

The carotenoid pathway in diatoms

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 15
der Johann Wolfgang Goethe -Universität
in Frankfurt am Main

von
Alba Blázquez Pla
aus Xàtiva

Frankfurt (2020)
(D30)

vom Fachbereich 15 der
Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Klimpel

Gutachter : Prof. Dr. Büchel, Prof. Dr. Sandmann

Datum der Disputation : 02/12/2020

Table of contents

Abbreviations.....	0
1. General background	1
1.1 Diatoms.....	1
1.2 Carotenoid pathway in diatoms.....	4
1.3 Transport of enzymes to the plastid	7
1.4 Regulation of the carotenoid pathway	8
2. Materials and methods.....	10
2.1 Cultivation of the organisms	10
2.1.1 Cultivation of <i>Phaeodactylum tricornutum</i>	10
2.1.2 Cultivation of <i>Cyclotella</i> and <i>Chlorella</i>	10
2.1.3 Cultivation of <i>Escherichia coli</i>	11
2.1.4 Cultivation of <i>Agrobacterium tumefaciens</i>	11
2.1.5 Cultivation of <i>Arabidopsis thaliana</i>	12
2.2 Isolation and preparation of genetic material.....	12
2.2.1 RNA extraction from <i>P. tricornutum</i>	12
2.2.2 RNA isolation from plant tissue	13
2.2.3 Preparation of cDNA from mRNA	14
2.2.4 DNA isolation from <i>P. tricornutum</i>	14
2.2.5 DNA isolation from plant tissue	14
2.2.6 Plasmid extraction from the cells	15
2.2.7 Amplification of DNA by PCR.....	15
2.2.8 Colony PCR	17
2.2.9 Restriction Assay	17
2.3 Cloning by the hot fusion method.....	17
2.4 Transformation of the organisms.....	18
2.4.1 Production of competent <i>E. coli</i> XL-1 cells.....	18
2.4.2 Transformation of <i>E. coli</i> by the heat shock method	19
2.4.3 Production of competent <i>Agrobacterium tumefaciens</i>	19
2.4.4 Transformation of <i>Agrobacterium tumefaciens</i>	20
2.4.5 Transformation of <i>Arabidopsis thaliana</i> and screening of transformants.....	20
2.4.6 Transformation of <i>Phaeodactylum tricornutum</i>	21

2.5 Extraction and quantification techniques	22
2.5.1 Pigment extraction from <i>P. tricornutum</i>	22
2.5.2 Pigment extraction from plant leaves	22
2.5.3 HPLC analysis	23
2.5.4 DNA quantification with Nanodrop or ImageJ	23
2.6. Cell encapsulation.....	24
2.6.1 Determination of cell viability	25
3.Improvement of fucoxanthin levels	26
3.1Introduction	26
3.1.1 Fucoxanthin	26
3.1.2 Metabolic engineering.....	27
3.1.3 Endogenous regulation of metabolism.....	29
3.1.4 Design-build-test-learn cycle.....	30
3.1.5 Approaches to enhance carotenoid levels	31
3.2 Objectives	36
3.3 Results	36
4.Screening system for the selection of mutants	46
4.1 Introduction	46
4.1.1 High-throughput screening of carotenoids	46
4.1.2 Cell encapsulation.....	47
4.1.3 Encapsulation in alginate beads	49
4.2 Objectives	50
4.3 Results	51
4.3.1 Bead stability in the culture media	51
4.3.2 Cell proliferation over time after encapsulation	52
4.3.3 Cell viability within the alginate beads.....	55
4.3.4 Determination of the cell occupation	55
4.3.5 Biomass monitoring	57
5. Unknown enzymes of the pathway and comparative genomics	59
5.1 Introduction	59
5.1.1 Carotenoid diversity in algae.....	59
5.1.2 Strategies for linking functions to sequences.....	64
5.1.3 Missing enzymes of the carotenoid pathway in diatoms	68
5.2 Objectives	69
5.3 Results	69

5.3.1 Characterization of the gene order in the MEP and carotenoid pathways.....	69
5.3.2 Characterization of gene co-expression in the MEP and carotenoid pathways.....	73
5.3.3 Identification of putative genes for the fucoxanthin synthase function	76
6. Functional characterization of the lut genes of <i>P. tricornutum</i>	82
6.1 Introduction	82
6.1.1 Classes of β -carotene hydroxylases	82
6.1.2 Carotene hydroxylases in plants	83
6.1.3 Carotene hydroxylation in diatoms.....	86
6.2 Objective	86
6.3 Results	87
7. Discussion	96
8. Conclusions.....	107
9. Zusammenfassung.....	108
10. Bibliography.....	113
11. Appendix	122
11.1 Sequence alignments.....	122
11.2 Gene Identity.....	124
11.3 List of primers.....	125
11.4 Vector information	128
11.5 List of organisms	130
11.6 Chromosome location of MEP and carotenogenesis genes of <i>P. tricornutum</i>	132
Acknowledgements	134
Curriculum vitae	135

Abbreviations

ABA: Abscisic acid

ALE: Adaptative laboratory evolution

bp: Base pair

CHY-B: β -Carotene hydroxylase

CMK: 4-(Cytidine 5'-diphospho)-2-C-methyl-D- erythritol kinase

CMS: 4-Diphosphocytidyl-2-C-methyl-D-erythritol synthase

COPAS: Complex Object Parametric Analyzer and Sorter

CRTISO: Carotenoid isomerase

CTAB: Cetrimonium bromide

CYP: Cytochrome P450

DDE: Diadinoxanthin de-epoxidase

DEP: Diatoxanthin epoxidase

dNTP: Deoxyribonucleotide triphosphate

DXR: 1-Deoxy-D-xylulose 5-phosphate reductoisomerase.

DXS: 1-Deoxy-D-xylulose 5-phosphate synthase

DMAPP: Dimethylallyl pyrophosphate

E: Einstein

EDTA: Ethylenediaminetetraacetic acid

ER: Endoplasmic reticulum

FAD: Flavin adenosine dinucleotide

FCP: Fucoxanthin-chlorophyll a/c protein

FDA: Fluorescein diacetate

G: L-Guluronic acid

Gent^r: Gentamycine resistant

GGDP: Geranylgeranyl diphosphate

GGPPS: Geranylgeranyl diphosphate synthase

G3P: Glyceraldehyde 3-phosphate

HDR: (E)-4-Hydroxy-3-methylbut-2-enyl diphosphate reductase

HDS: (E)-4-Hydroxy-3-methylbut-2-enyl diphosphate synthase

HPLC: High throughput liquid chromatography

IDI: Isopentenyl phosphate

Kb: Kilo base pair

LCYB: Lycopene b-cyclase

LHC: Light harvesting complex

M: D-Mannuronic acid

MCS: 2-C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase.

MEP: 2-C-methyl-D-erythritol 4-phosphate

MS: Murashige and Skoog Basal Salt Mixture

MVA: Mevalonate

Nat: Nourseothricine

NH: Non-heme

NLR: Nano liter reactor

PCI: Phenol-Chloroform-Isoamyl alcohol

PCR: Polymerase Chain Reaction

PDS: Phytoene desaturase

PPP: Pentose phosphate pathway

PSY: Phytoene synthase

ROS: Reactive oxygen species

RT: Reverse transcriptase

Sul: Sulfadiazine

Tet^r: Tetracycline resistant

VDE: Violaxanthin de-epoxidase

VDL: Violaxanthin de-epoxidase-like

WT: Wild type

ZDS: ζ-Carotene desaturase

ZEP: Zeaxanthin epoxidase

Z-ISO: 15-Cis-ζ-carotene isomerase

1. General background

1.1 Diatoms

Diatoms are unicellular eukaryotic microalgae. They are a very diverse group with more than 100.000 species that can be found in soil, freshwater and marine environments. They are widely distributed in the oceans, being present from the equator to the polar region. Diatoms are very well-adapted to fluctuating light conditions and some species can also resist extreme conditions of salinity, temperature or pH. They are one of the dominant groups present in phytoplankton and contribute to 20% of global carbon fixation¹.

Diatoms belong to the supergroup Chromalveolata, more specifically to the subgroup of Heterokonts or Stramenopiles, together with brown and golden algae (**Figure 1**). Heterokonts arose by a secondary endosymbiosis event in which a heterotrophic eukaryote engulfed a red alga (**Figure 2**). In addition, their genome shows that many gene transfers have occurred between diatoms and bacteria and green algae. Their complex genome with many genes of different origins explains the peculiar metabolism and cellular capabilities of diatoms. For example, diatoms present a complete urea cycle, normally present in ureotelic animals and several two-component signaling proteins typical of bacteria. Other differences in metabolism is for example that part of the glycolysis pathway takes places in the mitochondria and that they store carbon as chrysolaminarin, a soluble polymer stored inside vacuoles^{1,2}.

Most diatoms are autotrophic, but there are also mixotrophic and obligate heterotrophic diatoms. For example, *Nitzschia alba* is a heterotrophic diatom that lost its plastid and can grow in the presence of glucose¹.

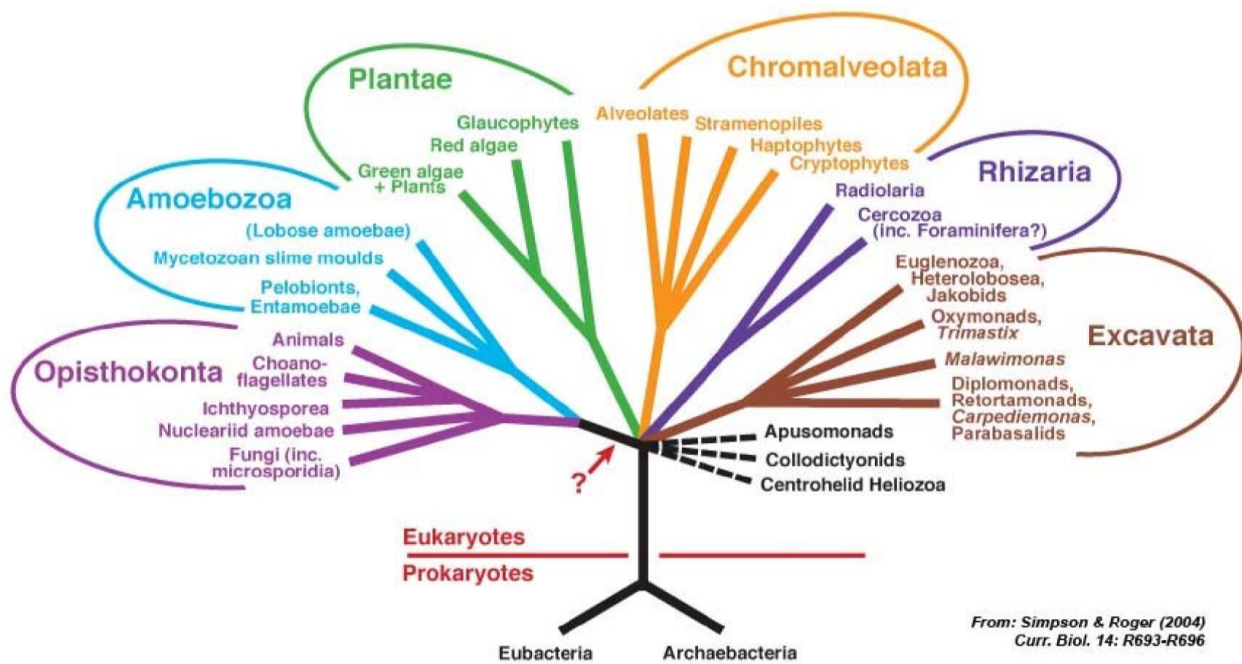


Figure 1. Tree of life. The main groups of the domain eukarya are represented. Diatoms belong to the superphylum stramenopiles or heterokonta, that in turn belongs to the supergroup Chromalveolata, together with cryptophytes, haptophytes and alveolates. In another group, the kingdom Plantae or Archaeplastida, are included green algae, plants, red algae and glaucophytes. Image from Simpson and Roger³.

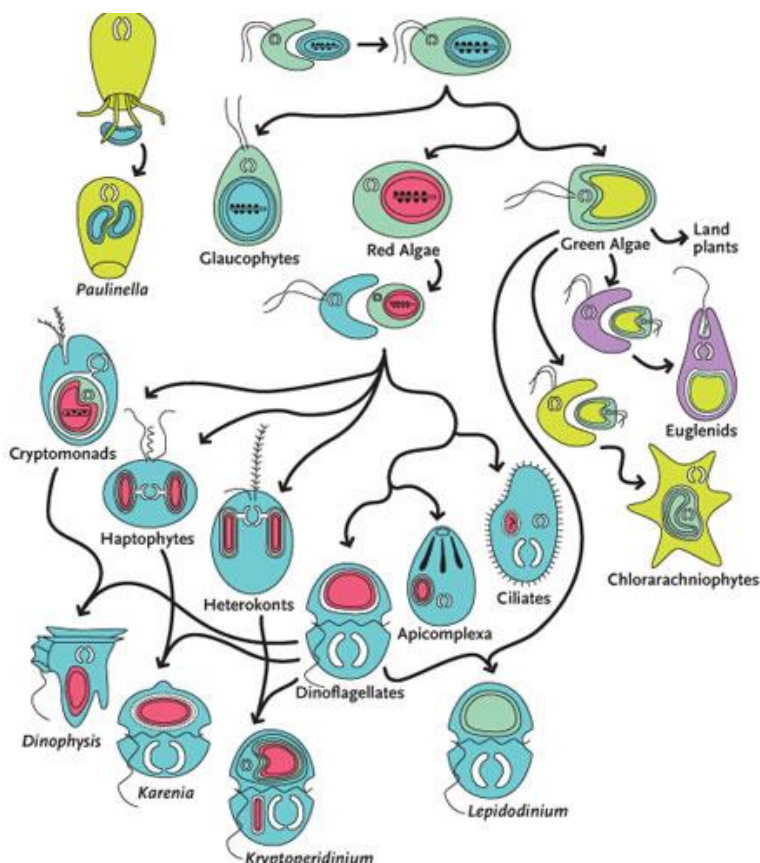


Figure 2. Endosymbiosis events. In the top of the figure the first endosymbiosis of a cyanobacteria by a eukaryote is shown. Three main groups derived from this event: glaucophytes, red algae and green algae. In some of these groups, also secondary and even tertiary endosymbiosis events took place. For instance, the secondary endosymbiosis of a red algae by a eukaryote gave rise to heterokonts. Image from Keeling⁴.

Autotrophic diatoms present some differences related to photosynthesis compared to green algae and plants. For example, diatoms present different pigments, such as chlorophyll c instead of chlorophyll b, and fucoxanthin as the main carotenoid. In addition to the violaxanthin-zeaxanthin cycle, diatoms present an additional xanthophyll cycle to dissipate excess light that could cause photodamage, the diadinoxanthin-diatoxanthin cycle. Unlike plants and green algae, diatoms do not possess state transition mechanisms that redistribute the excitation energy between the two photosystems. Other differences relate to the different architecture of their thylakoid membranes (e.g., they are not differentiated into grana and lamellae and instead form stacks of three) and affect redox signaling. Finally, the regulation of the Calvin cycle is different and the oxidative pentose phosphate cycle is not present in the plastid¹.

Due to its endosymbiosis origin, the plastid is surrounded by four membranes¹. Another feature of diatoms is the presence of an amorphous silica cell wall, that not only protects them against predators and UV radiations, but is also important in the carbon concentrating mechanism⁵. Some diatoms such as *Phaeodactylum tricorutum* do not deposit silica in their cell wall. In addition to their protective shell, other traits that contribute to the success of diatoms in the oceans are the presence of a water-filled central vacuole that controls their buoyancy and has a high capacity for the storage of nutrients⁶. In addition, the distinctive pigments of diatoms, fucoxanthin and chlorophyll c can absorb light in the blue-green region, which is not efficiently absorbed by chlorophyll a and b. This spectrum region penetrates deeper into the sea, so their absorption constitutes an advantage in this habitat⁷.

The most important applications of diatoms in biotechnology are the production of biodiesel and unsaturated fatty acids such as eicosapentaenoic acid for the food industry. Diatoms are also a source of bioactive compounds, such as the carotenoid fucoxanthin, with interesting applications for the pharmaceutical industry¹.

1.2 Carotenoid pathway in diatoms.

Carotenoids are pigments that are present in plants, algae, some fungi and bacteria that play an important role in photosynthesis participating in the absorption of light, photoprotection and cell signaling. They are also responsible for rendering color, which is determined by the number of double bonds. Animals do not synthesize carotenoids and need to acquire them in their diets. Some carotenoids such as β -carotene are precursors of vitamin A, an essential nutrient in humans⁸.

Carotenoids are tetraterpenoids (40C) that are derived from isoprenoid molecules (5C). The 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway is the carotenoid precursor pathway that forms the isoprenoid units starting from glycolysis intermediates. There are two pools of isoprenoids in diatoms: the cytosolic pool, derived from the mevalonate (MVA) pathway, and the plastidic pool formed by the MEP pathway. The MEP pathway reactions are described in **Figure 3**. The cytosolic isoprenoid pool is used for the formation of sterols, polyprenyls for protein prenylation and ubiquinone, while the plastid pool is used for the formation of carotenoids, the phytyl chain of chlorophylls and polyprenyls for plastoquinones (**Figure 4**). Transport of isoprenoids between the different compartments has been detected in plants, although export of precursors to the cytosol is most common⁹.

The common carotenogenesis pathway that is conserved across different lineages of algae, land plants and cyanobacteria is now described. It begins with the condensation of two geranylgeranyl diphosphate (GGDP) molecules to produce phytoene and it is catalyzed by phytoene synthase (PSY). Then, after four desaturation reactions catalyzed by phytoene desaturase (PDS) and ζ -carotene desaturase (ZDS), pro-lycopene is formed. Pro-lycopene is isomerized to form lycopene and two cyclization reactions at both ends of lycopene yield β -carotene. After that, oxygen atoms are introduced into β -carotene by hydroxylation to produce zeaxanthin. The enzyme that catalyzes this step is not known in diatoms¹⁰.

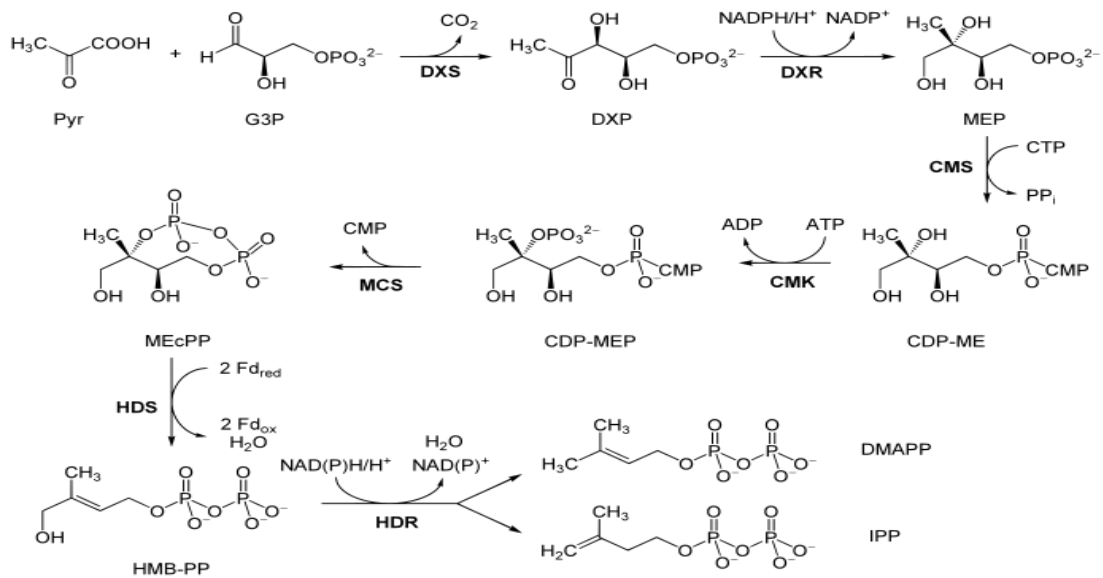


Figure 3. The Methylerythritol Phosphate (MEP) pathway. The MEP pathway produces the isoprenoid precursors of carotenoids, isopentenyl phosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). In the MEP pathway, the condensation of pyruvate and glyceraldehyde 3-phosphate (G3P) produces 1-deoxy-D-xylulose 5-phosphate (DOXP). The reductive isomerization of DOXP produces MEP. After that, a cytidyl group is added to MEP to form 4-(cytidine 5'-diphospho)-2-C-methyl-D erythritol (CDP-ME). Then CDP-ME is phosphorylated and during the synthesis of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP) the cytidyl group is eliminated. After two further reactions based on eliminations coupled with reduction steps, IPP and DMAPP are formed. These isoprenoids are interconvertible thanks to the action of the enzyme IPP: DMAPP isomerase (IDI)¹⁰. The names of the enzymes are in bold next to the arrows and the full name can be consulted in the Abbreviation section. Image from Wikipedia¹¹.

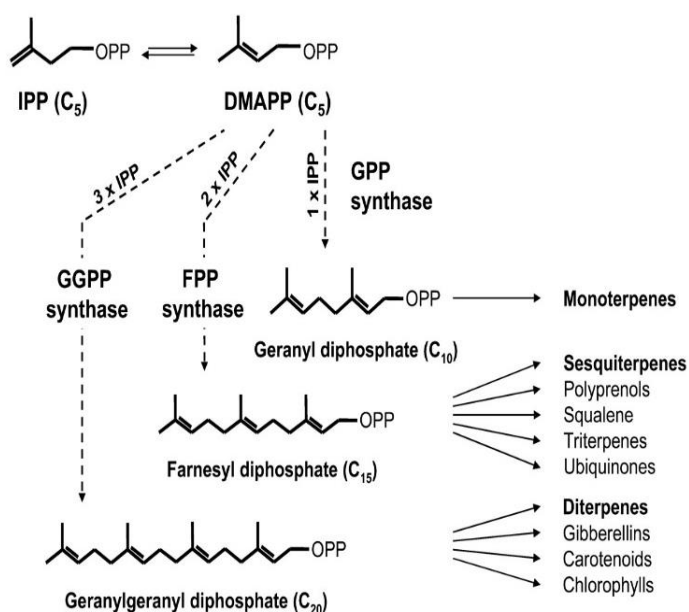


Figure 4. Formation of different terpenes. The end products of the MEP pathway are the building blocks for a wide array of terpenes. For instance, the sequential addition of four IPP molecules to a DMAPP molecule yields geranylgeranyl diphosphate (GGDP), that is the precursor of carotenoids, diterpenes and the phytol side chain of chlorophyll. This reaction is catalyzed by geranylgeranyl diphosphate synthase (GGPPS). Image modified from Schmidt et al¹².

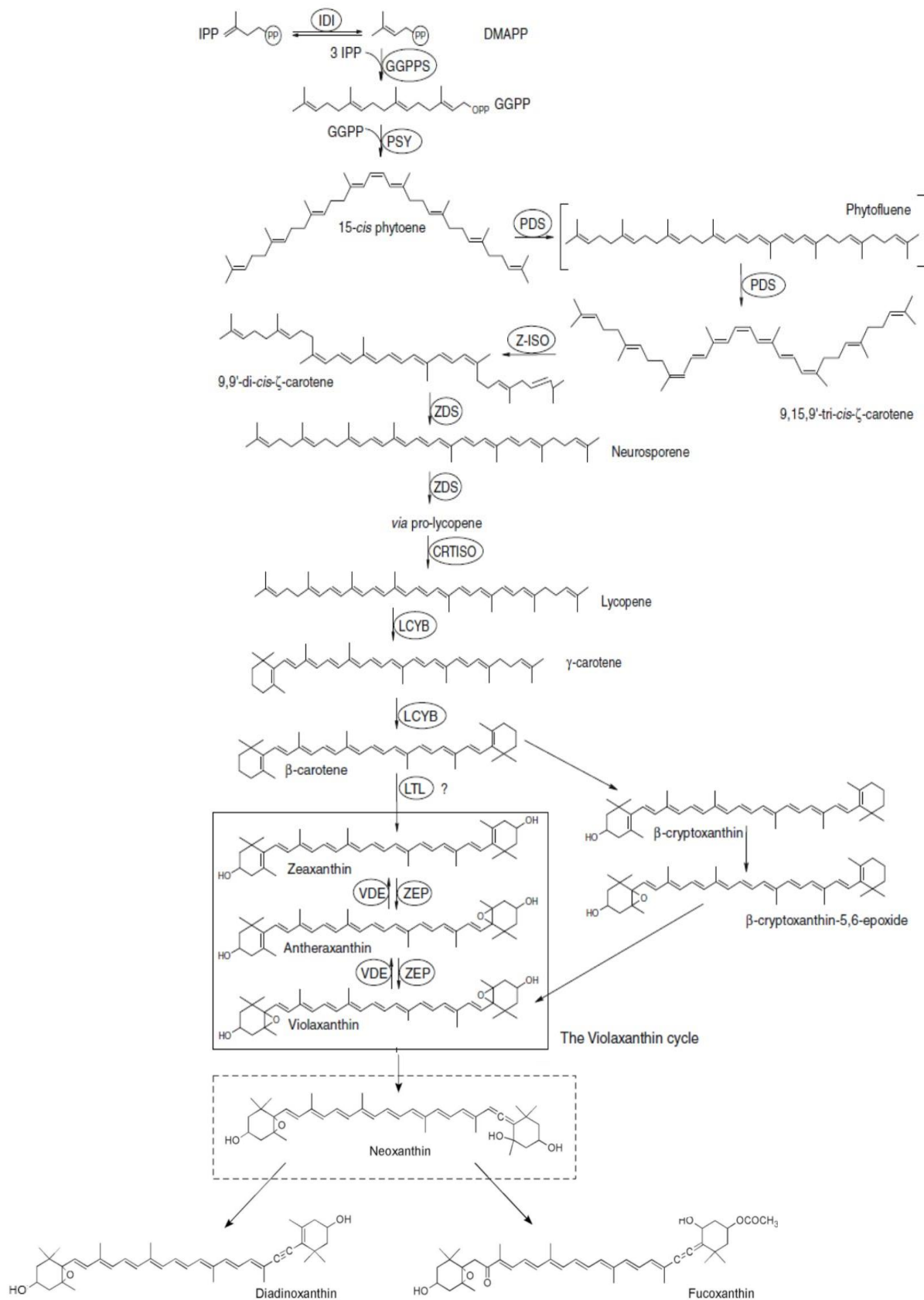


Figure 5. Carotenoid biosynthesis pathway in diatoms. Enzymes are shown inside a circle and their full name can be consulted in the Abbreviation section. Image adapted from Bertrand¹⁰ and Dambek et al¹³.

From this basic pathway, different organisms build additional carotenoids by specific reactions. For example, in some organisms from lycopene not only β -carotene is formed, but also α -carotene. This is the case of green algae, plants, cryptophytes and macrophytic red algae¹⁴.

In diatoms, the sequential epoxidation of both rings of zeaxanthin produces violaxanthin. Zeaxanthin epoxidase (ZEP) catalyzes this reaction. After this step, the sequence of reactions is not totally clear, but based on the structure of the carotenoids, the following order has been proposed. Violaxanthin is then converted into neoxanthin¹⁰. The enzyme violaxanthin de-epoxidase-like1 (VDL1) has been recently identified to catalyze this step¹⁵. From neoxanthin, diadinoxanthin and fucoxanthin are formed. The enzymes that participate in these reactions have not been identified. Diadinoxanthin can be transformed into diatoxanthin by de-epoxidation. This reaction is catalyzed by diadinoxanthin de-epoxidase (DDE)¹⁰. The carotenoid pathway of diatoms can be observed in **Figure 5**.

1.3 Transport of enzymes to the plastid

All the enzymes of the carotenogenesis and the MEP pathways are encoded by nuclear genes whose products are transported through the four membranes that enclose the plastid. These membranes are the result of the secondary endosymbiosis event that gave rise to diatoms. The outermost membrane is part of the endomembrane system of the host, that is continuous with the endoplasmic reticulum (ER), the second outermost membrane is derived from the plasma membrane of the red algae and it is called periplastidial membrane. Finally, the two inner membranes are the membranes of the plastid of the red algae¹⁶.

Plastid-targeted proteins in diatoms have a signal and a target peptide in their N-terminal region¹⁶. The import of plastid-targeted proteins is explained in **Figure 6**. There are prediction tools to detect the signal and the target peptides in proteins, e.g. AsaFind for the signal peptide and TargetP for the target peptide^{17,18}.

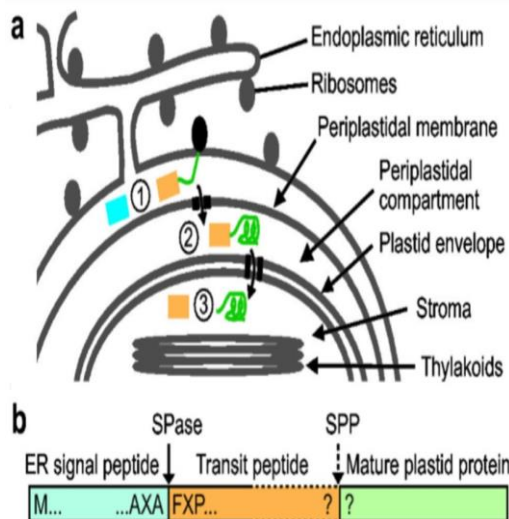


Figure 6. Import of plastid-targeted proteins. a) The diatom plastid is enclosed by 4 membranes. Proteins are synthesized in the cytosol and are co-translationally inserted in the ER lumen. There, the signal peptide is cleaved off by a peptidase (1). Afterwards, the target peptide leads the protein through the periplastidal (2) and the two inner membranes until the stroma (3). The target peptide is finally removed by a peptidase (3). b) Structure of a plastid -targeted protein precursor, that after the removal of the signal and target peptide becomes a mature protein¹⁶. Abbreviations: SPase, signal peptidase; SPP, stromal processing peptidase. Image from Huesgel et al¹⁶.

1.4 Regulation of the carotenoid pathway

The regulation of the MEP and the carotenoid pathway is poorly understood in diatoms, so the regulation in plants will be described when data is not available in diatoms, although one has to be cautious since there are many peculiarities of plants that are different in diatoms, so the regulatory information cannot be directly extrapolated. The regulatory process in plants is not fully understood as well.

In plants, the DXS enzyme that catalyzes the initial reaction of the MEP pathway is under regulation of the final products of the MEP pathway (IPP and DHAP) by feedback inhibition¹⁹. Redox control mediates its effect at the post-translational level and is present in the MEP pathway in plants, where DXR, HDS and HDR are controlled by thioredoxin, a protein oxidoreductase involved in the regulation of many proteins, some of them involved in photosynthesis. Thioredoxin is reduced when the light reactions of photosynthesis take place and, in this form, it upregulates different target proteins by reducing specific disulfide bonds. In addition, PDS and ZDS use plastoquinone as their hydrogen acceptor in plants. Therefore, their activities are dependent on the redox state of the photosynthetic electron transport chain, so light plays an important role in their activity⁸. Also, other plant enzymes like β -carotene hydroxylase (CHY-B) and zeaxanthin epoxidase are under redox control by the plastoquinone pool²⁰. In addition, the

redox state of the cell can affect the activity of several redox enzymes of the MEP and carotenoid pathway, since some enzymes require the flavin adenosine dinucleotide (FAD) cofactors⁸.

In the carotenoid pathway of diatoms, there are two xanthophyll cycles that play important roles in photoprotection: the diadinoxanthin and the violaxanthin cycles. The first one is the most important xanthophyll cycle in diatoms, while the latter only operates under prolonged high light conditions. Light controls the expression levels of genes participating in the xanthophyll cycles, such as diatoxanthin epoxidase (*dep*), diadinoxanthin de-epoxidase (*dde*), violaxanthin de-epoxidase (*vde*) and zeaxanthin epoxidase (*zep*) genes. High light induces the de-epoxidation of diadinoxanthin or violaxanthin to form diatoxanthin and zeaxanthin, respectively. When the light stress disappears, these reactions are reversed by the epoxidation of diatoxanthin and zeaxanthin. Also, changes in trans-thylakoid pH triggered by light affect the activity of their gene products^{8,21}.

It is important to remember that the levels of carotenoids not only depend on their synthesis, but also on their degradation. Degradation of carotenoids can occur in plants by photo-oxidation or by enzymatic oxidation. In plants, the formation of the photo-oxidation product of β -carotene, β -carotene-5,6-epoxide, correlates with light intensity. Enzymes might mediate the degradation of carotenoids in an unspecific way by oxidation, for example carried out by peroxidases and lipoxygenases, or in a specific way. In the latter case, carotenoid cleavage oxygenases can cleave particular carotenoids in plants to form apocarotenoids that have regulatory functions or are important in environmental responses^{8,22}.

In addition to the synthesis and degradation of carotenoids, storage has also a great influence in the levels of carotenoids. Plant chromoplasts have carotenoid-sequestering structures in which carotenoids are accumulated in a matrix of lipoproteins. An interesting case is the orange cauliflower, in which chromoplasts are formed due to the presence of a mutation in the *or* gene that encodes a DnaJ-like protein involved in targeting substrates to chaperones. This mutation triggers the accumulation of high levels of β -carotene compared to the wild type uncolored version of the cauliflower⁸.

2. Materials and methods

2.1 Cultivation of the organisms

2.1.1 Cultivation of *Phaeodactylum tricornutum*

The experiments were carried out with the diatom *Phaeodactylum tricornutum* UTEX 646 from the culture algae collection from the University of Texas. The diatom cultures were incubated at 18 ± 1 °C and 40 ± 2 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ in an agitator at 120 rpm and a light-dark cycle of 16:8 h, respectively. Cultures not used for experiment were reinoculated every 20 days. *P. tricornutum* has a duplication time of around 1,5 days and after inoculation it takes 2-3 days to reach exponential phase. All experiments were performed in this phase. *P. tricornutum* was grown autotrophically in the ASP medium, according to a modified version of Provasoli et al., 1957.

ASP medium (pH 7.7): NaCl (86 mM), Tris (4 mM), KCl (21 mM), $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$ (8.1 mM), NaNO_3 (11.8 mM), K_2HPO_4 (0.58 mM), H_3BO_3 (0.16 mM). After autoclaving, the following solutions were added: $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (2.72 mM), Fe(III) solution and trace element solution A.

The Fe(III) solution is composed of: EDTA (12.5 mM) and $\text{FeCl}_3 \cdot 6 \text{H}_2\text{O}$ (12.3 mM). The trace element solution A (pH 5.5) is made of: EDTA (80.6 mM), $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ (20.5 mM), ZnCl_2 (8.3 mM) and trace element solution B ($\text{CaCl}_2 \cdot 6 \text{H}_2\text{O}$ (1 mM), $\text{Na}_2\text{MoO}_4 \cdot 2 \text{H}_2\text{O}$ (0.5 mM), $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ (0.35 mM).

When antibiotics were used for selection, the following concentrations were used: Nourseothricine (Nat) (100 $\mu\text{g/ml}$), zeocin (100 $\mu\text{g/ml}$).

2.1.2 Cultivation of *Cyclotella* and *Chlorella*

The cultivation of *Cyclotella* and *Chlorella* took place at Fgen (Basel, Switzerland). In a first attempt, the cultures were grown by direct exposition to natural light (sunny days) at 21°C for 3 days under agitation. In two other rounds, the cultures were grown at 50 $\mu\text{E m}^{-2} \text{s}^{-1}$ using a LED lamp and at 21°C for 3 days under agitation.

ASP medium supplemented with 2 mM Silica was used for the cultivation of *Cyclotella*. N8 Medium was used for the cultivation of *Chlorella*.

N8 Medium: KH_2PO_4 (0.74 g/L), $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ (0.01 g/L), KNO_3 (1 g/L), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (0.1 g/L), $\text{Na}_2\text{HPO}_2 \cdot 12\text{H}_2\text{O}$ (0.26 g/L), Fe(III)-EDTA (9.2 mg/L), AZ-Solution 30X (1:30 dilution).

AZ- Solution 30X: $\text{Al}_2(\text{SO}_4)_3 \cdot 18\text{H}_2\text{O}$ (107.3 mg/L), KI (28 mg/L), KBr (28 mg/L), LiCl (28 mg/L), $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ (389 mg/L), H_3BO_3 (614 mg/L), $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ (55 mg/L), $\text{NiSO}_4 \cdot 6\text{H}_2\text{O}$ (55 mg/L), $\text{CoSO}_4 \cdot 6\text{H}_2\text{O}$ (55 mg/L), $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ (96 mg/L).

The Fe (III)-EDTA and the AZ- Solution were sterilized by filtration through a syringe and the rest of the components were autoclaved, each one separately to avoid interactions.

2.1.3 Cultivation of *Escherichia coli*

Escherichia coli (XL-1) was grown at 37°C under agitation (170 rpm). When using 5 ml tubes for culture, these were placed at a certain angle (45 °) for better agitation and aeration. *E. coli* was grown in LB medium.

LB medium (pH 7): Tryptone (10 g/l), Yeast extract (5 g/l), NaCl (5 g/l).

2.1.4 Cultivation of *Agrobacterium tumefaciens*

The strain of *Agrobacterium tumefaciens* GV3101 provided by Dr. Markus Fauth (Goethe University Frankfurt, Germany) contained the plasmid pSoup and pMP90 (Hellens et al., 2000), with the selection marker genes tetracycline and gentamycin, respectively. *A. tumefaciens* was grown in YEP medium at 28°C in

darkness under agitation.

YEP medium (pH 7): Peptone (10 g/l), Yeast extract (10 g/l), NaCl (5 g/l).

2.1.5 Cultivation of *Arabidopsis thaliana*

Arabidopsis thaliana was grown under controlled conditions of temperature (19-21°C) and light intensity (60-80 $\mu\text{E}/\text{m}^2\cdot\text{s}$) with a light cycle of 9 h light/ 15 h darkness. The conditions for the plants to flower were the same but with a longer photoperiod (16 h light/ 8 h dark). When planting seeds in soil, these were incubated for 2-3 days at 4°C in the dark for stratification of the seeds. The seeds of the triple knock-out mutant *chy1chy2lut5* were provided by Prof. Sandmann (Goethe University Frankfurt, Germany), that received them as a gift from Prof. Bassi (University of Verona, Italy).

2.2 Isolation and preparation of genetic material

2.2.1 RNA extraction from *P. tricornutum*

For the extraction of RNA, the cells were always harvested in the morning 1h after the onset of light. Since some genes of the carotenoid pathway are light-regulated, it is important to harvest the cells when they have had already at least one hour of light. Exponentially growing cells were harvested by centrifugation (5000 rpm, 5 min, 4°C) and cells are then broken by sonication. For that purpose, glass pearls were added (0.1/0.3 mm, 50:50) to the cell pellet and 0.5 ml of Phenol-Chloroform-Isoamyl alcohol (PCI) mix was added to the sample. The sonication took place inside a Tissue Lyser for 16 min at 50 Hz.

After that, a short spin in a hand-centrifuge was carried out and 1 ml of RNase-free water was added to the sample. After resuspending, the samples were let to rest for 5 min and then were centrifuged at 13000 rpm for 3 min. Then, the upper phase was kept, it was fill up to 1 ml with water and 1 ml of PCI mix was added.

