**Towards a synthetic nuclear transcription system in green algae:
Characterization of *Chlamydomonas reinhardtii* nuclear transcription factors and
identification of targeted promoters**

Melissa S. Anderson, Travis J. Muff, David R. Georgianna, Stephen P. Mayfield*

California Center for Algae Biotechnology and Division of Biological Sciences,
University of California, San Diego, 9500 Gilman Drive, La Jolla, California, United
States of America

*Corresponding author: smayfield@ucsd.edu

Abstract

Algae represent a unique opportunity to develop a sustainable production platform for many valuable products. A key hindrance to exploiting algae's potential as a bio-product manufacturing platform is our limited understanding of nuclear genetic regulatory systems. To begin to address this deficiency we identified prospective transcription factor–promoter pairs that have the potential for use in a controllable synthetic transcription system. We constructed a library of 92 recombinant nuclear transcription factors (TFs) from the model alga *Chlamydomonas reinhardtii*. Using a yeast one-hybrid assay, we identified TFs with the ability to activate transcription, and chose TF64 for further analysis. We generated two independent *C. reinhardtii* strains that constitutively expressed the gene encoding TF64 and used RNA-sequencing on these strains to identify transcripts that were altered by expression of the TF's gene. Many mRNAs were regulated by TF64, and we confirmed activation of transcription of nine genes from the *LHCBM* family by RT-qPCR. Overall, we were able to identify numerous potential TF–cognate promoter binding partners in *C. reinhardtii*, thus laying the groundwork to develop a controllable synthetic nuclear transcription system for microalgae, an essential tool for realizing their full biotechnological potential.

Key words: *Chlamydomonas*; transcription factor; nuclear gene expression; RNA-sequencing; yeast one-hybrid; synthetic biology.

1. Introduction

Algae are one of the most diverse organisms on the planet. They have evolved for over two billion years to adapt to a wide range of environmental habitats and conditions. This diversity, and the associated genetic complexity, make algae an attractive and potentially superior option for a plethora of modern day bio-manufacturing needs [1-4]. In addition to naturally producing many bio-active compounds such as antioxidants and omega-3 fatty acids [5], algae can be genetically engineered to produce an even greater array of recombinant proteins for animal and human therapeutics or industrial enzymes [6, 7], as well as lipid biofuel [8-10].

Chlamydomonas reinhardtii is a eukaryotic, green microalga and model organism for genetic research [11]. *C. reinhardtii* has been a successful resource for producing recombinant proteins including: cancer cell-targeted immunotoxins [12], malaria antigens for vaccines [13], and mammalian bioactive proteins useful as nutritional supplements [14], all from within the chloroplast. However, in order to fully realize the potential of this photosynthetic organism, genetic engineering tools for nuclear gene expression need to be developed. While this is an active area of research, with early work hinting promise [15-18], further studies are needed. Limited recombinant protein accumulation has notoriously hindered the field likely due to a combination of factors including: inefficient promoters, genome insertional effects on transgene expression (due to non-homologous recombination), and transgene silencing [6, 7]. Successes in this area would allow for greater post-translational modifications of proteins, eukaryotic

57 protein folding machinery, protein targeting to specific sub-cellular organelles, and
58 protein secretion [19-21].

59 In this study, we propose a new method for nuclear transgene expression
60 utilizing native transcription factors and endogenous promoters. Interestingly, only a
61 handful of nuclear transcription factors (TFs) and their regulatory mechanisms have
62 been characterized thus far [22-27]. With nearly 350 putative TFs encoded in the
63 nuclear genome [11, 22, 28], this is clearly an understudied area in a field with huge
64 potential and important implications.

65 We present data on the construction and analyses of the first recombinant library
66 of *C. reinhardtii* nuclear transcription factors (TFs). We studied the TFs' ability to be
67 produced, their potential DNA binding activity, and analyzed transcription profiles of the
68 genes they regulate in order to identify prospective native *C. reinhardtii* TF-cognate
69 promoter binding partners.

2. Materials and Methods

2.1 Algal Strains, Culture Conditions, and Genetic Transformation

Chlamydomonas reinhardtii cc1010 (*Chlamydomonas* Resource Center, St. Paul, NM) was used as the wild type strain for this study. Algal strains were cultured in TAP (Tris-Acetate-Phosphate) medium [29] at 23°C under constant illumination (100 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) provided by fluorescent lamps and with constant shaking (100 rpm). *C. reinhardtii* was transformed by electroporation as previously described [16] with the exception of the 40 mM sucrose supplement. Transformants were selected on TAP medium agar plates supplemented with 10 $\mu\text{g/ml}$ zeocin. Colonies were determined to be gene-positive after successful PCR amplification of the specific TF-encoding gene.

2.2 Generation of Transcription Factor Library

Initial gene models for 346 identified *C. reinhardtii* TFs were obtained from the PlnTFDB (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) [28, 30]. These were then cross-referenced by BLAST against the Phytozome database (<http://phytozome.jgi.doe.gov>) to obtain the most up-to-date and accurate gene models. Primers were designed to anneal to the 5' and 3' ends of each gene (Integrated DNA Technologies). RNA was isolated from cc1010 cultures grown to $\sim 1 \times 10^6$ cells per ml using PureLink Plant RNA Reagent (Ambion by Life Technologies) and cDNA libraries generated with Verso cDNA Synthesis Kit (Thermo Fisher Scientific). Gene coding sequences were amplified with Phusion Polymerase using the GC buffer (Thermo Fisher Scientific) supplemented with 0.5 to 1M Betaine (Sigma) with a touchdown PCR protocol [31]. Successfully amplified

CDSs were then cloned into the pENTR/D-TOPO vector in *E. coli* via TOPO cloning (Invitrogen). Resulting clones were sequence verified by Sanger sequencing. Silent mutations were deemed acceptable. In the case of non-silent mutations, these were allowed only after multiple independent clones were confirmed with the same difference(s) from the published gene model. Clones were transferred to pDEST22 (*S. cerevisiae* Y1H vector, Invitrogen) or pTM207 (*ble2A*–derived [16] for compatibility with GATEWAY cloning, *C. reinhardtii* nuclear expression vector) via Gateway LR-Clonase (Invitrogen).

2.3 Yeast Culture Conditions and Yeast One-Hybrid Assay

Culture conditions and mating of *Saccharomyces cerevisiae* YM4271 and Y1H assay were performed following the MATCHMAKER One-Hybrid System protocol (CLONTECH Laboratories, Inc.). Reporter plasmids were chromosomally integrated in the *S. cerevisiae* YM4271 genome. Briefly, Y1H library strains were inoculated into 96 well plates and cultured overnight (O/N). The OD₆₀₀ was measured from 100 µl of O/N culture. Using white 96 well plates (Greiner), 50 µl of O/N culture was combined with 50 µl of Luciferase Assay Reagent (Promega) using an injector on a Tecan plate reader (Tecan Infinite M200 PRO). Luminescence was measured five seconds post-injection. Luminescence was first normalized to the OD₆₀₀ and then for each TF normalized to the empty vector control. The data presented are a representative single experiment for each TF-promoter pair tested. A one sided Grubbs' test for outliers (0.05 level) was used to determine fold increases in luminescence that were outside the distribution. Assays were repeated with replicates for outlier samples. Values were determined

significant by Student's *t*-test and/or greater than two standard deviations from the mean of the empty vector luminescence control.

