Title

The sugar-responsive alga, Chromochloris zofingiensis provides insight into photosynthesis, sugar signaling, and thylakoid biogenesis

Author

Westcott, Daniel James

Publication Date

2020

The sugar-responsive alga, *Chromochloris zofingiensis* provides insight into photosynthesis, sugar signaling, and thylakoid biogenesis

By

Daniel Westcott

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:
Professor Krishna Niyogi, Co-chair
Dr. Melissa Roth, Co-chair
Professor Anastasios Melis
Professor Henrik Scheller
Professor Daniel Fletcher

Fall 2020

Abstract

The sugar-responsive alga, *Chromochloris zofingiensis* provides insight into photosynthesis, sugar signaling, and thylakoid biogenesis

by

Daniel Westcott

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Krishna Niyogi, Co-Chair

Dr. Melissa Roth, Co-Chair

Carbon fixation by photosynthetic organisms provides the bulk of energy that most other life depends on. Many photosynthetic organisms can shift metabolism in order to metabolize other sources of energy when necessary. The unicellular, Chlorophycean alga, *Chromochloris zofingiensis*, is one example of a photosynthetic organism with metabolic flexibility. When supplemented with glucose, *C. zofingiensis* will utilize both photosynthesis and glucose as an energy source. In specific culture conditions, supplemental glucose causes the total loss of photosynthetic capacity. Additionally, this phenomenon is reversible, and *C. zofingiensis* is able to rapidly regenerate the photosynthetic apparatus if glucose is removed. This observation, that glucose can trigger a reversible collapse of photosynthesis, provides the opportunity to investigate fundamental metabolic and signaling processes and the genes required to rebuild photosynthetic capacity. Chapter 1 introduces *C. zofingiensis,* its scientific history, and places it in the context of current research.

Chapter 2 presents results of a two-phase experiment, in which we drive *C. zofingiensis* into a state of photosynthetic repression using glucose in the first phase. In the second phase, we release that repression by removing glucose. We observe the severe impact on photosynthetic capacity, where oxygen evolution stops occurring and fluorescence signatures of a functional photosystem decrease to near zero. The proteins and lipids required for photosynthesis are greatly diminished. Cells grow larger in glucose, but thylakoid membranes in the chloroplast shrink, while storage molecules in the form of triacylglycerols and starch increase, as well as the keto-carotenoid, astaxanthin. In the second phase, when these cells are transferred into a glucose-free medium, photosynthetic repression is released, and the cells restart

photosynthesizing within 12 hours and continue to regreen over the course of 48 hours. We analyze and observe gene expression during critical metabolic transitions and find a wholesale transcriptional reordering in glucose conditions that results in the differential expression of 41% of the genes in the genome. We find that glycolysis, lipid metabolism, chlorophyll biosynthesis, and the expression of photosynthetic proteins are tightly controlled by the presence or absence of glucose. This chapter presents *C. zofingiensis* as a model system to understand metabolic flexibility between photoautotrophic and heterotrophic modes, furthering our understanding of the regulation of algal metabolism. This work will facilitate engineering efforts to reroute metabolism towards beneficial bioproducts for energy, food, and potentially pharmaceuticals.

In Chapter 3, we seek to understand genetic determinants of trophic flexibility in *C. zofingiensis.* We utilize a genetic selection to identify mutants that are unable to shift their trophic strategy when supplemented with glucose. UV-mutagenized cells were plated on medium containing the glucose analog, 2-deoxy-D-glucose, which induces photosynthetic repression, but is unable to be metabolized by *C. zofingiensis.* We identify potential causative mutations using whole-genome sequencing and find that all eight mutants generated and investigated have mutations in the critical glucose-phosphorylating enzyme, hexokinase (*HXK1*). We further characterize two of these independently generated mutants, which share the same *hxk1* mutation. When treated with glucose, photosynthetic capacity, oxygen evolution, and proteins related to photosynthesis of *hxk1* mutants are all much more similar to non-glucose-treated wild-type *C. zofingiensis.* The expression of genes representative of key pathways described in chapter 1, such as *HXK1,* a light-harvesting antenna protein (LHC16), A fatty acid de-esterase *(FAD2),* and the major lipid-droplet storage protein (*MLDP1)*, is unresponsive to glucose. This provides the basis for our model of glucose-induced repression of photosynthesis, where the cellular response to glucose is mediated by HXK1.

Finally, in Chapter 4, we describe the potential of *C. zofingiensis* as an inducible model system of oxygenic photosynthesis to study thylakoid biogenesis. Utilizing a comparative genomics framework coupled with extant co-expression data for genes of the model plant *Arabidopsis thaliana* and data from a mutant library in model green alga *Chlamydomonas reinhardtii*, which contains many potential photosynthetic defects, we identify genes likely to be essential for thylakoid biosynthesis. By comparing sets of *C. zofingiensis* genes that are differentially expressed during the transition from glucose to non-glucose conditions to the *A. thaliana* and *C. reinhardtii* gene sets, we identify sets of genes, many of which are well-known components of thylakoid biogenesis, whereas others have roles that have yet to be described in thylakoid biogenesis. This work utilizes a recently developed bioinformatic tool called OrthoLang and

presents a systematic approach to the integration of disparate data types, including RNA-seq data, a sequenced mutant library, and co-expression analysis.

In summary, *C. zofingiensis* is emerging as a powerful model organism to study multiple aspects of photosynthetic metabolism. Signals that rewire metabolism in the presence of glucose pass through the HXK1 and result in global photosynthetic repression. Release of that repression allows *C. zofingiensis* to rapidly recover photosynthesis, providing an additional window into thylakoid biogenesis.

Dedication

I have resisted the reductionist approach to life as strongly as I could. Life is a continuation of moments, each strung together with no discernible border between them. As much as we might like, life is not a series of discrete events that can be pulled apart examined in minute detail under a dissecting microscope. The hand that reaches through to the future is connected to the arm in the past. To become a reductionist is to specialize, and I have never wanted to be a specialist. It forces one to define a region of focus, thereby excluding other important, potentially critical parts of life from that region of focus.

So, I entered into my graduate studies as a committed generalist. A jack-of-all-trades; an amateur in many disciplines, but master of none. I've worked a variety of jobs. I've been a cook, office manager, grocer, home health care aid, landscaper, construction worker, and farmer. I play music. I can proficiently play guitar, drums, bass, upright bass, and less proficiently, piano. I play in bands and alone, for myself, and sometimes for others. I build things. I've remodeled a kitchen, built a 300-square foot studio, built furniture, instruments, toys, and bicycles. I love to be active. I skateboard, ride BMX, snowboard, hike, and occasionally run.

I care about things. Politics, the fate of the environment, climate, species diversity, my child's growth and development; I care. I want reach through to a future with more opportunities for love and learning than the past. Where creativity flourishes and the stifling power of poverty is abolished. I want to build this future with a trunk rooted in the lessons of history. And I want a present where each moment, a deep breath of ocean air, an arm outstretched in the warmth of the sun, is considered precious.

And, even with these lofty sentiments, I'm reminded daily of the impossible expectations we put on ourselves to achieve a better future. An individual, awash in a world of systems and structures, can seem to have little impact. It can feel futile to try to improve it at the best of times, and depressingly insane at others. The endless supply of grist continues to be ground, and progress is slow.

It was necessary to ask myself what purpose would volunteering for the reductionist, and incremental approach be? Would I be another drop in the ocean, minuscule, but additively increasing the breadth of our knowledge? How would that purpose feel after five or six years of in-depth specialization? Absent my participation, would my role have been filled by other eager students, unencumbered by these epistemological questions? Would the work have continued apace? Viewing one's own life through the lens of reductionism leads to the idea that I, myself can be reduced to my critical parts. I am only one example of a greater phenomenon; that the individual is unimportant, but the patterns of behavior of individuals *is.*

When you dissect a phenomenon into its component parts, you may understand those components. However, you've lost any emergent properties of the whole. This is the fundamental problem of reductionism. At what scale do you stop reducing? How tiny of a mechanism do you need to understand to re-assemble a coherent model of the whole? How large of a whole do you want to model? These questions vex me. Reductionism can best be analogized to the use of a flashlight in the dark. You can shine

a light – elucidate – a small region of the darkened word, and examine it. If you adjust the lens, and widen the beam, the dimmer light covers a wider field of view.

It is with this frame of mind that I come to the task at hand. Throughout this work, I may write about phenomena and mechanisms with a firm, near absolute, tone. But it must be said now that I am only working on the best available knowledge that I have, using the best available tools that I have had access to, and working with the best available minds that have been kind enough to think alongside me.

It is these minds that have allowed me to forge ahead in times of doubt, and to them who I dedicate this work. When I begin to distrust the power of our models, the broader impacts of our work, and the feasibility of implementing our ideas, it was my fellow scientists who brought me back to center. The engagement and curiosity of my peers reminds me that the scientific endeavor is not simply about making a difference. It is just as often about enabling curiosity. When I soured on my ability to speak with clarity on topics of my expertise, my family simply asked me how my work was going. If I could explain it in a way that my six-year-old understood, I knew that I was understanding too. I had reduced my work to smaller bits, and spoken about the purpose without confusion. Without the support of my family and my peers, I would not have been able to complete this effort. The natural world is crackling with phenomena. From water use in redwoods, to the dogged behavior of photons, there is so much to see and understand. But our minds cannot assimilate disparate bits of data all together. We must tell ourselves short stories to foster our understanding. Let me tell you the story of my life as a reductionist.

Finally, I must also dedicate this dissertation to the memory of my father, who, although he did not graduate high school, was an inspiration to my eternal curiosity. He was incredibly proud that I was able to pursue my doctorate degree. Sadly, he passed away two months before he could witness this goal realized.

Table of Contents

List of figures

Figure are listed in the format chapter.figure_number. For example, figure 3 from chapter 2 would be Figure 2.3

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of tables

Tables are listed in the format chapter.table number. For example, table 1 from chapter 3 would be Figure 3.1

Chapter 1

Chapter 3

Chapter 4

Chapter 1

An Introduction to *Chromochloris zofingiensis*

The unicellular alga *Chromochloris zofingiensis* is seemingly commonplace. It is small, ranging between 3 and 10 μm in diameter, spherical, and perches on an unremarkable branch in the tree of life (Figure 1). It does have a single trait that has consistently drawn the attention of a few curious scientists since the 1930s. While it is a green alga and fixes carbon using CO_2 through processes common to oxygenic photosynthetic organisms, it turns a rusty orange under certain conditions such as age, stress, or nutrient variation. While this simple color change may have been the initial motivation for the isolation and study of *C. zofingiensis,* the underlying biology of this organism is much more remarkable.

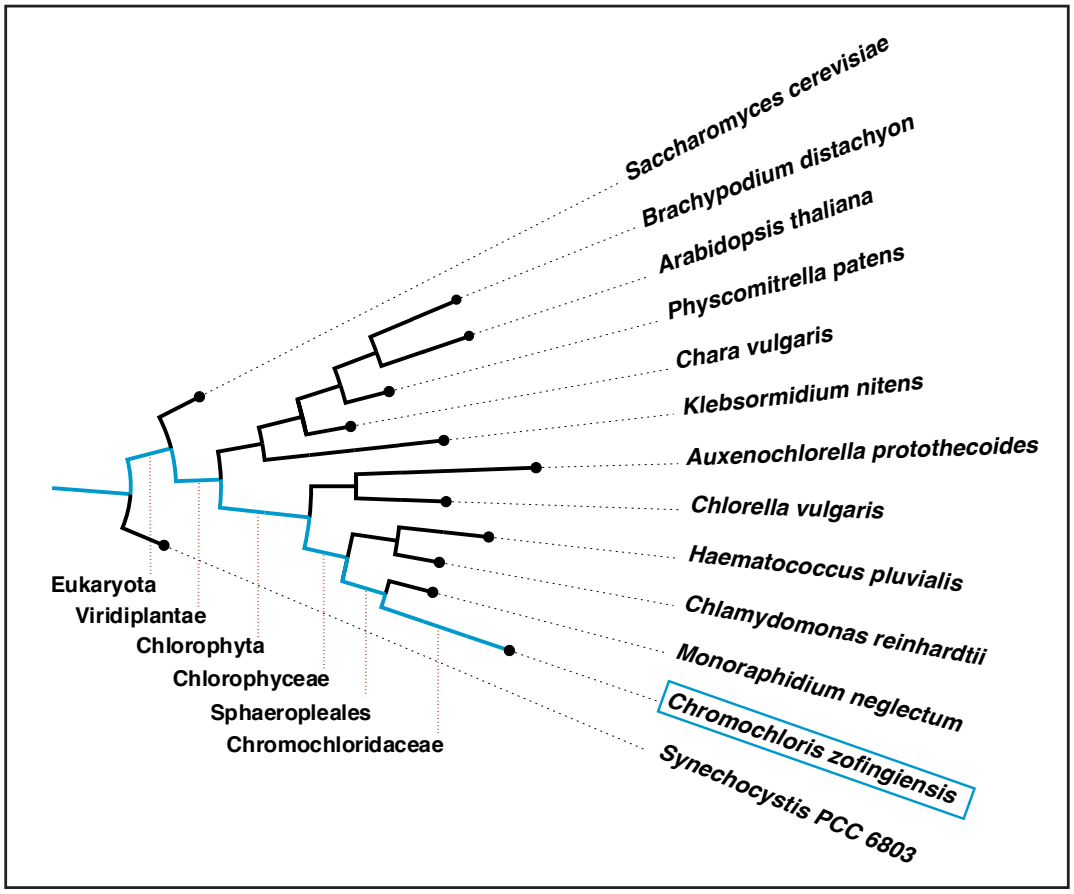


Figure 1. Species tree displaying the taxonomy of *Chromochloris zofingiensis,* which is the only known genus of the class Chromochloridaceae. Branch lengths are not to scale. Tree was built using publicly available data and the NCBI Common Taxonomy Tree tool. The image was generated using FigTree and further annotated in Adobe Illustrator.

In the original description of *C. zofingiensis*, it was the reddish color that initially drew the attention of its first documented investigator, Christ Dönz (Dönz, 1934).

> *During a field trip on February 1933 at the Ramooswald in Zofingen I saw on a sandy path conspicuously colored flecks in which I found a reddish, single-celled alga recognized under microscopic analysis.*

Dönz went on to describe this alga in detail, including information about the growth and physiological changes of *C. zofingiensis* on glucose. This dissertation presents the biological underpinnings of the trophic flexibility and pigment accumulation initially described by Dönz. I go on to relate this metabolic capacity to photosynthesis and cellular signaling pathways. And further, I expand into a novel direction for this alga by presenting it as a model organism to study thylakoid biogenesis. In 1934, Dönz likely used microscopes and culturing conditions that were state of the art to create his detailed drawings of *C. zofingiensis*, which are remarkably accurate. In this work, I use sets of tools and techniques available to the modern molecular biologist to further our understanding of photosynthesis, metabolic flexibility, and thylakoid biogenesis.

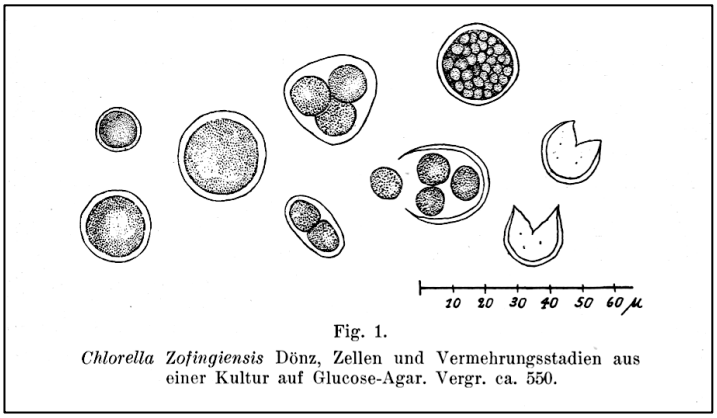


Figure 2. Drawings from Dönz's original isolation of *C. zofingiensis.*

An historical overview of the study *C. zofingiensis*

The ability of *C. zofingiensis* to accumulate significant amounts of energy-dense lipids and the valuable keto-carotenoid astaxanthin has driven a tremendous spike in scientific interest, but this attention is a relatively recent phenomenon. According to molecular phylogenetic analysis, this alga was independently isolated and added into culture collections on at least six occasions (Table 1) The original descriptive and taxonomic publication was in May of 1934 by Dönz who dubbed it *Chlorella zofingiensis* (SAG 211-14, UTEX 32, Dönz, 1934), taking its species name from the nearby town of Zofingen, Switzerland. Later, in September of that same year, and 200 km away, the same algal species was independently isolated and named *Chromochloris cinnabarina* (Kol & Chodat, 1934, 1934, SAG 221-2, UTEX 56, CCAP 221/2). Presumably, the algae reminded the authors of the bright brick red of mercury sulfide, or cinnabar. A poorly documented isolate was deposited in the SAG culture collection by a Johns Hopkins researcher, M.B. Allen, sometime before 1952 (SAG 4.80) In 1963, an isolate deemed *Muriella zofingiensis* was collected in Ortenberg, Germany by F.C. Czygan from the trunk of a deciduous tree (SAG 34.80, Kessler & Czygan, 1965). In France, a strain labelled *Bracteacoccus cinnabarinus* was isolated by GML Patterson from a polluted

stream (CCAP 211/51, Patterson, 1978). An additional strain was isolated from the Croatian Island of Lavsa in 1979 by Kurt Schwarz of the University of Innsbruck (Schwarz, 1979). This strain was termed *Bracteacoccus minutus*, and unlike previous isolates, Schwarz describes a motile phase of the life cycle. Although we report on the theoretical capacity of *C. zofingiensis* to engage in sexual reproduction, I have not specifically attempted to induce sexual reproduction and have not witnessed a sexual phase (Roth et al., PNAS 2017). Unfortunately, no type specimen of *B. minutus* exists.

Table 1. Isolation events, naming history, and available strains of *C. zofingiensis.* Strain origin details from Fučikova & Lewis, 2012. ***** Denotes heterotypic strain name. ****** *C. vulgaris* (SAG 4.80), while listed as a strain of *C. zofingiensis,* is likely to be misidentified. See text for details.

Current species	Isolated as	Isolated or reclassified by (i/r)	Where	When	SAG Strain	UTEX Strain	CCAP Strain
C. zofingiensis	*Chlorella zofingiensis*	Dönz (i)	Zofingen, Switzerland	1934	211-14	32	211/14
	Chromochloris cinnabarina *	Kol & Chodat (i)	Scuol, Switzerland	1934	221-2	56	221/2
	Dictyococcus zofingiensis	Vischer (r)	Scuol, Switzerland	1936	221-2	56	221/2
	Dictyococcus cinnabarinus	Starr (r)	Scuol, Switzerland	1955	221-2	56	221/2
	Mychonastes zofingiensis	Kalina & Punčochárová (r)	Zofingin, Switzerland	1987	211-14	32	211/14
	C. vulgaris **	M.B. Allen	freshwater	1952?	4.80 (previously C-1.2.1)		
	Muriella zofingiensis	Czygan	Ortenberg, Germany	1963	34.80 (previously 211-14b)		
	Bracteacoccus cinnabarinus *	Patterson	France, polluted stream	1978			211/51
	Bracteacoccus minutus *	Schwarz	Lavsa, Croatia	1979	No type specimen	No type specimen	No type specimen

The taxonomic history of *C. zofingiensis* reflects its multiple isolations and non-descript morphology. Shortly after isolation, *C. zofingiensis* was transferred to the genus *Dictyococcus* by Vischer (1936) and later to *Bracteacoccus* (Starr, 1955). Additional, independent isolations resulted in multiple names for the same organism. This confusing situation was clarified using sequence-based methods (18S, *rbcL*), and disparate isolates were collapsed into the single species *Chromochloris zofingiensis* (Fučiková & Lewis, 2012). Historical synonyms of *C. zofingiensis* include the homotypes

Chlorella zofingiensis, Mychonastes zofingiensis, and *Muriella zofingiensis,* and the heterotypes *Chromochloris cinnabarina, Dictyococcus cinnabarinus, Bracteacoccus minutus,* and finally *Bracteacoccus cinnabarinus* (NCBI TaxID 31302,Fučíková & Lewis, 2012). As a result of the multiple isolation events, maintenance in multiple culture collections, and taxonomic reclassification, there are currently nine strains available to the phycological community. However, to my knowledge, these only represent five independent isolations.

The trophic flexibility and the accumulation of orange pigments in *C. zofingiensis* has been remarked upon on many occasions. In 1961, *C. zofingiensis* was described as a facultative heterotroph able to grow on glucose (Parker et al., 1961). Shortly thereafter, reports focused on the heterotrophic capacity and carotenogenic phenotype of *C. zofingiensis* (as *Dictyococcus cinnabarinus*), providing biochemical details about carotenoid and lipid accumulation (Accadia Di et al., 1966; di Accadia et al., 1965). Notably, identification of orange-colored, ketolated carotenoids such as canthaxanthin, as well as β-carotene was reported. The authors had some difficulty conclusively determining all carotenoids. Di Accadia et al. (1965) described the accumulation of echinenone, which was not found in subsequent studies. This pigment profile was later expanded to include astacene, a ketocarotenoid with a nearly identical structure to astaxanthin (Sassu, 1972). The ketocarotenoid astaxanthin is responsible for the orange color of *C. zofingiensis* when grown in high light, glucose, or poor nutrient media (Breuer et al., 2012b; Roth et al., 2017). It is likely that astaxanthin was misidentified as astacene. Astaxanthin and astacene differ only in that two hydroxyl groups are deprotonated, leaving carbonyl groups at the 3 and 3' positions of the terminal carotenoid rings.

In the 1970s, the study of *C. zofingiensis* was taken up by D.L. Lynch's lab at Northern Illinois University (Lynch & Fenwick, 1974). Like others before, this group described the de-greening of *C. zofingiensis* (as *B. cinnabarinus,* UTEX-56) when grown in media containing glucose, observing an increase in carotenoid accumulation (Lynch & Fenwick, 1974). However, they further examined the effect of supplemental carbon when *C. zofingiensis* was grown on potassium acetate or sodium acetate in both light and dark conditions, reporting that potassium acetate did not support growth in either light or dark conditions while sodium acetate supported robust growth (Lynch & Fenwick, 1974). They followed up this descriptive work with a study of the ultrastructure of cells grown heterotrophically using electron microscopy, finding that potassium acetate-grown cells accumulate lipids and display lamellar disorganization in the chloroplast (Hornung et al., 1977; Rentas, 1980).

Interest in *C. zofingiensis* remained low, but persistent, over the next 20 years. An upper growth temperature of 28°C was determined (Erich Kessler, 1985). In an effort to resolve the phylogeny of multiple algal isolates, Kessler included four strains of *C. zofingiensis,* SAG 211-14a, SAG 211-14b, SAG 211-14c, and SAG 4.80 (referred to as C-1.2.1). Kessler noted that the three SAG 211 strains were orange when grown at 28°C, but SAG 4.80 was not. Deposition details are unclear about SAG 4.80, only stating that it was deposited sometime before 1952, but providing no information about

where. To my knowledge this strain has not been verified as *C. zofingiensis* using sequence-based techniques.

In the 1990s the pigment profile of *C. zofingiensis* under nitrogen stress was reported (Rise et al., 1994). This pigment profile appears to be consistent with later work (Mulders et al., 2015; Roth et al., 2017; Roth, Gallaher, et al., 2019). They additionally examined changes in pigment profiles under differing nitrogen conditions, demonstrating that carotenoids accumulate to their highest concentration when cells are grown in high light in nitrogen-limiting conditions. This lab produced some of the most detailed transmission electron microscopy (TEM) images of *C. zofingiensis* to date, both with and without nitrogen stress (Bar et al., 1995). Starch, lipid, and astaxanthin all accumulate under treatment conditions. Oddly, two variables were assessed in a single experiment, comparing the combination of high light and nitrogen stress to a control condition of low light, replete with nitrogen. It is difficult, therefore, to deconvolve the contribution of nitrogen or light stress to the described phenotype. Regardless, the authors put forward a three-stage model of lipid and pigment accumulation, suggesting two stages of carotenogenesis followed by a lipid accumulation phase, which is hypothesized to internally shade excess light (Bar et al., 1995).

An additional, and heretofore unconsidered, thermotolerance phenotype was described in 1997 showing that a heat-shocked strain of *C. zofingiensis* (SAG 34.80/211-14b) was able to grow under supra-optimal temperature conditions (Shen & Lee, 1997). This expanded the 28°C limit determined in previous work to 35°C for growth, with the ability to withstand shocks up to 70°C for 20 minutes (Kessler, 1985) Heat tolerance could be an important adaptive trait given that *C. zofingiensis* (as *B. minutus*) has been reported to be found in the soil crust of the central desert of Baja California, Mexico (Flechtner et al., 1998). There is reason to doubt this finding due to the non-descript morphology of *C. zofingiensis.* Microscopy was the sole technique used to discriminate samples in this work, and the presence of *C. zofingiensis* in desert ecosystems remains to be verified.

From its isolation in 1934 up to the turn of the 21st century, *C. zofingiensis* maintained small, but dedicated pockets of admirers. The first 70 years of study saw multiple, disconnected groups finding the trophic flexibility and pigment profile to be interesting biological phenomena worthy of study. The next 20 years of biological study brought the blossoming suite of experimental tools, techniques, and intellectual frameworks to bear, which were readily applied to *C. zofingiensis,* leading to a flowering of inquiry and understanding in this organism.

Advantages of *C. zofingiensis* as a model system

The properties of *C. zofingiensis* make it an ideal experimental organism for a number of different goals. First, the trophic transitions between photoautotrophic, mixotrophic, and heterotrophic modes provide a suitable model to study the fundamental cellular needs of each mode and the sets of shared or distinct genes that allow these transitions to occur. Second, its tendency to accumulate high concentrations

of lipids and astaxanthin make it an attractive candidate for industrial oil and astaxanthin production. Finally, the ability to regreen after nutrient stress provides an opportunity to study the biogenesis of thylakoid membranes and the photosynthetic apparatus. Current work on the biology of *C. zofingiensis* can generally fall into one of these four categories: trophic flexibility, signaling response to glucose, industrial production of triacylglycerols and astaxanthin, and thylakoid biogenesis.

Trophic flexibility in algae and *Chromochloris zofingiensis*

The ability of algae to shift metabolic strategies according to available energy sources is widespread. Interest in basic science, as well as applications in biofuels and bioengineering, have driven this research. Algae utilizing diverse carbon compounds were studied as far back as 1926, when growth enhancements were observed in *Scenedesmus costulatus* supplemented with glucose, galactose, maltose, or sucrose (Roach, 1926). Other carbon sources such as pyruvate, lactic acid, and acetate are also able to support growth in *Chlorella*, and acetate is commonly used to maintain cultures of *Chlamydomonas reinhardtii* (Eny, 1950; Eppley & MaciasR, 1962; Shihira & Krauss, 1965). More recent research demonstrates the importance of metabolic flexibility to robust growth in the fast growing desert alga, *Chlorella ohadii* (Treves et al., 2017).

C. zofingiensis, is more closely related to *C. reinhardtii* than *Chlorella* (Figure 1). However, its response to glucose is similar to that observed in *Chlorella.* The influence of glucose on the pigmentation of *Chlorella* was described in an early survey of 41 algal species (Shihira & Krauss, 1965). It was this work that provided the background for the investigation of a phenomenon that became known as “glucose bleaching.” Ishikawa & Hase (1964) reported that glucose-grown *Auxenochlorella* (formerly *Chlorella*) *protothecoides* showed a marked decrease in chlorophyll content, resulting in a pale phenotype. Chloroplast RNA degradation, chlorophyll biosynthesis, and lipid content were subsequently analyzed in these cells (Matsuka & Hase, 1965; Matsuka & Miyachi, 1974; Mihara et al., 1968; Oh-hama & Hase, 1975; Oh-hama & Senger, 1978).

Like other algae, *C. zofingiensis* can grow by photosynthetically fixing carbon from CO_2, as well as by using a variety of carbon sources. Our work has mainly focused on glucose, but I have also observed that *C. zofingiensis* can utilize other hexoses, as well as acetate (Table 2, unpublished data). In general, our research group uses *C. zofingiensis* Dönz, strain SAG 211-14.

Table 2. *C. zofingiensis* (strain SAG 211-14) is able to grow on alternative carbon sources on agar plates (unpublished data).

Carbon Source	Growth in the dark	Pigment change in light	Pigment change in dark
Glucose	✓	✓	✓
Fructose	✓	✓	✓
Sucrose	✓	✓	✓
Maltose	✓	✓	X
Galactose	✓	✓	X
Acetate	✓	X	X
Mannitol	X	X	X

Investigation of the physiological and mechanistic explanations of trophic flexibility has led to insights into the interplay between light energy and glucose. For example, the effect of light on glucose-fed cultures was studied, and it was observed that dark-grown *C. zofingiensis* accumulates 14% more lipids than light-grown, glucose-fed cultures (53.1% and 35.2% dry weight, respectively, Chen et al., 2015). This group has also described the role of oxidative stress in heterotrophic cultures, finding that the level of reactive oxidative species (ROS) correlates well with astaxanthin levels, independent of culture conditions, and that the inhibition of fatty acid synthesis results in increased astaxanthin accumulation (Z. Zhang et al., 2016). A transcriptional analysis of the response of *C. zofingiensis* during TAG and astaxanthin accumulation provides some preliminary insight into critical pathways. Unfortunately, time resolution is absent, and the authors' use of a non-public reference genome makes these findings difficult to replicate (W. Huang et al., 2016). While insightful, these studies give little attention to the specific role of photosynthesis during trophic transitions, focusing instead on the accumulation of biomass, astaxanthin, and the carotenoid biosynthetic pathway.

The transition from a photosynthetic growth mode to a mixotrophic or heterotrophic growth mode requires a wholesale reordering of cellular processes. Our work to understand what underpins the trophic flexibility of *C. zofingiensis* is addressed in detail in Chapter one (Roth, Gallaher, et al., 2019). We present a time-resolved analysis of *C. zofingiensis'* transition from photoautotrophic growth to mixotrophic growth. We use transcriptomics, lipidomics, protein analysis, electron microscopy, HPLC, and various photosynthetic parameters to assess this biological process. Additionally, we continue this analysis to observe the transition from heterotrophy back to photoautotrophic growth. This work spotlights the physiological transition between trophic modes and offers insight into potential engineering targets in *C. zofingiensis* and other algae. Additionally, the time-resolved expression data set serves as a rich source

of information for the advancement of basic scientific understanding, as demonstrated in chapter 3.

Signaling responses to glucose

The signaling pathways required to monitor and respond to varying energy sources in algae are not well understood. It has been shown previously that glucose induces the expression of some carotenogenic genes and that the use of glucose analogs can abolish this activity (Yantao Li et al., 2008). This behavior is characteristic of a signaling pathway. However, a deeper study of how *C. zofingiensis* monitors energy and nutrient status can serve to advance our basic understanding of sugar signaling and will provide critical, actionable insight if we aim to engineer improvements in growth or bioproducts.

Repression of photosynthesis in response to glucose was also observed in *Chlorella* when treated with the glucose analog 2-deoxy-D-glucose (2-DOG) (Semenenko, 1981; Vladimirova, 1976; Zvereva et al., 1981). As 2-DOG enters the glycolytic pathway, it becomes a metabolic dead end (Wick et al., 1956). Hence, 2-DOG can trigger a similar response to glucose without providing any useful energy. Use of this molecule and other glucose analogs has been the basis for efforts to determine specific mechanisms of glucose sensitivity in algae.

For example, 2-DOG was used to investigate the phenomenon known as glucose bleaching in *Chlorella* (Semenenko, 1981). *Chlorella sp.* K.1 (likely *C. vulgaris*) showed a decrease in the 77K fluorescence spectra of chlorophyll peaks characteristic of both PSII and PSI 8 hours after treatment with 2-DOG (Semenenko, 1981). Glucose induced a reduction of RNA synthesis as measured by the incorporation of ^{14}C-labeled uracil, which was reversible upon removal of glucose (Chemeris et al., 1991; Zvereva et al., 1981). Within 10 hours of the removal of 2-DOG, *Chlorella* began recovery of PSII-associated chlorophyll fluorescence. By 30 hours post-removal, it had recovered to near-untreated levels (Semenenko, 1981).

In plants, interest in this topic led to solid hypotheses of the mechanism of glucose-induced photosynthetic repression. But, to date this mechanism remains to be fully described. In the 1990s, the group of Jen Sheen began an investigation of the metabolic repression of photosynthesis in vascular plants. Initially, a maize protoplast system was used to demonstrate that photosynthesis-related genes are repressed in the presence of glucose (Sheen, 1990). In *Arabidopsis*, this repression was shown to be controlled by the glucose-phosphorylating kinase, hexokinase (HXK) (J. C. Jang & Sheen, 1994). Physiologically relevant concentrations of glucose resulted in decreased chlorophyll content in *Arabidopsis* (Xiao et al., 2000). Critical residues necessary for photosynthetic repression were identified in AtHXK1 (Moore et al., 2003). The crystal structure of AtHXK1 provided details about both the ATP binding pocket and residues necessary to coordinate glucose phosphorylation (Feng et al., 2015). This provided a structural body of support for the preceding genetics studies. Taken together, these results demonstrate that the kinase activity of AtHXK1 is not necessary to induce

photosynthetic repression suggesting that *At*HXK1 propagates a glucose-mediated signal.

In non-photosynthetic organisms, the role of HXK is understood in greater detail than that in plants (Ahuatzi et al., 2004, 2007; Fernández-García et al., 2012; Peláez et al., 2009; Rodríguez et al., 2001; Vega et al., 2016). The major *Saccharomyces cerevisiae* hexokinase, *Sc*Hxk2, moves into the nucleus when cells are grown in glucose (Fernández-García et al., 2012). This nuclear translocation is dependent upon the presence of the zinc-finger transcription factor Mig1 (Ahuatzi et al., 2004). *Sc*Hxk2 binds Mig1 in high-glucose conditions and regulates the transcription of glucose-responsive genes including its homolog *Sc*Hxk1 (Ahuatzi et al., 2007; Rodríguez et al., 2001). This shuttling between cytoplasm and nucleus is regulated by the phosphorylation state of *Sc*Hxk2 (Fernández-García et al., 2012). The regulatory and metabolic functions of this enzyme were demonstrated to be distinct using mutant alleles that lacked regulatory function or catalytic function (Peláez et al., 2009, 2012). It remains to be demonstrated how analogous this regulatory network is to algae and plants.

There have been compelling experiments that suggest that *At*HXK1, analogous to *S. cerevisiae*, also shuttles into the nucleus, where it impacts transcription of photosynthetic genes (Cho et al., 2006). According to this hypothesis, *At*HXK1 enters the nucleus where it interacts with two proteins, the β subunit of a vacuolar ATP-ase, VHA-B1, and a proteasomal regulatory particle, RPT5b (Cho et al., 2006). These proteins were not previously known to be localized to the nucleus, but in this study they were shown to interact, via yeast-2-hybrid screens, with two nuclear transcription factors (At1g5040, At3g11280, Cho, et al., 2006). This hypothesis underscores the signaling role of *At*HXK1 in glucose-induced photosynthetic repression.

However, an alternative hypothesis has been put forward that does not involve the shuttling of *At*HXK1 in and out of the nucleus (Balasubramanian et al., 2007, 2008; Karve et al., 2010). Using similar experimental techniques and the same antibodies and strains of plants, the authors came to a different understanding of the role of *At*HXK1. According to their hypothesis, *At*HXK1 is localized to the mitochondrial membrane via an N-terminal tether (Balasubramanian et al., 2007, 2008). When the N terminus of the protein is truncated, the protein freely diffuses throughout the cytoplasm, but it is not detectable in the nucleus. The authors explain the apparent difference in localization by pointing to the higher concentration of biomass used by Cho et al. and potentially less stringent cellular fractionation techniques (Cho et al., 2006).

The alternative hypothesis of glucose-based photosynthetic repression is based upon the interaction of *At*HXK1 with F-Actin (Balasubramanian et al., 2007, 2008). Fluorescently labelled *At*HXK1 in glucose-treated seedlings was shown to interact with actin. Short-term glucose-treated seedlings displayed disruptions in fine-actin filaments (Balasubramanian et al., 2007, 2008). As the description of the initial phenotype was seedling developmental arrest (Moore et al., 2003), the authors suggested that a disruption in fine mesh, F-actin filaments would explain this more adequately than a signaling defect. They propose two mechanistic possibilities that do not involve transcriptional control by *At*HXK1. First, depolymerization of actin may allow actin monomers to disrupt transcription, citing previously identified, nuclear-localized actin

related proteins (Kandasamy et al., 2004). Alternatively, a loss of actin-filaments would remove the substrate actin-associated polysomes or actin-tethered mitochondria, causing a disruption the translation of key genes or mitochondrial activity (Balasubramanian et al., 2007, 2008). Unfortunately, a similar analysis was not completed in plant strains with mutations in *At*HXK1, so it remains to be determined how the activity of this enzyme is responsible for photosynthetic repression (Balasubramanian et al., 2007, 2008). In algae, there has been some effort to bridge the gap between the levels of scientific understanding in *S. cerevisiae* and *A. thaliana.* HXK derived from the charophyte alga *K. nitens* is able to complement *A. thaliana gin2* mutants, but due to challenges in algal transformation, the generation of *hxk* mutants in *K. nitens* was not presented (Ulfstedt et al., 2018a). It remains an open question: does hexokinase induce photosynthetic repression directly by protein-protein interaction, or indirectly by generating downstream signaling molecules via glycolysis?

In Chapter two, we show that the repression of photosynthesis in *C. zofingiensis,* when grown in glucose, depends upon *Cz*HXK1. We use a forward genetics approach coupled with whole-genome sequencing to determine specific alleles of *CzHXK1* that do not repress photosynthesis in the presence of glucose. Additionally, we characterize the transcriptional response of specific marker genes involved in photosynthesis, glycolysis, lipid, and astaxanthin biosynthesis, as well as the lipid, protein, and pigment profiles, and we assess the photosynthetic capacity of these novel mutant strains. We demonstrate that *Czhxk1* mutants, unlike wild type, maintain photosynthesis when grown in glucose. Glucose is unable to diminish photosynthetic efficiency, induce carotenogenesis and lipid formation, or spark a transcriptional response in the absence of functional HXK.

Industrial production of astaxanthin and lipids

Although the accumulation of astaxanthin draws attention from industry, carotenoids play critical roles in photosynthesis, and astaxanthin is able to protect cells during stress conditions that generate excess oxidative species (X. Liu & Osawa, 2007). While not a specific focus of this dissertation, the potential use of *C. zofingiensis* as an industrial production strain must be remarked upon, as it has driven much of the biochemical pathway analysis work presented here.

The ketocarotenoid astaxanthin is of economic interest due to its use as a feed additive in aquaculture, food coloring, as well as its nutraceutical use as an antioxidant. Estimates vary, but the global market for astaxanthin was $250 million USD in 2009, and it was nearly $1 billion USD in 2019 (Jeffers & Roth, 2021; Voort et al., 2015). While synthetic astaxanthin is cheaply available, the stereochemistry of synthetic astaxanthin is a mix of enantiomers, whereas naturally produced astaxanthin is a single stereoisomer (Ishikawa et al., 2018; Jeffers & Roth, 2021, in press). The nature of astaxanthin stereoisomers, combined with the lack of esterification, results in a 50-fold reduction of the antioxidant activity of synthetic astaxanthin when compared to naturally occurring astaxanthin (Capelli et al., 2013; Ishikawa et al., 2018) (Capelli et al., 2014,

Ishikawa et al., 2019). Additionally, synthetic astaxanthin has not been approved for direct human consumption, precluding its use as a nutraceutical.

Two surveys of green algae completed in 2000 identified *C. zofingiensis* as a potential astaxanthin production strain due to its high culture density and high titer of astaxanthin upon high light and nitrogen starvation (Kopecký et al., 2000; Orosa et al., 2000). Stress-induced astaxanthin accumulation was further explored, demonstrating the effects of light, nitrogen concentration, temperature, and NaCl concentrations on the accumulation of astaxanthin (del Campo Castillo, 2000; Del Campo et al., 2004). These efforts achieved a maximal astaxanthin accumulation of 15.7 mg/L of culture. Other similar studies of the time report 12.5 and 10.3 mg/L maximum yield of astaxanthin (Ip et al., 2004; Ip & Chen, 2005).

Efforts to understand the particular mechanisms and enzymes related to the astaxanthin biosynthesis pathway revealed likely routes to astaxanthin accumulation, and potential differences between that of astaxanthin accumulation in *C. zofingiensis* and in *Haematococcus pluvialis* (Jeffers & Roth, 2021; Yantao Li et al., 2008, 2009). The common approach to this body of work was to change nutrient or lighting conditions in order to assess growth and carotenoid yield (Del Campo et al., 2004; Ip et al., 2004; Ip & Chen, 2005; Yantao Li et al., 2008, 2009; J. Liu et al., 2012; Orosa et al., 2000). Until recently, the highest astaxanthin content was reported under high-light conditions of 920 μmol photons m^{-2} s^{-1} while bubbling with 1% CO_2, resulting in 19 mg of astaxanthin per liter of culture (Del Campo et al., 2004). Other high levels of astaxanthin were achieved using glucose or molasses (13.4 and 14.14 mg/L respectively, Li et al., 2008) and nitrogen starvation (12.5 mg/L Ip et al., 2004). Remarkably, the addition of indole acetic acid to the culture medium increased biomass, lipid, and astaxanthin accumulation to levels higher than previously reported. This strategy resulted in astaxanthin levels of 89.9 mg/L and lipid accumulation to as much as 65% of the dry cell culture weight (J. hui Chen et al., 2020). Depending upon the culture conditions and the biomass accumulation of the particular experiment, these levels of astaxanthin make *C. zofingiensis* an attractive alternative source for natural astaxanthin.

Efforts to maximize culture density and pigment production are epitomized by strategies in which cultures are subjected to varying concentrations of glucose or light intensities in a nutrient broth (Tao Chen & Wang, 2013; Yantao Li et al., 2008; Sun et al., 2020; Z. Zhang et al., 2017). Others are interested in the biochemistry and genetic underpinnings of this biochemical pathway (W. Huang et al., 2016; Roth, Gallaher, et al., 2019; Z. Zhang et al., 2016, 2019; Zhong et al., 2011). The astaxanthin pathway and its variable expression during trophic transitions are addressed in Chapter 1.

The majority of natural astaxanthin is currently produced using batch cultures of *Haematococcus pluvialis*, a green alga that becomes rust-red upon high light and nitrogen starvation. Few organisms accumulate as high a level of astaxanthin as a percentage of dry weight, which can be as much as 4% (Boussiba, 2000), or relative to culture titer, which can reach levels as high as 175 mg/L (Kang et al., 2005). *C. zofingiensis* accumulates substantially less astaxanthin as a percentage of dry weight

(0.6%; Boussiba, 2000), but growth characteristics, growth density, and suitability for continuous culture make *C. zofingiensis* an enticing alternative astaxanthin producer. Additionally, the accumulation of astaxanthin in *C. zofingiensis* co-occurs with high levels of triacylglycerol (J. Liu et al., 2014a), *and C. zofingiensis* can grow to higher culture densities in shorter periods of time, meaning that the total productivity of *C. zofingiensis* rivals that of *H. pluvialis* (Jeffers & Roth, in press). It should also be noted that the genome of *C. zofingiensis* has fewer duplications and is much reduced compared to *H. pluvialis*, which could offer a less complex engineering system for increasing yield of both TAG and astaxanthin (Jeffers & Roth, in press). The carotenoid and lipid biosynthesis pathways of *C. zofingiensis* are addressed in Chapter 1.

In addition to astaxanthin biosynthesis, *C. zofingiensis* generates a high density of the biofuel precursor triacylglycerol (TAG, Breuer et al., 2013; Huang et al., 2016; J. Liu et al., 2014b; Mulders et al., 2014; Roth et al., 2017; Zhu et al., 2014).In a survey of 100 oil-producing algae, *C. zofingiensis* had a higher TAG production rate than 98 of the algae surveyed, producing 243 mg/L of TAGs per day after nitrogen depletion (Breuer et al., 2013). Others have reported similar productivity when using a mixotrophic strategy rather than nitrogen starvation, opening the possibility of continuous culture strategies (Z. Zhang et al., 2017). This high TAG productivity combined with the genetic resources developed in *C. zofingiensis* presents rich engineering targets for further increasing the viability of *C. zofingiensis* as a biofuel feedstock (Huang et al., 2016, Roth et al., 2017; Roth et al.,2020). In most culture conditions, those conditions that produce high levels of TAGs also produce high levels of astaxanthin, making *C. zofingiensis* an economically feasible alternative to the traditional astaxanthin feedstock *H. pluvialis.*

Thylakoid biogenesis

An additional area of investigation that is particularly well suited to *C. zofingiensis* is that of thylakoid biogenesis. The complex process by which an organism builds and organizes the membranes, proteins, and cofactors required to harvest light is critical to most life on Earth. However, our understanding of thylakoid biogenesis is incomplete.

To date, the majority of plant research in this area has used three types of experimental systems to investigate thylakoid biogenesis: etiolated seedlings, developmental gradients, and mutants with defects at stages of thylakoid development. Each experimental system has provided substantial insight into thylakoid development, and each system has inherent limitations.

In the case of etiolated seedlings, a newly germinated seed is kept in the dark for some period of time. During this dark acclimation, the seedlings generate a paracrystalline prolamellar body consisting of chlorophyll and membrane precursors (Lütz & Klein, 1979; Selstam & Sandelius, 1984; Sundqvist & Dahlin, 1997). This structure is rich in protochlorophyllide, galactolipids, carotenoids, and NADPH-dependent protochlorophyllide oxidoreductase, which are organized into an interconnected three-dimensional lattice (Murakami et al., 1985). When exposed to light,

this lattice begins to rapidly shift to prothylakoids and then to thylakoids on the time scale of hours (Virgin, 1963). Recently, a time-resolved, structural approach has shed light upon the distinct stages of the etioplast-to-chloroplast transition, showing that the structural organization of the prolamellar body is disrupted within milliseconds of the introduction of light (Kowalewska et al., 2016). Further, a structural examination of LPOR, the major protein component of the prolamellar body has revealed the biophysical mechanisms of this rapid shift (Nguyen et al., 2020)The etioplast and the prolamellar body are not universal to photosynthetic organisms. Therefore, insights gained in this system must be independently verified in mosses, algae, or cyanobacteria.

An additional model system used to understand thylakoid biogenesis is that of developmental gradients. This method requires the researcher to segment the growth stages of plants along a developmental axis, typically on a leaf. A study that typifies this approach uses *A. thaliana* leaves segmented perpendicular to the midvein which are observed using electron microscopy (Gügel & Soll, 2017). In this way, nascent chloroplasts can be compared to mature chloroplasts from the same leaf, demonstrating a continuum of thylakoid development. While this method is insightful in plants, similar to etioplast-based systems, there are significant challenges extrapolating findings to the thylakoid development of mosses, algae, and cyanobacteria.

The third common strategy to investigate thylakoid biogenesis uses mutant strains to seek phenotypes of interest. Mutant analyses have been important in determining the function of proteins such as VIPP1 and CURT1 in *A. thaliana*, CPSFL1 in *C. reinhardtii*, and a host of other proteins required for functional thylakoid development (García-Cerdán et al., 2020; Hertle et al., 2013; Kroll et al., 2001). This approach is broadly applicable to diverse types of photosynthetic organisms. However, mutations in genes critical to photosynthesis are often lethal, or can cause severe growth defects. If an organism is able to utilize alternate sources of carbon, this limitation can be overcome, but there may be additional effects introduced into the system when exogenous carbon is supplemented.

The metabolic flexibility of *C. zofingiensis* offers a novel, and complementary, approach to this topic. Similar to the system developed for the investigation of anoxygenic photosynthesis (Chory et al., 1984), *C. zofingiensis*' transition from heterotrophic to photoautotrophic growth modes is inducible. By acclimating cells to heterotrophic growth and subsequently transferring cells to a glucose-free medium, direct observation of thylakoid biogenesis in a wild-type genetic background can be achieved.

With this goal in mind, Chapter 3 discusses the potential of *C. zofingiensis* as a novel model system to investigate thylakoid biogenesis. Additionally, utilizing a newly developed comparative genomics tool called OrthoLang (cite Jeff's thesis), I use the rich transcriptional dataset described in Chapter 1 as a basis to identify genes critical to thylakoid biogenesis. By constraining these expression data to the time points during the re-greening process, I restrict the gene set to a suite of genes likely to be involved in thylakoid biogenesis. This set of genes is systematically compared using sequence alignment algorithms to *A. thaliana* genes that are co-expressed with genes known to

be essential to thylakoid biogenesis. Further, the *C. zofingiensis* expression data are compared to a newly sequenced *C. reinhardtii* mutant library consisting of mutants incapable of performing photosynthesis. Genes homologous to non-photosynthetic organisms are removed and the resultant sets of genes are described. As well as providing a framework for the comparison of disparate types of data, this approach identifies known and unknown genes related to thylakoid biogenesis that will further our understanding of this complex process.

Chapter 2

Regulation of Oxygenic Photosynthesis during Trophic Transitions in the Green Alga *Chromochloris zofingiensis*

Preface

As detailed in Chapter 1, the response of *C. zofingiensis* to glucose has continued to evoke interest from the scientific community. I was especially interested in the physiological response to glucose. While much attention has been focused on the increase of carotenoids and fatty acids during glucose, the effect of glucose on photosynthesis has largely been ignored in this organism. Our observations suggested that the cell completely reforms metabolism to accommodate this additional energy input. We had also observed a remarkable drop in photosynthetic capacity and chlorophyll fluorescence. The underlying physiological processes that explain these observations were not understood. Although efforts to understand glucose-induced chlorosis in other *Chlorella* were undertaken in the 1980s, modern molecular techniques have not previously been applied to understand this phenomenon (Semenenko, 1981; Zvereva et al., 1981). Here we report on a photosynthetic switch in *C. zofingiensis* which is tightly coupled to previous reports of metabolic flexibility

This work was originally published under a title of the same name in March of 2019 in the journal The Plant Cell. As this was a collaborative effort included in an individual dissertation, an explicit breakdown of the authors' contributions to the work is necessary. In the case of authors aside from myself, this statement is only as accurate as my memory, so kindly forgive any oversights.

Citation:

Regulation of Oxygenic Photosynthesis during Trophic Transitions in the Green Alga *Chromochloris zofingiensis*

Melissa S. Roth, Sean D. Gallaher, Daniel J. Westcott, Masakazu Iwai, Katherine B. Louie, Maria Mueller, Andreas Walter, Fatima Foflonker, Benjamin P. Bowen, Nassim N. Ataii, JunhaSong, Jian-Hua Chen, Crysten E. Blaby-Haas, Carolyn Larabell, Manfred Auer, Trent R.Northen, Sabeeha S. Merchant, Krishna K. Niyogi

The Plant Cell Mar 2019, 31 (3) 579-601; **DOI:** 10.1105/tpc.18.00742

Conception: Melissa Roth and Kris Niyogi conceived of the study.

Experimental design: The data from four main experiments were used in this study: 1) for RNASeq, 2) for soft x-ray tomography, 3) for transmission electron microscopy, and 4) for everything else including physiology, growth, HPLC, lipidomics, photosynthetic efficiency, oxygen evolution, Western Blots, and structured illumination microscopy.

Melissa Roth and Kris Niyogi designed experiments 1, 2, 3, and 4. Sean Gallaher assisted with designing experiment 1.

Experimental execution: Melissa Roth initiated and conducted the experiments for 1, 2, 3, and 4. Melissa Roth, Dagmar Lyska and Setsuko Wakao collected samples for experiment 1. Melissa Roth collected samples for experiments 2 and 3. **Melissa Roth and Daniel Westcott collected samples for experiment 4**. Melissa Roth measured culture density and OD. **For experiment 4, Daniel Westcott aliquoted appropriate volumes for later processing and either freeze those aliquots for downstream analysis or passed them along to colleagues for immediate processing.** Melissa Roth conducted photosynthetic efficiency and oxygen evolution during the time course. Masakazu Iwai processed and imaged samples for structured illumination microscopy during this time course.

Sample Processing: Melissa Roth extracted RNA for RNA-seq, extracted pigments and processed samples on the HPLC, extracted proteins and conducted the Western Blots, processed samples and conducted measurements for photosynthetic efficiency, and processed samples and conducted measurements for oxygen evolution. Melissa Roth and Andreas Walter processed samples for Soft-X-Ray tomography. Masakazu Iwai processed samples for structured illumination microscopy. Katherine Louie processed samples for MS/MS lipidomic and biomass analysis. Melissa Roth and Maria Mueller processed samples for electron microscopy.

Data Generation: Melissa Roth generated data for HPLC, photosynthetic capacity, oxygen evolution and protein analysis. Sean Gallagher generated data for RNA-seq. Katherine Louie generated lipidomic and biomass data. Masa Iwai generated structured illumination microscopy data. Andreas Walter generated Soft-X-Ray tomography data. Maria Mueller generated electron microscopy data. Crysten Blaby-Haas and Fatima Foflonker conducted gene annotation.

Data Analysis: Melissa Roth performed data analysis on data concerning cell culture (biomass, volume, and density), HPLC, photosynthetic capacity (Fv/Fm), oxygen evolution and consumption, protein, electron microscopy, soft-X-ray tomography, and RNA-seq. Sean Gallagher performed analysis on RNA-seq data including initial processing of raw reads, clustering analysis, principal components analysis, differential expression, and set analyses. Sean Gallagher also generated the underlying data used in the pathway analysis. **Daniel Westcott performed data analysis on lipidomics, RNA-seq expression data and cluster analysis, as well as generating the metabolic pathways used in Figures 10 and 11.** Benjamin Bowen aided in data analysis of lipidomics. Kris Niyogi and Melissa Roth generated the pathway used in Figure 9.

Figure generation: Melissa Roth generated Figures 1, 2 and 3. Melissa Roth, Maria Muller and Sean Gallaher generated Figure 4. Melissa Roth, Masakazu Iwai and Sean Gallaher generated Figure 5. Melissa Roth and Sean Gallaher generated Figures 6, 7 and 8. Kris Niyogi, Melissa Roth and Sean Gallaher generated Figure 9**. Daniel Westcott generated Figures 10 and 11 with input from Sean Gallaher, Melissa Roth and Kris Niyogi.**

Writing: Melissa Roth wrote the manuscript with input from all authors. **Results and Discussion sections related to enriched pathways and k-means clustering analysis were based on work by Daniel Westcott with editing and curation by Melissa Roth, Sean Gallagher, and Kris Niyogi. Results and Discussion sections related to lipidomics were based on work by Daniel Westcott with editing and curation by Melissa Roth and Kris Niyogi.** Individual experimental researchers were responsible for summarizing their methods sections.

Support and/or supervision: Kris Niyogi, Sabeeha S. Merchant, Trent R. Northen, Manfred Auer, and Carolyn Larabell provided material support, supervision, guidance, and editing.

Other: Nassim N. Ataii and Junha Song assisted with transmission electron microscopy segmentation, and Jian-Huan Chen assisted with soft x-ray tomography videos.

Abstract

Light and nutrients are critical regulators of photosynthesis and metabolism in plants and algae. Many algae have the metabolic flexibility to grow photoautotrophically, heterotrophically, or mixotrophically. Here, we describe reversible Glc-dependent repression/activation of oxygenic photosynthesis in the unicellular green alga *Chromochloris zofingiensis*. We observed rapid and reversible changes in photosynthesis, in the photosynthetic apparatus, in thylakoid ultrastructure, and in energy stores including lipids and starch. Following Glc addition in the light, *C. zofingiensis* shuts off photosynthesis within days and accumulates large amounts of commercially relevant bioproducts, including triacylglycerols and the high-value nutraceutical ketocarotenoid astaxanthin, while increasing culture biomass. RNA sequencing reveals reversible changes in the transcriptome that form the basis of this metabolic regulation. Functional enrichment analyses show that Glc represses photosynthetic pathways while ketocarotenoid biosynthesis and heterotrophic carbon metabolism are upregulated. Because sugars play fundamental regulatory roles in gene expression, physiology, metabolism, and growth in both plants and animals, we have developed a simple algal model system to investigate conserved eukaryotic sugar responses as well as mechanisms of thylakoid breakdown and biogenesis in chloroplasts. Understanding regulation of photosynthesis and metabolism in algae could enable bioengineering to reroute metabolism toward beneficial bioproducts for energy, food, pharmaceuticals, and human health

Introduction

Regulation of oxygenic photosynthesis and primary carbon metabolism in algae and plants is critical for net primary production of biomass on a global scale. Algae can regulate photosynthesis and metabolism in response to changes in light and nutrient availability. Many algae have flexible metabolism and can change from photoautotrophic to mixotrophic and/or heterotrophic growth. The popular model green alga *Chlamydomonas reinhardtii* (Chlorophyceae) grows autotrophically in the light and either mixotrophically with the carbon source acetate in the light or heterotrophically with acetate in the dark (Harris, 2009). *C. reinhardtii* remains green in the presence of acetate in both light and dark, and even dark-grown cells remain capable of carbon fixation via photosynthesis in the light (Merchant et al., 2007). The green alga *Auxenochlorella protothecoides* (Trebouxiophyceae), formerly known as *Chlorella protothecoides* (Huss et al., 1999), changes its pigmentation from green to yellow to colorless depending on the amount of nutrients and carbon source (Shihira and Krauss, 1963). *A. protothecoides* exhibits a dramatic degeneration of its chloroplast when grown with Glc, a phenomenon known as glucose-induced bleaching (Shihira-Ishikawa and Hase, 1964). More recently it was shown that Glc upregulates fatty acid (FA) synthesis genes and causes an accumulation of lipid droplets in *A. protothecoides* (Gao et al.,

2014). Insight into the molecular mechanisms that allow algae to increase biomass and accumulate high levels of lipids and starch when consuming exogenous reduced carbon will expand our knowledge of algal metabolism and improve the economic potential of algal biofuels and other bioproducts.

Sugars are critical metabolic and regulatory molecules in plants. Exogenous Glc, a preferred carbon source for many organisms, has been shown to repress photosynthesis while modulating plant growth and development (reviewed in Rolland et al., 2006; Sheen, 2014). Despite the early work on Glc-induced bleaching (Shihira and Krauss, 1963; Shihira-Ishikawa and Hase, 1964; Matsuka et al., 1969), a detailed understanding of Glc responses in green algae is presently lacking. There is renewed interest in improving algal biomass for both biofuels and bioproducts, and regulation and manipulation of algal growth under heterotrophic and mixotrophic conditions has become an active area of investigation.

Chromochloris zofingiensis (Chlorophyceae) is emerging as a model for investigating metabolic flexibility in green algae because it can be cultivated in multiple trophic modes (Chen et al., 2015b; Zhang et al., 2017) on a variety of carbon sources (Sun et al., 2008) and has a high-quality chromosome-level small (~58-Mbp) genome and transcriptome (Roth et al., 2017). Also, research on *C. zofingiensis* can benefit from its similarity to *C. reinhardtii*, a closely related alga that has been widely studied in terms of its biology, genetics, and biochemistry. *C. zofingiensis*, which was placed in Chlorella, Muriella, and Mychonastes before its current placement into *Chromochloris* (Fucikova and Lewis, 2012), is an ;4-mm unicellular, haploid, coccoid alga with a mitochondrial network and a single starch-containing chloroplast (Roth et al., 2017). In contrast to *C. reinhardtii*, *C. zofingiensis* can grow with Glc as its sole carbon source, and has an expanded number of hexose transporter genes, suggesting the ability to acclimate to a wide variety of environmental conditions (Suzuki et al., 2018). Under specific conditions such as heterotrophy, nitrogen deprivation, or high light, *C. zofingiensis* accumulates high amounts of the valuable secondary ketocarotenoid astaxanthin and biofuel precursors (Breuer et al., 2012; Liu et al., 2014; Mulders et al., 2014; Huang et al., 2016; Zhang et al., 2016; Roth et al., 2017). Astaxanthin has a variety of commercial applications such as in feed and cosmetics, and recent studies have highlighted its value as a pharmaceutical. Naturally produced astaxanthin is a potent antioxidant and anti-inflammatory agent that can improve human health and combat disease (Hussein et al., 2006; Yuan et al., 2011; Liu et al., 2014). *C. zofingiensis* has been recently recognized as one of the highest producers of the biofuel precursor triacylglycerol (TAG) (Breuer et al., 2012). This combination of valuable attributes has led to increased

biological and commercial interest in *C. zofingiensis*.