After centrifuging (13.000 rpm, 3 min), the upper phase was collected and 1 ml of chloroform was added to remove possible rests of phenol. The sample was centrifuged (13.000 rpm, 3 min) and the upper aqueous phase was collected. Then, the sample was incubated with 1 volume of LiCl₂ (5 M) for 1h at -20°C for RNA precipitation. After that, the tube was centrifuged (13.000 rpm, 20 min, 4°C). The precipitate was then washed twice with ethanol 70%, leaving an incubation time of 5 min. Then the pellet was dried at 42°C to evaporate the rests of ethanol. After drying, the pellet was resuspended in 42 µl of water, 5 µl of 10X DNaseI Buffer and 3 µl of DNaseI (Thermo Scientific) and this mixture was incubated for 3h at 37°C. After this, the sample was transferred to a new tube and the whole procedure of Phenol-Chloroform extraction was repeated.

2.2.2 RNA isolation from plant tissue

A few leaves of *A. thaliana* (100 mg) were collected in a tube and frozen immediately in N₂ (l). The disruption of the cells was done with a homogenizer (Janke and Kunkel) and then, 500 µl of TRI solution was added. This step was repeated. Then this mixture was centrifuged (8000 g, 10 min, 4°C). The supernatant was collected and 200 µl of chloroform were added. After an incubation time of 15 min, the sample was centrifuged (8000 g, 15 min, 4°C). The upper aqueous phase was taken and 500 µl of isopropanol were added to the sample and was incubated for 15 min at room temperature in order to precipitate the nucleic acids. After centrifugation (8000 g, 10 min, 4°C), the pellet was washed twice with 75% ethanol, with 5 min of incubation time. Then the pellet was dried at 42°C. A further step for degradation of DNA was carried out by the addition of 3 µl of DNaseI (Thermo Scientific), 5 µl of DNaseI buffer and 42 µl of water. The liquid-liquid extraction was performed again and at the end the pellet was resuspended in 15 µl of water.

TRI solution (pH 5): 0.4 M of ammonium thiocyanate, 0.8 M of guanidinium thiocyanate, 0.1 M of sodium acetate, 5% of glycerol (w/v) and 4.09 M of phenol.

2.2.3 Preparation of cDNA from mRNA

For the synthesis of the first strand of the cDNA, 1 µg of RNA was incubated together with 1 µl of oligo-dT18 and RNase-free water at 65°C for 5 minutes in a thermoblock. Then, 4 µl of RNase buffer, 2 µl of dNTPs (10 mM) and 1 µl of Reverse Transcriptase (RT) (Thermo Scientific) were added. The final volume was 20 µl. The sample was incubated for 1 h at 42°C. Then, the RT was deactivated by heat (70°C, 10 min). The second strand of the cDNA was then synthesized by a PCR reaction with specific primers for the desired gene.

2.2.4 DNA isolation from *P. tricornutum*

The isolation of DNA started with the harvesting of 50 ml of a densely grown culture of cells by centrifugation (5000 g, 3 min). Then the pellet was resuspended in 500 µl of cetrimonium bromide (CTAB) buffer. Glass beads (0.1/ 0.3 mm, 50:50) were added to the sample for the subsequent mechanical disruption of the cells by homogenization using a Tissue Lyser (Qiagen) (10 min, 50 Hz). After homogenization, 500 µl of chloroform: Isoamyl alcohol (24:1) were added and then the samples were centrifuged (13000 rpm, 2 min). The upper phase was transferred to a new tube and a volume of isopropanol was added to precipitate the nucleic acids. The samples were incubated at -20°C for 10 min, and then they were centrifuged (13000 rpm, 3 min). After that, the pellet was washed twice with 70% ethanol. Finally, the pellet was dried at 37°C and resuspended in 30 µl of Tris Buffer (10 mM).

CTAB buffer: 100 mM Tris, 1.4 M NaCl, 20 mM EDTA, 1% CTAB, 0.2% Mercaptoethanol.

2.2.5 DNA isolation from plant tissue

The isolation protocol of DNA from plant leaves is the same as in section 2.2.4, except for the homogenization procedure, in which a different mechanical

homogenizer (Janke and Kunkel) was used.

2.2.6 Plasmid extraction from the cells

A 5 ml culture was inoculated with either *E.coli* or *Agrobacterium*, and incubated overnight at 37°C or 28°C, respectively, under agitation (170 rpm). For the plasmid isolation, the GeneJet Plasmid Miniprep Kit (Fermentas) was used.

Firstly, the cells were harvested by centrifugation (15000 rpm, 10 s) and then the pelleted cells were resuspended in 250 µl of the Resuspension Solution (50 mM glucose, 10 mM EDTA, 25 mM Tris pH 8), containing RNaseA. Subsequently, 250 µl of Lysis solution (0.2 N NaOH, 1% SDS) was added and mixed by inversion. Then, 350 µl of Neutralization solution (3 M Potassium Acetate pH 6 in glacial acetate) was immediately added and mixed by inversion. Then the sample was centrifuged (15000 rpm, 5 min). The supernatant was then transferred to a GeneJET spin column and it was centrifuged (15000 rpm, 1 min). The flow-through was discarded and the column was then washed twice with 500 µl of ethanol 70%. Finally, the column was transferred to a new tube and the plasmid DNA bound to the column was eluted by the addition of 25 µl of Tris pH 8. The sample was incubated for 2 min and centrifuged (15000 rpm, 2 min). The eluate was placed back on top of the column and centrifuged again. Then the sample was stored at -20°C.

2.2.7 Amplification of DNA by PCR

The Polymerase Chain Reaction (PCR) method was used to specifically amplify segments of DNA. The components for carrying out this reaction when using Taq polymerase were: 1 µl of cDNA or gDNA, 1 µl of forward primer (10 mM), 1 µl of reverse primer (10 mM) and 7.5 µl of Taq master mix. For the sequences of the primers, see Appendix. When using Phusion Polymerase, the components were the same but instead of using the Taq master mix, the Phusion master mix was used (Thermo Scientific Phusion Flash High-Fidelity PCR Master Mix).

Taq master mix: 100 µl of Taq polymerase, 160 µl of 5 mM dNTPs, 130 µl of 3 mM betain, 150 µl of Xylenol Orange Solution (0,25% Xylenol Orange, 50% Glycerol), 150 µl of 10X buffer PCR (100 mM Tris, 40 mM MgCl₂, 500 mM KCl, pH 8.3). Final volume 750 µl.

PCR conditions using Taq polymerase:

-Initial Denaturation: 95°C, 5 min

-X Cycles (X=20-35):

-Denaturation step: 95°C, 30 s

-Annealing step: Annealing T (*), 30 s

-Extension step by DNA polymerase: 72°C, extension time (**)

-Final Extension: 72°C, 10 min

*The annealing temperature is usually 2°C less than the lower melting temperature of the primer pair.

**The extension time depends on the length of the gene to amplify. The rate of amplification of Taq polymerase is 1Kb/min)

PCR conditions using Phusion polymerase:

-Initial Denaturation: 98°C, 30 s

-X Cycles (X=20-35):

-Denaturation step: 98°C, 10 s

-Annealing step: Annealing T, 30 s

-Extension step: 72°C, extension time (*)

-Final Extension: 72°C, 10 min

*The extension time depends on the length of the gene to amplify. The rate of amplification of Phusion polymerase is 1 Kbp /30 s.

For every PCR, a negative control was always included, in which the same reactives were added to the PCR mixture, with exception of the gDNA or cDNA, that was replaced by water. The same PCR program was applied to the samples and to their respective negative control.

When a positive control was included, this contained the same reagents of the PCR mixture, but in this case around 10 ng of plasmid DNA, that had been previously tested to contain the gene of interest, was included. The same PCR program was applied to the positive control and to the samples.

2.2.8 Colony PCR

Colony PCR is a PCR in which the DNA is obtained directly from the colonies. A single colony was picked with the tip of the pipette and then spotted into a new culture plate with the corresponding antibiotic. The same tip was then submerged in the tube containing the PCR components. The plate with the colonies was incubated overnight at the appropriate temperature for the cultivation of the organism.

2.2.9 Restriction Assay

The vector was linearized by a restriction assay with one or two restriction enzymes (Thermo Scientific). 1 µg of vector was mixed with 1.5 µl of 10X of the appropriate digestion buffer, 1 µl of each restriction enzyme and water up to 15 µl of total volume. These reaction components were then incubated at 37°C for 14 h and after that, the enzyme was deactivated (80°C, 5 min). A control should be done with undigested vector. Then, part of the digested vector was run in an agarose gel to check if the sizes of the fragments were correct and to quantify them. 6 µl of the GeneRuler 1 kb DNA ladder (Thermo Scientific) was always loaded in all agarose gels for size and concentration determination.

2.3 Cloning by the hot fusion method

The principle of the hot fusion method is explained in Changlin et al., 2014. Firstly, the insert was amplified by PCR (Section 2.2.7) using Phusion polymerase (Thermo Scientific). The vector was linearized by a restriction assay with two

restriction enzymes that cut where the insert will be inserted (Section 2.2.9). In the Hot fusion PCR reaction, the insert and the vector were incubated together in the presence of the T5 exonuclease and the DNA polymerase. The incubation was for 1 h at 50°C, followed by a cool-down step up to 20°C at a rate of 0.1°C per second.

-Hot fusion PCR components: 0.015 pmol of vector, 0.035 pmol of insert, 2.5 µl of reaction buffer. Final volume of 5 µl.

-Reaction buffer: 100 µl of ISO buffer 5X, 6.25 µl of 2 U/ µl of Phusion polymerase (New England Biolabs), 0.2 µl of 10 units/ µl of T5 exonuclease (Epicentre). Final volume of 250 µl. This buffer was aliquoted and stored at -20°C.

-Isothermal reaction buffer (ISO) 5X: 25 % PEG-8000, Tris-HCl (500 mM, pH 7.5), MgCl₂ (50 mM), DTT (50 mM), dNTP (1 mM, for each), NAD (5 mM). This buffer was aliquoted and stored at -20°C.

After the hot fusion PCR, 1 µl of this mix was used for the transformation of *Escherichia coli* (Section 2.4.2). After that, colony PCR was carried out to find out if the selected colonies contained the inserted gene (Section 2.2.8). A 5 ml culture was inoculated with a positive colony in order to extract finally the recombinant vector by Miniprep (section 2.2.6).

2.4 Transformation of the organisms

2.4.1 Production of competent *E. coli* XL-1 cells

An overnight culture of *E. coli* cells (with 12.5 µg/ ml tetracycline) was inoculated into 200 ml of LB medium and the cells were grown at 37°C until exponential phase (optical density value at 600 nm of 0.4-0.6). Then the cells were sedimented by centrifugation (3500 g, 10 min, 4°C).

After cell harvesting, the cells were conferred competence by incubation with 40

ml of Transformation Buffer I (30 mM Potassium Acetate, 50 mM MnCl₂, 400 mM KCl, 10 mM CaCl₂ and 15% glycerine (v/v), pH 5.8, filtered) for 2 h in ice. Subsequently, the cells were sedimented by centrifugation (650 g, 5 min, 4°C) and resuspended in 8 ml of Transformation Buffer 2 (10 mM MOPS, 10 mM KCl, 75 mM CaCl₂, 15% glycerine (v/v), pH 7, filtered). The competent cells were aliquoted, immediately frozen in N₂ (l) and stored at -80°C.

2.4.2 Transformation of *E. coli* by the heat shock method

For the transformation of *E. coli*, 200 µl of competent cells were mixed with 200 µl of Transformation Buffer A (100 mM CaCl₂, 50 mM MgCl₂). Then, 200 µl of this mixture were incubated with 1 µl of the hot fusion PCR reaction, containing the recombinant vector, for 30 min on ice. Then the cells were exposed to a heat shock (42°C for 60-90 s), so the plasmid could enter the cells. After that, they were shortly cooled on ice and incubated with 1 ml of LB medium for 30 min at 37°C and 1000 rpm. Then the cells were pelleted by centrifugation (15000 rpm, 5s) and about 200 µl of supernatant was left to resuspend the cells, which were then plated in LB plates with the appropriate antibiotic. Ampicillin (100 µg/ml) was used when using the pPhaNR plasmid and kanamycin (50 µg/ml) was used for the pGreen plasmid. The plates were then incubated at 37°C for 16 h.

2.4.3 Production of competent *Agrobacterium tumefaciens*

A 5 ml pre-inoculum of *A. tumefaciens* in YEP medium was incubated for 20 h at 28°C in the dark and with an agitation of 160 rpm. This inoculum was used to inoculate a 250 ml culture of YEP medium with 0.5% (w/v) glucose. The culture was grown at 28°C until an optical density at 600 nm of 0.5-0.6. Cells were then harvested by centrifugation (4000 g, 5 min, 4°C) and resuspended in 2.5 ml of 150 mM NaCl. Then, the cell suspension was left in ice for 15 min, followed by a centrifugation step (4000 g, 5 min, 4°C). The cell pellet was resuspended in 2.5 ml of 20 mM CaCl₂. The competent cells were finally aliquoted (100 µl/ tube) and submerged in N₂(l) before storage at -80°C.

2.4.4 Transformation of *Agrobacterium tumefaciens*

First, 600 ng of plasmid were added to 100 µl competent *Agrobacterium tumefaciens* cells and then this mixture was incubated at 37°C for 5 min (heat shock). Then 1 ml of YEP medium was added and it was incubated for 2-4 h at 28°C and 160 rpm. After this period, the cells were harvested by centrifugation (4000 g, 10 min) and were resuspended in 100 µl of YEP medium. The cell suspension was finally plated on YEP plates with 50 µg/ml kanamycin, 25 µg/ml gentamycin and 12.5 µg/ml tetracycline and incubated at 28°C in darkness for 2-3 days.

2.4.5 Transformation of *Arabidopsis thaliana* and screening of transformants

First, 50 ml of YEP medium with the appropriate antibiotics (Kanamycin, gentamycin and tetracycline) were inoculated with a positive *A. tumefaciens* clone containing our plasmid of interest and incubated overnight at 28°C in darkness. After that, the culture was centrifuged (5000 g, 10 min, room temperature) and the pellet was resuspended in 50 ml of the dipping buffer. The flowers of *A. thaliana* that were still closed were then treated with that cell suspension and then incubated overnight in the dark. *A. tumefaciens* infects plants and transfers the T-DNA region containing the gene of interest into the plant. The plants were packaged in a foil to prevent the dipping buffer to get dry too quickly. The following day the plants were exposed to the light again. The whole procedure was repeated after one week. Finally, the seeds were collected and stored at 4°C.

The seeds collected after transformation of the plants were tested for the presence of the T-DNA insert. The screening was based on exposure to antibiotics, whose corresponding antibiotic resistance gene is present in the T-DNA. First the seeds were disinfected with 70% ethanol for 5 minutes and then washed 3 times with water. The seeds were transferred to MS plates with sulfadiazine (sul) (7.5 µg/ml in dimethyl sulfoxide). The plates were incubated for

1 day in the dark at 4°C for stratification and then exposed to light. Approximately 1-2 weeks after germination, it could be seen whether the plantlets were resistant to the antibiotic. The non-resistant plantlets got bent, did not grow further and they were of a pale green-yellow colour, while the resistant plantlets had an intense green colour and grew to a fully-developed plant.

Dipping-buffer: 5% Saccharose (w/v), 0.04 µM Benzylaminopurine, 0.01% Tween 20 (v/v), 2.2 mg/ml Murashige and Skoog Basal Salt Mixture (MS)-Salt (Serva).

(MS) Plates: 0.47% MS-Salt, 0.05% 2-(N-Morpholino)-ethane sulphonic acid (MES), 0.3% Gelrite. Adjust to pH 5.7 with KOH

2.4.6 Transformation of *Phaeodactylum tricornutum*

The microparticle bombardment method was used for the nuclear transformation of *P. tricornutum*. In this method, 10 ASP plates without antibiotics, containing 10^8 cells each in exponential phase, were prepared and cultivated overnight at 18°C. Shortly before transformation, the DNA was prepared to be shot. 1 µg of DNA was needed for every plate. The DNA (10 µg in maximum 40 µl) was mixed with 100 µl 2.5 M CaCl₂, 40 µl 0.1 M spermidine and 100 µl of tungsten particles (60 mg/ml), which were previously washed with ethanol 100% HPLC-grade and then with sterile water. The DNA sample was then vortexed for 1 min and was left to sediment for 10 min. After that, it was centrifuged for 10 seconds. The supernatant was taken off and the sediment was washed twice with ethanol 100%. Finally, the sediment containing the DNA-coated particles was suspended in 100 µl ethanol 100%.

For the transformation, 10 µl of the DNA-coated particle suspension were placed in the disc of the gene gun. The nitrogen valve and the pump were opened. Then the vacuum was set at 140 mbar. When that pressure was reached, the pump was closed and the DNA was shoot into the plate containing the cells. After shooting, the gas was allowed to go out. The transformation method was done under sterile conditions to avoid contamination. The plates were then incubated

at 18°C overnight and the next day, the cells of each plate were transferred with the help of 500 µl of ASP medium into two fresh ASP plates with the corresponding antibiotic. The cells were let to grow under light conditions at 18°C. After approximately 8 weeks, clones were visible.

2.5 Extraction and quantification techniques

2.5.1 Pigment extraction from *P. tricornutum*

A 50 ml culture was initially inoculated with 500.000 cells/ml of *P. tricornutum*. After 4 days, a 2 ml sample was collected and the cells were harvested by centrifugation (13.000 rpm, 1 min, 4°C). The pellet was immediately submerged in N₂(l) and it was stored at -80°C until analysis. For the extraction of the pigments, the cell pellet was resuspended in 200 µl of methanol and it was vortexed. Then, the cells were broken by incubation in an ultrasound bath during 15 min. Later, the cell debris was pelleted by centrifugation (13.000 rpm, 5 min, 4°C) and the supernatant containing the pigments was kept and filtered for subsequent analysis by HPLC. During the extraction and analysis of pigments, the samples were kept in ice in darkness conditions.

2.5.2 Pigment extraction from plant leaves

Plant leaves (without the stem) were immediately frozen in N₂(l) and then grinded in a mortar covered with N₂(l) until a fine powder was formed. Later the powder was transferred to a smaller mortar containing 2-5 ml of 80% acetone. Then the solution was transferred to 2 ml tubes and centrifuged (13.000 rpm, 5 min, 4°C). The supernatant was then transferred to a 1 ml tube. Additionally, 100 µl of a saturated NaCl solution and 150 µl of petrol ether with bp 50-70°C were added to the tube. If in the step before the pellet still contained pigments, the NaCl solution and the petrol ether would be added to the pellet as well, making sure that the pellet is well resuspended. The tube was shortly vortexed and then two phases appeared: an upper green organic phase with the pigments and a polar lower

phase. Finally, the upper phases of several tubes were transferred to a new tube. The pigments were subsequently analysed by HPLC after filtering or separated by thin layer chromatography.

2.5.3 HPLC analysis

The pigments were quantified by high-performance liquid chromatography (HPLC) (Elite LaChrome, Hitachi). A reverse-phase column (Merck, Chromolith, RP-18e, 25 cm x 6 mm) was used. The separation was carried out using a 2-step gradient of solvents A and B at a flow rate of 2 ml/min. A diode array detector (Hitachi) was used to detect the absorption spectrum (400-750 nm) of the different pigments that were leaving the column according to its polarity. The quantification of the pigments was determined by the area under the specific peak. The software EZChrome-Elite was used for analysis of the results and it provided the area under the peak of absorbance of a specific wavelength characteristic of every pigment. The area of a specific pigment is proportional to the amount of this pigment and a numerical correlation was determined based on calibration for this specific system by K. Gundermann.

Mobile phase solutions (A and B) and washing solutions (C and D):

- Solution A: 80% methanol, 1 mM Tris (pH 7.5)
- Solution B: 60% methanol, 40% acetone.
- Solution C: 20% methanol
- Solution D: 80% methanol

2 Step Gradient of solutions A and B:

- 0 min: 100% A, 0% B
- 3.5 min: 0% A, 100% B
- 6.7 min: 100% A, 0% B
- 8.0 min: 0% A, 100% B

2.5.4 DNA quantification with Nanodrop or ImageJ

DNA was quantified either using Nanodrop ND-1000 (Thermo Scientific) or by analysis using the software ImageJ of DNA bands present in an agarose gel.

2.6. Cell encapsulation

The number of cells to encapsulate were determined with the occupation prediction calculator. This depends on set-up parameters, like the flow rate and the frequency through the encapsulator, on the volume batch and on lambda, that is the number of cells per nanoliter reactor (NLR) that one wishes to have. Lambda was calculated according to the Poisson distribution, and for a defined lambda value a different ratio of empty, single, double or triple-occupied NLRs results. Ideally, just after encapsulation, every NLR would contain only 1 cell. A lambda value of 0,6 was used for all the experiments, which corresponds to 94.545 cells for the 500 μm NLRs. The cells were counted under the microscope using a counting chamber (Fast Read, 102).

The determined number of cells were resuspended in a final volume of 2 ml of its own medium. Since it already contains NaCl, it was not additionally supplemented. This resuspension was added to 8 ml of alginate 2,5% (v/v) and was vigorously mixed by tube inversion so that the cells were homogeneously distributed throughout the alginate. Vortex must be avoided. This mixture was introduced into a syringe maintaining sterile conditions and was let stand horizontally for 10 min until degassed.

Before the encapsulation, the encapsulator (Nisco) was washed and sterilized. Afterwards, the sample with the cells and the alginate was pumped through the encapsulator at a rate of 3.3 ml/min with the aid of a pump (kdScientific). The frequency for disrupting the flow into little droplets was 0.6 kHz. The beads flowed to an agitated container that had the hardening solution for the alginate, that is made of 100 mM CaCl_2 , buffered with 50 mM Tris pH 7.7. Initially when the drops started falling, the hardening container was blocked with a petri dish, and only when the flow-through was as desired, with only one straight flow without side flows, the block was removed. When 1 ml of the solution was left in the syringe, the hardening container was blocked again and the encapsulator was

immediately washed with water. Then, the NLRs were let to harden in the hardening solution under agitation during 10 min.

Once the alginate beads were hard, they were passed through a 100 µm filter (Easy strainer, Greiner bio-one). Then they were recovered with culture medium and they were filtered again. Finally, 1 g of NLRs was resuspended in 10 ml of medium. The culture flasks with the NLRs containing the algae were then cultured for 3 days in the conditions for cultivating the algae.

2.6.1 Determination of cell viability

Fluorescein diacetate (FDA) (dilution 1:1000) was used to keep track of the biomass in the cell sorter COPAS and to determine cell viability under the fluorescence microscope. The COPAS cell sorter was operated by the researchers of Fgen (Basel, Switzerland).

3. Improvement of fucoxanthin levels

3.1 Introduction

3.1.1 Fucoxanthin

Fucoxanthin is a carotenoid that can be found in most heterokonts (including diatoms, brown and golden algae), haptophytes and some dinoflagellates and it is responsible for the brown color of these organisms¹⁴. Fucoxanthin is present in the photosynthetic apparatus, more specifically in the thylakoid membrane of the chloroplast, where it forms the fucoxanthin-chlorophyll a/c protein (FCP) complex that function as antenna, harvesting light and transferring the excitation energy to chlorophyll²⁸. There are three groups of FCPs: Lhcf, the main antenna for light harvesting, Lhcr, that is specific of photosystem I, and Lhcx, with a photoprotection role. FCPs contain approximately 8 chlorophyll a, 2 chlorophyll c and 6 fucoxanthin molecules per monomer²³.

Fucoxanthin has been investigated for its health benefits in humans:

- Fucoxanthin has been investigated for its activity in cancer prevention and treatment, since it induces cell cycle arrest and apoptosis²⁴.
- Fucoxanthin can also help to reduce obesity by regulating enzymes involved in lipid metabolism such as acetyl-coA carboxylase, by inducing uncoupling protein 1 in white adipose tissue and thus dissipating the proton gradient in mitochondria, and regulating hormones, e.g. leptin, that control energy balance²⁵.
- Fucoxanthin, as many other carotenoids, has antioxidant activity that exerts by scavenging free radicals and quenching singlet oxygen. Antioxidant activity is important since oxidative stress can lead to many diseases, e.g. cardiovascular diseases²⁴.
- Fucoxanthin has also beneficial effects preventing Alzheimer disease. It has been shown that fucoxanthin can prevent the formation of β -amyloid aggregates and inhibits acetylcholinesterase in vitro^{26,27}.

Nowadays, fucoxanthin is commercialized as dietary supplements for its anti-obesity effects²⁸. These supplements are produced mostly by different Chinese companies and the extracts come from brown seaweeds, especially from discarded parts from the edible macroalgae *Laminaria japonica*²⁹. These and another macroalgae are widely used in East Asia for food consumption and are cultured for that purpose. Only in Japan, approximately 57000 tones of cultured *Laminaria* (generally known as kombu) are discarded every year in the food processing industry³⁰. Therefore, it is logical that in these countries in which macroalgae are cultivated, the wasted parts of kombu are given other uses, like the production of fucoxanthin- containing extracts.

Elsewhere in the world, it would be more beneficial to produce fucoxanthin from microalgae, since macroalgae contain very low concentrations of fucoxanthin, ranging from 0.02 to 0.58 mg/g fresh weight. In contrast, microalgae produce at least ten times more fucoxanthin, in mg/g dry weight, than macroalgae³¹. Algatech, an Israeli company, launched for the first time in 2018 the first extract containing fucoxanthin and eicosapentaenoic acid coming from microalgae, specifically from the diatom *P. tricornutum*, sold for its benefits to the liver. The algae are cultivated in photobioreactors, proving a closed and controlled environment that can be grown all the year without being influenced by seasonal fluctuations³².

Therefore, diatoms can become an interesting cell factory for the production of fucoxanthin for the nutraceutical and pharmaceutical market. The yield of fucoxanthin could be further increased by the development of mutant strains that produce higher fucoxanthin levels in order to make the process more economically-feasible³³.

3.1.2 Metabolic engineering

Industrial biotechnology is an alternative means of production to the conventional petrochemical synthesis that involves the adaptation and/or modification of

organisms found in nature or parts of them, eg. enzymes, with the goal to produce chemicals, drugs, fuels, materials or energy³⁴.

The use of whole cells as catalysts inspired the concept of cell factory. In this approach, scientists take advantage of the metabolic network present in the cells to produce a compound of interest. A common example is the production of beer by fermentation³⁵. Classically, increments in the production yields were based on strain improvement by methods that create natural biodiversity such as random mutagenesis, hybridization, protoplast fusion or directed evolution, followed by a high-throughput screening for the more productive strains. Several cycles of natural variation and screening are usually needed³⁶.

The development of genetic engineering tools in the last decades has opened a new paradigm of possibilities, with the introduction of metabolic pathways that are not present in the production organism (heterologous pathways) and the fine tuning of these pathways to improve the product yield³⁵. Metabolic engineering is the discipline that makes use of directed genetic modifications for the modulation of metabolic pathways in order to improve the properties of a production strain. It aims at redirecting the metabolic fluxes towards the formation of a product of interest^{37,38}.

The process is different depending whether the pathway is native in the host organism or it is not. When introducing heterologous pathways, it is important to fine tune the expression levels of the enzymes of the pathway, for example by adjusting the promoter strength, the expression copy number or the ribosome binding site strength. Additional decisions can be taken when building heterologous pathways, like the colocalization of enzymes for metabolic channeling or the localization of the pathway in a specific cell compartment when using eukaryotic hosts. These strategies allow to increase the local concentration of metabolites and also in the latter case, to isolate the pathway from competing reactions³⁹.

Common strategies in metabolic engineering are the overexpression of genes that encode enzymes that catalyze bottleneck steps, deleting or downregulating genes at branching points from the pathway, increasing the supply of the precursors and enzyme cofactors, improving the transport system and product

sequestration or export, among others⁴⁰. These strategies have led in many cases to a successful improvement in the productivity. However, sometimes the decision of the target genes to modify are intuitive and do not always yield the expected cellular phenotype.

With the construction of genome-scale models for many important host strains and the development of approaches like flux balance analysis and prediction algorithms it is possible the identification of potential targets for genetic modification in silico. This approach is of particular interest since it can provide targets that are non-obvious by looking exclusively at the pathway of interest⁴¹.

Another approach that is useful for the identification of potential targets for metabolic engineering purposes is Metabolic Control Analysis. This methodology can determine in a quantitative basis which enzyme of the pathway exerts more control on the flux. Contrary to the classical idea of one "rate-limiting reaction", metabolic control analysis explains why the control of the flux is shared by all enzymes in the pathway⁴².

Finally, it is important to consider the possible burden effects on the cell by a given genetic modification. This aspect has also to be taken into account when selecting the strategy of choice, since it can affect cell growth and therefore productivity⁴³.

3.1.3 Endogenous regulation of metabolism

In order to maintain homeostasis, cells have evolved a complex regulatory system to manage their carbon and energy sources for the synthesis of all cellular components. This regulation can be a problem when trying to manipulate metabolism by redirecting the flux to a given pathway. For this reason, it is very important to have a clear picture of the regulation of the pathway of interest before performing any genetic modification. Omic techniques can help to unravel pathway regulation. On the other hand, the homeostasis achieved by regulation is an advantage, since robust strains operate stably in fermenters, where varying conditions can take place^{38,44}.

To overcome the problem of regulation, one can de-regulate key regulatory enzymes in the pathway by protein engineering. One example of this strategy is the de-regulation of acetyl-CoA carboxylase, an important enzyme involved in the synthesis of lipids, by making a point mutation at the site in which is phosphorylated, avoiding in this way its inactivation. Another option is to introduce a whole heterologous pathway that is not influenced by the endogenous host regulatory system³⁸.

A new trend in the field of metabolic engineering is introducing control into the pathway in a similar way that is done in control engineering. This strategy allows to carry out a dynamic control on the metabolic fluxes, so they are rebalanced according to the metabolic state of the cell, on the different cell growth phase and on the different culture conditions that can arise especially at large scale fermentors³⁹. The introduction of dynamic regulation helps to reach a more efficient trade-off between production and cell growth, avoids the accumulation of intermediate metabolites and can prevent the deleterious effects on production caused by cell stress⁴⁵.

Genetic circuits have been developed for the dynamic regulation of pathways. For example, circuits have been designed that employ sensors of the accumulation of a given metabolite and respond by optimizing the flux accordingly. These circuits are built following the basic principles of synthetic biology of modularity, standardization and characterization. Genetic circuits can exert control at the transcriptional, translational or cell signaling level^{46,47}.

3.1.4 Design-build-test-learn cycle

Metabolic engineering usually involves several rounds of the design-build-test-learn cycle. In the design stage, the host organism is first selected according to the raw materials available, the processing conditions and purification steps of the product, the toxicity of the product for the organism and the requirement of specific modifications in the product, like the glycosylation pattern in the production of recombinant proteins. After the selection of the host, the metabolic targets for improving the yield should be identified and optionally, dynamic control

mechanisms could be also applied. After the design, the metabolic strategies are carried out in the laboratory in the building phase with the aid of molecular biology techniques. After that, the host strain should be tested in terms of physiology and productivity³⁸.

In the test stage, omic techniques can be of great help for understanding the effects of the genetic modifications in cell physiology in the context of the whole metabolic network. Transcriptomic analysis of the producer strain compared to the wild type strain can reveal for instance stress responses, such as the unfolded protein response, energy or redox imbalances and limitations in the transport and secretion process. Another approach useful for the characterization of the producer strain is metabolic profiling. This technique can help to monitor the carbon flow and can reveal intermediate metabolites that accumulate or unwanted byproducts that have been formed⁴⁸.

Finally, from the knowledge gathered in the test stage with regard to the physiology of the producer strain and the production of the desired compound, new targets for improvement can be detected and a new round of the metabolic cycle can take place. This iterative process is repeated until a robust strain that produces the product at the desired yield is obtained³⁸.

3.1.5 Approaches to enhance carotenoid levels

Several approaches have been adopted in microalgae, plants and fungi for increasing the carotenoid levels. Depending on the type of product, it is important to consider also other factors, like public acceptance, when selecting the approach. While directed genetic modifications provide more control over the elicited changes, this strategy is sometimes limited by the regulatory mechanisms of the cell and it does not count with consumer acceptance, especially when it comes to genetically modified food in Europe. In contrast, strategies like random mutagenesis or environmental or stress conditions are not legally considered as genetic modification, so they can count with more public acceptance, but one has often poor control over the relying process, what sometimes leads to a big variability in the observed responses²⁰.

Here the main strategies for the enhancement of carotenoid levels is summarized:

- Stress conditions

One strategy that has been widely applied to increase carotenoid levels is the application of stress conditions to the organisms⁴⁹. The main sources of stress that have shown positive effects for the accumulation of given carotenoids are:

-Nutrient stress (nitrate, phosphor or sulfur limitation)⁴⁹.

-Environmental and/or photo-oxidative stress (hydrogen peroxide, increased salinity, increased temperature, high light)⁴⁹.

Reactive oxygen species (ROS) are created due to the stress conditions and promote carotenogenesis via redox signaling²⁰. ROS can be very harmful at high concentrations and cause damage in cell structures, but at low concentrations they can activate signaling pathways that initiate certain physiological processes, for example by oxidation of cysteine residues of proteins⁵⁰.

The results of stress conditions vary widely depending on the algal species considered. Stress conditions pose restrictions on growth, limiting the productivity of the carotenoids. To circumvent this problem a 2-stage cultivation system could be adopted, with a first stage with the optimal growing conditions to achieve the desired biomass concentration, and a second stage in which the stress condition is applied to increase the production of the carotenoid of interest⁴⁹. As an example, a two-stage mixotrophic culture using a perfusion system, in which light intensity was progressively increasing (150-450 $\mu\text{E}/\text{m}^2\cdot\text{s}$) during the second stage, resulted in a high accumulation of astaxanthin (602 mg/L) in the green algae *H. pluvialis*⁵¹.

An alternative strategy to prevent the slow growth of the organisms under stress conditions is adaptative laboratory evolution (ALE). This technique consists in adapting organisms to unfavorable conditions through many rounds of growth and selection of the organisms with the best phenotypes that arise through mutations⁴⁹.

Finally, it is interesting to take into account that many of the stress conditions increase also lipid levels, being able to make more profitable the production by the simultaneous enhancement of both carotenoids and lipids⁴⁹.

- Random mutagenesis

Random mutagenesis consists in producing natural variability by using mutagenic agents. It is important to notice that *P. tricornutum* cells are diploid. Several mutagenic agents have been used in order to increase carotenoid levels. For example, Ethyl methanesulfonate is a chemical that produces point mutations by guanine alkylation. Then, during replication a thymine group is placed complementary to the alkylated guanine, instead of cytosine. After conducting chemical mutagenesis with ethyl methanesulfonate in the diatom *P. tricornutum* the best mutant enhanced the levels of fucoxanthin by 69.3% compared to wild type⁵².

UVC radiation is a physical agent that can promote mutagenesis and DNA lesions, such as pyrimidine dimers in which covalent bonds are formed between adjacent pyrimidine bases and the DNA structure is therefore altered. By exposing *P. tricornutum* cells to UVC radiation, the top mutant had 1.7 times more fucoxanthin levels than wild type⁵³.

- Metabolic engineering

Several approaches have been conducted to increase the levels of carotenoids by genetic engineering methods.

The overexpression of the first enzymes of the pathway that direct the mass flux into that pathway has yielded good results. For instance, the single overexpression of the *psy* and *dxs* genes in the diatom *P. tricornutum* produced an increase of 2.4 and 1.8-fold, respectively, in fucoxanthin levels in comparison to wild type⁵⁴.

Also, the over-expression or direct evolution of the rate-limiting enzymes of the pathway is another common approach to increase carotenoid levels. PDS, GGPPS and IDI have been proved to be rate-limiting. In a study, the overexpression of endogenous mutated PDS increase carotenoid levels by 32.1% in the green algae *C. zofingiensis*. In another study, the directed evolution of the IDI

enzyme resulted in an 1.8-fold increase of the lycopene production in *E.coli*. However, it is well known that only one enzyme does not control the flux of a pathway, and for this reason it is more efficient to overexpress several rate-limiting enzymes at the same time. In addition, a common challenge is to eliminate the feedback inhibition of the MEP pathway exerted by the accumulation of IPP/DMAPP^{55,56}.

An interesting approach has been to increase the carotenoid storage capacity of the organisms. Carotenoids are accumulated in membranes and in some organisms, they are stored in lipid bodies. In several studies in *E.coli*, the membrane area was increased by the over-expression of enzymes responsible for the formation of membrane lipids, what resulted in the increase of β -carotene levels by 43%⁵⁶.

Another strategy has been to address the problem by modules, trying to optimize the different modules that contribute to the increase of carotenoids and then balancing the different modules with the aid of the Multivariant Modular Metabolic Engineering technique. The proper balancing of the different modules is essential to avoid the accumulation of precursors that could cause toxicity and result in inefficient bottlenecks. The modules considered were: the carotenoid module, the cofactor module, the central metabolism module and the carotenoid precursor module (MEP pathway). Since many strategies to increase the carotenoid and the MEP modules have been previously described, the central metabolism and the cofactor modules will be now addressed⁵⁶.

For the central metabolism module, the main task is to solve the availability of the precursors of the MEP pathway: pyruvate and glyceraldehyde-3-phosphate (G3P). These metabolites are part of the glycolysis pathway, but the problem lies in that the flux through this pathway goes into the formation of pyruvate, so the available levels of G3P are not as high. Since both precursors are needed in equal amounts, the approach consisted in redirecting the flux for the formation of G3P by the down-regulation of G3P dehydrogenase enzyme or by over-expression of the phosphoenol pyruvate synthase. These strategies increased the lycopene production in *E.coli* by 45 and 250%⁵⁶.

The cofactor module is also of big importance, since the MEP and carotenoid pathways are anabolic pathways that require reducing power and energy. To produce one β -carotene molecule, 16 NADPH, 8 ATP and 8 CTP molecules are needed following the MEP pathway. NADPH is obtained by the pentose phosphate pathway (PPP) and the Krebs cycle. These pathways have been manipulated in order to increase the levels of NADPH. ATP is produced by the electron transfer chain and the ATP-synthase. Analogously, the manipulation of the enzymes involved in the synthesis of ATP led to a 120% increase of β -carotene levels in *E.coli*⁵⁶.

Finally, it is interesting to know that a genome-scale metabolic model of *P. tricornutum* has been constructed and that it has been already used to predict in silico the genetic modifications that would yield a higher production of carotenoids using constraint-based modeling, flux balance analysis and the OptFlux software^{57,58}. The knockout reactions proposed by the program mostly involved the pyruvate, glutamine/glutamate, nitrogen assimilation pathways and transport reactions. For instance, knockouts of nitrite reductase, glutamine and glutamate synthase, pyruvate-phosphate dikinase, G3P dehydrogenase and triose-phosphate isomerase. The authors proposed that since the 3 first enzymes are involved in amino acid biosynthesis, the knockout of these enzymes would leave more reducing agents available for the use by the carotenoid pathway. In vivo that strategy might not be very realistic since amino acids are vital to produce proteins. The other knockouts proposed are in line with the strategies previously presented in the central carbon module of redirecting the flux towards the formation of G3P, although an approach involving down-regulation of these enzymes would be more beneficial than a knockout in order to keep cellular fitness⁵⁸.

However, the fact that some enzymes of the carotenoid pathway in diatoms, such as fucoxanthin synthase, have not been identified yet, together with the lack of knowledge about the regulation of the pathway, pose restrictions for developing successful strategies using metabolic engineering approaches for the accumulation of carotenoids in diatoms¹³.