2.4 Immunoblotting

Cells were cultured until mid to late log phase ($\sim 1 \times 10^6$ cells per ml), washed in PBS-T (Phosphate-Buffered Saline-Tween) buffer, and lysed by sonication. Total soluble protein was normalized to 100 ug per lane as a loading control. Protein pellets were resuspended in SDS-PAGE loading buffer. Boiled samples were separated on a 12% SDS-PAGE gel, transferred to nitrocellulose for 1-2 hrs in transfer buffer, and probed with anti-GAL4-AD antibody (Sigma) for *S. cerevisiae* or anti-FLAG monoclonal antibody conjugated to alkaline phosphatase (Sigma) for *C. reinhardtii*.

2.5 RNA Purification

RNA was extracted from *C. reinhardtii* strains of interest after 3–4 days of growth ($\sim 1 \times 10^6$ cells per ml) in TAP medium under constant light using PureLink Plant RNA Reagent (Ambion by Life Technologies) according to the manufacture's protocol. RNA was treated with 4 U of TURBO DNase (Thermo Fisher Scientific) for 30 min at 37°C.

2.6 RNA-Sequencing and Analysis

RNA from three biological replicates (independent liquid cultures) for each strain analyzed (cc1010::TF64-7, cc1010::TF64-9, and cc1010::GFP) was sent to the Institute for Genomic Medicine at the University of California, San Diego for Next-Generation Sequencing on an Illumina HiSeq2500. Single-end 50 bp reads were generated. Reads

were aligned to the latest reference index (Chlre4_Augustus5_transcripts.fasta) downloaded from the Joint Genome Institute (JGI) at www.phytozome.net using TopHat open software on Galaxy (usegalaxy.org) [32-34]. Differential expression analysis was performed using Cufflinks also on Galaxy. For gene identification, *C. reinhardtii* strain 503 was used as a reference strain due to the lack of a published sequence for strain cc1010. The average log₂ (fold change) of all FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values ≥ 1.0 for the experimental strain (transcription factor constitutive-expression) compared to the control strain (GFP constitutive-expression strain) was plotted.

2.7 Reverse Transcriptase Quantitative Polymerase Chain Reaction

1 μ g of total RNA was reverse transcribed using the Verso cDNA synthesis kit (Thermo Fisher Scientific). cDNA was diluted 1:2 for qPCR analysis using Power SYBR Green PCR Master Mix (Applied Biosystems). qPCR was performed on a My iQ thermocycler (Bio Rad). Two biological replicates were performed each with technical triplicates. The $\Delta\Delta$ Ct method was used for relative quantification of gene expression [35]. *RACK1* was used as an internal standard. The mean log₂ (fold change) from each biological replicate and the mean was plotted.

2.8 Promoter Motif Identification

Promoter sequences were obtained from NCBI. DNA sequences were analyzed using the software programs MEME [36, 37], AME [38], and Jalview [39].

3. Results

3.1 Construction of a putative transcription factor library

One of our main goals with this project was to narrow down the list of potential cognate TF–promoter pairs, i.e., which TFs bind and regulate which nuclear promoters, in *C. reinhardtii*. An understanding of the global network of regulatory interactions within the nuclear genome is critical for the engineering of synthetic transcription systems, a long-range goal for our laboratory. Therefore, we set out to construct a library of recombinant *C. reinhardtii* nuclear transcription factors (TFs). Just after the *C. reinhardtii* genome sequence was completed [11], putative TFs, as well as transcription regulators (TRs), were identified by presence of homology to known TF/TR domains and available at the Plant Transcription Factor Database (PInTFDB) [28, 30]. In order to have the most up-to-date gene model for the TFs and TRs, we took fragments from the identified genes and used a BLAST search against the latest gene models from Phytozome. The TF/TR library (referred to simply as the TF library from here on) was generated using TOPO cloning such that the gene encoding each TF was PCR amplified from *C. reinhardtii* cc1010 cDNA and ligated into the pENTR/D-TOPO vector, followed by transformation into *Escherichia coli* (see Materials and Methods). We were able to successfully construct plasmid vectors encoding 92 different putative TFs predicted in the *C. reinhardtii* genome (Supp. Table 1) (from a total of over 300 TFs identified by bioinformatics). Our library contains TFs belonging to multiple TF families including but not limited to: High Mobility Group (HMG) box, basic Helix-Loop-Helix (bHLH), Cys₂His₂ zinc finger (C2H2), Cys₃His zinc finger (C3H), Forkhead-associated (FHA), basic

Leucine Zipper (bZIP), MYB (myeloblastosis), Gcn5-related *N*-acetyltransferase (GNAT), Tubby bipartite (TUB), Tumor necrosis factor receptor-associated (TRAF), SET (histone methyltransferases), and CCAAT-enhancer-binding proteins (CCAAT). A complete list of each TF and relevant information can be found in Supplemental Table 1.

3.2 Putative transcription factors initiate transcription from *C. reinhardtii* promoters in yeast

As an initial screen for potential DNA-binding activity, we performed a high-throughput yeast one-hybrid (Y1H) assay to test our TFs' ability to activate transcription from known *C. reinhardtii* promoters [40, 41]. We transferred our entire pENTR-TF library to the Y1H vector pDEST22 via Gateway LR-transferase which allowed the TFs to be fused to the yeast GAL4 transcription activation domain [42]. Separately, "bait" promoters of interest were cloned (in 300 base pair (bp) fragments, labeled A, B, and C (5' to 3'), for a total of 900 bps per promoter (Supp. Table 2)) 5' to a yeast minimal promoter element followed by the reporter gene *Gaussia luciferase* [43]. Each TF-vector was transformed into separate haploid *Saccharomyces cerevisiae* YM4271 cells and crossed against the opposite mating type of strains harboring DNA bait promoters of interest. *S. cerevisiae* strains producing each TF were also cultured so whole cells could be processed for immunoblot analysis of TF protein production (Supp. Fig. 1).

We assayed all 92 TFs against five *C. reinhardtii* nuclear promoters: *LCIC*, *LCI5*, *SEBP1*, *Nar1.2*, and *LHCBM5* (Fig. 1). *LCIC*, *LCI5*, and *Nar1.2* are low CO₂-induced genes that play roles in the CO₂-concentrating mechanism (CCM) [44-46]. *SEBP1*

encodes sedoheptulose-1,7-bisphosphatase which functions during the Calvin cycle [47]. *LHCBM5* encodes a component of light harvesting complex II and is involved in photosynthesis [48]. These genes were chosen because they were identified from a published RNA-sequencing dataset as highly expressed under laboratory conditions in *C. reinhardtii* [49].

TFs 2, 3, 9, 28, 34, 45, 64, 69, and 81 each activated transcription from *LC/C* promoter fragment C (Fig. 1). TF64 activated transcription from *LC/5* promoter fragment A; TFs 39 and 78 activated transcription from *LC/5* promoter fragment C. TFs 3, 6, 27, 30, and 64 activated transcription from *SEBP1* promoter fragment A; TF64 activated transcription from *SEBP1* promoter fragment B; TFs 27, 30, 56, and 64 activated transcription from *SEBP1* promoter fragment C. TFs 10, 30, and 64 activated transcription from *Nar1.2* promoter fragment C. Finally, TF34 activated transcription from *LHCBM5* promoter fragment C (Fig. 1). Note that *LHCBM5* promoter fragment B was unable to be cloned (due to repeat sequences) and therefore was not assayed here. (See Materials and Methods for statistical information on Y1H assay.)

To summarize these Y1H assays, our data provide information on 1,288 TF–promoter potential binding interactions, 26 of which were positive hits. TF64 was the most active in this assay, activating transcription with four of the five promoters tested. TFs 3, 30, and 34 each activated transcription from two promoters. Note that some TFs bound multiple fragments of the same promoter. Many TFs however did not show activity with any of the five *C. reinhardtii* promoters we assayed. These data are summarized in Table 1.

