). In particular, genes encoding isoforms of rate-determining enzymes such as DXS, HDS and HDR were included in the magenta module of the first WGCNA approach and the pink module of the second, both showing a very marked induction at 28 hpi (Figure 4.14A). Genes encoding DXR and another HDS isoforms, despite not been included in these modules, also showed a significant up-regulation at this period. Different genes encoding GGPPS isoforms are predicted to

be present in the *N. benthamiana* genome. From them, a few are encoded by genes that also showed this up-regulation profile and were included in the magenta and pink modules (Figure 4.14C).

Figure 4.14: Isoprenoid-related gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes encoding enzymes from MEP pathway (A), MVA pathway (B), GGPPS, FPPS and IDI (C), and tocopherol biosynthesis (D). Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

4.2.3 MVA pathway

Most genes encoding the enzymes that produce IPP and DMAPP in the cytosol through the MVA pathway showed the same pattern of expression, being included in the brown module from the first WGCNA approach and the brown or turquoise modules from the DESeq2 approach (Figure 4.14B). They showed a significant downregulation from at the beginning of the time course (events (i) and (ii), corresponding to the first three time-points), changing to an up-regulation at the very end, with no significant differences of expression compared to the control at the middle lag period. Genes encoding for isomerases converting IPP to DMAPP (IPPI)

and FPP synthases (FPPS) showed an expression profile similar to that described for MVA pathway genes (Figure 4.14C).

4.2.4 Tocopherols, plastoquinones and phylloquinones

There are different pathways that use the plastidial pool of GGPP derived from the MEP pathway, including those producing tocopherols, plastoquinones and phylloquinones (Figure 1.1). *VTE* genes from the tocopherol pathway showed no relevant expression changes, each one clustering in a different module (Figure 4.14D). Genes encoding enzymes for the biosynthesis of plastoquinone and phylloquinone are lacked a clear pattern of differential expression (Figure 4.15). The only exception was *DHNAT*, which encodes a peroxisomal enzyme upstream the integration of GGPP derived products into the phylloquinone pathway. This gene was overexpressed at 28 hpi and it was included in magenta and pink modules.
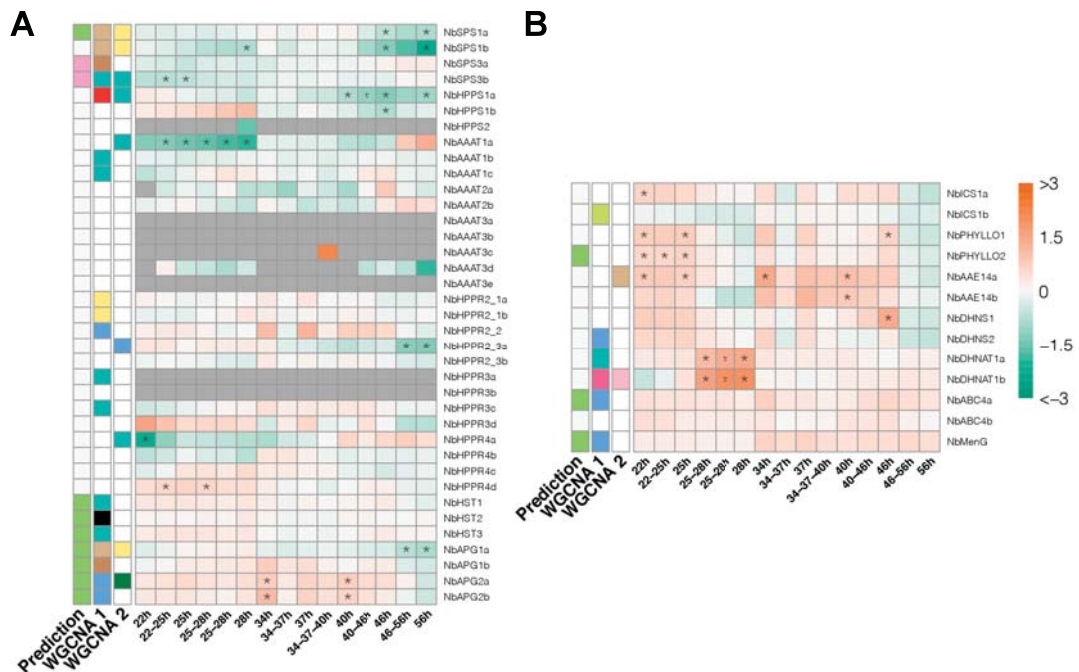
Figure 4.15: Plastoquinone and phylloquinone biosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes encoding enzymes from plastoquinone (A) and phylloquinone (B) biosynthetic pathways. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

4.2.5 Fibrillins, stromules and membrane remodelling proteins

Fibrillins are structural proteins that have been associated to chromoplast differentiation and carotenoid storage (Figure 1.2). Their protein level was found to increase in (p)crtB samples (Figure 3.4). Consistently, a few genes encoding FBN2,

FBN6 and FBN9 isoforms showed a weak but significant induction during the late stages of the time course (Figure 4.16A). Stromules are structures that expand from plastids and showed a different morphology in (p)crtB samples compared to the GFP controls (Figure 3.4C). However, most of the genes associated to stromules showed a very low expression in all samples, they were not included in any module from the first WGCNA approach (that only had a filter regarding the low expression of genes), and most of their expression was not differentially significant in any time-point (Figure 4.16B). Exceptions include *CHUP1* homologs that were slightly but significantly induced during event (ii) as well as *kinesin 4* homologs belonging to Group 8 of genes showing a slight down-regulation att the beginning, a neutral expression in the middle and an up-regulation by the end of the process (Figure 4.16B). Endosomal sorting complexes required for transport (ESCRT) comprise cytosolic proteins that constitute a machinery involved in the membrane remodeling, related to different cellular processes. Many of the genes encoding ESCRT proteins showed a common pattern of expression, being included in the turquoise module of the first WGCNA approach and showing an initial down-regulation in the first time-points of the time course, changing to the up-regulation at the very end (Supplementary Figure 27).
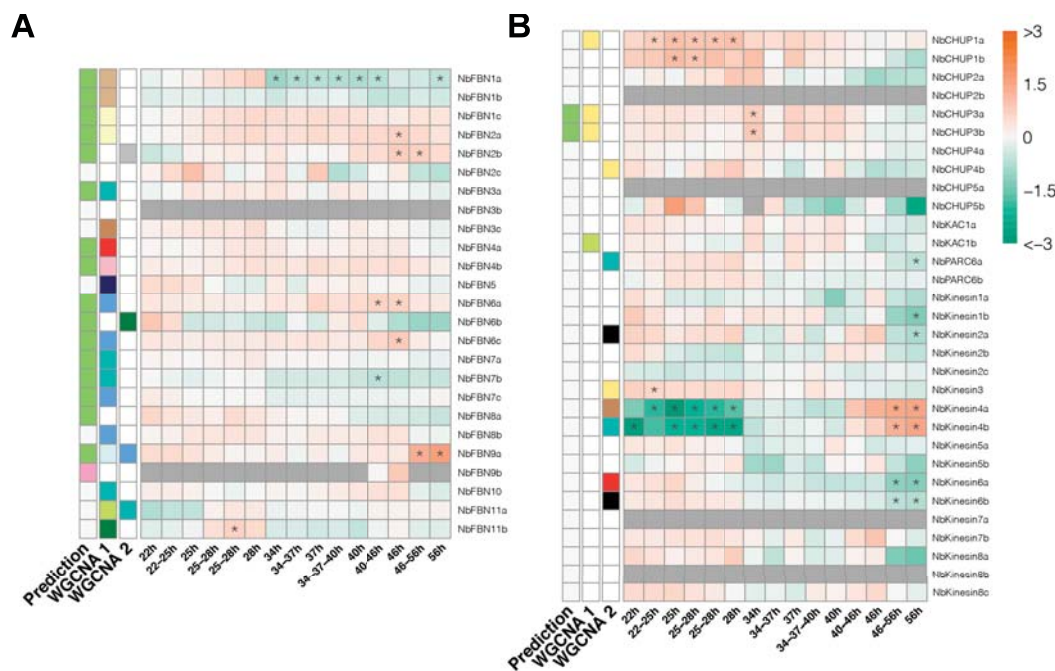
Figure 4.16: Fibrillins biosynthesis and stromules-related gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes encoding enzymes from fibrillins biosynthesis (A) and related to stromules (B). Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

4.2.6   Cell wall

Cell wall related genes were affected at different levels in the exploring approaches (e.g., comparing the system in *N. benthamiana* to others, included in overlapped groups of genes from different time-points or in specific modules from the WGCNA analysis). Exploring the genes that encode biosynthetic enzymes of the cell wall, most were observed to have a very low expression, not been included in any module (Supplementary Figure 28A). Among the expressed genes, a significant overexpression was observed at the 25-28 hpi in many of them, although they were not included in the magenta-pink (previous-later) modules as it would have been expected.

4.2.7   Circadian clock, protein import, PIFs, light receptors and autophagy

Different results suggested that the artificial chromoplast differentiation in leaves of *N. benthamiana* might impact the circadian clock, hence creating a mismatch between GFP and (p)crtB samples. Circadian clock associated genes were identified, but most of their expression was too low and they were not included in any module of WGCNA analysis (Supplementary Figure 28B). Those that were more highly expressed did not show a significant pattern of expression nor were included in any of the meaningful modules from any WGCNA approach. Protein import from cytosol to plastids was another interesting functional group that we wanted to study, since chromoplast differentiation relies on it (Figure 1.2). Genes encoding translocons on the inner and outer chloroplast membranes (TIC and TOC, respectively) were studied, but any of them showed a significant expression at any time-point or showed a co-expressed pattern, since they were spread in many different modules (Figure 4.17A). Interestingly, a *N. benthamiana* homolog of *SP1* (Figure 1.2) was found to be weakly but significantly up-regulated at 28 hpi (Figure 4.17A). Phytochromes, cryptochromes and phytochrome interacting factors (PIFs) are known to regulate chloroplast development and hence were thought to potentially modulate the chromoplast differentiation process. However, most of the genes encoding them showed very low levels of expression or lacked any particular profile associated to (p)crtB samples (Figure 4.17B-C). Another biological group that was retrieved from previous GO terms analyses was "autophagy". Autophagy-regalated (*ATG*) genes were identified and studied in *N. benthamiana*, and most of them showed a co-expression associated to the turquoise module of the first WGCNA approach, but they were not associated to (p)crtB samples since they were not included in DESeq2 modules or showed any differentially significant expression at any time-point (Supplementary Figure 28C).
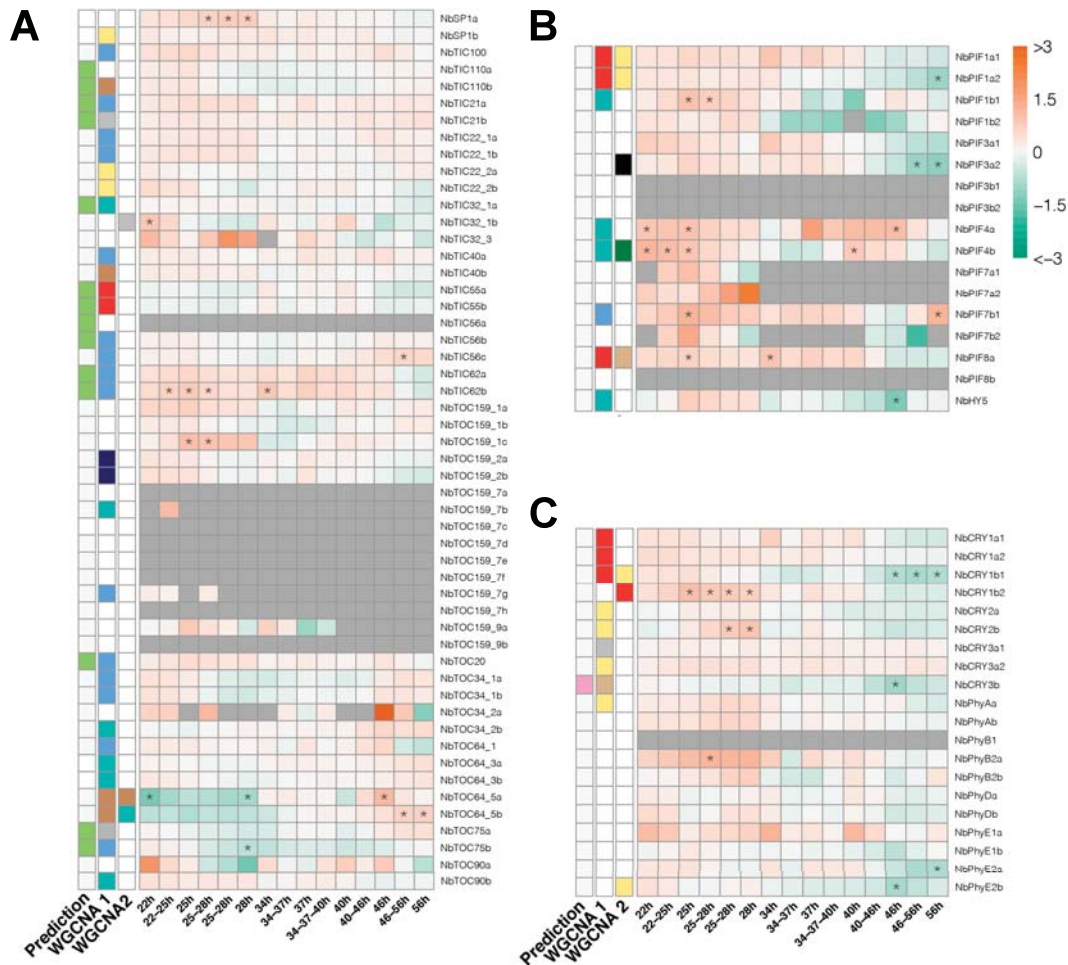
Figure 4.17: Protein import, PIFs and light receptors gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes encoding proteins from the inner and outer translocons (TIC and TOC) (A), phytochrome interacting factors (PIFs) (B) and light receptors (phytochromes and cryptochromes) (C). Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

4.2.8 Protein folding

Previous approaches showed an enrichment in protein folding related terms representing downregulating genes throughout the process. Chaperones are proteins involved in protein folding and unfolding, so we were interested in identifying the different families of chaperones in *N. benthamiana*. Some genes from the family of BAG chaperones showed up-regulations at the beginning of the time-course, while others were always down-regulated (Supplementary Figure 32A). Most of the genes encoding chaperonin-60, DNAJ proteins, prefoldins and rotamases did not show an affected expression at any time-point, but those that did were downregulated and associated to pink-purple (previous-later) modules, that were those

consisting in down-regulated genes (Supplementary Figure 32, Supplementary Figure 33, Supplementary Figure 34A). DNAJ includes Orange and Orange-like (OR and ORL, respectively) proteins, none of the *N. benthamiana* homolog genes showed differences between (p)crtB samples and the GFP control at any time-point (Supplementary Figure 33). The expression of cyclophilin family genes was mostly not affected (Supplementary Figure 34B). On the other hand, many of the genes encoding small chaperonins (Cpn10), Hsp20, Hsp90, Hsp90 co-chaperones and Hsp100 showed a significant down-regulation in most of the time-points and were included in the pink-purple modules (Figure 4.18A-D, Supplementary Figure 35). Most of the genes encoding Hsp70 family showed the same down-regulation profile as the previous families, but there was a group of genes encoding for proteins known as luminal binding proteins (BiPs) that were included in the blue module of the DESeq2 WGCNA analysis, showing a significant up-regulation in the second half of the time-course (Supplementary Figure 36A). That profile was also observed in genes encoding protein disulfide-isomerase (PDI) family, containing genes clustered in the same module and showing the same significant up-regulation (Supplementary Figure 36B).

4.2.9 Chlorophylls

Although chlorophyll levels did not change during artificial chromoplastogenesis in *N. benthamiana* leaves unlike in other chromoplast differentiation processes, we were interested on the genes encoding enzymes from the biosynthesis and degradation of chlorophylls. Genes involved in biosynthesis of chlorophyll a showed a common pattern of expression. Most of them were included in the blue module of the first approach of WGCNA but their expression was not significantly different between (p)crtB and control samples according to DESeq2 analysis (they were not included in any module in this approach) (Supplementary Figure 37A). Genes encoding enzymes involved in the chlorophyll cycle between chlorophyll a and chlorophyll b did not show any significant difference (Supplementary Figure 37B). Genes encoding enzymes of the degradation of chlorophylls were generally not affected either. However, genes encoding isoforms of pheophytinase (PPH), that catalyze the conversion of pheophytin to pheophorbide a and showed an overexpression at 25-28 hpi, included in the magenta-pink (previous-later) modules of WGCNA (Supplementary Figure 37C).

4.2.10 Photosynthesis and antenna proteins

Despite the decrease on photosynthetic activity during the artificial chromoplastogenesis process, that ended up with no activity at all by the end of it, some terms related to photosynthesis were associated to induced genes in previous analysis. Genes encoding proteins from the antenna complex were identified, and they showed a co-expression pattern, included in turquoise and blue modules in the first
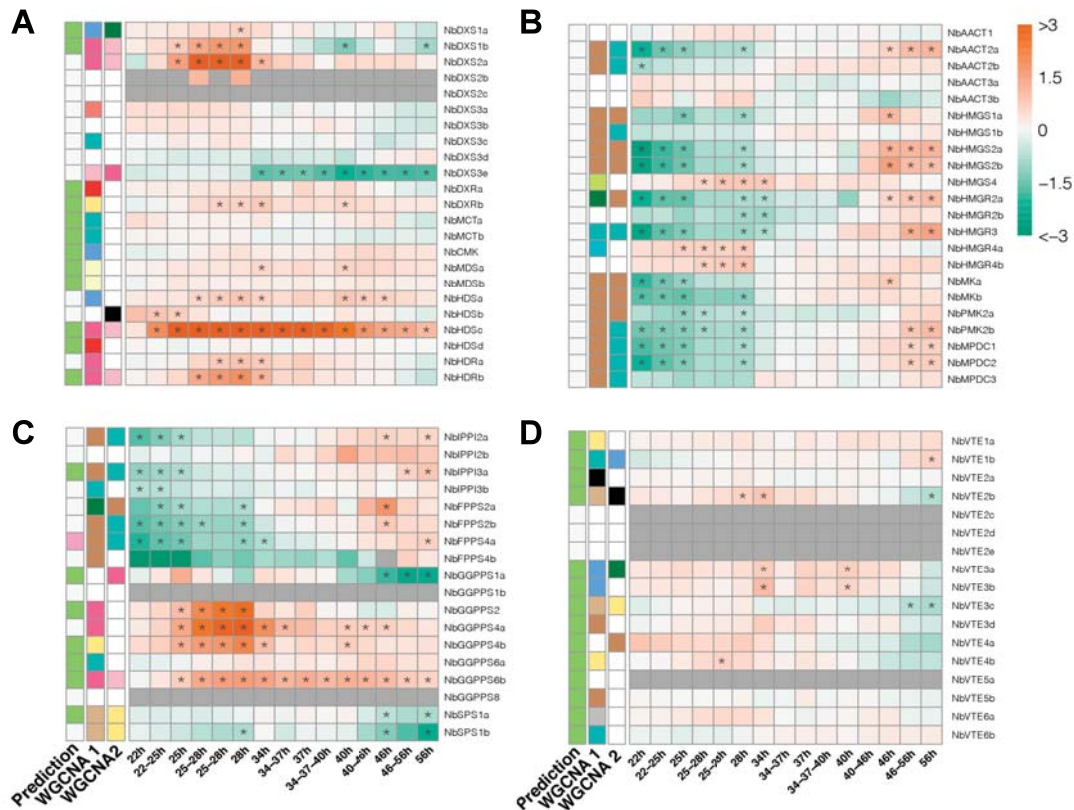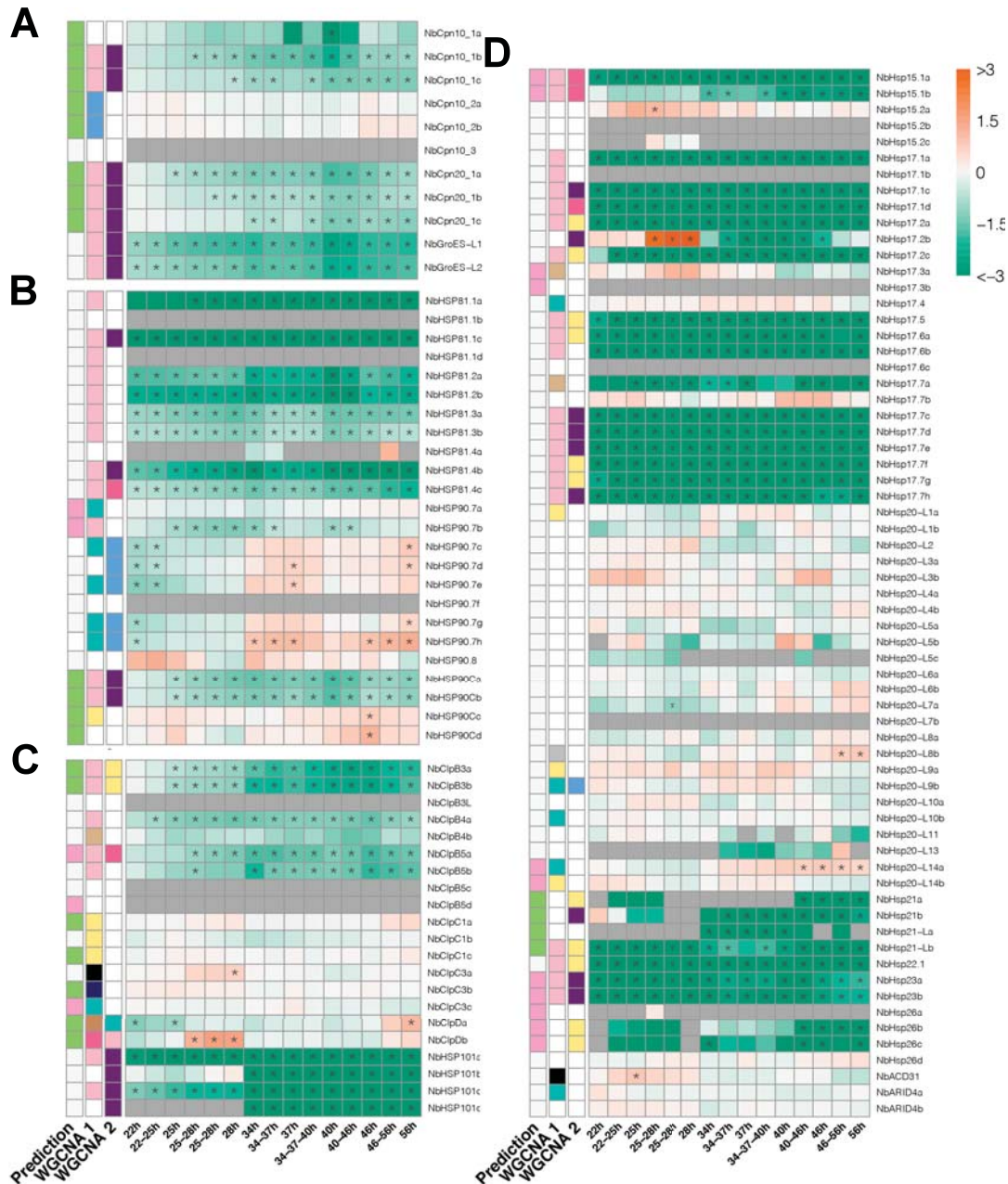
Figure 4.18: Cpn10, Hsp90, Hsp100 and Hsp20 gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to Cpn10 (A), and Hsp90 (B), Hsp100 (C) and Hsp20 (D) chaperone families. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

WGCNA approach (Supplementary Figure 38). Many of them showed a significant up-regulation at the 40-46 hpi, that disappeared at the 56 hpi (Supplementary Figure 38). Photosynthesis-related candidate genes were also identified and showed the same profile, most of them included in the blue module of the first WGCNA approach (Supplementary Figure 39). However, a portion of these genes were also

significantly up-regulated still at the 56 hpi.

4.2.11 Calvin-Benson cycle and primary metabolism

The huge network of genes involved in primary metabolism and the Calvin-Benson cycle was estimated by looking for the candidates homologous in *N.benthamiana* in order to see specific profiles of expression associated to the previous information analyzed. Initially, not all genes were identified for their specific function, due to the massive amount. Once the expression values in the time course were assigned to each candidate, they were clustered according to that (Supplementary Figure 40, Supplementary Figure 41). In heatmaps were observed clusters of genes with a plastid signal peptide showing the same pattern of expression as the photosynthetic and antenna complex genes, assigned to the blue and turquoise modules of the first WGCNA approach and to any or green modules of the DESeq2 approach, exhibiting a general significant up-regulation throughout the time course, except for the last time-point. Most of these genes encode members of the Calvin cycle (e.g., subunits of Rubisco, transketolases or aldolases) or other genes related to trioses and pentoses phosphate like glyceraldehyde dehydrogenase or chloroplastidial malic enzyme. Other clusters showed a profile similar to the MVA pathway, with genes included in modules from the Group 8 of the DESeq2 approach (blue, brown and turquoise), exhibiting an early down-regulation from the 22 to the 28 hpi, that was diluted after that and many of them being up-regulated by the last time-point at 56 hpi (Supplementary Figure 40, Supplementary Figure 41). Most of the enzymes encoded by these genes were predicted to be cytosolic or mitochondrial and were identified to play a core role in metabolism. In general, these data show a scenario marked by a general transcriptional shut down of primary metabolism at the very beginning of the time course, including glycolysis and tricarboxylic acid (TCA) cycle (Figure 5.1). Finally, the other cluster interesting for our purpose was one that showed an up-regulation at 25-28 hpi time-points and some of their genes were included in the magenta module of the first WGCNA approach, like MEP pathway or JA related genes (Supplementary Figure 41). Some of these genes encode enzymes involved in the biosynthesis of serine and tetrahydropholate (THF) related reactions, that participate in the metabolism of amino acids and nucleotides. Also reactions related to transport of metabolites through compartments like 2-oxoglutarate to glutamate, pyruvate to malate or the biosynthesis of citrate were represented (Figure 5.1).

4.2.12 Hormones

Biosynthetic genes from different families of hormones were identified in order to study their possible association to chromoplast differentiation. Most of the genes involved in the synthesis of abscisic acid (ABA), auxins, brassinosteroids and

cytokinins showed low levels of expression, but some of those that were more expressed showed an up-regulation at the 25-28 hpi, despite only few were included in the magenta-pink modules (previous-later analysis of the WGCNA) (Supplementary Figure 29A-B, Supplementary Figure 30A-B). The genes related to the synthesis of strigolactones, did not appear to be affected, since most of them were very lowly expressed and the others showed a disperse pattern of expression (Supplementary Figure 30C). Genes encoding the first to enzymes of the GGPP-dependent gibberellin pathway, CPS and KS, were down-regulated from 25 hpi (Supplementary Figure 31A), whereas those encoding enzymes that produce active gibberellins from inactive intermediates (GA20ox and GA3ox) were included in the Group 8 (subsection 4.1.6). Similarly, genes encoding some isoforms of the pathways producing salicylic acid and ethylene were included in the Group 8 (Supplementary Figure 31B-C).
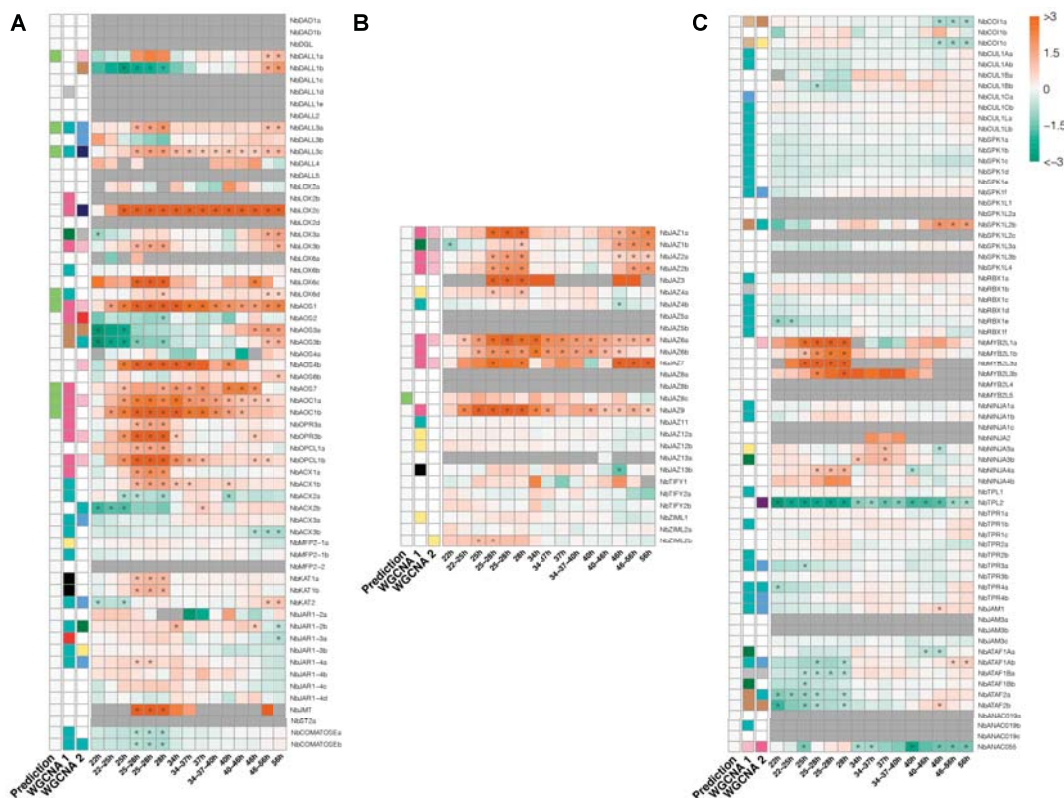
Figure 4.19: Jasmonic acid related gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to jasmonate biosynthesis (A), JAZs (B) and jasmonate response (C). Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Jasmonic acid (JA) was the hormone that showed the most consistently affected profile in the expression of its biosynthetic genes. Many of them were highly up-regulated and associated to magenta-pink modules (previous-later approaches of

WGCNA), with a peak of expression during event (ii) at 25-28 hpi (Figure 4.19A). Genes encoding JAR1 isoforms, that catalyze the synthesis of the only known JA derivative needed for JA signaling, JA-isoleucine (JA-Ile), were however not affected. The expression of JAZ proteins (also named Tify), transcription factors involved in the repression of the JA response, showed a very similar profile, with a high up-regulation and included in the same modules (Figure 4.19B). The SCF complex, comprised by COI1, CUL1, SPK and RBX, involved in the degradation of JAZ proteins, did not show any change in the expression of their genes (Figure 4.19C). By contrast, some MYB2 related genes showed a particular overexpression at 25-28 hpi (Figure 4.19C). Other genes involved in the response to JA did not show very remarkable changes on their expression, with the exception of some genes encoding ATAF, that included the NAC domain and showed a down-regulation at the first time-points (Figure 4.19C).

4.3   Hormone quantification

Due to the affections observed in genes encoding biosynthetic enzymes from some hormone families (e.g., gibberellins and jasmonates), a new time course experiment was performed following the same procedure as before, i.e., infiltrating *GFP* and *(p)crtB* constructs in the same leaf of *N. benthamiana* and collecting samples at 25, 35, 45 and 96 hpi. Collected samples were then used to extract and quantify representative hormones, including auxin (indole-3-acetic acid [IAA]), gibberellins (gibberellin A1 [GA1] and GA4), cytokinins (isopentenyl-adenine [iP], trans-zeatin [tZ] and dihydrozeatin [DHZ]), jasmonic acid (JA), abscisic acid (ABA) and salicylic acid (SA). Hormones directly derived from MEP pathway products, such as DMAPP-derived citokinins (iP, tZ) and GGPP-derived gibberellins (GA1 and GA4) already showed lower levels in (p)crtB samples compared to GFP in the first time point (Figure 4.20). However, iP stayed at low levels in (p)crtB samples in the following time points, while the others recovered. Actually, gibberellins increased at later time-points (45 hpi) and GA4 showed over 2-fold increase levels in (p)crtB at the end of the time course (Figure 4.20). IAA, SA, ABA and JA showed an increase profile throughout the time course, but it was only significant in the case of JA, that showed a ca. 4-fold increase in (p)crtB at 96 hpi (Figure 4.20).

Figure 4.20: Hormone quantification. (A) Levels of nine different hormones from six different families in GFP (green) and (p)crtB (orange) samples during a time course. (B) Relative levels of same values in (p)crtB/GFP ratio per biological replicate. Auxins: IAA (indole-3-acetic acid); gibberellins: GA1 (gibberellin A1) and GA4; cytokinins: iP (isopentenyladenine), tZ (trans-zeatin) and DHZ (dihydrozeatin); JA (jasmonic acid); ABA (abscisic acid); SA (salicylic acid).

Discussion

5.1   crtB triggers chromoplast differentiation in leaves

Leaves are ubiquitous organs in the plant kingdom. Despite their diversity in shapes, sizes and number (from the pointed needles in pine trees to the huge buoyant leaves of *Victoria amazonica*), most of them share a common function: photosynthesis (Nikolov et al., 2019). This feature is easily recognized from their green color, caused by the abundance of chlorophylls in the thylakoids, that are in charge of light harvesting. In leaves, plastid transitions are only observed when chloroplasts are made (from proplastids during leaf development or etioplasts during de-etiolation) or destroyed (during senescence, when chloroplasts transition to gerontoplasts). Hardly ever chloroplasts modify their identity in healthy mature leaves (Dhami and Cazzonelli, 2020; Sun et al., 2017; Torres-Montilla and Rodriguez-Concepcion, 2021). The differentiation of chromoplasts in leaves of *N. benthamiana* from pre-existing chloroplasts described in this thesis work hence represents a novel and innovative system.

In general, carotenoid content is not altered in chloroplasts, probably to keep a proper balance with chlorophylls for effective light harvesting and protection against oxidative stress (Ruiz-Sola and Rodríguez-Concepción, 2012). The introduction of crtB, an external phytoene synthase that produced phytoene in massive amounts, was found to force the accumulation of this metabolite and feed the endogenous carotenoid pathway. Phytoene is reported to be hardly present under normal conditions in mature leaves. As the first committed intermediate of the pathway, it acts as a bottleneck regulating the flux, and it is consumed as soon as it is produced (Ruiz-Sola and Rodríguez-Concepción, 2012; Rodriguez-Concepcion et al., 2018). *PSY* overexpression was reported to show no phenotype in *Arabidopsis* leaves (Zhou et al., 2015), while it caused dwarfism in tomato plants caused by a reduction in gibberellins due to the redirection of metabolites (Fray et al., 1995). The transient expression of *Arabidopsis PSY* and tomato *PSY2* in leaves of *N. benthamiana* did not show any phenotype either, and only caused a minor phytoene accumulation (Llorente et al., 2020) (Figure 3.2). Only in systems that are not specialized in photosynthetic activity (e.g., fruits, roots or callus), *PSY* overexpression was reported to trigger an increase in carotenoid content that could eventually lead to chromoplast differentiation (Fraser et al., 2007; Schaub et al., 2018; Maass et al., 2009). These results suggest that plant PSY enzymes might have a very strict regulatory system

that controls their activity, both transcriptionally and post-transcriptionally, adjusting the production of phytoene in a very strict manner. As a candidate, it has been already described an E3 ubiquitin ligase in tomato named Plastid Protein Sensing RING E3 ligase 1 (PPSR1) that regulates carotenoid biosynthesis by mediating the degradation of PSY1 by ubiquitination (Wang et al., 2020). This putative regulatory system would act only in mature chloroplasts of photosynthetic tissues, attenuating its control in conditions where plastids need to transit from one type to another, like de-etiolation, fruit ripening or petal formation. Additionally, the Orange protein (OR) is an important post-transcriptional regulator of PSY (Li et al., 2001; Lu et al., 2006; Sun et al., 2020). OR is a chaperone that has been reported to prevent PSY misfolding and degradation, hence ensuring its proper activity and stability (Park et al., 2016; Welsch et al., 2018; Zhou et al., 2015). While OR is the only protein proposed to directly promote chromoplast differentiation, it is not essential for chromoplast differentiation in *N. benthamiana* leaves, since crtB-mediated chromoplastogenesis was also reported in *Arabidopsis ator atorl* defective double mutants completely devoid of OR activity (Llorente et al., 2020). Furthermore, expression of genes for *OR* homologs was not found to change during chloroplast-to-chromoplast transition in *N. benthamiana* (Supplementary Figure 32). These results suggest that the role of OR might be mostly related to PSY protection against the regulatory system that impedes its activity, but it is not require for chromoplastogenesis *per se* when another enzyme produces enough phytoene (e.g., crtB).