3.2 Objectives

The main objective is to enhance the production of fucoxanthin in diatoms by genetic engineering approaches or by stress conditions in order to reduce costs in the biotechnological production system. The research will be conducted using the model organism *Phaeodactylum tricoratum*, for which the whole genome sequence is available.

The impact of stress conditions, such as increased salinity and high light, will be evaluated in the accumulation of fucoxanthin. In addition, the effect of the overexpression of rate-limiting or key enzymes in the carotenoid and MEP pathway will be investigated. In the beginning these effects will be observed in single mutants, but in the end the aim it is to create a mutant overexpressing several rate-limiting enzymes, since only one enzyme does not have the control of the flux of the pathway and in this way also bottlenecks could be avoided.

3.3 Results

First, it was tried to increase the fucoxanthin levels by applying stress to the diatom *P. tricoratum*, since carotenoids are usually accumulated under these conditions. For that, cells were inoculated at the same cell concentration ($5 \cdot 10^5$ cells/ml) at the exponential phase but with different culture conditions. The different conditions were the concentration of salts in the medium (0.1, 0.2, 0.3 and 0.4 M of NaCl) and the comparison of high light ($100 \mu\text{E}/\text{m}^2 \cdot \text{s}$) versus low light ($40 \mu\text{E}/\text{m}^2 \cdot \text{s}$). Usually the ASP medium used for the cultivation of *P. tricoratum* contains 0.086 M of NaCl, so the sample at 0.1 M of NaCl can serve as an indicator of the standard growth. The different cultures were grown for 14 days, and samples for HPLC were taken at days 4, 8 and 14.

As it can be seen in **Figure 7**, the higher the concentration of NaCl, the lower are the fucoxanthin levels. The cells grown in 0.1 M of salt showed the highest fucoxanthin levels. Only the levels of fucoxanthin in the cells grown at 0.3 M NaCl at day 14 do not follow this trend, but probably this is due to an extraction mistake.

In the same way, also low light was more beneficial for the accumulation of fucoxanthin than high light.

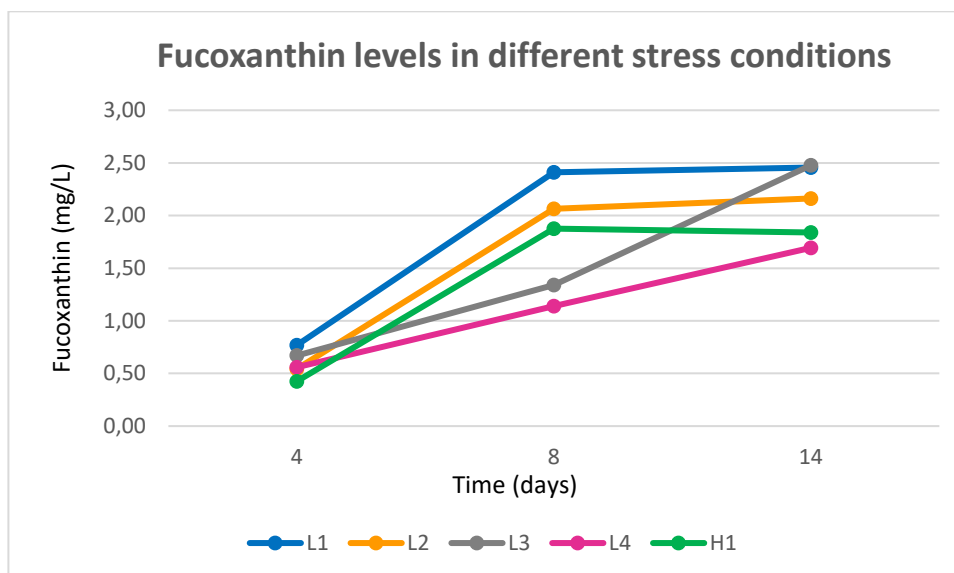


Figure 7. Fucoxanthin levels in *P. tricornutum* under different stress conditions. *P. tricornutum* cultures, initially inoculated with the same cell number, were grown for 14 days under different conditions. L1, L2, L3 and L4 cultures were grown in low light conditions ($40 \mu\text{E}/\text{m}^2\cdot\text{s}$) with 0.1, 0.2, 0.3 and 0.4 M of NaCl, respectively. The H1 culture was grown in high light conditions ($100 \mu\text{E}/\text{m}^2\cdot\text{s}$) with 0.1 M of NaCl.

Genetic manipulation was also used to increase fucoxanthin levels in *P. tricornutum*. For this approach, the rate-limiting enzymes of the pathways were cloned and additionally inserted into the genome. The plasmids pPhaNR-Nat containing the *psy* gene (Gene ID: 7199567) and pPhanNR-Zeocin with the *dxs* gene (Gene ID: 7204829) were provided by Dr. Eilers. For the visualization of the components pPhanNR vector, see **Appendix**. In the pPhanNR vector, the genes of interest are under the control of the nitrate reductase promoter and terminator. Since the ASP medium contains nitrate, the genes will be constitutively expressed. For producing enough amount of the plasmids for the transformation of *P. tricornutum*, the plasmids were first transformed in *E.coli* and then isolated by Miniprep. The presence of the *psy* and *dxs* genes in the plasmids was checked by a restriction assay in which the insert was cut out of the plasmid with the restriction enzymes EcoRI and HindIII (**Figure 8**). Finally, both plasmids were co-transformed in *P. tricornutum* via biolistic transformation.

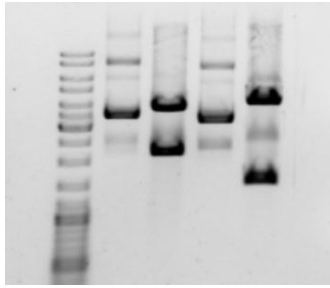


Figure 8. Restriction analysis of the vector pPhaNR-*dxs* and PhaNR-*psy*. From left to right: Gene ruler 1kb, control pPhaNR-*dxs* unrestricted, pPhaNR-*dxs* restricted, pPhaNR-*psy* unrestricted, pPhaNR-*psy* restricted. Sizes: pPhaNR (3851bp), pPhaNR-*dxs* (5901 bp), *dxs* (2197 bp), pPhaNR-*psy* (5545 bp), *psy* (1694 bp).

The genes *idi* (Gene ID: 7201899), *ggps* (Gene ID: 7197706) and *pds2* (Gene ID: 7198353) were cloned in the pPhaNR-Zeocin plasmid by the hot fusion method (Figure 9).

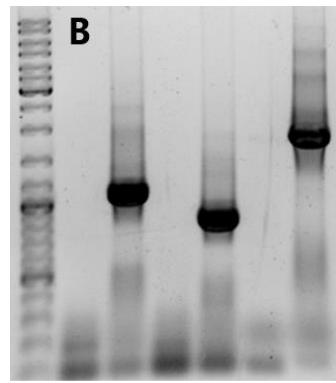


Figure 9. Verification of the presence of the genes by PCR. A) Amplification by PCR of the signal and target peptide of the *zds* gene from cDNA of *P.tricornutum* (C-, negative control) . B) Amplification by PCR of the genes *ggpps*, *idi* and *pds2* after Miniprep. From left to right: GeneRuler 1 kb, negative control of *ggpps*, *ggpps*, negative control of *idi*, *idi*, negative control of *pds2* and *pds2*.

In diatoms, two *idi* genes exist in the cell (Gene ID: 7201899 and Gene ID: 7201112). One codes for an isopentenyl diphosphate isomerase in the cytoplasm participating in the MVA pathway and one in the plastid taking part in the MEP pathway. Both enzymes catalyze the same reaction and have a high sequence similarity (Appendix). The *idi* gene (Gene ID:7201112) is not well annotated and the sequence at the beginning of the gene is not complete. For that reason, the *idi* gene (Gene ID: 7201899) was cloned. However, no signal peptide was detected using the SignalP server, indicating that this might be the cytoplasmic enzyme. Therefore, the signal and target peptide of a plastid-localized protein, coded by the gene zeta-carotene desaturase (*zds*), was fused in front of the coding sequence of *idi*, maintaining the frame.

After cloning, the *idi*, *ggps* and the *pds2* genes were transformed into

Phaeodactylum tricornutum cells via biolistic transformation that leads to random integration into the genome.

The clones obtained transformed with *dxs-psy* and *ggps* were checked by PCR using genomic DNA with primers specific for the construct in the vector (**Figure 10**). In this way, the forward primer binds the end of the nitrate reductase promoter in the vector and the reverse primer binds the cloned gene. Therefore, although the wild type contains naturally the genes that have been additionally inserted, the DNA amplified by this reaction is not present in the wild type because the native genes are not under the control of the nitrate reductase promoter. Therefore, only the clones that have been transformed will give a positive result in the PCR. A negative control was also included.

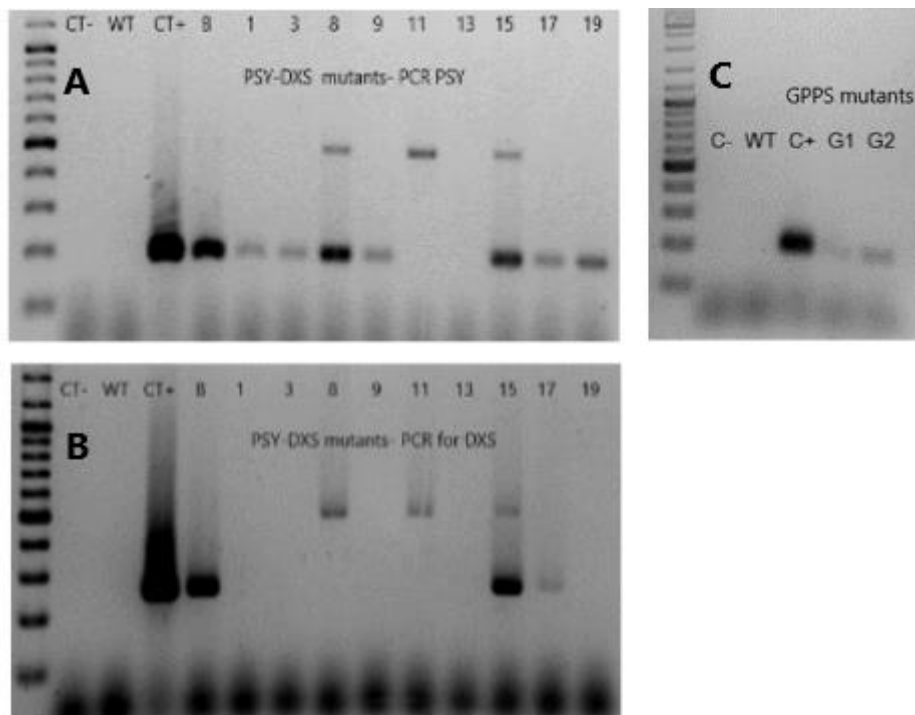


Figure 10. Verification of the presence of the inserted *psy* and *dxs* genes in the diatom mutants. A and B) The clones of *P. tricornutum* that were simultaneously transformed with the genes *psy* and *dxs* were tested by 2 PCR reactions, one to detect the presence of *psy* (A) and the other of *dxs* (B). C) The clones that were only transformed with the gene *ggpps* were tested by PCR. Amplification size for all the genes ranges from 200-300 bp. In all PCRs, construct-specific primers have been used. C-: negative control. C+: positive control. WT: wild type.

As it can be seen in **Figure 10** only clones B, 15 and 17 are double mutants being

transformed with both *psy* and *dxs*. In **Figure 10** it can be seen that both clone 1 and 2 have being transformed with *ggpps*.

We can see in **Figure 11** that the fucoxanthin levels in the mutants are very similar to the levels in wild type, varying between 1.00-1.20 mg/L of fucoxanthin after 4 days of growth. The GGPS1 clone has a higher volumetric productivity of fucoxanthin per million cells (28% more productive), since it produces slightly lower levels of fucoxanthin but it grows less. This clone is therefore not interesting from a biotechnology point of view, but we can see that the cells are more efficient producing fucoxanthin, but due to the reduced growth, the total levels are not higher than the levels in wild type.

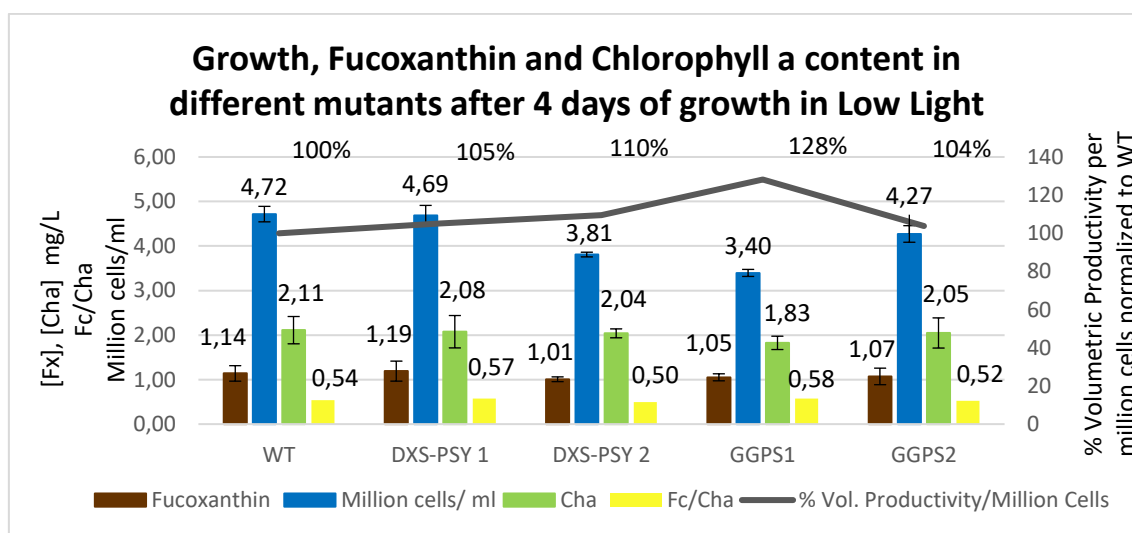


Figure 11. Characterization of the DXS-PSY and GGPPS mutants. The fucoxanthin (Fc) and chlorophyll a (Cha) levels in the mutants grown under low light conditions ($50 \mu\text{E m}^{-2} \text{s}^{-1}$) for 4 days were quantified by HPLC and compared to wild type (WT) levels. In addition, the cell number, the fucoxanthin to chlorophyll a ratio and the volumetric productivity per million cells are represented. The originally named clones DXS-PSY17 and DXS-PSYB were renamed DXS-PSY1 and DXS-PSY2, respectively, for simplicity.

The clones obtained transformed with *pds2* and *idi* were checked by PCR using genomic DNA with primers specific for the construct in the vector. For the *idi* gene, the forward promoter binds the nitrate reductase promoter and the reverse primer binds to the end of the *idi* gene. For the *pds2* gene, the forward primer binds the beginning of the *pds2* gene and the reverse primer binds the nitrate reductase

terminator. Only one clone, clone 7, resulted positive, while 2 clones showed up to be positive for the *idi* gene, clone 3 and Y (**Figure 12**).

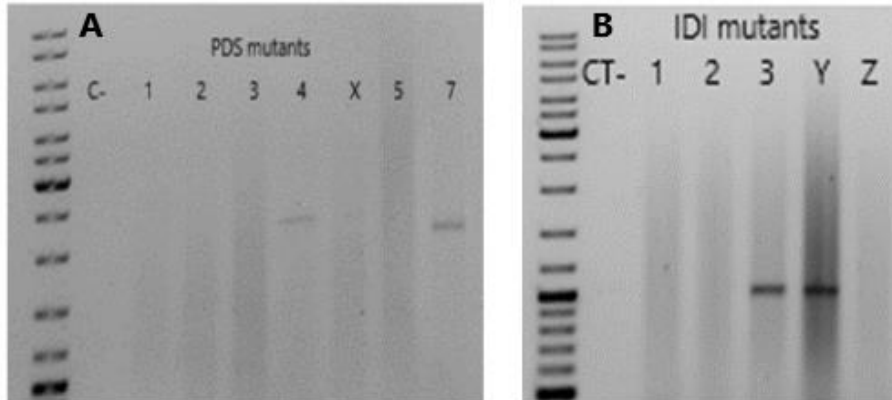


Figure 12. Verification of the presence of the inserted *pds2* and *idi* genes in the diatom mutants. The clones of *P. tricornutum* that resisted the antibiotic selection after transformation were tested by a PCR reaction, in which construct-specific oligonucleotides were used. A and B) PCR to detect the presence of the transformed *pds2* (A) and *idi* (B) genes. C-: negative control.

In **Figure 13** it can be seen that also the PDS and IDI mutants do not produce higher fucoxanthin levels than wild type. Usually, Cha levels are used as an indirect measure of growth, but in this case, some of the mutants show the same or a slightly increased growth (eg. PDS7 or IDI Y) with respect to wild type, but however, the levels of Cha are lower than in the wild type. The volumetric productivity per million cells is much lower in the mutants than in the wild type, since the growth of the mutants is not negatively affected but they have lower fucoxanthin levels.

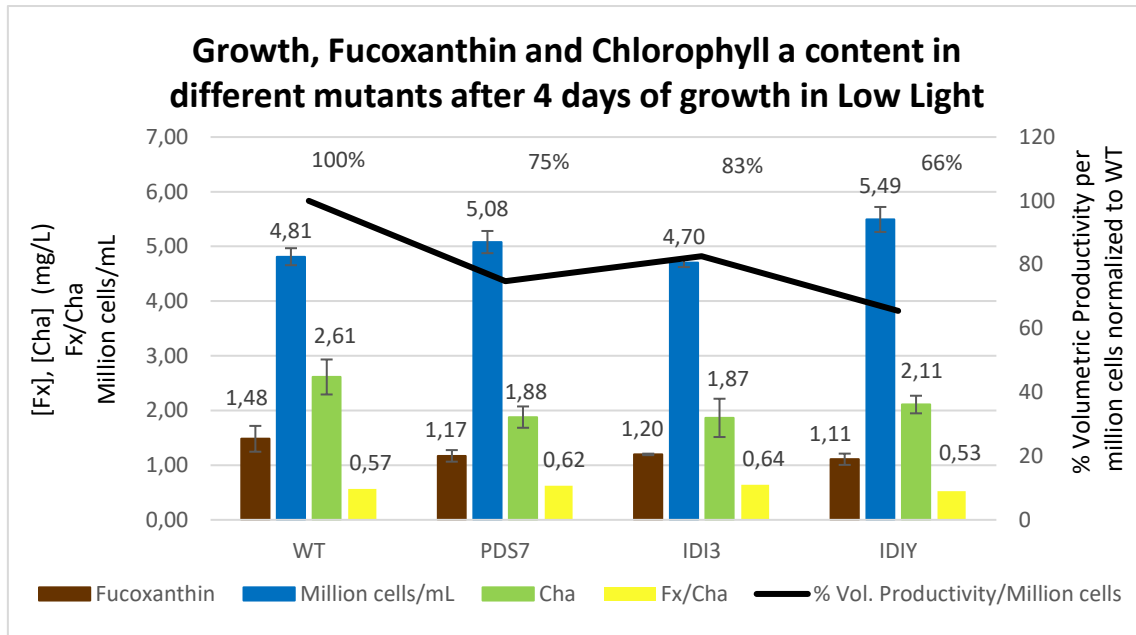


Figure 13. Characterization of the PDS and GGPPS mutants. The fucoxanthin (Fx) and chlorophyll a (Cha) levels in the mutants grown under low light conditions ($50 \mu\text{E m}^{-2} \text{s}^{-1}$) for 4 days were quantified by HPLC and compared to wild type (WT) levels. In addition, the cell number, the fucoxanthin to chlorophyll a ratio and the volumetric productivity per million cells are represented.

For the construction of the mutant with additional copies of genes that code for several rate-limiting enzymes of the carotenoid pathway, the whole cassette (promoter-gene-terminator) of the gene *ggpps* was amplified and cloned into the vector *pPhanNR-psy*, which had been previously restricted with *NheI*. Then, the vector *pPhanNR-psy-ggpps-Zeocin* and the vector *pPhanNR-dxs-Nat* were co-transformed by biolistics. Two clones were obtained, clone A and B, that were tested by PCR with construct-specific primers (**Figure 14**).

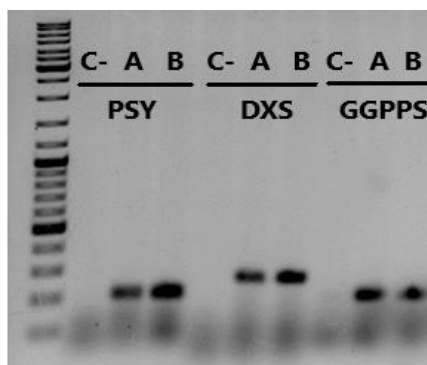


Figure 14. Verification of the presence of the inserted *psy*, *dxs* and *ggpps* genes in the diatom mutants. The 2 clones of *P. tricornutum* (clone A and B) that were obtained after the simultaneous transformation with 3 genes (*psy*, *dxs* and *ggpps*) were tested by PCR, in which construct-specific primers were used. Amplification size ranges from 200-300 bp. C-: negative control.

In the beginning, the mutants in which additional copies of the genes *psy*, *dxs* and *ggpps* had been introduced were growing very slowly and they looked very stressed, since they were growing as clumps of cells. For that reason, they were transferred from low light conditions ($50 \mu\text{E m}^{-2} \text{s}^{-1}$) to very low light conditions ($10 \mu\text{E m}^{-2} \text{s}^{-1}$) and then the mutants started to grow in solution.

The pigment levels of the mutants were quantified by HPLC and compared to those in wild type (**Figure 15**). For this experiment, the algae were grown for 4 days at $10 \mu\text{E m}^{-2} \text{s}^{-1}$. It is important to notice that with such low light intensities the light is not equally distributed among the different flasks. This was verified later with the aid of a photometer. This may have caused that for the same culture some biological replicates grew, although very slowly, and others not enough to have a pellet after centrifugation. That is why only 1 biological replicate of the wild type and 2 biological replicates for every mutant could be analyzed.

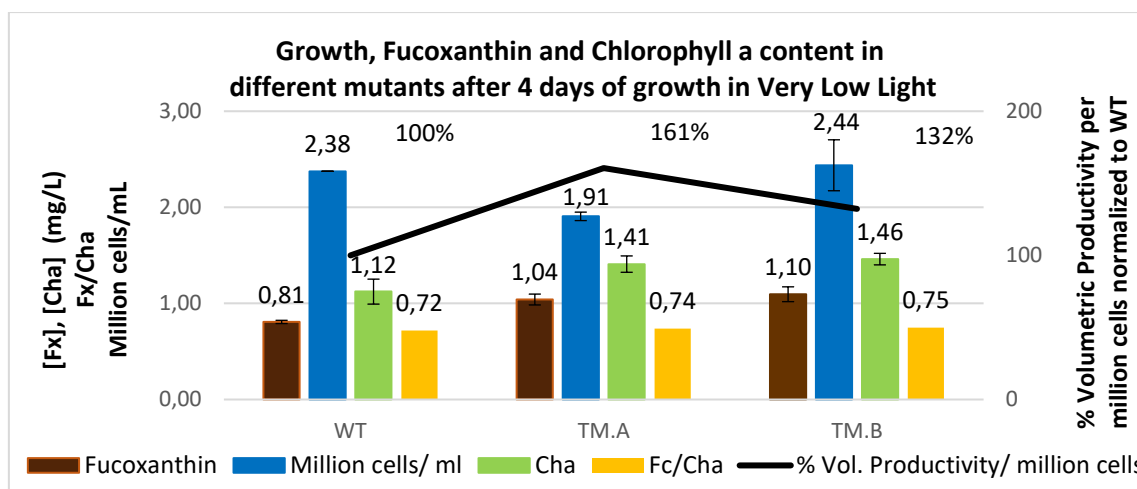


Figure 15. Characterization of the DXS-PSY-GGPPS mutants under very low light. The fucoxanthin and chlorophyll a levels in the mutants that were simultaneously transformed with the genes *psy*, *dxs* and *ggpps* were quantified by HPLC and compared to wild type (WT) levels. These cultures were grown under very low light conditions ($10 \mu\text{E m}^{-2} \text{s}^{-1}$) for 4 days. In addition, the cell number, the fucoxanthin to chlorophyll a ratio and the volumetric productivity per million cells are represented. The 2 clones obtained (A and B) are called TM, that stands for triple mutant.

Comparing **Figure 15** with **Figure 13**, it is clear that as expected the growth, fucoxanthin and Cha levels in the wild type were considerably reduced when grown in very low light in comparison to low light. The fucoxanthin to chlorophyll

(Fx/Cha) ratio is increased in very low light conditions (Fx/Cha ratio of 0.72 in very low light, versus a ratio of 0.57 in low light).

Figure 15 shows that the fucoxanthin levels were higher in the mutants with additional copies of *psy*, *dxs* and *ggpps* genes than in the wild type under very low light conditions. However, they were not higher than the levels of the wild type under low light conditions. Therefore, these mutants are of no biotechnological interest. The mutants also present higher Cha content compared to the wild type, so the Fx/Cha ratios are very similar among them. The volumetric productivity per million cells is considerably higher in the mutants than in the wild type (61% and 32% higher productivity in clones A and B, respectively, compared to wild type). However, again it can be seen that there isn't a comparable correlation between number of cells and chlorophyll levels, since clone A has the lower cell number, but on the contrary possesses higher chlorophyll a levels than wild type.

After that experiment, the cultures of the triple mutants were transferred back to low light conditions to reassess their ability to grow in suspension, in an attempt to increase their growth. However, **Figure 16** indicates that after 6 days in these conditions the growth of the mutants is very reduced in comparison to the wild type. The fucoxanthin levels were not analyzed by HPLC since it could be seen by the naked eye that the levels were much lower in the mutants with respect to the wild type. This is possible since fucoxanthin is responsible for the brown color of these algae, and the difference among the color intensities in the mutants can be detected even by the naked eye when the differences in color intensity are big.

In conclusion, it can be said that even in the clones in which it was achieved to increase the cell productivity for the accumulation of fucoxanthin, the total levels of fucoxanthin were not enhanced in comparison to the wild type, due to a reduction of growth in some cases or to other causes which remain unknown.

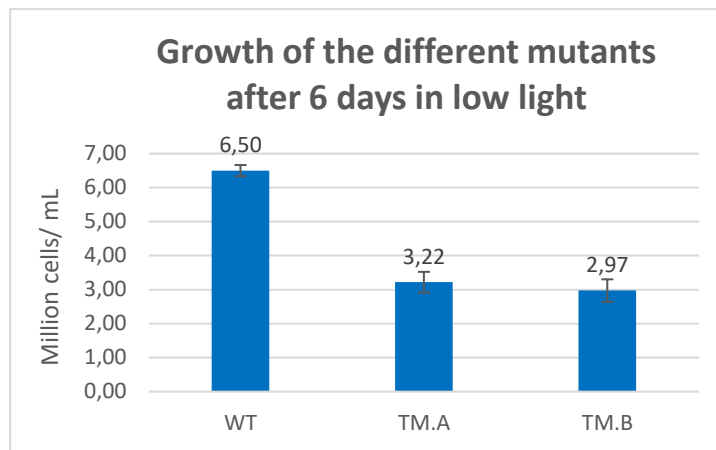


Figure 16. Characterization of the DXS-PSY-GGPPS mutants under low light. The cell density of the mutants that were simultaneously transformed with the genes *psy*, *dxs* and *ggpps* was determined and compared to the wild type (WT). The 2 clones obtained (A and B) are called TM, that stands for triple mutant. These cultures were grown under low light conditions ($50 \mu\text{E m}^{-2} \text{s}^{-1}$) for 6 days.

4. Screening system for the selection of mutants

4.1 Introduction

4.1.1 High-throughput screening of carotenoids

Random mutagenesis consists of a mutagenesis step followed by a screening process to select the mutants with the best phenotype. In this work, that would be the mutants with a higher production of carotenoids, more specifically, of fucoxanthin.

Usually the screening of hundreds to thousands of mutants is necessary to find phenotypes of interest. This implies that the use of a high-throughput screening method is vital in order to make the process viable regarding time and costs.

An important aspect to consider is the cultivation of the mutants. Due to the high number of mutants produced, it is not always possible their cultivation in flasks due to limitations in space availability. In addition, the handling and culturing of such a big number of mutants would make the process very laborious.

Therefore, the miniaturization of cultures in microplates is often the strategy of choice to solve the above-mentioned issues. However, one should be careful when using this approach and think that the results might not be comparable to those in flasks. For instance, different turbulence patterns are created in microplates and in flasks, due to the different geometry of the vessel and on the volume. Since the agitation has an important influence in culture aeration, the cell growth and cell morphologies might not be comparable in the different vessels. In addition, the low volumes of culture can give rise to artifacts due to unequal evaporation across the plate⁵⁹.

Regarding the quantification of carotenoids, the high-pressure liquid chromatography (HPLC) is recognized as the golden method for an accurate identification and quantification of carotenoids once they have been extracted

out of the cells. But for the screening of a mutant library, HPLC is not a good option due to the high time and economic cost involved. An alternative method is UV/visible light spectrophotometry, in which the measured absorbance values are used to calculate the total carotenoid levels through a formula, since they are directly proportional. This technique is not as accurate as HPLC and it does not allow to distinguish the levels of the different carotenoids, but supposes a good compromise because it enables the fast and simultaneous measure of many samples by the use of a microplate reader⁶⁰. Another approach is based on the high correlation between carotenoid and chlorophyll-a levels during exponential growth. In this strategy, a first screening process based on fluorescence detection of chlorophyll a is carried out and the most promising mutants are then selected for a final screening in which the pigments are extracted and the carotenoid levels analyzed⁵².

For the screening of mutants based on carotenoid levels, an extraction step is sometimes necessary for the posterior quantification. The extraction of pigments in algae usually involves a centrifugation step for the concentration of the cells, and then the use of organic solvents together with a cell disruption strategy, such as sonication, to extract the pigments and finally another centrifugation step to separate the cell debris from the supernatant containing the pigments. Thus, this additional extraction step also makes the high-throughput screening of the mutants more challenging in terms of time. Therefore, quantification approaches that do not require the extraction of the pigments would make the screening method more feasible.

4.1.2 Cell encapsulation

Cell encapsulation is one cell immobilization method in which cells are entrapped inside a semipermeable membrane. This membrane allows the pass of nutrients and oxygen and the elimination of waste products, but it does not permit the pass of bigger molecules, such as antibodies.

There are several crucial factors in the design of this method. It is necessary that the immobilization technique is non-toxic for the cells and that the

membrane is biocompatible. The size of the microbeads is also important since it must allow a good diffusion of nutrients in any point of the bead. The beads must be also mechanically stable to avoid that they burst⁶¹.

Cell encapsulation has found multiple applications in different biotechnology sectors and in biomedicine. Some applications of cell encapsulation are the following:

- The use of bioreactors using immobilized cells. The main advantages are: a high cell density is possible, protection of the cells from shear forces (agitation) and potential culture stresses, cell wash-out is avoided in continuous systems and facilitates the downstream processes. One has to make sure though that there are no mass transfer limitations⁶².
- Environmental applications. Immobilized cells are used in wastewater treatment plants due to the same benefits outlined before⁶³.
- Food industry. As an example, cell encapsulation of probiotic bacteria allows to increase their viability during the production of dairy products and to optimize the delivery in the gastrointestinal tract⁶¹.
- Applications in cell library and drug screening. For instance, after the directed evolution of a given enzyme, the cell library was encapsulated. The screening of the best variants was based in fluorescence-activated cell sorting.⁶⁴
- Applications in biomedicine. For instance, transplantation of encapsulated hepatocytes has been used to treat acute liver failure. Encapsulation of the cells avoids immune rejection, since the immune cells cannot reach the transplanted cells, while the transplanted hepatic cells contribute to the hepatic function⁶⁵. In addition, the beads provide a 3D environment that simulate better the geometric conditions of the extracellular matrix in the cell than 2D culture systems⁶⁶.

Recapitulating with the previous section about the screening of a cell library, encapsulation of the cell library would avoid the limitations encountered with the miniaturization of the cultures in microplates. In this system, every mutant cell would be encapsulated in a microbead and all the encapsulated cells would be cultured in the same flask. This approach makes the cultivation of

the mutants easier and provides homogeneity in the culture conditions. The combination of this method with a high-throughput detection system would make the screening of a mutant cell library much more feasible than conventional methods.

4.1.3 Encapsulation in alginate beads

One of the mechanisms for the formation of the microbeads is ionotropic gelation using alginate. Alginate is a polysaccharide that comes from the cell wall of brown algae and some bacteria. It is a hydrophilic linear polymer of D-mannuronic acid (M) and L-guluronic acid (G). The polymer consists of homo-polymeric blocks (M-blocks or G-blocks) and hetero-polymeric blocks (MG-blocks), with about 20 monomers per block⁶⁶.

When there are several polymer chains and 2 G-blocks are in front of each other, each in a polymer chain, they adapt a structure that can bind divalent cations, such as calcium. The calcium ions fit into the guluronic structures like an egg into an egg-box (**Figure 17**). Therefore, it is called the egg-box model of cross-linking with calcium ions. In this way, the different alginate chains are connected by the calcium ions, what leads to the gelation of the alginate⁶⁶.

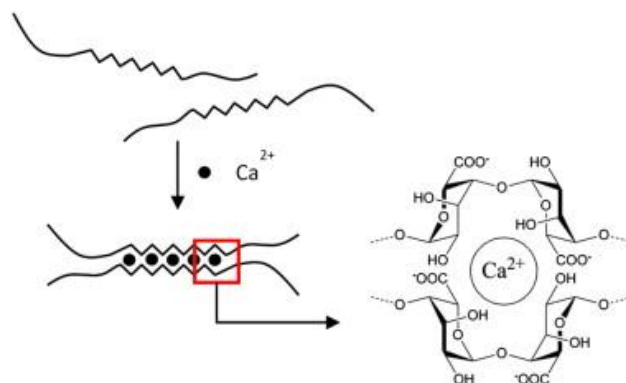


Figure 17. Mechanism of gelation of alginate. In the presence of calcium, two polymer linear chains bind calcium atoms and form a structure that resembles an egg-box. In this way, the cross-linking of the different polymer chains results in the gelation of alginate. Image from Paques et al⁶⁷.

4.2 Objectives

The objective is to set-up a high throughput method to screen the carotenoid levels in diatoms following the procedure represented in **Figure 18**. This is an important step to be able to screen a library of mutants in an efficient manner.

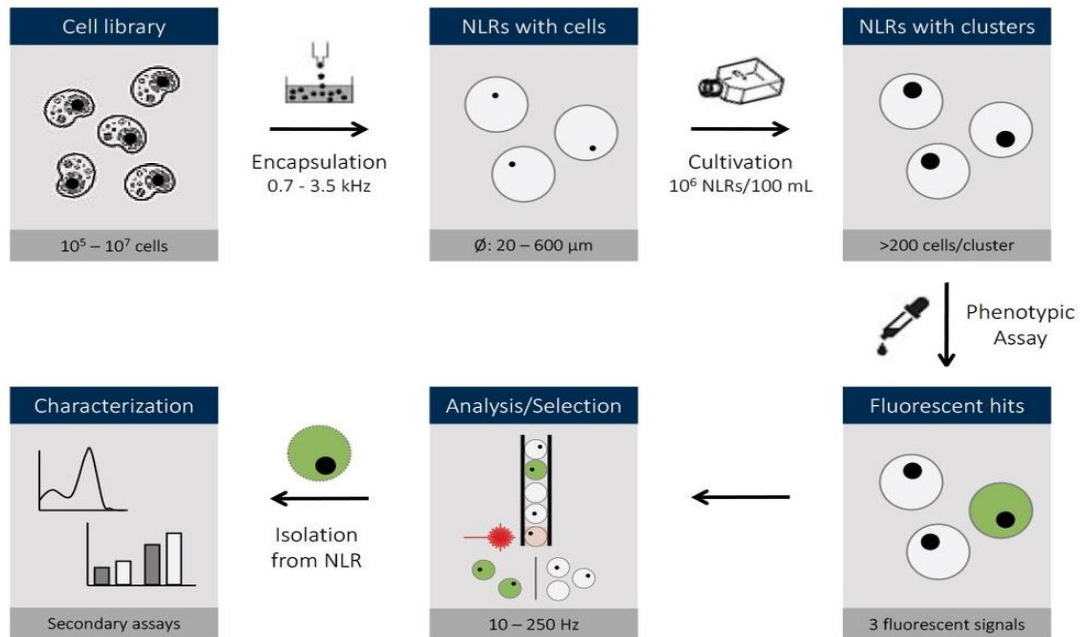


Figure 18. Screening of a cell library. The screening platform is based on the encapsulation of the cells of the cell library within alginate beads, also called nano-liter reactors (NLR). The cells are then cultivated and, in the end, they are analyzed with a flow cytometer in order to select the best-producing cells. This screening platform is implemented by the company Fgen (Switzerland), where this work was carried out. Figure from Fgen webpage⁶⁸.

The strategy was to encapsulate the diatoms within alginate beads (ideally 1 cell per bead) to have one pot culture where every cell can grow under the same conditions in the same flask. Only wild type cells are tested, since this is only a proof of concept work, but the idea is that if the protocol works, this method could potentially be used for the screening of a mutant cell library. The screening method consisted in the analysis of the biomass or the chlorophyll levels, measured by fluorescence emission by a flow cytometer for large particles (COPAS). This could serve as a first round of screening and once the promising

cells are identified, later they could be analyzed by HPLC to determine the levels of specific carotenoids. Unfortunately, carotenoids do not emit fluorescence and therefore, they cannot be quantified directly by the COPAS.

This work is only a preliminary effort that can serve as a proof of concept to determine if the encapsulation of diatoms can be carried out without negative effects on cell proliferation. For reaching this aim, 2 factors are mainly assessed:

- The stability of the alginate beads in the cell culture medium of the organisms.
- The proliferation and cell viability of the diatoms within the alginate beads.

4.3 Results

4.3.1 Bead stability in the culture media

First, the stability of the alginate beads in the standard culture medium of the microalgae was tested. If the beads are not stable, modifications in the composition of the medium will be carried out. The medium may contain ion chelants or phosphates that might interfere with the stability of the beads, which rely on calcium ions. A compromise should be reached between the growth of the algae and the correct performance of the high-screening system.

Alginate beads (diameter: 200 μm) were incubated with the different media for 20 minutes and then the bead diameter was measured with aid of a microscope. To have a representative measurement, the diameter of 100 beads was measured (**Table 1**). N8 is the medium of the green algae *Chlorella*, ASP is the medium of the diatom *Phaeodactylum* and silica-supplemented ASP is the medium of the diatom *Cyclotella*. CaCl_2 (10 mM) was used as a control to have a sample in which the beads are stable.

The N8 medium was very stable, showing a similar diameter to the control. In contrast, the beads in the ASP medium were 29.5% bigger than the control and the beads incubated in silica-supplemented ASP shrunk 31.6%. The problem is that bigger beads have higher chances to burst, while smaller beads might not

provide enough space for the cells to divide. Possible strategies to correct the bead diameter are to decrease the phosphates in the medium that might be interacting with the calcium or to increase the CaCl₂ content. However, further encapsulation experiments will use the same cultivation medium as long as no problems associated with bead burst or cell proliferation were observed.