3.3 Putative transcription factors initiate transcription from orthologous promoters from multiple species

We also assayed our TF library with bait promoters from algal species *Volvox carteri* and *Chlorella vulgaris*, as well as from the plant species *Arabidopsis thaliana* and *Zea mays*. Again, we tested promoters *LCIC*, *LCI5*, *SEBP1*, *Nar1.2*, and *LHCBM5* (Table 1, Supp. Fig. 2). Like the *C. reinhardtii* promoter data, TF64 was the most active in activating transcription in combination with promoter fragments from other species, specifically *V. carteri* *LCIC*, *LCI5*, *SEBP1*, and *LHCB5*; *C. vulgaris* *SEBP1*; and *Z. mays* *SEBP1* and *LHCBM5* (Table 1, Supp. Fig. 2). In full we analyzed 49 promoter fragments against 92 TFs for a total of 4,508 potential binding interactions. We found 65 positive hits and, most importantly, 28 TFs with potential DNA binding activity.

3.4 Analysis of potential TF64-binding promoters identified from the Y1H assay

Utilizing the collection of our Y1H data, we hypothesized we could identify commonalities among promoters which may function as specific motifs or binding sites important for gene regulation. We chose to analyze the promoter fragments that activated transcription in combination with TF64 because it provided us with the largest sample size, 13 promoter fragments in total. We used the software program MEME (Multiple Em (Expectation maximization) for Motif Elicitation) [36, 37] to search for enriched DNA motifs. Unfortunately, no statistically significant motifs were identified. The top motif found was an 11 nucleotide sequence, TGNGCANNTNN (Supp. Fig. 3A). Interestingly, this motif does contains remnants of the canonical binding site, CANNTG

(nucleotides 5–10) (Supp. Fig. 3B), typical for the basic Helix-Loop-Helix family of transcription factors that TF64 belongs to [50, 51].

3.5 Constitutive expression of the TF library in *C. reinhardtii*

We next attempted to study our TF library expressed in *C. reinhardtii* cc1010. The gene encoding each TF was cloned from the pENTR vector into a *ble*-2A expression vector [16], pTM207 (see Fig. 2B). This expression vector results in co-transcription of a gene of interest along with the *ble* gene (conferring zeocin resistance) followed by post-translation cleavage into two proteins at the 2A linker peptide site. Each pTM207 plasmid encoding a unique TF under control of the constitutive promoter P_{AR1} was electroporated into the *C. reinhardtii* nuclear genome. However, we were unable to obtain colonies of *C. reinhardtii* constitutively expressing the genes encoding most TFs. While we attempted transformation of all 92 TFs, gene-positive colonies were only recovered for 59 TFs, and only 21 TFs (1, 2, 4, 5, 14, 22, 31, 34, 38, 40, 41, 47, 52, 53, 55, 62, 63, 64, 75, 76, 84) had over 20% gene-positive colonies of those tested (data not shown). Immunoblot analyses of whole cell lysates were performed to verify production of the TFs, however protein was detected only in strains transformed with TFs 1, 2, 5, 13, 22, 31, 40, and 64 (data not shown).

In deciding which TF to carry forward with our study, we considered our Y1H data concurrently with our limited ability to produce the recombinant TFs in *C. reinhardtii*. TFs 2 and 64 both showed potential DNA binding activity and were capable of being constitutively produced in *C. reinhardtii*. Ultimately, we chose TF64 to continue our study of TF–promoter binding partners in *C. reinhardtii*.

3.6 Production of TF64 in *C. reinhardtii*

Basic Helix-Loop-Helix (bHLH) transcription factor family members, like TF64, are highly conserved in their functional and DNA-binding domains, even across distantly related species and genera [50-53]. They recognize a canonical binding site, CANNTG (called the E-box), in promoters of genes they regulate [51, 53]. A BLAST search of the PInTFDB TF64 sequence showed conservation in DNA binding, E-box specificity site, and dimerization interface domains among top hits of TF-like proteins from other microalgal species (Fig. 2A). The remainder of the TF64 protein sequence is highly variable with the exception of a conserved ACT domain in the C-terminus of unknown function typically found in bacterial species [54] (Supp. Fig. 4).

We generated multiple strains of cc1010 that constitutively produced TF64 (cc1010::TF64-4, -7, -8, -9, and -11) shown by immunoblot (Fig. 2B, 2C). Total protein from whole cell lysates was normalized to 100 ug per lane for a loading control. The pTM207 vector encodes an N-terminal 3xFLAG-tag fused to each TF, and the TF64 proteins were detected using antibodies against FLAG-tag. TF64 is predicted to be a 33 kDa protein (Fig. 2C). The 3xFLAG-tag adds 3 kDa to the protein product. The higher molecular weight band is the Ble-2A-3xFLAG-tag-TF64 fusion product prior to 2A cleavage, approximately 50 kDa (the *ble* protein product is 12 kDa, 2A encodes a 2 kDa peptide). Through multiple immunoblot analyses, strain cc1010::TF64-7 appeared to produce the least amount of transcription factor protein, and strain cc1010::TF64-9 appeared to produce the most amount of protein (Fig. 2C).

As a control, we also used the pTM207 vector to generate a strain that constitutively produced 3xFLAG-tag-GFP fusion protein under control of P_{AR1} (Fig. 2B). The full length Ble-2A-3xFLAG-tag-GFP protein (~44 kDa) from a whole cell lysate of strain cc1010::GFP is shown on the immunoblot in Figure 2C.

Growth curves were performed on strains cc1010::TF64-7, cc1010::GFP, and wild type cc1010 cultured in TAP medium under constant light ($100 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) for four days to determine if either mutation had an effect on growth (Fig. 2C, 2D). While cc1010::TF64-7 did exhibit an extended lag phase during culture, it was capable of reaching an OD_{750} similar to that of the wild type cc1010 strain and cc1010::GFP (Fig. 2C, 2D).

3.7 TF64 regulates many endogenous nuclear genes

To identify the genes/promoters TF64 regulates in *C. reinhardtii*, we performed an RNA-sequencing experiment on two independent strains, cc1010::TF64-7 (referred to as the low-constitutive strain) and cc1010::TF64-9 (referred to as the high-constitutive strain), along with our control strain cc1010::GFP (Fig. 3, Complete RNA-Seq dataset in Supp. Table 3). RNA from three biological replicates for each strain was sequenced at the UCSD Institute for Genomic Medicine. Transcript abundance and differential expression analysis for each TF64-producing strain was compared to the GFP-producing strain (Fig. 3A). The data indicate that approximately 2.4% and 1.0% of the genome was affected at least 10-fold ($\log_2 \geq 3.33$) either directly or indirectly in strains cc1010::TF64-7 and cc1010::TF64-9, respectively. An approximately equal number of genes were activated as were inhibited in both datasets. While in a general sense,

genes that were activated in one dataset were also activated in the other, and genes that were inhibited in one dataset were also inhibited in the other (Fig. 3B, purple trend line slope = 0.56), the two datasets exhibit distinct variability (Fig. 3B, $R^2 = 0.498$). Furthermore, a greater range of regulation was observed in the low-constitutive strain (TF64-7) compared to the high-constitutive strain (TF64-9) (Fig. 3A, 3B, 3C).

The most highly regulated genes, both activated and inhibited, from the low-constitutive and high-constitutive TF64-producing strains were identified by bioinformatics using the BLASTx search function from NCBI (Supp. Table 4). Inhibited genes were mostly uncharacterized and showed little similarity in function. Activated genes, particularly from the low-constitutive TF64-7 dataset, fell into relatively distinct functional categories including: photosynthesis, cell structure, cell cycle, and metabolism. Table 2 lists the top 20 activated genes (that have also been previously characterized) identified from the TF64-7 RNA-Seq data. These data suggest TF64, like many bHLH transcription factor family members [55, 56], regulates genes involved in a wide variety of developmental and cellular processes in *C. reinhardtii*.

