Centro de Investigación Científica de Yucatán, A.C.
Posgrado en Ciencias Biológicas

MOLECULAR CLONING AND CHARACTERIZATION OF A GENE ENCODING A DIACYLGLYCEROL ACYLTRANSFERASE LIKE PROTEIN IN THE GREEN MICROALGA *Chlorella saccharophila*

Tesis que presenta

LUIS ERNESTO OSUNA ROSALES

En opción al título de
MAESTRO EN CIENCIAS
(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México
Julio 2013

Centro de Investigación Científica de Yucatán, A.C.

Posgrado en Ciencias Biológicas

MOLECULAR CLONING AND
CHARACTERIZATION OF A GENE ENCODING A
DIACYLGLYCEROL ACYLTRANSFERASE LIKE
PROTEIN IN THE GREEN MICROALGA Chlorella
sacchariophila

Tesis que presenta

LUIS ERNESTO OSUNA ROSALES

En opción al título de

MAESTRO EN CIENCIAS

(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México

Julio 2013

RECONOCIMIENTO

Por medio de la presente, hago constar que el trabajo de tesis titulado: "Molecular cloning and characterization of a gene encoding a Diacylglycerol Acyltransferase like protein in the green microalga *Chlorella saccharophila*", a cargo del estudiante de maestría Luis Ernesto Osuna Rosales, fue realizado en los laboratorios de la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, A.C. bajo la dirección del Dr. Santy Peraza Echeverria, dentro de la Opción Biotecnología perteneciente al Programa de Posgrado en Ciencias Biológicas de este Centro.

Atentamente

Dr. Felipe A. Vázquez Flota

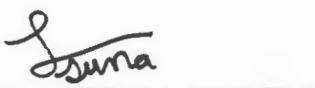
Coordinador de Docencia

Centro de Investigación Científica de Yucatán, A.C.

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos Experimentales, los Resultados y Discusión de este documento proviene de las actividades de experimentación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se regirán en todo caso por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.

Mérida, Yucatán, México a 09 de Julio de 2013



Luis Ernesto Osuna Rosales

AGRADECIMIENTOS

Con especial atención al Dr. Santy Peraza Echeverria, quien además de fungir como director de esta tesis, también fue un guía académico y personal en todo este posgrado.

Igualmente a la Dra. Virginia Aurora Herrera Valencia, cuya paciencia hacia mi persona y colaboración en este proyecto fue crucial para su óptimo desarrollo.

Y a ambos, porque su constante apoyo no sólo se limitó al trabajo sino también a enfrentar las vicisitudes diarias encontradas durante todo este viaje.

A mi comité tutorial integrado por el Dr. Santy Peraza Echeverria, la Dra. Virginia Aurora Herrera Valencia, la Dra. Luisa Alhucema López Ochoa y el Dr. Víctor Suárez Solís, cuyas observaciones y revisiones permitieron finalizar este trabajo adecuadamente.

A los revisores de la tesis, que incluye además de los integrantes del comité tutorial al Dr. Jorge Santamaría, por sus valiosos comentarios y observaciones.

A la Q.B.A. Ileana Cecilia Borges Argáez por el apoyo técnico prestado para la realización de este trabajo.

Al Dr. Amit Dhingra por recibirme en su laboratorio para la realización de una estancia académica en la Washington State University de Julio 15 a Octubre 12 de 2012, así como por su guía, amistad y apoyo durante esta estancia.

Al Centro de Investigación Científica de Yucatán (CICY), la Unidad de Biotecnología y el Posgrado en Ciencias Biológicas del CICY, por las instalaciones y el apoyo prestados para la realización de esta tesis.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por la beca (No. 254372) otorgada para los estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el financiamiento del proyecto de Ciencia Básica 169217.

A la M. en C. Melissa L. Casais Molina por el plásmido pCrGPDH3c.

Al grupo del laboratorio BTD04 de la Unidad de Biotecnología, grupo de "Biotecnología de Microalgas y Cultivos Tropicales": Rodrigo, Maru, Ángeles, Anayeli, Melissa, Patricia, Merly, Alejandra y Miguel, por su amistad, compañerismo y ayuda durante toda mi estancia en la maestría.

A mis padres y hermanos, por ser mi incansable soporte.

TABLE OF CONTENTS

ABBREVIATION LIST

ACCase	Acetyl-CoA carboxylase
AGRF	Australian Genome Research Facility
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
cDNA	Complementary Deoxyribonucleic acid
CoA	Coenzyme A
CO_2	Carbon Dioxide
CTAB	Cetyl trimethylammonium bromide
DAG	Diacyglycerol
DGAT	Diacylglycerol acyltransferase
DHAP	Dihydroxyacetone 3 phosphate
DOE	Department of Energy
ER	Endoplasmic reticulum
EST	Expressed Sequence Tag
FAEE	Fatty Acids Ethyl Esters
FAME	Fatty Acids Methyl Esters
FUE	Far Upstream Elements
G3P	Glycerol-3-phosphate
gDNA	Genomic Deoxyribonucleic acid
GE	General Electric
GPDH	Glycerol-3-phosphate dehydrogenase
HSP	Heat Shock Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JGI	Joint Genome Institute
LB	Luria Bertani
MALDI TOF	Matrix-Assisted Laser Desorption And Ionization Time-Of-Flight
MDB	Membrane Desalting Buffer
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
NADH	Nicotinamide Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
ND	Nitrogen Deficiency
NEB	New England Biolabs
NUE	Near Upstream Elements

ORF	Open Reading Frames
PDAT	Phospholipid:Diacylglycerol Acyltransferase
PLACE	Plant Cis-Acting Regulatory DNA Elements
PLE	Poultry Litter Extract
psi	pounds per square inch
QPCR	Quantitative Polymerase Chain Reactin
RBCS	RuBisCO
RLM RACE	RNA Ligase-Mediated Rapid Amplification Of Cdna Ends
RNA	Ribonucleic Acid
rpm	Revolutions per minute
RuBisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase
SDS	Sodium dodecyl sulfate
TAP	Tris Acetate Phosphate
TE	Tris EDTA
TEN	Tris EDTA NaCl
UADY	Universidad Autónoma de Yucatán
UTR	Untranslated Region
WT	Wild Type

FIGURE INDEX

TABLE INDEX

ABSTRACT

The environmental and health problems caused by the excessive use of fossil fuels and the constant rise on their prices have created the necessity to find new sources of alternative energy. Microalgae have shown great potential as a solution, not only because they contribute to the capture of environmental CO_2 but they are also efficient generators of biomass and highly energetic molecules, such as triacylglycerides (TAGs). These lipidic molecules are used as raw material for the production of biodiesel. There are several enzymes involved in the biochemical route of TAGs synthesis, among them Diacylglycerol acyltransferase (DGAT) is an enzyme committed to perform the last step in the formation of TAGs. The aim of the present study was the molecular characterization of at least one DGAT homologue gene in *Chlorella saccharophila*, a green microalga native of the state of Yucatan in Mexico. This microalga is capable of reaching high biomass and TAGs productivities, with an adequate fatty acid profile for biodiesel production. Using a set of degenerate primers, one putative DGAT gene was isolated and named as *CsDGAT*. From the bioinformatic analysis it was concluded that this gene belongs to the type 2 of DGAT sequences. Both 5' and 3' cDNA termini were fully mapped, and its analysis suggested possible post-transcriptional regulation of this gene. Furthermore, we evaluated the expression of *CsDGAT* during nitrogen starvation and we found the presence of transcripts under this condition. In addition, we subcloned *CsDGAT* in an expression vector under the control of a strong, constitutive promoter and used it to transform the green microalga *Chlamydomonas reinhardtii*. We obtained four transgenic lines of *C. reinhatdtii* which will be used in further studies to evaluate the effect of over-expressing this gene on the levels of TAGs in the transgenic lines.

RESUMEN

Los daños ambientales y a la salud causados por el uso excesivo de combustibles fósiles y el aumento constante en sus precios han creado la necesidad de encontrar nuevas fuentes de energía alternativa. Las microalgas han mostrado tener un gran potencial como solución, no sólo porque ayudan a la captura de CO_2 ambiental sino también porque son eficientes generadoras de biomasa y moléculas altamente energéticas tales como los triacilglicéridos (TAGs). Estas moléculas lipídicas son usadas como materia prima para la producción de biodiesel. Existen varias enzimas que participan en la ruta bioquímica de la síntesis de TAGs, entre las cuales está la Diacilglierol Aciltransferasa (DGAT). Dicha enzima está comprometida a llevar a cabo el último paso en la formación de TAGs. El objetivo del presente estudio fue la caracterización molecular de un gen DGAT en *Chlorella saccharophila*, una microalga verde nativa del estado de Yucatán en México. Esta microalga es capaz de alcanzar altas productividades de biomasa y TAGs, con un perfil adecuado de ácidos grasos para la producción de biodiesel. Usando una estrategia de RT-PCR degenerada se logró obtener un producto del tamaño esperado. De acuerdo al análisis bioinformático se concluyó que este gen pertenece a las secuencias DGAT tipo 2 y se le nombró como *CsDGAT*. Los extremos 5' y 3' del correspondiente ADNc fueron secuenciados en su totalidad, y su estructura sugirió una posible regulación post-transcripcional para este gen. Asimismo, se encontró que la expresión de *CsDGAT* estuvo presente en condiciones de deficiencia de nitrógeno. Adicionalmente, subclonamos a *CsDGAT* en un vector de expresión bajo el control de un promotor constitutivo fuerte, y fue usado para transformar a la microalga modelo *Chlamydomonas reinhardtii*. Obtuvimos cuatro líneas de *C. reinhardtii* las cuales serán utilizadas para evaluar el efecto de sobre expresar este gen en los niveles de TAGs de las líneas transgénicas.

INTRODUCTION

Humanity has relied on the use of fossil fuels as energy source for the last century. Unfortunately, this continued and massive use of such petroleum-based chemicals has resulted on dangerous environmental pollution, an increase on CO_2 levels and dependence on politically unstable regions to fill energy demands. To address these issues, including the projected increase on oil prices, it will be necessary to develop new and sustainable strategies to satisfy the global demand for fuel [Chisti 2007; Radakovits *et al.* 2010]. One of the most promising approaches to face this challenging future is the utilization of biological entities as raw material to produce fuels [Schenk, 2008]. A clear example of this is biodiesel produced by transesterification of vegetable oils or animal fat, which has several advantages over its chemical counterpart: the released carbon produced by its combustion does not represent a new addition to the atmosphere because it had been previously fixed during its production, biodegradability, low or zero toxicity and fewer residual amounts such as carbon monoxide or sulfur compounds are some examples of the environmental-friendly properties of this kind of biofuel [Durret 2008; Greenwell 2010].

However, despite the advantages offered by vegetable crops as a souce of oils for biodiesel, there are also a number of issues to consider. The use of arable land is a serious limitation as well as their water needs, and importantly, many of them compete with human food necessities. Due to these disadvantages, other strategies are being considered. One of the most promising is the use of microalgae. A key feature of these microorganisms is that their culture do not compete with land otherwise used for feed crops, having the alternative to be grown on fresh, salt or even waste waters [Wang 2010]; In addition, they have the ability to be either phototrophic (use of light as a source of energy and CO_2 as a carbon source), heterotrophic (organic compounds as source of energy and carbon) or mixotrophic (a combination of both). Besides, the accumulation of their biomass can be achieved more efficiently than with ordinary crops [Cristi 2007].

A remarkable characteristic of these photosynthetic microorganisms is their capacity to be used for the production of different kinds of fuels, such as ethanol, hydrogen and methane [Wijifels 2010]. All these benefits make microalgae an attractive option for the production of energy. On the 1970's a renewed interest on microalgae emerged owed to an energy crisis at that time. As a result, the US Department of Energy carried out an algae study for several years which concluded with a collection of over 3,000 specimens [Sheehan *et al.* 1998]. Several green microalgae from this and other collections have been studied as possible source of oil for biodiesel production,

including *Chlamydomonas reinhardtii*, *Dunaliella tertiolecta* and *Chlorella protothecoides*, among others [Chen *et al.* 2011].

The genus *Chlorella* has been extensively used as an experimental organism. It has been employed as a model organism to study the heterotrophic growth and its impact on lipid profiles, where the CO_2 fixation rates of *Chlorella* sp. were measured [Chen and Johns 1991], as well as the efficiency of light transmission when grown on photobioreactors [Hirata *et al.* 1995]. The green microalga *Chlorella saccharophila* has been previously employed on a diverse range of studies. Maruyama *et al.* [1994] applied the electroporation technique to transform *C. saccharophila* with plasmid pBI221 carrying the gene GUS, which was used as a reporter and determined the optimum conditions to introduce foreign genes into *C. saccharophila*. Jonsson *et al.* [2001] examined *C. saccharophila* as a biomarker for the insecticide pyridaphention in order to stablish a toxicity database. Recently, *C. saccharophila* was proposed as a suitable source of oil for biodiesel production due to its high biomass productivity and levels of TAGs [Herrera-Valencia *et al.* 2011]. The study by Herrera-Valencia *et al.* [2011] was the first one to consider *C. saccharophila* as a biofuel producer.

Nevertheless, several limitations need to be overcome in order to scale-up the microalgae culture to an industrial level; such as the need to develop economically feasible methods to harvest cells, the difficult and variable conditions on open-pond environments, light penetrance, the constant threat of invasive species, and importantly, an efficient production of Triacilglycerides (TAGs), which are the raw material for biodiesel production. There is no doubt that vegetable crops have been of vital importance for the development of civilizations. Human civilizations have had the time to modify the characteristics of those organisms for their own benefit. The same strategy is necessary to be applied to microalgae, but using genetic engineering tools to accelerate the process [Radakovits *et al.* 2010].

Understanding and modifying the metabolic route for the production of TAGs on plants and microalgae have been of special interest on recent years, because of their importance in the production of biodiesel [Huang *et al.* 2009; Durrett *et al.* 2008]. In this regard, the last step in biosynthesis of TAGs is catalyzed by the enzyme Diacylglycerol acyltransferase (DGAT). Genetic modifications of this enzyme have proved to improve levels of TAGs on both plants [Lung and Weselake 2006; Lardizabal *et al.*, 2008] and microalgae [Wagner *et al.* 2010].

The aim of the present study was to isolate and characterize one *DGAT* homologue gene from the microalga *C. saccharophila*, and to clone it in an expression cassette under a constitutive strong promoter for further functional analysis to determine its role on TAG synthesis in this microalga.

CHAPTER I

1.1 BACKGROUND

1.1.1 *Chlorella saccharophila*: morphology and taxonomy

Members of the genus *Chlorella* belong to the green algae, found as single organisms or clusters in salt or fresh water, soil and snow. They are spherically-shaped and have one large chloroplast. These organisms have been proposed and used as a food source, because of its nutrient composition and its ability to produce large biomass, since one *Chlorella* individual can divide itself four times every 20 hours [Encyclopedia Britannica 2011; Moronta 2006]. The morphology and taxonomy of *C. saccharophila* is shown in Figure 1.1.

Eukaryota
Kingdom Viridiplantae
Phylum Chlorophyta
Class Trebouxiophyceae
Order Chlorellales
Family Chlorellaceae
Genus *Chlorella*
Species *Chlorella saccharophila*

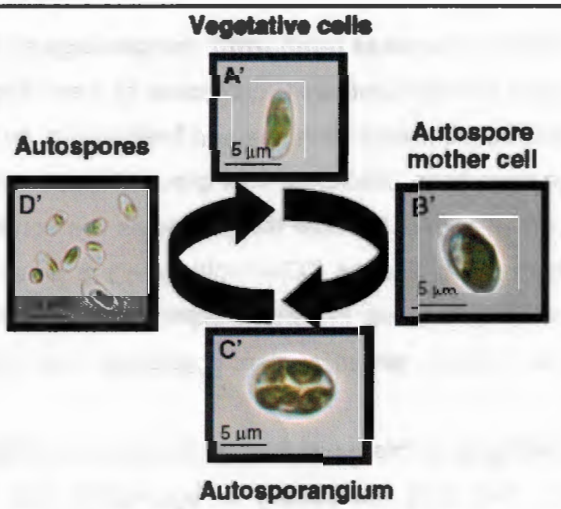


Figure 1.1 *Chlorella saccharophila* taxonomy, morphology and cell cycle [Zuppini *et al.* 2010]. This alga has different morphologic characteristics depending on the cycle stage. Image by Zuppini [2010]

1.1.2 *C. saccharophila* scientific studies

Although the genus *Chlorella* has been subject of numerous studies, the particular species *C. saccarophila* is not quite well understood as others. *Chlorella vulgaris* for example, has its chloroplast genome already sequenced, and the complete genome of *Chlorella variabilis* is currently being assembled [Wakasugi 1997; Entrez Genome Database 2011].

In one study, the fatty acid composition of *C. saccharophila* was analyzed when grown on different conditions. It was observed that when using glucose as the sole carbon

source, the lipid content increased compared to the photosynthetic growth. It was noticed that the concentration of this sugar also altered the lipid profiles, having the highest one with the minimum concentration of glucose [Tan 1991].

On another research, this microalga was used to understand the Programmed Cell Death on unicellular organisms. It was observed that heat stress induced a PCD pathway involving characteristic features such as DNA fragmentation, cell shrinkage and detachment of the plasma membrane from the cell wall [Zuppini 2007].

To date there is only one report that correlates *C. saccharophila* with biofuel production [Herrera-Valencia *et al.* 2011]. Using nitrogen deprivation and salinity treatments it was possible to increase the yield of fatty acid methyl esters (FAME) without a negative impact on biomass productivity, and therefore it was suggested an increase in accumulation of TAGs compared to the normal growth in this microalga.

1.1.3 *Chlamydomonas reinhardtii*: morphology and taxonomy

The genus *Chlamydomonas* is composed of green flagellates. These organisms can be encountered worldwide from sea and freshwaters, soil and snow where more than 500 species have been described. This group of algae was considered as the most ancient of the green plants, but now these species are considered non-ancestral members of the "chlorophyte" lineage ("Chlorophyceae") of green algae. Several *Chlamydomonas* species have become important experimental organisms in fields such as cell and molecular biology, genetics, plant physiology, and biotechnology [Harris 2009; O'Kelly 2004]

Cells belonging to this genus can be encountered with their cell wall on two states: thin or thick. The cells are solitary or aggregated into randomly-organized "palmelloid" colonies [Harris 2009; O'Kelly 2004].

On general basis, the photosynthetic mechanism is very similar in plants and algae, but is more efficient on the late ones because of their unicellular condition. This feature also gives them a better CO_2 sequestration and water acquisition. Due to these reasons the oil production on algae is considered much more efficient than on vegetable crops [Sheehan *et al.* 1998].

C. reinhardtii is a unicellular green alga. Ancient relative of land plants, but its lineage diverged from them over 1 billion years ago. These cells are haploid, and have the ability to grow on a simple medium of inorganic salts, having its energy demands provided by photosynthesis. During nitrogen deprivation, haploid cells of opposite mating types can fuse to form a diploid zygospore which forms a hard outer wall that protects it from adverse environmental conditions. Once the conditions are restored to

their status quo, the diploid zygote undergoes meiosis and releases four haploid cells that resume the vegetative life cycle [Harris 2001].

The reason that makes this species the predominant laboratory member of the *Chlamydomonas* genus relies on its ability to grow heterotrophically with acetate as its sole carbon source [Harris 2001]. The morphology and taxonomy of *C. reinhardtii* is shown in Figure 1.2.

Eukaryota
Kingdom Viridiplantae
Phylum Chlorophyta
Class Chlorophyceae
Order Chlamydomonadales
Family Chlamydomonadaceae
Genus *Chlamydomonas*
Species *Chlamydomonas reinhardtii*

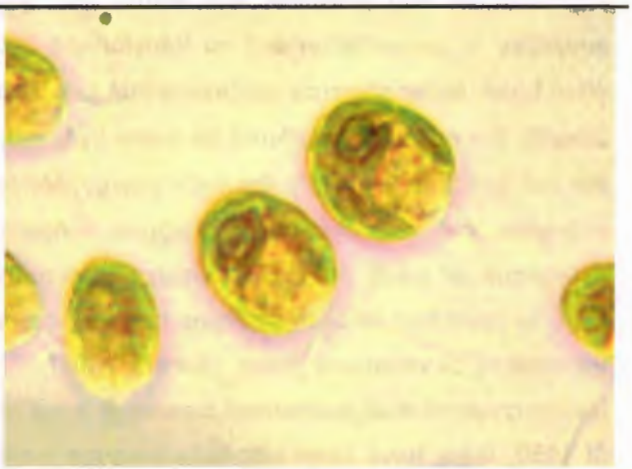


Figure 1.2 *C. reinhardtii* taxonomy (O'Kelly 2004) and morphology. Photo by Rodrigo Rivera

1.1.4 *C. reinhardtii* scientific studies

C. reinhardtii has numerous features that make it an ideal study model, it is a subject to tetrad analyses, its three genomes (nuclear, mitochondrial and chloroplastic) are amenable to specific transformation, and it can be grown photoautotrophically, heterotrophically or mixotrophically (which is a combination of the former) thus making the machinery for photosynthetic activity unessential for cell viability. Therefore, it can be studied on two ways: photosynthesis-defective mutants which may be grown heterotrophically, and respiratory-deficient mutants that need to perform photosynthesis. *C. reinhardtii* also has become a model system in the study of chloroplast biogenesis, structure and assembly of flagella and interactions between chloroplasts and mitochondria in unicellular organisms [Funes *et al.* 2007].

C. reinhardtii is an ideal model system for mitochondria of unicellular photosynthetic organism studies because it allows to combine biochemical approaches with genetical analysis of mutants. It can be used to elucidate the non-orthodox respiratory components of photosynthetic organisms, including several alternate nicotinamide

adenine dinucleotide (NADH) dehydrogenases and at least two alternative terminal oxidases [Funes *et al.* 2007].

1.1.5 Microalgae as a lipid source for biofuels production

A biofuel can be defined as a material used as fuel composed of or made from biological matter. The first generation of biofuels comprises those synthesized using vegetable oil and products, like starch and sugars. These biomolecules can be subjected to fermentation and so transformed into bioethanol. Lipids and fat on the other hand, suffer chemical processes that turn them into biodiesel [Biofuel 2010].

Despite the advantages offered by these systems, they have not been (and certainly will not be) able to satisfy the high energy demand due to serious limitations, both economic and environmental. The biggest concern lies on the competence that exists for agricultural lands; in fact a diminishment of natural lands with the objective of using them to plant biofuel-oriented crops has already ocurred, as well as an increment on the cooking oil varieties's prices. [Schenk 2008]

Research on microalgae-derived bioenergy is not a recent matter. Since the beginnings of 1950, there have been efforts to produce methane gas from algae. This process tried to use waste water as both medium and nutrient source for these organisms. During 1970, the US Department of Energy (DOE) and its coworkers financed these studies for the waste water management and energy production. It was a way to cope with two issues: environmental and energetic [Sheehan *et al.* 1998]

As previously stated, the molecules needed to make biofuels are sugars (Starch) and lipids. When it comes to the second ones, the primarly used are the Triglycerides (TAGs). Microalgae are able to produce these kinds of lipids, each species has a different lipid profiles and it depends on environmental conditions how would these profiles perform. Different strategies can be carried on in order to modify such profiles [Sheehan *et al.* 1998]. TAGs consist of a glycerol joined to three long-chain fatty acids; which must be treated with an energy-demanding process called transesterification to produce Fatty Acids Methyl/Ethyl Esters (FAME or FAEE respectively): Here, using temperature and NaOH as catalyst, the fatty acids are split from the glycerol and rejoined to an acyl-acceptor (ethanol or methanol) to obtain biodiesel [Korus *et al.* 2001]. Table 1.1 shows some of the species of microalgae employed on studies focused on production of TAGs or on total lipid productivity.

Table 1.1 Some species of microalgae with high TAGs or total lipid productivities

Microalgae	TAGs or total lipid productivity	Reference
Chlorella species		
C. saccharophila	Lipid content of 36.5% when using 2.5 g/L glucose	Tan *et al.* 1991
	TAGs production of ~70% dry weight (dw) on Nitrogen deficiency (ND)	Herrera-Valencia *et al.* 2011
C. vulgaris	Effect of carbon dioxide and sodium nitrate concentrations on biomass, lipid and cellulose content.	Aguirre *et al.*, 2013.
Neochloris oleobundans	Lipid productivity of 3.8 g/day. TAGs reached 37% dw under ND	Pruvost *et al.*, 2009
Scenedesmus sp.	Oil concentration of 15.12% when adding 10 g/L glycerol, under ND	Makareviciene *et al.*, 2012

1.1.6 Lipids, fatty acids and triglyceride biosynthesis on microalgae

The lipid pathways on microalgae are of great interest for biofuel production, because its raw material is the result of those pathways. Both synthesis and catabolism of lipids are not well studied on microalgae as they are on higher plants, but it has been suggested that they could be similar [Radakovits *et al.* 2010]. Figure 1.3 shows a schematic representation of enzymes and molecules that participate on fatty acids and biosynthesis of TAGs [Radakovits *et al.* 2010].

The lipid biosynthesis on higher plants has been extensively studied, and so the enzymes that participate are well characterized and the knowledge regarding the whole metabolism is quite extensive. Since *C. reinhardtii*'s genome has been sequenced, a number of Expressed Sequence Tags (EST's), and genetic tools are available, it has been possible to accurately predict some functions of the enzymes involved in lipid biosynthesis.

One difference on lipid biosynthesis between plant and microalgae relies on their organization. Plants, being pluricellular organisms; have tissue-specific regulations of their metabolism. Microalgae on the other hand, store TAGs as lipid droplets on the whole cell (primarily cytoplasm) as high energy storage specially when faced to unfavorable growth conditions [Moellering and Benning 2010]. Moellering and Benning [2010] discovered a unique protein to green microalgae identified on such lipid bodies

and called it "Major Lipid Droplet Protein". A repression of that protein with RNA interference led to increased lipid droplet size, but no change in TAGs metabolism or TAGs content.

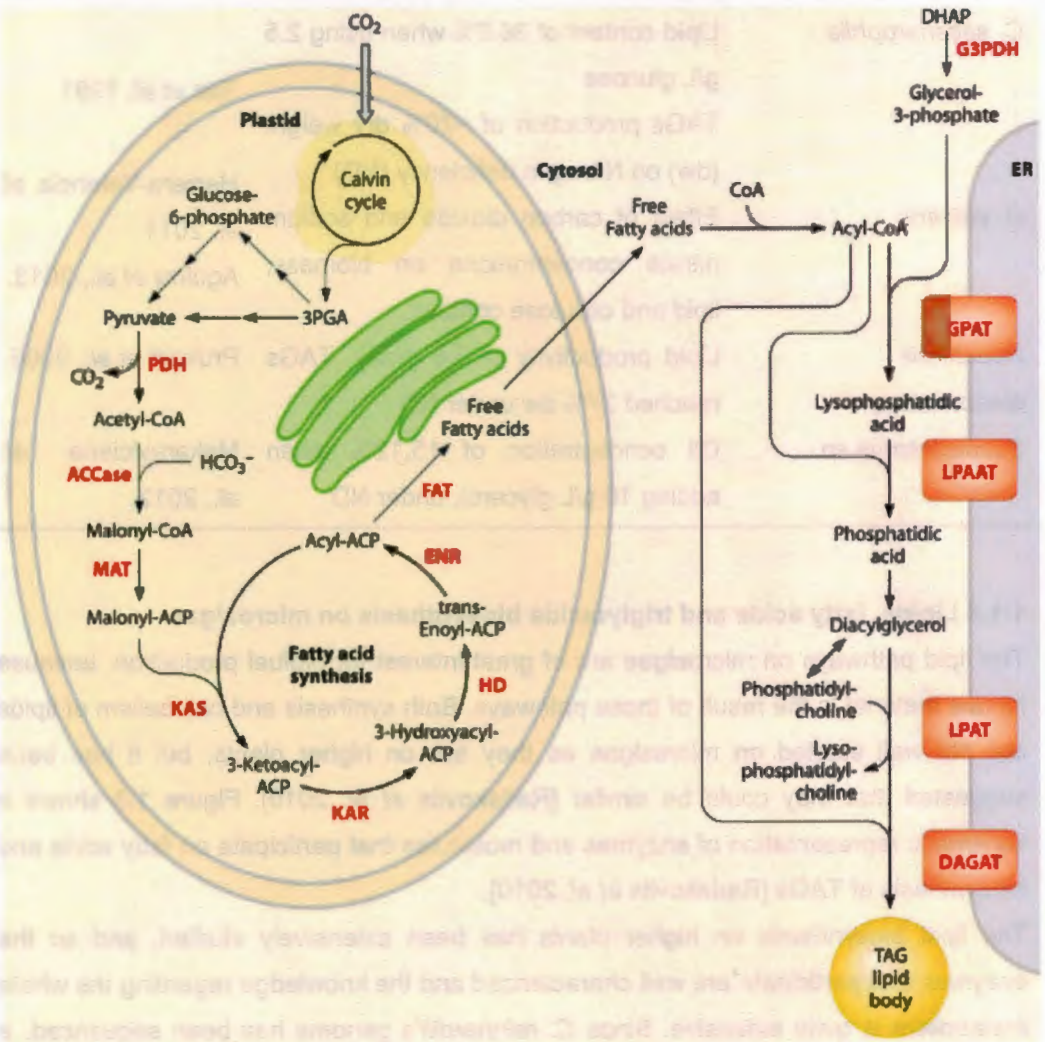


Figure 1.3 Simplified overview of the metabolites and representative pathways in plant lipid biosynthesis. ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; CoA, coenzyme A; DGAT, diacylglycerol acyltransferase; DHAP, dihydroxyacetone phosphate; ENR, enoyl-ACP reductase; FAT, fatty acyl-ACP thioesterase; G3PDH, gycerol-3-phosphate dehydrogenase; GPAT, glycerol-3-phosphate acyltransferase; HD, 3-hydroxyacyl-ACP dehydratase; KAR, 3-ketoacyl-ACP reductase; KAS, 3-ketoacyl-ACP synthase; LPAAT, lyso-phosphatidic acid acyltransferase; LPAT, lyso-phosphatidylcholine acyltransferase; MAT, malonyl-CoA:ACP transacylase; PDH, pyruvate dehydrogenase complex; TAG, triacylglycerols. Image from Radakovits *et al.* [2010]

The lipid composition of the microalga *C. reinhardtiii* was dilucidated by means of matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). It was revealed to be very similar as higher plants, but different compared to the diatom *Cyclotella meneghiniana* in lipid classes and acyl compositions [Vieler *et al.* 2007].

Recently, three acyltransferases implicated on nitrogen starvation-induced TAGs accumulation on *C. reinhardtii* were characterized [Boyle *et al.* 2012]. These acyltransferases consisted on Diacylglycerol acyltransferase (annotated as DGAT and DGTT on the genome of *C. reinhardtii*, and a phospholipid:diacylglycerol acyltransferase, PDAT). It was revealed that all acyltransferases were up-regulated during nitrogen starvation, but gene *DGTT1* was the more sensitive to this condition. Additionally, it was found that *C. reinhardtii* possesses a pathway for the biosynthesis of TAGs on the chloroplast [Fan *et al.* 2011]. This pathway is believed to rely on the synthesis of fatty acids, and TAGs are stored on both the chloroplast and the cytosol.

There is still large information about lipid synthesis on microalgae that need to be completed, as much of the information available still require functional essays to determine gene roles more accurately.

1.1.7 Enhanced lipid synthesis on higher plants

Several of the enzymes that participate on lipid biosynthesis have been subject of different studies with the objective of over-expressing them to increase lipid content. For instance, the seeds of *Arabidopsis thaliana* produce the largest quantity of TAGs, so this structure has been subject of different genetic modification approaches in order to understand and improve its production [Jako *et al.* 2001]. These authors accomplished an over expression of the enzyme Diacylglycerol acyltransferase (DGAT) on these seeds, therefore increasing the oil content from 10% to as high as 70%. Furthermore, it was possible to complement a DGAT-impaired activity on *A. thaliana* with transformation with its own DGAT cDNA sequence. In contrast, over-expressing an enzyme involved in the fatty acid pathway does not always increase lipid accumulation. For example the enzyme 2-Keto-Acyl Carrier Protein Synthase III (initiates the fatty acid synthesis in plants by catalyzing the condensing reaction of acetyl-CoA and malonyl-acyl carrier protein) from *Spinacia oleracea* (spinach) was expressed in *Nicotiana tabacum*, *B. napus* and *A. thaliana*, but it did not affect oil composition considerably [Dehesh 2001].

Another attempt was to increase the lipid content on *Solanum tuberosum* (potato), a normally starch-rich crop; by over-expressing an Acetyl-CoA carboxylase (this enzyme catalyzes the irreversible carboxylation of acetyl-CoA to produce malonyl-CoA through

its two catalytic activities, biotin carboxylase and carboxyltransferase. The most important function of ACC is to provide the malonyl-CoA substrate for the biosynthesis of fatty acids) from *Arabidopsis* on potato tubers amyloplasts. An increase in the fatty acid synthesis and TAGs production by 5-fold was possible using this approach. This demonstrated that ACCase may be playing an important role on lipid biosynthesis in tissues that normally do not use lipids as energy storage [Klaus *et al.* 2004]. However, when this approach was used on microalgae, it did not have the same results. Dunahay *et al.* [1995] managed to overexpress native ACCase in the diatom *C. cryptica*. Even though there was a 2- or 3-fold increase in ACCase activity, no increased lipid production could be observed.

Interestingly, while over-expression of genes related to fatty acid biosynthesis may not have a significant role increasing TAGs content, the over expression of genes involved in TAGs biosynthesis pathway has had promising results. For example, *Brassica napus* (rapeseed) over-expressing a yeast gene coding for glycerol-3-phosphate deshydrogenase (GPDH) had a three- to fourfold increase on the activity on this enzyme and the seed-oil content was also increased up to 40% [Vigeolas *et al.* 2007]. G3PDH reduces dihydroxyacetone 3 phosphate (DHAP) to glycerol-3-phosphate (G3P), which is a precursor for the glyceride synthesis. Transgenic expression of the enzyme DGAT has also had success increasing TAGs content. Lardizabal *et al.* [2008] introduced a fungal *Umbelopsis rumanniana* DGAT2A gene on *Glycine max* (soy) and expressed it during seed development. The absolute oil increase was 1.5% by weight on mature seed.

1.1.8 Enhanced lipid synthesis on microalgae

According to Hu *et al.* [2008], TAGs are primarily a source of energy on microalgae and does not have structural functions like other lipids. It is also suggested that biosynthesis of TAGs may respond to stress conditions. And so, most of the work done to date to increase the biomass/lipid content has been focused on modifying or altering microalgae culturing and harvesting conditions, as well as the effect on several stress factors. Table 1.2 shows some examples.