Using cellular physiology analyses, transcriptomics, lipid analyses, transmission electron microscopy (TEM), soft x-ray tomography (SXT), and structured illumination microscopy (SIM), we provide insight into Glc-dependent repression and activation of oxygenic photosynthesis that occurs in *C. zofingiensis* during trophic transitions. Within

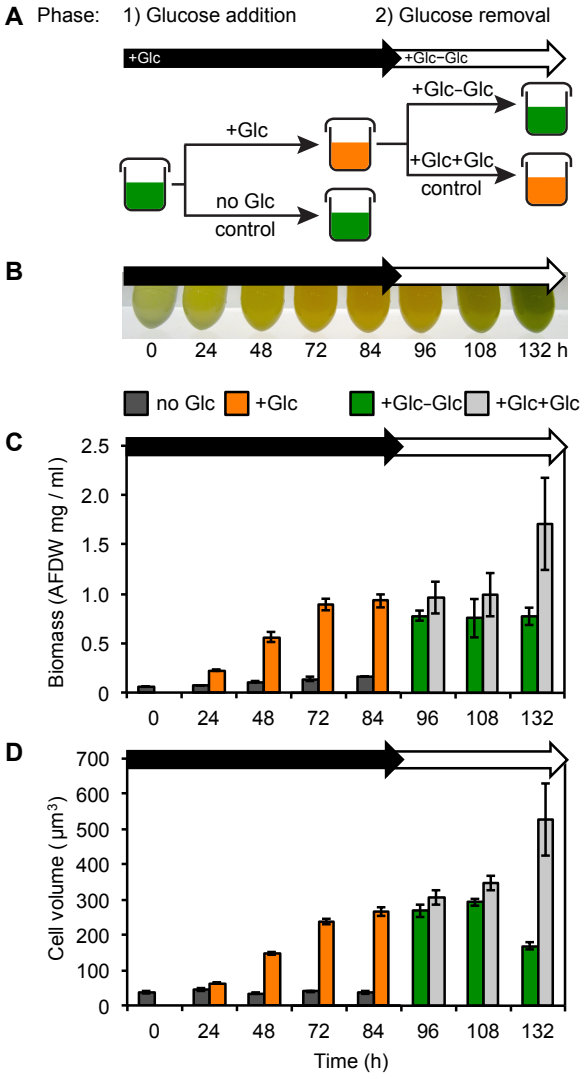


Figure 1. Overview of Glc Addition and Removal Experiment. (A) Schematic of the experimental design, showing the Glc addition (black arrows) and Glc removal (white arrows) phases (see "Methods"). The first phase of the experiment (0 to 84 h) involves Glc addition (+Glc, orange), with photoautotrophic controls with no Glc added (no Glc, dark gray). In the second phase of the experiment (84 to 132 h), heterotrophic cells (+Glc) are pelleted and resuspended in fresh medium without Glc (+Glc–Glc, green). Heterotrophic controls (no Glc removal) are pelleted and resuspended in fresh medium containing Glc (+Glc+Glc, light gray). **(B)** Representative concentrated C. zofingiensis cultures during the Glc addition/removal experiment. **(C)** Culture biomass (AFDW) dynamics during Glc addition and removal. **(D)** Cell density dynamics during Glc addition and removal. **(E)** Individual cell volume dynamics during Glc addition and removal. The following color scheme is used for (C) to (E): no Glc, dark gray; +Glc, orange; +Glc-Glc, green; +Glc+Glc, light gray. Data for **(C)** to **(E)** represent means 6 SD (n = 3 to 4 biological replicates).

a few days of Glc addition in the light, photosynthetic efficiency declines to zero, and oxygen evolution in the light ceases. There is a loss of the photosynthetic apparatus, and thylakoid ultrastructure changes dramatically as membranes become reduced and overall thylakoid volume decreases. We observed an increase in energy stores, including starch and lipids and in particular TAGs. These changes were reversed upon removal of Glc within 2 d. Genome-wide analyses over a 12-h time course of Glc addition and removal revealed broad, reversible transcriptomic changes involving a third of all gene models. Glc represses photosynthetic pathways and upregulates ketocarotenoid and FA biosynthesis and modification. Our work highlights the utility of *C. zofingiensis* as a simple emerging model system to investigate the regulation of

photosynthesis, thylakoid membrane dynamics, sugar responses, and changing metabolic modes in the green lineage.

METHODS

Culture Strain and Growth

We used the Chromochloris zofingiensis strain SAG 211-14 obtained from the Culture Collection of Algae (Göettingen University), whose genome was published recently (Roth et al., 2017). The cells were grown in liquid cultures shaking at 100 to 150 rpm under diurnal conditions (16-h-light/8-h-dark cycle) at a light intensity of 100 μmol photons m^{-2} s^{-1} (cool white spectrum) at 25°C, as previously described (Roth et al., 2017). Cells were grown in Proteose medium (UTEX Culture Collection of Algae, University of Texas, Austin) with Chu's micronutrient solution (2 mL/L; UTEX Culture Collection of Algae). For experiments, cells were grown in large, liquid, photoautotrophic preculture until cells reached exponential growth (~4 x 10^6 cells/mL). The cells were mixed and distributed into separate beakers the day prior to the experimental start time. Unless otherwise specified, cells were collected by centrifugation (at 3000 to 20,000g, for 2 to 10 min). Three to four biological replicates were used, and parallel samples started at the same time were used at the different time points. Cells were counted and sized with the Multisizer 3 Coulter counter (Beckman Coulter).

Two-Phase Glc Addition and Removal Experiment

Cells were grown as described above and distributed into 66 separate beakers. For the first phase of the experiment (0 to 84 h), Glc was added to photoautotrophic cultures in the light. To maintain Glc-replete conditions, 10 mM Glc was added every 24 h. Photoautotrophic cultures were maintained as no-Glc controls. Samples were collected at 0, 24, 48, 72, and 84 h. For the second phase of the experiment (84 to 132 h), Glc was removed from heterotrophic cultures in the light. After 84 h of Glc, heterotrophic cells were collected by centrifugation, the supernatant was discarded, and cells were resuspended in fresh medium without Glc. Heterotrophic cultures were also collected by centrifugation, the supernatant was discarded, and cells were resuspended in fresh medium with Glc as heterotrophic controls. Subsequently, 10 mM Glc was added every 24 h to heterotrophic controls. Samples were collected at 96, 108, and 132 h from the beginning of the experiment. For the Glc addition RNA-seq analysis, the liquid preculture of photoautotrophically grown cells was distributed into 44 separate beakers 1 d before the start of the experiment: enough for four biological replicates of each time point and condition. Samples were collected at 0.5, 1, 3, 6, and 12 h after Glc addition to photoautotrophic cultures and at 0, 0.5, 1, 3, 6, and 12 h in photoautotrophic controls. For Glc removal, the liquid preculture of photoautotrophically grown cells was distributed into 44 separate beakers 1 d before Glc was added, and cultures grew with Glc for 72 h. Samples were then collected from four biological replicate cultures at 0.5, 1, 3, 6, and 12 h after Glc removal from heterotrophic cultures and 0, 0.5, 1, 3, 6, and 12 h in heterotrophic controls. Glc was added or removed ;3 h after the start of the light

phase (8:00 AM), and all samples were collected during the light phase.

Measurement of Photosynthetic Efficiency

Photosynthetic efficiency was measured with the FMS2 system (Hansatech Instruments) as previously described (Roth et al., 2017), with modifications. After cells were dark acclimated while shaking for 30 min, 3.5 x 10^6 cells were collected onto a glass fiber filter that was then placed into the instrument's leaf clip. The maximum efficiency of PSII, $(F_m - F_o)/F_m = F_v/F_m$, was measured using a 0.5-s saturating pulse (at >2000 μmol photons m^{-2} s^{-1}).

Measurement of Oxygen Consumption and Net Oxygen Evolution

Oxygen consumption and net oxygen evolution were measured with the Oxygraph Plus System (Hansatech Instruments) as previously described (Peers et al., 2009), with modifications. Cell densities were adjusted to 10^7 cells/mL and dark acclimated for >30 min. Oxygen consumption was measured over 1 min after reaching steady state in the dark at a constant temperature of 25°C. Net oxygen evolution was also measured over 1 min after reaching steady state under actinic light of 100 μmol photons m^{-2} s^{-1} at a constant temperature of 25°C. For cultures with cell volumes larger than 100 mm3, additional samples with a cell density of 3.5 x 10^6 cells/mL were used in order to achieve an OD750 of ;0.2 to 0.4. Oxygen consumption and net oxygen evolution were normalized to cell volume and chlorophyll.

Identification of Photosynthetic Pigments

Photosynthetic pigments were identified by HPLC (1100 HPLC; Agilent) as previously described (Roth et al., 2017). Cells were pelleted and homogenized with acetone and lysing matrix D (MP Biomedical) for 3 3 60 s at 6.5 m s^{-1} with the FastPrep-24 5G High Speed homogenizer (MP Bio- medical). Cell debris was pelleted by centrifugation (at 20,000g, for 3 min), and the supernatant was collected. To ensure complete extraction, acetone extractions were repeated twice. Pigments were normalized to total chlorophyll per cell and to cell volume because the cell volume changed during the experiment.

Immunoblot Analysis

Cells were pelleted by centrifugation and homogenized in solubilization buffer (Tris-HCl, pH 6.8, 3.5% w/v SDS, 6% w/v urea, and 10% w/v glycerol) and lysing matrix D for 3 X 60 s with a FastPrep-24 5G High Speed homogenizer (at 6.5 m s–1; MP Biomedical). Chloroform:methanol protein purification was performed as previously described (Wessel and Flügge, 1984). Proteins were resolubilized in solubilization buffer containing 25 mM DTT. For analysis, 10 mg of protein was run with loading buffer on Any kD Mini-PROTEAN TGX gels (Bio-Rad), transferred to polyvinylidene difluoride membranes, and immunoblotted with anti-PsbD (1:5000; Agrisera), anti-PsbC (1:1000;

Agrisera), anti-PsaA (1:1000; Agrisera), anti-PetB (1:10,000; Agrisera), anti-P17 (1:5000; Bassi and Wollman, 1991), anti-Lhca2 (1:5000; Agrisera), or anti-AtpB (1:5000; Agrisera) antibodies. Because the AtpB antibody detects both chloroplast and mitochondrial F1b subunits, Expasy compute was used to discriminate between these two variants. CzCPg01090 was predicted to be ~50 kD and Cz03g32200 was predicted to be ~65 kD. Additionally, both PredAlgo and TargetP prediction software predicted an N-terminal mitochondrial targeting sequence for Cz03g32200. Proteins were visualized with an anti-rabbit IgG antibody (1:10,000; GE Healthcare) and ChemiDoc MP imaging system (Bio-Rad). Protein concentration was determined using the DC protein assay (BioRad), and equal protein levels were confirmed by Coomassie Brilliant Blue R 250 staining. Two biological replicates for each primary antibody were used. Supplemental Table 1 includes catalog information for antibodies.

Transmission Electron Microscopy

Cells were concentrated by centrifugation, loaded into 100-mm-deep specimen carriers, and ultra-rapidly frozen using a Bal-Tec HPM 010 high-pressure freezer without a filler. Specimen carriers containing the samples were transferred onto the surface of the frozen freeze-substitution medium (1% OsO4 and 0.1% uranyl acetate) and placed inside an EM AFS2 freeze substitution processor (Leica) equilibrated to 290°C using the following protocol: 38 h at 290°C, 32.5-h linear warm up to 225°C, 12 h at 225°C followed by a 5-h linear warm up to 0°C. After three to four quick acetone rinses over 5 to 10 min, the samples were removed from the carriers and cut into smaller pieces if needed. The samples were introduced to a graduated Epon Araldite resin series (6.2 g of Eponate 12, 4.4 g of Araldite 502, 12.2 g of dodecenyl succinic anhydride, and 0.8 mL of benzyl-dimethyl amine [BDMA; only added for the final 100% changes]) of 25, 50, and 75% resin: acetone. The following infiltration protocol was used: 25% for 1 h, 50% for 2 h, 75% for 4 h, 1 3 10 min rinse with 100% resin, 100% overnight, 2 3 10 min 100% + BDMA, followed by a final 4 h with 100% + BDMA. After infiltration, the samples were transferred to PELCO flat embedding molds and polymerized for 2 d at 60°C. Sections were cut 70 nm thick on a Reichert-Jung Ultracut E ultramicrotome, picked up on Formvar-coated slot grids, and poststained for 7 min in 2% aqueous uranyl acetate and 4 min in Reynold's lead citrate. Sections were examined and imaged on a FEI Tecnai 12 transmission electron microscope operated at 120 kV, and images were acquired on a Gatan 2Kx2K Ultrascan 1000 digital camera. Approximately six to eight technical replicates were used for 2D electron micrographs, and three technical replicates were used for 3D reconstructions. Electron micrographs of serial sections were recorded using Serial EM software package (Digital Micrograph 1.80.58 for GMS 1.8.0). To process the images, electron micrographs of serial sections were aligned using the 3dmod 4.7.12 software package. Segmentation and 3D reconstructions of cell organelles and internal membranes were performed using Amira (v5.3.3) manual segmentation tools. The histogram of thylakoid membrane length was binned as 0 to 50 nm, 51 to 100 nm, and so on and labeled as 50 to represent the number of membranes with lengths 0 to 50 nm.

Cryo-Soft X-Ray Tomography

SXT of cryopreserved C. zofingiensis was done as previously described (Le Gros et al., 2012, 2014). Briefly, cells were pelleted by centrifugation and loaded into custom-made, thin-walled glass capillaries that were previously dipped in a solution of 100-nm gold nanoparticles and subsequently used as fiducial markers for alignment of the x-ray projections. Capillaries with cells were cryopreserved by plunging the tip into an ~90K reservoir of liquid propane at 2 m s^{-1} using a custom-made fast-freezing apparatus. Soft x-ray tomographic data were acquired using the cryogenic soft x-ray microscope in the National Center for X-ray Tomography at the Advanced Light Source in Berkeley, California, as previously described (Roth et al., 2017). Projection images were collected at 517 eV using a Fresnel zone plate with a resolution of ~50 nm as the objective lens. For each data set, 90 projection images were acquired spanning a range of 180°. During data acquisition, the specimen was kept in a stream of helium gas that had been cooled to liquid nitrogen temperatures to maintain cryopreservation of the sample. Depending on the thickness of the specimen, exposure times for each projection image varied between 200 and 350 ms. 3D reconstructions of the x-ray projections were calculated using the software package IMOD (v4.7.12) after manually tracking fiducial markers on adjacent images for alignment. AMIRA (v5.3.3) was used on technical replicates (n = 8 to 11) to semiautomatically segment the 3D volumetric reconstructions into subcellular compartments (lipid droplets, chloroplasts, starch, and mitochondria) based on their different gray-level ranges. Segmentation of the nucleus was performed manually.

Structured Illumination Microscopy

To stain neutral lipids, the pelleted cells were resuspended with 5 mg/mL BODIPY 493/503 dye (Thermo Fisher Scientific) in Proteose medium as previously described (Iwai et al., 2018). The cells were incubated in the dark for 10 min and washed (three times) with 1 mL of Proteose medium and centrifuged (at 3000g, for 1 min). After washing, the pelleted cells were resuspended with 50 to 200 mL of 0.5% low-melting-point agarose prepared with Proteose medium, and then 4 mL of resuspended cells was immediately mounted between two microscope cover slips. After the agarose solidified, the mounted cover slips were placed in the Attofluor cell chamber (Thermo Fisher Scientific). C. zofingiensis live cells (technical replicates n = 8) were observed using a Zeiss Elyra PS.1 structured illumination microscope with objective lens Plan-APOCHROMAT 1003/1.46 (Zeiss). Chlorophyll and BODIPY were excited by 642 and 488 nm lasers, respectively, and fluorescence from each fluorophore was acquired through 650 to 730 nm and 505 to 550 nm band-pass filters, respectively. The image acquisition was done as fully controlled by ZEN software (v2012 SP5; Zeiss). Raw images were processed to reconstruct super-resolution 3D images using ZEN software.

Ash-Free Dry Weight

To measure the total amount of organic material, ash-free dry weight (AFDW) was determined. Pellets were transferred to pre-weighed aluminum pans and then weighed after drying at 80°C for 2 d and after combusting in an oven at 450°C for 4 h. The difference in pellet weight between these two drying conditions was used to calculate the AFDW (Schagerl and Müller, 2006).

Lipid Extraction

For lipid extraction, LC-MS grade solvents were obtained from Sigma Aldrich and HPLC grade chloroform from VWR. Prior to lipid extraction, cell pellets were lyophilized to dryness (FreeZone 2.5 Plus; Labconco) and powdered by bead beating with a 3.2-mm stainless steel bead in a bead beater (Mini-Beadbeater-96; BioSpec Products) for 5 s (three times). To extract lipids, a chloroform-based lipid extraction was performed using a modified Bligh–Dyer approach (Bligh and Dyer, 1959). Water (60 mL) was added to the powdered pellet, vortexed, and 600 mL of 2:1 methanol:CH3Cl (final ratio of 2:1:0.8 MeOH:CH3Cl:H2O) was added; this was followed by brief vortexing and incubating for 15 min in a sonic water bath (Scientific Aquasonic Water Bath model 150HT; VWR). An additional 200 mL of CH3Cl and 300 mL of water were added (final ratio of 1:1:0.9 MeOH:CH3Cl:H2O), followed by brief vortexing and incubating for 10 min in the sonic water bath. Samples were centrifuged (at 2655g, for 2 min), and 200 mL of the bottom lipid-enriched chloroform phase was transferred to a new tube. An additional 200 mL of chloroform was added, followed by another round of sonication and centrifugation, and the bottom chloroform phase was combined with the previously collected extract. Chloroform extracts were dried in a SpeedVac (SPD111V; Thermo Fisher Scientific) and stored at 220°C.

Lipid Analysis

LC-MS/MS was performed on lipid extracts. Extracted lipids were re-suspended in 3:3:4 isopropanol:acetonitrile:methanol (IPA:ACN:methanol), with the resuspension volume varied for each sample to normalize by AFDW and the resuspension solvent containing a 4 mM internal standard mixture of deuterium-labeled lipids (1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1'-rac-glycerol-1',1',2',3',3'-d5), 1,2-dipalmitoyl-sn-glycero-3-O-4'-[N,N,N-trimethyl(d9)]-homoserine, oleic acid, 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-[phospho-L-serine], 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphate, 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphoethanolamine, 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphocholine, 1,3-di-(9Z-hexadecenoyl)-2-hydroxy-sn-glycerol-d5, 1,3(d5)-diheptadecanoyl-2-(10Z-heptadecenoyl)-glycerol; Avanti Polar Lipids, Inc.; Supplemental Table 1). Ultra HPLC (UHPLC) reverse phase chromatography was performed using an Agilent 1290 LC stack, with MS and MS/MS data collected using a Q Exactive Orbitrap MS (Thermo Fisher Scientific). Full MS spectra were collected from m/z 80 to 1200 at 70,000 resolution in both positive and negative modes, with MS/MS fragmentation data acquired using stepped 10-, 20-, and 30-eV collision energies at 17,500 resolution. Chromatography was performed using a

C18 column (ZORBAX Eclipse Plus C18, Rapid Resolution HD, 2.1 3 50 mm, 1.8 mm; Agilent) at a flow rate of 0.4 mL/min with a 2-mL injection volume. Samples were analyzed using a gradient on the C18 column (55°C) that was first equilibrated with 80% buffer A (40:60 water:acetonitrile with 5 mM ammonium acetate and 0.1% formic acid) for 1.5 min, buffer A reduced to 45% with buffer B (90:10 IPA:ACN with 5 mM ammonium acetate and 0.1% formic acid) over 2.5 min, then to 20% A over 6 min where it was held for isocratic elution for 2 min, then reduced to 0% A over 1.5 min, and finishing with isocratic elution in 100% buffer B for 3.5 min. Samples consisted of four biological replicates each and four extraction controls, with sample injection order randomized and an injection blank (2 mL of 3:3:4 IPA:ACN:methanol) run between each sample.

Lipid Identification and Analyses

Exact mass and retention time coupled with MS/MS fragmentation spectra were used to identify lipids. For a detected compound, its lipid class was determined based on characteristic fragment ions or neutral loss, coupled with exact mass to determine specific lipid identity (number of carbons in FA tails and degree of unsaturation) (Murphy, 2015). XCalibur software (Thermo Fisher Scientific) was used to visualize raw data, and MetAtlas, a custom software created in-house, was used to extract peak heights and corresponding MS/MS spectra of detected lipids from the raw data. In positive mode, DGTS lipids ionized as [M+H]+ with a characteristic fragment ion, m/z 236.1; MGDG and DGDG lipids as [M+NH4]+ with a neutral loss of 179 D and 341 D, respectively; and TAGs as [M+NH4]+ with FA tails detected in the MS/MS spectra. In negative mode, SQDG ionized as [M2H]2 with a characteristic fragment ion, m/z 225.01. Lipid standards were run separately for analysis of MGDG, DGDG, and SQDG lipids, and deuterated internal standards DGTS and TAG were used to verify fragmentation pattern and elution times of each lipid class. Supplemental Table 1 includes catalog numbers of standards. MS results of identified lipids are summarized in Supplemental Data Set 4. Distribution frequencies of each lipid class (e.g., MGDG 34:1) were calculated for each time point. Those with a distribution of less than 0.005 D, (0.5%) at all time points, were excluded from analysis. Lipids were normalized to AFDW.

RNA Preparation and Quality Assessment

Total RNA was extracted from C. zofingiensis as previously described (Roth et al., 2017; Gallaher and Roth, 2018). Cells were collected by centrifugation (at 3200g, for 5 min at 4°C) during daytime, frozen in liquid nitrogen, washed with cold (-20°C) ethanol on dry ice, and ethanol was removed by centrifugation (at 2200g, for 3 min at 4°C). To break the cells open, they were homogenized with lysing matrix D on dry ice for 2 3 60 s with the FastPrep-24 homogenizer (6.0 m s–1; MP Biomedicals). Lysis buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM EDTA, 2% SDS, and 1 mg/mL Proteinase K) was added, samples were vortexed and incubated for 3 min at room temperature, and cell debris was pelleted by centrifugation (at 20,000g, for 3 min). One milliliter of sample

was added to 10 mL of TRIzol in a MaXtract high-density tube and incubated for 3 min at room temperature. To extract RNA, 1/5 volume of chloroform was added, samples were vigorously shaken, incubated for 5 min at room temperature, and phases were separated by centrifugation (at 800g, for 5 min at 22°C) and decanting. Total RNA was precipitated by adding cold ethanol on the aqueous phase and purified using the miRNeasy mini kit (Qiagen). RNA was eluted with diethylpyrocarbonate-treated water and cleaned using an ethanol precipitation step (100% ethanol and 85 mM sodium acetate, pH 8.0), centrifugation (at 15,000g, for 5 min at 4°C), and ethanol washing. The pellet was briefly air-dried and resuspended in diethylpyrocarbonate-treated water. RNA concentration and integrity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and Agilent 2100 bioanalyzer (RNA-seq samples).

RNA-Seq Analysis

RNA-seq analysis was performed as described previously (Roth et al., 2017; Gallaher and Roth, 2018), with some modifications. In brief, RNA was depleted of rRNA by means of the RiboZero rRNA Removal Kit for plant leaves (Illumina). The remaining RNA was used as input for RNA-seq library preparation by means of the KAPA stranded RNA-seq kit (KAPA Bio-systems). Libraries were sequenced with 50-nucleotide single-end reads on a HiSeq 2500 sequencer (Illumina).

Adaptor sequence was trimmed from the resulting sequencing reads by means of scythe (v0.981), and 39 base calls with Phread scores < 30 were trimmed with sickle (v1.210). The trimmed reads were then aligned to the C. zofingiensis reference genome (v.5.2.3.2) available from Phytozome (http:// phytozome.net) by means of RNA-STAR (v2.4.0j). Default settings were used except for –alignIntronMax 5000. Aligned reads were then assigned to C. zofingiensis gene models with the featureCounts program of the R subread package (v1.12.6) within the R statistical computing platform (v3.4.4). The following settings were used: countMultiMappingReads= TRUE, fraction=TRUE, ignoreDup=FALSE, strandSpecific=2, useMetaFeatures= TRUE, fracOverlap=0.5. Adjusted counts per gene, DEGs, and regularized log -transformed read counts were calculated using the DESeq2 package 2 (v1.18.1) in R.

During the analysis, a number of samples were determined to be outliers due to problems that occurred during the library construction. Contaminating DNA in RNA-seq libraries was identified by a high percentage of intergenic reads and reads on the noncoding strand. This was quantified with Rsubread featureCounts. Overamplification of the libraries was identified as a high percentage of nuclear read duplicates, which were calculated using Picard tools MarkDuplicates (v1.138). Samples with more than 50% reads aligned intergenically, or with more than 1 SD above the mean of nuclear read duplicates were excluded from further analysis. Collectively, these filters excluded 11 of 88 samples.

In the course of this analysis, we identified an ~40-kb region of chromosome 3 (approximately chr03:1,425,408-1,465,327) that appears to be erroneously duplicated on chromosome 9 (approximately chr09:2,290,532-2,339,298) in the current genome assembly (v5.2.3.2). This duplication would suggest that Cz09g23020 and Cz03g14080

should instead be a single gene located on chromosome 3. Because both loci are identical, the RNA-seq reads that had aligned to these regions mapped equally well to either locus. To address this, we masked the duplicated portion of chromosome 9 and excluded it from the analysis. Expression estimates in terms of FPKMs were calculated in R. PCAs of the regularized $\log_2$ -transformed counts for the 500 genes (Supplemental 2 Data Set 5) with the highest variance were performed using the prcomp feature in R and plotted with ggplot2 (v2.2.1). Heatmaps were generated in R with the ggplot2. k-means clustering was performed in R with the amap package (v0.8-14). BioCyc pathway annotations were downloaded from Phytozome's BioMart (https://phytozome.jgi.doe.gov/biomart). Functional enrichments of BioCyc pathways were calculated in R using the hypergeometric distribution with BH multiple testing correction and a false discovery rate of 0.05. Venn diagrams were produced in R with the VennDiagram package (v1.6.20).

Functional annotations for pathway analysis in Figures 9, 10, and 11 were performed manually based on relatedness to annotated proteins from *Arabidopsis* (*Arabidopsis thaliana*) and *Chlamydomonas reinhardtii*, unless noted otherwise in Supplemental Data Set 3. When possible, gene/protein symbols were chosen based on orthology to named *C. reinhardtii* proteins.

In cases of gene duplication or ambiguous orthology, making one-to-one correspondence difficult, the root of the protein symbol from *C. reinhardtii* was used and homologs were numbered. Some exceptions apply and are noted in Supplemental Data Set 3. To highlight chloroplastic processes in Figure 14, lipid biosynthesis and FA modification exclude predicted mitochondrial isoforms. When appropriate, transit peptide predictions from PredAlgo (for details, see Roth et al., 2017) were used to assess localization (Supplemental Data Set 3).

Data Availability and Accession Number

The raw metabolomics data will be made available at the Joint Genome Institute Genome Portal (https://genome.jgi.doe.gov/portal/). The raw sequencing data and analyzed RNA-seq results are available at the National Center for Biotechnology Information (https://www.ncbi.nlm.nih. gov/) Gene Expression Omnibus archive (https://www.ncbi.nlm.nih.gov/ geo/) and can be accessed using the accession number GSE92519.

RESULTS

Two-Phase Glc Addition and Glc Removal Experiment

To assess the effects of Glc on photosynthesis in *C. zofingiensis*, we conducted a two-phase experiment of Glc addition and subsequent removal in the light (Figure 1A). In the first phase (0 to 84 h), we added Glc to photoautotrophic cultures and maintained parallel cultures without Glc as photoautotrophic controls (see "Methods"). In the second phase (84 to 132 h), we removed Glc from heterotrophic cultures and provided fresh

media without Glc. Additionally, we maintained parallel cultures where we removed Glc and provided fresh media with Glc as heterotrophic controls. Cultures were maintained in diurnal light (at 100 mmol photons m^{-2} s^{-1}) on a 16-h-light/8-h-dark cycle. Green cells turned orange in the presence of Glc due to an increase in ketocarotenoids, and these cells reverted to green after Glc removal (Figure 1B). Because previous studies have characterized Glc induction of ketocarotenoid biosynthesis and, in particular, astaxanthin accumulation in *C. zofingiensis* in both the light and dark (Ip et al., 2004; Ip and Chen, 2005; Li et al., 2008; Sun et al., 2008), we focused here on Glc-induced changes in photosynthesis, metabolism, and energy stores in the light. Whereas culture biomass and cell size increased with Glc, cell density in cultures increased more slowly with Glc, suggesting a decrease in the rate of cell division during the 0 to 84 h time course of our experiment (Figures 1C to 1E). Physiological, lipid, transcriptomic, and structural changes were analyzed during both Glc addition and Glc removal phases.

Glucose Causes a Reversible Decline in Photosynthesis and the Photosynthetic Machinery

Glc caused a rapid decline in photosynthetic activity that recovered after Glc was removed. The maximum efficiency of PSII (F_V/F_m) decreased from 0.66 ± 0.01 to 0.01 ± 0.01 with Glc within 84 h (Figure 2A) and recovered after Glc removal. Furthermore, net oxygen evolution (production in the light) decreased from 9.4 ± 0.3 to 21.9 ± 0.4 nmol O_2 min^{-1} mm^{-3} with Glc (Figure 2B, results normalized to cell volume). Within 48 h, oxygen consumption in heterotrophic cultures surpassed oxygen production in the light. The changes in net oxygen evolution were reversed after Glc removal. There was a decrease in chlorophyll per unit cell volume (Figure 2D), and there was also a loss of net oxygen evolution per unit of chlorophyll (Figure 3A). These data suggest that residual chlorophyll is not associated with functional reaction centers. Respiration (oxygen consumption in the dark) increased quickly upon Glc addition and remained high after Glc removal (Figure 2C), suggesting it may be governed by a different mechanism than the regulation of photosynthetic activity. The net oxygen evolution in cultures following Glc removal was lower than in photoautotrophic cultures because respiration remained high after removing Glc. Using the classical calculation of net photosynthesis (net oxygen evolution in the light minus oxygen evolution in the dark; Kirk, 2010), photosynthesis decreased from 11.1 ± 0.07 to 0.3 ± 0.07 nmol O_2 min^{-1} $mm^{-3}$with Glc within 84 h and then recovered to 4.9 ± 0.6 nmol O_2 min^{-1} mm^{-3} 132 h after Glc removal.

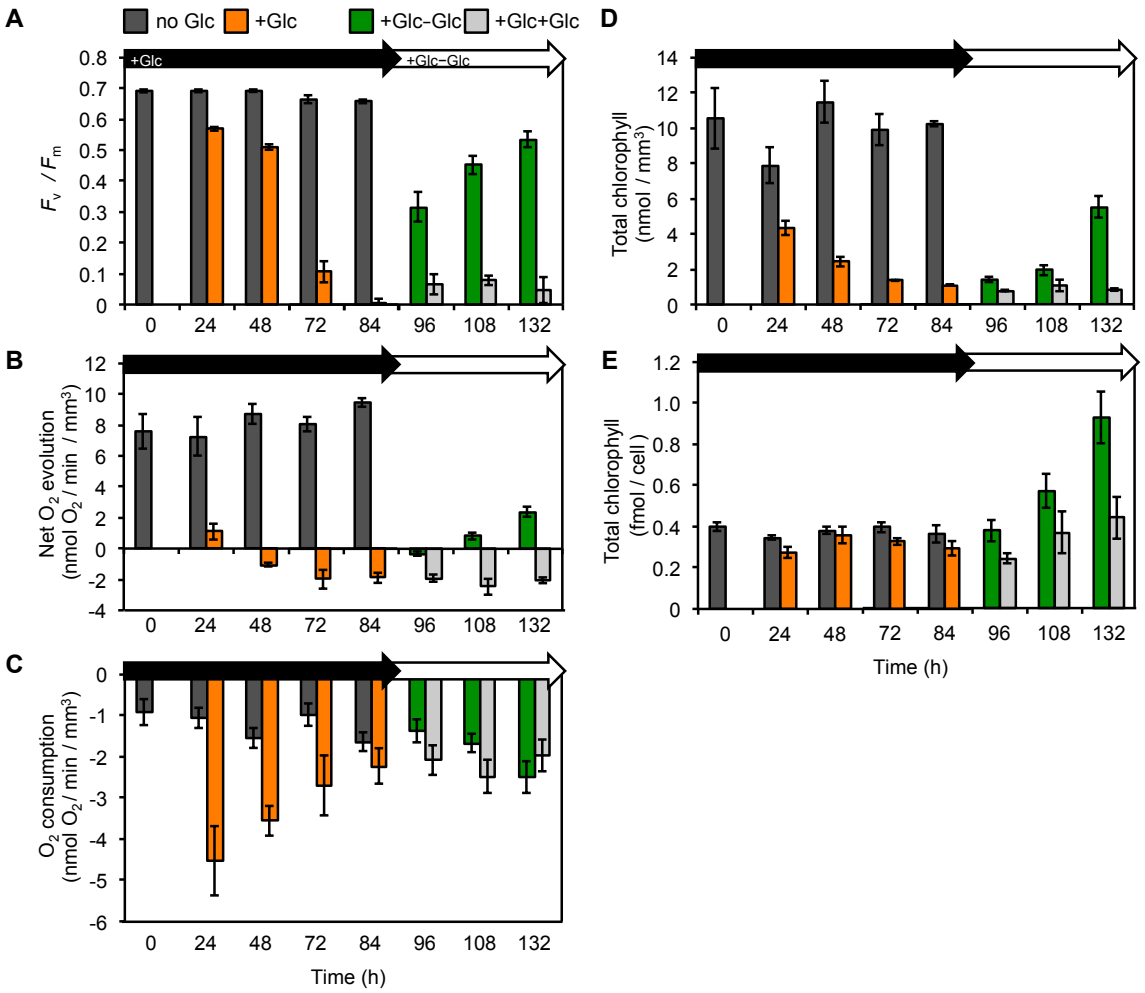


Figure 2. **Glc Causes Reversible Repression of Photosynthesis. (A)** Maximum photosynthetic efficiency (Fv/Fm) dynamics during Glc addition and removal. **(B)** Dynamics of net oxygen evolution (oxygen production in the light, at 100 mmol photons m22 s21) during Glc addition and removal (data normalized to cell volume). See Supplemental Figure 1 for data normalized to chlorophyll. **(C)** Dynamics of oxygen consumption in the dark during glucose addition and removal. See Supplemental Figure 1 for data normalized to chlorophyll. **(D)** Total chlorophyll (chlorophylls a and b) dynamics during glucose addition and removal normalized by cell volume. See Supplemental Figure 1D for HPLC chromatograms. **(E)** Total chlorophyll (chlorophylls a and b) dynamics during glucose addition and removal normalized per cell. Data were tested for significance ($P < 0.05$) using two-tailed t-test in JMP (v8.0). See Supplemental Figure 1D for HPLC chromatograms. (F) Photosynthetic pigments relative to total chlorophyll during Glc addition and removal. Data were tested for significance ($P < 0.05$) using two-tailed t test in JMP (v8.0). See Supplemental Figure 1D for HPLC chromatograms. The following color scheme is used for **(A)** to **(F)**: no Glc, dark gray; +Glc, orange; +Glc-Glc, green; +Glc+Glc, light gray. Data for **(A)** to **(F)** represent means 6 SD (n = 3 to 4 biological replicates).

Underlying the reversible changes in photosynthesis were substantial changes to components of photosynthetic complexes. When normalized to cell volume, total

chlorophyll (chlorophylls a and b) at 84 h decreased nearly 10-fold from 10.3 nmol chlorophyll mm^{-3} in photoautotrophic cells to 1.1 nmol chlorophyll mm^{-3} in heterotrophic cells (Figure 2D). However, at the same 84-h time point, the amount of total chlorophyll per cell was not significantly different between photoautotrophic and heterotrophic cultures (0.36 ± 0.04 and 0.29 ± 0.03 fmol $cell^{-1}$, respectively; two-tailed t test, P = 0.056; Figure 2E). These data suggest that chlorophyll was diluted rather than degraded as cells grew larger. After Glc removal, the amount of chlorophyll (normalized per cell or cell volume) increased substantially by the end of the Glc removal phase, although the ratio of chlorophyll a:b was relatively constant during the experiment (Figure 3C).

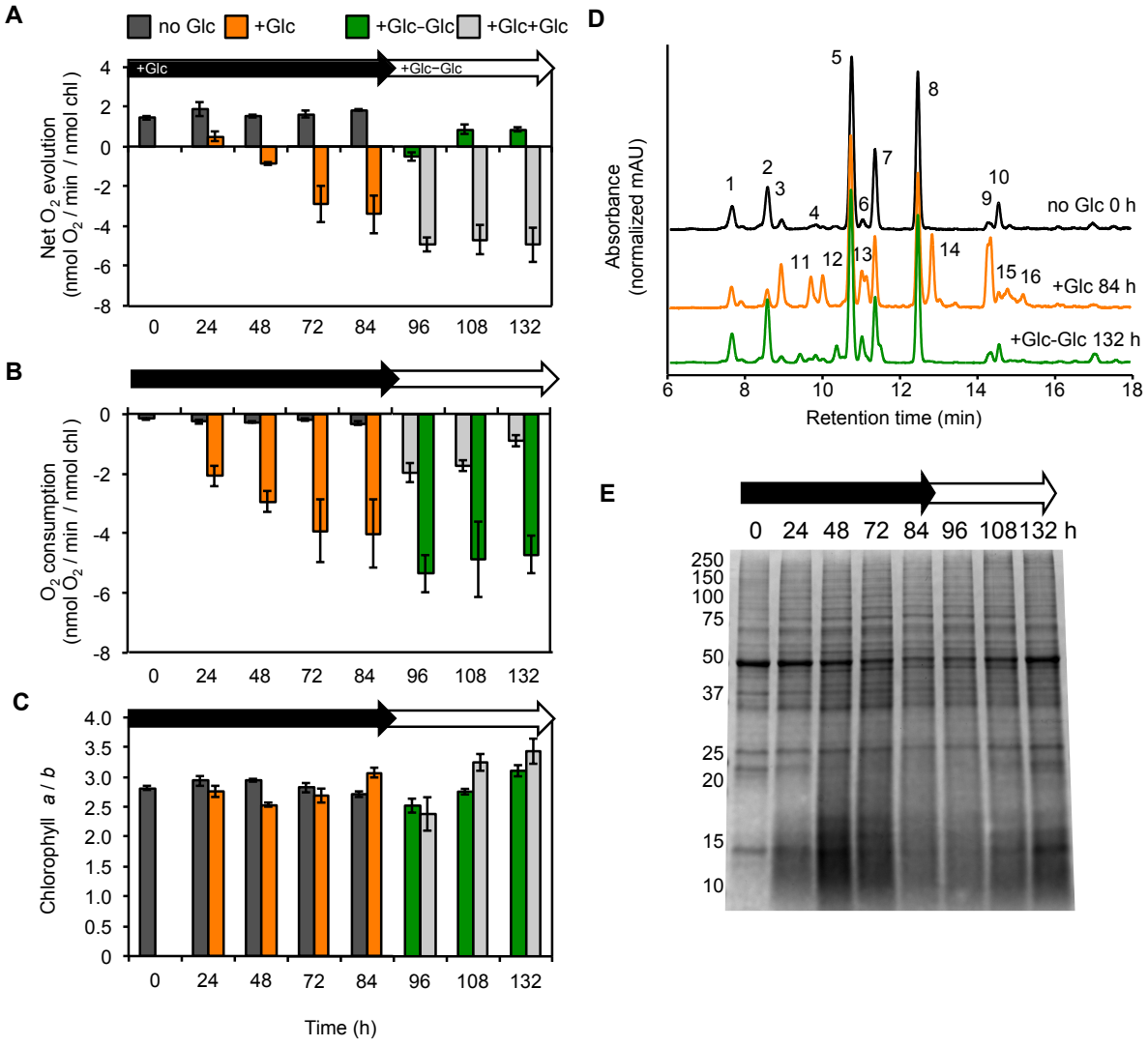


Figure 3. Glucose induces a reversible photosynthetic switch. (A) Dynamics of net oxygen evolution (oxygen production in the light, 100 µmol photons m–2 s–1) during glucose addition and removal. Data represent means ± SD (n = 3-4). Data are normalized to total chlorophyll. See Figure 1 for data normalized by cell volume. (B) Dynamics of oxygen consumption in the dark during glucose addition and removal. Data are normalized to total chlorophyll. Data represent means ± SD (n = 3-4). See Figure 1 for data normalized by cell volume. (C) Chlorophyll a / b dynamics during glucose addition and removal. Data represent means ± SD (n = 3-4). (D) HPLC chromatograms of representative samples during glucose addition and removal.1, neoxanthin; 2, violaxanthin; 3, 9, 11, 12, 13, 14, 15, 16 unidentified ketocarotenoid; 4, antheraxanthin; 5, lutein; 6, zeaxanthin; 7, chlorophyll b; 8, chlorophyll a; 10, β-carotene (E) SDS-PAGE gel analysis of total protein extracted from whole cells and stained with Coomassie brilliant blue. Samples were normalized to total protein and 10 µg of protein were loaded.

There was less violaxanthin and more zeaxanthin in the presence of Glc, but the total pool size of xanthophyll cycle pigments (violaxanthin, antheraxanthin, and zeaxanthin) was not significantly different between photoautotrophic and heterotrophic cultures (two-tailed t test, P = 0.95; Figure 2F; Figure 3D). After Glc removal, the xanthophyll cycle pool size increased due to an increase in violaxanthin. The level of b-carotene was similar between treatments (Figure 2F; Figure 3D). Immunoblot analysis was used to determine changes in protein abundance of the photosynthetic apparatus (Figure 4). With Glc there was a steady decline in PSII core proteins D2 and CP43 and PSI core protein PsaA. After Glc removal, the levels of PSI and PSII proteins recovered. The large subunit cytochrome b_6 of the cytochrome b_6f (Cyt b_6f) complex also declined rapidly with Glc and recovered after Glc removal. The PSI, PSII, and Cyt b_6f complexes showed a reduction and a subsequent recovery in protein content within 24 h. Furthermore, proteins in light-harvesting complexes (LHCs) of both PSI and PSII showed reversible declines with Glc. By contrast, chloroplastic and mitochondrial ATP synthase (AtpB) subunits CF_1b and F_1b, respectively, were maintained during Glc addition and removal. SDS-PAGE analysis of total protein extracts stained with Coomassie brilliant blue showed the reversible decline of the large subunit of the CO - fixing enzyme ribulose-2-1,5-bisphosphate carboxylase/oxygenase (Rubisco) at 53 kD (Figure 3E). The changes in the protein subunits of the photosynthetic apparatus were concurrent with the changes in photosynthetic activity during the experiment.

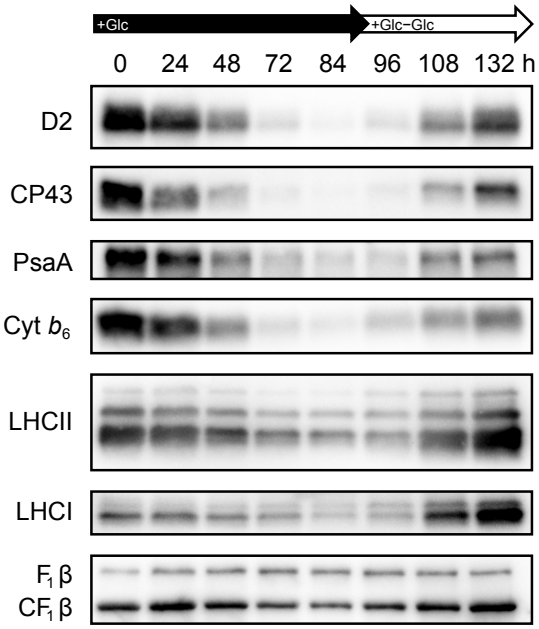


Figure 4. Glc Causes reversible loss of the photosynthetic apparatus. Immunoblot analysis of PSII (D2, CP43), PSI (PsaA), Cyt b_6f (Cyt b_6), LHC of PSII (LHCII), LHC of PSI (LHCI), and AtpB subunits. The global anti-AtpB antibody detected both the chloroplastic and mitochondrial subunits CF_1b and F_1b (faster and slower bands, respectively; see "Methods"). Samples were normalized to total protein, and 10 mg of protein was loaded per lane. See Figure 3E for stained total protein gel

To investigate changes in thylakoid ultrastructure, we used TEM. In photoautotrophic cells, thylakoid membranes were visible as stacked appressed membrane regions (Figure 7A). In heterotrophic cells, there was a reduction in size and layers of these membrane regions at 84 h. To quantify these changes, we generated three-dimensional (3D) reconstructions of serial sections from TEM (Figures 4A to 4C). The overall volume of thylakoids declined nearly 14-fold, and relative thylakoid volume

(as a percentage of the chloroplast) decreased from 16 to 2%.

Thylakoid membranes of heterotrophic cells had a reduced number of layers in appressed membrane regions compared to photoautotrophic cells, and the mode of layers decreased from 3 to 2 (Figure 7D). Appressed membrane regions with ≥6 layers were observed only in photoautotrophic cells. Furthermore, thylakoid membranes of heterotrophic cells were shorter in length than those of photoautotrophic cells, and the mode of thylakoid membrane lengths decreased substantially from 350-400 nm to 50-100 nm, respectively (Figure 7C). Photoautotrophic cells had a broad distribution of thylakoid membrane lengths, including some $1 mm, and some were nearly 5-fold longer than those from glucose-treated cells. Having both shorter membranes and a reduced number of layers contributed to the diminished volume of thylakoid membranes in Glc-treated cells (Figure 7B). Live-cell imaging using super-resolution SIM revealed changes in chloroplast structure as well as a decrease in chlorophyll fluorescence (Figure 8A). These results are consistent with the Glc-induced decline in chlorophyll per cell volume (Figure 2D). During Glc removal, layers of appressed membrane regions increased, and original thylakoid structure recovered (Figures 4A and 5A). These results showed that Glc has major impacts on thylakoid membrane ultrastructure.

We used liquid chromatography-tandem mass spectrometry (LC-MS/MS) to determine changes in the abundance of various lipid species. The most abundant thylakoid membrane lipids, monogalactosyldiacylglycerol (MGDG) and digalactosyldiacyl-glycerol (DGDG), declined 10-fold and 4-fold, respectively, with Glc at 84 h (Figure 7F; Figure 5). After Glc removal, MGDG and DGDG increased fivefold and threefold, respectively, by the end of the experiment. We also observed declines and recovery in the conserved thylakoid membrane lipid sulfoquinovosyldiacylglycerol (SQDG) and in diacylglyceryltrimethylhomoserine (DGTS), a common lipid that can be found in chloroplasts and the endoplasmic reticulum in algae (Figure 7F; Supplemental Figures 2 and 3; Thompson, 1996). MGDG, DGDG, and SQDG associate with PSII (Boudière et al., 2014), and their changes were consistent with the reversible repression and activation of photosynthesis during Glc addition and removal, respectively. Glc Induces an Increase in Starch and Lipid Energy Stores.

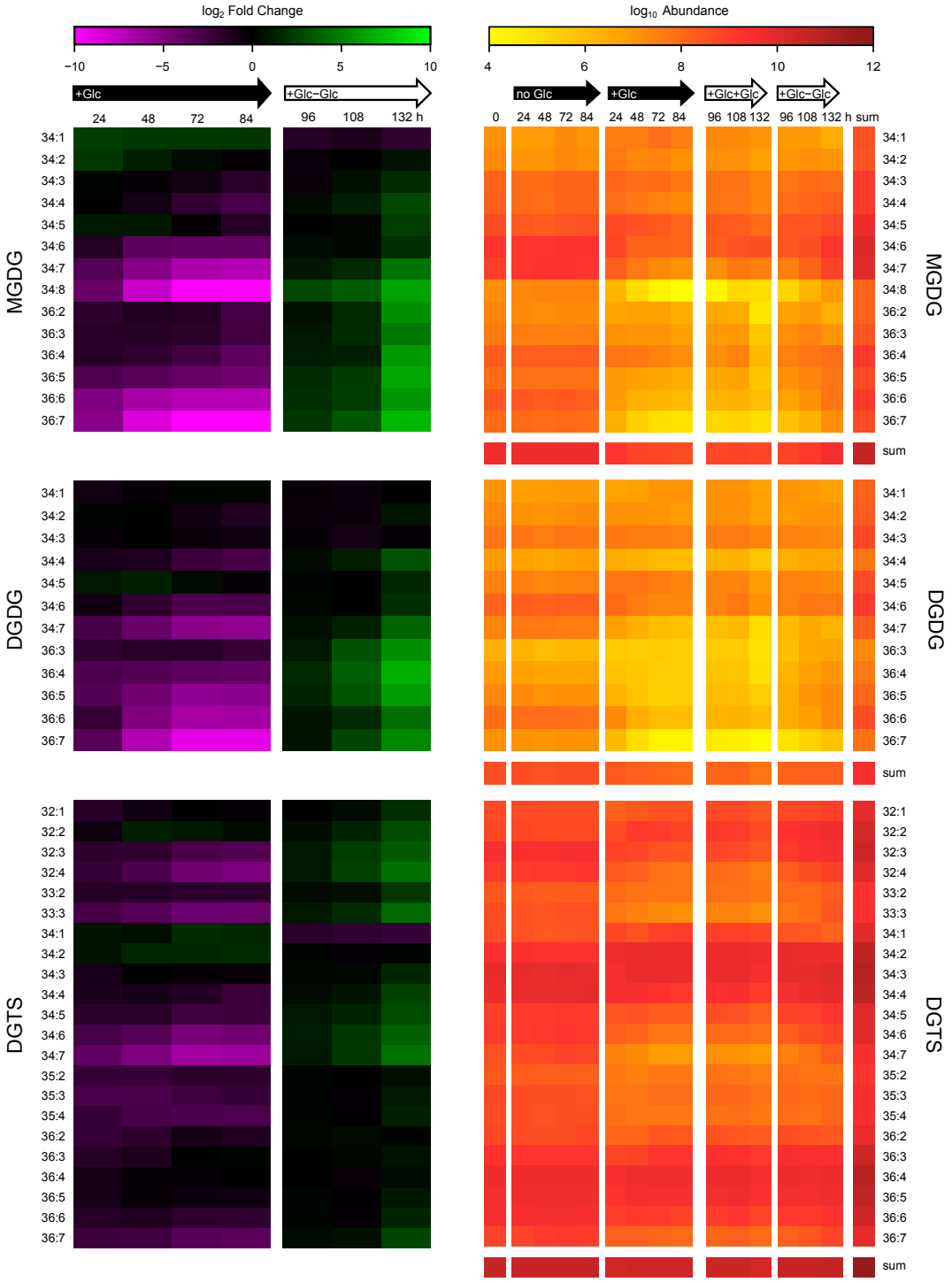


Figure 5. Changes in MGDG, DGDG and DGTS lipid species during glucose addition and removal. Left side: log_2-transformed fold change in lipid species abundance for MGDG, DGDG and DGTS upon glucose addition and subsequent removal relative to time-matched control sample plotted as a heatmap. Right side: A heatmap of the lipid species abundance for MGDG, DGDG and DGTS normalized to AFDW is plotted for photoautotrophic control, glucose addition, heterotrophic control and glucose removal. Data represent means (n = 3-4).

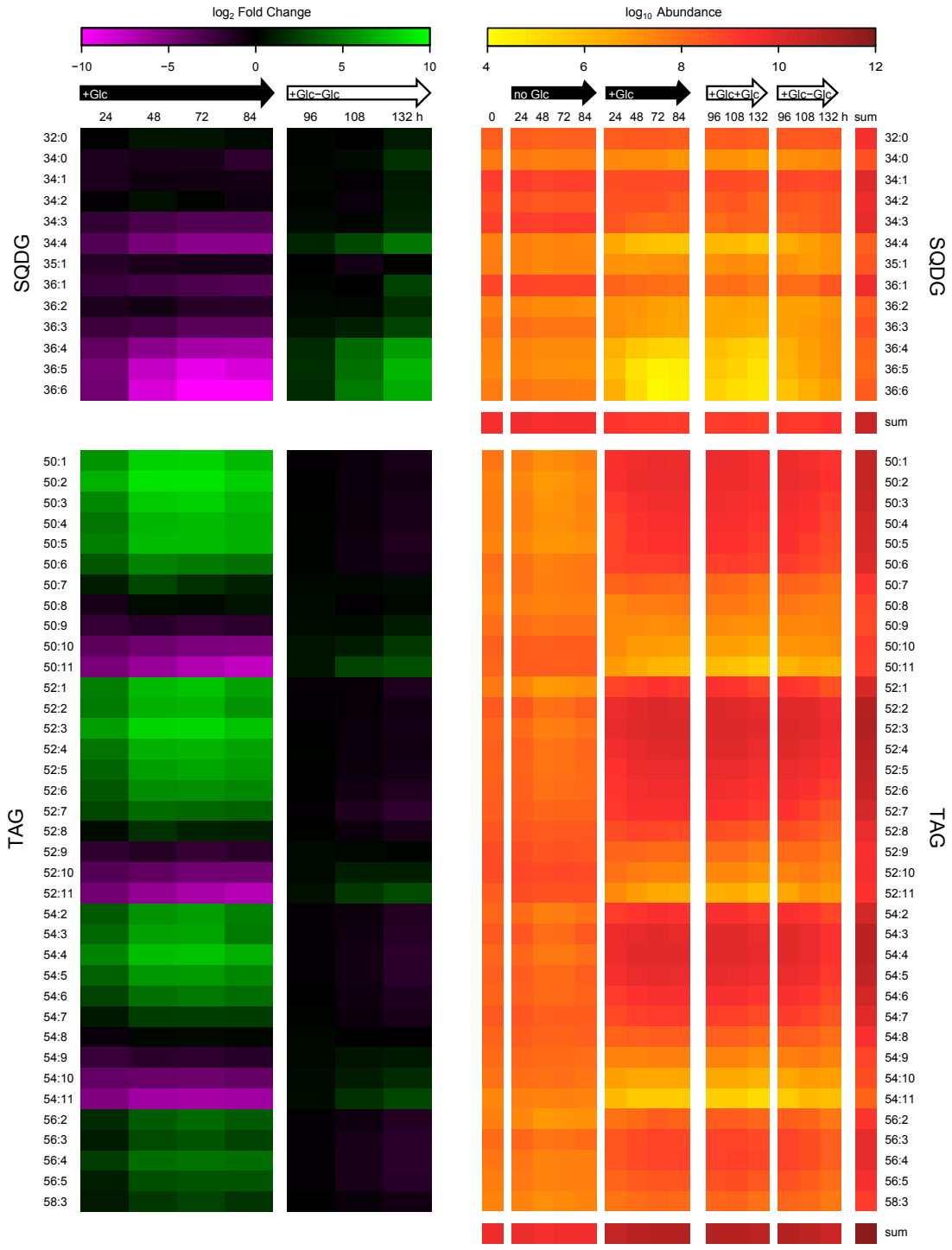


Figure 6. Changes in SQDG and TAG lipid species during glucose addition and removal. On the left, the $\log_2$-transformed fold change in lipid species abundance for SQDG and TAG upon glucose addition and subsequent removal relative to time-matched control sample plotted as a heatmap. On the right, a heatmap of the lipid species abundance for SQDG and TAG normalized to AFDW is plotted for photoautotrophic control, glucose addition, heterotrophic control and glucose removal. Data represent means (n = 3-4).

We used cryo-SXT and SIM to investigate changes in cellular components with Glc. Cryo-SXT and 3D reconstructions were used to quantify volumetric changes in cell organelles and storage bodies with Glc (Figures 5B and 5C; Supplemental Movies 1, 2, and 3). As cell size increased, the chloroplast increased approximately twofold in size and occupied the same percentage of the cell volume, 34 ± 4 and 33 ± 7% in photoautotrophic and heterotrophic cells, respectively. Starch granules, located within the chloroplast, increased nearly fivefold in heterotrophic cells, ultimately representing 30 ± 9% of the chloroplast volume and 10 ± 2% of the cell. Consistent with the cryo-SXT results, observations using TEM also showed an increase in starch in heterotrophic cells (Figure 7A).

One of the largest changes in Glc-treated cells was an increase in TAGs (Figure 8; Figure 6; Supplemental Movie 2). The volume of cytoplasmic lipid bodies of heterotrophic cells increased >16-fold and reached 14 ± 3% of the volume of the cell (Figure 8C). The majority of lipid bodies were detected near the plasma membrane in heterotrophic cells but were not observed in photoautotrophic cells (Figure 8B). SIM of live cells stained for neutral lipids confirmed this result (Figure 8A). Using LC-MS/MS, we detected a >20-fold increase in TAGs in Glc-treated cells after 84 h (Figure 8D; Figure 6). Among lipids generally, and among TAGs specifically, polyunsaturated acyl chains were depleted in Glc-treated cells, while saturated lipid species increased (Supplemental Figures 2 and 3). These results suggest that de novo FA synthesis rather than rearrangement of existing acyl chains was more likely occurring.

With both cryo-SXT and SIM, we observed a decline in the presence of lipid bodies after Glc removal (Figures 5A to 5C). Accordingly, LC-MS/MS data showed a twofold decline in TAGs (Figure 8D). Despite the increase in respiration in Glc-treated cells, the relative mitochondrial volume per cell remained similar between photoautotrophic and heterotrophic cells at 2.8 ± 1.2 and 2.2 ± 1.2%, respectively (Figure 8C). Altogether, Glc addition induced major changes in cell growth and cellular components, and these data provide evidence for a Glc-driven metabolic change resulting in increased energy stores, including starch granules and lipid bodies.

A principal component analysis (PCA) of the regularized $\log_2$-transformed counts from the resulting transcriptome profiles was plotted for both the Glc addition and Glc removal phases (Figure 9A). For Glc addition, PC1, which accounted for 59% of the variance, corresponded with the time of day, while PC2, accounting for 25% of the variance, corresponded to Glc addition. Replicate samples were tightly clustered. In contrast to Glc addition, we could not disentangle the effects of time from treatment in Glc removal. The most significant PC, representing 79% of the variance, combined both time and treatment and was observed as a rightward shift along PC1 for the Glc removal samples relative to their time-matched controls. This pattern suggests that the transcriptomic changes that are occurring in Glc removal samples also occur in heterotrophic control samples, just later in the day. The second most significant component was smaller (9% of variance) and distinguished midday time points (3 and 6 h) from early and late time points (0.5, 1, and 12 h). Taken together, our PCA demonstrates that time of day has a major effect on the transcriptomic profile of the cultures. This result is consistent with previously reported high light experiments (Roth

et al., 2017). To distinguish the effects of diurnal periodicity from the effects of Glc, all analyses of the RNA-seq experiments focused on differences compared with time-matched controls.