3.8 Bioinformatic analysis of promoters of genes regulated by TF64

We chose three sets of promoters, TF64-activated, TF64-inhibited, and TF64-non-regulated, from the low-constitutive TF64-7 RNA-Seq dataset to analyze for common motifs. Promoters included 1,000 bps 5' to the ATG translation start site of the 30 top activated, inhibited, and non-regulated ($\log_2 = 0$) genes. Most genes did not have annotated 5' UTRs. Promoters from each regulatory category were analyzed by MEME to identify any common motifs, however no statistically significant sequences were

found for any group. Additionally, we used the program AME (Analysis of Motif Enrichment) [38] to determine if the bHLH canonical binding site, CANNTG, was present with statistical significance, and it was not for any of the three promoter categories.

We further analyzed the promoter groups using the alignment software Jalview [39]. Promoters were aligned without gaps and all CANNTG sequences were identified (Supp. Fig. 5) for each group. Analysis of CANNTG composition as well as relative location within the promoter did not reveal significant differences among the three promoter groups analyzed. These data suggest that the CANNTG sequence is ubiquitous throughout the *C. reinhardtii* genome. While this motif may play a role in TF64-DNA binding, it is not solely responsible for the gene regulation observed in the TF64–constitutive expression strains. It is likely that other co-factors and/or regulatory elements are important for transcription of the genes we identified to be regulated by TF64, further underscoring the complex nature of nuclear gene regulation in eukaryotic microalgae.

3.9 TF64 activates transcription of light harvesting complex II components

To validate our RNA-Seq analysis, we performed reverse transcriptase quantitative PCR (RT-qPCR) on selected genes (primer sequences listed in Supp. Table 5). Strains cc1010::TF64-7 and cc1010::GFP were cultured in TAP medium under constant light for three days until mid-log phase growth was reached. RNA was isolated from cells and cDNA was synthesized for RT-qPCR analysis. Among the top activated genes from the TF64-7 RNA-Seq dataset were *LHCBM7*, *LHCBM8*, *LHCBM4*, and

LHCBM1 (Table 2) of light harvesting complex II (PSII) [48]. We were able to confirm that transcripts from these genes were approximately 16 times (for *LHCBM7*), four times (for *LHCBM8* and *LHCBM4*), and eight times (for *LHCBM1*) more abundant in the TF64–producing strain compared to the GFP–producing strain by RT-qPCR (Fig. 4A). Furthermore, genes *LHCBM5*, *LHCBM2*, *LHCBM3*, *LHCBM6*, and *LHCBM9* also of PSII [48] were additionally analyzed and found to be activated in the TF64–producing strain (Fig. 4A). Interestingly, the promoter of gene *LHCBM5* was assayed in our Y1H screen but was not detected to activate transcription with TF64 in yeast. Figure 4A shows transcript abundance data for each of these genes by RNA-Seq and RT-qPCR. These data indicate TF64 plays a role in activating PSII components and possibly regulation of photosynthesis.

The nine PSII promoters were analyzed similarly to those previously discussed. Again, MEME did not identify any new motifs, CANNTG was not present with statistical significance determined by AME (data not shown), and CANNTG composition and location were not different from any group of promoters analyzed from the RNA-Seq selected promoters (Supp. Fig. 6).

3.10 Transcription analysis of Y1H–assayed genes

We also investigated transcription of the genes whose promoters were found to activate transcription with TF64 by Y1H (i.e., *LCI5*, *SEBP1*, *LCIC*, and *Nar1.2*). RNA-Seq data indicated that each of these genes were down-regulated in *C. reinhardtii* cells constitutively expressing the gene encoding TF64 (Fig. 4B, Supp. Table 4). By RT-qPCR, we confirmed that transcription of the genes *LCI5*, *SEBP1*, and *LCIC* were in fact

390 inhibited by constitutive expression of the gene encoding of TF64. *Nar1.2*, however,
391 was activated in our RT-qPCR analysis (Fig 4B). Overall, these data support our RNA-
392 Seq analysis.

393 Collectively, these results highlight the nature of high-throughput screens, like the
394 Y1H, and high-throughput sequencing data, as generated here by RNA-sequencing:
395 they produce large amounts of data that can serve as an excellent starting point for
396 narrowing down potential molecular interactions of interest. Here, we successfully used
397 these two screens to identify potential TF-promoter binding partners in *C. reinhardtii*.

398

4. Discussion

In this study, we sought to present research on the early stages of a novel method for constructing a synthetic nuclear transcription system in *C. reinhardtii*, i.e., using native transcription factor-promoter pairs to regulate transgene expression. We successfully constructed a recombinant transcription factor library that includes 92 (nearly one third of the putative) transcription factors (TFs) encoded by the nuclear genome of *C. reinhardtii*. To date, relatively few TFs have actually been characterized from this species of microalgae [22]. We analyzed the 92 TFs' ability to activate transcription via a yeast one-hybrid screen, studied the TFs' abilities to be constitutively expressed in their native organism *C. reinhardtii*, and finally assessed transcription profiles by RNA-Seq from two independent strains constitutively expressing one specific TF (TF64). These high-throughput studies were designed to narrow down the vast amount of hypothetical transcription factor–promoter binding pairs in *C. reinhardtii* (~350 TFs x ~15,000 nuclear genes = 5,250,000 potential interactions). Our results establish a clear direction for investigation of direct binding partners that could be used in an engineered synthetic nuclear transcription system in green algae.

Using a yeast one-hybrid assay [41], we were able to analyze 4,508 potential binding interactions between TFs and promoter fragments. Sixty-five of these were found to be positive hits correlating with 28 TFs with potential DNA binding activity. We assayed five promoters (*LCIC*, *LCI5*, *SEBP1*, *Nar1.2*, and *LHCBM5*) in different combinations from *C. reinhardtii*, *V. carteri*, *C. vulgaris*, *A. thaliana*, and *Z. mays*. The ability to activate transcription from unique DNA sequences by a number of the putative

TFs analyzed support the bioinformatic data [28] suggesting these proteins are in fact functional transcription factors, capable of regulating transcription in *C. reinhardtii*.

Compiling the yeast one-hybrid data, we sought to identify common motifs among promoter fragments found to activate transcription in combination with an individual TF. The promoters, however, proved to be more cryptic than anticipated. We studied TF64-associated promoters, 13 sequences in total, and were unable to identify commonalities by bioinformatics. It may be that a larger number of promoters need to be analyzed before such a characterization is possible. In the future, it would be interesting to compare DNA sequences from a larger dataset of *C. reinhardtii* promoters and also determine if identified motifs were conserved in the promoters of other closely or distantly related species.

Our TF library was cloned into a *C. reinhardtii* constitutive expression vector for production in *C. reinhardtii*. To our knowledge, this was the first attempt to constitutively produce a recombinant library of native TFs in *C. reinhardtii*. Of the 92 TF-encoding vectors that were transformed, only eight resulted in successful production of protein under the conditions attempted. As almost all of the TFs produced protein in *S. cerevisiae*, the algae expression data suggest that the failure for most TFs to produce protein in *C. reinhardtii* is possibly due either to adverse effects of constitutively expressing their genes or transgene silencing. It is possible these TFs could be produced under more tightly controlled experimental conditions, or when placed under inducible or conditional expression systems.

TF64 was our most successful TF in that it was able to be produced in multiple strains of *C. reinhardtii* and it was the most active TF in the yeast one-hybrid assay.