Table 1.2 Different approaches used in order to alter lipid content and growth rates in microalgae

Organism	Experiment	Results	Reference
C. vulgaris	CO_2 concentration Nitrogen starvation Harvesting method	With higher CO_2 the lipid production increased (it affects growth directly). This also happens with N starvation, but the growth was compromised. During lipid extraction, high temperatures resulted on a decreased TAGs content.	Widjaja *et al.* 2009.
Scenedesmus sp.	Phosphorus limitation Nitrogen limitation	Lipid accumulation reached 30% for N and 53% for P, but the productivity/unit volume of culture was not increased.	Xin *et al.* 2010.
Monodus subterraneus	Phosphate starvation	Very long chain polyunsaturated fatty acid biosynthesis decreased, and TAGs content increased from 6.5% to 39.3%.	Khozin – Goldberg and Cohen 2006.
Chlorella protothecoides	Glucose and glycerol as carbon sources	0.1% glucose increases oil accumulation by an order of magnitude. Glycerol enhances growth, making the process more efficient.	Sayre 2010.
Chlamydomonas globosa, Chlorella minutissima and *Scenedesmus bijuga*	Mixotrophic growth on wasterwaters and poultry litter extract (PLE) with modifications.	Waster waters and PLE supplemented with glucose and nitrogen resulted in 2-7 times increase in biomass.	Bhatnagar *et al.* 2011.

1.1.9 Diacylglycerol acyltransferase (DGAT)

Diacylglycerol acyltransferase (DGAT; E.C number 2.3.1.20) is the enzyme responsible for transfering an acyl group from Acyl-CoA to a diacilglycerol (DAG), forming a triacylglycerol (TAG) as a result, on a route known as the "Kennedy" pathway [Kennedy

1956]. It has been reported that this enzyme is committed in performing such activity. To date, three different DGAT families have been described (Table 1.3).

Table 1.3 Organisms where each DGAT family were initially discovered

	Discovered on	Localization	Reference
DGAT1	*Mus musculus*	ER	Cases *et al.*, 1998
DGAT2	*Mortierella ramanniana*	ER	Lardizabal *et al.*, 2001
DGAT3	*Arachis hypogaea*	Cytosol	Saha *et al.*, 2006

In one study carried out to prove the importance of DGAT on TAGs formation, Jako *et al.* [2001] performed a complementation with a wild-type *DGAT1* on *A. thaliana* that had been impaired by a sited directed mutagenesis. This mutation resulted on a seed oil reduction, a decreased in enzyme activity and a retarded seed development. When these mutants had their *DGAT* gene restored by transformation, these parameters returned to an activity comparable to the wild type. In addition, the authors observed an increase on TAGs formation when they performed this transformation on the wild type with an increase on the enzyme activity [Jako 2001].

Regarding DGAT2, a recent research by Banilas *et al.* [2011] revealed that DGAT2 on *Olea europaea* (olive) may be a key mediator of higher oil yields in ripening mesocarps, where oil droplets increase in size and TAGs are still accumulating. It was also observed that this enzyme was expressed in leaves and in vascular and tapetum cells of flowers.

The differences between these two DGATs depend on the organism where they are compared. For example, on *Vernicia fordii* (Tung tree) DGAT2 is induced in developing seeds at the time the oil biosynthesis begins [Shockey *et al.* 2006] and DGAT1 is expressed in several tissues at similar levels. On the other hand, the study performed with olive [Banilas *et al.* 2011] showed that DGAT1 contributes most of TAGs deposition in seeds.

In contrast to these two families of enzymes, which are bound to membranes, DGAT3 was discovered to be cytosolic. Also, when it was compared to its DGAT1 and 2 counterparts, DGAT3 showed an identity below 10%. This enzyme was proposed to be part of both biosynthesis of TAGs and wax ester formation [Saha 2006].

1.1.10 Structure and modeling of DGAT

As each of the different DGAT families was unraveled, studies focused on their structure have been developed. Different functional motifs have been elucidated for

each family. In the case of DGAT 1, a motif FYXDWWN was found to be implicated on fatty acyl-CoA binding. Other two important motifs for DGAT type 1 were found on *Tropaeolum majus.* For the structure of DGAT, a motif HKW-XX-RH-X-Y-X-P has been found to be the beginning of a highly hydrophobic transmembranal region [Yen *et al.*, 2008; Xu *et al.*, 2008]. Mutations found on either of these motifs have been correlated to a loss or diminishment of the enzyme function.

At least two fundamental motifs were described by Liu *et al.* [2011] on DGAT type. These motifs consist of HPHG and RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q). To find them, these authors compared several DGAT type 2 sequences belonging to fungi, animals and plants and searched for conserved regions, and so these two motifs were found. For the first one, it was found that the two initial aminoacids may change depending on the organism (EPHS for plants). This motif is present on the active site of DGAT type 2, and when mutated the functionality of the enzyme is abolished. The second motif RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q) is present on a transmembranal domain, thus having structural importance.

For DGAT type 2, there is avaible a 3d modeling of its protein, performed by Mischra *et al.* [2009]. They isolated a DGAT type 2 sequence of *M. ramanniana* from Gen Bank and compared it to *Cucurbita moschata*, a plant; and retrieved two sequences (named A and B). Using a software called MODELLER, they generated two models, one for each gene. Fig 1.4 shows these modelings:

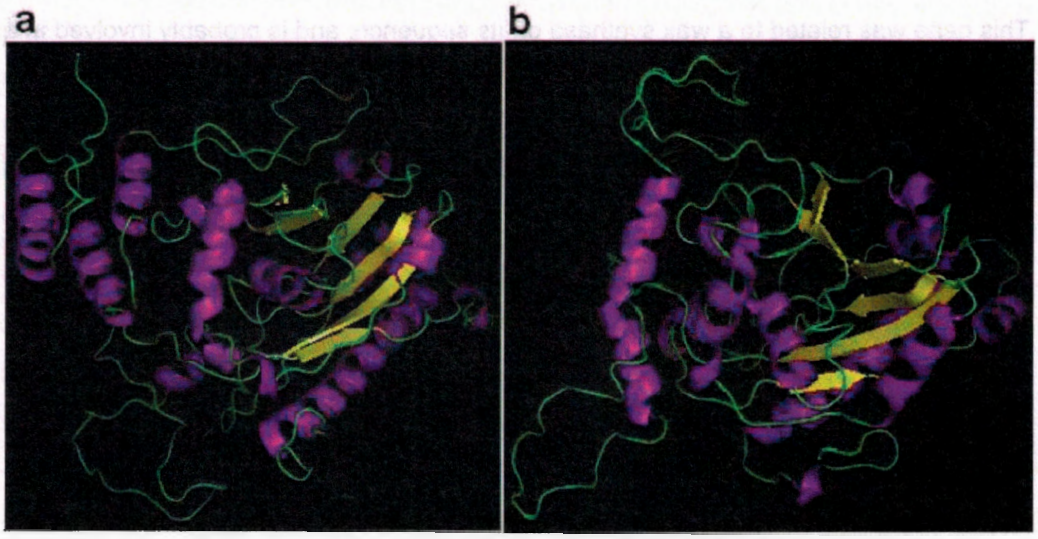


Figure 1.4. Model of *M. ramanniana* DGAT (a) type 2A and (b) type 2B. Image from Mischra *et al.* [2009]

It is stated that the modeling of DGAT type 2B is more accurate than 2A, because of available information on both structure and function. The authors claim that more

insights need to be done in order to achieve a better modeling of both DGAT type 2 versions for this and other fungal organisms.

The last family of DGAT to be discovered was the DGAT type 3, on peanut [Saha *et al.*, 2006]. Both DGAT 1 and 2 were found to have transmembrane domains, but DGAT type 3 was observed to be cytosolic. Fig 1.5 shows a hydrophathy plot with very low values indicating no transmembrane domains.

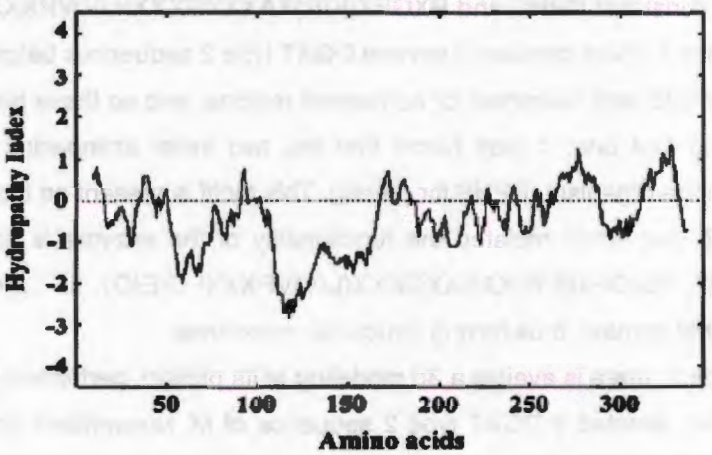


Figure 1.5. Hydropathy plot of cytosolic peanut DGAT. No transmembrane domain was predicted. Image from Saha *et al.* [2006]

This gene was related to a wax synthase on its sequence, and is probably involved in a pathway of both TAG and wax ester. DGAT type 3 homolog in *Arabidopsis thaliana* is expressed during early seed development. The authors state that this phenomenon occurs with the rest of the DGAT families, which is the temporal and spatial regulation of each one.

1.1.11 Phylogeny of Diacylglycerol acyltransferase

As mentioned before, three families of DGAT have been found and reported so far. Recently, a comprehensive analysis of the evolutionary history of DGAT 1 and 2 was performed [Turchetto *et al*, 2011]. Since information about DGAT 3 is still scarce, this family was not included. It was found that these two genes (DGAT 1 and 2) are almost ubiquitous in eukaryotes, but evolved separately and converged on their function. Figure 1.6 depicts a phylogenetic tree where two separate clades are formed, each one corresponding to a different DGAT family.

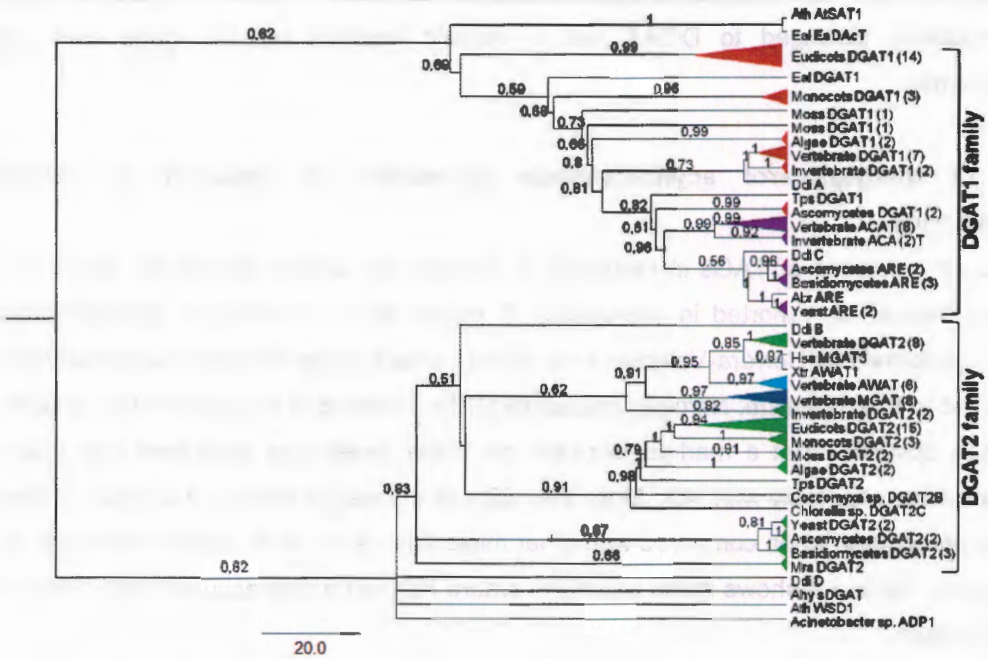


Figure 1.6. Phylogenetic relationships between DGAT1 and DGAT2 gene families reconstructed by the Bayesian method. Numbers within brackets correspond to the number of species within each group. *Alternaria brassicicola* (Abr), *Arabidopsis thaliana* (Ath), *Dictyostelium discoideum* (Ddi), *Euonymus alatus* (Eal), *Homo sapiens* (Hsa), *Mortierella ramanniana* (Mra), *Thalassiosira pseudonana* (Tps), *Xenopus tropicalis* (Xtr). Image from Turchetto *et al.* [2011].

1.1.12 Diacylglycerol acyltransferase on microalgae

To this day there is only one report involving expression of microalgae DGATs on heterologous systems [Wagner *et al.* 2010]. These authors used the three different DGAT sequences reported to date (DGAT1 from *A. thaliana*, DGAT2 from *M. rommaniana* and DGAT3 from *A. hypogaea*) as queries, and found three possible DGAT sequences in *Ostreococcus tauri*. Based on structure and phylogeny analysis, it was concluded that the three alleles were from the DGAT2 family. Using those possible DGAT sequences, they were able to reestablish the TAGs biosynthesis on a DGAT-mutant *Saccharomyces cereviseae*, proving their functionality as DGAT.

Rismani-Yazdi *et al.* [2011] produced a comprehensive transcriptome database corresponding to *Dunaliella tertiolecta* when grown on osmotic and nitrogen conditions (and their counterparts also). After filtering and processing the data, they gave annotations to each one of their sequences which form part of both starch and TAGs biosynthesis.They could reconstruct the entire TAGs pathway and discovered that it was very similar to the one described for *C. reinhardtii* and several plants. One of the

annotations belonged to DGAT, but a specific analysis on that gene was not performed.

1.1.13 Diacylglycerol acyltransferase expression in response to stress conditions

The accumulation of TAGs in response to nutrient deprivation and media stress has been extensively reported in microalgae. A recent report involving *C. saccharophila* was performed by Herrera-Valencia *et al.* [2011], where three different treatments were applied to this organism: Nitrogen deficiency (ND), salinity and the combination of both. It was observed that a marked increase on TAGs levels was increased with these treatments, especially with ND. More than 60% of dry weight belonged to lipids, a very high percentage when compared with other microalgae such as *N. oleobundans* and *C. vulgaris.* Table 1.4 shows some examples where ND had a lipid accumulating effect in microalgae.

Table 1.4 Effect of ND on lipid metabolism on different microalgae

Species	Lipid profile change	Reference
C. reinhardtii	Total lipid increase	Dean *et al.* 2010
Nannochloropsis oculata	Lipid increase by15.3%	Converti *et al.* 2009
C. vulgaris	Lipid increase by 16.41%	Converti *et al.* 2009
Chlorella sp	Lipid productivity of 53.96 ± 0.63 mg/L d	Praveenkumar *et al.* 2012

A correlation between the increase on DGAT expression and ND was made by Guihéneuf *et al.* [2011]. These authors discovered two *DGAT* type 1 (named *PtDGAT1long* and *PtDGAT1short*, based on their length) on *Phaeodactylum tricornutum*, a marine diatom. This organism was subjected to nutrient deprivation for over the course of 13 days, and discovered that the expression of both genes was increased after the treatment. Nevertheless, the expression of the *PtDGAT1short* decreased after eight days, suggesting some kind of temporal regulation (Fig. 1.7).

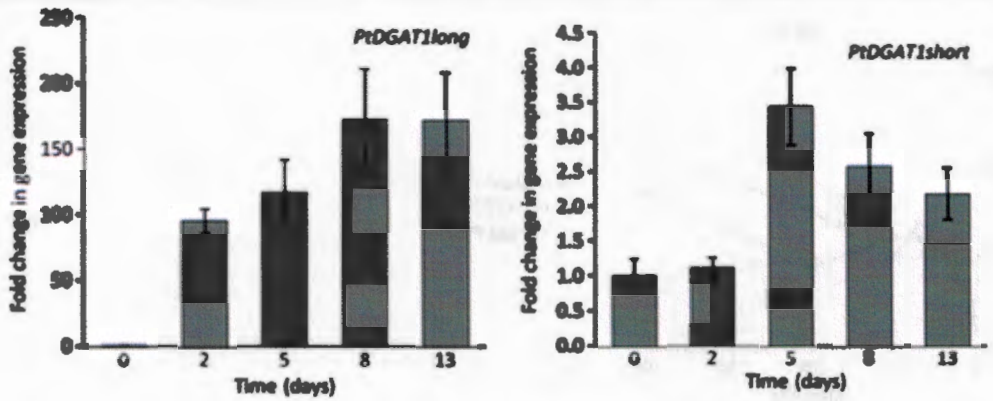
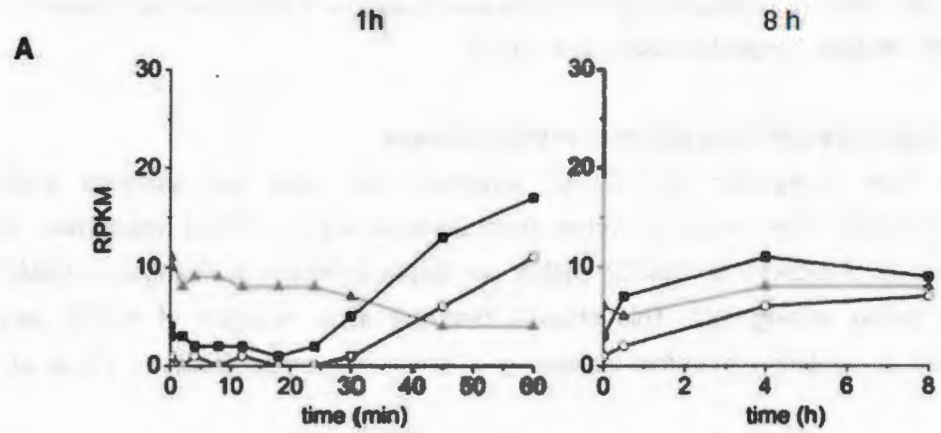


Figure 1.7 Expression of two DGATs from *P. tricornutum* over a 13 day time course. Image from Guihéneuf *et al.* [2011]

A study performed by Boyle *et al.* [2012] evaluated three different genes on *C. reinhardtii* under ND: *PDAT* coding for a Phospholipid Diacylglycerol Acyltransferase, and two genes coding for different Diacylglycerol acyltransferases, *DGAT1* coding for a DGAT type 1 and *DGTT* coding for a DGAT type 2. These authors performed a time course analysis of different sets, from 1 to 48 h. They observed the expression of both *DGAT1* and *DGTT1*, and even thought the accumulation of mRNA is not linear nor steady (it shows some decreases), overall the increase in expression is achieved along time for 48 hours (Fig. 1.8).



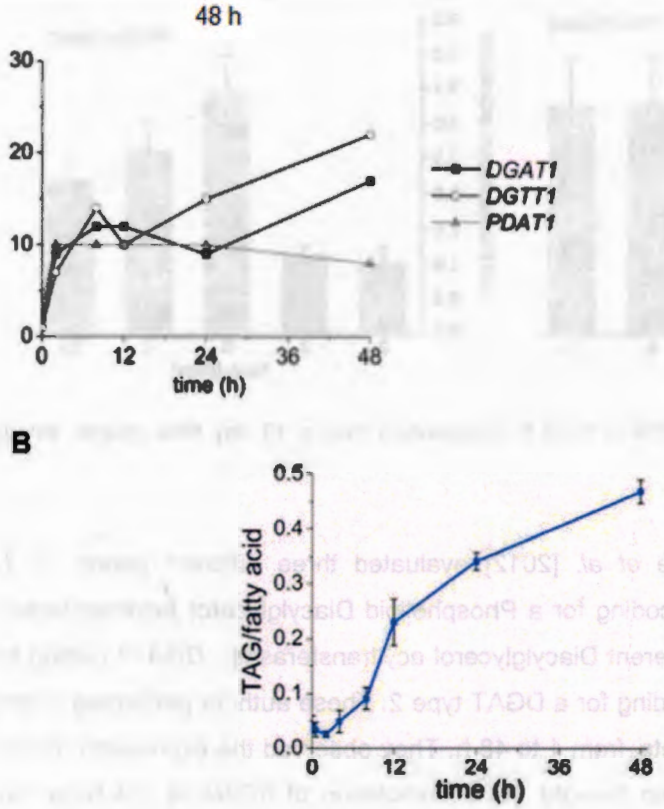


Figure 1.8 Measurement of the expression of three acyltransferases in response to ND. (A) Increased expression of DGAT1, DGTT1, and PDAT1 encoding three nitrogen-deficiency responsive acyltransferases. mRNA abundance reported in reads per kilobase per million reads (RPKM) in a 1 h timecourse, 8 h timecourse and 48 h timecourse after the onset of nitrogen starvation. (B) Timecourse of increased accumulation of TAGs in nitrogen starved *C. reinhardtii* wild type. Images from Boyle *et al.* [2012]

1.1.14 Regulation of Diacylglycerol acyltransferase

It has been suggested that DGAT enzymes may also be regulated post-transcriptionally. This evidence comes from observations in 3T3-L1 adipocytes, in which levels of DGAT1 or DGAT2 mRNA are disproportionate to changes in DGAT activity during adipogensis. This showed that the same quantity of mRNA was subjected to regulation, therefore increasing or diminishing enzyme activity. [Yu *et al.*, 2002].

Another case of post transcriptional regulation of DGAT was found on castor seeds. After pollination, *RcDGAT* transcript was measured and appeared after 12 days and reached the highest level 26 days later, and declined rapidly after that. However, the RcDGAT protein started to accumulate 26 days later, reached its peak at day 47, then remained at this high level until 54 days after pollination passed. The significant

difference between the expression of mRNA and protein indicates that gene expression of *RcDGAT* in maturing castor seeds is controlled at the posttranscriptional level [He *et al.*, 2004]. However, no deeper insights were done to measure or understand these mechanisms.

1.1.15 Patents on Diacylglycerol acyltransferase

Patents focused on DGAT are not only oriented on biofuel-related products or processes, but also they target the food industry. In one patent application, the DGAT gene derived from *Trepaoleum majus* (Garden Nasturtium) is claimed to be used for an enhanced production of TAGs in oilseed crops, as well as a method to mutate it and reduce such lipids in order to obtain healthier food [Taylor, 2008].

Interestingly, one patent application claims the use of DGAT (isolated from the fungus *M. rommanniana* or *S. cerevisiae*) for a variety of applications including human wellbeing. It comprises a diagnose assay to evaluate the levels of DGAT associated to altered conditions. This patent also claims, but is not limited to the use of this enzyme, parts of it or peptides derived from it in a variety of plant crops (peanut, corn, rapeseed, among others). It is important to point out that this patent belongs to Monsanto®, and the methods claimed here are very broad [Lardizabal *et al.* 2010].

There is one patent application [Sayre *et al.* 2010] that aims to protect an integrated system to enhance and economize biofuel production. Based on several chemical and genetic engineering strategies, these authors claim different approaches and their combination in order to alter the biofuel production. For instance, one of their claims is the use of *DGAT* and *ACCase* genes from *C. prototecoides* for this purpose.

1.2 JUSTIFICATION

According to genome projects and different studies, the pathway that leads to accumulation of TAGs must be similar in plants and microalgae, having the same major enzymes involved, such as DGAT. Furthermore, over expression of *DGAT* genes has been shown to increase the production of TAGs in plants and one microalga. Therefore, *DGAT* is a good candidate to study in order to increase the accumulation of TAGs in microalgae. Nevertheless, in order to alter the expression of this gene on a given microalga, it is necessary to first identify and characterize this type of genes on that microalga, and not many microalgae have had their genome sequence available. Recently, it was proposed that the microalga *C. saccharophila* could be considered as a suitable source of oil for biodiesel production. However, *DGAT* homologues have not been identified and characterized in *C. saccharophila*. The genetic manipulation of this key gene involved in lipid synthesis may lead to improve the accumulation of TAGs for biodiesel production in this microalga.

1.3 HYPOTHESIS

The green microalga *Chlorella saccharophila* has in its genome at least one Diacylglycerol acyltransferase (*DGAT*) homologue gene.

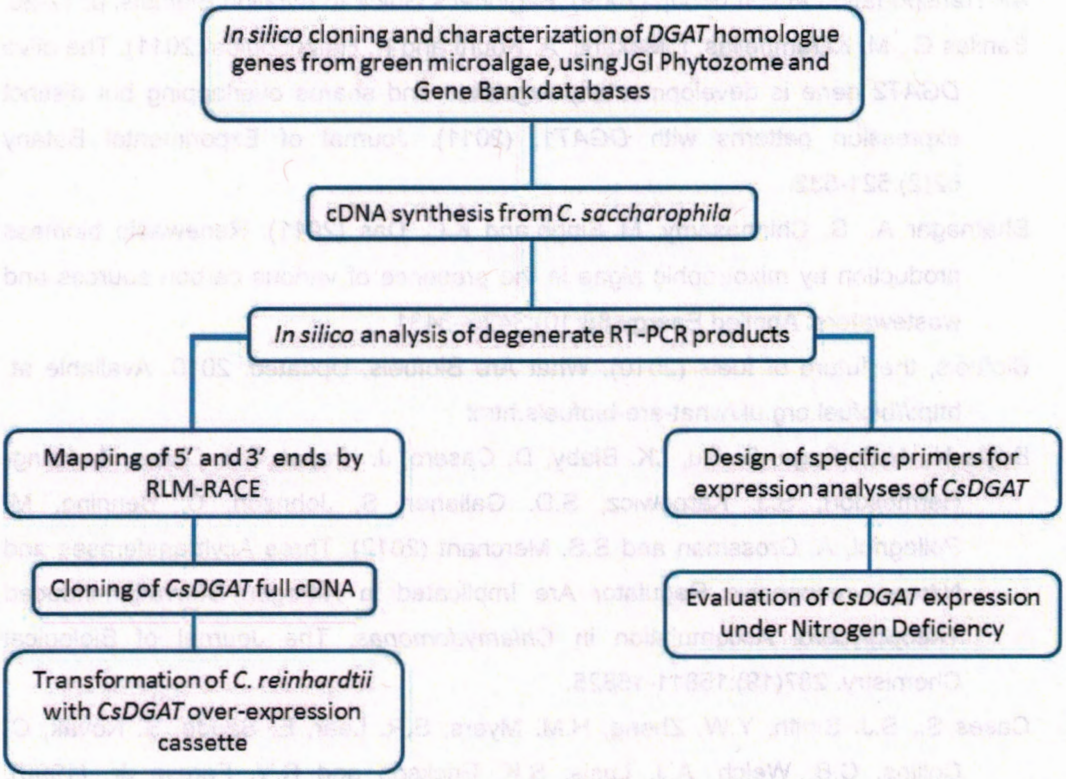
1.4 GENERAL OBJECTIVE

To isolate and characterize one *DGAT* homologue gene in *C. saccharophila.*

1.5 OBJECTIVES

1. To isolate and characterize the structure and phylogeny of one *DGAT* homologue gene from *C. saccharophila*.
2. To evaluate the expression of one *DGAT* homologue gene in response to nitrogen depletion in *C. saccharophila*.
3. To transform *C. reinhardtii* with a construction harboring the *CsDGAT* gene under the control of a strong constitutive promoter.

1.6 EXPERIMENTAL STRATEGY



1.7 REFERENCES

Aguirre A. and A. Bassi (2013). Investigation of biomass concentration, lipid production, and cellulose content in *Chlorella vulgaris* cultures using response surface methodology. Biotechnology Bioengineering. 110(8):2114–2122.

Air Transportation Action Group (2009). Beginner's Guide to Aviation Biofuels. p: 17-20

Banilas G., M. Karampelias, I. Makariti, A. Kourti and P. Hatzopoulos (2011). The olive *DGAT2* gene is developmentally regulated and shares overlapping but distinct expression patterns with *DGAT1*. (2011). Journal of Experimental Botany 62(2):521-532.

Bhatnagar A., S. Chinnasamy, M.,Singh,and K.C. Das (2011). Renewable biomass production by mixotrophic algae in the presence of various carbon sources and wastewaters. Applied Energy 88(10):3425–3431

Biofuels, the future of fuels (2010). What Are Biofuels. Updated: 2010. Available at: http://biofuel.org.uk/what-are-biofuels.html

Boyle N., M.D. Page, B. Liu, I.K. Blaby, D. Casero, J. Kropat, S.J. Cokus, A. Hong-Hermesdorf, S.J. Karpowicz, S.D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman and S.S. Merchant (2012). Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in *Chlamydomonas*. The Journal of Biological Chemistry. 287(19):15811-15825.

Cases S., S.J. Smith, Y.W. Zheng, H.M. Myers, S.R. Lear, E. Sande, S. Novak, C. Collins, C.B. Welch, A.J. Lusis, S.K. Erickson and R.V. Farese Jr. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences. 95(22):13018–13023.

Cases S., S. Novak, Y.W. Zheng, H.M. Myers, S.R. Lear, E. Sande, C.B. Welch, A.J. Lusis, T.A. Spencer, B.R. Krause, S.K. Erickson, R.V. Farese Jr. (1998). ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase, Its cloning, expression and characterization. The Journal of Biological Chemistry. 273(41):26755-26764.

Chen F. and Johns M.R. (1991). Effect of C/N ratio and aeration on the fatty acid composition of heterotrophic *Chlorella sorokiniana*. Journal of Applied Phycology 3(3):203-209.

Chisti Y. (2007). Biodiesel from microalgae. Biotechnol. Adv. 25:294–306.

Chlamydomonas Center: An Online Informatics Resource for *Chlamydomonas*. http://www.chlamy.org/. Updated: January 6, 2011

Contreras-Pool P. (2010). Análisis del perfil de los ésteres metílicos obtenidos de dos microalgas verdes: *Chlamydomonas reinhardtii* y *Chlorella saccharophila*, y su variación en respuesta a condiciones de estrés. Tesis de Maestría. Centro de Investigación Científica de Yucatán.

Converti A., A.A. Casazza, E.Y. Ortiz, P. Perego and M. Del Borghi (2009). Effect of temperature and nitrogen concentration on the growth and lipid content of *Nannochloropsis oculata* and *Chlorella vulgaris* for biodiesel production. Chemical Engineering and Process. 48(6):1146-1151.

Dean A.P., D.C. Sigee, B. Estrada and J. K. Pittman (2010) Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. Bioresource Technology. 101(12):4499–4507.

Dehesh K., H. Tai, P. Edwards, J. Byrne and J. Jaworski (2001). Overexpression of 3-ketoacyl-acyl-carrier protein synthase IIIs in plants reduces the rate of lipid synthesis. Plant Physiology 125(2):1103-1114.

Durrett T.P., C. Benning and J. Ohlrogge (2008). Plant triacylglycerols as feedstocks for the production of biofuels. The Plant Journal, 54(4):593-607.

Dunahay T.G., E.E. Jarvis and P.G. Roessler (1995). Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. Journal of Phycology 31(6):1004-1011.

Encyclopedia Britanica. *Chlorella*. (2011). In Encyclopædia Britannica Online. Available at: http://www.britannica.com/EBchecked/topic/113518/Chlorella. Access 14 March, 2011.

Entrez Genome Project, National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi. Updated constantly. Access 20 March, 2011.

Fan J., C. Andre and X. Changcheng (2011). A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. FEBS Letters 585(12):1985–1991.

Funes S., F. Lar-Gunnar and D. González-Halphen (2007). *Chlamydomonas reinhardtii*: the model of choice to study mitochondria from unicellular photosynthetic organisms. Methods in Molecular Biology 372:137-149.

Greenwell H.C., L.M.L. Laurens, R.J. Shields, R.W. Lovitt and K.J. Flynn (2010). Placing microalgae on the biofuels priority list: a review of the technological challenges. Journal of the Royal Society Interface. 7(406):703-726.

Guiheneuf F., S. Leu, A. Zarka, I. Khozin-Goldberg, I. Khalilov and S. Boussiba (2011). Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol

acyltransferase 1-like gene (*PtDGAT1*) from the diatom *Phaeodactylum tricornutum*. The FEBS Journal. 278(19):3651-3666.

Harris E.H. (2001). *Chlamydomonas* as a model organism. Annu Rev Plant Physiol Plant Mol Biol 52:363-406.

Harris E.H. (2009). The Chlamydomonas sourcebook, 2nd edition, Introduction to *Chlamydomonas* and its laboratory use Vol.1, Academic Press, Elsevier

He X., G. Chen, J.T. Lin and T. McKeon (2004). Regulation of diacylglycerol acyltransferase in developing seeds of castor. Lipids 39(9):865-871.

Hirata S., M. Hayashitani, M. Taya and S. Tone (1996). Carbon Dioxide Fixation in Batch culture of *Chlorella sp*. Using a Photobioreactor with a Sunlight-Collection Device. Journal of Fermentation and Bioengineering. 81(5):470-472.

Hu Q., M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibertl and A. Darzins (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. The Plant Journal. 54:621–639.

Herrera-Valencia V.A., Y.P. Contreras-Pool, S.J. López-Adrián, S. Peraza-Echeverría and L.F. Barahona-Pérez (2011).The green microalga *Chlorella saccharophila* as a suitable source of oil for biodiesel production. Current Microbiology 63(2):151-157.

Jako C., A. Kumar, Y. Wei, J. Zou, D. Barton, E. Giblin, P Covello and D. Taylor (2001). Seed-Specific Over-Expression of an *Arabidopsis* cDNA Encoding a Diacylglycerol Acyltransferase Enhances Seed Oil Content and Seed Weight. Plant Physiology 126(2):861-874.

Jonsson C.M., L.C. Paraiba, M.T. Mendoza, C. Sabater and J.M. Carrasco (2001). Bioconcentration of the insecticide pyridaphenthion by the green algae *Chlorella saccharophila*. Chemosphere 43(3):321-325.

Kennedy E.P. and S.B. Weiss (1956). The function of cytidine coenzymes in the biosynthesis of phospholipids. Journal of Biological Chemistry. 222(1):193-214.

Khozin-Goldberg I. and Z. Cohen (2010). Unraveling algal lipid metabolism: Recent advances in gene identification. Biochimie 93(1):91-100.

Klaus D., J.B. Ohlrogge, H.E. Neuhaus and P. Dormann (2004). Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. Planta 219(3):389-396.

Korus R., D. Hoffman, N. Bam, C. Peterson and D. Drown (2001). Transesterification Process to Manufacture Ethyl Ester of Rape Oil. Department of Chemical Engineering, University of Idaho, Moscow, Idaho, USA. Available at: http://www4.zetatalk.com/docs/Biodiesel/Rape-Seed-Oil-Conversion_2001.pdf Acces 20 August 2011

Lardizabal K., M. Jennifer, N. Wagner, A. Wyrick, T. Voelker and D. Hawkins (2001). DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. Journal of Biological Chemistry 276(42):38862–38869.

Lardizabal K., R. Effertz, C. Levering, J. Mai, M.C. Pedroso, T. Jury, E. Aasen, K. Gruys and K. Bennett (2008). Expression of *Umbelopsis ramanniana* DGAT2A in seed increases oil in soybean. Plant Physiology. 148(1):89–96.

Lardizabal K., G. Thompson and D. Hawkins (2010). Diacylglycerol Acyltransferase Proteins. WO 98/55631, WO 00/01713. Publication number US7741532 B2. Available at: https://www.google.com/patents/US7741532

Liu Q., R. Siloto, C. Snyder and R. Weselake (2011). Functional and Topological Analysis of Yeast Acyl-CoA:Diacylglycerol Acyltransferase 2, an Endoplasmic Reticulum Enzyme Essential for Triacylglycerol Biosynthesis. Journal of Biological Chemistry. 286(15):13115-13126.