Sequence differences between plant PSY and bacterial crtB (Supplementary Figure 42) might allow the bacterial enzyme to escape from the plant regulatory system even in photosynthetic leaves (Kato et al., 2016; Misawa et al., 1990). There have been previous approaches using crtB as a strategy for carotenoid biofortification in different plant tissues, but the dramatic effect that we observed upon its constitutive expression in leaves was never reported. Bacterial genes (e.g., *crtE*, *crtB* or *crtI*) have been widely used in multiple organisms (including plants) for production of carotenoids (Aluru et al., 2008; Diretto et al., 2007; Ravanello et al., 2003). It was speculated that these bacterial enzymes might physically interact to form a metabolon when they act together in plant cells, channeling the universal isoprenoid precursors (IPP and DMAPP) to the synthesis of carotenoids (Nogueira et al., 2013; Andersen et al., 2021). When *crtB* was also expressed with a fruit specific promoter and fused to the plastid transit peptide from *SlPSY1* in tomato, increased levels of carotenoids were observed without changes in the expression of the endogenous carotenoid biosynthetic genes or other related isoprenoids (Fraser et al., 2002). This strategy was also used in combination with *crtE* and *crtI* expression, observing perturbations in the expression of the carotenoid pathway genes, phytoene accumulation in plastoglobules, and changes in membrane components (Nogueira et al., 2013). Under a fruit specific promoter, crtB only induces remarkable changes in non-photosynthetic tissues, in like manner than *PSY* overexpression (Fraser et al., 2007; Lätari et al., 2015; Maass et al., 2009). Using the *35S* constitutive promoter for *crtB* expression

and pea rbcS transit peptide to target the enzyme to plastids, a notable increase in carotenoids levels, mostly sequestered in crystal forms, and lycopene accumulation (absent in wild-type lines) were reported in citrus calli (Cao et al., 2012).

The first time that we observed crtB to trigger unusual changes in plant leaves was in a work in which *crtB* was expressed with the *35S* promoter in plant leaf tissue with no other transgenes (Majer et al., 2017). The purpose of that work was the production of lycopene in the plant cell cytosol using the bacterial enzymes crtE, crtB and crtI in their original form, i.e., without added plastid-targeting peptides. The approach of expressing *crtB* alone was thought just as a control for the agroinfiltration, but a strong yellow phenotype in crtB-producing leaf areas was surprisingly found. The yellow phenotype is a consequence of the increase in the levels of carotenoids, in particular lutein (yellow) and β-carotene (orange). Chlorophylls were reported to decrease in viral infections (Majer et al., 2017), but transient expression of *crtB* by agroinfiltration did not show this effect (Llorente et al., 2020), showing that chlorophyll levels are not directly impacted by crtB production (Figure 3.1). It was previously described that chlorophylls degradation is not required for chromoplast differentiation during fruit ripening in tomato (Cheung et al., 1993).

Different hypotheses were thought to explain how crtB induces the yellow phenotype. (1) The first one was the possibility that phytoene was produced in the cytosol, triggering some effect from there. This hypothesis was soon rejected since crtB was only found to induce the yellow phenotype when localized to the plastids (Llorente et al., 2020) (Figure 3.4). Although the exclusively (c)crtB version was found to be active in the cytosol, the levels of accumulated phytoene when co-expressed with crtE were not very high (Figure 3.3). These results could be a consequence of the lower abundance of IPP and DMAPP derived from the MVA pathway in the cytosol compared to the plastidial pool derived from the MEP pathway (Ruiz-Sola and Rodríguez-Concepción, 2012). In support to this conclusion, when (c)crtB works together with a truncated version of HMGR (the main regulatory enzyme of the MVA pathway) and crtE (a bacterial GGPP synthase), it is able to accumulate 20-fold higher levels of phytoene in the cytosol compared to crtB in plastids without major effects on endogenous downstream carotenoid contents or leaf color (Andersen et al., 2021). (2) The second possibility was that phytoene accumulation inside plastids could *per se* cause the observed effects. This possibility was also discarded since norflurazon (NF), an inhibitor of phytoene desaturation, has been frequently used in leaves causing phytoene accumulations but not downstream carotenoid accumulation or leaf yellowing (Andersen et al., 2021; Simkin et al., 2000; Simkin et al., 2003). (3) The third hypothesis was the crtB-derived phytoene could feed the endogenous carotenoid pathway. This theory was supported by the obervation that NF prevents the crtB phenotype to develop (Llorente et al., 2020). (4) The last hypothesis is related to the consumption of the plastidial pool of GGPP, derived from

the MEP pathway. Besides carotenoids, this pool is used for the biosynthesis of different metabolites like gibberellins, tocopherols, plastoquinone, phylloquinones or chlorophylls (Figure 1.1) (Ruiz-Sola and Rodríguez-Concepción, 2012).

The confirmation that chloroplasts differentiate into chromoplasts upon the expression of *crtB* in leaves was made attending to the results from different approaches, including electron microscopy (TEM), chlorophyll fluorescence (PAM), Western-blot and RNA-seq. The definition of a chromoplast is ambiguous, englobing any plastid with no photosynthetic activity, synthesizing and accumulating carotenoids and showing a disorganized morphology without functional thylakoids (Li et al., 2016; Sun et al., 2017; Torres-Montilla and Rodriguez-Concepcion, 2021). The plastid ultrastructure observed in crtB agroinfiltrated leaves did not correspond to the natural chromoplasts observed in leaves of *Buxus sempervirens*, since these last ones showed a massive proliferation of huge electrodense plastoglobules spread inside the plastids (Koiwa et al., 1986). The reversible differentiation of chloroplasts to chromoplasts in textitB. sempervirens was described as a response to light stress, accumulating these plastids in cell layers just below the epidermis (Hormaetxe et al., 2005). By contrast, the artificial chromoplast found in crtB-producing leaves never reverted to chloroplasts (Llorente et al., 2020), suggesting that the mechanism controlling this two different situations might not be the same. In any case, our results suggest a general mechanism for chromoplast differentiation that might be applied to all plant systems (Llorente et al., 2020). We propose that the process can be divided in two phases. In the first one (phase I), chloroplasts must become competent (i.e., preconditioned) by lowering their photosynthetic capacity, hence weakening their chloroplast identity. In a second stage (phase II), an increased production of carotenoids is necessary and sufficient to complete the differentiation process.

5.2 Heterogeneity of chloroplast-to-chromoplast systems

While the proposed two-phase mechanism to transform chloroplasts into chromoplasts is likely general among plants, our work also reveals that the molecular pathways involved in each step of the process can vary widely among plant species and organs. The RNA-seq analyses performed during this thesis illustrate a learning path in bioinformatics, starting from simple and assisted approaches focused on obtaining DEGs lists and the GO terms representing them, to the application of clustering methods and phylogenetic analysis to extract more resolutive information. In the last decades, transcriptomics have evolved from the initial Sanger methods to the next generation sequencing (NGS) that improved the price and speed of the analysis, arriving to the current third generation sequencing, that are capable to obtain a higher quality by obtaining much longer reads (Dijk et al., 2018). The two RNA-seq experiments analyzed in this thesis were made using Illumina sequencing (NGS) by two different services. Despite the differences in the sequence procedure, the results should be similar from the technical point of view, since they were relativized to the

control (GFP) samples, agroinfiltrated in the same leaf used for the (p)crtB treatment. A proof of concept performed with the data from two tomato fruit ripening RNA-seq experiments performed in different labs showed that the enriched GO terms were quite homogenous among them (Figure 4.8). By contrast, tomato fruit ripening systems did not share many similarities compared to pepper fruit ripening. Despite both species belong to the *Solanaceae* family and show differentiation from chloroplasts to chromoplasts during their ripening, tomato is climacteric while pepper is nonclimacteric (Paran and Van Der Knaap, 2007), although it has been also reported an induction of ethylene biosynthesis genes in pepper (Osorio et al., 2012). Both tomato and pepper ripening share the up-regulation of carotenoid biosynthesis and cell wall disassembly, so both species may conserve ethylene signaling components but differ in hormonal regulation (Osorio et al., 2012). But also many other differences have been reported between both systems, such as sugar metabolism, amino acid content, polyphenols, and redox metabolism (Calumpang et al., 2020; Osorio et al., 2012; Rödiger et al., 2020). Since two apparently similar processes including the ripening-dependent chloroplast to chromoplast differentiation disguise many differences, it should not be surprising the relatively small amount of shared GO terms between natural (tomato, pepper) and artificial (*N. benthamiana*) chromoplast differentiation systems (Figure 4.8). Interestingly, the overexpression of OR^{His} in tomato, that accelerates the differentiation of chloroplasts into chromoplasts, showed more similarities with *N. benthamiana* than with tomato fruit ripening. The fact that the process of senescence in leaves of *Arabidopsis* was the most different system also emphasizes the importance of those GO terms that were shared by most of the sets in the chromoplastogenic systems. In summary, chloroplast-to-chromolast differentiation seems to be a very heterogenous process that differs depending on the system where it occurs. The most conserved terms in all studied chromoplastogenesis systems in this study were protein refolding, isoprenoid biosynthesis and chorismate biosynthesis (end product of the shikimate pathway and precursor for the phenylalanine, tryptophan, tyrosine, p-aminobenzoic acid, plastoquinone and other quinones, and salicylic acid) (Buchanan et al., 2015) among induced processes (Supplementary Figure 8), and anthocyanin biosynthesis, transmembrane transport, and regulation of transcription as repressed (Supplementary Figure 9).

5.3   Description of the chloroplast-to-chromoplast transition in leaves of *N. benthamiana*

Chromoplast differentiation in leaves of *N. benthamiana* has been eventually characterized in this thesis work from the point of view of the different families of genes that were identified. All the information obtained from the different analyses applied to the data were summarized in a heatmap for each family, plotting the fold-change (FC) values for each comparison and indicating if the gene was differentially expressed according to DESeq2 analysis (Love et al., 2014). Also, the

modules to which each gene was clustered were included, attending to two different approaches for WGCNA analysis, using all expressed genes and just including those that were differentially expressed according to construct and time variables. WGCNA has been recently used for many transcriptomic analysis, especially when several samples are analyzed and they are affected by different conditions, such as genotypes, time-points or treatments (Liu et al., 2021; Loudya et al., 2021; Shinozaki et al., 2018). FC based analyses just give a view of the comparison of two conditions, while WGCNA allows to understand how was the pattern of expression of a gene in the full experiment, and which other genes were co-expressed (Langfelder and Horvath, 2008).

Despite the increase in different carotenoids downstream of crtB-produced phytoene, notably lutein or β-carotene (Figure 4.1), the expression of the biosynthetic genes hardly changed during the time course (Figure 4.13). This suggests that the capacity of the endogenous biosynthetic enzymes was not saturated, since they were able to synthesize more product when provided with an enhanced supply of substrate without requiring an overexpression. On the other hand, many of the carotenogenic genes showed a slight but statistically significant up-regulated expression in (p)crtB samples compared to the GFP controls at 96 hpi. Summarizing, the transition from chloroplasts to chromoplasts in leaves caused by the overproduction of phytoene involves an increased production of downstream carotenoids that, initially, does not require an up-regulated expression of endogenous carotenoid biosynthetic genes. However, once the chromoplasts have already being established, these genes increase their expression, hence contributing to an increased metabolic flux through the carotenoid pathway. Despite the stagnation of carotenogenic genes in the initial steps of the chloroplast-to-chromoplast transition, there was an upstream response from both MVA and MEP pathway for the synthesis of IPP and DMAPP. The response, however, was different in each compartment, probably related to the primary metabolic precursors of each pathway. The MVA pathway and primary metabolism suffered a very early response, downregulating most of the steps from glycolysis to the TCA cycle. By contrast, genes involved in photosynthesis, photorespiration and Calvin cycle were not affected at this early point in which photosynthesis was not affected yet. This means that gene expression was responding to the initial production of phytoene or the presence of crtB protein itself (there were already high levels of crtB transcript, so crtB protein was probably already produced) or to something else not measured. Our main hypothesis is that the consumption of GGPP by crtB in the plastid somehow results in a downregulation of cytosolic pathways consuming pyruvate, one of the precursors of the MEP pathway (Figure 5.1). The rapid crtB-dependent decrease in MEP pathway products is supposed by the low levels of MEP-derived cytokinins (Kasahara et al., 2004; Sakakibara, 2006) and gibberellins (Hedden, 2020) hormones at 25 hpi also showed decrease levels at 25

hpi (Figure 4.20). The presumed retrograde pathway connecting an increased consumption of GGPP in plastids with changes in nuclear gene expression remains unknown. Leaves are the main photosynthetic machinery, representing the supply of energy and carbon for the plant (Nikolov et al., 2019; Paul and Pellny, 2003). Therefore, turning off the primary metabolism out of plastids might be a stress response to an unexpected deprivation of precursors in plastids in a theoretical source tissue.
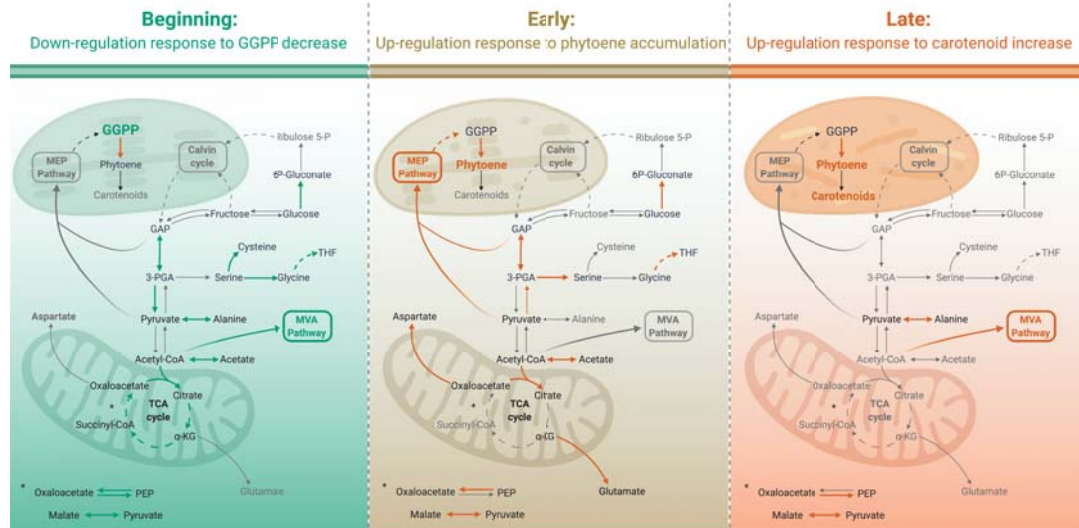
Figure 5.1: Metabolism diagram. Diagram showing representative reactions of metabolism in chloroplast, cytoplasm and mitochondria at three different events of the time course: Beginning or event (i) (22 hpi), early or event (ii) (25-28 hpi) and late of event (iv) (46-56 hpi). Green arrows represent down-regulated reactions, and orange arrows represent up-regulated reactions. Green GGPP represents its consumption, while orange phytoene and carotenoids represents their over-production and accumulation.

After the first changes in gene expression characterized by down-regulation, there was a new episode represented by genes included in the magenta-pink modules (first and second approaches of the WGCNA, respectively), showing an up-regulation at 25-28 hpi and probably responding to the onset of phytoene accumulation. Among other processes, genes from the MEP pathway, primary metabolism and jasmonic acid (JA) biosynthesis and signal transduction were particularly represented in this list (Figure 4.14, Figure 4.19,Figure 5.1). Genes encoding isoforms of *DXS*, *HDS* and *HDR* were responding from the MEP pathway, probably to activate the flux towards IPP and DMAPP synthesis. The induction of the MEP pathway expression was observed in other plastid transitions involving an activated production of cartotenoids like chloroplast to chromoplast in fruit ripening or etioplast to chloroplast in de-etiolation (Chenge-Espinosa et al., 2018; Lois et al., 2000; Rodríguez-Villalón et al., 2009; Shinozaki et al., 2018). The sequential response of MVA pathway before and MEP pathway after was also observed during de-etiolation in rice (Xin et al., 2021). Also genes involved in primary metabolism were included in this induction, most of them related to amino acid biosynthesis and the synthesis of molecules involved in transport of carbon between cellular compartments, like 2-oxoglutarate, malate, trioses phosphate, phosphoenolpyruvate, citrate, glutamate and aspartate

(Buchanan et al., 2015). All these metabolites were found to be accumulated in crtB agroinfiltrated leaves at 96 hpi (Llorente et al., 2020), and some of them were also reported to accumulate during tomato fruit ripening (Carrari and Fernie, 2006; Quinet et al., 2019). Actually, citrate and malate were reported to be the two main organic acids accumulated in both tomato and pepper ripe fruits and they were the only two intermediaries of the TCA cycle that showed increased levels at 96 hpi, so they emerge as important organic acids in chromoplast development (Llorente et al., 2020; Osorio et al., 2012).

On the other hand, the JA response was very intense, englobing biosynthetic genes and JAZ genes (Figure 4.19). JAZ proteins repress JA responses by blocking MYB2, the main activator of JA-responsive genes (Huang et al., 2017; Wasternack and Song, 2017). JA-Ile, an isoleucine conjugated of JA, mediates JAZs degradation by the SCF complex, leaving MYB2 free of the inhibition, which activates JA-responsive genes (Figure 5.2). However, *JAR1* homolog genes putatively encoding the enzyme that produces Ile-JA in the last step of the biosynthetic pathway (Wasternack and Hause, 2013; Wasternack and Song, 2017) were hardly induced during the artificial chromoplastogenesis in *N. benthamiana*.The *JMT* gene, involved in the synthesis of the methyl ester of jasmonates (MeJA), by contrast, was strongly induced (Ruan et al., 2019). In a time course performed in *Arabidopsis* where MeJA was applied to the leaves, there was no induction of MEP pathway genes at any time point (data obtained from Hickman et al., 2017), so both responses might be reacting to the same stimulus, not one to the other. While the MEP pathway is fed by the plastidial pool of pyruvate and glyceraldehyde 3-phosphate, JA is synthesized from α-linolenic acid derived from the plastidial shikimate pathway (Rodriguez-Concepcion et al., 2018; Ruiz-Sola and Rodríguez-Concepción, 2012; Wasternack and Hause, 2013; Wasternack and Hause, 2019). Although the induction of expression of JA-related genes started at 25 hpi, some of them stayed induced until the end of the time course. Actually, the levels of JA were especially elevated in (p)crtB samples at 45 and 96 hpi, despite there were no induced genes at 96 hpi. The most prominent action of JAs is response to stress, but they are also involved in many processes like seed germination, growth, stamen development, and senescence (Wasternack and Song, 2017). Regarding leaf development, JAs repress leaf expansion and plant growth (Yang et al., 2012; Zhang and Turner, 2008). However, most of the action of JAs require their interaction with COI and the SFC complex to degrade JAZs, in order to stop their inhibition of the response (Huang et al., 2017; Wasternack and Song, 2017). Most JAZ-encoding genes were also up-regulated during event (ii), i.e., at 25 and 28 hpi, and also at event (iv), from 40-46 hpi, suggesting that they might respond to drops in photosynthetic activity (Figure 4.19). A lower photosynthetic performance under the same light coditions usually leads to oxidative stress, a condition that might be linked to the production of JA, while JAZ proteins might be degraded at protein level and the induction of the expression be a response to that degradation. JAs compete with gibberellins, inhibiting many processes that the

other promote and vice versa (Huang et al., 2017). Therefore, the induction of JAs could also be a response to the early decrease in gibberellin content as their antagonist.
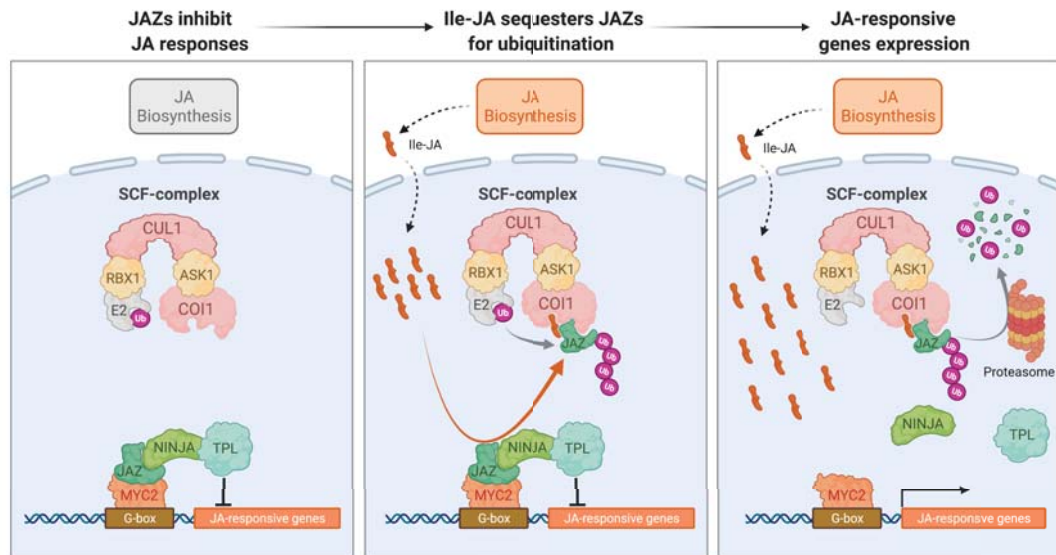
Figure 5.2: JA signaling model. Simplified model of JA-Ile perception and signaling via the SCF^{COI1}–JAZ co-receptor complex. JAZs repress positive regulators of JA signaling MYC2 together with NINJA and TOPLESS (TPL). When Jasmonate biosynthesis is induced and JA-Ile is produced, it enters into nucleus and recruit JAZs to the SCF-complex, which ubiquitinates JAZs, tagging them for their degradation throught the 26S proteasome. COI1, ASK2, CULLIN1, Rbx, and E2 are components of the SCF^{COI1} complex. Ub, ubiquitin. [Modified from Wasternack and Song, 2017].

The induction episode at 25-28 hpi was concomitant to the first affections in photosynthetic activity, which was reduced in crtB samples. Nevertheless, genes encoding for photosynthesis or light complexes did not show any repression; on the contrary, their expression was induced until 46 hpi (Supplementary Figure 38, Supplementary Figure 39). This is consistent with the idea that plastids were loosing their identity as chloroplasts, but the cell tried to avoid it by inducing the expression of the photosynthesis-related genes. The second phase of photosynthesis drop was at 46 hpi, and the ϕPSII values were already 0 at 56 hpi. At this point, the induction of photosynthesis-related genes stopped, assuming the impossibility to return to the previous "normality" as a chloroplast. That corresponded to the second peak of changes in gene expression at event (iv). Most of the photosynthesis-related and light complexes-related genes were assigned to blue and turquoise models of the first WGCNA analysis (Supplementary Figure 38, Supplementary Figure 39). Most of chlorophyll biosynthesis genes, MVA pathway genes, Calvin-Benson genes and many genes encoding primary metabolism enzymes were also included in these modules (Figure 4.14B, Supplementary Figure 37A, Supplementary Figure 40, Supplementary Figure 41). The general profile of these genes was down-regulation in events (i) and (ii) for the genes encoding proteins with non-plastidial prediction, turning to an up-regulation at event (iv), while those encoding plastidial predicted

proteins showed an up-regulation at 40-46 hpi. These dynamics of gene expression illustrate a coordinate process in a compartment-dependent manner.

Chaperones showed a particular pattern of expression during the chromoplast transition in *N. benthamiana*. In particular, genes encoding heat-shock proteins (Hsp20s, Hsp70s, Hsp90s and Hsp100s) and small chaperonins (Chap20s and Chap60s) were in general down-regulated in all time-points of the time course (from 22 to 56 hpi), but then they showed a remarkable up-regulation at 96 hpi (Figure 4.18, Supplementary Figure 36). Chaperones participate in signaling and developmental processes (Jacob et al., 2017; Xu et al., 2012). A complex formed by the Hsp90 co-chaperon SGT1b, Hsp70 and Hsp90 was reported to be involved in JA signaling by stabilizing COI1 (Zhang et al., 2015). The two *SGT1b* homologs in *N. benthamiana* showed a downregulation at the beginning of the time course and most of the genes encoding Hsp70s and Hsp90s were constantly downregulated, while none of the three *COI* isogenes showed any perturbance in their expression, compared to GFP samples (Figure 4.19C). Hsp70s and Hsp100s are involved in protein folding and degradation of enzymes from major metabolic pathways, including MEP pathway and carotenoids (Llamas et al., 2017; Pulido et al., 2016; Rodriguez-Concepcion et al., 2019). While plastidial Hsp100 chaperones of the ClpC and ClpD type unfold protein for delivery to the Clp protease complex, which mediates the degradation of DXS and PSY, ClpB-type Hsp100 can refold them back to their active form. Most of these plastidial chaperones and their adaptors are induced during tomato fruit ripening, while a reduced activity of the Clp protease complex actually impairs chromoplast differentiation (D'Andrea et al., 2018; D'Andrea and Rodriguez-Concepcion, 2019; Sato et al., 2012; Shinozaki et al., 2018). However, most of the *N. benthamiana* homologs decrease rather than increase during the time course. Strikingly, experiments of pathogen infection in tomato leaves showed that different chaperones are down-regulated in infected samples compared with mock treatments (Genevestigator data from Badet et al., 2017; Yang et al., 2015). Since the expression of chaperones was affected from the first time-point of the time course, they might also be responding to the presence of (p)crtB or the consumption of the plastidial pool of GGPP. As a bacterial enzyme, crtB protein structure might be inducing a pathogen infection response. This response, related to an unexpected stress in the plant, might cause the reduction of the protein folding machinery that might be contributing to the loss of the chloroplast identity and boosting the transition to chromoplasts. But when the chromoplasts were already formed, this pattern is reversed and the expression of chaperones is induced, probably to maintain proteostasis in a similar way of a chromoplast from a tomato ripe fruit.

In summary, our transcriptomic analyses represent two different stages of the process: the transition from a chloroplast to a chromoplast, and the establishment of a chromoplast in a leaf. The first one is represented in Figure 5.3 and was characterized by an unexpected shock of an exogenous enzyme that consumed the plastidial
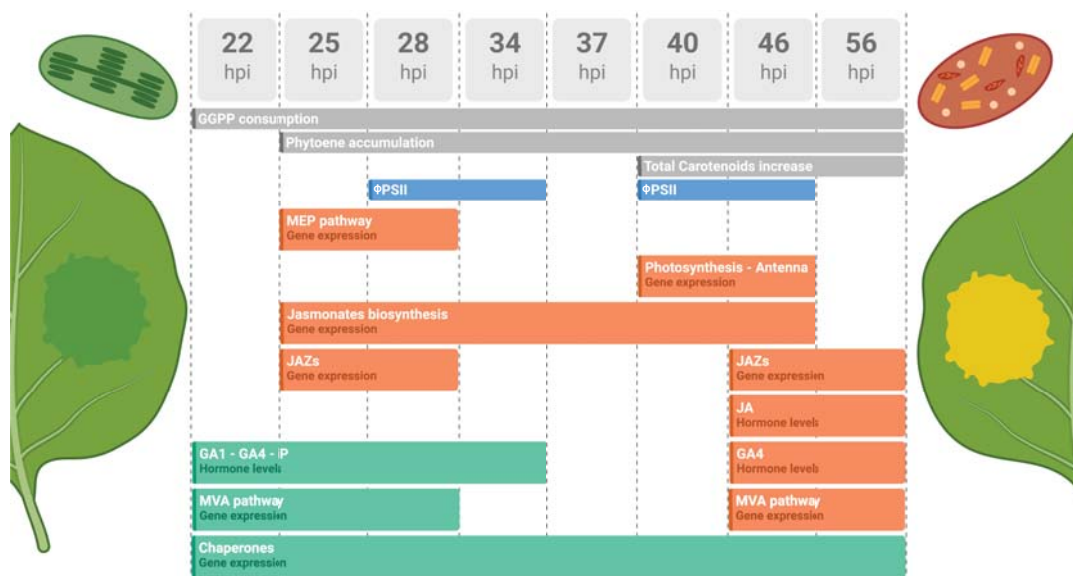
Figure 5.3: Model of chloroplast-to-chromolast transition events in *N. benthamiana*. Schematic representation of events during chloroplast-to-chromoplast transition in *N. benthamiana* leaves. In orange are highlighted up-regulated events, while down-regulated are highlighted in green.

pool of GGPP and produced phytoene, turning off primary metabolism in the cytosol and the chaperone system immediately . Then, biosynthetic genes for MEP pathway, JAs and specific primary metabolites (e.g., organic acids) induced their expression as phytoene was accumulated. That created a first affection to photosynthesis, that nuclear expression tried to avoid by inducing photosynthesis-related genes. That first shock was followed by a reduction of gene expression, but the transition did not stop since crtB was still producing phytoene, so a second peak of expression was overlapped with the 2-fold increase of carotenoids and the total dissipation of photosynthesis. This transition took part with no changes in the expression of carotenoid biosynthetic genes and the constant repression of the chaperone system. On the other hand, the expression profile of most of the genes was the complete opposite once chromoplasts were already established.

5.4 Functional annotation of the *N. benthamiana* genome

The annotation of the genome of *N. benthamiana* in the course of this thesis allowed us to describe the system with an unprecedent resolution. The previously published genome was Niben.v1.0.1, which was publicly released in 2014 in the Solgenomics webpage (Bombarely et al., 2012; Fernandez-Pozo et al., 2015). This version was not included in most of the widely used bioinformatic tools like Mapman (RRID:SCR_003543), KEGG or Panther (Kanehisa, 2019; Kanehisa and Goto, 2000; Thomas et al., 2003), and most of their genes were not properly mapped or/and annotated. The release of the new Niben.v2.6.1 assembly by Prof. Aureliano Bombarely allowed a higher coverage of the genome (61,411 sequences vs. 56,094 in the

previous version), including 83 sequences from the plastome and assigning the nuclear sequences to each of the 19 chromosomes. While this new genome remains unavailable for the general scientific community, we had access to it thanks to the collaboration with Prof. Bombarely initiated by my 3-month research stay in his lab at the University of Milan.

The lack of agreement for genome annotations have generated many issues that hinder the scientific advance and progress. Since the first plant genome was sequenced in 2000, more than 300 plant species have been sequenced (Kersey, 2019; The Arabidopsis Genome Initiative, 2000) and the aim is to increase these number to reach the 10.000 plant species sequenced and even all eukaryotes in the next decade (Lewin et al., 2018; Twyford, 2018). Many of these plants have been sequenced more than once by different scientific teams, using different techniques and annotation codes. Moreover, there are different online databases to compile assemblies and annotations from different plants, some of them generalistic (e.g., NCBI, Ensembl Plants or Phytozome) while others are restricted for subsets of plants species (e.g., The Arabidopsis Information Resource or Sol Genomics Network), each one using their own annotations (Agarwala et al., 2018; Fernandez-Pozo et al., 2015; Goodstein et al., 2012; Kersey et al., 2018; Lamesch et al., 2012). This diversity extremely hampers the unification of all available data. Most of the published genome assemblies have been more focused in gene discovery than gene functional annotation, which requires an extra effort in curation (Kersey, 2019). One of the most widely used tools for functional studies are Gene Ontologies (GO), that associate genes to processes, functions and compartments in an universal language of terms (Carbon et al., 2017). The ideal scenario would be that all genes were associated to GO terms, and that most of these associations were validated by experimental evidence. For *Arabidopsis*, the best known plant species, >94% of the protein-coding genes have been associated to GO terms, and half of the total genes include validated annotations (Kersey, 2019). In stack contrast, the majority of the sequenced plant species (including *N. benthamiana*) show a much lower percentage of genes associated to GO terms and include very limited -if any- experimental validation. The functional annotation of the new *N. benthamiana* genome carried out in this thesis work facilitated access to available tools for transcriptomic analyses, but GO term enrichment analysis includes many assumptions that made them useful for preliminary analysis only, not for studies in detail. In order to obtain the most accurate information from specific functional groups in the absence of any experimental data, phylogenetic analyses using well annotated species are the most robust approach to identify the best candidate genes. To illustrate it with an example, the functional annotation based on GO terms associated twelve *N. benthamiana* genes to the "Carotenoid Biosynthesis Process" term (GO:0016117), while the following construction of the phylogeny of the carotenoid biosynthetic pathway revealed that it actually contained forty different genes. This fact illustrates the inconvenience of basing any analysis just in GO enrichment, that would imply the loss of more than 2/3 of the information. For the

last years, efforts have been done regarding the unification of data and the generation of tools that allow the easy and efficient curation of genes for different species. The last frontier is the involvement of the scientific community and the aim to make things better.

Conclusions

1. The bacterial phytoene synthase crtB triggers chloroplast-to-chromoplast differentiation in leaves of *N. benthamiana*, but only when it localizes in plastids, producing phytoene from the pool of GGPP derived from the MEP pathway.

2. The chromoplastogenesis process is associated with carotenoid overaccumulation and a yellow phenotype that varies depending on the age of agroinfiltrated leaves and position of the agroinfiltrated area on the leaf.

3. Chromoplast differentiation is not a fixed developmental program *per se* but a very heterogenous process in different plant systems, showing many differences and particularities at the gene expression level.

4. In *N. benthamiana* leaves, crtB-dependent artificial chromoplastogenesis shows an early burst on gene expression linked to an initial drop in photosynthesis followed by a lag period and a second peak of gene expression changes and loss of photosynthetic activity.

5. The expression of genes encoding carotenogenic enzymes is not affected during the chloroplast-to-chromoplast transition, but it is induced once chromoplasts are developed.

6. The consumption of GGPP by crtB at the beginning of the process is linked to a decrease in MEP-derived gibberellins and cytokinins, and a down-regulation of genes encoding enzymes of related cytosolic pathways.

7. Chaperone gene expression is globally down-regulated all over the chloroplast-to-chromoplast transition.

8. The initial photosynthesis drop associated to the accumulation of phytoene correlates with an induction of genes encoding MEP pathway enzymes and genes involved in the production and signaling of jasmonates via JAZs. Jasmonic acid accumulates later during chromoplastogenesis.

9. Many of the genes that are down-regulated during the initial stages are later up-regulated during the second and last photosynthesis drop, suggesting antagonistic responses. After this, many genes also show opposite expression trends as chromoplast differentiation is completed.

10. The new Niben261 genome provides a manually curated annotation of over 50 gene families, available for the scientific community.