445 From RNA-sequencing data on strains constitutively producing TF64, compared to a
446 GFP-constitutive strain, we were able to determine that TF64 likely plays a role in
447 regulating transcription of genes involved in multiple cellular and developmental
448 processes in wild type *C. reinhardtii*. Constitutive production of TF64 led to an increase
449 in transcript levels of a subset of genes functioning in photosynthesis and the cell cycle,
450 as well as many others. Follow-up studies on the biological role of TF64 should prove to
451 be interesting from a basic science perspective, leading to greater insights into the *C.*
452 *reinhardtii* lifecycle.

453   Here, for our first description and characterization of our TF library, we cultured
454 *C. reinhardtii* in rich medium. It is a possibility that rich medium up-regulates the many
455 cellular processes identified in our RNA-Seq experiment, like photosynthesis, the cell
456 cycle, and metabolism. Comparing the TF64-expressing strains to the GFP-expressing
457 strain, however, in both the RNA-Seq and RT-qPCR experiments helps to minimize this
458 caveat. Furthermore, it is of note that while such processes like photosynthesis and the
459 cell cycle were up-regulated in the TF64-producing strains, we did not observe an
460 increase in growth rate. There are at least two non-mutually exclusive possibilities as to
461 why: 1) constitutive expression of a TF can slow growth, and 2) many of the down-
462 regulated genes identified by RNA-Seq were uncharacterized and we do not know their
463 function which could also act to slow growth. In follow-up studies we would modify
464 culture conditions to various types of minimal medium to analyze a particular TF-
465 promoter pair or as an alternative method to help with expression of the TFs that we
466 were unable to express in rich medium.

Our goal with this study was to identify potential cognate transcription factor–promoter pairs from *C. reinhardtii* that, once validated, could be used in a synthetic nuclear transcription system. From our yeast one-hybrid data, we identified 28 TFs with possible DNA binding activity. Further studies are required to confirm these interactions *in vivo* in *C. reinhardtii*. Specifically focusing on TF64, we were able to verify the activation of transcription of nine genes, *LHCBM1–9*, by both RNA-Seq and RT-qPCR. It is yet to be determined if this gene activation is in fact due to a direct TF–promoter binding interaction.

While these data are clearly just the beginning to fully understanding regulatory networks among nuclear genes in *C. reinhardtii*, they do lay the groundwork for the construction of a synthetic transcription system. In the future, it is our goal to continue this line of work to provide the scientific community the necessary tools for sophisticated and robust genetic engineering in microalgae.

Figure Legends

Figure 1. *C. reinhardtii* TF library tested for transcription activation from select promoters via yeast one-hybrid assay. Y1H assay performed with all 92 TF library proteins against five *C. reinhardtii* promoters (*LCIC*, *LCI5*, *SEBP1*, *Nar1.2*, and *LHCBM5*), each in 300 bp fragments (labeled A, B, and C). Functional read out was expression of the *lux* gene. Red data points indicate statistical significance of increased *lux* expression compared to an empty vector control (see Materials and Methods). x axes: TF, transcription factor library number.

Figure 2. Basic Helix-Loop-Helix transcription factor alignment, strain construction and growth. A) Protein sequence alignment of TF64-related proteins. The *C. reinhardtii* TF64 sequence from the PlnTFD was used as a query in a BLAST search for related proteins. Selected top hits are shown. *C. reinhardtii* strain 503 (in bold, used as a reference strain in this study due to the lack of a published sequence for strain cc1010) was among the top hits. Proteins from other related algal species are also shown. Alignment is focused on the basic Helix-Loop-Helix region. Functionally important conserved residues are indicated by color. B) Schematic of the pTM207 vectors used to constitutively express the genes encoding TF64 and GFP. The *b1e* gene confers zeocin resistance and 2A encodes a linker peptide that is cleaved post-translationally. The pTM207 vector also encodes an N-terminal 3xFLAG-tag fused to each gene of interest. C) Immunoblot of whole cell lysates of wild type (WT) *C. reinhardtii* and engineered strains producing TF64 (64-4, 64-7, 64-8, 64-9, 64-11) or

GFP, separated by SDS-PAGE and probed with anti-FLAG antibody. Total protein was normalized to 100 ug per lane. For the TF64 sample lanes, the higher molecular weight products are the full length fusion protein Ble-2A-3xFLAG-tag-TF64 (~50 kDa), prior to 2A cleavage. Cleaved 3xFLAG-tag-TF64 is ~36 kDa. Protein in the GFP lane is full length Ble-2A-3xFLAG-tag-GFP (~44 kDa). D) Growth curves of wild type (cc1010) *C. reinhardtii* and strains producing TF64 (TF64-7) or GFP, cultured for four days in TAP medium under constant light ($100 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$). Growth was measured at OD₇₅₀. Data is plotted from three biological replicates with the SEM for each strain. The “Exponential Growth” graph indicates the slope of the line during log phase growth for each strain by color.

Figure 3. RNA-sequencing data from two strains constitutively producing either low or high amounts of TF64. A) Differential transcription analysis of strains cc1010::TF64-7 and cc1010::TF64-9 compared to cc1010::GFP by RNA-sequencing. For each unique read, the log₂ of the fold change FPKM value of the experimental strain over the control strain (i.e., FPKM value for the TF64-producing strain divided by the FPKM value for the GFP-producing strain) was plotted (see Materials and Methods, as well as Supplemental Table 3). B) Comparison of RNA-Seq data from each TF64-producing strain (TF64-7 and TF64-9). Each data point represents a unique read. The log₂ (fold change) was plotted. Purple line represents the best-fit line for all data, $R^2 = 0.498$, slope = 0.560. C) Heat map of expression profiles from the top 20 activated and inhibited genes and Y1H-assayed genes in strains cc1010::TF64-7 and cc1010::TF64-9 compared to cc1010::GFP. Units for heat map key values are log₂ (fold change). Genes

of interest are labeled below the heat map. RNA-sequencing data was compiled from three biological replicates.

Figure 4. Transcription regulation of light harvesting complex II components and Yeast One-Hybrid–assayed genes by TF64. Expression data for A) genes *LHCBM1–9* and B) genes *LCI5*, *SEBP1*, *LCIC*, *NAR1.2* from strain cc1010::TF64-7 compared to cc1010::GFP analyzed by RT-qPCR and RNA-Seq. The log₂ (fold change) was plotted. RT-qPCR data is from two biological replicates; individual data points are shown with the mean. RNA-Seq data is the average of three biological replicates. Note that there were multiple unique reads for certain genes.

537 **Table 1.** Yeast one-hybrid data summary

Species	Promoter	Transcription Factor
<i>Chlamydomonas reinhardtii</i>	<i>SEBP1</i>	3,6,27,30,56,64
<i>Chlamydomonas reinhardtii</i>	<i>LCI5</i>	39,64,78
<i>Chlamydomonas reinhardtii</i>	<i>LCIC</i>	2,3,9,28,34,45,64,69,81
<i>Chlamydomonas reinhardtii</i>	<i>NAR1.2</i>	10,30,64
<i>Chlamydomonas reinhardtii</i>	<i>LHCBM5</i>	34
<i>Volvox carteri</i>	<i>SEBP1</i>	64
<i>Volvox carteri</i>	<i>LCI5</i>	2,64
<i>Volvox carteri</i>	<i>LCIC</i>	2,21,45,57,64,69
<i>Volvox carteri</i>	<i>NAR1.2</i>	2,3,4,5,13
<i>Volvox carteri</i>	<i>LHCBM5</i>	58,64
<i>Chlorella vulgaris</i>	<i>SEBP1</i>	64
<i>Chlorella vulgaris</i>	<i>LCIC</i>	10
<i>Chlorella vulgaris</i>	<i>NAR1.2</i>	7
<i>Chlorella vulgaris</i>	<i>LHCBM5</i>	2,7,18,27,51
<i>Zea mays</i>	<i>SEBP1</i>	30,64
<i>Zea mays</i>	<i>LHCBM5</i>	2,6,14,28,37,64,76
<i>Arabidopsis thaliana</i>	<i>SEBP1</i>	56
<i>Arabidopsis thaliana</i>	<i>LHCBM5</i>	85