Lung S.-C. and R. Weselake (2006). Diacylglycerol Acyltransferase: A Key Mediator of Plant Triacylglycerol Synthesis. Lipids, 41(12):1073-1089.

Makareviciene V., V. Skorupskaitė and V. Andrulevičiūtė (2012). Biomass and Oil Production of Green Microalgae *Scenedesmus sp.* Using Different Nutrients and Growth. Environmental Research, Engineering and Management. 62:4

Maruyama M., I. Horakova, H. Honda, X. Xing, N. Shiragami and H. Unno (1994). Introduction of foreign DNA into *Chlorella saccharophila* by electroporation. Biotechnology Techniques 8:821-826.

Mishra S., S. Dwivedi, N. Dwivedi, A. Kumar, A. Rawat and Y. Kamisaka (2009). A molecular model for diacylglycerol acyltransferase from *Mortierella ramanniana var. angulispora*. Bioinformation 3(9):394-398.

Moellering E. and C. Benning (2010). RNA Interference Silencing of a Major Lipid Droplet Protein Affects Lipid Droplet Size in *Chlamydomonas reinhardtii*. Eukaryotic Cell. 9(1):97-106.

Moronta R., R. Mora and E. Morales (2006). Respuesta de la microalga *Chlorella sorokiana* al pH, salinidad y temperatura en condiciones axénicas y no axenicas. Revista de Facultad de Agronomía. 23:28-43.

O'Kelly C. (2004). Protist Image Database. *Chlamydomonas*. (Updated March 30 2004). Available: http://megasun.bch.umontreal.ca/protists/chlamy/summary.html

Praveenkumar R., K. Shameera, G. Mahalakshmi, M.A. Akbarsha and N. Thajuddin (2012). Influence of nutrient deprivations on lipid accumulation in a dominant

indigenous microalga *Chlorella sp.*, bum11008: Evaluation for biodiesel production. Biomass Bioenerg. 37:60–66.

Pruvost J., G. Van Vooren, G. Cogne and J. Legrand (2009). Investigation of biomass and lipids production with *Neochloris oleoabundans* in photobioreactor. Bioresource Technology 100(23):5988-5995.

Radakovits R., R. Jinkerson, A. Darzins and M. Posewitz (2010). Genetic Engineering of Algae for Enhanced Biofuel Production. Eukaryotic Cell 9(4):486-501.

Rismani-Yazdi H., B. Haznedaroglu, K. Bibby and J. Peccia (2011). Transcriptome sequencing and annotation of the microalga *Dunaliella tertiolecta*: Pathway description and gene discovery for production of next-generation biofuels. BMC Genomics 12(148):1-17

Saha S., B. Enugutti, S. Rajakumari and R. Rajasekharan (2006). Cytosolic Triacylglycerol Biosynthetic Pathway in Oilseeds. Molecular Cloning and Expression of Peanut. Plant Physiology, 141(14):1533-1543.

Sayre R. and S. Pereira (2010). Molecular Approaches for the Optimization of Biofuel Production. United States Patent Application Publication. Appl. No. 12/743, 434. Pub. No. US 2010/0317073 A1

Schenk P.M., S. Thomas-Hall, E. Stephens, U. Marx, J. Mussgnug, C. Posten, O. Kruse and B. Hankamer (2008). Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. BioEnergy Research 1:20-43.

Sheehan J., T. Dunahay, J. Benemann, P. Roessler (1998). A Look Back at the U.S Department of Energy's Aquatic species. Golden, Colorado. National Renewable Energy Laboratory

Shockey J., S. Gidda, D. Chapital, J.C. Kuan, P.K. Dhanoa, J.M. Bland, S.J. Rothstein, R.T. Mullen and J.M. Dyera (2006). Tung Tree *DGAT1* and *DGAT2* Have Nonredundant Functions in Triacylglycerol Biosynthesis and Are Localized to Different Subdomains of the Endoplasmic Reticulum. The Plant Cell 18(9):2294-2313.

Tan C. and M. Johns (1991). Fatty acid production by heterotrophic *Chlorella saccharophila*. Hydrobiologia 215(1):13-19.

Taylor D., J. Xu, E. Mietkiewska and T. Francis (2008). Acyl-CoA-Dependent Diacylglycerol Acyltransferase1 (*DGAT1*) gene from *Tropaeolum Majus*, Protein Encoded thereby and uses thereof. WO/2008/006207. Available at: http://www.google.com/patents/WO2008006207A3?cl=en

Vieler A., C. Wilhelm, R. Goss, R. Süss and J. Schiller (2007). The lipid composition of the unicellular green alga *Chlamydomonas reinhardtii* and the diatom *Cyclotella*

meneghiniana investigated by MALDI-TOF MS and TLC. Chemistry and Physics of Lipids 150(2):143-155.

Vigeolas H., P. Waldeck, T. Zank and P. Geigenberger (2007). Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seedspecific promoter. Plant Biotechnol. J. 5(3):431–441.

Wagner M., K. Hoppe, T. Czabany, M. Heilmann, G. Daumb, I. Feussner and M. Fulda (2010). Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (*DGAT2*) gene from the microalga *O. tauri*. Plant Physiology and Biochemistry 48(6):407-416.

Wakasugi T., T. Nagai, M. Kapoor, M. Sugita, M. Ito, S. Ito, J. Tsudzuki, K. Nakashima, T. Tsudzuki, T. Suzuki, A. Hamada, T. Ohta, A. Inamura, K. Yoshinaga and M. Sugiura (1997). Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. Proceedings of the National Academy of Sciences 94(11):5967-5972.

Wang L., M. Min, Y. Li, P. Chen, Y. Chen, Y. Liu, Y. Wang and R. Ruan (2010). Cultivation of Green Algae *Chlorella sp.* in Different Wastewaters from Municipal Wastewater Treatment Plant. Applied Biochemistry and Biotechnology. 162(4):1174-1186.

Widjaja A., C.C. Chien and Y.H. Ju (2009). Study of increasing lipid production from fresh water microalgae *Chlorella vulgaris*. Journal of the Taiwan Institute of Chemical Engineers 40:13-20.

Wijffels R. and M. Barbosa (2010). An Outlook on Microalgal Biofuels. Science 329(5993):796-799.

Xin L., H.Y. Hu, G. Ke and Y.X. Sun (2010). Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid accumulation of a freshwater microalga *Scenedesmus sp.* Bioresource Technology 101(14):5494-54500.

Xu J., T. Francis, E. Mietkiewska, E. Giblin, D. Barton, Y. Zhang, M. Zhang and D. Taylor (2008). Cloning and characterization of an acyl-CoA-dependent diacylglycerol acyltransferase 1(DGAT1) gene from *Tropaeolum majus*, and a study of the functional motifs of the DGAT protein using site-directed mutagenesis to modify enzyme activity and oil content. Plant Biotechnology Journal 6(8):799-818.

Yen, C.-L. E.S. Stone, S. Koliwad, C. Harris and R. Farese (2008). Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. Journal Of Lipid Research 49(11):2283-2300.

Yu Y.H., Y. Zhang, P. Oelkers, S.L. Sturley, D.J. Rader and H.N. Ginsberg (2002). Posttranscriptional control of the expression and function of diacylglycerol acyltransferase-1 in mouse adipocytes. Journal of Biological Chemistry. 277(52):50876–50884.

Zuppini A., C. Andreoli and B. Baldan. (2007). Heat stress: an inducer of programmed cell death in *Chlorella saccharophila*. Plant & Cell Physiology 48(7):1000-1009.

Zuppini A., C. Gerotto and B. Baldan (2010). Programmed Cell Death and Adaptation: Two Different Types of Abiotic Stress Response in a Unicellular Chlorophyte. Plant & Cell Physiology. 51(6):884–895.

CHAPTER II

CLONING AND MOLECULAR CHARACTERIZATION OF *CsDGAT*: A *DGAT* HOMOLOGUE GENE IN THE GREEN MICROALGA *Chlorella saccharophila*

2.1 INTRODUCTION

Diacylglycerol acyltransferase (DGAT) is a key enzyme on the TAGs biosynthetic pathway in plants. Initially found on mouse [Cases 1998], this enzyme is present in many eukaryots such as microalgae, fungi, plants and animals. Three DGAT types have been identified, named in order as they were discovered: DGAT1, DGAT2 and DGAT3. The two first types are transmembranal proteins and type 3 is cytosolic. Not all these enzymes are present ubiquitously or uniformly. For instance, yeast (e.g. *Candida albicans, Saccharomyces cereviseae*) and basidiomycetes (e.g. *Laccaria bicolor*) lack the presence of DGAT type 1 [Turchetto 2011]. Functional and bioinformatic-based analyses have shown that microalgae mostly have DGAT type 2 sequences. For example, Wagner *et al.* [2010] identified and characterized DGAT type 2 sequences on *Ostreococcus tauri*, and Turchetto *et al.* [2011] found only this type of sequences on the genomes of several microalgae such as *Chlorella spp* and *Ostreococcus lucimarinus*.

DGAT type 2 is mainly located at the endoplasmic reticulum (ER), where biosynthesis of TAGs takes place [Lung and Weselake 2006]. Because of this cellular localization and possible enzyme conformation, several putative transmembrane domains were analysed on *S. cereviseae* DGAT2 [Liu *et al.* 2011]. Based on topology predictions and functional analysis, it was determined that at least three motifs were conserved among the members of DGAT2: YFP, HPHG and RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E /Q). Multiple sequence alignments and phylogenetic analysis showed that these regions were conserved among diverse organisms. This finding provides a useful guide for analysis of DGAT homologues in microalgae.

Recently, there has been a renewed interest on microalgae because of their capacity to produce a variety of high-value products, the ability to remediate contaminated water and their potential to be used as a source of raw materials for biofuels [Greenwell *et al.* 2010].

Microalgae can be found in a wide variety of environmental aquatic conditions, from freshwater ponds to cold ecosystems. These organisms have thrived in such divergent conditions because their ability to adapt to them including nutrient deficiency conditions. Regarding their lipid metabolism, the amount of lipid content on microalgal biomass growing on optimal conditions ranges between 5 and 20% of their dry weight [Schenk *et al.* 2012]. Limiting the nutrients available on the media severely affects microalgal growth and hence the rate and production of its biomolecules. Interestingly, and in contrast to biomass accumulation, the synthesis of lipids is highly increased during nutrient limitation, as long as there is enough light and a carbon source provided [Schenk *et al.* 2012]. At the time when nutrient availability becomes limited and the cell growth (divisions) is stopped, microalgae shift lipid use otherwise destined for cell wall and organelles to production and accumulation of TAGs. Therefore, in general there is an inverse relationship between lipid content and nitrogen concentration [Gouveia and Oliveira 2009]. This trend of accumulation of TAGs under nitrogen starvation has been observed to cause lipid accumulation in different microalgae: *C. reinhardtii*, *Scenedesmus suspicatus* [Dean *et al.*, 2010], *Nannochloropsis occulata* [Converti *et al.* 2010], *C. vulgaris* and *Chlorella sp* [Yeh *et al.* 2011; Praveenkumar *et al.* 2012].

Recently, the green microalga *C. saccharophila* was proposed to have high potential as a source of TAGs for biodiesel [Herrera-Valencia *et al.* 2011]. Because of this promising finding, the need for a better understanding of this microalga has raisen, however, this microalga has not been widely studied for biofuel production purposes, and its genome has not been sequenced yet. In contrast, the complete genome sequence of some microalgae have become publicly available, such as the sequence from the model microalga *C. reinhardtii,* available at the Joint Genome Institute [Merchant *et al.* 2007], and *Chlorella variabilis* NC64A [Blanc *et al.* 2010], which provides a valuable working platform for bioinformatic analysis of genes related with lipid metabolism in these microorganisms.

Although it is known that TAGs represent highly energetic molecules that can be transesterified to biodiesel, their biosynthetic pathway on microalgae is not completely understood, and particularly in *C. saccharophila* there are not studies regarding TAGs metabolism. Therefore, the aim of this chapter was to isolate and characterize a DGAT homologue gene from *C. saccharophila*.

2.2 MATERIALS AND METHODS

2.2.1 Microalgal culture and cell counting

C. saccharophila (UADY-PRIORI-014-FMVZ-04 strain) was kindly donated by Q.F.B. Silvia J. López Adrián from the microalgae collection of the "Alfredo Barrera Marín" herbarium, Universidad Autónoma de Yucatán (Yucatan Autonomous University, UADY). Cells were kept on TAP medium on agitation at 140 rpm. Harvesting was carried out after 10 days of growth according to Contreras [2010]. Tris-Acetate-Phosphate medium [Harris 1989] was prepared as shown on the Appendix (Table A and B) and pH adjusted to 7 (using HCl and KOH 1M). Media was sterilized on autoclave at 1.5 atm for 20 minutes. Semisolid TAP was prepared by adding 15 g/L Merck agar-agar [Contreras 2010]. Cell number was estimated using a Hausser Scientific Hemocytometer following a method by Rouge [2002], and visualized using a Carl Zeiss Primo Star Optical Microscope. The hemocytometer had two chambers, each one with a microscopic grid etched on the surface. A glass coverslip was placed 0.1 mm above the chambers. To fix cells, 0.1 mL Lugol's iodine was added to 0.9 mL culture. This mixture was then loaded into the chamber (approximately 15 µL). Capillary action was enough to fill the space completely. Counting: The full grid on the hemocytometer contains nine squares, each of which is 1 mm^2. The central area (used to count) has 25 large squares, each one has 16 smaller ones.The squares used to count the cells were those located on each corner and the middle, as shown on figure 2.1.

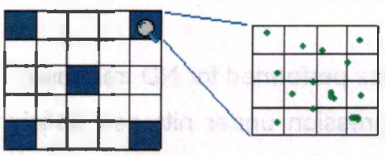


Figure 2.1 Squares where the cells were counted using a Hausser Scientific Hemocytometer

After performing this count twice, the mean was calculated. This value then was subjected to the next formulas, in order to obtain the number of cells per milliliter and total number of cells in a flask:

$$Mean \times 5 \times 10{,}000 = Cells/_{milliliter}$$

$$(Number\ of\ cells)\ X\ (milliliters\ in\ original\ culture) = Number\ of\ cells\ in\ original\ culture$$

To manipulate cells it was necessary to have them concentrated. To achieve this, the next formula was applied:

$$\frac{\textit{Number of cells in original culture}}{\textit{Millions cells required}} = \textit{Milliliters to concentrate cells}$$

With this value, the next step was to concentrate the culture: first it was transferred from the flask to a plastic tube and then centrifuged at 3,200 rpm for 5 minutes. The cell pellet was then resuspended on the medium left from the previous centrifugation. The now concentrated cells were distributed on Safe Lock® Tubes (Eppendorf) and centrifuged at 3,000 rcf for 5 minutes. The supernatant was then discarded and the remaining cell pellets were frozen on liquid Nitrogen (-196 °C) and stored at -80°C.

2.2.2 Nitrogen deficiency treatment

First, a single colony from solid culture was transferred to liquid TAP, and cells were grown until they reached the end of logarithmic phase (approximately 10 days), to obtain a starter culture. This was used as inoculum for the experimental samples, inoculating an initial concentration of 10,000 cells per milliliter on 250 mL flasks with 50 mL of TAP medium. These cultures were allowed to grow for 10 days and then the nitrogen deficiency (ND) treatment was applied. To apply this treatment, cells were harvested by centrifugation, washed once with TAP-N (TAP with limited nitrogen source, trace metals as the only source of nitrogen with 0.011 mg N per liter), and then transferred to 50 mL of fresh TAP-N. For the control, cells were washed with TAP and then transferred to fresh TAP.

A time based experiment was performed for ND treatment, in order to analyze temporal responses in *CsDGAT* expression under nitrogen deficiency. Five different samples were analyzed: a control consisting on cells at 10 d of culture on TAP before starting the experiment, and then samples collected at 6, 12, 24 and 48 hours after ND treatment was applied. As a control, *CsDGAT* expression was also analyzed at the same times described before but using TAP medium (no ND treatment). *CsActin* espression was also analyzed for the same samples described above as a positive control of all the experiments. Cells were collected on 2 mL tubes with 200 million cells per tube, frozen with liquid Nitrogen and stored at -80 °C until use for expression analysis. The method used to isolate RNA was CTAB and 5 µg of RNA was employed on all samples to synthesize cDNA, as it will be described later.

2.2.3 Nucleic acid extraction

DNA extraction was carried out using a protocol from the General Electric Healthcare Illustra Nucleon Phytopure Genomic DNA Extraction Kit: Cell lysis: 600 µL Reagent 1 and 10 µL mercaptoethanol were added, with all ingredients fully dissolved. The tube was mixed quickly by inverting it several times. 200 µL Reagent 2 were applied. The tube was inverted again several times until a homogeneous mixture was obtained. Next, the mixture was incubated at 65°C in a water bath for 10 min with regular manual agitation during the incubation. Completing the incubation, the sample was placed on ice for 20 min. After that 500 µL cold chloroform (-20°C) were added followed by 100 µL of Nucleon Phytopure resin suspension (which was vigorously resuspended on a vortex), mixed by inverting the tube several times and incubated at room temperature for 10 min with regular manual agitation. Then it was centrifuged at 6000 rpm for 10 min. Without disturbing the Nucleon resin suspension layer, the upper DNA-RNA containing phase (above the brown resin layer) was transferred into a fresh tube; spun at 12000 rpm/5 min and the upper phase (without disturbing the pellet) transferred into a new tube. An equal volume of isopropanol (room temperature) was added and gently inverted several times. To pellet the DNA-RNA it was spun at 12000 rpm for 5 minutes. The supernatant was discarded and the pellet (now visible) washed twice with 1 ml of room temperature 70% ethanol (12000 rpm/2 min each spin). The supernatant was removed and the pellet air-dried for 30 min at room temperature. DNA was resuspended on 50 µL buffer TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0), heated at 65°C for 10 minutes and then chilled on ice for 3 min. The sample was then gently pipetted up and down a few times and subjected to a spin (12000 rpm/ 1 min) to remove any remaining debris and finally transferred to a fresh tube. To remove RNA, RNase A was added and incubated at 37°C for 20 minutes. For long term, it was stored at -20°C.

DNA was also isolated using a protocol based on Dellaporta *et al.* [1983]: Sample was prepared by adding 650 µl TEN buffer + preheated (65°C) β-mercaptoethanol (0.014 volumes β-mercaptoethanol per 1 volume buffer TEN), vortexed 4 minutes and incubated at 65°C 10 minutes. 43 µl Sodium dodecyl sulfate (SDS) was added and vortexed, followed by a 65 °C incubation for 25 minutes with a vortex every 3 minutes. After samples reached room temperature, chloroform: isoamilic alcohol (24: 1) was added. Samples were then vigorously shaken and centrifuged at 14,000 rpm for 5 minutes; the supernatant was separated and this treatment repeated twice. An equal volume of isopropanol is added, gently mixed for 3 minutes and centrifuged at 13,500

rpm for 5 minutes. The remaining pellet was washed twice with cold 70% ethanol and centrifuged at 13,500 rpm for 5 minutes. Finally, after this pellet was air-dried (approximately 10 minutes), it was resuspended with 40 µl Tris HCl 10 mM pH 7.5 and incubated 10 minutes at 65°C.

RNA extraction was carried out using a CTAB Protocol [Kieffer 2000]: Samples were collected and stored on liquid Nitrogen for transportation. During the experiment they were kept on ice. CTAB reagent was heated to 65°C, and added to the frozen samples. The mixture was homogenized thoroughly with a pestle for one minute. Next, 24 µL β-mercaptoethanol were added. The mixture was homogenized with a vigorous vortex for another minute. Samples were then incubated at 65 °C for 10 minutes, applying a vortex every 2 min. Next, a chloroform-isoamilic alcohol (24:1) solution was added. In order to separate phases, the samples were centrifuged at 14000 rpm for 5 minutes. After recovering the supernatant, 600 µL of cold isopropanol and 100 µL of sodium acetate were incorporated. After that, the samples were incubated on ice for 15 minutes and centrifuged at 14000 rpm for 10 minutes; the supernatant was discarded, leaving a visible pellet at the bottom. 800 µL of cold 70% ethanol were added, followed by a 14000 rpm centrifugation for 5 minutes. The supernatant was discarded and the pellet was air-dried and resuspended on 39 µL ultrapure water. 10 µL DNAse (RQ1 RNase-free DNase 1 µg/ µL; Promega), 10 µl DNAse buffer (RQ1 DNase 10X Reaction Buffer; Promega) and 1 RNAse inhibitor (RNaseOUT 40U/ µL; Invitrogen) were added and incubated at 37°C for 30 minutes. To purify the sample, 600 µL SSTE + 600 µL Chloroform-isoamilic alcohol (24:1) and 100 µL sodium acetate 3M were added, the mixture was homogenized and centrifuged at 14000 rpm for 5 minutes. The supernatant was recovered on a fresh tube and 600 µL cold (-20°C) isopropanol were added, next the sample was placed on ice for 5 minues. Next, it was centrifuged at 14000 rpm for 10 minutes and the supernatant discarded, leaving a visible pellet that was washed with 800 µL 70% cold ethanol (-20°C) and centrifuged at 14000 rpm for 5 minutes. The supernatant was discarded and the pellet was air-dried. Finally, it was resuspended on 60 µL ultrapure water.

RNA was also isolated using a protocol from General Electric illustra RNAspin Mini RNA Isolation Kit: Cell lysis was done by adding 350 µL buffer RA1 and 3.5 µL ß-mercaptoethanol to the cell pellet. It was centrifuged for 1 minute at 11000x g on a RNAspin Mini Filter units to filtrate it. 350 µL 70% ethanol was added and centrifuged 30 seconds at 8000 g on a RNAspin Mini column. After adding 350 µL Membrane Desalting Buffer (MDB) and 11000x g for 1 minute centrifugation was applied. The

flow-through was discarded. To digest DNA, 95 µL DNase reaction mixture (reconstituted DNase I and DNase reaction buffer 1:10) was added. The reaction was incutabed at room temperature for 15 minutes. Three different washes were applied: the first one with 200 µL RA2 buffer centrifuged for 1 minute at 11000x g on an RNAspin Mini column. The second, on the same column (same recollection tube) with 600 µL buffer RA3 and centrifuged for 1 minute at 11000x g. The last wash was performed using 250 µL buffer RA3 and centrifuged for 2 minutes at 11000x g on the same RNAspin Mini column. RNA was eluted with 100 µL nuclease-free H_2O and centrifuged at 11000x g for 1 minute. Samples were kept at -80°C for long term storage.

2.2.4 Synthesis of cDNA and RT-PCR

For the synthesis of cDNA 10 µg total RNA was mixed with 1 µL oligo dT primer (50 µM; Invitrogen) in a final volume of 20 µL ultrapure water, then a a 75°C incubation for 5 minutes was used to linearize messenger RNAs and then the tube was chilled on ice for 3 minutes. Then, 8µ 5X First Strand Buffer, 4 µL Dithiothreitol 0.01M, 4 µL dNTP's 10 mM and 2 µL RNAse inhibitor 40 U/µL were added. The final volume of 38 µL were divided into two 19 µL tubes, which were then mixed with 1 µL Super Script III 200 U/µL reverse transcriptase onto one of the tubes, and 1 µL water to the second one (to function as a negative genomic DNA control). Reactions were incubated at 50°C for 90 minutes, inactivating the enzyme afterwards at 75°C for 15 minutes. After that, 4 µL RNAse A 1 mg/µL were added followed by a 37°C incubation for 30 minutes. Finally, the mixture was diluted 1:5 by adding 80 µL ultrapure water and conserved at -20°C.

This cDNA was employed for two different reactions: Degenerate PCRs and 3' mapping with RACE. On the degenerate PCRs, this cDNA was used directly as template for the degenerate primers (it will be explained later). For the RACE, this cDNA functions as template for PCR and Nested PCR using specific primers for the 3' end. cDNA using 5 µg of total RNA was employed for the expression analyses. The same amount of RNA was employed on all samples to normalize.

Also, cDNA was synthesized to map the 5' end,using a method called RLM-RACE (Invitrogen) in which, as described by the manufacturer: full-length mRNAs are selected by treating total or poly(A) RNA with calf intestinal phosphatase (CIP) to remove the 5'-phosphate from all molecules which contain free 5'-phosphates (ribosomal RNA, fragmented mRNA, tRNA, and contaminating genomic DNA). Full-length mRNAs are unaffected. The RNA is then treated with tobacco acid pyrophosphatase (TAP) to remove the cap structure from the full-length mRNA leaving a 5'-monophosphate. A synthetic RNA adapter is ligated to the RNA population - only

molecules containing a 5'-phosphate - the uncapped, full-length mRNAs - will accept the adapter. Then, these mRNAs are decapped and ligated to a RNA oligo at the 5'. The cDNA is then synthesized and used as template for PCR and nested PCR.

2.2.5 Bioinformatic resources

So far, NCBI reports 15 green alga genome sequence projects, and there are also numerous efforts and projects to complete plastid genomes. All these studies provide valuable tools for the genetic manipulation of these organisms, and opportunities to compare such data [Radakovits *et al.* 2010].

In order to obtain putative DGAT nucleic and amino acidic sequences, a revision of entries and a Basic Local Alignment Search Tool (BLAST) was performed on the Department of Energy (DOE) Joint Genome Institute (JGI) database (Phytozome). *A. thaliana* DGAT1 (Accesion AF051849), DGAT2 (NP_566952) and *Arachis hypogaea* (AAX62735) were used as queries.

Predicted protein sequences were aligned using ClustalX software [Thompson *et al.* 1997; Larkin *et al.* 2011] with default settings. Boxshade server was used to highlight identical and similar segments between sequences in these aligments. DNAStar Lasergene 7 version 7.2.1 [Burland 2000] package was used for several applications: EditSeq to manipulate and edit sequences, MegAlign to perform multiple alignments and phylogenetic trees and SeqMan to construct and assemble sequences. BioEdit Sequence Alignment Editor was used to translate DNA and schematize open reading frames [Hall 1999].

2.2.6 Searches of microalgae and plant *DGAT* genes in databases, multiple sequence alignment and phylogenetic analysis

In order to acquire a large number of putative DGAT sequences from databases, two different searches were performed. One was done manually by revising JGI and NCBI digital collection of genes, on the lipid metabolism section. Another search was achieved by means of BLAST using 3 different characterized DGAT sequences.

Using the sequences obtained from the search of DGAT genes previously described, a multiple sequence alignment was performed with the Clustal W method. This analysis was then employed to construct a phylogenetic tree by the method of Hein [1990].

Phylogenetic tree using the complete CsDGAT, DGATs from microalgae and characterized DGATs from plants was inferred using the Neighbor-Joining method [Saitou and Nei 1987]. The optimal tree with the sum of branch length = 8.01839193 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches

[Felsenstein 1985]. The evolutionary distances were computed using the p-distance method [Nei and Kumar 2000] and are in the units of the number of amino acid differences per site. The analysis involved 31 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 57 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [Tamura *et al.* 2007]

2.2.7 Degenerate PCR strategy for cloning DGAT sequences from *C. sacharophila*

All PCRs were performed on a BIORAD's C1000 Thermal Cycler. A set of degenerate primers were designed in order to amplify the most probable DGAT sequences from *C. saccharophila* nuclear genome. The following strategy was followed in order to design these degenerate primers: First, three different DGAT sequences (types 1, 2 from *A. thaliana* and type 3 from *Arachis hypogaea*) already characterized were used as queries to find and isolate DGAT sequences from green microalgae in the bioinformatics databases Phytozome and Gene Bank. These sequences were then aligned and used to construct a phylogenetic tree with the software MegAlign from Lasergene 7 with default conditions (Clustal W and Jotunheim algorithms, respectively). A subsequent alignment was performed with the aminoacidic sequences belonging to each of the clades (groups) formed on the tree, employing the same conditions. On this alignment, regions with the highest similarity were selected to design the degenerate primers.

2.2.8 Design of DGAT degenerate primers

A multiple sequence alignment was performed using sequences of DGAT from microalgae, and from this alignment a set of degenerate primers was designed. Given the numerous sequences used, only the regions with the highest identity and similarity were used for the primer design. To do so, aminoacidic sequencs were used in order to avoid the genetic code degeneracy. In this alignment, we identified regions with the highest similarity and selected them to design the primers. The sequence of each primer corresponds to the codons of every aminoacid found on these regions. Because each codon vary on the third base, special degenerate bases were incorporated on the *in silico* sequence of the primer design. These bases represent every possible nucleotide that could possible occupy such space. For example, degenerate base "R" represents a possible adenine or guanine, base "D" represents guanine, adenine or timine. Four pairs of degenerate primers were designed.

2.2.9 Degenerate primer design to amplify an actin sequence from *C. saccharophila*

As mentioned before, *C. saccharophila* genome is not sequenced so there is no information available about reported oligonucleotides or sequences that could be used as positive controls. Since it is known that actin sequences are amongst the most conserved sequences in eukaryots [Muller *et al.*, 2005], this was the gene of choice to use as a positive control for PCRs. Therefore, degenerate primers had to be designed to amplify a fragment of an actin gene from *C. saccharophila*.

To design the degenerate primers, several *actin* genes were isolated from databases (JGI Phytozome and GeneBank), an *actin* sequence from *C. reinhardtii* ID 603700 was used as query on a BLASTX search. With this strategy, *actin* genes of other green microalgae could be retrieved from JGI Phytozome: *Volvox carteri* (ID 109972 and 27374), *Chlorella sp.* NC64A (ID 136976) and *Ostreococcus tauri* (ID 29599). These sequences were aligned and searched for a highly conserved region, using the same strategy as stated above, to amplify a fragment of approximately 200 bp.

The PCR conditions for the degenerate primers for the actin gene were as followed: 35 cycles consisting of 1 minute at 95 °C for denaturalization step, 30 seconds at 45 °C for annealing, and 1 minute at 72°C for extension. For these PCRs both genomic and complementary DNA were tested using 500 ng of each as templete, 10 mM dNTPs, 10 mM $MgCl_2$, 100 mM of each degenerate primer, 1X buffer and 1 U µLTaq Polimerase.

A fragment from cDNA was obtained and used to further design and synthesize specific oligos for *CsActin*, to be used as positive control in further experiments.

2.2.10 Mapping the 5' and 3' ends

In this study, a degenerate PCR strategy was employed to amplify a putative *DGAT* gene. Therefore it was necessary to elucidate its complete and correct sequence in order to clone it on an expression vector. RLM-RACE and RACE were used to map *CsDGAT* 5' and 3' ends, respectively. The RLM-RACE method has been employed to fully map a gene whose sequence is not available on public databases, it functions to reveal key features on 5' (and also 3') ends, such as start codons and poly adenine sequences. This technique is based on RNA ligase-mediated (RLM-RACE) and oligo-capping amplification of cDNA ends (RACE) methods, and comprises the selective ligation of an RNA oligonucleotide to the 5' ends of decapped full-length mRNA using T4 RNA ligase. This procedure increases substantially the probability to select and

amplify only complete mRNA molecules (not truncated) which are further retro-transcribed to complementary DNA.

Using the products of both 5' and 3' RACE, Touchdown PCR was performed to amplify the full cDNA. For the 5' end, a forward primer from the Invitrogen Kit (5'-CGACTGGAGCACGAGGACACTGA-3', 23 bases, Tm 74°C) and a reverse specific primer for the *CsDGAT* gene (5'-GCCAATAGTCGAACACCTGAGT-3', 22 bases, Tm 64°C) were used for the PCR. For the nested PCR, a forward primer from the Invitrogen Kit (5'- GGACACTGACATGGACTGAAGGAGTA -3', 26 bases, Tm 78°C) and a reverse specific primer for the *CsDGAT* gene (5'-AAGGACTGAGGCCAATAGTCG -3', 21 bases, Tm 64°C) were used. For the 3' end, a reverse primer from the Invitrogen Kit (5'- GCTGTCAACGATACGCTACGTAACG- 3', 25 bases Tm 76 °C) and a forward specific primer for the *CsDGAT* gene (5'-TTGCTCGCAAGAACAGAGTG -3', 20 bases, Tm 64.3 °C) were used for the PCR. For the nested PCR, a reverse primer from the Invitrogen Kit (5' - CGCTACGTAACGGCATGACAGTG- 3', 23 bases, Tm 72 °C), and a forward specific primer for the *CsDGAT* gene (5'- AACTCCTGTGCCGCGCAAGG -3', 20 bases, Tm 73.7 °C) were used. PCR conditions for both reactions of 5' and 3' RACE were as following: The first five cycles consisted of 1 minute at 95 °C denaturation, 30 seconds at 65° annealing and 1 minute at 72 °C extension. The second round of 35 cycles consisted of 1 minute at 95 °C denaturation, 30 seconds at 60° annealing and 1 minute at 72 °C extension.

2.2.11 Purification of PCR products

At first, fragments from PCRs were purified using two methods depending on their specificity: one method used directly the liquid PCR products and the other one used excised gel bands from electrophoresis products.

Roche High Pure PCR Product Purification Kit. Total volume of PCR products were adjusted to a total volume of 100 µL followed by an addition of 500 µL binding buffer. The mixture was transferred to a filter tube and centrifuged for 1 minute at 14,000 rpm. The flowthrough solution was discarded and the filter tube washed twice with 500µL and 200 µL wash buffer each with a centrifugation at 14000 rpm for 1 minute. The flowthrough solution was again discarded and the DNA was recovered by adding 50µL elution buffer and centrifuged for 1 minute at 14000 rpm.

Qiagen Qiaquick Gel Extraction Kit Protocol. Initially, target bands were excised from gel with a scalpel and placed on a 1.5 mL microcentrifuge tube. Next, 3 volumes of QG buffer to 1 volume of agarose were added, and incubated at 50°C for 10 minutes with regular vortexing, followed by an addition of 1 volume isopropanol. The whole mixture

was then applied to a Qiaquick column and centrifuged at 13000 rpm for 1 minute. After discarding the flowthrough, the column was washed with 750 µL PE buffer + ethanol and centrifuged at 13000 rpm for 1 minute. The flowthrough was discarded and the column centrifuged again with the same conditions. After placing the column on a clean tube, the purified fragments were eluted with 50 µL 10 mM Tris HCl and centrifuged at 13000 rpm for 1 minute.

2.2.12 Cloning and sequencing of PCR products

To clone the PCR products, the ligation reaction was performed as described by manufacturer: On a 1.5 mL tube, 7 µL purified insert obtained from the PCR purification (see above), 1 µL ligase buffer 10X, 1 µL pGEM-T-easy Vector 50 ng/µL, 0.2 µL T4 DNA ligase 1U/µL New England Biolabs were added and the final volume adjusted to 10 µL. The reaction was gently mixed and incubated for 16 hours at 16 °C and the enzyme inactivated for 10 minutes at 65 °C.

To transform *Escherichia coli* DH10B cells, 5 µL of the ligation product was mixed with 200 µL cells and incubated on ice for 30 minutes. After a 42°C heat shock for 45 seconds, SOC medium was added and incubated at 37°C for 90 minutes at 180 rpm. From this culture, 100 µL cells were applied on LB solid medium with 10 µg/mL ampicilin, 0.1 M IPTG and 80 µg/µL X-GAL. The plates were incubated overnight at 37 °C, blue and white colonies were observed afterwards.