Table 1. Bead stability test in the culture medium of the algae.

Bead Diameter	Mean (µm)	St Dev (µm)
CaCl ₂ 10 mM (Control)	274.06	7.49
N8 medium	277.51	8.56
ASP medium	354.84	11.85
ASP medium with silica	189.59	7.74

4.3.2 Cell proliferation over time after encapsulation

Cell proliferation was assessed after the encapsulation of the microalgae *Chlorella* and *Cyclotella* in alginate beads. The cell proliferation of *Phaeodacylum* could not be assessed since the culture got contaminated. Algae in an exponential stage of growth were used for the encapsulation. This experiment was repeated 3 times. In the first attempt, the cultures were grown by direct exposition to natural light in summer sunny days at 21°C for 3 days. In 2 other rounds, the cultures were grown at 50 µE m⁻² s⁻¹ using a LED lamp and at 21°C for 3 days.

For the quantification of cell proliferation over time, different variables were measured: colony diameter and number of cells. At least the growth in 20 beads per sample was assessed. This quantification was useful for the first day after

encapsulation, but as colonies got bigger, they adopted different irregular shapes with a high cell number that made the quantification by colony diameter or cell number very difficult. In **Figure 19** the proliferation of the microalgal growth under direct sunlight is represented by pictures that represent the different cell or colony populations at that time point.

4.3.2.1 *Chlorella* encapsulation and growth

At time zero, the cells that were in exponential phase were not forming colonies.

When *Chlorella* was grown under sun light conditions after encapsulation (**Figure 19**):

-One day after the cell encapsulation, there were three different populations: single cells, colonies of 6 cells and colonies of approximately 10-50 cells. Of those, the colonies of 6 cells were by far the most abundant. Colonies have a rounded shape.

-At day 2 after encapsulation, there were 4 populations: single cells, colonies of 5-20 cells, colonies of around 20-50 cells and colonies of more than 50 cells. Now, the colonies of 20-50 cells were the most abundant, followed by the single cells. The colonies of more of 50 cells were very scarce.

-At day 3 after encapsulation, most of the cells were still forming part of colonies of 20-50 cells and single cells, although more colonies of 50 cells could be seen.

In contrast, when *Chlorella* were grown at $50 \mu\text{E m}^{-2} \text{s}^{-1}$, the cells proliferated during the following days after encapsulation forming colonies, although at day 3, only colonies with around 20-50 cells were seen.

4.3.2.2 *Cyclotella* encapsulation and growth

At time zero, the cells are not forming colonies or they have divided into 2 cells that are still together.

When *Cyclotella* was grown under sun light conditions after encapsulation (**Figure 19**):

-At day 1 after encapsulation, the cells have divided and now form colonies of 4-8 cells that lie adjacent to each other in a row. 70% of the colonies are in the stage of 4 cells and 30% are of 7-8 cells.

-At day 2 after encapsulation, there are different populations of colonies with diverse colony shapes, but almost all colonies are beyond the stage of 4-8 cells.

-At day 3 after encapsulation, the colonies are considerably bigger with respect to the previous day. Many are close to the wall of the NLR, so probably they are going to be released out of the alginate bead.

In comparison, when *Chlorella* cells were grown at $50 \mu\text{E m}^{-2} \text{s}^{-1}$, cells proliferate over time and form colonies, but colonies do not reach such a high cell number compared with the encapsulation in which cells grow under direct solar light.

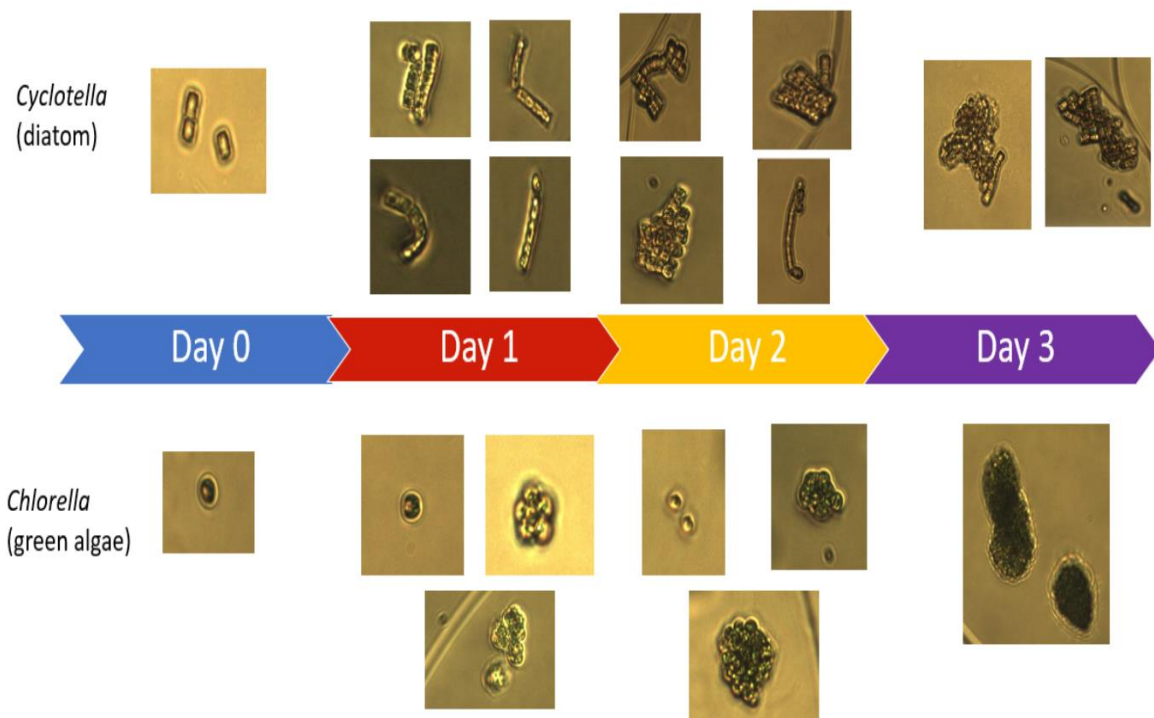


Figure 19. Cell proliferation of the encapsulated microalgae. After encapsulation in alginate beads, the proliferation of the microalgae *Cyclotella* and *Chlorella*, grown under direct sunlight, was monitored for 3 days. cell proliferation.

4.3.3 Cell viability within the alginate beads

The viability of the cells after cell encapsulation was monitored every day. For this experiment, the cells were grown at $50 \mu\text{E m}^{-2} \text{s}^{-1}$. To assess cell viability, the stain fluorescein diacetate (FDA) was added, so living cells were stained green. At least, 30 cells per sample were considered for the determination of cell viability.

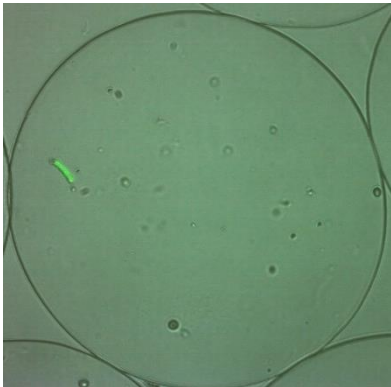


Figure 20. Cell viability monitoring. In this picture, a colony of several cells of *Cyclotella* encapsulated within an alginate bead shows green fluorescence after being stained with FDA. The colonies that show fluorescence indicate that they are alive.

In *Cyclotella*, at time zero, 97.5% of the cells were alive. The first three days after encapsulation, 100% of the cells were alive. For *Chlorella*, the strong fluorescent background hindered the precise cell viability determination the first days. At day 3 after encapsulation 81.25% of *Chlorella* cells were alive.

4.3.4 Determination of the cell occupation

Ideally only one cell would be encapsulated per alginate bead, also called nano-liter reactor (NLR), so proliferation would occur in every bead from only 1 cell.

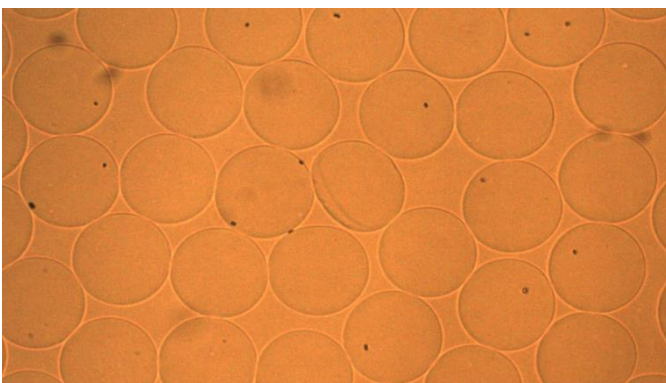


Figure 21. Determination of cell occupation. In this microscope image several NLR can be seen. At this magnification, the cells or cell colonies inside them are seen as black dots. The number of cells per NLR were counted after encapsulation to determine the cell occupation.

However, in reality, many NLRs are empty and in some NLRs more than one cell is encapsulated. For that reason, the cell occupation must be optimized so that most NLRs contain after encapsulation only one cell.

The Poisson distribution is a probability function that is used to describe the distribution of cells per NLR. A key parameter of this distribution is lambda, that it is the average number of cells per NLR. A lambda of 0.6 was fixed.

The cell occupation was determined in two different encapsulations by counting the number of cells present in each NLR after encapsulation (**Figure 21**). 100 events were considered. In **Table 2**, the theoretical cell occupation following the Poisson distribution with a $\lambda=0.6$ can be seen, while in **Table 3** and **Table 4** the observed cell occupations during two cell encapsulations are shown.

Table 2. Theoretical cell occupation following the Poisson distribution with a $\lambda=0.6$.

Cells/NLR	0	1	2	3	4
Theoretical	54.9%	32.9%	9.9%	2.0%	0.3%

Table 3. Cell occupation in the first encapsulation.

Cells/NLR	0	1	2	3
Chlorella	56.84%	26.32%	14.74%	2.10%
Cyclotella	40.00%	27.73%	26.89%	19.83%

For Chlorella, the distribution can be approximated to $\lambda=0.6$, but there is a little redistribution between the events with 1 and 2 cells/NLR, maybe due to insufficient homogenization of the cells with the alginate.

In contrast, for Cyclotella the cell occupation is higher than for $\lambda=0.6$, around $\lambda=0.8$, if we consider percentage of empty NLRs. There is also a redistribution between the groups of 1 and 2 cells per NLR. A possible explanation for the

higher occupation could be the incorrect counting of the cells to be added to the alginate.

Table 4. Cell occupation in the second encapsulation.

Cells/NLR	0	1	2	3
Chlorella	53.84%	40.38%	5.8%	-
Cyclotella	58.62%	27.59%	13.79%	-

In the second encapsulation, the Poisson distribution could be approximated to a $\lambda=0.6$ for Chlorella and $\lambda=0.5-0.6$ in Cyclotella, as intended.

4.3.5 Biomass monitoring

The biomass of Cyclotella was monitored over time by staining a sample with the live staining FDA and the fluorescent measurement by the flow cytometer (COPAS). Since the biomass increased over time within the NLR, therefore an increase of fluorescence is expected over time.

Firstly, for every sample, the correct population of NLRs (non- fused NLRs) is selected based on the graph "Count/TOF" (Data not shown). After that, for the measurement of the fluorescence, the graph "Count/Green Peak Height" is examined. At time zero, just after encapsulation, a first peak can be seen, which represents the empty NLRs (background fluorescence) and the rest of the peaks (R4), that show higher fluorescence, are the NLRs with the encapsulated cells (**Figure 22**). The mean fluorescence at time zero is of 1317 fluorescence units. After one day, the mean fluorescence has increased until 1877 fluorescence units (**Figure 22**). At day 3, the fluorescence increased to 2183 fluorescence units (**Figure 22**).

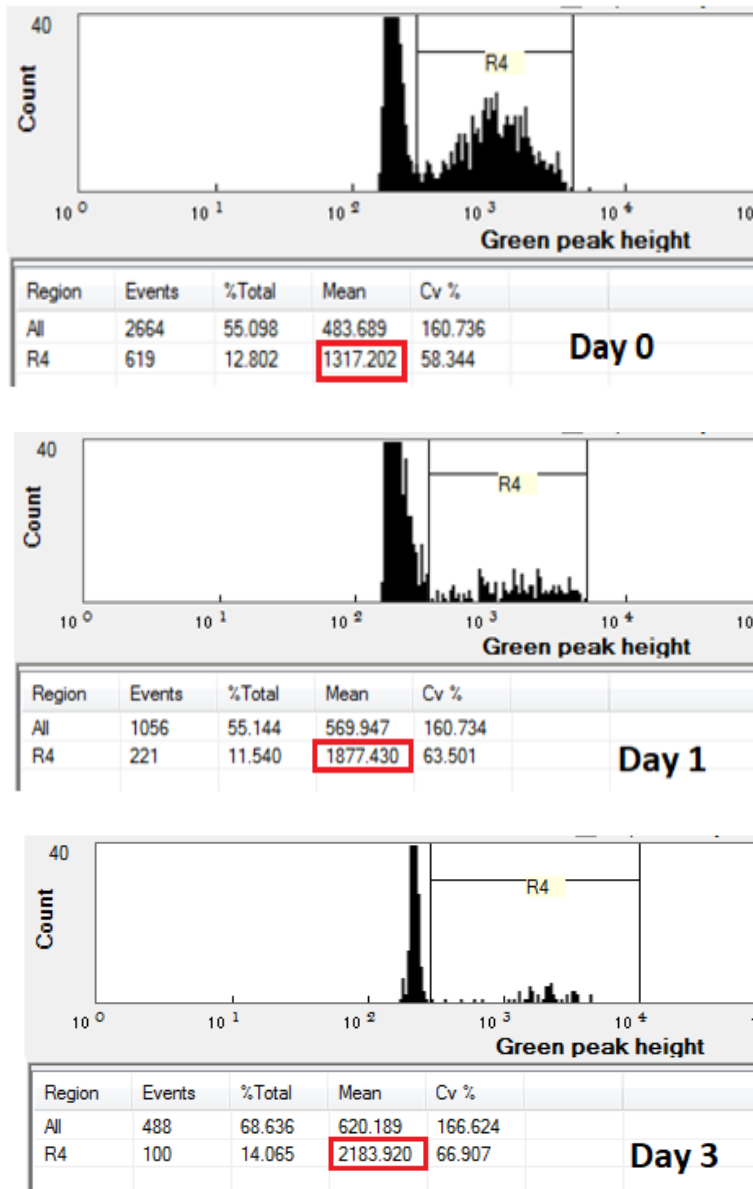


Figure 22. Biomass monitoring over time of the encapsulated cells. After staining with FDA, the biomass of encapsulated *Cyclotella* was registered at different time points for 3 days by fluorescence in the flow cytometer COPAS. Only the counts corresponding to occupied NLRs were taken into account (R4), while the background fluorescence was excluded. The Y-axis of the graphs represents the counts, that is the number of NLRs (with encapsulated microalgae) emitting a given intensity of fluorescence. The X-axis shows the intensity of green fluorescence in units of green peak height. The mean fluorescence of the biomass has been highlighted with a square.

5. Unknown enzymes of the pathway and comparative genomics

5.1 Introduction

The carotenoid pathway in diatoms is not fully characterized. One of the unknown enzymes in diatoms is β -carotene hydroxylase. This function has been assigned to several enzymes in plants, so homology searches could facilitate its search. In addition, the enzyme(s) that catalyze the formation of fucoxanthin is clade-specific, since fucoxanthin is only produced by heterokonts, haptophytes and some dinophyta and it has not been identified so far³⁰. The same applies for the enzyme diadinoxanthin synthase, that is also clade-specific, whose identity remains unknown.

5.1.1 Carotenoid diversity in algae

It is important to have an overview of the diversity of carotenoids in algae to try to identify carotenoids that are structurally similar or with the same functional groups that our carotenoid of interest, whose biosynthetic pathway is unknown. For that, it is also important to do this while considering the phylogenetic relationship among algae (**Figure 23**). A short review of the carotenoid distribution in algae is now presented.

We can only find α -carotene and its derivatives (eg. lutein, linoxanthin, siphonaxanthin) in some cyanobacteria, multicellular Rhodophyta, Cryptophyta, Chlorarachniophyta and Chlorophyta. In contrast, β -carotene is present across all algae lineages¹⁴.

- The main carotenoids in **cyanobacteria** are β -carotene, its hydroxyl derivatives (zeaxanthin, nostoxanthin), its keto derivatives (echinenone, canthaxanthin) and carotenoid glycosides (myxol 2'-glycoside, 4-ketomyxol 2'-glycoside). These carotenoids are not present in all cyanobacteria. Different species contain different carotenoids⁶⁹.

- **Red algae** have chlorophyll a (not b nor c) and do not present unique carotenoids. Unicellular red algae contain only β -carotene and zeaxanthin, while the macrophytic red algae contain also α -carotene and lutein¹⁴. Besides, violaxanthin (part of the xanthophyll cycle in plants and green algae) has been found in a very few red algae orders⁷⁰.
- **Heterokonts, Haptophyta and Dinophyta** have chlorophyll a and c and contain, except for Eustigmatophyceae, diadinoxanthin and diatoxanthin. These acetylenic carotenoids form a photoprotective xanthophyll cycle in these algae. Fucoxanthin is found in some heterokonts, dinoflagellates and haptophytes, while peridinin and its derivatives are only present in dinoflagellates¹⁴.

It is interesting to observe how these organisms, that come from a secondary endosymbiosis of a red algae, contain some carotenoids derived from zeaxanthin, like violaxanthin and neoxanthin, that are also present in green algae, but not in red algae. In addition, other carotenoid proteins are more similar to the green algae homolog than to the red algae one. Therefore, it has been proposed that some green algae genes were captured by the chromist host before the endosymbiosis of the red algae took place⁷¹.

- Like heterokonts, **Cryptophyta** contain chlorophyll a and c. But unlike them, cryptophyta have α -carotene and acetylenic derivatives unique to this lineage, such as crocoxanthin and monadoxanthin¹⁴.
- **Euglenophyta, Chlorarachniophyta and Chlorophyta and land plants** contain chlorophyll a and b, β -carotene and its derivatives (violaxanthin and 9'-cis neoxanthin) and lutein, that is an α -carotene derivative. In addition, some of the algal classes contain unique carotenoids produced from lutein, such as loroxanthin, siphonaxanthin and prasinoxanthin¹⁴.

Division Class	Carotene		Xanthophyll											Chlorophyll		
	β	α	Ze	Vi	Ne	Da	Dd	Fx	Va	Lu	Lo	Sx	Other xanthophyll(s)	a	b	c
Cyanophyta	H	L	H										No, L; Ec, H; My, H	H	L	
Glaucophyta	H		H											H		
Rhodophyta																
Unicellular type	H		H											H		
Macrophytic type	L	L	H	L				L		H				H		
Cryptophyta		H	L										Al, L; Cr, L; Mo, L	H		H
Heterokontophyta																
Chrysophyceae	H		L			L	L	H	L					H		H
Raphidophyceae	H		H	L		L	L	L						H		H
Bacillariophyceae	H		L			L	L	H						H		H
Phaeophyceae	H		H	H		L	L	H						H		H
Xanthophyceae	H		L			H	H					Va-FA, L		H		H
Eustigmatophyceae	H			H					L					H		
Haptophyta	H		L			L	H	H					Fx-FA, L	H		H
Dinophyta			L		L			L	H	L			Pe, H		H	H
Euglenophyta			H		L		L	L	H				L, L		H	H
Chlorarachniophyta			H		L	L	L					L, L	Lo-FA, L		H	H
Chlorophyta																
Prasinophyceae	H		L	L	H	H					L	L	H	Pr, L; Lo-FA, L; Sx-FA, H	H	H
Chlorophyceae	H		H	L	H	H					H	L	L	Sx-FA, L		H
Ulvophyceae	H		L	L	H	H					L	L	L	Sx-FA, H		H
Trebouxiophyceae	H			L	H	H					H					H
Charophyceae	H			L	H	H					H					H
Land Plants			H	L	L	H	H						H			H

Figure 23. Carotenoids in different algae and in land plants. Abbreviations: H, High content; L, Low content in most species. a, α -carotene; β , β -carotene; Al, alloxanthin; Cr, crocoxanthin; Da, diatoxanthin; Dd, diadinoxanthin; Ec, echinenone; -FA, fatty acid ester; Fx, fucoxanthin; Lo, loroxanthin; Lu, lutein; Mo, monodoxanthin; My, myxol glycosides and oscillol glycosides; Ne, neoxanthin; No, nostoxanthin; Pe, peridinin; Pr, prasinoxanthin; Sx, siphonaxanthin; Va, vaucheriaxanthin; Vi, violaxanthin; Ze, zeaxanthin. Red, α -carotene and its derivatives. Image from Takaichi¹⁴.

After this overview of the carotenoids in algae, now it is time to take a look at the carotenoids that present the same functional groups as fucoxanthin, diadinoxanthin and diatoxanthin.

Fucoxanthin has a unique structure that contains an allene group, an acetyl group and a keto group in the polyene chain.

- Allene groups (C=C=C) are found also in neoxanthin (in land plants, green algae, diatoms...), 19'-acyloxyfucoxanthin (in Haptophyta and some dinoflagellates), peridinine and dinoxanthin (in dinoflagellates) and vaucheriaxanthin (in some lineages of heterokonts) (**Figure 24**). All these carotenoids are derived from neoxanthin, and therefore contain this functional group¹⁴.
- Acetyl groups (OCOCH₃) are also present in peridinine, dinoxanthin and pyrroxanthin (in dinoflagellates) (**Figure 24**)¹⁴.
- Keto groups (C=O) are present in several carotenoids. For instance, astaxanthin, cantaxanthin and echinone contain keto groups in the carotenoid ring. Keto groups are also part of the lactone ring that is present in some carotenoids such as uriolide, pyroxanthin, peridinine and peridinol. However, the keto group of fucoxanthin is present in the C8, that is directly in the polyene chain. In the same position a keto group is also found in the carotenoids prasincoxanthin and siphonoxanthin, present in some green algae (**Figure 25**)⁷¹.

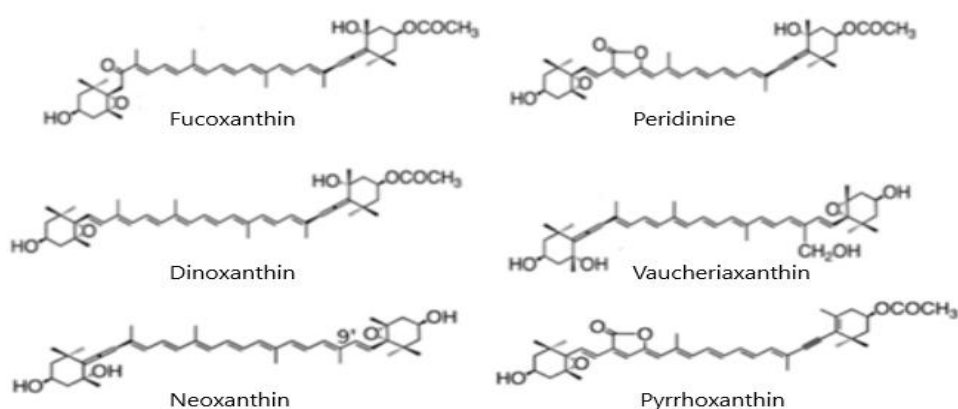


Figure 24. Carotenoids with either allene or acetyl groups, or both. Image adapted from Takaichi¹⁴.

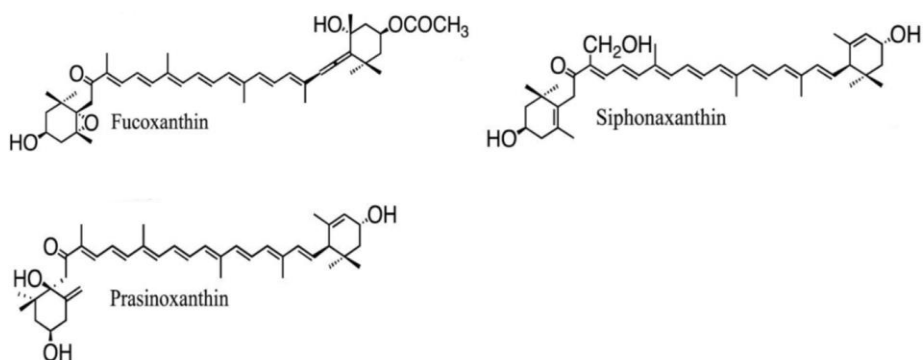


Figure 25. Carotenoids containing a keto group in the C8. Adapted from Takaichi¹⁴.

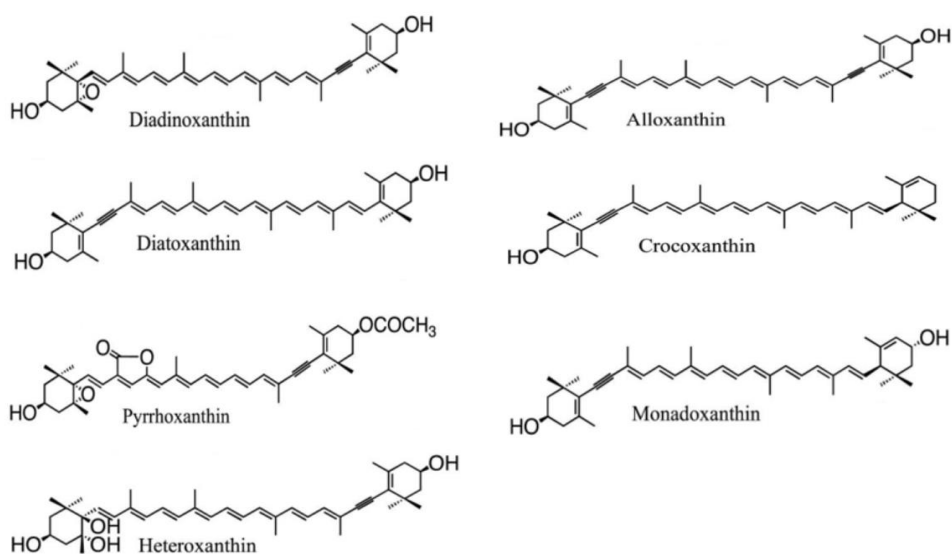


Figure 26. Carotenoids containing an acetylene group. Adapted from Takaichi¹⁴.

Diadinoxanthin and **diatoxanthin** have a unique structure containing an acetylene group ($C\equiv C$). These carotenoids are present in heterokonts, Haptophyta, Dinophyta and Euglenophyta. Other carotenoids containing an acetylene group are alloxanthin, crocoxanthin and monadoxanthin, that are present in Cryptophyta, and heteroxanthin, present in Xantophyceae, a lineage from heterokonts (**Figure 26**)¹⁴.

Unfortunately, the enzymes necessary for the formation of the above-mentioned carotenoids with the same functional groups as fucoxanthin, diadinoxanthin and diatoxanthin have also not been identified. But maybe once one of these enzymes

is discovered, the other enzymes catalyzing the same reaction in the other organisms are easily identified.

5.1.2 Strategies for linking functions to sequences

The simplest and most effective method to link functions to sequences is based on **homology searches**. Proteins with similar amino acid sequences are predicted to have similar structures. Regions (or protein domains) with similar structure often perform similar or identical functions in different organisms that come from a common ancestor⁷². Therefore, when the function of the protein of interest has been assigned before in another organism, a sequence homology search can be carried out with programs such as BLAST between the known protein sequence in a given organism and the protein sequence database of the organism of interest.

However, homologous-based methods sometimes produce incorrect annotations of genes and in addition they cannot be used for enzymes whose sequence is unknown in any organism (orphan enzymes). In this situation, **comparative genomics** might provide helpful hints. Like homology-based searches, comparative genomics relies on the "guilt by association" principle, that is based on the prediction of the gene function based on some type of association between known and unknown genes of the same pathway (**Figure 27**)⁷². Some useful genomic associations are now described:

- Gene clustering

Gene clustering can be very useful for the identification of enzymes. This approach is especially helpful in prokaryotes, where genes coding for enzymes of the same pathway or with a similar function are sometimes arranged in operons. The genes in the operon are transcribed from a single promoter that controls the expression of all of them. However, in eukaryotes gene location in the genome is not random, and genes with a similar or a coordinated expression can be clustered or in proximity in the genome⁷². The reason for that lies in that location in the genome affects expression and in the role of chromatin structure

in the activation or silencing of genes. It is important to consider that cluster sizes vary widely across eukaryotes, depending on the genome structure of the organism. While in some organisms like yeast the genes in a cluster are adjacent to each other, in other organisms gene clusters can span from tens to hundreds of kb, with many other unrelated genes in between⁷³.

- Gene fusions

In case that a gene of undetermined function forms a gene fusion with another gene whose function is known, a hint in the function of the unknown gene could be predicted. This is due to the fact that fused genes are commonly between genes of the same pathway or between genes that code for proteins that will form a complex⁷².

- Shared regulatory sites

Transcription factors recognize specific DNA sequences present in the promoter of different genes and regulate the function of those genes. Genes of the same pathway are often regulated by the same transcription factor. Thus, finding shared regulatory sites is a good approach to find functional associations among genes⁷².

- Co-expression

Often the genes of the same metabolic pathway present a similar expression pattern under different conditions. The construction of co-expression networks, based on transcriptomic techniques such as DNA microarrays and RNA-sequencing, can help in the identification of unknown genes⁷².

- Protein-protein interactions

In addition to transcriptomics, proteomics can also reveal functional associations based on protein-protein interaction experiments⁷².

- Phylogenetic occurrence profiles

This approach is based in the principle that enzymes present in the same pathway are either all present or absent in an organism. Considering the

presence/absence patterns of genes in the genomes in different organisms can help to identify candidates for orphan enzymes⁷².

- Protein structures

Structural genomic projects have aided the determination of many enzyme functions. Sometimes two proteins do not show homology at the sequence level, but their three-dimensional structure reveals some kind of homology that cannot be detected at the sequence level, for instance by fold recognition. In addition, knowledge about the substrate binding site structure can be very useful to assign functions. Therefore, protein structures can provide basic information about function, although often they cannot tell exactly the precise function. Even when the structure of a protein is not available, it can be inferred by structure prediction algorithms⁷².

- Phenomic data

Large scale knockout libraries together with databases on knockout phenotype can associate the function of a gene to a particular phenotype. However, not all gene deletions produce a visible phenotype and appropriate high throughput techniques should be selected. In addition, if the gene is essential the organism will die and then no phenotype conclusion can be drawn, other than it is an important gene that is essential for a given function⁷².

- Metabolic reconstruction

Having a genome-scale metabolic reconstruction, one can detect easily pathways that are not connected by an enzyme. There are several gap-filling algorithms available based in different principles that can help to assign gene functions by using the genome-scale model⁷².

- Enzyme location

Finally, in eukaryotes information about the subcellular location of the unknown enzyme can rule out putative sequences based on the presence or absence of organelle-specific target peptides⁷².

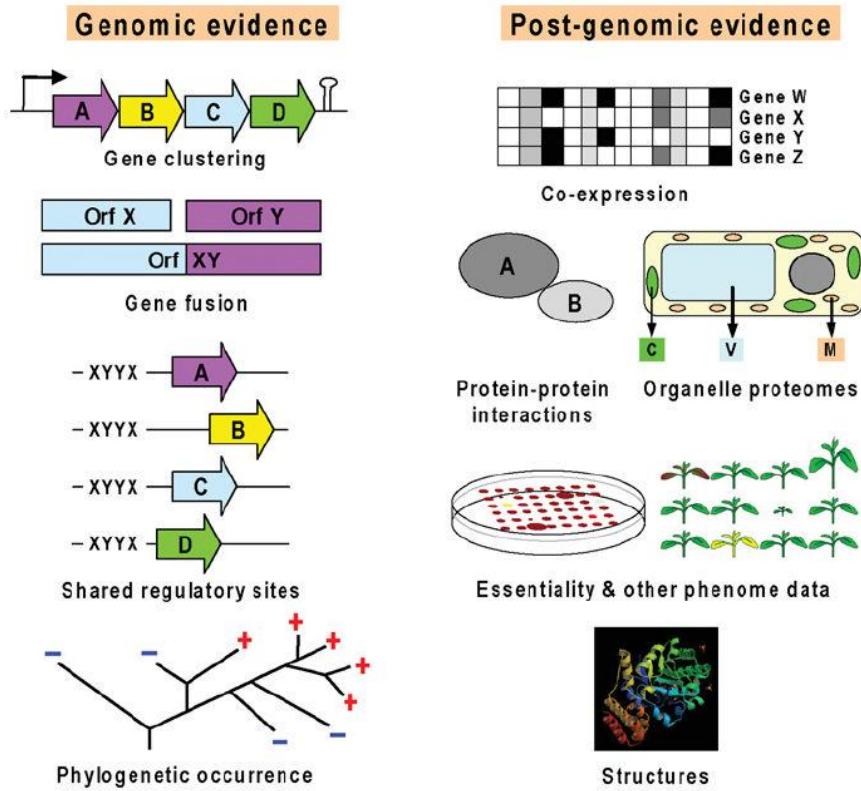
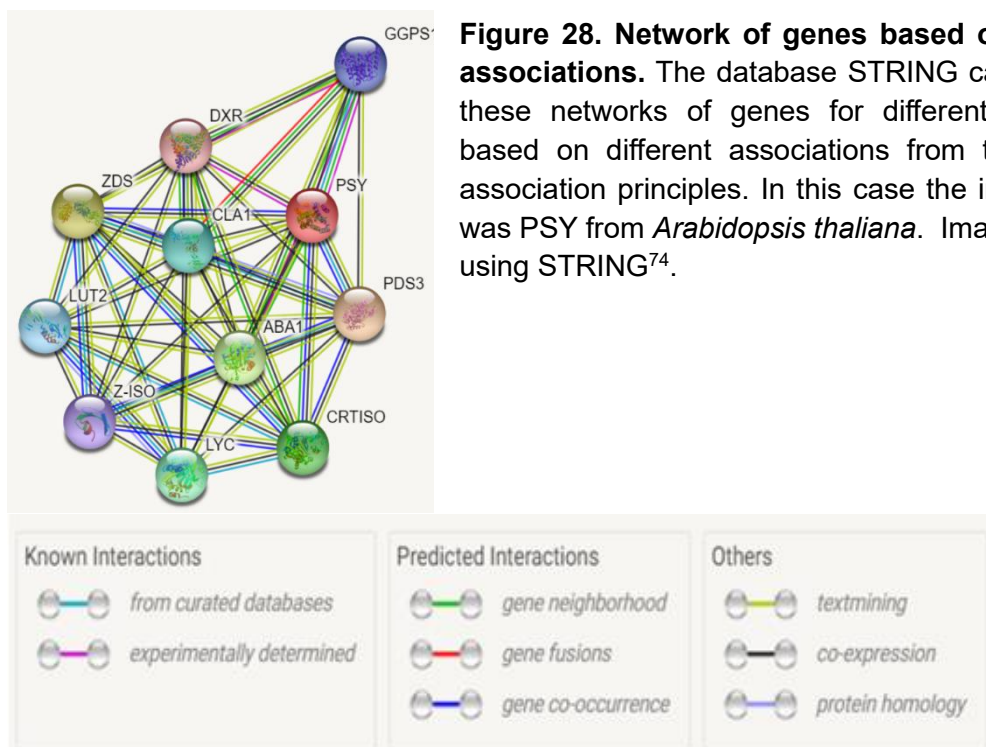


Figure 27. Guilt by association principles for the identification of unknown enzymes based on genomic and post-genomic evidences. Image from Hanson et al⁷².



Some of these associations, coming from experimental or bioinformatic data, are collected into the database STRING, which upon search of a given gene by the user provides a functional network of genes based upon these associations (**Figure 28**)⁷⁵.

5.1.3 Missing enzymes of the carotenoid pathway in diatoms

Several functions of the carotenogenesis pathway in diatoms have not been assigned to a sequence. The enzymes that have not been yet identified are: β -carotene hydroxylase, fucoxanthin synthase and diadinoxanthin synthase.

The enzymes that catalyze the formation of fucoxanthin and diadinoxanthin are clade-specific, since these carotenoids are only present in heterokonts, haptophytes and some dinoflagellates. In this case, sequence homology methods cannot be applied, since the enzymes have not been identified in any organism.

The formation of fucoxanthin from neoxanthin involves a ketolation reaction at C8, that could first consist in a hydroxylation reaction followed by a keto-enol tautomerism. The latter could take place spontaneously. Another reaction is also needed for the transformation of neoxanthin in fucoxanthin, and that is the acetylation at C3' (**Figure 29**)¹³.

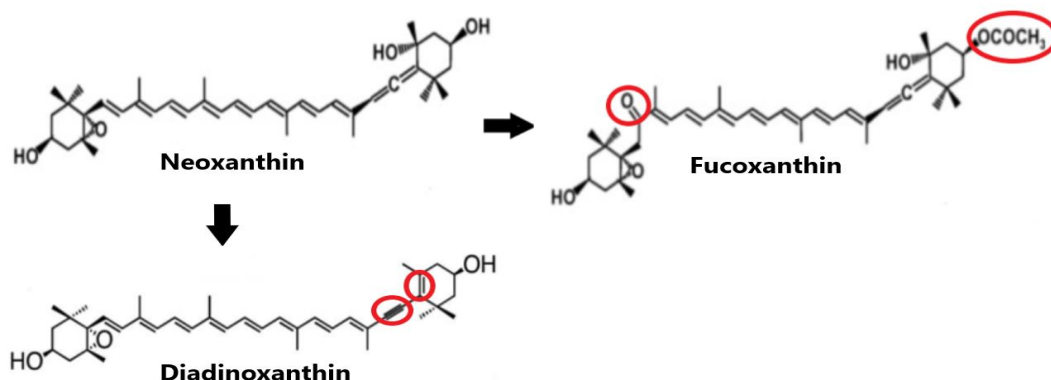


Figure 29. Formation of fucoxanthin and diadinoxanthin from neoxanthin. The chemical groups introduced or the chemical modifications created by these reactions are shown in a circle. Image adapted from Tackaichi et al¹⁴.

The transformation of neoxanthin into diadinoxanthin involves the creation of an acetylenic bond from an allene group by elimination of the 5-hydroxyl group as water (**Figure 29**)¹³.

The reaction of β -carotene hydroxylase is not described in this section, because it will be reviewed in the next chapter.

5.2 Objectives

The objective is to find candidate genes that could catalyze the reactions for the formation of fucoxanthin and diadinoxanthin in the diatom *Phaeodactylum tricorutum*. The approach was to apply the guilt by association principle, more specifically the gene order and the gene co-expression concepts. The first step consists in the characterization of the carotenoid and the MEP pathway to find out if the genes of these pathways share these principles, i.e. if they are closely located in the chromosome and if they are co-expressed under the same conditions. Then, the next step is to find candidate genes belonging to the same metabolic pathway based on these principles and additionally considering the enzymatic function for the reactions under investigation and other features shared by the genes of the pathway, like enzyme location and presence in different phylogenetic groups.

5.3 Results

5.3.1 Characterization of the gene order in the MEP and carotenoid pathways

The information about the gene location of the genes from the MEP and the carotenoid pathway was collected from the ncbi/gene database. In addition, also it was also investigated the location of the genes from the main central metabolic pathways (Glycolysis, Krebs Cycle, Photosynthesis (including Calvin Cycle),

Pentose Phosphate Pathway (PPP) and Lipid biosynthesis) and from pathways related to the carotenoid function, such as chlorophyll biosynthesis, in order to find out if they are clustered with the genes of the carotenoid pathway or with themselves.