538

539 **Table 2.** Top 20 up-regulated genes in *C. reinhardtii* cc1010::TF64-7

No.	Gene ID	Log2 Fold Change	Gene Symbol	Function	Category
1	jgi Chlre4 513883 au5.g4042_t1:0-146	7.58	<i>LHCBM7</i>	Chlorophylla- <i>b</i> binding protein of LHCII	Photosynthesis
2	jgi Chlre4 523567 au5.g13085_t1:285-1460	7.09	<i>LHCBM8</i>	Chlorophylla- <i>b</i> binding protein of LHCII	Photosynthesis
3	jgi Chlre4 523561 au5.g13079_t1:149-1280	6.81	<i>LHCBM4</i>	Chlorophylla- <i>b</i> binding protein of LHCII	Photosynthesis
4	jgi Chlre4 518507 au5.g8360_t1:204-4111	6.80	<i>FAP211</i>	Flagellar associated protein	Motility
5	jgi Chlre4 521087 au5.g10761_t1:39-2944	6.66	<i>METE</i>	Cobalamin-independent methionine synthase	Metabolism

6	jgi Chlre4 521595 au5.g11226_t1:266-2760	6.09	<i>SAH1</i>	S-Adenosyl homocysteine hydrolase	Metabolism
7	jgi Chlre4 515402 au5.g5474_t1:537-2854	5.79	<i>PHC13</i>	Cell wall protein pherophorin-C13	Cell structure
8	jgi Chlre4 520083 au5.g9823_t1:3664-4112	5.77	<i>GCP3</i>	Gamma tubulin interacting protein	Cell structure
9	jgi Chlre4 520120 au5.g9859_t1:2-129	5.39	<i>LHCBM1</i>	Chlorophylla- <i>b</i> binding protein of LHCII	Photosynthesis
10	jgi Chlre4 524285 au5.g13771_t1:3002-3795	5.34	<i>MCM4</i>	Minichromosome maintenance protein 4	Cell cycle
11	jgi Chlre4 512084 au5.g2359_t1:10431-10587	5.15	<i>DCL2</i>	Dicer-like protein	Regulation
12	jgi Chlre4 512529 au5.g2782_t1:35-1932	5.11	<i>GAP1</i>	Glyceraldehyde 3-phosphate dehydrogenase	Metabolism
13	jgi Chlre4 518966 au5.g8779_t1:1773-1883	5.09	<i>SYPT2</i>	Qc-SNARE protein, SYP7-family	Localization
14	jgi Chlre4 519390 au5.g9173_t1:283-2258	5.07	<i>FTSZ1</i>	Plastid division protein	Cell cycle
15	jgi Chlre4 515943 au5.g5981_t1:176-2507	4.99	<i>FTSZ2</i>	Plastid division protein	Cell cycle
16	jgi Chlre4 520083 au5.g9823_t1:4197-4255	4.79	<i>GCP3</i>	Gamma tubulin interacting protein	Cell structure / Localization
17	jgi Chlre4 523024 au5.g12580_t1:45-2087	4.75	<i>EFG8</i>	Mitochondrial translation factor Tu	Translation
18	jgi Chlre4 513496 au5.g3676_t1:1531-1934	4.73	<i>GLN3</i>	Glutamine synthetase	Metabolism
19	jgi Chlre4 513333 au5.g3525_t1:167-1848	4.69	<i>MIND1</i>	Chloroplast septum site-determining protein	Cell cycle
20	jgi Chlre4 520302 au5.g10033_t1:278-1558	4.66	<i>TEF13</i>	Aminoacyl-tRNA synthase CAAD domain	Localization

540

541

Acknowledgments: The authors gratefully acknowledge Dr. Jose Pruneda-Paz (UCSD) for assistance with the Yeast One-Hybrid assay design and use of robotic equipment. We also thank members of the Mayfield laboratory, particularly Joseph Ostrand and William Ansari, for insightful discussion.

Author Contributions: MSA, TJM, DRG, SPM conceived the research; MSA, TJM, DRG performed the experiments; MSA, TJM, DRG analyzed the data; MSA wrote the manuscript; TJM, DRG, SPM edited the manuscript.

Funding: This work was supported by a grant from the US Department of Energy (DE-EE0003373; SPM).

References

1. Blunt JW, Copp BR, Keyzers RA, Munro MH, Prinsep MR. Marine natural products. *Nat Prod Rep.* 29(2):144-222. Epub 2011/12/24. doi: 10.1039/c2np00090c. PubMed PMID: 22193773.
2. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, et al. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 2008;9(5):R90. Epub 2008/05/30. doi: gb-2008-9-5-r90 [pii] 10.1186/gb-2008-9-5-r90. PubMed PMID: 18507822; PubMed Central PMCID: PMC2441476.
3. Parker MS, Mock T, Armbrust EV. Genomic insights into marine microalgae. *Annu Rev Genet.* 2008;42:619-45. Epub 2008/11/06. doi: 10.1146/annurev.genet.42.110807.091417. PubMed PMID: 18983264.
4. Gimpel JA, Specht EA, Georgianna DR, Mayfield SP. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr Opin Chem Biol.* 17(3):489-95. Epub 2013/05/21. doi: S1367-5931(13)00061-6 [pii] 10.1016/j.cbpa.2013.03.038. PubMed PMID: 23684717.
5. Cardozo KH, Guaratini T, Barros MP, Falcao VR, Tonon AP, Lopes NP, et al. Metabolites from algae with economical impact. *Comp Biochem Physiol C Toxicol Pharmacol.* 2007;146(1-2):60-78. Epub 2006/08/12. doi: S1532-0456(06)00153-0 [pii] 10.1016/j.cbpc.2006.05.007. PubMed PMID: 16901759.
6. Rosales-Mendoza S, Paz-Maldonado LM, Soria-Guerra RE. *Chlamydomonas reinhardtii* as a viable platform for the production of recombinant proteins: current status and perspectives. *Plant Cell Rep.* 31(3):479-94. Epub 2011/11/15. doi: 10.1007/s00299-011-1186-8. PubMed PMID: 22080228.
7. Specht E, Miyake-Stoner S, Mayfield S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol Lett.* 32(10):1373-83. Epub 2010/06/18. doi: 10.1007/s10529-010-0326-5. PubMed PMID: 20556634; PubMed Central PMCID: PMC2941057.
8. Jones CS, Mayfield SP. Algae biofuels: versatility for the future of bioenergy. *Curr Opin Biotechnol.* 23(3):346-51. Epub 2011/11/23. doi: S0958-1669(11)00709-9 [pii] 10.1016/j.copbio.2011.10.013. PubMed PMID: 22104720.
9. Stephens E, Ross IL, King Z, Mussgnug JH, Kruse O, Posten C, et al. An economic and technical evaluation of microalgal biofuels. *Nat Biotechnol.* 28(2):126-8. Epub 2010/02/09. doi: nbt0210-126 [pii] 10.1038/nbt0210-126. PubMed PMID: 20139944.
10. Georgianna DR, Mayfield SP. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature.* 488(7411):329-35. Epub 2012/08/17. doi: nature11479 [pii] 10.1038/nature11479. PubMed PMID: 22895338.
11. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science.* 2007;318(5848):245-50. Epub 2007/10/13. doi: 318/5848/245 [pii]

10.1126/science.1143609. PubMed PMID: 17932292; PubMed Central PMCID: PMC2875087.