White colonies, which harbor the gen of interest; were picked and grown on liquid LB medium with 5 µL ampicilin (100 mg/µL) and incubated at 37°C overnight at 215 rpm. Afterwards, the plasmids were extracted using the High Pure Plasmid Isolation Kit (Roche). Cells were collected by centrifugation at 9,000 rpm for 30 seconds and their pellet treated with 250 µL lysis buffer and incubated at room temperature for 5 minutes. Then, 350 µL chilled (4°C) binding buffer were added and incubated on ice for another 5 minutes; following a 14,000 rpm centrifugation. The resulted supernatant was transferred to a spin filter tube and centrifuged at 14,000 rpm for 1 minute, washing with 700 µL buffer II and removing it with another 14,000 rpm centrifugation for 30 seconds. Finally, purified plasmids were collected with 50 µL elution buffer.

Sequencing reactions were based on the Sanger method, using dideoxinucleotides, which lack the 3' OH. The reaction mixture consisted of 5 µL plasmid (approximately 1 µg), 1 µL Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), 3.5 µL buffer 5X, 1 µL oligonucleotidce (sense or antisense) and 9.5 µL water, to reach a volume of 20 µL. PCR cycles consisted of 95°C for 30 seconds, 35 cycles of 95°C 30 seconds, 50°C 30 seconds and 60°C 4 minutes, with a final extension at 60°C for 20 minutes. The reaction was purified afterwards with the addition of a solution consisting

of 0.2 mM Mg_2SO_4 and ethanol, which was incubated at room temperature for 15 minutes. Next, it was centrifuged at 14000 rpm for 20 minutes, and the supernatant discarded. Finally, the pellet was air-dried and processed by the Australian Genome Research Facility (AGRF).

2.2.13 Expression analysis of *CsDGAT* by RT-PCR

Once a *CsDGAT* sequence was identified, a set of specific primers were designed to amplify and analyze gene expression of this gene: Forward 5'-ACCTGTGTATCACTTCGGCAAC -3' and Reverse 5'- CACCACTTCAGCGTGAATCT C -3'. An amplicon of 238 bp was expected when using cDNA as template.

PCR conditions were as follows: one cycle of 3 minutes at 95°C, 35 cycles of a denaturalization step at 95°C for 30 seconds, an annealing step of 55°C for 30 seconds and an extension step of 72°C for 1 minute, and a final cycle of 72°C for 5 minutes.

A similar strategy was employed for *CsActin*. With a degenerate fragment sequenced, a pair of specific primers was created to identify specifically this gene, to be used as positive control on future reactions. The sequences of such primers were: 5'-TGCTGTGGTGGTGAAGCTG -3' for the forward and 5'- TCAACGCACCAGCAATCT ATG -3' for the reverse primer. PCR conditions were the same as the ones used for *CsDGAT* (mentioned above). A fragment of 233 bp was expected to be amplified.

2.3 RESULTS

2.3.1 DNA and RNA extractions

Since there was not a protocol reported for DNA and RNA extraction from *C. saccharophila*, two protocols for DNA and three protocolos for RNA extraction were tested. The results for the DNA extraction using a protocol of Dellaporta [1983] and a protocol of General Electric's extraction kit (illustra Nucleon Phytopure) are shown in Figure 2.2. Table 2.1 summarizes these results. Based on the obtained results, the illustra Nucleon Phytopure kit was selected as the main method to extract DNA on the following experiments.

Table 2.1 Comparison of different methods for DNA extraction. Values of number of cells, yield and purity are given.

	Cell number (million)	Yield (µg/mL)	Purity (A260/A280)
Dellaporta	20	81.42	1.31
	40	123.99	1.56
illustra Nucleon Phytopure	20	224.8	1.31
	40	252.83	1.56

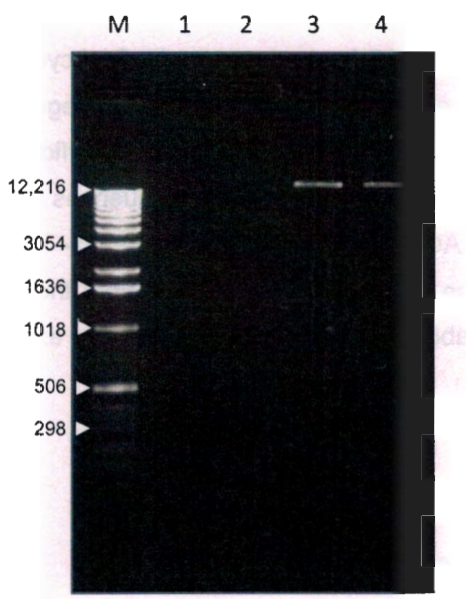


Figure 2.2 DNA extraction method comparisons. Lines 1 and 2: technique based on Dellaporta, 20 and 40 million cells respectively. Lines 3 and 4: General Electric Extraction Kit, 20 and 40 million cells respectively

The results for the RNA extraction using the protocol of the Trizol reagent, the GE illustra RNAspin Mini kit (Fig. 2.3) and a CTAB protocol are shown in Figure 2.4. Table 2.2 summarizes these results. The highest RNA concentration was obtained with CTAB method: 428.92 µg/mL (Fig 2.4). It was observed on previous results that 100 million cells consistently gave the highest RNA concentration, so this number of cells was processed using CTAB. The effect of adding DNAse is also showed (Fig 2.4). As showed, CTAB method gave the highest yields so this procedure was applied to the experimental samples that followed.

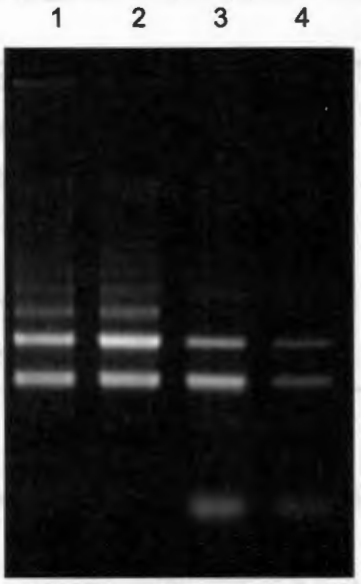


Figure 2.3 RNA extraction. GE's Kit (lines 1 and 2) and Trizol method (lines 3 and 4)

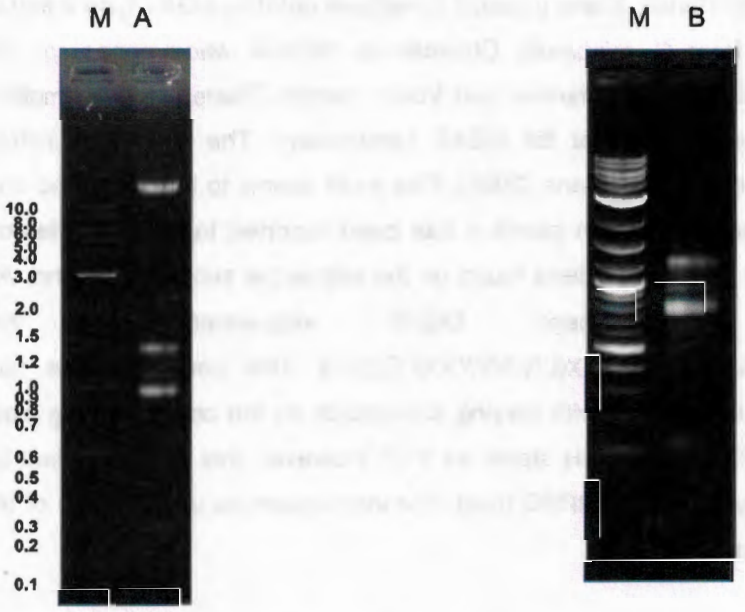


Figure 2.4 CTAB method RNA, before (A) and after (B) DNAse treatment. MWM: 2-Log (NEB)

Table 2.2 Comparison of different methods for RNA extraction. Values of number of cells, yield and purity are given

	Cell number (million)	Yield (µg/mL)	Purity (A260/A280)
Trizol	50	78.38	1.49
	100	86.61	1.54
illustra RNAspin Mini	50	283	1.84
	100	304.73	1.79
CTAB	100	428.92	1.73

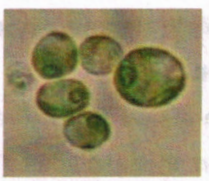
2.3.2 Searches of microalgae and plant *DGAT* genes in the GenBank and Phytozome Databases

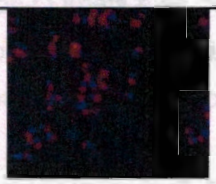
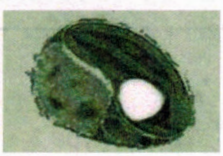
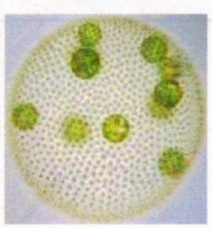
The public databases employed in this study comprise numerous sequences from plants and microalgae (among other organisms) that have industrial, medical or research importance. It was possible to retrieve putative DGAT type 2 and unannotated sequences from *C. reinhardtii, Chlorella sp.* NC64A, *Micromonas sp., Ostreococcus tauri, Ostreococcus lucimarinus* and *Volvox carteri*. There are three motifs reported to be significantly important for DGAT functionality. The first motif HPHG was first discovered on mouse [Stone, 2006]. This motif seems to be conserved on DGAT type 2 sequences, although in plants it has been reported to have a different sequence; EPHS. This motif was indeed found on the microalgal sequences found. Another motif observed on these DGAT sequences is the large RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q). This could also be found on the microalgae sequences, with varying aminoacids on the corresponding places (marked as an X). The third motif is stated as YFP. However, this region has not been given a high importance as the HPHG motif. For the sequences used, not all of them showed this particular YFP motif.

The results of this search are shown on Table 2.3. It can be noted that no DGAT type 3 sequence could be found on any microalgae genome sequenced so far. To date, there has been no report of a cytosolic (type 3) DGAT on microalgae nor a prediction for such enzyme on either database (NCBI and JGI). Regarding DGAT type 1, the sequence from *A. thaliana* used showed similarity with another putative DGAT from *Chlorella sp.* NC64A. However such enzyme is annotated on JGI database as a possible Sterol O-acyltransferase / diacylglycerol acyltransferase. In fact, *Chlorella sp* NC64A DGAT is shown to form a distinct clade with a *V. carteri* DGAT which is also

annotated as a sterol O-acyltransferase / diacylglycerol acyltransferase (Figure 2.5 B). The most abundant type of DGAT found was the type 2, which seemed to be more related to the microalgae sequences than the other types.

Table 2.3 Sequences retrieved from plants and microalgae. Their identification number is shown

Gene	ID	Organism
		Arabidopsis thaliana
DGAT type 1	AF051849	
DGAT type 2	NP_566952	
		Arachis hypogaea
DGAT type 3	AAX62735	
		Ricinus communis
DGAT type 3	XP_002519339	
	10860	*Chlorella sp.* NC64A
	138958	
DGAT	25078	
	52783	
	57660	
Sterol-O acyl transferase	28812	
	141301	*Chlamydomonas reinhardtii*
	180240	
DGAT	184281	
	188937	
	190539	
	108405	*Micromonas sp.*
	58460	
DGAT	61819	
	82855	

	99714	
		Ostreococcus lucimarinus
DGAT	13176 16489 26790 27747	
		Ostreococcus tauri
DGAT	34368 35844 8948 9124	
	103301	*Volvox carteri*
DGAT	104223 77655 89744 89965 93670	
Sterol-O acyl transferase	69078	

2.3.2.1 Multiple sequence alignment and phylogenetic analysis of microalgae and plant genes

This tree showed 5 different clades formed only by DGAT type 2 sequences, and so based on these clades (Fig. 2.5) a multiple alignment of each clade (group) was performed separately (Fig. 2.6). Motifs HPHG, RXGFX(K/R)XAXXXGXXX(L/V)VPXXX FG(E/Q and YFP were marked where found. It was observed that motif HPHG was also found in plants as EPHS, which was expected given that microalgae and plants are evolutively and morphologically related. Of all the groups aligned, group 4 showed the smallest divergence having several residues shared among its members. For the rest of the sequences, the biggest similarity among them resided on these mentioned motifs. For group 5, none of these conserved regions could be observed so it was not considered for further analysis.

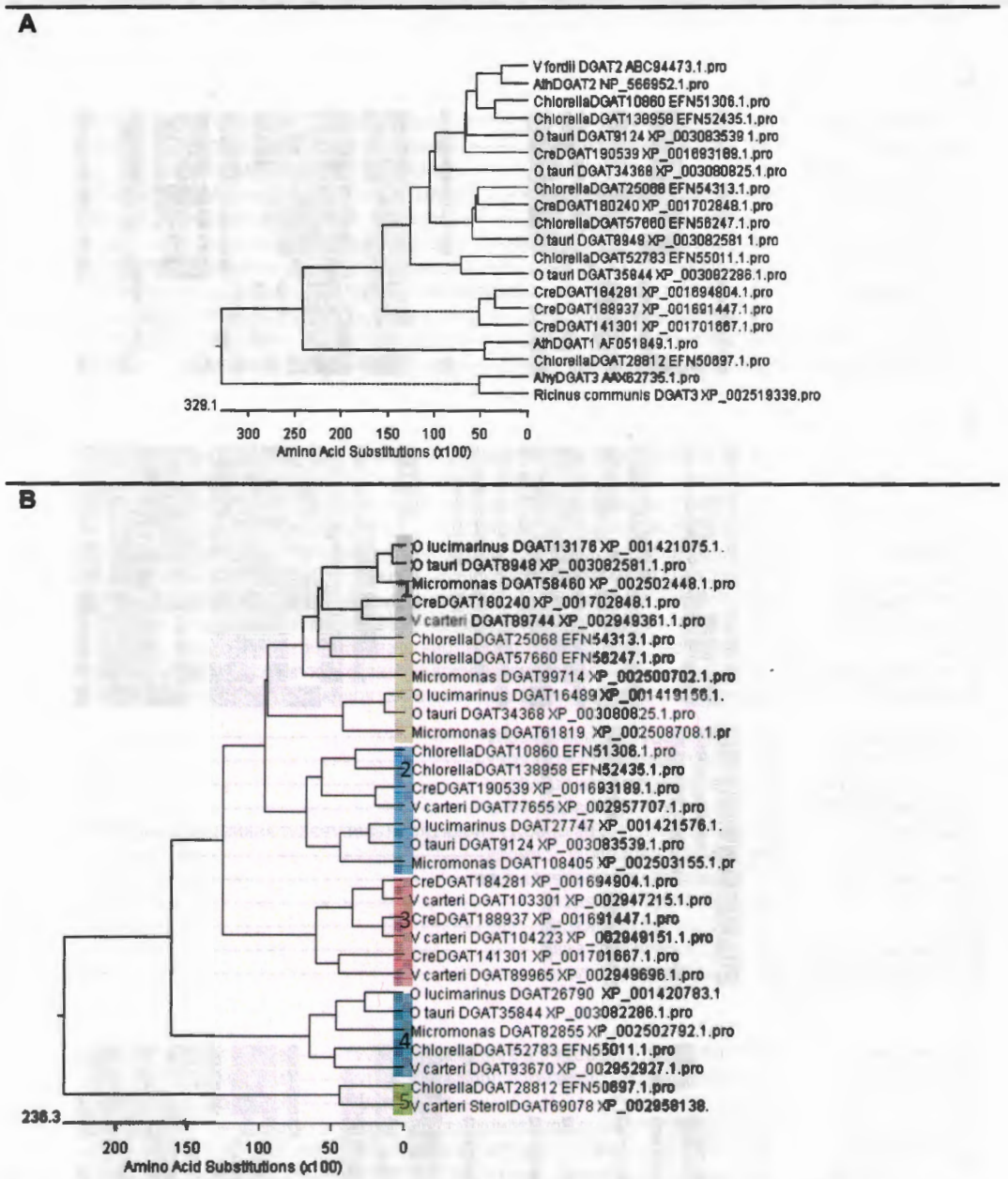
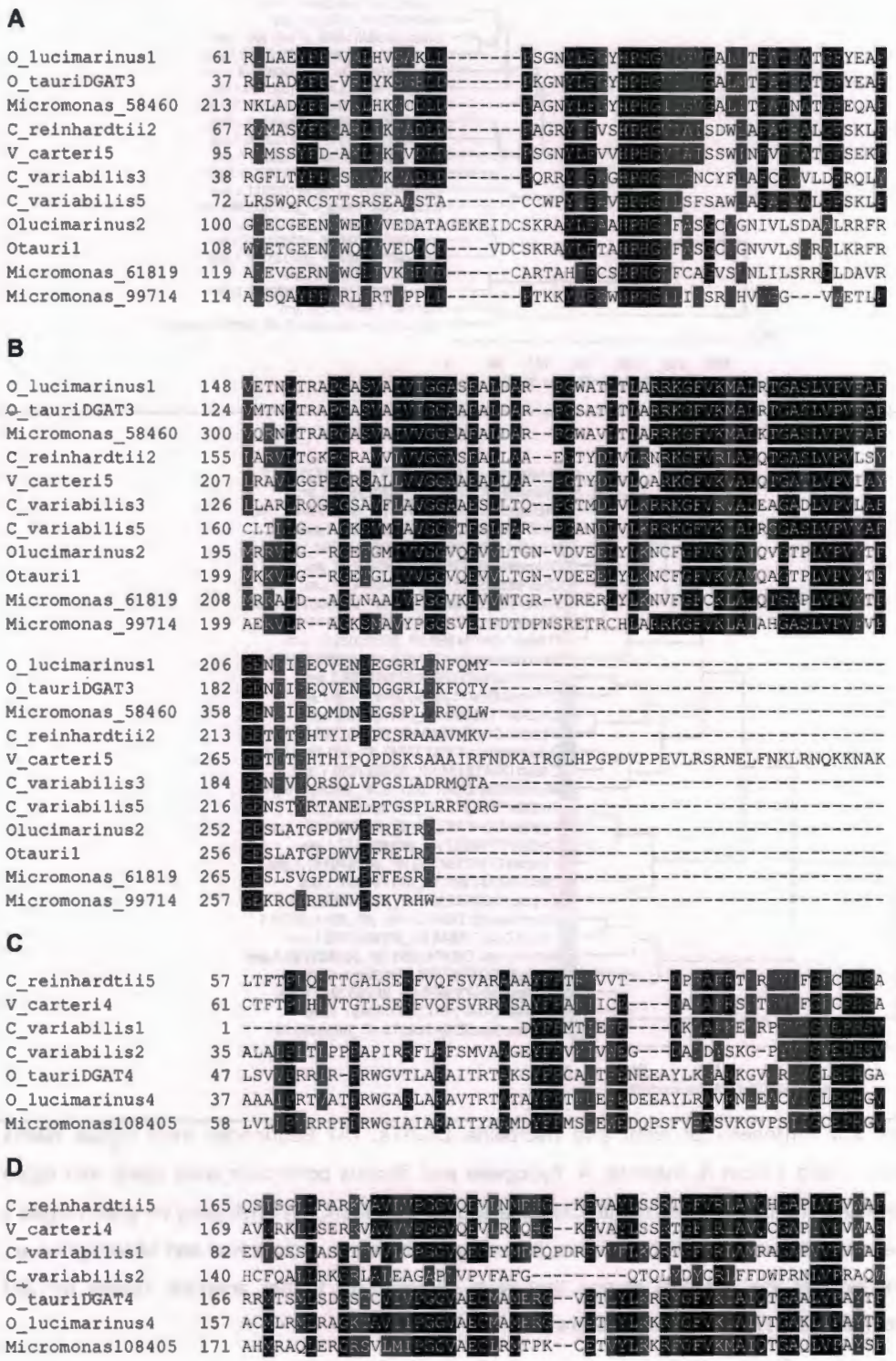


Figure 2.5 Phylogeny of plant and microalga DGATs. (A) Sequences from higher plants DGAT1, 2 and 3 from *A. thaliana, A. hypogaea* and *Ricinus communis* were used, and algae putative DGAT's from *C. reinhardtii, Chlorella spp.* and *O. tauri*. (B) Phylogeny for green algae's sequences retrieved from JGI database, including *V. carteri*, *O. lucimarinus* and *Micromonas sp.* Colors indicate the different groups that were selected for further analysis. Codes for JGI (microalgae) and NCBI's databases are shown


A
O_lucimarinus1 61
O_tauriDGAT3 37
Micromonas_58460 213
C_reinhardtii2 67
V_carteri5 95
C_variabilis3 38
C_variabilis5 72
Olucimarinus2 100
Otauri1 108
Micromonas_61819 119
Micromonas_99714 114
B
O_lucimarinus1 148
O_tauriDGAT3 124
Micromonas_58460 300
C_reinhardtii2 155
V_carteri5 207
C_variabilis3 126
C_variabilis5 160
Olucimarinus2 195
Otauri1 199
Micromonas_61819 208
Micromonas_99714 199
O_lucimarinus1 206
O_tauriDGAT3 182
Micromonas_58460 358
C_reinhardtii2 213
V_carteri5 265
C_variabilis3 184
C_variabilis5 216
Olucimarinus2 252
Otauri1 256
Micromonas_61819 265
Micromonas_99714 257
C
C_reinhardtii5 57
V_carteri4 61
C_variabilis1 1
C_variabilis2 35
O_tauriDGAT4 47
O_lucimarinus4 37
Micromonas108405 58
D
C_reinhardtii5 165
V_carteri4 169
C_variabilis1 82
C_variabilis2 140
O_tauriDGAT4 166
O_lucimarinus4 157
Micromonas108405 171

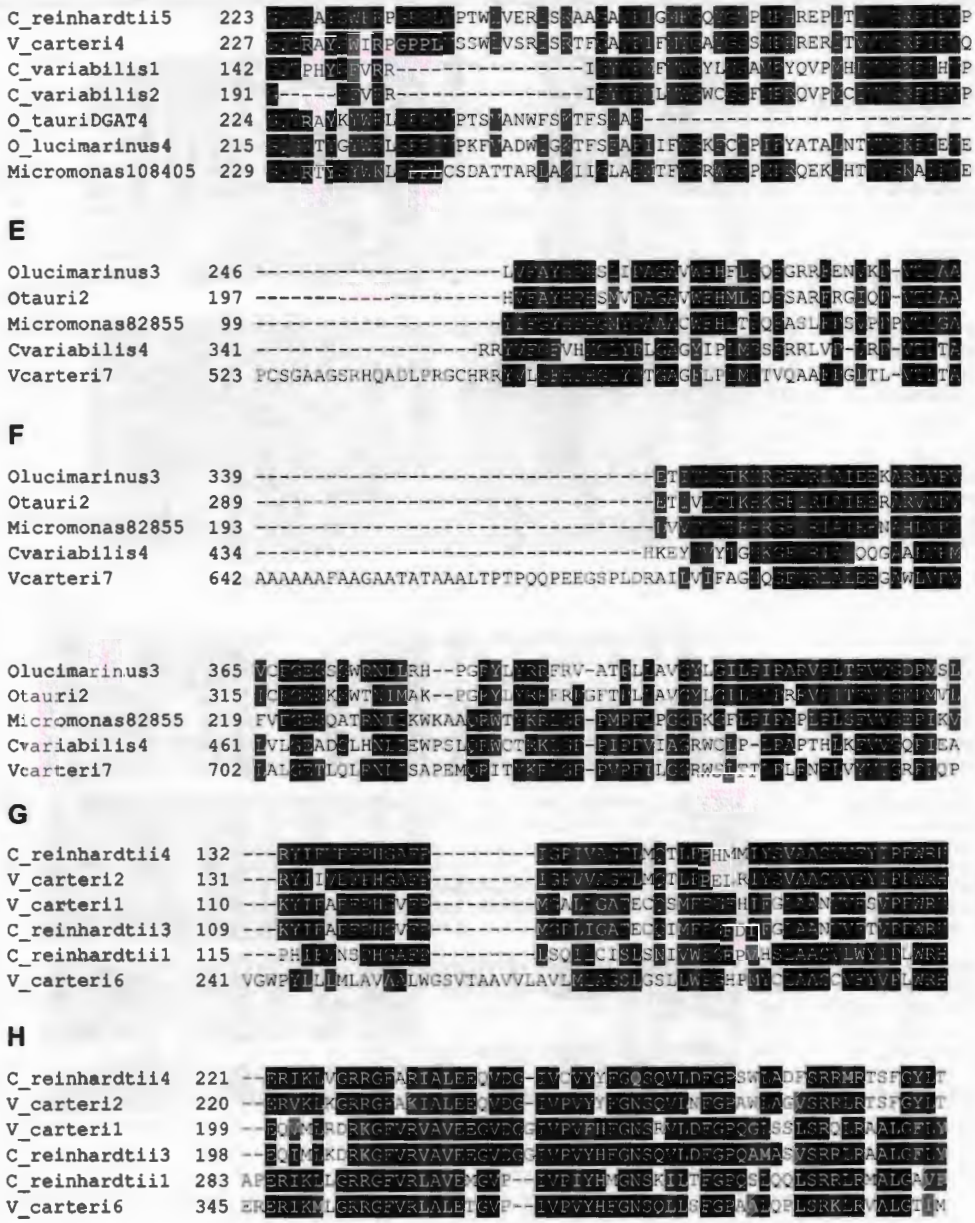
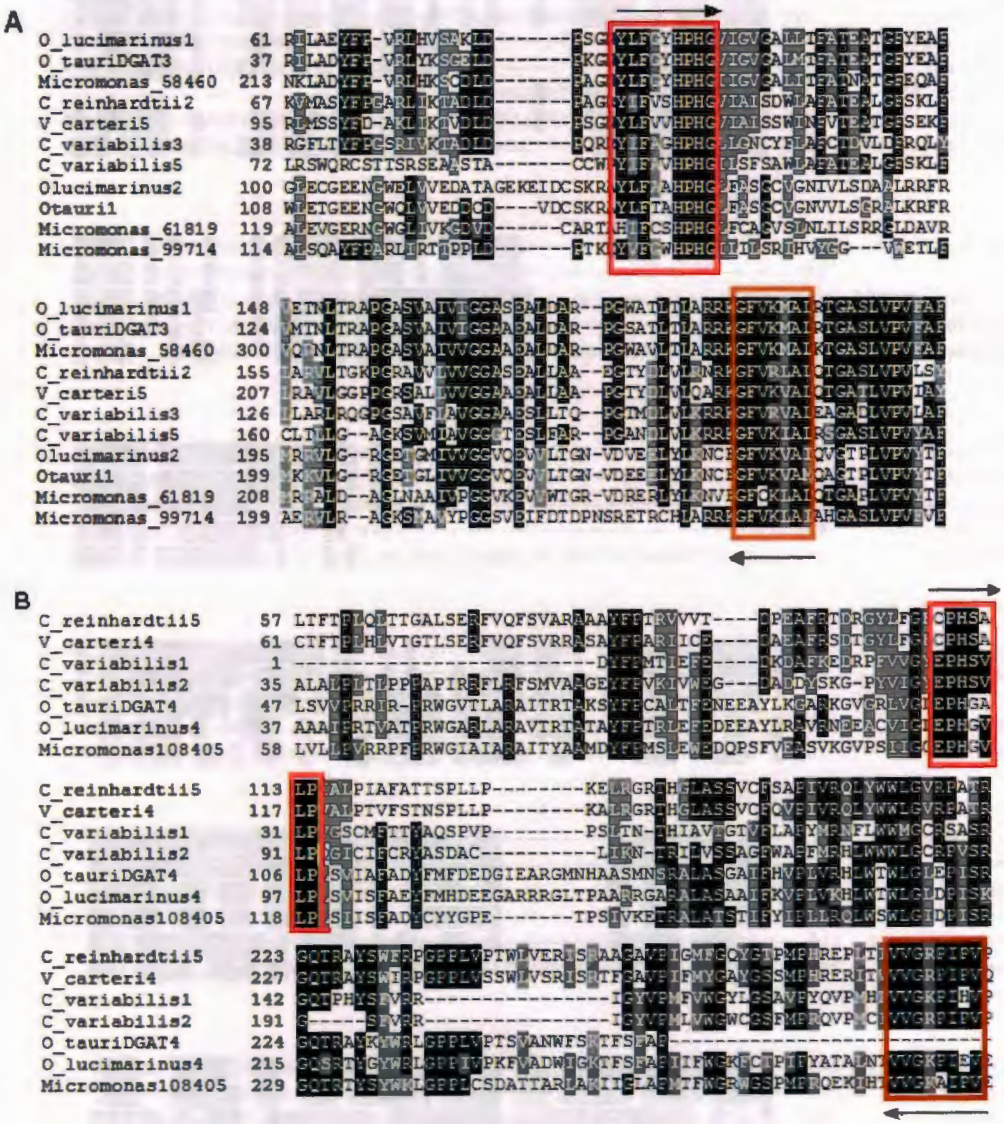


Figure 2.6 Multiple sequence alignments. (A & B) group 1, (C & D) group 2, (E & F) group 3 and (G & H) group 4. Conserved motifs marked in blue squares

2.3.3 Design of DGAT degenerate primers

As described on materials and methods, four different pairs of degenerate primers were synthesized (Table 2.4). The alignment of aminoacidic sequences performed to design these primers where highest similarity regions were employed (Fig. 2.7).


A
O_lucimarinus1 61
O_tauriDGAT3 37
Micromonas_58460 213
C_reinhardtii2 67
V_carteri5 95
C_variabilis3 38
C_variabilis5 72
Olucimarinus2 100
Otauri1 108
Micromonas_61819 119
Micromonas_99714 114
O_lucimarinus1 148
O_tauriDGAT3 124
Micromonas_58460 300
C_reinhardtii2 155
V_carteri5 207
C_variabilis3 126
C_variabilis5 160
Olucimarinus2 195
Otauri1 199
Micromonas_61819 208
Micromonas_99714 199
B
C_reinhardtii5 57
V_carteri4 61
C_variabilis1 1
C_variabilis2 35
O_tauriDGAT4 47
O_lucimarinus4 37
Micromonas108405 58
C_reinhardtii5 113
V_carteri4 117
C_variabilis1 31
C_variabilis2 91
O_tauriDGAT4 106
O_lucimarinus4 97
Micromonas108405 118
C_reinhardtii5 223
V_carteri4 227
C_variabilis1 142
C_variabilis2 191
O_tauriDGAT4 224
O_lucimarinus4 215
Micromonas108405 229

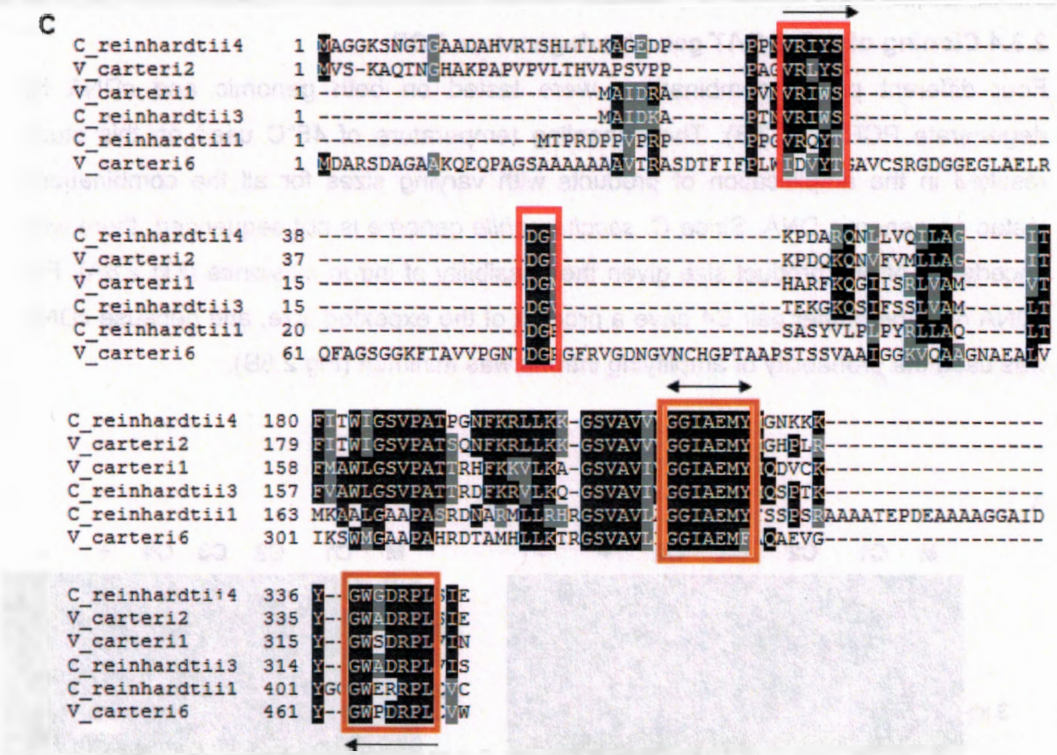


Figure 2.7 High similarity regions selected on groups resulted from the phylogenetic analysis to design primers, (A) group 1, (B) group 2 and (C) group 3. Middle section of group 3 was used both for Forward and Reverse primers. Motifs HPHG and RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q are included

Table 2.4 Name, sequence and origin of degenerate primers designed and used in this study

Primer ID	Sequences	Origin
C1	5'- TAYHTITTYGBIIIICAYCCICAYGG -3'	Group 1
	5'- ADIGCIAIIYKIACRAAICC -3'	
C2	5'- KRICCICAYISIGYIYTICC -3'	Group 2
	5'- RYIKSDATIGGIYIICCIACIAC -3';	
C3	5'- GTIMGIHTITRBWSIGAYGG -3'	Group 3
	5'- TACATYTCIGCDATICCICC -3'	
C4	5'- GGIGGIATHGCIGARATG -3'	Group 3
	5'- ARIGGICKRTCIIICCAICC -3'	

2.3.4 Cloning of a *CsDGAT* gene by degenerate PCR

Four different primer combinations were tested on both genomic and cDNA by degenerate PCR (Fig 2.8). The annealing temperature of 45°C used on this study resulted in the amplification of products with varying sizes for all the combinations tested on genomic DNA. Since *C. saccharophila* genome is not sequenced, there was uncertainty of any product size given the possibility of intron existence (Fig 2.8A). For cDNA only the primer pair C4 gave a product of the expexted size, and because cDNA was used the probability of amplifying introns was minimun (Fig 2.8B).