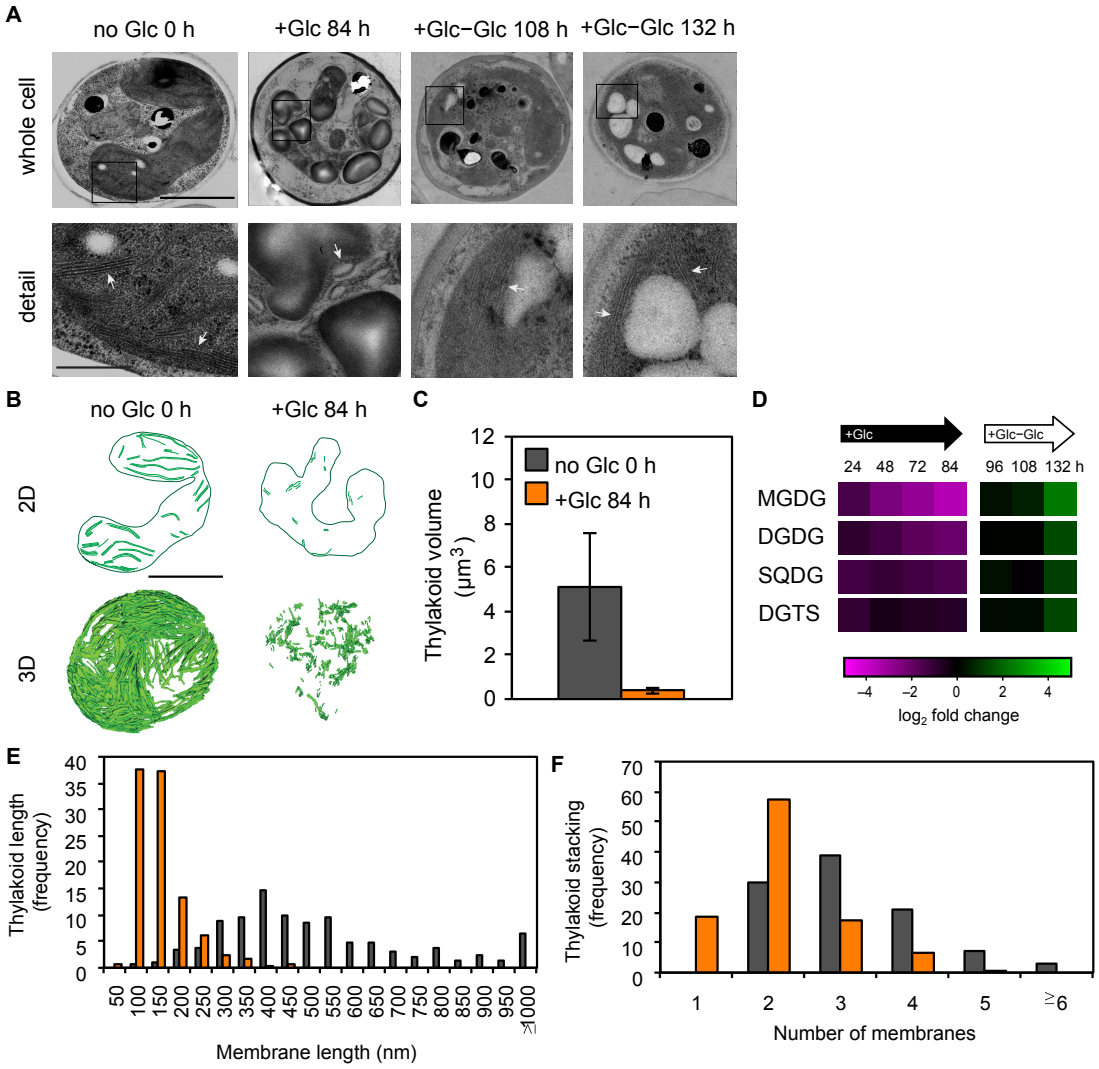


Figure 7. Thylakoid Membranes and Lipids Decrease during Glc-Induced Repression of Photosynthesis. (A) Representative images of cell 2D TEM with thylakoid detail. White arrows point to thylakoid membranes. Whole-cell bar = 2 mm; detail bar = 0.5 mm. **(B)** Thylakoid volume without and with Glc quantified by 3D reconstruction of TEM serial sections (n = 3 technical replicates). **(C)** Cartoon of a representative 2D TEM slice of chloroplast and thylakoid membranes outlined and the 3D reconstruction of the whole cell. Bar = 2 mm. **(D)** Histogram of the number of membranes per thylakoid stack with Glc. Data represent all technical replicates (n = 3). **(E)** Histogram of thylakoid length (nm) with Glc. Membrane length labels on the x-axis are the end point of the category (e.g., 50 represents the number of membranes with lengths 0 to 50 nm). Data represent all technical replicates (n = 3). **(F)** Heatmap showing thylakoid membrane lipids (MGDG, DGDG, SQDG, and DGTS) plotted as log -transformed fold change in comparisons of Glc versus 2photoautotrophic control and Glc removal versus heterotrophic control at each time point. Lipids are normalized to AFWD. Data represent means (n = 3 to 4 biological replicates). Supplemental Figures 2 and 3 contain abundance and fold change for each lipid species. The following color scheme is used for **(B)**, **(D)**, and **(E)**: no glc 0 h, dark gray; +glc 84 h, orange.

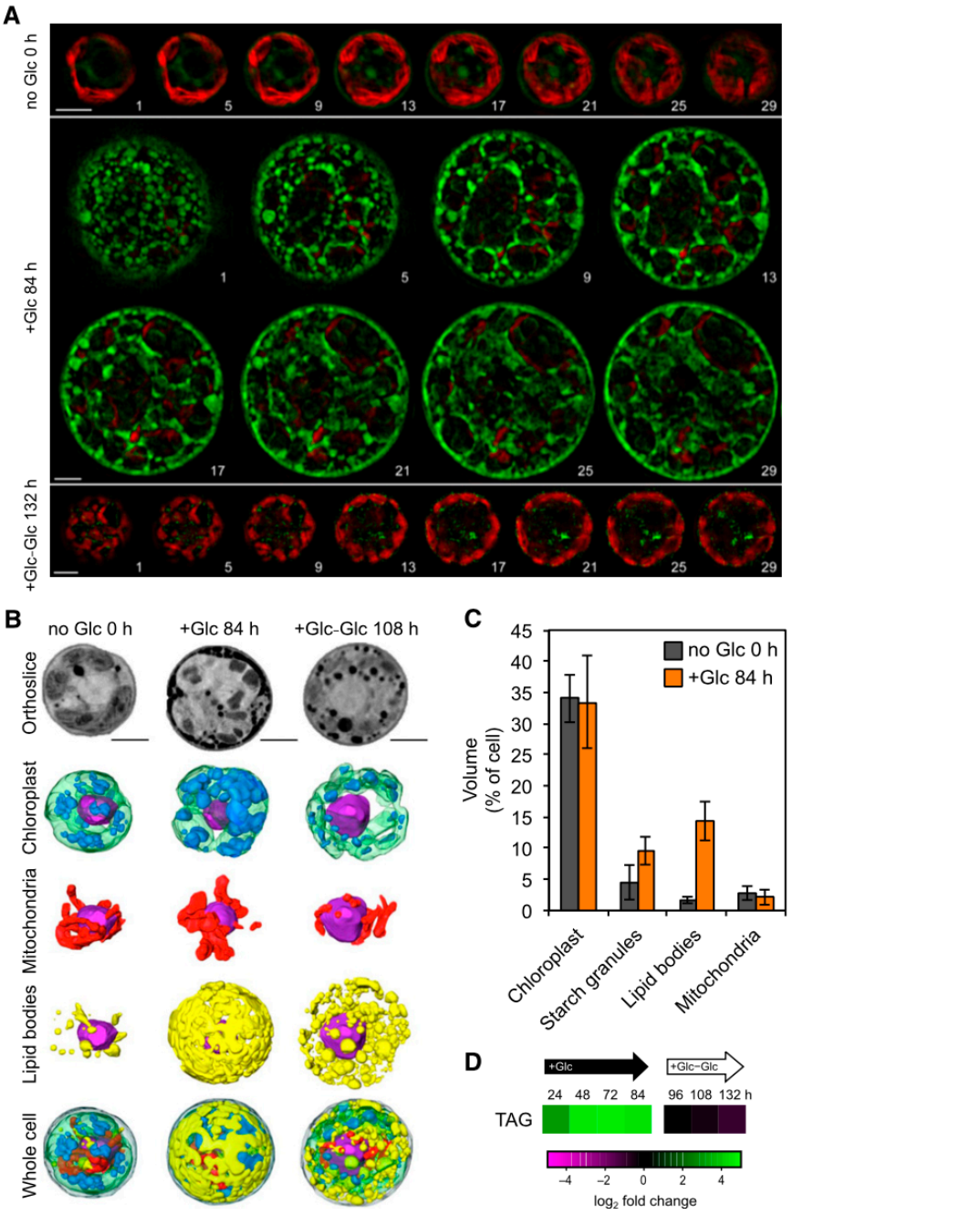


Figure 8. TAG and Starch Energy Stores Increase during Glc-Induced Metabolic Changes. (A) Representative images of live-cell SIM showing merged chlorophyll autofluorescence (red, 650 to 730 nm) and BODIPY neutral lipid dye fluorescence (green, 505 to 550 nm). Each number corresponds to the selected focal plane of serial optical sections moving from surface of the cell to the center with a z-step of 101 nm. Bar = 2 mm. **(B)** Representative cryo-SXT of cells showing orthoslices, segmented chloroplasts containing starch granules and nucleus, mitochondrial networks and nuclei, lipid bodies and nuclei, and fully segmented cells (Supplemental Movies 1, 2, and 3). Nucleus, purple; chloroplast, green; mitochondria, red; lipid bodies, yellow; starch granules within the chloroplast, blue. Bar = 2 mm. **(C)** Changes in intracellular morphology with Glc as determined by cryo-SXT. Data represent means 6 SD (n = 8 to 11 technical replicates). **(D)** Heatmap TAG abundance plotted as log -transformed fold change in comparisons of Glc versus photoautotrophic control and Glc removal versus heterotrophic control at each time point. Lipids are normalized to AFDW. Data represent means (n = 3 to 4 biological replicates). Figure 6 contains abundance and fold change for each lipid species.

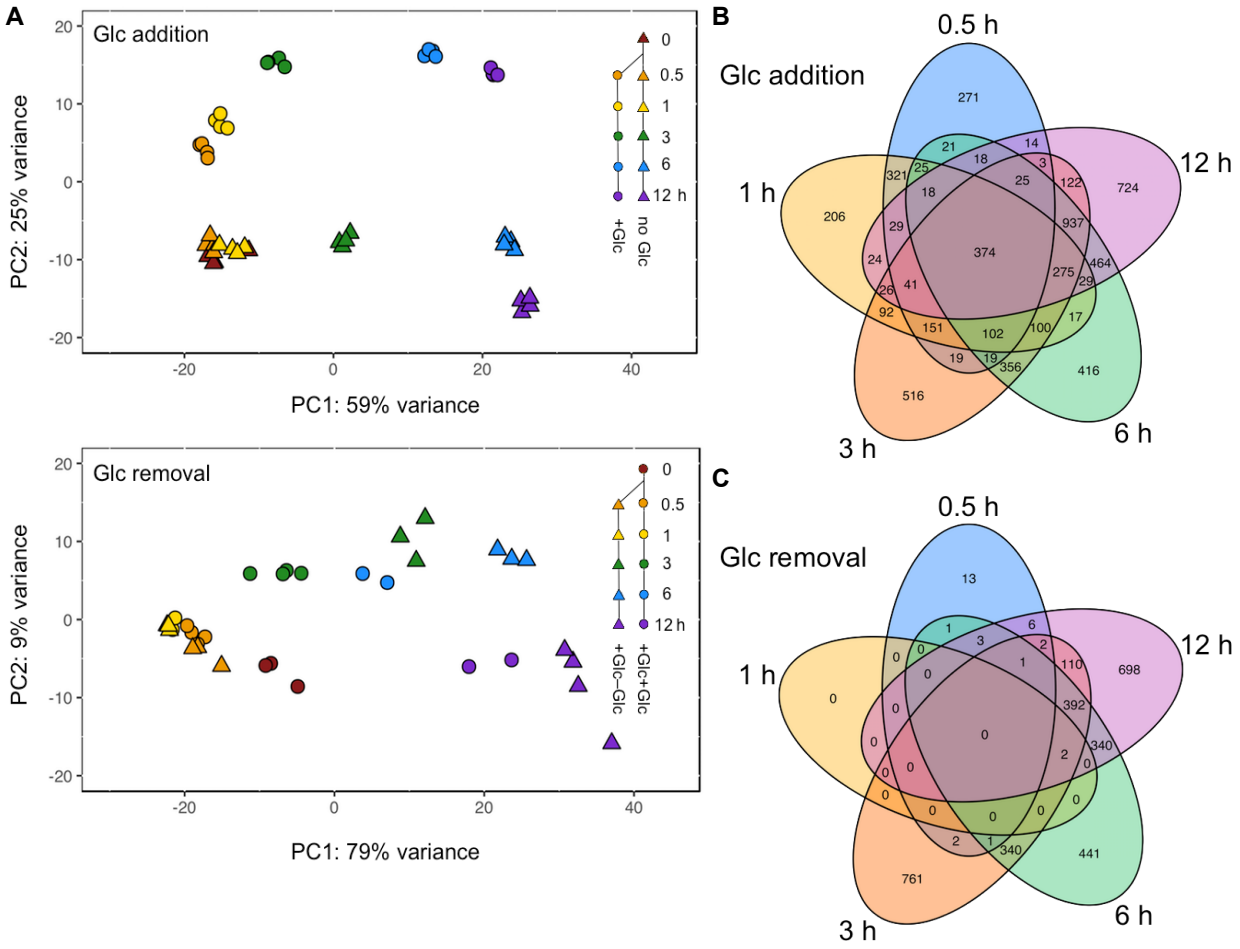


Figure 9. Transcriptome Response to Glc Addition and Removal.

RNA abundance changes during Glc addition and removal. At t = 0, Glc addition or Glc removal was initiated. Samples were collected in quadruplicate at 0, 0.5, 1, 3, 6, and 12 h for photoautotrophic and heterotrophic controls and at 0.5, 1, 3, 6, and 12 h for Glc addition and removal. Transcript abundances for each sample were determined by RNA-seq. DEGs are more than twofold up- or downregulated relative to time-matched controls with BH-adjusted P < 0.01. See also Figure 3 for DEGs during Glc addition and removal and Venn diagram showing reversible differential expression during Glc addition and removal. **(A)** A PCA of regularized log -transformed counts for the 500 genes with the highest variance during Glc addition (top panel) and removal (bottom panel) is plotted. Supplemental Data Set 5 contains the genes used in the PCA. For Glc addition, the photoautotrophic controls (circles) and the Glc addition samples (triangles) are displayed with time points indicated by color. The two most significant components, accounting for 84% of the variation, are shown. For Glc removal, the heterotrophic control (circles) and the Glc removal samples (triangles) are displayed with time points indicated by color. **(B)** Venn diagram comparing the number of DEGs at each time point (0.5, 1, 3, 6, and 12 h) for Glc addition. **(C)** Venn diagram comparing the number of DEGs at each time point (0.5, 1, 3, 6, and 12 h) for Glc removal.

To examine the effects of Glc addition and removal on gene expression, we identified differentially expressed genes (DEGs) between treatment and control at each time point. Genes were considered differentially expressed if they showed a greater than twofold change in transcript abundance (Benjamini–Hochberg [BH]-adjusted $P < 0.01$) in either direction between Glc and photoautotrophic control cultures or Glc removal and heterotrophic control cultures (Figures 6 and 7). More than a third of the genes in the transcriptome were differentially expressed in response to Glc at one or more time points in this study (6356 of 15,369 genes, Figure 10). Consistent with the PCA, more genes were differentially expressed upon Glc addition than upon Glc removal (5755 and 3113, respectively), and the two treatments shared 2512 DEGs. There were 374 DEGs shared across all time points in Glc addition experiments (Figure 9B). By contrast, Glc removal had no DEGs shared across all time points (Figure 9C). There were two distinct stages of gene expression in Glc removal: an early stage (0.5 and 1 h), which had minimal changes (31 DEGs combined), and a mid-late stage, which had a large number of DEGs. The later time points (3, 6, and 12 h) each had >1500 DEGs and shared 395 DEGs (Figures 6 and 7). It is possible that the delayed response was due to intercellular stores of Glc that had not yet been depleted. With Glc addition, 64% of the DEGs were down-regulated, whereas the opposite was true with Glc removal, as 63% of genes were upregulated. Seventy-five percent of the DEGs that were upregulated upon Glc removal were also downregulated with Glc addition (Figure 10). Conversely, 68% of downregulated DEGs upon Glc removal were upregulated DEGs with Glc addition. Functional Enrichment Analysis Reveals Pathways Involved in Photosynthetic and Metabolic Changes

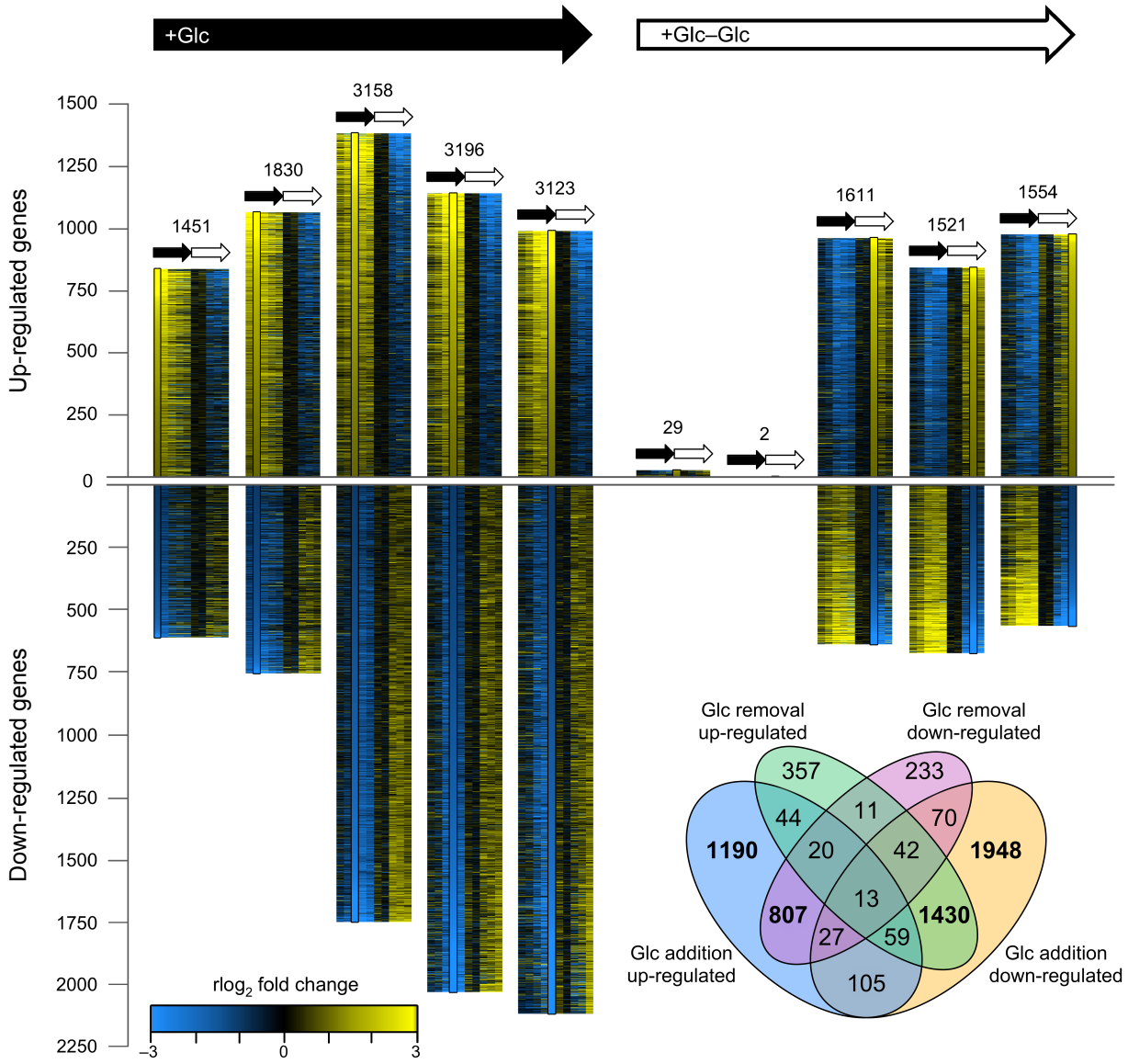


Figure 10. Genome-Wide Transcriptome Responses during Glc Addition and Removal Reveal Specific Reversible Patterns. Gene expression changes during Glc addition and removal. Genes were considered to be DEGs if they were more than twofold up- or downregulated relative to time-matched control samples with a BH-adjusted $P < 0.01$. On left is plotted the number of DEGs in a comparison of Glc-treated samples versus photoautotrophic controls at each time point. On right, the same comparison was made for Glc removal samples versus heterotrophic controls. The number of DEGs is indicated by the height of the bar above the line for upregulated DEGs or below the line for downregulated DEGs, and the total is indicated above each bar. The difference between treatment and its respective control was calculated in terms of regularized (r)$\log_2$-transformed fold change of counts and plotted in the black box for each time point using the included color scale. Data represent means (n = 2 to 4 biological replicates). For comparison, the fold change of each DEG at all other time points in both Glc addition and removal is presented flanking the black box. The columns below the small black and white arrows indicate, from left to right, 0.5, 1, 3, 6, and 12 h after Glc addition and removal, respectively. Inset: A Venn diagram is used to demonstrate the overlap in the numbers of DEGs upregulated or downregulated during Glc addition and Glc removal. Regions with >500 DEGs are highlighted in bold.

To gain insight into the pathways that respond after Glc addition and subsequent removal, all expressed genes (mean fragments per kb of exon per million mapped reads [FPKM] >/=1 across all time points; 13,428 total) were grouped based on their coexpression using k-means clustering analysis with six centers (Figure 11). Next, we used enrichment of BioCyc pathway annotations as an unbiased approach to identify key functions in each of the clusters (Supplemental Data Set 2).

Cluster 1 was characterized as highly up-and downregulated upon Glc addition and Glc removal, respectively (Figure 11). Pathways enriched in cluster 1 included glycolysis, FA and biotin biosynthesis, and acetyl-CoA generation (Supplemental Data Set 2). Cluster 1 also included genes encoding hexose phosphate isomerases, which had highly dynamic expression patterns, although their associated pathways were not enriched. Cluster 2 was characterized as moderately up-and downregulated upon Glc addition and Glc removal, respectively, and as in cluster 1, FA metabolism pathways were enriched. Furthermore, putative Suc biosynthesis, starch metabolism, and the methylerythritol phosphate pathway for isoprenoid biosynthesis were enriched. Because both starch synthesis and degradation pathways were enriched, there is potential for a futile starch cycle during Glc addition. Pathways related to the metabolism of very long chain FAs were also enriched.

In contrast to clusters 1 and 2, cluster 6 showed the opposite trend: gene expression was highly down and upregulated upon Glc addition and Glc removal, respectively. Pathways enriched in cluster 6 included photosynthesis, chlorophyll biosynthesis, b-oxidation of FAs, and the glyoxylate cycle (Supplemental Data Set 2). Pathways enriched related to FA metabolism included catabolism and long-chain b-ketoacyl-synthase–related pathways. Lipid catabolism by peroxisome-mediated b-oxidation of FAs and the glyoxylate cycle were both upregulated during Glc removal, suggesting a mechanism for TAG breakdown and mobilization and resynthesis of Glc from acetyl-CoA.

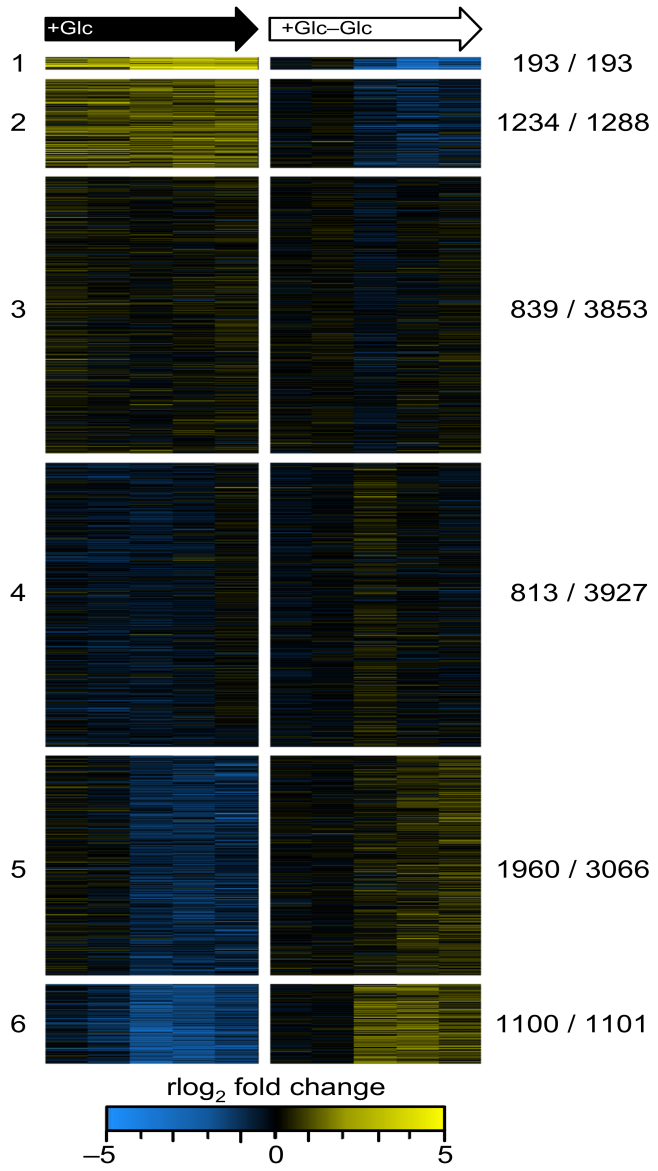


Figure 11. Heatmap of *k*-Means Clustering of Gene Coexpression Reveals Reversible Glucose-Induced Patterns.

The fold changes between Glc addition versus the photoautotrophic control and between Glc removal versus the heterotrophic control were calculated for samples at 0.5, 1, 3, 6, and 12 h posttreatment for all genes with >/= 1 FPKMs mean expression across all samples (13,307 of 15,369 total genes). Data represent means (n = 2 to 4 biological replicates). The resulting data were subjected to k-means clustering with six centers and plotted as a heatmap of regularized (r)$\log_2$ -transformed fold change for each gene in each cluster. The number of DEGs (more than two fold up-regulated or down-regulated) and the total number of genes for each cluster are indicated to the right.

Cluster 5 had a pattern of slight decreasing and increasing upon Glc addition and Glc removal, respectively, but no pathways were functionally enriched (Supplemental Data Set 2). Clusters 3 and 4 did not show clear trends with respect to changes in gene expression during Glc addition and removal, and only one pathway was enriched in each cluster. Mitochondrial genes were present in clusters 3 and 4, and chloroplast genes were split among clusters 3,4, 5, and 6.

We also conducted functional enrichment analyses on four groups of DEGs: upregulated upon Glc addition, downregulated upon Glc addition, upregulated upon Glc removal, and down-regulated upon Glc removal (Supplemental Data Set 2). We found a similar set of metabolic pathways enriched by this analysis as were revealed by the k-means cluster analysis, but this grouping allowed us to distinguish pathways that have similar initial kinetics but demonstrate divergent behavior. This approach identified additional enriched pathways related to nitrate reduction and TAG biosynthesis.

In summary, the most striking enriched pathways from both approaches in genes whose expression increased upon Glc addition and decreased after Glc removal were related to glycolysis, starch, and FA metabolism including TAG biosynthesis. The most notable pathways enriched in genes whose expression is decreased upon Glc addition and increased with Glc removal were those encoding enzymes of b-oxidation, the glyoxylate cycle, and photosynthesis including chlorophyll biosynthesis and other related activities.

Reversible Glc-Induced Upregulation of Ketocarotenoid Biosynthesis Gene Expression

To further characterize expression of carotenoid biosynthesis genes, we mapped RNA abundance changes upon Glc addition and removal onto the carotenoid biosynthetic pathway (Figure 12; Supplemental Data Set 3). These data in combination with the analysis of photosynthetic pigments (Figures 1 and 2; Figure 3D) revealed a preference for changes in expression of the b-carotene branch over the a-carotene branch of the pathway upon Glc addition: lycopene beta-cyclase (LCYB1) was upregulated, while lycopene epsilon-cyclase (LCYE1) was downregulated. Beta-carotene hydroxylase (CHYB1) and beta ketolase (BKT)1, which are critical for the production of astaxanthin in *C. zofingiensis* (Li et al., 2008; Roth et al., 2017), were upregulated in Glc-treated cells, consistent with the increase in ketocarotenoids (Figures 1B and 2F; Figure 3D). Moreover, expression of the gene encoding the chlorophycean violaxanthin de-epoxidase (CVDE1), which converts violaxanthin to zeaxanthin (Li et al., 2016), was upregulated, while zeaxanthin epoxidase (ZEP1), the gene responsible for the reverse reaction, was downregulated. These expression changes correlated with an increase in the proportion of zeaxanthin to violaxanthin (Figure 2F; Figure 3D). *C. zofingiensis* also has an ortholog of plant-type VDE1 whose transcript abundance is similar to that of CVDE1, but its expression was insensitive to Glc addition. Alternatively, VDE1 was upregulated upon Glc removal, suggesting that its gene product could be utilized under distinct conditions relative to CVDE1 or that it has another role in *C. zofingiensis*. Upon Glc removal, the changes in gene expression and ultimately pigment abundances were reversed (Figures 1, 2, and 9; Figure 3D). Although these changes were modest, our results showed that Glc caused a reversible change in expression of genes associated with the production of ketocarotenoids.

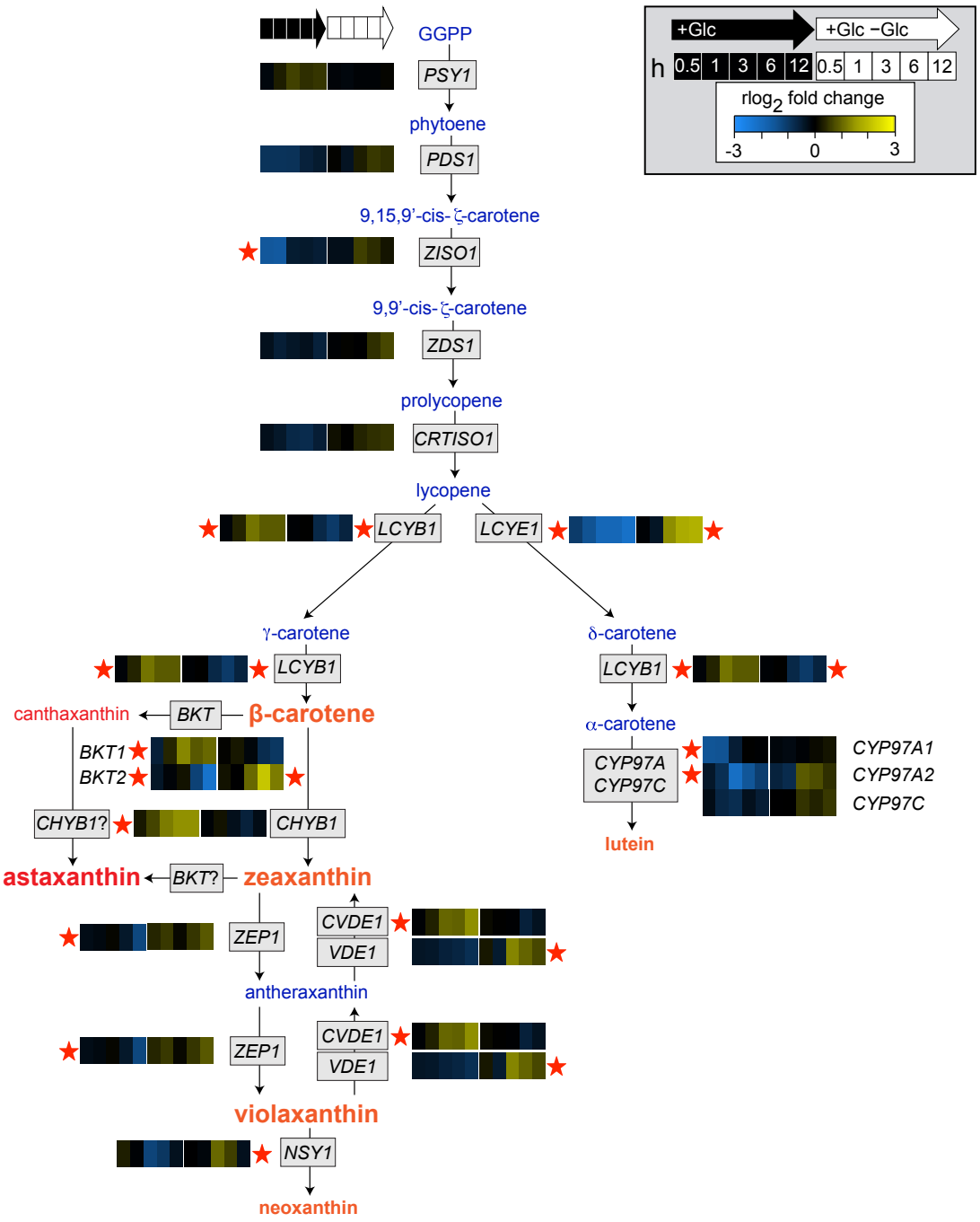


Figure 12. Glc Causes Reversible Increases in Ketocarotenoid Gene Expression. The carotenoid pathway with regularized (r)$\log_2$ -transformed fold change in gene expression upon Glc addition and subsequent removal relative to time-matched control sample plotted as a heatmap. Data represent means (n = 2 to 4 biological replicates). DEGs are indicated by a red star (more than twofold up-or downregulated, BH-adjusted $P < 0.01$). Fold-change values and gene IDs are in Supplemental Data Set 3.

Reversible Glc-Induced Downregulation of Chlorophyll Biosynthesis Genes

We mapped the dynamics of the expression of genes encoding chlorophyll biosynthesis and degradation functions upon Glc addition and subsequent removal (Figure 13; Supplemental Data Set 3). Genes involved in chlorophyll biosynthesis were differentially downregulated with Glc. This trend was observed for nearly all steps of tetrapyrrole biosynthesis, magnesium chelation, and phytol addition. Transcripts of enzymes related to chlorophyll degradation or phytol biosynthesis did not show trends toward coordinated regulation and were insensitive to Glc treatment. Chlorophyll biosynthesis genes that were differentially downregulated with Glc were generally upregulated upon Glc removal. These results demonstrate that Glc caused changes in chlorophyll biosynthesis gene expression, and these changes correspond to the observed changes in chlorophyll abundance (Figures 2D and 2E).

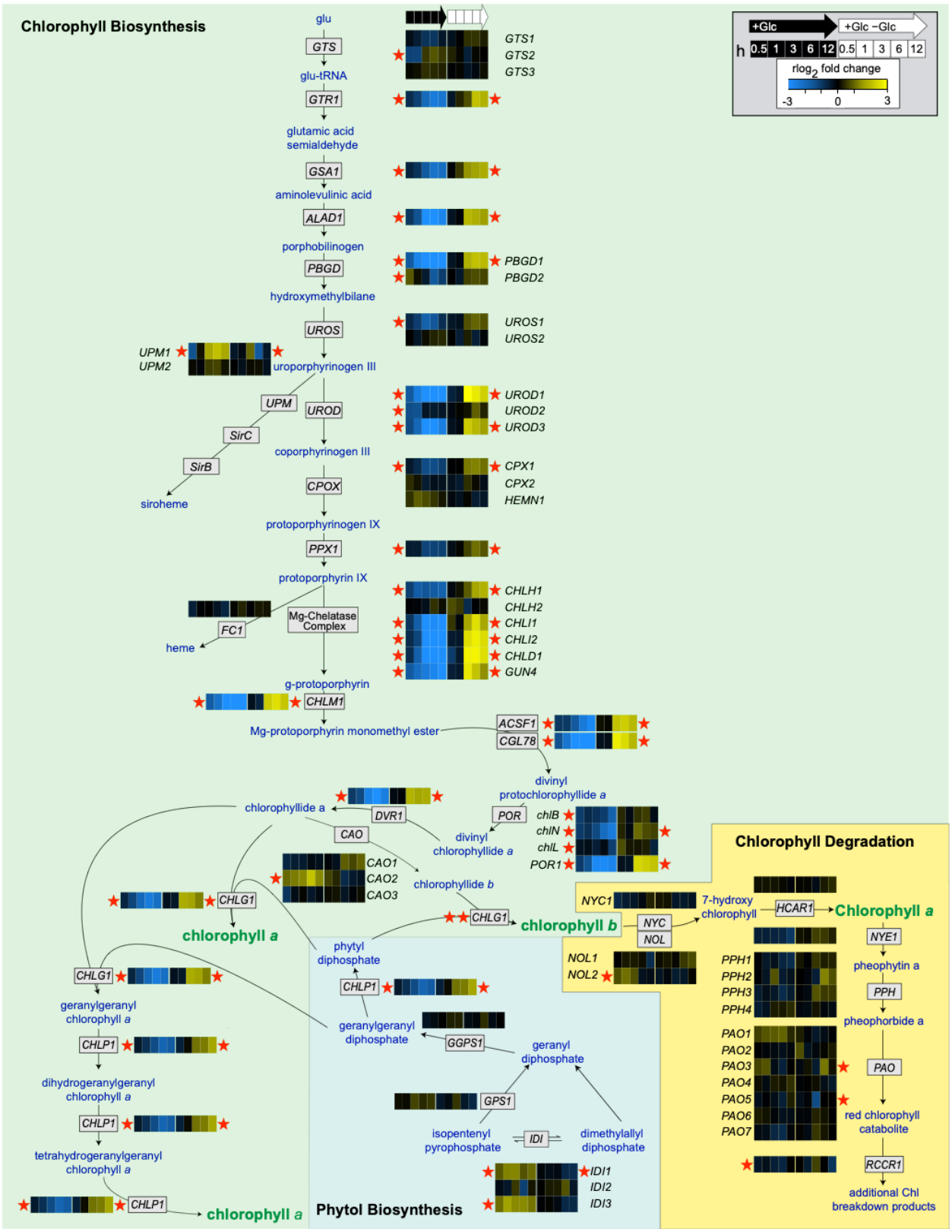


Figure 13. Glc Causes Reversible Decreases in Chlorophyll Biosynthesis Gene Expression. The chlorophyll biosynthesis and degradation pathways with regularized (r)$\log_2$ -transformed fold change in gene expression upon Glc addition and 2 subsequent removal relative to time-matched control sample plotted as a heatmap. Data represent means (n = 2 to 4 biological replicates). DEGs are indicated by a red star (more than twofold up- or downregulated, BH-adjusted P < 0.01). Fold-change values and gene IDs are in Supplemental Data Set 3.

Reversible Glc-Induced Changes in Glycolysis and FA Biosynthesis Gene Expression

To examine metabolic changes that may lead to increased accumulation of TAGs during heterotrophy (Figure 8D; Figure 6), we mapped RNA abundance changes onto glycolysis and FA metabolism pathways upon Glc addition and removal (Figure 14; Supplemental Data Set 3). The abundances of transcripts for most enzymes in glycolysis were increased in Glc-treated cells. RNA encoding the first step in glycolysis, HXK1, rapidly increased 30-fold by 0.5 h and 40-fold by 1 h after Glc addition. Upon Glc removal, the abundance of transcripts encoding glycolytic enzymes was unchanged or decreased.

Pyruvate generated from glycolysis is further oxidized to form acetyl-CoA by the pyruvate dehydrogenase (PDH) complex. After Glc addition, genes encoding the three subunits of the plastid-localized PDH complex were reversibly upregulated in response to Glc (Figure 14; Supplemental Data Set 3). The PDH components E1a, E1b, E2, and E3 (encoded by pyruvate decarboxylase [PDC2], pyruvate dehydrogenase [PDH2], dihydrolipoamide acetyltransferase [DLA2], and dihydrolipoamide dehydrogenase DLD2, respectively) were regulated in near synchrony, demonstrating a tightly regulated response.

Other genes encoding components of FA biosynthesis were also reversibly regulated by Glc (Figure 14; Supplemental Data Set 3), including those encoding acetyl-CoA carboxylase (ACC), acyl carrier protein (ACP2), malonyl-CoA:ACP transacylase (MCT1), subunits of the FA synthase complex (beta-ketoacyl ACP synthase [KAS1-3], beta-hydroxyacyl ACP dehydrase/dehydratase [HAD1], beta-ketoacyl ACP reductase [KAR1], enoyl ACP reductase [ENR1]), and a fatty acid desaturase (FAD; SAD1). By contrast, transcript abundance of genes required for the synthesis of thylakoid lipids MGDG, DGDG, SQDG, and DGTS were relatively insensitive to Glc addition and removal. We did, however, observe dynamic regulation of plastid galactolipid degradation (PGD1), whose homolog is responsible for MGDG degradation in *C. reinhardtii* (Li et al., 2012). FAT1, encoding an enzyme responsible for the export of free FA from the chloroplast, was reversibly upregulated with Glc.

Many extra-plastidic enzymes involved in lipid biosynthesis were differentially expressed upon Glc addition and removal (Figure 14; Supplemental Data Set 3). Genes involved in TAG biosynthesis, GPAT2, LPAAT2, PAP4, and DGATs (DGTT3, DGTT6, DGTT7), were reversibly upregulated with Glc. The gene encoding major lipid droplet protein (MLDP1), a protein scaffold found in lipid droplets whose abundance is correlated with cellular TAG content (Moellering and Benning, 2010), was particularly responsive during Glc addition and removal. We also identified a potential TAG lipase gene (LIP1) that was dynamically regulated with Glc in accordance with TAG accumulation with Glc. FADs, which determine the position and degree of saturation of lipid acyl chains, also had reversible upregulation with Glc. FAD6B and FAD7B, whose products are predicted to be chloroplast localized, had the strongest response to Glc. The reversible RNA abundance changes in TAG biosynthesis genes were consistent with observed changes in TAG abundance and the presence of cytoplasmic lipid

droplets (Figure 8).

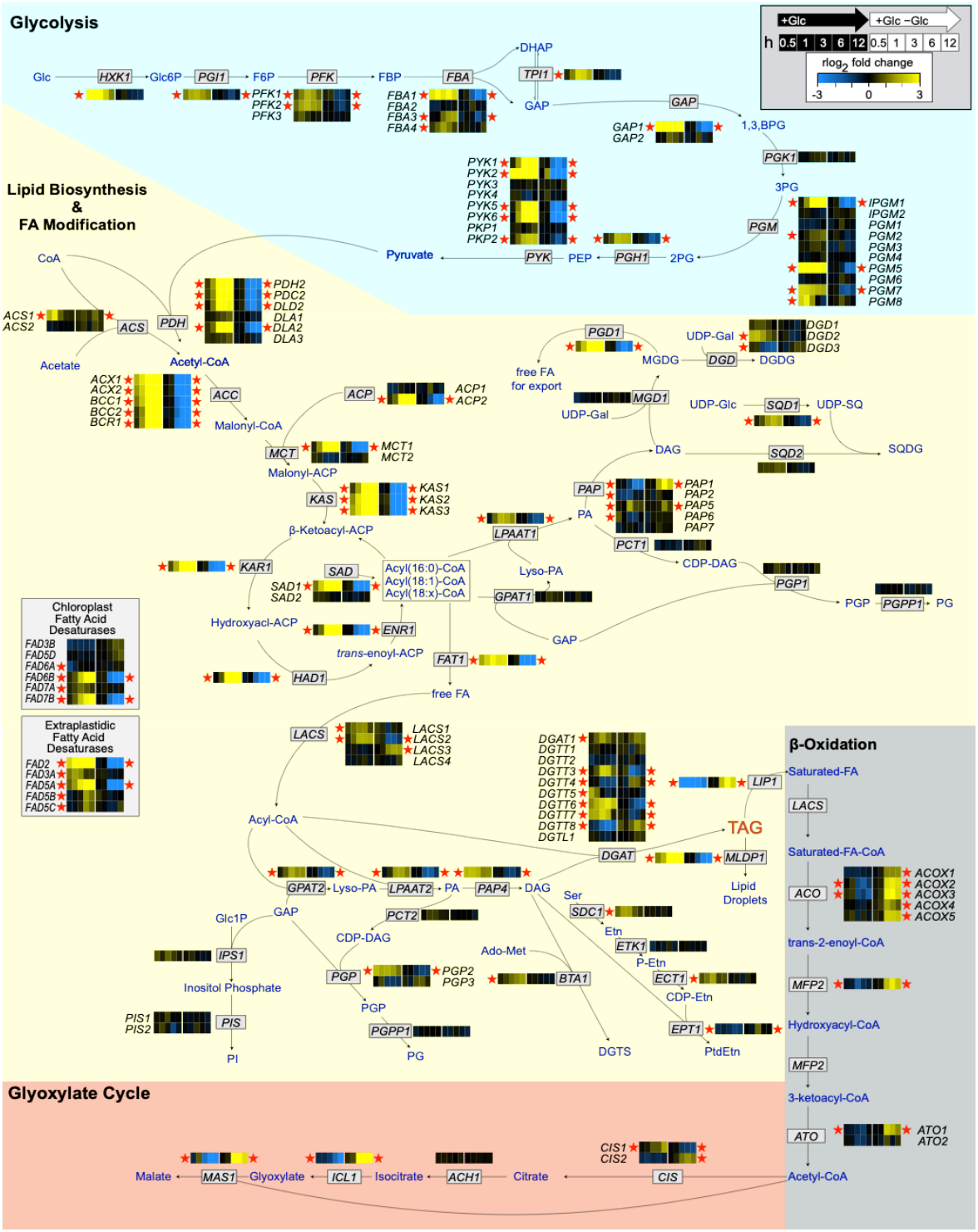


Figure 14. Gene Expression Changes Associated with Altered Carbon Flux with Glc Indicate Increased Glycolysis and FA Biosynthesis. Glycolysis and FA metabolism pathways with regularized (r)log-transformed fold change in gene expression upon Glc addition and subsequent removal relative to time-matched control sample plotted as a heatmap. Data represent means (n = 2 to 4 biological replicates). DEGs are indicated by a red star (more than twofold up- or downregulated, BH-adjusted P < 0.01). Fold-change values and gene IDs are in Supplemental Data Set 3.

DISCUSSION

In this study, we have characterized a trophic transition in a green alga and identified biological processes involved in disassembly and assembly of photosynthesis and associated changes in heterotrophic carbon metabolism in the light. We have integrated cellular physiology studies, lipid analyses, transcriptomics, TEM, cryo-SXT, and SIM to provide a detailed characterization of the tightly controlled, reversible regulation of photosynthesis and metabolism. It is notable that the repression and activation of photosynthesis characterized in this study occurs in the light. In addition to the interesting biology of *C. zofingiensis* during its metabolic changes during trophic transitions, recent interest in this organism has grown because of its commercial potential: it can amass large amounts of TAGs in parallel with the high-value ketocarotenoid astaxanthin while accumulating biomass.

The first report of a visual color change in algae upon addition of sugars was from Krüger (Krüger, 1894) when he isolated *A. protothecoides* from tree sap and cultured it in a variety of conditions. In the 1960s and 1970s, Hase and coauthors further explored this phenomenon, and they described algal cells as “glucose-bleached” (summarized in Hase, 1975). Hase and coauthors focused on the dark degeneration and specifically the regeneration of chloroplasts in the light. Their work in *A. protothecoides* provides a good comparison with this study. However, the trebouxiophyceaen *A. protothecoides* is distinct from the chlorophyceaen *C. zofingiensis* in that it does not accumulate astaxanthin during growth on Glc. Coordinated Assembly and Disassembly of Photosynthesis

While many algae have flexible metabolism, it is unusual to completely turn off photosynthesis during trophic transitions (Shugarman and Appleman, 1966; Semenenko, 1981). Here, we have shown that *C. zofingiensis* transitions from photoautotrophy through mixotrophy to heterotrophy with the addition of exogenous Glc in the light. Photosynthesis reversibly turns off, and the photosynthetic machinery is degraded and regenerated during these trophic shifts. Within a few days, we observed complete loss of PSII activity and cessation of oxygen evolution in the light (Figure 2). In conjunction with the absence of photosynthetic activity, the abundance of reaction center proteins of PSI and PSII decreases dramatically (Figure 4), the abundance of dominant thylakoid membrane lipids associated with PSII decreases (Figure 7F), thylakoid membranes are truncated (Figures 4D and 4E), and thylakoid volume decreases 14-fold (Figure 7B). The abundance of PSII core proteins appears to be affected more than that of PSI core proteins, and core proteins of both PSs are affected more than LHCs (Figure 7). As with *C. zofingiensis*, differential degradation of PSI and PSII was observed with Glc in A. protothecoides (Hase, 1975). PSI and PSII may become decoupled during the transition from photoautotrophy to heterotrophy, and residual PSI could retain some activity within the time scale of this experiment. Consistent with the changes in Rubisco abundance that we observed (Figure 3E), declines in Rubisco activity with Glc were reported previously for *C. zofingiensis* (Zhang et al., 2017) and *A. protothecoides* (Oshio and Hase, 1972). Glc removal quickly restores photosynthetic activity and oxygen production, although respiration of cells

remains high, suggesting that some physiological differences in cells remain within the time frame of this experiment. Similarly, during Glc removal, CO_2 fixation and Rubisco activity recovered quickly in *A. protothecoides*, and light was necessary to reach high rates of photosynthesis (Oh-Hama et al., 1965; Oshio and Hase, 1972). Additionally, in *A. protothecoides*, photosynthetic and CO_2 fixation enzymes are degraded under heterotrophic conditions, reflecting the changes in photosynthesis (Gao et al., 2014).

C. zofingiensis thylakoid membranes in photoautotrophic conditions showed substantial variation in both length, ranging from 70 to 2100 nm, and stacking, ranging from two to six membranes (Figure 7D). In addition to the loss of thylakoid volume as measured by serial-section TEM (Figures 4B and 4C), the degeneration of thylakoid ultrastructure with Glc was visible from the changes in chlorophyll fluorescence using SIM (Figure 8A). Degradation of thylakoid membranes was previously seen, but was not quantified in *A. protothecoides* (Shihira-Ishikawa and Hase, 1964; Oh-Hama et al., 1965). In *C. zofingiensis*, in addition to the decrease in thylakoid membranes, the chloroplast grew in proportion to the cell size with Glc. During Glc removal, light is necessary for the organization of fully developed thylakoid structure in *A. protothecoides* (Oh-Hama et al., 1965), and a carbohydrate reserve rather than photosynthesis was hypothesized to contribute to the regeneration of thylakoid membranes (Hase, 1975). Using antibiotics, it was shown that cell division is not necessary for chloroplast regeneration and the formation of chlorophyll, RNA, and protein in *A. protothecoides* (Aoki and Hase, 1965).

Changes in photosynthesis correlated with decreases in the expression of chlorophyll biosynthesis genes rather than increases in the expression of chlorophyll degradation genes (Figure 13). The amount of chlorophyll per cell remained constant during this experiment (Figure 2E). It is unknown where the remaining chlorophyll molecules are located, but it is likely that they are contained in LHCs. LHCs are the most abundant thylakoid membrane protein and contain chlorophyll and, while the abundance of LHCs was reduced with Glc, LHCI and LHCII remained detectable over the time-course of this experiment (Figure 4). Additionally, it may be possible for chlorophyll to be sequestered through autophagy, a process of membrane trafficking in which damaged or toxic cellular components are recycled and degraded, as was observed in *C. reinhardtii* (Pérez-Pérez et al., 2017). Future investigations into the location of the remaining chlorophyll molecules and/or what are they bound to could resolve this interesting issue. With Glc addition, the cells grew larger, which resulted in a decrease in chlorophyll when compared to total cell volume (Figure 2D). It is likely that if the Glc phase of the experiment continued, the reduction of chlorophyll biosynthesis gene expression would have ultimately led to a decrease in chlorophyll on a per cell basis. During Glc removal, there was increased expression of chlorophyll biosynthesis genes, resulting in an increase in chlorophyll within the cells. Future research could determine whether the remaining chlorophyll was recycled or synthesized de novo during thylakoid biogenesis. Similarly, a reversible decline in chlorophyll absorbance was observed over five days with Glc in *A. protothecoides* (Shihira-Ishikawa and Hase, 1964). It was also noted in Chlorella that nutrient concentrations affected the changes in chlorophyll and chloroplast structures (Shihira-Ishikawa and Hase, 1964), and this is likely the case for

C. zofingiensis, providing an important area for future investigations.

Previous work on *Chlorella* sp. K showed that dGlc, a Glc analog that can be transported into the cell but not metabolized, causes inhibition of photosynthesis, suppression of both chlorophyll biosynthesis and Calvin cycle enzymes, and degradation of the chloroplast ultrastructure (Semenenko, 1981; Zvereva et al., 1981). Because these studies used a Glc analog rather than Glc, these data suggest the repression of photosynthesis response is due to Glc signaling rather than Glc metabolism. Future research will determine whether Glc and/or a downstream metabolite is involved in the Glc-dependent effects on metabolism in algae. Our current study describes controllable repression and activation of photosynthesis that will enable future studies on mechanisms of regulation of photosynthesis and thylakoid assembly and disassembly in green algae.

Time-Resolved Transcriptomics Reveal Biological Processes Induced during Trophic Transitions

High-temporal resolution sampling during both the transition from photoautotrophy to heterotrophy and from heterotrophy to photoautotrophy allowed us to characterize in detail the changes in RNA abundance during algal metabolic acclimation. Not surprisingly, our transcriptome results showed a strong influence of the diurnal cycle, as has been well characterized in C. *reinhardtii* (Zones et al., 2015). However, using time-matched controls, we were able to distinguish the effects of the diurnal cycle and trophic transitions. Repressing and activating photosynthesis, and transitioning to different trophic modes, involved a global change including differential regulation of one-third of all gene models. This response was considerably larger (approximately sixfold) than the transcriptomic response to high light in *C. zofingiensis* (Roth et al., 2017). Glc addition caused differential gene expression of a substantially larger number of genes than did Glc removal (5755 and 3113 DEGs, respectively). These experiments highlighted the reversibility of gene expression between the two treatments, suggesting that these are likely to be transcriptionally regulated biological processes.

The photosynthetic and metabolic changes in the cell were driven by changes in gene expression. The large-scale, reversible changes in the transcriptome correlated with the addition and removal of Glc and were consistent with significant physiological modifications. Enriched pathways in our analyses included glycolysis, starch synthesis and degradation, chlorophyll biosynthesis, FA metabolism and TAG biosynthesis, photosynthesis, b-oxidation, and the glyoxylate cycle. The increase in astaxanthin and FA biosynthesis genes with Glc is consistent with findings of a previous study (Huang et al., 2016). Similarly, sugars in plants and algae play a major role in regulating thousands of genes, and studies have shown sugar-mediated repression of photosynthetic genes, and induction of carbon storage genes (Rolland et al., 2006; Gao et al., 2014; Sheen, 2014; Fan et al., 2015). This study provides a high-resolution transcriptome analysis revealing the biological processes that underlie the reversible,

Glc-mediated transitions between photoautotrophy and heterotrophy in *C. zofingiensis*. This data set and our co-expression analyses will also enable future studies to investigate other processes during trophic transitions in green algae.

Transcriptome Dynamics Correspond with Changes in Energy Reserves and Ketocarotenoids

The repression and activation of photosynthesis that we observed coincided with a redirection of carbon metabolism toward increased energy reserves in the form of starch and lipids, specifically TAGs. We observed a nearly fivefold increase in starch volume within 84 h (Figures 5B and 5C), and the starch pathway was functionally enriched in cluster 2 (Figure 11), which contained genes that were upregulated with Glc addition and downregulated with Glc removal. Surprisingly, starch degradation was also functionally enriched; therefore, the observed starch accumulation may be due to posttranscriptional mechanisms that promote the anabolism. This study, consistent with previous studies, showed that Glc could induce production of TAGs and astaxanthin in *C. zofingiensis* (Huang et al., 2016; Zhang et al., 2016). However, *C. zofingiensis* contrasts with *Chlorella pyrenoidosa*, which accumulates starch during heterotrophy but lipids during photoautotrophy (Fan et al., 2015).

Combining lipid analyses and transcriptional changes, our study provides an in-depth look at genes involved in altering lipid abundances and provides insights into mechanisms controlling TAG accumulation. We observed increases in TAGs and decreases in thylakoid membrane lipids, in particular MGDG. Transcriptome analysis showed significant increases in the gene expression of nearly all genes involved in glycolysis, and this likely provides sufficient pyruvate for increased lipid biosynthesis.

With Glc, multiple steps of FA biosynthesis and elongation were upregulated, similar to what was previously reported at 96 h with Glc under low light (Huang et al., 2016). In *A. protothecoides*, an increase in FAs was observed with Glc, and it was hypothesized that Glc leads to an increase in glucose-6-phosphate and ultimately acetyl-CoA, which serve as precursors for FAs and lipid biosynthesis (Matsuka et al., 1969).

RNA abundance of genes leading to production of MGDG, DGDG, SQDG, and DGTS was relatively unresponsive or downregulated during the experiment. By contrast, *PGD1* was actively upregulated with Glc (Figure 14). *PGD1* encodes a lipase whose homolog in *C. reinhardtii* degrades MGDG by removing an acyl chain which is either recycled into MGDG or exported and added to TAGs (Li et al., 2012; Du et al., 2018). Thus, our data provide evidence that the decrease in MGDG abundance results from active lipid degradation rather than decreased expression of biosynthetic enzymes and that PGD1-derived free FAs may be precursors to TAGs. During Glc removal, *PGD1* is downregulated, and MGDG abundance increases and TAG abundance decreases. Moreover, genes involved directly in TAG biosynthesis were upregulated with Glc. MLDP is a green algal–specific protein that was shown in *C. reinhardtii* to form a proteinaceous coat to stabilize mature TAG lipid droplets (Moellering and Benning,

2010; Tsai et al., 2015). In our study, *MLDP* was strongly upregulated with Glc and downregulated with Glc removal, providing further support of the function of this protein across different green algal species. With Glc, we also observed significant reversible downregulation of the gene encoding a potential TAG lipase (*LIP1*). This putative TAG lipase likely removes a fatty acyl chain for transport to the peroxisome leading to breakdown to acetyl-CoA via b-oxidation. This may be further processed to malate via the glyoxylate cycle. Our results identify this lipase as a potential bioengineering target for increasing the accumulation of TAGs in *C. zofingiensis*. Glc removal led to restoration of photosynthesis and the photosynthetic machinery within a day, and this was followed by a decrease in TAGs and an increase in thylakoid membrane lipids concomitant with changes in RNA abundance for genes involved in glycolysis, lipid biosynthesis, and FA modification (Figure 14).

We also observed an increase in ketocarotenoids with Glc, as can be seen both by the change in culture color from green to orange and by ketocarotenoid peaks in the high-performance liquid chromatography (HPLC) chromatogram (Figure 1B; Figure 3D). This result is consistent with results from multiple studies quantifying these pigments (e.g., Ip et al., 2004; Ip and Chen, 2005; Li et al., 2008; Sun et al., 2008). The upregulation of *LCYB1* and downregulation of *LCYE1* gene expression with Glc support a preference for the b-carotene branch, which produces the ketocarotenoids, similar to a previous study conducted under low light and longer timescales (Huang et al., 2016). Huang et al. (2016) reported an increase in RNA abundance of both *BKT1* and *BKT2* transcripts at 72 and 96 h, whereas our study within 12 h showed upregulation of *BKT1*. The function of *BKT2* and whether it is involved in astaxanthin biosynthesis remain unknown, as *bkt1* mutants do not make astaxanthin under high light (Roth et al., 2017). A previous study suggested that Glc induces ketocarotenoid biosynthesis genes *BKT1* and *CHYB1* through different signaling pathways but that both are necessary for astaxanthin accumulation (Li et al., 2008). Recently, the connection between the accumulation of TAGs and astaxanthin in *C. zofingiensis* has become of interest to improve commercialization of this alga. Either Glc addition or nitrogen starvation stimulates the accumulation of TAGs and astaxanthin (this study; Breuer et al., 2012; Mulders et al., 2014; Huang et al., 2016; Zhang et al., 2016). Astaxanthin accumulates primarily as astaxanthin esters with mono or diesterified FAs. It is possible that the esterified forms of astaxanthin colocalize with TAGs in cytoplasmic lipid droplets near the cell membrane, but further work is needed to test this hypothesis. It has been hypothesized that ketocarotenoid and FA biosynthesis compete for carbon precursors and reducing power (Zhang et al., 2016). In *Haematococcus pluvialis* (Chlorophyceae), there is coordination between astaxanthin and FA biosynthesis at the metabolite level, but not the transcript level (Chen et al., 2015a). Additionally, both free and esterified forms of astaxanthin are made in the endoplasmic reticulum in *H. pluvialis* (Chen et al., 2015a). In *C. zofingiensis*, it remains unknown where astaxanthin is formed and whether b-carotene or zeaxanthin is the precursor to astaxanthin.

Glucose-Induced Changes in Cell Division

Proper regulation of cell growth and division not only is a critical aspect of development but also has the potential to be exploited to improve production of microalgae. Not surprisingly, in this study we observed a six-fold increase in biomass of *C. zofingiensis* cultures with Glc and light compared with photoautotrophic cultures within 84 h (Figure 1C). The increased culture biomass is primarily due to increasing cell size rather than cell division. Heterotrophic cell density increased slowly (Figure 1D), but cells reach greater than seven-fold the volume of photoautotrophic cells (Figure 1E). These data suggest that when *C. zofingiensis* is grown with Glc, cells invest in growing larger rather than in cell division, at least initially. Although *C. zofingiensis* can mature into cells with multiple nuclei (Fucikova and Lewis, 2012), our study was conducted with young cells that contain a single nucleus and chloroplast. By contrast, Glc removal promoted cell division, resulting in smaller cells. Thus, our study shows that Glc is able to affect cell cycle regulation and control cell division, albeit through unknown mechanisms. Like *C. reinhardtii*, *C. zofingiensis* exhibits multiple fission (Dönz, 1934; Roth et al., 2017), in which a cell undergoes multiple rounds of reductive cell division to produce 2^n daughter cells (Umen, 2018). Multiple fission in *C. zofingiensis* results in similar-sized daughter cells regardless of whether they came from photoautotrophic or heterotrophic cells. Similarly, in A. protothecoides, Glc delays DNA synthesis and cell division (Hase, 1975). Future studies with *C. zofingiensis* should provide insight into nutritional influences on cell growth and division.