12. Tran M, Van C, Barrera DJ, Pettersson PL, Peinado CD, Bui J, et al. Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc Natl Acad Sci U S A*. 110(1):E15-22. Epub 2012/12/14. doi: 1214638110 [pii] 10.1073/pnas.1214638110. PubMed PMID: 23236148; PubMed Central PMCID: PMC3538218.

13. Gregory JA, Li F, Tomosada LM, Cox CJ, Topol AB, Vinetz JM, et al. Algae-produced Pfs25 elicits antibodies that inhibit malaria transmission. *PLoS One*. 7(5):e37179. Epub 2012/05/23. doi: 10.1371/journal.pone.0037179 PONE-D-12-00773 [pii]. PubMed PMID: 22615931; PubMed Central PMCID: PMC3353897.

14. Gimpel JA, Hyun JS, Schoepp NG, Mayfield SP. Production of recombinant proteins in microalgae at pilot greenhouse scale. *Biotechnol Bioeng*. 112(2):339-45. Epub 2014/08/15. doi: 10.1002/bit.25357. PubMed PMID: 25116083.

15. Neupert J, Karcher D, Bock R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J*. 2009;57(6):1140-50. Epub 2008/11/28. doi: TPJ3746 [pii] 10.1111/j.1365-313X.2008.03746.x. PubMed PMID: 19036032.

16. Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP. Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One*. 7(8):e43349. Epub 2012/09/01. doi: 10.1371/journal.pone.0043349 PONE-D-12-11628 [pii]. PubMed PMID: 22937037; PubMed Central PMCID: PMC3427385.

17. Barahimipour R, Strenkert D, Neupert J, Schroda M, Merchant SS, Bock R. Dissecting the contributions of GC content and codon usage to gene expression in the model alga *Chlamydomonas reinhardtii*. *Plant J*. 2015;84(4):704-17. doi: 10.1111/tpj.13033. PubMed PMID: 26402748; PubMed Central PMCID: PMCPMC4715772.

18. Strenkert D, Schmollinger S, Schroda M. Heat shock factor 1 counteracts epigenetic silencing of nuclear transgenes in *Chlamydomonas reinhardtii*. *Nucleic Acids Res*. 2013;41(10):5273-89. doi: 10.1093/nar/gkt224. PubMed PMID: 23585280; PubMed Central PMCID: PMCPMC3664811.

19. Lingg N, Zhang P, Song Z, Bardor M. The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J*. 7(12):1462-72. Epub 2012/07/26. doi: 10.1002/biot.201200078. PubMed PMID: 22829536.

20. Corchero JL, Gasser B, Resina D, Smith W, Parrilli E, Vazquez F, et al. Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnol Adv*. 31(2):140-53. Epub 2012/09/19. doi: S0734-9750(12)00151-6 [pii] 10.1016/j.biotechadv.2012.09.001. PubMed PMID: 22985698.

21. Rasala BA, Chao SS, Pier M, Barrera DJ, Mayfield SP. Enhanced genetic tools for engineering multigene traits into green algae. *PLoS One*. 9(4):e94028. Epub 2014/04/09. doi: 10.1371/journal.pone.0094028

643 PONE-D-13-44402 [pii]. PubMed PMID: 24710110; PubMed Central PMCID:
644 PMC3978050.

645 22. Riano-Pachon DM, Correa LG, Trejos-Espinosa R, Mueller-Roeber B. Green
646 transcription factors: a chlamydomonas overview. *Genetics*. 2008;179(1):31-9. Epub
647 2008/05/22. doi: 179/1/31 [pii]
648 10.1534/genetics.107.086090. PubMed PMID: 18493038; PubMed Central PMCID:
649 PMC2390610.

650 23. Yoshioka S, Taniguchi F, Miura K, Inoue T, Yamano T, Fukuzawa H. The novel
651 Myb transcription factor LCR1 regulates the CO₂-responsive gene *Cah1*, encoding a
652 periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii*. *Plant Cell*.
653 2004;16(6):1466-77. Epub 2004/05/25. doi: 10.1105/tpc.021162
654 tpc.021162 [pii]. PubMed PMID: 15155888; PubMed Central PMCID: PMC490039.

655 24. Ibanez-Salazar A, Rosales-Mendoza S, Rocha-Urbe A, Ramirez-Alonso JI, Lara-
656 Hernandez I, Hernandez-Torres A, et al. Over-expression of Dof-type transcription
657 factor increases lipid production in *Chlamydomonas reinhardtii*. *J Biotechnol*. 184:27-38.
658 Epub 2014/05/23. doi: S0168-1656(14)00231-4 [pii]
659 10.1016/j.jbiotec.2014.05.003. PubMed PMID: 24844864.

660 25. Tsai CH, Warakanont J, Takeuchi T, Sears BB, Moellering ER, Benning C. The
661 protein Compromised Hydrolysis of Triacylglycerols 7 (CHT7) acts as a repressor of
662 cellular quiescence in *Chlamydomonas*. *Proc Natl Acad Sci U S A*. 111(44):15833-8.
663 Epub 2014/10/15. doi: 1414567111 [pii]
664 10.1073/pnas.1414567111. PubMed PMID: 25313078; PubMed Central PMCID:
665 PMC4226073.

666 26. Camargo A, Llamas A, Schnell RA, Higuera JJ, Gonzalez-Ballester D, Lefebvre
667 PA, et al. Nitrate signaling by the regulatory gene NIT2 in *Chlamydomonas*. *Plant Cell*.
668 2007;19(11):3491-503. doi: 10.1105/tpc.106.045922. PubMed PMID: 18024571;
669 PubMed Central PMCID: PMCPMC2174885.

670 27. Kropat J, Tottey S, Birkenbihl RP, Depege N, Huijser P, Merchant S. A regulator
671 of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that
672 recognizes the GTAC core of copper response element. *Proc Natl Acad Sci U S A*.
673 2005;102(51):18730-5. doi: 10.1073/pnas.0507693102. PubMed PMID: 16352720;
674 PubMed Central PMCID: PMCPMC1311908.

675 28. Riano-Pachon DM, Ruzicic S, Dreyer I, Mueller-Roeber B. PlnTFDB: an
676 integrative plant transcription factor database. *BMC Bioinformatics*. 2007;8:42. Epub
677 2007/02/09. doi: 1471-2105-8-42 [pii]
678 10.1186/1471-2105-8-42. PubMed PMID: 17286856; PubMed Central PMCID:
679 PMC1802092.

680 29. Gorman DS, Levine RP. Cytochrome f and plastocyanin: their sequence in the
681 photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A*. 1965;54(6):1665-9. Epub 1965/12/01. PubMed PMID: 4379719; PubMed
682 Central PMCID: PMC300531.

683 30. Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B,
684 Mueller-Roeber B. PlnTFDB: updated content and new features of the plant
685 transcription factor database. *Nucleic Acids Res*. 38(Database issue):D822-7. Epub
686 2009/10/28. doi: gkp805 [pii]

688 10.1093/nar/gkp805. PubMed PMID: 19858103; PubMed Central PMCID:
689 PMC2808933.

690 31. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in
691 PCR amplification. *Nat Protoc.* 2008;3(9):1452-6. Epub 2008/09/06. doi: nprot.2008.133
692 [pii]
693 10.1038/nprot.2008.133. PubMed PMID: 18772872.

694 32. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for
695 supporting accessible, reproducible, and transparent computational research in the life
696 sciences. *Genome Biol.* 11(8):R86. Epub 2010/08/27. doi: gb-2010-11-8-r86 [pii]
697 10.1186/gb-2010-11-8-r86. PubMed PMID: 20738864; PubMed Central PMCID:
698 PMC2945788.