As it can be seen in **Figure 30** some of the carotenogenesis and the MEP genes of *P. tricornutum* are clustered or located at a close distance. Generally, unlike in prokaryotes, genes are not directly adjacent, but are separated by many genes, except for *vde-zep3* and *zep1-vdl2*. Genes of the carotenoid and MEP pathways occur with an average of 3-4 genes per cluster, with each cluster located in a different chromosome. These clusters are interspersed with genes from other pathways, some of them with related functions, like e.g. genes for the chlorophyll biosynthetic pathway or encoding photosynthetic proteins, but also from genes with a completely different function.

However, not all the genes from the carotenoid and MEP pathways are found in clusters or at close distance of other related genes. For instance, the phytoene synthase gene and the violaxanthin de-epoxidase-like1 genes are located in chromosome 5 and 9, respectively, and they are not at a close distance (less than 100 kb) to other carotenogenesis gene. See **Appendix** for the location of all the genes of the carotenogenesis and MEP pathways in *P. tricornutum*.

In **Figure 30** the location of some genes from the central metabolic pathways that are closely located in the chromosome are shown. As it can be seen, the genes from the Calvin pathway are often located in the neighborhood of genes coding for proteins of the photosynthetic apparatus.

The fact that some genes from the carotenogenic and MEP pathways are clustered in *P. tricornutum* raises the question about the conservation of gene order for the genes of these pathways among other photosynthetic algae. Particularly, it would be interesting to observe if there is gene clustering in other organisms that produce fucoxanthin. This could also aid to identify the missing enzymes in the carotenoid pathway. Unfortunately, the chromosome location of the genes in some diatoms (eg, *Fragilariopsis*), brown algae (eg, *Ectocarpus*) or haptophyte (eg, *Emiliana huxleyi*) is not known. The organisms in which the

clustering of the carotenogenic and MEP genes was assessed were: *T. pseudonana*, *C. merolae* and *O. tauri*.

T. pseudonana is a representative of another class of diatoms. Three “clusters” could be identified, that are, however, not identical to those shown for *P. triornutum*. The word cluster for these findings is not very appropriate, since in some cases only 2 genes are closely located and the distance among the genes is not that close, in some cases more than 100 Kb apart, for a eukaryote with a compacted genome.

- Chromosome 2: *zep* (1117 Kb) – *vdI2* (1187 Kb)- *lcyb* (1230 Kb).
- Chromosome 5: *hdr* (1227 Kb) - *psy* (1330 Kb)
- Chromosome 8: *dde* (841 Kb) - *hds* (998 Kb)

For red algae only the gene location for one organism is available, the primitive algae *C. merolae*. Here, only two genes each are found in close proximity.

- Chromosome 8: *cmk* (133 Kb) - *cms* (278 Kb)
- Chromosome 11: *lcyb* (143 Kb) – *ggpps* (167 Kb).

O. tauri is a member of the group of Prasinophyceae of green algae. Here three “clusters” could be found.

- Chromosome 12: *ggpps* (211 Kb)- *crtiso* (303 Kb).
- Chromosome 14: *lcy* (117 Kb)-*crtiso* (207 Kb)
- Chromosome 16: *zds* (90 Kb) –*vde* (103 Kb) – *zep* (105 Kb)

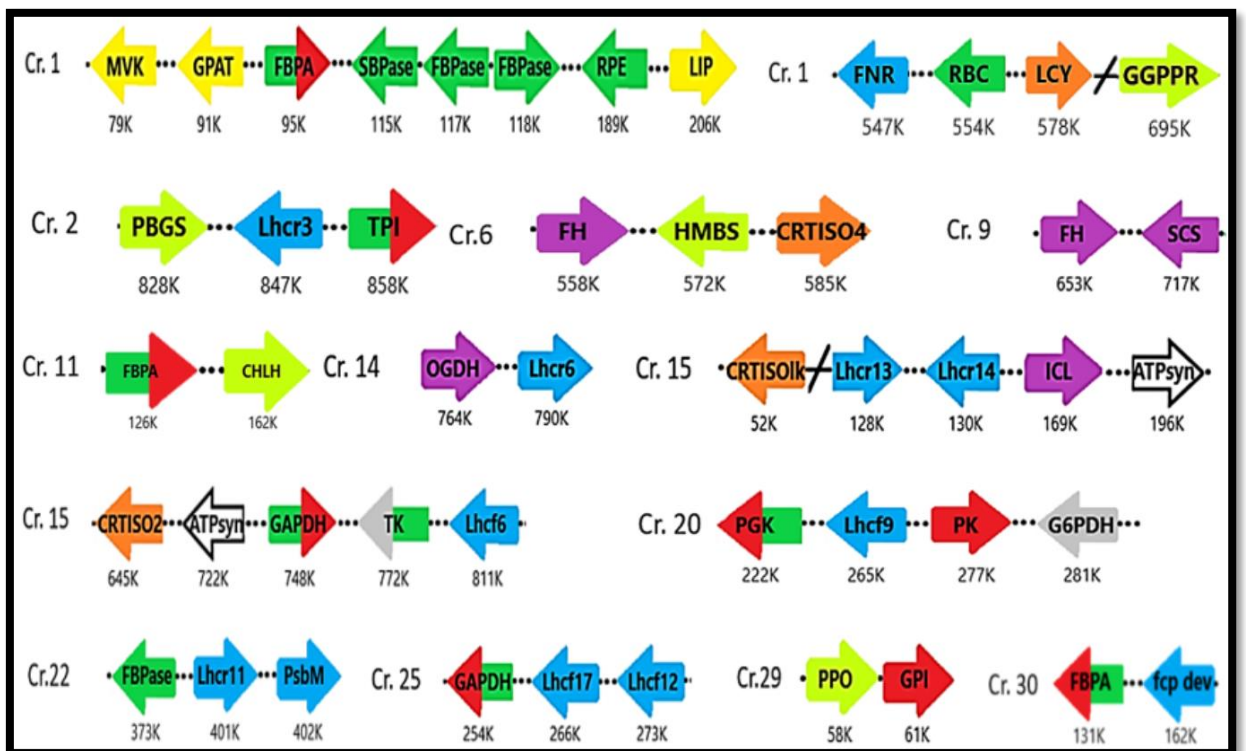
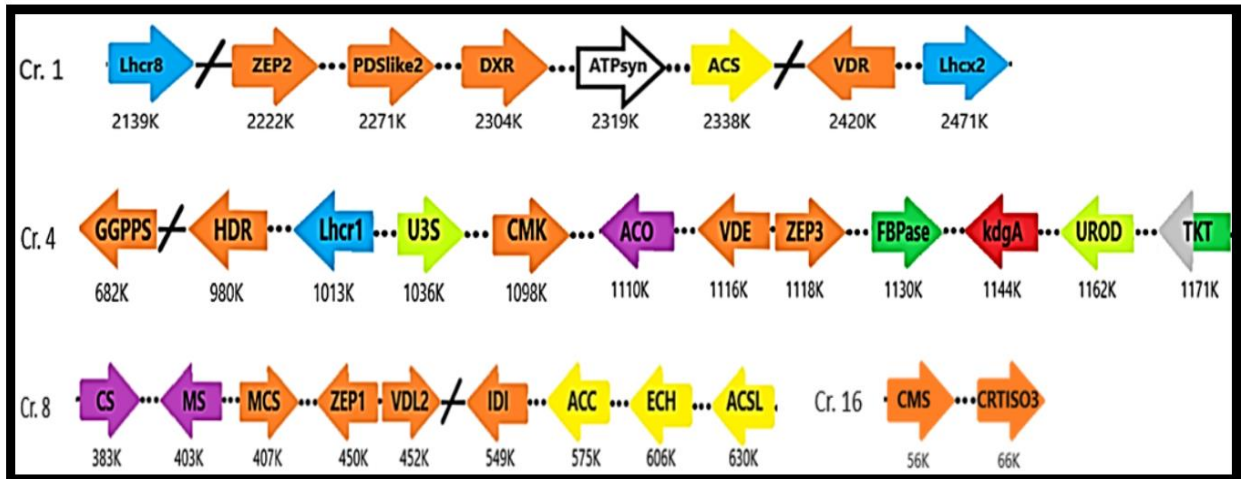


Figure 30. Gene clustering in *P. tricornutum*. In the top figure, the gene clusters in the chromosome (Cr.) from the genes involved in the MEP and carotenoid pathways are represented. In the bottom figure, some clusters of genes from the central metabolic pathways are displayed. Arrows represent genes involved in the following metabolic pathways: Carotenogenesis and MEP (**orange**), Photosynthesis (**blue**), Calvin cycle (**dark green**), Chlorophyll biosynthesis (**pale green**), lipid biosynthesis (**yellow**), Krebs cycle (**purple**), Glycolysis (**red**), Pentose phosphate pathway (**grey**). The abbreviations of the genes of the carotenoid and MEP pathways can be found in the Abbreviation section. Abbreviations of other genes: **Lhc**, light harvesting complex; **ATPsyn**, ATP synthase; **U3S**, Uroporphyrinogen III synthase; **Aco**, Aconitase; **FBPase**, Fructose 1,6-bisphosphatase; **kdgA**, 2-dehydro-3-deoxy-phosphogluconate aldolase; **UROD**, Uroporphyrinogen decarboxylase; **TKT**, transketolase; **CS**, Citrate synthase; **MS**, Malate synthase; **ACC**, Acetyl-CoA carboxylase; **ECH**, Enoyl-coa hydratase; **ACSL**, Long chain acyl-CoA synthetase; **MVK**, Mevalonate kinase; **GPAT**, glycerol 3P acyltransferase;

FBPA, Fructose 1,6- biphosphate aldolase; **SBPase**, Sedoheptulose biphosphatase; **RPE**, Ribulose-phosphate-3 epimerase; **LIP**, lipase 3; **FNR**, Ferredoxin oxidoreductase; **RBC**, rubisco; **GGPPR**, Geranylgeranyl diphosphate reductase; **PBGS**, porphobilinogen synthase; **TPI**, Triosephosphate isomerase; **FH**, fumarase; **HMBS**, Porphobilinogen deaminase; **SCS**, Succinyl-coA synthetase; **CHLH**, Protoporphyrin X Mg chetalase; **OGDH**, oxoglutarate dehydrogenase; **ICL**, Isocitrate Lyase; **GAPDH**, Glyceraldehyde 3P dehydrogenase; **PGK**, Phosphoglycerate kinase; **PK**, Pyruvate kinase; **G6PDH**, Glucose 6-P dehydrogenase; **PPO**, Protoporphyrinogen oxidase; **GPI**, Glucose-6-phosphate isomerase. Data from gene location in chromosome from ncbi/gene webpage⁷⁶.

5.3.2 Characterization of gene co-expression in the MEP and carotenoid pathways

The RNA-seq data from Dr. König⁷⁷ was used to assess the co-expression of the carotenogenic and MEP genes. The conditions of the cells from which the RNA was isolated was dark acclimation followed by 1h of blue light. As outlined in the introduction, light plays an important role in the regulation of the carotenogenesis pathway.

DXS	3,8			
DXR	1,3			
CMS	1,3			
CMK	NK			
MCS	-1,1			
HDS	0,2			
HDR	NK			
IDI pl	5,2			
GGPPS	5,0	-2,2	0,1	
PSY	2,2			
PDS	3,0	-0,4	-1,5	
ZDS	3,1	-0,6		
CRTISO	5,3	1,9	1,6	1,3
	0,9	-4,7	-5,0	-6,3
LCYB	1,7			
LUT	4,4	0,6		
ZEP	5,8	3,4	NK	
VDE	4,4	-0,1	-0,7	
VDL	4,3	NK		

Figure 31. Expression levels of the genes from the MEP and carotenoid pathways after a shift from dark to blue light. The normalized log₂ fold changes for the genes in the MEP (left) and carotenoid (right) pathways are showed. The genes appear in the order in which their coding products participate in the pathway. When there is more than one gene for the same function, they are arranged one next to the other horizontally. Red (Blue) values represent a big increase (decrease) in expression levels in blue light compared to dark conditions, while white values represent either genes with unknown (NK) values or genes whose expression levels are not affected from this shift in conditions. This data has been extracted from a RNA-seq experiment conducted by Dr. Sarah König⁷⁷.

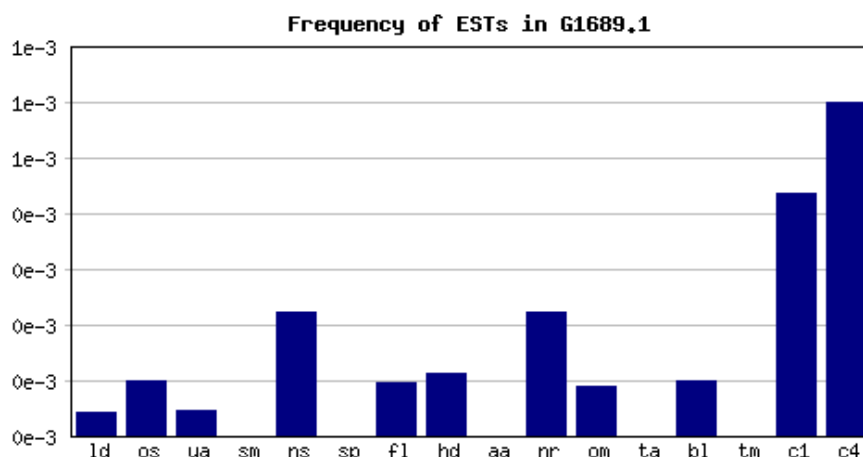
As seen in **Figure 31**, all the genes from the carotenogenesis pathway are highly over-expressed when dark-acclimated cells are exposed to light in comparison to

cells in dark conditions. Not all the isoenzymes are regulated by light, but at least one of them it is. Surprisingly, the *psy* gene, catalyzing the first enzyme of the pathway shows one of the lowest fold change compared to the other genes of the pathway. Among the highest expressed genes under the above-mentioned conditions are the *zep* and *crtiso* genes.

In contrast to the carotenogenesis pathway, not all the genes that code for the enzymes of the MEP pathway are over-expressed after light exposure compared to dark conditions. Only the genes *dxs*, *idi* and *ggpps* are over-expressed, that are the genes that code for the enzymes participating in the first and last reactions of the pathway. Some of the genes whose gene products catalyze the middle reactions of the MEP pathway were not present in the cDNA library, while others were not over-expressed (**Figure 31**).

In addition to blue light, one can assess other environmental conditions for which the genes of the MEP and carotenogenesis pathway are co-expressed (**Figure 32**). For that, one can use transcriptomics data from the public repository Gene Expression OMNIBUS⁷⁸ or from the diatom expressed sequence tag database⁷⁹.

For instance, the genes *cms*, *idi*, *ggpps* and *psy* have very high levels of expression under iron limitation, followed by other genes, like *hds*, *zds*, *vde* and *zep1* with high levels. Although not every gene from the carotenogenesis pathway shows high levels under these conditions, many of them do. Therefore, the expression under iron limitation could be another clue to make functional associations.



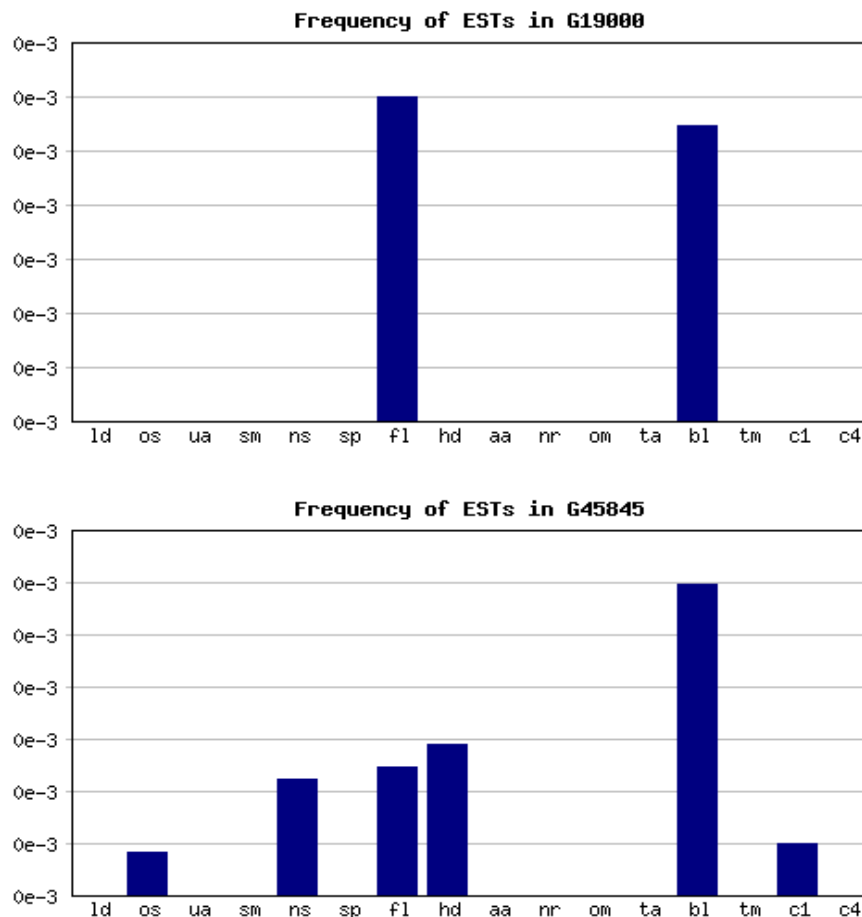


Figure 32. Expression levels of the genes *dxs*, *ggpps* and *zep1* under different environmental conditions or with different phenotypes. The frequency of expressed sequence tags (EST) is shown for the genes *dxs* (top graph), *ggpps* (middle graph) and *zep1* (bottom graph) under different conditions or with different phenotypes. **Ld**, Low decadienal treated (0.5 µg/ml decadienal for 6h); **os**, original 12000 standard (Scala et al. 2002); **ua**, urea adapted (50 µM of urea); **sm**, silica minus (artificial seawater); **ns**, nitrate starved (50 µM nitrate for 3 days in chemostat culture); **sp**, silica plus (350 µM metasilicate in artificial seawater); **fl**, iron limited (5 nM of iron); **hd**, high decadienal treated (5 µg/ml decadienal for 6h); **aa**, ammonium adapted (75 µM of ammonium); **nr**, nitrate repleted (1.12 mM of nitrate in chemostat culture); **om**, oval morphology (English strain Pt3 (CCAP1052/1B) in low salinity (10%); **ta**, tropical accession Pt9 (Cold stress, grown at 15°C); **bl**, blue light (48 hours dark-adapted cells exposed to 1 hour of blue light); **tm**, triradiate morphotype (Canadian strain PT8); **c1**: CO₂ high 1 day (230 µM/kg of CO₂ for 1 day in chemostat culture); **c4**: CO₂ high 4 day (230 µM/kg of CO₂ for 4 days in chemostat culture). These graphs have been retrieved from the EST database for *P.tricornutum*⁷⁹.

For the high CO₂ condition, many genes of the MEP pathway were highly expressed, like *dxs*, *cms*, *hds* and *idi*, but in contrast, from the genes of the carotenogenesis pathway only *vdI2* showed very high levels under these conditions. Thus, this condition cannot be used for functional associations.

5.3.3 Identification of putative genes for the fucoxanthin synthase function

In order to look for candidates of the fucoxanthin synthase enzyme, the first thing to be done was a search in NCBI/gene by the general class of enzymes based on their **function**, in this case hydroxylase and acyltransferase or more specifically, acetyltransferase, for *P. triornutum*. The candidates were short-listed, excluding all the genes for which the reaction they catalyze is known. In the case of the acyltransferase, most of the candidates were excluded since they were N-acyltransferases, but the formation of fucoxanthin requires an O-acyltransferase.

The short-listed candidates were then assessed in terms of guilt-by-association criteria, which have been corroborated in Sections 5.3.1 and 5.3.2 that are shared by many carotenogenesis genes, like high expression under blue light after dark adaptation or gene clustering in the chromosome. In addition, the organisms for which homologs exist and the presence of a signal and target peptide for plastid localization was also assessed. The inclusion/ exclusion criteria for different guilt-by-association principles are now described:

- **Transcriptomics:** Genes that were under-expressed at blue light conditions after dark adaptation were excluded and genes with positive, but low values of expression under these conditions were also excluded if other functional association criteria were absent.
- **Gene clustering:** Genes that were not in the vicinity of other genes from the MEP and carotenogenesis pathway or from other functional related proteins, like the ones involved in photosynthesis or from the chlorophyll biosynthesis pathway, were not excluded, since as seen in Section 5.3.1, not all carotenogenesis genes are clustered. However, priority was given to genes that were in one of the carotenoid clusters in the chromosome.
- **Presence in different phylogenetic groups:** Genes with a broad homology across different taxa, especially in non-photosynthetic organisms like animals, were excluded. Genes present in closely related organisms, like other algae, were not excluded, since that enzyme might have evolved to a more specific function from a common enzyme present

anciently in an ancestor of these organisms. Also, genes present in bacteria were not excluded, since a big number of genes have been transferred between diatoms and bacteria by horizontal transfer. However, more relevance was given to the candidates present only in the groups in which fucoxanthin is produced.

- **Presence in the plastid.** The candidates were not excluded based on the absence of the signal and target peptide for the plastid, since the programs used to detect them are not always correct. However, preference was given to the genes that were predicted to have them.

In **Table 5** the result of these searches is shown. As it can be seen, none of the candidate gene products are predicted to be imported into the plastid, so the expectations of these being the true candidates can be set low, although as stated before, the predictions can be erroneous. From these set of genes, there is no candidate with a very high gene expression under blue light, typical of carotenogenesis genes. Candidates that are worth to mention from this set are the ones with gene ID:7197805 and ID: 7197754. The first one is predicted to have a carotene β -ring hydroxylase domain. The hydroxylation of fucoxanthin does not take place in the ring, but in the aliphatic chain. Therefore, this gene might not catalyze the synthesis of fucoxanthin, but maybe of another carotenoid. In addition, this gene is present in the carotenoid cluster of chromosome 4 and has a positive log₂ value under blue light conditions. The case of the gene candidate with ID: 7197754 is curious, since it presents a high degree of homology with the putative enzyme for β -carotene hydroxylase, is over-expressed under blue conditions and is in the vicinity of another carotenogenesis gene. However, it seems very unlikely that this enzyme catalyzes the formation of fucoxanthin, since it is only present in diatoms. A scenario in which diatoms, brown algae and haptophytes evolved different enzymes for that function independently is not very plausible.

In a different approach, the vicinities in the chromosome of the carotenoid clusters were investigated for the identification of genes with unknown function that could fit under the inclusion criteria for the guilt by association principles. In **Table 6** we can see the candidate genes found.

Table 5. List of candidate genes for fucoxanthin synthase based on their function and some guilt by association principles.

Gene ID	Predicted function	Log2BL	Chromosome location and closest genes	Homology	SignalP
7203920	Carotene β-ring hydroxylase domain	0.81	Chr3 (1264K) Lhcf14 , lipid synthesis	D,W,G,R,P	No SP
7197805	Carotene β-ring hydroxylase domain	1.63*	Chr4 (916K) HDR, CMK, VDE, ZEP3	D,W,G,P	No SP
7200710	Fatty acid hydroxylase superfamily	0.47	Chr7 (348K) CRTISO1	D,H,F, G,R,B,A	No SP
7203855	Fatty acid hydroxylase superfamily	1.96*	Chr3 (898K) lipid synthesis	D,G,B,M	No SP
7201479	Fatty acid desaturase superfamily**	0.84	Chr9 (792K) FdOR	D,H,G, C	No SP
7197754 (partial)	Homology to LUT1.2 and other CYPs.	2.87*	Chr4 (660K) GGPPS	D	No SP
7202115	Membrane-bound O-acyl transferase	2.68*	Chr12 (224K) TCA	D,P,R,F,H	No SP
7194891	Membrane-bound O-acyl transferase	1.62*	Chr18 (455K)	D,P,F	No SP
7201336	Membrane-bound O-acyl transferase	1.76*	Chr9 (254K) Chlorophyll	D,G,R	No SP

Log2BL represents the log-ratio of the expression values of a gene in onset of light versus dark-adapted cells. Data extracted from a RNA-seq experiment by Dr. König⁷⁷. The chromosome location is in brackets (K: kbp) and below the presence of genes of the carotenoid pathway or of pathways of the primary metabolism are shown. When a pathway is mentioned means that a gene of that pathways is nearby the gene candidate. **Bold** letters represent good associations with the carotenoid pathway or a function that is in accordance with the function carried out by fucoxanthin synthase. Abbreviations: **D**, diatoms; **C**, cyanobacteria; **W**, brown algae; **G**, green algae; **P**, plants; **B**, bacteria; **H**, haptophytes; **F**, dinoflagellates; **G**, fungi; **A**, archaea; **M**, cryptomonad; **?**, not known; **SP**, signal peptide. * Significant value of log2 fold change. ** Includes β -carotene ketolase, β -carotene hydroxylase, among others.

In this search, the gene with ID: 7197626 is of special interest since belongs to the fatty acid desaturase-like family, which includes carotenoid hydroxylases and ketolases and in addition it is highly expressed under blue conditions, it is in the carotenoid cluster and it is predicted to go to the plastid. However, this enzyme does not show homology to any enzyme in brown algae, which also synthesize fucoxanthin.

Table 6. List of candidate genes for fucoxanthin synthase based on proximity to the carotenoid pathway and other association criteria.

Gene ID	Predicted function	Log2BL	Chromosome location and closest genes	Homology	SignalP
7197626	Fatty acid desaturase-like**	3.14*	Chr4 (1072K) Carotenoid cluster	D,H,F,G,B	Yes, plastid HC
7197629 (partial)	?	2.58*	Chr4 (1088K) Carotenoid cluster	D,W,H,M,G	No SP

Log2BL represents the log-ratio of the expression values of a gene in onset of light versus dark-adapted cells. Data extracted from a RNA-seq experiment by Dr. König⁷⁷. The chromosome location is in brackets (K: kbp). **Bold** letters represent good associations with the carotenoid pathway or a function that is in accordance with the function carried out by fucoxanthin synthase. Abbreviations: **D**, diatoms; **C**, cyanobacteria; **W**, brown algae; **G**, green algae; **P**, plants; **B**, bacteria; **H**, haptophytes; **F**, dinoflagellates; **G**, fungi; **A**, archaea; **M**, cryptomonad; **?**, not known; **SP**, signal peptide; **HC**, high confidence.* Significant value of log2 fold change. ** Includes β -carotene ketolase, β -carotene hydroxylase, among others.

Finally, one last approach for the identification of putative enzymes was conducted based on the transcriptomic data. In this strategy, all the genes with a log2BL value lower than 2.5 were discarded, which were the vast majority. The rest of the genes, with values higher than 2.5, and which lacked a functional annotation, were initially classified into several groups based on log2BL ranges and then they were assessed by the guilt by association principles following the same inclusion/ exclusion criteria. In **Table 7** the candidates found by this approach are listed.

Table 7. List of candidate genes for fucoxanthin synthase based on the values of log2 fold change and other association criteria.

Gene ID	Predicted function	Log2BL	Chromosome location and closest genes	Homology	SignalP
7197629 (partial)	?	2.58*	Chr4 (1088K) In carotenoid cluster	D,W,H,M,G	No SP
7196017	?	3.82*	Chr1 (756K) GGPPR	D	Plastid LC
7199733	?	4.89*	Chr5 (177K) LUT1.2	D	Plastid HC
7197067 (partial)	?	5.45*	Chr2 (31K)	D,W,B	Plastid HC
7199026 (partial)	?	2.92*	Chr 28 (122K) Glycolysis	D,W	Plastid HC
7197592	?	5.78*	Chr2 (22K)	D,W,H,C	Plastid HC
7195400	?	5.21*	Chr20 (285K) PPP, Glycolysis, Lhcf	D,W,H	Plastid HC
7203461	?	2.85*	Chr16 (192K)	D,W,H	Plastid HC
7198588	?	2.64*	Chr25 (141K) Chlorophyll	D,W,H	Plastid HC

Log2BL represents the log-ratio of the expression values of a gene in onset of light versus dark-adapted cells. Data extracted from a RNA-seq experiment by Dr. König⁷⁷. The chromosome location is in brackets (K: kbp) and below the presence of genes of the carotenoid pathway or of pathways of the primary metabolism are shown. When a pathway is mentioned means that a gene of that pathways is nearby the gene candidate. **Bold** letters represent good associations with the carotenoid pathway or a function that is in accordance with the function carried out by fucoxanthin synthase. Abbreviations: **D**, diatoms; **C**, cyanobacteria; **W**, brown algae; **G**, green algae; **P**, plants; **B**, bacteria; **H**, haptophytes; **F**, dinoflagellates; **G**, fungi; **A**, archaea; **M**, cryptomonad; **?**, not known; **SP**, signal peptide; **LC**, low confidence; **HC**, high confidence. * Significant value of log2 fold change. ** Includes β -carotene ketolase, β -carotene hydroxylase, among others.

The gene with gene ID: 7197629 is present in both Table 6 and 7, since it shows both a high expression in light compared to dark conditions and it is located in the chromosome close to other carotenoid biosynthetic genes.

In general, the candidates show high expression in blue light, almost all of them are predicted to be localized in the plastid and some of them are only present in the organisms in which fucoxanthin is produced. However, no functional annotation is present in the gene databases, so their function should be determined experimentally. However, one must be careful, since some of these genes are very short in length (less than 1 kb), so they might be partial. Therefore, it would be recommended to assess bioinformatically other stop codons present downstream the predicted stop codon or in case of doubt, to perform a rapid amplification of cDNA ends experiment.

Finally, as a curiosity, as a result of a STRING search a gene with ID:7196107 was found, with unknown function. A search for homologs in other organisms was carried out, that resulted in homologs in diatoms, brown and green algae, plants, moss, bacteria and haptophytes. In the plant homologs, the enzyme is expressed in the chloroplast and it works as a β -carotene isomerase, an enzyme that converts trans-carotene into cis- carotene, an intermediate metabolite for the synthesis of strigolactones (phytohormones)⁸⁰. A further examination of the gene in diatoms yielded that this gene is highly expressed under blue light after dark adaptation ($\log_2BL=4.10$) and it is located in chromosome 1, relatively close to the carotenogenesis gene *zds*.

In conclusion, gene by association parameters, such as gene expression and gene clustering could be helpful for the identification of orphan enzymes. The integration of this information with further sources of bioinformatic data could improve the gene predictions. In the end, it would be necessary to test the candidates experimentally.

6. Functional characterization of the *lut* genes of *P. tricornutum*

6.1 Introduction

6.1.1 Classes of β -carotene hydroxylases

Zeaxanthin is formed by the introduction of a hydroxyl group in both β -rings of β -carotene. The reaction is catalyzed by β -carotene hydroxylase⁸¹. Two classes of β -carotene hydroxylases have been identified to date:

- **Non heme (NH) di-iron hydroxylases** need iron, molecular oxygen, ferredoxin and ferredoxin oxidoreductase to function and they seem to be located in membranes due to the presence of putative transmembrane helices. These enzymes contain conserved histidine motifs that are typical of fatty acid and sterol desaturases⁸¹.

Aside from these regions, different groups of NH di-iron hydroxylases share low primary structure, although they perform a similar reaction mechanism. There are four main groups: NH di-iron hydroxylases from plants and green algae, from non-photosynthetic bacteria, from cyanobacteria and from archaea. The NH hydroxylases from cyanobacteria have more protein identity to bacterial β -ketolases than to the other groups of β -hydroxylases. In primitive red algae, a putative gene related to the β -hydroxylases of cyanobacteria has been identified, although it is not present in other members of red algae⁸¹.

In heterokonts, no gene related to this class of enzymes, except for the centric diatom *T. pseudonana* in which there is a gene with homology to the NH di-iron hydroxylases of plants and bacteria, although this gene was not found in the expressed sequence tag (EST) database, so it might be a pseudogene. In contrast, no gene of the pennate diatom *P. tricornutum* has homology to NH di-iron enzymes⁸².

- The other class of carotene hydroxylases belong to the **cytochrome P450 (CYP) monooxygenase** superfamily. CYPs are hemoproteins that are present in all kingdoms of life and they participate in the oxidative metabolism of many substances. The electrons needed for the reaction are provided by NAD(P)H via another redox species (ferredoxin, FAD or FMN)⁸³.

The proteins encoded by CYP97 genes have been related to the hydroxylation of β - and α -carotene. There are three groups of CYP97 proteins: CYP97a, CYP97b and CYP97c. In higher plants and green algae, members of the CYP97a and CYP97c subfamilies participate in the hydroxylation of β - and α -carotene. Higher plants also contain the CYP97b subfamily, although they are not targeted to the chloroplast. The CYP97a and CYP97c subfamilies are not present in the genome of red algae and heterokonts. These organisms present members of the CYP97b subfamily. On the contrary, the genome of cyanobacteria does not contain any gene homologous to the CYP97 family⁸¹.

6.1.2 Carotene hydroxylases in plants

In plants, the pathway is divided in two branches after the cyclase reaction from lycopene: in one branch β -carotene is formed and in the other one, α -carotene. α -carotene contains a β - and a ϵ -ring, while β -carotene contains 2 β -rings. And the xanthophylls derived from β - or α -carotene are called β - or α -xanthophylls, respectively. The first hydroxylation of β -carotene produces cryptoxanthin and after the second hydroxylation, zeaxanthin is formed. Analogously, the first hydroxylation of α -carotene produces zeinoxanthin and after the second hydroxylation, lutein is formed⁸⁴.

The enzymes that participate in the hydroxylation of β - and/or α -carotene are: CHY1, CHY2, LUT1 (CYP97c1) and LUT5 (CYP97a3). A quadruple knockout mutant for these enzymes was lethal in soil, but it could grow in a culture medium

supplemented with a carbon source. β - and α -xanthophylls were not detected in these mutants by HPLC, thereby confirming that these 4 proteins are the whole set of enzymes that catalyze these reactions, and therefore no other enzyme participates⁸⁴.

The above-mentioned carotene hydroxylases preferentially hydroxylate either the β - or ϵ -rings of carotene, showing also preference for either α - or β -carotene. However, there is a functional overlap among them⁸⁴.

Now the reactions involved in the hydroxylation of carotene in *Arabidopsis thaliana* are described (**Figure 33**).

- Hydroxylation of the β -rings of β -carotene

This reaction is preferentially catalyzed by the NH di-iron hydroxylases **CHY1** and **CHY2**. Single knock-out mutants for these enzymes revealed a phenotype that was not growth-impaired and with a not very relevant effect on the levels of xanthophylls, showing that these enzymes are redundant in their function⁴⁷. In the double null mutant (*chy1 chy2*) the levels of β -xanthophylls were reduced by 76%. This implies that CHY1 and CHY2 are the main enzymes catalyzing this step, although another enzyme can carry out this function⁸⁴.

The other enzyme that can catalyze this reaction is CYP97a3 (**LUT5**), although β -carotene is not its preferred substrate. The triple knock-out mutant *chy1 chy2 lut5* only produces 2% of the wild-type β -xanthophyll level, indicating that LUT1 has barely no affinity towards β -carotene⁸⁴.

- Hydroxylation of the β -ring of α -carotene

The **LUT5** enzyme presents the highest affinity towards the β -ring of α -carotene and it is the main responsible for this reaction. However, the *lut5* knock-out mutant still produces 78% of lutein, compared to wild-type levels⁸⁴.

The **CHY1** and **CHY2** enzymes can also recognize the β -ring of α -carotene

and catalyze this reaction, although α -carotene is not their preferred substrate. This can be deduced from the *lut1lut5* knock-out mutant in which zeinoxanthin accumulates⁸⁴.

Similarly, although **LUT1** (CYP97c1) has a strong preference for the ϵ -ring of α -carotene, it has also affinity to the β -ring of α -carotene. This can be seen from the *bch1 bch2 lut5* knockout mutant, in which lutein levels are only reduced an 26% in comparison to wild type⁸⁴.

- Hydroxylation of the ϵ -ring of α -carotene

This reaction is catalyzed by **LUT1**. The *lut1* knock-out produces only the 3% of the lutein levels compared to wild type⁸⁴.

The extent to which the different enzymes participate in these reactions can be seen by the comparison of the carotenoid levels in the single, double and triple knockout mutants with respect to the wild type levels (**Table 8**)⁸⁴.

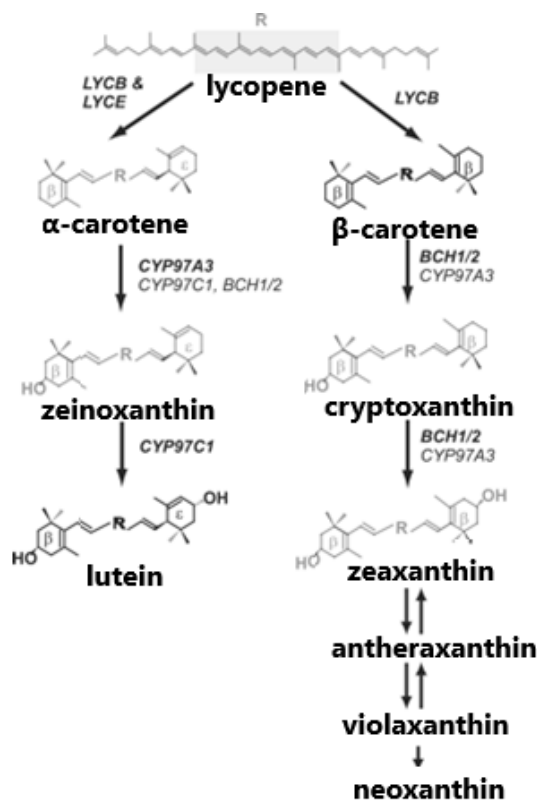


Figure 33. Carotenoid pathway from lycopene of *A.thaliana*. Image modified from Kim et al⁸⁴.

Table 8. Comparison of carotenoid levels between different genotypes of *A. thaliana* in which one or more genes encoding a carotene hydroxylase have been inactivated, and wild type (Col-O and Ws). Table from Kim et al⁸⁴.