C. zofingiensis Is an Emerging Model Organism

C. zofingiensis is emerging as a new model for oleaginous algae. With a small genome of about 58 Mbp (Roth et al., 2017) and a readily controlled photosynthetic and metabolic switch, as shown by this study, *C. zofingiensis* can be exploited as an experimental organism for investigating Glc responses, carbon metabolism and partitioning, disassembly and reassembly of the photosynthetic apparatus, astaxanthin biosynthesis, as well as TAG accumulation and remobilization from cytoplasmic oil bodies. Ultimately, understanding these fundamental processes in *C. zofingiensis* will enable bioengineering approaches to enhance production of biofuel precursors and high-value products from microalgae.

ACKNOWLEDGMENTS

We thank Dagmar Lyska and Setsuko Wakao for help with sample collection; Kent McDonald for help and advice with TEM; Winnie Wong for help with cryo-SXT segmentation; Tim Jeffers, Chris Baker, Chris Gee, and Setsuko Wakao for helpful discussions; and Tim Jeffers for comments on the manuscript. Analysis of physiological experiments and RNA-seq data was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (award DE-

SC0018301). The U.S. Department of Agriculture National Institute of Food and Agriculture supported sequencing through the Agriculture and Food Research Initiative Competitive Grant (2013-67012-21272 to M.S.R.). Lipid analysis was performed at the U.S. Department of Energy Joint Genome Institute, a Department of Energy Office of Science User Facility (contract no. DE-AC02-05CH11231). Cryo-SXT was supported by the U.S. Department of Energy, Office of Science, through the Photosynthetic Systems program in the Office of Basic Energy Sciences. The National Center for X-ray Tomography is supported by the National Institute of General Medical Sciences of the National Institutes of Health (grant P41GM103445) and the U.S. Department of Energy, Office of Biological and Environmental Research (grant DE-AC02-05CH11231). Carl Zeiss Elyra PS.1 SIM was supported by the National Institutes of Health S10 program (grant no. 1S10OD018136-01). M.A. acknowledges support for electron microscopy analysis from the Department of Energy Joint BioEnergy Institute, Office of Biological and Environmental Research (grant no. DE-AC02-05CH11231) and the National Institutes of Health National Institute of General Medical Sciences (grant P01GM051487). D.J.W. was supported by a National Science Foundation Graduate Research Fellowship. K.K.N. is an investigator of the Howard Hughes Medical Institute. The U.S. Government and the publisher, by accepting the article for publication, acknowledge that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this article, or allow others to do so, for U.S. Government purposes.

In Conclusion

The unicellular green alga, C. zofingiensis displays flexible metabolism. When supplemented with glucose in the media, it fully represses photosynthetic metabolism and shifts to a heterotrophic mode. This process is reversible. However, the underlying signaling mechanisms remain to be described. In Chapter 3, we use forward genetics to reveal the genetic determinants of this metabolic shift. This work provides the foundational knowledge necessary to bridge gaps of understanding between our understanding of glucose signaling in plant and animal literature.

Chapter 3

Hexokinase is necessary for glucose-mediated photosynthesis repression and lipid accumulation in a green alga

Preface

As shown in Chapter 2, my work, and the work of my collaborators, has thoroughly detailed the physiological response of C. zofingiensis to glucose. We have described the basic biology of glucose-responsiveness of C. zofingiensis. We demonstrate that C. zofingiensis is able to repress photosynthesis when given glucose, as evidenced by a complete loss of photosynthetic capacity, as well as the reduction in expression of photosynthetic genes, and the loss of structural components required for photosynthesis (Roth et al., 2019). What we have not yet addressed, by what means C. zofingiensis processes information about glucose status. My work up to this point had been unable to describe what mechanisms allow for such a glucose-based response.

To that end, we endeavored to uncover the mechanisms and genetic determinants of this observed metabolic rewiring in C. zofingiensis. Although we had the capacity to apply forward genetics, we could not use classical genetic techniques to determine any causal mutations that explain this metabolic shift. To our knowledge, C. zofingiensis is not easily crossed, and we have not observed sexual recombination. Therefore, our approach was to use a classical genetic selection combined with whole-genome sequencing to identify mutations which could inform our understanding of glucose-induced photosynthetic repression. This proved to be fruitful.This work was originally published under a title of the same name in September of 2019 in the journal Communications Biology. As this was a collaborative effort submitted in support of an individual doctoral degree, an explicit breakdown author contributions to the work is below.

Citation:
Roth, M.S., Westcott, D.J., Iwai, M. *et al.* Hexokinase is necessary for glucose-mediated photosynthesis repression and lipid accumulation in a green alga. *Commun Biol* **2,** 347 (2019). https://doi.org/10.1038/s42003-019-0577-1

Melissa Roth and Kris Niyogi conceived of the study.

Experimental details:

Melissa Roth generated mutant strains, measured photosynthetic capacity, oxygen evolution, pigment analysis (extraction and HPLC), and performed western blots (protein extraction, SDS-PAGE, antibody optimization).

Daniel Westcott performed mutant analysis (DNA extraction, sequencing analysis, strain confirmation), TLC of lipids (lipid extraction, solvent optimization, and qRT-PCR (RNA extraction, primer design and verification and data analysis).

Masakazu Iwai performed Structured Illumination Microscopy and helped to optimize western blots.

Figures were generated collaboratively, however all plots excluding HPLC data were generated by Daniel Westcott

Writing:

Melissa Roth wrote the paper, in part based on research summaries generated by Daniel Westcott. Daniel Westcott, Masakazu Iwai and Kris Niyogi provided edits on the manuscript.

Kris Niyogi provided Resources, supervision, and editorial oversight.

Abstract

Global primary production is driven largely by oxygenic photosynthesis, with algae as major contributors. The green alga Chromochloris zofingiensis reversibly switches off photosynthesis in the presence of glucose in the light and augments production of biofuel precursors (triacylglycerols) and the high-value antioxidant astaxanthin. Here we used forward genetics to reveal that this photosynthetic and metabolic switch is mediated by the glycolytic enzyme hexokinase (CzHXK1). In contrast to wild-type, glucose-treated hxk1 mutants do not shut off photosynthesis or accumulate astaxanthin, triacylglycerols, or cytoplasmic lipid droplets. We show that CzHXK1 is critical for the regulation of genes related to photosynthesis, ketocarotenoid synthesis and fatty acid biosynthesis. Sugars play fundamental regulatory roles in gene expression, physiology, metabolism, and growth in plants and animals, and we introduce a relatively simple, emerging model system to investigate conserved eukaryotic sugar sensing and signaling at the base of the green lineage.

Introduction

Global net primary production (NPP) depends on regulation of oxygenic photosynthesis and primary carbon metabolism. Algae and cyanobacteria are responsible for approximately half of global NPP (Field et al., 1998). Algae can have flexible metabolism, and they can grow either photoautotrophically or heterotrophically with an exogenous reduced carbon source (X. Johnson & Alric, 2012; Shihira & Krauss, 1965). Many algae, including the reference alga *Chlamydomonas reinhardtii*, which can grow on a reduced carbon source, will also maintain photosynthesis and are thus capable of mixotrophic growth in the light (Merchant et al., 2007). In contrast, *Chromochloris zofingiensis*, a unicellular coccoid green alga that is closely related to *C. reinhardtii* (Roth et al., 2017), reversibly turns off photosynthesis, degrades the photosynthetic apparatus, and reduces thylakoid membranes in the presence of glucose in the light (Roth, Gallaher, et al., 2019). *C. zofingiensis* can be cultivated on a variety of sugars and in multiple trophic modes, and it can produce high amounts of biofuel precursors and astaxanthin under conditions such as heterotrophy, nitrogen deprivation, and high light (Breuer et al., 2012a; W. Huang et al., 2016; Mulders et al., 2015; Roth et al., 2017; Roth, Gallaher, et al., 2019). Recent commercial interest in this alga has grown due to its economically promising attributes (J. Liu et al., 2014a). The easily controlled, glucose-induced photosynthetic switch in *C. zofingiensis* offers the opportunity to gain insight into the molecular players and mechanisms regulating photosynthesis and carbon metabolism in algae, which will enable rerouting and engineering of metabolism to improve commercial prospects of algal bioproducts.

Sugars are main carbon and energy sources of cells. In many organisms, sugars have regulatory roles that affect metabolism, growth, aging, and stress resistance. In plants, sugars are key metabolic and regulatory molecules, and their roles in signaling have been extensively investigated (F. Rolland & Sheen, 2005; Sheen, 2014). The preferred carbon source for many organisms is glucose. Exogenous glucose has been shown to repress photosynthesis in plants while modulating growth and development (F. Rolland & Sheen, 2005; Sheen, 2014). However, a knowledge of the molecular players and mechanistic understanding of glucose signaling in green algae, a sister group to land plants, is presently lacking, likely due to the inability of *C. reinhardtii* to grow on glucose as a sole carbon source (Harris, 2009). Identification of the molecular players and mechanisms underlying photosynthetic and metabolic changes in response to glucose and other sugars in green algae will provide a phylogenetic perspective on sugar signaling and reveal insights into fundamental mechanisms present at the base of the green lineage.

In this study, we investigated the regulation of photosynthesis and metabolism in *C. zofingiensis* in response to exogenous glucose, building on the foundation provided by the recent high-quality, chromosome-level genome and transcriptome of *C. zofingiensis* (Roth et al., 2017). Through a forward genetics screen, we uncovered hexokinase1 (CzHXK1) as a master regulator of photosynthesis, carbon metabolism, and ketocarotenoid (astaxanthin) biosynthesis in *C. zofingiensis*. The hxk1 mutants are deficient in turning off photosynthesis and accumulating astaxanthin and lipids in the

presence of glucose. HXK, a key enzyme in carbon metabolism, generates glucose-6-phosphate (G6P) that can be used in a variety of pathways including glycolysis, the oxidative pentose phosphate pathway, and starch and cell wall biosynthesis (Claeyssen & Rivoal, 2007). HXK is also known to be an evolutionarily conserved eukaryotic glucose sensor in yeast and plants (Filip Rolland et al., 2006; Sheen, 2014). In contrast to yeast and plants, *C. zofingiensis* has a single gene encoding hexokinase (Roth et al., 2017). Thus, our work establishes *C. zofingiensis* as a relatively simple system to investigate conserved mechanisms of sugar sensing and signaling in the green lineage.

METHODS

Strain and growth conditions. We used the *Chromochloris zofingiensis* strain SAG 211-14 obtained from the Culture Collection of Algae at Goettingen University, whose genome was published recently (Roth et al., 2017). The cells were grown in liquid cultures shaking at 100–150 rpm under diurnal conditions (16 h light, 8 h dark) at a light intensity of 100 µmol photons m^{-2} s^{-1} (cool white spectrum) and at 25 °C as previously described (Roth et al., 2017; Roth, Gallaher, et al., 2019). Cells were grown in Proteose medium (UTEX Culture Collection of Algae) with Chu's micronutrient solution (2 mL/L, UTEX Culture Collection of Algae). For the glucose experiment, cells were grown until exponential phase (~4 × 10^6 cells/mL) and distributed into separate beakers for acclimation 1 day prior to each experiment. For physiological measurements, the glucose addition was conducted by adding glucose to cultures at a final concentration of 35 mM, and photoautotrophic controls were maintained in parallel. Measurements were conducted after 84 h of treatment. For qRT-PCR measurements, the glucose addition experiment was conducted by adding glucose to cultures at a final concentration of 10 mM, and photoautotrophic controls were maintained in parallel. Cells were collected at 0.5 h and 12 h after the glucose addition during daylight hours. Unless otherwise specified, cells were collected by centrifugation (3200 × g for 5 min) and 3–4 biological replicates were used. Cells were counted and sized with the Multisizer 3 Coulter Counter (Beckman Coulter).

Mutant generation and identification.

A non-targeted, forward genetics lethal screen (selection) was used to generate hxk1 mutants. Exponentially growing cultures (2–5 × 10^6 cells/mL) were subjected to ultraviolet radiation (80,000 µJoules) in the Stratalinker 1800UV crosslinker (Agilent Genomics), and cells (~35,000/plate) were plated onto 5 plates with selective medium (Proteose with 5 mM 2-deoxyglucose). Cells that survived were isolated, and whole-genome sequencing was used to identify the mutations.

Genomic DNA was isolated using a cetyltrimethylammonium bromide (CTAB)-based phenol-chloroform extraction as previously described (Roth et al., 2017). Pelleted cells were resuspended in sodium dodecyl sulfate-Edward's Buffer (SDS-EB; 20 mM Tris/HCl, 10 mM EDTA, 100 mM NaCl, 2% SDS) and incubated with slight agitation for

2 h at 65 °C. In all, 170 µL of 5 M NaCl and 135 µL CTAB solution (274 mM CTAB, 700 mM NaCl) was added, and the solution incubated for 15 min at 65 °C. In all, 400 µL of phenol-chloroform solution (25:24:1 phenol:chloroform: isoamyl alcohol) was added, and the total mixture was vortexed before centrifugation (19,283 × g for 5 min). The aqueous phase was collected and incubated with 5 units RNAse A (Qiagen, Cat no. 19101) at 37 °C for 20 min. Two additional phenol-chloroform extractions were performed before adding 400 µL of 24:1 chloroform:isoamyl alcohol and vortexing. After centrifugation (19,283 × g for 5 min), the aqueous phase was transferred to a new tube where 0.1x volumes of 5 M NaCl and 0.7x volumes of isopropanol was added. This was mixed gently by inversion 4–6 times and centrifuged (19,283 × g for 15 min at 4 °C). The pellet was washed twice with 500 µL 4 °C 70% ethanol and briefly air-dried before being resuspended in 100 µL MilliQ water. DNA was incubated for 2 h at 25 °C followed by 12 h at 4 °C before quantification with a NanoDrop 2000 (Thermo Scientific).

Illumina library preparation was performed by UC Berkeley's Functional Genomics Laboratory. Double-stranded DNA was fragmented using Covaris sonication, and size distribution was checked using an Agilent 2100 Bioanalyzer. Sequencing and barcoding adapters were added, and the size distribution was confirmed by Bioanalyzer. Illumina sequencing was performed by UC Berkeley's Vincent J. Coates Sequencing Facility. 100 bp, paired-end reads were performed on an Illumina Hi-Seq 4000. High sequencing coverage (70–120×) was obtained for the eight mutant strains (Table 1).

Whole-genome analysis of mutants was performed using a combination of the Broad Institute's Genome Analysis Took Kit's best practices and SnpEff. Read quality was assessed, and low-quality reads were trimmed(Cingolani et al., 2012; DePristo et al., 2011). FastQ reads were aligned using Burrows Wheeler Aligner using the published reference genome (H. Li & Durbin, 2009; Roth et al., 2017). The output SAM files were sorted, indexed, and deduplicated using Picard Tools (GATK). GATK was used to generate a realigned BAM file from which variants were called with the Haplotype Caller tool. This naive dataset was used to bootstrap GATK's Base Quality Score Recalibrator tool. Variants were called, extracted, and filtered on the recalibrated scores. SnpEff was used to call variants of high impact. By examining SNPs shared among the eight mutant strains and not in WT, a list of likely candidate genes was generated. Mutations in CzHXK1 were confirmed by Sanger sequencing. SNPs, INDELS, mutation rates, and transition/transversion rates between mutants were generated using GATK to verify that each mutant line represents an independently generated allele.

Photosynthetic efficiency.

Photosystem II efficiency was measured with the Hansatech FMS2 system as previously described (Roth, Gallaher, et al., 2019). Cells were dark-acclimated while shaking for 30 min. In all, 3.5×10^6 cells were collected onto a glass fiber filter, which was placed into the instrument's leaf clip. The maximum efficiency of photosystem II, $(F_m - F_o)/F_m = F_v/F_m$, was measured using a 0.5 s saturating pulse (>2000 µmol quanta m^{-2} s^{-1}).

Oxygen consumption and net oxygen evolution.

Oxygen consumption and net oxygen evolution were measured with the Oxygraph Plus System (Hansatech Instruments) as previously described (Roth, Gallaher, et al., 2019). Cells were adjusted to 1×10^7 cells/mL and dark-acclimated for >30 min. Oxygen consumption was measured over 1 min after reaching steady-state in the dark at a constant temperature of 25 °C. Net oxygen evolution was also measured over 1 min after reaching steady-state under actinic light of 100 µmol photons m^{-2} s^{-1} and at a constant temperature of 25 °C. For cultures with cell volume larger than 100 μm^3, additional samples were run with a cell density of ≤5×10^6 cells/mL to achieve OD at 750 nm of ~0.2–0.4. Oxygen consumption and net oxygen evolution were normalized to cell volume, because cell volume changed during the experiment.

Immunoblot analysis.

Immunoblot analysis was conducted as previously described (Roth, Gallaher, et al., 2019). Cells were pelleted by centrifugation and homogenized with solubilization buffer (Tris-HCl, pH 6.8, 3.5% SDS, 6% urea, 10% glycerol) and lysing matrix D for 3 × 60 s with the FastPrep-24 5 GTM High Speed Homogenizer (6.5 m s^{-1}, MP Biomedical). Chloroform-methanol protein purification was performed as previously described (Wessel & Flügge, 1984). Proteins were resolubilized with solubilization buffer with 25 mM DTT. 10 µg of protein were run with loading buffer on Mini-PROTEAN TGX gels (Bio-Rad), transferred to PVDF membranes, and immunoblotted with anti-PSBD (1:5000; Agrisera), anti-PsbC (1,1000; Agrisera), anti-PsaA (1:1000; Agrisera), anti-PetB (1:10,000; Agrisera), anti-P17 (1:5000; from Bassi and Wollman 1991), anti-LHCA2 (1:5000; Agrisera), or anti-AtpB (1:5000; Agrisera) antibodies (Bassi & Wollman, 1991). Because the AtpB antibody detects both chloroplast and mitochondrial $F_1\beta$ subunits, Expasy compute was used to discriminate these two variants. CzCPg01090 was predicted to be ~50 kDa and Cz03g32200 was predicted to be ~65 kDa. In addition, both PredAlgo and TargetP prediction software predicted an N-terminal mitochondrial targeting sequence for Cz03g32200. Proteins were visualized with anti-Rabbit IgG (1:10,000; GE Healthcare) and a ChemiDoc MP Imaging System (Bio-Rad). Protein concentration was determined by DC Protein Assay (Bio-Rad), and equal protein levels were confirmed by Coomassie Brillant Blue stain. Two biological replicates for each primary antibody were used.

Pigments.

Pigments were determined by high performance liquid chromatography (HPLC; 100 HPLC, Agilent) as previously described (Roth et al., 2017; Roth, Gallaher, et al., 2019). Cells were pelleted and homogenized with acetone and lysing matrix D for 3 × 60 s with the FastPrep-24 5 GTM High Speed Homogenizer (6.5 m s^{-1}, MP Biomedical). The cell debris was pelleted by centrifugation (20,000 × g for 3 min) and the supernatant was collected. To ensure complete extraction, acetone extractions were repeated three times. Absorbance at wavelengths 445 nm and 520 nm were used for pigment

quantification.

Structured illumination microscopy.

To stain neutral lipids, the pelleted cells were resuspended with 5 µg/mL BODIPY^{TM} 493/503 dye (Thermo Fisher Scientific) in Proteose medium as previously described (Iwai et al., 2018; Roth, Gallaher, et al., 2019). The cells were incubated in the dark for 10 min and washed (3×) with 1 mL of Proteose medium and centrifuged (3000 × g for 1 min). After washing, the pelleted cells were resuspended with 50–200 µL of 0.5% low melting-point agarose prepared with Proteose medium. 4 µL of resuspended cells were immediately mounted between two microscope coverslips. After the agarose solidified, the mounted coverslips were placed in the Attofluor cell chamber (Thermo Fisher Scientific). C. zofingiensis live cells (technical replicates $n \geq 8$) were observed using Zeiss Elyra PS.1 structured illumination microscopy with objective lens Plan-APOCHROMAT 100 ×/1.46 (Zeiss). Chlorophyll and BODIPY were excited by 642 and 488 nm lasers, respectively, and fluorescence from each fluorophore was acquired through 650–730 nm and 505–550 nm bandpass filters, respectively. The image acquisition was done as fully controlled by ZEN software (Zeiss). Raw images were processed to reconstruct super-resolution 3D images using ZEN software.

Lipid extraction and thin-layer chromatography.

Lipids were determined using thin-layer chromatography (TLC). In all, 3.3×10^6 μm^3 of cells were flash frozen in liquid N_2, and samples were normalized to culture biomass using the Multisizer 3 Coulter Counter (Beckman Coulter). Frozen pellets were broken in an MP Biosciences MP-24 bead beater using lysing matrix D in a 2-mL screw-capped tube (6.5 m/s, 60 s, 3 times, –20 °C). In all, 1 mL of 2:1 chloroform:methanol was added to the disrupted cells and placed on a vortex for 5 min. 266 µL of 0.73% (w/v) NaCl solution was added, and the mixture was inverted 5-6 times to mix. Samples were then centrifuged for 5 min at 10,000 × g (9,000 RCF). The lower, solvent phase was removed. In all, 200 µL of this total lipid extract was dried under an N_2 stream and resuspended to 33.3 µL in chloroform. In total, 10 µL of the concentrated lipid extract was loaded onto a clean silica TLC plate. These were developed in either acetone:toluene:water (91:30:8) or hexane:diethyl ether:glacial acetic acid (91:39:1.3). Lipids were visualized by sulfuric acid spray and charring (25% H_2SO_4 in 50% ethanol, 100 °C for 6 min). Standards are as follows: Oleic acid (cis-9-Octadecenoic acid, Sigma Aldrich catalog number O1008, 125 µg loaded), TAG (Glyceryl trilinoleate, Sigma Aldrich catalog number T9517, 50 µg loaded), MGDG (monogalactosyldiacylglycerol, Avanti Polar Lipids catalog number 840523, 0.50 µg loaded), DGDG (digalactosyldiacylglycerol, Avanti Polar Lipids catalog number 850524, 40 µg loaded). Store-bought olive oil was extracted as described above, and 2.5 µL of extract was loaded. TLC was conducted for two biological replicates of each treatment.

qRT-PCR.

RNA was extracted from WT, *hxk1-1*, and *hxk1-2* cells as described (Gallaher & Roth, 2018; Roth et al., 2017; Roth, Gallaher, et al., 2019) with the modification of increased DNAse treatment time to 60 min. A cDNA library was generated using the OmniScript RT Kit (Qiagen) with slight modifications to the manufacturer's instructions. Two microgram RNA template was added to a total reaction volume of 20 µL. Both random nonamers (15n, 10 µM final concentration) and oligo dT (15 T, 1 µM final concentration) were used to ensure efficient reverse transcription without biasing towards polyadenylated transcripts. Four units of Omni RT enzyme were added as well as 10 units of RiboLock RNAse inhibitor and mixed by inversion. Reactions were briefly spun down and then incubated at 37 °C for 2 h. Negative RT control was performed identically by adding water in place of the Omni RT enzyme. qRT-PCR primers were designed with a T_m of 60 °C using Primer3 (Table 2). qRT-PCR primers were designed to span exon-exon junctions and were biased towards the 3′ end of the transcript where possible. Primer sets were optimized to have a doubling efficiency between 95–101%. qRT-PCR was performed using SybrGreen according to the manufacturer's recommendations on an Applied Biosystems 7500 instrument. In all, 5 µL of 1/5x RT reaction was used as a template. qRT-PCR expression level was quantified using the $\Delta\Delta C_T$ method (Livak & Schmittgen, 2001). Cz04g37020 (Aspartate-prephenate aminotransferase) was used as an internal reference gene due to its robust FPKM level and low variability in RNA-Seq data (Roth, Gallaher, et al., 2019).

Statistics and reproducibility.

Data are presented as means ± SD (n = 3–4) with individual data points shown for photosynthetic efficiency, oxygen consumption, net oxygen evolution, and pigments. Immunoblots and thin-layer-chromatography from two biological replicates were conducted. For structured illumination microscopy, technical replicates (n ≥ 8) were observed. For qRT-PCR, data represent means ± SD of three biological replicates consisting of three technical replicates each.

Table 2. Primers used for qRT-PCR analyses.

Gene ID (v5.2.3.2)	Gene name	Forward	Reverse
Cz13g07170	*HXK1*	CCGTGTGGTTTATGTGAAGCTG	TCGGCCGGTGTATTTGTCAATA
Cz12g12230	*PSAH1*	AAGGTCTGCGCAAAGTATGGCG	GGTTCTGCAGGGGGTTGTAACG
Cz16g19040	*PSBO1*	AGGAGAACAACAAGAGCGCA	ATGGACTCGAATACACCGGC
Cz13g03240	*LHC16*	AATGAGAACCTGGTGCACGCAC	CAACCTGTCCAGACCCTCACCA
Cz13g13100	*BKT1*	CCGTGTGGTTTATGTGAAGCTG	TCGGCCGGTGTATTTGTCAATA
Cz17g13100	*RBCS1*	GCCTACACTGCCCAGTACAA	TTGGCAGCTTCCACATAGTC
Cz02g32130	*COX10*	TGCCTTGTTCTTCTGGCAGATG	CTATGTTTGACAGCATGCGGAAG
Cz07g18160	*SPM1*	CTTTCGTCATTGGCCAGTGC	GAAGAGCTGCGCAATGAACA
Cz01g21060	*PYK2*	CCGCCATCACTTTGACTTCC	GACAGGCTGCCATTCTTTGT
Cz05g34160	*GAP1*	GGCACCGTAGAAGGCACAAAGG	CATGGGATCTTGGTGGGGTCCA
Cz04g09090	*SAD1*	AAAGACAGCTGTCCATAGGGTG	GTCCATTGCATCAGATGAGGGA
Cz03g33220	*FAD2*	TGTTCAGTCAGCTGCCACATTA	GCAGATGTTGGTTTCTTGCCAT
Cz04g29220	*MLDP1*	CATCCTTTGGACCTGTCTCGAA	CAGGAAGGGACACGATGGAATT
Cz06g03010	*GLK1*	CATGTTCCTGGCCATTGTGGG	ACGATCCAGCAACCGAGGCATA
Cz04g37020	*APPS1*	CAGGGCTTTGGGTCGGTG	ACCAAATGCATCCCCGGG

Results

Genetics identifies hexokinase as a regulator of photosynthesis.

To uncover the molecular players involved in the glucose-dependent photosynthetic switch in *C. zofingiensis*, we used forward genetics to identify mutants that maintain photosynthesis in the presence of glucose. Using UV mutagenesis followed by selection on the glucose analog, 2-deoxy-D-glucose (2-DOG), we generated eight independent mutant strains that did not shut off photosynthesis and grew in the presence of 2-DOG with light. 2-DOG is transported into the cell and phosphorylated but cannot be metabolized by glycolysis (Yantao Li et al., 2008; Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Wick et al., 1956). 2-DOG has been utilized to investigate sugar sensing in a variety of organisms (Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Wick et al., 1956). Wild-type (WT) *C. zofingiensis* cells sense the presence of 2-DOG and repress photosynthesis, resulting in cell death because 2-DOG is phosphorylated to a dead-end metabolic product. Mutants that are resistant to 2-DOG survive, because they do not shut off photosynthesis. To eliminate glucose transport mutants, we tested the growth of mutants in the dark. All eight mutants were able to grow in the dark on glucose, providing evidence that they can still take up and metabolize glucose. A detailed experiment on WT and two mutants showed that growth on glucose in the light is slightly faster than growth on glucose in the dark (Figure 1).

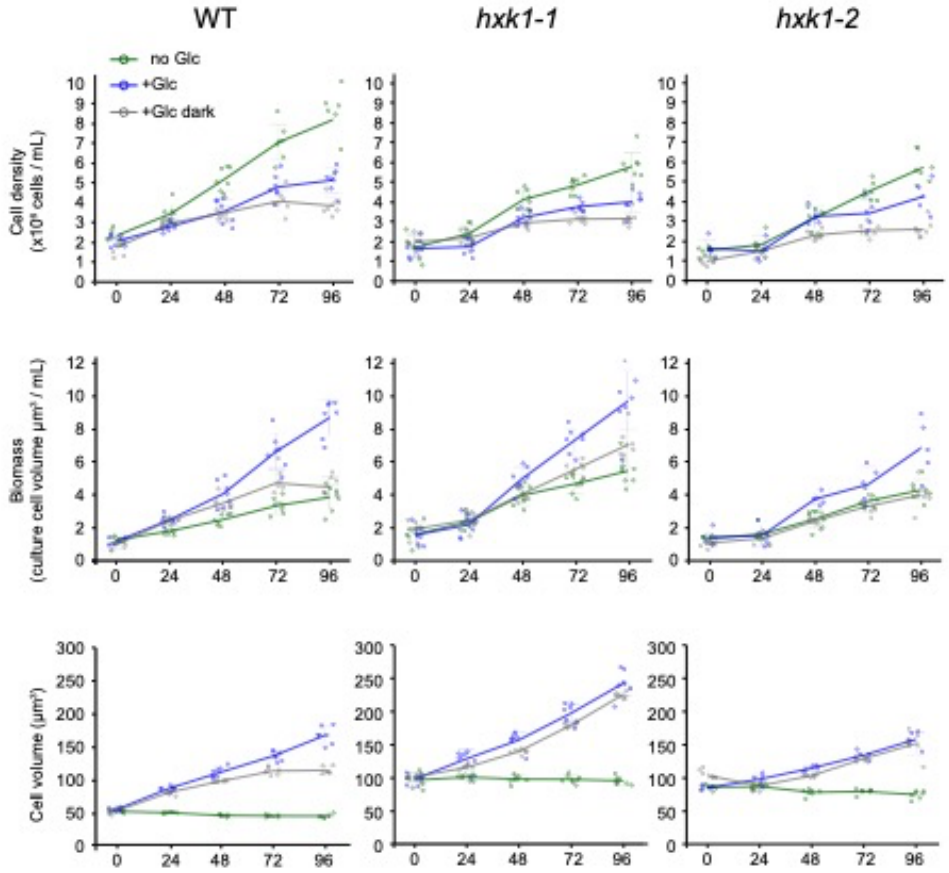


Figure 1. Growth of WT and *hxk1* mutants under various conditions. Cell density, culture biomass, and individual cell volume of cells grown without glucose in the light (green), with glucose in the light (blue), and with glucose in the dark (grey). Data represent means ± SD (n = 5-8 biological
replicates, individual data points shown).

Next, we used a whole-genome sequencing pipeline to identify potentially causative mutations (Cingolani et al., 2012; DePristo et al., 2011). High-coverage genome sequencing (average 94x coverage) revealed that all eight mutants had disruptive mutations in the single gene encoding hexokinase (Cz*HXK1*) (Figure 2a, Table 1). These mutations were confirmed by Sanger sequencing. HXK1 is both a sugar kinase and a well-studied glucose sensor and regulator in plants and yeast (Moore et al., 2003; Sheen, 2014; Vega et al., 2016). Three classes of *hxk1* mutants were found: class 1 with an insertion (C471CG) that causes an early stop codon in the second exon; class 2 with a base pair change (C3426T) that causes an early stop codon in exon 7; and class 3 mutants with a T4795G mutation that disrupts a splice site between exons 9 and 10 (Table 1). Each of the eight mutants represents an independently generated allele, as verified by analysis of the number and locations of non-causative mutations, including SNPs and INDELs (Table 1). The ability of the *hxk1* mutants to grow in the

dark on glucose is likely due to the presence of a glucokinase gene (Cz06g03010) in the genome (Roth et al., 2017).

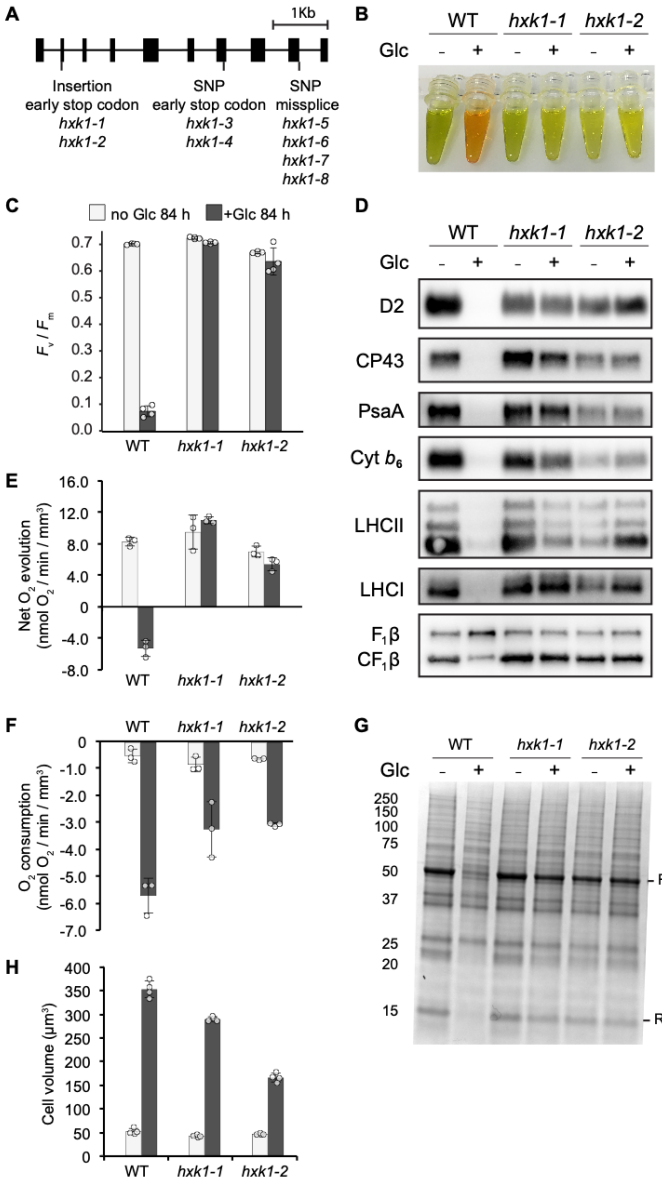


Figure 2. Mutants reveal hexokinase1 as regulator of photosynthesis.

a Gene structure of CzHXK1 mutation location and type of mutations as determined by genome sequencing (see Methods). b Image of extracted pigments from representative samples of WT and hxk1 mutants after 84 h with and without glucose (Glc). c Maximum photosystem II efficiency (Fv/Fm) of hxk1 mutants is insensitive to glucose. Data represent means ± SD (n = 4 biological replicates, individual data points shown). d Photosynthetic apparatus of hxk1 mutants is insensitive to glucose. Immunoblot analysis of PSII (D2, CP43), PSI (PsaA), cytochrome b6f (Cyt b6), light-harvesting complex of PSII (LHCII), light-harvesting complex of PSI (LHCI) and ATP synthase (AtpB) subunits after 84 h with and without glucose. The global AtpB antibody detected both the chloroplastic AtpB at ~50 kDa and the mitochondrial AtpB at ~70 kDa. Samples were normalized to total protein, and 10 µg of protein were loaded in each well. See Figure 2g for stained total protein gel loading control. Images of the detected chemiluminescent signal on the complete membranes are available in Figure 3. e Net oxygen evolution (oxygen production in the light, 100 µmol photons m–2 s–1) of hxk1 mutants is insensitive to glucose. Data represent means ± SD (n = 3 biological replicates, individual data points shown). f Respiration (oxygen consumption in the dark) of hxk1 mutants is sensitive to glucose. Data represent means ± SD (n = 3 biological replicates, individual data points shown). g SDS-PAGE analysis of total protein extracted from whole cells and stained with Coomassie brilliant blue. Samples were normalized to total protein, and 10 µg of protein were loaded in each well. h Cell volume of hxk1 mutants increases with glucose. Data represent means ± SD (n = 4 biological replicates, individual data points shown)

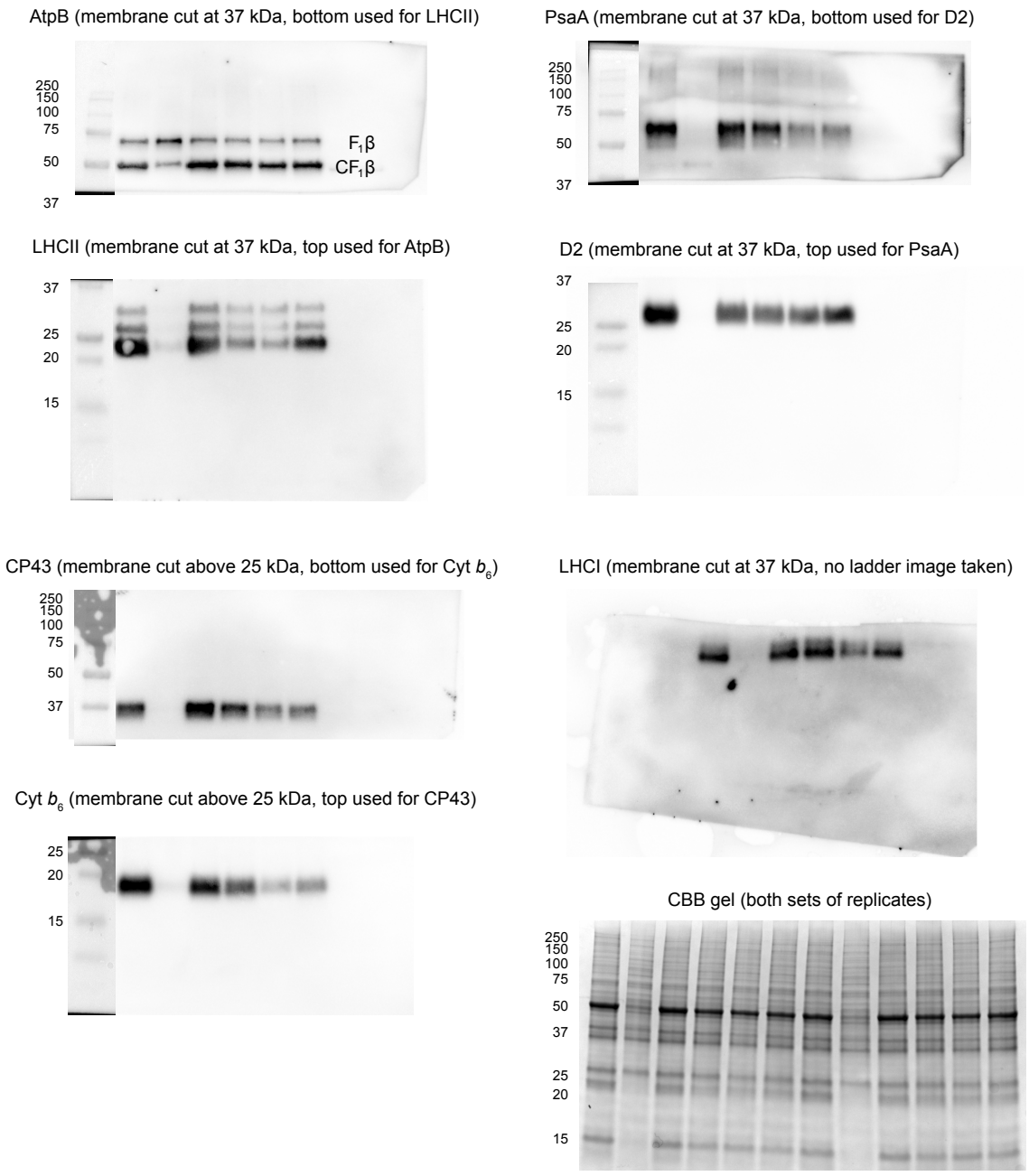


Figure 3. Supporting material for immunoblots in Figure 2D.

Immunoblot methods including antibody concentrations are described in Methods. Samples were normalized to total protein, and 10 µg of protein were loaded in each well. After protein transfer, the membranes were cut to use with antibodies of different sizes. Unspecific immunoreactions of incorrect size were observed with the PsaA antibody. The protein marker (Bio-Rad, Precision Plus ProteinTM All Blue Prestained Protein Standards #1610373) was used as a reference and imaged with white illumination

Table 1 Summary of *hxk1* mutants

Mutant Strain	Class	Mutation	Type	Predicted Result	Coverage (x)	Total Number of Variants	Number of Potential Effects	Variant Rate	SNPs	Insertions	Deletions
hxk1-1	1	C471CG	Insertion	Early Stop Codon Exon 2	*116.4*	717	2835	1.23E-05	497	105	115
hxk1-2	1	C471CG	Insertion	Early Stop Codon Exon 2	*85.8*	625	2517	1.08E-05	400	99	126
hxk1-3	2	C3426T	missense	Early Stop Codon Exon 7	92.4	622	2511	1.07E-05	366	119	137
hxk1-4	2	C3426T	missense	Early Stop Codon Exon 7	94.9	663	2686	1.14E-05	412	111	140
hxk1-5	3	T4795G	missense	Missplice	69.1	634	2618	1.09E-05	434	78	122
hxk1-6	3	T4795G	missense	Missplice	92.6	778	3028	1.33E-05	555	106	117
hxk1-7	3	T4795G	missense	Missplice	80.9	646	2514	1.11E-05	441	89	116
hxk1-8	3	T4795G	missense	Missplice	102.3	733	2922	1.26E-05	499	106	128
Sum						5418	21631	9.30E-05	3604	813	1001
Average					91.8	677.3	2703.9	0.0	450.5	101.6	125.1

Table 1 continued. Summary of *hxk1* mutants

Mutant Strain	Transitions	Transversion	Ts/Tv Ratio	A>C	A>G	A>T	C>A	C>G	C>T	G>A	G>C	G>T	T>A	T>C	T>G
hxk1-1	323	174	1.856	16	67	11	16	32	78	121	21	25	25	57	28
hxk1-2	238	162	1.469	12	61	24	14	25	56	66	26	66	15	55	20
hxk1-3	213	153	1.392	21	59	11	16	29	46	60	24	12	18	48	22
hxk1-4	240	172	1.395	25	63	16	18	25	62	57	22	24	18	58	24
hxk1-5	268	166	1.614	21	59	15	24	24	69	77	23	19	15	63	25
hxk1-6	340	215	1.581	22	65	29	27	29	102	105	25	27	18	68	28
hxk1-7	275	166	1.657	21	54	14	22	32	74	81	18	12	27	66	20
hxk1-8	322	177	1.819	23	78	22	21	26	85	90	19	13	25	69	28
Sum	2219	1385	1.602	161	506	142	158	222	572	657	178	198	161	484	195
Average	277.4	173.1	1.6	20.1	63.3	17.8	19.8	27.8	71.5	82.1	22.3	24.8	20.1	60.5	24.4

Hexokinase is required for the photosynthetic switch.

To characterize the response of *hxk1* mutants to glucose, we added glucose to photoautotrophic cultures of *hxk1* mutants and WT and conducted various physiological analyses after 84 h, when WT exhibits a complete shut-off of photosynthesis (Roth, Gallaher, et al., 2019). For these experiments, we used two independent class 1 mutants, *hxk1-1* and *hxk1-2*. These strains share an identical mutation that produces an early stop codon in CzHXK1, and thus they are most likely null mutations affecting HXK activity, but they contain different background mutations in their genomes that resulted from UV mutagenesis (Table 1). Therefore, the shared phenotypes can be attributed specifically to the shared mutation in CzHXK1 and are not likely explained by UV-induced mutations in other genes.

In response to glucose, WT changed color dramatically from green to orange, while *hxk1* mutants with and without glucose remained similar in color (Figure 2b). The *hxk1* mutants did not show a decrease in photosynthesis with glucose (Figure 2c–g). In contrast to WT, which exhibited a severe decline of photosystem (PS) II efficiency at 84 h, Fv/Fm of *hxk1-1* and *hxk1-2* did not decline with glucose (Figure 2c). Furthermore, *hxk1* mutants continued to evolve oxygen in the light at comparable rates in the presence and absence of glucose, whereas WT decreased oxygen evolution to the point of net oxygen consumption in the light (Figure 2e). Results were normalized to cell volume, because both WT and hxk1 cells grew larger with glucose (Figure 2h). As the glucose-treated cells grew larger in volume, their rate of cell division decreased (Supplementary Figure 2; Roth et al. 2019). Our previous study showed that the decreases in PSII efficiency and oxygen evolution in the light in WT are reversible with the removal of glucose within 24h (Roth, Gallaher, et al., 2019). Respiration (oxygen uptake in the dark) with glucose increased in both WT and hxk1 cells (Figure 2f). The PSII efficiency and oxygen evolution data suggest that CzHXK1 is necessary for glucose to shut off photosynthesis.

We used immunoblot analysis to characterize the changes in abundance of protein subunits of the photosynthetic apparatus. In WT, subunits of PSII (D2 and CP43), PSI (PsaA), cytochrome b_6f complex (Cyt b_6), and light-harvesting complexes (LHCI and LHCII) decreased substantially with glucose (Figure 2d). This glucose-induced loss of the photosynthetic apparatus is reversible with the removal of glucose within 24h (Roth, Gallaher, et al., 2019). Chloroplastic and mitochondrial AtpB subunits ($CF_1\beta$ and $F_1\beta$, respectively) were present in WT with and without glucose. With glucose, there was potentially a slight increase in the ratio of mitochondrial to chloroplastic AtpB, which is consistent with the increase in respiration. In contrast, immunoblot analysis of the mutants revealed that PSI, PSII, LHCs, Cyt b_6f complex, and ATP synthase were relatively insensitive to glucose and showed no consistent increases or decreases with and without glucose (Figure 2d). The loading control, an SDS-PAGE analysis of total protein extracts stained with Coomassie brilliant blue, also showed a dramatic decrease in the large and small subunits of the CO_2-fixing enzyme Rubisco at 53 kDa and 15

kDa, respectively, in WT with glucose (Figure 2g). These data suggest that the loss of the photosynthetic apparatus with glucose requires CzHXK1.

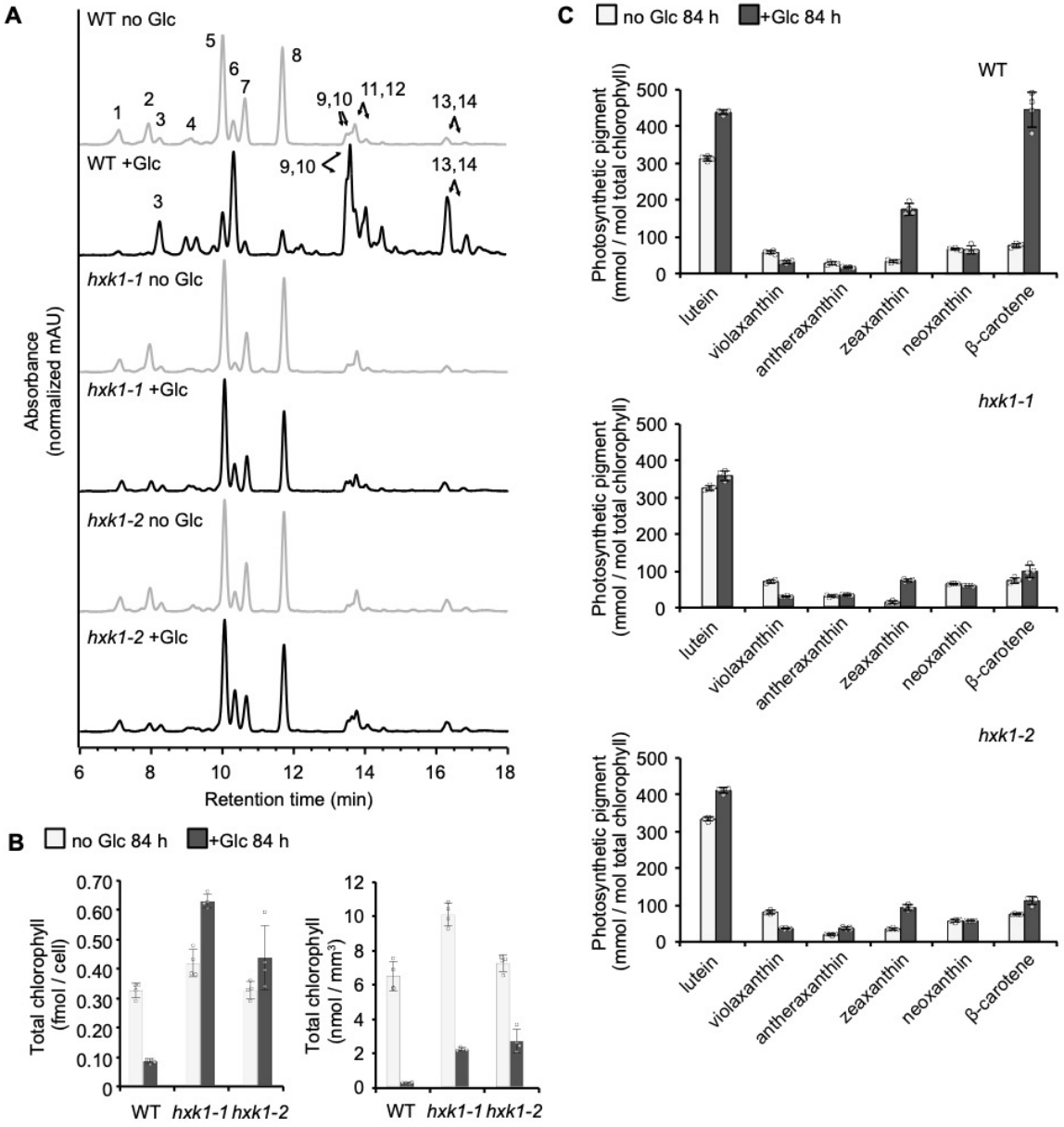


Figure 4. The pigment response of hexokinase1 mutants to glucose is attenuated.

A. Representative HPLC chromatograms of WT and hxk1 mutants after 84 h with (black lines) and without (gray lines) glucose (Glc). 1, neoxanthin; 2, violaxanthin; 3, free astaxanthin; 4, antheraxanthin; 5, lutein; 6, zeaxanthin; 7, chlorophyll b; 8, chlorophyll a; 9 and 10, esterified astaxanthin; 11 and 12, β-carotene; 13 and 14, esterified astaxanthin. b Total chlorophyll (chlorophyll a and chlorophyll b) of WT and hxk1 mutants after 84 h with and without glucose. Left graph is normalized per cell and right graph is normalized by cell volume. Data represent means ± SD (n = 4 biological replicates, individual data points shown). c Pigments relative to total chlorophyll of WT (top), hxk1-1 (middle), and hxk1-2 (bottom) after 84 h with and without glucose. Data represent means ± SD (n = 4 biological replicates, individual data points shown)

Levels of total chlorophyll (chlorophylls a and b) per cell decreased in WT but increased in the hxk1 mutants with glucose (Figure 4a, b). However, because the cells grew larger with glucose (Figure 2h), the abundance of chlorophyll per cell volume decreased in both WT and the mutants (Figure 4b). Live-cell imaging using super-resolution structured illumination microscopy (SIM) revealed a dramatic decrease in chlorophyll fluorescence intensity in WT with glucose, whereas chlorophyll fluorescence intensity remained comparable with and without glucose in hxk1 mutants (Figure 5). Altogether, measurements of photosynthesis and the photosynthetic apparatus provide evidence that HXK1 is necessary for the glucose-induced photosynthetic switch.

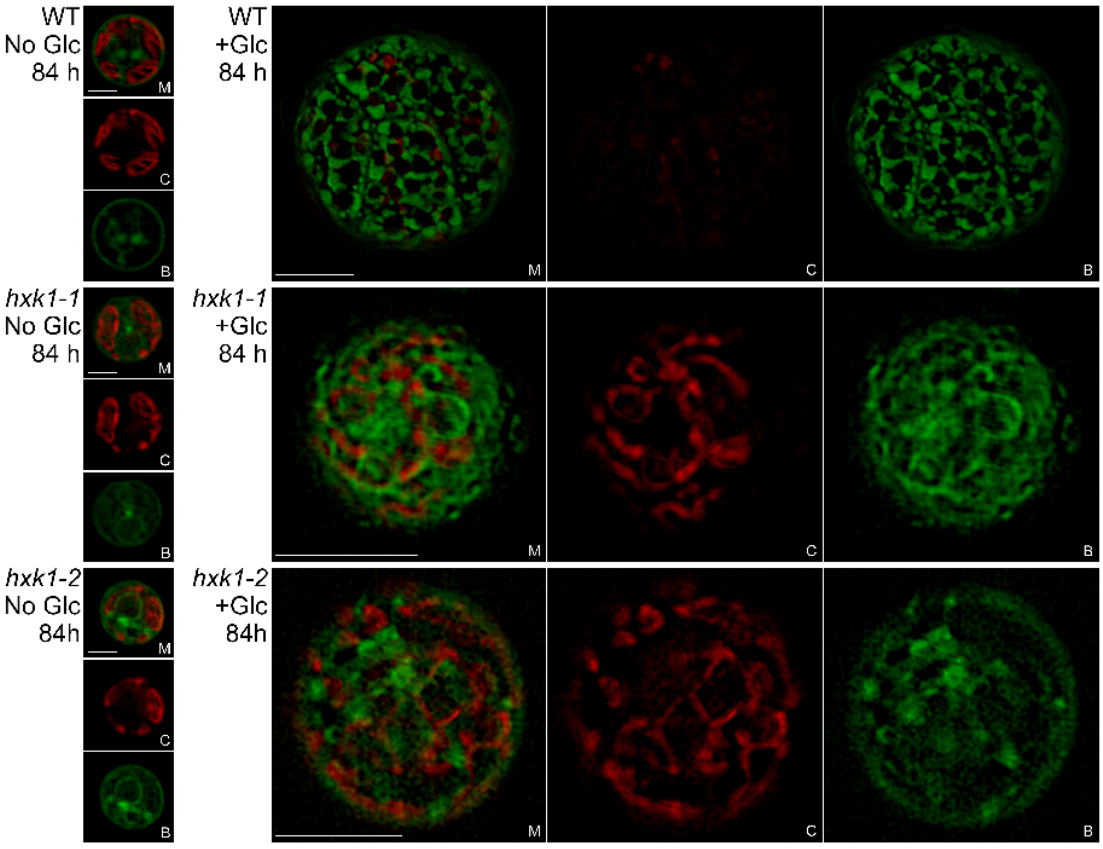


Figure 5 Hexokinase1 mutants are deficient in lipid droplet accumulation.

Representative images of live-cell super-resolution structured illumination microscopy (SIM) of WT and hxk1 mutants with and without glucose (Glc) showing (M) merged chlorophyll autofluorescence (650–730 nm) and BODIPY fluorescence (505–550 nm), (C) individual channel showing chlorophyll autofluorescence, and (B) individual channel showing BODIPY fluorescence. Scale bars, 2µm for No Glc and 5µm for +Glc

Hexokinase is required for high astaxanthin accumulation.

Under specific conditions, *C. zofingiensis* is well known to accumulate high amounts of astaxanthin (Breuer et al., 2012a; W. Huang et al., 2016; Mulders et al., 2015; Roth et al., 2017; Roth, Gallaher, et al., 2019). Astaxanthin accumulates in esterified forms, first as fatty acyl mono-esters and secondarily as di-esters (Bar et al., 1995). In WT with glucose, the astaxanthin mono-and di-esters accumulated as the culture changed from green to orange (Figures 2b, 3a). In contrast, the hxk1 mutants remained green with glucose, and we did not observe an increase in astaxanthin (Figures 2b, 3a). Similar to WT, *hxk1* mutants did show an increase in the ratio of zeaxanthin to violaxanthin and an increase in β-carotene with glucose (Figure 4d–f). Zeaxanthin and/or β-carotene are hypothesized to be precursors of astaxanthin in *C. zofingiensis* (J. Liu et al., 2014a; Roth et al., 2017). These data support that CzHXK1 is critical for the glucose-dependent accumulation of large amounts of astaxanthin.

Hexokinase is required for accumulation of cytoplasmic lipid droplets. Neutral lipid staining of live cells followed by SIM provided insight into the formation of lipid droplets in the presence of glucose. In WT, a network of cytoplasmic lipid droplets accumulated near the cell membrane in glucose-treated cells (z-stack 1 in WT+glc 84 h, Figure 5), which is consistent with previous work (Iwai et al., 2018; Roth, Gallaher, et al., 2019). However, the cytoplasmic lipid droplets observed in WT were missing in hxk1 mutants (Figure 5). Instead of clearly defined lipid droplets as in WT, the lipid staining in hxk1 mutants was diffuse. Overall, these data suggest that CzHXK1 is essential for the accumulation of cytoplasmic lipid droplets near the plasma membrane.

Hexokinase is required for accumulation of TAGs.

Commercial interest in C. zofingiensis has grown recently because it can accumulate high amounts of the preferred lipid precursor for biofuels, triacylglycerol (TAG), under nitrogen deprivation or with glucose (Breuer et al., 2012a; W. Huang et al., 2016; Mulders et al., 2014; Roth, Gallaher, et al., 2019). Lipid content assayed using thin-layer chromatography showed that while WT accumulates high levels of TAG with glucose, the hxk1 mutants do not accumulate TAG with glucose (Figure 6). It is likely that in WT, TAG accumulates in the neutral cytoplasmic lipid droplets near the plasma membrane as visualized by SIM (Figure 5). In addition, WT with glucose showed a reduction in the most abundant thylakoid lipids, monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), as previously observed (Roth, Gallaher, et al., 2019). In contrast, the hxk1 mutants did not decrease thylakoid lipids with glucose, and this result is consistent with the maintenance of the photosynthetic apparatus (Figure 2). Taken together, these data provide evidence that CzHXK1 is important for TAG accumulation with glucose.

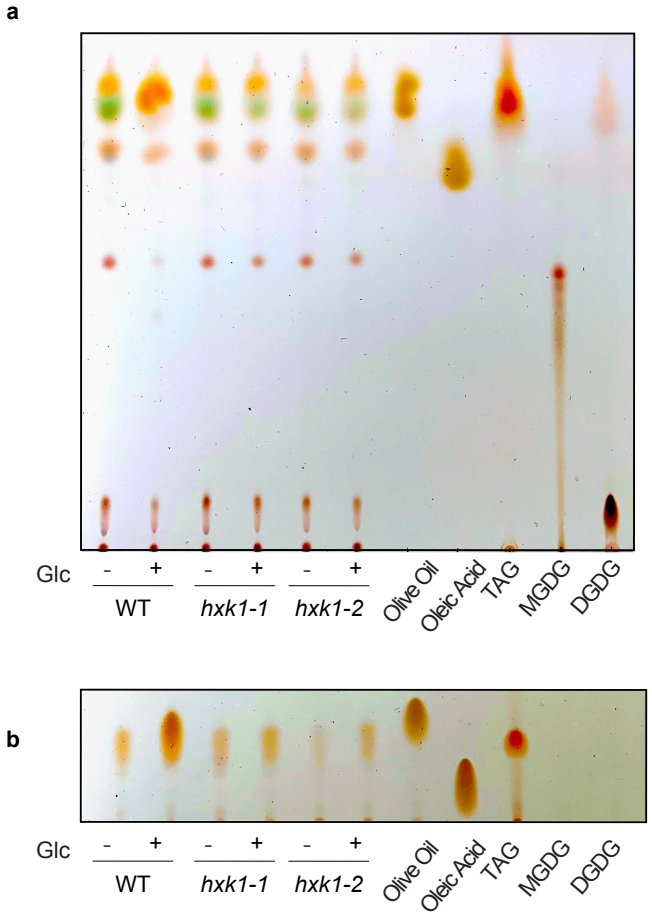


Figure 6 Hexokinase1 mutants are deficient in triacylglycerol accumulation.

Lipid extracts of WT and hxk1 mutants after 84 h with and without glucose (Glc) developed on silica TLC plates with a acetone:toluene:water (91:30:8) and b hexane:diethyl ether:glacial acetic acid (91:39:1.3). Solvent used in a shows more lipid classes and solvent used in b shows greater separation of TAG. Samples were normalized to culture biomass. 125 µg oleic acid, 50 µg TAG, 50 µg MGDG, 40 µg DGDG, and 2.5 µL of olive oil extract were loaded as standards.

Hexokinase regulates photosynthetic and metabolic gene expression.