699 33. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al.
700 Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol.*
701 Chapter 19:Unit 19 0 1-21. Epub 2010/01/14. doi: 10.1002/0471142727.mb1910s89.
702 PubMed PMID: 20069535; PubMed Central PMCID: PMC4264107.

703 34. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy:
704 a platform for interactive large-scale genome analysis. *Genome Res.* 2005;15(10):1451-
705 5. Epub 2005/09/20. doi: gr.4086505 [pii]
706 10.1101/gr.4086505. PubMed PMID: 16169926; PubMed Central PMCID:
707 PMC1240089.

708 35. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-
709 time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001;25(4):402-8.
710 Epub 2002/02/16. doi: 10.1006/meth.2001.1262
711 S1046-2023(01)91262-9 [pii]. PubMed PMID: 11846609.

712 36. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME
713 SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server
714 issue):W202-8. Epub 2009/05/22. doi: gkp335 [pii]
715 10.1093/nar/gkp335. PubMed PMID: 19458158; PubMed Central PMCID:
716 PMC2703892.

717 37. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to
718 discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2:28-36. Epub
719 1994/01/01. PubMed PMID: 7584402.

720 38. McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an
721 evaluation on ChIP data. *BMC Bioinformatics.* 11:165. Epub 2010/04/02. doi: 1471-
722 2105-11-165 [pii]
723 10.1186/1471-2105-11-165. PubMed PMID: 20356413; PubMed Central PMCID:
724 PMC2868005.

725 39. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2-
726 -a multiple sequence alignment editor and analysis workbench. *Bioinformatics.*
727 2009;25(9):1189-91. Epub 2009/01/20. doi: btp033 [pii]
728 10.1093/bioinformatics/btp033. PubMed PMID: 19151095; PubMed Central PMCID:
729 PMC2672624.

730 40. Reece-Hoyes JS, Marian Walhout AJ. Yeast one-hybrid assays: a historical and
731 technical perspective. *Methods.* 57(4):441-7. Epub 2012/08/14. doi: S1046-
732 2023(12)00183-1 [pii]

10.1016/j.ymeth.2012.07.027. PubMed PMID: 22884952; PubMed Central PMCID: PMC3448804.

41. Gaudinier A, Zhang L, Reece-Hoyes JS, Taylor-Teeple M, Pu L, Liu Z, et al. Enhanced Y1H assays for Arabidopsis. *Nat Methods*. 8(12):1053-5. Epub 2011/11/01. doi: nmeth.1750 [pii] 10.1038/nmeth.1750. PubMed PMID: 22037706; PubMed Central PMCID: PMC3821074.

42. Wilson TE, Fahrner TJ, Johnston M, Milbrandt J. Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science*. 1991;252(5010):1296-300. Epub 1991/05/31. PubMed PMID: 1925541.

43. Verhaegent M, Christopoulos TK. Recombinant Gaussia luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. *Anal Chem*. 2002;74(17):4378-85. Epub 2002/09/19. PubMed PMID: 12236345.

44. Yamano T, Tsujikawa T, Hatano K, Ozawa S, Takahashi Y, Fukuzawa H. Light and low-CO₂-dependent LCIB-LCIC complex localization in the chloroplast supports the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant Cell Physiol*. 51(9):1453-68. Epub 2010/07/28. doi: pcq105 [pii] 10.1093/pcp/pcq105. PubMed PMID: 20660228.

45. Turkina MV, Blanco-Rivero A, Vainonen JP, Vener AV, Villarejo A. CO₂ limitation induces specific redox-dependent protein phosphorylation in *Chlamydomonas reinhardtii*. *Proteomics*. 2006;6(9):2693-704. Epub 2006/03/31. doi: 10.1002/pmic.200500461. PubMed PMID: 16572472.

46. Mariscal V, Moulin P, Orsel M, Miller AJ, Fernandez E, Galvan A. Differential regulation of the *Chlamydomonas* Nar1 gene family by carbon and nitrogen. *Protist*. 2006;157(4):421-33. Epub 2006/08/15. doi: S1434-4610(06)00058-7 [pii] 10.1016/j.protis.2006.06.003. PubMed PMID: 16905358.

47. Hahn D, Kaltenbach C, Kuck U. The Calvin cycle enzyme sedoheptulose-1,7-bisphosphatase is encoded by a light-regulated gene in *Chlamydomonas reinhardtii*. *Plant Mol Biol*. 1998;36(6):929-34. Epub 1998/03/31. PubMed PMID: 9520283.

48. Stauber EJ, Fink A, Markert C, Kruse O, Johanningmeier U, Hippler M. Proteomics of *Chlamydomonas reinhardtii* light-harvesting proteins. *Eukaryot Cell*. 2003;2(5):978-94. Epub 2003/10/14. PubMed PMID: 14555480; PubMed Central PMCID: PMC219354.

49. Fang W, Si Y, Douglass S, Casero D, Merchant SS, Pellegrini M, et al. Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO₂-concentrating mechanism regulator CIA5/CCM1. *Plant Cell*. 24(5):1876-93. Epub 2012/05/29. doi: tpc.112.097949 [pii] 10.1105/tpc.112.097949. PubMed PMID: 22634760; PubMed Central PMCID: PMC3442575.

50. Pireyre M, Burow M. Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol Plant*. 8(3):378-88. Epub 2015/02/11. doi: S1674-2052(14)00048-3 [pii] 10.1016/j.molp.2014.11.022. PubMed PMID: 25667003.

51. Robinson KA, Lopes JM. SURVEY AND SUMMARY: *Saccharomyces cerevisiae* basic helix-loop-helix proteins regulate diverse biological processes. *Nucleic Acids Res*.

779 2000;28(7):1499-505. Epub 2000/03/10. doi: gkd 276 [pii]. PubMed PMID: 10710415;
 780 PubMed Central PMCID: PMC102793.
 781 52. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative
 782 analysis of MYB and bHLH plant transcription factors. *Plant J.* 66(1):94-116. Epub
 783 2011/03/30. doi: 10.1111/j.1365-313X.2010.04459.x. PubMed PMID: 21443626.
 784 53. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix-loop-
 785 helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol.* 2004;36(2):189-
 786 204. Epub 2003/12/04. doi: S1357272503002115 [pii]. PubMed PMID: 14643885.
 787 54. Lang EJ, Cross PJ, Mittelstadt G, Jameson GB, Parker EJ. Allosteric ACTion: the
 788 varied ACT domains regulating enzymes of amino-acid metabolism. *Curr Opin Struct*
 789 *Biol.* 29:102-11. Epub 2014/12/30. doi: S0959-440X(14)00145-6 [pii]
 790 10.1016/j.sbi.2014.10.007. PubMed PMID: 25543886.
 791 55. Zhao H, Li X, Ma L. Basic helix-loop-helix transcription factors and epidermal cell
 792 fate determination in *Arabidopsis*. *Plant Signal Behav.* 7(12):1556-60. Epub 2012/10/18.
 793 doi: 22404 [pii]
 794 10.4161/psb.22404. PubMed PMID: 23073001; PubMed Central PMCID: PMC3578892.
 795 56. Castilhos G, Lazzarotto F, Spagnolo-Fonini L, Bodanese-Zanettini MH, Margis-
 796 Pinheiro M. Possible roles of basic helix-loop-helix transcription factors in adaptation to
 797 drought. *Plant Sci.* 223:1-7. Epub 2014/04/29. doi: S0168-9452(14)00041-7 [pii]
 798 10.1016/j.plantsci.2014.02.010. PubMed PMID: 24767109.
 799

800

801

802

807

808

809