Chapter 4 Functional genomics of *Chlamydomonas reinhardtii*

Ian K. Blaby, Maria J. Soto, and Crysten E. Blaby-Haas

CHAPTER CONTENTS

- **I. Introduction**
- **II. The *Chlamydomonas* genome: structural genomics**
- **III. An introduction to functional annotation**
- **IV. State of functional annotations in *Chlamydomonas***
- **V. Functional genomics of *Chlamydomonas*: Overview**
- **VI. Transcriptomics**
- **VII. Proteomics**
- **VIII. Phylogenomics**
- **IX. Genome-wide phenotype screens**
- **X. Use of mutant collections for reverse-genetics**
- **XI. Databases and resources for *Chlamydomonas* genomics and functional genomics**
- **XII. Outlook and future directions**

ABSTRACT

Access to the *Chlamydomonas reinhardtii* genome provides the means to pursue fundamental genomics-based research of relevance to algae, land plants and animal lineages. Of the 17741 gene models, the function of most is only of limited knowledge, thus undermining our ability to truly understand the biology of this alga, and plants at large. Transcriptomes, proteomes and phenomes can each provide clues by presenting a global snapshot image of the behavior of the cell under specific conditions. Herein we describe the history of functional genomics, how it is being applied to Chlamydomonas and discuss the resources available for the experimental validation of function predictions.

I. INTRODUCTION

The *Chlamydomonas* genome project was initiated during a number of genomic breakthroughs in the late 90s and early 2000s (Grossman et al., 2003), with the first bacterial, eukaryote, plant, mammalian and human draft genomes all published within less than a decade (Arabidopsis Genome, 2000; Fleischmann et al., 1995; Goffeau et al., 1996; Lander et al., 2001; Mouse Genome Sequencing et al., 2002). As an understanding of the genetic components underpinning biological processes became easier to attain, these watershed undertakings and achievements marked a new period of biological research. Access to the *Chlamydomonas* genome enabled the community to identify gene families shared with plant and animal lineages while leveraging sequence-based analyses to contextualize gene/protein-specific knowledge gained from other reference organisms and apply that information to *Chlamydomonas* genes/proteins. A new appreciation for biological complexity dawned since many coding sequences lacked similarity to proteins of known function. Although two decades have passed since the *Chlamydomonas* genome project was initiated, close to half of predicted *Chlamydomonas* proteins still fall into this “gene of unknown function” category. At the same time, the accuracy of many *predicted* functions remains unknown, while available functional annotations can be vague. Fortunately, as new genome-wide technologies have accelerated gene-specific data acquisition and decreasing costs have democratized acquisition, an increasing body of information is available to link genes and proteins with function and generate hypotheses for further testing. This chapter aims to describe the present state of functional knowledge associated with the *Chlamydomonas* genome, the role of post-genomics data types (transcriptomics, proteomics, phylogenomics, and genome-wide mutant screens) in providing functional

information, and the resources that have been developed to increase our understanding of the functional potential encoded in the *Chlamydomonas* genome.

II. The *CHLAMYDOMONAS* GENOME: STRUCTURAL GENOMICS

Following two early releases, the publication of the third version of the *Chlamydomonas* genome constituted a landmark achievement in algal research: a genomic touchstone became available to link multiple fields of biology from ciliary biogenesis to photosynthesis (Merchant et al., 2007). Constituting 120Mb spread across 1557 scaffolds with an average coverage of 30x Sanger-based sequences, the draft assembly provided the basis for post-genomics experiments and the broad application of bioinformatics and reverse genetics. Over the following decade a series of updates to both the draft genome sequence assembly and gene models was enabled with significant sequencing input (of both genomic DNA and transcripts), culminating in the fifth genome (v5) release in 2012. This release consisted of 17 chromosomes plus an additional 37 unanchored scaffolds totaling 111Mb of DNA.

The raw genome sequence gives little insight into biology, and post-assembly analyses are needed to identify genomic elements, such as genes and protein-coding regions. Two releases of gene models were generated on v5 of the genome, namely, 5.3.1 and 5.5, which amplified previous releases by incorporating additional transcriptomic (i.e., RNA-Seq) data (I. K. Blaby et al., 2014). A revised draft genome sequence, v6, and accompanying gene models are actively being worked on with a view to release in 2022. For a comprehensive discussion of *Chlamydomonas* structural genomics (i.e., broadly, the raw genome sequence, assembly, and gene model prediction), the reader is referred to Chapter V of this volume.