To assess the role of CzHXK1-dependent gene expression in the glucose-dependent photosynthetic and metabolic switch, we measured changes in RNA abundance upon glucose addition to *hxk1* mutants as compared to WT using quantitative reverse transcription-PCR (qRT-PCR). We selected photosynthetic and metabolic genes identified by RNA-Seq (Roth, Gallaher, et al., 2019), and investigated RNA abundance fold changes with and without glucose at two time points after adding glucose, 0.5 h and 12 h (Figures 7-9). Our previous study showed that CzHXK1 was rapidly up-regulated 30-fold within 30 min of glucose addition to the culture (Roth, Gallaher, et al., 2019). Similarly, in this experiment we observed a rapid increase in CzHXK1 mRNA in WT at 0.5 h, but not in the *hxk1* mutants (Figure 7). Photosynthesis-related nuclear genes including PSAH1, PSBO1, LHC16, and RBCS1, which were strongly down-regulated in WT at 12 h, were insensitive to glucose in hxk1 mutants (Figure 7). The mitochondrial respiration gene COX10 was slightly up-regulated in WT at 0.5h and slightly down-regulated in WT at 12 h and minimally affected in *hxk1* at both time points (Figure 7, Figure 8), suggesting that changes in respiration with glucose are

not transcriptionally regulated. Likewise, glucose-responsive genes in sugar and lipid metabolism including glyceraldehyde-3-phosphate dehydrogenase (GAP1), fatty acid desaturase (FAD2), stearoyl ACP desaturase (SAD1), a putative sugar transporter (SPM1), and the major lipid droplet protein gene MLDP1, were up-regulated in WT but not induced in *hxk1* mutants (Figure 7) at 12 h. Results for photosynthetic, sugar and lipid metabolism genes at 0.5 h were similar to 12 h except for SPM1 in *hxk1-1*, which was slightly up-regulated at 0.5h (Figure 8). Overall, these data provide strong evidence that CzHXK1 plays a critical role in repressing photosynthesis and up-regulating lipid metabolism with glucose.

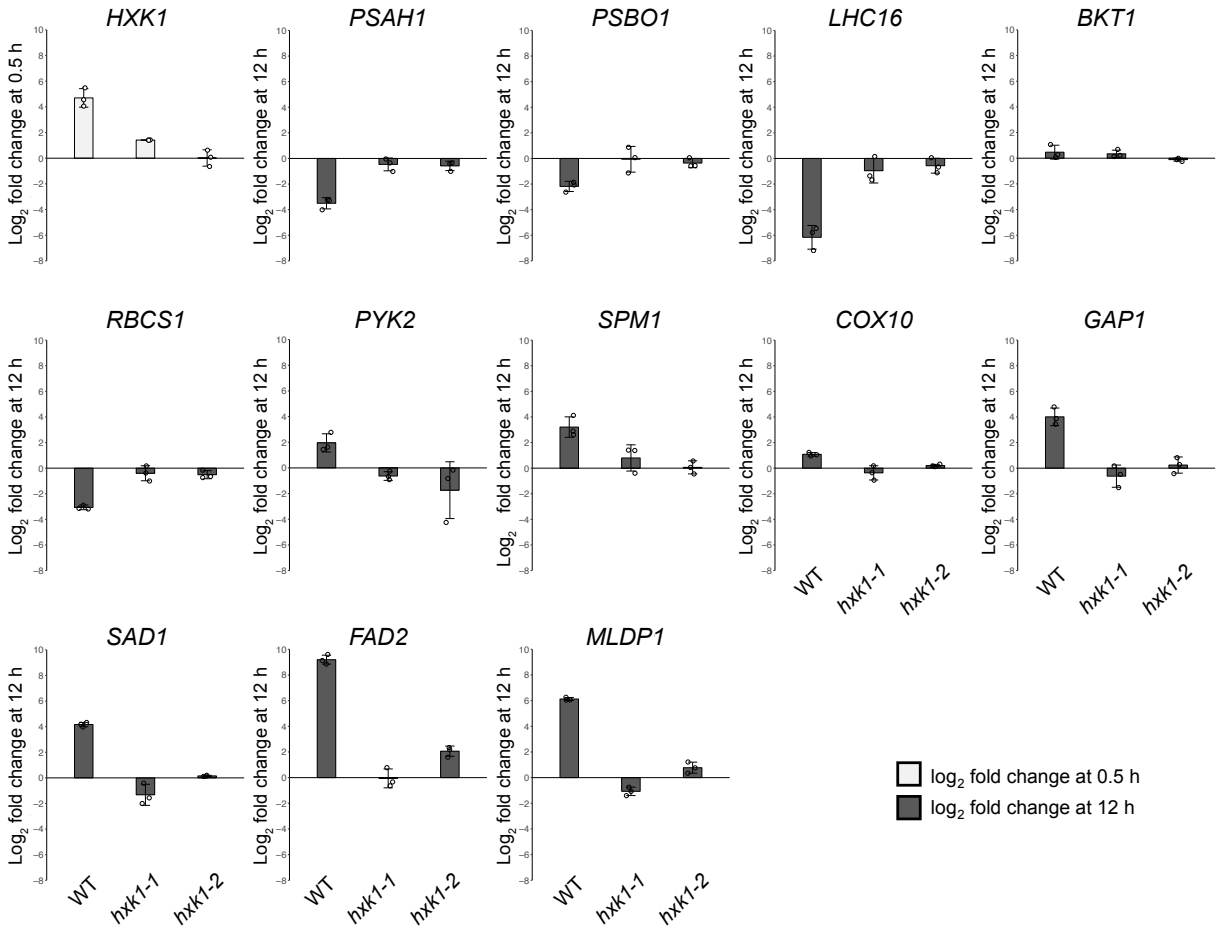


Figure 7. Mutants reveal CzHXK1-dependent transcriptional changes.

qRT-PCR analysis of mRNA levels of select photosynthetic and metabolic genes (identified by RNA-Seq; (Roth, Gallaher, et al., 2019).) in WT and hxk1 mutants. The log2-transformed fold change of mRNA level with glucose relative to a time-matched control without glucose at 0.5 h for CzHXK1 and at 12 h for all other genes. Additional time points and genes are shown in Figure 8. Data represent means ± SD (n = 3 biological replicates, individual data points shown). Raw ΔC_T data and biological replicates are shown in Figure 9.

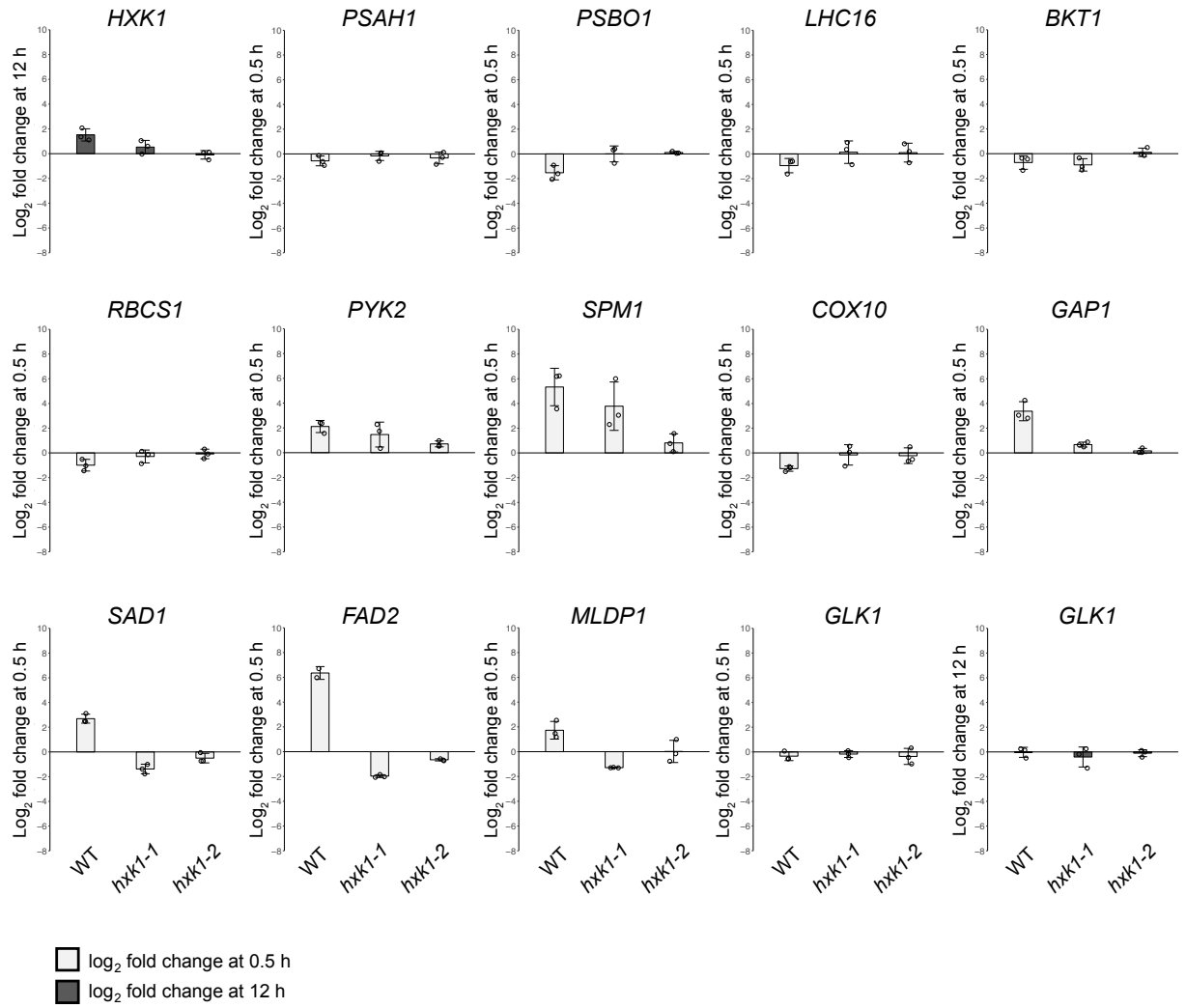


Figure 8. The transcriptional response of *hxk1* mutants to glucose is attenuated.

qRT-PCR analysis of mRNA levels of select photosynthetic and metabolic genes (identified by RNA-Seq6) in WT and *hxk1* mutants. The log2-transformed fold change of mRNA level with glucose relative to a time-matched control without glucose at 12 h for *HXK1* and *GLK1*, and 0.5 h for all other plots. Additional time points are shown in Figure 7. Data represent means ± SD (n = 3 biological replicates, individual data points shown). Raw ΔCT data and biological replicates are shown in Figure 9.

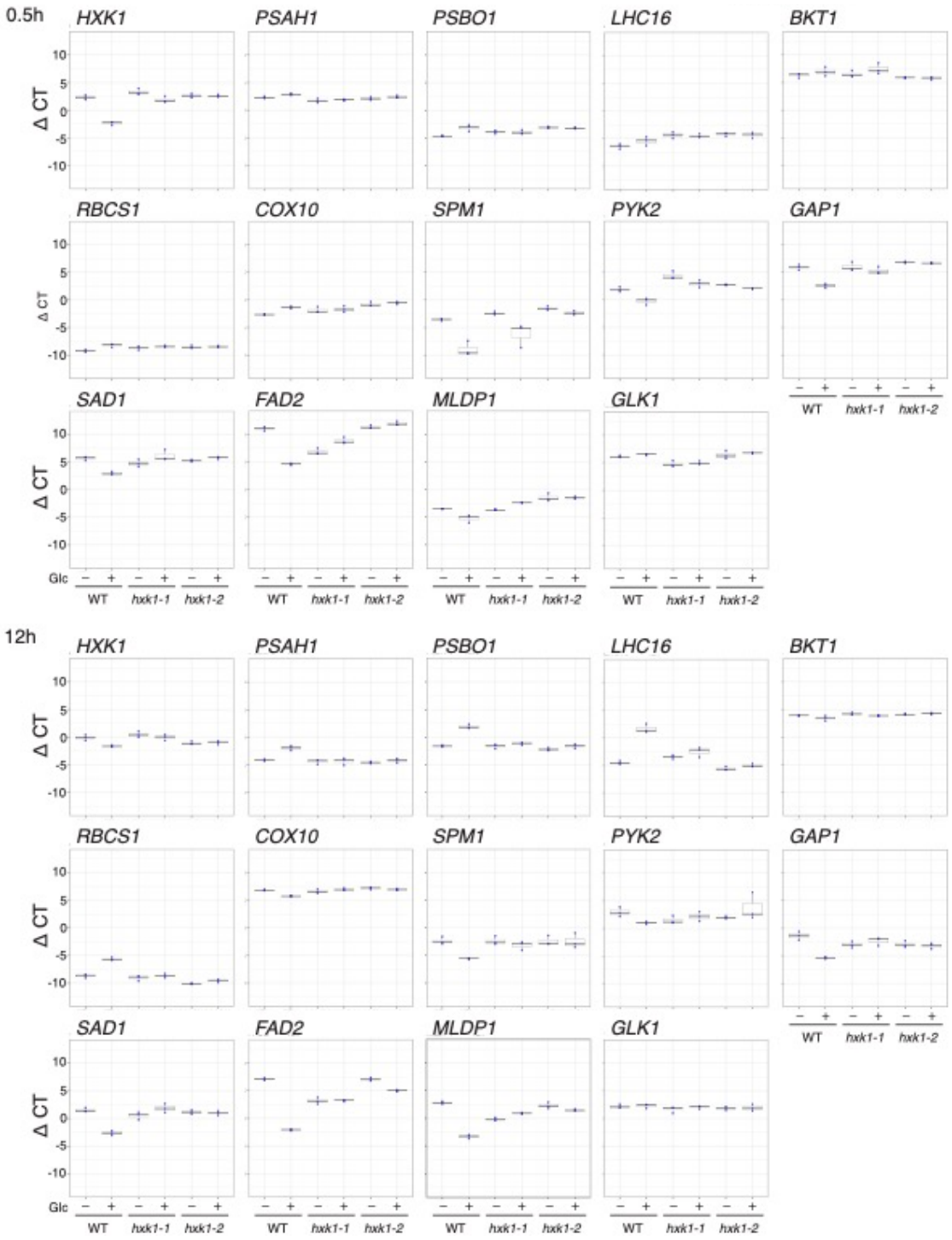


Figure 9. Raw qRT-PCR data. qRT-PCR raw ΔC_T (see Methods) of biological triplicates at 0.5 h and 12 h with (+) and without (–) glucose.

Box is centered on the mean; the thick line represents the median; the bottom of the box represents the 25th percentile; the top of the box represents the 75th percentile; upper and lower whiskers are ± 1.5 interquartile range; and dots represent biological triplicates.

We propose a pathway of CzHXK1-dependent signaling in *C. zofingiensis* based on our physiological, microscopic, and qRT PCR analyses of glucose responses in WT and hxk1 mutants (Figure 10). In the mutants, the lack of response of nuclear photosynthetic and lipid metabolism genes is consistent with the absence of changes in photosynthesis, astaxanthin and TAG accumulation, and cytoplasmic lipid droplet formation. These data suggest that HXK1 plays a direct or indirect role in repressing nuclear-encoded photosynthetic genes and activating astaxanthin and lipid accumulation. Evidence that hxk1 mutants consume and metabolize glucose includes the increased respiration and cell volume with glucose and the growth in the dark on glucose. Therefore, these data suggest that respiration, cellular division, and cell volume growth are regulated independently from CzHXK1.

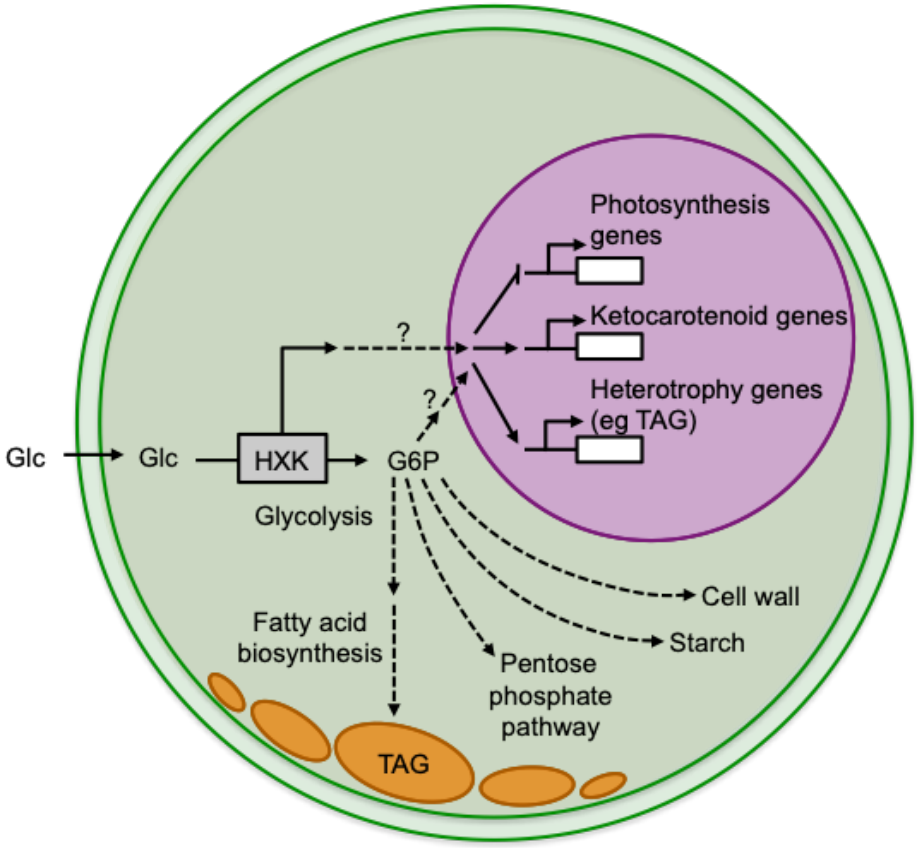


Figure 10. Model of hexokinase-dependent glucose pathways in C. zofingiensis.

Hexokinase (CzHXK) catalyzes the phosphorylation of glucose (Glc) to glucose-6-phosphate (G6P) in the first step of glycolysis. G6P can be used in a variety of downstream processes, including glycolysis, starch and cell wall biosynthesis, and the pentose phosphate pathway. There are two possible mechanisms for CzHXK1-dependent glucose responses: (1) glucose-bound CzHXK1 initiates a signaling cascade with unknown intermediaries to regulate gene expression and (2) metabolic intermediates or derivatives of the glycolytic pathway are required to induce the glucose phenotype. These pathways may also work in tandem for glucose signaling and glucose responses. CzHXK-derived G6P feeds into glycolysis and fatty acid biosynthesis, resulting in the accumulation of TAGs and cytoplasmic lipid droplets.

Discussion

Sugars induce a wide range of biological changes in plants and algae that manifest at all levels of cellular activity from transcription and translation to protein stability and activity. The sugar-signaling transcriptional network impacts thousands of genes, including those involved in cell cycle, DNA and protein synthesis, amino acid metabolism, nucleotide synthesis, cell wall synthesis, glycolysis, TCA cycle, electron transport chain, carotenoid biosynthesis, fatty acid biosynthesis, protein degradation, lipid degradation, amino acid degradation, gluconeogenesis, carbohydrate metabolism, and photosynthesis (Baena-González et al., 2007; L. Li & Sheen, 2016; Yunhai Li et al., 2006; Osuna et al., 2007; Roth, Gallaher, et al., 2019; Xiong et al., 2013). In plants, there are three glucose-modulated master regulators, the glucose sensor HXK1, the energy sensor kinases KIN10/KIN11 which are inactivated by glucose, and the glucose-activated target of rapamycin (TOR) kinase (Sheen, 2014). Many photosynthetic organisms repress photosynthesis, decrease chlorophyll, reduce Calvin-Benson cycle enzymes, and/or accumulate starch and/or lipids in response to the preferred and most studied sugar, glucose (Ahuatzi et al., 2004; Matsuka & Miyachi, 1974; Filip Rolland et al., 2006; Roth, Gallaher, et al., 2019; Sheen, 2014; Shihira-ishikawa & Hase, 1964a; Shihira & Krauss, 1965). Although plants and algae share many responses to glucose, there are some key differences. The most prominent difference relates to the multicellularity of plants. Sugar produced in photosynthetic tissues (source) is transported to other tissues (sink) to maintain nutrient and energy homeostasis at an organismal level (Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Filip Rolland et al., 2006). This tissue heterogeneity makes sugar signaling complex, because plants must perceive, respond, and signal changes in different tissues, cell compartments, and developmental stages (Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Filip Rolland et al., 2006). Multiple sugar receptors, regulators, and transduction pathways add to the complexity, leaving many questions about sugar sensing and signaling unanswered. Many algae, including *C. zofingiensis*, are unicellular, reducing the level of complexities and potentially offering new insights into sugar sensing and signaling.

In *C. zofingiensis*, glucose induces reversible changes in photosynthesis, the photosynthetic apparatus, cellular ultrastructure, lipids including thylakoid lipids and TAG, starch, astaxanthin, and gene expression pathways relating to photosynthesis, keto-carotenoid biosynthesis, metabolism, and fatty acid biosynthesis (Roth, Gallaher, et al., 2019). In this study, we used a forward genetics selection based on resistance to a glucose analog (2-DOG) to identify mutants that are insensitive to several glucose responses. In contrast to WT, the mutants did not (a) shut off photosynthesis, (b) accumulate high amounts of astaxanthin, TAG, or cytoplasmic lipid bodies, or (c) decrease thylakoid lipids upon addition of glucose. Photosynthesis-, metabolism-, and lipid-related genes that were differentially regulated in WT with glucose were unresponsive in glucose-treated mutants. Similar to WT, the mutants did increase cell volume and respiration, providing evidence that they did show some glucose-induced responses.

Whole-genome sequencing of the mutants revealed that CzHXK1 is necessary for the photosynthetic and metabolic switch during glucose-induced trophic transitions in *C. zofingiensis*. Intriguingly, all mutants that survived selection on 2-DOG contained a disruptive mutation in the single-copy *CzHXK1* gene at one of three locations in the gene. The fact that all the recovered mutations were in *CzHXK1* could be due to the lethality of mutations in other glucose signaling genes and/or genetic redundancy. For example, the lack of mutations in glucose transporter genes is likely due to redundancy, because *C. zofingiensis* has multiple annotated glucose transporter genes.

HXK is an evolutionarily conserved key enzyme in carbon metabolism. In the first step of the glycolytic pathway, HXK converts glucose to G6P for use in glycolysis, starch, and cell wall biosynthesis and the pentose phosphate pathway (Claeyssen & Rivoal, 2007). In addition to changes in cell size and respiration, the hxk1 mutants grew in the dark with glucose, providing evidence that they are still able to metabolize glucose. *C. zofingiensis* has glucokinase (Roth et al., 2017), which a higher affinity for glucose than CzHXK1, but has a slightly lower Vmax (Pasquini & Blaby, unpublished correspondence). In contrast, plants lack glucokinase, however they have multiple copies of HXK.

In plants, the master regulator HXK functions as both a conserved glucose sensor and glucose-metabolizing enzyme (Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Filip Rolland et al., 2006; Sheen, 2014). In this study, the rapid induction of *CzHXK1* RNA abundance by glucose is consistent with the established roles of HXK in glucose metabolism, and possibly also in signaling. One challenge in understanding sugar signaling is dissociating the metabolic and signaling aspects. We present a model in which CzHXK1 plays a central role in inhibiting the transcription of nuclear-encoded photosynthetic genes, while activating genes related to heterotrophy, lipids, and ketocarotenoids resulting in the accumulation of TAGs and astaxanthin (Figure 10). However, there are two possible mechanisms for CzHXK1-dependent changes: (1) a direct role in which glucose-bound CzHXK1 initiates a signaling cascade with unknown intermediaries that regulate nuclear gene expression, or (2) an indirect role through metabolic intermediates or derivatives of the glycolytic pathway, which are required to induce the glucose phenotype. It is possible that these two pathways work in tandem, allowing the cell to respond rapidly and robustly to changing metabolic requirement when glucose is added. This study identifies CzHXK1 as a critical molecular player required for the glucose-induced photosynthetic and metabolic switch in *C. zofingiensis*, but our data cannot distinguish whether CzHXK1 plays a signaling or metabolic role in inducing the responses. To further describe the role of HXK1 in both sensing and metabolizing glucose, it will be necessary to adapt successful efforts from yeast and plants using catalytically inactive forms of the enzyme to dissect the impact of these distinct roles in *C. zofingiensis*.

This study also suggests that the majority of accumulated TAG in glucose-grown cells begins as glucose and passes through CzHXK1 during glycolysis as it feeds into the lipid biosynthesis pathway via pyruvate. This model is consistent with our previous analysis of the glycolytic and lipid biosynthesis pathway (Roth, Gallaher, et al., 2019). Future work with catalytically inactive CzHXK1 will help to determine any regulatory role

of CzHXK1 in TAG or astaxanthin accumulation. Because the *C. zofingiensis* genome encodes a single CzHXK, it provides a simple system to study HXK function, localization, interacting partners, and additional downstream signaling nodes.

In contrast to *C. zofingiensis*, the well-studied plant and yeast systems have multiple copies of HXK. In Arabidopsis, which has six encoded *HXKs*, *AtHXK1* is thought to play dual roles in signaling and metabolism as well as the integration of intrinsic and extrinsic regulatory signals (Paulina Aguilera-Alvarado & Sanchez-Nieto, 2017; Filip Rolland et al., 2006; Sheen, 2014). However, there are still unknown details regarding HXK localization and its regulatory network in plants. Mechanistic details in yeast, which has two HXKs, are better understood, and it is often used as a eukaryotic model for glucose sensing. ScHXK2 shuttles in/out of the nucleus, where it stabilizes a repressor complex, preventing transcription of genes necessary to utilize carbon alternatives to glucose (Ahuatzi et al., 2004). Although the transcriptional system and interacting players are well characterized in yeast (Vega et al., 2016), how much of the glucose response is due to HXK signaling or catalysis is still debated (Lane et al., 2018). Assays demonstrate that the catalytic activity of HXKs, but not their signaling activity, is interchangeable between plants and yeast (J. Jang et al., 1997). Although instrumental in demonstrating the separate roles of plant HXKs, yeast is a poor model for glucose signaling in algae and plants, which tightly regulate glucose flux and gluconeogenesis in accordance with photosynthetic rates (Claeyssen & Rivoal, 2007). Recently, a single-copy KnHXK1 gene was identified in the multicellular charophyte green alga *Klebsormidium nitens*, and it was shown that its signaling role is dependent on its catalytic activity when expressed in *Arabidopsis* (Ulfstedt et al., 2018b). *C. zofingiensis* is positioned at an evolutionary branch basal to land plants such as Arabidopsis and can be exploited as a unicellular model system for plants as well as a conceptual bridge between yeast and plants. This study shows that HXK1 is critical for sugar regulation of photosynthesis and metabolism in algae. Future studies of glucose responses at the foundation of the green lineage will improve general understanding of the critical mechanisms underpinning sugar sensing and signaling in eukaryotes.

In Conclusion

In this chapter, we show that the glucose responsiveness described in Chapter 2 is hexokinase dependent. Chapter 2 describes what occurs during this reversible metabolic shift, and Chapter 3 describes by what means this shift might occur, namely via signaling that depends on hexokinase. Chapter 4 builds upon these findings by further investigating the reversibility of this metabolic shift. In short, when C. zofingiensis rebuilds photosynthetic capacity, what genes are critical to this process?

Chapter 4

ReGreenCuts: Identifying possible components of thylakoid biogenesis using comparative genomics and *Chromochloris zofingiensis*

Abstract

The process of thylakoid biogenesis is only partially understood. Here I use a combination of expression data, co-expression analysis, and a recently sequenced mutant collection to identify novel genes of interest in the process of thylakoid biogenesis. Previously generated and published RNA-seq data were constrained to generate a set of differentially expressed genes in *Chromochloris zofingiensis* that are upregulated during re-greening after removal of glucose from the growth medium (*Cz*DEGs). Further, *A. thaliana* genes known to be necessary during thylakoid biogenesis were used as seed genes to generate a list of co-expressed genes using extant data (*At*CEGs). Finally, the *Cz*DEGs were compared to a list of genes that was generated from an unpublished mutant library known as the *Chlamydomonas reinhardtii* acetate-requiring collection (CrARGs). Using a recently developed software suite, OrthoLang, these gene lists were compared using three sequence alignment algorithms and a variety of search approaches. Using *E. coli, S. cerevisiae, and Melainabacteria* as non-photosynthetic outgroups, genes that were not unique to photosynthetic organisms were removed. The resultant lists of genes, called ReGreenCuts, are presented here. As expected, many genes encoding light-harvesting complexes, photosystem subunits, and enzymes of the chlorophyll biosynthesis pathway are well represented among these cuts. A number of additional genes of interest, which have not previously been shown to play a role in thylakoid biogenesis, are identified and discussed. This work presents a systematic approach to integrating three disparate genetics approaches, comparative genomics, expression analysis, and mutant libraries and identifies novel genes likely to be of importance in the process of thylakoid biogenesis.

Introduction

During growth and development, organisms that perform oxygenic photosynthesis must first build the capacity to do so. The photosynthetic apparatus requires tight coordination between distinct cellular compartments in order to assemble photosystems which require lipids, pigments, cofactors, and proteins (Adam et al., 2011; Bastien et al., 2016; Mechela et al., 2019; Rast et al., 2015b). In eukaryotic organisms, nuclear encoded proteins must be translated, transcribed, and transported through multiple membranes, and proteins must be processed before they are fully functional. Lipids that originate in the chloroplast are exported, modified in the endoplasmic reticulum, and reimported into the chloroplast (Figure 1). Importantly, the excitation of chlorophyll must be confined until either quenching capacity has been built or the downstream electron transport chain is functional. Excited triplet state chlorophyll can produce singlet oxygen, leading to oxidative damage of membranes and proteins (Cazzaniga et al., 2013).

The process of thylakoid biogenesis is multi-faceted and difficult to study. The cellular coordination required makes identifying individual components challenging. Mutations in one part of the process, galactolipid biosynthesis for example, are likely to negatively impact other aspects, such as the accumulation of functional light-harvesting complexes (LHCs, Kobayashi et al. 2007). Additionally, the initial stages of thylakoid biogenesis occur within minutes. For example, in dark-acclimated etiolated protoplasts, PSI activity is detectable within 20 min of illumination (Filippis et al., 1980). This requires a tightly controlled, time-resolved, experimental system that is amenable to observations at a sub-micron scale.

It should not then be surprising that the fundamental process of thylakoid biogenesis remains only partially understood. This partial understanding is not due to a lack of interest among the scientific community. Much effort has gone into generating appropriate experimental systems to study this process, which generally exist of three types. The first uses dark-grown, etiolated seedlings exposed to light, the second depends on developmental gradients, and the third utilizes mutant strains (Asakura et al., 2008; Kobayashi et al., 2013; Kowalewska et al., 2016; Gügel & Soll, 2017). These strategies have provided enough data to generate a partial model in which overlapping processes combine to form nascent thylakoids (Rast et al., 2015b). In the case of etiolated seedlings, an excellent example of the strengths and weaknesses of these approaches was recently presented. A cryo-EM investigation of etiolated runner bean chloroplasts during development describes clear stages of thylakoid development (Kowalewska et al., 2016). However, this approach requires etiolation, which is not common to all photosynthetic organisms, and while common among angiosperms, is not a necessary stage of thylakoid biogenesis.

An excellent example of the use of developmental gradients in *A. thaliana* shows the range of thylakoid morphologies over the course of development (Gügel & Soll, 2017). This approach divides the leaf into mature and immature tissues and examines samples from each section. Unfortunately, this approach has little explanatory power over the

initial stages of biogenesis. Additionally, even when done well, developmental gradients segment a continuous gradient into discrete sections, imposing the researcher's interpretation upon the study at the outset.

Mutant analyses have been highly informative research approaches. Much of our understanding of thylakoid biogenesis and myriad other plant processes are based upon this approach (J. Huang et al., 2006; Kawata & Cheung, 1990; Meinke, 2020; Page & Grossniklaus, 2002) While informative, mutant analyses should be treated with caution as such a core biological process as thylakoid biogenesis may be prone to confounding pleiotropic effects. For example, the Δ*Lhcb2* mutant in *A. thaliana* has a 60% decrease in Lhcb1 and Lhcb2 which, non-intuitively, leads to lower levels of cytochrome b_6f and ATP synthase respective to WT, allowing the plant to continue to assimilate near wild-type levels of CO_2 (Bielczynski et al., 2020). This hypothesis remains unsettled, as previous work on this mutant suggests that other antennae proteins are increased allowing for the compensation of photosynthesis (Ruban et al., 2003).

As discussed in previous reviews on the topic of thylakoid biogenesis, there are generally three models that attempt to explain how new thylakoid membranes form (Adam et al., 2011; Bastien et al., 2016; Mechela et al., 2019; Rast et al., 2015b): (1) protein-mediated transport of new proteins, lipids, and cofactors, (2) vesicle-mediated transport of new proteins, lipids, and cofactors, and (3) a continuous membrane, or "bridge", between the inner chloroplast envelope membrane and nascent thylakoids. Evidence exists for each of these models, and it is likely that these processes work in tandem to form functional thylakoids *de novo*. Each model will be introduced in turn.

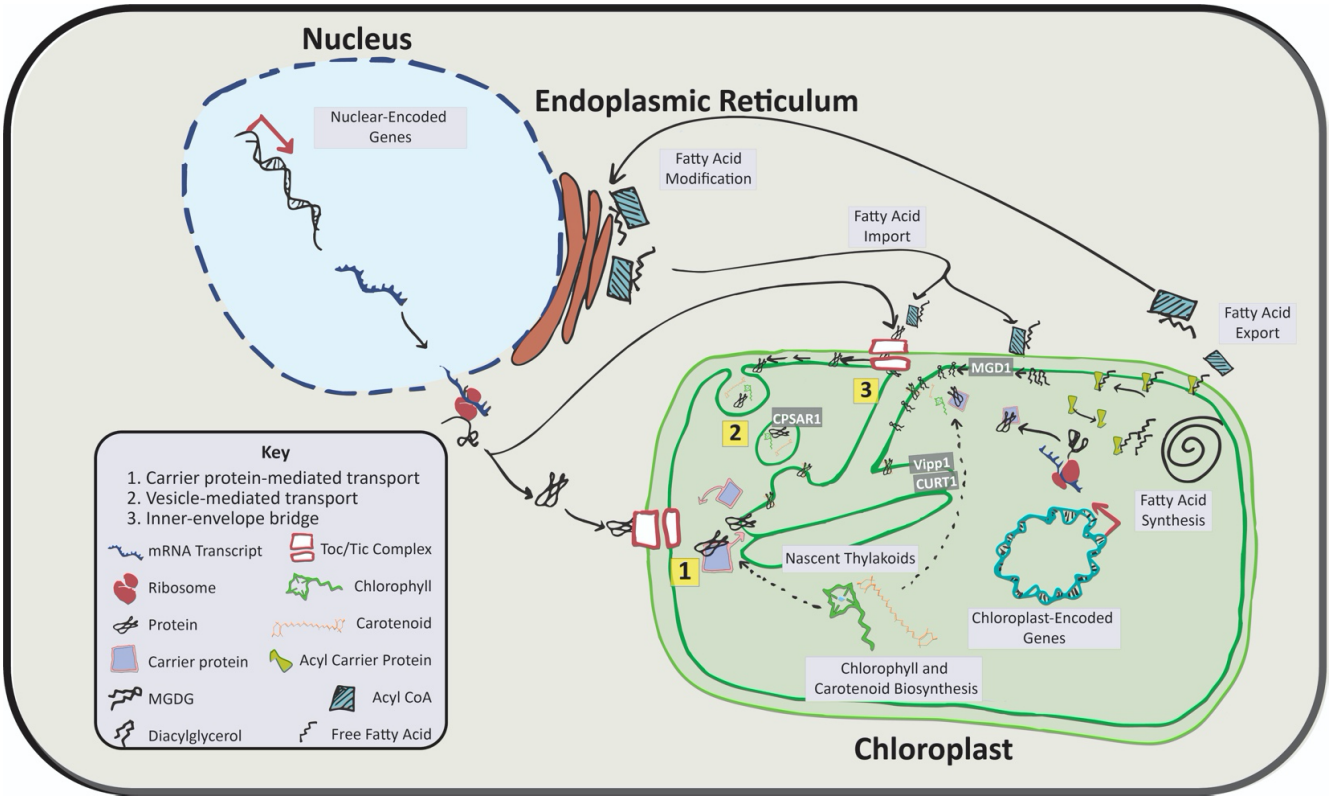


Figure 1. An overview of thylakoid biogenesis. Three working models explain this process. 1. Carrier protein-mediated transport, where nuclear-encoded thylakoid proteins are chaperoned to the thylakoid by protein carriers. 2. Vesicle budding from the inner-envelope membrane shuttles pigment, protein, and lipid cargo to the nascent thylakoids. 3. An inner-envelope bridge creates a continuous inner-membrane space with nascent thylakoids, and MGD1 is a critical lipid biosynthetic enzyme. VIPP1 and CURT1 are associated with thylakoid biogenesis and have been localized to highly curved thylakoid membranes while CPSAR1 has a role in vesicle-mediated transport during thylakoid biogenesis (Figure 1 based on the following references: Kobayashi et al. 2007; Kroll et al. 2001; Armbruster et al. 2013; Garcia et al. 2010; Mechela, Schwenkert, and Soll 2019; Bastien et al. 2016; Rast, Heinz, and Nickelsen 2015; Adam et al. 2011)

Protein-mediated transport

In the first model, nuclear-encoded proteins, cofactors, and are brought to nascent thylakoids by carrier proteins. Nucleus-encoded proteins, such as light-harvesting complex (LHC) antenna proteins, enter the chloroplast via the TOC/TIC complex (Froehlich & Keegstra, 2011). From there, they are processed and targeted to the thylakoid membrane by cpSRP43 and cpSRP54 via the thylakoid translocation channel, cpSecY (Lee et al., 2017).

CpSRP54 is also necessary to target the chloroplast-encoded, core PSII protein, D1 (Zhang & Aro, 2002). It is possible that this particular mechanism of D1 targeting is most active during the D1 repair cycle and not vital during thylakoid biogenesis, as it has

been shown that *Arabidopsis thaliana* (*A. thaliana*) mutants deficient in cpSRP54 continue to be able to undergo thylakoid biogenesis (Amin et al., 1999).

Other chloroplast-encoded proteins must also be integrated into nascent thylakoids. In some cases, this likely occurs by membrane-localized ribosomes translating new polypeptides directly into the thylakoid membrane (Hristou et al., 2019). Recent cryo-electron tomography data demonstrate that a subset of ribosomes are associated with biogenic regions of cyanobacterial thylakoid membranes, supporting the idea of direct translation into new thylakoid membranes (Rast et al., 2019). How and where these proteins come into contact with pigments and carotenoids to become functional photosystems continues to be an active area of inquiry (Bučinská et al., 2018; Komenda et al., 2016; Proctor et al., 2018; Zoschke et al., 2017).

Vesicle-mediated transport

It has been suggested that vesicle-mediated transport plays a role in thylakoid maintenance and biogenesis. Chloroplast vesicle accumulation was observed when various plants (pea, soybean, spinach, and tobacco) were subjected to 30 or 60 min of cold stress (Morré et al., 1991). Under these conditions, vesicles accumulate, possibly due to hindered vesicle fusion at cold temperatures. In contrast, *Arabidopsis* mutants defective in VIPP1 do not accumulate vesicles under cold conditions, suggesting a role for this conserved protein in vesicle formation (Kroll et al., 2001). Similar biogenic arrest has been observed in the embryo-lethal *Arabidopsis cpsar1* mutants in which plastids are undeveloped. However, in this case, vesicles continue to accumulate (Garcia et al., 2010). These authors report that CPSAR1 protein was found in cold-induced vesicles, suggesting a connection between thylakoid development, vesiculation, and CPSAR1. Recently, CPSFL1, an *A. thaliana* gene with a seedling lethal mutant phenotype, has been shown to be necessary for the appearance of vesicles in chloroplasts and chloroplast development (Hertle et al., 2020). Similarly, in *C. reinhardtii, cpsfl1* mutant strains are unable to survive photoautotrophically and have reduced numbers and volumes of plastoglobules, suggesting that the function of this protein may be conserved from *C. reinhardtii* to *A. thaliana* (García-Cerdán et al., 2020).

An examination of differentiating plastids in the shoot apical meristem of *Arabidopsis* found budding vesicle-like structures that appear to be trafficking between the inner envelope membrane of the chloroplast and the early thylakoid membrane (Charuvi et al., 2012). In *Arabidopsis mgd1* mutants, thylakoid biogenesis is arrested, however invaginations of the membrane can be observed under phosphate stress (Kobayashi et al., 2013). Taken together, the specific role that vesicle transport plays in thylakoid biogenesis may not yet be clear, but it appears to be a critical aspect of this process.

A membrane bridge

Membrane invaginations that connect the inner chloroplast envelope with nascent thylakoid membranes make intuitive sense in thylakoid biogenesis, and this concept is common in the published literature (Bastien et al., 2016; Liberton et al., 2006; Mechela et al., 2019; Rast et al., 2015b). Biosynthesis of the major lipid component of thylakoids, MGDG, occurs at the inner envelope (Awai et al., 2001). DGDG, the second most abundant thylakoid lipid is synthesized on the periplasmic region of the outer envelope membrane (Froehlich et al., 2001). Lipophilic molecules such as carotenoids and quinones could be transported in a lipid matrix towards nascent thylakoids via a membrane bridge.

However, direct evidence of a bridge-like structure is not as abundant as protein-mediated or vesicle-mediated transport of thylakoid precursors. In the case of etiolated seedlings, thylakoids appear to develop inside the proplastid and reach the inner envelope during later stages of development, suggesting that a bridge is unnecessary under these conditions (Kowalewska et al., 2016). In an exhaustive review of glycerolipid trafficking, both protein-mediated and vesicle-mediated lipid transport are discussed at length, however no evidence is presented to suggest such a bridge between the inner envelope and thylakoid membranes (Jouhet et al., 2007).

In a recent report on the cyanobacterium *Synechocystis* PCC 6803, direct observations of a convergence zone, as observed by *in-situ* cryo-electron tomography, suggest that, at least in prokaryotes, such a lipid bridge exists (Rast et al., 2019). However, this conclusion contrasts with similar work in this organism (Liberton et al., 2006). Additionally, the *Arabidopsis* protein FZL, which is in the dynamin family of proteins that mediate membrane fusion, has been shown to localize to both the inner envelope and the thylakoid membrane, providing additional evidence for the existence of a membrane bridge in plants (Gao et al., 2006). In *C. reinhardtii,* CRISPR-generated mutants in the Fzl homolog are light sensitive and display large, plastid associated vesicles, suggesting Fzl plays a role in membrane fusion in *C. reinhardtii* as well (Findinier et al., 2019).However, additional experiments will be necessary to support this model as vesicular transport may also explain these observations.

Chromochloris zofingiensis as a model system to study thylakoid biogenesis

The physiological response of *C. zofingiensis* to glucose makes it an attractive model system for the investigation of thylakoid biogenesis. Its ability to metabolize glucose and subsequent degradation of the photosynthetic apparatus in these conditions provides an experimental system in which a photosynthetic organism can be grown in a thylakoid-free state. Removal of glucose, and with it the repression of photosynthesis, allows *C. zofingiensis* to build thylakoid membranes anew.

In 1984, an inducible system of anoxygenic photosynthetic membrane biogenesis was presented to the scientific community (Chory et al., 1984). This system allowed for a temporally resolved, structural description of the process of membrane biogenesis in the anoxygenic phototroph, *Rhodobacter sphaeroides.* After induction of photosynthesis

in this bacterium, cellular growth is arrested, and the cells shift quickly to structurally reorganize, resulting in the establishment of photosynthesis within 15 h. We propose that *C. zofingiensis* can serve as an appropriate analog to study organisms performing oxygenic photosynthesis. As demonstrated in Chapter 1, *C. zofingiensis* can be grown in glucose conditions that repress photosynthesis and result in degradation of thylakoid membranes. This repression can be released by removing glucose, allowing the molecular components of photosynthesis to regenerate (Roth, Gallaher, et al. 2019, Figure 1). This provides an environmentally inducible system of thylakoid biogenesis in a wild-type organism that performs oxygenic photosynthesis. Additionally, the Chlorophyceae *C. zofingiensis*, a eukaryote, shares many cellular features common to land plants, but its single-celled morphological simplicity makes it suitable to investigations that may be confounded in a multi-cellular organism. *C. zofingiensis* is an emerging model to investigate thylakoid biogenesis that can provide new insight into this fundamental biological process that nearly all life depends upon.

Identifying thylakoid biogenesis genes using comparative genomics

Comparative genomics, simply defined, is utilizing genomic data from one organism to make inferences about another. As the availability of genomic data increases year upon year, comparative genomics offers a methodology for leveraging that which is known to understand that which is unknown. One of the first examples of this method in photosynthetic organisms was termed the GreenCut, which identified genes and subsets of genes that were unique to plants, and even to plastids (Merchant et al., 2007). This straightforward yet powerful approach looks at sets of orthologous genes that are shared among organisms with certain characteristics, while excluding genes that are shared with distinct organisms that lack these characteristics. These types of comparisons can be made for any organism for which sequencing data is available.

Given the infancy of *C. zofingiensis* as a model system, few publications have utilized the genomic and transcriptomic datasets that have been published (Roth, Gallaher, et al., 2019; Roth, Westcott, et al., 2019). To inform our experimental work, it is necessary to draw on the experience and lessons of more established model systems such as *Chlamydomonas reinhardtii* and *Arabidopsis thaliana*. To this end, I have used comparative genomics to identify potential genes of interest related to thylakoid biogenesis in *C. zofingiensis.* I have taken the approach of comparative genomics and adapted it to identify subsets of genes not simply based on their presence or absence in an organism, but on their expression patterns. This approach marries co-expression analysis with comparative genomics to identify high-priority genes of interest among a subset of co-expressed genes.

While *C. zofingiensis* shares many core features with other photosynthetic organisms, it is important to keep in mind the diversity of photosynthetic organisms when drawing comparisons about thylakoid biogenesis. The prokaryotic cyanobacteria

have a different set of morphological and environmental constraints than does the redwood. The architecture of thylakoid membranes can be relatively simple, as in *Synechococcus* PCC 7942 where thylakoids are arrayed around the periphery of the cell. Alternatively, this architecture can be quite complex as in *Arabidopsis thaliana* where thylakoids have distinct features such as grana stacks, grana margins, and connecting stroma lamellae. Additionally, the evolution of primary and secondary plastids necessitates alternative regulatory and biophysical constraints. The specific biological pathways and environmental responses in one organism are not necessarily conserved in the next. In light of this diversity, genes that are highly conserved between algae and plants imply that they play a critical role in photosynthetic metabolism. Comparative genomics provides the framework to seek these conserved genes, and a recently developed bioinformatic tool dubbed OrthoLang provides the method.

OrthoLang has specifically been developed to facilitate comparative genomics (Johnson 2019; https://github.com/jefdaj/ortholang). OrthoLang allows sets of genes to be systematically compared while ensuring that the results are repeatable and robust. Additionally, a diverse set of sequence alignment sequence alignment tools is integrated into the program. OrthoLang allows common input gene sets to be compared using different search algorithms with outputs that share common metrics. Iterating between varied sequence analysis tools provides a robustness to results, which helps to minimize any skew of the results that might be inherent in a single tool. This approach identifies genes in *C. zofingiensis* that are of importance in the process of thylakoid biogenesis by making comparisons to constrained sets of genes from *Arabidopsis* and *Chlamydomonas,* while excluding genes that are common to non-photosynthetic organisms (Figure 2). In *C. zofingiensis,* thylakoid biogenesis is a phenomenon of regreening after glucose treatment. Sequence-based orthology cuts that identify genes of interest in this process are dubbed ReGreenCut genes. This work presents multiple sets of genes, each its own ReGreenCut

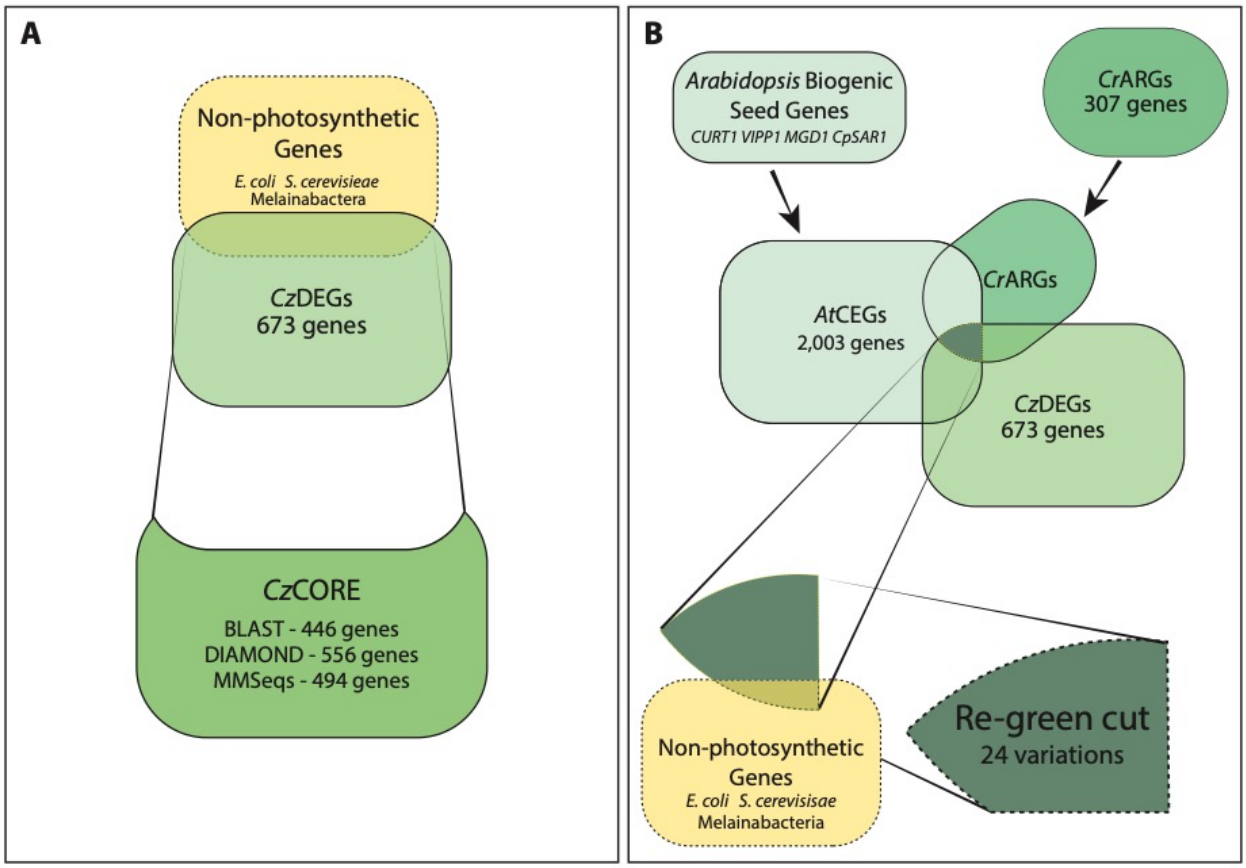


Figure 2. A schematic representation of the process of the set relationships used to generate *Cz*Core and RegreenCuts A The generation of *Cz*CORE. Differentially upregulated genes upon glucose removal at 1,3, and 6 hours comprise the 673 genes in *Cz*DEGs. Genes which share homology to non-photosynthetic genes with a sequence alignment score of 1 x 10-5 or lower were removed from this set. Results from BLAST, DIAMOND, and MMSeqs are listed. **B.** *Cz*DEGS is constrained according to the following set relationships. A gene is considered to be a RegreenCut gene if it is in *Cz*DEGs, AND has orthologs in A. thaliania co-expressed genes (*At*CEGs), AND has orthologs in C. rienhardtii acetate requiring genes (*Cr*ARGs), AND does not have orthologs in the non-photosynthetic gene sets. *At*CEGs was generated by using known thylakoid biogenesis genes (*CURT1*, *MGD1*, *VIPP1*, and *CPSAR1*) as seed genes for co-expression analysis. *Cr*ARGs are genes whose disruption requires supplementation with acetate for growth in *C. reinhardtii* (Wakao et al., manuscript in preparation). The definition of an ortholog is dependent upon search strategy, the specific algorithm (BLAST, DIAMOND BLAST, or MMseqs), and sequence-similarity thresholds. These parameters are varied and resultant gene sets are compared. Depending upon the stringency of the search, genes that are common among search strategies, or are present in any strategy are deemed to be of further interest.

METHODS

Data sources and preparation

The data for this approach was gathered from three sources. An initial set of candidate genes was generated using expression data from our previous work (Roth, Gallaher, et al., 2019). Briefly, glucose was added to photosynthetic cultures over 72 h. Photosynthesis rapidly declines during this process as described in Chapter 1. Cultures were centrifuged and re-suspended in a glucose-free medium. Samples were taken at 0, 0.5, 1, 3, 6, and 12 hours after glucose removal for RNA sequence analysis. This transcriptional dataset served as the broader set of genes from which to compare to other organisms. An additional set of genes was derived from publicly available expression data in *Arabidopsis thaliana* using published co-expression analysis data. Further, a set of genes identified in *Chlamydomonas reinhardtii* as critical for normal photosynthetic function was included (Wakao, Niyogi et al., unpublished). Each of these datasets will be detailed below.

C. zofingiensis expression data

The transcriptional data generated in our previous approach was constrained according to the following parameters (Roth, Gallaher, et al., 2019). Genes expressed at 30 min post-glucose removal were excluded. Internal concentrations of glucose may not be fully consumed within 30 min. Additionally, there may be residual glucose in the medium during the +glucose to -glucose transition.

As previously shown *C. zofingiensis* is able to partially recover photosynthetic activity 12 h post glucose (Roth, Gallaher, et al., 2019). Therefore, the 12-h time point was also excluded in order to remove genes that are likely to be involved in general photosynthetic growth, rather than the initial biogenesis of thylakoids. Genes that were down-regulated were also excluded, as this effort seeks to identify proteins involved in the biophysical and structural process of thylakoid biogenesis. While the absence of a protein may aid in restructuring a biological system, the difficulty in discriminating such genes from those whose functions are dedicated to heterotrophic growth is beyond the scope of this approach. Additionally, genes with very low FPKM values (below 2 FPKM), and those that did not vary in expression (Log_2 fold change less than 1.5), were also excluded from this analysis (Figure 3).

The resultant list of genes represents those that are upregulated during the transition from heterotrophy to photoautotrophy.

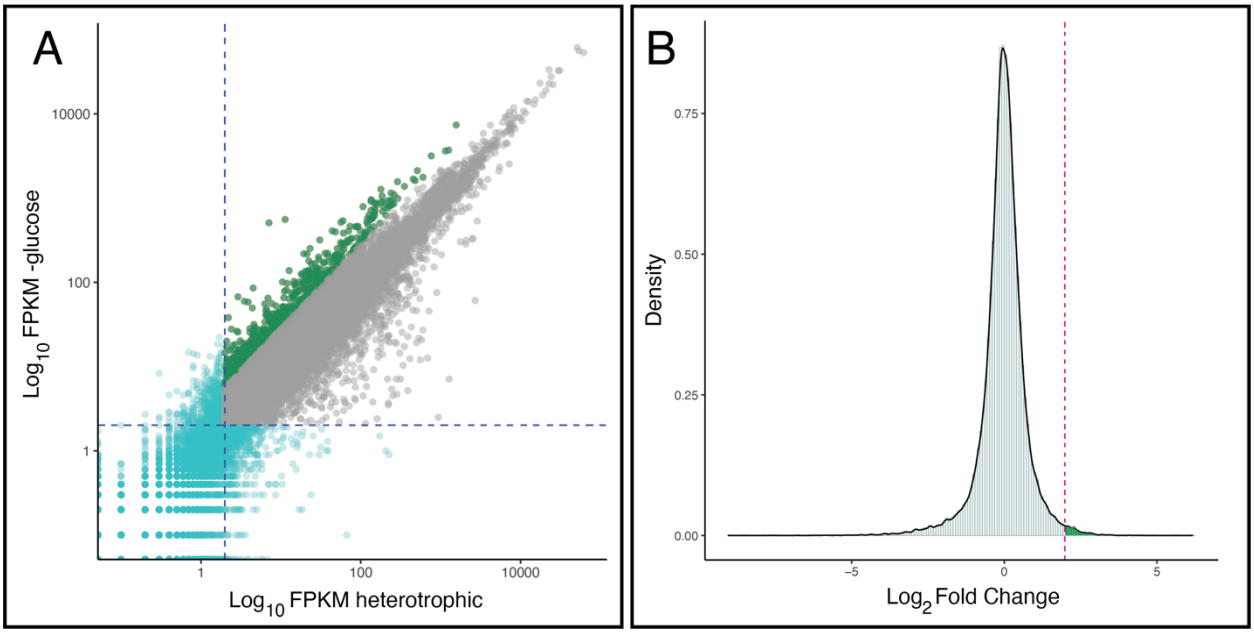


Figure 3. **A. Raw FPKM values of each gene over 1h, 3h, and 6h for treatment (+glucose/-glucose) and control (+glucose/+glucose, heterotrophic).** Genes with Log_2 fold change of 1.5 or greater (green) were included in this analysis. Genes with fewer than 2 FPKM (cyan) at all time points were excluded, as well as those with Log_2 fold change less than 1.5. **B.** Density plot of each expressed gene over 1h, 3h, and 6h time points. The 673 included genes are on the right side of the dotted line colored in green.

Of the 15,369 *C. zofingiensis* gene models used at the time of this expression analysis, 15,240 had a measurable level of expression at 1, 3, or 6 hours after glucose removal, in either treatment or control groups. 12,948 gene models had an expression level greater than 2 FPKM at any time point in either treatment or control conditions. 1,885 genes had a log_2 fold-change greater than, or equal to 1.5, but of those, only 673 genes were both expressed above the threshold of 2 FPKM and had a log_2 fold-change greater than or equal to 1.5 at any time point here considered. These 673 genes comprise the core group of genes used to identify thylakoid biogenesis genes. These genes are termed ***Cz*DEGs** for *C. zofingiensis* differentially expressed genes.

Arabidopsis thaliana co-expression data

A set of genes that are likely to be expressed during chloroplast development and thylakoid biogenesis was derived for *A. thaliana* by using a co-expression analysis. Four bait genes were used, *CURT1*, *MGD1*, *VIPP1,* and *CPSAR1*. There is experimental evidence describing the role of each of these genes during thylakoid biogenesis. For example, thylakoid architecture is highly altered in *curt1* mutants, and *Arabidopsis curt* mutant plants are unable develop grana stacks (Armbruster et al., 2013). *curT* mutants in cyanobacteria also have aberrant morphology and are unable to form typical convergence zones (Heinz et al., 2016). While *curt* mutants are able to perform photosynthesis despite altered morphology, mutants in the critical galactolipid synthesis gene, *MGD1*, result in severe dwarf and albino phenotypes, stemming from the

disrupted biosynthesis of the key thylakoid lipid, monogalactosyldiacylglycerol (MGDG, Kobayashi et al. 2007). Similarly impaired in photosynthesis are mutants in *VIPP1*, also known as *IM30*. *A. thaliana* plants with mutations in *VIPP1* are unable to develop a robust thylakoid network, with knockdowns showing a dose-dependent, thylakoid-defective phenotype (Kroll et al., 2001; Lingang Zhang et al., 2012). Finally, *CPSAR1*, a GTPase localized to the inner chloroplast envelope membrane, is likely to be involved in vesiculation during thylakoid biogenesis (Bang et al., 2009; Chigri et al., 2009; Garcia et al., 2010). These experimentally validated genes were used as bait for co-expression analysis.

Using the Expression Angler tool, I identified genes co-expressed with each bait gene (Austin et al., 2016). I used an r-value cutoff range of 0.8 – 1.0. As there are numerous expression data sets for *A. thaliana*, I included three expression data sets to ensure that a broad and robust gene set was derived. These included the AtGeneExpress Plus Extended Tissue Compendium, the Bioanalytic Resource Compendium, and the Seed Compendium (Bassel et al., 2008; Redman et al., 2004; Schmid et al., 2005; Toufighi et al., 2005). The lists from each bait gene were combined and deduplicated. The resultant list contained 2,003 unique *A. thaliana* gene models. Amino acid sequences were retrieved using the Arabidopsis Information Resource (TAIR, www.arabidopsis.org). For the purposes of this work, this set of genes will be referred to as *A. thaliana* coexpressed genes (***At*CEG**).