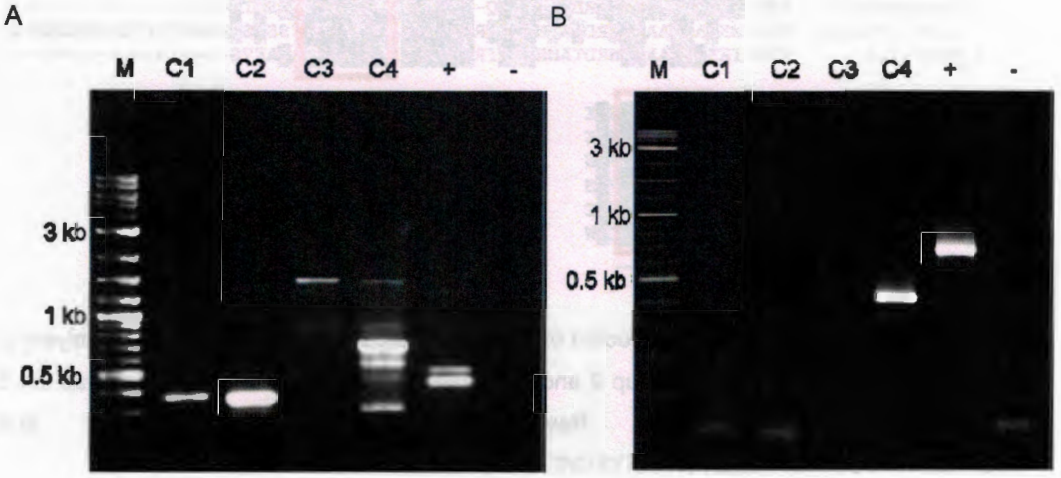


Figure 2.8 Degenerate PCR with (A) gDNA and (B) cDNA as template. Actin was used as positive control (+), and water as negative control (-). Each lane contains a result from each combination of primers used (C1 to C4). M = Molecular marker 2-log (New England Biolabs)

All the amplicons from C1, C2 and C3 were subsequently cloned in pGEM-T easy and sequenced. Since an amplicon of the expected size (400 bp approximately) was obtained for C4 on cDNA, this amplicon was also cloned and sequenced.

In order to corroborate whether the sequenced clones corresponded to DGAT-type genes, a BLASTX search was performed. Only the amplicon obtained using the C4 combination of primers on cDNA proved to be a DGAT type sequence, so this sequence was chosen for further analysis. The searches performed for the sequences from C1, C2 and C3 failed to show any resemblance to this gene family whatsoever, so they were not further investigated. The most significant BLAST results are shown on Table 2.5. Based on these findings, it could be confirmed that the sequence amplified by the degenerate PCR was a DGAT-type gene so it was named *CsDGAT*. The

sequences with the most similar results were from other green microalgae, *V. carteri* and *C. reinhardtii*, whose genomes are sequenced. Figure 2.9 shows the sequence from combination 4 found and regions of interest. Additionally, the location of the set of primers designed to map the 5' and 3' ends of the cDNA, using RLM-RACE and RACE, respectively, are shown.

Table 2.5 BLAST results using the sequence resulting from C4 combination of primers on cDNA. Number of ID according to Phytozome or NCBI

Sequence	Organism	ID	% Identities	% Positives	Score	E
Diacylglycerol acyltransferase	*V. carteri*	XM_002947169	58	75	151	$2.2\ e^{-45}$
Diacylglycerol acyltransferase	*C. reinhardtii*	XM_001694852	57	75	147	$1\ e^{-44}$
Diacylglycerol acyltransferase	*C. reinhardtii*	XM_001691395	51	69	92.3	$1\ e^{-38}$
Diacylglycerol acyltransferase	*V. carteri*	XM_002949105	48	72	87.2	$2\ e^{-37}$
Diacylglycerol acyltransferase	*V. carteri*	XM_002949650	46	67	84	$8\ e^{-33}$

```
1     GGG GGG ATT GCG GAG ATG TAC ATG CAG CAC AAG CGC AAG GAA CGC   45
1      G   G   I   A   E   M   Y   M   Q   H   K   R   K   E   R   15

46    ATC AAG CTG CTA GAC AGG AAA GGG TTT GTC AAG ATT GCT GTT GAG   90
16     I   K   L   L   D   R   K   G   F   V   K   I   A   V   E   30

91    GAG GGC CTG GAT GGA GGG ATC ATA CCT CTG TAT CAC TTC GGC AAC   135
31     E   G   L   D   G   G   I   I   P   V   Y   H   F   G   N   45
          5' RACE nested <------  5' RACE ------
136   ACT CAG GTG TTC GAC TAT TGG CCT CAG TCC TTT GAG AAA TTT GCT   180
46     T   Q   V   F   D   Y   W   P   Q   S   F   E   K   F   A   60
      3' RACE ------>
181   CGC AAG AAC AGA GTG GGT GTG GGA TTC TTG GTC GGA CGA TGG GGA   225
61     R   K   N   R   V   G   V   G   F   L   V   G   R   W   G   75
      3' RACE nested ------>
226   ACT CCT GTG CCG CGC AAG GTG CCC TTG TAC ATG GTG AGT GGC AAG   270
76     T   P   V   P   R   K   V   P   L   Y   M   V   S   G   K   90

271   CCC ATT CCA GTG CCC AAG GTC GAC AAG AAT GAT ACC GAG AAG TTC   315
91     P   I   P   V   P   K   V   D   K   N   D   T   E   K   F   105

316   AAC CAG ACT GTA GAT GAG ATT CAC GCT GAA GTG GTG CAA CAG CTG   360
106    N   Q   T   V   D   E   I   H   A   E   V   V   Q   Q   L   120
```

```
361   CAG GAC CTC TAT GAC AGA CAC AAG GCA AGC TAT GGC TGG CCC GAC   405
121    Q   D   L   Y   D   R   H   K   A   S   Y   G   W   P   D    135

406   CGC CCC CTA   414
136    R   P   L
```

Figure 2.9 Sequence obtained from degenerate RT-PCR from C4 primer combination, corresponding to a partial ORF of *CsDGAT*. Motif highlighted in gray RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q). Arrows above the sequence indicate the position and orientation of the 5'RACE and 3'RACE primers used in this study

2.3.5 Degenerate primer design to amplify an *actin* gene from *C. saccharophila*

As stated before, the complete genome of *C. saccharophila* has not been sequenced yet, so it was necessary to find and map a house keeping gene to be used as positive control in PCR reactions. The selected gene was *actin*, since other microalgae sequences are readily available and found to be conserved.

The BLASTX search resulted on four *actin* sequences belonging to three diferent microalgae. Table 2.6 shows the identification numbers and organisms of the actins retrieved.

Table 2.6 Organisms and identifications (ID) for actin genes

ID	Organism
109972	*V. carteri*
127374	
136976	*Chlorella sp.*
29599	*O. tauri*

An alignment of the actin sequences was performed (Fig 2.10). As expected, all sequences were highly conserved, having only minimal differences. A set of degenerate primers were designed on these conserved regions and used as positive controls for future analysis. Table 2.7 shows the name and sequence of these primers

Figure 2.10 Multiple sequence alignment of actin genes. Position of degenerate primers are indicated with red arrows

Table 2.7 Name and sequence of actin degenerate primers

ID	Sequence
Control + F	5'- ACIAAYTGGGAYGAYATGG -3'
Control + R	5'- MSIAMCATIGTIGTICCICC -3'.

2.3.6 Cloning of a *CsActin* gene by degenerate RT-PCR

Using cDNA as template, a degenerate PCR was performed in order to amplify and isolate a possible actin gene from *C. saccharophila* (Fig. 2.11). The sequence obtained with the *actin* degenerate primers was also submitted to the GenBank using the BLASTX algorithm in order to search for homologues. The BLASTX results indicated that this *C. saccharophila* sequence correspond to an *actin* type of gene (Table 2.8).

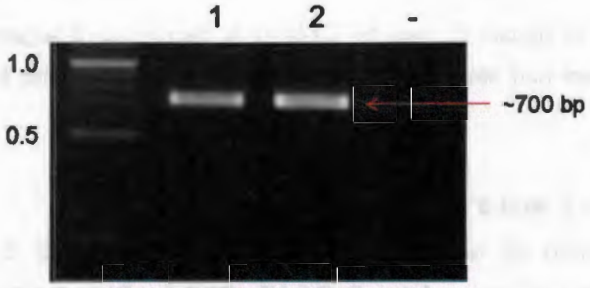


Figure 2.11 Degenerate PCR of *CsActin*. Lanes 1 and 2 show degenerate PCR result with cDNA of *C. saccharophila* as template. Amplifications of ~700 bp are shown with arrow. Negative control (-) with water and no cDNA is included. Molecular marker 2-log (NEB)

Table 2.8 BLAST results using the sequence obtained from actin primers in a degenerate PCR. Number of ID according to Phytozome or NCBI

Sequence	Organism	ID	% Identities	% Positives	Score	E
Actin	*C. vulgaris*	U66585	99	100	538	$5\ e^{-150}$
Actin	*C. reinhardtii*	XM_001699016	94	98	514	$8\ e^{-143}$
Actin	*V. carteri*	XM_002955490	94	97	510	$1\ e^{-141}$

A set of specific primers for *CsActin* were designed based on the sequence resulted from the degenerate PCR. These primers were used on the remaining experiments, to amplify a fragment of 233 bp. Figure 2.12 shows the sequence and location of primers.

A

GCGAACATGGTGGTGCCGCCAGACAGCACGATGTTGCTGTACAGATCCTTCCTGATATCAACAT
CACACTTCATGATGGAGTTGAAAGTGGTATCATGGATACCGTTGGCTTCCATGCCCACCATGCT
GGGGTTGAACAGCACCTCAGGGCAGCGGAAGCGCTCGTTACCAATGGTGATCACCTGTCCGTCT
GGCAGCTCATAGTTCTTCTCCAGTGCGCTGCTAGACATTGATGTTGCCATCTCCTGCTCAAAGT
CCAGGGCTACATAGGCCAGCTTCTCCTTGATGTCACGAACAATTTCACGCTCTGCTGTGGTGGT
GAAGCTGTACCCGCGCTCCATGAGGATCTTCATGAGCCAGTCCGTCAGGTCACGACCTCCCACC
TCTAGACGCAGGATAGCATGTGGCAGGGCATAACCCTCGTAGATAGGCACTGTGTGGGTGACAC
CATCACCAGAGTCCAGCACAATACCAGTGGTACGACCGCTGGCGTACAGAGACAGCACGGCCTG
GATGGCAA[redacted]TGAGGTCTCAAACATGATCTGTGTCATCTTCTCCCTG
TTGGCCTTGGGGTTCAGGGGCGCCTCTGTCAGCAGCACTGGGTGTTCCTCTGGTGCAACACGCA
GCTCATTATAGAAGGTGTGGTGCCAGATCTTCTCCATGTCAT

B

ID	Sequence
Control + Fwd	TGCTGTGGTGGTGAAGCTG
Control + Rev	TCAACGCACCAGCAATCTATG

Figure 2.12 Design of specific primers for *CsActin*. A. Sequence of degenerate *CsActin* gene, localization of primers and length of fragment (233 bp). B. Name and sequence of specific primers

2.3.7 Mapping the 5' and 3' ends of the *CsDGAT* cDNA

Using a combination of universal and gene-specific primers it was possible to determine the 5' and 3' ends of the *CsDGAT* cDNA by RLM-RACE . Fig 2.13 shows that for the 5' end, one prominent band of approximately 500 bp was amplified, and for the 3' end at least four bands were amplified, being the two more intense bands of 907 and 384 bp. All these prominent bands were cloned and sequenced. The weak amplifications obtained (Fig. 2.13 B) were not considered for cloning and sequencing.

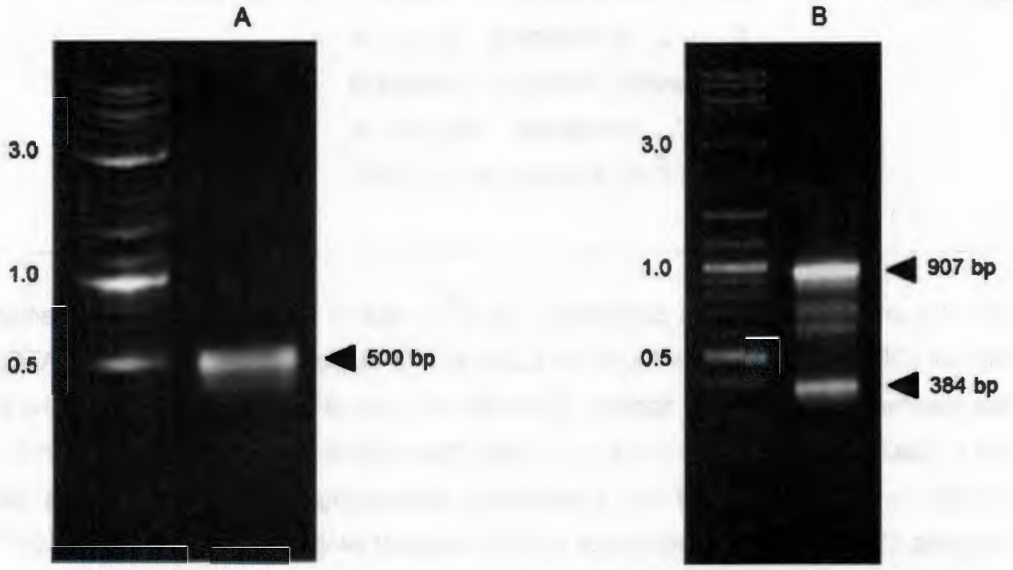


Figure 2.13 RACE reactions of 5' (A) and 3' ends (B). Molecular marker 2-log (NEB).

Sequences of both 5' and 3' ends were edited to remove artificial sequences (RLM-RACE primers and plasmid) and assembled to evaluate possible incomplete or unspecific amplifications; also to revise for possible alternative splicing. Several features were found on both ends of this gene, which are described as follows.

2.3.7.1 Features of *CsDGAT* 5' end

The sequence resulted from the RLM-RACE for the 5' end (Fig. 2.14) was introduced into the PLACE bioinformatic tool (see materials and methods) to search for possible regulatory elements located in *CIS*. Table 2.9 shows some of the major components found in *CsDGAT* 5' UTR.

Table 2.9 Predicted *CIS* elements using PLACE for *CsDGAT* 5' UTR

Element	Function	Reference
Intron-exon splice junction	TGCAGG is a consensus sequence for plant intron. 3' intron-exon junction. Position in *CsDGAT*: 205	Brown 1986.
CCAAT Box	Sequence found in 5' UTR of eukaryotic genes. Position in *CsDGAT*: 168	Rieping *et al.* 1992

EECCRCAH1	Consensus motif of EE-1 and EE-2, two enchancers found in periplasmic carbonic anhydrase in *C. reinhardtii*. Sequence: GANTTNC. Position in *CsDGAT*: 152	Kucho *et al.* 2003

With the analysis of 5' UTR completed, we did a search for possible Open Reading Frames (ORFs) that could be found for *CsDGAT*. To do so, a manual search of ATGs and bioinformatic tools were applied. The main features of the ORF presence in the 5' end of *CsDGAT* are showed in Fig 2.14. Both the manual searches for ORFs and the analysis by BLAST pointed out a sequence comprising 855 bp (when using the complete *CsDGAT*) as the sequence with the highest probability to code for a DGAT protein. Nevertheless, this would imply a large fraction of the 5' of approximately 300 bp to be non-coding.

Upstream this possible ORF 1, the analysis performed by FGenesH found a putative intron in the middle of the cDNA sequence, which could mean an event of intron retention as a mechanism to regulate expression. The existence of this intron would certainly modify the open reading frame to a big extent, either by staying on the final version or by being removed. Therefore, two possible scenarios regarding the predicted intron could be considered:

1) Intron removal: if it is spliced out from the mature RNA, a functional protein could be expressed having the specific and important motifs for DGAT.
2) Intron retention: when not spliced, this sequence adds a stop codon to the otherwise complete ORF.

5'ATATTCGTGTGTCCTCTTTTGTGTT[ATGTTC]CGCTC[AAATTGTCATCAAGATGGCATGCCCAGCTTAGC

TGTGCCTGTCTGTTGCACCTGCAAACGAGGACATGAACG**ATG**CTTTGCAAGTCAGAATTGGATTGCCTTTGTGATGAC

TTACTTTGGACAGCGCAATATCTGtaagccagcccgacgttttgttgcactgcagg**ATC**TATTCTGATGGCTTCAGCA

AGGAGCACAAGCAGTCGTTCATCTCATGGCTTGTGGCAATTGTCACCCTTACCATCTACACAGGATGG**ATG**CACATAC

TGATAGGGTTGACTGTGGCCAGTTTGTTCAGTAGGACATGCCTCTGCATCCTCGTAGCCATATGGGGAACGGTCTTCC

TGCCGCCAAAGCCAGTGTTGTGC[ATGC]TTCTGCAA 3'

Figure 2.14 Possible ORFs predicted for *CsDGAT* 5` region .ATG: first initiation codon found, with a putative intron on its sequence (in red) detected by FGENESH. ORF starting in ATG does

not have an intron. Primer locations for amplification of a fragment of *CsDGAT* containing a predicted intron are highlighted in gray

If the existence of such intron is true, therefore at least two populations of mRNAs would be present: one carrying an intron and another not carrying it. So, a pair of primers was designed flanking this intron which could result in two possible amplifications: one of 400 and another of 368 bp (Fig. 2.15).

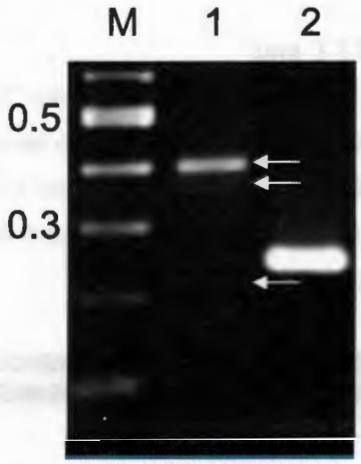


Figure 2.15 RT-PCR to investigate the presence of a potential intron in the 5´region of *CsDGAT*. 1) Amplification of a fragment of *CsDGAT* cDNA containing a predicted intron, 2) Primers for *CsActin* were used as positive control. Molecular marker 2-log (NEB)

As expected, more than one group of cDNAs was amplified (Fig. 2.15). The expected ~400 bp amplicon had the most prominent band, while two other bands of a little less than 400 bp (~360 bp) and ~200 bp were visible but at a much weaker intensity. The stronger amplification suggests a bigger proportion of mRNAs with the potential intron retained. Further studies will be needed to investigate whether the retention of the intron can be modified by variation in the microalgal culture conditions. The ~360 bp band appers to correspond to mRNAs with the intron removed. Furthermore, there could be a protein translated from mRNAs containing a different start codon, which could correspond to the ~200 bp observed in Figure 2.15. The presence or absence of the intron in the cDNA may be causing a different start codon for *CsDGAT*. Probably a new start codon could be formed as a result of the removal of the putative intron(Fig. 2.16).

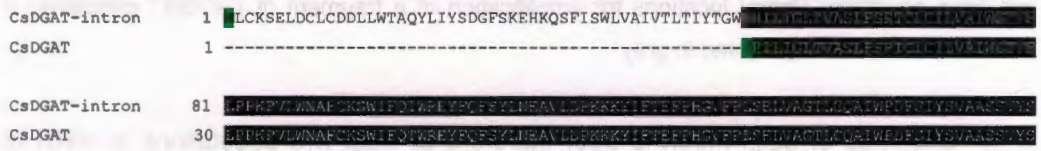


Figure 2.16. Alignment of proteins coded by both versions of *CsDGAT*. Differences in start codon highlighted with green.

2.3.7.2 Features of *CsDGAT* 3' end

Using RACE, the 3' end of *CsDGAT* was also mapped. Two fragments were amplified and sequenced, and both were named based on the length of their UTRs: Short (153 bp) and Long (716 bp). Afterwards, coding sequences and important features specific for their UTR regions, such as poliadenilation sites, were searched. The most important features are shown on Figure 2.17.

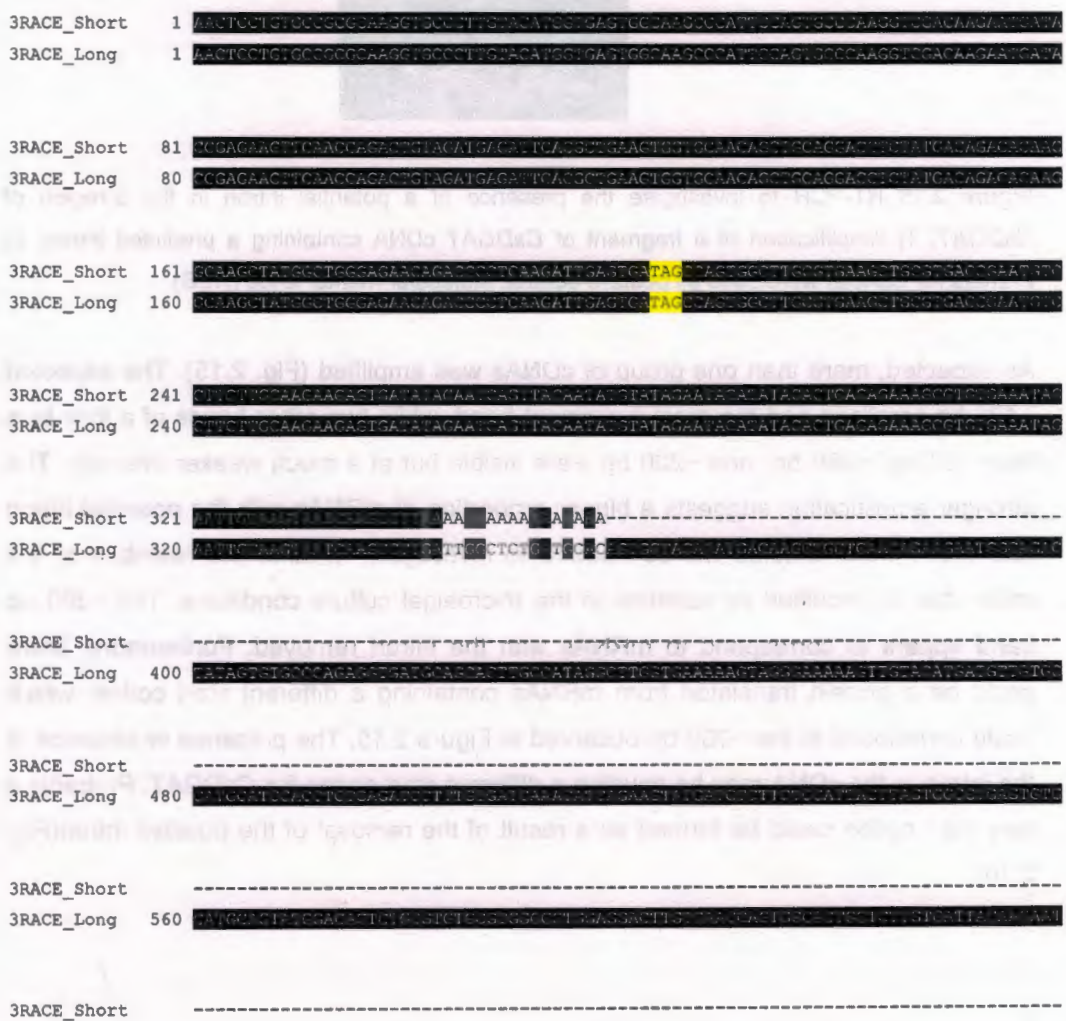


Figure 2.17 Alignment of *CsDGAT* 3' ends and its UTRs. Stop codon highlighted in yellow

As seen on Figure 2.17, the coding sequence of both fragments were the same and had the same stop codon, whereas the main difference relied on their 3' UTR, where major differences on length were found. This may suggest a regulation at a post-transcriptional level. A variable number of specific sequences found on 3' UTRs, which are involved in gene regulation, has been reported in microalgae and plants [Shen *et al.* 2008]. These sequences are known as: starting from the poly A, Near Upstream Elements (NUE), Far Upstream Elements (FUE) and poly A signals. These sites were searched on both *CsDGAT* short and long 3' UTRs, and their localization is shown on Figure 2.18.

A

```
410      TAG CCA GTG GCT TGC GTG AAG CTG CGT GAC CGA ATG TCC TTC   450
451  TTG CAA GAA GAG TGA TAT AGA ATC AGT TAC AAT ATG TAT AGA ATA   495
496  CAT ATA CAC TGA GAG AAA GCG TGT GAA ATA GAA TTG TAA GTA ATC   540
541  AAG CCT AAA AAA AAA AAA AAA
```

B

```
496  CAT ATA CAC TGA GAG AAA GCG TGT GAA ATA GAA TTG TAA GTA ATC   540
541  AAG CCT TGA TTG GCT CTG ATG CAC AGT GGT ACA TAT GAC AAG CTG   585
586  CTG AAC AGA AGA ATG GCA GAC TAT AGT GTG CCC AGA GGT GAT GCA   630
631  GCA CTT GCC ATA GGC TTG GTG AAA AAG TAG GGA AAA AGT TGT ACA   675
676  TGA TGT TGT GCC ATG CAT CCA CCT GTG AGA TCC TTT GTG ATC AAA   720
721  CAT CAG AAG TAT GAT GTC TTY TGC ATT ACT TAG TGG CGT GAG TGT   765
766  GCA TGC AGT ATG GAC AGT CTG CGT GTG CGT GGG GGT GGA GGA CTT   810
811  GTG AAC CAG TGG CAC TGG TCT TGT GAT TAC AAG AAT ATG TTT ACA   855
856  GCG TTA ATA TTG CAT ACT GTT GGT TCG AGG TGG TTA TGC AAG TGC   900
901  AAG ATC AGT CGG CCA TGC GGT GTA TCA GGT GCC TAT CGC AAT TTG   945
```

```
946   TGT ATG CTG AAC TGC AGC CAG CTA CTG TTT GCA ATC CTG CAT CAC   990
991   AAG ATA TGA GTG TTT TTG TCA ATT GTG TGG CTT GAC CAG GCT TGC   1035
1036  CTT CAA AGG TGC ACA TTG ATT TTG AAC CAG TAC AAC TGC ACT GGC   1080
1081  GTA CTG GCA AGG AGT AAT TAA AGA AAC CAA TTC AAA AAA AAA AAA   1125
1126  AAA
```

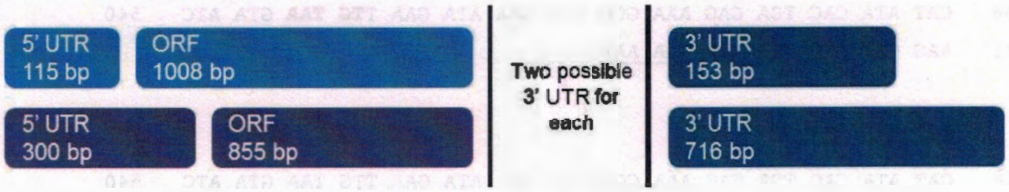
Figure 2.18 Features of *CsDGAT* 3' UTR A) Short and B) Long. Far Upstream Elements (FUE) highlighted in blue, Near Upstream Elements (NUE) in red. Poly Adenines underlined

Because of the large difference on the length of both 3' termini studied here, two versions of *CsDGAT* were considered based on their 3' UTR length: short (153 bp) and long (716 bp). Interestingly, some of the regulatory conserved sequences were found only on the longer UTR. A NUE exists on both 3' UTRs, although with defferent sequences (TGTAA in the short version and AATTAA in the long version), indicating an alternate poyadenylation site. The implication of this phenomenon will be discussed later in this chapter.

2.3.8 Complete mapping of *CsDGAT* cDNA

With both 5' and 3' ends fully mapped, the complete sequence of *CsDGAT* cDNA was determined. Based on its nucleotide sequence further analyzed by FGenesH, two possible ORFs were considered: 1) one starting on the first ATG identified and considering the removal of the putative intron, and 2) another starting on the second ATG identified. The schematic representation and nucleotide sequence of both sequences are shown on Fig. 2.19.

A)



B)

```
1     ATG CTT TGC AAG TCA GAA TTG GAT TGC CTT TGT GAT GAC TTA CTT   45
1      M   L   C   K   S   E   L   D   C   L   C   D   D   L   L   15

46    TGG ACA GCG CAA TAT CTG ATC TAT TCT GAT GGC TTC AGC AAG GAG   90
16     W   T   A   Q   Y   L   I   Y   S   D   G   F   S   K   E   30

91    CAC AAG CAG TCG TTC ATC TCA TGG CTT GTG GCA ATT GTC ACC CTT   135
31     H   K   Q   S   F   I   S   W   L   V   A   I   V   T   L   45

136   ACC ATC TAC ACA GGA TGG ATG CAC ATA CTG ATA GGG TTG ACT GTG   180
```

```
 46    T   I   Y   T   G   W   M   H   I   L   I   G   L   T   V    60

181   GCC AGT TTG TTC AGT AGG ACA TGC CTC TGC ATC CTC GTA GCC ATA   225
 61    A   S   L   F   S   R   T   C   L   C   I   L   V   A   I    75

226   TGG GGA ACG GTC TTC CTG CCG CCA AAG CCA GTG TTG TGG AAT GCC   270
 76    W   G   T   V   F   L   P   P   K   P   V   L   W   N   A    90

271   TTC TGC AAA AGC TGG ATA TTC CAG ACC TGG CGC GAG TAT TTC CAA   315
 91    F   C   K   S   W   I   F   Q   T   W   R   E   Y   F   Q   105

316   TTC AGC TAC TTG AAT GAA GCG GTG CTG GAC CCC AAG AAG AAG TAC   360
106    F   S   Y   L   N   E   A   V   L   D   P   K   K   K   Y   120

361   ATA TTC ACC GAG TTC CCT CAT GGA GTC TTC CCT CTT AGC GAA CTA   405
121    I   F   T   E   F   P   H   G   V   F   P   L   S   E   L   135

406   GTT GCA GGA ACG CTG TGT CAA GCC ATC TGG CCG GAC TTC AGC ATA   450
136    V   A   G   T   L   C   Q   A   I   W   P   D   F   S   I   150

451   TAT TCA GTG GCA GCC AGC AGT GTG TAC AGC ATT CCC TTC TGG AGG   495
151    Y   S   V   A   A   S   S   V   Y   S   I   P   F   W   R   165

496   CAC TTC ATC GCT TGG CTT GGT GCA GTC CCA GCT ACA GCA GGC AAT   540
166    H   F   I   A   W   L   G   A   V   P   A   T   A   G   N   180

541   TTC AAG AAA ATG ATG AAG AGG GGA AGC TTG GCC GTC ATT GTG GGA   585
181    F   K   K   M   M   K   R   G   S   L   A   V   I   V   G   195

586   GGC ATT GCT GAG ATG TAC ATG CAG CAC AAG CGC AAG GAA CGC ATC   630
196    G   I   A   E   M   Y   M   Q   H   K   R   K   E   R   I   210

631   AAG CTG CTA GAC AGG AAA GGG TTT GTC AAG ATT GCT GTT GAG GAG   675
211    K   L   L   D   R   K   G   F   V   K   I   A   V   E   E   225

676   GGC CTG GAT GGA GGG ATC ATA CCT GTG TAT CAC TTC GGC AAC ACT   720
226    G   L   D   G   G   I   I   P   V   Y   H   F   G   N   T   240

721   CAG GTG TTC GAC TAT TGG CCT CAG TCC TTT GAG AAA TTT GCT CGC   765
241    Q   V   F   D   Y   W   P   Q   S   F   E   K   F   A   R   255

766   AAG AAC AGA GTG GGT GTG GGA TTC TTG GTC GGA CGA TGG GGA ACT   810
256    K   N   R   V   G   V   G   F   L   V   G   R   W   G   T   270

811   CCT GTG CCG CGC AAG GTG CCC TTG TAC ATG GTG AGT GGC AAG CCC   855
271    P   V   P   R   K   V   P   L   Y   M   V   S   G   K   P   285

856   ATT CCA GTG CCC AAG GTC GAC AAG AAT GAT ACC GAG AAG TTC AAC   900
286    I   P   V   P   K   V   D   K   N   D   T   E   K   F   N   300

901   CAG ACT GTA GAT GAG ATT CAC GCT GAA GTG GTG CAA CAG CTG CAG   945
301    Q   T   V   D   E   I   H   A   E   V   V   Q   Q   L   Q   315

946   GAC CTC TAT GAC AGA CAC AAG GCA AGC TAT GGC TGG GAG AAC AGA   990
316    D   L   Y   D   R   H   K   A   S   Y   G   W   E   N   R   330

991   CCC CTC AAG ATT GAG TGA                                    1008
331    P   L   K   I   E   *
```

C)

```
1     ATG CAC ATA CTG ATA GGG TTG ACT GTG GCC AGT TTG TTC AGT AGG   45
1      M   H   I   L   I   G   L   T   V   A   S   L   F   S   R    15

46    ACA TGC CTC TGC ATC CTC GTA GCC ATA TGG GGA ACG GTC TTC CTG   90
16     T   C   L   C   I   L   V   A   I   W   G   T   V   F   L    30

91    CCG CCA AAG CCA GTG TTG TGG AAT GCC TTC TGC AAA AGC TGG ATA   135
31     P   P   K   P   V   L   W   N   A   F   C   K   S   W   I    45

136   TTC CAG ACC TGG CGC GAG TAT TTC CAA TTC AGC TAC TTG AAT GAA   180
46     F   Q   T   W   R   E   Y   F   Q   F   S   Y   L   N   E    60

181   GCG GTG CTG GAC CCC AAG AAG AAG TAC ATA TTC ACC GAG TTC CCT   225
61     A   V   L   D   P   K   K   K   Y   I   F   T   [redacted]    75

226   CAT GGA GTC TTC CCT CTT AGC GAA CTA GTT GCA GGA ACG CTG TGT   270
76     [redacted]   V   F   P   L   S   E   L   V   A   G   T   L   C    90

271   CAA GCC ATC TGG CCG GAC TTC AGC ATA TAT TCA GTG GCA GCC AGC   315
91     Q   A   I   W   P   D   F   S   I   Y   S   V   A   A   S   105

316   AGT GTG TAC AGC ATT CCC TTC TGG AGG CAC TTC ATC GCT TGG CTT   360
106    S   V   Y   S   I   P   F   W   R   H   F   I   A   W   L   120

361   GGT GCA GTC CCA GCT ACA GCA GGC AAT TTC AAG AAA ATG ATG AAG   405
121    G   A   V   P   A   T   A   G   N   F   K   K   M   M   K   135

406   AGG GGA AGC TTG GCC GTC ATT GTG GGA GGC ATT GCT GAG ATG TAC   450
136    R   G   S   L   A   V   I   V   G   G   I   A   E   M   Y   150

451   ATG CAG CAC AAG CGC AAG GAA CGC ATC AAG CTG CTA GAC AGG AAA   495
151    M   Q   H   K   R   K   E   R   I   K   L   L   D   [redacted]   165

496   GGG TTT GTC AAG ATT GCT GTT GAG GAG GCC CTG GAT GGA GGG ATC   540
166    [redacted]   E   [redacted]   G   [redacted]   180

541   ATA CCT GTG TAT CAC TTC GGC AAC ACT CAG GTG TTC GAC TAT TGG   585
181    [redacted]   T   Q   V   F   D   Y   W   195

586   CCT CAG TCC TTT GAG AAA TTT GCT CGC AAG AAC AGA GTG GGT GTG   630
196    P   Q   S   F   E   K   F   A   R   K   N   R   V   G   V   210

631   GGA TTC TTG GTC GGA CGA TGG GGA ACT CCT GTG CCG CGC AAG GTG   675
211    G   F   L   V   G   R   W   G   T   P   V   P   R   K   V   225

676   CCC TTG TAC ATG GTG AGT GGC AAG CCC ATT CCA GTG CCC AAG GTC   720
226    P   L   Y   M   V   S   G   K   P   I   P   V   P   K   V   240

721   GAC AAG AAT GAT ACC GAG AAG TTC AAC CAG ACT GTA GAT GAG ATT   765
241    D   K   N   D   T   E   K   F   N   Q   T   V   D   E   I   255

766   CAC GCT GAA GTG GTG CAA CAG CTG CAG GAC CTC TAT GAC AGA CAC   810
256    H   A   E   V   V   Q   Q   L   Q   D   L   Y   D   R   H   270

811   AAG GCA AGC TAT GGC TGG GAG AAC AGA CCC CTC AAG ATT GAG TGA   855
271    K   A   S   Y   G   W   E   N   R   P   L   K   I   E   *   285
```

Figure 2.19 Nucleotide and deduced amino acid sequences of CsDGAT. A) Schematic representation of the different versions of *CsDGAT* depending on the presence or abscense of the putative intron. Also, two different forms of the 3' UTR are shown. B) Nucleotide and amino acid sequence of *CsDGAT* ORF starting at the first ATG identified and including intron removal

and C) Nucleotide and amino acid sequence of *CsDGAT* ORF starting at the second ATG identified, considering intron as part of the 5' UTR sequence.
Motifs HPHG and RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q) highlighted in gray

From the mapping of the 5' and 3' UTRs, and the complete sequencing of the *CsDGAT* cDNA found in this study, where two different 3' UTR were identified, we propose two versions of *CsDGAT* cDNA, based on their total length and varying on the span of their 3' UTRs, as shown in Figure 2.20. These two versions of *CsDGAT* cDNA were considered based primarily on the differences of their total length. Further studies are necessary to elucidate whether there are in fact two 5' UTRs and therefore two ORFs.