Leaf carotenoid composition (mmol pigment mol⁻¹ Chl) from soil-grown plants of the indicated genotypes

	Col-0	Ws	<i>cyp97c1</i>	<i>cyp97a3</i>	<i>bch1 bch2</i>	<i>cyp97a3</i> <i>cyp97c1</i>	<i>bch1 bch2</i> <i>cyp97c1</i>	<i>bch1</i> <i>cyp97a3</i>	<i>bch1 bch2</i> <i>cyp97a3</i>
Violaxanthin	31 ± 4.6 (12)	28 ± 3.1 (12)	55 ± 4.0 (23)	21 ± 2.0 (9)	11 ± 1.0 (4)	41 ± 2.7 (17)	10 ± 3.0 (4)	14 ± 0.8 (7)	0.2 ± 0.1 (0)
Antheraxanthin	2.5 ± 0.6 (1)	1.8 ± 0.4 (0.7)	24 ± 3.0 (10)	2.3 ± 2.0 (1)	1.1 ± 0.3 (0)	22 ± 2.5 (9)	8.6 ± 0.6 (4)	4.3 ± 1.6 (2)	0.6 ± 0.3 (0)
Zeaxanthin	ND	ND	8.1 ± 1.4 (3)	ND	ND	7.0 ± 1.3 (3)	4.3 ± 0.4 (2)	ND	ND
Neoxanthin	32 ± 3.0 (12)	30 ± 2.5 (13)	26 ± 1.8 (11)	18 ± 2.6 (8)	3.8 ± 1.0 (2)	20 ± 2.4 (8)	2.0 ± 0.9 (1)	4.1 ± 0.7 (2)	0.5 ± 0.9 (0)
β-Xanthophylls ^a	66 ± 8.2 (25)	60 ± 6.0 (26)	113 ± 10 (47)	41 ± 5.6 (18)	16 ± 2.3 (6)	90 ± 8.9 (37)	25 ± 4.9 (11)	22 ± 3.1 (11)	1.3 ± 1.3 (0)
Lutein	133 ± 10 (52)	122 ± 8.9 (51)	4.0 ± 0.4 (2)	104 ± 6.7 (46)	156 ± 8.0 (65)	3.2 ± 0.4 (1)	7.2 ± 0.7 (3)	94 ± 6.2 (43)	99 ± 14 (48)
α-Car-OH	0.8 ± 0.3 (0)	0.6 ± 0.1 (0.2)	61 ± 5.8 (26)	5.4 ± 0.6 (2)	1.0 ± 0.3 (0)	53 ± 5.2 (22)	101 ± 3.8 (43)	8.8 ± 0.7 (4)	10 ± 5.9 (5)
α-Carotene	1.6 ± 0.6 (1)	1.5 ± 0.2 (0.6)	1.4 ± 0.3 (1)	49 ± 0.8 (22)	1.5 ± 0.2 (1)	54 ± 3.1 (23)	1.6 ± 0.3 (1)	59 ± 6.3 (27)	56 ± 9.2 (27)
β-Carotene	56 ± 3.6 (22)	54 ± 3.7 (23)	57 ± 4.7 (24)	26 ± 3.1 (12)	67 ± 2.5 (28)	38 ± 3.5 (16)	99 ± 4.7 (42)	33 ± 5.4 (15)	39 ± 6.9 (19)
Total carotenoids	257 ± 21	238 ± 16	236 ± 17	227 ± 10	241 ± 11	237 ± 18	234 ± 7.6	217 ± 16	205 ± 20

^aSum of violaxanthin, antheraxanthin, zeaxanthin and neoxanthin.

The values shown are the mean of at least three biological replicates analyzed in triplicate ± SD. The numbers in parentheses indicate the percentage of total carotenoids for each carotenoid in the indicated genotype.

ND, below the detection limit of HPLC (0.5 ng); α-Car-OH, monohydroxy α-carotene.

6.1.3 Carotene hydroxylation in diatoms

In contrast to plants, in diatoms only β-carotene is formed from lycopene. α-carotene is not present in diatoms. In the diatom *Phaeodactylum tricornutum* there are two enzymes of the CYP97b family, the lutein deficient-like enzyme LUT1-1 and LUT1-2, that show sequence homology to the LUT1 and LUT5 carotene hydroxylases in plants and are targeted to the plastid. It has been suggested that the hydroxylation of β-carotene by CYP proteins in diatoms instead of the NH-diiron hydroxylases might be advantageous for diatoms due to the low levels of iron in the oceans⁸².

6.2 Objective

The objective is to functionally characterize the putative genes for β-carotene hydroxylase of *Phaeodactylum tricornutum*, *lut1-1* and *lut1-2*. For that purpose, a *chy1chy2lut5* knockout mutant (**Figure 34 and 35**) of *Arabidopsis thaliana* created by the group of Prof. Bassi was used as a background organism to perform a complementation assay by the transformation in the plant of the diatom putative genes of β-carotene hydroxylases, *lut1-1* and *lut1-2*. It would be

expected that if they carried out this function, the low levels of xanthophylls in the knockout mutant (**Figure 34**) due to the absence of genes involved in the hydroxylation of β -carotene would be restored to wild-type levels.

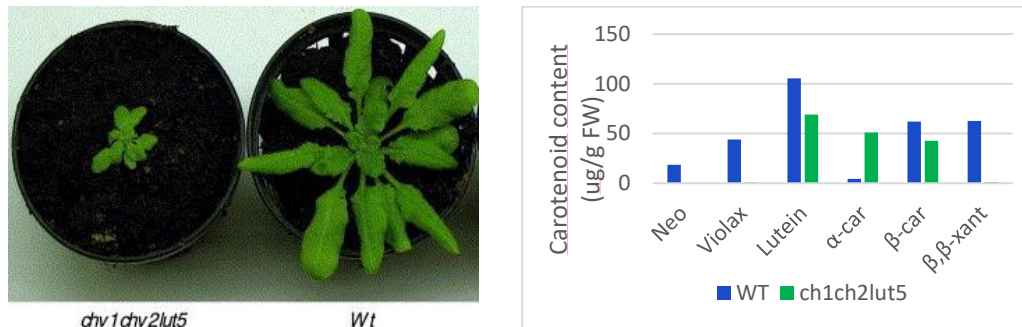


Figure 34. Phenotype and carotenoid levels of the *chy1chy2lut5* mutant compared to wild type. The left image and the data for the graph were taken from Fiore et al⁸⁵.

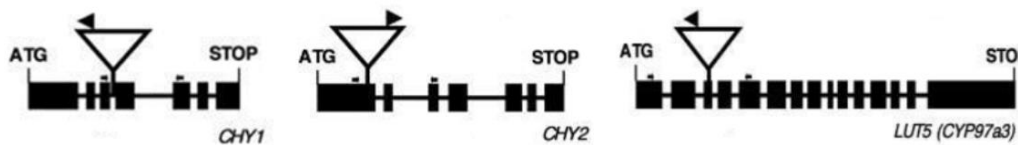


Figure 35. T-DNA insertion of the *chy1*, *chy2* and *lut5* genes of the *chy1chy2lut5* mutant. Images from Alessia Fiore et al⁸⁵.

6.3 Results

The *P. tricornutum* genes *lut1-1* and *lut1-2*. were amplified by PCR with a high-fidelity DNA polymerase (Phusion Polymerase) from the pUC18-*lut1-1* and pUC18-*lut1-2* vectors, that were a gift of Dr. Breitenbach.

Since the aim is that their gene products are in the end transported into the plastids of *A. thaliana*, it is necessary to amplify the genes without the sequence of the signal and target peptide of diatoms and introduce in frame directly upstream of the gene a plastid target peptide specific for *Arabidopsis*.

So, first, the target peptide of the gene *zep* of *Arabidopsis* was amplified from the

pET-*zep* vector, that was a gift of Dr. Eilers (**Figure 36**). The identification of the target peptide was possible thanks to the algorithm TargetP. Then, the gene was cloned by the hot-fusion method into the pGreen plasmid that had been previously restricted with XbaI and BamHI.

In order to identify the signal and target peptide of *P.tricornutum* the programs SignalP 3.0 and ASAFind were used. For both proteins, it was predicted with SignalP to contain a signal peptide with a cleavage site between the amino acid 25 and 26. However, ASAFind predicted that their target peptides do not lead the proteins to the plastid. However, although these algorithms are very helpful, they are not fully trustworthy. For instance, the enzyme phytoene synthase of *P. tricornutum* is not predicted to have a signal peptide by SignalP3.0, while a location experiment based on protein fusion with the Green Fluorescent Protein (GFP) showed that the protein is present in the plastids⁸⁶.

Based on the average lengths of the signal peptide and target peptide in diatoms¹⁶, two versions for every gene were amplified by PCR, in which a short- or a long-version of the signal and target peptide was excluded from the amplified sequence (**Figure 36**). They were called *lut1-1_144*, *lut1-1_270*, *lut1-2_180* and *lut1-2_360*, whereby the number after the hyphen represents the number of base pairs removed at the beginning of the coding sequence of the genes. It was checked that the active site of the enzyme was maintained, and the number of base pairs removed was multiple of three to conserve the frame.

The *lut* genes were cloned by hot fusion into the pGreen vector containing the target peptide of the *zep* gene. The vector was restricted before the cloning step with BamHI and EcoRI (for cloning the *lut1-1_144* and *lut1-1_270* genes) or with SmaI and SacI (for cloning the *lut1-2_180* and *lut1-2_360* genes). The cloning of the *lut* genes was carried out by Marina Freyberg, at that time a bachelor student under my supervision.

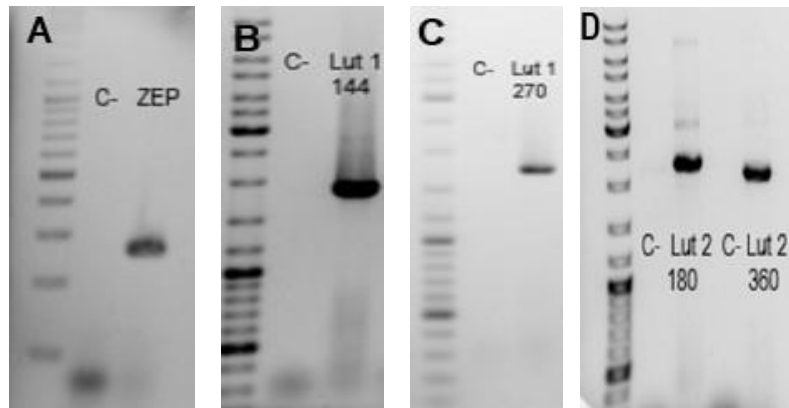


Figure 36. Amplification by PCR of the putative β -carotene hydroxylase genes of *P. tricornutum* and a plant target peptide. The target peptide of the *zep* gene (A), and the *lut* genes *lut1* (B and C) and *lut2* (D) were amplified by PCR. C-: negative control.

The *Arabidopsis* knock-out mutant had been already transformed with a vector containing the *bar* gene, that confers resistance to the herbicide glufosinate ammonium (BASTA), and with the *nptII* gene, that provides resistance to the antibiotic kanamycin. Therefore, a new selection marker had to be introduced into the pGreen vector. The sulfadiazine (*sul*) gene was isolated by DNA extraction and PCR (**Figure 37**) from the *A. thaliana* line 833H12 (GABI-kat) that contains integrated in the genome the plasmid pAC106 that possess the *sul* gene. Then, this gene was cloned in the multiple cloning site of an empty pGreen vector that had been previously restricted with HindIII and EcoRI. After that, the construct containing the 35S promoter, the *sul* gene and the 35S terminator was amplified by PCR (**Figure 37**) and cloned into the final pGreen vector, linearized with NheI, that contains the *zep* target peptide and the *lut* gene.

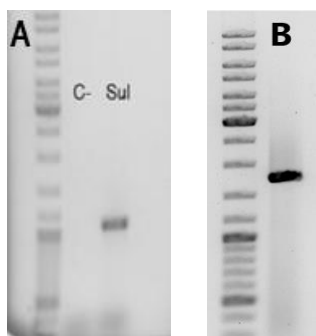


Figure 37. Amplification by PCR of the selective marker. (A) The *sul* gene was amplified by PCR from a mutant plant that contain it. C-: negative control. (B) After cloning this gene, the construct containing 35S promoter, *sul* gene and 35S terminator was amplified by PCR in order to clone it into the final vector containing the *lut* gene.

The final pGreen vectors, each containing either the *lut1-1_144*, *lut1-1_270*, *lut1-2_180* or *lut1-2_360* genes, were individually transformed by the heat-shock method into competent *Agrobacterium tumefaciens*, strain GV 3101, containing the pSoup (Tet^r) and pMP90 (Gent^r) plasmids (**Figure 38**). The transformation of *A. tumefaciens* was carried out by Marina Freyberg. For the picture of the final pGreen vector containing *lut1-2_360* see **Appendix**.

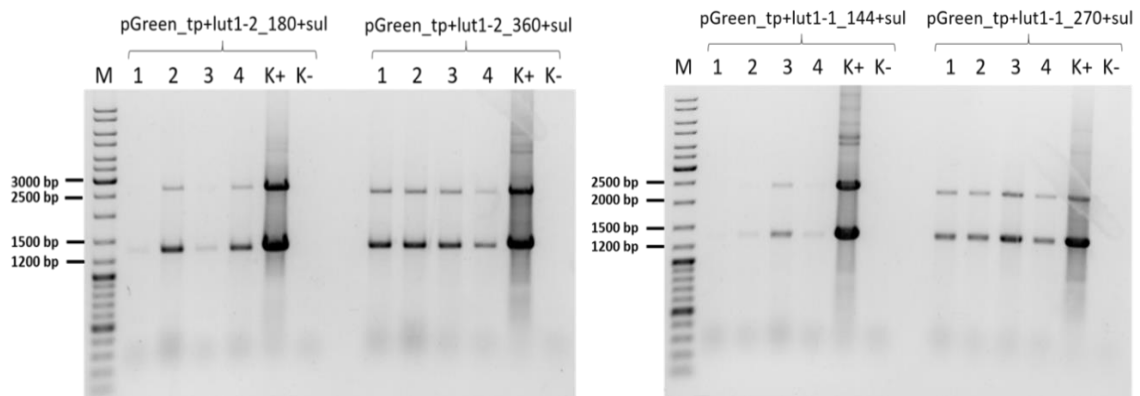


Figure 38. Colony PCR for the *lut* and *sul* genes after transformation in *A. tumefaciens*. The pGreen vector containing the *lut1-1* or *lut1-2*, each in its two length versions, and the selective marker *sul* were cloned into *A. tumefaciens*. The transformation was checked with colony PCR to detect the presence of these genes. The same primers were used for all the PCRs, and they bind the 35S-promoter and the 35-S terminator. Therefore, both *lut* and *sul* genes are expected to be amplified. The expected lengths are for *sul*, 1323bp; for *lut1-1_144*, 2341bp; for *lut1-1_270*, 2215bp; for *lut1-2_180*, 2692bp and for *lut1-2_360*, 2512bp. K+: Positive control. K-: Negative control. This figure is from Marina Freyberg.

Then, the transformed *Agrobacterium* cells was suspended in a specific buffer, and this suspension was used to transform the *chy1 chy2 lut5* knockout mutants of *A. thaliana* by the floral-dip transformation method. The transformation of *A. thaliana* was carried out by Marina Freyberg. The seeds of these plants were collected and screened in MS-plates with sulfadiazine. Of the 4 constructs, only one plant transformed with the *lut1-2_360* (mutant A) could grow under sulfadiazine conditions and was clearly positive by looking at its intense green color. Other plants were also further tested because their phenotype was dubious, since they grew and they were not so yellowish as the other plants that are negative, but also not as green as the plant selected for *lut1-2_360*. The

selected plants were finally transferred to soil and the seeds were collected and planted to obtain the second generation from the transformed plant, from which further analysis were conducted. By PCR, it was confirmed that only the plant with the clear phenotype was positive (mutant A) and possessed the diatom *lut1-2_360* gene (**Figure 39**).

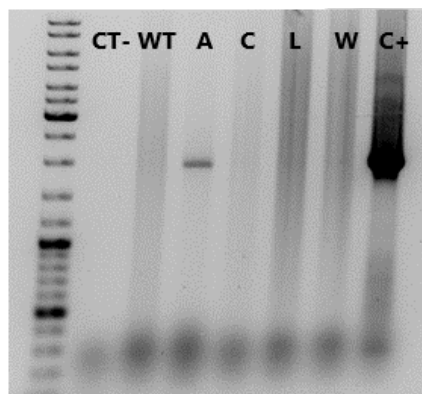


Figure 39. Screening of the knockout plants for the presence of the *lut2_360* gene from *P. tricornutum*. After the transformation of the *chy1chy2lut5* plants with the *lut2_360*, the plants were selected by resistance to sulfadiazine and in the next generation, the presence of the gene *lut2_360* was checked by PCR. Selected plants to test: A, C, L, W; C-: negative control; WT: wild type; C+: positive control.

Then, the pigment content was analyzed from this mutant and it was compared to the pigments in the wild type and in the original background phenotype, that is the *chy1 chy2 lut5* knockout that was used for the transformation of the diatom *lut* genes (**Figure 40**, **Figure 41** and **Figure 42**) With this information, one could see that the original phenotype was not as expected (**Figure 34**). However, this phenotype can also be used in this complementation experiment, since the levels of β -xanthophylls are much lower than in the wild type, so it could be seen by the complementation with the diatom *lut* gene if the β -xanthophylls levels are restored.

The ratio between zeaxanthin to chlorophyll a was calculated (**Figure 43**). It can be seen that the ratio in the mutant transformed with the diatom *lut1-2_360* is higher to the one in the original triple knock-out plant, although it does not reach the levels of the wild type. Therefore, it could be implied that the *lut1-2* gene product plays a role in the hydroxylation of β -carotene into zeaxanthin.

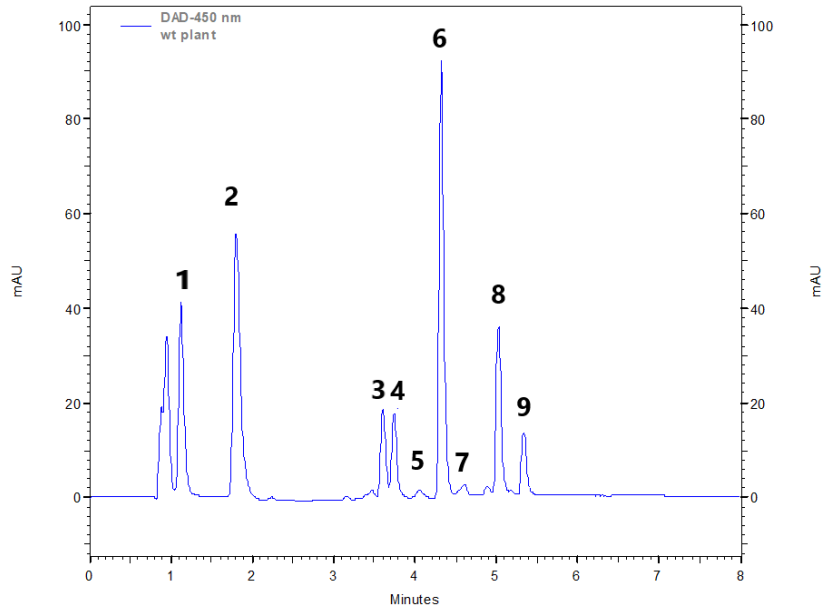


Figure 40. HPLC elution profile at 450 nm of the pigment levels in *Arabidopsis thaliana* Col-O (wild type). Peak 1 and 2: Chlorophyllide, 3: Neoxanthin, 4: Violaxanthin, 5: Antheraxanthin, 6: Lutein, 7: Zeaxanthin, 8: Chlorophyll b, 9: Chlorophyll a.

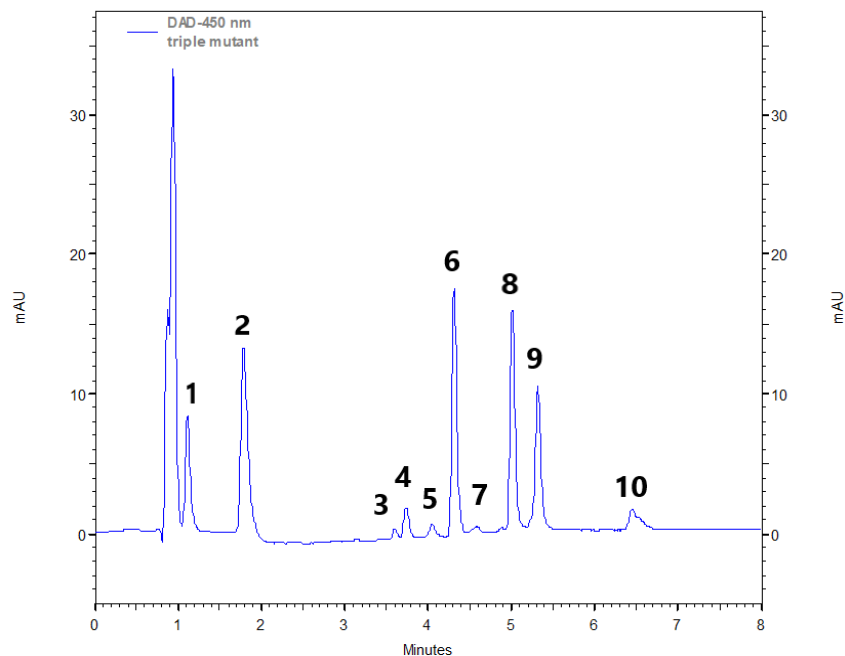


Figure 41. HPLC elution profile at 450 nm of the pigment levels in *Arabidopsis thaliana* with reduced levels of β -xanthophylls. This genotype was the background in which posteriorly the *lut* genes of *P. tricornutum* were cloned. Peak 1 and 2: Chlorophyllide, 3: Neoxanthin, 4: Violaxanthin, 5: Antheraxanthin, 6: Lutein, 7: Zeaxanthin, 8: Chlorophyll b, 9: Chlorophyll a.

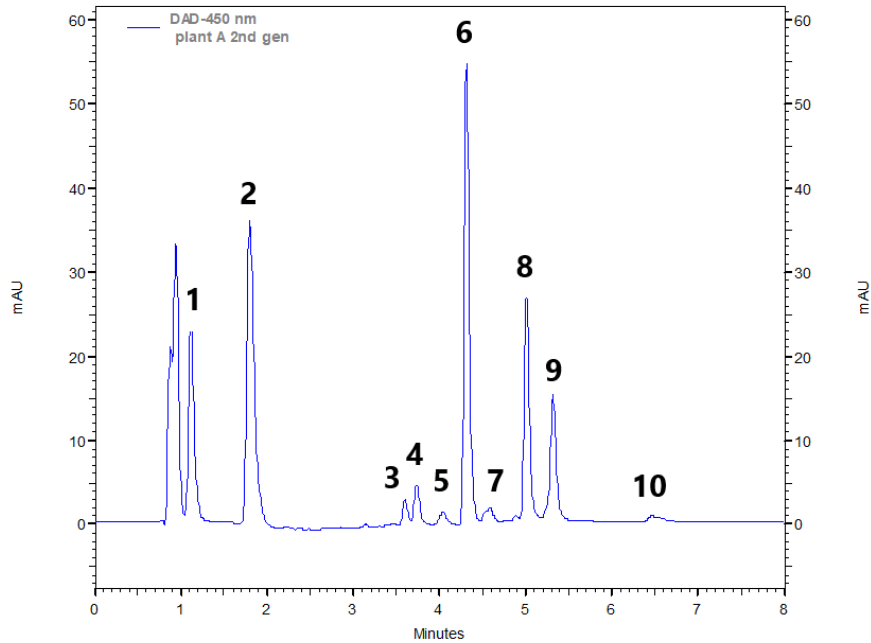


Figure 42. HPLC elution profile at 450 nm of the pigment levels of *A.thaliana* with reduced levels of β -xanthophylls complemented with the diatom *lut1-2* gene. Peak 1 and 2: Chlorophyllide, 3: Neoxanthin, 4: Violaxanthin, 5: Antheraxanthin, 6: Lutein, 7: Zeaxanthin, 8: Chlorophyll b, 9: Chlorophyll a, 10: β -carotene.

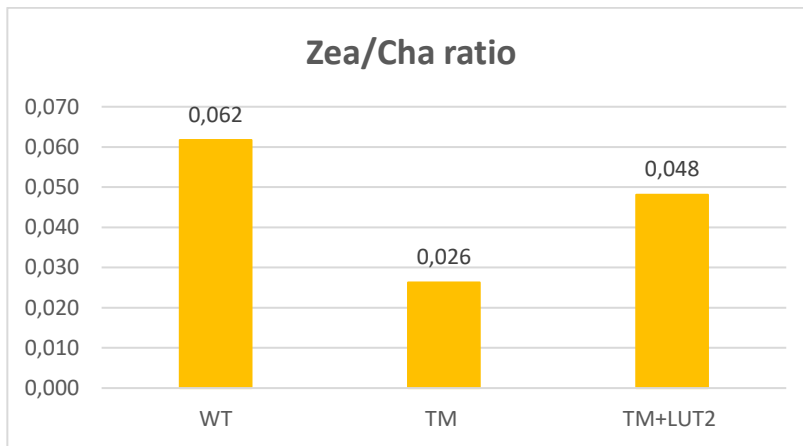


Figure 43. Comparison of the Zeaxanthin/ Chlorophyll a ratio in different genotypes. The pigments were extracted and quantified by HPLC from the following genotypes: wild type (WT), the plant with reduced levels of β -xanthophylls (TM), and the plant with reduced levels of β -xanthophylls complemented with the gene *lut1-2* from *P. tricornutum* (TM+LUT2).

The genotype of the transformed plants with respect to the genes *chy1*, *chy2* and *lut5* of *A. thaliana* was checked. They were supposed to be homozygous for the knockouts in these genes. The absence of the expression of these genes was examined by PCR from cDNA samples of these plants. The quality of the cDNA was previously verified by the amplification of the *fk3* gene. As seen in the **Figure 44**, the *chy1* and *lut5* genes are not expressed, as expected because of the T-DNA insertion in the beginning of the gene. However, the *chy2* gene was expressed.

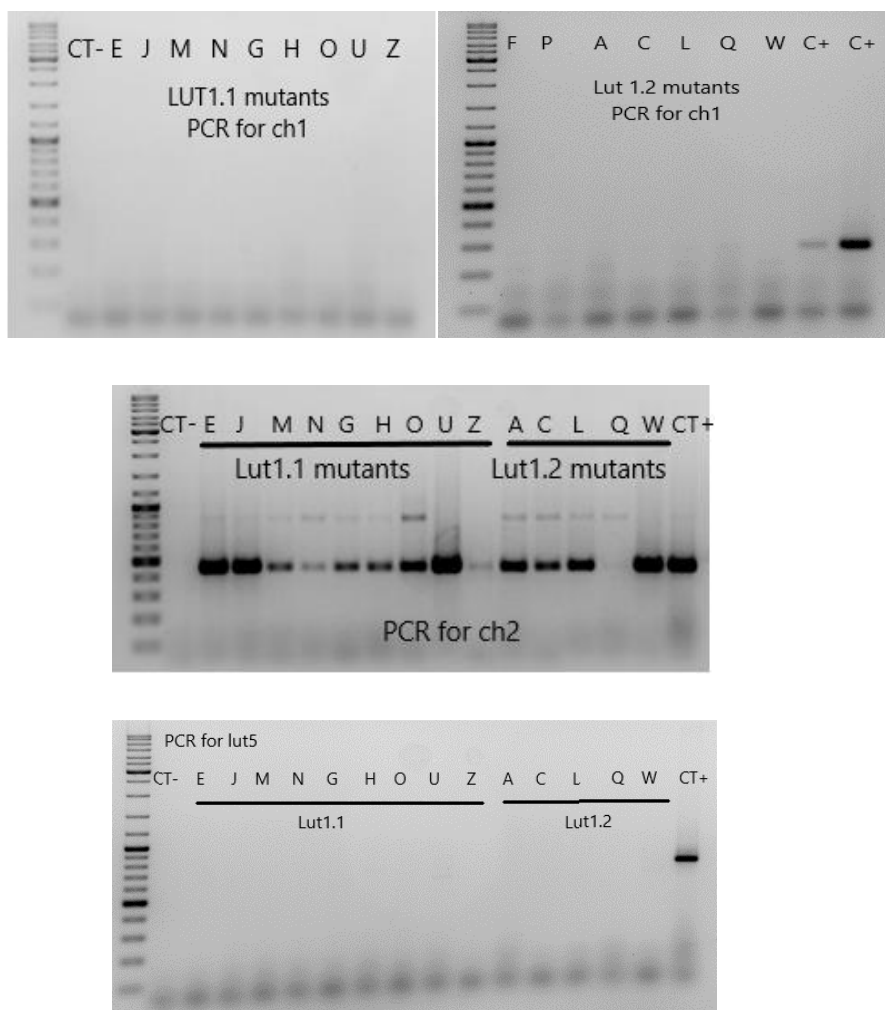


Figure 44. Corroboration of the expression levels of *chy1*, *chy2* and *lut5* genes in the assumed *chy1chy2lut5* plant mutant. RNA was extracted from different plant mutants (named with letters from A-Z) with the *chy1chy2lut5* genotype that were transformed with either the *lut1-1* or the *lut1-2* genes. Then, cDNA was prepared from RNA and the quality of all cDNA samples was tested by amplification of the *fk3* gene. After that, the expression of the plant genes *chy1* (top figures), *chy2* (middle figure) and *lut5* (bottom figure) was checked by PCR. C-: negative control, C+: positive control.

To investigate if the T-DNA insert was present in the genome, a PCR reaction was carried out from gDNA of the plants using two primers, one binds on the T-DNA insert and the other binds the gene in which it is inserted. As it can be seen in **Figure 45**, the plant that was used to transform the *lut* genes contained in its genome the T-DNA insertions in the genes *chy1*, *chy2* and *lut5*.

Therefore, it can be said that the plants are homozygous for the deletion of the *chy1* and *lut5* genes, but the *chy2* gene is still expressed although the T-DNA insertion is present. It could be argued, that maybe the gene it is still partly expressed but then the protein is truncated. However, in the publication where the triple knockout was characterized the *chy2* gene was not expressed at all. Furthermore, looking at the pigment levels, β -xanthophylls are still produced in these plants, although in low levels, so one of the β -carotene hydroxylase must be expressed. Therefore, it could be hypothesized that they are heterozygous for the T-DNA insertion in *chy2*. Anyways, as commented above, this genotype, although it is not the expected one, could be used for the complementation assay since the β -xanthophyll levels are considerably inferior to the levels in wild type.

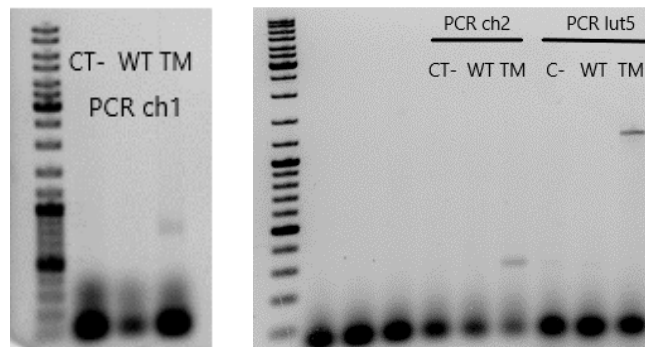


Figure 45. Detection of the T-DNA insertion in the *chy1chy2lut5* plants. After the extraction of gDNA of the *chy1chy2lut5* plant (TM) and the wild type (WT) of *A. thaliana*, the presence of the T-DNA insertion was checked by PCR with specific primers, in which one of them binds the T-DNA and the other the gene that has been inserted. The T-DNA insertion in the genes *chy1* (left figure), *chy2* and *lut 5* (right figure) was confirmed. C-: negative control.

7. Discussion

Oceans contain a wide diversity of organisms that produce an immense array of molecules, some of which can encounter many applications, e.g. in the pharmaceutical, food or cosmetic industry. Therefore, oceans provide a natural resource of products, most of them unexplored yet, that can advance the progress of society, e.g. in the treatment of diseases⁸⁷. Marine products and biomass have a high potential for becoming a key player in the creation of the bioeconomy, a concept that aims to replace production of chemicals and energy from fossil fuels by a biological production from renewable biological raw materials. This radical change in the production systems could be crucial in the future to face some current problems of society, such as depletion of fossil fuels, shortage of food in an ever-growing world population and environmental damage caused by our current production systems⁸⁷. For the bioeconomy to become a reality, bioprocesses must be optimized in order to be efficient and economically competitive. Therefore, advances in the fields of biocatalysis, metabolic engineering and bioprocess engineering are key for the competitiveness of this production approach and will determine if the bioeconomy will just remain as a utopian concept or it will become true in the future.

Fucoxanthin is a compound derived from some marine algae that has several applications in the pharmaceutical and in the nutraceutical industries. In Asian countries, fucoxanthin extracts are produced from the wasted parts of edible macroalgae. However, it is known that microalgae produce at least 10 times more fucoxanthin, per dry weight, than macroalgae³¹. Prior studies that aimed to increase the production yield of fucoxanthin in the microalgae *P. tricornutum* by genetic engineering produced an increase of 2.4 and 1.8-fold compared to wild type, by the over-expression of the *psy* and *dxs* genes, that are the first genes of the carotenoid and MEP pathway, and therefore their gene products direct the flux into these pathways⁵⁴.

In **Chapter 3**, it was aimed to produce further increases in the fucoxanthin yield in *P. tricornutum* by considering other enzymes that are known in other organisms

that are rate-limiting for the production of carotenoids. These are IDI, that is the last enzyme of the MEP pathway, GGPPS, the enzyme involved in the sequential addition of the 5C-precursors of the MEP pathway to form the 20C molecule GGDP, and PDS, that catalyzes the second reaction of the carotenoid pathway. In a first attempt, their genes were additionally introduced into the diatom cells to see what the result of this single manipulation was.

An important fact to bear in mind is that the accumulation of the end products of the MEP pathway is known to produce feedback inhibition of the first enzyme of the MEP pathway in plants¹⁹. Therefore, the increase of the levels of rate-limiting enzymes in the MEP pathway would only be beneficial if it is coupled to an increase in the rate of the carotenoid pathway, to avoid the accumulation of intermediary products that can cause feedback inhibition. However, this regulatory effect has not been studied in diatoms, so the regulation could be different than the situation encountered in plants.

Another key point to consider is that the isoprenoid pool feeds the formation of many cellular compounds. IPP and DMAPP are not only the precursors of carotenoids, but also of monoterpenes, sesquiterpenes, polyprenols, triterpenes and ubiquinones. In turn, GGDP is not only an exclusive precursor of carotenoids, but also of diterpenes and chlorophylls^{10,88,89}. The above-mentioned compounds are formed either from the cytoplasmic or the plastid pool of isoprenoids, formed by the MVA or the MEP pathway, respectively. Although no information about the transport of isoprenoids is known in diatoms, in plants the transport of isoprenoid precursors between the cytosol and the plastid has been detected⁹.

Since both carotenoids and the phytyl group of chlorophylls are formed from the isoprenoid pool of the plastid, it could be thought that the manipulation of the MEP or the carotenoid pathway might influence the synthesis of chlorophyll. However, it must be mentioned that due to the higher ratio of chlorophyll: carotenoid molecules, this effect might not have a high relevance.

Following a very simplistic view, it could be considered that when only the MEP pathway and GGPPS are modified to yield a higher amount of their end products,

the expected result (assuming no feedback inhibition) is that more precursors are available for both the chlorophyll and carotenoid pathways. In contrast, when the carotenoid pathway is modified to be more productive, it might be expected that it consumes more precursors and hence, the available pool for the synthesis of chlorophyll is decreased. However, cell regulation adds control systems to guaranty that all their building blocks are wisely distributed to guarantee the synthesis of all essential products. Consequently, the results of genetic engineering modifications are not easily predicted unless the regulation of the pathway is well-characterized. Therefore, in this study, both carotenoid and chlorophyll levels are expected to be modified. For that reason, the widely used parameter, fucoxanthin to chlorophyll ratio, is not being considered as a normalized measure of the fucoxanthin levels, since chlorophyll levels can also vary as a result of the genetic modification. Instead, the volumetric productivity of fucoxanthin per million cells ($(\mu\text{g Fx}/(\text{cells}\cdot\text{L}\cdot\text{day})\cdot 10^6\text{cells})$) is in this case the measure of choice to normalize the levels of fucoxanthin in the different mutants.

After these considerations, now the results of the insertion of additional copies of the gene *idi* in the genome are discussed. Only 2 positive mutants were obtained. In these clones, both the fucoxanthin and the chlorophyll levels were considerably lower than the wild type. While the cell growth in the mutant IDI3 was very similar to the wild type, the clone IDIY showed an increased cell growth. These results cannot be explained by the hypothesis of the feedback inhibition, since the enzyme IDI does not increase the amount of IPP and DMAPP, but it interconverts IPP into DMAPP. Since both IPP and DMAPP are known in plants to inhibit DXS by displacing its cofactor group, then in principle the increase of levels of IDI would not contribute to the hypothetical feedback inhibition effect. Therefore, since the cell density is not diminished in these mutants, then the reason for the decrease in fucoxanthin and chlorophyll levels is not clear.

Regarding the mutants in which additional copies of *ggpps* were inserted, it could be said that the mutant GGPPS2 behaves similarly to wild type in terms of pigment levels and growth. In contrast, the GGPS1 clone shows slightly reduced levels of fucoxanthin and chlorophyll than wild type, but it has a reduced growth. Therefore, the fact that almost the same levels of fucoxanthin and chlorophyll are

produced with considerably less cells must mean that every cell produces higher amounts of these pigments. The possible causes for the reduced growth are explained later. Although at this stage, this mutant is not interesting for biotechnology purposes, it could in principle accumulate higher levels of fucoxanthin than wild type if measures were taken to avoid the growth retarding effects. In any case, the higher fucoxanthin yield per cell must be confirmed by the production of more mutants with additional copies of *ggpps*, since these effects were only observed in one mutant.

The results of the only positive clone obtained by inserting further copies in the genome of the gene *pds* show that both the fucoxanthin and chlorophyll levels are decreased, while the growth is slightly higher than in wild type. The result of this genetic modification does not have any positive effect in the accumulation of carotenoids.

Since in previous studies by Ulrike et al.⁵⁴ the over-expression of the genes *psy* and *dxs* resulted in an increase in fucoxanthin levels, in this work mutants were created in which additional copies of *psy* and *dxs* were inserted in the genome. The mutant DXS-PSY1 shows very similar levels of pigments and of growth with respect to the wild type. In contrast, in mutant DXS-PSY2 the final cell density is considerably lower, while the fucoxanthin and chlorophyll levels are only slightly decreased. The volumetric productivity per million cells is a 10% higher than in the wild type, so the cells of this mutant produce more fucoxanthin than wild type, but since growth is slower the total fucoxanthin levels are lower than wild type. The increase productivity per million cells is in agreement with the previous studies, since in theory coupling the increase in the rate of the MEP pathway with the carotenoid pathway would avoid the accumulation of intermediaries and would lead to a higher production of carotenoids.

Up to now, only the mutants DXS-PSY2 and GGPPS1 produce an increase in the productivity per million cells, while the mutants of IDI and PDS did not produce any increase. Therefore, triple mutants in which additional copies of the genes *dxs*, *psy* and *ggpps* were inserted into the genome were created, to see if together the increase in productivity per million cells of DXS-PSY (10%) and of

GGPPS (28%) would produce a synergistic increase in the productivity per million cells in the triple mutant. The results of the triple mutant growing in very low light show that indeed the mutants produce higher fucoxanthin levels than wild type and the volumetric productivity per million cells is increased 61% and 32% compared to wild type. These could mean that GGPPS is an important target for the accumulation of carotenoids, possibly because it is a rate-limiting enzyme, like in other organisms, and because it connects the MEP and carotenoid pathway, since this enzyme has as substrates the end products of the MEP pathway and produces the substrate of the carotenoid pathway. Again, cells from this triple mutant are more productive, but in this case due to the culture conditions of very low light that were required to alleviate stress on the mutants, cell density of the mutants and the wild type was approximately half of the cell density in low light conditions. Another fact to consider, is that we can see that the fucoxanthin to chlorophyll ratio is considerably increased in very low light compared to low light conditions, what it is very beneficial for our goal, but since growth is retarded, then in the end the total fucoxanthin levels are lower than in low light.