Materials and methods

7.1 Plant Material and Growth Conditions

N. benthamiana wild type and RDR6i were grown under standard greenhouse conditions (14 h light at $26 \pm 1°C$ and 10 h dark at $22 \pm 1°C$, with 50-60% of relative humidity). *Arabidopsis* lines used in this thesis were in the Columbia-0 (Col-0) background, grown under long day (LD) conditions (8 h in darkness and 16 h under fluorescent white light at a photosynthetic photon flux density (PPFD) of 60 $molm^{-2}s^{-1}$). *Arabidopsis* seeds were sterilized immersing them in 1 mL 70% ethanol and shaking for 5 minutes, then they were immersed in 1 mL 100% ethanol for 1 minute and finally they were dried in a laminar flow hood cabinet. Seeds were sown on sterile Murashige and Skood with 1% agar and no sucrose, adding the corresponding antibiotic for the selection of transgenic lines and mutants. Sown plates were stratified for 2 days at 4°C in the dark before transferring them to LD grow chambers. Different approaches were performed for the induction with 17β-estradiol (Sigma), applying it at the growing medium plates, spraying plants of different growths, using a brush to apply it to the leaves and transferring plants to hydroponic mediums of liquid Murashige and Skood with the 17β-estradiol, everything at different concentrations, from 5 to 100 °M. Heat-shock inductions were performed transferring plants to 42°C chambers for 1 h.

7.2 Transient expression assays

For transient expression, *Agrobacterium tumefaciens* strain GV3101 carrying vectors of interest were grown at 28°C shaking overnight, and optical density (OD) was measured at 600 nm. Volume was adjusted for a final OD of 0.5, centrifuge 10 min at 4000 rpm and resuspended in 10 mL agroinfiltration solution (10 mM MES-NaOH, 150μM acetosyringone and 10 mM $MgCl_2$). The cultures were incubated shaking 2h at 28°C and infiltrated with 5 mL syringes. 3rd and 4th leaves were used, infiltrating both hemispheres of the leaf at the proximal region (one construct per hemisphere) in small wounds previously made with a needle. When *N. benthamiana* wild-type plants were used, gene silencing was prevented co-agroinfiltrating with a strain carrying the helper component protease (HC-pro) of the *Watermelon mosaic virus* (WMV) cloned in vector pGWB702, kindly provided by Juan José Lopez Moya

and Maria Luisa Domingo Calap. Material was collected at corresponding time cutting the agroinfiltrated area with scissors and transferred it to 2 mL Eppendorf tube, freezing it immediately in liquid nitrogen, followed by 24 h freeze-drying.

7.3 Gene Constructs

Table 7.1: DNA constructs

Construct[a]	Vector name[a]	Bacterial selection	Marker for plant
eGFP	*pB7FWG2*	Spc	Km
crtB	*pDONR207*	Genta	-
(p)crtB	*pDONR207*	Genta	-
1S-crtB	*pDONR207*	Genta	-
crtB-Ctag	*pDONR207*	Genta	-
(p)crtB-Ctag	*pDONR207*	Genta	-
SlPSY2-Ctag	*pDONR207*	Genta	-
pHsp70	*pDONR221-P1-P4*	Km	-
(p)crtB	*pDONR221*	Km	-
tNOS	*pDONR221-P3-P2*	Km	-
35S:crtB	*pGWB405*	Spc	Km
35S:(p)crtB	*pGWB405*	Spc	Km
35S:1S-crtB	*pGWB405*	Spc	Km
35S:(c)crtB	*pGWB506*	Spc	Hyg
35S:crtB	*pGWB605*	Spc	BASTA
35S:(p)crtB	*pGWB605*	Spc	BASTA
XVE:crtB	*pER8*	Spc	Hyg
pHsp70:crtB	*pGWB501*	Spc	Hyg
35S:AtPSY	*pGWB405*	Spc	Km
35S:SlPSY2	*pGWB405*	Spc	Km

[a]In orange are highlighted those constructs generated in this study. In black are construct previously generated in the lab.

7.4 Gene Expression Analysis

Total RNA was extracted from grinded lyophilized tissue ($\sim$4 mg) using the Maxwell RSC Plant RNA Kit (Promega). RNA was quantified using a NanoDrop (Thermo Scientific) and its integrity was analyzed running 2 μL in an agarose gel electrophoresis. The cDNA synthesis was performed according to the recommendations of the Transcriptor First Strand cDNA synthesis Kit (Roche) or NZY First-Strand cDNA Synthesis Kit (NZYtech). RT-qPCR was done in a total reaction of 20 μL in 96 wells plates or 10 μL in 384 wells plates using LightCycler 480 SYBR Green I Master (Roche) on a LightCycler 480 Real-Time PCR System (Roche). Relative expression of target genes was calculated using NbACT expression as reference.

<div align="center">

Table 7.2: cDNA synthesis

</div>

Reagent	Volume (μl)	Thermocycling conditions
Template RNA (1 μg)	Up to 11	5 minutes 55°C
Oligo dT	1	
Nuclease-free water	Up to 13	
Buffer (1X)	4	30 minutes 4°C
dNTPs (1 mM)	2	5 minutes 65°C
RNase Inhibitor (4U/ μl)	0.5	
RTase (10U)	0.5	
Final volume	20	

<div align="center">

Table 7.3: qPCR mix reaction

</div>

Reagent	Volume (μl) for 96 plates	Volume (μl) for 384 plates
SYBR Green I Master Mix	10	5
Primer Fw (300 nM)	0.6	0.3
Primer Rv (300 nM)	0.6	0.3
Template cDNA (50 ng)	1	2
Nuclease-free water	7.8	2.4
Final volume	20	10

<div align="center">

Table 7.4: qPCR program

</div>

Step	Temperature	Time	Cycles
Taq activation	95°C	10 minutes	45
Primer Fw (300 nM)	95°C	10 seconds	
Primer Rv (300 nM)	60°C	30 seconds	

<div align="center">

Table 7.5: qPCR primers

</div>

Gene	AGI Locus	Primers used for qPCR (Fw/Rv)	Reference
ACT	Niben261Chr01g0820008.1	5' TAAGGTTGTTGCACCACCAG 3' 5' ACATCTGCTGGAATGTGCTG 3'	This study
Chap60B	Niben261Chr03g1204001.1	5' TCCCTGCGATTAAGGACAAG 3' 5' TTCCTGCATTTTGGGCTATC 3'	This study
Hsp18	Niben261Chr12g0008007.1	5' TGCTGCAGATTAGGGGAGAG 3' 5' GGCAACCTAAATCGCCTAAG 3'	This study
Hsp3	Niben261Chr04g1789009.1	5' CAGTGGGGAGAGAAGCAGAG 3' 5' TCTGGCAGCCTAAACCTTCTC 3'	This study

7.5 Metabolite Analysis

Leaf carotenoids, tocopherols and chlorophylls were extracted as described by Morelli, 2021: 4 mg of freeze-dried leaf tissue was used in 2 mL Eppendorf tubes, adding 375 μL of methanol as extraction solvent and 25 μL of 10% (w/v) solution of canthaxanthin in chloroform (Sigma) as internal standard. 10 seg shaking in vortex were followed by lysing of the tissue with 4 mm glass beads for 1 min at 30 Hz in a TissueLyser II (QIAGEN), the addition of 400 μL of Tris-NaCl pH 7.5, 1 min at 30 Hz in TissueLyser II, the addition of 800 μL of chloroform and another final 1 min at 30 Hz in TissueLyser II. Samples were centrifuged 5 min at 13,000 rpm in cold (4μC). The lower organic phase was transferred in a new 1.5 mL Eppendorf tube and

evaporated for 1 h in a SpeedVac system (Eppendorf Cooncentrator plus). 200 μL of acetone were added for the solution of the dry metabolites, sonicating them for 15 seg seconds to collect all material from the walls of the tube, and finally the solution was filtered with 0.2 μm filters into amber-colored 2 mL glass vials. A 10 μL aliquot of each sample was then injected onto an Agilent Technologies 1200 series HPLC system. A C30 reverse-phase column (YMC Carotenoid, 250 x 4.6 mm x 3 μm) was used, with three mobile phases consisting of methanol (A), water/methanol (20/80 v/v) containing 0.2% ammonium acetate (w/v) (B), and tert-methyl butyl ether (C). Metabolites were separated with the following gradient: 95% A, 5% B isocratically for 12 min, a step-up to 80% A, 5% B, 15% C by 12 min, followed by a linear gradient up to 30% A, 5% B, 65% C by 30 min. The flow rate was maintained at 1 ml/min. The HPLC equipment was coupled to a Photometric Diode Array (PDA) detector (Santa Clara, CA) allowing the detection of the full UV-visible absorption spectra of the different metabolites. Peak areas of chlorophylls at 650 nm, carotenoids at 472 nm (lycopene, lutein, carotene, violaxanthin, neoxanthin, canthaxanthin) or 290 nm (phytoene) were determined using the Agilent ChemStation software. A fluorescence detector at 330 nm was used for tocopherols identification. For quantification, samples only containing canthaxanthin were ran at the begging and end of the experiment, in order to calculate the average of material that was degraded during the run. The peak of each metabolite was relativized to the peak of canthaxanthin of each sample, the average canthaxanthin calculated for degradation, and the dried weight of each sample, to obtain the final values in μg of metabolite per mg dry weight of tissue.

For hormones extraction, 20 mg of grinded lyophilized agroinfiltrated leaves of *N. benthamiana* were suspended in 80% methanol – 1% acetic acid, containing internal standards and mixed by shaking during one hour at 4°C. The extract was kept at -20°C overnight and then centrifuged, collecting the supernatant and drying it in a vacuum evaporator. The dry residue was dissolved in 1% acetic acid and run through an Oasis HLB (reverse phase) column, as described in ((Seo et al., 2011).

For gibberellins, auxins, abscisic acid, salicylic acid and jasmonic acid (JA) quantification, the dried eluate was dissolved in 5% acetonitrile – 1% acetic acid, and hormones were separated using an autosampler and a reverse phase UHPLC chromatography (2.6 μg Accucore RP-MS column, 100 mm length x 2.1 mm i.d.; ThermoFisher Scientific) with a 5 to 50% acetronitrile gradient containing 0.05% acetic acid, at 400 μL/min over 21 min.

For cytokinins, the extracts were additionally passed through an Oasis MCX (cationic exchange) and eluted with 60% methanol – 5% NH4OH to obtain the basic fraction containing cytokinins. The final eluate was dried and dissolved in 5% acetonitrile – 1% acetic acid, to finally separate cytokinins with a 5 to 50% acetonitrile gradient over 10 min. The hormones were analyzed with a Q-Exactive mass spectrometer (Orbitrap detector; ThermoFisher Scientific) by targeted Selected Ion

Monitoring (SIM). The concentrations of hormones in the extracts were determined using embedded calibration curves and the Xcalibur 4.0 and TraceFinder 4.1 SP1 software. The internal standards for quantification of each of the different plant hormones were the deuterium-labelled hormones, except for JA, for which dhJA was used.

7.6 Protein Analysis

Total protein extracts were obtained from $\sim$5 mg of *N. benthamiana* grinded lyophilized leaves. The tissue was resuspended in 2 mL Eppendorf tube in 300 μL cold TKMES homogenization buffer (100 mM Tricine-potassium hydroxide pH 7.5, 10 mM KCl, 1 mM MgCl2, 1mM EDTA, and 10% [w/v] sucrose) supplemented with 0.2% (v/v) Triton X-100, 1 mM DTT, 100 μg/mL PMSF, 3 μg/mL E64, and 1X Sigma plant protease inhibitor. The sample was vortexed for resuspension and homogenization, and centrifuged at 12,000 rpm for 10 minutes at 4°C. The supernatant was transferred to a new Eppendorf tube, and another centrifuge was performed in the same conditions. This process was repeated 3-4 times, until no pellet was observed, collecting a final volume of $\sim$200 μL. Protein concentration was determined using the kit Coomasie Plus (Bradford) Assay Kit (Thermo Scientific). Measurements were performed in a total volume of 200 μL of Bradford with 2 μL of 1:4 dilution of protein extract in a 96 well plate using the SpectraMax M3 multi-mode microplate reader (Molecular Devices). A standard curve was performed using diluted albumin (BSA) standards from the kit, as follows in the kit protocol. 4 technical replicates for each measure were made. After quantification, loading buffer 4X was added (250 mM Tris-HCl pH 6.8, 4% 2-mercaptoethanol, 8% [v/v] sodium dodecyl sulfate [SDS], 0.02% [w/v] bromophenol blue, and 40% [v/v] glycerol, in milli-Q) and the solution was boiled at 96°C for 5 min. Samples were run at 120 V in 1.5 mm SDS-PAGE gel (4% upper stacking gel solution: 4% [v/v] acrylamide, 25% [v/v] upper buffer pH 6.8, 0.1% [v/v] TEMED, 0.05% [v/v] ammonium persulfate [APS]; 10% running gel solution: 10% [v/v] acrylamide, 25% [v/v] lower buffer pH 8.8, 0.05% [v/v] TEMED, 0.05% [v/v] ammonium persulfate [APS]). Proteins were electrotransferred to hybond-P polyvinylidene difluoride membranes (Amersham) using a Trans-Blot Turbo Transfer System (Bio-Rad) 12 min at 2.5 A constant and 25V. Membranes were blocked for 1 h in 5% (w/v) powder milk in 1% (v/v) phosphate buffered saline (PBS) + 0.5% (v/v) Tween 20, and then incubated overnight at 4°C with the corresponding primary antibody. Incubation with the horseradish peroxidase-conjugated secondary antibody (1:10,000) was performed for 1 h at room temperature. Detection of protein bands was performed using the WesternBright ECL Western blotting detection kit (Advansta). Chemiluminescent signals were visualized using ChemiDoc System (Bio-Rad).

Table 7.6: Antibodies used

Antibody	Dilution	Origin	Protein concentration	Supplier
ClpB3	1:2,000	Rabbit	30 μg	Agrisera
AtpB	1:5,000	Rabbit	5 μg	
PsbA	1:2,000	Rabbit	5 μg	
PsaB	1:2,000	Rabbit	5 μg	
Fibrillin	1:4,000	Rabbit	5 μg	

7.7 Confocal Microscopy Analysis

GFP and chlorophyll fluorescence were determined with a Leica TCS SP5 confocal microscope for subcellular localization, using an argon laser for excitation at 488 nm and 450-490 nm filter for detection of GFP fluorescence, and 610-700 nm filter for detection of chlorophyll fluorescence.

7.8 Transmission Electron Microscopy

Agroinfiltrated regions of *N. benthamiana* leaves were cut in small rectangles ($\sim$0.5 x 0.2 mm) and fixed in 2.5% glutaraldehyde and 2% paraformaldehyde in 100 mM sodium cacodylate buffer pH 7.4. in vacuum for 4 hours. Samples were post-fixated in 1% buffered OsO_4 for 1 h, rinsed in water, dehydrated in a gradient of acetone, and embedded in Spurr's resin. After microtomy, ultrathin sections were stained with 2% uranyl acetate for 30 min followed by 2.6% lead citrate for 10 min. Samples were observed using a FEI Tecnai G2 F20 electron microscope operating at 200 keV and recorded using FEI Eagle 4 k x 4 k CCD camera.

7.9 Photosynthetic measurements

Chlorophyll fluorescence measurements were carried out as explained by Morelli, 2021: using a MAXI-PAM fluorometer (Heinz Walz GmbH). *N. benthamiana* leaves were placed under the camera and maximum quantum yield of PSII, Fv/Fm, was calculated as (Fm - F_0)/Fm, where Fm and F_0 are respectively the maximum and the minimun fluorescence of dark-adapted samples, after a pulse of actinic light. The light intensity used for the measurements was 21 PAR for all leaf areas analyzed as the last value able to generate a response in the (p)crtB-infiltrated areas before having a null photosynthetic activity. Each value is the average result of three technical replicates in the same leaf and three biological replicates from different plants. Effective quantum yield of photosystem II (PSII) under growth light, ϕPSII, was measured as ΔF/Fm', where ΔF corresponds to Fm' - F (the maximum minus the minimum fluorescence of light-exposed plants), applying pulses of actinic light every 30 second until the values where stabilized, taking the last measure of the serie.

7.10 RNA-seq Analysis

RNA was extracted as described for gene expression analysis. Quality was measured obtaining the RNA Integrity Number (RIN) with the Agilent 2100 Bioanalyzer System using the Plant RNA Nano program. Sequentia Biotech SL (Barcelona, Spain) performed the 96 hpi analysis. Indexed libraries were prepared from 1 μg/ea purified RNA with TruSeq Stranded mRNA Sample Prep Kit (Illumina), according to the manufacturer's intructions. Libraries were quantified using the TapeStation 4200 (Agilent Technologies) and pooled such that each index-tagged sample was present in equimolar amounts, with final concentration of the pooled samples of 2 nM. The pooled samples were subject to cluster generation and sequencing using a NextSeq 500 System (Illumina) in a 2x75 paired end format at a final concentration of 1.8 pmol. The raw sequences files generated (.fastq files) underwent quality control analysis using FastQC (`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`). Data analysis was performed with the online platform AIR (`www.transcriptomics.cloud`) using the SolGenomics Network N. benthamiana v1.01 (Niben v101) reference genome (Bombarely et al., 2012). For time course analysis, strand-specific RNA-seq library preparations were performed as described in (Zhong et al., 2011) with 24 independently bar-coded samples multiplexed and sequenced (75bp, single end) on one lane of the Illumina NextSeq 500/500 platform.

7.10.1 Trimming

To clean the data, fastq-mcf command was used to discarding sequences with a quality lower than 30 and shorter than 50 bp. A fasta file containing Illumina adapter sequences was used to identify them. Compressed fasta files were generated as output with .fq.gz extension.

This is an example of the code used for the trimming:

Listing 7.1: Trimming

```
fastq-mcf -q 30 -l 50 -o filename_q30_l50.fq.gz
    adapterfile.fasta filename.fastq.gz
```

7.10.2 Mapping

Trimmed sequences were mapped to the reference genome using STAR (Dobin et al., 2013). First, the reference genome was indexed, and then each sequence was mapped to it. Then, .fq.gz trimmed files were aligned to the reference genome using the index, generating the corresponding BAM file.

This is an example of the code used to create the index of the genome and for the mapping of trimmed sequences from a sample:

Listing 7.2: Mapping

```
STAR --runThreadN 8 --runMode genomeGenerate
    --genomeDir indexdirectory
    --genomeFastaFiles referencegenome.fasta
    --sjdbGTFfile referencegenome.gtf
    --genomeSAindexNbases 12

STAR --runThreadN 8 --genomeDir indexdirectory
    --readFilesIn filename_q30_l50.fq.gz
    --outFileNamePrefix prefixforoutputfile
    --readFilesCommand "gunzip -c"
    --outSAMtype BAM SortedByCoordinate
```

7.10.3   Quantification

Gene counts were obtained from BAM files using HTSeq (Anders et al., 2015).

This is an example of the code used to quantify genes in one sample:

Listing 7.3: Quantification

```
htseq-count -f bam -m union -i ID -a 10 --type=mRNA
    --stranded no filename.bam $referencegff
    filename.counts
```

7.10.4   Differential Expression Analysis

To obtain lists of DEGs, DESeq2 package was used in R following the in-struction of the manual (Love et al., 2014). For the lists of DEGs in time points or alternative group comparisons, ∼Construct design was used within the `DESeqDataSetFromHTSeqCount` function, using GFP samples as reference. The different groups obtained from these comparisons were those attending to the time-points *per se*, but also some creating alternative groups depending of the clustering in the PCA plots (Table 7.7)

To obtain the list of DEGs according to Time and Constructs variables (i.e., those genes differentially expressed in (p)crtB samples compared to GFP throughout the time course) used for the second WGCNA analysis, all samples were used and the design was changed to ∼Construct+TimePoint+Construct:TimePoint.

Table 7.7: Samples included in DESeq2 comparisons

Comparison	GFP samples	(p)crtB samples
22h	22h-1, 22h-2, 22h-3	22h-1, 22h-2, 22h-3
22h-25h	22h-1, 22h-2, 22h-3 25h-1, 25h-2	22h-1, 22h-2, 22h-3 22h-1, 25h-2
25h	25h-1, 25h-2, 25h-3	25h-1, 25h-2, 25h-3
25h-28h	25h-1, 25h-2, 25h-3 28h-1, 28h-2, 28h-3	25h-1, 25h-2, 25h-3 28h-1, 28h-2, 28h-3
25h-28h	25h-3 28h-1, 28h-2, 28h-3	25h-3 28h-1, 28h-2, 28h-3
28h	28h-1, 28h-2, 28h-3	28h-1, 28h-2, 28h-3
34h	34h-1, 34h-2, 34h-3	34h-1, 34h-2, 34h-3
34h-37h	34h-2, 34h-3 37h-1, 37h-3	34h-2, 34h-3 37h-1, 37h-3
37h	37h-1, 37h-2, 37h-3	37h-1, 37h-2, 37h-3
34h-37h-40h	34h-3 37h-1, 37h-3 40h-1	34h-3 37h-1, 37h-3 40h-1
40h	40h-1, 40h-2, 40h-3	40h-1, 40h-2, 40h-3
40h-46h	40h-2, 40h-3 46h-2, 46h-3	40h-2, 40h-3 46h-2, 46h-3
46h	46h-1, 46h-2, 46h-3	46h-1, 46h-2, 46h-3
46h-56h	46h-1 56h-1, 56h-2, 56h-3	46h-1 56h-1, 56h-2, 56h-3
56h	56h-1, 56h-2, 56h-3	56h-1, 56h-2, 56h-3

7.10.5 WGCNA

For the creation of the gene-coexpression modules, WGCNA package was used in RStudio (Langfelder and Horvath, 2008). Tutorials from the authors were followed, using the automatic, one-step network construction and module detection with a power of 12, according to the recommendations of the developers (our sample size was 48 samples), TOMtype= "signed", maxBlockSize= 20000, minModuleSize=30, reassignThreshold= 0, mergeCutHeight= 0.25 and verbose= 3 arguments for the `blockwiseModules` function.

7.10.6 GO enrichment

At the beginning of this work, GO enrichment analyses were performed using Parametric Analysis of Gene Set Enrichment (PAGE) from AgriGO v2.0 webpage, using list of DEGs together with their FC values.

Later, GO enrichment analyses were performed in RStudio using topGO package (Alexa and Rahnenfuhrer, 2021), following the manual instructions, using `algorithm= "weight01"` and `statistic= "fisher"` within the `runTest` function. Some examples of these analysis are included in the GitHub repository.

7.11   Genome functional annotation

7.11.1   General annotation

To obtain the functional annotation of the new *N. benthamiana* genome (Niben261) and to re-annotate tomato (ITAG4.0) and Pepper (Capsicum_annuum.ASM51225v2 from Ensembl Plants) genomes, assigning a predictive name to each gene and associating GO terms to them, first a BLAST was run against three different databases (Araport11, Swissprot and Trembl/GeneBank) using DIAMOND program. This is an example:

Listing 7.4: Diamond

```
diamond blastp --outfmt 6 --threads 100
    --db /directory/uniprot_trembl --out
    my_genome.pep.diamond.trembl.o6.txt
    --query my_genome.pep.fasta --more-sensitive
```

After that, Interproscan was run with protain domains:

Listing 7.5: Interproscan

```
interproscan.sh --cpu 100 --pathways --iprlookup --goterms
    --output-file-base my_genoma.pep.ipr
    --formats TSV,XML,GFF3 -i my_genoma.pep.fasta
```

Then, all data was integrated using Automated Assignment of Human Readable Descriptions (AHRD) program. The `ahrd_annotation_conf.yml` file was used with default parameters including our file names, and the following command was run:

Listing 7.6: AHRD

```
java -Xmx64g -jar /data/00_software/AHRD/dist/ahrd.jar
    ahrd_annotation_conf.yml
```

7.11.2   Specific manual annotation of gene families

To annotate specific gene families more accurately, *Arabidopsis* was used as reference. The desired genes were researched in the bibliography, obtaining the protein sequences from Araport11. A BLAST was run against tomato, pepper and *N. benthamiana* proteomes, obtaining three lists of candidates homologs. A smooth filter was applied to these lists, keeping only those sequences with an percentage of identical matches equal of above 35%. The four list of sequences (the original list from *Arabidopsis* and the three lists of candidates in tomato, pepper and *N. benthamiana*)

were aligned. For that, all sequences were concatenated in a single fasta file, and then Guidance program (Sela et al., 2015) was run using ClustalW alingment:

Listing 7.7: Guidance

```
perl /directory/guidance.pl --seqFile AllSequences.fasta
    --msaProgram CLUSTALW --seqType aa
    --outDir AllSequences --proc_num 64
```

Then, iqtree ((Nguyen et al., 2015)) was run for a first time to select the best model fitting the aligment to create the phylogenetic tree:

Listing 7.8: iqtree 1

```
/directory/iqtree -s AllSequences/MSA.CLUSTALW.aln.With_Names
    -m MF -nt AUTO
```

Once the model was selected (it was recorded in the `AllSequences/MSA.CLUSTALW.aln.With_Names.iqtree` file), iqtree was run for a second time with that model:

Listing 7.9: iqtree 2

```
/directory/iqtree -s AllSequences/MSA.CLUSTALW.aln.With_Names
    -m ModelSelected -nt AUTO -bb 1000
    -pre AllSequences/MSA.CLUSTALW.aln.With\_Names.Model.
        1000bootstraps
```

Finally, the *.treefile file was openned with FigTree v1.4.4, and all genes were annotated one by one based on their clustering with *Arabidopsis*. Tomato genes were checked in SolGenomics website, and *N. benthamiana* genes were annotated based on them. The normal clustering was 1:1:2 (tomato-pepper-*N. benthamiana*, respectively).

7.11.3 Subcellular prediction

The subcellular location of annotated genes of *N. benthamiana* was predicted using TargetP 2.0 (Armenteros et al., 2019) with the following command:

Listing 7.10: TargetP

```
targetp -2.0/bin/targetp -batch 500
    -fasta Niben_GeneFamily.fasta -org p
```

7.12 Data Visualization

7.12.1 General plots

All plots presented in this study were created in RStudio (RStudio Team, 2020). Among many others, the most used packages to transform data and visualize it were those included in the Tidyverse package (Wickham et al., 2019), such as tidyr, dplyr and ggplot2 (Wickham, 2021; Wickham et al., 2021; Wickham, 2016).

An example of how the bar plot of HPLC results from Figure 3.1 is showed. The data obtained in the experiment is shown in Table 8 and Table 9. In R, ggplot2 and tidyverse packages were activated with the `library` function. Then, the data was read using `read.delim` function, indicating that fields were separated by tabulators, each column contained a header (first row) and decimals were written with commas. This data was assinged to the object "HPLC.1".

The data transformed and assigned to a new object, "HPCL.2", selecting just some of the columns with the `select` function, and the data was arranged into new columns: "Sample", that contained previous column names, and "values", including the μg/mg of each metabolite, using the `gather` function. Then, the new column "Sample" was divided into two new columns, "Construct" and "Replicate", attending to the "_" chactacter, using `separate` function. Finally, two rows were excluded using the `filter` function (discarding rows where Metabolite column contained values included in ("Gammatocopherol", "Alphatocopherol") vector. Then, values from Metabolite and Construct columns were transformed to factor and sorted in specific order, using `factor` function.

The object "theme_HPLC" was created in order to specify some features of the final plot.

Finally, the plot was created using `ggplot` function, assigning the corresponding variables to x axis, y axis and fill parameters of the aesthetics. The different aspects of the plot were defined and finally it was saved as a PDF.

```r
# First, we load the libraries that we will need to used

library(ggplot2)
library(tidyverse)

# Then, we read the file of the measures. We assign it to
# an object (HPCL). We indicate that fields are separated by
# tab and it contains column names (first row)

HPLC.1 <- read.delim("/compile/Appendices/HPLC.txt",
sep="\t", header= TRUE, dec= ",")
```

```r
# Then I transform the data

HPLC.2 <- HPLC.1 %>%
  gather(Sample, values, -Metabolite) %>%
  separate(Sample, into= c("Construct", "Replicate") %>%
  filter(!Metabolite %in% c("Gammatocopherol", "Alfatocopherol"))

# Values are ordered for some columns

HPLC.2$Metabolite <- factor(HPLC.2$Metabolite,
    levels = c("Phytoene", "Lycopene", "TotalBetacarotene",
        "Lutein","Violaxanthin", "Neoxanthin", "Chlorophylla",
        "Chlorophyllb", "Total_Carotenoids",
        "Total_Chlorophylls", "Total_Tocopherols", "Ratio"),
    labels = c("Phytoene", "cis-Lycopene", "Betacarotene",
    "Lutein", "Violaxanthin", "Neoxanthin", "Chlorophyll-a",
    "Chlorophyll-b", "Total Carotenoids", "Total Chlorophylls",
    "Total Tocopherols", "Ratio Car/Chl"))

HPLC.2$Construct <- factor(HPLC.2$Construct,
    levels= c("GFP", "(c)crtB", "1S-crtB", "crtB", "(p)crtB"))

# I create an object including some settings for the plot

theme_HPLC <- theme(plot.title = element_text(hjust = 0.5,
                        face = "bold"),
                    axis.title = element_text(face = "bold"),
                    strip.text = element_text(face = "bold"),
                    panel.background = element_blank(),
                    legend.background = element_blank(),
                    legend.key = element_blank(),
                    panel.grid = element_blank(),
                    axis.text = element_text(colour = "black"),
                    axis.line = element_line(colour = "black"),
                    axis.ticks.x = element_blank()
)



# Finally, I create the plot

ggplot(HPLC.2,
    aes(x = Construct, y = values, fill = Construct))+
```

```r
geom_bar(stat = "summary", fun.y = "mean",
    position = position_dodge())+
geom_point(shape = 20, size = 3, aes(col = Construct),
    position = position_dodge(width = 0.9))+
stat_summary(fun.data = mean_sdl, geom = "errorbar",
    fun.args = list(mult = 1), width = 0.4,
    position = position_dodge(width=0.9))+
ylab(expression(mu*"g/mg dry weight"))+
scale_y_continuous(expand = expand_scale(mult = c(0, .2)))+
scale_x_discrete(labels = c("GFP", "(c)crtB",
    "1S-crtB", "crtB", "(p)crtB"))+
scale_fill_manual(values = c("GFP"= "#66c2a5",
                            "(c)crtB"= "#9cb794",
                            "1S-crtB"= "#c1ab83",
                            "crtB"= "#e19e73",
                            "(p)crtB" = "#fc8d62"))+
scale_color_manual(values = c("GFP"= "#1b9e77",
                            "(c)crtB"= "#73935e",
                            "1S-crtB"= "#9d8545",
                            "crtB"= "#bd752b",
                            "(p)crtB" = "#d95f02"))+
labs(fill = "Constructs", col = "Constructs")+
facet_wrap(~Metabolite, scales = "free_y")+
theme_HPLC+
theme(axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    title = element_text(face = "bold", size = 12),
    legend.title = element_text(face = "bold", size = 10),
    legend.text = element_text(face = "bold", size = 8),
    strip.text = element_text(size = 8),
    legend.position = "bottom",
    panel.grid.major.y = element_line(linetype = "dotted",
    colour = "grey"))+
annotate("segment", x=-Inf, xend=Inf, y=-Inf, yend=-Inf)+
annotate("segment", x=-Inf, xend=-Inf, y=-Inf, yend=Inf)
```

7.12.2 Heatmaps

All heatmaps were created in RStudio using pheatmap package (Kolde, 2019). This is an example, using a dataframe (df) containing FC values of different genes, where rownames were the name of the gene and all values were numeric:

```r
library(pheatmap)

Color <- colorRampPalette(c("#1b9e77","#8ecdb6","#f7f7f7",
        "#fba882","#d95f02"))(n=200)
Breaks <- c(seq(-3,0, length.out = 100),
        seq(0.01,3,length.out = 100))

pheatmap(df,
        border_color = "grey60",legend= TRUE,
        color=Color,
        cluster_rows =FALSE,
        breaks = Breaks,
        legend_breaks = c(-3, -1.5, 0, 1.5, 3),
        legend_labels = c("<-3", "-1.5", "0", "1.5", ">3"),
        cluster_cols = FALSE,
        show_colnames=T, treeheight_col = 35,
        fontsize=12,
        scale = "none",
        show_rownames = TRUE,
        fontsize_row = 7,
        fontsize_col = 8,
        labels_col = c("22h", "22-25h", "25h",  "25-28h",
        "25-28h", "28h", "34h","34-37h", "37h",
        "34-37-40h","40h","40-46h","46h", "46-56h","56h"),
        cex = 1,
        angle_col=45,
        main = "Title",
        annotation_legend = TRUE,
        display_numbers = df_DEG
        fontsize_number = 15,
        na_col = "darkgrey"
)
```

7.12.3 PCA plots

PCA plots were created using ggplot2 package as showed in GitHub. The block normalization for phenotypic data (pigment content from HPLC and ϕPSII from PAM) was carried out using the `blockScale` function from prospectr package (Stevens and Ramirez-Lopez, 2020).

Bibliography

Adami, M., P. D. Franceschi, F. Brandi, A. Liverani, D. Giovannini, C. Rosati, et al. (2013). "Identifying a Carotenoid Cleavage Dioxygenase (ccd4) Gene Controlling Yellow / White Fruit Flesh Color of Peach". In: *Plant Molecular Biology Reporter* 31, pp. 1166–1175. DOI: 10.1007/s11105-013-0628-6.

Agarwala, R., T. Barrett, J. Beck, D. A. Benson, C. Bollin, E. Bolton, et al. (2018). "Database resources of the National Center for Biotechnology Information". In: *Nucleic Acids Research* 46.D1, pp. D8–D13. ISSN: 13624962. DOI: 10.1093/nar/gkx1095.

Alexa, A. and J. Rahnenfuhrer (2021). *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.44.0.

Alexander, L. and D. Grierson (2002). "Ethylene biosynthesis and action in tomato: A model for climacteric fruit ripening". In: *Journal of Experimental Botany* 53.377, pp. 2039–2055. ISSN: 00220957. DOI: 10.1093/jxb/erf072.

Aluru, M., Y. Xu, R. Guo, Z. Wang, S. Li, W. White, et al. (2008). "Generation of transgenic maize with enhanced provitamin A content". In: *Journal of Experimental Botany* 59.13, pp. 3551–3562. DOI: 10.1093/jxb/ern212.

Anders, S., P. T. Pyl, and W. Huber (2015). "HTSeq-A Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu638.

Andersen, T. B., B. Llorente, L. Morelli, S. Torres-Montilla, G. Bordanaba-Florit, F. A. Espinosa, et al. (2021). "An engineered extraplastidial pathway for carotenoid biofortification of leaves". In: *Plant Biotechnology Journal*, pp. 0–2. DOI: 10.1111/pbi.13526.

Armenteros, J. J. A., M. Salvatore, O. Emanuelsson, O. Winther, G. Von Heijne, A. Elofsson, et al. (2019). "Detecting sequence signals in targeting peptides using deep learning". In: *Life Science Alliance* 2.5, pp. 1–14. ISSN: 25751077. DOI: 10.26508/lsa.201900429.

Badet, T., D. Voisin, M. Mbengue, M. Barascud, J. Sucher, P. Sadon, et al. (2017). "Parallel evolution of the POQR prolyl oligo peptidase gene conferring plant quantitative disease resistance". In: *PLoS Genetics* 13.12, pp. 1–24. ISSN: 15537404. DOI: 10.1371/journal.pgen.1007143.

Barsan, C., P. Sanchez-Bel, C. Rombaldi, I. Egea, M. Rossignol, M. Kuntz, et al. (2010). "Characteristics of the tomato chromoplast revealed by proteomic analysis". In:

Journal of Experimental Botany 61.9, pp. 2413–2431. ISSN: 14602431. DOI: `10.1093/jxb/erq070`.

Berman, J., U. Zorrilla-López, V. Medina, G. Farré, G. Sandmann, T. Capell, et al. (2017). "The Arabidopsis ORANGE (AtOR) gene promotes carotenoid accumulation in transgenic corn hybrids derived from parental lines with limited carotenoid pools". In: *Plant Cell Reports* 36.6, pp. 933–945. ISSN: 07217714. DOI: `10.1007/s00299-017-2126-z`.