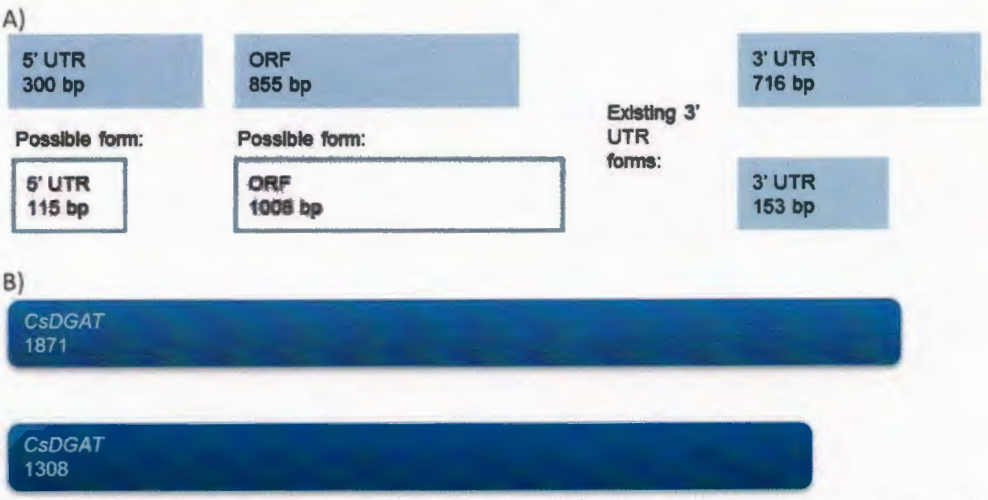


Figure 2.20 Complete structure of *CsDGAT* cDNAs, and schematic representation of its full mapping. A) Existing and possible forms of 5' UTR and ORF for *CsDGAT*. Both version of 3' UTR are shown. B) Based on length, two existing forms of *CsDGAT*. Length influenced by 3' UTR lenghts

2.3.9 Phylogenetic analysis of *CsDGAT*

A phylogenetic tree was constructed using CsDGAT and DGAT amino acidic sequences from plant and microalgae. Sequences from plants have been characterized (Fig. 2.21). Three main clades were formed: One clade grouped all the DGAT type 2 sequences from plants and microalgae. The two sequences from plants, *A. thaliana* and *V. fordii*; have been previously characterized as DGAT type 2. The sequences from microalgae, such as *C. reinhardtii* and *V. carteri* are annotated on the public databases as DGAT type 2. CsDGAT was grouped on this clade, showing to be related to this type of DGAT (CsDGAT 2). A second clade consisted on DGAT type 3 sequences. A DGAT type 3 entry from *A. hypogaea* (the first DGAT type 3 to be

discovered and characterized) is part of this clade, as well as another sequence belonging to *R. communis*. Finally, DGAT type 1 sequences formed a third clade. One functionally characterized sequence DGAT type 1 from *A. thaliana* is grouped here with another sequence from *Chlorella sp.*

As observed, DGAT types 1 and 3 from plants are separated from their type 2 counterpart. No microalgae sequence grouped with a DGAT3 and only one from *Chlorella sp.* was similar to a DGAT1. The clade formed with DGAT2 sequences is separated from the rest of the sequences as a root is formed, indicating distance between them.

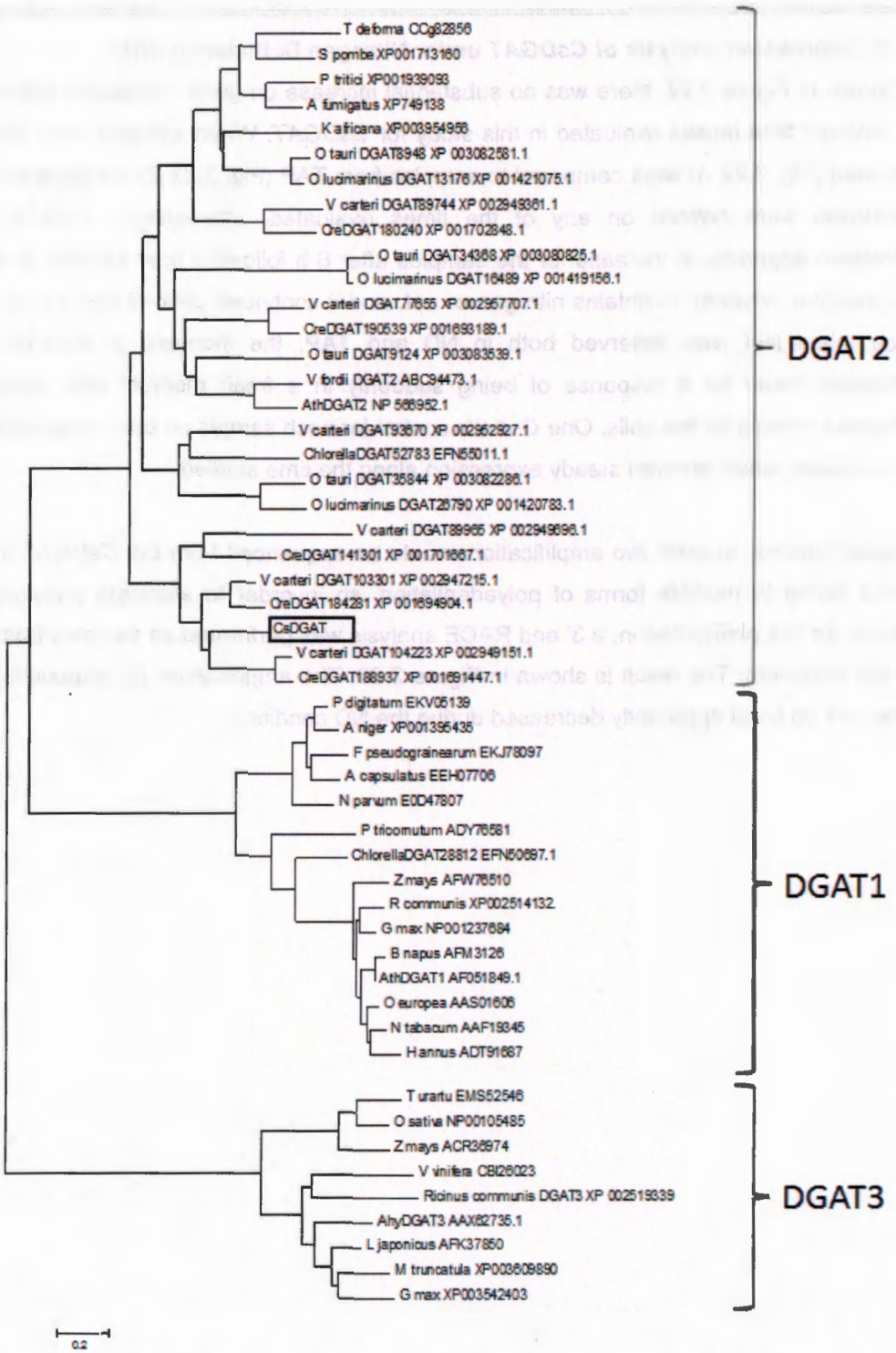


Figure 2.21 Phylogenetic tree of DGAT sequences from microalgae, fungi and plants. Each type of DGAT (1, 2 and 3) forms one clade. CsDGAT (marked with a square) is grouped on DGAT type 2 clade

2.3.10 Expression analysis of *CsDGAT* under Nitrogen Defficiency (ND)

As shown in Figure 2.22, there was no substantial increase on gene expression along the different time lapses evaluated in this study for *CsDGAT*. When samples from ND treatment (Fig. 2.22 A) were compared to samples from TAP (Fig. 2.22 B), no apparent differences were noticed on any of the times evaluated. Interestingly, *CsDGAT* expression appeared to increase for the samples after 6 h following their transfer to a new medium, whether it contains nitrogen or not, and it continued until 48 hours later. Since this effect was observed both in ND and TAP, the increase in *CsDGAT* expression could be a response of being suddenly in a fresh medium with more nutrients available for the cells. One *CsActin* control for each sample on each treatment was included, which showed steady expression along the time studied.

As stated before, at least two amplifications could be sequenced from the *CsDGAT* 3' termini owing to multiple forms of polyadenilation, so in order to elucidate possible reasons for this phenomenon, a 3' end RACE analysis was performed on samples from the ND treatment. The result is shown in Figure 2.23. The amplification corresponding to the 384 bp band apparently decreased during the ND condition.

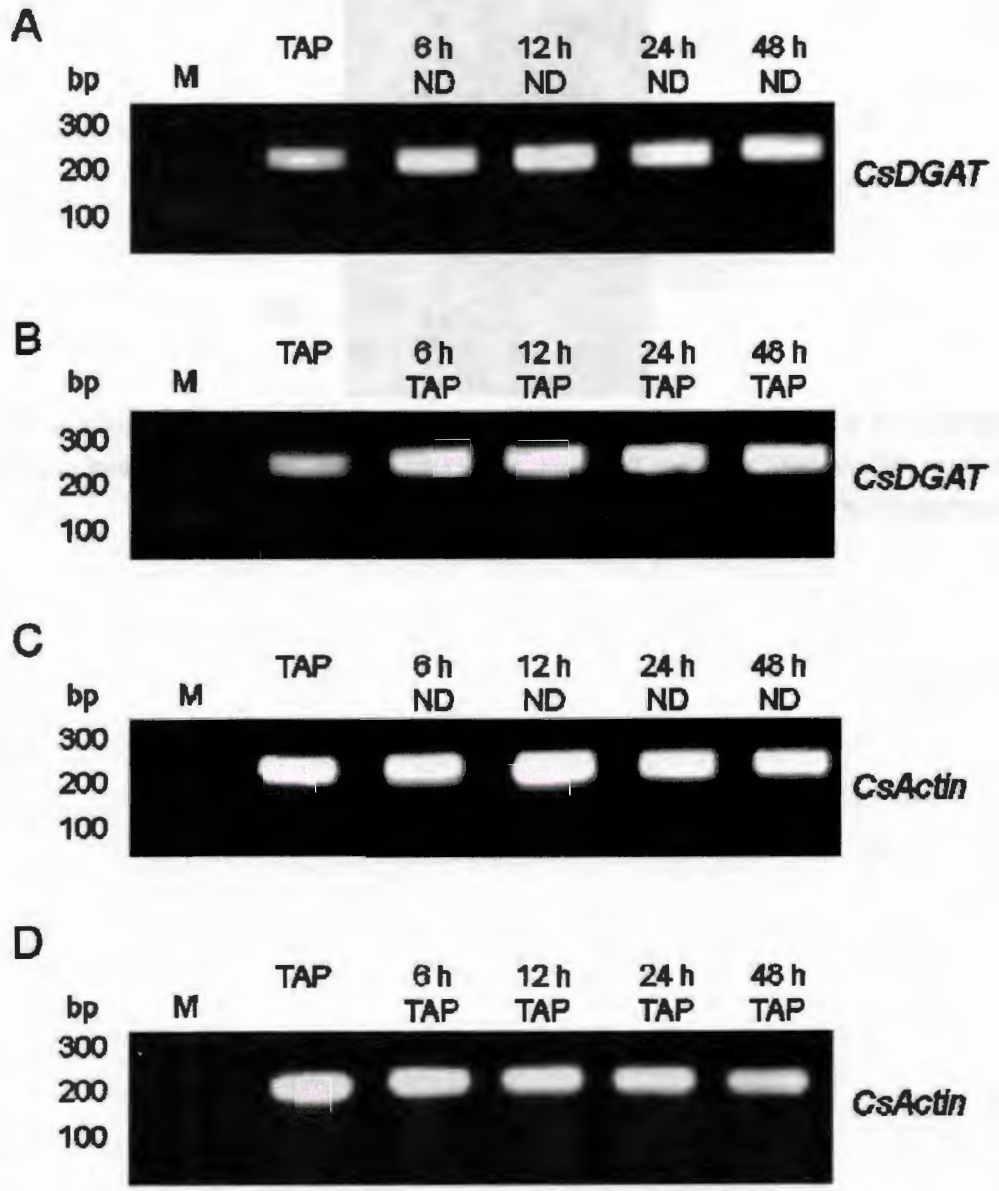


Figure 2.22 Expression analysis of *CsDGAT* over a time course under nitrogen deficiency (ND) or on complete TAP medium (TAP). The first line corresponds to a sample from TAP at day 10 of cultivation. A) *CsDGAT* expression after 6, 12, 24 and 48 h under ND treatment. B) *CsDGAT* expression after 6, 12, 24 and 48 h under complete TAP medium, as a control. *CsActin* expression after 6, 12, 24 and 48 h under ND treatment, as a positive control for samples under ND treatment. D) *CsActin* expression after 6, 12, 24 and 48 h under complete TAP medium, as a positive control on complete TAP medium

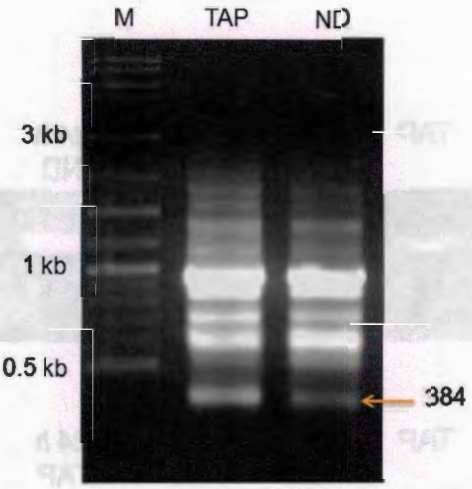


Figure 2.23 3' RACE analysis of CsDGAT under nitrogen deficiency. TAP: complete TAP medium; ND: nitrogen deficiency treatment. Arrow shows a band of 384 corresponding to the short version of 3' end. Molecular marker 2-log (NEB)

2.4 DISCUSSION

2.4.1 *C. saccharophila* harbors in its genome at least one homologue of a type 2 *DGAT* gene

The green microalga *C. saccharophila* has been shown to be a promising source to produce biodiesel, owing to its high biomass and TAGs productivities [Herrera-Valencia *et al,* 2011]. However, this microalga is not as extensively studied as *C. reinhardtii* and it has not been subjected to large genomic projects; the complete genome sequence of *C. saccharophila* has not been obtained and there is limited information in public databases regarding its genome and the genes it contains. Since DGAT is an enzyme that catalyzes the last step in TAG formation, a strategy to isolate and sequence a homologue of a *DGAT* gene in *C. saccharophila* was used.

Using characterized DGAT type 1, 2 and 3 sequences as queries, only DGAT type 2 sequences could be found in several green microalgae by means of searching bioinformatic databases. To date, both functional and evolutionary studies on these genes have found only DGAT type 2 on these microorganisms. La Russa *et al.* [2012] performed an early bioinformatic and functional analysis of these genes on *C. reinhardtii* using putative sequences reported on JGI. These authors concluded that only DGAT type 2 were found according to their analysis, and designed over-expression cassettes for three *CrDGAT* type 2 genes and transformed *C. reinhardtii.* Even though an increase on mRNA levels between 1.7 and 29.1 times was found for *CrDGAT* in transgenic lines in comparison to the wild type, no increase on TAGs levels was noted. Although importantly, it should be noted that the study by these authors lacked the mapping of the full cDNA sequence, including 5' and 3' UTRs, so all results were obtained using the predicted ORF from the NCBI database.

In the present study, the complete cDNA of a homologue of DGAT type 2 was obtained from *C. saccharophila* by a combination of degenerate RT-PCR and RLM-RACE, an ORF was predicted and it was named *CsDGAT*. BLAST searches confirmed that *CsDGAT* was part of the DGAT family, and phylogenetic analysis indicated this gene to be more related to DGAT type 2. Similar results were found for other green microalgae such as *O. tauri* and *C. reinhardtii* [Wagner *et al.* 2010; La Russa *et al.* 2012; Deng *et al.* 2012]. Three DGAT sequences were isolated for *O. tauri*, while five sequences were isolated for *C. reinhardtii*, and all of them belonged to the DGAT type 2 family. The approach used by Wagner *et al.* [2010] and La Russa *et al.* [2012] to find and isolate

DGAT genes on these microalgae, was to employ annotated sequences as queries and to perform BLAST searches on the genomes of *O. tauri;* their bioinformatic analyses showed two DGAT type 2 genes found for *O. tauri*, but in the heterologous expression on *S. cerevisiae* only one gene (named *OtDGAT2A)* proved to have DGAT function. In the case of *C. reinhardtii,* five DGAT type 2 sequences were isolated *in silico*. Of these five sequences, two were not correlated to lipid content when they were silenced with RNAi constructs, implying a diverse internal mechanism for the regulation of lipid content. The other three genes were positively correlated to TAG content, in silencing and overexpression analyses [Deng *et al.* 2012]. For *C. saccharophila* is not yet know what could be the case. This thesis is the first study where a *CsDGAT* gene has been isolated and cloned from *C.saccharophila*, thereby establishing the basis for future functional analysis for this gene, similarly to La Russa *et al.* [2012] whom started by finding candidate genes based on another work performed by Miller *et al.* [2010]. These authors correlated the upregulation of three *DGAT* genes from *C. reinhardtii* to nitrogen starvation, and concluded that these three sequences were important for lipid and TAGs accumulation, therefore settling an important basis for future works. In *C. saccharophila*, however, since the bioinformatics tools available for this microalga are still very limited to this date, other molecular approaches may be used to determine whether other DGAT sequences can be found for this microalga.

For other green microalgae, more than one DGAT sequence has been found, and so it can be the same case for *C. saccharophila*. The complete sequence of its genome would allow this task to be completed more easily. If not, an analysis of total mRNA constructing a library could be performed. As more and more genomes of microalgae are completed, more bioinformatics tools and information are available to work with. With the basis established on this thesis, in conjunction with data on public databases, it will be possible to isolate and characterize further this and other *DGAT* sequences on *C. saccharophila.* Turchetto *et al.* [2011] made a comprehensive evolutionary study on DGAT's, including several sequences from organisms ranging from fungi, yeasts, to microalgae, plants and animals. These authors suggested that DGAT type 1 and 2 probably evolved separately, and converged on their function. This can be observed on the little similarity between their sequences, and on the phylogenetic trees constructed using these sequences, where DGAT types 1 and 2 appear to be clearly separated on its common ancestor. Interestingly, it was reported by these authors that normally yeasts, animals and other organisms only have one copy of each DGAT gene, except for plants and microalgae which can have from two or three copies (e.g. *A. thaliana*) or five (e.g. *C. reinhardtii*). In fact, these authors found three to five copies of DGAT2 on

microalgae such as *Chlorella sp, O. tauri, O. lucimarinus* and *C. reinhardtii*, which could have been probably originated from duplication events. Turchetto *et al.* [2011] also noted the divergence of sequence in the motif HPHG (important for DGAT) in microalgae, which it was found as EPHS, a motif sequence also present in plants. This motif was encountered on *CsDGAT* as FPHG, a possible alternate version of this motif. Other variants of this motif were described by Liu *et al.* [2011] on fungi, plants and animals, where the first aminoacid does not remain conserved between these organisms, but these authors stated that the aminoacids P and H on the second and third position respectively are fundamental for the functionality of a DGAT type 2, since mutations on the Histidine abolish the enzyme activity.

2.4.2 Cloning the complete cDNA and mapping the 5' and 3' UTRs of *CsDGAT* cDNA

Using a set of degenerate primers, a fragment of a putative *DGAT* gene was amplified by RT-PCR, cloned and sequenced. From this sequence specific primers were design in order to fully map its cDNA. By means of RLM-RACE and RACE, the complete sequence of *CsDGAT* cDNA was obtained. FGenesH was then used to search for possible ORFs in *CsDGAT* cDNA sequence and two apparent versions were considered.

In one version, a putative intron (located 179 bp downstream the transcription initiation site) was apparently retained within its sequence, which codes for an early stop codon resulting in a truncated protein. A set of primers flanking this putative intron were designed in order to corroborate the existence of such sequence. At least two different mRNAs were detected with this strategy, thus proving that another sequence besides the expected one of ~400 bp corresponding to the target sequence of those primers was also present. A similar approach was performed by Ner-Gaon *et al.* [2004] where the phenomenon of intron retention was studied in *A. thaliana*. These authors reported that at least 2% of the total genes in *A. thaliana* presented this condition and many (if not all) of them can be subjected to regulation via intron retention caused by environmental stresses, or changes in development. Microalgae are subjected to stressful and changing environments, and so they must be able to adapt to such possible alterations. The presence of an intron could be an important part of a stress-induced regulation, as it was observed in durum wheat [Mastrangelo *et al.* 2005], where an expression library was screened during low temperature stress. Two genes were found whose mature mRNAs contained introns that acted as targets for regulation. Their retention was upregulated during this stress, and the transcription factors that recognized them were either degraded or blocked due to cold stress.

Further studies are necessary to investigate whether the intron found in *CsDGAT* plays indeed a role in some post-transcriptional regulation of this gene in *C. saccharophila*.

To the date of this thesis, there are few studies focused on untranslated regions on mRNA from microalgae. Kuo *et al.* [2013], recently studied the transcriptome of the alga *Eutreptiella sp.* and discovered that 5' UTRs tend to be short (~21 bp) in this organism. In plants these sequences are very variable, ranging from dozens of nucleotides to hundreds [Kochetov *et al.* 2002]. However, they determined that the most common length of 5' UTR vary from 50 to 150 nucleotides. The 5' UTR from *C. saccharophila* was found to vary between 150 and 300 bp, depending on the putative intron found if it is indeed excised from the final sequence. So this sequence can be considered longer than usual when compared to both plants and one alga.

In *C. reinhardtii*, sequences found on the 5' of genes *rbcL* and *atpB* have been implied to be crucial on mRNA stability and control of expression [Anthonisen *et al.* 2001]. Mutations on these sequences, when coupled with reporter genes such as GUS; affected negatively the expression of transcripts to a 60-70%. Is it unknown whether these regions act only at a structural level or indeed work as sites for protein bingind that confer stability to mRNAs. A more recent work focused on structural importante [Leinass 2008] found that the addition of bases on the 5' UTR of *rbcL*, instead of mutating the existing ones; also diminished mRNA life and stability. It was unclear however if it is necessary to modify that many bases or just one nucleotide could affect the expression. A similar condition could be the case for *C. saccharophila,* where different sequences found on the two UTRs found may decide the expression of either form. As shown by Leinass, the presence of more nucleotides on these regions not necessarily contribute to the stability of the transcript.

Unfortunately, until now there are no studies centered on massive analysis of 5' UTRs in microalgae at a level comparable to plants or animals so a comparison between *C. saccharophila* and several groups of other green microalgae is not yet possible.

The 3' RACE technique permitted the identification and sequencing of the 3' end of *CsDGAT* cDNA. Two fragments were amplified by PCR, suggesting two possible forms (lengths) of the same mRNA, this difference was encountered on the 3' UTR and consisted of a 563 bp difference between a long 3' UTR (716 bp) and a short one (153 bp). This could be the result of alternative splicing, an event in which a single gene can create multiple mRNA transcripts [Barbazuk 2008]. Alternative splicing has been observed in mammals such as human and had been related to message stability and translation efficiency, as well as the diversity of proteins [Andreassi *et al.* 2009].

Another possibility for the existence of these different mRNAs is the reprogramming of 3' UTRs by alternative polyadenylation. A key feature of transcript regulation relies on the 3' UTRs, as these regions are susceptible to cis-acting elements involved on stability and expression. These events have been found to occur on both plant [Xing and Li 2010] and animal [Ji and Tian 2009] cells. In animal cells it was observed that the 3' UTR length was correlated to the state of pluripotent cells, and that alternative polyadenylation was highly regulated during embryonic stage. Although in plant cells this mechanism is still poorly understood, it is known that the transcriptomes related to flowering time control and stress responses are regulated by this kind of event, which takes place on the 3' UTR. Major differences were encountered between the two 3' UTR mapped for *CsDGAT* in terms of both length and presence/absence of regulatory sequences, suggesting a possible event of this nature.

For the alternate polyadenilation event to occur, the cell needs different recognition sites, named polyadenilation signals which are recognized by a group of polyadenilation factors [Shen *et al.* 2008]. Several motifs have been described that fulfill this role [Zhao *et al.* 1999]. These motifs include the canonical sequence AAUAAA found upstream the poly A as a near upstream element, which has been identified in yeast, plant and mammalian cells [Shen *et al.* 2008]. In *C. reinhardtii* this motif was described but its signal was found to be weaker than the motif UGUAA; this motif was found on 52% of the analysed sequences and it was suggested that it is positively correlated with higher gene expression levels [Shen *et al.* 2008]. It is important to note that half of all genes studied by Shen *et al.* [2008] presented at least two polyadenylation signals, a characteristic apparently present on *CsDGAT*. When the 3' UTR of *CsDGAT* was analyzed *in silico*, two probable/predicted/putative polyadenylation signals were found: UGUAA and AAUAAA. This discovery agrees with other reports [Shen *et al.* 2008; Ji and Tian 2009], where it is stated that the existence of two or more signals for polyadenylation is correlated to alternative processing of such event.

2.4.3 Cloning of a *CsActin* gene by degenerate PCR

The partial sequence of an *Actin* gene on *C. saccharophila* was amplified with a set of degenerate primers. This strategy differs from Sugase *et al.* [1996], whom isolated an actin gene from *C. reinhardtii* (*CrActin*). Their strategy was to use an *actin* probe from *Aedes aegypti* on a genomic library of *C. reinhardtii* in combination with Southern blot. They found only one copy of this gene, the same used in the present thesis as query for the BLASTx analyses.

For *C. saccharophila* only one *CsActin* gene was isolated, as was the case for *C. reinhardtii*. For other green microalgae, such as *V. carteri*, at least two genes annotated as actin can be found on public databases. For *C. saccharophila* it is not yet known the total number of actin genes present in its genome, and so a similar approach as the one by Sugase *et al.* [1996] could be carried our in order to elucidate this.

CsActin was not fully mapped in the present study. The reason was because it was not necessary to know all regulatory sequences, since it was employed mainly as control for PCR and expression analyses. However, it would be interesting to fully map its sequence and add it to the information contained in public databases with the aim of facilitating future analyses, such as a tool to find any other possible gene or to generate more positive controls for other assays.

2.4.4 *CsDGAT* expression under Nitrogen Defficiency

It has been demostrated that nitrogen deficiency triggers the accumulation of TAGs in microalgae [Guihéneuf 2011; Herrera-Valencia *et al.* 2011]. Being deprived of a major nutrient, the cell is forced to make-up large reservoirs of energy in the form of TAG, a very energetic molecule, in order to maintain its homeostasis. DGAT is a key enzyme in the formation of TAGs and so, it would be expected that the genes coding for these enzymes be part of some regulation when the cell is under a stress that involves the accumulation of TAGs. For example, when the bacteria *Mycobacterium tuberculosis* were placed under stress conditions such as hypoxia, static and acidic conditions, the TAG levels were increased [Sirakova *et al.* 2006]. Interestingly, 15 putative diacylglycerol acyltransferase genes (named *tg)* were found but only one (*tg1*) was correlated with TAGs accumulation provoked by stressful conditions [Sirakova *et al.* 2006].

The nature of microalgae allows them to strive on a very wide variety of environments. Their ability to adapt to the environmental changes is reflected also on its gene expression; as a result, different expression responses are achieved depending on the conditions or stress applied. It is known that severe nutrient deficiency can be accounted as an effective way to affect both cellular growth and lipid production in a broad way [Schenk *et al.* 2012]. This approach has been recently and widely used as a method to increase total lipid production, and interesting results were achieved. Herrera-Valencia *et al.* [2011] found that total lipid content in *C. saccharophila* could reach ~65% of its dry weight when subjected to nitrogen deficiency stress, compared to the control treatment which showed 40%. It was also noted that salinity treatment could increase these levels, reaching up to more than 55% of its dry weight.

Since TAG formation pathway in microalgae is believed to be similar to that of plants, many of the main participating enzymes in this pathway are being characterized in these microorganisms. DGAT enzyme is well known for catalyzing the last step in TAG formation in plants, however, studies in microalgae are still very limited. Recently, Guihéneuf *et al.* [2011] characterized a DGAT type 1 on the diatom *Phaeodactylum tricornutum* (*PtDGAT1*) and found that its expression increased under nitrogen deficiency stress.

In the present thesis, a different result was found. The expression of *CsDGAT* was constitutive during nitrogen deficiency. Since this measure was performed 48 hours after the stress application, a more detailed analysis was considered. A time course experiment was carried out, following a similar approach as Boyle *et al.* [2012]. These authors designed and performed a stress application over a timer course: 8, 12, 24 and 48 hours. Since the first hours, an increase on the expression was achieved and it remained on that state for the rest of the experiment. No decline on transcript accumulation over time was noticed. In the present thesis however, a small increase in *CsDGAT* expression was notice in both treatment and control samples, so it is currently unknown whether the accumulation on mRNA was due to treatments or a phenomenon present in regular cell growth. Both Boyle *et al.* [2012] and Guihéneuf *et al.* [2011] did not clarified this event since control samples were considered only at the beginning of the experiment and not on every time course.

Guihéneuf *et al.* [2011] and Boyle *et al.* [2012] did carried out a more precise and detailed analysis of mRNA abundance which was performed in order to stablish a difference between control and treatment samples. Methodologies such as QPCR, (employed by Guihéneuf *et al.* 2011) or estimation of relative RNA abundance calculated by QPCR (employed by Boyle *et al.* 2012) could be applied to *C. saccharophila* in order to clarify the outcomes of ND treatment on *CsDGAT* expression and to determine a possible role of post-transcriptional regulation of this gene.

Based on these results, we found that *CsDGAT* was expressed constitutively under the stress conditions evaluated. A possible explanation could be that this gene is not involved in TAG formation and that there must be at least another *DGAT* gene involved in this process in *C. saccharophila*. Another possible explanation could be that this gene is indeed involve in TAG formation, and since TAGs are not normally accumulated, this gene may be regulated at another level, for example it could be under post-transcriptional regulation, either at its 5' or 3' end. A third explanation could be that due to the use of a qualitative RT-PCR it was not possible to distinguish between subtle differences in gene expression, thus real time RT-PCR should be used

in further studies to quantitative analyze the expression of *CsDGAT* under the treatments evaluated in this study.

An interesting result has been found on a *DGAT* gene from the tree *Ricinus communis* (*RcDGAT*) [He *et al.* 2004]. These authors found a considerable difference between the abundance of mRNA and the protein levels. Also, the expression of *RcDGAT* was strongly regulated by growth cycles and was tissued dependent. Similarly, the authors found signals of post-transcriptional regulation on both UTRs, which coupled with experimental data, suggested this type of regulation for *RcDGAT*. To investigate the roles of the UTRs found in *CsDGAT* future studies could carry out a similar assay to the one used by Merritt *et al.* [2008]. These authors designed a gene construction using UTRs from *C. elegans* and the reporter gene GFP, and were able to detect temporal and spatial regulation of this gene according to the UTRs used. Applying this same strategy, it could be determined the importance of the putative regulatory sequences found on both 5' and 3' UTRs of *CsDGAT*.

Evidence showing that *DGAT* genes can undergo regulation was found in the study by Boyle *et al.* [2012], where a promoter-binding protein domain transcription factor named SQUAMOSA was correlated with the increase in gene expression and thus it was proposed as a candidate regulator of the nitrogen deficiency responses. However, it is yet unknown the details about its interaction with the *DGAT* gene.

It has been observed that *DGAT* genes respond differentialy to nutrient limitation on a temporal manner [Miller *et al.* 2010; Msanne *et al.* 2012]. In *C. reinhardtii, DGAT1* was upregulated during nitrogen deficiency, but other *DGAT* genes showed little or no response to this condition. Msanne *et al.* [2012] suggested that the expression of all *DGAT* genes during ND is time dependent. For *C. reinhardtii,* they argue that *DGTT1* may be expressed on the early stages of TAG accumulation and that other *DGAT* genes participate on the following steps. A similar scenario could be the case for *CsDGAT*. We found that this gene was expressed constitutively during ND, thus it could be possible that the version of *CsDGAT* found in this study is not as responsive to ND as other genes that *C. saccharophila* may harbor in its genome, as it has been seen on *C. reinhardtii*.

2.5 REFERENCES

Andreassi C. and A. Riccio (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. Trends in Cell Biology 19(9):465-474

Anthonisen I.L., M. Salvador and U. Klein (2001). Specific sequence elements in the 5' untranslated regions of rbcL and atpB gene mRNas stabilize transcripts in the chloroplast of *Chlamydomonas reinhardtii*. RNA New York Ny 7:1024-1033

Barbazuk W.B., Y. Fu, and K.M. McGinnis (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Research 18(9):1381-1392.

Blanc G., G. Duncan, I. Agarkova, M. Borodovsky, J. Gurnon, A. Kuo, E. Lindquist, S. Lucas, J. Pangilinan, J. Polle, A. Salamov, A. Terry, T. Yamada, D. D. Dunigan, I.V. Grigoriev, J.M. Claverie and J.L. Van Etten. (2010). The *Chlorella variabilis* NC64A Genome Reveals Adaptation to Photosymbiosis, Coevolution with Viruses, and Cryptic Sex. The Plant Cell, 22(9):2943-2955.

Boyle N., M.D. Page, B. Liu, I.K. Blaby, D. Casero, J. Kropat, S.J. Cokus, A. Hong-Hermesdorf, S.J. Karpowicz, S.D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman and S.S. Merchant (2012). Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in *Chlamydomonas*. The Journal of Biological Chemistry. 287(19):15811-15825.