Therefore, the main issue in the creation of mutants is that even if they show higher productivity per million cells, the retarded growth hampers the attempts to increase the total fucoxanthin levels. The reduction in growth may be caused by the overload that is imposed to the cell by the production and transport of additional amounts of proteins across the 4 membranes of the plastid. This overload might cause a saturation for example of the transport systems in the plastid. In addition, the MEP and carotenoid pathway are anabolic pathways that require high amounts of energy and reduction power, further contributing to the overload of the cells. In this aspect, omic techniques could reveal clues of what is happening at the cellular level in these mutants to explain their stress and slow growth and based on these results, plan what to do to alleviate the cell stress.

A possible strategy to try to avoid the detriment on biomass during the production of fucoxanthin could be to establish a 2 step culture, in which in a first stage the cells are grown under optimal conditions for growth, and after that the production of the rate-limiting enzymes of the carotenoid pathway is induced by the addition

of an inductor. In diatoms, the nitrate reductase promoter is available which is induced by the presence of nitrates. Thus, in the first stage the diatoms could be grown under a different nitrogen source other than nitrates, and hence our genes of interest would be only expressed at low levels (residual expression), followed by the addition of nitrates after an optimal biomass concentration is reached.

Another plausible explanation for the slow growth of the mutants is derived from the fact that in *P. tricornutum* non-homologous end joining is the preferred method over homologous recombination to integrate the DNA material into the genome⁹⁰. Therefore, upon biolistic transformation the plasmid containing the gene of interest is integrated randomly into the nuclear genome. Therefore, every clone is not identical since the gene of interest is integrated into a different location in each case, what leads to very different expression levels depending on the place of insertion (position effect). This random insertion could in some clones cause the disruption of important genes for cell physiology, hence retarding growth⁹⁰. These deleterious effects could be avoided by the use of alternative transformation methods that have been employed in the latest years in diatoms, such as conjugation, in which the episomes with the gene of interest are transferred from bacteria to diatoms⁹¹, and once in diatoms the episome is replicated as an independent unit and therefore the expression of the gene of interest is mainly influenced by its own promoter and disruptions in the chromosome are avoided.

Additionally, the conjugation method shows a higher transformation yield compared to the biolistic method⁹¹, in which usually, a low number of clones (1-20) is obtained after a transformation round. Therefore, when using the biolistic method, the possibility of finding a mutant that grows well and accumulates high amounts of fucoxanthin is low.

A different point to consider is carotenoid storage. Fucoxanthin is a component present in protein complexes of the photosynthetic machinery of diatoms in the thylakoid membranes. It remains unknown whether fucoxanthin is stored in lipid droplets when it accumulates in high amounts, as it occurs with other carotenoids in other organisms. Storage capacity plays an important role in carotenoid

accumulation, and it might be a limiting factor in metabolic engineering projects⁸.

Considering now the results of applying oxidative stress to the cells by the cultivation of diatoms under different concentrations of salts or under high light resulted in the reduction of growth and fucoxanthin levels. A possible explanation for these results is that although fucoxanthin is an antioxidant, its primary role is the harvesting of light for photosynthesis.

An important aspect to take into account is that the measurements reflect the synthesis minus the degradation of carotenoids. Therefore, in stress conditions, where the carotenoid degradation is enhanced, the synthesis must increase more than in non-stress conditions in order to see carotenoid accumulation. Then, it could be possible that the synthesis of carotenoids increased under stress conditions, but it did not compensate for the degradation of carotenoids.

In any case, the methodology used was not the most appropriate to conduct this kind of experiments. It would have been better to carry out this experiment as a 2-step culture, in which in a first step high biomass is accumulated and then the stress is applied. Another method could have been the Adaptive Laboratory Evolution (ALE), in which the strain is over the time adapted to an unfavorable environmental condition, such as oxidative stress, in which carotenoids, as antioxidants, might accumulate. A different strategy consists in the improvement of photosynthetic efficiency as a means to increase fucoxanthin production⁹². This strategy is based on the correlation of fucoxanthin with photosynthetic efficiency and in any case, would lead to the increase of biomass, which is always beneficial to increase productivity.

Regarding the results of **Chapter 4**, it was shown that the microalgae *Chlorella* and *Cyclotella* could be encapsulated without important hurdles concerning the stability of the alginate beads in the culture media. At least during the cultivation time of the microalgae for the screening, that was of 3 days, the capsules remained stable. In addition, the encapsulated cells could proliferate and showed a high viability during this time.

An important point to consider is that the growth inside the encapsulated cells for the screening process must reflect as much as possible the growth of the algae

in the conditions used in the production process. So, if later on, microalgae are also encapsulated in a bioreactor in the production phase then there is no problem since the screening and the production take place using encapsulated algae. However, if in the production free microalgae are used, then one must conduct different experiments to guarantee that the screening method is reflecting the conditions later used in production.

As an example of the different growth of encapsulated and free cells, Lau et al.⁹³ determined that encapsulated *Chlorella* and *Scenedesmus* show a longer lag phase than in a free state, although they reach a similar maximum growth rate. In contrast to this situation, Yoshitomi et al.⁹⁴ observed that in the case of *Chlamydomonas*, encapsulation promoted cell proliferation, and consequently the stationary phase was reached earlier than free cells. In addition to the growth curve, another difference regarding cells in free culture is that usually the formation of cell colonies is promoted. In this study, *Cyclotella* free cells were not forming colonies after 7 days of culture. One could only see cells in the stage of the cell division together. While, for *Chlorella* one could also see colonies of 5-6 cells after 1 week of culture, although they were a small minority. In contrast, encapsulated *Cyclotella* and *Chlorella* cells were growing in colonies that reached a high cell number. The same phenomenon has been observed in *Chlamydomonas* and many other organisms⁹⁴. On the contrary, cell morphology does not seem to be altered in encapsulated algae⁹³.

In order to determine the optimal parameters for cell encapsulation compared to free cells, one could test for instance different polymer porosities to adapt the permeability, so that similar gas and nutrient exchanges are reached compared to the free cells. In other studies it has been determined that alginate does not interfere with light penetration and has little effect on diffusion of nutrients and other small molecules⁹³. However, when reaching high cell densities inside of the alginate beads, then the diffusion of light to the cells could be hampered⁹³. For these purposes, one should select the highest diameter of the bead, eg, 500 μm , that allows good diffusion while allowing more cells to grow inside them not in such a concentrated state as in smaller beads.

So, in the case in which cell growth inside the alginate beads is comparable to the free state, the encapsulation of cells represents a high advantage since it

allows the cultivation of different mutants in the same vessel. A one-pot culture presents many benefits, such as the homogenization of environmental conditions and a higher efficiency regarding time consumption, culture space and economical costs.

The subsequent high-throughput method based on a flow cytometer that sorts the different mutants based on chlorophyll levels based on their emitted fluorescence could be a first screening round in the detection of the best carotenoid accumulating mutants. After the selection of the best mutants, a second round would take place for the determination of carotenoid levels. The advantages of performing this first screening round is that allows a fast screening and selection of a high amount of mutants, although it lacks of specificity, since although chlorophyll levels have a good correlation to carotenoid levels in algae, they are an indirect measure and it is always preferable to measure directly the carotenoids.

So far, the results about the experiments to increase fucoxanthin levels via genetic manipulations and the exploration of a high-throughput method for the screening of mutants have been discussed. However, genetic engineering approaches need of as much information as possible about the pathway and its regulation. In this regard, the regulation of the carotenoid pathway in diatoms is poorly understood and the pathway has not been completely characterized. Enzymes like fucoxanthin and diadinoxanthin synthase have not been identified and for these clade-specific enzymes, no homology search is possible¹⁰. In these cases, one could opt for an experimental approach based on essentiality or phenomic data or for a bioinformatic approach that could guide the search for putative enzymes, that should be verified experimentally. The bioinformatic approach could be based on different kinds of information, such as protein structure considering the substrate-binding site. In **Chapter 5**, the guilt by association principle, more specifically gene clustering and the gene expression patterns was investigated as a bioinformatic means of searching for candidate enzymes for unknown functions. Among these principles, the gene clustering and the gene expression patterns were assessed in the MEP and carotenoid pathway of diatoms.

Regarding gene clustering, in the diatom *P. tricornutum* some of the carotenoid and the MEP genes seem to be grouped in the genome in three main “clusters”, each spanning tens of thousands of bp. This phenomenon does not seem to be an exception in this diatom, since many genes with functions related to photosynthesis and the synthesis of chlorophyll appear to be in close proximity in the genome. The implications of the clusters in the regulation of the carotenoid pathway is unknown, but there might be an association based on chromatin structure, histone modifications or another mechanism of gene regulation operating at a relatively long distance⁷³. As commented before, genome clustering could provide a clue, together with other evidence, that could help in the search for unknown enzymes of the carotenoid pathway⁷². In the end, the function of the candidate enzymes must be always tested experimentally.

Considering gene expression patterns, it was clear that almost all the genes from the carotenoid pathway were upregulated from a shift from dark conditions to blue light. In addition, also under iron limitation conditions many genes of the MEP and carotenoid pathway were highly expressed, although not all of them.

Combining several pieces of evidence from the data on gene clustering, gene expression patterns, presence in different phylogenetic groups and protein location, a list of candidate genes for the function of fucoxanthin synthase was created. Although it was not performed in this work, it is usual to give scores for the prediction of the enzyme candidates. When dealing with several association data from the guilt by association principles, one should normalize the scores of the different types of associations⁹⁵. For the integration of the scores one could apply the direct likelihood ratio to assess the probability distributions of the association scores. In this method it is assumed that individual association scores are independent and that the likelihood of association increases monotonically with the value of the score. Basically, in this method the total likelihood ratio of the candidate enzyme carrying out a given function is calculated as a product of likelihood ratios for every association score⁹⁵. An alternative approach to the likelihood ratio is the Adaboost method, that uses machine learning and it does not assume the independence of the association scores. This method is based on alternating decision trees⁹⁵.

In **Chapter 6** it was investigated if the bioinformatically predicted β -carotene hydroxylase candidates did in fact catalyse the hydroxylation reactions converting β -carotene into zeaxanthin. These candidates had been identified by Coesel et al.⁸² based on homology searches, since the enzymes encoding this function are known in plants.

The approach followed for testing the function of these enzymes was a complementation assay of an *Arabidopsis* mutant showing strongly reduced levels of β -xanthophylls due to the inactivation of some of the genes that can hydroxylate β -carotene. After the transformation of the candidate genes of diatoms in the *Arabidopsis* mutant and the screening process, only was mutant was obtained containing the *lut1-2* gene of *P. tricornutum*. Compared with the *Arabidopsis* mutant, the complemented plant showed a higher zeaxanthin to chlorophyll a ratio, partially restoring the levels of the wild type. Therefore, it could be assumed that the LUT1-2 enzyme of *P. tricornutum* is involved in the formation of zeaxanthin. Recently, Cui et al.⁹⁶ have also confirmed experimentally the function of the *lut1-1* and *lut1-2* genes of *P. tricornutum* based on the transformation of these genes in β -carotene-accumulating *E.coli* mutants and the detection of zeaxanthin in the resulting strain. However, it remains unclear whether this is the full set of enzymes that can catalyse this function and if the LUT1-1 and LUT1-2 enzymes have just overlapping functions or one of them is the preferred enzyme under given conditions in vivo in *P. tricornutum*. The creation of single and double knockouts of these enzymes in diatoms using the CRISPR-Cas technology, recently adapted for diatoms by Nymark et al.⁹⁷, could shed some light on these questions. Besides, the fact that the *lut1.1* gene has a low expression (log₂ value of 0.6) after the shift from dark to blue light conditions, while the *lut1.2* gene is highly expressed in these conditions (log₂ value of 4.4) suggests that the expression of these genes are differentially regulated and thus they might act in different environmental conditions in the cells.

8. Conclusions

The conclusions of this work are the following:

- The genes *psy*, *dxs* and *ggpps* seem to be the best candidates, among the tested genes, to increase the yield of fucoxanthin in diatoms.
- The simultaneous insertion of additional copies of the *psy*, *dxs* and *ggpps* genes into the genome led to the highest volumetric productivity per million cells. However, the slow growth of these mutants make the final productivity lower than the wild type. Therefore, further research is needed to reveal the cause for this decrease in biomass and to apply measures in order to alleviate this problem.
- More information about the regulation of the carotenoid pathway would be highly beneficial to be able to predict the result of genetic modifications.
- The strategy of applying oxidative stress to the cells, such as high light or increased salt concentration, is counterproductive for the accumulation of fucoxanthin.
- The encapsulation of diatoms could serve as an efficient means of cultivating a mutant library for the subsequent high-throughput screening of these mutants, for instance, with a cell sorter. The diatom *Cyclotella* is viable and proliferates forming colonies inside alginate beads, which are mechanically stable in *Cyclotella*'s medium after 3 days.
- The guilt by association principles are a powerful approach to search for candidate enzymes for a given reaction, especially in prokaryotes, but also in eukaryotes with compact genomes, like diatoms. In this work, a list of candidate genes for fucoxanthin synthase was collected by this method, although they need to be tested experimentally.
- The study of complementation of plants seems to point that the *lut1-2* gene of *P. tricornutum* codes for a protein that works as a β -carotene hydroxylase. Further research is needed to clarify the roles of both LUT1-1 and LUT1-2 enzymes *in vivo* in *P. tricornutum*, and to reveal if they constitute the whole set of β -carotene hydroxylases in diatoms.

9. Zusammenfassung

Carotinoide sind Pigmente, die in Pflanzen, Algen, einigen Pilzen und Bakterien vorkommen. Sie spielen eine wichtige Rolle bei der Photosynthese durch Absorption von Licht und beim Lichtschutz. Sie sind verantwortlich für die braunen, roten, orangen und gelben Farben von Obst, Gemüse, Herbstblättern und die Farbe einiger Blumen und Algen. Tiere können keine Carotinoide synthetisieren, daher ist ihre Anwesenheit auf die Nahrungsaufnahme zurückzuführen. Carotinoide sind Tetraterpenoide (40C), die aus Isoprenoidmolekülen (5C) synthetisiert werden. Der Methylerythritol-phosphatweg ist der Carotinoid-Vorläuferweg, der die Isoprenoideinheiten bildet. Carotinoide haben aufgrund ihrer gesundheitlichen Vorteile das Interesse der Nutrazeptika-Industrie geweckt ⁶.

Fucoxanthin ist ein Carotinoid, das nur in Kieselalgen, Braunalgen, Haptophyten und einigen Dinoflagellaten vorkommt. Aufgrund seiner Vorteile zur Vorbeugung von Krebs, kognitiven Erkrankungen und Fettleibigkeit sowie seiner antioxidativen Eigenschaften ist Fucoxanthin ein sehr interessantes Molekül für die Nutrazeptikabranche ¹⁷.

Fucoxanthin hat eine komplexe chemische Struktur mit einer Allenbindung und einer Epoxyketogruppe. Daher wäre seine chemische Synthese kompliziert, da es auch eine stereokontrollierte Synthese erfordert ⁸⁶. Aus diesem Grund ist die Extraktion aus Makroalgen oder Mikroalgen die Methode der Wahl für die kommerzielle Herstellung von Fucoxanthin ^{22,25}.

In dieser Arbeit bestand das Ziel darin, die Fucoxanthin-Produktivität in Kieselalgen mit gentechnischen Methoden zu steigern, damit die Zellen mehr Fucoxanthin produzieren. Zu diesem Zweck wurde der Effekt der Insertion zusätzlicher Kopien von Genen in das Genom untersucht, die für geschwindigkeitsbestimmende oder Schlüsselenzyme im Carotinoid- und MEP-Weg kodieren.

Zu Beginn wurden diese Effekte bei einzelnen Mutanten beobachtet. Letztendlich ist es jedoch das Ziel, eine Mutante zu erzeugen, die mehrere geschwindigkeitsbestimmende Enzyme überexprimiert, um auf diese Weise

Engpässe zu vermeiden. In früheren Studien erreichten Eilers et al.⁵⁴ durch die einmalige Überexpression der *psy*- und *dxs*-Gene in der Kieselalge *P. tricornutum* einen 2.4- und 1.8-fachen Anstieg der Fucoxanthin-Spiegel.

In dieser Arbeit führte die Insertion zusätzlicher Kopien der Gene *idi* und *pds2* nicht dazu, dass die Zellen mehr Fucoxanthin produzieren. Im Gegensatz dazu erreichten die Mutanten mit zusätzlichen Kopien der Gen *ggpps* und mit zusätzlichen Kopien sowohl von *psy* als auch von *dxs* seine um 28% bzw. 10% höhere Fucoxanthin-Produktivität pro Million Zellen. Bei diesen Mutanten ist die Gesamtproduktivität jedoch geringer als beim Wildtyp, da ihr Wachstum langsamer als beim Wildtyp ist.

Unter Berücksichtigung der besten Zielgene wurden Mutanten erzeugt, die gleichzeitig zusätzliche Kopien von *psy*, *dxs* und *ggpps* enthielten. Die Mutanten hatten unter sehr niedrigen Lichtbedingungen eine um bis zu 61% höhere Produktivität pro Million Zellen als der Wildtyp. Ausnahmsweise wurden diese Mutanten bei sehr schwachem Licht ($10 \mu\text{E m}^{-2} \text{s}^{-1}$) gezüchtet, da sie sehr gestresst waren und als Zellklumpen wuchsen. Obwohl die Gesamt-Fucoxanthin-Spiegel in diesen Mutanten unter diesen Bedingungen höher sind als im Wildtyp, sind sie daher niedriger als die Fucoxanthin-Spiegel bei den in anderen Experimenten verwendeten Lichtbedingungen ($50 \mu\text{E m}^{-2} \text{s}^{-1}$). Als Ergebnis dieser Experimente kann gesagt werden, dass die Belastung der Zellen nach den genetischen Veränderungen untersucht werden muss, da dies zu einer Abnahme der Biomasse und folglich zu einer Abnahme der Fucoxanthinproduktion führt. Alternativ könnte auch eine 2-Stufen-Kultur etabliert werden, in der in einem ersten Schritt eine hohe Biomasse erreicht wird und im zweiten Schritt die Expression der interessierenden Gene induziert wird.

Aufgrund der antioxidativen Eigenschaften von Carotinoiden besteht eine übliche Strategie zur Akkumulation von Carotinoiden darin, die Zellen unter oxidative Stressbedingungen zu setzen. Diese Strategie ist jedoch nicht wirksam für die Anreicherung von Fucoxanthin unter hohen Salzkonzentrationen oder hohen Lichtbedingungen. Bessere Versuchspläne könnten jedoch eine 2-Stufen-Kultur oder adaptive Laborbedingungen gewesen sein.

Eine andere mögliche Strategie zur Erhöhung des Fucoxanthinspiegels wäre die Durchführung einer zufälligen Mutagenese der Zellen. Auf diese Weise sind keine Vorkenntnisse über den Carotinoidsyntheseweg und seine Regulation erforderlich und es kann zu Veränderungen in Genen führen, die keine offensichtlichen Ziele sind.

Experimente mit zufälliger Mutagenese erfordern ein Hochdurchsatz-Screeningsystem, da Hunderte oder sogar Tausende von Mutanten erhalten werden. Eine mögliche Strategie, um die Kultivierung der hohen Anzahl von Mutanten zu vereinfachen, ist die Einkapselung dieser Mutanten in Alginatkügelchen. Auf diese Weise können alle Mutanten in demselben Gefäß kultiviert werden. Die eingekapselten Zellen können dann beispielsweise mit einem Durchflusszytometer auf große Partikel durch Fluoreszenz- oder Absorptionsmessungen gescreent werden.

In dieser Arbeit wurde gezeigt, dass sich die Kieselalge *Cyclotella* in Alginatkügelchen vermehren und lebensfähig sein kann. Zusätzlich waren die Alginatkügelchen nach 3 Tagen im Kulturmedium von *Cyclotella* (ASP-Medium) mechanisch stabil.

Ein wichtiger Aspekt bei der Durchführung von metabolischen Engineering-Ansätzen ist die vollständige Charakterisierung des Stoffwechselwegs. In dieser Hinsicht ist die Regulation des Carotinoidsyntheseweges von Kieselalgen weitgehend unerforscht. Zusätzlich sind noch mehrere Enzyme des letzten Teils des Carotinoidsyntheseweges von Kieselalgen zu identifizieren. Die unbekannt Enzyme sind Fucoxanthin-Synthase (wahrscheinlich 2 Enzyme) und Diadinoxanthin-Synthase. Die beiden letztgenannten Enzyme sind stammspezifisch und wurden bisher in keinem Organismus identifiziert.

Zu Beginn dieser Doktorarbeit wurden die Enzyme β -Carotinhydroxylase von Kieselalgen nur bioinformatisch anhand der Homologie zu den β -Carotinhydroxylasen von Pflanzen identifiziert. β -Carotinhydroxylase ist ein Enzym, das die Hydroxylierung an beiden Ringen von β -Carotin katalysiert, um Zeaxanthin zu bilden.

In Pflanzen gibt es zwei Arten von β -Carotinhydroxylasen: Nicht-Häm (NH) -Di-Eisenhydroxylasen und Cytochrom P450 (CYP)-Monooxygenasen⁷⁴.

Nicht-Häm (NH)-Di-Eisenhydroxylasen enthalten konservierte Histidinmotive, die typisch für Fettsäure- und Sterol-Desaturasen sind. Abgesehen von diesen Regionen teilen verschiedene Gruppen von NH-Di-Eisenhydroxylasen eine niedrige Primärstrukturähnlichkeit, obwohl sie einen ähnlichen Reaktionsmechanismus ausführen⁷⁴.

CYPs sind Hämoproteine, die in allen Lebensbereichen vorhanden sind und am oxidativen Metabolismus vieler Substanzen beteiligt sind. Die CYP97-Enzyme wurden mit der Hydroxylierung von β - und α -Carotin in Verbindung gebracht⁷⁶.

In Pflanzen katalysieren die NH-Di-Eisenhydroxylasen CHY1 und CHY2 bevorzugt die Hydroxylierung von β -Carotin. Obwohl β -Carotin kein bevorzugtes Substrat ist, kann das CYP97-Enzym LUT5 diese Reaktion ebenfalls katalysieren⁷⁷.

In der Kieselalge *P. tricornutum* wurde kein Gen gefunden, das Homologie zu den NH-Di-Eisenhydroxylasen zeigte. In der Kieselalge *Phaeodactylum tricornutum* gibt es zwei Enzyme der CYP97-Familie, LUT1-1 und LUT1-2, die eine Sequenzhomologie zu den Carotinhydroxylasen LUT1 und LUT5 in Pflanzen aufweisen⁷⁵.

Um experimentell zu testen, ob die *lut1-1* und *lut1-2* Gene diese Funktion erfüllen, wurde ein Komplementationstest durchgeführt. Dieser Test bestand darin, Pflanzen mit den *lut*-Genen von Kieselalgen zu transformieren, in der mehrere Enzyme, die für die Hydroxylierung von β -Carotin verantwortlich sind, inaktiviert waren (*chy1chy2lut5* knock-out Pflanzen). Daher enthält diese mutierte Pflanze sehr geringe Mengen an β -Xanthophyllen, und durch Insertion der *lut*-Gene aus Kieselalgen, wenn es sich tatsächlich um β -Carotinhydroxylasen handelt, würden wir erwarten, dass die normalen Mengen wiederhergestellt werden. Nach der Transformation wurde nur eine Pflanze mit dem *lut1-2*-Gen aus Kieselalgen erhalten. In dieser Mutante wurden die Spiegel von Zeaxanthin, dem Produkt der Hydroxylierungsreaktion, im Vergleich zur Knock-out-Pflanze wiederhergestellt, wenn auch nicht auf Wildtyp-Niveau. Daher ist das LUT1-2-Enzym an der Hydroxylierung von β -Carotin beteiligt. Cui et al.⁹⁶ zeigen auch in Experimenten mit β -Carotin produzierendem *E. coli*, dass LUT1-1 und LUT-2 die Funktion einer β -Carotinhydroxylase erfüllen. Es bleibt jedoch unklar, ob dies der vollständige

Satz von Enzymen ist, die diese Funktion katalysieren können, und ob die Enzyme LUT1-1 und LUT1-2 nur überlappende Funktionen haben oder ob sie unter verschiedenen Umgebungsbedingungen in *P. tricornutum* unterschiedlich exprimiert werden.

Wie bereits erwähnt, gibt es im Carotinoid-Weg Enzyme, die nicht identifiziert wurden und in denen keine Homologiesuche durchgeführt werden kann. Dies ist der Fall bei der Fucoxanthin-Synthase. In diesen Fällen können "Guilt by Association Principles" hilfreich sein, um sie zu identifizieren. Das Prinzip basiert auf der Tatsache, dass Gene desselben Stoffwechselweges wahrscheinlich viele Gemeinsamkeiten haben, die helfen können, unbekannte Enzyme zu identifizieren. Zu diesen Prinzipien gehören beispielweise Genclustering, Expressionsprofile, Genfusionen, Proteininteraktionen und phylogenetische Zugehörigkeit⁶¹.

In dieser Arbeit wurde festgestellt, dass einige Gene des Carotinoidsyntheseweges und seiner Vorläufer aus Kieselalgen nahe zueinander auf einem Chromosom liegen und dass ihre Expression im blauen Licht gegenüber Dunkelheit stark gesteigert ist.

Je nach enzymatischer Funktion, Expressionsgrad und Position auf dem Chromosom wurden verschiedene Suchanfragen durchgeführt, um das Enzym Fucoxanthin-Synthase zu identifizieren. Darüber hinaus wurden andere Parameter berücksichtigt, wie die Ortung im Plastid und das Vorhandensein in verschiedenen Organismen. Als Ergebnis wurde eine Liste von Kandidatenenzymen gesammelt, die bioinformatisch weiter untersucht und schließlich experimentell getestet werden können.

10. Bibliography

1. Hildebrand M, Davis AK, Smith SR, Traller JC, Abbriano R. The place of diatoms in the biofuels industry. *Biofuels*. 2012;3(2):221-240.
2. Bowler C, Allen AE, Badger JH, et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*. 2008; 456 :239-244.
3. Simpson AGB, Roger AJ. The real “kingdoms” of eukaryotes. *Curr Biol*. 2004;14(17):693-696.
4. Keeling PJ. Diversity and evolutionary history of plastids and their hosts. *Am J Bot*. 2004;91(10):1481-1493.
5. De Tommasi E, Congestri R, Dardano P, et al. UV-shielding and wavelength conversion by centric diatom nanopatterned frustules. *Sci Rep*. 2018;8(1):1-14.
6. Hansen AN, Visser AW. The seasonal succession of optimal diatom traits. *Limnol Oceanogr*. 2019;64(4):1442-1457.
7. Wang W, Yu LJ, Xu C, et al. Structural basis for blue-green light harvesting and energy dissipation in diatoms. *Science*. 2019;363(6427): eeav0365.
8. Ruiz-Sola MÁ, Rodríguez-Concepción M. Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. *Arab B*. 2012;10:e0158.
9. Lohr M, Schwender J, Polle JE. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. *Plant Sci*. 2012;185-186:9-22.
10. Bertrand M. Carotenoid biosynthesis in diatoms. *Photosynth Res* (2010) 106:89-102.
11. https://en.wikipedia.org/wiki/Non-mevalonate_pathway.
12. Schmidt A, Wächtler B, Temp U, Krekling T, Séguin A, Gershenzon J. A bifunctional geranyl and geranylgeranyl diphosphate synthase is involved in terpene oleoresin formation in *Picea abies*. *Plant Physiol*. 2010;152(2):639-655.
13. Dambek M, Eilers U, Breitenbach J, Steiger S, Büchel C, Sandmann G.

- Biosynthesis of fucoxanthin and diadinoxanthin and function of initial pathway genes in *Phaeodactylum tricornutum*. *J. Exp. Bot.* 2012;63(15):5607-5612.
14. Takaichi S. Carotenoids in algae: Distributions, biosyntheses and functions. *Agro Food Ind Hi Tech.* 2013;24(1):55-58.
 15. Dautermann O, Lyska D, Andersen-Ranberg J, et al. An algal enzyme required for biosynthesis of the most abundant marine carotenoids. *Sci Adv.* 2020;6(10):eaaw9183.
 16. Huesgen PF, Alami M, Lange PF, et al. Proteomic Amino-Termini Profiling Reveals Targeting Information for Protein Import into Complex Plastids. *PLoS One.* 2013;8(9):1-12.
 17. <http://www.cbs.dtu.dk/services/SignalP/>.
 18. <http://www.cbs.dtu.dk/services/TargetP/>.
 19. Banerjee A, Wu Y, Banerjee R, Li Y, Yan H, Sharkey TD. Feedback inhibition of deoxy-D-xylulose-5-phosphate synthase regulates the methylerythritol 4-phosphate pathway. *J Biol Chem.* 2013;288(23):16926-16936.
 20. Fanciullino AL, Bidel LPR, Urban L. Carotenoid responses to environmental stimuli: Integrating redox and carbon controls into a fruit model. *Plant, Cell Environ.* 2014;37(2):273-289.
 21. Jahns P, Latowski D, Strzalka K. Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. *Biochim Biophys Acta - Bioenerg.* 2009;1787(1):3-14.
 22. Sandmann G, Römer S, Fraser PD. Understanding carotenoid metabolism as a necessity for genetic engineering of crop plants. *Metab Eng.* 2006;8(4):291-302.
 23. Hohmann-Marriott MF. *The Structural Basis of Biological Energy Generation.*; 2014:31.
 24. Mikami K, Hosokawa M. Biosynthetic Pathway and Health Benefits of

- Fucoxanthin , an Algae-Specific Xanthophyll in Brown Seaweeds. *Int. J. Mol. Sci.* 2013; 14:13763-13781.
25. Gammone MA, D'Orazio N. Anti-obesity activity of the marine carotenoid fucoxanthin. *Mar Drugs.* 2015;13(4):2196-2214.
 26. Xiang S, Liu F, Lin J, et al. Fucoxanthin Inhibits β -Amyloid Assembly and Attenuates β -Amyloid Oligomer-Induced Cognitive Impairments. *J Agric Food Chem.* 2017;65(20):4092-4102.
 27. Lin J, Huang L, Yu J, et al. Fucoxanthin, a marine carotenoid, reverses scopolamine-induced cognitive impairments in mice and inhibits acetylcholinesterase in vitro. *Mar Drugs.* 2016;14(4):67.
 28. <http://bggworld.com/thinogentm-fucoxanthin/>.
 29. <https://www.made-in-china.com/products-search/hot-china-products/Fucoxanthin.html>.
 30. Kanazawa K, Ozaki Y, Hashimoto T, et al. Commercial-scale Preparation of Biofunctional Fucoxanthin from Waste Parts of Brown Sea Algae *Laminaria japonica*. *Food Sci Technol Res.* 2008;14(6):573-582.
 31. Xia S, Wang K, Wan L, Li A, Hu Q, Zhang C. Production, Characterization, and Antioxidant Activity of Fucoxanthin from the Marine Diatom *Odontella aurita*. *Mar. Drugs.* 2013; 11:2667-2681.
 32. <https://www.algatech.com/algatech-triples-fucovital-production-capacity/>.
 33. Kim SM, Jung Y, Kwon O. A Potential Commercial Source of Fucoxanthin Extracted from the Microalga *Phaeodactylum tricornutum*. *Appl Biochem Biotechnol* 2012; 166: 1843-1855.
 34. OECD. *The Application of Biotechnology to Industrial Sustainability-A Primer*. <https://www.oecd.org/sti/biotech/1947629.pdf>.
 35. Lee SY, Mattanovich D, Villaverde A. Systems metabolic engineering , industrial biotechnology and microbial cell factories. *Microb cell fact.* 2012; 11: 156.
 36. Dequin S. The potential of genetic engineering for improving brewing, wine-

- making and baking yeasts. *Appl Microbiol Biotechnol*. 2001;56(5-6):577-588.
37. Stephanopoulos G. Synthetic biology and metabolic engineering. *ACS Synth Biol*. 2012;1(11):514-525.
 38. Nielsen J, Keasling JD. Engineering Cellular Metabolism. *Cell*. 2016;164(6):1185-1197.
 39. Liu Y, Shin H dong, Li J, Liu L. Toward metabolic engineering in the context of system biology and synthetic biology: advances and prospects. *Appl Microbiol Biotechnol*. 2014;99(3):1109-1118.
 40. Yang S-T. *Bioprocessing for Value-Added Products from Renewable Resources.*; 2006:94
 41. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245-248.
 42. Moreno-Sánchez R, Saavedra E, Rodríguez-Enríquez S, Olín-Sandoval V. Metabolic Control Analysis: A tool for designing strategies to manipulate metabolic pathways. *J Biomed Biotechnol*. 2008. 2008: 597913.
 43. He F, Murabito E, Westerhoff H V. Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering. *J R Soc Interface*. 2016;13(117):20151046.
 44. Vemuri GN, Aristidou A a. Metabolic Engineering in the -omics Era: Elucidating and Modulating Regulatory Networks. *Microbiol Mol Biol Rev*. 2005;69(2):197-216.
 45. Castranova V, Asgharian B, Sayre P, Virginia W, Carolina N. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol J*. 2016;10(9):1922-2013.
 46. Farmer WR, Liao JC. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat Biotechnol*. 2000;18(5):533-537.
 47. Brophy J, Voigt C. Principles of Genetic Circuit Design. *Nature Methods*. 2014;11(5):508-520.

48. Dien S Van. *Metabolic Engineering for Bioprocess Commercialization.*; 2016:39.
49. Sun XM, Ren LJ, Zhao QY, Ji XJ, Huang H. Microalgae for the production of lipid and carotenoids: A review with focus on stress regulation and adaptation. *Biotechnol Biofuels.* 2018;11(1):1-16.
50. Schieber M, Chandel NS. ROS function in redox signaling and oxidative stress. *Curr Biol.* 2014;24(10):453-462.
51. Park JC, Choi SP, Hong ME, Sim SJ. Enhanced astaxanthin production from microalga, *Haematococcus pluvialis* by two-stage perfusion culture with stepwise light irradiation. *Bioprocess Biosyst Eng.* 2014;37(10):2039-2047.
52. Yi Z, Su Y, Xu M, et al. Chemical mutagenesis and fluorescence-based high-throughput screening for enhanced accumulation of carotenoids in a model marine diatom *phaeodactylum tricornutum*. *Mar Drugs.* 2018;16(8):272.
53. Yi Z, Xu M, Magnúsdóttir M, Zhang Y et al. Photo-Oxidative Stress-Driven Mutagenesis and Adaptive Evolution on the Marine Diatom *Phaeodactylum tricornutum* for Enhanced Carotenoid Accumulation. *Mar Drugs.* 2015; 13 (10): 6138-6151.
54. Eilers U, Bikoulis A, Breitenbach J, Büchel C, Sandmann G. Limitations in the biosynthesis of fucoxanthin as targets for genetic engineering in *Phaeodactylum tricornutum*. *J Appl Phycol.* 2016; 28:123-129.
55. Saini DK, Chakdar H, Pabbi S, Shukla P. Enhancing production of microalgal biopigments through metabolic and genetic engineering. *Crit Rev Food Sci Nutr.* 2019; 60(3):1-15.
56. Li C, Swofford CA, Sinskey AJ. Modular engineering for microbial production of carotenoids. *Metab Eng Commun.* 2020;10:e00118.
57. Levering J, Broddrick J, Dupont CL, et al. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One.* 2016;11(5):1-22.

58. Bauer C., Vilaça P., Ramlov F. et al. *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference.* ; 2018:139-148.
59. Haire TC, Bell C, Cutshaw K, Swiger B, Winkelmann K, Palmer AG. Robust microplate-based methods for culturing and in vivo phenotypic screening of *Chlamydomonas reinhardtii*. *Front Plant Sci.* 2018; 9:235.
60. Torres PB, Chow F, Furlan CM, Mandelli F, Mercadante A, dos Santos DYAC. Standardization of a protocol to extract and analyze chlorophyll a and carotenoids in *Gracilaria tenuistipitata* var. *liui*. zhang and xia (rhodophyta). *Brazilian J Oceanogr.* 2014;62(1):57-63.
61. Kwiecień I, Kwiecień M. Application of Polysaccharide-Based Hydrogels as Probiotic Delivery Systems. *Gels.* 2018;4(2):47.
62. Genisheva Z, Teixeira JA, Oliveira JM. Immobilized cell systems for batch and continuous winemaking. *Trends Food Sci Technol.* 2014;40(1):33-47.
63. Zhang X, You S, Ma L, Chen C, Li C. The Application of Immobilized Microorganism Technology in Wastewater Treatment. *Mmeceb.* 2016;103-106.
64. Vanella R, Ta DT, Nash MA. Enzyme-mediated hydrogel encapsulation of single cells for high-throughput screening and directed evolution of oxidoreductases. *Biotechnol Bioeng.* 2019;116(8):1878-1886.
65. Iansante V, Dhawan A, Masmoudi F, et al. A new high throughput screening platform for cell encapsulation in alginate hydrogel shows improved hepatocyte functions by mesenchymal stromal cells co-encapsulation. *Front Med.* 2018;5:216.
66. Ghidoni I, Chlapanidas T, Bucco M, et al. Alginate cell encapsulation: New advances in reproduction and cartilage regenerative medicine. *Cytotechnology.* 2008;58(1):49-56.
67. Paques JP, Van Der Linden E, Van Rijn CJM, Sagis LMC. Preparation methods of alginate nanoparticles. *Adv Colloid Interface Sci.* 2014;209:163-171.
68. <http://fgen.ch/en/technologies>.