Biais, B., C. Bénard, B. Beauvoit, S. Colombié, D. Prodhomme, G. Ménard, et al. (2014). "Remarkable reproducibility of enzyme activity profiles in tomato fruits grown under contrasting environments provides a roadmap for studies of fruit-metabolism". In: *Plant Physiology* 164.3, pp. 1204–1221. ISSN: 15322548. DOI: `10.1104/pp.113.231241`.

Bian, W. (2012). "The chloroplast-to-chromoplast transition in tomato fruit". In.

Bombarely, A., H. G. Rosli, J. Vrebalov, P. Moffett, L. A. Mueller, and G. B. Martin (2012). "A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research". In: *Molecular Plant-Microbe Interactions* 25.12, pp. 1523–1530. ISSN: 08940282. DOI: `10.1094/MPMI-06-12-0148-TA`.

Britton, G. (1995). "Structure and properties of carotenoids in relation to function". In: *FASEB Journal* 9, pp. 1551–1558. DOI: `10.1096/fasebj.9.15.8529834`.

Buchanan, B. B., W. Guissem, and R. L. Jones (2015). *Biochemistry and molecular biology of plants (2nd ed.)* Ed. by B. B. Buchanan, W. Guissem, and R. L. Jones. 2nd. Wiley-Blackwell.

Calumpang, C. L. F., T. Saigo, M. Watanabe, and T. Tohge (2020). "Cross-species comparison of fruit-metabolomics to elucidate metabolic regulation of fruit polyphenolics among solanaceous crops". In: *Metabolites* 10.5. ISSN: 22181989. DOI: `10.3390/metabo10050209`.

Cao, H., J. Zhang, J. Xu, J. Ye, Z. Yun, Q. Xu, et al. (2012). "Comprehending crystalline -carotene accumulation by comparing engineered cell models and the natural carotenoid-rich system of citrus". In: *Journal of Experimental Botany* 63.12, pp. 4403–4417. DOI: `10.1093/jxb/err115`.

Carbon, S., H. Dietze, S. E. Lewis, C. J. Mungall, M. C. Munoz-Torres, S. Basu, et al. (2017). "Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium". In: *Nucleic Acids Research* 45.D1, pp. D331–D338. ISSN: 13624962. DOI: `10.1093/nar/gkw1108`.

Carrari, F. and A. R. Fernie (2006). "Metabolic regulation underlying tomato fruit development". In: *Journal of Experimental Botany* 57.9, pp. 1883–1897. ISSN: 00220957. DOI: `10.1093/jxb/erj020`.

Chedea, V. S. and M. Jisaka (2013). "Lipoxygenase and carotenoids: A co-oxidation story". In: *African Journal of Biotechnology* 12.20, pp. 2786–2791. ISSN: 1684-5315. DOI: `10.5897/AJB12.2944`.

Chenge-Espinosa, M., E. Cordoba, C. Romero-Guido, G. Toledo-Ortiz, and P. León (2018). "Shedding light on the methylerythritol phosphate (MEP)-pathway: long

hypocotyl 5 (HY5)/phytochrome-interacting factors (PIFs) transcription factors modulating key limiting steps". In: *Plant Journal* 96.4, pp. 828–841. ISSN: 1365313X. DOI: `10.1111/tpj.14071`.

Cheung, A. Y., T. McNellis, and B. Piekos (1993). "Maintenance of chloroplast components during chromoplast differentiation in the tomato mutant green flesh". In: *Plant Physiology* 101.4, pp. 1223–1229. ISSN: 00320889. DOI: `10.1104/pp.101.4.1223`.

D'Andrea, L. and M. Rodriguez-Concepcion (2019). "Manipulation of Plastidial Protein Quality Control Components as a New Strategy to Improve Carotenoid Contents in Tomato Fruit". In: *Frontiers in Plant Science* 10.September, pp. 1–7. ISSN: 1664462X. DOI: `10.3389/fpls.2019.01071`.

D'Andrea, L., M. Simon-moya, B. Llorente, E. Llamas, M. Marro, P. Loza-alvarez, et al. (2018). "Interference with Clp protease impairs carotenoid accumulation during tomato fruit ripening". In: *Journal of Experimental Botany* 69.7, pp. 1557–1567. DOI: `10.1093/jxb/erx491`.

Demurtas, O. C., R. de Brito Francisco, G. Diretto, P. Ferrante, S. Frusciante, M. Pietrella, et al. (2019). "ABCC Transporters Mediate the Vacuolar Accumulation of Crocins in Saffron Stigmas". In: *The Plant cell* 31.11, pp. 2789–2804. ISSN: 1532298X. DOI: `10.1105/tpc.19.00193`.

Demurtas, O. C., S. Frusciante, P. Ferrante, G. Diretto, N. H. Azad, M. Pietrella, et al. (2018). "Candidate enzymes for saffron crocin biosynthesis are localized in multiple cellular compartments". In: *Plant Physiology* 177.3, pp. 990–1006. ISSN: 15322548. DOI: `10.1104/pp.17.01815`.

Deruere, J., S. Romer, A. D'Harlingue, R. A. Backhaus, M. Kuntz, and B. Camara (1994). "Fibril assembly and carotenoid overaccumulation in chromoplasts: A model for supramolecular lipoprotein structures". In: *Plant Cell* 6.1, pp. 119–133. ISSN: 10404651. DOI: `10.1105/tpc.6.1.119`.

Dhami, N. and C. I. Cazzonelli (2020). "Environmental impacts on carotenoid metabolism in leaves". In: *Plant Growth Regulation* 92.3, pp. 455–477. ISSN: 15735087. DOI: `10.1007/s10725-020-00661-w`. URL: `https://doi.org/10.1007/s10725-020-00661-w`.

Dhar, M. K., S. Mishra, A. Bhat, S. Chib, and S. Kaul (2020). "Plant carotenoid cleavage oxygenases : structure-function relationships and role in development and metabolism". In: *Briefings in Functional Genomics* 19.1, pp. 1–9. DOI: `10.1093/bfgp/elz037`.

Dijk, E. L. van, Y. Jaszczyszyn, D. Naquin, and C. Thermes (2018). "The Third Revolution in Sequencing Technology". In: *Trends in Genetics* 34.9, pp. 666–681. ISSN: 13624555. DOI: `10.1016/j.tig.2018.05.008`. URL: `https://doi.org/10.1016/j.tig.2018.05.008`.

Diretto, G., O. Ahrazem, Á. Rubio-Moraga, A. Fiore, F. Sevi, J. Argandoña, et al. (2019). "UGT709G1: a novel uridine diphosphate glycosyltransferase involved in the biosynthesis of picrocrocin, the precursor of safranal in saffron (Crocus

sativus)". In: *New Phytologist* 224.2, pp. 725–740. ISSN: 14698137. DOI: 10.1111/nph.16079.

Diretto, G., S. Al-Babili, R. Tavazza, V. Papacchioli, P. Beyer, and G. Giuliano (2007). "Metabolic engineering of potato carotenoid content through tuber-specific over-expression of a bacterial mini-pathway". In: *PLoS ONE* 2.4, pp. 1–8. ISSN: 19326203. DOI: 10.1371/journal.pone.0000350.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al. (2013). "STAR: Ultrafast universal RNA-seq aligner". In: *Bioinformatics* 29.1, pp. 15–21. ISSN: 13674803. DOI: 10.1093/bioinformatics/bts635.

Domonkos, I., M. Kis, Z. Gombos, and B. Ughy (2013). "Carotenoids, versatile components of oxygenic photosynthesis". In: *Progress in Lipid Research* 52.4, pp. 539–561. ISSN: 01637827. DOI: 10.1016/j.plipres.2013.07.001.

Egea, I., C. Barsan, W. Bian, E. Purgatto, A. Latché, C. Chervin, et al. (2010). "Chromoplast Differentiation : Current Status and Perspectives Mini Review". In: *Plant and Cell Physiology* 51.10, pp. 1601–1611. DOI: 10.1093/pcp/pcq136.

Eggersdorfer, M. and A. Wyss (2018). "Carotenoids in human nutrition and health". In: *Archives of Biochemistry and Biophysics* 652.May, pp. 18–26. ISSN: 10960384. DOI: 10.1016/j.abb.2018.06.001. URL: https://doi.org/10.1016/j.abb.2018.06.001.

Fernandez-Pozo, N., N. Menda, J. D. Edwards, S. Saha, I. Y. Tecle, S. R. Strickler, et al. (2015). "The Sol Genomics Network (SGN)-from genotype to phenotype to breeding". In: *Nucleic Acids Research* 43.D1, pp. D1036–D1041. ISSN: 13624962. DOI: 10.1093/nar/gku1195.

Ferrero, S., R. E. Grados-Torrez, P. Leivar, M. Antolín-Llovera, C. López-Iglesias, N. Cortadellas, et al. (2015). "Proliferation and morphogenesis of the endoplasmic reticulum driven by the membrane domain of 3-hydroxy-3-methylglutaryl coenzyme a reductase in plant cells". In: *Plant Physiology* 168.3, pp. 899–914. ISSN: 15322548. DOI: 10.1104/pp.15.00597.

Fraser, P. D., E. M. Enfissi, J. M. Halket, M. R. Truesdale, D. Yu, C. Gerrish, et al. (2007). "Manipulation of phytoene levels in tomato fruit: Effects on isoprenoids, plastids, and intermediary metabolism". In: *Plant Cell* 19.10, pp. 3194–3211. ISSN: 10404651. DOI: 10.1105/tpc.106.049817.

Fraser, P. D., S. Romer, C. A. Shipton, P. B. Mills, J. W. Kiano, N. Misawa, et al. (2002). "Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.2, pp. 1092–1097. ISSN: 00278424. DOI: 10.1073/pnas.241374598.

Fray, R. G., A. Wallace, P. D. Fraser, D. Valero, Hedden, P. M. Bramley, et al. (1995). "Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway". In: *The Plant Journal* 8.5, pp. 693–701. DOI: https://doi.org/10.1046/j.1365-313X.1995.08050693.x.

Gao, L., I. Gonda, H. Sun, Q. Ma, K. Bao, D. M. Tieman, et al. (2019). "The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor". In: *Nature Genetics* 51.6, pp. 1044–1051. ISSN: 15461718. DOI: `10.1038/s41588-019-0410-2`.

Gehlenborg, N. (2019). *UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets*. R package version 1.4.0. URL: `https://CRAN.R-project.org/package=UpSetR`.

Giuliano, G. (2017). "Provitamin A biofortification of crop plants: a gold rush with many miners". In: *Current Opinion in Biotechnology* 44.Figure 1, pp. 169–180. ISSN: 18790429. DOI: `10.1016/j.copbio.2017.02.001`. URL: `http://dx.doi.org/10.1016/j.copbio.2017.02.001`.

Gómez-Gómez, L., V. Parra-Vega, A. Rivas-Sendra, J. M. Seguí-Simarro, A. Prieto, and O. Ahrazem (2017). "Unraveling Massive Crocins Transport and Accumulation through Proteome and Microscopy Tools during the Development of Saffron Stigma". In: *International Journal of Molecular Sciences* 18.76, pp. 1–22. DOI: `10.3390/ijms18010076`.

González-Gordo, S., R. Bautista, M. G. Claros, A. Cañas, J. M. Palma, and F. J. Corpas (2019). "Nitric oxide-dependent regulation of sweet pepper fruit ripening". In: *Journal of Experimental Botany* 70.17, pp. 4557–4570. ISSN: 14602431. DOI: `10.1093/jxb/erz136`.

Gonzalez-Jorge, S., S.-H. Ha, M. Magallanes-Lundback, L. U. Gilliland, A. Zhou, A. E. Lipka, et al. (2013). "CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds". In: *Plant Cell* 25, pp. 4812–4826. DOI: `10.1105/tpc.113.119677`.

Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, et al. (2012). "Phytozome: A comparative platform for green plant genomics". In: *Nucleic Acids Research* 40.D1, pp. 1178–1186. ISSN: 03051048. DOI: `10.1093/nar/gkr944`.

Grilli-Caiola, M. and A. Canini (2004). "Ultrastructure of chromoplasts and other plastids in Crocus sativus L. (Iridaceae)". In: *Plant Biosystems* 138.1, pp. 43–52. ISSN: 11263504. DOI: `10.1080/11263500410001684116`.

Gruszecki, W. I. (1999). "Carotenoids in Membranes". In: *The Photochemistry of Carotenoids*. Ed. by H. A. Frank, A. J. Young, G. Britton, and R. J. Cogdell. Chap. 20, pp. 363–379. DOI: `10.1007/0-306-48209-6_20`.

Hansmann, P., R. Junker, H. Sauter, and P. Sitte (1987). "Chromoplast Development in Daffodil Coronae during Anthesis". In: *Journal of Plant Physiology* 131, pp. 133–143. DOI: `10.1016/S0176-1617(87)80274-2`.

Harris, W. M. and A. R. Spurr (1969). "Chromoplasts of Tomato Fruits. Ii. the Red Tomato". In: *American Journal of Botany* 56.4, pp. 380–389. ISSN: 0002-9122. DOI: `10.1002/j.1537-2197.1969.tb07547.x`.

Hashimoto, H., C. Uragami, and R. J. Cogdell (2016). "Carotenoids and Photosynthesis". In: *Carotenoids in Nature*. Ed. by C. Stange. Chap. 4. ISBN: 9783319391267. DOI: `10.1007/978-3-319-39126-7`.

Havaux, M. (2014). "Carotenoid oxidation products as stress signals in plants". In: *Plant Journal* 79.4, pp. 597–606. ISSN: 1365313X. DOI: 10.1111/tpj.12386.

Havaux, M. (2020). "Plastoquinone In and Beyond Photosynthesis". In: *Trends in Plant Science* 25.12, pp. 1252–1265. ISSN: 13601385. DOI: 10.1016/j.tplants.2020.06.011. URL: https://doi.org/10.1016/j.tplants.2020.06.011.

Hedden, P. (2020). "The current status of research on gibberellin biosynthesis". In: *Plant and Cell Physiology* 61.11, pp. 1832–1849. ISSN: 14719053. DOI: 10.1093/pcp/pcaa092.

Hickman, R., M. C. Van Verk, A. J. Van Dijken, M. P. Mendes, I. A. Vroegop-Vos, L. Caarls, et al. (2017). "Architecture and dynamics of the jasmonic acid gene regulatory network". In: *Plant Cell* 29.9, pp. 2086–2105. ISSN: 1532298X. DOI: 10.1105/tpc.16.00958.

Hormaetxe, K., J. M. Becerril, I. Fleck, M. Pintó, and J. I. García-Plazaola (2005). "Functional role of red (retro)-carotenoids as passive light filters in the leaves of Buxus sempervirens L.: Increased protection of photosynthetic tissues?" In: *Journal of Experimental Botany* 56.420, pp. 2629–2636. ISSN: 00220957. DOI: 10.1093/jxb/eri255.

Hormaetxe, K., A. Hernández, J. M. Becerril, and J. I. García-Plazaola (2004). "Role of red carotenoids in photoprotection during winter acclimation in Buxus sempervirens leaves". In: *Plant Biology* 6.3, pp. 325–332. ISSN: 14358603. DOI: 10.1055/s-2004-817883.

Hou, X., J. Rivers, P. León, R. P. McQuinn, and B. J. Pogson (2016). "Synthesis and Function of Apocarotenoid Signals in Plants". In: *Trends in Plant Science* 21.9, pp. 792–803. ISSN: 13601385. DOI: 10.1016/j.tplants.2016.06.001. URL: http://dx.doi.org/10.1016/j.tplants.2016.06.001.

Howitt, C. A. and B. J. Pogson (2006). "Carotenoid accumulation and function in seeds and non-green tissues". In: *Plant, Cell and Environment* 29.3, pp. 435–445. ISSN: 01407791. DOI: 10.1111/j.1365-3040.2005.01492.x.

Huang, H., B. Liu, L. Liu, and S. Song (2017). "Jasmonate action in plant growth and development". In: *Journal of Experimental Botany* 68.6, pp. 1349–1359. ISSN: 14602431. DOI: 10.1093/jxb/erw495.

Jacob, P., H. Hirt, and A. Bendahmane (2017). "The heat-shock protein/chaperone network and multiple stress resistance". In: *Plant Biotechnology Journal* 15.4, pp. 405–414. ISSN: 14677652. DOI: 10.1111/pbi.12659.

Jeffery, J., A. Holzenburg, and S. King (2012). "Physical barriers to carotenoid bioaccessibility . Ultrastructure survey of chromoplast and cell wall morphology in nine carotenoid-containing fruits and vegetables". In: *Journal of the Science of Food and Agriculture* May 2011, pp. 2594–2602. DOI: 10.1002/jsfa.5767.

Joyard, J., M. Ferro, C. Masselon, D. Seigneurin-Berny, D. Salvi, J. Garin, et al. (2009). "Chloroplast proteomics and the compartmentation of plastidial isoprenoid biosynthetic pathways". In: *Molecular Plant* 2.6, pp. 1154–1180. ISSN: 17529867. DOI: 10.1093/mp/ssp088.

Kanehisa, M. (2019). "Toward understanding the origin and evolution of cellular organisms". In: *Protein Science* 28.11, pp. 1947–1951. ISSN: 1469896X. DOI: 10.1002/pro.3715.

Kanehisa, M. and S. Goto (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1, pp. 27–30. DOI: 10.1093/nar/28.1.27.

Kasahara, H., K. Takei, N. Ueda, S. Hishiyama, T. Yamaya, Y. Kamiya, et al. (2004). "Distinct Isoprenoid Origins of cis- and trans-Zeatin Biosyntheses in Arabidopsis". In: *Journal of Biological Chemistry* 279.14, pp. 14049–14054. ISSN: 00219258. DOI: 10.1074/jbc.M314195200. URL: http://dx.doi.org/10.1074/jbc.M314195200.

Kato, S., S. Takaichi, T. Ishikawa, M. Asahina, S. Takahashi, and T. Shinomura (2016). "Identification and functional analysis of the geranylgeranyl pyrophosphate synthase gene (crtE) and phytoene synthase gene (crtB) for carotenoid biosynthesis in Euglena gracilis". In: *BMC Plant Biology* 16.1, pp. 1–12. ISSN: 14712229. DOI: 10.1186/s12870-015-0698-8. URL: http://dx.doi.org/10.1186/s12870-015-0698-8.

Kersey, P. J. (2019). "Plant genome sequences: past, present, future". In: *Current Opinion in Plant Biology* 48, pp. 1–8. ISSN: 13695266. DOI: 10.1016/j.pbi.2018.11.001. URL: https://doi.org/10.1016/j.pbi.2018.11.001.

Kersey, P. J., J. E. Allen, A. Allot, M. Barba, S. Boddu, B. J. Bolt, et al. (2018). "Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species". In: *Nucleic Acids Research* 46.D1, pp. D802–D808. ISSN: 13624962. DOI: 10.1093/nar/gkx1011.

Kilambi, H. V., R. Kumar, R. Sharma, and Y. Sreelakshmi (2013). "Chromoplast-specific carotenoid-associated protein appears to be important for enhanced accumulation of carotenoids in hp1 tomato fruits". In: *Plant Physiology* 161.4, pp. 2085–2101. ISSN: 00320889. DOI: 10.1104/pp.112.212191.

Kim, J. E., K. H. Rensing, C. J. Douglas, and K. M. Cheng (2010). "Chromoplasts ultrastructure and estimated carotene content in root secondary phloem of different carrot varieties". In: *Planta* 231.3, pp. 549–558. ISSN: 00320935. DOI: 10.1007/s00425-009-1071-7.

Klie, S., S. Osorio, T. Tohge, M. F. Drincovich, A. Fait, J. J. Giovannoni, et al. (2014). "Conserved changes in the dynamics of metabolic processes during fruit development and ripening across species". In: *Plant Physiology* 164.1, pp. 55–68. ISSN: 15322548. DOI: 10.1104/pp.113.226142.

Koiwa, H., T. Ikeda, and Y. Yoshida (1986). "Reversal of Chromoplasts to Chloroplasats in Buxus Leaves". In: *Botanical Magazine, Tokyo* 99, pp. 233–240.

Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. R package version 1.0.12. URL: `https://CRAN.R-project.org/package=pheatmap`.

Koschmieder, J., F. Wüst, P. Schaub, D. Álvarez, D. Trautmann, M. Krischke, et al. (2021). "Plant apocarotenoid metabolism utilizes defense mechanisms against reactive carbonyl species and xenobiotics". In: *Plant Physiology* 185.2, pp. 331–351. ISSN: 0032-0889. DOI: `10.1093/PLPHYS/KIAA033`. URL: `https://academic.oup.com/plphys/article/185/2/331/6015243`.

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, et al. (2012). "The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools". In: *Nucleic Acids Research* 40.D1, pp. 1202–1210. ISSN: 03051048. DOI: `10.1093/nar/gkr1090`.

Langfelder, P. and S. Horvath (2008). "WGCNA: An R package for weighted correlation network analysis". In: *BMC Bioinformatics* 9. ISSN: 14712105. DOI: `10.1186/1471-2105-9-559`.

Lätari, K., F. Wüst, M. Hübner, P. Schaub, K. G. Beisel, S. Matsubara, et al. (2015). "Tissue-specific apocarotenoid glycosylation contributes to carotenoid homeostasis in arabidopsis leaves". In: *Plant Physiology* 168.4, pp. 1550–1562. ISSN: 15322548. DOI: `10.1104/pp.15.00243`.

Leitner-Dagan, Y., M. Ovadis, E. Shklarman, Y. Elad, D. R. David, and A. Vainstein (2006a). "Expression and functional analyses of the plastid lipid-associated protein CHRC suggest its role in chromoplastogenesis and stress". In: *Plant Physiology* 142.1, pp. 233–244. ISSN: 00320889. DOI: `10.1104/pp.106.082404`.

Leitner-Dagan, Y., M. Ovadis, A. Zuker, E. Shklarman, I. Ohad, T. Tzfira, et al. (2006b). "CHRD, a plant member of the evolutionarily conserved YjgF family, influences photosynthesis and chromoplastogenesis". In: *Planta* 225.1, pp. 89–102. ISSN: 00320935. DOI: `10.1007/s00425-006-0332-y`.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington, K. A. Crandall, et al. (2018). "Earth BioGenome Project: Sequencing life for the future of life". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.17, pp. 4325–4333. ISSN: 10916490. DOI: `10.1073/pnas.1720115115`.

Li, J., J. Yang, B. Zhu, and G. Xie (2019). "Overexpressing OsFBN1 enhances plastoglobule formation, reduces grain-filling percent and jasmonate levels under heat stress in rice". In: *Plant Science* 285.January, pp. 230–238. ISSN: 18732259. DOI: `10.1016/j.plantsci.2019.05.007`.

Li, L., D. J. Paolillo, M. V. Parthasarathy, E. M. DiMuzio, and D. F. Garvin (2001). "A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (Brassica oleracea var. botrytis)". In: *Plant Journal* 26.1, pp. 59–67. ISSN: 09607412. DOI: `10.1046/j.1365-313X.2001.01008.x`.

Li, L. and H. Yuan (2013). "Chromoplast biogenesis and carotenoid accumulation". In: *Archives of Biochemistry and Biophysics* 539.2, pp. 102–109. ISSN: 00039861. DOI: `10.1016/j.abb.2013.07.002`.

Li, L., H. Yuan, Y. Zeng, and Q. Xu (2016). "Plastids and Carotenoid Accumulation". In: *Carotenoids in nature: biosynthesis, regulation, and function*. Ed. by C. Stange. Vol. 79. Chap. 10, pp. 377–414. ISBN: 978-3-319-39124-3. DOI: 10.1007/978-3-319-39126-7_10. URL: http://link.springer.com/10.1007/978-3-319-39126-7.

Ling, Q., N. M. Sadali, Z. Soufi, Y. Zhou, B. Huang, Y. Zeng, et al. (2021). "The chloroplast-associated protein degradation pathway controls chromoplast development and fruit ripening in tomato". In: *Nature Plants* 7.May. DOI: 10.1038/s41477-021-00916-y.

Liu, B., K. Seong, S. Pang, J. Song, H. Gao, C. Wang, et al. (2021). "Functional specificity, diversity, and redundancy of Arabidopsis JAZ family repressors in jasmonate and COI1-regulated growth, development, and defense". In: *New Phytologist* 231.4, pp. 1525–1545. ISSN: 14698137. DOI: 10.1111/nph.17477.

Ljubesic, N., M Wrischer, and Z Devidé (1996). "Chromoplast structures in Thunbergia flowers". In: *Protoplasma* 193, pp. 174–180.

Ljubesic, N., M. Wrischer, and Z. Devide (1991). "Chromoplasts - the last stages in plastid development". In: *International Journal of Developmental Biology* 35, pp. 251–258.

Llamas, E., P. Pulido, and M. Rodriguez-Concepcion (2017). *Interference with plastome gene expression and Clp protease activity in Arabidopsis triggers a chloroplast unfolded protein response to restore protein homeostasis*. Vol. 13. 9. ISBN: 1111111111. DOI: 10.1371/journal.pgen.1007022.

Llorente, B., J. F. Martinez-Garcia, C. Stange, and M. Rodriguez-Concepcion (2017). "Illuminating colors: regulation of carotenoid biosynthesis and accumulation by light". In: *Current Opinion in Plant Biology* 37, pp. 49–55. ISSN: 13695266. DOI: 10.1016/j.pbi.2017.03.011. URL: http://dx.doi.org/10.1016/j.pbi.2017.03.011.

Llorente, B., S. Torres-Montilla, L. Morelli, I. Florez-Sarasa, J. T. Matus, M. Ezquerro, et al. (2020). "Synthetic conversion of leaf chloroplasts into carotenoid-rich plastids reveals mechanistic basis of natural chromoplast development". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.35, pp. 21796–21803. ISSN: 10916490. DOI: 10.1073/pnas.2004405117.

Lohmann, A., M. A. Scho, C. Bréhélin, F. Kessler, R. Bock, and E. B. Cahoon (2006). "Deficiency in Phylloquinone (Vitamin K1) Methylation Affects Prenyl Quinone Distribution , Photosystem I Abundance , and Anthocyanin Accumulation in the Arabidopsis AtmenG Mutant". In: *Journal of Biological Chemistry* 281.52, pp. 40461–40472. DOI: 10.1074/jbc.M609412200.

Lois, L. M., M. Rodríguez-Concepción, F. Gallego, N. Campos, and A. Boronat (2000). "Carotenoid biosynthesis during tomato fruit development: Regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase". In: *Plant Journal* 22.6, pp. 503–513. ISSN: 09607412. DOI: 10.1046/j.1365-313X.2000.00764.x.

López, A. J., S. Frusciante, E. Niza, O. Ahrazem, Á. Rubio-Moraga, G. Diretto, et al. (2021). "A new glycosyltransferase enzyme from family 91, ugt91p3, is responsible for the final glucosylation step of crocins in saffron (Crocus sativus l.)" In: *International Journal of Molecular Sciences* 22.16, pp. 1–17. ISSN: 14220067. DOI: 10.3390/ijms22168815.

Lopez, A. B., J. Van Eck, B. J. Conlin, D. J. Paolillo, J. O'Neill, and L. Li (2008). "Effect of the cauliflower or transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers". In: *Journal of Experimental Botany* 59.2, pp. 213–223. ISSN: 00220957. DOI: 10.1093/jxb/erm299.

Lopez-Juez, E. and K. A. Pyke (2005). "Plastids unleashed: Their development and their integration in plant development". In: *International Journal of Developmental Biology* 49.5-6, pp. 557–577. ISSN: 02146282. DOI: 10.1387/ijdb.051997el.

Loudya, N., P. Mishra, K. Takahagi, Y. Uehara-Yamaguchi, K. Inoue, L. Bogre, et al. (2021). "Cellular and transcriptomic analyses reveal two-staged chloroplast biogenesis underpinning photosynthesis build-up in the wheat leaf". In: *Genome Biology* 22.1, pp. 1–30. ISSN: 1474760X. DOI: 10.1186/s13059-021-02366-3.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12, pp. 1–21. ISSN: 1474760X. DOI: 10.1186/s13059-014-0550-8.

Lu, S., J. V. Eck, X. Zhou, A. B. Lopez, D. M. O'Halloran, K. M. Cosman, et al. (2006). "The Cauliflower Or Gene Encodes a DnaJ Cysteine-Rich Domain-Containing Protein That Mediates High Levels of b -Carotene Accumulation". In: *Plant Cell* 18.December, pp. 3594–3605. DOI: 10.1105/tpc.106.046417.

Maass, D., J. Arango, F. Wüst, P. Beyer, and R. Welsch (2009). "Carotenoid Crystal Formation in Arabidopsis and Carrot Roots Caused by Increased Phytoene Synthase Protein Levels". In: *PLoS ONE* 4.7. DOI: 10.1371/journal.pone.0006373.

Majer, E., B. Llorente, M. Rodríguez-Concepción, and J.-A. Daròs (2017). "Rewiring carotenoid biosynthesis in plants using a viral vector". In: *Scientific Reports* January, pp. 1–10. DOI: 10.1038/srep41645. URL: http://dx.doi.org/10.1038/srep41645.

Martí, M., G. Diretto, V. Aragonés, S. Frusciante, O. Ahrazem, L. Gómez-Gómez, et al. (2020). "Efficient production of saffron crocins and picrocrocin in Nicotiana benthamiana using a virus-driven system". In: *Metabolic Engineering* 61.December 2019, pp. 238–250. ISSN: 10967184. DOI: 10.1016/j.ymben.2020.06.009.

Mehrshahi, P., C. Johnny, and D. DellaPenna (2014). "Redefining the metabolic continuity of chloroplasts and ER". In: *Trends in Plant Science* 19.8, pp. 501–507. ISSN: 13601385. DOI: 10.1016/j.tplants.2014.02.013. URL: http://dx.doi.org/10.1016/j.tplants.2014.02.013.

Mercadante, A. Z., D. B. Rodrigues, F. C. Petry, and L. R. B. Mariutti (2017). "Carotenoid esters in foods - A review and practical directions on analysis and occurrence". In: *Food Research International* 99, pp. 830–850. ISSN: 0963-9969. DOI:

10.1016/j.foodres.2016.12.018. URL: http://dx.doi.org/10.1016/j.foodres.2016.12.018.

Misawa, N., M. Nakagawa, K. Kobayashi, S. Yamano, Y. Izawa, K. Nakamura, et al. (1990). "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functional analysis of gene products expressed in Escherichia coli". In: *Journal of Bacteriology* 172.12, pp. 6704–6712. ISSN: 00219193. DOI: 10.1128/jb.172.12.6704-6712.1990.

Mohamed, H. E., A. M. Van De Meene, R. W. Roberson, and W. F. Vermaas (2005). "Myxoxanthophyll is required for normal cell wall structure and thylakoid organization in the cyanobacterium Synechocystis sp. strain PCC 6803". In: *Journal of Bacteriology* 187.20, pp. 6883–6892. ISSN: 00219193. DOI: 10.1128/JB.187.20.6883-6892.2005.

Morelli, L. (2021). "Exploring an artificial chromoplast system for enrichment of plant leaves in isoprenoid vitamins". PhD thesis.

Moreno, J. C., J. Mi, Y. Alagoz, and S. Al-Babili (2020). "Plant apocarotenoids : from retrograde signaling to interspecific communication". In: *The Plant Journal*, pp. 1–25. DOI: 10.1111/tpj.15102.

Mulisch, M. and K. Krupinska (2013). "Ultrastructural Analysese of Senescence Associated Dismantling of Chloroplasts Revisited". In: *Plastid Development in Leaves During Growth and Senescence*. Ed. by B. Briswal, K. Krupinska, and U. C. Biswal. Chap. 14, pp. 307–335. ISBN: 9789400757240. DOI: 10.1007/978-94-007-5724-0.

Neta-Sharir, I., T. Isaacson, S. Lurie, and D. Weiss (2005). "Dual role for tomato heat shock protein 21: Protecting photosystem II from oxidative stress and promoting color changes during fruit maturation". In: *Plant Cell* 17.6, pp. 1829–1838. ISSN: 10404651. DOI: 10.1105/tpc.105.031914.

Nguyen, L. T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh (2015). "IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies". In: *Molecular Biology and Evolution* 32.1, pp. 268–274. ISSN: 15371719. DOI: 10.1093/molbev/msu300.

Nikolov, L. A., A. Runions, M. Das Gupta, and M. Tsiantis (2019). *Leaf development and evolution*. 1st ed. Vol. 131. Elsevier Inc., pp. 109–139. ISBN: 9780128098042. DOI: 10.1016/bs.ctdb.2018.11.006. URL: http://dx.doi.org/10.1016/bs.ctdb.2018.11.006.

Nisar, N., L. Li, S. Lu, N. C. Khin, and B. J. Pogson (2015). "Carotenoid metabolism in plants". In: *Molecular Plant* 8.1, pp. 68–82. ISSN: 17529867. DOI: 10.1016/j.molp.2014.12.007.

Nogueira, M., L. Mora, E. M. Enfissi, P. M. Bramley, and P. D. Fraser (2013). "Subchromoplast sequestration of carotenoids affects regulatory mechanisms in tomato lines expressing different carotenoid gene combinations". In: *Plant Cell* 25.11, pp. 4560–4579. ISSN: 10404651. DOI: 10.1105/tpc.113.116210.

Osorio, S., R. Alba, Z. Nikoloski, A. Kochevenko, A. R. Fernie, and J. J. Giovannoni (2012). "Integrative comparative analyses of transcript and metabolite profiles from pepper and tomato ripening and development stages uncovers species-specific patterns of network regulatory behavior". In: *Plant Physiology* 159.4, pp. 1713–1729. ISSN: 15322548. DOI: 10.1104/pp.112.199711.

Paolillo, D. J., D. F. Garvin, and M. V. Parthasarathy (2004). "The chromoplasts of Or mutants of cauliflower (Brassica oleracea L. var. botrytis)". In: *Protoplasma* 224, pp. 245–253. DOI: 10.1007/s00709-004-0059-1.

Paran, I. and E. Van Der Knaap (2007). "Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper". In: *Journal of Experimental Botany* 58.14, pp. 3841–3852. ISSN: 00220957. DOI: 10.1093/jxb/erm257.

Park, H., S. S. Kreunen, A. J. Cuttriss, D. DellaPenna, and B. J. Pogson (2002). "Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis". In: *Plant Cell* 14.2, pp. 321–332. ISSN: 10404651. DOI: 10.1105/tpc.010302.

Park, S., H. S. Kim, Y. J. Jung, S. H. Kim, C. Y. Ji, Z. Wang, et al. (2016). "Orange protein has a role in phytoene synthase stabilization in sweetpotato". In: *Scientific Reports*, pp. 1–12. DOI: 10.1038/srep33563.

Park, S. C., S. H. Kim, S. Park, H. U. Lee, J. S. Lee, W. S. Park, et al. (2015). "Enhanced accumulation of carotenoids in sweetpotato plants overexpressing IbOr-Ins gene in purple-fleshed sweetpotato cultivar". In: *Plant Physiology and Biochemistry* 86, pp. 82–90. ISSN: 09819428. DOI: 10.1016/j.plaphy.2014.11.017. URL: http://dx.doi.org/10.1016/j.plaphy.2014.11.017.

Paul, M. J. and T. K. Pellny (2003). "Carbon metabolite feedback regulation of leaf photosynthesis and development". In: *Journal of Experimental Botany* 54.382, pp. 539–547. ISSN: 00220957. DOI: 10.1093/jxb/erg052.

Phillips, M. A., P. León, A. Boronat, and M. Rodríguez-Concepción (2008). "The plastidial MEP pathway: unified nomenclature and resources". In: *Trends in Plant Science* 13.12, pp. 619–623. ISSN: 13601385. DOI: 10.1016/j.tplants.2008.09.003.

Pokhilko, A., J. Bou-torrent, P. Pulido, and M. Rodr (2015). "Mathematical modelling of the diurnal regulation of the MEP pathway in Arabidopsis". In: *New Phytologist* 206, pp. 1075–1085. DOI: 10.1111/nph.13258.