Brown J. (1986). A catalog of splice junction and putative branch point sequences from plant introns. Nucleic Acids Research 14(24):9549-9559.

Burland G. (2000). DNASTAR´s Lasergene sequence analysis software. Methods in Molecular Biology Clifton. 132:71-91.

Cases S., S.J. Smith, Y.W. Zheng, H.M. Myers, S.R. Lear, E.Sande, S.Novak, C. Collins, C.B. Welch, A.J. Lusis, S.K. Erickson and R.V. Farese Jr. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences. USA 95(22):13018-13023.

Contreras-Pool P. (2010). Análisis del perfil de los ésteres metílicos obtenidos de dos microalgas verdes: *Chlamydomonas reinhardtii* y *Chlorella saccharophila*, y su variación en respuesta a condiciones de estrés. Tesis de Maestría. Centro de Investigación Científica de Yucatán.

Converti A., A.A. Casazza, E.Y. Ortiz, P. Perego and M. Del Borghi (2009). Effect of temperature and nitrogen concentration on the growth and lipid content of

Nannochloropsis oculata and *Chlorella vulgaris* for biodiesel production. Chemical Engineering and Processing, 48(6):1146–1151.

Dean A.P., D.C. Sigee, B. Estrada, and J.K. Pittman (2010). Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. Bioresource Technology, 101(12):4499–4507.

Deng X.D., B. Gu, Y.J. Li, X.W. Hu, J.C. Guo and X.W. Fei (2012). The Roles of acyl-CoA: Diacylglycerol Acyltransferase 2 Genes in the Biosynthesis of Triacylglycerols by the Green Algae *Chlamydomonas reinhardtii*. Molecular Plant 5(4):945-947

Dellaporta S.L, J. Wood and J.B Hicks (1983). A Plant DNA Minipreparation: Version II. Plant Molecular Biology Reporter 1:19-21.

Gouveia, L. and A.C. Oliveira (2009). Microalgae as a raw material for biofuels production. Journal of Industrial Microbiology & Biotechnology. 36(2):269-274.

Guihéneuf F., S. Leu, A. Zarka, I. Khozin-Goldberg, I. Khalilov and S. Boussiba (2011). Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom *Phaeodactylum tricornutum*. The FEBS Journal. 278:3651–3666.

Greenwell H.C., L.M.L. Laurens, R.J. Shields, R.W. Lovitt and K.J. Flynn (2010). Placing microalgae on the biofuels priority list: a review of the technological challenges. Journal of the Royal Society Interface. 7(406):703-726.

Hall T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98.

Harris E.H. (1989). The Chlamydomonas sourcebook: a comprehensive guide to biology and laboratory use. Academic Press, San Diego, 780pp.

He X., C. Turner, G. Chen, J-T. Lin and T. McKeon. (2004). Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids 39(4):311-318.

Hein J.J. (1990). A Unified Approach to Phylogenies and Alignments. Methods in Enzymology. 183:625-644.

Herrera-Valencia V.A., Y.P. Contreras-Pool, S.J. López-Adrián, S. Peraza-Echeverría and L.F. Barahona-Pérez (2011).The green microalga *Chlorella saccharophila* as a suitable source of oil for biodiesel production. Current Microbiology 63(2):151-157.

Higgins D. and P. Sharp (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. 73:237-244.

Ji Z. and B. Tian (2009). Reprogramming of 3' Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. PLoS ONE 4:13.

Kieffer E., W. Heller and D. Ernst (2000). A simple and efficient protocol for isolation of functional RNA from plant tissues rich in secundary metabolites. Plant Molecular Biology Reporter, 18:33-39.

Kochetov A., O. Symik, I. Rogozin, G. Glazko, M. Komarova and V. Shumnyi (2002). Context organization of mRNA 5'-untranslated regions of higher plants. Molecular Biology (Mosk). 36:649-56.

Kucho K., S. Yoshioka, F. Taniguchi, K. Ohyama and H. Fukuzawa. (2003). Cis-acting elements and DNA-binding proteins involved in CO_2-responsive transcriptional activation of Cah1 encoding a periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiology. 133(2):783-793.

Kuo R., H. Zhang, Y. Zhuang, L. Hannick and S. Lin (2013). Transcriptomic Study Reveals Widespread Spliced Leader Trans-Splicing, Short 59-UTRs and Potential Complex Carbon Fixation Mechanisms in the Euglenoid Alga *Eutreptiella* sp. Plos One. 8:4.

La Russa M., C. Bogena, A. Uhmeyera, A. Doebbea, E. Filipponeb, O. Krusea and J.H. Mussgnuga (2012). Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. Journal of Biotechnology, 162(1):13-20.

Larkin M., G. Blackshields, N. Brown, R. Chenna, P. McGettigan, H. McWilliam, F. Valentin, I. Wallace, A. Wilm, R. Lopez, J. Thompson, T. Gibson and D. Higgins (2007). Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947-2948.

Leinaas M.N. (2008). Destabilitation of Chlamydomonas reinhardtii rbcL transcripts by changes in their 5'UTR. Tesis de Maestría. University of Oslo.

Liu Q., R. Siloto, C. Snyder and R. Weselake (2011). Functional and Topological Analysis of Yeast Acyl-CoA:Diacylglycerol Acyltransferase 2, an Endoplasmic Reticulum Enzyme Essential for Triacylglycerol Biosynthesis. The Journal of Biological Chemistry, 286(15):13115-13126.

Lung S.C. and R. Weselake (2006). Diacylglycerol Acyltransferase: A Key Mediator of Plant Triacylglycerol Synthesis. Lipids, 41(12): 1073 - 1089.

Mastrangelo A. M., S. Belloni, S. Barilli, B. Ruperti, , N. Di Fonzo, A. M. Stanca and L. Cattivelli (2005). Low temperature promotes intron retention in two e-cor genes of durum wheat. Planta, 221(5):705-715.

Merchant S.S., S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, A. Terry, A. Salamov, L.K. Fritz-Laylin, L. Maréchal-Drouard, W.F. Marshall, L.H.

Qu, D.R. Nelson, A.A. Sanderfoot, M.H. Spalding, V.V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S.M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.L. Chen, V. Cognat, M.T. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa, D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P.A. Lefebvre, S.D. Lemaire, A.V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J.V. Moroney, J. Moseley, C. Napoli, A.M. Nedelcu, K. Niyogi, S.V. Novoselov, I.T. Paulsen, G. Pazour, S. Purton, J.P. Ral, D.M. Riaño-Pachón, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S.L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.J .Chen, M. Elias, K. Gendler, C. Hauser, M.R .Lamb, H. Ledford, J.C. Long, J. Minagawa, M.D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A.M. Terauchi, P. Yang, S. Ball, C. Bowler, C.L. Dieckmann, V.N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R.T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y.W. Huang, J. Jhaveri, Y. Luo, D. Martínez, W.C. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I.V. Grigoriev, D.S. Rokhsar and A.R. Grossman (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 12:245-250.

Merritt C., D. Rasoloson, D. Ko and G. Seydoux (2008). 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. Current Biology 18(19):1476-82.

Miller R., G. Wu, R.R. Deshpande, A. Vieler, K. Gärtner, X. Li, E.R. Moellering, S. Zäuner, A.J. Cornish, B. Liu, B. Bullard, B.B. Sears, M.H. Kuo, E.L. Hegg, Y. Shachar-Hill, S.H. Shiu and C. Benning (2010). Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. Plant Physiology. 154(4):1737–1752.

Msanne J., D. Xu, A. R. Konda, J. A. Casas-Mollano, T. Awada, E. B. Cahoon and H. Cerutti (2012). Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae *Chlamydomonas reinhardtii* and *Coccomyxa sp.* C-169. Phytochemistry 75:50–59.

Muller J., Y. Oma, L. Vallar, E. Friederich, O. Poch and B. Winsor (2005). Sequence and Comparative Genomic Analysis of Actin-related Proteins. Molecular Biology of the Cell. 16(12):5736-5748.

Nei M. and S. Kumar (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Ner-Gaon H., R. Halachmi, S. Savaldi-Goldstein, E. Rubin, R. Ophir and R. Fluhr (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. The Plant Journal 39(6):877-885.

Praveenkumar R., K. Shameera, G. Mahalakshmi, M.A. Akbarsha and N. Thajuddin (2012). Influence of nutrient deprivations on lipid accumulation in a dominant indigenous microalga *Chlorella sp.*, bum11008: Evaluation for biodiesel production. Biomass Bioenerg. 37:60–66.

Radakovits R., R. Jinkerson, A. Darzins and M. Posewitz (2010). Genetic Engineering of Algae for Enhanced Biofuel Production. Eukaryotic Cell. 9(4):486-501.

Rieping M. and F. Schoffl (1992). Synergistic effect of upstream sequences, CCAAT box elements, and sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. Molecular Genetics and Genomics. 231(2):226-232.

Rouge M. (2002). Counting Cells with a Hemacytometer. Updated: September 2 2002. Available:
http://www.vivo.colostate.edu/hbooks/pathphys/reprod/semeneval/hemacytometer.html

Saitou N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Shen Y., Y. Liu, L. Liu, C. Liang and Q. Li (2008). Unique Features of Nuclear mRNA Poly(A) Signals and Alternative Polyadenylation in *Chlamydomonas reinhardtii*. Genetics 179(1):167-176.

Sirakova T.D., V. Dubey, C. Deb, J. Daniel, T. Korotkova, B. Abomoelak and P. Kolattukudy (2006). Identification of a diacylglycerol acyltransferase gene involved in accumulation of triacylglycerol in *Mycobacterium tuberculosis* under stress. Microbiology Reading England 152(9):2717-2725.

Stone S.J., M.C. Levin and R.V. Farese (2006). Membrane topology and identification of key functional amino acid residues of murine Acyl-CoA: Diacylglycerol acyltransferase-2. Journal of Biological Chemistry, 281(52):40273-40282.

Sugase Y., M. Hirono, K. Kindle and R. Kamiya (1996). Cloning and characterization of the actin-encoding gene of *Chlamydomonas reinhardtii*. Gene 168(1):117-121.

Tamura K., J. Dudley, M. Nei and S. Kumar (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Thompson J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24:4876–4882.

Turchetto-Zolet A.C., F.S. Maraschin, G.L. De Morais, A. Cagliari, C. Andrade, M. Margis-Pinheiro and R. Margis (2011). Evolutionary view of acyl-CoA

diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. BMC Evolutionary Biology 11:263.

Wagner M., K. Hoppe, T. Czabany, M. Heilmann, G. Daumb, I. Feussner and M. Fulda (2010). Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (*DGAT2*) gene from the microalga *O. tauri*. Plant Physiology and Biochemistry 48(6):407-416.

Xing D. and Q. Li (2011). Alternative polyadenylation and gene expression regulation in plants. Wiley Interdisciplinary Reviews RNA 2(3):445-458.

Yeh K.L. and J.S. Chang (2011). Nitrogen starvation strategies and photobioreactor design for enhancing lipid production of a newly isolated microalga *Chlorella vulgaris* esp-31: Implications for biofuels. Biotechnol. J. 6(11):1358-1366.

Zhao J., L. Hyman and C. Moore (1999). Formation of mRNA 39 ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiology and Molecular Biology Reviews. 63(2):405-445.

CHAPTER III

TRANSFORMATION OF *Chlamydomonas reinhardtii* WITH *CsDGAT* GENE BY MICROPARTICLE BOMBARDMENT

3.1 INTRODUCTION

Diacilglycerol acyltransferase is an enzyme that participates in the committed and last step of TAGs formation, so it has been well accepted that the understanding of its function and its modification are key aspects to improve biofuel production. Several approaches have been followed to both study the function and to over-express *DGAT* genes with the aim to increase TAG levels. Wagner *et al.* [2010] were successful on expressing a DGAT type 2 from *Ostreococcus tauri* on a TAG deficient yeast strain. The same approach was followed by Guihéneuf *et al.* [2011] with the difference that they isolated and overexpressed a DGAT type 1 from *Phaeodactylum tricornutum*, a diatom. In both cases, the production of TAGs was restored on the mutant yeast, proving the function of this enzyme.

On another study, two distinct DGAT genes were cloned from *Brassica napus* cell cultures and then expressed on *Saccharomyces cerevisiae*. Not only a functional enzyme was produced, but also a three fold increase compared to wild types on the total levels of TAGs was achieved [Nykiforuk *et al.* 2002]. In contrast, three different DGAT type 2 sequences from *Chlamydomonas reinhardtii* were isolated and overexpressed on this same microalga. However, their efforts were unsuccessful when trying to increase TAG levels [La Russa *et al.* 2012]. These authors suggested a possible regulation at a post-transcriptional or post-translational level that inhibited the protein accumulation.

In this thesis a homologue gene of DGAT was isolated and characterized from *C. saccharophila* genome: *CsDGAT*, and three differente versions of this gene were proposed. Thus, it will be interesting to investigate the role of this gene in microalgal TAG biosynthesis. Recently, incorporation of foreing genes into *Chlorella* sp. cells was achieved by *Agrobacterium tumefaciens* using hygromycin as the selection agent and beta-glucuronidase (GUS) as reporter gene on plasmid pCAMBIA1304 [Cha *et al.* 2011]. However, to date there are no transformation protocols available for *C. saccharophila*. In contrast, transformation protocols have been developed for some microalgae like *C. reinhardtii*, some *Chlorella* species and some diatoms [Gong *et al.* 2011]. The plasmid pSP124 [Lumbreras *et al.* 1998] is a valuable tool to accomplish transformation of *C. reinhardtii,* although it has not been used in other microalgae. This plasmid carries the *ble* resistante gene which confers resistance to the antibiotics

phelomycin, zeocyn and bleomycin. Also, an important factor to achieve the overexpression of an endogenous gene is the use of a promoter that drives the expression efficiently, and the promoter *HSP70A* fused upstream to promoter *RBCS2* demonstrated to enhance transgene expression when placed together upstream other promoters, such as *RBCS2* or *βTUB* (tubuline) [Schroda *et al.* 2000].

Therefore, the purpose of this chapter was to obtain a gene construct with three versions of the *CsDGAT* gene under the control of the fusion promoter *HSP70A/RBCS2*, which is a strong constitutive promoter, in the plasmid pSP124S, and use it for nuclear transformation of *C. reinhardtii.*

3.2 MATERIALS AND METHODS

3.2.1 Cloning vector and oligonucleotides

CsDGAT gene was cloned in plasmid pCrGPDH3c (Fig. 3.1) kindly donated by MSc Melissa Casais. The plasmid had the following components:

- Plasmid pSP124S. The best version of pSP124 described in Lumbreras *et al.* (1998) had been additionaly modify to give pSP124S, which was obtain from the Chlamydomonas Center (www.chlamy.org). pSP124S has the plasmid pBluescript SK- as backbone which contains the bacterial selection gene for resistance to ampicillin (*amp*). For microalgae it uses the selection gene *ble*, which confers resistance to zeocin [Stevens *et al.* 1996]. It is 4,132 bp long and has been successfully used to transform *C. reinhardtii.*
- Fusion promoter HSP70A/RBCS2. Promoter HSP (Heat Shock Protein) confers high inducibility by itself and can be induced by both, heat or light. When fused with promoter RBCS2 (RuBisCO) it improves transgene expression greatly (Schroda *et al.* 2000).
- 5' UTR from *RBCS2*. Transformation rate is heavily influenced by the presence of this sequence. If neglected, this rate is severely decreased. Also, negative regulatory sequences have been found on positions -740 to -300. The removal of this portion has increased expression efficiency threefold (Lumbreras *et al.* 1998).
- 3' UTR from *RBCS2*. Although 3' UTR alone showed no real contribution for the expression of transgene (including regulatory sequences such as polyadenilation signals), its combination with 5' UTR proved to be fundamental for optimum transgene expression [Lumbreras *et al.* 1998; Eichler-Stahlberg *et al.* 2008].

- Introns 1 and 3 from RBCS2. The inclusion of these sequences has been reported to increase both expression and stability of genes in *Chlamydomonas*. Regulatory sequences (enhancer) are present in these regions (Lumbreras *et al.* 1998).

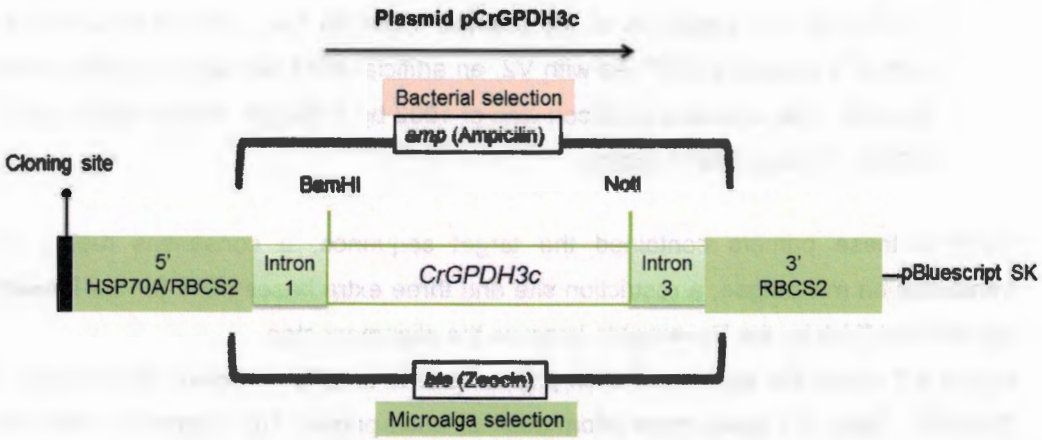


Figure 3.1 Schematic representation of plasmid pCrGPDH3c

As mentioned earlier in this study, the presence (retention) or absence of a putative intron in the cDNA may be causing a different start codon for *CsDGAT*, therefore more than one posible form of *CsDGAT* had to be considered. For that reason, three different version of CsDGAT were amplified. In order to do so, three forward primers in combination with one reverse primer were used for cloning this gene in three different versions: *CsDGAT* V1, *CsDGAT* V2 and *CsDGAT* V3 as described as follows, and as shown in Figure 3.2 and Table 3.1.

- *CsDGAT* V1 (Version 1 of *CsDGAT*). This version considers the putative intron to be part of the ORF, and upon removal a complete DGAT ORF sequence is formed. The primer forward was designed on the first ATG codon. The expected amplicon was of 1040 bp in length. Primer combination: DGAT_F1Clon/ DGAT_Rclon.

- *CsDGAT* V2 (Version 2 of *CsDGAT*). This is the smallest version considered, and begins exactly where the sequence of the putative intron ends. As with V1, this version also has DGAT motifs within its sequence. Since there are no ATGs

found on the original *CsDGAT* sequence in this part, an articial ATG codon was added on the primer in order to stablish a complete ORF on the further cloning steps. The expected amplicon was of 948 bp in length. Primer combination: DGAT_F2Clon/ DGAT_Rclon.

- *CsDGAT* V3 (Version 3 of *CsDGAT*). It consists of a sequence which does not incorporate the sequence of the putative intron as such, but its sequence as part of a complete ORF. As with V2, an artificial ATG was added to the primer forward. The expected amplicon was of 1002 bp in length. Primer combination: DGAT_F3Clon/ DGAT_Rclon.

Each of these primers contained the target sequence, a consensus region for translation on microalgae, a restriction site and three extra bases (TAT for the forward primers and TAA for the Reverse) to improve the alignment step.

Figure 3.2 shows the position of each primer used to amplify and clone all versions of *CsDGAT*. Table 3.1 gives more information of each primer: Tm, restriction sites and artificial ATGs added.

ATATTCGTGTGTCCTCTTTTGTGTTTGATGTTCTCGCTCCTTGAAATTGTCATCAAGATGGCATGCCCAG
CTTAGCTGTGCCTGTCTGTTGCACCTGCAAACGAGGACATGAACGATGCTTTGCAAGTCAGA
GCCTTTGTGATGACTTACTTTGGACAGCGCAATATCTG
ATCTATTCTGATGGCTTCAGCAAGGAGCACAAGCAGTCGTTCATCTCATGGCTTGTGGCAATTGTCACCC
TTACCATCTACACAGGATGGATGCACATACTGATAGGGTTGACTGTGGCCAGTTTGTTCAGTAGGACATG
CCTCTGCATCCTCGTAGCCATATGGGGAACGGTCTTCCTGCCGCCAAAGCCAGTGTTGTGGAATGCCTTC
TGCAAAAGCTGGATATTCCAGACCTGGCGCGAGTATTTCCAATTCAGCTACTTGAATGAAGCGGTGCTGG
ACCCCAAGAAGAAGTACATATTCACCGAGTTCCCTCATGGAGTCTTCCCTCTTAGCGAACTAGTTGCAGG
AACGCTGTGTCAAGCCATCTGGCCGGACTTCAGCATATATTCAGTGGCAGCCAGCAGTGTGTACAGCATT
CCCTTCTGGAGGCACTTCATCGCTTGGCTTGGTGCAGTCCCAGCTACAGCAGGCAATTTCAAGAAAATGA
TGAAGAGGGGAAGCTTGGCCGTCATTGTGGGAGGCATTGCTGAGATGTACATGCAGCACAAGCGCAAGGA
ACGCATCAAGCTGCTAGACAGGAAAGGGTTTGTCAAGATTGCTGTTGAGGAGGGCCTGGATGGAGGGATC
ATACCTGTGTATCACTTCGGCAACACTCAGGTGTTCGACTATTGGCCTCAGTCCTTTGAGAAATTTGCTC
GCAAGAACAGAGTGGGTGTGGGATTCTTGGTCGGACGATGGGGAACTCCTGTGCCGCGCAAGGTGCCCTT
GTACATGGTGAGTGGCAAGCCCATTCCAGTGCCCAAGGTCGACAAGAATGATACCGAGAAGTTCAACCAG
ACTGTAGATGAGATTCACGCTGAAGTGGTGCAACAGCTGCAGGACCTCTATGACAGACACAAGGCAAGCT
ATGGCTGGGAGAACAGACCCCTCAAGATTGAGTGATAGCCAGTGGCTTGCGTGAAGCTGCGTGACCGAAT
GTCCTTCTTGCAAGAAGAGTGATATAGAATCAGTTACAATATGTATAGAATACATATACACTGAGAGAAA
GCGTGTGAAATAGAATTGTAAGTAATCAAGCCTTAAAAAAAAAAAAAAAAA

Figure 3.2 Position of primers designed to clone different versions of *CsDGAT*. The forward primer for Version 1 (*CsDGAT* V1) is marked with grey. The forward primer for Version 2 (*CsDGAT* V2) is marked with yellow and the forward primer for Version 3 (*CsDGAT* V3) is

marked with cyan. Reverse primer for the three versions is underlined. Putative intron is highlighted in black

Table 3.1 Primers for the cloning of *CsDGAT* gene. Bam HI restriction site (GGATCC), EcoRI restriction site (GAATTC), NotI restriction site (GCGGCCGC), consensus translation site for microalgae (GCAACA) and artificial translation initiation codons (ATG) were added. Codons highlighted in green indicate first codon encountered on a possible ORF. Stop codon highlighted in yellow (TCA)

Primer ID	Sequence	Tm °
DGAT_F1Clon	TATGAATTCGGATCCGCAACAATGCTTTGCAAGTCAGAATTGGATTG	84.8
DGAT_F2Clon	TATGAATTCGGATCCGCAACAATGATCTATTCTGATGGCTTCAGCAAGG	84.8
DGAT_F3Clon	TATGAATTCGGATCCGCAACAATGCTTACTTTGGACAGCGCAATATC	83.9
DGAT_RcIon	TAAGCGGCCGCTCACTCAATCTTGAGGGGTCTGTTCTC	83.8

3.2.2 Cloning of three versions of *CsDGAT*

Each one of the three versions of *CsDGAT* ORF was amplified from *C. saccharophila* cDNA by PCR using Expand Long Template PCR System (Roche) and cloned on pGEM-T Easy (Promega). To extract and purify plasmids from *E. coli cultures*, the High Pure Plasmid Isolation Kit (Roche) was employed. The amplicons were then sequenced and aligned with previous results (sequences assembled from *CsDGAT* from chapter II) in order to verify that no mutations occurred.

After confirming the integrity of all three ORF sequences, each one of these were subcloned into pCrGPDH3c, to originate the transformation plasmids: pCsDGAT V1, pCsDGAT V2 and pCsDGAT V3. Briefly, each ORF was excised from pGEM-T Easy using two restriction enzymes: NotI and BamHI. The conditions used on restriction reactions were as followed: 5 U restriction enzyme, 2 µL reaction buffer (1X), 10 µL plasmid pGEM-T Easy with *CsDGAT* sequences on a volume of 20 µL. Plasmid pCrGPDH3c was also digested with the same enzymes, in order to leave sticky ends used on the subsequent subclonings and to the release of gene *CrGPDH3c*. Each one of the three digested versions of *CsDGAT* was then ligated overnight into the empty pCrGPDH3c using 10 U DNA T4 Ligase (NEB), multiplied in *E. coli* strain DH10B, and the resuling plasmids were purified using the High Pure Plasmid Isolation Kit (Roche). Finally, the resulting plasmids (pCsDGAT V1 to V3) were digested with NotI and BamHI to examine that all fragments had been correctly introduced.

3.2.3 *C. reinhardtii* transformation method and selection media

C. reinhardtii was transformed using microparticle bombardment (biolistics). First, 55 mg of 0.6 µm gold particles (BioRad ®) were washed with 100% ethanol and vortexed

for 3 minutes, pelleted by centrifugation for 2-5 minutes and the supernatant was discarded. 70% ethanol was added and vortexed for 2 minutes. The mixture was incubated for 15 min at room temperature and mixed three times. Then it was centrifuged at maximum speed for 3 min and the supernatant discarded. The particles were resuspended on 1mL of ethanol and allowed to settle down for 1 min at room temperature. Then, samples were centrifuged at maximum speed for 2 min and the supernatant discarded. 50% v/v sterile glycerol was added to the gold particles at a final concentration of 50 mg/mL.

For the bombardment: 50 µL of the gold particle solution was transferred into a 1.5 mL tube. The following components were added in this order: first, a maximum of 10 µg Plasmid DNA, then 50 µL of 2.5 mM $CaCl_2$, followed by 20 µL of 0.1 M spermidine. This mixture was then vortexed for 20-30 min at 4°C. Then, 200 µL absolute ethanol were added, vortexed, centrifuged 45 seconds for 1300 rpm and the supernatant was discarded. This step was repeated three more times. After the final step, the pellet was resuspended in 30 µL of absolute ethanol. 5 µL of this solution was spread on each carrying membrane, and five membranes (one for each shot) were used for every gold particle solution (five shots per plasmid).

Cell preparation for particle bombardments were as follows: a single colony was taken and transferred to 50 mL of liquid TAP on a 250 mL Erlenmeyer flask and grown for seven days until reaching the end of logarithmic phase. Cells were then counted and collected by centrifugation at 3,220 g for five minutes at 4°C. 40 x 10^6 cells were resuspended on 250 µL TAP and then plated on solid TAP on Petri dishes (on the center) and let to dry the excess of liquid medium on the laminar flow cabinet. Bombardments were performed using Helium as the propulsor gas on a microparticle bombardment cabinet, 0.6 µm size for the gold particles and 1100 pounds per square inch (psi) as pressure for membrane disruption.

After performing the bombardments, the microalgae plates were stored overnight at low light intensity (2 - 5 µE/m^2/s) and then transferred to selection TAP solid medium (20 µg/mL zeocin) and incubated at photoperiod (16/8 hours light/dark) 25 - 30 µE/m^2/s of light intensity. After aproximately 7-10 days the first colonies appeared. The colonies recovered from the transformation plates were transferred to flasks containing liquid TAP medium (no antibiotic) for culture.

3.2.4 Analysis of colonies recovered after microparticle bombardment

For the detection of the *CsDGAT* gene on colonies recovered from microparticle bombardment, each colony was grown on liquid TAP under regular light conditions for

seven days to reach the end of logarithmic phase. DNA was extracted using the protocol by Dellaporta *et al.* [1983] and a fragment of *CsDGAT* was amplified by PCR using gene specific primers. The forward primer was complementary to the ORF sequence of *CsDGAT*: 5' ACCTGTGTATCACTTCGGCAAC 3'. The reverse primer was generated to align to the sequence: 5'-ACGAGCGCCTCCATTTACAC-3' found in the 3' UTR of *RBCS2* in pCsDGAT. PCR conditions were: denaturation step at 95°C for 1 minutes, annealing step at 60°C for 30 seconds and 1 minute at 72 °C for extension. This PCR was used to detect both a successful transformation with the plasmid and the incorporation of the *CsDGAT* gene.

3.3 RESULTS

3.3.1 Amplification and cloning of *CsDGAT* V1 to V3

Three versions of *CsDGAT* (*CsDGAT* V1, V2 and V3) were amplified by PCR using the primers described in Materials and Methods (Table 3.2), and the amplicons are shown in Figure 3.3A. Each amplicon was cloned into pGEM-T Easy (Fig. 3.3B) to form plasmids pGCsDGAT V1 to V3, and sequenced to verify the integrity of the sequence.

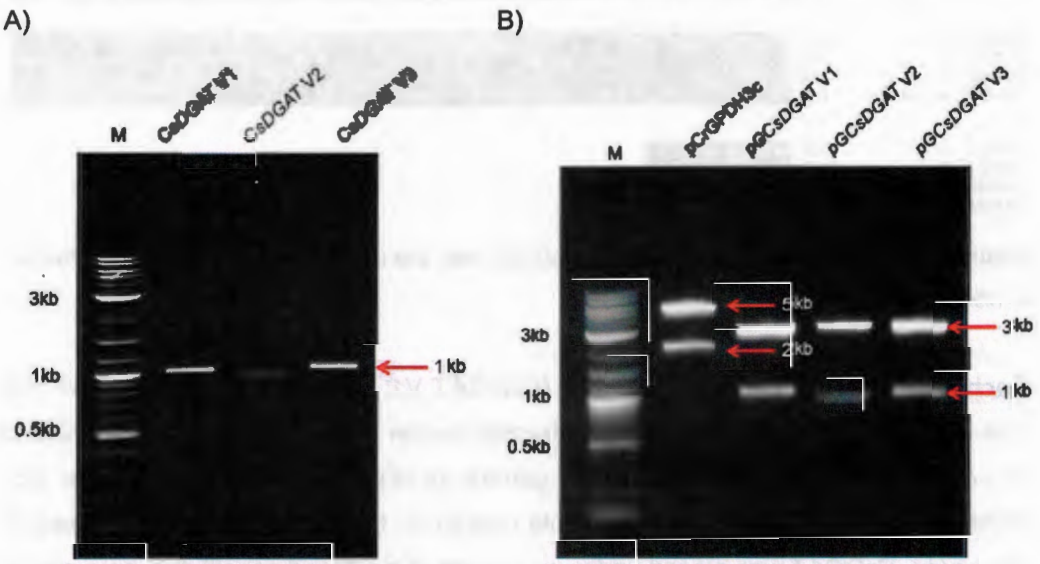


Figure 3.3 Cloning of three different versions of *CsDGAT*. A) PCR amplifications of three versions of *CsDGAT* (*CsDGAT* V1, V2 and V3) from *C. saccharophila* cDNA. B) Analysis by restriction enzymes *Not*I and *Bam*HI of the three cloned version of *CsDGAT* cloned in pGEM-T Easy. pCrGPDH3: digestion of plasmid, 5kb band corresponds to plasmid and 2kb band corresponds to *CrGPDH3* gene originally cloned in this plasmid. pGCsDGAT V1 to V3: digestion of plasmids, 3kb band corresponds to pGEM-T Easy, and bands of approximately 1kb are the different versions of *CsDGAT*. Molecular marker 2-log (NEB)

The sequence obtained for each cloned version of *CsDGAT* was aligned to the originally isolated and characterized sequence from chapter II, in order to assure that no changes in nucleotides that could affect aminoacid changes had occurred. As shown in Figure 3.4, the integrity of all three *CsDGAT* versions was maintained. Amino acid sequences were employed in order to avoid discrepancies caused by genetic code degeneration. The only differences encountered were on the first 22 aminoacids, caused by the presence of the putative intron, as well as the artificial ATGs added on *CsDGAT* V2 and V3.

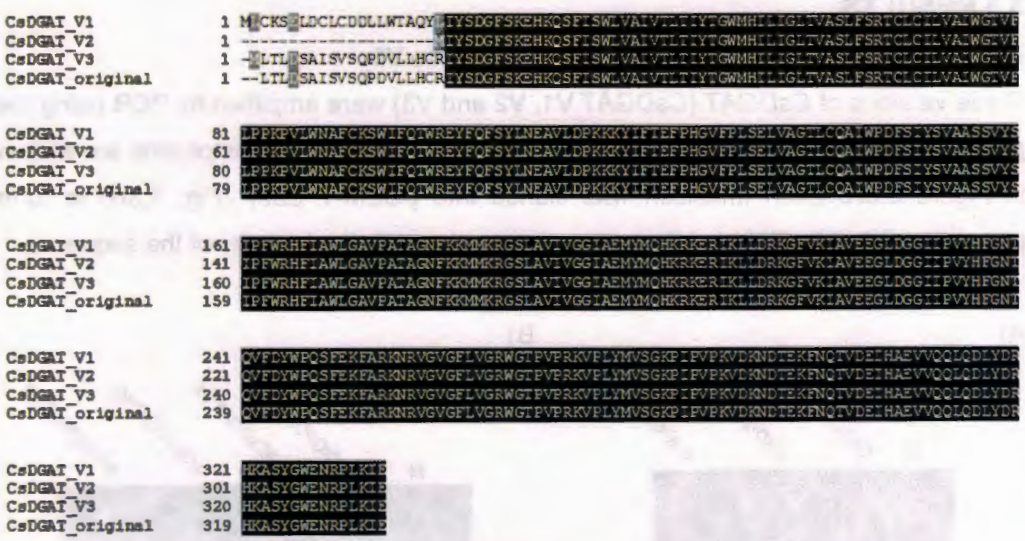


Figure 3.4 Alignment of three versions of *CsDGAT* with the originally isolated and characterized *in silico* sequence of *CsDGAT*

Each of the three versions of *CsDGAT* (*CsDGAT* V1, *CsDGAT* V2 and *CsDGAT* V3) was cloned in pCrGPDH3c, to form the transformation plasmids pCsDGAT V1, V2 and V3. A squematic representation of the general pCsDGAT is depicted in Figure 3.5, which summarizes the major components present in the expression cassette used in this study. *CsDGAT* was placed under the control of the fused constitutive promoter of the Heat Shock Protein 70A and the Ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco). Two introns were included, which have been reported to improve the gene expression [Schroda *et al.* 2000; Lumbreras *et al.* 1998].