69. Takaichi S, Mochimaru M. Carotenoids and carotenogenesis in cyanobacteria: Unique ketocarotenoids and carotenoid glycosides. *Cell Mol Life Sci.* 2007;64(19-20):2607-2619.
70. Schubert N, García-Mendoza E, Pacheco-Ruiz I. Carotenoid composition of marine red algae. *J Phycol.* 2006;42(6):1208-1216.
71. Frommolt R, Werner S, Paulsen H, et al. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol Biol Evol.* 2008;25(12):2653-2667.
72. Hanson AD, Pribat A, de Crécy-Lagard V. “Unknown” proteins and “orphans” enzymes: The missing half of the engineering part list - And how to find it. *Biochem J.* 2010;425(1):1-11.
73. Hurst LD, Pál C, Lercher MJ. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.* 2004;5(4):299-310.
74. <https://string-db.org/>.
75. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:607-613.
76. <https://www.ncbi.nlm.nih.gov/gene>.
77. Knig S, Eisenhut M, Brutigam A, Kurz S, Weber APM, Bchel C. The Influence of a Cryptochrome on the Gene Expression Profile in the Diatom *Phaeodactylum tricornutum* under Blue Light and in Darkness. *Plant Cell Physiol.* 2017;58(11):1914-1923.
78. <https://www.ncbi.nlm.nih.gov/geo/>. <https://www.ncbi.nlm.nih.gov/geo/>.
79. <http://www.diatomics.biologie.ens.fr/EST3/est3.php>.
80. Harrison PJ, Newgas SA, Descombes F, Shepherd SA, Thompson AJ, Bugg TDH. Biochemical characterization and selective inhibition of β -carotene cis-trans isomerase D27 and carotenoid cleavage dioxygenase CCD8 on the strigolactone biosynthetic pathway. *FEBS J.*

- 2015;282(20):3986-4000.
81. Cui H, Yu X, Wang Y, et al. Evolutionary origins, molecular cloning and expression of carotenoid hydroxylases in eukaryotic photosynthetic algae. *BMC Genomics*. 2013;14(1):457.
 82. Coesel S, Oborník M, Varela J, Falciatore A, Bowler C. Evolutionary Origins and Functions of the Carotenoid Biosynthetic Pathway in Marine Diatoms. *PLoS ONE*. 2008;3(8): e2896.
 83. Stange C. *Carotenoids in Nature: Biosynthesis, Regulation and Function.*; 2016: 24-25.
 84. Kim J, Smith JJ, Tian L, DellaPenna D. The evolution and function of carotenoid hydroxylases in arabidopsis. *Plant Cell Physiol*. 2009;50(3):463-479.
 85. Fiore A, Dall'Osto L, Fraser PD, Bassi R, Giuliano G. Elucidation of the β -carotene hydroxylation pathway in *Arabidopsis thaliana*. *FEBS Lett*. 2006;580(19):4718-4722.
 86. Kadono T, Kira N, Suzuki K, et al. Effect of an introduced phytoene synthase gene expression on Carotenoid biosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Mar Drugs*. 2015;13(8):5334-5357.
 87. Querellou J, Cadoret J, Allen MJ, Collén J. Introduction to Marine Genomics. *Introd to Mar Genomics*. 2010:287-313.
 88. Prestegard SK, Erga SR, Steinrücken P, Mjøs SA, Knutsen G, Rohloff J. Specific metabolites in a *Phaeodactylum tricornutum* strain isolated from Western Norwegian fjord water. *Mar Drugs*. 2016;14(1):1-17.
 89. Massé G, Belt ST, Rowland SJ, Rohmer M. Isoprenoid biosynthesis in the diatoms *Rhizosolenia setigera* (Brightwell) and *Haslea ostrearia* (Simonsen). *Proc Natl Acad Sci U S A*. 2004;101(13):4413-4418.
 90. Angstenberger M, Krischer J, Aktaş O, Büchel C. Knock-Down of a ligIV Homologue Enables DNA Integration via Homologous Recombination in the Marine Diatom *Phaeodactylum tricornutum*. *ACS Synth Biol*. 2019;8(1):57-69.

91. Karas BJ, Diner RE, Lefebvre SC, et al. Designer diatom episomes delivered by bacterial conjugation. *Nat Commun.* 2015;6: 6925.
92. Wang S, Verma SK, Hakeem Said I, Thomsen L, Ullrich MS, Kuhnert N. Changes in the fucoxanthin production and protein profiles in *Cylindrotheca closterium* in response to blue light-emitting diode light. *Microb Cell Fact.* 2018;17(1):1-13.
93. Pineda A, Pinilla-Agudelo G, Montenegro-Ruiz LC, Melgarejo LM. Does the nutrient concentration of water ecosystems affect growth rates and maximum PSII quantum yield in calcium alginate-encapsulated *Scenedesmus ovalternus* and *Chlorella vulgaris*? *Limnetica.* 2017;36(2):405-425.
94. Yoshitomi T, Kaminaga S, Sato N, Toyoshima M, Moriyama T, Yoshimoto K. Formation of Spherical Palmelloid Colony with Enhanced Lipid Accumulation by Gel Encapsulation of *Chlamydomonas debaryana* NIES-2212. *Plant Cell Physiol.* 2020;61(1):158-168.
95. Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM. Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics.* 2006;7:1-16.
96. Cui H, Ma H, Cui Y, Zhu X, Qin S, Li R. Cloning, identification and functional characterization of two cytochrome P450 carotenoids hydroxylases from the diatom *Phaeodactylum tricornutum*. *J Biosci Bioeng.* 2019;128(6):755-765.
97. Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep.* 2016;6: 24951.

11. Appendix

11.1 Sequence alignments

- Alignment of the partial protein XP_002180941.1 (gene ID: 7201899) and the partial protein XP_002180259.1 (gene ID: 7201112) from *P. tricornutum*.

CLUSTAL O(1.2.4) multiple sequence alignment

```
XP_002180941.1      MTTTTLTPETPSWGVTQNDMMKEDTVLVLDERDNVIGS--ASKKTSHFNAQQPHGIL      58
XP_002180259.1      -----GVNMDQEAMMESDRLIAVNENDRLVRNVNLSKRNGHTFNKETPRAAL      47
                      **   * : *** * : : : * . * : : .   ** : . * . * * : * : . *

XP_002180941.1      HRAFSVVFVFERQSSRMLLQQRRAHSKITFFPNVWNTNTCCSHPLHGMDPNEVDTPADVADGSV  118
XP_002180259.1      HRAFSFFLFDDQ-DRMLLTQRAGTKITFFPNVWNTNTCCSHPLYDMTPNEVDVADD-AYPMF  105
*****.*:*: * .**** * * :*****:*****:.* *****. * * .

XP_002180941.1      RGVKHAAMRKLQHELGI DPALDLHGFRFLTRLHYWAADTVTHGPDAPWGEHEIDYVLFY      178
XP_002180259.1      PGIKHAAVRKCKHELGIAPENLPKEDIQFLTRFHYWAADTVTYGDDTAWGEHEIDYILFL      165
*:*:*:*:* * :***** * * ..:*****:*****:* * : *****:*

XP_002180941.1      VVDKIEALPLLPHPEEVDTRWVTLQELQTMRRDDNLLFSPWFRLLIVEKWLTEWWQNLDV      238
XP_002180259.1      QVNG--QVPVDANPDEVSDYKYVSMEEKMDMEDPDLWSPWFRGIMNRGGFEWWADLR-      222
* :      * :      * : * . * : : : : * : * : * : * * * * * * * * * * *

XP_002180941.1      CMRPDPQQN                247
XP_002180259.1      -----                222
```

- Alignment of the protein LUT1-1 (XP_002185034.1) and LUT1-2 (XP_002178724.1) from *P. tricornutum*.

CLUSTAL O(1.2.4) multiple sequence alignment

```
XP_002185034.1      -----MYSITFLSLTAVVVVACFF---PKRTQSLVVP-FPSSAVRCCFGIH      42
XP_002178724.1      MRSSDYSRAPRWESFVVLAWATASAVLVGNVFOQALPQPVAFGPALVPHLSRRCDLIQR      60
                      :  : : : : * : * * * : * : : . * : * * : :

XP_002185034.1      RQSRVLTSA LFSTTEDKT-----DEQTDNKAGEPATMKG-----EVASSVIT-TEK      87
XP_002178724.1      RVSSDLEDVDVDVDNDAILGSNNIPRRLRIQGRGIPTRKDP IQPLDSMTYESDLIKTWEQ      120
* * * . . . . : * . : * * : . * : * . * :

XP_002185034.1      NP-DAEGLPWWWEV-----VWDL-----DI-----MQVGKSGEEI-----      116
XP_002178724.1      DPSRQKGFDEWIEIKLRRYFAGLRMRDDGVVWRQPSFFDFLVSKSRSDPGNAPRPVGLVDV      180
* :      * : * * * * * * * * * * * * * * * * * * * * *

XP_002185034.1      -----SFGDSA---NVLRTNIEQIYGGFPSLDGCP LAEGELADIGDGMFIGL      161
XP_002178724.1      VKLVL TNSLTSVGLG PALGMAAVPNAVIQKYEGSFFSFI-KGVLGGDLQTLAGGPLFLL      239
. : * :      * : : * : * * * : : * : * . . * : * : *

XP_002185034.1      QNYRNYGSPYKLCFGPKSFLVISDPVQAKHILKDAN-TNYDKGVLAEILEPIMGKGLIP      220
XP_002178724.1      NKYFEVYGPIFNLSFGPKSFLVSDPVMARHVLRETS PDQYCKGMLAEILD PIMGKGLIP      299
: : : . * * : : * . * * * * * * * * * * * * * * * * * * * * *
```

XP_002185034.1	ADPETWSIRRRQIVPAFHKAWEHIVLGFYCNQPLIDTLNKRVDGDKGVEMESLFCSSVA	280
XP_002178724.1	ADPATWKRVRRAIVPSFHKRWLNRMITLFAERAELIADDLQPKSAKQVVMDEERFCVST	359
	*** **.:*** ***:*** **::: ** . : * * * : : . * :*. ****:	
XP_002185034.1	LDIIGLSVFNVEFGSVTQESPVIKAVYSALVEAEHRSMTPAPYWNLPANQLVPRLRKFN	340
XP_002178724.1	LDIIGKAVFNVDYDFGVTDESPIIKAVYRVLREAEHRSSSFI PYWNLPYADQWGGQVFR	419
	***** :****:*****:***:***** .* ***** : ***** *: * : :.*	
XP_002185034.1	SDLKLLNDVLDLITRAKQTRTVEDIEELENRNYNEVQDPSLLRFLVDMRGADIDNKQLR	400
XP_002178724.1	KDMTMLDDILADLINKAVSTRREASIEELEKRE--NEDDPDLLRFLVGMRGEDLSSMVL	477
	.*:.*:.*: * **.:* .* .*****: * : :*****.*** *:. **	
XP_002185034.1	DDLMTMLIAGHETTAAVLTWALFELTK-NPEIMKELQDEIDEV-VGDRMPNYEDIKKMKF	458
XP_002178724.1	DDLMTMLIAGHETTAAMLTWLFLSRGDPGLLKEVQAEVTVLKGKERPDYDDIVAMKK	537
	*****:*****:***:*****: : * :***: * * : * *.. *:*:** **	
XP_002185034.1	LRLVVAETLRMYPEPPLLIIRRCRTPELDPQGAGR---EAKVIRGMDIFMAVYNIHRDERF	515
XP_002178724.1	LRYSLIEALRLYPEPPLLIIRARTEDNLPAGSSDLKSGVKVLRGTDMFISTWNLHRSPDL	597
	** : *:*:*****.*** *:* *:. .**:* *:*:..*:* . :	
XP_002185034.1	WPSPTDFDLRFTRSHSNPDVPGWAGFDPKKWEKLYPNEVASDFALPFGGGARKCVGD	575
XP_002178724.1	WENPEVFDPTRWDRPFNAGIPGWSGNPDKVSGLYPSENAADFALPFGGGQRKCVGD	656
	* .*:.* * : * ..* .:***:*.:. * . * **.* * :***** *****	
XP_002185034.1	EFAILEATVTLAMVLRREFSFDSEKFEKDDILSSAQGLNHPVGMRTGATIHTRNLHL	635
XP_002178724.1	QFAMMEATVTMALMIKKYDFDFAIP-----AEDVGMKTGATIHTMNLMLM	701
	:***:*****:***:***:*. * . ***:***** ** :	
XP_002185034.1	VVEKRGVPK-----	644
XP_002178724.1	RARQVNEPEPVQSAEGYWEMQHLKRGLNANGRPYTTEEEAVWQTSERLSHKKKEEKPNGE	761
	..: . :	
XP_002185034.1	----- 644	
XP_002178724.1	GGCPMHKM 769	

- Alignment of the proteins LUT1 from *A. thaliana* (NP_190881.2), LUT5 from *A.thaliana* (NP_564384.1), LUT1-1 from *P. tricornutum* (XP_002185034.1),LUT1-2 from *P. tricornutum* (XP_002178724.1)

CLUSTAL O(1.2.4) multiple sequence alignment

NP_190881.2	-----MESSLFSPPSSSYSSLFTAKPTRLLSPPKPKFTF	33
NP_564384.1	-----MAMAFPLSYTPTITVKPVTYSRRSNFVVF	29
XP_002185034.1	-----MYSITFLSLTAVVVVACFF---PKRTQSLVVP-FPSSAVRCCFGIH	42
XP_002178724.1	MRSSDYSRAPRWESFVVLAWATASAVLVGNVQFQALPQPVQAFGPALVPHLSRRCDLIQR	60
	:	
NP_190881.2	SIRSSIE-----	40
NP_564384.1	SSSSNGRD-----PLEENS-----VPNGVKSLKQLQ-----EE	57
XP_002185034.1	RQSRVLTSAFSTTEDKT-----DEQTDNKAGEPATMKG-----EVASSVIT-TEK	87
XP_002178724.1	RVSSDLEDVDVDVNDAILGSNNIPRRLRIQGRGIPTRKDP IQPLDSMTYESDLIKTWEQ	120
NP_190881.2	KP-----KPKLETNSS-----KSQSW-----VS---	58
NP_564384.1	KR-----RAELSARIA-SGAFTVRKSSFST-----VKNGLSKIGI---	92
XP_002185034.1	NP-DAEGLPWWWEV-----VWDL-----DI-----MQVGKSGEEI-----	116
XP_002178724.1	DPSRQKGFDEIEKLRRYFAGLRMRDDGVVWRQPSFFDFLVSKSRSDPGNAPRPVGLVDV	180
	.	
NP_190881.2	-----PDWLTTLTRTLSSGKNDESGIPIANAKLDDVADLLGGALF	98
NP_564384.1	-----PSNVLDFMFDWTGSDQD---YPKVPEAKGSIQAVRNEAFF	129
XP_002185034.1	-----SFGDSA---NVLRTNIEQIYGGFPSS---LDGCPLAEGELADIGDTMF	158
XP_002178724.1	VKLVLTNLSLTSVGLGPALGMAAVPNAVIQKYEGSFFS---FI-KGVLGGDLQTLAGGFLF	236
	..: . : *:	

NP_190881.2	LPLYKWMNEYGPIYRLAAGPRNFVIVSDPAIAKHVLRNY--PKYAKGLVAEVSEFLFGSG	156
NP_564384.1	IPLYELFLTYGGIFRLTFGPKSFLVISDPVQAKHILKDAN-KAYSKGILAEIILDFVMGKG	188
XP_002185034.1	IGLQNYRNYGSPYKLCFGPKSFLVISDPVQAKHILKDAN-TNYDKGVLAELILEPIMGKG	217
XP_002178724.1	LLLNKYFEVYGPIFNLSFGPKSFLVSDPVMARHVLRETSPPDYCKGMLAEIILDPIMGKG : * : ** :.* **:..:*** **:*: : * **:*: : : :.*	296
NP_190881.2	FAIAEGPLWTARRRAVPSLHRRYLSVIVERVFCKCAERLVEKLPYAEDGSAVNMEAKF	216
NP_564384.1	LIPADGEIWRRRRAIVPALHQKYVAAMI-SLFGEASDRLCQKLDAALKEEVEMESLF	247
XP_002185034.1	LIPADPETWSIRRRQIVPAFHKAWEHIV-GLFGYCNQPLIDTLNKRVDGDKVEMESLF	276
XP_002178724.1	LIPADPATWKVRRRAIVPSFHKRWLNRMI-TLFAERAEILADDLQPKSAKQVVDMEERF : * : * ** *:***: : : : : * : * : * : . ** : * *	355
NP_190881.2	SQMTLDVIGLSLFNFDLSLTDSVPVIEAVYTALKEAELRSTDLLPYWKIDALCKIVPRQ	276
NP_564384.1	SRLTLDIIGKAVFNFDLSLTNDTVIEAVYTVLREAEDRSVSPIPVWDIPIWKDISPRQ	307
XP_002185034.1	CSVALDIIGLSVFNVEFGSVTQESPVIKAVYSALVEAEHRSMTPAPYWNPLANQLVPRRL	336
XP_002178724.1	CSVTLDIIGKAVFNDFGSVTDSPVIEAVYTVLREAEDRSVSPIPYWNLPYADQWGGQ . : ** : * : ** : * : * : * : : : : * * * * * * * * . * . : .	415
NP_190881.2	VKAEKAVTLIRETVEDLIAKCKEIVERE-GERINDEEYVNDADPSILRFLLA-SREEVSS	334
NP_564384.1	RKVATSLKLINDTLDLIATCKRMVEEE-ELQ-FHEEYMNERDPSILHFLLA-SGDDVSS	364
XP_002185034.1	RKFNSDLKLLNDVLDLITRAKQTRTVEDEELENRNYNEVDPSLLRFLVDMRGADIDN	396
XP_002178724.1	VEFRKDMTMLDDILADLINKAVSTREASIEELEKRE--NEDDPSLLRFLVGMRGEDLSS : . : : : : : * * . . . : : * * : * * : : : . . .	473
NP_190881.2	VQLRDDLLSMLVAGHETTGSVLTWTLYLLSK-NSSALRKAQEEVDRV-LEGRNPAFEDIK	392
NP_564384.1	KQLRDDLMTMLIAGHETSAAVLTWTFYLLTT-EPSVVAKLQEEVDSV-IGDRFPTIQDMK	422
XP_002185034.1	KQLRDDLMTMLIAGHETTAAVLTWALFELTK-NPEIMKELQDEIDEV-VGDRMPNYEDIK	454
XP_002178724.1	MVLRDDLMTMLIAGHETTAAMLTWLTFELSRGDPGLLKEVQAEVTVLKGKERPDYDDIV *****: ** : * * * * : : : : : * : : : : * * * * * . * : * :	533
NP_190881.2	ELKYITRCINESMRLYPHPPVLIIRRAQVPDILPG-----NYKVNTGQDIMISVYNIHR	445
NP_564384.1	KLKYTRVMNESLRLYPQPPVLIIRRSIDNDI-LG-----EYPIKRGEDIFISVWNLHR	474
XP_002185034.1	KMKFLRLVVAETLRMYPEPPLLIRRCRTPDELPGAGR---EAKVIRGMDIFMAVYNIHR	511
XP_002178724.1	AMKKLRYSLIEALRLYPEPPLLIRRTEDNLPAGSSDLKSGVKVLRGDTMDFISTWNLHR : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :	593
NP_190881.2	SSEVWEKAEFLPERFDIDG-----AIPNETNTDFKFI PFSGGPRK	486
NP_564384.1	SPLHWDDAEKFNPERWPLDG-----PNPNETNQNFSYLFFGGGPRK	515
XP_002185034.1	DERFWSPDTFDPLRFTRSHSNPDVPGWAGFDPKKWEGKLYPNEVASDFAFLPFGGGARK	571
XP_002178724.1	SPDLWENPEVFDPTRWRPFPNAGIPGWSGYNPKVSG-LYPSENAADFALPFGGGQRK . * . : * * * : * . * : * : * * * * *	652
NP_190881.2	CVGDQFALMEAIVALAVFLQRLNVELVPD-----Q-TISMTTGATIHTTN	530
NP_564384.1	CIGDMFASFENVVAIAMLIIRRFNFQIAPG-----APPVKMTTGATIHTTE	560
XP_002185034.1	CVGDQFALMEATVTLAMVLRREFSFDSEKFEKDDILSSAQGLNHPVGMRTGATIHTRN	631
XP_002178724.1	CVGDQFAMMEATVTMALMIKKYDFDFAIP-----AEDVGMKTGATIHTMN *: ** * * : * **: * : : : : : : : : : : * * * * * * : :	697
NP_190881.2	GLYMKVSQR-----	539
NP_564384.1	GLKLTVTKRTPKPLDIPSVPII-PMDTSRDEVSS-----ALS-----	595
XP_002185034.1	GLHLVVEKRGVPK-----	644
XP_002178724.1	GLMMRARQVNEPEVQSAEGYWEMOHLKRGLNANGRPYTTEEEAVWQTSERLSHKKEEEK ** : . :	757
NP_190881.2	-----	539
NP_564384.1	-----	595
XP_002185034.1	-----	644
XP_002178724.1	PNGEGGCPMHKM	769

11.2 Gene identity

Table 9. Genes used in the experiments and its identity.

Gene	Organism	Gene ID
<i>dxs</i>	<i>P. tricornutum</i>	7204829
<i>psy</i>	<i>P. tricornutum</i>	7199567
<i>ggpps</i>	<i>P. tricornutum</i>	7197706
<i>zds</i>	<i>P. tricornutum</i>	7196365
<i>idi</i>	<i>P. tricornutum</i>	7201899
<i>pds2</i>	<i>P. tricornutum</i>	7198353
<i>zep</i>	<i>A. thaliana</i>	836838
<i>lut1-1</i>	<i>P. tricornutum</i>	7198820
<i>lut1-2</i>	<i>P. tricornutum</i>	7199887
<i>chy1</i>	<i>A. thaliana</i>	828675
<i>chy2</i>	<i>A. thaliana</i>	835334
<i>lut5</i>	<i>A. thaliana</i>	840067

11.3 List of primers

- The primers used for the cloning of the MEP and carotenoid genes are the following:

dxs gene of *P. tricornutum*:

Fw: TCGAATTCATGCGTCTATCCAGCGCTC

Rv: AGTAAGCTTCTAGTCCTGCAATTGCGGAAC

psy gene of *P. tricornutum*:

fw: GAATTCGATGAAAGTTTCGACAAAGCTCTG

rv: AAGCTTTCATACTTGATCCAATTGGACC

ggpps gene of *P. tricornutum*:

fw: TCACCACTTGTGCGAACGGAATTCATGCGCGTTTCTCTGTTG

rv: TGCATGCCTGCAGGTGCGACTTAATTCTTACGATTGAT

Signal and target peptide of the gene *zds* of *P. tricornutum*:

fw: TCACCACTTGTGCGAACGGAATTCATGAGGTTGTTGTTTGCT

rv: GACTCTAGAGGATCCCCGGGTACCGCGAGCCATGAGATTTT

idi gene of *P. tricornutum*:

fw: CAAAATCTCATGGCTCGCGGTACCATGACGACGACGACGACG

rv: CTGCAGGTCGACTCTAGATTAGTGATGGTGATGGTGATGGTTCTGTTGG

GGATCGGGCC

pds2 gene of *P. tricornutum*:

fw: GTCACCACTTGTGCGAACGGAATTCATGAAGCTTGTGTTCTCG

rv: CCTGCAGGTCGACTCTAGACTAGTGATGGTGATGGTGATGCACAACGA

CGGCCCCCTC

ggpps cassette:

Fw: TGTAAGTACTGAGAGTGACCCACATATGCGGAAGTACTG

Rv: TACAGTCACTTCCGCATATGCGTTCATTTTAGATCCT

Target peptide of *zep* gene of *A. thaliana*:

fw: AGGACAGCCCAAGCTTCGACTCTAGAATGGGTTCAACTCCGTTTTG

rv: CCGAATTCGAGCTCGCCCGGGGATCCCTCCTTCTCAACTAACGCCG

lut1-1 gene (amplification of *lut1-1_144*) of *P. tricornutum*:

fw: GTTAGTTGAGAAGGAGGGATCCACATCTGCCTTGTCTCCAC

rv: TGGTGATTTTCAGCGTACCGAATTCTTAGTGATGGTGATGGTGATGTTTC

GGTACCCACGCTTC

lut1-1 gene (amplification of *lut1-1_270*) of *P. tricornutum*:

fw: GTTAGTTGAGAAGGAGGGATCCGCCGAGGGTCTCCCATGGTG

rv: TGGTGATTTTCAGCGTACCGAATTCTTAGTGATGGTGATGGTGATGTTTC

GGTACCCACGCTTC

lut1-2 gene (amplification of *lut1-2_180*) of *P. tricornutum*:

fw: AGTTGAGAAGGAGGGATCCCCGGGTCGCGTCTCGTCCGACCTG

rv: TGGTGATTTTCAGCGTACCGAGCTCTTAGTGATGGTGATGGTGATGCAT

CTTGTGCATTGGACA

lut1-2 gene (amplification of *lut1-2_360*) of *P. tricornutum*:

fw: AGTTGAGAAGGAGGGATCCCCGGGTGACCCTTCCCGACAAAAG

rv: TGGTGATTTTCAGCGTACCGAGCTCTTAGTGATGGTGATGGTGATGCAT

CTTGTGCATTGGACA

sul selective marker:

fw: TCATTTGGAGAGGACAGCCCAAGCTTATGGCTTCTATGATATCCTC

rv: GTGATTTGAGCGTACCGAATTCGAGCTCCTAGGCATGATCTAACCCCTC

sul cassette:

fw: TATTGCGCCGCTCTTAGACG

rv: GCTGAGTGCATAACCACCAG

- Primers used for the screening of the diatom mutants:

psy gene:

fw: TCAAAGTCACCACTTGTGCGAACG

rv: AACTTTGAGGCGGCGTCGCCAG

dxs gene:

fw: TCAAAGTCACCACTTGTGCGAACG

rv: GCTTGAGTTGCCTCATGTCAAG

ggpps gene:

fw: TCAAAGTCACCACTTGTGCGAACG

rv: CAAAGGCAGCTTATCCGC

pds2 gene:

Fw: GTCACCACTTGTGCGAACGGAATTCATGAAGCTTGTGTTCTCG

Rv: CCTAACAGTGATCCTCATCGTCG

idi gene:

Fw: TCAAAGTCACCACTTGTGCGAACG

Rv: CTGCAGGTCGACTCTAGATTAGTGATGGTGATGGTGATGGTTCTG
TTGGGGATCGGGCC

- Primers used for the detection of the T-DNA insertions in the *chy1chy2lut5* plants:

chy1-tdna:

Fw: CATTCAAACCACTCCACCGC

Rv: TAGCATCTGAATTTTCATAACCAAT

chy2-tdna:

Fw: TAGCATCTGAATTTTCATAACCAAT

Rv: GCCCAAACCTCCATCCCAA

lut5-tdna:

Fw: CGTACTCTCGGAGATCGAAC

Rv: TGGTTCACGTAGTGGGCCATCG

- Primers used to detect whether the *chy1*, *chy2* and *lut5* genes of *A. thaliana* are expressed in the *chy1chy2lut5* plants:

chy1 gene

Fw: CATAGAGCTCTGTGGCACGC

Rv: AGGGACGTCGGCGATGGGAC

chy2 gene

Fw: AGCAGGACTATCAACAATCG

Rv: GCCCAAACCTCCATCCCAA

lut5 gene

Fw: CGTACTCTCGGAGATCGAAC

Rv: TATCAAGTGTCAAACGAGAG

11.4 Vector information

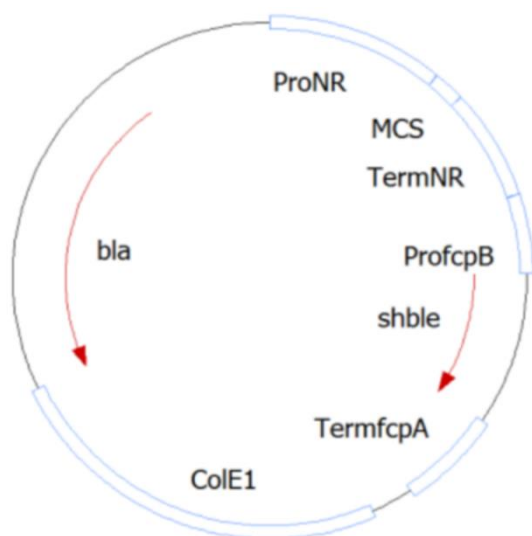


Figure 47. pPhanNR vector. This vector (3860bp), in which a gene of interest was cloned in it, was used for the transformation of the diatom *P. tricornutum*. Abbreviations: **bla**, Ampicillin resistance; **shble**, Zeocin resistance; **ColE1**, Replication origin; **MCS**, Multiple cloning site; **Pro**, Promoter; **Term**, Terminator; **NR**, nitrate reductase (Hempel et al., 2009).

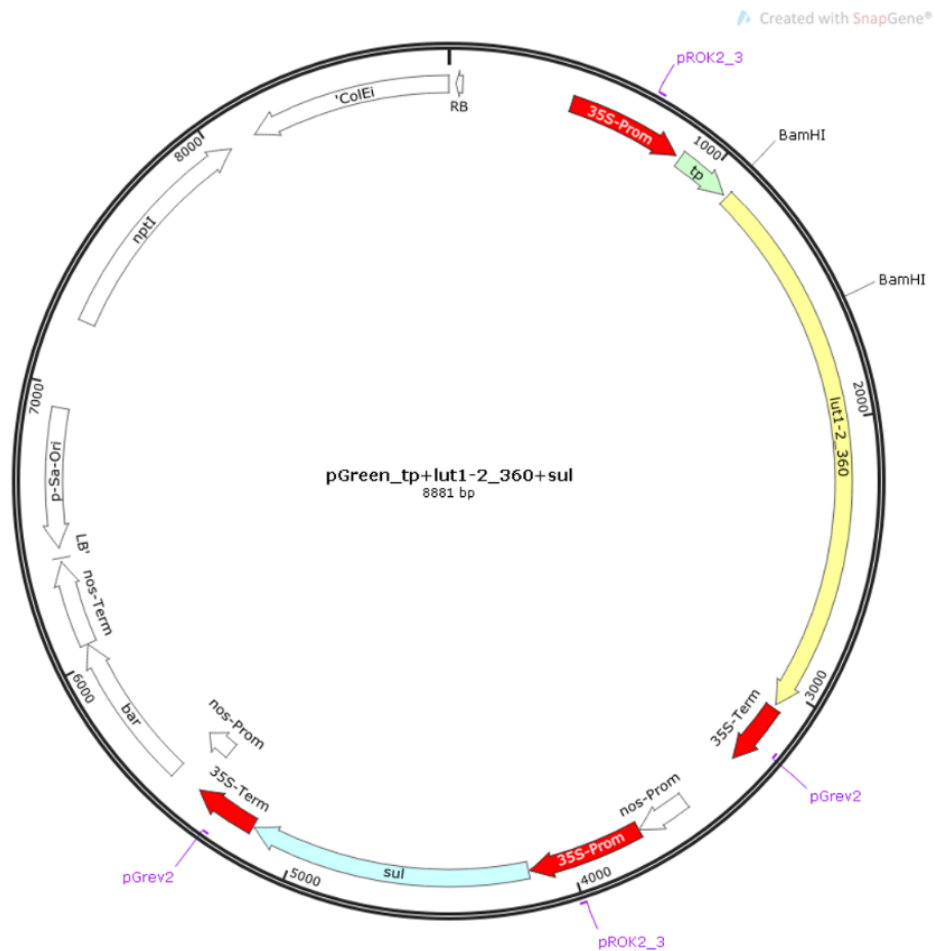


Figure 48. pGreen vector with the *lut1-2* and *sul* genes. Abbreviations: **Prom**, Promoter; **Term**, Terminator; **tp**, target peptide; **bar**, Phosphinothricin N-acetyltransferase; **RB/LB**, right/left border; **ColE1**, Replication origin; **nos**, nitric oxide synthase; **nptI**, kanamycin resistance gene.

Table 10. List of vectors created or used during the thesis.

Vector	Person who created it
pPhaNR-Zeocin- <i>psy</i>	Ulrike Eilers
pPhaNR- <i>nat-dxs</i>	Ulrike Eilers
pPhaNR-Zeocin-ggpps	Alba Blázquez
pPhaNR-Zeocin- <i>idi</i>	Alba Blázquez
pPhaNR-Zeocin- <i>pds2</i>	Alba Blázquez
pPhaNR-Zeocin- <i>psy-ggpps</i>	Alba Blázquez
pUC18- <i>lut1-1</i>	Jürgen Breitenbach

pUC18-<i>lut1-2</i>	Jürgen Breitenbach
pGreen- <i>zep</i> TP*-<i>lut1-1_144-sul</i>	Alba Blázquez; Marina Freyberg
pGreen-<i>zep</i> TP-<i>lut1-1_270-sul</i>	Alba Blázquez; Marina Freyberg
pGreen-<i>zep</i> TP-<i>lut1-2_180-sul</i>	Alba Blázquez; Marina Freyberg
pGreen-<i>zep</i> TP-<i>lut1-2_360-sul</i>	Alba Blázquez; Marina Freyberg

*TP: target peptide.

11.5 List of organisms

Table 11. List of organisms created, derived from *E.coli* XL1.

Name	Organism description	Person who created it
<i>E.coli-ggpps</i>	<i>E.coli</i> transformed with pPhaNR-Zeocin- <i>ggpps</i>	Alba Blázquez
<i>E.coli-idi</i>	<i>E.coli</i> transformed with pPhaNR-Zeocin- <i>idi</i>	Alba Blázquez
<i>E.coli-pds</i>	<i>E.coli</i> transformed with pPhaNR-Zeocin- <i>pds2</i>	Alba Blázquez
<i>E.coli-psy-ggpps</i>	<i>E.coli</i> transformed with pPhaNR-Zeocin- <i>psy-ggpps</i>	Alba Blázquez
<i>E.coli-lut1-1_144</i>	<i>E.coli</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-1_144-sul</i>	Marina Freyberg
<i>E.coli-lut1-1_270</i>	<i>E.coli</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-1_270-sul</i>	Marina Freyberg
<i>E.coli-lut1-2_180</i>	<i>E.coli</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-2_180-sul</i>	Marina Freyberg
<i>E.coli-lut1-2_360</i>	<i>E.coli</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-2_360-sul</i>	Marina Freyberg

Table 12. List of organisms created, derived from *P. tricornutum* UTEX 646.

Name	Organism description	Person who created it
<i>Ptr-psy-dxs-17</i>	<i>P. tricornutum</i> transformed with pPhaNR-Zeocin- <i>psy</i> and pPhaNR-nat- <i>dxs</i>	Alba Blázquez
<i>Ptr-ggpps-2</i>	<i>P. tricornutum</i> transformed with pPhaNR-Zeocin- <i>ggpps</i>	Alba Blázquez
<i>Ptr-idi3</i>	<i>P. tricornutum</i> transformed with pPhaNR-Zeocin- <i>idi</i>	Alba Blázquez
<i>Ptr-pds2</i>	<i>P. tricornutum</i> transformed with pPhaNR-Zeocin- <i>pds2</i>	Alba Blázquez
<i>Ptr-psy-dxs-ggpps</i>	<i>P. tricornutum</i> transformed with pPhaNR-Zeocin- <i>psy-ggpps</i> and pPhaNR-nat- <i>dxs</i>	Alba Blázquez

Table 13. List of organisms created, derived from *A. tumefaciens* GV3101.

Name	Organism description	Person who created it
<i>A.tum-lut1-1_144</i>	<i>A. tumefaciens</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-1_144-sul</i>	Marina Freyberg
<i>A.tum-lut1-1_270</i>	<i>A. tumefaciens</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-1_270-sul</i>	Marina Freyberg
<i>A.tum-lut1-2_180</i>	<i>A. tumefaciens</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-2_180-sul</i>	Marina Freyberg
<i>A.tum-lut1-2_360</i>	<i>A. tumefaciens</i> transformed with pGreen- <i>zep</i> TP- <i>lut1-2_360-sul</i>	Marina Freyberg

Table 14. List of *A. thaliana* organisms used.

Name	Organism description	Person who created it
<i>At chy1 chy2 lut5</i>	<i>A. thaliana</i> with T-DNA insertions in <i>chy1 chy2 lut5</i>	Gift from Prof. Bassi (Univ. Verona)
<i>At-TM +lut360A</i>	<i>A. thaliana</i> with T-DNA insertions in <i>chy1 chy2 lut5</i> transformed with pGreen- <i>zep TP-lut1-2_360-sul</i>	Marina Freyberg, Alba Blázquez

11.6 Chromosome location of MEP and carotenogenesis genes of *P. tricornutum*

The location is shown in K (kilo base pairs).

-Chromosome 1: *lcy*: 578-580K; *zds*: 925-927K; *crtiso-5*: 1892-1894K, *zep2*: 2222-2224K; *pds-like2*: 2271-2273K; *dxr*: 2304-2306K; *vdr*: 2420-2422K.

-Chromosome 3: *pds-like1*: 645-647K

-Chromosome 4: *ggpps*: 682-683K; *hdr*: 980-982K; *cmk*: 1098-1100K; *vde*: 1116-1117K; *zep3*: 1118-1120K

-Chromosome 5: *lut1-2*: 196-199K; *psy*: 447-449K; *hds*: 765-767K

-Chromosome 6: *crtiso-4*: 585-587K

-Chromosome 7: *crtiso-1*: 329- 331K

-Chromosome 8: *mcs*: 407-408K; *zep1*: 450-452K; *vdI2*: 452-454K; *idi*: 549-550K

-Chromosome 9: *vdI1*: 433-435K

-Chromosome 10: Squalene synthase/*psy*: 88-89K

-Chromosome 15: *crtiso*: 52-56K; *crtiso-2*: 645-648K

-Chromosome 16: *cms*: 56-57K; *crtiso-3*: 66-68K

-Chromosome 23: *pds-like3*: 468-469K

-Chromosome 27: *lut1-1*: 41-43K

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Büchel for all her support and advice throughout the project. I truly appreciate her understanding and reasonable personality. In addition, I am also very thankful to Prof. Sandmann for the proposal of ideas, the discussion of results and the analysis of samples. His expertise and his interest in solving research questions about the carotenoid pathway were a big source of motivation for me. He represents the kind of scientists I admire, the scientists who even at an advanced age, and after all their important contributions to science, have not lost the motivation typical of young people, are still curious and have fun with science. I am extremely thankful to Dr. Breitenbach, Dr. Dietzel, Dr. Angstenberger, Dr. Schmidt and Nico Herrmann for all their advice on carotenoid extraction, cultivation and transformation of plants, molecular biology and HPLC. Their experience in the different techniques and our discussions have been crucial for this thesis. I am also very thankful to Kerstin Pieper for all her help in different lab protocols and for being like our mother in the lab, always showing interest in how we are doing. I would like to thank Holger Schranz for his advice in plant cultivation, especially when the mutant plants were very weak. And in general, I would like to thank all my research group for being always very collaborative with each other and being always ready to help.

I am also very thankful to Dr. Meyer for welcoming me in his company Fgen for the algae encapsulation experiment and to Dr. Dabene for teaching me the encapsulation procedure and being always very nice.

Finally, I would like to thank my friends, for making these 3 years in Frankfurt an unforgettable experience and my parents, for being always there for me.

Curriculum vitae

Personal Information

Alba Blázquez Pla

Xàtiva - Spain

Alba5vlc@gmail.com

Education

-Master's Degree in Advanced Biotechnology.

Sept. 2015 – July 2016. Autonomous University of Barcelona.

Master Thesis: A More Environmentally-Friendly Production of Fatty Acid Esters with Potential Applications as Biocides.

-Bachelor's Degree in Biotechnology.

Sept. 2010 – July 2015. University of Valencia.

Erasmus Program in Belgium for 1 year

Bachelor Thesis: Therapeutic Approaches for the Treatment of Autosomal Dominant Polycystic Kidney Disease in Animal Models.

Work Experience

-Early Stage Researcher / EU Marie Curie Program- SE2B (Solar Energy to Biomass) in Goethe Universität Frankfurt.

October 2016- Sept. 2019. Institute of Molecular Biosciences. Frankfurt am Main.

Duties: Molecular Biology and Analytical Techniques (HPLC). Culture and Transformation of Microalgae and Plants.

-Research Intern in Leitat Technological Center.

December 2015- July 2016. Biotechnology Department. Terrassa.

Duties: Optimization of Enzymatic Reaction by Design of Experiment. Up-scaling to a 2L Bioreactor. Microbiological Assays in Different Formulations.

-Research Intern in the Institute for Health Research La Fe.

February- May 2015. Group of Regeneration and Cardiac Transplant. Valencia.

Duties: Cell culture. Histology. Immunohistochemistry.

Languages

-Spanish, Catalan: Mother tongues.

-English: Fluent.

-German: Intermediate.

Further Training

-GxP Course (GMP, GLP, GCP, PV)

23-28 Sept. 2019. Pharmaakademie.

-Summer School in Synthetic and Systems Biology.

25-29 July 2018. Siena.