Popova, A. V. and A. S. Andreeva (2013). "Carotenoid-Lipid Interactions". In: *Advances in Planar Lipid Bilayers and Liposomes*. 1st ed. Vol. 17. Copyright © 2013 Elsevier Inc. All rights reserved. Chap. Eight, pp. 215–236. ISBN: 9780124115163. DOI: 10.1016/B978-0-12-411516-3.00008-5. URL: http://dx.doi.org/10.1016/B978-0-12-411516-3.00008-5.

Pulido, P., E. Llamas, B. Llorente, S. Ventura, L. P. Wright, and M. Rodríguez-Concepción (2016). "Specific Hsp100 Chaperones Determine the Fate of the First Enzyme of the Plastidial Isoprenoid Pathway for Either Refolding or Degradation by the Stromal Clp Protease in Arabidopsis". In: *PLoS Genetics* 12.1. ISSN: 15537404.

DOI: `10.1371/journal.pgen.1005824`. URL: `http://dx.doi.org/10.1371/journal.pgen.1005824`.

Qin, G., H. Gu, L. Ma, Y. Peng, X. W. Deng, Z. Chen, et al. (2007). "Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll , carotenoid , and gibberellin biosynthesis". In: *Cell Research* 17, pp. 471–482. DOI: `10.1038/cr.2007.40`.

Quian-Ulloa, R. and C. Stange (2021). "Carotenoid Biosynthesis and Plastid Development in Plants : The Role of Light". In: *International Journal of Molecular Sciences* 22.1184. DOI: `10.3390/ijms22031184`.

Quinet, M., T. Angosto, F. J. Yuste-Lisbona, R. Blanchard-Gros, S. Bigot, J. P. Martinez, et al. (2019). "Tomato Fruit Development and Metabolism". In: *Frontiers in Plant Science* 10.November, pp. 1–23. ISSN: 1664462X. DOI: `10.3389/fpls.2019.01554`.

Ramel, F., S. Birtic, C. Ginies, L. Soubigou-taconnat, and C. Triantaphylidès (2012). "Carotenoid oxidation products are stress signals that mediate gene responses to singlet oxygen in plants". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.14. DOI: `10.1073/pnas.1115982109`.

Ravanello, M. P., D. Ke, J. Alvarez, B. Huang, and C. K. Shewmaker (2003). "Coordinate expression of multiple bacterial carotenoid genes in canola leading to altered carotenoid production". In: *Metabolic Engineering* 5.4, pp. 255–263. ISSN: 10967176. DOI: `10.1016/j.ymben.2003.08.001`.

Rey, P., B. Gillet, S. Römer, F. Eymery, J. Massimino, G. Peltier, et al. (2000). "Overexpression of a pepper plastid lipid-associated protein in tobacco leads to changes in plastid ultrastructure and plant development upon stress". In: *Plant Journal* 21.5, pp. 483–494. ISSN: 09607412. DOI: `10.1046/j.1365-313X.2000.00699.x`.

Rödiger, A., B. Agne, D. Dobritzsch, S. Helm, F. Müller, N. Pötzsch, et al. (2020). "Chromoplast differentiation in bell pepper (Capsicum annuum) fruits". In: *Plant Journal*, pp. 1–12. DOI: `10.1111/tpj.15104`.

Rodriguez-Concepcion, M., J. Avalos, M. L. Bonet, A. Boronat, L. Gomez-Gomez, D. Hornero-Mendez, et al. (2018). "A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health". In: *Progress in Lipid Research* 70.February, pp. 62–93. ISSN: 18732194. DOI: `10.1016/j.plipres.2018.04.004`.

Rodriguez-Concepcion, M., L. D'Andrea, and P. Pulido (2019). "Control of plastidial metabolism by the Clp protease complex". In: *Journal of Experimental Botany* 70.7, pp. 2049–2058. ISSN: 14602431. DOI: `10.1093/jxb/ery441`.

Rodríguez-Villalón, A., E. Gas, and M. Rodríguez-Concepción (2009). "Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings". In: *Plant Journal* 60.3, pp. 424–435. ISSN: 09607412. DOI: `10.1111/j.1365-313X.2009.03966.x`.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. URL: `http://www.rstudio.com/`.

Ruan, J., Y. Zhou, M. Zhou, J. Yan, M. Khurshid, W. Weng, et al. (2019). "Jasmonic Acid Signaling Pathway in Plants". In: *International Journal of Molecular Sciences* 20.

Ruban, A. V., M. P. Johnson, and C. D. P. Duffy (2012). "The photoprotective molecular switch in the photosystem II antenna". In: *Biochimica et Biophysica Acta* 1817.1, pp. 167–181. ISSN: 0005-2728. DOI: 10.1016/j.bbabio.2011.04.007. URL: http://dx.doi.org/10.1016/j.bbabio.2011.04.007.

Ruiz-Sola, M. Á., M. V. Barja, D. Manzano, B. Llorente, B. Schipper, J. Beekwilder, et al. (2016a). "A Single Arabidopsis Gene Encodes Two Differentially Targeted Geranylgeranyl Diphosphate Synthase Isoforms". In: *Plant Physiology* 172, pp. 1393–1402. DOI: 10.1104/pp.16.01392.

Ruiz-Sola, M. Á., D. Coman, G. Beck, M. V. Barja, M. Colinas, A. Graf, et al. (2016b). "Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids". In: *New Phytologist* 209, pp. 252–264. DOI: 10.1111/nph.13580.

Ruiz-Sola, M. Á. and M. Rodríguez-Concepción (2012). "Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway". In: *The Arabidopsis Book* 10, e0158. ISSN: 1543-8120. DOI: 10.1199/tab.0158. URL: http://www.bioone.org/doi/abs/10.1199/tab.0158.

Sadali, N. M., R. G. Sowden, Q. Ling, and R. P. Jarvis (2019). "Differentiation of chromoplasts and other plastids in plants". In: *Plant Cell Reports* 38.7, pp. 803–818. ISSN: 1432203X. DOI: 10.1007/s00299-019-02420-2. URL: https://doi.org/10.1007/s00299-019-02420-2.

Sakakibara, H. (2006). "Cytokinins: Activity, biosynthesis, and translocation". In: *Annual Review of Plant Biology* 57, pp. 431–449. ISSN: 15435008. DOI: 10.1146/annurev.arplant.57.032905.105231.

Sato, S., S. Tabata, H. Hirakawa, E. Asamizu, K. Shirasawa, S. Isobe, et al. (2012). "The tomato genome sequence provides insights into fleshy fruit evolution". In: *Nature* 485.7400, pp. 635–641. ISSN: 14764687. DOI: 10.1038/nature11119.

Schaub, P., M. Rodriguez-Franco, C. I. Cazzonelli, D. Alvarez, F. Wüst, and R. Welsch (2018). "Establishment of an Arabidopsis callus system to study the interrelations of biosynthesis, degradation and accumulation of carotenoids". In: *PLoS ONE* 13.2, pp. 1–28. ISSN: 19326203. DOI: 10.1371/journal.pone.0192158.

Schweiggert, R. M., C. B. Steingass, A. Heller, P. Esquivel, and R. Carle (2011). "Characterization of chromoplasts and carotenoids of red- and yellow-fleshed papaya (Carica papaya L.)" In: *Planta* 234, pp. 1031–1044. DOI: 10.1007/s00425-011-1457-1.

Sela, I., H. Ashkenazy, K. Katoh, and T. Pupko (2015). "GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters". In: *Nucleic Acids Research* 43.W1, W7–W14. ISSN: 13624962. DOI: 10.1093/nar/gkv318.

Seo, M., Y. Jikumaru, and Y. Kamiya (2011). "Profiling of Hormones and Related Metabolites in Seed Dormancy and Germination Studies". In: *Seed Dormancy.*

Methods in Molecular Biology (Methods and Protocols). Ed. by A. R. Kermode. Vol. 773. Chap. 7, pp. 99–111. ISBN: 978-1-61779-230-4. DOI: `10.1007/978-1-61779-231-1_7`. URL: `http://link.springer.com/10.1007/978-1-61779-231-1`.

Shinozaki, Y., P. Nicolas, N. Fernandez-Pozo, Q. Ma, D. J. Evanich, Y. Shi, et al. (2018). "High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening". In: *Nature Communications* 9.1. ISSN: 20411723. DOI: `10.1038/s41467-017-02782-9`. URL: `http://dx.doi.org/10.1038/s41467-017-02782-9`.

Shumskaya, M. and E. T. Wurtzel (2013). "The carotenoid biosynthetic pathway: Thinking in all dimensions". In: *Plant Science* 208, pp. 58–63. ISSN: 01689452. DOI: `10.1016/j.plantsci.2013.03.012`. URL: `http://dx.doi.org/10.1016/j.plantsci.2013.03.012`.

Simkin, A. J., J. Breitenbach, M. Kuntz, and G. Sandmann (2000). "In vitro and in situ inhibition of carotenoid biosynthesis in Capsicum annuum by bleaching herbicides". In: *Journal of Agricultural and Food Chemistry* 48.10, pp. 4676–4680. ISSN: 00218561. DOI: `10.1021/jf0006426`.

Simkin, A. J., J. Gaffé, J. P. Alcaraz, J. P. Carde, P. M. Bramley, P. D. Fraser, et al. (2007). "Fibrillin influence on plastid ultrastructure and pigment content in tomato fruit". In: *Phytochemistry* 68.11, pp. 1545–1556. ISSN: 00319422. DOI: `10.1016/j.phytochem.2007.03.014`.

Simkin, A. J., A. M. Labouré, M. Kuntz, and G. Sandmann (2003). "Comparison of carotenoid content, gene expression and enzyme levels in tomato (Lycopersicon esculentum) leaves". In: *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 58.5-6, pp. 371–380. ISSN: 09395075. DOI: `10.1515/znc-2003-5-615`.

Simonyi, M., Z. Bikádi, F. Zsila, and J. Deli (2003). "Supramolecular exciton chirality of carotenoid aggregates". In: *Chirality* 15.8, pp. 680–698. ISSN: 08990042. DOI: `10.1002/chir.10282`.

Singh, D. K. and T. W. McNellis (2011). "Fibrillin protein function: the tip of the iceberg?" In: *Trends in Plant Science* 16.8, pp. 432–441. ISSN: 1360-1385. DOI: `10.1016/j.tplants.2011.03.014`. URL: `http://dx.doi.org/10.1016/j.tplants.2011.03.014`.

Solymosi, K., J. Lethin, and H. Aronsson (2018). "Diversity and Plasticity of Plastids in Land Plants". In: *Plastids*. Ed. by E. Maréchal. Springer Protocols. Chap. 4, pp. 55–72. ISBN: 9781493986545. DOI: `10.1007/978-1-4939-8654-5_4`.

Solymosi, K. and B. Schoefs (2010). "Etioplast and etio-chloroplast formation under natural conditions: The dark side of chlorophyll biosynthesis in angiosperms". In: *Photosynthesis Research* 105.2, pp. 143–166. ISSN: 01668595. DOI: `10.1007/s11120-010-9568-2`.

Stevens, A. and L. Ramirez-Lopez (2020). *An introduction to the prospectr package*. R package version 0.2.1.

Sun, T. and L. Li (2020). "Toward the 'golden' era: The status in uncovering the regulatory control of carotenoid accumulation in plants". In: *Plant Science* 290.November 2019. ISSN: 18732259. DOI: 10.1016/j.plantsci.2019.110331.

Sun, T., H. Yuan, H. Cao, M. Yazdani, Y. Tadmor, and L. Li (2017). "Carotenoid Metabolism in Plants: The Role of Plastids". In: *Molecular Plant*, pp. 1–17. ISSN: 16742052. DOI: 10.1016/j.molp.2017.09.010. URL: http://linkinghub.elsevier.com/retrieve/pii/S1674205217302733.

Sun, T., H. Yuan, C. Chen, D. K. Kadirjan-Kalbach, M. Mazourek, K. W. Osteryoung, et al. (2020). "ORHis, a Natural Variant of OR, Specifically Interacts with Plastid Division Factor ARC3 to Regulate Chromoplast Number and Carotenoid Accumulation". In: *Molecular Plant* 13.6, pp. 864–878. ISSN: 17529867. DOI: 10.1016/j.molp.2020.03.007.

Sundqvist, C. and C. Dahlin (1997). "With chlorophyll pigments from prolamellar bodies to light-harvesting complexes". In: *Physiologia Plantarum* 100.4, pp. 748–759. ISSN: 00319317. DOI: 10.1034/j.1399-3054.1997.1000402.x.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc (2011). "Revigo summarizes and visualizes long lists of gene ontology terms". In: *PLoS ONE* 6.7. ISSN: 19326203. DOI: 10.1371/journal.pone.0021800.

The Arabidopsis Genome Initiative (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana". In: *Nature* 408, pp. 796–815. DOI: 10.1038/35048692.

Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, et al. (2003). "PANTHER: A library of protein families and subfamilies indexed by function". In: *Genome Research* 13.9, pp. 2129–2141. ISSN: 10889051. DOI: 10.1101/gr.772403.

Tian, T., Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, et al. (2017). "AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update". In: *Nucleic Acids Research* 45.W1, W122–W129. ISSN: 13624962. DOI: 10.1093/nar/gkx382.

Torres-Montilla, S. (2017). "Desvelando el mecanismo que inicia la diferenciación de cromoplastos en hojas de Nicotiana benthamiana". PhD thesis.

Torres-Montilla, S. and M. Rodriguez-Concepcion (2021). "Making extra room for carotenoids in plant cells: New opportunities for biofortification". In: *Progress in Lipid Research* 84, p. 101128. ISSN: 01637827. DOI: 10.1016/j.plipres.2021.101128. URL: https://doi.org/10.1016/j.plipres.2021.101128.

Tóth, T. N., V. Chukhutsina, I. Domonkos, J. Knoppová, J. Komenda, M. Kis, et al. (2015). "Carotenoids are essential for the assembly of cyanobacterial photosynthetic complexes". In: *Biochimica et Biophysica Acta - Bioenergetics* 1847.10, pp. 1153–1165. ISSN: 18792650. DOI: 10.1016/j.bbabio.2015.05.020.

Twyford, A. D. (2018). "The road to 10,000 plant genomes". In: *Nature Plants* 4.6, pp. 312–313. ISSN: 20550278. DOI: 10.1038/s41477-018-0165-2. URL: http://dx.doi.org/10.1038/s41477-018-0165-2.

Tzuri, G., X. Zhou, N. Chayut, H. Yuan, V. Portnoy, A. Meir, et al. (2015). "A 'golden' SNP in CmOr governs the fruit flesh color of melon (Cucumis melo)". In: *Plant Journal* 82, pp. 267–279. DOI: 10.1111/tpj.12814.

Van Wijk, K. J. and F. Kessler (2017). "Plastoglobuli: Plastid Microcompartments with Integrated Functions in Metabolism, Plastid Developmental Transitions, and Environmental Adaptation". In: *Annual Review of Plant Biology* 68, pp. 253–289. ISSN: 15452123. DOI: 10.1146/annurev-arplant-043015-111737.

Vidi, P.-A., M. Kanwischer, S. Baginsky, J. R. Austin, G. Csucs, P. Dörmann, et al. (2006). "Tocopherol Cyclase (VTE1) Localization and Vitamin E Accumulation in Chloroplast Plastoglobule Lipoprotein Particles". In: *The Journal of Biological Chemistry* 281.16, pp. 11225–11234. DOI: 10.1074/jbc.M511939200.

Wang, R., M. Lammers, Y. Tikunov, A. G. Bovy, G. C. Angenent, and R. A. de Maagd (2020). "The rin, nor and Cnr spontaneous mutations inhibit tomato fruit ripening in additive and epistatic manners". In: *Plant Science* 294.January, p. 110436. ISSN: 18732259. DOI: 10.1016/j.plantsci.2020.110436. URL: https://doi.org/10.1016/j.plantsci.2020.110436.

Wang, Y.-Q., Y. Yang, Z. Fei, H. Yuan, T. Fish, T. W. Thannhauser, et al. (2013). "Proteomic analysis of chromoplasts from six crop species reveals insights into chromoplast function and development". In: *Journal of Experimental Botany* 64.4, pp. 949–961. ISSN: 00220957. DOI: 10.1093/jxb/ers375.

Wasternack, C. and B. Hause (2013). "Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany". In: *Annals of Botany* 111, pp. 1021–1058. DOI: 10.1093/aob/mct067.

Wasternack, C. and B. Hause (2019). "The missing link in jasmonic acid biosynthesis". In: *Nature Plants* 5.8, pp. 776–777. ISSN: 20550278. DOI: 10.1038/s41477-019-0492-y. URL: http://dx.doi.org/10.1038/s41477-019-0492-y.

Wasternack, C. and S. Song (2017). "Jasmonates: biosynthesis , metabolism, and signaling by proteins activating and repressing transcription". In: *Journal of Experimental Botany* 68.6, pp. 1303–1321. DOI: 10.1093/jxb/erw443.

Watkins, J. L. and B. J. Pogson (2020). "Prospects for Carotenoid Biofortification Targeting Retention and Catabolism". In: *Trends in Plant Science* 25.5, pp. 501–512. ISSN: 13601385. DOI: 10.1016/j.tplants.2019.12.021. URL: https://doi.org/10.1016/j.tplants.2019.12.021.

Welsch, R., X. Zhou, H. Yuan, D. Álvarez, T. Sun, D. Schlossarek, et al. (2018). "Clp Protease and OR Directly Control the Proteostasis of Phytoene Synthase, the Crucial Enzyme for Carotenoid Biosynthesis in Arabidopsis". In: *Molecular Plant* 11.1, pp. 149–162. ISSN: 17529867. DOI: 10.1016/j.molp.2017.11.003.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Wickham, H. (2021). *tidyr: Tidy Messy Data*. R package version 1.1.4. URL: https://CRAN.R-project.org/package=tidyr.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: `10.21105/joss.01686`.

Wickham, H., R. François, L. Henry, and K. Müller (2021). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.7. URL: `https://CRAN.R-project.org/package=dplyr`.

Woo, H. R., H. J. Koo, J. Kim, H. Jeong, J. O. Yang, I. H. Lee, et al. (2016). "Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in arabidopsis". In: *Plant Physiology* 171.1, pp. 452–467. ISSN: 15322548. DOI: `10.1104/pp.15.01929`.

Wurtzel, E. T. (2019). "Changing form and function through carotenoids and synthetic biology". In: *Plant Physiology* 179.3, pp. 830–843. ISSN: 15322548. DOI: `10.1104/pp.18.01122`.

Xin, X., C. Baysal, M. Drapal, Y. Sheng, X. Huang, W. He, et al. (2021). "The Coordinated Upregulated Expression of Genes Involved in MEP, Chlorophyll, Carotenoid and Tocopherol Pathways, Mirrored the Corresponding Metabolite Contents in Rice Leaves during De-Etiolation". In: *Plants (Basel)* 10.1456. DOI: `10.3390/plants10071456`.

Xu, Z. Y., K. H. Lee, T. Dong, J. C. Jeong, J. B. Jin, Y. Kanno, et al. (2012). "A vacuolar β-Glucosidase homolog that possesses glucose-conjugated abscisic acid hydrolyzing activity plays an important role in osmotic stress responses in Arabidopsis". In: *Plant Cell* 24.5, pp. 2184–2199. ISSN: 10404651. DOI: `10.1105/tpc.112.095935`.

Yang, D. L., J. Yao, C. S. Mei, X. H. Tong, L. J. Zeng, Q. Li, et al. (2012). "Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.19. ISSN: 00278424. DOI: `10.1073/pnas.1201616109`.

Yang, Y. X., M. M. Wang, Y. L. Yin, E. Onac, G. F. Zhou, S. Peng, et al. (2015). "RNA-seq analysis reveals the role of red light in resistance against Pseudomonas syringae pv. tomato DC3000 in tomato plants". In: *BMC Genomics* 16.1, pp. 1–16. ISSN: 14712164. DOI: `10.1186/s12864-015-1228-7`.

Yazdani, M., Z. Sun, H. Yuan, S. Zeng, T. W. Thannhauser, J. Vrebalov, et al. (2019). "Ectopic expression of ORANGE promotes carotenoid accumulation and fruit development in tomato". In: *Plant Biotechnology Journal* 17.1, pp. 33–49. ISSN: 14677652. DOI: `10.1111/pbi.12945`.

Ytterberg, A. J., J.-B. Peltier, and K. J. Van Wijk (2006). "Protein Profiling of Plastoglobules in Chloroplasts and Chromoplasts . A Surprising Site for Differential Accumulation of Metabolic Enzymes". In: *Plant Physiology* 140.March 2006, pp. 984–997. DOI: `10.1104/pp.105.076083.984`.

Zbierzak, A. M., M. Kanwischer, C. Wille, P.-A. Vidi, P. Giavalisco, A. Lohmann, et al. (2010). "Intersection of the tocopherol and plastoquinol metabolic pathways at the plastoglobule". In: *Biochemical Journal* 425.2, pp. 389–399. ISSN: 14708728. DOI: `10.1042/BJ20090704`.

Zhang, M., B. Yuan, and P. Leng (2009). "The role of ABA in triggering ethylene biosynthesis and ripening of tomato fruit". In: *Journal of Experimental Botany* 60.6, pp. 1579–1588. ISSN: 00220957. DOI: 10.1093/jxb/erp026.

Zhang, X.-C., Y. A. Millet, Z. Cheng, J. Bush, and F. M. Ausubel (2015). "Jasmonate signalling in Arabidopsis involves SGT1b–HSP70–HSP90 chaperone complexes". In: *Nature Plants* 1.5, pp. 1–18. ISSN: 2055-0278. DOI: 10.1038/nplants.2015.49.

Zhang, Y. and J. G. Turner (2008). "Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis". In: *PLoS ONE* 3.11. ISSN: 19326203. DOI: 10.1371/journal.pone.0003699.

Zheng, X., G. Giuliano, and S. Al-Babili (2020). "Carotenoid biofortification in crop plants: citius, altius, fortius". In: *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1865.11, p. 158664. ISSN: 18792618. DOI: 10.1016/j.bbalip.2020.158664. URL: https://doi.org/10.1016/j.bbalip.2020.158664.

Zheng, Y., C. Jiao, H. Sun, H. G. Rosli, M. A. Pombo, P. Zhang, et al. (2016). "iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases". In: *Molecular Plant* 9.12, pp. 1667–1670. ISSN: 17529867. DOI: 10.1016/j.molp.2016.09.014.

Zhong, S., J. G. Joung, Y. Zheng, Y. R. Chen, B. Liu, Y. Shao, et al. (2011). "High-throughput illumina strand-specific RNA sequencing library preparation". In: *Cold Spring Harbor Protocols* 6.8, pp. 940–949. ISSN: 15596095. DOI: 10.1101/pdb.prot5652.

Zhou, F., C. Y. Wang, M. Gutensohn, L. Jiang, P. Zhang, D. Zhang, et al. (2017). "A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.26, pp. 6866–6871. ISSN: 10916490. DOI: 10.1073/pnas.1705689114.

Zhou, X., R. Welsch, Y. Yang, D. Álvarez, M. Riediger, H. Yuan, et al. (2015). "Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.11, pp. 3558–3563. ISSN: 10916490. DOI: 10.1073/pnas.1420831112.

Supplemental
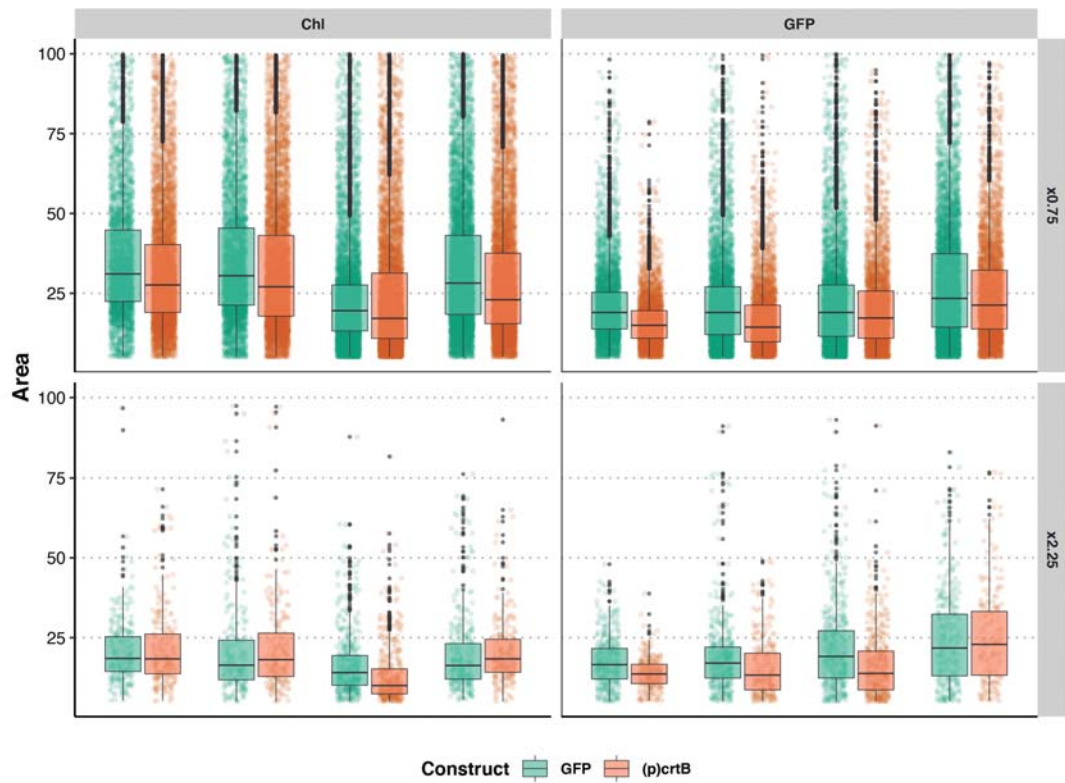
Supplementary Figure 1: (p)crtB causes plastid area decrease. Plastid area measured by chlorophyll and GFP fluorescence at two different magnifications, in plastid localized GFP with and without (p)crtB co-agroinfiltrated.

Table 8: HPLC data 1

Metabolite	(p)crtB_1	(p)crtB_2	(p)crtB_3	1S-crtB_1	1S-crtB_2	1S-crtB_3	crtB_1
Phytoene	0.084504846	0.055690455	0.05516304	0.005906206	0.004296199	0.004042581	0.070208465
Violaxanthin	0.270146711	0.316603557	0.302980241	0.213993446	0.170457507	0.206643919	0.304781101
Neoxanthin	0.125481583	0.136991424	0.111869128	0.123386302	0.100973427	0.126180069	0.116374056
Lutein	1.311620941	1.354641096	1.160106046	0.811358919	0.694817882	0.819750396	1.246294496
TotalBetacarotene	0.948629834	0.895791674	0.786586815	0.396271345	0.323844693	0.375185641	0.779124815
Lycopene	0.005302379	0.005427558	0.004134191	0	0	0	0.005493064
Chlorophyllb	0.391642019	0.401375468	0.328314074	0.373980674	0.294752098	0.376351074	0.352189938
Chlorophylla	0.957168849	0.997233167	0.833106692	0.921408839	0.770790165	0.91916265	0.870378018
Gammatocopherol	0.05742899	0.041408921	0.067842247	0.015057132	0.018940165	0.015564953	0.070514521
Alphatocopherol	0.297519111	0.294732598	0.319592473	0.250960674	0.243995872	0.228116304	0.359597266
Total_Carotenoids	2.745686294	2.765145764	2.420839462	1.550916218	1.294389708	1.531802605	2.522275997
Total_Chlorophylls	1.348810868	1.398608636	1.161420766	1.295389513	1.065542264	1.295137724	1.222567956
Total_Tocopherols	0.354948101	0.336141519	0.38743472	0.266017806	0.262936037	0.243681257	0.430111787
Ratio	2.035634764	1.977068991	2.084377629	1.197258587	1.214770875	1.182390103	2.063096766