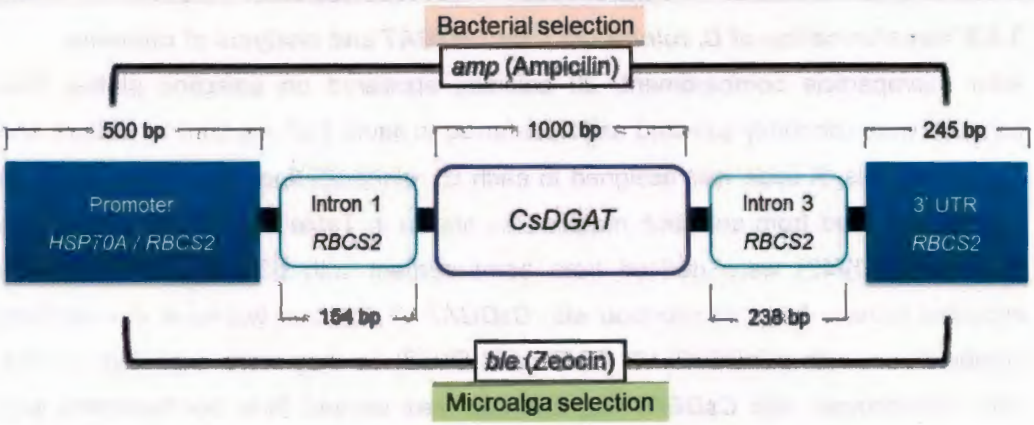


Figure 3.5 Schematic representation of transformation plasmid (pCsDGAT) containing one of the three versions of *CsDGAT* (*CsDGAT* V1, *CsDGAT* V2 or *CsDGAT* V3) and regulatory sequences

Three clones from each version of pCsDGAT were analyzed by restriction enzyme digestion with BamHI and NotI to corroborate that *CsDGAT* had been in introduced into the plasmid (Fig. 3.6). Two bands of the expected size were visible for each clone, one band of 5kb corresponding to the plasmid and a 1kb band corresponding to the insert *CsDGAT*.

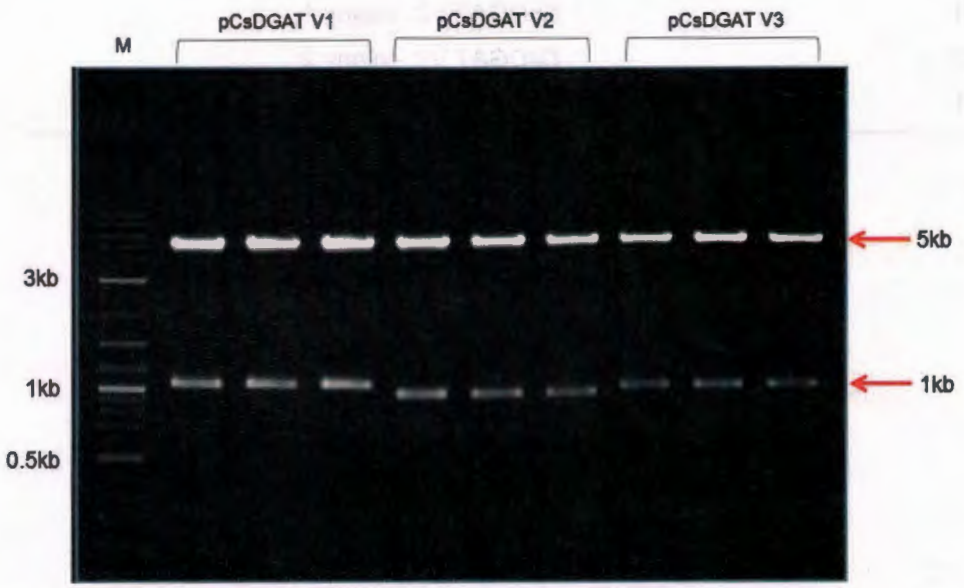


Figure 3.6 Retriction analysis of plasmids pCsDGAT V1, V2 and V3 with *Not*I and *Bam*HI. 5kb band corresponds to pCsDGAT and 1kb bands correspond to each version of *CsDGAT* excised from the plasmid. Marker used: 2-log (NEB)

3.3.2 Transformation of *C. reinhardtii* with *CsDGAT* and analysis of colonies

After microparticle bombardment, 20 colonies appeared on selection plates. Five colonies were randomly selected and transferred to liquid TAP medium for culture and further analysis. A code was assigned to each *C. reinhardtii* line derived from a single colony recovered from selection medium, as shown in Table 3.2. Two of these lines (DV11 and DV12) were derived from bombardment with pCsDGAT V1 so were expected to have been transformed with *CsDGAT* V1, another two were derived from bombardment with pCsDGAT V2 (DV21 and DV22) so they were expected to have been transformed with *CsDGAT* V2, and one was derived from bombardment with pCsDGAT V3 (DV31) so it was expected to have been transformed with *CsDGAT* V3. Total DNA was isolated from each of the five lines described above and its integrity was evaluated by electrophoresis on an agarose gel (Fig. 3.7).

Table 3.2 Code assignation and description for colonies recovered after microparticle bombardment

Code	*C. reinhardtii* line description
DV11	*CsDGAT* V1, colony 1.
DV12	*CsDGAT* V1, colony 2.
DV21	*CsDGAT* V2, colony 1.
DV22	*CsDGAT* V2, colony 2.
DV31	*CsDGAT* V3, colony 1.

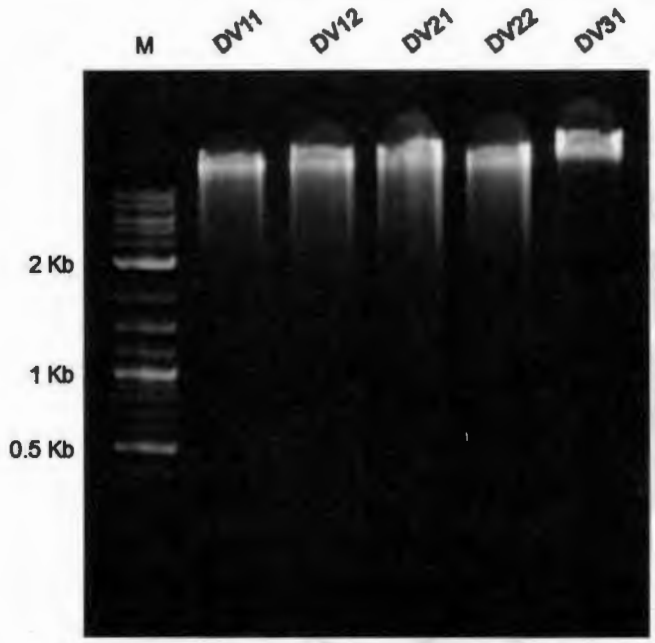


Figure 3.7 Visualization of genomic DNA belonging to *C. reinhardtii* colonies recovered from bombardment. Each lane correspond to a line derived from a single colony. Codes from Table 3.2. Molecular marker 2-log (NEB)

The five lines of *C. reinhardtii* recovered after microparticle bombardment, were analyzed by PCR to determine whether the *CsDGAT* gene was present in their genomic DNA (Fig. 3.8). The appropriate transformation plasmid was used as a positive control for its respective line. The empty plasmid and the gDNA of the wild type *C. reinhardtii* were used as negative controls.

As shown in Figure 3.8, at least two positive results for each *CsDGAT* V1 and V2 were obtained, where the expected band of ~1kb was obtained. As expected, no amplification was visible on either the empty plasmid or the wild type strain, while the expected band was visible in all the positive controls. Unfortunately, it was not possible to obtain a positive transformant harboring *CsDGAT* V3.

In total, two transformed colonies were obtained for *CsDGAT* V1 and another two for *CsDGAT* V2. These transgenic lines will need to be functionally studied to clarify the role of the putative intron and to investigate the effect of the overexpression of *CsDGAT* on the levels of TAGs.

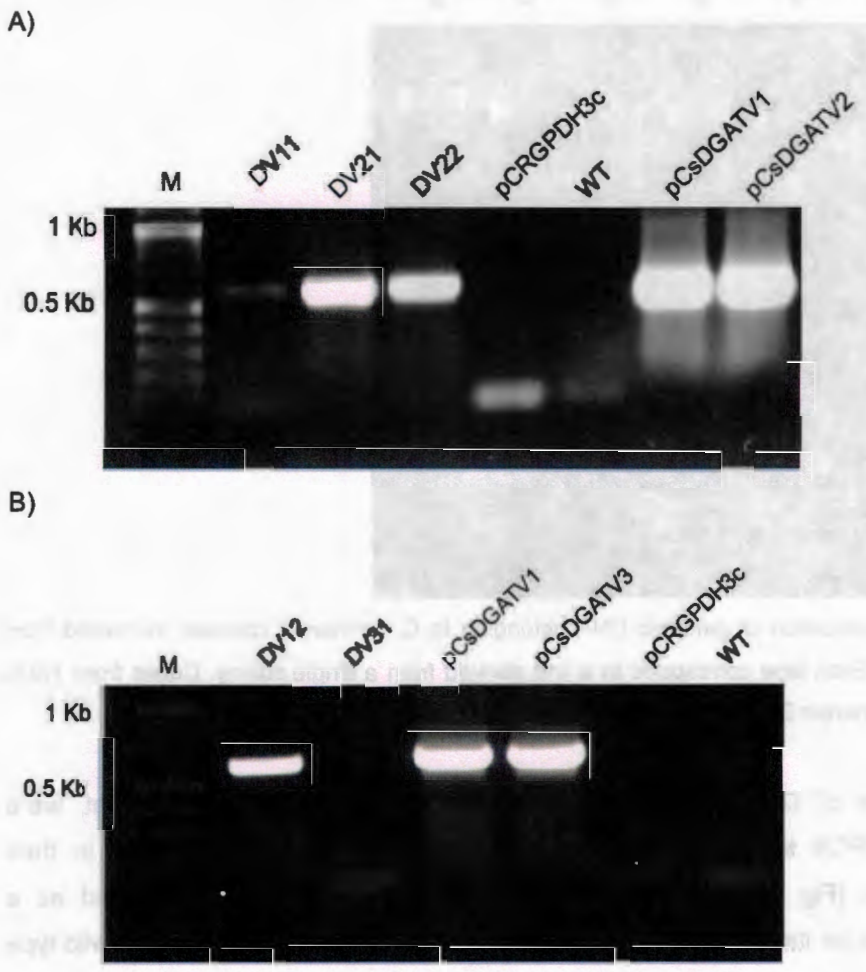


Figure 3.8 Analysis of putative transformants by PCR. A) Detection of *CsDGAT* Versions 1 and 2 (DV11, DV21 and DV22). B) Detection of *CsDGAT* Versions 1 and 3 (DV12 and DV31) Plasmids pCsDGAT of each version were used as positive controls. Empty plasmid pCsDGAT and wild type strains (WT) were used as negative controls. Molecular marker 2-log (NEB)

3.4 DISCUSSION

In this chapter, three versions of *CsDGAT* (*CsDGAT* V1 to V3) were cloned into a transformation plasmid and two of them (*CsDGAT* V1 and V2) were successfully transformed into *C. reinhardtii*. To do so, first they were introduced into pGEM-T easy and their sequence was verified, and then subcloned into plasmid pCrGPDH3c to form three different versions of transformation plasmid pCsDGAT (V1 to V3). This plasmid was used in transformation of *C. reinhardtii* with biolistics and then putative transgenic colonies were analyzed. In total, 20 colonies were recovered; a very low number when compared with results obtained by Stevens *et al.* [2002], since these authors could recover nearly 500 colonies per plate. However, in their methodology, based on Kindle *et al.* [1994], Stevens and collaborators [2002] used a glass bead technique with cell wall deficient strains of *C. reinhardtii*, which enhances their transformation efficiency. Even though this glass bead strategy has proven to give high number of colonies, this was not the case for Walker [2003], when carrying the same experiment a reduced number of *C. reinhardtii* transformed colonies were obtained: 11 colonies per every 10 plates used. A number of intrinsic parameters of each method influence its efficiency, including quality and age of DNA and microalgae cultures, quantity of DNA introduced and the manipulation itself of the method employed [Walker *et al.* 2005]. The plasmid used in this thesis (pSP124S with a constituve promoter) has been used for biolistics essays on *C. reinhardtii* before. Yamasaki *et al.* [2008] used this plasmid with the *RBCS2* promoter to introduce interference RNA into *C. reinhardtii* to silence the gene *aad* which codes for an aminoglycoside 3'-adenyltransferase, present on the strain 19-P[1030] strain. These authors do not mention how many colonies they obtained on their results, but they do mention that they randomly selected 59 colonies growing on zeocin media.

In the present study, five putative transformed colonies growing on selection media with zeocin were randomly picked and their DNA was analysed by PCR. However, one of these colonies (DV31) did not have the *CsDGAT* fragment, even though the same procedure and PCR conditions were applied as with the other colonies. Khachatouians *et al.* [2002] described a possible event of non-transformant false positives (also called "escapes") of a physical nature, in plants. In this event the epidermal cells contribute to the formation of these escapes since they are not transformed, and so the elimination of these cells is performed before transformation. Although *C. reinhardtii* is a unicellular organism and does not have epidermal cells, it has been observed in our laboratory that microalgae tend to agglomerate when forming colonies and so the microalgae that

grow on top of others, can be protected from the antibiotic. Since the colonies were first recovered from solid medium but then multiplied on liquid TAP lacking any antibiotic, these non-transformant protected colonies could have been able to grow in spite of their DNA having no transgene integration.

An important obstacle for the transformation of microalgae has been the difficult to find vectors and regulatory sequences to be correctly used, and not only extrapolate data from plants. For that reason, several attempts are currently being developed to circumvent these issues. Many of these rely on the recent discoveries made mainly on the model microalga *C. reinhardtii*. Therefore, these new findings regarding regulation in microalgae transformation come from and are being used in this organism [Lumbreras *et al*, 1998]. In the present study, several elements used for the transgene expression on *C. reinhardtii* were found into the plasmid pSP124S, including the selection gene *ble* that confers resistance to zeocin. Schroda *et al.* [2000] characterized the promoter *HSP70A* and discovered that it induced a better expression when fused with other elements (such as *RBCS2* or *βTUB*) than those sequences alone. This fusion constitutive promoter (*HSP70A / RBCS2*) was used on our constructions with *CsDGAT*, since it has been shown to improve total transgene expression constitutively when compared to other strategies, such as the promoter *RBCS2* alone [Kindle 1990; Blankenship *et al.* 1993; Heitzer and Zshoerning 2007].
The addition of introns from *RBCS2* has been described as elements that stimulate the rate of transformation and expression levels. Initially described by Lumbreras *et al.* [1998], more recent reports used this strategy for the expression of reporter genes such as luciferase [Heitzer and Zschoerning 2007; Eichler-Stahlberg *et al.* 2009]. For this study, is expected that the inclusion of these sequences will indeed improve the expression of *CsDGAT* on *C. reinhardtii*. On the report by Lumbreras *et al.* [1998] the regulatory sequences present at the 5' and 3' UTR are also considered. They concluded that the presence of these kinds of regions is fundamental for the expression of transgenes since their deletion ends in the reduction of transformation rate by two orders of magnitude.

In this chapter, three different version of the same gene *CsDGAT* were amplified and introduced to the plasmid pCrGPDH3c, and subsequently transformed into *C. reinhardtii* by means of biolistics. Two versions of *CsDGAT* (V1 and V2) were successfully inserted on *C. reinhardtii*'s genome, as shown on the PCR results. Since the number of copies of a given gene introduced by biolistics is unknown, it would be necessary to perform a Southern blot essay, to determine this condition. However,

these results reflect only the integration of *CsDGAT* gene but not its expression. Several analyses need to be performed in order to elucidate this phenomenon, such as RT-PCRs specific for this gene, and Real Time PCRs to quantify its expression. The functional expression of this transgene also needs to be verified, by measuring TAG levels. It is widely accepted that the optimum expression of transgenes in microalga has been very difficult to achieve. Even though the regulatory sequences contemplated here effectively are proven to improve this task, a general solution for every transgene is not readily available. Recently, a possible solution for this issue was investigated by Neupert *et al.* [2009]. These authors interfered with the inherent mechanism of *C. reinhardtii* that regulate gene expression, and achieved a high expression of both a native gene and a reporter (GFP). To do so, they transformed *C. reinhardtii* with the *CRY1-1* gene which confers insensitivity to the translational inhibitor emetine. When recovering resistant colonies, they subjected them to UV mutagenesis in order to interfere with these transgene regulation mechanisms. The inclusion of these kinds of techniques for the study of *CsDGAT* could represent an interesting approach to improve its expression.

All three different versions of *CsDGAT* should be analyzed further. The main difference of these versions rely on the presence/absence of the putative intron on the mature RNA, thus their functional insights could give strong evidence of intron retention events in microalgae. Guihéneuf *et al.* [2011] gave the firsts data regarding this occurrence in microalgae, but the information is still scarce specially when compared to plants, where it is a well studied phenomenon [Mastrangelo *et al.* 2005].

As mentioned earlier, *C. reinhardtii* is a model organism extensively used for microalgae research. For that reason, it was the microalga of choice for this study as well. The transgenic lines of *C. reinhardtii* with the three versions of *CsDGAT* will be fundamental on the study of the overexpression of *CsDGAT*, and its effect on total TAGs present on the cell, as well as its response to stress conditions. Also, the predicted presence and remotion of the intron found could be clarified, a post-transcriptional regulation not yet described in green microalgae.

3.5 REFERENCES

Blankenship J.E. and K.L. Kindle (1992). *Exp*ression of chimeric genes by the light-regulated cabII-1 promoter in *Chlamydomonas reinhardtii*: a cabII-1/nit1 gene functions as a dominant selectable marker in a nit1- nit2-strain. Mol. Cell. Biol. 12:5268-5279.

Cha T.S., W. Yee and A. Aziz (2011). Assessment of Factor Affecting Agrobacterium-mediated Genetic Transformation of the Unicellular Green Alga, *Chlorella vulgaris*. World Journal of Microbiology and Biotechnology 28(4):1771-1779.

Dellaporta S.L., J. Wood and J.B. Hicks (1983). A Plant DNA Minipreparation: Version II. Plant Molecular Biology Reporter 1:19-21.

Eichler-Stahlberg A., W. Weisheit, O. Ruecker and M. Heitzer (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta 229(4):873-883.

Gong Y., H. Hu, Y. Gao, X. Xu and H. Gao (2011). Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. Journal Industrial Microbiology Biotechnology 38(12):1879-1890.

Guihéneuf F., S. Leu, A. Zarka, I. Khozin-Goldberg, I. Khalilov and S. Boussiba (2011). Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (*PtDGAT1*) from the diatom *Phaeodactylum tricornutum*. The FEBS Journal 278(19):3651-3666.

Heitzer M. and B. Zschoernig (2007). Construction of modular tandem expression vectors for the green alga *Chlamydomonas reinhardtii* using the Cre/lox-system. Biotechniques 43(3):324-332.

Herrera-Valencia V.A., Y.P. Contreras-Pool, S.J. López-Adrián, S. Peraza-Echeverría and L.F. Barahona-Pérez (2011).The green microalga *Chlorella saccharophila* as a suitable source of oil for biodiesel production. Current Microbiology 63(2):151-157.

Khachatouians G., A. McHughen, R. Scorza, W. Nip and Y.H. Hui (2002). Transgenic Plants and Crops. CRC Press. p 870-880.

Kindle K.L. (1990). High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences. USA 87:1228-1232.

La Russa M., C. Bogena, A. Uhmeyera, A. Doebbea, E. Filipponeb, O. Krusea and J.H. Mussgnuga (2012). Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. Journal of Biotechnology 162(1):13-20.

Lumbreras V., D.R. Stevens and S. Purton (1998). Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. The Plant Journal 14(4):441-448.

Mastrangelo A.M., S. Belloni, S. Barilli, B. Ruperti, N. Di Fonzo, A.M. Stanca and L. Cattivelli (2005). Low temperature promotes intron retention in two e-cor genes of durum wheat. Planta 221(5):705-715.

Neupert J., D. Karcher and R. Bock (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. The Plant Journal 57(6):1140-1150.

Nykiforuk C., T. Furukawa-Stoffer, P. Huff, M. Sarna, A. Laroche, M. Moloney and R. Weselake (2002). Characterization of cDNAs encoding diacylglycerol acyltransferase from cultures of *Brassica napus* and sucrose-mediated induction of enzyme biosynthesis. Biochimica et Biophysica Acta 1580(2-3):95-109.

Schroda M., D. Blöcker and C.F. Beck (2000). The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. The Plant Journal 21(2):121-131.

Stevens, D.R., J.D Rochaix and S. Purton (1996). The bacterial phleomycin resistance gene ble as a dominant selectable marker in *Chlamydomonas*. Molecular Genetics and Genomics. 251(1):23-30.

Wagner M., K. Hoppe, T. Czabany, M. Heilmann, G. Daumb, I. Feussner and M. Fulda (2010). Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (*DGAT2*) gene from the microalga *O. tauri*. Plant Physiology and Biochemistry 48(6):407-416.

Walker T.L. (2003). The Development of Microalgae as a Bioreactor System for the Production of Recombinant Proteins. Tesis de Doctorado. 194-204 pp

Walker T., S. Purton, D. Becker and C. Collet (2005). Microalgae as bioreactors. Plant Cell Reports 24:629-641.

Yamasaki T., H. Miyasaka and T. Ohama (2008). Unstable RNAi Effects Through Epigenetic Silencing of an Inverted Repeat Transgene in *Chlamydomonas reinhardtii*. Genetics 180:1927–1944.

CHAPTER IV

GENERAL DISCUSSION AND CONCLUSION

4.1 DISCUSSION

In previous studies, at least three different types of DGAT have been characterized: DGAT 1, found in *Arabidopsis* for the first time, DGAT 2 which was characterized firstly in mouse but has been reported for several yeast species, plants and animals [Cases *et al.*,1998]. Both DGAT 1 and 2 have been isolated from microalgae and both have been related to lipid synthesis, specifically TAG formation. The third type of DGAT was initially found in peanut [Saha *et al.* 2006], and sub sequentially discovered in plant specimens such as *Arabidopsis* and some animals, but to date there are no reports involving this type of DGAT on microalgae [Wagner *et al.* 2010; La Russa *et al.* 2012]. Functions among these different kinds of DGAT vary greatly among individuals, especially in plants. In the case for the types 1 and 2, their expression has been associated to specific tissues, circadian cycles and/or growth phases. However, it has also been annotated as species and tissue dependent and no general correlation has been made so far [Shockey *et al.* 2006; Deng *et al.* 2012].

What gives each family a punctual distinction from another one, is the presence of specific motifs responsible for their location or activity. The family type 2 of DGAT present at least two important motifs that give the enzyme its transmembrane location: motifs HPHG and RXGFX(K/R)XAXXXGXXX(L/V)VPXXXFG(E/Q) [Liu *et al.* 2011]. These sequences were found on the *CsDGAT* gene studied in this thesis. Furthermore, it showed the most similitude with the type 2 family when included on phylogenetic and bioinformatics analyses.

On previous studies, Wagner *et al.* [2010] followed a similar approach with the microalga *Ostreococcus tauri* which is a chlorophyta (green alga) as well. These authors concluded that the gene characterized belonged to the type 2 family, based on similar aspects as the present study: Specific motif presence and phylogenetic position. Recently, La Russa *et al.* [2012] performed a search for *DGAT* genes on the genome of *Chlamydomonas reinhardtii* and found 5 candidate sequences. They compared several DGAT type 1 and 2 and discovered that all the genes retrieved from *C. reinhardtii* belonged to the type 2 family. Including *Chlorella saccharophila* studied here, there have been 3 different chlorophytae whose DGAT sequences were found to be classified as DGAT type 2 sequences.

Based on the regulatory sequences found at both 5' and 3' UTR, it could be suggested that *CsDGAT* may be post-transcriptionally regulated. In addition, a possible event of intron retention was found, a phenomenon described on superior organisms such as plants. In *Arabidopsis,* Ner-Gaon *et al.* [2006] described an event of this nature on its genome and reported that 2% of *Arabidopsis* genes fall on this category. However, more detailed analyses for microalgae are necessary to further understand this regulation.

Even more possibilites of regulation regulation at 3' UTR were noticed. One of these regulations could be attained to alternative splicing, since more than two forms of 3' UTR could be amplified. This type of regulation have been described extensively on plants and animals, and have a correlation with mRNA stability and protein diversity [Barbazuk 2008]. Another possible regulation based on 3' UTR occurs as alternative polyadenylation. This mechanism occurs in both animals and plants animal [Ji and Tian 2009; Xing and Li 2010] where the regulation of cell differentiation, stress responses and flowering control take place on the 3' UTR. Additionally, for the polyadenylation event to take place, several signals on the RNA need to be present. Shen *et al.* [2008] characterized *in silico* several putative sequences on *C. reinhardtii* as responsible sequences to form a poly-A. Sequences of these nature were found on *CsDGAT* and agrees with other authors who state that the existence of more than one signal can be associated with this alternative forms of poly-A's.

There are several studies that have correlated an increase of DGAT expression by lack of nutrients, a condition reported for bacteria [Sirakova *et al.* 2006] and microalgae [Deng *et al.* 2012]. However, it has been pointed out that the strongest inductor of this gene is the deprivation of nitrogen. For that reason, in this report the expression of *CsDGAT* was examined during a nitrogen deficiency treatment. A complete nitrogen deprivation stress was not performed in these experiments. Instead, a very low amount of this nutrient was employed compared to optimum growth conditions. This change in media composition has been reported for *C. saccharophila* [Herrera-Valencia *et al.* 2011] as triggering effect for an increase in total lipid and TAG accumulation. Hence, this same procedure was applied for the measure of gene expression.

A steady *CsDGAT* expression was detected following different time courses up to 48 hours of the stress onset, that means, a constitutive expression was found. Boyle *et al.* [2012] used a similar experimental strategy, with the genes from *C. reinhardtii*. It was possible for them to quantify the expression of *CreDGAT* genes under nitrogen starvation, where a gradual increase was noticed and measured. On the present work, *CsDGAT* did not increase its expression, and it is yet unknown whether it presents a

similar pattern as its *C. reinhardtii* orthologue or follows another direction since the total accumulation did not change from the first six hours to the final forty eight. Further studies using real time RT-PCR are necessary to investigate this issue for *CsDGAT*.

Several attempts have been made to improve the transgene expression in microalgae [Lumbreras *et al.* 1998; Schroda *et al.* 2000; Heitzer and Zschoerning 2007; Eichler-Stahlberg *et al.* 2009], such as the inclusion of regulatory sequences like UTRs and introns. Therefore, these regions were included on the plasmid pSP124S which has been reported to be successfully transformed into *C. reinhardtii,* and harbor the gene *ble* that confers resistance to zeocin. In the present thesis, three versions of *CsDGAT* (called V1, V2 and V3 respectively) were cloned into a transformation plasmid under a constitutive strong fusion promoter in order to achieve high levels of transgene. These plasmids were transformed into *C. reinhardtii* by biolistics. Five random colonies were selected and their DNA extracted with the purpose of analyzing transgene integration. Of all the versions of *CsDGAT* employed, V1 and V2 were found on two colonies each, but no positive results were generated for V3. By the end of these experiments, four transgenic lines of *C. reinhardtii* were generated harboring the gene *CsDGAT* V1 and V2. The subsequent assays need to prove the functionality of this gene, by analyzing both mRNA expression, southern blot to determine number of copies integrated and finally, the effect on TAG levels.

4.2 GENERAL CONCLUSIONS

1. A full cDNA sequence encoding a protein with homology to the DGAT family was isolated from *C. saccharophila* and named as *CsDGAT*. Additionally, a partial cDNA sequence of an *Actin* like gene was obtained in the present study for *C. saccharophila* to be used as a control in RT-PCR assays. The expression of this gene was constitutive.

2. The phylogenetic analysis showed that the CsDGAT sequence grouped within the clade of DGAT type 2 proteins.

3. The 5' and 3' ends of *CsDGAT* cDNA were mapped. A possible intron retention event was found near to the translation initiation codon, and events of alternative splicing were discovered for the 3'UTR.

4. The expression of *CsDGAT* is present in TAP medium and nitrogen deficiency conditions.

5. Four transgenic lines of *Chlamydomonas reinhardtii* harboring two different version of *CsDGAT* (*CsDGAT* V1 and V2) under the control of a strong constitutive promoter were generated for the functional characterization of this gene.

4.3 PERSPECTIVES

The present study provided the first insigths of a *DGAT* homologue gene in the microalga *C. saccharophila*. In order to expand the knowledge on this sequence and other posible homologues, as well as the potential use of this type of gene to increase the TAG content in this microalga, we propose the following studies:

1. Sequencing analysis of *CsDGAT* transcripts to validate the prediction of the intron near to translation initiation site and also to study other posttranscriptional regulation events.

2. Transcriptional analysis of *CsDGAT* using real time RT-PCR to validate whether there are differences on the expression levels of *CsDGAT* under TAP and nitrogen deficiency conditions.

3. Expression and copy number analyses of ten trangenic lines of *C. reinhardtti* harboring the *CsDGAT* gene.

4. Assessing the effect of overexpressing *CsDGAT* on the TAG levels of *C. reinhardtti* transgenic lines.

5. Trancriptomic analysis of *C. saccharophila* using the technology of RNA-seq on TAP and Nitrogen deficiency growing conditions. This high-throughput technology should allow us determining whether other DGAT homologues genes are present in *C. saccharophila*, and also to quantify their expression. This analysis will also provide us with a broad picture of other genes involved in TAG metabolism.

4.4 REFERENCES

Barbazuk W.B., Y. Fu and K.M. McGinnis (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Research 18(9):1381-1392.

Boyle N., M.D. Page, B. Liu, I.K. Blaby, D. Casero, J. Kropat, S.J. Cokus, A. Hong-Hermesdorf, S.J. Karpowicz, S.D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman and S.S. Merchant (2012). Three Acyltransferases and Nitrogen-responsive Regulator Are Implicated in Nitrogen Starvation-induced Triacylglycerol Accumulation in *Chlamydomonas*. The Journal of Biological Chemistry. 287(19):15811-15825.

Cases S., S.J. Smith, Y.W. Zheng, H.M. Myers, S.R. Lear, E. Sande, S. Novak, C. Collins, C.B. Welch, A.J. Lusis, S.K. Erickson and R.V. Farese Jr. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences. 95(22):13018–13023.

Cases S., S. Novak, Y.W. Zheng, H.M. Myers, S.R. Lear, E. Sande, C.B. Welch, A.J. Lusis, T.A. Spencer, B.R. Krause, S.K. Erickson, R.V. Farese Jr. (1998). ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase, Its cloning, expression and characterization. The Journal of Biological Chemistry. 273(41):26755-26764.

Deng X.D., B. Gu, Y.J. Li, X.W. Hu, J.C. Guo and X.W. Fei (2012). The Roles of acyl-CoA: Diacylglycerol Acyltransferase 2 Genes in the Biosynthesis of Triacylglycerols by the Green Algae *Chlamydomonas reinhardtii*. Molecular. Plant 5(4):945-947.

Eichler-Stahlberg A., W. Weisheit, O. Ruecker and M. Heitzer (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. Planta 229(4):873-883.

Heitzer M. and B. Zschoernig (2007). Construction of modular tandem expression vectors for the green alga *Chlamydomonas reinhardtii* using the Cre/lox-system. Biotechniques 43(3):324-332.

Herrera-Valencia V.A., Y.P. Contreras-Pool, S.J. López-Adrián, S. Peraza-Echeverría and L.F. Barahona-Pérez (2011). The green microalga *Chlorella saccharophila* as a suitable source of oil for biodiesel production. Current Microbiology 63(2):151-157.

Ji Z. and B. Tian (2009). Reprogramming of 3′ Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. PLoS ONE 4:13.

La Russa M., C. Bogena, A. Uhmeyera, A. Doebbea, E. Filipponeb, O. Krusea and J.H. Mussgnuga (2012). Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. Journal of Biotechnology, 162(1):13-20.

Lumbreras V., D.R. Stevens and S. Purton (1998). Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. The Plant Journal 14(4):441-448.

Liu Q., R. Siloto, C. Snyder and R. Weselake (2011). Functional and Topological Analysis of Yeast Acyl-CoA:Diacylglycerol Acyltransferase 2, an Endoplasmic Reticulum Enzyme Essential for Triacylglycerol Biosynthesis. The Journal of Biological Chemistry 286(15):13115-13126.

Lumbreras V., D.R. Stevens and S. Purton (1998). Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. The Plant Journal 14(4):441-447.

Ner-Gaon H., R. Halachmi, S. Savaldi-Goldstein, E. Rubin, R. Ophir and R. Fluhr (2004). Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. The Plant Journal 39(6):877-885.

Saha S., B. Enugutti, S. Rajakumari and R. Rajasekharan (2006). Cytosolic Triacylglycerol Biosynthetic Pathway in Oilseeds. Molecular Cloning and Expression of Peanut Plant Physiology, 141(14):1533-1543.

Schroda M., D. Blöcker and C.F Beck (2000). The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. The Plant Journal 21(2):121-131.

Shen Y., Y. Liu, L. Liu, C. Liang and Q. Li (2008). Unique Features of Nuclear mRNA Poly(A) Signals and Alternative Polyadenylation in *Chlamydomonas reinhardtii*. Genetics 179(1):167-176.

Shockey J., S. Gidda, D. Chapital, J.C. Kuan, P.K. Dhanoa, J.M. Bland, S.J. Rothstein, R.T. Mullen and J.M. Dyera (2006). Tung Tree *DGAT1* and *DGAT2* Have Nonredundant Functions in Triacylglycerol Biosynthesis and Are Localized to Different Subdomains of the Endoplasmic Reticulum. The Plant Cell 18(9):2294-2313.

Sirakova T.D., V. Dubey, C. Deb, J. Daniel, T. Korotkova, B. Abomoelak and P. Kolattukudy (2006). Identification of a diacylglycerol acyltransferase gene involved in accumulation of triacylglycerol in *Mycobacterium tuberculosis* under stress. Microbiology Reading England 152(9):2717-2725.

Wagner M., K. Hoppe, T. Czabany, M. Heilmann, G. Daumb, I. Feussner and M. Fulda (2010). Identification and characterization of an acyl-CoA:diacylglycerol

acyltransferase 2 (*DGAT2*) gene from the microalga *O. tauri*. Plant Physiology and Biochemistry 48(6):407-416.

Xing D. and Q. Li (2011). Alternative polyadenylation and gene expression regulation in plants. Wiley Interdisciplinary Reviews RNA 2(3):445-458.

4.5 APPENDIX

Table A. Solutions for TAP medium

Stock solution	For 1 Liter (mL)
1M tris base	20
Phosphate buffer II	1
Hutner trace metals	1
Solution A	10
Glacial acetic acid	1

Table B. Components for solutions used in TAP medium

Phosphate buffer II		Hutner trace metals	
Components	**For 100 mL (g)**	**Components**	**For 500 LM (g)**
K_2HPO_4	10.8	BO_3H_3	5.7
KH_2PO_4	5.6	$ZnSO_4 \cdot 7\ H_2O$	11
Solution A		$MnCl_2 \cdot 4\ H_2O$	2.53
Components	**For 500 mL (g)**	$FeSO_4 \cdot 7\ H_2O$	2.495
NH_4Cl	20	$CoCl_2 \cdot 6\ H_2O$	0.085
$MgSO_4 \cdot 7\ H_2O$	5	$CuSO_4 \cdot 5\ H_2O$	0.785
$CaCl_2 \cdot 2\ H_2O$	2.5	$Mo_7O_{24}(NH_4)_6 \cdot 4\ H_2O$	0.5