It should be noted that, although the *C. zofingiensis* genome contains homologs to the bait genes used to generate the *At*CEG gene set, it does not necessarily follow that these homologs are present in the *Cz*DEG gene set. In the experimental conditions used to generate these expression data, these genes are not differentially expressed in *C. zofingiensis*, and thus, did not pass through the initial constraints outlined above.

Critical genes required for photoautotrophic growth in *Chlamydomonas reinhardtii*.

I further expanded the orthogonal search for genes critical to thylakoid biogenesis using sequencing results from a *Chlamydomonas reinhardtii* mutant library. This strain collection was recently described as the Acetate-Requiring Collection, or ARC (Wakao, Niyogi et al., manuscript in preparation). This mutant collection contains strains that are severely growth limited in the absence of supplemental carbon in the form of acetate. In collaboration with the Joint Genome Institute, this mutant collection was subjected to low-coverage, short-read sequencing. The sequence data were analyzed and further curated to identify likely causative mutations for each strain among the collection. The resultant list contains 307 genes likely to be required for photoautotrophic growth. These genes will be referred to as the *C. reinhardtii* acetate-requiring genes (***Cr*ARG)**.

Non-photosynthetic organisms

To eliminate genes that were common to all groups, yet unlikely to be involved in thylakoid biogenesis, I compared resultant gene lists to the whole genomes of non-photosynthetic organisms. I chose to use *Escherichia coli* (strain 0157:H7) as a representative prokaryotic bacterium due to its thoroughly characterized genome. I included *Saccharomyces cerevisiae* (strain S288C) to represent a non-photosynthetic organism capable of shifting trophic modes and surviving on glucose in a hexokinase-dependent manner. Finally, I chose to include *Melainabacteria*, which are closely related to *Cyanobacteria*, yet lack the capacity for photosynthesis (Soo et al., 2017). This analysis included three *Melainabacteria* strains isolated from gut flora for which complete genomes have been published (Mel_A1, Mel_B2, and Mel_C2, Di Rienzi et al. 2013).

Approach

OrthoLang allows for the design and iterative modification of sequence-based search strategies which generate lists of genes meeting defined criteria. The program includes stable versions of many popular sequence search algorithms, meaning that algorithms are not altered or updated within an instance of OrthoLang, and thus results are reproducible over time. All OrthoLang scripts used in this work have been deposited in a github repository which can be found at https://github.com/daniel-westcott/dissertation.

In this instance, I used OrthoLang to compare three different search algorithms, BLAST, MMSeqs-search, and DIAMOND (Altschul et al., 1990; Buchfink et al., 2014; Hauser et al., 2016). I applied these algorithms in eight different ways. First, I generated a broad set of potential genes of interest by removing non-photosynthetic homologs. Second, I used one-way sequence alignment searches with a constant expected value (eVal) threshold. It should be noted that eVals are not necessarily comparable on a one-to-one basis from one algorithm to the next. However, in the absence of established guidelines for comparison, I chose to use the same eVal among all algorithms in this strategy. I increased the stringency of this approach by using five eVal thresholds, $1x10^{-10}$, $1x10^{-20}$, $1x10^{-50}$, $1x10^{-100}$, and $1x10^{-200}$, resulting in five unique gene sets. Next, I used a species-specific (SS) search strategy where eVals are set by considering the evolutionary relationship among organisms. Finally, I utilized a strategy based on reciprocal best hits. In total, this resulted in 8 gene sets per algorithm, or 24 primary gene sets in total. The overview of search strategies is presented in Table 1. These strategies will be described in turn. An overview of the different approaches is presented in Figure 4.

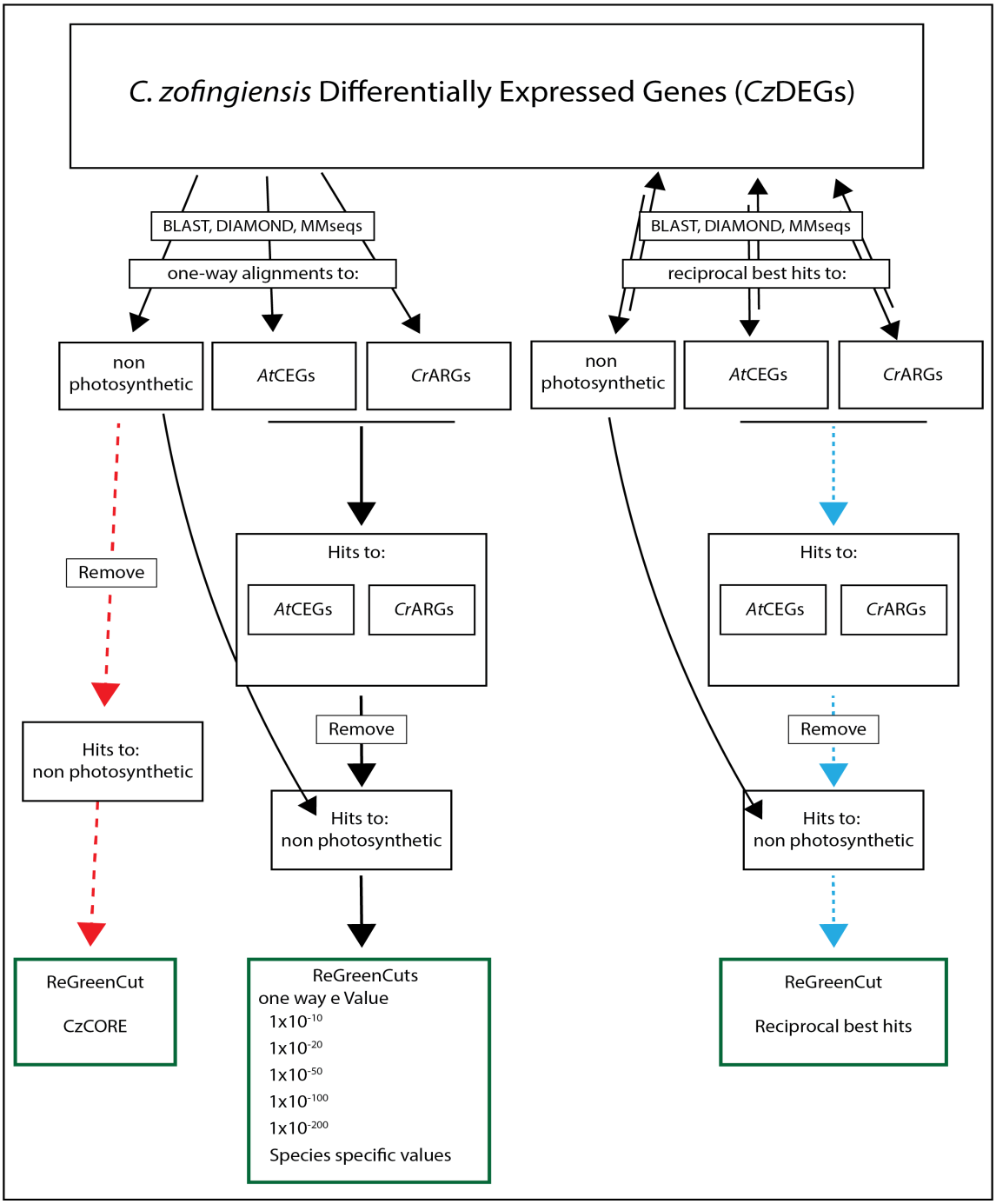


Figure 4. An overview of the strategies used to generate 24 unique ReGreenCuts. All cuts begin with *C. zofingiensis* differentially expressed genes during regreening (*Cz*DEGs). Using three search algorithms (BLAST, DIAMOND, and MMseqs), sequence alignments were generated by using *Cz*DEGs as queries against *A. thaliana* co-expressed genes (*At*CEGs), *C. reinhardtii* Acetate-Requiring Genes (*Cr*ARGs), and non-photosynthetic gene sets. In the most straightforward case, hits to non-photosynthetic organisms are removed, creating the *Cz*CORE ReGreenCut (Left dashed arrows). In the second case, subsets of *Cz*DEGs are generated by comparing to *Cz*DEGs to *At*CEGs and *Cr*ARGs. Non-photosynthetic genes are removed resulting in five additional ReGreenCuts depending on the eValue threshold (Black arrows, center). A similar workflow using a reciprocal best hits search approach results in the final ReGreenCut (blue dashed arrows, right).

Core RegreenCut (*Cz*CORE)

The most straightforward approach to identify genes involved in thylakoid biogenesis is to remove homologs to non-photosynthetic organisms from the *Cz*DEGs gene set. In order to do so, I used *Cz*DEGS as a query set against the non-photosynthetic organisms in one-way sequence alignment searches. A permissive eVal threshold of $1x10^{-5}$ was set in an effort to identify as many potential homologs as possible. Genes which passed this threshold in any of the three search algorithms were removed from *Cz*DEGs resulting in an exclusively photosynthetic gene set dubbed *Cz*CORE. This gene set removes evolutionarily conserved genes that may have diverged over time. I have chosen not to use *Cz*CORE as the basis for additional searches because I do not wish to exclude those outright, but I have included the results of this approach in Table 2 and highlight genes of interest in the discussion. This table is too cumbersome to include in-line with the text and has been appended to the end of this chapter.

One-way sequence alignment searches

This approach seeks homologs from various organisms using *Cz*DEGs as queries and either whole genomes, *At*CEG, or *Cr*ARG as targets. A consistent eVal was used for all searches in this approach, which was repeated at increasingly stringent eVals ($1x10^{-10}$, $1x10^{-20}$, $1x10^{-50}$, $1x10^{-100}$, and $1x10^{-200)}$. *Cz*DEGs were searched against *At*CEGs, *Cr*ARGs, *E. coli*, *S. cerevisiae* and three *Melainabacteria* strains, and a RegreenCut gene must have orthologs in *At*CEGs or *Cr*ARGs and must not have orthologs in the non-photosynthetic gene set (*E. coli*, *S. cerevisiae*, or Melainabacteria). Once homologs that meet the eVal threshold are identified from each target gene set, these genes are passed through the comparative pipeline as described in Figure 2.

Briefly, a set of genes common to the *Cz*DEGs, *At*CEGs, and *Cr*ARGs was generated. Further, a set of genes common to *Cz*DEGs and the non-photosynthetic organisms *E. coli*, *S. cerevisiae,* and *Melainabacteria* was generated. Non-photosynthetic homologs were removed from the *Cz*DEG set. An additional set of genes common to *Cz*DEGs and other photosynthetic organisms, *A. thaliana, A. protothecoides, and C. reinhardtii* was also generated. The photosynthetic genes are retained for comparison, but are retained in the final gene list.

Species-specific eValues.

This approach follows the same pipeline as described above for the one-way search with one exception. In an effort to generate a stricter non-photosynthetic gene set, and a more tightly related thylakoid biogenesis gene set, variable eVals were used. Potentially distant homologs among the non-photosynthetic set were retained by allowing permissive eVals during those searches. Genes that may have diverged over

time were discarded by choosing stricter eVals for closely related organisms (e.g. *C. zofingiensis* and *C. reinhardtii*).

These eVals were generated empirically using sequence alignment searches based on conserved proteins. Proteins associated with the large and small subunits of the ribosome were identified for each comparative species. These sequences were queried against the *C. zofingiensis* genome and scored. eVal thresholds were determined by using the second quartile of the distribution of eVals. In other terms, query genes that result in an eVal less than, or equal to the bottom 25% of the ribosomal protein eVal distribution were considered homologs for the purpose of this search strategy (Figure 5 overview.

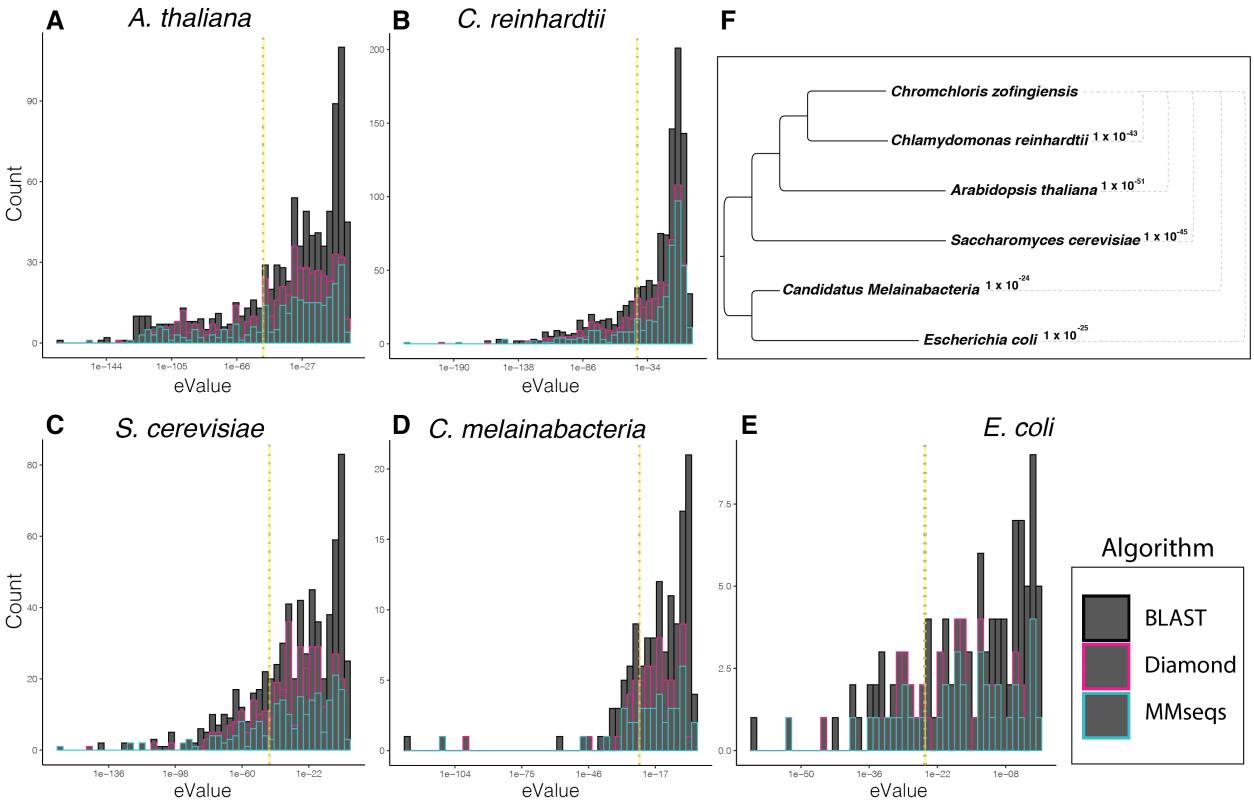


Figure 5. A-E: **Species-specific e-Value thresholds were derived by subjecting ribosome-associated proteins to BLAST, Diamond, and MMseqs sequence searches to the *C. zofingiensis* genome.** F: A phylogenetic overview of the species relationships including the resultant e-value thresholds used in the species-specific search strategies.

Reciprocal best hits

An alternative approach to one-way blast searches is a reciprocal best hits (RBH) approach. RBH analysis queried each gene in a genome against a genome target. The reciprocal search then used each gene that was previously among the genome target as queries against the previous genome. Results from both searches were compared, and the common, best hit in both searches was considered the RBH. This strategy

identifies conserved homologs of high sequence identity between organisms. Taken as a single strategy, it is potentially too restrictive to be useful. However, used in conjunction with more permissive sequence search strategies, it will help to identify any genes that are highly likely to be functional homologs. Any genes that are present in both the RBH set, and one-way searches are strong candidates for further investigation.

Algorithms

In addition to varying search strategies, OrthoLang enables the exchange of algorithms in an existing bioinformatic pathway. The common output of OrthoLang is a blast hits table (BLAST out format 6). Search algorithms that may not natively result in a common output format do so in OrthoLang, making results comparable between alternative algorithms. By varying search algorithms and stringencies, multiple gene lists were created, which were then compared to identify genes resistant to perturbation. The results of this variation on final lists can be seen in Table 1.

Results

Core RegreenCut (*Cz*CORE)

Combining the overlap between three algorithms (BLAST, DIAMOND, and MMseqs) when using one-way searches resulted in a shared set of 116 non-photosynthetic genes (Fig 5 – *Cz*CORE overview). This combined set of non-photosynthetic genes was removed from *Cz*DEGS, resulting in a set of genes that are likely to be specific to photosynthetic modes of growth. These genes are termed *Cz*CORE.

Table 1. RegreenCut final results. Eight search strategies per algorithm were used, resulting in 24 total RegreenCuts. The total number of genes that result from each search strategy, reciprocal best hits (RBH), Species specific eValues (SS), and static eValues are described for each algorithm. Unique represents the number of genes that appear in the results from any search algorithm. Intersection represents the number of genes present in results from all three algorithms.

	Cz*CORE*	*RBH*	*SS*	*e-10*	*e-20*	*e-50*	*e-100*	*e-200*
BLAST	466	41	55	95	82	55	34	11
DIAMOND	556	50	55	83	75	51	24	4
MMseqs	494	47	53	93	83	53	30	5
Unique genes	557	75	65	107	90	57	38	12
Intersection	465	24	33	70	69	50	21	4

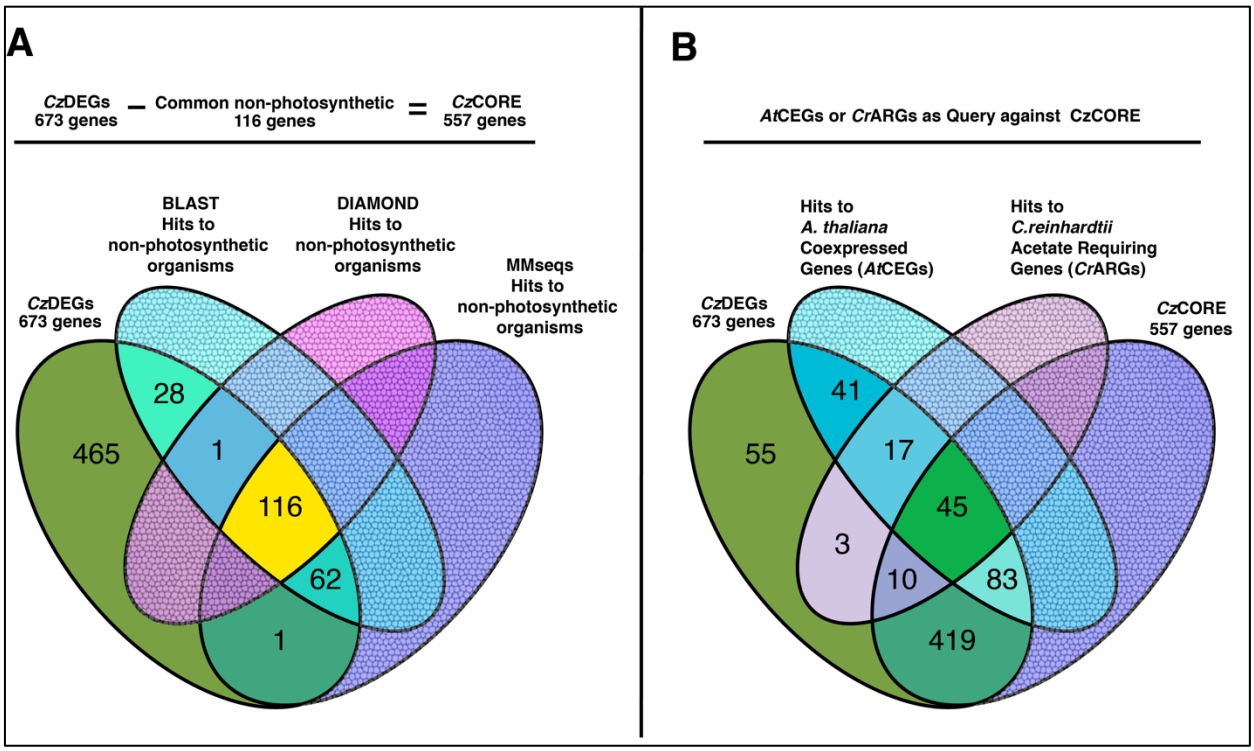


Figure 6. *Cz*CORE overview. A. Non-photosynthetic homologs were identified and removed from the greater *Cz*DEGS resulting in CzCORE (557 genes). Genes that were present in all three search algorithms with an eValue threshold of $1x10^{-5}$ were considered robust homologs (yellow). B. One-way search results from BLAST, DIAMOND, and MMseqs (eVal = $1x10^{-5}$) against *Cz*DEGs using either *At*CEGs or *Cr*ARGs as query.

The 557 remaining genes in *Cz*CORE represent the broadest set of genes that are both up-regulated in *C. zofingiensis* during thylakoid biogenesis and are also absent in the non-photosynthetic organisms queried. When subjecting amino acid sequences to domain prediction programs Interpro Scan 5, Gene Ontology, KEGG Ontology, EGGNOG, and HMMER, 421 could be assigned annotation based on sequence and domain prediction (Ashburner et al., 2000; Finn et al., 2011; Huerta-Cepas et al., 2019; Jones et al., 2014; Kanehisa & Sato, 2020; The Gene Ontology Consortium, 2019). A comprehensive table including all genes and their inclusion in gene sets is provided in Table 2. These genes encode photosystem subunits, light-harvesting complexes, and components of the chlorophyll biosynthesis pathway.

Of these, a permissive sequence alignment search using an eVal of $1x\ 10^{-5}$ finds 83 homologs exclusively shared among the *At*CEG gene set, 10 exclusively shared among the *CrARG* gene set, and 45 common to both *At*CEG and *CrARG* (Figure 6). The details of this gene set are in Table 2. Of these 45 genes, 27 are well-established genes either involved in chlorophyll biosynthesis, directly involved in photosynthesis (PSI, PSII, LHCs, or RUBISCO), or are otherwise known to be important to thylakoid biogenesis (Tic62, pTAC2, LUT5, MCG1). This leaves 21 genes with little or no experimental evidence describing their role in thylakoid biogenesis (Table 3). This list represents high-priority discovery targets whose role in thylakoid biogenesis has not yet been

described. Among these is Cz14g09130, a predicted TAG lipase that we have previously identified as a potential engineering target is included in this gene list (Roth, Gallaher, et al., 2019). This set of genes represents the most straightforward application of this pathway, as described in Figure 4 (left dashed arrows).

Table 3. Genes uncharacterized in the process of thylakoid biogenesis resulting from considering one-way BLAST searches of *At*CEGs and *CrARG* against *Cz*CORE.

Cz GeneID	*A.th* Homolog	*At* Gene Name	*C.re* Homolog	*C.re* Gene Name	GO terms
Cz01g05240	AT3G04650.1	N/A	Cre03.g207150.t1.1	N/A	hypothetical protein MNEG_13559
Cz01g22130	AT1G74850.1	PTAC2	Cre10.g437150.t1.2	PPR6	Pentatricopeptide repeat-containing
Cz01g30130	AT3G23080.2	N/A	Cre03.g174850.t1.2	CGL111	lipid-binding START
Cz01g31110	AT1G08860.1	BON3	Cre03.g170200.t1.1	GEM-4	calcium-dependent phospholipid-binding Copine family
Cz01g37110	AT1G54780.1	TLP18.3	Cre03.g182150.t1.2	TEF8	UPF0603 At1g54780, chloroplastic
Cz02g00170	AT2G45700.1	N/A	Cre09.g399907.t1.1	N/A	DNA cross-link repair 1A
Cz02g00290	AT5G53490.2	TL17	Cre09.g404000.t1.1	CPLD59	Thylakoid lumenal kDa , chloroplastic
Cz04g04250	AT5G20140.1	HBP5	Cre16.g666550.t1.1	N/A	SOUL heme-binding
Cz04g33130	AT3G15290.1	N/A	Cre12.g534350.t1.2	N/A	3-hydroxybutyryl- dehydrogenase
Cz05g24080	AT1G53140.1	DRP5A	Cre12.g529450.t1.2	CGL111	dynamin-related GTPase
Cz06g31130	AT1G12250.3	TL20	Cre17.g721700.t1.2	CPLD44	thylakoid lumenal
Cz09g07060	AT3G48420.1	N/A	Cre03.g206550.t1.2	CPLD2	haloacid dehalogenase-like hydrolase domain-containing
Cz09g17160	AT3G15620.1	UVR3	Cre06.g278251.t1.1	N/A	UVR3 3-like 6-4 DNA photolyase
Cz09g22110	AT4G29190.1	OZF2	Cre07.g326150.t1.1	N/A	zinc finger CCCH domain-containing 30

One-way BLAST searches with increasing stringency identify a reduced set of thylakoid biogenesis candidate genes

In this strategy, *Cz*DEGS are used as queries to seek homologs among *At*CEGs and *Cr*ARGs in a unidirectional strategy using fixed eVals of increasing stringency. Non-photosynthetic genes are not removed at the outset, but rather later in the pipeline as described in Figure 2. Unlike *Cz*CORE, this allows homologs to non-photosynthetic organisms to be included, so long as they both meet the eVal threshold and are found in both *At*CEGs and *Cr*ARGs. This provides the opportunity to identify genes of interest that may be related to non-photosynthetic progenitors, but whose function has specialized divergently over time.

As shown in Table 1, the number of hits varies between algorithm and eVal threshold. The most relaxed search strategy, using an eVal of $1x10^{-10}$, results in 95, 83, and 93 genes using BLAST, DIAMOND, and MMseqs respectively. The total set of these genes contains 107 genes, and the overlapping genes among these three

algorithms contains 70 genes (Figure 7A). As the eVal threshold is reduced, the numbers of genes that meet these more stringent criteria are also reduced. At $1x10^{-20}$, 90 total genes are included, and 69 are common to all algorithms (Figure 7B). At a further reduction to $1x10^{-50}$, 57 total genes result with 50 common to all algorithms (Figure 7C). When the threshold is quite strict, at $1x10^{-100}$, 38 total genes are included, and 21 are common to all algorithms (Figure 7D). At the most stringent threshold utilized in this study, $1x10^{-200}$, only 12 total genes are included and 4 are common to all algorithms (Figure 7E). Among these four genes are two subunits of the magnesium chelatase complex, subunits D and H. Additionally, two genes heretofore undescribed in the process of thylakoid biogenesis are included when using the most restrictive strategy here employed. This includes Cz01g05200, a von Willebrand factor, and Cz18g04130, a catalase. The full details of the unidirectional gene sets are described in Table 2 and Figure 4 (center solid arrows).

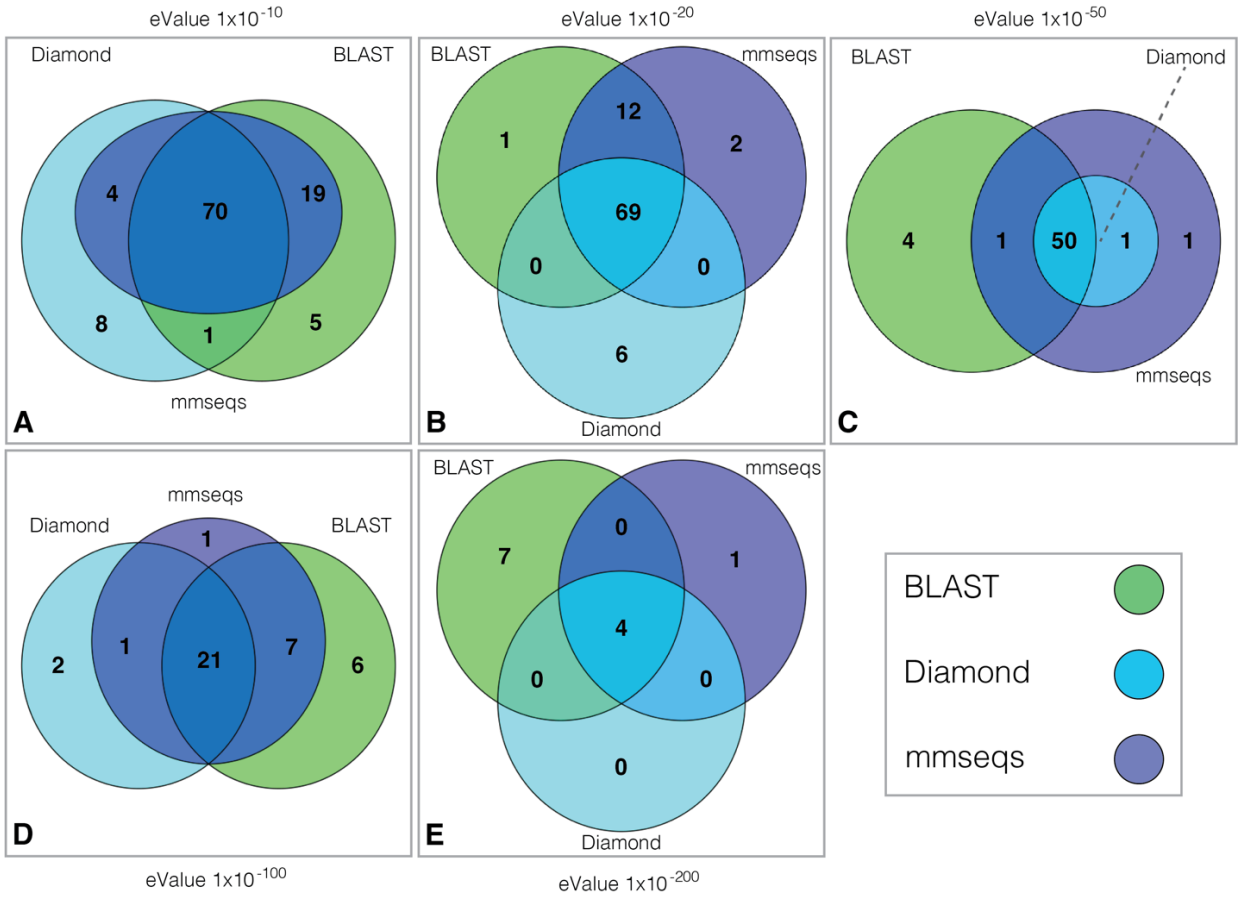


Figure 7. Numbers of resultant RGC genes depending using unidirectional search strategies with consistent eVal thresholds. A. eValue threshold of $1x10^{-10}$. B. . eValue threshold of $1x10^{-20}$ C. . eValue threshold of $1x10^{-50}$. D. eValue threshold of $1x10^{-100}$. E. . eValue threshold of $1x10^{-200}$.

Species-specific eValues identify an alternative set of genes of interest.

Using eVals that vary depending upon evolutionary relationships results in alternate sets of genes. As described in the one-way sequence alignment searches, non-photosynthetic genes are not removed at the outset in this search strategy. Species specific eVals and their derivation are described above (Figure 5).

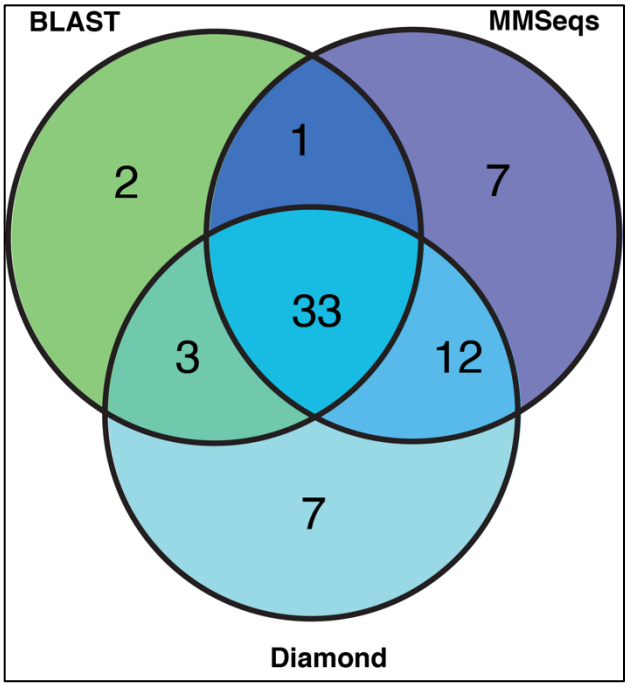


Figure 8. Species specific search results by algorithm. A total of 65 genes were identified, 33 of which were shared among all three sequence alignment algorithms. This shared set of genes is detailed in Table 4.

65 total genes were identified among all three search algorithms, with 33 genes in common (Figure 7, Table 2). Contained in this total gene set, 42 are known genes related to chlorophyll biosynthesis, photosystem subunits, LHCs, and carbon fixation. 21 of these 42 are common to the set of genes that were identified in Figure 6B, central overlap (45 genes, dark green). Of the additional genes not identified previously that are known to be involved in thylakoid biogenesis or photosynthesis, of note are *LCYE*, *POR, CPSRP43,* homologs to *AtGDC1 and AtALB3* (Cz05g36130 and Cz07g20040, respectively*),* and a ferredoxin (Cz13g08015).

Of the 23 remaining genes, two are included in the genes of interest described in Table 3 (UNPLg00445 & Cz05g10190). Interestingly, among the 21 remaining genes, two are not present in the *Cz*CORE gene set; Cz02g00170 is a homolog to *Arabidopsis AtSNM1 (*AT3G26680), and is annotated to be involved in DNA repair. Cz10g09190 contains a kinesin motor domain, which is homologous to the *Arabidopsis gene AtKIN14h* (AT3G44730). Their inclusion among the non-photosynthetic gene set in other search strategies suggests that these genes may have evolved to play a role in photosynthesis, but they retain high identity to their evolutionary precursors. These 21 potentially novel high-interest genes are detailed in Table 4.

Table 4. Genes of interest resulting from a species-specific search strategy.

Cz GeneID	*At* Homolog	*At* Gene Name	*Cr* Homolog	*Cr* Gene Name	GO terms
Cz01g31110	AT1G08860.1	BON3	Cre03.g170200.t1.1	GEM-4	calcium-dependent phospholipid-binding Copine family
Cz01g37110	AT1G54780.1	TLP18.3	Cre03.g182150.t1.2	TEF8	UPF0603 At1g54780, chloroplastic
Cz01g22130	AT1G74850.1	PTAC2	Cre10.g437150.t1.2	PPR6	Pentatricopeptide repeat-containing
Cz16g12070	AT4G27700.1	N/A	Cre16.g650050.t1.2	N/A	rhodanese-like domain-containing 14, chloroplastic
Cz01g05240	AT3G04650.1	N/A	Cre03.g207150.t1.1	N/A	hypothetical protein MNEG_13559
Cz10g10230	AT2G40140.1	CZF1	Cre07.g326150.t1.1	N/A	zinc finger CCCH domain-containing 30-like
Cz09g22110	AT4G29190.1	OZF2	Cre07.g326150.t1.1	N/A	zinc finger CCCH domain-containing 30
Cz02g00290	AT5G53490.2	TL17	Cre09.g404000.t1.1	CPLD59	Thylakoid lumenal kDa , chloroplastic
Cz06g31130	AT1G12250.3	TL20	Cre17.g721700.t1.2	CPLD44	thylakoid lumenal
Cz13g01110	AT1G08230.2	N/A	Cre03.g168550.t1.1	N/A	GABA transporter 1-like
Cz01g30130	AT3G23080.2	N/A	Cre03.g174850.t1.2	CGL111	lipid-binding START
Cz04g33130	AT3G15290.1	N/A	Cre12.g534350.t1.2	N/A	3-hydroxybutyryl- dehydrogenase
Cz09g07060	AT3G48420.1	N/A	Cre03.g206550.t1.2	CPLD2	haloacid dehalogenase-like hydrolase domain-containing
Cz15g23070	AT1G79450.1	ALIS5	Cre06.g278300.t1.2	N/A	ALA-interacting subunit 1-like
Cz10g02250	AT5G36890.2	BGLU42	Cre07.g320850.t1.2	N/A	glycoside hydrolase family 1
Cz05g24080	AT1G53140.1	DRP5A	Cre12.g529450.t1.2	CGL111	dynamin-related GTPase
UNPLg00129	AT4G08850.2	MIK2	Cre04.g216050.t1.1	N/A	surface antigen
Cz09g17160	AT3G15620.1	UVR3	Cre06.g278251.t1.1	N/A	UVR3 3-like 6-4 DNA photolyase
Cz04g04250	AT5G20140.1	HBP5	Cre16.g666550.t1.1	N/A	SOUL heme-binding
Cz02g00170	AT2G45700.1	N/A	Cre09.g399907.t1.1	N/A	DNA cross-link repair 1A
Cz10g09190	AT1G09170.4	N/A	Cre05.g233100.t1.2	N/A	Os03g0301800, partial

A reciprocal best hits approach highlights genes of high sequence similarity

In order to identify genes among *Cz*DEGS with the tightest homology to *At*CEGs and *CrARG*, a RBH approach was employed (Figure 4, right dashed arrows). The resultant gene list contains a total of 75 genes when results from all search algorithms are combined (Figure 9). Among this total gene set, 52 have been identified in previous

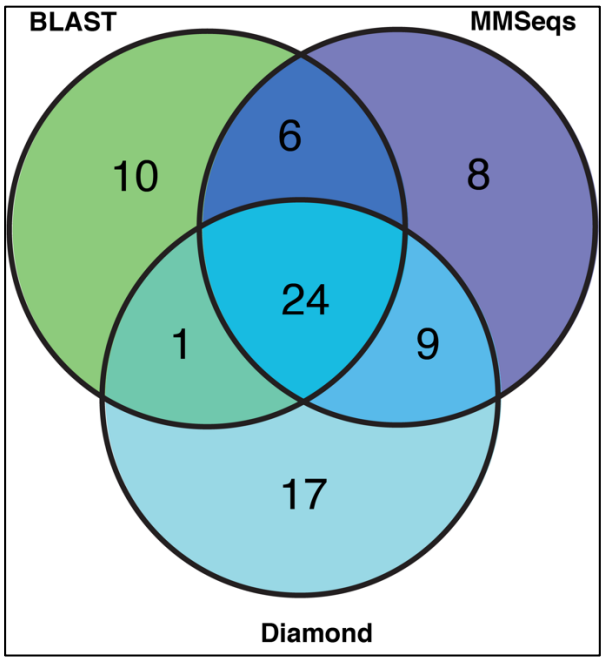


Figure 9. Numbers of genes per algorithm using a reciprocal best hits approach. 24 genes common to BLAST, MMseqs, and DIAMOND represent genes with tight homology between *C. zofingiensis, A. thaliana, and C. reinhardtii.*

search strategies presented above. The intersection of these three search strategies results in 24 genes highly conserved among *C. reinhardtii, A. thaliana, and C. zofingiensis* (Figure 9, center). Like other intersections presented above, this conserved gene set contains components of the chlorophyll biosynthesis pathway, photosystem subunits, LHCs, and proteins related to carbon fixation. These genes are also included in previously described gene sets above.

There are four genes worth remarking upon here that have not been presented in other gene sets above. Cz04g01140 is a poorly annotated gene which shares homology to *A. thaliana* TIC62. The protein encoded at the locus Cz05g10300 contains a DnaJ molecular chaperone domain. Cz05g10190 is a glyoxylase family protein which may potentially be homologous to yaeR, a methylglyoxal detoxification enzyme important in during heightened glycolysis (Sankaranarayanan et al., 2017). Finally, Cz16g01140 is a potential TAG lipase with a truncated gene sequence due to an error in the current gene model.

Perturbation-resistant genes

Genes that continue to appear in search results, despite variation in search approaches are deemed perturbation-resistant. In total, I employed 8 search strategies using a total of 3 sequence alignment algorithms resulting in 24 primary gene cuts (Table 1). When viewed as a whole, many genes appear in more than one set, and a few genes appear in many sets (Figure 9). Genes that appear in nine or more searches are likely to be involved in thylakoid biogenesis. 77 genes meet these criteria (Table 2b). The largest class of these genes are LHCs (16 genes) and proteins related to chlorophyll biosynthesis (13 genes) which can be found in Table 2b. There are 5 subunits of PSI (*PSAD, PSAE, PSAF, PSAL, and PSAO*), and 5 subunits of PSII (*PSBO, PSBP1, PSBP3, PSBR, and PSB27A*). Additionally, there are 4 genes related to lipid metabolism. Cz14g09130 is a putative TAG lipase, Cz02g33030 is annotated as a plastid-lipid associated fibrillin, and Cz01g31110 is a likely homolog to the *Arabidopsis thaliana BONZAI 3* gene, which encodes a calcium-dependent phospholipid-binding protein. Other genes that are considered perturbation-resistant are detailed in Figure 11, and full gene details can be found in Table 2a.

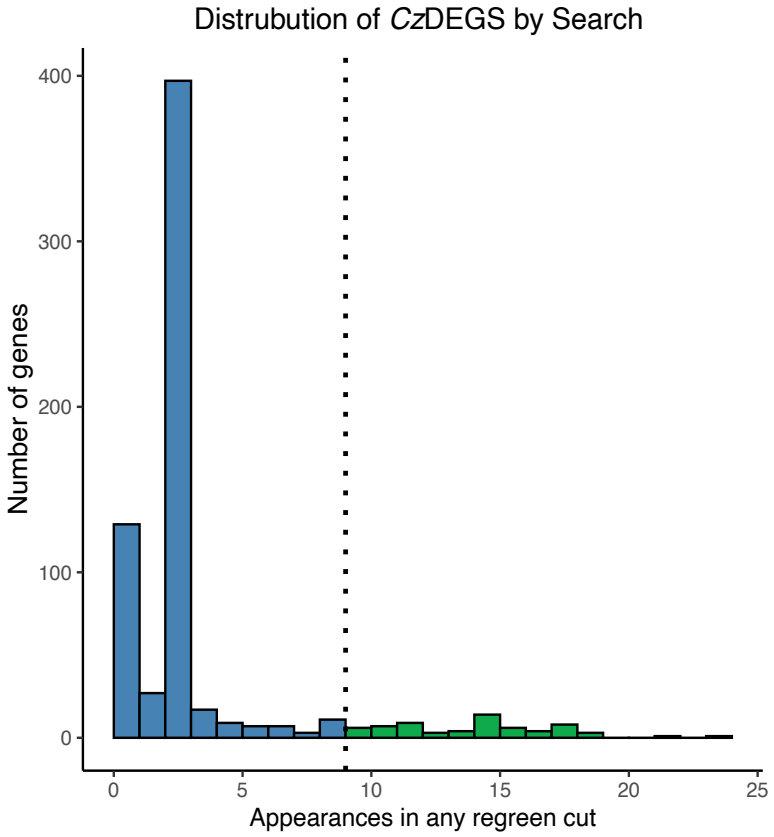
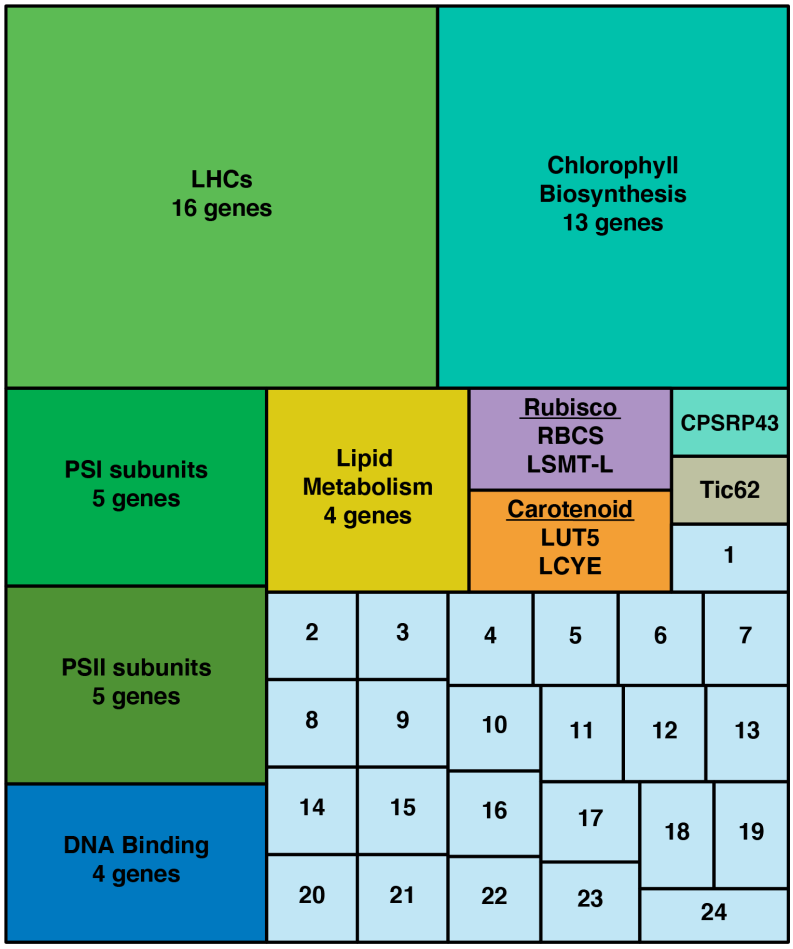


Figure 10. Distribution of the appearance of each gene in *Cz*DEGs in resultant ReGreenCuts. 77 genes appear in more than nine search results and are considered perturbation-resistant (green).

Components of the chlorophyll biosynthesis pathway appear in many search strategies. *CHLH1,* subunit H of the magnesium chelatase complex, appears in all 24 searches employed here. Other genes that appear with high frequency are *CRD1*, a Mg-Protoporphyrin Monomethyl Ester Cyclase (22/24 searches); *CHLM, a* Mg-Protoporphyrin IX methyltransferase (19/24 searches); Chlorophyll synthase (*CHLG,* 19/24 searches); and *DVR,* a Divinyl Chlorophyllide a 8-Vinyl Reductase (19/24). Genes that appear in less than 19 searches are detailed in Table 2, but a few genes of note are *CPSRP43, Tic62, ALB3, GDC1 (LTD* in *C. reinhardtii), PDE343, TLP18.3,* and *REC1.* These genes have been previously identified to have roles in either chloroplast development or the photosynthesis repair cycle. See discussion section for details and full citations. In most cases, these gene names are based on homology and annotation and so require further validation.



	Annotation	Gene ID
1	dynamin-related GTPase	Cz05g24080
2	glycosyl hydrolase	Cz10g02250
3	glyoxylase	Cz05g10190
4	oxidoreductase	Cz04g33130
5	rhodanese	Cz16g12070
6	TLP18	Cz01g37110
7	ALB3	Cz05g36130
8	GABA transporter	Cz13g01110
9	ALIS1	Cz15g23070
10	Unknown	Cz03g27110
11	HAD domain	Cz09g07060
12	TTLL	Cz04g31190
13	APE2	Cz08g12270
14	LHCP translocation defect	Cz07g20040
15	Unknown - Rossman fold	Cz01g05240
16	VWA2 - Von Willebrand factor	Cz01g05200
17	FdC1 - Ferredoxin	Cz13g08015
18	Unknown	Cz04g30100
19	Calcium sensing receptor	Cz02g14030
20	Unknown - thylakoid lumenal	Cz06g31130
21	Unknown - thylakoid lumenal	Cz02g00290
22	PDE343,PTAC2, MCA1	Cz01g22130
23	AtREC1, CrFAP	Cz04g36140
24	DWA3	Cz04g17300

Figure 11. Genes that appear in nine or more search strategies are considered perturbation-resistant. These genes contain LHCs, chlorophyll biosynthesis genes, subunits of PSII and PSII, genes with DNA binding domains, genes involved in lipid metabolism and carotenoid biosynthesis, as well as 26 genes that do not fit in neat categories. Two of these genes, CPSRP43 and Tic62 have previously been identified in *C. zofingiensis,* however 24 additional, unnamed, *zofingiensis* genes are also present among this group of perturbation-resistant genes.

Discussion

Comparative genomics is a powerful tool that is most often used to compare the whole genomes of distinct organisms in an effort to identify patterns of synteny, evolutionary history, or to identify genes that are common to organisms that perform a particular biological process. The work presented here marries a whole-genome approach to longstanding methods of RNA expression and mutant analyses. Utilizing tightly controlled experimental protocols in *C. zofingiensis* has provided a rich gene expression dataset which provides the basis of comparison to *Arabidopsis* co-expression data, *Chlamydomonas* mutant analysis, as well as whole-genome comparison to non-photosynthetic outgroup organisms.

There are likely to be additional genes that may be of interest contained among the RegreenCuts. It is not possible to exhaustively cover each possible gene of interest. Instead, these cuts stand as an additional set of information from which to identify genes or proteins for further study. This discussion will, however, highlight some genes and pathways of particular relevance.

Confirmation of known genes

The use of OrthoLang to generate sets of genes provides a standardized method to vary search strategies while producing reproducible results with comparable metrics. I used OrthoLang to identify sets of genes that are likely to be critical to thylakoid biogenesis. A large number of genes identified through this approach are well-known components of PSI, PSII, or chlorophyll biosynthesis. For example, the most well-represented gene among the 24 unique search strategies is *CHLH1*, a subunit of the magnesium chelatase subunit, which was present in 23 of 24 search strategies. In addition to readily identifiable components of photosynthesis such as LHCs or components of the chlorophyll biosynthesis pathway, the presence of genes that have previously been shown to have a role in thylakoid biogenesis such as *CPSRP43, Tic62, GDC1, ALB3,* and *PTAC2* supports the validity of this approach.

Four of these gene products, *CPSRP43, Tic62, GDC1,* and *ALB3,* facilitate the transport and insertion of new LHC proteins into the thylakoid membrane. LHC targeting passes through the TIC-TOC complex. Tic62*,* is thought to be light or redox sensitive, and it facilitates chloroplast envelope passage of LHCs (Benz et al., 2009). Once released from the TIC-TOC complex, GDC1 binds to the nascent LHC and delivers it to CPSRP43 (Cui et al., 2011; Ouyang et al., 2011). Lack of this gene in *C. reinhardtii* results in a reduction in the level of PSI and a reduced antenna size (Jeong et al., 2018). In *A. thaliana*, along with CPSRP54, CPSRP43 targets nuclear-encoded LHCs to the thylakoid membrane (Klimyuk et al., 1999; Stengel, Katharina F. et al., 2008). CPSRP43 then transports nascent LHCs to ALB3, a thylakoid membrane-bound insertase via interaction with the ALB3 C-terminal sequence (Falk et al., 2010). *CPSRP43, Tic62, GDC1,* and *ALB3* are all among those genes identified as

perturbation resistant, appearing in 10, 15, 14, and 10 of the 24 primary gene lists presented here, respectively. This includes reciprocal best hits, species-specific, and one-way alignment search strategies. This demonstrates that their persistent presence among the RegreenCuts is supported by experimental validation.

The presence of *PTAC2* in multiple cuts is also well supported by experimental evidence. In *A. thaliana, PTAC2,* encodes a protein required for the efficient translation of *psbA*. *A. thaliana* PTAC2 protein associates with plastid-encoded polymerases and can be co-purified with transcriptionally active plastid chromosomes (Pfalz et al., 2006). The *C. zofingiensis* copy of *PTAC2* is homologous to a gene in both *At*CEGs and *Cr*ARGs (AT1G74850, and Cre08.g358250, respectively), and is present in 11 of 24 primary gene searches. The orthologous gene in *C. reinhardtii* is known as *MCA1* and is among the *Cr*ARGs. In *C. reinhardtii*, MCA1 performs a role in cytochrome *f* synthesis by stabilizing *petA* mRNA. The discrepancy between the roles of *A. thaliana* PTAC2 and *C. reinhardtii* MCA1 is interesting, and will require additional understanding to fully reconcile how this highly conserved protein may perform different functions in the biosynthesis of photosynthetic complexes.

Other genes, whose mutant phenotypes have been described but whose role is only partially characterized include *TLP18.3, REC1/FLL2, DRP5A,* and the calcium sensor receptor protein *CAS.* Thylakoid lumen protein 18.3 (*TLP18.3*) is present in 75% of the primary gene sets generated during this work, meaning that it is represented in *Cz*CORE, the reciprocal best hits, species-specific, and one-way sequence alignment strategies. While the *C. reinhardtii* ortholog is not present among *Cr*ARGs, the *A. thaliana* homolog, AT1G54780 is present among *At*CEGs. This gene was initially identified in a proteomic analysis of the *A. thaliana* thylakoid lumen (Schubert et al., 2002). It was further found to be enriched in a proteomic screen of thylakoid-associated nascent chain polysomes isolated from *A. thaliania* leaves (Sirpiö et al., 2007). Plants with mutations in this gene have a stunted growth phenotype under fluctuating light, and *TLP18.3* has been implicated in the D1 repair cycle (Wu et al., 2011). This gene is conserved in the green lineage and has been experimentally shown to have an active phosphatase domain (Karpowicz et al., 2011; Wu et al., 2011). Its phosphatase activity is at an optimum between pH 3.5 and 4.5, suggesting that TLP18.3 activity could be at its peak when the thylakoid lumen is acidified during high light stress. The expression pattern of *A. thaliana* plants compared to *tlp18.3* mutant plants was investigated, and although these authors did not find marked differences in photosynthetic quantum yield, they did find a general dysregulation of many genes linked to pathogen stress (Järvi et al., 2015). A role for *TLP18.3* in thylakoid biogenesis has not previously been described, however its presence among 18 of the 24 RegreenCut gene sets makes it an attractive candidate for further study.

REC1, alternatively known as *FLL2,* is found in both *At*CEGs and *Cr*ARGs. In *C. reinhardtii,* it is annotated as a flagellar associated protein. It is conserved from cyanobacteria to plants (Nakazato et al., 2003) and was first identified as a homolog to the bacterial recombination factor, *RecA,* which facilitates homologous recombination in bacteria. In *A. thaliana,* plants with mutations in this gene have altered chloroplast size. sugar metabolism, leaf size, chlorophyll content, aberrant glucose levels, and starch

assimilation (Larkin et al., 2016; Sato et al., 2020). While it is tempting to propose a regulatory role of REC1 in chloroplast gene expression, insufficient evidence exists to support any such hypothesis currently, and further research must be done to identify the function of this gene and its potential role in thylakoid biogenesis.

Cz05g24080 is homologous to *A. thaliana DRP5A*, which is a GreenCut2 dynamin-related GTPase (Karpowicz et al., 2011). It is present in 9 of 24 searches, including reciprocal best hits, species specific, and one-way sequence alignment approaches. DRP5A is absent among the *C. reinhardtii* acetate-requiring genes, however the *C. zofingiensis* gene is a clear homolog of the *C. reinhardtii* gene (Cre12.g529450, BLAST eVal = 0, 77% sequence identity). While the related protein, DRP5B has been established to play an important role in chloroplast division, DRP5A only shares 20% sequence similarity to DRP5B, and the role of DRP5A has not yet been fully described (Hong et al., 2003; Sakamoto et al., 2008).

One of the more interesting genes present among the perturbation resistant genes is the calcium-sensing receptor *CAS* (Cz02g14030). CAS is a chloroplast-localized calcium sensor, modulator, and a member of the GreenCut2 proteins (Karpowicz et al., 2011). In *A. thaliana,* CAS is phosphorylated under high light conditions by STN7 (Cutolo et al., 2019; Vainonen et al., 2008). CAS was indirectly shown to modulate photosynthetic electron transport rate by controlling transpiration via stomatal aperture (Z. Wang & Benning, 2012; Weinl et al., 2008). Others report a direct function of CAS during *A. thaliana* de-etiolation, showing that antisense *cas* plants have significantly less chlorophyll and a delayed greening phenotype (S. S. Huang et al., 2012). Recently, *cas* mutant plants were shown to have prolonged excitation of PSI, and CAS is required for an efficient high light response (Cutolo et al., 2019). In *C. reinhardtii*, CAS is mobilized from the thylakoid membrane to the pyrenoid under low CO_2 conditions, at which point the expression of CCM-associated genes is altered (L. Wang et al., 2016). In both *A. thaliana* and *C. reinhardtii*, CAS is likely to be supported by calmodulin (S. S. Huang et al., 2012; L. Wang et al., 2016). Intriguingly, among *Cz*CORE, Cz16g12010, is annotated as containing a calmodulin domain orthologous to *A. thaliana* CML38, and four *C. zofingiensis* genes (Cz05g09090, Cz14g13070, Cz01g24040, and Cz03g31060) share significant homology to the same *C. reinhardtii* calmodulin-dependent kinase Cre17.g719450, which is among the *Cr*ARGs gene set.

As CAS modulates calcium concentration, other calcium-dependent proteins are likely affected. For example, *A. thaliana* BON3 is a calcium-dependent, lipid-binding member of the copine family. BON3 is best known for its role in repressing cell death and interacting with R genes during pathogen response (Yongqing Li et al., 2009; Yang et al., 2006). Recently BON3 has been suggested to impact pollen germination and to interact with plasma-membrane localized calcium pumps (Yun Li et al., 2018). The inclusion of *CAS* among 9 RegreenCut search strategies and other calcium-dependent proteins further supports a pivotal role of calcium during thylakoid biogenesis. Calmodulin-mediated membrane fusion has been suggested, but no experimental evidence exists describing any specific roles that calcium, calmodulins, BON3, and CAS play during thylakoid biogenesis (Mechela et al., 2019; Westphal et al., 2001).

Genes necessary for biliverdin biosynthesis are well represented among RegreenCut searches

Apart from LHCs and photosystem subunits, the most well represented class of genes among the RegreenCuts are those related to chlorophyll biosynthesis. The chlorophyll biosynthesis pathway shares many common precursors to the heme biosynthetic pathway, and the biliverdin branch of this pathway has been shown to be important for light responses in both *A. thaliana* and *C. reinhardtii* (Duanmu et al., 2013; H.-J. Lee et al., 2012). For example, the ortholog to the *A. thaliana* heme-binding protein, HBP5, is among the most well-represented genes and is present in 18 of 24 primary searches. In *A. thaliana, HBP5* encodes a chloroplast-localized heme-binding protein that is a member of a larger, conserved heme-binding protein family. *A. thaliana* mutants lacking a functional HBP5 experience increased oxidative stress (Lee et al, 2012).

A. thaliana HBP5 interacts with the HY1 heme oxygenase in vitro, forming a complex that uses heme as a substrate (H.-J. Lee et al., 2012). HY1 cleaves heme, producing the linear biliverdin tetrapyrrole biliverdin-IXα (BV-IXα), which is further reduced by biliverdin reductase to produce either phytochromobilin or phycocyanobilin, two chromophores utilized by phytochromes. Among the perturbation-resistant gene set is an uncharacterized gene UNPLg00445, which is annotated as a biliverdin reductase. The homologous gene in *A. thaliana* (AT2G34460) is an uncharacterized Greencut2 protein (Karpowicz et al., 2011). This suggests that a conserved biliverdin biosynthesis pathway in both *A. thaliana* and *C. zofingiensis* is important to thylakoid biogenesis.

This is supported by studies of biliverdin metabolism in *C. reinhardtii.* Although *C. reinhardtii* and many other chlorophytes lack phytochromes, these organisms maintain a functional biliverdin biosynthetic pathway, which is regulated by light and oxidative stress (Duanmu et al., 2013). This pathway is necessary for light-dependent photoacclimation (Wittkopp et al., 2017). *C. reinhardtii* mutants lacking the heme-oxygenase *HMOX1* phenocopy chlorophyll biosynthesis mutants due to a severe decrease in PSI biosynthesis, which can be rescued by exogenous addition of BV-IXα (Wittkopp et al., 2017). A functional biliverdin biosynthetic pathway has been shown to have a large impact on light-dependent transcriptional responses (Duanmu et al., 2013; Wittkopp et al., 2017). Additionally, BV-IXα interacts with the light-dependent protochlorophyllide oxidoreductase LPOR via PCYA1, a ferredoxin-dependent biliverdin reductase (W. Zhang et al., 2018)). Unsurprisingly, *Cz*POR1 is well represented among the RegreenCuts, being present in 18 of 24 primary gene lists. The ferredoxin *FDX4* is also present among the perturbation-resistant genes. The *C. zofingiensis* homolog to *C. reinhardtii PCYA1,* UNPLg00517, while expressed at detectable levels, is not differentially expressed during Regreening, and is not included in *Cz*DEGS.

The specific mechanism by which BV-IXα and its derivatives alter light responses is unclear. It has been proposed that an uncharacterized protein has the capacity to bind a BV-IXα derivative, acting similarly to phytochromes in plant species (Wittkopp et al.,

2017). Three *C. reinhardtii* genes were put forward as possible proposed *chlorochromes* (Cre12.g552150, Cre14.g628200, and Cre17.g739250), however none of these genes are among *Cr*ARGs (Wittkopp et al., 2017). Additionally, their most similar orthologs in *C. zofingiensis* are not among *Cz*DEGs, suggesting that if a chlorochrome exists in *C. reinhardtii* it is not likely one of these proposed genes. Other *C. zofingiensis* proteins annotated as containing the necessary GAF domain for billiverdin binding are also absent from *Cz*DEGs. This area of research requires additional experimental work to clarify the exact nature of biliverdin derivatives and their role in thylakoid biogenesis.