Table 9: HPLC data 2

Metabolite	crtB_2	crtB_3	GFP_1	GFP_2	GFP_3	(c)crtB_1	(c)crtB_2	(c)crtB_3
Phytoene	0.046335618	0.076051885	0.002014343	0.002621568	0.003532602	0.004312487	0.002593294	0.003604422
Violaxanthin	0.281062367	0.314504784	0.227330384	0.201171688	0.234683303	0.199700403	0.185342429	0.235217937
Neoxanthin	0.098835365	0.122534159	0.113748813	0.09646818	0.118098941	0.114801789	0.111127842	0.124583404
Lutein	1.110514232	1.310936395	0.693265951	0.614777994	0.752563873	0.865793588	0.842629627	0.933251822
TotalBetacarotene	0.650295085	0.844682575	0.389998633	0.34093117	0.413208159	0.410255386	0.368236603	0.421966659
Lycopene	0.003958764	0.004456007	0	0	0	0	0	0
Chlorophyllb	0.265863185	0.354879494	0.342978425	0.268068563	0.346426957	0.339760514	0.319292741	0.348897767
Chlorophylla	0.706296022	0.907374114	0.870561981	0.709755067	0.890732951	0.845517233	0.804746144	0.904669762
Gammatocopherol	0.128132087	0.061569677	0.015498832	0.016615545	0.015260023	0.016866105	0.01283771	0.014872882
Alphatocopherol	0.34904836	0.362498204	0.253502068	0.261434692	0.23750843	0.191281158	0.166179795	0.224937735
Total_Carotenoids	2.191001431	2.673165804	1.426358124	1.255970599	1.522086877	1.594863655	1.509929795	1.718624243
Total_Chlorophylls	0.972159208	1.262253608	1.213540406	0.97782363	1.237159907	1.185277747	1.124038885	1.253567529
Total_Tocopherols	0.477180447	0.424067881	0.2690009	0.278050237	0.252768453	0.208147263	0.179017505	0.239810617
Ratio	2.253747549	2.117772361	1.175369288	1.284455152	1.23030731	1.345561121	1.343307438	1.370986567
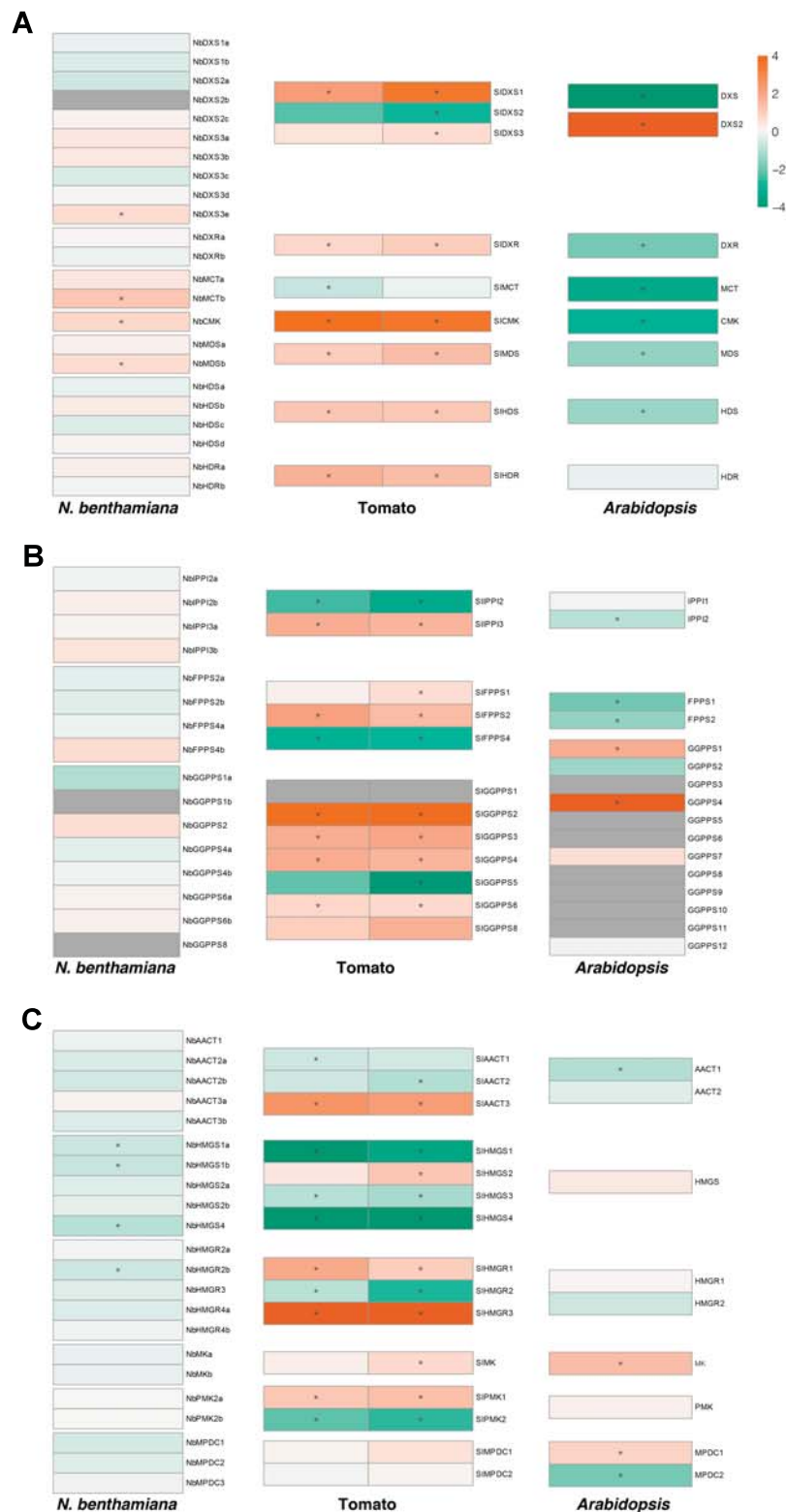
Supplementary Figure 2: Isoprenoid precursors pathways gene expression in *N. ben-thamiana*, tomato and *Arabidopsis*. Gene expression of carotenoid biosynthetic pathways in log$_2$FC, in *N. benthamiana* comparison of (p)crtB vs GFP, tomato fruit ripening comparisons of OR vs MG and RR vs MG and *Arabidopsis* comparison of senescent vs non-senescent leaves.
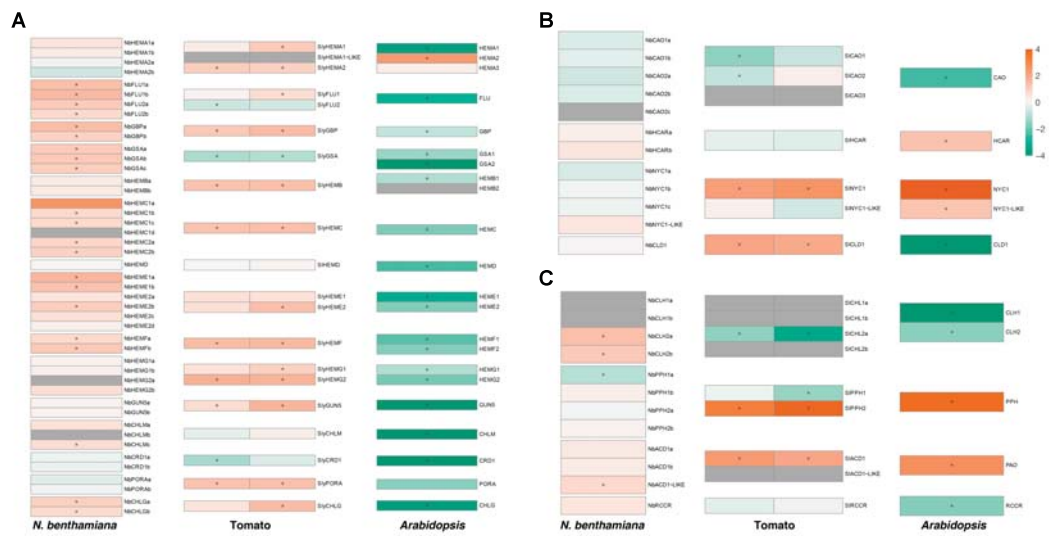
Supplementary Figure 3: Chlorophylls pathways gene expression in *N. benthamiana*, tomato and *Arabidopsis*. Gene expression of chlorophyll biosynthesis (A), cycle (B) and degradation (C) pathways in $\log_2$FC, in *N. benthamiana* comparison of (p)crtB vs GFP, tomato fruit ripening comparisons of OR vs MG and RR vs MG and *Arabidopsis* comparison of senescent vs non-senescent leaves.
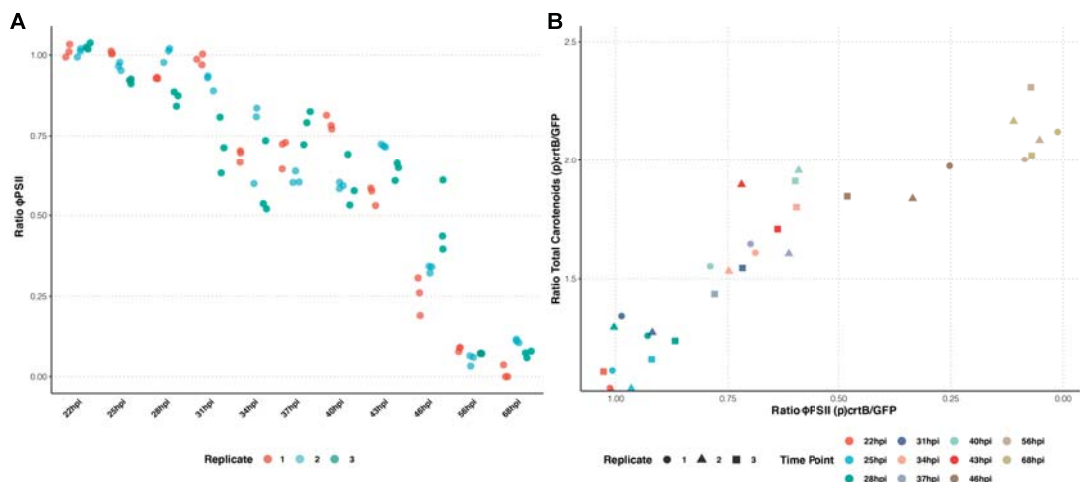
Supplementary Figure 4: ϕPSII and carotenoids vs. photosynthesis during time course. (A) (p)crtB/GFP ratio of ϕPSII values in in leaves of *N. benthamiana* in all replicates. (B) (p)crtB/GFP ratio of total carotenoid levels vs. (p)crtB/GFP ratio of ϕPSII values in all samples.
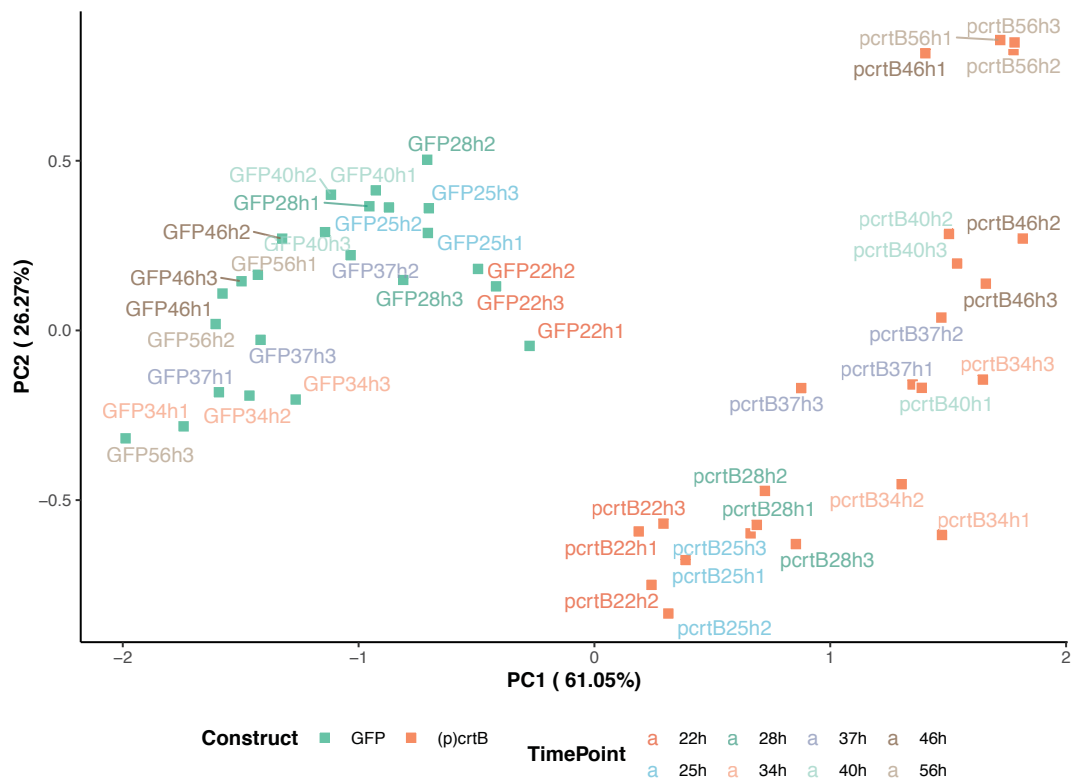
Supplementary Figure 5: PCA plot of transcriptomic and phenotypic data from time course. Distribution of GFP and (p)crtB samples in a PCA, including transcriptomic and phenotypic data from HPCL and ϕPSII after block normalization.
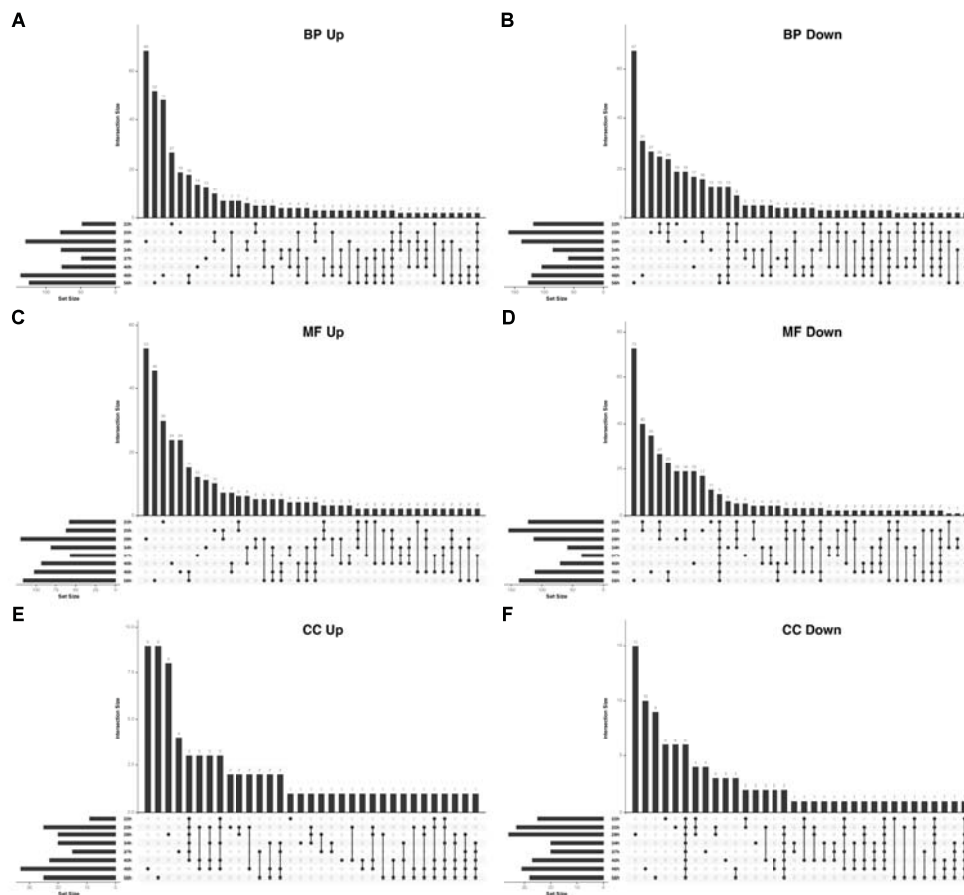
Supplementary Figure 6: UpSet diagram of GOs in time course. Distribution of up- (A) and down-regulated (B) BPs, MFs (C-D) and CCs (E-F) of the different time points. Groups were distributed according to their abundance (from left to right). Number of GO terms in each group are represented in bar plots on the top; time points including in each group are highlighted below; number of GO terms in each time points are represented in horizontal bar plots on the left.

A

B
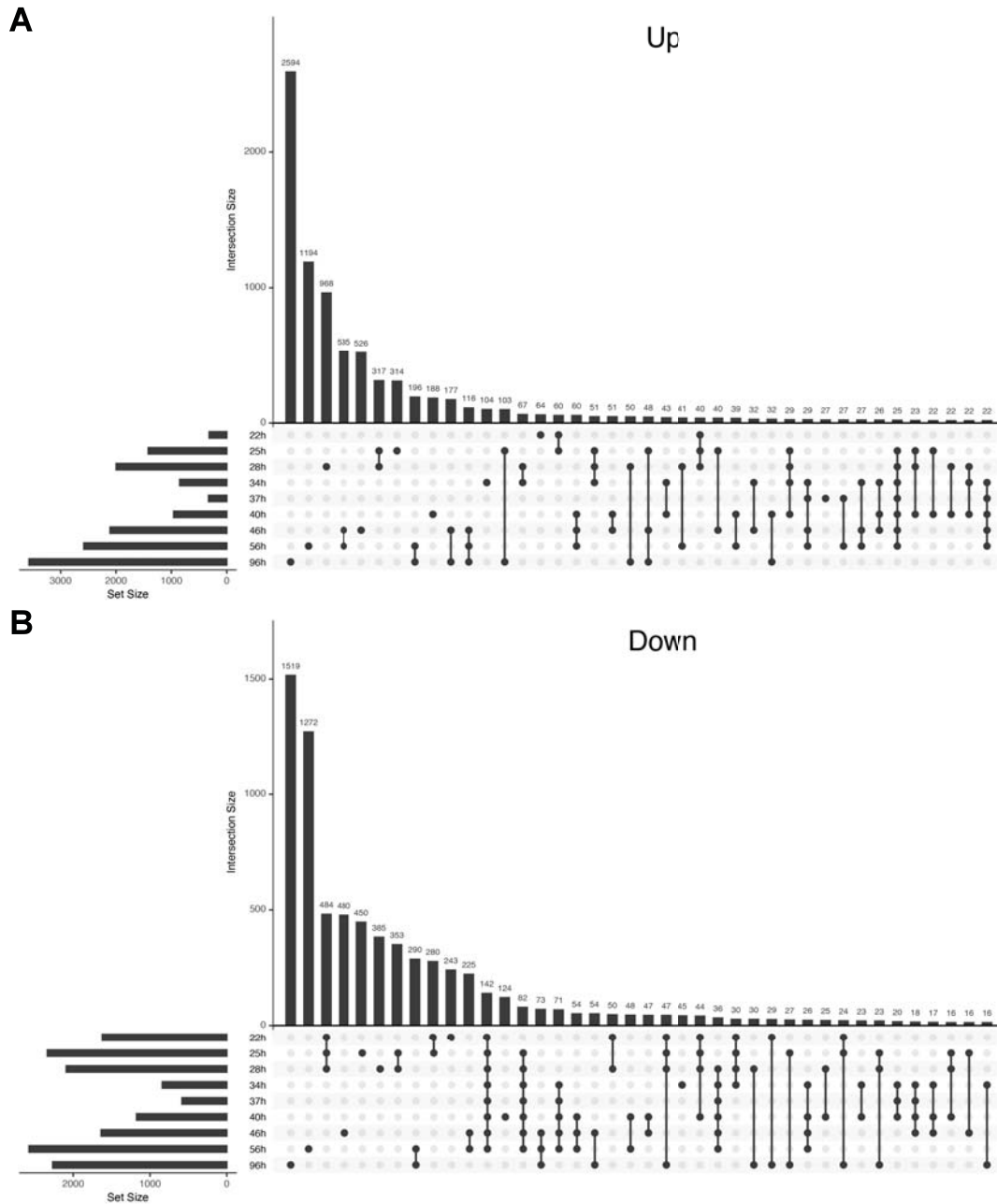
Supplementary Figure 7: UpSet diagram of DEGs in time course and end-point. Distribution of up- (A) and down-regulated (B) DEGs of the different time points, including endpoint at 96 hpi. Groups were distributed according to their abundance (from left to right). Number of DEGs in each group are represented in bar plots on the top; time points including in each group are highlighted below; number of DEGs in each time points are represented in horizontal bar plots on the left.

Supplementary Figure 8: Up-regulated "Spread" BPs. Distribution of all up-regulated BPs included "Spread" category. GO terms are divided in 3 different groups: Early (green), late (orange) and disperse (blue).
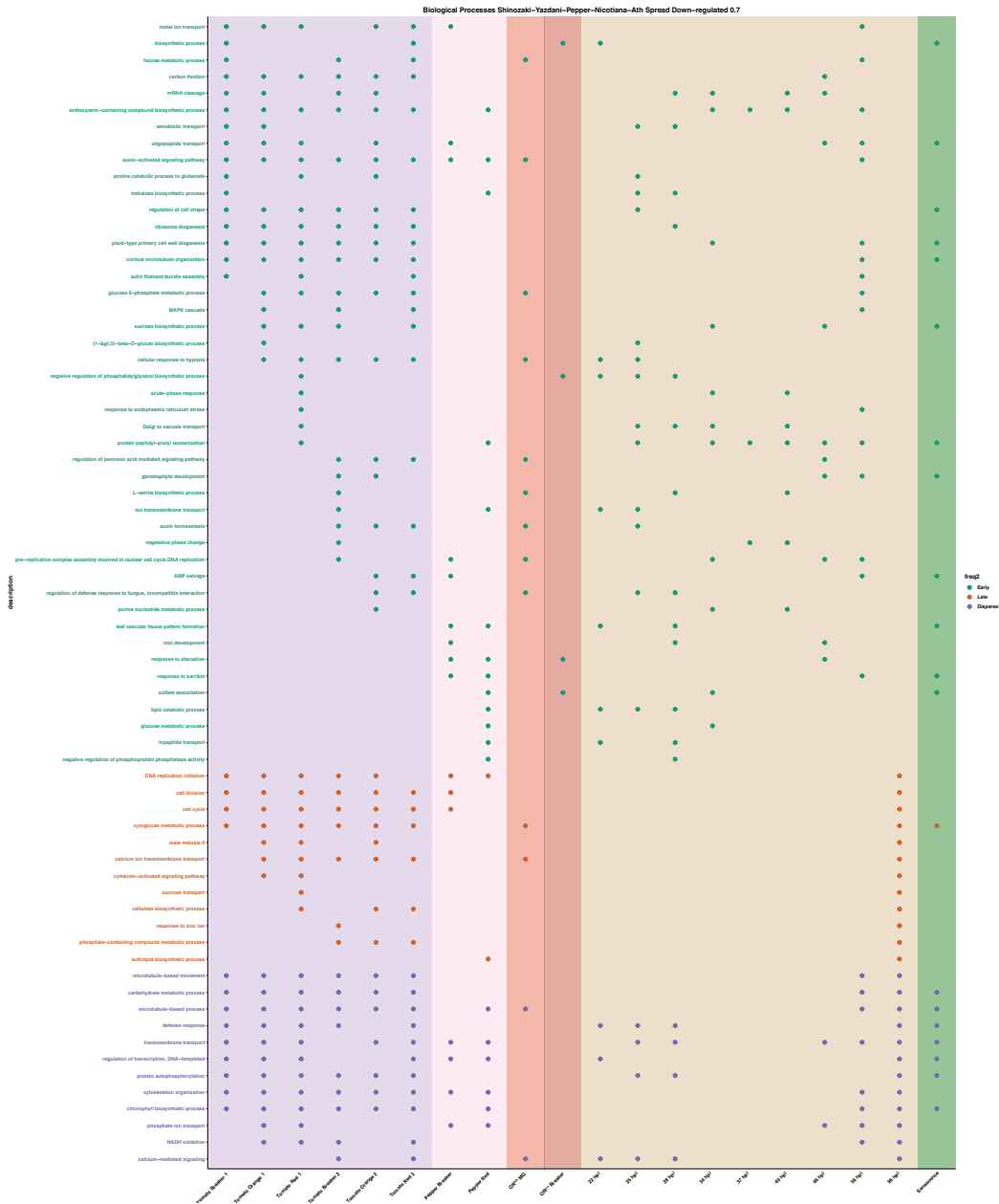
Supplementary Figure 9: Down-regulated "Spread" BPs. Distribution of all down-regulated BPs included "Spread" category. GO terms are divided in 3 different groups: Early (green), late (orange) and disperse (blue).
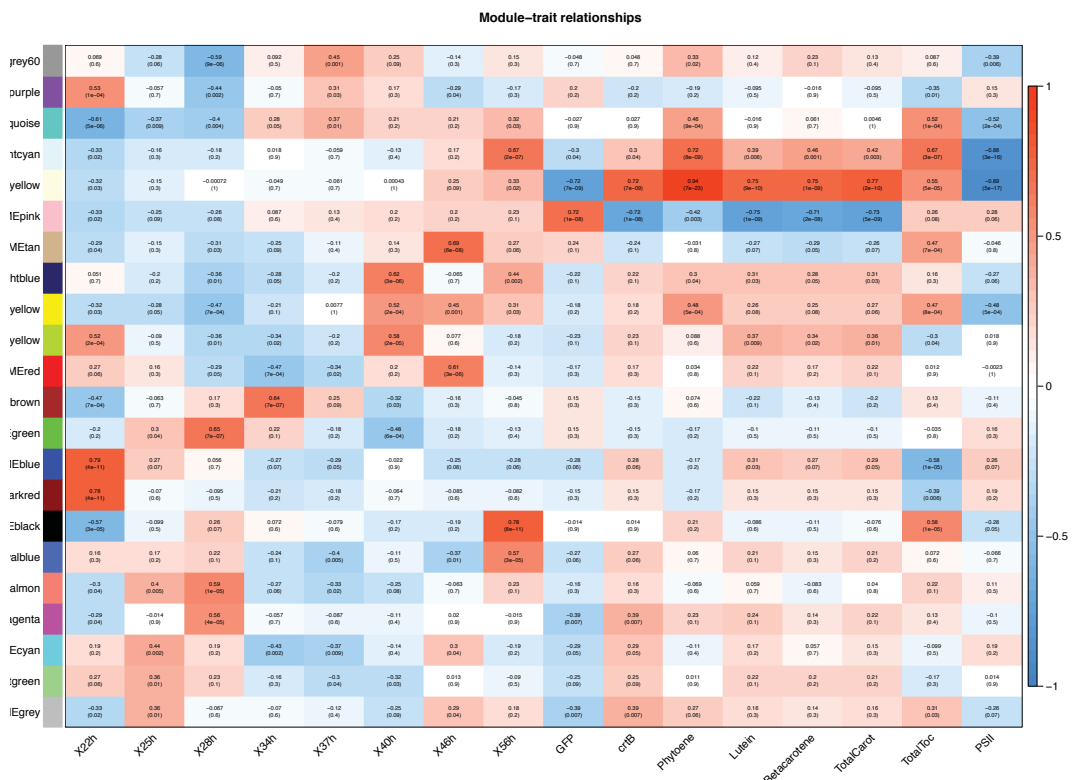
Supplementary Figure 10: Heatmap of correlations modules-phenotypic data with all genes. Correlation of eigengenes values in modules from WGCNA analysis using all genes from time course and phenotypic data (time points, construct, carotenoid levels and ϕPSII values.
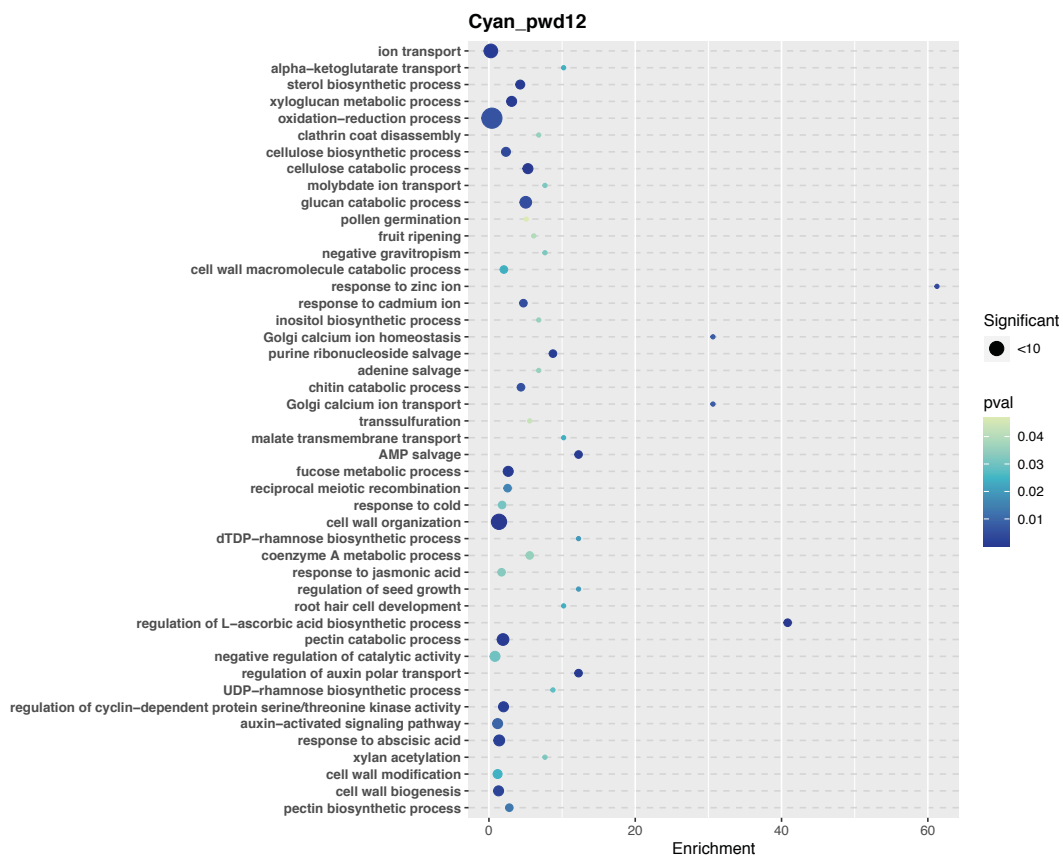
Supplementary Figure 11: GO enrichement in Cyan Module First WGCNA approach. Enrichment of BPs included in cyan module. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
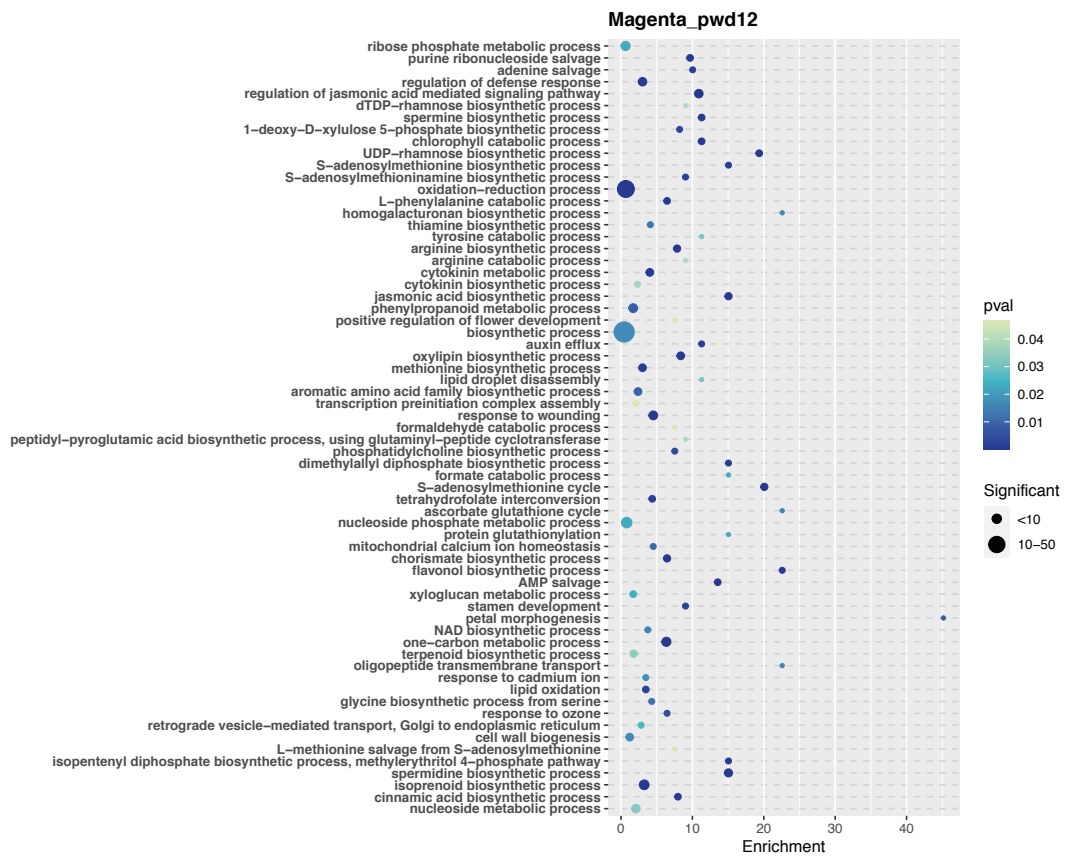
Supplementary Figure 12: GO enrichement in Magenta Module First WGCNA approach.
Enrichment of BPs included in magenta module. Number of genes included in each term
corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
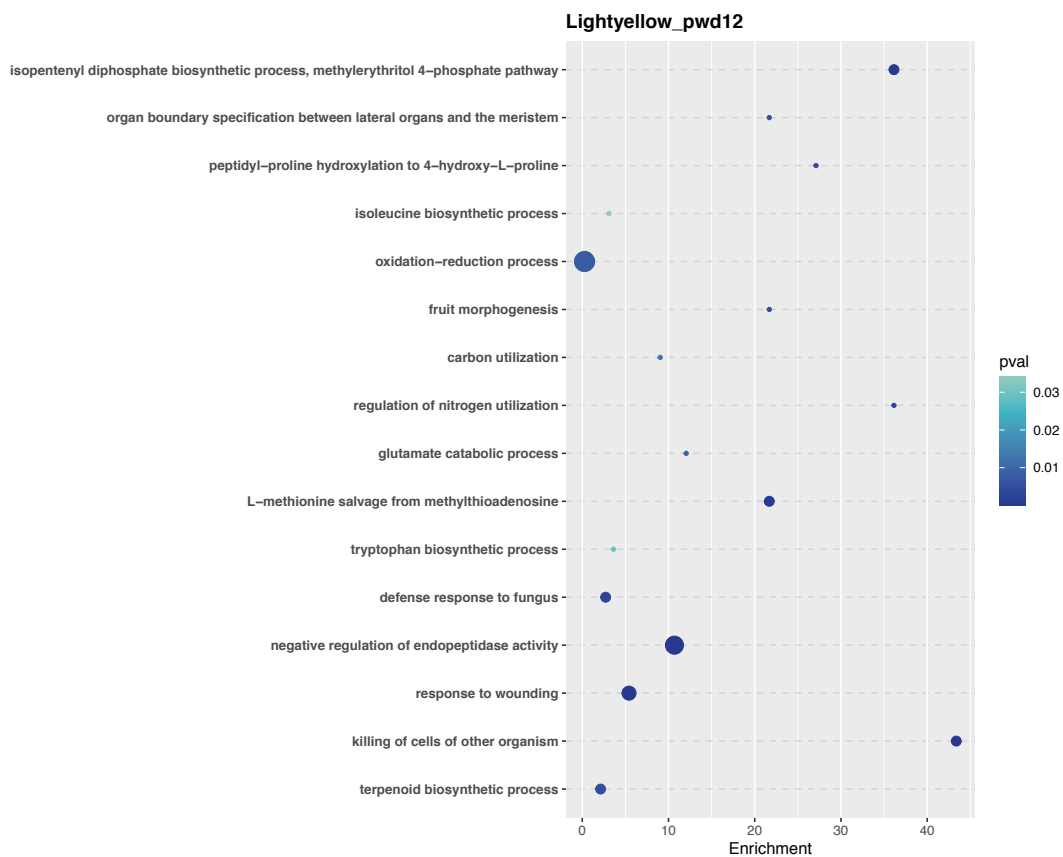
Lightyellow_pwd12

Supplementary Figure 13: GO enrichement in Lightyellow Module First WGCNA approach. Enrichment of BPs included in lightyellow module. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
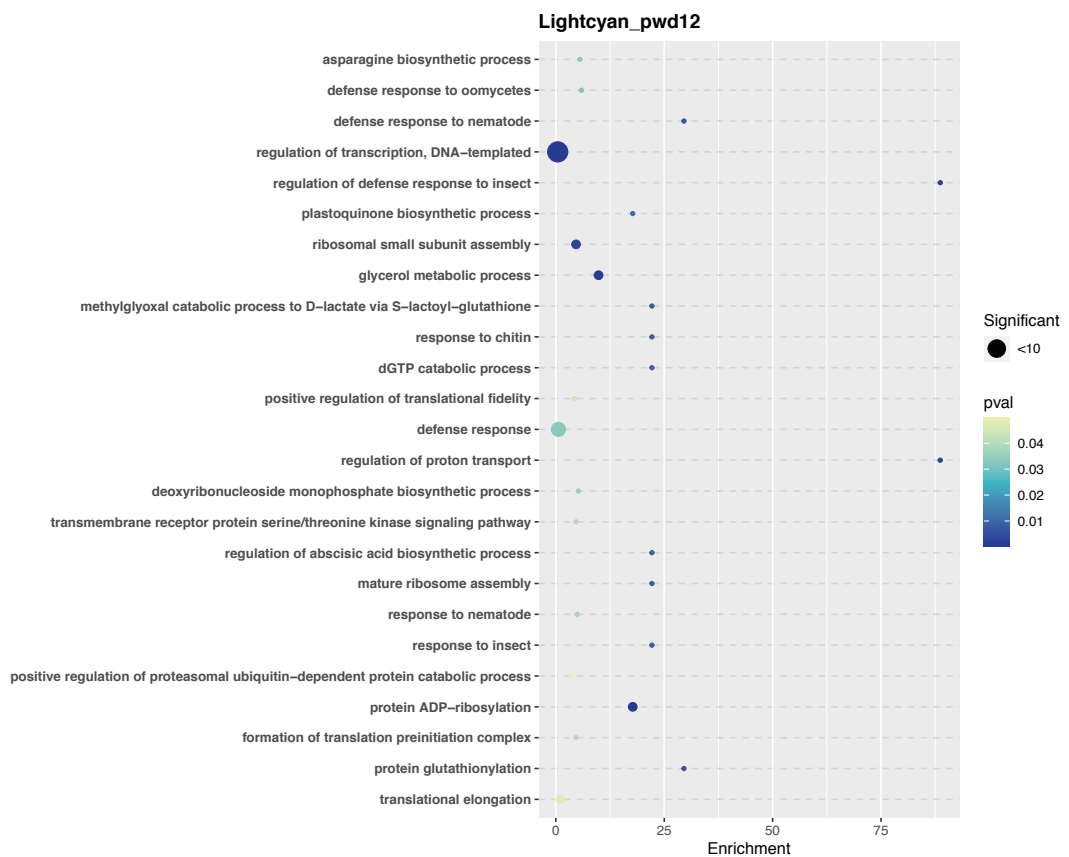
Supplementary Figure 14: GO enrichement in Lightcyan Module First WGCNA approach. Enrichment of BPs included in lightcyan module. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
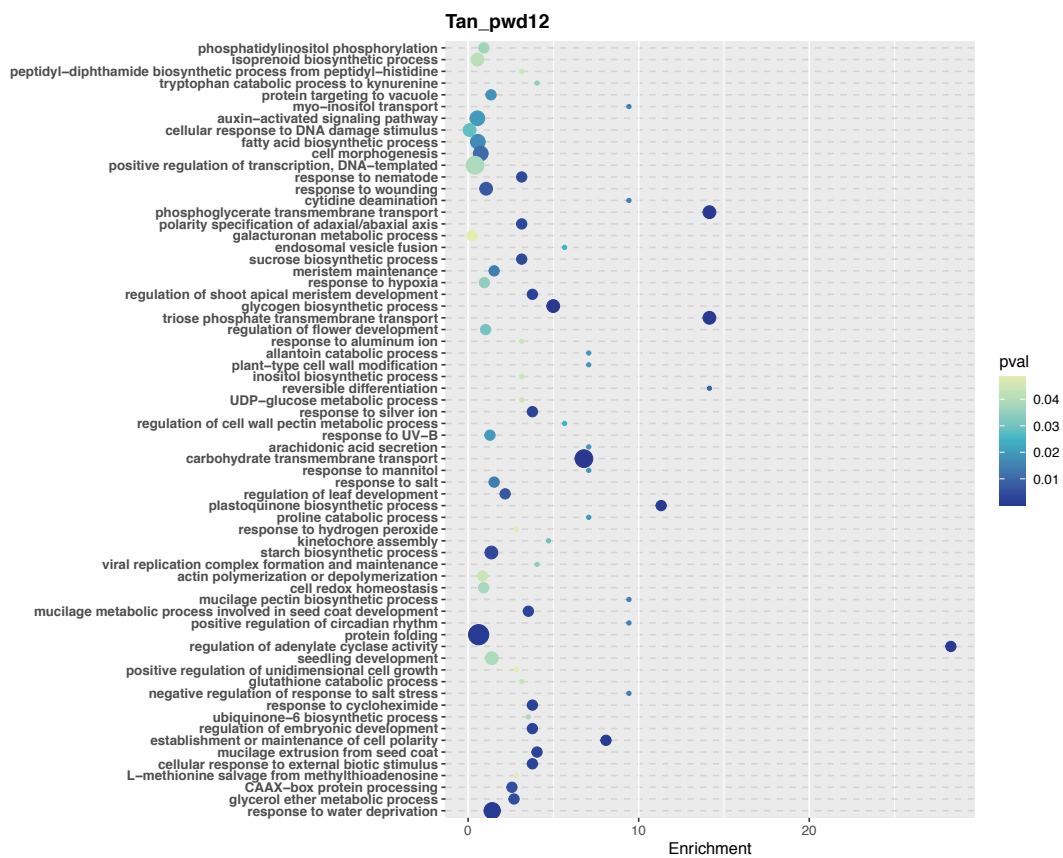
Supplementary Figure 15: GO enrichement in Tan Module First WGCNA approach. Enrichment of BPs included in tan module. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
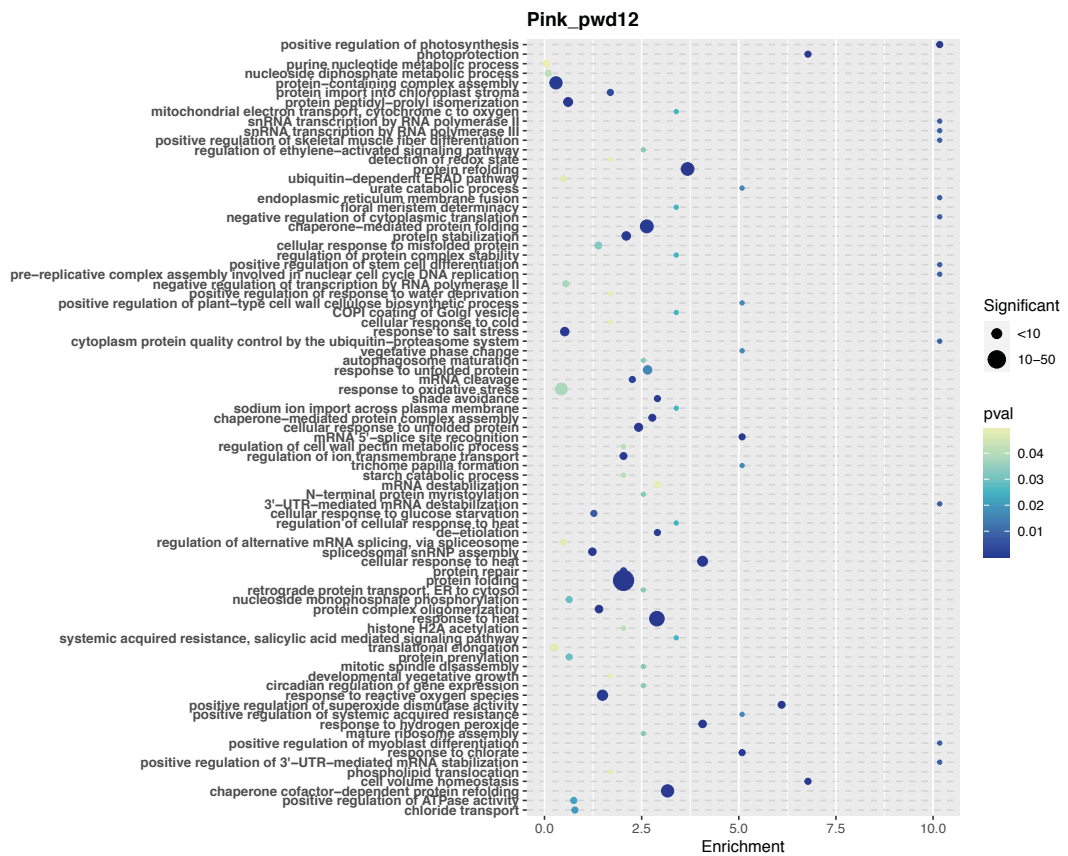
Supplementary Figure 16: GO enrichement in Pink Module First WGCNA approach. Enrichment of BPs included in pink module. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.