Known genes with a potentially novel role in thylakoid biogenesis

Among the perturbation-resistant genes, some are homologs of well-known genes that have not been previously reported to play a role in thylakoid biogenesis. These are listed in Figure 11 and include homologs to *AtTPT, AtALIS1*, *AtDWA3*, as well as two unnamed genes annotated as a glyoxalase and GABA transporter (Cz05g10190 and Cz13g01110 respectively). A brief explanation of these genes follows.

Triose phosphate transporter (*TPT/APE2)* is a gene whose mutant has been shown to have a reduction in maximum photosynthetic efficiency (Walters et al., 2003). In *A. thaliana,* this gene encodes a sugar transporter. There is a representative ortholog in *Cr*ARGs, *At*CEGs, and *Cz*DEGS (Cre11.g467754, AT5G46110, and Cz08g12270 respectively). Both the *A. thaliana* and *C. reinhardtii* orthologs are among the Greencut2 proteins suggesting that this particular sugar transporter is involved in photosynthesis. (Karpowicz et al., 2011). ALIS1 is named for its interactions with ALA and forms a stable complex with ALA3 which together are able to translocate lipids when co-expressed in *S. cerevisiae* (Poulsen et al., 2008). In *A. thaliana,* DWA3 is negative repressor of abscisic acid (ABA) responses, and *dwa3* plants are hypersensitive to ABA (J. H. Lee et al., 2011). DWA3 is thought to mediate ABA responses via an E3-ubiquitin ligase complex (Seo et al., 2014). ABA is a well-known phytohormone responsible for, among other things, the promotion of seed dormancy (Kumar et al., 2019). The presence of *DWA3* among many RegreenCuts (12/24) suggests that DWA3 is functioning to release ABA repression of chloroplast development. Previous work has shown that ABA treatment represses chloroplast-targeted genes that are both nuclear and chloroplast-encoded (Kusnetsov et al., 1998; Yamburenko et al., 2013).

Present in 16 gene lists, the putative *C. zofingiensis* glyoxalase, Cz05g10190 is highly represented among the 24 RegreenCuts. This gene has homologous genes in both *Cr*ARGs and *At*CEGs, which are both among the Greencut2 proteins (AT5G57040, Cre10.g466500, respectively, (Karpowicz et al., 2011). Much focus has been placed on the role of glyoxalases in detoxifying the photosynthetic byproduct methylglyoxal and minimizing damage from reactive oxidative species (Schmitz et al., 2018) However recent work has expanded the functional role of these enzymes into other biological pathways such as seed germination and cell division (Singla-Pareek et al., 2020).

Finally, of note among the perturbation resistant genes is Cz13g01110, annotated as a GABA transporter. This gene is present in 10 of 24 searches and is orthologous to the uncharacterized *A. thaliana* gene at locus AT1G61270. GABA is an amino acid that is not utilized in proteins but serves as a metabolite and stress signaling molecule (Bown & Shelp, 2016; Shelp et al., 2012). What role GABA plays in thylakoid biogenesis is unclear, but it may be related to increased reactive oxygen species during photosynthesis.

Twelve genes pass the most stringent eValue searches

As shown on Figure 7E and in Table 1, when the results from all algorithms are combined, there are 12 genes that pass the most stringent eVal threshold ($1x10^{-200}$). Six of these genes can confidently be assigned to tetrapyrrole biosynthesis. Of the remaining six genes, Cz04g03070 is an ortholog of *At*HCEF1, a chloroplast localized fructose-1,6-bisphosphate. *A. thaliana* plants with lesions in *HCF1* have elevated levels of cyclic electron flow potentially due to the accumulation of metabolites that stimulate cyclic electron flow (Livingston et al., 2010). There are two likely catalases among this narrow gene set, Cz18g04130 and UNPLg00177, which both have sequence alignment hits to the *A. thaliana* catalase, *SEN2/CAT3,* with an eValue of zero. Catalases are critical enzymes in the management of H_2O_2 levels, both detoxifying and controlling H_2O_2 signaling (Vandenabeele et al., 2004), and they are likely performing a similar role during thylakoid biogenesis. During initial photosystem assembly, imbalances between excited chlorophyll and downstream carbon fixation that may lead to an overproduction of reactive oxidative species. The presence of two catalases among the strict RegreenCut suggests that the mitigation of H_2O_2 is an important aspect of thylakoid biogenesis.

Additionally, a curious gene with a Von-Willebrand domain is present among this strict set of genes (Cz01g05200). With an alignment hit only among *Cr*ARGs (Cre03.g207400), this gene may not be common to plants. This is supported by an NCBI-hosted BLAST search against the non-redundant protein database, which contains 3.23×10^8 sequences to date. When using a threshold eVal of 5×10^{-2} in a BLASTP query using this Von Willebrand-domain containing protein sequence, 794 hits result. Of these, approximately 91% of these hits are to bacteria and 7.3% of these hits are among the Chlorophyta. There are no hits suggesting homology of this gene to red algae or vascular plants. Among algal photosynthesis researchers, this gene may be well worth a deeper investigation.

Lastly, among these 12 genes are two which do not contain enough sequence similarity to annotated genes to allow credible speculation. Cz12g24200 has a kinesin motor domain that is homologous to an *A. thaliana* kinesin (AT3G45850), which has not been the specific focus of any research. Finally, Cz16g05180 is an NADH ubiquinone reductase that has sequence alignment hits to a gene in *At*CEGs and *Cr*ARGs (LCI18/NDA3 NAD(P)H dehydrogenase, which is annotated as mitochondrially localized

in both *A. thaliana* and *C. reinhardtii,* but is predicted to be chloroplast-targeted in *C. zofingiensis.*

Comparison to GreenCut2

The GreenCut and the subsequent update known as the GreenCut2 represent an important body of work in the field of comparative genomics (Karpowicz et al., 2011; Merchant et al., 2007). In an effort to provide additional context to the gene lists generated here, I sought genes that were among the broader *Cz*DEGs gene set that had clear homologs to genes in *A. thaliana* or *C. reinhardtii* and compared those homologs to the genes in the GreenCut2.

Among genes shared between *At*CEGs, *Cr*ARGs, and *Cz*CORE, there are 12 orthologs that are included in the GreenCut2. These genes are *PSBP1*, *PSAE, PSB27, PSAD, PSBP, RBCS, CPSRP43, UROD4, CHLD,* Cz05g10190, Cz08g12270, and Cz12g11010.

The three unnamed genes, Cz05g10190, Cz08g12270, and Cz12g11010 are annotated as a glyoxylase family protein (yaeR), a solute carrier with homologs in both *A. thaliana* and *C. reinhardtii, APE2,* and a gene homologous to both *AtAFT* and *CrANK22b,* respectively. *APE2* is discussed above, however to my knowledge, *yaeR* has not been well characterized in photosynthetic organisms, although its presence in the GreenCut2 gene set demonstrates its evolutionary conservation (Karpowicz et al., 2011). This yaeR homolog likely plays a role in the glyoxylate cycle, which has implications on lipid catabolism (Theodoulou & Eastmond, 2012). *At*AFT is of a particular interest with respect to thylakoid biogenesis as it has been shown to be a necessary protein to target proteins to the chloroplast outer envelope membrane by binding the thylakoid lipids MGDG and phosphatidylglycerol (Bae et al., 2008; Kim et al., 2014). *AtAFT2* works in conjunction with a small heat-shock protein, HSP 17.2 (Kim et al., 2011). While HSP 17.2 was not included in *AtCEGs,* there is an homologous gene among *Cz*COREs (Cz05g36020) that may serve the same role in *C. zofingiensis.*

In addition to the 12 genes shared among *A. thaliana* and *C. reinhardtii* GreenCut2 proteins, there are other sets of RegreenCut genes shared among the GreenCut2. These are *C. reinhardtii* and *A. thaliana* genes that are present in the GreenCut2, which did not have matching homologs in *At*CEGs and *Cr*ARGs respectively. This does not imply that these genes are any more or less important, only that they were not present in the starting gene lists, and therefore would not be found in the resultant lists.

There are two sets of genes to consider, those GreenCut2 genes common to *C. reinhardtii* and *CzDEGs* and those GreenCut2 common to among *A. thaliana* and *CzDEGs*. There are five genes shared between *C. reinhardtii* Greencut2 genes and *CzDEGs*. Of these, four are part of the chlorophyll biosynthesis pathway, *UROD3, CHLI,* and *GUN4*, and an additional, unannotated copy of *CHLI.* This second *CHLI* is split across 2 gene models (UNPLg00023 & UNPLg00024, version 5.2.3.2) which likely prevented its correct annotation when gene model prediction was performed (Roth et

al., 2017). The final gene among these GreenCut2 genes is ELIP3. Early light-induced proteins are likely involved in photoprotection during thylakoid biogenesis (Dall'Osto et al., 2015; Van Buren et al., 2019).

There are 43 GreenCut2 genes shared among *At*CEGs and *Cz*DEGs, substantially more than in *Cr*ARGs, owing to the larger number of genes in the *At*CEGs when compared to *Cr*ARGs (2003 to 307, respectively). Unsurprisingly, about 67% of genes included in this set are related to PSI, PSII, LHCs, chlorophyll biosynthesis, or are other known components of photosynthesis such as plastocyanin (Table 5). Two genes, TLP18.3 and GDC1, were previously remarked upon in this discussion. This leaves 14 GreenCut2 genes which are either completely uncharacterized, or have no obvious role in the process of thylakoid biogenesis.

Table 5. *Cz*DEGs genes common to *Arabidopsis thaliana* GreenCut2 genes

Cz GeneID	*Cz* Name	*At* GeneID	*At* Name	*At* Description
Cz01g23030	LHC1	AT3G47470.1	LHCA4,CAB4	LHCP complex I subunit A4
Cz01g30070	CYP26	AT1G74070.1	N/A	no full name available
Cz01g35130	PETE	AT1G76100.1	PETE1	plastocyanin 1
Cz01g37110	N/A	AT1G54780.1	TLP18.3,AtTLP18.3	thylakoid lumen protein 18.3
Cz02g00290	N/A	AT5G53490.3	N/A	no full name available
Cz02g14020	N/A	AT5G23060.1	CaS	calcium sensing receptor
Cz02g14140	LHC8	AT3G47470.1	LHCA4,CAB4	LHCP complex I subunit A4
Cz02g18180	PSAL	AT4G12800.1	PSAL	PS I subunit l
Cz02g24270	CHLM	AT4G25080.3	CHLM	magnesium-protoporphyrin IX methyltransferase
Cz02g33030	N/A	AT5G19940.1	N/A	no full name available
Cz03g01030	N/A	AT1G71500.1	PSB33	PS B protein 33
Cz03g06230	DVR	AT5G18660.1	PCB2	pale-green and chlorophyll b reduced 2
Cz04g04130	LHC2	AT3G47470.1	LHCA4,CAB4	LHCP complex I subunit A4
Cz04g04250	N/A	AT5G20140.2	HBP5,AtHBP5	heme-binding protein 5
Cz04g30100	N/A	AT2G05310.1	N/A	no full name available
Cz05g09100	N/A	AT4G21280.2	PSBQ-1,PSBQA,PSBQ	PS II subunit QA, PS II subunit Q-1, PS II Q
Cz05g24080	N/A	AT3G19720.1	ARC5,DRP5B	Dynamin related protein 5B,accumulation and replication of chloroplast 5
Cz05g29140	N/A	AT1G14030.1	LSMT-L	lysine methyltransferase (LSMT)-like
Cz05g32290	N/A	AT2G34460.1	N/A	no full name available
Cz06g26010	PSAK	AT1G30380.1	PSAK	PS I subunit K
Cz06g28160	POR	AT4G27440.1	PORB	protochlorophyllide oxidoreductase B
Cz06g31120	N/A	AT1G12250.1	N/A	no full name available
Cz06g31280	PSAG2	AT1G55670.1	PSAG	PS I subunit G
Cz07g01040	N/A	AT2G44920.2	N/A	no full name available
Cz07g20040	N/A	AT1G50900.1	GDC1,LTD	Grana Deficient Chloroplast 1,LHCP translocation defect
Cz08g11020	N/A	AT3G07670.1	N/A	no full name available
Cz08g11160	LHCB5	AT4G10340.1	LHCB5	light harvesting complex of PS II 5
Cz08g28070	CHLG	AT3G51820.1	G4,CHLG,ATG4,PDE325	PIGMENT DEFECTIVE 325
Cz09g07060	N/A	AT3G48420.1	N/A	no full name available
Cz09g18310	LCYE	AT5G57030.1	LUT2	LUTEIN DEFICIENT 2
Cz11g10100	N/A	AT5G53580.1	AtPLR1,PLR1	pyridoxal reductase 1
Cz11g26110	N/A	AT5G58330.1	NADP-MDH	NADP-dependent Malate Dehydrogenase
Cz12g12230	PSAH	AT1G52230.1	PSI-H,PSAH-2,PSAH2	PS I subunit H2, PS I subunit H-2
Cz13g06010	PSAF	AT1G31330.1	PSAF	PS I subunit F
Cz13g08015	N/A	AT4G14890.1	FdC1	ferredoxin C 1
Cz15g01010	PSAO	AT1G08380.1	PSAO	PS I subunit O
Cz15g17110	LHC7	AT3G54890.1	LHCA1	PS I light harvesting complex gene 1
Cz16g08310	CRD1	AT3G56940.1	CRD1,ACSF,CHL27	COPPER RESPONSE DEFECT 1
Cz16g14150	LHC9	AT1G61520.1	LHCA3	PS I light harvesting complex gene 3
Cz16g19040	PSBO	AT5G66570.1	OEE1,PSBO-1,MSP-1, OE33,OEE33,PSBO1	PS II oxygen-evolving complex 1, oxygen evolving complex 33 Kda
Cz17g15140	PSBY	AT1G67740.1	YCF32,PSBY	PS II BY
Cz19g05110	PSBW	AT2G30570.1	PSBW	PS II reaction center W
UNPLg00445	N/A	AT2G34460.1	N/A	no full name available

SPL13 is an Under-characterized Regulatory Element Present in *Cz*DEGs, *At*CEGs, and *Cr*ARGs

There are a number of genes among these results that likely play regulatory roles, either as transcription factors, or through other DNA binding activity. Of note is Cz15g08240, which appears in 9 of 24 gene lists. This gene is homologous to a Squamosa-promoter-binding like gene in *A. thaliana* (*SPL13*) and *C. reinhardtii* (Cre10.g433350). *SPL* genes have been demonstrated to play a developmental role in *A. thaliana,* aiding in the control of flowering, and *SPL13,* in particular contains a distinguishing transcriptional feature in its 3' UTR (Schmid et al., 2005). When using this *C. zofingiensis* sequence as a BLAST query against the non-redundant protein database with an eValue of 1×10^{-5}, there are no hits to organisms outside of the viridiplantae, suggesting that this gene may have a regulatory role specific to green photosynthetic eukaryotes.

Potential limitations of this work

This approach is able to confidently identify known genes related to thylakoid biogenesis, as well as other which have not previously been highlighted as important to this process. However, there are significant limitations based upon the assumptions of the researcher. For example, in the case presented here, the presence or absence of a gene among the *A. thaliana* co-expressed genes was entirely dependent upon choices made. The bait genes used for co-expression analysis were chosen based on a survey of the literature, which may leave room for systematic improvement. Additionally, what value to use as a co-expression threshold was not derived empirically, leaving room for a somewhat arbitrary inclusion or exclusion from the input *At*CEGs gene list.

Additionally, I chose to utilize the reference organism, *A. thaliana*, because of its vast community of researchers and the wide range of resources they have built. However, *A. thaliana* is only one species of plant and is not necessarily representative of all photosynthetic organisms. All the necessary caveats stand when transferring findings from one organism to another. Similar does not mean same. Fortunately, I was able to include a dataset generated in *C. reinhardtii*, which like *A. thaliana* has a robust research community and sets of resources. While a wider taxonomy of organisms should be included in any future comparative genomics and comparative expression analysis, this work demonstrates that sufficient evolutionary conservation exists between *C. zofingiensis, C. reinhardtii ,* and *A. thaliana* to generate functional datasets.

An additional limitation to this work stems from the initial constraints placed upon the *C. zofingiensis* transcriptional dataset, *Cz*DEGs. While I chose to restrict timepoints to 1, 3, and 6 h post glucose removal, there exists data at time point zero, 0.5, and 12 h. It is possible that this constraint prevents the identification of additional genes of interest. A similar case could be made for the exclusion of genes that were

downregulated post glucose removal. Similarly, a set of genes that would be entirely absent from this approach are those that are translated at steady-state levels, but function is altered by post-translational modification.

Conclusions and Future Directions

The application of a comparative genomics framework to expression data and a mutant library represents an integration of data-types that can provide deeper insight than each of these approaches alone. While analysis of expression data can enable identification of actively expressed or repressed genes given a particular stimulus, the interpretation of such data sets is usually limited to the single organism studied. Comparative genomics is also a powerful approach to examine evolutionary relationships among organisms, providing results on the presence or absence of genes that are correlated with any relationship of interest such as evolutionary adaptation or morphological phenotype. Comparative genomics is unable, however, to resolve the activity of genes so identified. Mutant libraries are arguably the most powerful tool for the reliable identification of single genes correlated with a mutant phenotype. However, mutant libraries are unlikely to capture all genes related to a given phenotype due to lack of saturation during mutant screening or redundancy of such genes. By restricting gene sets using expression or mutant phenotype and then comparing those genes to the whole genomes of non-photosynthetic outgroup organisms, the approach presented here combines expression data, a mutant library and whole-genome comparisons to generate gene lists with multiple supporting data-types. Further, utilizing OrthoLang makes these results robust, repeatable, and able to be iterated as more information comes to light.

The inclusion of many critical genes related to photosynthesis repair and chloroplast biogenesis, as well genes previously identified in the GreenCut and GreenCut2 lends support to this approach (Karpowicz et al., 2011; Merchant et al., 2007). Uncharacterized genes presented on these lists remain to be validated experimentally, but they represent high-priority experimental targets to further our understanding of thylakoid biogenesis

Table 2. ReGreenCuts. Genes included in each cut are listed below. Column headers to primary gene cuts are shaded in grey. This is a subset of a fully annotated table which is available at dryad.org https://doi.org/10.6078/D1X995

Gene ID	Gene Name	CzCORE	Non Photosynthetic Homolog	BLAST *Cz*CORE	Diamond *Cz*CORE	MMseqs *Cz*CORE	*Cz*CORE/*At*CEG/*Cr*ARC overlap	Perturbation-resistant	BLAST RBH	Diamond RBH	MMseqs RBH	BLAST Species Specific	Diamond Species Specific	MMseqs Species Specific	BLAST eVal 10E-10	Diamond eVal 10E-10	MMseqs eVal 10E-10	BLAST eVal 10E-20	Diamond eVal 10E-20	MMseqs eVal 10E-20	BLAST eVal 10E-50	Diamond eVal 10E-50	MMseqs eVal 10E-50	BLAST eVal 10E-100	Diamond eVal 10E-100	MMseqs eVal 10E-100	BLAST eVal 10E-200	Diamond eVal 10E-200	MMseqs eVal 10E-200
Cz09g03180	ALAD		×																										
Cz09g17290	ARC6-2	×			×	×	×																						
Cz01g24010	ARC6-3	×		×	×	×																							
Cz04g11250	BKT2	×		×	×	×																							
Cz08g03270	CHLD		×	×				×		×																			
Cz08g28070	CHLG	×		×	×	×		×	×	×	×																		
Cz14g15070	CHLH1	×		×	×	×	×	×	×	×	×	×																	
Cz06g16260	CHLI	×		×	×		×	×	×	×	×	×	×																
Cz02g24270	CHLM	×		×	×	×		×	×	×	×	×	×	×															
Cz10g07010	CPSRP43		×					×	×	×	×	×	×	×	×														
Cz01g30070	CYP26	×			×			×	×	×	×	×	×	×	×	×													
Cz16g08310	CRD1	×		×	×	×		×	×	×	×	×	×	×	×	×	×												
Cz03g14080	DGAT1C	×			×																								
Cz03g06230	DVR	×		×	×	×		×	×	×	×	×	×	×	×	×	×	×											
Cz14g24190	ELIP3	×		×	×	×				×	×	×	×	×	×	×	×	×	×										
Cz05g01180	FBP		×																										
Cz11g03180	FDX3	×		×	×	×																							
Cz12g26150	FTSY		×																										
Cz11g23070	GSA		×																										
Cz03g02090	GTR3		×																										

Cz11g24180	GUN4	x		x	x	x		x	x	x	x				x	x	x	x	x	x	x	x	x						
Cz07g04220	ICL		x																										
Cz05g30060	LACS3		x																					x	x	x			
Cz09g18310	LCYE	x		x	x	x		x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Cz01g23030	LHC1	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz13g03240	LHC16	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Cz07g02110	LHC17	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x		x			
Cz07g02100	LHC18	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x		x			
Cz15g09060	LHC19	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Cz04g04130	LHC2	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
UNPLg00080	LHC20	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Cz04g09210	LHC3	x		x	x	x	x	x					x	x	x	x	x	x	x	x	x	x	x	x					
Cz09g32070	LHC4	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz10g10160	LHC5	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz12g16070	LHC6	x		x	x	x	x	x					x	x	x	x	x	x	x	x	x								
Cz15g17110	LHC7	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz02g14140	LHC8	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz16g14150	LHC9	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x					
Cz03g26240	LHCB4	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz08g11160	LHCB5	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz14g25060	MAS1		x																										
Cz12g06020	MDH4		x																							x			
Cz12g08110	NIR1	x			x																								
Cz03g22200	PBGD1		x																					x	x	x			
Cz01g05160	PCK		x																										
Cz01g35130	PETE	x		x	x	x									x	x	x												
Cz06g28160	POR	x		x	x	x		x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
Cz05g36030	PSAD	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						
Cz04g01180	PSAE	x		x	x	x	x	x	x	x	x				x	x	x	x	x	x									

Cz13g06010	PSAF	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x								
Cz06g31280	PSAG2	x		x	x	x									x		x												
Cz12g12230	PSAH	x		x	x	x									x		x												
Cz06g26010	PSAK	x		x	x	x				x					x	x	x												
Cz02g18180	PSAL	x		x	x	x		x	x	x	x		x	x	x	x	x	x	x	x			x						
Cz17g10010	PSAN1	x		x	x	x			x		x				x		x												
Cz15g01010	PSAO	x		x	x	x		x	x	x	x				x	x	x	x	x	x									
Cz04g12110	PSB27-H1-1	x		x	x	x	x	x	x	x	x				x	x	x	x	x	x									
Cz16g19040	PSBO	x		x	x	x		x				x	x	x	x	x	x	x	x	x	x	x	x	x	x				
Cz02g18040	PSBP1	x		x	x	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x					
Cz07g15040	PSBP3	x		x	x	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x	x	x					
Cz09g21140	PSBR	x		x	x	x		x	x	x	x				x	x	x	x	x	x									
Cz19g05110	PSBW	x		x	x	x			x						x														
Cz17g16010	PSBX	x		x	x	x			x		x																		
Cz17g15140	PSBY	x		x	x	x					x				x		x												
Cz17g13100	RBCS	x		x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x						
Cz12g01050	UROD2	x		x	x	x																							
Cz14g23190	UROD3		x																					x		x			
Cz16g13070	UROD4		x																					x	x	x	x		
Cz19g06290	YCF54	x		x	x	x																							
Cz01g00140	N/A		x												x		x												
Cz01g01180	N/A	x		x	x	x																							
Cz01g01190	N/A	x			x		x								x		x												
Cz01g05200	N/A	x			x			x		x						x			x		x	x	x	x	x	x	x	x	x
Cz01g05240	N/A	x		x	x	x		x	x	x	x			x	x	x	x	x	x	x									
Cz01g05275	N/A	x		x	x	x			x																				
Cz01g07180	N/A		x																		x	x	x						
Cz01g07200	N/A	x		x	x	x																							

Cz01g07230	N/A		x																											
Cz01g08080	N/A		x																											
Cz01g08100	N/A	x		x	x	x																								
Cz01g08110	N/A	x		x	x	x																								
Cz01g10020	N/A	x			x	x																								
Cz01g11180	N/A	x		x	x	x																								
Cz01g11210	N/A	x			x	x																								
Cz01g12150	N/A	x		x	x	x																								
Cz01g15080	N/A	x		x	x	x																								
Cz01g16070	N/A		x																											
Cz01g19210	N/A	x		x	x	x																								
Cz01g20200	N/A	x		x	x	x																								
Cz01g22130	N/A	x		x	x	x	x	x					x	x	x	x	x	x	x	x										
Cz01g24030	N/A	x		x	x	x																								
Cz01g24040	N/A		x																x											
Cz01g28030	N/A	x		x	x	x																								
Cz01g28050	N/A	x		x	x	x																								
Cz01g29110	N/A	x		x	x	x																								
Cz01g30040	N/A	x			x	x																								
Cz01g30130	N/A	x		x	x	x								x	x		x	x		x										
Cz01g30160	N/A	x		x	x	x																								
Cz01g31110	N/A	x		x	x	x		x				x	x	x	x	x	x	x	x	x	x	x	x	x		x				
Cz01g31220	N/A	x		x	x	x																								
Cz01g32230	N/A	x			x																									
Cz01g32280	N/A	x			x	x																								
Cz01g34300	N/A	x		x	x	x																								
Cz01g35190	N/A		x																											
Cz01g35200	N/A		x																											
Cz01g37100	N/A	x			x																									

Cz01g37110	N/A	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						
Cz01g38010	N/A	x		x	x	x																							
Cz01g38280	N/A	x		x	x	x																							
Cz01g41040	N/A	x		x	x	x																							
Cz01g42030	N/A		x																										
Cz01g44190	N/A		x																										
Cz01g46020	N/A		x																										
Cz01g46080	N/A	x		x	x	x																							
Cz01g46120	N/A	x		x	x	x																							
Cz01g46130	N/A	x		x	x	x																							
Cz01g46280	N/A	x		x	x	x									x		x	x		x									
Cz02g00170	N/A		x									x									x	x	x						
Cz02g00180	N/A	x		x	x	x																							
Cz02g00290	N/A	x		x	x	x		x		x	x			x	x	x	x	x		x	x								
Cz02g01200	N/A	x		x	x	x																							
Cz02g02080	N/A	x		x	x	x																							
Cz02g02120	N/A	x		x	x	x																							
Cz02g03170	N/A		x																										
Cz02g04130	N/A	x		x	x	x																							
Cz02g07300	N/A	x		x	x	x																							
Cz02g08020	N/A	x		x	x	x																							
Cz02g08090	N/A	x		x	x	x																							
Cz02g11180	N/A	x		x	x	x																							
Cz02g12180	N/A	x			x													x		x									
Cz02g13240	N/A	x			x	x																							
Cz02g13250	N/A	x		x	x	x																							
Cz02g14010	N/A	x		x	x	x																							
Cz02g14020	N/A	x		x	x	x																							
Cz02g14030	N/A	x		x	x	x		x		x					x	x	x	x		x									

Cz02g14200	N/A	x		x	x	x																								
Cz02g15090	N/A	x		x	x	x																								
Cz02g15220	N/A	x		x	x	x																								
Cz02g16260	N/A	x		x	x	x																								
Cz02g17080	N/A	x		x	x	x																								
Cz02g17090	N/A	x		x	x	x																								
Cz02g24230	N/A		x																											
Cz02g27170	N/A	x			x	x																								
Cz02g30140	N/A	x		x	x	x																								
Cz02g30190	N/A	x		x	x	x																								
Cz02g32020	N/A	x			x	x					x																			
Cz02g33030	N/A	x		x	x	x		x		x	x				x	x	x	x	x	x										
Cz02g33240	N/A	x		x	x	x																								
Cz02g35020	N/A	x		x	x	x																								
Cz02g36040	N/A		x																											
Cz02g36170	N/A	x			x																									
Cz02g37020	N/A	x			x	x																								
Cz02g37030	N/A	x		x	x	x																								
Cz02g37280	N/A	x		x	x	x																								
Cz02g38290	N/A	x		x	x	x																								
Cz02g39090	N/A	x		x	x	x																								
Cz02g39300	N/A	x			x																									
Cz02g41100	N/A	x		x	x	x																								
Cz02g41120	N/A	x		x	x	x																								
Cz03g01030	N/A	x		x	x	x			x		x																			
Cz03g02250	N/A	x		x	x	x																								
Cz03g03170	N/A	x		x	x	x																								
Cz03g04020	N/A	x			x																									
Cz03g05280	N/A	x		x	x	x																								

Cz03g08220	N/A		x						x	x																			
Cz03g09130	N/A	x		x	x	x																							
Cz03g10030	N/A	x		x	x	x																							
Cz03g10070	N/A	x			x	x																							
Cz03g12070	N/A	x		x	x	x																							
Cz03g14220	N/A		x																										
Cz03g14230	N/A		x																										
Cz03g15290	N/A	x		x	x	x																							
Cz03g16130	N/A	x			x	x																							
Cz03g16220	N/A	x		x	x	x																							
Cz03g21010	N/A	x		x	x	x																							
Cz03g21130	N/A		x																										
Cz03g21140	N/A	x		x	x	x																							
Cz03g21250	N/A	x		x	x	x																							
Cz03g22010	N/A		x																										
Cz03g22100	N/A	x		x	x	x																							
Cz03g24250	N/A	x		x	x	x																							
Cz03g27010	N/A	x		x	x	x																							
Cz03g27110	N/A	x		x	x	x		x	x	x	x				x	x	x												
Cz03g29040	N/A	x		x	x	x																							
Cz03g29240	N/A		x						x																				
Cz03g31060	N/A	x			x		x																						
Cz03g31170	N/A	x		x	x	x																							
Cz03g31210	N/A	x		x	x	x																							
Cz03g35060	N/A		x							x	x											x	x	x					
Cz03g35210	N/A		x																										
Cz03g40190	N/A	x		x	x	x																							
Cz04g00130	N/A	x		x	x	x																							
Cz04g00220	N/A	x		x	x	x																							

Cz04g01120	N/A	x			x																								
Cz04g01140	N/A	x			x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x						
Cz04g01170	N/A		x																										
Cz04g03070	N/A		x																					x	x	x	x		
Cz04g04050	N/A	x		x	x	x																							
Cz04g04160	N/A	x		x	x	x																							
Cz04g04250	N/A	x		x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x						
Cz04g06030	N/A	x			x																								
Cz04g07220	N/A		x																										
Cz04g08100	N/A	x		x	x	x																							
Cz04g08190	N/A	x		x	x	x																							
Cz04g10030	N/A	x		x	x	x																							
Cz04g10080	N/A		x																										
Cz04g12210	N/A	x			x																								
Cz04g13010	N/A	x		x	x	x																							
Cz04g13180	N/A	x		x	x	x																							
Cz04g13260	N/A	x		x	x	x																							
Cz04g17300	N/A	x			x		x	x								x	x	x	x	x	x	x	x	x	x	x			
Cz04g18040	N/A	x		x	x	x																							
Cz04g19270	N/A	x		x	x	x																							
Cz04g21210	N/A	x		x	x	x																							
Cz04g21260	N/A	x		x	x	x																							
Cz04g24170	N/A	x		x	x	x																							
Cz04g24300	N/A	x		x	x	x																							
Cz04g26060	N/A	x		x	x	x																							
Cz04g26200	N/A	x		x	x	x																							
Cz04g29200	N/A	x			x	x									x		x												
Cz04g30100	N/A	x		x	x	x		x	x	x	x				x	x	x												
Cz04g30160	N/A	x		x	x	x																							

Cz04g31010	N/A	x		x	x	x																							
Cz04g31070	N/A	x		x	x	x																							
Cz04g31190	N/A	x			x	x		x		x	x					x	x	x	x	x									
Cz04g32070	N/A	x		x	x	x																							
Cz04g33080	N/A	x		x	x	x																							
Cz04g33130	N/A	x			x			x		x			x	x	x	x	x	x	x	x									
Cz04g34020	N/A		x								x																		
Cz04g35200	N/A		x																										
Cz04g36060	N/A	x		x	x	x																							
Cz04g36140	N/A	x		x	x	x	x	x							x	x	x	x	x	x		x	x						
Cz04g38180	N/A	x		x	x	x																							
Cz04g40020	N/A	x		x	x	x																							
Cz05g00070	N/A	x		x	x	x																							
Cz05g00190	N/A	x		x	x	x																							
Cz05g01240	N/A	x		x	x	x					x																		
Cz05g02010	N/A	x			x																								
Cz05g04080	N/A	x		x	x	x																							
Cz05g04150	N/A		x																										
Cz05g04170	N/A	x		x	x	x																							
Cz05g04180	N/A		x																										
Cz05g04190	N/A		x																										
Cz05g06200	N/A		x								x																		
Cz05g09090	N/A		x																										
Cz05g09100	N/A	x		x	x	x				x	x				x														
Cz05g09280	N/A	x		x	x	x																							
Cz05g10190	N/A	x			x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x						
Cz05g10300	N/A	x		x	x	x	x		x																				
Cz05g11100	N/A	x		x	x	x																							
Cz05g11170	N/A	x		x	x	x																							

Cz05g15080	N/A	x		x	x	x																								
Cz05g15270	N/A	x		x	x	x																								
Cz05g16190	N/A	x		x	x	x																								
Cz05g17010	N/A	x		x	x	x																								
Cz05g18080	N/A	x		x	x	x																								
Cz05g20250	N/A	x		x	x	x																								
Cz05g21080	N/A		x																	x									
Cz05g21230	N/A		x																											
Cz05g22150	N/A	x		x	x	x																								
Cz05g24080	N/A	x			x			x			x	x	x			x		x	x	x	x								
Cz05g24150	N/A	x		x	x	x																								
Cz05g25180	N/A		x																											
Cz05g26010	N/A	x		x	x	x																								
Cz05g26120	N/A	x			x																									
Cz05g26170	N/A	x			x																									
Cz05g28030	N/A	x		x	x	x																								
Cz05g28210	N/A	x			x																									
Cz05g28220	N/A	x		x	x	x																								
Cz05g29140	N/A	x		x	x	x		x				x	x	x	x	x	x	x	x	x	x	x	x						
Cz05g29160	N/A		x																											
Cz05g30080	N/A	x		x	x	x																								
Cz05g31220	N/A	x		x	x	x																								
Cz05g32080	N/A	x		x	x	x																								
Cz05g32090	N/A	x		x	x	x																								
Cz05g32100	N/A	x		x	x	x																								
Cz05g32280	N/A	x			x																									
Cz05g32290	N/A	x		x	x	x																								
Cz05g34170	N/A	x		x	x	x																								
Cz05g34220	N/A	x		x	x	x																								

Cz05g36020	N/A	x			x	x																							
Cz05g36130	N/A	x			x			x		x			x			x		x	x	x	x	x	x						
Cz06g00200	N/A	x		x	x	x																							
Cz06g01100	N/A	x		x	x	x																							
Cz06g05030	N/A	x		x	x	x																							
Cz06g05220	N/A	x		x	x	x																							
Cz06g11080	N/A	x		x	x	x																							
Cz06g12190	N/A	x		x	x	x																							
Cz06g12240	N/A	x			x	x																							
Cz06g13190	N/A		x																										
Cz06g13290	N/A	x		x	x	x																							
Cz06g14190	N/A	x			x	x																							
Cz06g15040	N/A	x		x	x	x																							
Cz06g15190	N/A	x		x	x	x																							
Cz06g15200	N/A	x		x	x	x																							
Cz06g15210	N/A	x		x	x	x																							
Cz06g16060	N/A	x		x	x	x																							
Cz06g17210	N/A	x		x	x	x																							
Cz06g17220	N/A	x		x	x	x																							
Cz06g17230	N/A		x																										
Cz06g19080	N/A		x							x												x	x	x					
Cz06g20100	N/A		x																										
Cz06g21320	N/A	x		x	x	x																							
Cz06g22060	N/A	x		x	x	x																							
Cz06g22160	N/A	x		x	x	x																							
Cz06g22210	N/A		x																						x	x	x		
Cz06g26110	N/A	x		x	x	x																							
Cz06g27080	N/A	x			x																								
Cz06g30110	N/A	x			x		x		x																				

Cz06g30220	N/A	x		x	x	x																								
Cz06g31120	N/A	x		x	x	x																								
Cz06g31130	N/A	x		x	x	x		x						x	x	x	x	x		x										
Cz06g31140	N/A	x		x	x	x																								
Cz06g31230	N/A	x		x	x	x																								
Cz06g33080	N/A	x		x	x	x																								
Cz06g33100	N/A	x		x	x	x																								
Cz06g33110	N/A	x		x	x	x																								
Cz06g33190	N/A	x		x	x	x																								
Cz06g34100	N/A	x			x																									
Cz06g36270	N/A		x																											
Cz06g36290	N/A	x			x																									
Cz06g37190	N/A	x		x	x	x																								
Cz07g00160	N/A		x																											
Cz07g01040	N/A	x			x	x									x		x	x		x										
Cz07g01160	N/A	x		x	x	x																								
Cz07g01170	N/A	x		x	x	x																								
Cz07g03160	N/A	x		x	x	x																								
Cz07g03170	N/A	x		x	x	x																								
Cz07g05030	N/A	x		x	x	x																								
Cz07g05040	N/A	x		x	x	x																								
Cz07g07100	N/A		x																											
Cz07g07110	N/A	x		x	x	x																								
Cz07g07280	N/A	x		x	x	x																								
Cz07g08060	N/A	x		x	x	x																								
Cz07g08070	N/A	x		x	x	x																								
Cz07g08150	N/A		x																											
Cz07g08160	N/A		x																											
Cz07g08180	N/A		x																											

Cz07g08190	N/A		x																										
Cz07g11080	N/A	x		x	x	x																							
Cz07g12010	N/A	x		x	x	x																							
Cz07g12210	N/A	x		x	x	x																							
Cz07g14090	N/A	x			x		x	x	x				x	x		x	x	x	x	x									
Cz07g15190	N/A	x		x	x	x																							
Cz07g17220	N/A	x		x	x	x																							
Cz07g18040	N/A	x			x																								
Cz07g19040	N/A	x		x	x	x																							
Cz07g20040	N/A	x		x	x	x		x	x	x	x		x	x	x	x	x	x	x	x									
Cz07g21050	N/A	x		x	x	x																							
Cz07g22130	N/A	x		x	x	x																							
Cz07g22170	N/A		x																										
Cz07g28050	N/A	x		x	x	x									x	x	x			x									
Cz07g28100	N/A	x		x	x	x																							
Cz07g28110	N/A	x		x	x	x																							
Cz07g29010	N/A	x		x	x	x																							
Cz07g29030	N/A	x		x	x	x																							
Cz07g30200	N/A	x		x	x	x																							
Cz07g31160	N/A		x																										
Cz07g32070	N/A		x																										
Cz07g33160	N/A	x		x	x	x			x																				
Cz08g01110	N/A	x		x	x	x																							
Cz08g01190	N/A	x		x	x	x																							
Cz08g04060	N/A	x		x	x	x																							
Cz08g04120	N/A	x			x																								
Cz08g04130	N/A		x																										
Cz08g07210	N/A	x		x	x	x																							
Cz08g08110	N/A	x		x	x	x																							

Cz08g08170	N/A	x		x	x	x																								
Cz08g08190	N/A		x																											
Cz08g08240	N/A		x																											
Cz08g09120	N/A		x																											
Cz08g09130	N/A		x															x	x	x										
Cz08g09150	N/A	x		x	x	x																								
Cz08g11010	N/A	x		x	x	x																								
Cz08g11020	N/A	x		x	x	x	x		x		x				x		x													
Cz08g12110	N/A	x		x	x	x																								
Cz08g12270	N/A	x			x	x	x	x	x	x	x				x	x	x	x	x	x	x	x	x						
Cz08g13060	N/A	x			x	x																								
Cz08g14070	N/A	x		x	x	x																								
Cz08g15300	N/A	x		x	x	x																								
Cz08g17080	N/A		x																											
Cz08g17090	N/A	x		x	x	x																								
Cz08g17120	N/A	x		x	x	x																								
Cz08g19010	N/A		x						x																					
Cz08g19210	N/A	x		x	x	x																								
Cz08g20170	N/A	x		x	x	x																								
Cz08g22220	N/A	x		x	x	x																								
Cz08g23100	N/A	x			x	x																								
Cz08g25130	N/A	x		x	x	x																								
Cz08g25190	N/A	x		x	x	x																								
Cz08g27140	N/A	x		x	x	x																								
Cz08g27150	N/A		x																											
Cz08g27160	N/A		x																											
Cz08g28050	N/A	x		x	x	x																								
Cz08g28130	N/A	x		x	x	x																								
Cz08g28140	N/A	x		x	x	x																								

Cz08g29120	N/A	x		x	x	x																							
Cz08g29240	N/A	x		x	x	x																							
Cz08g29250	N/A	x		x	x	x																							
Cz08g30210	N/A	x		x	x	x																							
Cz08g32160	N/A	x		x	x	x																							
Cz09g00350	N/A	x		x	x	x																							
Cz09g01010	N/A	x		x	x	x																							
Cz09g01150	N/A	x		x	x	x																							
Cz09g06150	N/A	x		x	x	x																							
Cz09g06200	N/A		x						x	x	x																		
Cz09g07060	N/A	x			x			x					x			x	x	x	x	x	x	x	x	x					
Cz09g08080	N/A	x		x	x	x																							
Cz09g08090	N/A	x		x	x	x																							
Cz09g10050	N/A	x		x	x	x																							
Cz09g10060	N/A	x		x	x	x																							
Cz09g10070	N/A	x		x	x	x																							
Cz09g11020	N/A	x			x	x																							
Cz09g11030	N/A	x			x	x				x					x														
Cz09g11140	N/A	x		x	x	x																							
Cz09g11150	N/A	x		x	x	x																							
Cz09g12170	N/A	x		x	x	x																							
Cz09g13200	N/A	x		x	x	x																							
Cz09g13210	N/A	x		x	x	x																							
Cz09g17160	N/A	x			x								x			x			x										
Cz09g17300	N/A	x		x	x	x																							
Cz09g19020	N/A	x			x										x														
Cz09g19030	N/A	x		x	x	x																							
Cz09g22110	N/A	x		x	x	x		x					x	x	x	x	x	x	x	x									
Cz09g22160	N/A	x		x	x	x																							

Cz09g24080	N/A	x		x	x	x																								
Cz09g24270	N/A	x		x	x	x																								
Cz09g25160	N/A	x			x																									
Cz09g26060	N/A	x		x	x	x																								
Cz09g26180	N/A	x		x	x	x																								
Cz09g27120	N/A	x		x	x	x																								
Cz09g28080	N/A		x																											
Cz09g29180	N/A	x		x	x	x																								
Cz09g30060	N/A	x			x																									
Cz10g02090	N/A	x		x	x	x																								
Cz10g02250	N/A	x			x			x					x			x			x		x	x	x	x		x				
Cz10g03100	N/A	x		x	x	x																								
Cz10g06080	N/A		x																											
Cz10g08020	N/A	x		x	x	x																								
Cz10g08080	N/A	x		x	x	x																								
Cz10g09190	N/A		x									x										x	x	x						
Cz10g09260	N/A		x																											
Cz10g10070	N/A		x																							x				
Cz10g10230	N/A	x		x	x	x		x					x	x	x	x	x	x	x	x										
Cz10g10250	N/A	x		x	x	x																								
Cz10g11150	N/A		x																x											
Cz10g11200	N/A	x			x																									
Cz10g11210	N/A	x		x	x	x																								
Cz10g14320	N/A	x		x	x	x																								
Cz10g15090	N/A	x		x	x	x				x					x															
Cz10g19120	N/A	x		x	x	x																								
Cz10g20280	N/A		x																											
Cz10g24040	N/A	x		x	x	x																								
Cz10g24050	N/A	x		x	x	x																								

Cz10g26110	N/A		x																										
Cz10g26120	N/A		x																										
Cz10g26170	N/A	x		x	x	x																							
Cz10g28090	N/A	x		x	x	x																							
Cz10g28100	N/A	x		x	x	x																							
Cz10g28160	N/A	x		x	x	x																							
Cz10g29030	N/A		x																					x					
Cz11g03190	N/A	x			x																								
Cz11g03280	N/A	x		x	x	x																							
Cz11g04220	N/A	x		x	x	x																							
Cz11g04250	N/A	x		x	x	x																							
Cz11g05030	N/A		x																										
Cz11g05190	N/A	x		x	x	x																							
Cz11g07070	N/A	x		x	x	x																							
Cz11g09340	N/A	x		x	x	x																							
Cz11g10100	N/A	x			x						x																		
Cz11g10110	N/A	x		x	x	x																							
Cz11g10120	N/A	x		x	x	x																							
Cz11g10130	N/A	x		x	x	x																							
Cz11g11020	N/A		x																										
Cz11g11070	N/A		x																					x	x	x			
Cz11g12230	N/A		x																										
Cz11g15130	N/A		x																										
Cz11g16120	N/A	x		x	x	x																							
Cz11g17200	N/A	x		x	x	x																							
Cz11g19030	N/A	x		x	x	x																							
Cz11g19120	N/A	x		x	x	x																							
Cz11g19140	N/A		x																										
Cz11g20020	N/A	x			x													x		x									

Cz11g20130	N/A	x		x	x	x																								
Cz11g20190	N/A	x		x	x	x																								
Cz11g20280	N/A		x																											
Cz11g23180	N/A	x		x	x	x																								
Cz11g24090	N/A	x		x	x	x																								
Cz11g24170	N/A	x		x	x	x																								
Cz11g26100	N/A	x		x	x	x									x		x													
Cz11g26110	N/A	x		x	x	x								x	x		x	x		x										
Cz11g26160	N/A	x		x	x	x																								
Cz12g02250	N/A	x		x	x	x																								
Cz12g02260	N/A	x		x	x	x																								
Cz12g03020	N/A	x		x	x	x																								
Cz12g03030	N/A	x		x	x	x																								
Cz12g05150	N/A	x		x	x	x																								
Cz12g07220	N/A	x		x	x	x																								
Cz12g08070	N/A	x		x	x	x																								
Cz12g08090	N/A		x																											
Cz12g08100	N/A	x		x	x	x																								
Cz12g08200	N/A	x		x	x	x																								
Cz12g09070	N/A	x		x	x	x																								
Cz12g10160	N/A	x		x	x	x																								
Cz12g10190	N/A	x		x	x	x																								
Cz12g11010	N/A	x				x	x								x	x	x	x	x	x										
Cz12g12080	N/A	x		x	x	x																								
Cz12g12270	N/A	x		x	x	x																								
Cz12g13090	N/A	x		x	x	x																								
Cz12g14130	N/A	x		x	x	x																								
Cz12g15010	N/A	x		x	x	x																								
Cz12g15040	N/A	x		x	x	x																								

Cz12g15050	N/A	x		x	x	x																							
Cz12g15240	N/A	x		x	x	x																							
Cz12g20020	N/A	x		x	x	x																							
Cz12g21040	N/A	x		x	x	x																							
Cz12g24020	N/A	x		x	x	x																							
Cz12g24200	N/A		x																						x	x	x		
Cz12g25170	N/A	x		x	x	x																							
Cz12g25180	N/A		x																										
Cz12g27090	N/A		x																										
Cz12g28210	N/A	x		x	x	x																							
Cz13g00060	N/A	x		x	x	x																							
Cz13g00070	N/A	x		x	x	x																							
Cz13g00190	N/A	x		x	x	x																							
Cz13g00220	N/A	x		x	x	x																							
Cz13g01110	N/A	x		x	x	x		x	x		x			x	x		x	x		x								
Cz13g01160	N/A	x		x	x	x																							
Cz13g04150	N/A	x		x	x	x																							
Cz13g04250	N/A	x		x	x	x																							
Cz13g05080	N/A	x		x	x	x																							
Cz13g08015	N/A	x			x			x		x			x	x	x	x	x	x	x	x									
Cz13g08170	N/A	x		x	x	x																							
Cz13g11160	N/A	x		x	x	x																							
Cz13g11170	N/A		x																										
Cz13g12100	N/A	x		x	x	x																							
Cz13g13130	N/A	x		x	x	x																							
Cz13g13150	N/A	x		x	x	x																							
Cz13g14050	N/A	x		x	x	x																							
Cz13g14150	N/A	x		x	x	x																							
Cz13g15210	N/A	x		x	x	x																							

Cz13g17110	N/A	x			x					x																				
Cz13g17170	N/A	x		x	x	x																								
Cz13g20030	N/A	x		x	x	x																								
Cz13g20040	N/A	x		x	x	x			x																					
Cz14g03080	N/A	x		x	x	x																								
Cz14g05040	N/A	x		x	x	x																								
Cz14g05140	N/A	x		x	x	x																								
Cz14g09020	N/A	x		x	x	x																								
Cz14g09130	N/A	x		x	x	x	x	x	x	x	x				x	x	x	x	x	x										
Cz14g10050	N/A	x		x	x	x																								
Cz14g10070	N/A	x			x										x		x													
Cz14g10110	N/A	x		x	x	x																								
Cz14g11270	N/A	x		x	x	x		x	x		x				x		x	x		x										
Cz14g12150	N/A	x		x	x	x																								
Cz14g13070	N/A		x																											
Cz14g14210	N/A		x																											
Cz14g17180	N/A	x		x	x	x																								
Cz14g18070	N/A	x		x	x	x																								
Cz14g18130	N/A	x			x	x																								
Cz14g18250	N/A	x		x	x	x																								
Cz14g18280	N/A	x		x	x	x																								
Cz14g19040	N/A	x		x	x	x																								
Cz14g19080	N/A	x		x	x	x																								
Cz14g21080	N/A	x		x	x	x																								
Cz14g23040	N/A	x		x	x	x																								
Cz14g23250	N/A	x		x	x	x																								
Cz14g23270	N/A	x		x	x	x																								
Cz14g23320	N/A		x																											
Cz14g24160	N/A	x		x	x	x																								

Cz14g24170	N/A		x																										
Cz15g01140	N/A	x		x	x	x																							
Cz15g02090	N/A	x		x	x	x																							
Cz15g07110	N/A	x		x	x	x																							
Cz15g07120	N/A	x		x	x	x																							
Cz15g07260	N/A	x		x	x	x																							
Cz15g08030	N/A	x			x					x	x							x											
Cz15g08240	N/A	x		x	x	x	x	x	x	x	x				x	x	x												
Cz15g09220	N/A	x		x	x	x																							
Cz15g09230	N/A	x		x	x	x																							
Cz15g10010	N/A	x			x																								
Cz15g10080	N/A	x		x	x	x																							
Cz15g12110	N/A	x		x	x	x																							
Cz15g16010	N/A	x		x	x	x																							
Cz15g17090	N/A	x		x	x	x																							
Cz15g20030	N/A		x																										
Cz15g21300	N/A	x		x	x	x																							
Cz15g23040	N/A	x			x	x																							
Cz15g23070	N/A	x			x			x	x	x	x	x	x			x			x		x	x	x						
Cz16g01010	N/A	x			x																								
Cz16g01030	N/A	x		x	x	x																							
Cz16g01100	N/A	x		x	x	x																							
Cz16g01110	N/A	x		x	x	x																							
Cz16g01130	N/A	x		x	x	x																							
Cz16g01140	N/A	x		x	x	x	x		x	x	x																		
Cz16g03180	N/A	x		x	x	x																							
Cz16g04250	N/A	x		x	x	x																							
Cz16g05120	N/A	x		x	x	x																							
Cz16g05130	N/A	x		x	x	x																							

Cz16g05180	N/A		x							x	x													x	x	x			x
Cz16g05240	N/A	x			x																									
Cz16g07120	N/A	x			x																									
Cz16g07130	N/A		x																											
Cz16g07140	N/A		x																											
Cz16g07190	N/A	x		x	x	x																								
Cz16g07340	N/A	x		x	x	x																								
Cz16g09270	N/A	x		x	x	x																								
Cz16g10020	N/A		x																											
Cz16g10170	N/A	x		x	x	x																								
Cz16g11010	N/A	x			x	x								x		x														
Cz16g12010	N/A	x		x	x	x																								
Cz16g12070	N/A	x		x	x	x		x		x			x	x	x	x	x	x	x	x										
Cz16g13060	N/A	x		x	x	x																								
Cz16g13130	N/A	x		x	x	x																								
Cz16g14110	N/A	x			x																									
Cz16g19120	N/A		x																											
Cz16g21020	N/A	x		x	x	x																								
Cz16g22020	N/A	x		x	x	x								x		x														
Cz17g02010	N/A	x		x	x	x																								
Cz17g13190	N/A	x		x	x	x																								
Cz17g14040	N/A	x		x	x	x																								
Cz17g14150	N/A	x			x																									
Cz17g17130	N/A	x		x	x	x																								
Cz18g00170	N/A	x			x																									
Cz18g04010	N/A	x		x	x	x																								
Cz18g04060	N/A	x		x	x	x																								
Cz18g04130	N/A		x																									x	x	x
Cz18g10020	N/A	x		x	x	x																								

Cz18g10180	N/A	x		x	x	x																							
Cz18g11080	N/A		x																										
Cz18g11090	N/A		x																										
Cz19g01190	N/A	x		x	x	x																							
Cz19g02190	N/A	x		x	x	x																							
Cz19g03010	N/A	x			x																								
Cz19g04040	N/A	x		x	x	x																							
Cz19g05230	N/A	x		x	x	x																							
Cz19g06170	N/A	x		x	x	x																							
Cz19g07080	N/A	x		x	x	x																							
Cz19g09090	N/A	x		x	x	x																							
UNPLg00003	N/A	x			x																								
UNPLg00023	N/A	x		x	x	x	x	x						x	x	x	x	x	x	x									
UNPLg00024	N/A	x			x		x	x					x		x	x	x	x	x	x	x	x	x	x	x	x			
UNPLg00129	N/A	x			x								x			x		x	x	x									
UNPLg00177	N/A		x																						x		x		
UNPLg00188	N/A	x		x	x	x																							
UNPLg00189	N/A	x		x	x	x																							
UNPLg00190	N/A	x		x	x	x																							
UNPLg00207	N/A	x			x		x																						
UNPLg00262	N/A	x		x	x	x																							
UNPLg00275	N/A	x		x	x	x																							
UNPLg00281	N/A	x		x	x	x																							
UNPLg00282	N/A	x		x	x	x																							
UNPLg00285	N/A	x		x	x	x																							
UNPLg00286	N/A	x			x																								
UNPLg00288	N/A		x																										
UNPLg00290	N/A	x		x	x	x																							
UNPLg00294	N/A	x		x	x	x																							

UNPLg00298	N/A	x		x	x	x																							
UNPLg00299	N/A	x			x																								
UNPLg00301	N/A		x																										
UNPLg00303	N/A	x		x	x	x																							
UNPLg00344	N/A	x		x	x	x																							
UNPLg00401	N/A	x		x	x	x																							
UNPLg00440	N/A	x		x	x	x																							
UNPLg00445	N/A	x			x		x	x				x	x	x	x	x	x	x	x	x	x	x	x						
UNPLg00492	N/A	x			x	x		x		x					x	x	x	x	x	x									
UNPLg00650	N/A	x		x	x	x																							
UNPLg00662	N/A	x		x	x	x																							
UNPLg00663	N/A	x		x	x	x																							
UNPLg00673	N/A	x			x																								
UNPLg00703	N/A	x		x	x	x																							
UNPLg00705	N/A	x		x	x	x																							
UNPLg00747	N/A	x		x	x	x																							

References

Accadia Di, F., Gribanovski-Sassu, O., Romagnoli, A. & Tuttobello, L. (1966). Isolation and identification of carotenoids produced by a green alga (Dictyococcus cinnabarinus) in submerged culture . *Biochemical Journal*, *101*(3), 735–740. https://doi.org/10.1042/bj1010735

Adam, Z., Charuvi, D., Tsabari, O., Knopf, R. R. & Reich, Z. (2011). Biogenesis of thylakoid networks in angiosperms: Knowns and unknowns. *Plant Molecular Biology*, *76*(3–5), 221–234. https://doi.org/10.1007/s11103-010-9693-5

Ahuatzi, D., Herrero, P., De La Cera, T. & Moreno, F. (2004). The Glucose-regulated Nuclear Localization of Hexokinase 2 in Saccharomyces cerevisiae Is Mig1-dependent. *Journal of Biological Chemistry*, *279*(14), 14440–14446. https://doi.org/10.1074/jbc.M313431200

Ahuatzi, D., Riera, A., Peláez, R., Herrero, P. & Moreno, F. (2007). Hxk2 regulates the phosphorylation state of Mig1 and therefore its nucleocytoplasmic distribution. *Journal of Biological Chemistry*, *282*(7), 4485–4493. https://doi.org/10.1074/jbc.M606854200

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amin, P., Sy, D. A. C., Pilgrim, M. L., Parry, D. H., Nussaume, L. & Hoffman, N. E. (1999). Arabidopsis mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes. *Plant Physiology*, *121*(1), 61–70. https://doi.org/10.1104/pp.121.1.61

Armbruster, U., Labs, M., Pribil, M., Viola, S., Xu, W., Scharfenberg, M., Hertle, A. P., Rojahn, U., Jensen, P. E., Rappaport, F., Joliot, P., Dörmann, P., Wanner, G. & Leister, D. (2013). Arabidopsis CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. *Plant Cell*, *25*(7), 2661–2678. https://doi.org/10.1105/tpc.113.113118

Asakura, Y., Kikuchi, S. & Nakai, M. (2008). Non-identical contributions of two membrane-bound cpSRP components, cpFtsY and Alb3, to thylakoid biogenesis. *Plant Journal*, *56*(6), 1007–1017. https://doi.org/10.1111/j.1365-313X.2008.03659.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Austin, R. S., Hiu, S., Waese, J., Ierullo, M., Pasha, A., Wang, T. T., Fan, J., Foong, C., Breit, R., Desveaux, D., Moses, A. & Provart, N. J. (2016). New BAR tools for mining expression data and exploring Cis-elements in Arabidopsis thaliana. *Plant Journal*, *88*(3), 490–504. https://doi.org/10.1111/tpj.13261

Awai, K., Maréchal, E., Block, M. A., Brun, D., Masuda, T., Shimada, H., Takamiya, K. I., Ohta, H. & Joyard, J. (2001). Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(19), 10960–10965. https://doi.org/10.1073/pnas.181331498

Bae, W., Lee, Y. J., Kim, D. H., Lee, J., Kim, S., Sohn, E. J. & Hwang, I. (2008). AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. *Nature Cell Biology*, *10*(2), 220–227. https://doi.org/10.1038/ncb1683

Baena-González, E., Rolland, F., Thevelein, J. M. & Sheen, J. (2007). A central integrator of transcription networks in plant stress and energy signalling. *Nature*, *448*(7156), 938–942. https://doi.org/10.1038/nature06069

Balasubramanian, R., Karve, A., Kandasamy, M., Meagher, R. B. & Moore, B. D. (2007). A role for F-actin in hexokinase-mediated glucose signaling. *Plant Physiology*, *145*(4), 1423–1434. https://doi.org/10.1104/pp.107.108704

Balasubramanian, R., Karve, A. & Moore, B. D. (2008). Actin-based cellular framework for glucose signaling by Arabidopsis hexokinase1. *Plant Signaling and Behavior*, *3*(5), 322–324. https://doi.org/10.4161/psb.3.5.5319

Bang, W. Y., Hata, A., Jeong, I. S., Umeda, T., Masuda, T., Chen, J., Yoko, I., Suwastika, I. N., Kim, D. W., Im, C. H., Lee, B. H., Lee, Y., Lee, K. W., Shiina, T. & Bahk, J. D. (2009). AtObgC, a plant ortholog of bacterial Obg, is a chloroplast-targeting GTPase essential for early embryogenesis. *Plant Molecular Biology*, *71*(4–5), 379–390. https://doi.org/10.1007/s11103-009-9529-3

Bar, E., Rise, M., Vishkautsan, M. & Arad, S. (Malis). (1995). Pigment and Structural Changes in Chlorella zofingiensis upon Light and Nitrogen Stress. *Journal of Plant Physiology*, *146*(4), 527–534. https://doi.org/10.1016/S0176-1617(11)82019-5

Bassel, G. W., Fung, P., Chow, T. F. F., Foong, J. A., Provart, N. J. & Cutler, S. R. (2008). Elucidating the germination transcriptional program using small molecules. *Plant Physiology*, *147*(1), 143–155. https://doi.org/10.1104/pp.107.110841

Bassi, R. & Wollman, F.-A. (1991). The chlorophyll-a/b proteins of photosystem II in Chlamydomonas reinhardtii. *Planta*, *183*(3), 423–433. https://doi.org/10.1007/BF00197742

Bastien, O., Botella, C., Chevalier, F., Block, M. A., Jouhet, J., Breton, C., Girard-Egrot, A. & Maréchal, E. (2016). New Insights on Thylakoid Biogenesis in Plant Cells. In *International Review of Cell and Molecular Biology* (Vol. 323). Elsevier Inc. https://doi.org/10.1016/bs.ircmb.2015.12.001

Benz, J. P., Stengel, A., Lintala, M., Lee, Y. H., Weber, A., Philippar, K., Gügel, I. L.,

Kaieda, S., Ikegami, T., Mulo, P., Soll, J. & Bölter, B. (2009). Arabidopsis Tic62 and ferredoxin-NADP(H) oxidoreductase form light-regulated complexes that are integrated into the chloroplast redox poise. *Plant Cell*, *21*(12), 3965–3983. https://doi.org/10.1105/tpc.109.069815

Bielczynski, L. W., Schansker, G. & Croce, R. (2020). Consequences of the reduction of the Photosystem II antenna size on the light acclimation capacity of Arabidopsis thaliana. *Plant Cell and Environment*, *43*(4), 866–879. https://doi.org/10.1111/pce.13701

Boussiba, S. (2000). Carotenogenesis in the green alga Haematococcus pluvialis: Cellular physiology and stress response. *Physiologia Plantarum*, *108*(2), 111–117. https://doi.org/10.1034/j.1399-3054.2000.108002111.x

Bown, A. W. & Shelp, B. J. (2016). Plant GABA: Not Just a Metabolite. *Trends in Plant Science*, *21*(10), 811–813. https://doi.org/10.1016/j.tplants.2016.08.001

Breuer, G., Evers, W. A. C., de Vree, J. H., Kleinegris, D. M. M., Martens, D. E., Wijffels, R. H. & Lamers, P. P. (2013). Analysis of fatty acid content and composition in microalgae. *Journal of Visualized Experiments : JoVE*, *5*(80), 1–9. https://doi.org/10.3791/50628

Breuer, G., Lamers, P. P., Martens, D. E., Draaisma, R. B. & Wijffels, R. H. (2012a). Bioresource Technology The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresource Technology*, *124*, 217–226. https://doi.org/10.1016/j.biortech.2012.08.003

Breuer, G., Lamers, P. P., Martens, D. E., Draaisma, R. B. & Wijffels, R. H. (2012b). The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresource Technology*, *124*, 217–226. https://doi.org/10.1016/j.biortech.2012.08.003

Buchfink, B., Xie, C. & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bučinská, L., Kiss, É., Koník, P., Knoppová, J., Komenda, J. & Sobotka, R. (2018). The ribosome-bound protein pam68 promotes insertion of chlorophyll into the CP47 subunit of photosystem II. *Plant Physiology*, *176*(4), 2931–2942. https://doi.org/10.1104/pp.18.00061

Capelli, B., Bagchi, D. & Cysewski, G. R. (2013). Synthetic astaxanthin is significantly inferior to algal-based astaxanthin as an antioxidant and may not be suitable as a human nutraceutical supplement. *Nutrafoods*, *12*(4), 145–152. https://doi.org/10.1007/s13749-013-0051-5

Cazzaniga, S., Dall'Osto, L., Kong, S. G., Wada, M. & Bassi, R. (2013). Interaction between avoidance of photon absorption, excess energy dissipation and zeaxanthin synthesis against photooxidative stress in Arabidopsis. *Plant Journal*, *76*(4), 568–579. https://doi.org/10.1111/tpj.12314

Charuvi, D., Kiss, V., Nevo, R., Shimoni, E., Adam, Z. & Reich, Z. (2012). *Gain and Loss of Photosynthetic Membranes during Plastid Differentiation in the Shoot Apex of Arabidopsis*. *24*(March), 1143–1157. https://doi.org/10.1105/tpc.111.094458

Chemeris, Y. K., Heck, O. J., Vasilev, S. S. & Venediktov, P. S. (1991). INACTIVATION OF PHOTOSYSTEM-II IN THE PROCESS OF HETEROTROPHIC GROWTH IN CHLORELLA. *SOVIET PLANT PHYSIOLOGY*, *38*(3), 322–328.

Chen, J. hui, Wei, D. & Lim, P. E. (2020). Enhanced coproduction of astaxanthin and lipids by the green microalga Chromochloris zofingiensis: Selected phytohormones as positive stimulators. *Bioresource Technology*, *295*(August 2019), 122242. https://doi.org/10.1016/j.biortech.2019.122242

Chen, Tao & Wang, Y. (2013). Optimized astaxanthin production in Chlorella zofingiensis under dark condition by response surface methodology. *Food Science and Biotechnology*, *22*(5), 1–8. https://doi.org/10.1007/s10068-013-0221-7

Chen, Tianpeng, Liu, J., Guo, B., Ma, X., Sun, P., Liu, B. & Chen, F. (2015). Light attenuates lipid accumulation while enhancing cell proliferation and starch synthesis in the glucose-fed oleaginous microalga Chlorella zofingiensis. *Scientific Reports*, *5*(October), 1–10. https://doi.org/10.1038/srep14936

Chigri, F., Sippel, C., Kolb, M. & Vothknecht, U. C. (2009). Arabidopsis OBG-Like GTPase (AtOBGL) is localized in chloroplasts and has an essential function in embryo development. *Molecular Plant*, *2*(6), 1373–1383. https://doi.org/10.1093/mp/ssp073

Cho, Y. H., Yoo, S. D. & Sheen, J. (2006). Regulatory Functions of Nuclear Hexokinase1 Complex in Glucose Signaling. *Cell*, *127*(3), 579–589. https://doi.org/10.1016/j.cell.2006.09.028

Chory, J., Donohue, T. J. & Varga, A. R. (1984). Biochemical and morphological studies of the induction of the photosynthetic membrane of Rhodopseudomonas sphaeroides. *Federation Proceedings*, *43*(6), 540–554.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Claeyssen, É. & Rivoal, J. (2007). Isozymes of plant hexokinase: Occurrence, properties and functions. *Phytochemistry*, *68*(6), 709–731. https://doi.org/10.1016/j.phytochem.2006.12.001

Cui, Y. L., Jia, Q. S., Yin, Q. Q., Lin, G. N., Kong, M. M. & Yang, Z. N. (2011). The GDC1 gene encodes a novel ankyrin domain-containing protein that is essential for grana formation in arabidopsis. *Plant Physiology*, *155*(1), 130–141. https://doi.org/10.1104/pp.110.165589

Cutolo, E., Parvin, N., Ruge, H., Pirayesh, N., Roustan, V., Weckwerth, W., Teige, M., Grieco, M., Larosa, V. & Vothknecht, U. C. (2019). The High Light Response in Arabidopsis Requires the Calcium Sensor Protein CAS, a Target of STN7- and STN8-Mediated Phosphorylation. *Frontiers in Plant Science*, *10*(July), 1–15. https://doi.org/10.3389/fpls.2019.00974

Dall'Osto, L., Bressan, M. & Bassi, R. (2015). Biogenesis of light harvesting proteins.

Biochimica et Biophysica Acta - Bioenergetics, *1847*(9), 861–871. https://doi.org/10.1016/j.bbabio.2015.02.009

del Campo Castillo, J. A. (2000). *Selección de Estirpes De Microalgas Para La Producción de Carotenoides de Interés Comercial. Producción de Luteína Por la Clorofícea Muriellopsis sp.*

Del Campo, J. A., Rodríguez, H., Moreno, J., Vargas, M. Á., Rivas, J. & Guerrero, M. G. (2004). Accumulation of astaxanthin and lutein in Chlorella zofingiensis (Chlorophyta). *Applied Microbiology and Biotechnology*, *64*(6), 848–854. https://doi.org/10.1007/s00253-003-1510-5

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

di Accadia, F. D., Gribanovski-Sassu, O., Romagnoli, A. & Tuttobello, L. (1965). Production of carotenoid by a green alga. *Nature*, *208*(5017), 1342.

Di Rienzi, S. C., Sharon, I., Wrighton, K. C., Koren, O., Hug, L. A., Thomas, B. C., Goodrich, J. K., Bell, J. T., Spector, T. D., Banfield, J. F. & Ley, R. E. (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *ELife*, *2*, 1–25. https://doi.org/10.7554/elife.01102

Dönz, C. (1934). Chlorella zofingiensis, eine neue Bodenalga. *Ber. Schweiz. Bot. Ges*, *43*, 127–131.

Duanmu, D., Casero, D., Dent, R. M., Gallaher, S., Yang, W., Rockwell, N. C., Martin, S. S., Pellegrini, M., Niyogi, K. K., Merchant, S. S., Grossman, A. R. & Lagarias, J. C. (2013). Retrograde bilin signaling enables Chlamydomonas greening and phototrophic survival. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3621–3626. https://doi.org/10.1073/pnas.1222375110

Eny, D. M. (1950). Respiration Studies on Chlorella. I. Growth Experiments With Acid Intermediates. *Plant Physiology*, *25*(3), 478–495. https://doi.org/10.1104/pp.25.3.478

Eppley, R. W. & MaciasR, F. . (1962). Rapid Growth of Sewage Lagoon Chlamydomonas with Aeetate. *Physiologia Plantarum*, *15*, 72–79.

Falk, S., Ravaud, S., Koch, J. & Sinning, I. (2010). The C terminus of the Alb3 membrane insertase recruits cpSRP43 to the thylakoid membrane. *Journal of Biological Chemistry*, *285*(8), 5954–5962. https://doi.org/10.1074/jbc.M109.084996

Feng, J., Zhao, S., Chen, X., Wang, W., Dong, W., Chen, J., Shen, J. R., Liu, L. & Kuang, T. (2015). Biochemical and structural study of Arabidopsis hexokinase 1. *Acta Crystallographica Section D: Biological Crystallography*, *71*, 367–375. https://doi.org/10.1107/S1399004714026091

Fernández-García, P., Peláez, R., Herrero, P. & Moreno, F. (2012). Nuclear import of the yeast hexokinase 2 protein requires α/β-importin-dependent pathway. *Journal of Biological Chemistry*, *287*(5), 3518–3529. https://doi.org/10.1074/jbc.M111.317230

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, *281*(5374), 237 LP – 240. https://doi.org/10.1126/science.281.5374.237

Filippis, L. F. De, Hampp, R., Ziegler, H., Filippis, L. F. D. E., Hampp, R. & Ziegler, H. (1980). *Protoplasts as a Means of Studying Chloroplast Development In vitro Published by : American Society of Plant Biologists (ASPB) Linked references are available on JSTOR for this article : Protoplasts as a Means of Studying Chloroplast Development in*. *66*(1), 1–7.

Findinier, J., Delevoye, C. & Cohen, M. M. (2019). The dynamin-like protein fzl promotes thylakoid fusion and resistance to light stress in chlamydomonas reinhardtii. *PLoS Genetics*, *15*(3), 1–30. https://doi.org/10.1371/journal.pgen.1008047

Finn, R. D., Clements, J. & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl_2), W29–W37. https://doi.org/10.1093/nar/gkr367

Flechtner, V. R., Johansen, J. R. & Clark, W. H. (1998). Algal composition of microbiotic crusts from the Central Desert of Raja California, Mexico. *Great Basin Naturalist*, *58*(4), 295–311.

Froehlich, J. E., Benning, C. & Dörmann, P. (2001). The Digalactosyldiacylglycerol (DGDG) Synthase DGD1 Is Inserted into the Outer Envelope Membrane of Chloroplasts in a Manner Independent of the General Import Pathway and Does Not Depend on Direct Interaction with Monogalactosyldiacylglycerol Synthase for. *Journal of Biological Chemistry*, *276*(34), 31806–31812. https://doi.org/10.1074/jbc.M104652200

Froehlich, J. E. & Keegstra, K. (2011). The role of the transmembrane domain in determining the targeting of membrane proteins to either the inner envelope or thylakoid membrane. *Plant Journal*, *68*(5), 844–856. https://doi.org/10.1111/j.1365-313X.2011.04735.x

Fučíková, K. & Lewis, L. A. (2012). Intersection of Chlorella, Muriella and Bracteacoccus: Resurrecting the genus Chromochloris Kol et Chodat (Chlorophyceae, Chlorophyta). *Fottea*, *12*(1), 83–93. https://doi.org/10.5507/fot.2012.007

Gallaher, S. D. & Roth, M. S. (2018). RNA Purification from the Unicellular Green Alga, Chromochloris zofingiensis. *Bio-Protocol*, *8*(7), e2793. https://doi.org/10.21769/BioProtoc.2793

Gao, H., Sage, T. L. & Osteryoung, K. W. (2006). FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(17), 6759–6764.

https://doi.org/10.1073/pnas.0507287103

García-Cerdán, J. G., Schmid, E. M., Takeuchi, T., McRae, I., McDonald, K. L., Yordduangjun, N., Hassan, A. M., Grob, P., Xu, C. S., Hess, H. F., Fletcher, D. A., Nogales, E. & Niyogi, K. K. (2020). Chloroplast Sec14-like 1 (CPSFL1) is essential for normal chloroplast development and affects carotenoid accumulation in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(22), 12452–12463. https://doi.org/10.1073/pnas.1916948117

Garcia, C., Khan, N. Z., Nannmark, U., Aronsson, H. & Academy, T. S. (2010). *The chloroplast protein CPSAR1 , dually localized in the stroma and the inner envelope membrane , is involved in thylakoid biogenesis*. 73–85. https://doi.org/10.1111/j.1365-313X.2010.04225.x

Gügel, I. L. & Soll, J. (2017). Chloroplast differentiation in the growing leaves of Arabidopsis thaliana. *Protoplasma*, *254*(5), 1857–1866. https://doi.org/10.1007/s00709-016-1057-9

Harris, E. (Ed.). (2009). *The Chlamydomonas Sourcebook* (2nd ed.). Academic Press.

Hauser, M., Steinegger, M. & Söding, J. (2016). MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics*, *32*(9), 1323–1330. https://doi.org/10.1093/bioinformatics/btw006

Heinz, S., Rast, A., Shao, L., Gutu, A., Gügel, I. L., Heyno, E., Labs, M., Rengstl, B., Viola, S., Nowaczyk, M. M., Leister, D. & Nickelsen, J. (2016). Thylakoid membrane architecture in Synechocystis depends on CurT, a homolog of the granal CURVATURE THYLAKOID1 proteins. *Plant Cell*, *28*(9), 2238–2260. https://doi.org/10.1105/tpc.16.00491

Hertle, A. P., Blunder, T., Wunder, T., Pesaresi, P., Pribil, M., Armbruster, U. & Leister, D. (2013). PGRL1 Is the Elusive Ferredoxin-Plastoquinone Reductase in Photosynthetic Cyclic Electron Flow. *Molecular Cell*, *49*(3), 511–523. https://doi.org/10.1016/j.molcel.2012.11.030

Hertle, A. P., García-Cerdán, J. G., Armbruster, U., Shih, R., Lee, J. J., Wong, W. & Niyogi, K. K. (2020). A Sec14 domain protein is required for photoautotrophic growth and chloroplast vesicle formation in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(16), 9101–9111. https://doi.org/10.1073/pnas.1916946117

Hong, Z., Bednarek, S. Y., Blumwald, E., Hwang, I., Jurgens, G., Menzel, D., Osteryoung, K. W., Raikhel, N. V., Shinozaki, K., Tsutsumi, N. & Verma, D. P. S. (2003). A unified nomenclature for Arabidopsis dynamin-related large GTPases based on homology and possible functions. *Plant Molecular Biology*, *53*(3), 261–265. https://doi.org/10.1023/B:PLAN.0000007000.29697.81

Hornung, R. L., Hanzely, L. & Lynch, D. L. (1977). Occurrence of microbodies in the green algaBracteacoccus cinnabarinus grown heterotrophically. *Protoplasma*, *93*(2–3), 135–145.

Hristou, A., Gerlach, I., Stolle, D. S., Neumann, J., Bischoff, A., Dünschede, B.,

Nowaczyk, M. M., Zoschke, R. & Schünemann, D. (2019). Ribosome-Associated Chloroplast SRP54 Enables Efficient Cotranslational Membrane Insertion of Key Photosynthetic Proteins. *The Plant Cell*, *31*(11), 2734–2750. https://doi.org/10.1105/tpc.19.00169

Huang, J., Taylor, J. P., Chen, J. G., Uhrig, J. F., Schnell, D. J., Nakagawa, T., Korth, K. L. & Jones, A. M. (2006). The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in Arabidopsis. *Plant Cell*, *18*(5), 1226–1238. https://doi.org/10.1105/tpc.105.037259

Huang, S. S., Chen, J., Dong, X. J., Patton, J., Pei, Z. M. & Zheng, H. L. (2012). Calcium and calcium receptor CAS promote Arabidopsis thaliana de-etiolation. *Physiologia Plantarum*, *144*(1), 73–82. https://doi.org/10.1111/j.1399-3054.2011.01523.x

Huang, W., Ye, J., Zhang, J., Lin, Y., He, M. & Huang, J. (2016). Transcriptome analysis of Chlorella zofingiensis to identify genes and their expressions involved in astaxanthin and triacylglycerol biosynthesis. *Algal Research*, *17*, 236–243. https://doi.org/10.1016/j.algal.2016.05.015

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C. & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Ip, P. F. & Chen, F. (2005). Production of astaxanthin by the green microalga Chlorella zofingiensis in the dark. *Process Biochemistry*, *40*(2), 733–738. https://doi.org/10.1016/j.procbio.2004.01.039

Ip, P. F., Wong, K. H. & Chen, F. (2004). Enhanced production of astaxanthin by the green microalga Chlorella zofingiensis in mixotrophic culture. *Process Biochemistry*, *39*(11), 1761–1766. https://doi.org/10.1016/j.procbio.2003.08.003

Ishikawa, M., Hirai, S., Yoshida, T., Shibuya, N., Hama, S., Takahashi, Y., Fukuta, T., Tanaka, T., Hosoi, S. & Kogure, K. (2018). Carotenoid stereochemistry affects antioxidative activity of liposomes co-encapsulating astaxanthin and tocotrienol. *Chemical and Pharmaceutical Bulletin*, *66*(7), 714–720. https://doi.org/10.1248/cpb.c18-00035

Iwai, M., Roth, M. S. & Niyogi, K. K. (2018). Subdiffraction-resolution live-cell imaging for visualizing thylakoid membranes. *The Plant Journal*, *96*(1), 233–243. https://doi.org/https://doi.org/10.1111/tpj.14021

Jang, J. C. & Sheen, J. (1994). Sugar Sensing in Higher Plants\n10.1105/tpc.6.11.1665. *Plant Cell*, *6*(11), 1665–1679. https://doi.org/10.1105/tpc.6.11.1665

Jang, J., León, P., Zhou, L. & Sheenl, J. (1997). *Hexokinase as a Sugar Sensor in Higher Plants*. *9*(January), 5–19.

Järvi, S., Suorsa, M. & Aro, E. M. (2015). Photosystem II repair in plant chloroplasts - Regulation, assisting proteins and shared components with photosystem II

biogenesis. *Biochimica et Biophysica Acta - Bioenergetics*, *1847*(9), 900–909. https://doi.org/10.1016/j.bbabio.2015.01.006

Jeffers, T. L. & Roth, M. S. (2021). Revealing mechanisms of algal astaxanthin production and bioengineering potential using multi-omics. In G. A. Rivishankar & R. . Ambati (Eds.), *Global perspective of astaxanthin: From industrial production to food, health and pharmaceutical applications*. Elsevier.

Jeong, J., Baek, K., Yu, J., Kirst, H., Betterle, N., Shin, W., Bae, S., Melis, A. & Jin, E. (2018). Deletion of the chloroplast LTD protein impedes LHCI import and PSI-LHCI assembly in Chlamydomonas reinhardtii. *Journal of Experimental Botany*, *69*(5), 1147–1158. https://doi.org/10.1093/jxb/erx457

Johnson, J. (UC B. (2019). Something new under the sun: transposon sequencing and phylogenomics shed light on unstudied photosynthetic genes [University of California, Berkeley]. In *Something new under the sun: transposon sequencing and phylogenomics shed light on unstudied photosynthetic genes*. https://digitalassets.lib.berkeley.edu/etd/ucb/text/Johnson_berkeley_0028E_19158.pdf

Johnson, X. & Alric, J. (2012). Interaction between starch breakdown, acetate assimilation, and photosynthetic cyclic electron flow in Chlamydomonas reinhardtii. *The Journal of Biological Chemistry*, *287*(31), 26445–26452. https://doi.org/10.1074/jbc.M112.370205

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jouhet, J., Maréchal, E. & Block, M. A. (2007). Glycerolipid transfer for the building of membranes in plant cells. *Progress in Lipid Research*, *46*(1), 37–55. https://doi.org/10.1016/j.plipres.2006.06.002

Kandasamy, M. K., Deal, R. B., McKinney, E. C. & Meagher, R. B. (2004). Plant actin-related proteins. *Trends in Plant Science*, *9*(4), 196–202. https://doi.org/10.1016/j.tplants.2004.02.004

Kanehisa, M. & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science : A Publication of the Protein Society*, *29*(1), 28–35. https://doi.org/10.1002/pro.3711

Kang, C. D., Lee, J. S., Park, T. H. & Sim, S. J. (2005). Comparison of heterotrophic and photoautotrophic induction on astaxanthin production by Haematococcus pluvialis. *Applied Microbiology and Biotechnology*, *68*(2), 237–241. https://doi.org/10.1007/s00253-005-1889-2

Karpowicz, S. J., Prochnik, S. E., Grossman, A. R. & Merchant, S. S. (2011). The greenCut2 resource, a phylogenomically derived inventory of proteins specific to the plant lineage. *Journal of Biological Chemistry*, *286*(24), 21427–21439. https://doi.org/10.1074/jbc.M111.233734

Karve, R., Lauria, M., Virnig, A., Xia, X., Rauh, B. L. & Moore, B. D. (2010). Evolutionary lineages and functional diversification of plant hexokinases. *Molecular Plant*, *3*(2), 334–346. https://doi.org/10.1093/mp/ssq003

Kawata, E. & Cheung, A. Y. (1990). Molecular analysis of an aurea photosynthetic mutant (Su/Su) in tobacco: LHCP depletion leads to pleiotropic mutant phenotypes. *EMBO Journal*, *9*(12), 4197–4203. https://doi.org/10.1002/j.1460-2075.1990.tb07644.x

Kessler, E & Czygan, F. C. (1965). Chlorella zofingiensis Dönz: Isolierung neuer Stämme und ihre physiologisch-biochemischen Eigenschaften. *Ber. Dtsch. Bot. Ges.*, *78*, 342–347.

Kessler, Erich. (1985). Upper limits of temperature for growth in Chlorella (Chlorophyceae). *Plant Systematics and Evolution*, *151*(1–2), 67–71. https://doi.org/10.1007/BF02418020

Kim, D. H., Park, M. J., Gwon, G. H., Silkov, A., Xu, Z. Y., Yang, E. C., Song, S., Song, K., Kim, Y., Yoon, H. S., Honig, B., Cho, W., Cho, Y. & Hwang, I. (2014). An ankyrin repeat domain of AKR2 drives chloroplast targeting through coincident binding of two chloroplast lipids. *Developmental Cell*, *30*(5), 598–609. https://doi.org/10.1016/j.devcel.2014.07.026

Kim, D. H., Xu, Z. Y., Na, Y. J., Yoo, Y. J., Lee, J., Sohn, E. J. & Hwang, I. (2011). Small heat shock protein Hsp17.8 functions as an AKR2A cofactor in the targeting of chloroplast outer membrane proteins in arabidopsis. *Plant Physiology*, *157*(1), 132–146. https://doi.org/10.1104/pp.111.178681

Klimyuk, V. I., Persello-Cartieaux, F., Havaux, M., Contard-David, P., Schuenemann, D., Meiherhoff, K., Gouet, P., Jones, J. D. G., Hoffman, N. E. & Nussaume, L. (1999). A chromodomain protein encoded by the Arabidopsis CAO gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting. *Plant Cell*, *11*(1), 87–99. https://doi.org/10.1105/tpc.11.1.87

Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M. & Ohta, H. (2007). *Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis , photosynthesis , and embryogenesis*.

Kobayashi, K., Narise, T., Sonoike, K., Hashimoto, H., Sato, N., Kondo, M., Nishimura, M., Sato, M., Toyooka, K., Sugimoto, K., Wada, H., Masuda, T. & Ohta, H. (2013). Role of galactolipid biosynthesis in coordinated development of photosynthetic complexes and thylakoid membranes during chloroplast biogenesis in Arabidopsis. *Plant Journal*, *73*(2), 250–261. https://doi.org/10.1111/tpj.12028

Kol, E. & Chodat, F. (1934). Quelques algues nouvelles des sols et de la neige du Parc National Suisse Engadine. *Bull. Soc. Bot. Genève*, *25*, 250–263.

Komenda, J., Norling, B., Zhang, L., Knoppova, J. & Sela, T. T. (2016). *Photosystem II Assembly Steps Take Place in the Thylakoid Membrane of the Cyanobacterium Synechocystis sp . PCC6803*. *1*(November 2015), 95–104. https://doi.org/10.1093/pcp/pcv178

Kopecký, J., Schoefs, B., Loest, K., Štys, D. & Pulz, O. (2000). Microalgae as a source for secondary carotenoid production: a screening study. *Algological Studies/Archiv Für Hydrobiologie, Supplement Volumes*, *98*(June), 153–168. https://doi.org/10.1127/algol_stud/98/2000/153

Kowalewska, Ł., Mazur, R., Suski, S., Garstka, M. & Mostowska, A. (2016). Three-dimensional visualization of the tubular-lamellar transformation of the internal plastid membrane network during runner bean chloroplast biogenesis. *Plant Cell*, *28*(4), 875–891. https://doi.org/10.1105/tpc.15.01053

Kroll, D., Meierhoff, K., Bechtold, N., Kinoshita, M., Westphal, S., Vothknecht, U. C., Soll, J. & Westhoff, P. (2001). VIPP1, a nuclear gene of Arabidopsis thaliana essential for thylakoid membrane formation. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 4238–4242. https://doi.org/10.1073/pnas.061500998

Kumar, M., Kesawat, M. S., Ali, A., Lee, S. C., Gill, S. S. & Kim, H. U. (2019). Integration of abscisic acid signaling with other signaling pathways in plant stress responses and development. *Plants*, *8*(12), 1–20. https://doi.org/10.3390/plants8120592

Kusnetsov, V., Herrmann, R. G., Kulaeva, O. N. & Oelmüller, R. (1998). Cytokinin stimulates and abscisic acid inhibits greening of etiolated Lupinus luteus cotyledons by affecting the expression of the light-sensitive protochlorophyllide oxidoreductase. *Molecular and General Genetics*, *259*(1), 21–28. https://doi.org/10.1007/s004380050784

Lane, S., Xu, H., Oh, E. J., Kim, H., Lesmana, A., Jeong, D., Zhang, G., Tsai, C. S., Jin, Y. S. & Kim, S. R. (2018). Glucose repression can be alleviated by reducing glucose phosphorylation rate in Saccharomyces cerevisiae. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-20804-4

Larkin, R. M., Stefano, G., Ruckle, M. E., Stavoe, A. K., Sinkler, C. A., Brandizzi, F., Malmstrom, C. M., Osteryoung, K. W. & Chory, J. (2016). REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), E1116–E1125. https://doi.org/10.1073/pnas.1515741113

Lee, D. W., Lee, J. & Hwang, I. (2017). Sorting of nuclear-encoded chloroplast membrane proteins. *Current Opinion in Plant Biology*, *40*, 1–7. https://doi.org/10.1016/j.pbi.2017.06.011

Lee, H.-J., Mochizuki, N., Masuda, T. & Buckhout, T. J. (2012). Disrupting the bimolecular binding of the haem-binding protein 5 (AtHBP5) to haem oxygenase 1 (HY1) leads to oxidative stress in Arabidopsis. *Journal of Experimental Botany*, *63*(16), 5967–5978. https://doi.org/10.1093/jxb/ers242

Lee, J. H., Terzaghi, W. & Deng, X. W. (2011). DWA3, an Arabidopsis DWD protein, acts as a negative regulator in ABA signal transduction. *Plant Science*, *180*(2), 352–357. https://doi.org/10.1016/j.plantsci.2010.10.008

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, L. & Sheen, J. (2016). Dynamic and diverse sugar signaling. *Current Opinion in Plant Biology*, *33*, 116–125. https://doi.org/10.1016/j.pbi.2016.06.018

Li, Yantao, Huang, J., Sandmann, G. & Chen, F. (2008). Glucose sensing and the mitochondrial alternative pathway are involved in the regulation of astaxanthin biosynthesis in the dark-grown Chlorella zofingiensis (Chlorophyceae). *Planta*, *228*(5), 735–743. https://doi.org/10.1007/s00425-008-0775-4

Li, Yantao, Huang, J., Sandmann, G. & Chen, F. (2009). High-light and sodium chloride stress differentially regulate the biosynthesis of astaxanthin in chlorella zofingiensis (chlorophyceae). *Journal of Phycology*, *45*(3), 635–641. https://doi.org/10.1111/j.1529-8817.2009.00689.x

Li, Yongqing, Pennington, B. O. & Hua, J. (2009). Multiple R-like genes are negatively regulated by BON1 and BON3 in arabidopsis. *Molecular Plant-Microbe Interactions*, *22*(7), 840–848. https://doi.org/10.1094/MPMI-22-7-0840

Li, Yun, Guo, J., Yang, Z. & Yang, D. L. (2018). Plasma membrane-localized calcium pumps and copines coordinately regulate pollen germination and fertility in arabidopsis. *International Journal of Molecular Sciences*, *19*(6). https://doi.org/10.3390/ijms19061774

Li, Yunhai, Lee, K. K., Walsh, S., Smith, C., Hadingham, S., Sorefan, K., Cawley, G. & Bevan, M. W. (2006). Establishing glucose- and ABA-regulated transcription networks in Arabidopsis by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Research*, *16*(3), 414–427. https://doi.org/10.1101/gr.4237406

Liberton, M., Howard Berg, R., Heuser, J., Roth, R. & Pakrasi, H. B. (2006). Ultrastructure of the membrane systems in the unicellular cyanobacterium Synechocystis sp. strain PCC 6803. *Protoplasma*, *227*(2–4), 129–138. https://doi.org/10.1007/s00709-006-0145-7

Liu, J., Huang, J., Jiang, Y. & Chen, F. (2012). Molasses-based growth and production of oil and astaxanthin by Chlorella zofingiensis. *Bioresource Technology*, *107*, 393–398. https://doi.org/10.1016/j.biortech.2011.12.047

Liu, J., Sun, Z., Gerken, H., Liu, Z., Jiang, Y. & Chen, F. (2014a). Chlorella zofingiensis as an alternative microalgal producer of astaxanthin: Biology and industrial potential. *Marine Drugs*, *12*(6), 3487–3515. https://doi.org/10.3390/md12063487

Liu, J., Sun, Z., Gerken, H., Liu, Z., Jiang, Y. & Chen, F. (2014b). *Chlorella zofingiensis as an Alternative Microalgal Producer of Astaxanthin: Biology and Industrial Potential*. 3487–3515. https://doi.org/10.3390/md12063487

Liu, X. & Osawa, T. (2007). Cis astaxanthin and especially 9-cis astaxanthin exhibits a higher antioxidant activity in vitro compared to the all-trans isomer. *Biochemical and Biophysical Research Communications*, *357*(1), 187–193. https://doi.org/10.1016/j.bbrc.2007.03.120

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Livingston, A. K., Cruz, J. A., Kohzuma, K., Dhingra, A. & Kramer, D. M. (2010). An arabidopsis mutant with high cyclic electron flow around photosystem i (hcef) involving the nadphdehydrogenase complex. *Plant Cell*, *22*(1), 221–233. https://doi.org/10.1105/tpc.109.071084

Lütz, C. & Klein, S. (1979). Biochemical and Cytological Observations on Chloroplast Development VI. Chlorophylls and Saponins in Prolamellar Bodies and Prothylakoids Separated from Etioplasts of Etiolated Avena sativa L. Leaves. *Zeitschrift Für Pflanzenphysiologie*, *95*(3), 227–237. https://doi.org/10.1016/s0044-328x(79)80236-6

Lynch, D. L. & Fenwick, M. G. (1974). Factors Affecting the Growth of Two Green Algae. *The American Biology Teacher*, *36*(8), 494–495.

Matsuka, M. & Hase, E. (1965). Metabolism of glucose in the process of “glucose-bleaching” of chlorella protothecoides. *Plant and Cell Physiology*, *6*(4), 721–741. https://doi.org/10.1093/oxfordjournals.pcp.a079144

Matsuka, M. & Miyachi, S. (1974). *Photosynthetic metabolism of 14 CO 2 in the process*. *926*, 919–926.

Mechela, A., Schwenkert, S. & Soll, J. (2019). A brief history of thylakoid biogenesis. *Open Biology*, *9*(1). https://doi.org/10.1098/rsob.180237

Meinke, D. W. (2020). Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *New Phytologist*, *226*(2), 306–325. https://doi.org/10.1111/nph.16071

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-laylin, L. K., Maréchal-drouard, L., Marshall, W. F., Qu, L., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V, Ren, Q., Ferris, P., Lindquist, E., … Grimwood, J. (2007). *The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions*. *318*(October), 245–252.

Mihara, S., Hase, E. & Koichi, K. (1968). Studies on ribonucleic acids from chlorellaprotothecoides with special reference to the degradation of chloroplast rna during the process of “glucose-bleaching1. *Plant and Cell Physiology*, *9*(1), 87–102. https://doi.org/10.1093/oxfordjournals.pcp.a079332

Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W. H., Liu, Y. X., Hwang, I., Jones, T. & Sheen, J. (2003). Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science*, *300*(5617), 332–336. https://doi.org/10.1126/science.1080585

Morré, D. J., Selldén, G., Sundqvist, C. & Sandelius, A. S. (1991). Stromal low temperature compartment derived from the inner membrane of the chloroplast envelope. *Plant Physiology*, *97*(4), 1558–1564. https://doi.org/10.1104/pp.97.4.1558

Mulders, K. J. M., Weesepoel, Y., Bodenes, P., Lamers, P. P., Vincken, J. & Martens,

D. E. (2015). *Nitrogen-depleted Chlorella zofingiensis produces astaxanthin , ketolutein and their fatty acid esters : a carotenoid metabolism study*. 125–140. https://doi.org/10.1007/s10811-014-0333-3

Mulders, K. J. M., Weesepoel, Y., Bodenes, P., Lamers, P. P., Vincken, J. P., Martens, D. E., Gruppen, H. & Wijffels, R. H. (2014). Nitrogen-depleted Chlorella zofingiensis produces astaxanthin, ketolutein and their fatty acid esters: a carotenoid metabolism study. *Journal of Applied Phycology*, *27*(1), 125–140. https://doi.org/10.1007/s10811-014-0333-3

Murakami, S., Yamada, N., Nagano, M. & Osumi, M. (1985). Three-dimensional structure of the prolamellar body in squash etioplasts. *Protoplasma*, *128*(2–3), 147–156. https://doi.org/10.1007/BF01276336

Nakazato, E., Fukuzawa, H., Tabata, S., Takahashi, H. & Tanaka, K. (2003). Identification and expression analysis of cDNA encoding a chloroplast recombination protein REC1, the chloroplast RecA homologue in Chlamydomonas reinhardtii. *Bioscience, Biotechnology and Biochemistry*, *67*(12), 2608–2613. https://doi.org/10.1271/bbb.67.2608

Nguyen, H. C., Melo, A. A., Kruk, J., Frost, A. & Gabruk, M. (2020). Photocatalytic plant LPOR forms helical lattices that shape membranes for chlorophyll synthesis. *BioRxiv*, 2020.08.19.257774. https://doi.org/10.1101/2020.08.19.257774

Oh-hama, T. & Hase, E. (1975). *Syntheses of d-aminolevulinic acid and chlorophyll during chloroplast formation in Chlorella protothecoides*. *303*, 297–303.

Oh-hama, T. & Senger, H. (1978). *Spectral effectiveness in chlorophyll and 5-aminolevulinic acid formation during regreening of glucose-bleached cells of Chlorella protothecoides*. *19*(7), 1295–1299.

Orosa, M., Torres, E., Fidalgo, P. & Abalde, J. (2000). Production and analysis of secondary carotenoids in green algae. *Journal of Applied Phycology*, *12*(3–5), 553–556. https://doi.org/10.1023/a:1008173807143

Osuna, D., Usadel, B., Morcuende, R., Gibon, Y., Bläsing, O. E., Höhne, M., Günter, M., Kamlage, B., Trethewey, R., Scheible, W.-R. & Stitt, M. (2007). Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *The Plant Journal : For Cell and Molecular Biology*, *49*(3), 463–491. https://doi.org/10.1111/j.1365-313X.2006.02979.x

Ouyang, M., Li, X., Ma, J., Chi, W., Xiao, J., Zou, M., Chen, F., Lu, C. & Zhang, L. (2011). LTD is a protein required for sorting light-harvesting chlorophyll-binding proteins to the chloroplast SRP pathway. *Nature Communications*, *2*(1). https://doi.org/10.1038/ncomms1278

Page, D. R. & Grossniklaus, U. (2002). The art and design of genetic screens: Arabidopsis thaliana. *Nature Reviews Genetics*, *3*(2), 124–136. https://doi.org/10.1038/nrg730

Parker, B. C., Bold, H. C. & Deason, T. R. (1961). Facultative heterotrophy in some chlorococcacean algae. *Science*, *133*(3455), 761–763.

Patterson, G. (1978). *Culture Collection of Algae and Protozoa Strain Information:*

Strain 211/51. Https://Www.Ccap.Ac.Uk. https://www.ccap.ac.uk/strain_info.php?Strain_No=211/51

Paulina Aguilera-Alvarado, G. & Sanchez-Nieto, S. (2017). Plant Hexokinases are Multifaceted Proteins. *Plant and Cell Physiology*, *58*(7), 1151–1160. https://doi.org/10.1093/pcp/pcx062

Peláez, R., Fernández-García, P., Herrero, P. & Moreno, F. (2012). Nuclear import of the yeast hexokinase 2 protein requires α/β-importin-dependent pathway. *Journal of Biological Chemistry*, *287*(5), 3518–3529.

Peláez, R., Herrero, P. & Moreno, F. (2009). Nuclear export of the yeast hexokinase 2 protein requires the Xpo1 (Crm-1)-dependent pathway. *Journal of Biological Chemistry*, *284*(31), 20548–20555. https://doi.org/10.1074/jbc.M109.013730

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J. & Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell*, *18*(1), 176–197. https://doi.org/10.1105/tpc.105.036392

Poulsen, L. R., López-Marqués, R. L., McDowell, S. C., Okkeri, J., Licht, D., Schulz, A., Pomorski, T., Harper, J. F. & Palmgren, M. G. (2008). The Arabidopsis P4-ATPase ALA3 localizes to the golgi and requires a β-subunit to function in lipid translocation and secretory vesicle formation. *Plant Cell*, *20*(3), 658–676. https://doi.org/10.1105/tpc.107.054767

Proctor, M. S., Chidgey, J. W., Shukla, M. K., Jackson, P. J., Sobotka, R., Hunter, C. N. & Hitchcock, A. (2018). Plant and algal chlorophyll synthases function in Synechocystis and interact with the YidC/Alb3 membrane insertase. *FEBS Letters*, *592*(18), 3062–3073. https://doi.org/10.1002/1873-3468.13222

Rast, A., Heinz, S. & Nickelsen, J. (2015a). Biochimica et Biophysica Acta Biogenesis of thylakoid membranes ☆. *BBA - Bioenergetics*, *1847*(9), 821–830. https://doi.org/10.1016/j.bbabio.2015.01.007

Rast, A., Heinz, S. & Nickelsen, J. (2015b). Biogenesis of thylakoid membranes. *Biochimica et Biophysica Acta - Bioenergetics*, *1847*(9), 821–830. https://doi.org/10.1016/j.bbabio.2015.01.007

Rast, A., Schaffer, M., Albert, S., Wan, W., Pfeffer, S., Beck, F., Plitzko, J. M., Nickelsen, J. & Engel, B. D. (2019). Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nature Plants*, *5*(4), 436–446. https://doi.org/10.1038/s41477-019-0399-7

Redman, J. C., Haas, B. J., Tanimoto, G. & Town, C. D. (2004). Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. *Plant Journal*, *38*(3), 545–561. https://doi.org/10.1111/j.1365-313X.2004.02061.x

Rentas, A. G. (1980). *The effect of potassium ion and light on the growth and ultrastructure of bracteacoccus cinnabarinus*.

Rise, M., Cohen, E., Vishkautsan, M., Cojocaur, M., Gottlieb, H. E. & Arad, (Malis), S. (1994). Accumulation of Secondary Carotenoids in Chlorella zofingiensis. *Plant Physiology*, *144*, 287–292.

Roach, M. B. (1926). On the Relation of Certain Soil Algae to some Soluble Carbon Compounds. *Annals of Botany*, *40*(157), 149–201. https://www.jstor.org/stable/43237033

Rodríguez, A., Cera, T. D. La, Herrero, P. & Moreno, F. (2001). The hexokinase 2 protein regulates the expression of the *GLK1*, *HXK1* and *HXK2* genes of *Saccharomyces cerevisiae*. *Biochemical Journal*, *355*, 625–631. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1221776/pdf/11311123.pdf

Rolland, F. & Sheen, J. (2005). Sugar sensing and signalling networks in plants. *Biochemical Society Transactions*, *33*(1), 269–271. https://doi.org/10.1042/BST0330269

Rolland, Filip, Baena-Gonzalez, E. & Sheen, J. (2006). SUGAR SENSING AND SIGNALING IN PLANTS: Conserved and Novel Mechanisms. *Annual Review of Plant Biology*, *57*(1), 675–709. https://doi.org/10.1146/annurev.arplant.57.032905.105441

Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C. A., Merchant, S. S., Pellegrini, M. & Niyogi, K. K. (2017). Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(21), E4296–E4305. https://doi.org/10.1073/pnas.1619928114

Roth, M. S., Gallaher, S. D., Westcott, D. J., Iwai, M., Louie, K. B., Mueller, M., Walter, A., Foflonker, F., Bowen, B. P., Ataii, N. N., Song, J., Chen, J. H., Blaby-Haas, C. E., Larabell, C., Auer, M., Northen, T. R., Merchant, S. S. & Niyogia, K. K. (2019). Regulation of oxygenic photosynthesis during trophic transitions in the green alga chromochloris zofingiensis. *Plant Cell*, *31*(3), 579–601. https://doi.org/10.1105/tpc.18.00742

Roth, M. S., Westcott, D. J., Iwai, M. & Niyogi, K. K. (2019). Hexokinase is necessary for glucose-mediated photosynthesis repression and lipid accumulation in a green alga. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-019-0577-1

Ruban, A. V., Wentworth, M., Yakushevska, A. E., Andersson, J., Lee, P. J., Keegstra, W., Dekker, J. P., Boekema, E. J., Jansson, S. & Horton, P. (2003). Plants lacking the main light-harvesting complex retain photosystem II macro-organization. *Nature*, *421*(6923), 648–652. https://doi.org/10.1038/nature01344

Sakamoto, W., Miyagishima, S. & Jarvis, P. (2008). Chloroplast Biogenesis: Control of Plastid Development, Protein Import, Division and Inheritance. *The Arabidopsis Book*, *6*, e0110. https://doi.org/10.1199/tab.0110

Sankaranarayanan, S., Jamshed, M., Kumar, A., Skori, L., Scandola, S., Wang, T., Spiegel, D. & Samuel, M. A. (2017). Glyoxalase goes green: The expanding roles of glyoxalase in plants. *International Journal of Molecular Sciences*, *18*(4). https://doi.org/10.3390/ijms18040898

Sassu, O. G. (1972). Effect of diphenylamine on carotenoid synthesis in Dictyococcus cinnabarinus. *Phytochemistry*, *11*(11), 3195–3198.

https://doi.org/https://doi.org/10.1016/S0031-9422(00)86372-4

Sato, N., Kihira, M., Matsushita, R., Kaneko, C., Ishii, Y., Yin, Y. G., Kawachi, N., Teramura, H., Kusano, H. & Shimada, H. (2020). AtFLL2, a member of the FLO2 gene family, affects the enlargement of leaves at the vegetative stage and facilitates the regulation of carbon metabolism and flow. *Bioscience, Biotechnology and Biochemistry*, *00*(00), 1–10. https://doi.org/10.1080/09168451.2020.1812374

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. & Lohmann, J. U. (2005). A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, *37*(5), 501–506. https://doi.org/10.1038/ng1543

Schmitz, J., Rossoni, A. W. & Maurino, V. G. (2018). Dissecting the physiological function of plant glyoxalase I and glyoxalase I-like proteins. *Frontiers in Plant Science*, *871*(November), 1–7. https://doi.org/10.3389/fpls.2018.01618

Schubert, M., Petersson, U. A., Haas, B. J., Funk, C., Schröder, W. P. & Kieselbach, T. (2002). Proteome map of the chloroplast lumen of Arabidopsis thaliana. *Journal of Biological Chemistry*, *277*(10), 8354–8365. https://doi.org/10.1074/jbc.M108575200

Schwarz, K. (1979). Neue Bodenalgen aus Dalmatien. *Plant Systematics and Evolution*, *131*, 193–209.

Selstam, E. & Sandelius, A. S. (1984). A comparison between prolamellar bodies and prothylakoid membranes of etioplasts of dark-grown wheat concerning lipid and polypeptide composition. *Plant Physiology*, *76*(4), 1036–1040. https://doi.org/10.1104/pp.76.4.1036

Semenenko, V. E. (1981). Metabolite Regulation of Chloroplast Genome Expression and of the Activity of the Photosynthetic Apparatus. In G. Akoyunoglou (Ed.), *Proceeding of the 5th International Congress on Photosynthesis* (pp. 767–776).

Seo, K. I., Lee, J. H., Nezames, C. D., Zhong, S., Song, E., Byun, M. O. & Deng, X. W. (2014). ABD1 is an Arabidopsis DCAF substrate receptor for CUL4-DDB1-based E3 ligases that acts as a negative regulator of abscisic acid signaling. *Plant Cell*, *26*(2), 695–711. https://doi.org/10.1105/tpc.113.119974

Sheen, J. (1990). *Metabolic Repression of Transcription in Higher Plants*. *2*(October), 1027–1039.

Sheen, J. (2014). Master Regulators in Plant Glucose Signaling Networks. *Journal of Plant Biology = Singmul Hakhoe Chi*, *57*(2), 67–79. https://doi.org/10.1007/s12374-014-0902-7

Shelp, B. J., Bozzo, G. G., Trobacher, C. P., Zarei, A., Deyman, K. L. & Brikis, C. J. (2012). Hypothesis/review: Contribution of putrescine to 4-aminobutyrate (GABA) production in response to abiotic stress. *Plant Science*, *193–194*, 130–135. https://doi.org/10.1016/j.plantsci.2012.06.001

Shen, H. & Lee, Y. K. (1997). Thermotolerance induced by heat shock in Chlorella. *Journal of Applied Phycology*, *9*(5), 471–475. https://doi.org/10.1023/A:1007956507345

Shihira-ishikawa, I. & Hase, E. (1964a). Nutritional control of cell pigmentation in

chlorella protothecoides with special reference to the degeneration of chloroplast induced by glucose. *Plant and Cell Physiology*, *5*(2), 227–240. https://doi.org/10.1093/oxfordjournals.pcp.a079037

Shihira-ishikawa, I. & Hase, E. (1964b). *NUTRITIONAL CONTROL OF CELL PIGMENTATION IN CHLORELLA PROTOTHECOIDES WITH SPECIAL REFERENCE TO THE DEGENERATION OF CHLOROPLAST INDUCED BY GLUCOSE 1. 5*.

Shihira, I. & Krauss, R. W. (1965). *Chlorella: physiology and taxonomy of forty-one isolates,*. University of Maryland.

Singla-Pareek, S. L., Kaur, C., Kumar, B., Pareek, A. & Sopory, S. K. (2020). Reassessing plant glyoxalases: large family and expanding functions. *New Phytologist*, *227*(3), 714–721. https://doi.org/10.1111/nph.16576

Sirpiö, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N. & Aro, E. M. (2007). TLP18.3, a novel thylakoid lumen protein regulating photosystem II repair cycle. *Biochemical Journal*, *406*(3), 415–425. https://doi.org/10.1042/BJ20070460

Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W. & Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, *355*(6332), 1436 LP – 1440. https://doi.org/10.1126/science.aal3794

Stengel, Katharina F., Holdermann, I., Cain, P., Robinson, C., Wild, K. & Sinning, I. (2008). Recognition Particle Protein cpSRP43. *Science*, *321*(July), 253–256. https://doi.org/10.1126/science.1158640

Sun, H., Ren, Y., Lao, Y., Li, X. & Chen, F. (2020). A novel fed-batch strategy enhances lipid and astaxanthin productivity without compromising biomass of Chromochloris zofingiensis. *Bioresource Technology*, *308*(February), 123306. https://doi.org/10.1016/j.biortech.2020.123306

Sundqvist, C. & Dahlin, C. (1997). With chlorophyll pigments from prolamellar bodies to light-harvesting complexes. *Physiologia Plantarum*, *100*(4), 748–759. https://doi.org/10.1034/j.1399-3054.1997.1000402.x

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055

Theodoulou, F. L. & Eastmond, P. J. (2012). Seed storage oil catabolism: A story of give and take. *Current Opinion in Plant Biology*, *15*(3), 322–328. https://doi.org/10.1016/j.pbi.2012.03.017

Toufighi, K., Brady, S. M., Austin, R., Ly, E. & Provart, N. J. (2005). The botany array resource: e-Northerns, expression angling, and promoter analyses. *Plant Journal*, *43*(1), 153–163. https://doi.org/10.1111/j.1365-313X.2005.02437.x

Treves, H., Murik, O., Kedem, I., Eisenstadt, D., Meir, S., Rogachev, I., Szymanski, J., Keren, N., Orf, I., Tiburcio, A. F., Alcázar, R., Aharoni, A., Kopka, J. & Kaplan, A. (2017). Metabolic Flexibility Underpins Growth Capabilities of the Fastest Growing Alga. *Current Biology*, *27*(16), 2559-2567.e3.

https://doi.org/10.1016/j.cub.2017.07.014

Ulfstedt, M., Hu, G.-Z., Eklund, D. M. & Ronne, H. (2018a). The Ability of a Charophyte Alga Hexokinase to Restore Glucose Signaling and Glucose Repression of Gene Expression in a Glucose-Insensitive Arabidopsis Hexokinase Mutant Depends on Its Catalytic Activity. *Frontiers in Plant Science*, *9*(December), 1–15. https://doi.org/10.3389/fpls.2018.01887

Ulfstedt, M., Hu, G. Z., Eklund, D. M. & Ronne, H. (2018b). The ability of a charophyte alga hexokinase to restore glucose signaling and glucose repression of gene expression in a glucose-insensitive Arabidopsis hexokinase mutant depends on its catalytic activity. *Frontiers in Plant Science*, *871*(December), 1–15. https://doi.org/10.3389/fpls.2018.01887

Vainonen, J. P., Sakuragi, Y., Stael, S., Tikkanen, M., Allahverdiyeva, Y., Paakkarinen, V., Aro, E., Suorsa, M., Scheller, H. V., Vener, A. V. & Aro, E. M. (2008). Light regulation of CaS, a novel phosphoprotein in the thylakoid membrane of Arabidopsis thaliana. *FEBS Journal*, *275*(8), 1767–1777. https://doi.org/10.1111/j.1742-4658.2008.06335.x

Van Buren, R., Pardo, J., Wai, C. M., Evans, S. & Bartels, D. (2019). Massive tandem proliferation of elips supports convergent evolution of desiccation tolerance across land plants. *Plant Physiology*, *179*(3), 1040–1049. https://doi.org/10.1104/pp.18.01420

Vandenabeele, S., Vanderauwera, S., Vuylsteke, M., Rombauts, S., Langebartels, C., Seidlitz, H. K., Zabeau, M., Van Montagu, M., Inzé, D. & Van Breusegem, F. (2004). Catalase deficiency drastically affects gene expression induced by high light in Arabidopsis thaliana. *Plant Journal*, *39*(1), 45–58. https://doi.org/10.1111/j.1365-313X.2004.02105.x

Vega, M., Riera, A., Fernández-Cid, A., Herrero, P. & Moreno, F. (2016). Hexokinase 2 Is an intracellular glucose sensor of yeast cells that maintains the structure and activity of mig1 protein repressor complex. *Journal of Biological Chemistry*, *291*(14), 7267–7285. https://doi.org/10.1074/jbc.M115.711408

Virgin, H. I. (1963). The Physiology of chlorophyll formation in relation to structural changes in chloroplasts. *Photochemical and Photobiological Sciences1*, *2*, 83–91.

Vladimirova, M. G. (1976). Changes of Ultrastructural Organization During Functional Reorganization of the Cell in Chlorella Sp.K. *Fiziologiya Rastenii*, *23*(6), 1180–1187.

Voort, M. P. J. van der, Vulsteke, E. & Visser, C. L. M. de. (2015). Macro-economics of algae products. *Marco-Economics of Algae Products Public Output Report WP2A7.02 of the EnAlgae Project*, 47.

Walters, R. G., Shephard, F., Rogers, J. J. M., Rolfe, S. A. & Horton, P. (2003). Identification of mutants of arabidopsis defective in acclimation of photosynthesis to the light environment. *Plant Physiology*, *131*(2), 472–481. https://doi.org/10.1104/pp.015479

Wang, L., Yamano, T., Takane, S., Niikawa, Y., Toyokawa, C., Ozawa, S. I., Tokutsu,

R., Takahashi, Y., Minagawa, J., Kanesaki, Y., Yoshikawa, H. & Fukuzawa, H. (2016). Chloroplast-mediated regulation of CO2-concentrating mechanism by Ca2+-binding protein CAS in the green alga Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12586–12591. https://doi.org/10.1073/pnas.1606519113

Wang, Z. & Benning, C. (2012). *Chloroplast lipid synthesis and lipid trafficking through ER – plastid membrane contact sites*. 457–463. https://doi.org/10.1042/BST20110752

Weinl, S., Held, K., Schlücking, K., Steinhorst, L., Kuhlgert, S., Hippler, M. & Kudla, J. (2008). A plastid protein crucial for Ca2+-regulated stomatal responses. *New Phytologist*, *179*(3), 675–686. https://doi.org/10.1111/j.1469-8137.2008.02492.x

Wessel, D. & Flügge, U. I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry*, *138*(1), 141–143. https://doi.org/10.1016/0003-2697(84)90782-6

Westphal, S., Soll, J. & Vothknecht, U. C. (2001). A vesicle transport system inside chloroplasts. *FEBS Letters*, *506*(3), 257–261. https://doi.org/10.1016/S0014-5793(01)02931-3

Wick, A. N., Drury, D. R., Nakada, H. I. & Wolfe, J. B. (1956). Localization of the Primary Metabolic Block Produced by 2-Deoxyglucose. *Journal of Biological Chemistry*, *224*(2), 963–969.

Wittkopp, T. M., Schmollinger, S., Saroussi, S., Hu, W., Zhang, W., Fan, Q., Gallaher, S. D., Leonard, M. T., Soubeyrand, E., Basset, G. J., Merchant, S. S., Grossman, A. R., Duanmu, D. & Lagarias, J. C. (2017). Bilin-dependent photoacclimation in Chlamydomonas reinhardtii. *Plant Cell*, *29*(11), 2711–2726. https://doi.org/10.1105/tpc.17.00149

Wu, H. Y., Liu, M. Sen, Lin, T. P. & Cheng, Y. S. (2011). Structural and functional assays of AtTLP18.3 identify its novel acid phosphatase activity in thylakoid lumen. *Plant Physiology*, *157*(3), 1015–1025. https://doi.org/10.1104/pp.111.184739

Xiao, W., Sheen, J. & Jang, J. C. (2000). The role of hexokinase in plant sugar signal transduction and growth and development. *Plant Molecular Biology*, *44*(4), 451–461. https://doi.org/10.1023/A:1026501430422

Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C. & Sheen, J. (2013). Glucose–TOR signalling reprograms the transcriptome and activates meristems. *Nature*, *496*(7444), 181–186. https://doi.org/10.1038/nature12030

Yamburenko, M. V., Zubo, Y. O., Vanková, R., Kusnetsov, V. V., Kulaeva, O. N. & Börner, T. (2013). Abscisic acid represses the transcription of chloroplast genes. *Journal of Experimental Botany*, *64*(14), 4491–4502. https://doi.org/10.1093/jxb/ert258

Yang, S., Yang, H., Grisafi, P., Sanchatjate, S., Fink, G. R., Sun, Q. & Hua, J. (2006). The BON/CPN gene family represses cell death and promotes cell growth in Arabidopsis. *Plant Journal*, *45*(2), 166–179. https://doi.org/10.1111/j.1365-313X.2005.02585.x

Zhang, Lingang, Kato, Y., Otters, S., Vothknecht, U. C. & Sakamoto, W. (2012). Essential role of VIPP1 in chloroplast envelope maintenance in Arabidopsisw. *Plant Cell*, *24*(9), 3695–3707. https://doi.org/10.1105/tpc.112.103606

Zhang, Lixin & Aro, E. M. (2002). Synthesis, membrane insertion and assembly of the chloroplast-encoded D1 protein into photosystem II. *FEBS Letters*, *512*(1–3), 13–18. https://doi.org/10.1016/S0014-5793(02)02218-4

Zhang, W., Zhong, H., Lu, H., Zhang, Y., Deng, X., Huang, K. & Duanmu, D. (2018). Characterization of ferredoxin-dependent biliverdin reductase PCYA1 reveals the dual function in retrograde bilin biosynthesis and interaction with light-dependent protochlorophyllide oxidoreductase LPOR in Chlamydomonas reinhardtii. *Frontiers in Plant Science*, *9*(May), 1–13. https://doi.org/10.3389/fpls.2018.00676

Zhang, Z., Huang, J. J., Sun, D., Lee, Y. & Chen, F. (2017). Two-step cultivation for production of astaxanthin in Chlorella zofingiensis using a patented energy-free rotating floating photobioreactor (RFP). *Bioresource Technology*, *224*, 515–522. https://doi.org/10.1016/j.biortech.2016.10.081

Zhang, Z., Sun, D., Mao, X., Liu, J. & Chen, F. (2016). The crosstalk between astaxanthin, fatty acids and reactive oxygen species in heterotrophic Chlorella zofingiensis. *Algal Research*, *19*, 178–183. https://doi.org/10.1016/j.algal.2016.08.015

Zhang, Z., Sun, D., Zhang, Y. & Chen, F. (2019). Glucose triggers cell structure changes and regulates astaxanthin biosynthesis in Chromochloris zofingiensis. *Algal Research*, *39*(October 2018), 101455. https://doi.org/10.1016/j.algal.2019.101455

Zhong, Y. J., Huang, J. C., Liu, J., Li, Y., Jiang, Y., Xu, Z. F., Sandmann, G. & Chen, F. (2011). Functional characterization of various algal carotenoid ketolases reveals that ketolating zeaxanthin efficiently is essential for high production of astaxanthin in transgenic Arabidopsis. *Journal of Experimental Botany*, *62*(10), 3659–3669. https://doi.org/10.1093/jxb/err070

Zhu, S., Huang, W., Xu, J., Wang, Z., Xu, J. & Yuan, Z. (2014). Bioresource Technology Metabolic changes of starch and lipid triggered by nitrogen starvation in the microalga Chlorella zofingiensis. *BIORESOURCE TECHNOLOGY*, *152*, 292–298. https://doi.org/10.1016/j.biortech.2013.10.092

Zoschke, R., Chotewutmontri, P. & Barkan, A. (2017). Translation and co-translational membrane engagement of plastid-encoded chlorophyll-binding proteins are not influenced by chlorophyll availability in maize. *Frontiers in Plant Science*, *8*(March), 1–13. https://doi.org/10.3389/fpls.2017.00385

Zvereva, M. G., Klimova, L. A. & Semenko, V. E. (1981). Repression of RNA synthesis and breakdown in activity of chloroplast photochemical systems under the influence of 2-deoxy-d-glucose and hypertrophic accumulation of assimilates in Chlorella cells. *Soviet Plant Physiology*.