Module–trait relationships
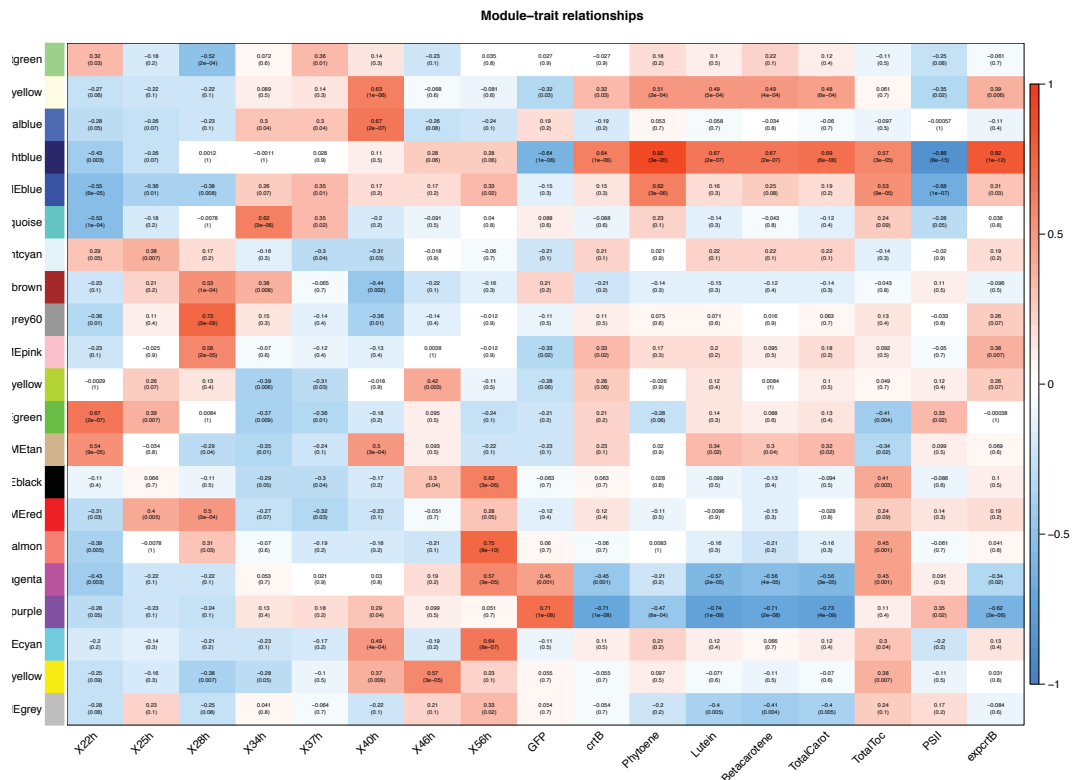
Supplementary Figure 17: Heatmap of correlations modules-phenotypic data with DEGs. Correlation of eigengenes values in modules from WGCNA analysis using all genes from time course and phenotypic data (time points, construct, carotenoid levels and ϕPSII values.
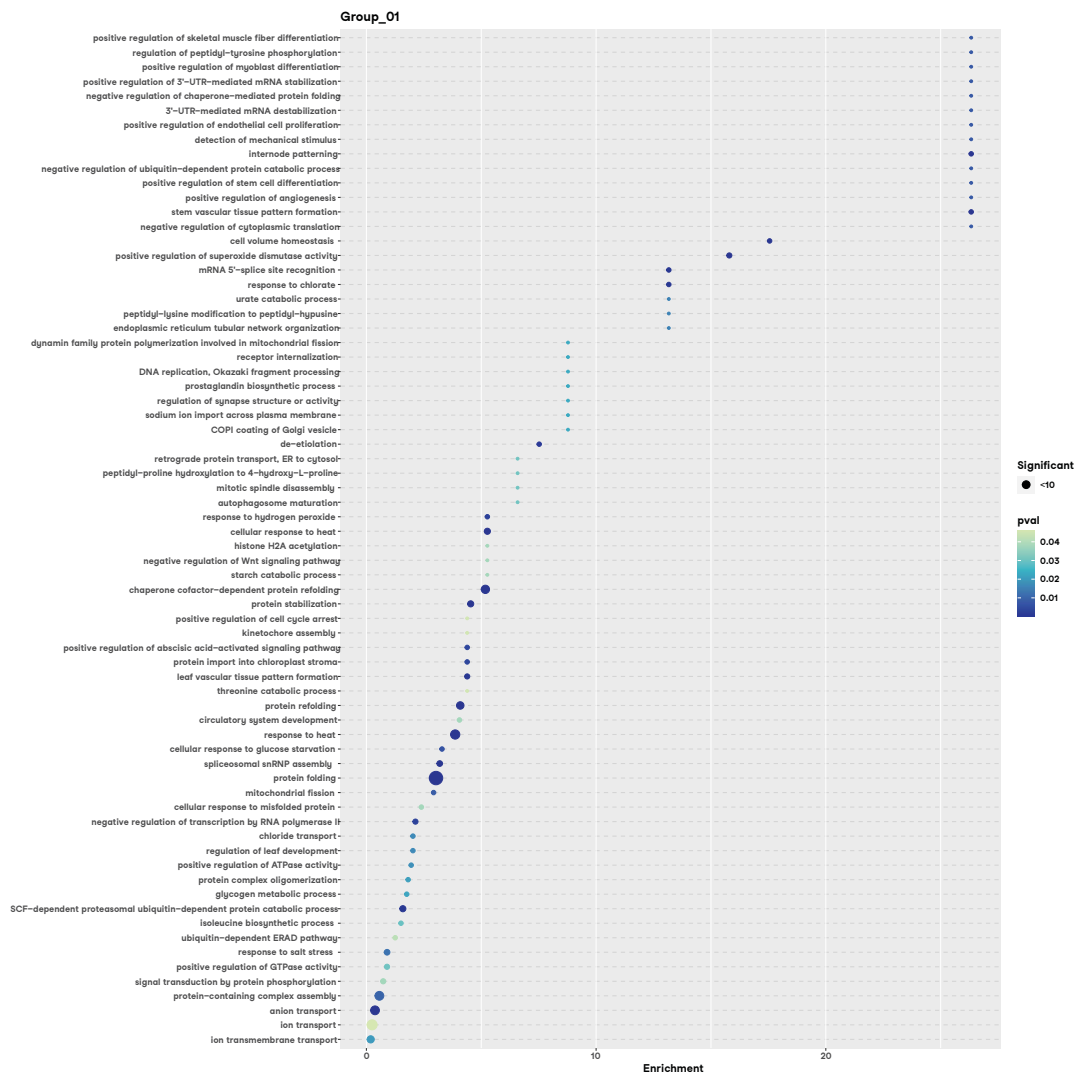
Supplementary Figure 18: GO enrichement in Group 1 of second WGCNA approach. Enrichment of BPs included in group 1 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
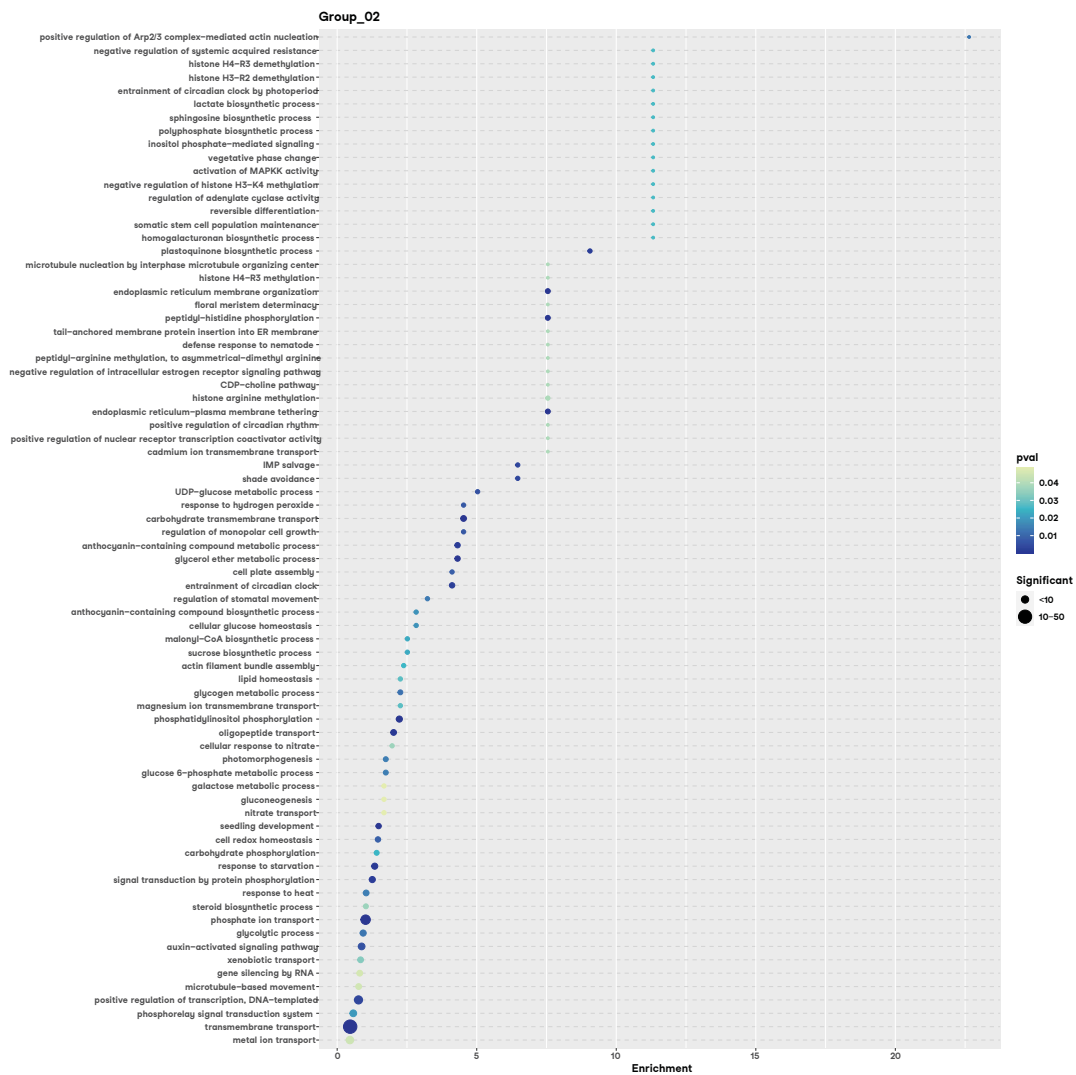
Supplementary Figure 19: GO enrichement in Group 2 of second WGCNA approach. Enrichment of BPs included in group 2 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
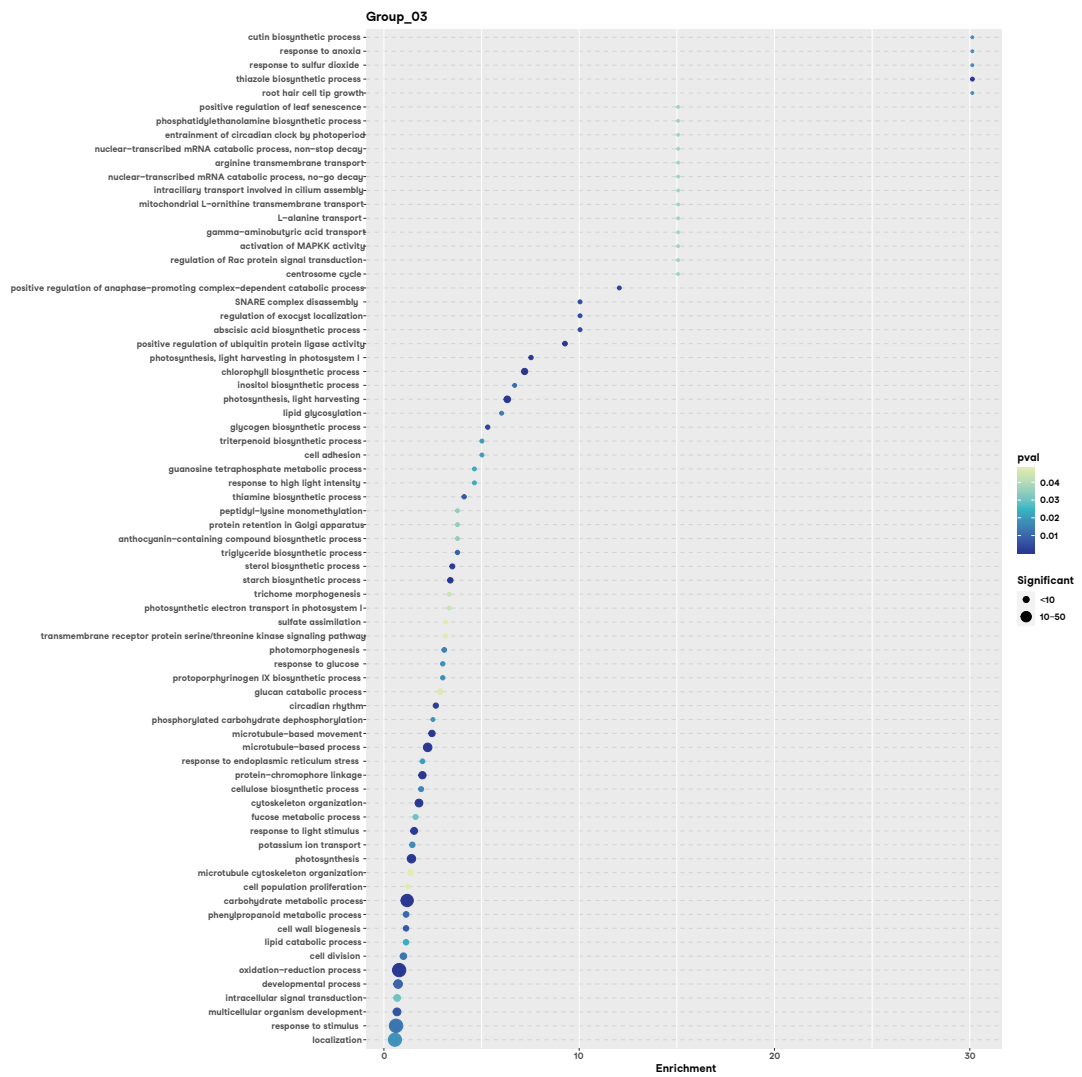
Supplementary Figure 20: GO enrichement in Group 3 of second WGCNA approach. Enrichment of BPs included in group 3 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
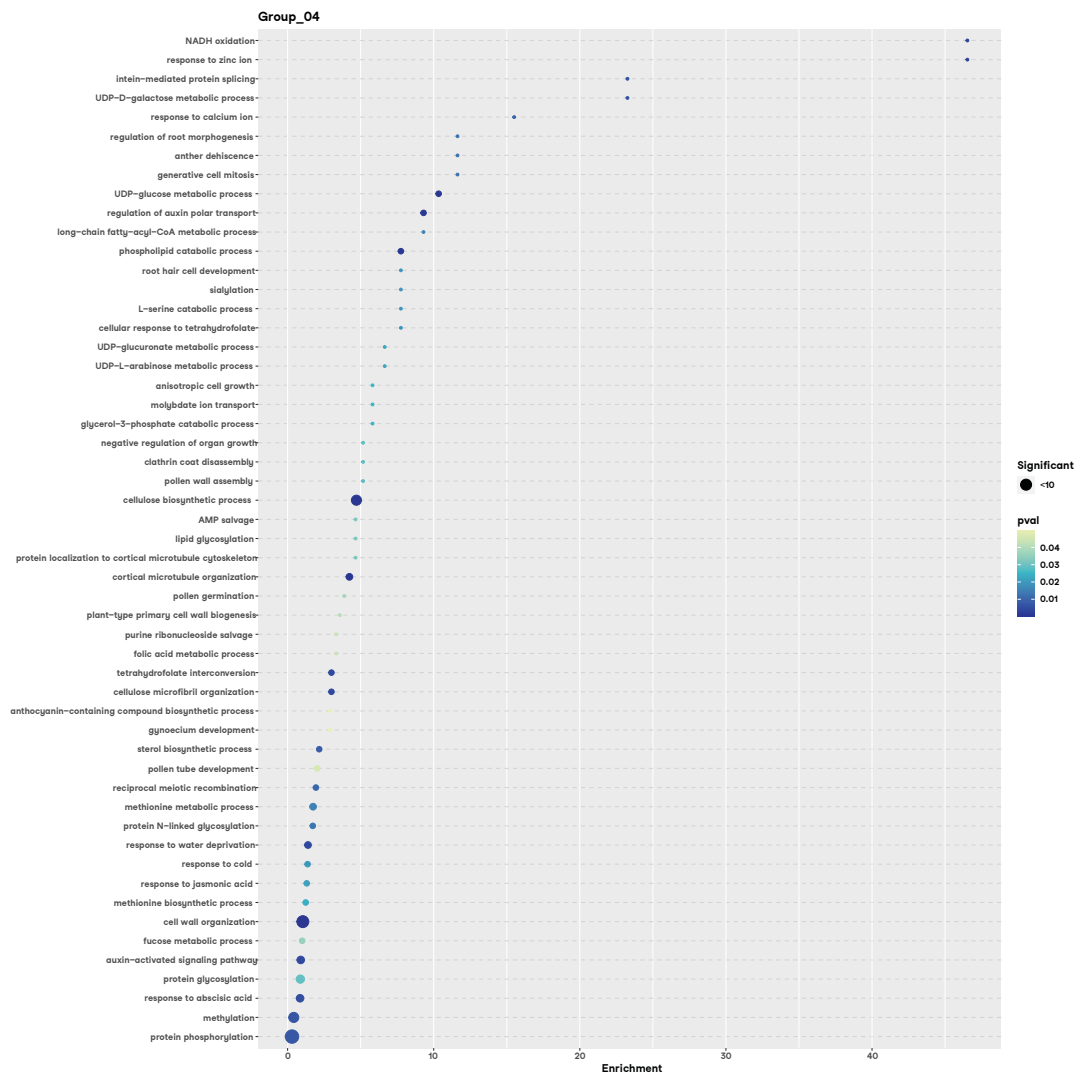
Supplementary Figure 21: GO enrichement in Group 4 of second WGCNA approach. Enrichment of BPs included in group 4 of the second WGCNA analysis, including only DEGs. Number of genes included in each term correspond to point size; p-values of gene enrichment analysis correspond to color scale.
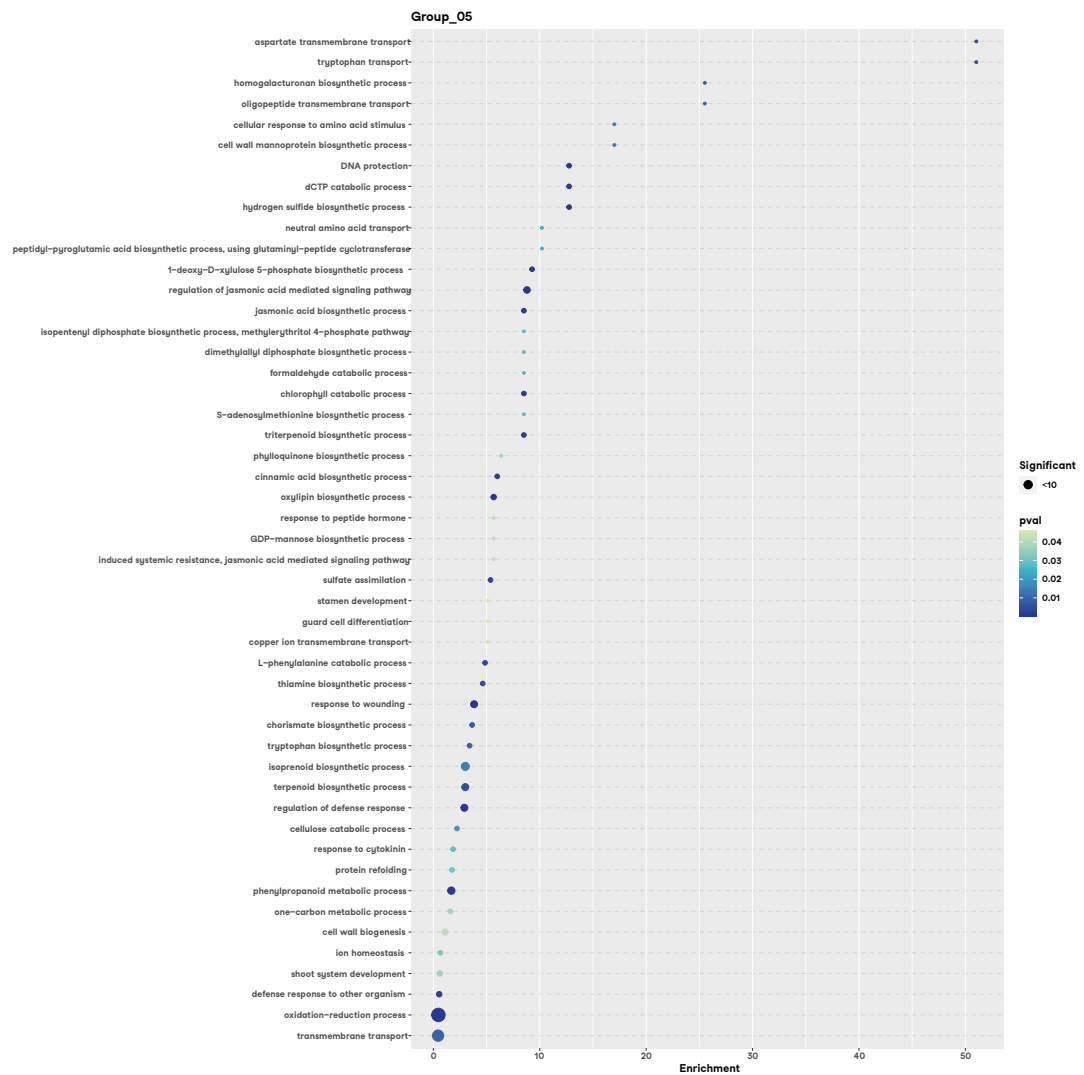
Supplementary Figure 22: GO enrichement in Group 5 of second WGCNA approach. Enrichment of BPs included in group 5 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
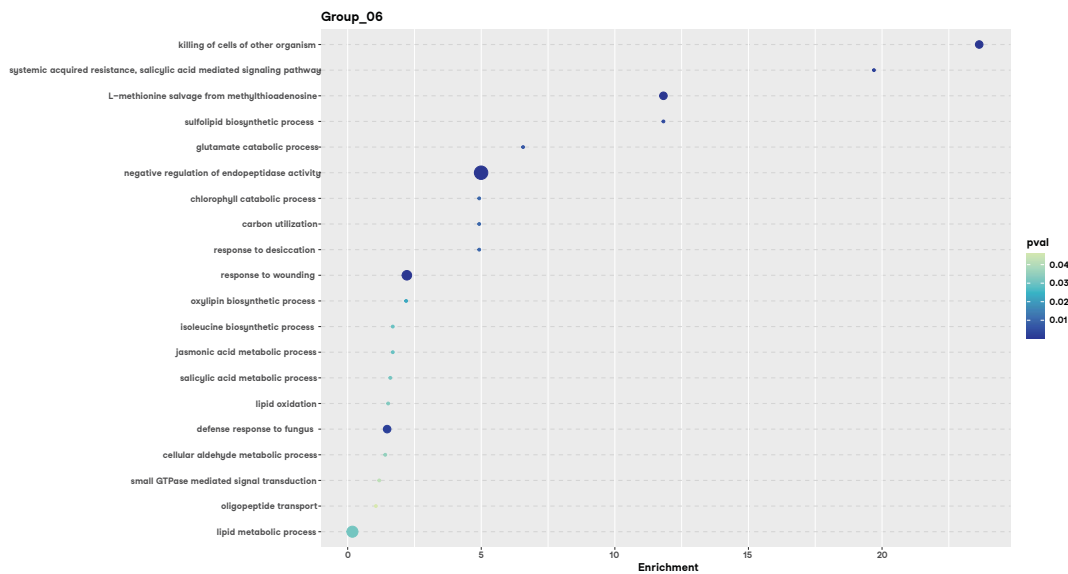
Supplementary Figure 23: GO enrichement in Group 6 of second WGCNA approach. Enrichment of BPs included in group 6 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
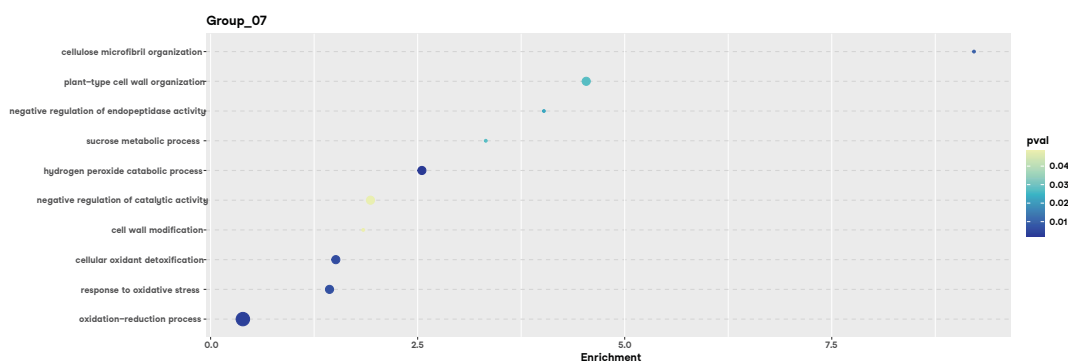
Supplementary Figure 24: GO enrichement in Group 7 of second WGCNA approach. Enrichment of BPs included in group 7 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.
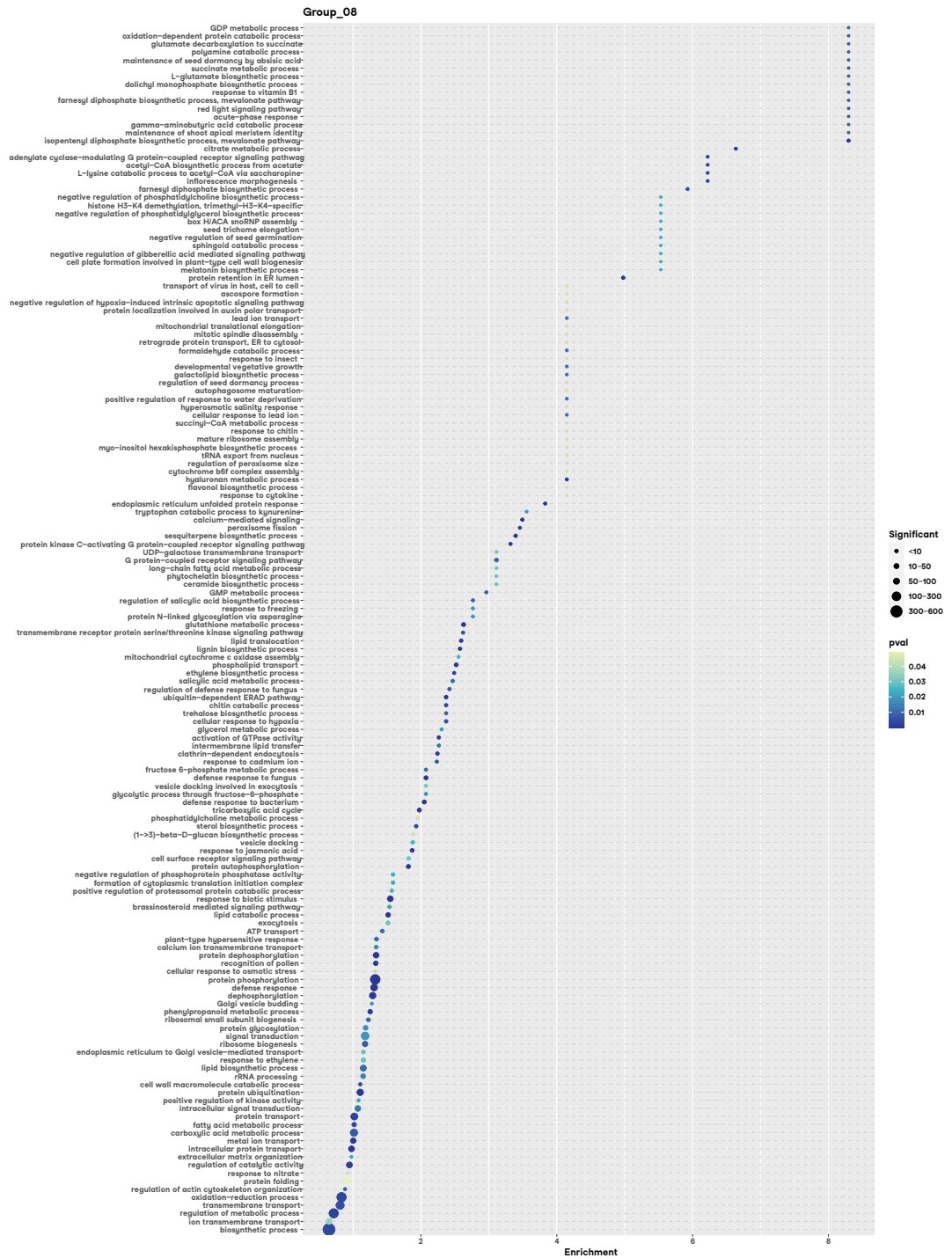
Supplementary Figure 25: GO enrichement in Group 8 of second WGCNA approach. Enrichment of BPs included in group 8 of the second WGCNA analysis, including only DEGs. Number of genes included in each term corresponds to point size; p-values of gene enrichment analysis correspond to color scale.

Supplementary Figure 26: UpSet diagram of GOs in groups from second WGCNA analysis. Distribution of BPs representing the genes included in the eight groups created from the second WGCNA analysis. Groups were distributed according to their abundance (from left to right). Number of GO terms in each group are represented in bar plots on the top; groups including in each group are highlighted below; number of GO terms in each time points are represented in horizontal bar plots on the left.

Supplementary Figure 27: ESCRT-related gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to ESCRT system. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 28: Cell wall biosynthesis, clock and autophagy related gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to cell wall biosynthesis (A), clock (B) and autophagy (C). Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
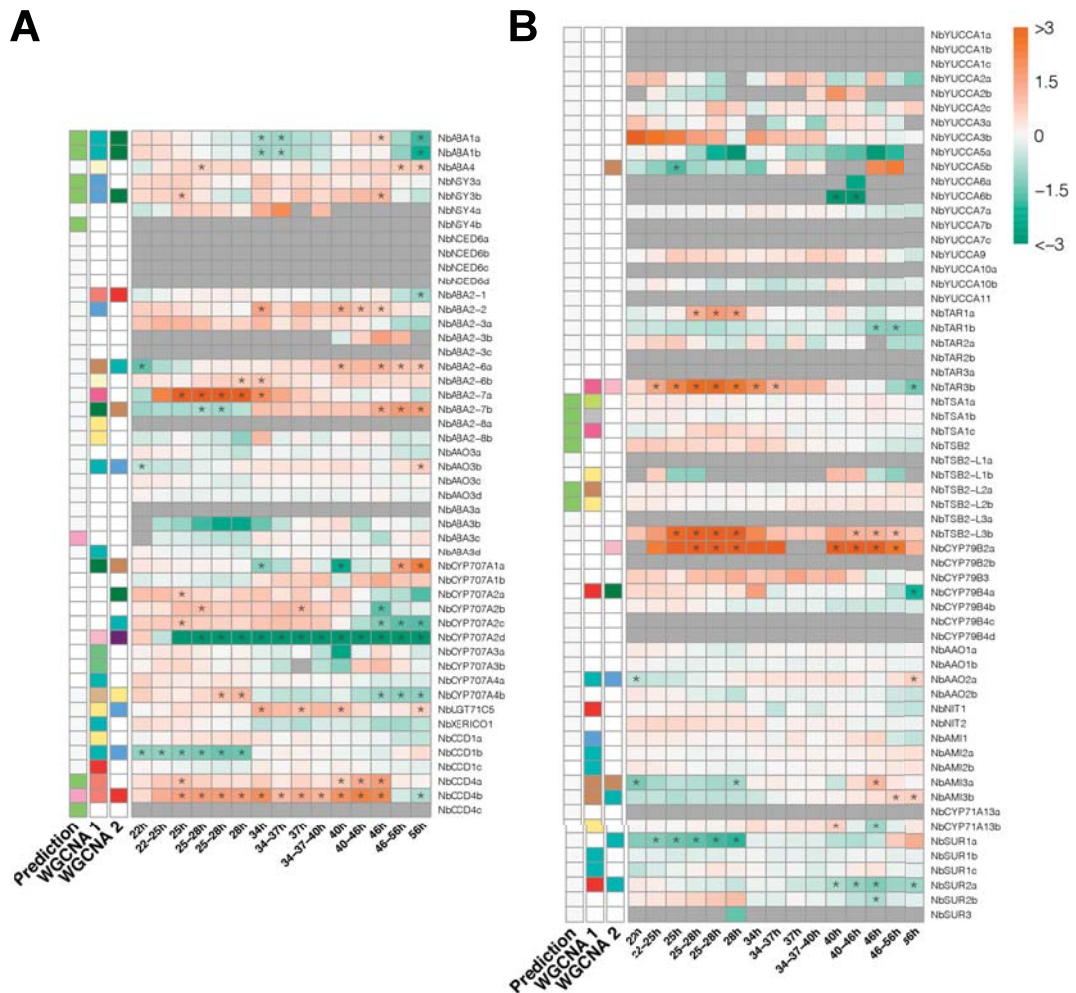
Supplementary Figure 29: ABA and auxin biosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to ABA (A) and auxin (B) biosynthesis. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
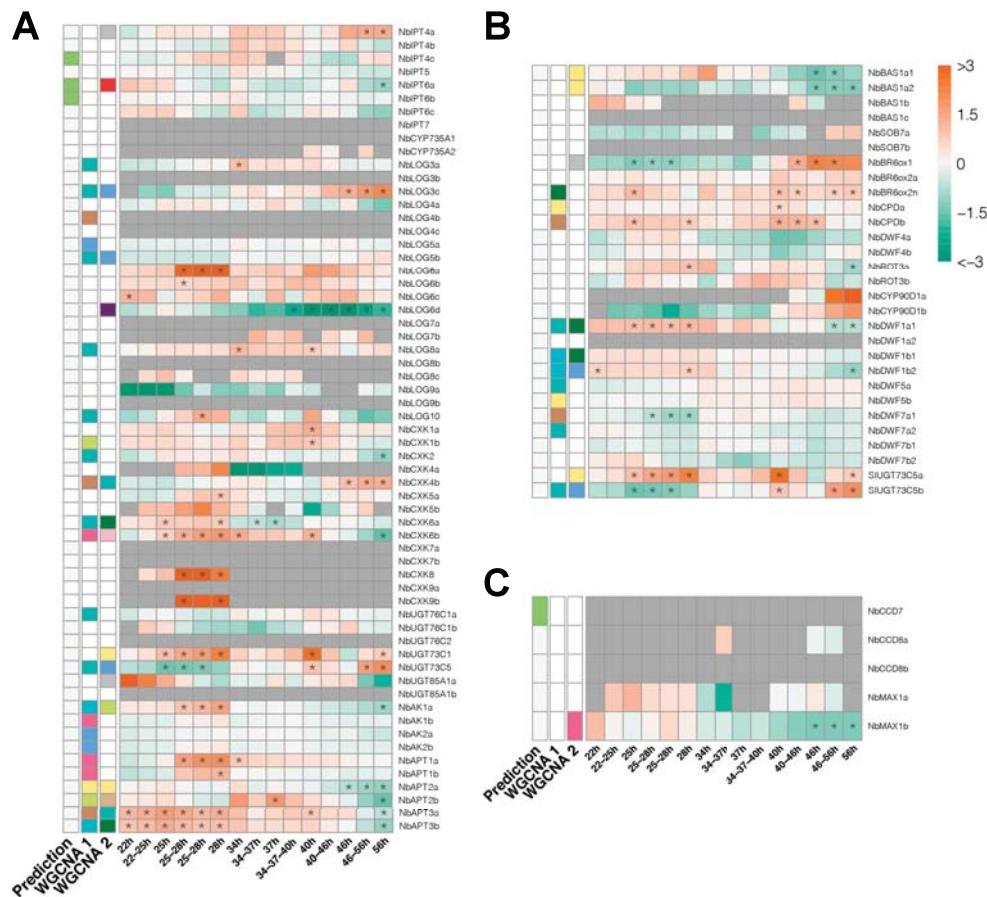
Supplementary Figure 30: Cytokinins, brassinosteroids and strigolactones biosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to cytokinins (A), brassinosteroids (B) and strigolactones (C) biosynthesis. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 31: Gibberellins, salicylic acid and ethylene biosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to gibberellins (A), salicylic acid (B) and ethylene (C) biosynthesis. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 32: BAG, Prefoldin and Cpn60 gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to BAG (A), prefolding (B) and chaperonin-60 (C) chaperone families. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 33: DNAJ gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to DNAJ chaperone families. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 34: Rotemase and cyclophilin gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to rotemase and cyclophilin chaperone families. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
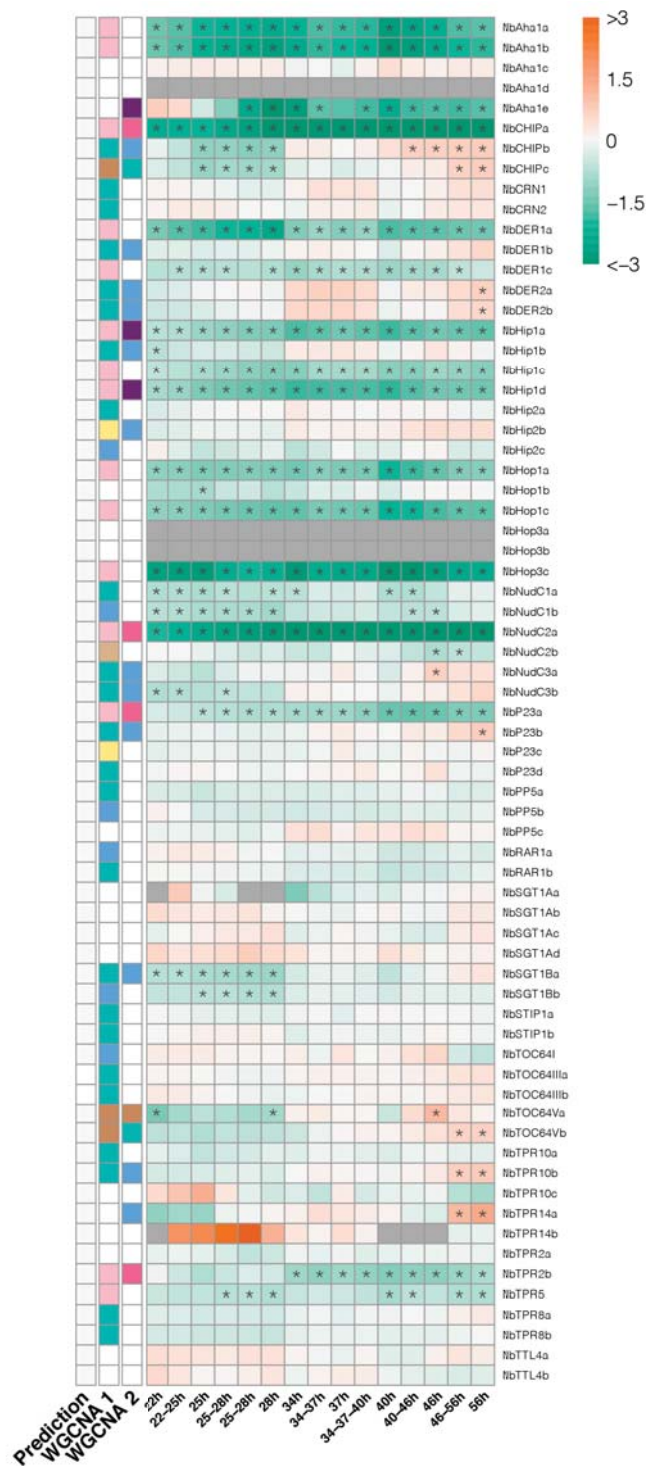
Supplementary Figure 35: Hsp90 co-chaperones gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to Hsp90 co-chaperones family. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
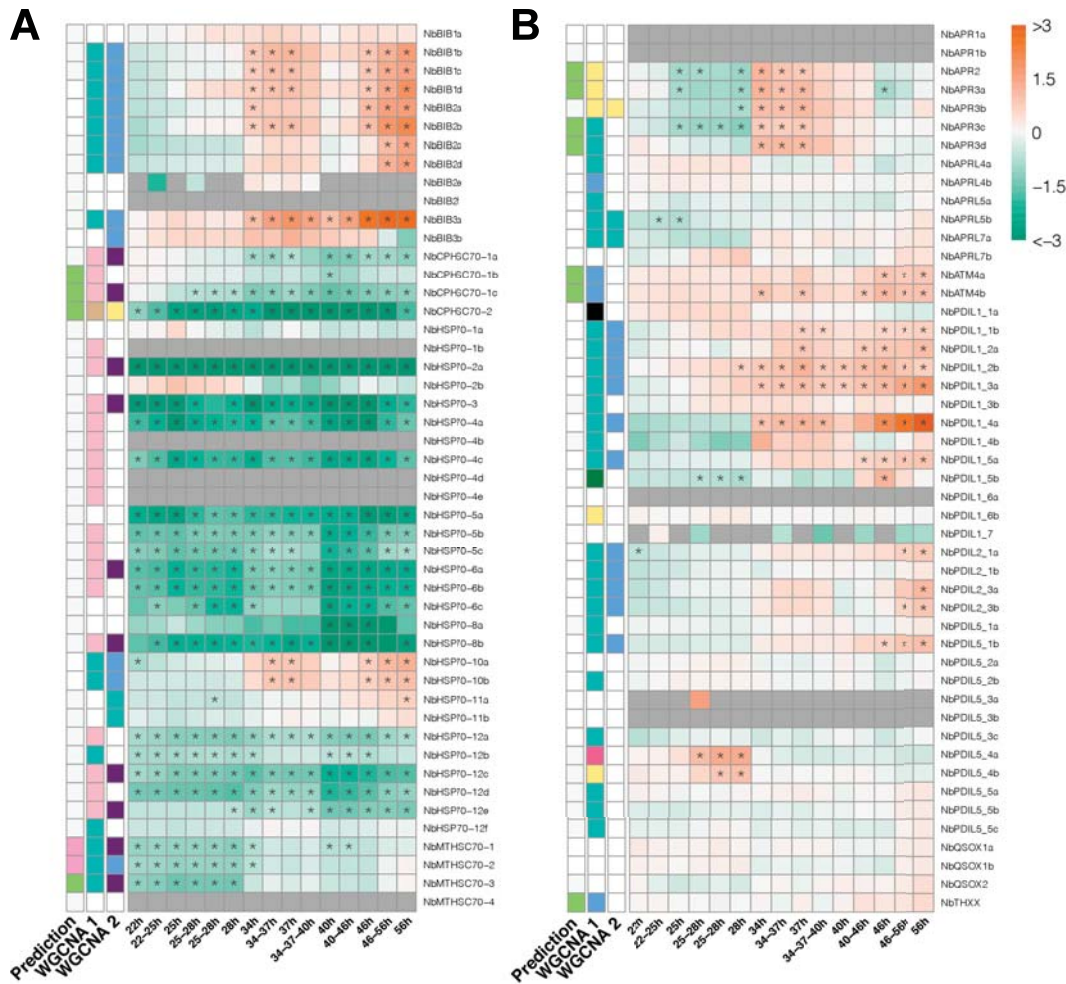
Supplementary Figure 36: Hsp70 and PDI gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to Hsp70 (A) and protein disulfide-isomerase (PDI) (B) chaperone families. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

Supplementary Figure 37: Chlorophylls gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to chlorophylls biosynthesis (A), chlorophyll cycle (B) and chlorophyll degradation (C) pathways. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
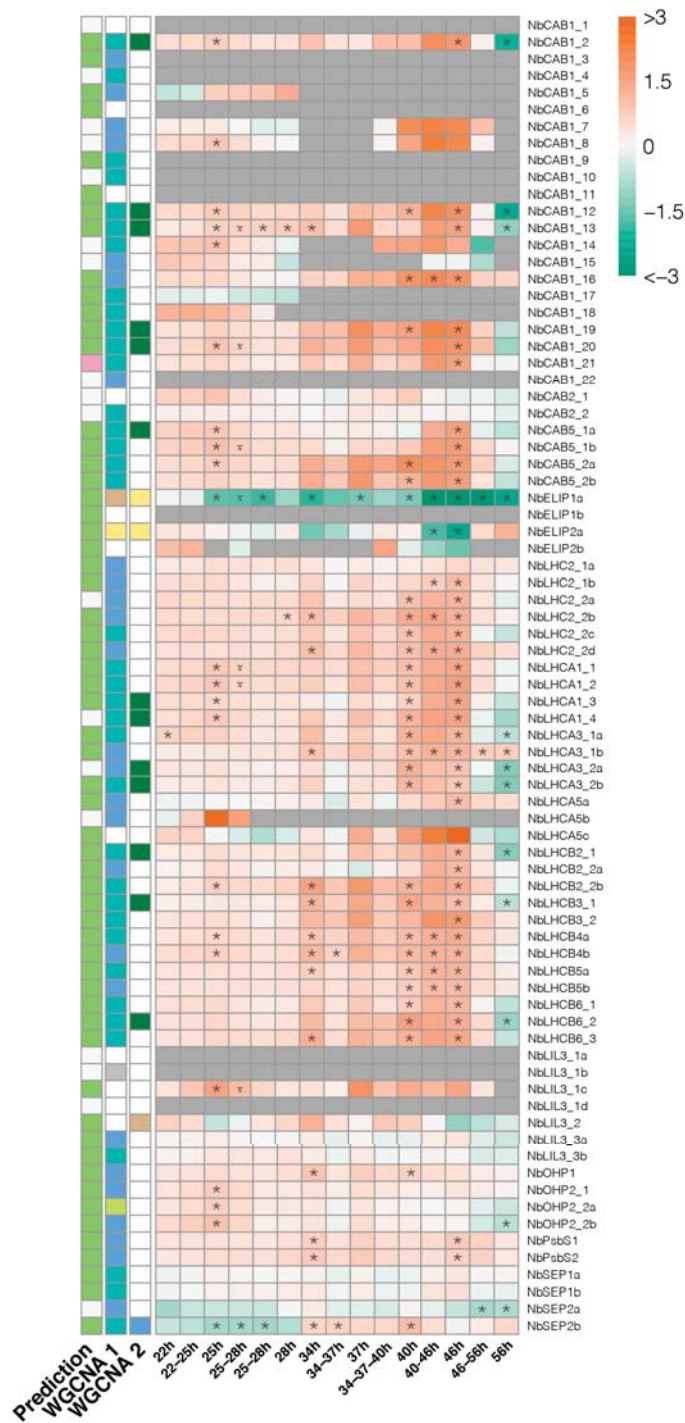
Supplementary Figure 38: Antenna complex gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of genes related to proteins from the antenna complex for light harvesting. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
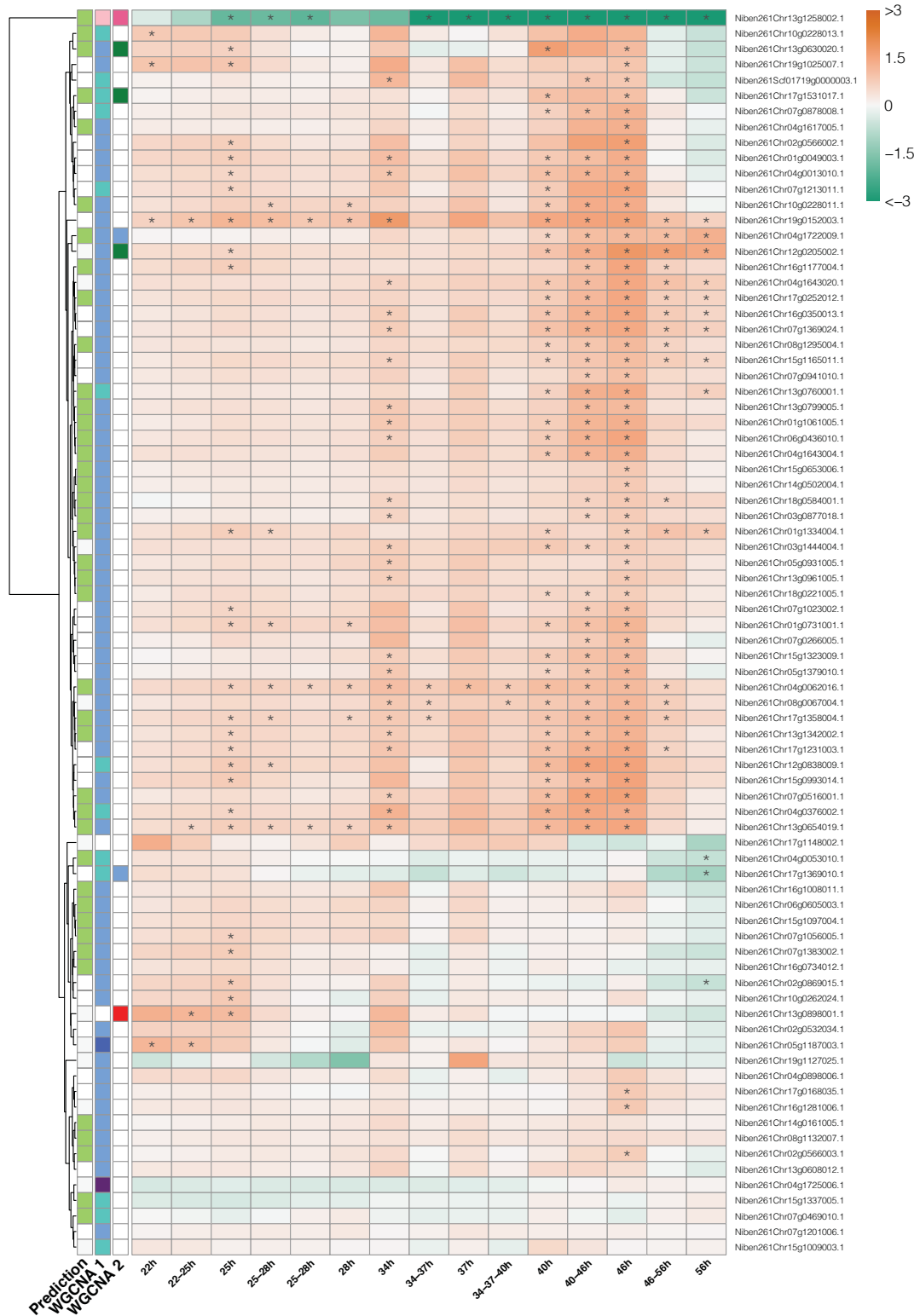
Supplementary Figure 39: Photosynthesis gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of candidate genes related to photosynthesis. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
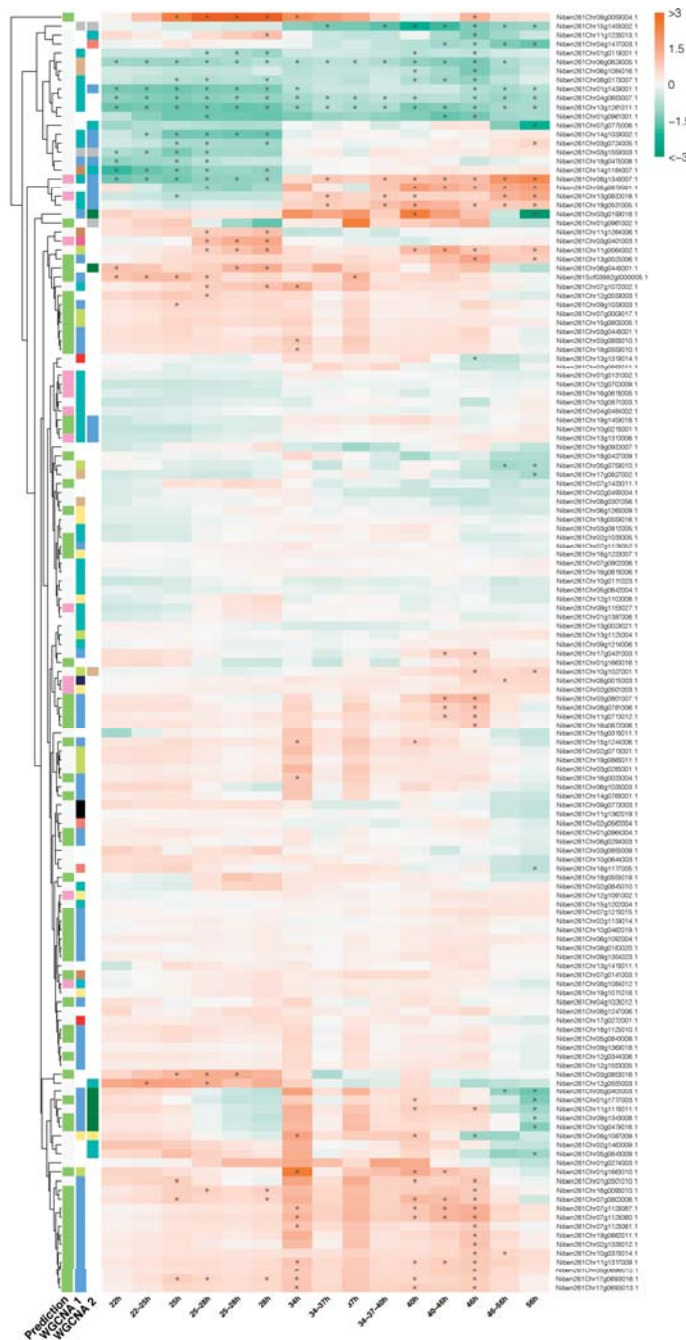
Supplementary Figure 40: Calvin-Benson cycle gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of candidate genes related to Calvin-Benson cycle. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.
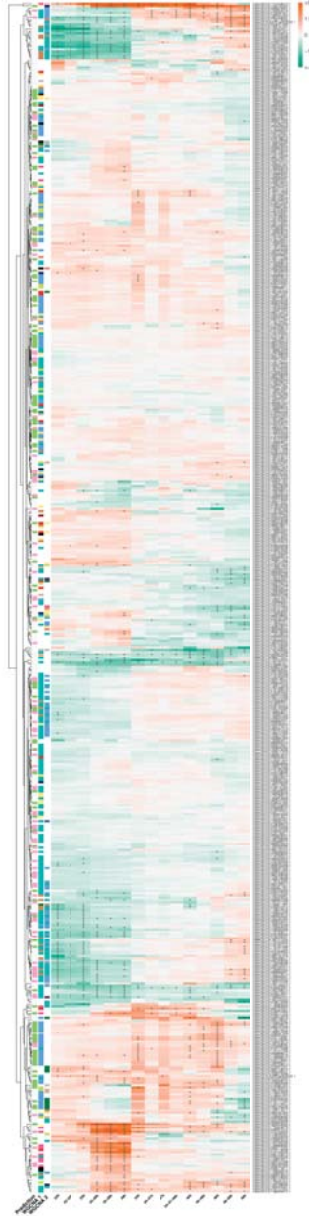
Supplementary Figure 41: Primary carbon metabolism gene expression in the time course. FC values of (p)crtB vs. GFP comparison in each comparison (time points and alternative groups based on PCA plot) of candidate genes related to primary carbon metabolism. Asterisks indicate a gene was significantly expressed in that specific comparison. Rows are labelled indicating if a gene was included in a module from the first WGCNA analysis, where all expressed genes were included (WGCNA 1), in a module from the second WGCNA analysis, where only DEGs were included (WGCNA 2), and the predicted subcellular location, according to TargetP 2.0 software (Prediction). Labels are showed in Figure 4.13.

```
            1         10                  20              30      70        80              90
S1PSY3    ..MCPATLSYSSN............SWLNARNEIFS.......SKKCKTKLLIRVD....WKQTNVTNEYLCCKNP.....RFDPMFLD
CaPSY3    ..MCPTTLSYSSN............SCINARNEIFS.......YQKCRRKLLIRA.....WKQTHVVNDLLCCRNP.....SFDPVFLD
NbPSY3b   ..MCPTILPYSAN............SCLRAGNRRLS.......SKKCRGKLHITA....WRQTHVTND..CYRKP.....SFDPMFLH
NbPSY3a   ..MCPAILPYSAN............SCLRAGNRRFL.......SQKCRGKLNITA.....YKELLTLIN.AFATKP.....SFDPMFLD
S1PSY2    ..MSVALLWVVSPNSEVSYG.TGFLDSVREGNRGLESSRFPSRDRNSMWKGGFKKGGRQGKRHLIS.TDDIQVKP..DIVLPGNIGLLS
CaPSY2    ..MSVALLWVVSPNSEVSNG.TGFLDSVREGNR......VPGRDRNSMWKERIKKGGRQRKRQLRS.TDDLEVKP..DIVLPGNIGLLS
NbPSY2b   .................................MK....KMQLRS.TDDLEVKP..DIVISENIGLLS
NbPSY2a   MSMSVALLWVVSPTSEVSNG.TGFLDSVREGNRVFVSSRFLARDRNLKWNGRIMKGGRQRKRQLRS.TDELEVKP..DIVVPGNIGLLS
NbPSY1b   ..MSVALLWVVSP.CEVSNG.TGFLDSIREGNRVFD....LSRYRNLVCNERNKRG....KRQLRS.TDDLEVKP..EIPLQGNISLLS
NbPSY1a   ..MSVALLWVVSP.CEVSNG.TGFLDSIREGNRVFD....WSRHRNLVCNERNKRG....KRQLRS.TDDLEVKP..EIPLPGNISLLS
S1PSY1    ..MSVALLWVVSP.CDVSNG.TSFMESVREGNRFFD....SSRHRNLVSNERINRGG...KRQLRS.TNELEVKP..DIPIPGNIGLLS
CaPSY1    ..MSVALLWVVSP.CDVSNG.TGFLVSVREGNRIFD....SSGRRRLNACNERIKRGG..KRQLRS.TDELDVKK..DIPIPGTLGLLS
PSY       MSSSVAVLWVATSSLNPDPMNNCGLVRVLESSRLFS......PCQNQRLNKGKKKQIPTWNKQLRSSSYDLDVKKPQDVVLPGSLSLLG
crtB      .............................................SKSFATASKLFDAKTR......RSVLMLY

            100        110       120                    130      140          150        160
S1PSY3    EAYELCRKICEEYAKTFYLGTKLMTEERQKAIWAIY.............VWCRRTDELVDGPNADYMNNSVLDRWEQRLEDIFNN.KPY
CaPSY3    EAYELCRKICAEYAKTFYLGTKLMTEERQKAIWAIY.............VWCRRTDELVDGPNADYMNNSVLDRWEE...ERLEDIFKN.KPY
NbPSY3b   EAYELCRKICEEYAKTFYL..............................ERLEDIFNC.RPY
NbPSY3a   EAYELCRKICAEYGKTFYL...................DELVDGPNADYMNNSVLDRWEE...ERLEDIFNC.RPY
S1PSY2    EAYDRCGEVCAEYAKTFYLGTMLMTPDRRRAIWAIY.............VWCRRTDELVDGPNASHITPQALDRWEARLEDIFNG.RPF
CaPSY2    EAYDRCGEVCAEYAKTFYLGTLMTPDRRRAIWAIY.............VWCRRTDELVDGPNASHITPQALDRWEARLEDIFSG.RPF
NbPSY2b   EAYDRC..........................................SDESYHSPTFPYIKV.............
NbPSY2a   EAYDRCGEVCAEYAKTFYLGTKLMTPERRRAIWAIY.............VWCRRTDELVDGPNASHITPQALDRWEARLEDIFSG.RPF
NbPSY1b   EAYDRCSEVCAEYAKTFYLGTMLMTPERRRAIWAIY.............VWCRRTDELVDGPNASHITPQGLDRWEDLLEDVFSG.RPF
NbPSY1a   EAYDRCSEVCAEYAKTFYLGTMLMTPERRRAIWAIY.............VWCRRTDELVDGPNASHITPQALDRWEDRLEDVFSG.RPF
S1PSY1    EAYDRCGEVCAEYAKTFNLGTMLMTPERRRAIWAIY.............VWCRRTDELVDGPNASYITPAALDRWENRLEDVFNG.RPF
CaPSY1    EAYDRCSEVCAEYAKTFYLGTMLMTPERRKAIWAIY.............VWCRRTDELVDGPNASHITPAALDRWEDRLEDVFSG.RPF
PSY       EAYDRCGEVCAEYAKTFYLGTLLMTPERRKAIWAIYVMLKVDFYKQSIVALVWCRRTDELVDGPNASHITPMALDRWEARLEDLFRG.RPF
crtB      AWCRHCDDVIDDQTLGFQAR...............QPALQTPEORLMQLEMKTRQAYAGSQMH

            170        180       190       200       210       220       230        240       250
S1PSY3    DMLDAALTDTICKFPLDIKPFKDMIDGMRMDTRKSRYANFQELYMYCYCVAGTVGLMSVPIMGIAPESCVSAQTVYNAALHLGIGNQLT
CaPSY3    DMLDAALTDTICKFPLDIKPFKDMIDGMRMDTRKSRYANFQELYMYCYCVAGTVGLMSVPIMGIAPECPVSAQTVYNAALHLGIGNQLT
NbPSY3b   DILDSALTDTISNFPLDIKPFKDMIEGMRMDTRKSRYANFQELYMYCCVAGTVGLMSVPIMGIAPECPLSAQTVYNSALHLGIGNQLT
NbPSY3a   GMLDSALTDTISNFPLDIKPFKDMIEGMRMDTRKSRYANFQELYMYCYCVAGTVGLMSVPIMGIAPECPVSAQTVYNSALHLGIGNQLT
S1PSY2    DMLDAALSDTVSRFPVDIQPFRDMVEGMRMDLWKSRYNNFDELYLYCYYVAGTVGLMSVPIMGIAPESKATTESVYNAALALGIANQLT
CaPSY2    DMLDAALSDTVSRFPVDIQPFRDMVEGMRMDLWKSRYMNFDELYLYCYYVAGTVGLMSVPIMGIAPESKATTESVYNAALALGIANQLT
NbPSY2b   ................................................................
NbPSY2a   DMLDAALSDTVSRFPVDIQPFRDMIEGMRMDLWKSRYNNFDELYLYCYYVAGTVGLMSVPVMGIAPESKATTESVYNAALALGIANQLT
NbPSY1b   DMLDAALSDTVSKFPVDIQPFRDMIEGMRMDLRKSRYRNFDELYLYCYYVAGTVGLMSVPIMGIAPDSKATTENVYNAALALGIANQLT
NbPSY1a   DMLDAALSDTVSKFPVDIQPFRDMIEGMRMDLRKSRYRNFDELYLYCYYVAGTVGLMSVPIMGIAPESKATTESVYNAALALGIANQLT
S1PSY1    DMLDGALSDTVSNFPVDIQPFRDMIEGMRMDLRKSRYKNFDELYLYCYYVAGTVGLMSVPIMGIAPESKATTESVYNAALALGIANQLT
CaPSY1    DMLDAALSDTVSKFPVDIQPFRDMIEGMRMDLRKSRYRNFDELYLYCYYVAGTVGLMSVPIMGIAPESKATTESVYNAALALGIANQLT
PSY       DMLDAALADTVARYPVDIQPFRDMIEGMRMDLKKSRYQNFDDLYLYCYYVAGTVGLMSVPVMGIDPKSKATTESVYNAALALGIANQLT
crtB      EPAFAAFQEVAMAHDIAPAYAFDHLEGFAMDVREAQYSQLDDTLRYCYHVAGVVGLMMAQIMGVRDN.....ATLDRACDLGLAFQLT

            260
S1PSY3    NILRDVG.....
CaPSY3    NILRDVG.....
NbPSY3b   NILRDVG.....
NbPSY3a   NILRDVG.....
S1PSY2    NILRDVG.....
CaPSY2    NILRDVG.....
NbPSY2b   .......
NbPSY2a   NILRDVG.....
NbPSY1b   NILRDVGEDARRGRVYLPQDELAHAGLSDDDIFAGKVTDKWRSFMKKQIQRARKFFDEAEEGVTQLSSASRWPPFRDMIEGMRMDLRKSRY
NbPSY1a   NILRDA.....
S1PSY1    NILRDVG.....
CaPSY1    NILRDVG.....
PSY       NILRDVG.....
crtB      NIARDIV.....

                                                            270       280       290
S1PSY3    ..........................................EDALRGRVYLPQDELAQFGIRDEDVFARKVTD
CaPSY3    ..........................................EDALRGRVYLPQDELAQYGIRDEDVFARNVTD
NbPSY3b   ..........................................EDALRGRVYLPQDELAQYGICDEDVFSRKVTD
NbPSY3a   ..........................................EAALRGRVYLPQDEQAQYGLCDEGVFSRKV..
S1PSY2    ..........................................EDARRGRVYLPQDELAQAGLSDEDIFAGKVTD
CaPSY2    ..........................................EDARRGRIYLPQDELAQAGLSGEDIFAGRVTD
NbPSY2b   ................................
NbPSY2a   ..........................................EDARRGRVYLPQDELAQAELSDEDIFAGKVTD
NbPSY1b   RNFDELYLYCYYVAGTVGLMSVPIMGIAPDSKATTENVYNAALALGIANQLTNILRDVGEDARRGRVYLPQDELAHAGLSDDDIFAGKVTD
NbPSY1a   ..........................................GLSDDDIFTGKVTD
S1PSY1    ..........................................EDARRGRVYLPQDELAQAGLSDEDIFAGRVTD
CaPSY1    ..........................................EDARRGRVYLPQDELAQAGLSDEDIFAGKVTD
PSY       ..........................................EDARRGRVYLPQDELAQAGLSDEDIFAGKVTD
crtB      ..........................................DDAHAGRCYLPASWLEHEGLNKENYAAPENRQ

            300        310       320       330       340       350       360       370       380
S1PSY3    QWREFMKEQIRRARFYFNLAEEGASHLNKASRWPVWSSLILYMKILDAIEENDYDNLTKRAYYGRAKKLVTLPVSYARALSLPSLAIQ.
CaPSY3    QWRGFMKEQIRRARFYFNLAEEGASHLNKASRWPVWSSLILYRKILDAIEENDYDNLTKRAYYGRAKKLATLPVSYARALSLPSLAIQ.
NbPSY3b   QWREFMKEQIRRARFYFNLAEEGASHLNKASRWPVWSSLILYRKILDAIEENDYDNLTKRAYYGRAKKLVTLPISYARALSVPSLAIQ.
NbPSY3a   ...............WSSLILYRKILDAIEENDYDNLTKRAYYGRAKKLVTLPVSYARALSVPSLAIQ.
S1PSY2    KWRIFMKKQIQRARKFFDEAEKGVTELSSASRWPVLASLLLYRKILDEIEANDYNNFTRRAYVSKPKLLTLPIAYARSLVPPKSTSSP
CaPSY2    KWRIFMKKQIQRARKFFDQAEKGVAELSSASRWPVLASLLLYRKILDEIEANDYNNFTRRAYVSKPKKLLTLPIAYARSLVPPKLTSSS
NbPSY2b   ................................
NbPSY2a   KWRNFMKKQIQRARKFFDESEKGVTELDSASRWPVLAALLLYRKILDEIEANDYNNFTRRAYVSKPKKLLTLPIAYAKSLVPPNRTSSP
NbPSY1b   KWRSFMKKQIQRARKFKNEAEEGVTQLSSASRWPVWASLLLYRQILDEIEANDYNNFTKRAYYSKSKKLISLPIAYAKSLVPPTRTLVS
NbPSY1a   KWRSFMKKQIQRARKFNEAEEGVTQLSSASRWPVWASLLLYRQILDEIEANDYNNFTKRAYYSKSKKLISLPIAYAKSLVPPTKTASL
S1PSY1    KWRIFMKKQIHRARKFFDEAEKGVTELSSASRWPVLASLLLYRRILDEIEANDYNNFTKRAYVSKPKKLIALPIAYAKSLVPPTKTASL
CaPSY1    KWRIFMKKQIQRARKFFDEAEKGVTELSAASRWPVLASLLLYRRILDEIEANDYNNFTKRAYVSKPKKLIALPIAYAKSLVPSTRT...
PSY       KWRNFMKMQLKRARMFFDEAEKGVTELSAAWRWPVLASLLLYRRILDEIEANDYNNFTKRAYYGKVKKIAAPLAYAKSVLKTSSSRLS
crtB      ALSRIARRLVQEAEPYYLSATAGLAGLPLRSAWAIATAKQVYRKIGVKVEQAGQQAWDQRQSTTTPEKLTLLLAASGQALTSRMRAHPP

S1PSY3    ..........
CaPSY3    ..........
NbPSY3b   ..........
NbPSY3a   ..........
S1PSY2    LAKT......
CaPSY2    LTKT......
NbPSY2b   ..........
NbPSY2a   LAKT......
NbPSY1b   SS........
NbPSY1a   SIAKA.....
S1PSY1    QR........
CaPSY1    ..........
PSY       I.........
crtB      RPAHLWQRPL
```

Supplementary Figure 42: Phytoene synthase protein sequences alignment. Alignment of protein sequences from PSY in *Arabidopsis*, crtB in *Pantoea ananatis* and predicted PSYs of tomato, pepper and *N. benthamiana*.