III. AN INTRODUCTION TO FUNCTIONAL ANNOTATION

Once structural annotations are available, the next step is functional annotation. Functional annotation is the curation of genes/proteins with biological information. However, there are no strict guidelines on how to functionally annotate a genome, and different research groups and different genome databases have their own approaches. The Gene Ontology Consortium (http://geneontology.org/docs/go-consortium/) curates gene/protein function with controlled vocabulary and categories based on 1) “molecular function,” a description of gene product’s activities that occur at the molecular level; 2) “biological process,” a description of the process to which the gene product (often in concert with other

gene products) contributes; and 3) “cellular component,” a description of the gene product’s location that can be a cellular compartment or macromolecular complex (Ashburner et al., 2000). This particular vocabulary-based structure aids in identifying enriched terms present in a genome-wide study, such as transcriptomics or proteomics.

In most other genome databases, functional annotations are presented as short deflines that may describe a characterized function, present a predicted function, or provide some type of functional information, such as family membership or the phenotype observed with the gene is disrupted. At one end of the spectrum, a functional annotation may be limited to the observation that the gene is upregulated under a certain condition, implicating the role of the encoded protein in response to that condition. At the other end, a fully characterized enzyme would be associated with knowledge of the reaction performed, in what pathway and in which cellular compartment the enzyme participates, the regulation of the gene and protein, and (ultimately) a mechanistic understanding as to how the reaction is catalyzed. Often this extent of functional annotation is not provided by a typical genome database and requires sourcing the primary literature.

IV. STATE OF FUNCTIONAL ANNOTATIONS IN *CHLAMYDOMONAS*

Based on community-driven curation efforts, 9% of *Chlamydomonas* genes are associated with a publication. In many cases, these publications present expert-derived bioinformatic analyses that place *Chlamydomonas* proteins within conserved families. Since the manual curation and experimental validation of each of the remaining 91% of genes/proteins is presently an impractical task, functional annotations, as with most other genomes, are largely based on genome-wide searches of sequence similarity to proteins in available databases, usually in an automated fashion via BLAST or sequence models that enable placing a protein within a conserved family (Altschul, Gish, Miller, Myers, & Lipman, 1990; Altschul et al., 1997).

Estimating a confidence level for such predicted functional annotations is difficult, and this type of functional annotation can lead to mis- and over-annotation (Danchin, Ouzounis, Tokuyasu, & Zucker, 2018; Promponas, Iliopoulos, & Ouzounis, 2015; Schnoes, Brown, Dodevski, & Babbitt, 2009). Specific metrics are not available, and there is no standard definition of what constitutes protein function (Rhee & Mutwil, 2014). Equally, the veracity of using sequence similarity alone to functionally annotate proteins is hotly debated (Radivojac et al., 2013) for a number of reasons: i) the source of the original, experimentally derived, annotation is only rarely preserved, thus obfuscating the evolutionary distance

between the experimentally characterized protein(s) and the protein being annotated; ii) functional annotations are often not transferred by similarity from the experimentally characterized progenitor of the annotation but from a sequence whose annotation is also computationally derived; iii) the E-value statistic used by BLAST is dependent on a number of parameters, including database size and protein length; hence using an arbitrary E-value cutoff as a proxy for functional conservation is a misappropriation of this statistic; iv) relatively minor amino acid changes can alter aspects of function, such as substrate-binding or localization (Rost, 2002; Zallot, Harrison, Kolaczkowski, & Crécy-Lagard, 2016).

To mitigate some of these challenges, Phytozome 13 furnishes automatically generated predictions derived from a handful of databases and preserves the name of those databases, the database identifiers, and provides a schematic of the detected similarity between each protein and the database-derived sequence models (Goodstein et al., 2012). Annotations tools such as MapMan (Thimm et al., 2004) and Mercator (Lohse et al., 2014) rely on manually curated classifications originating from green algae and land plants, which limits annotation error resulting from annotating *Chlamydomonas* proteins from very distantly related organism.

Since only a small proportion of *Chlamydomonas* genes/proteins are experimentally characterized, where is functional information available for annotating *Chlamydomonas* genes/proteins using bioinformatics? This is a difficult question to answer. Currently, only limited capabilities exist to automate data extraction from the primary literature via natural language processing techniques. As a result, reliable database entries have to be manually curated and updated with new discoveries and datasets as they are published (Ching et al., 2018; Pestian et al., 2007; Zhao, Su, Lu, & Wang, 2020). The best curated functional annotation database in AmiGO houses automated and curated GO terms and associated provenances (Carbon et al., 2009). Mining this database provides a snapshot of the state of functional annotations across the major phylogenetic lineages (Fig. 1). As part of the GO Consortium's Reference Genome Project, a comprehensive set of manually curated GO terms are available for the human genome and eleven reference organisms: *Arabidopsis thaliana, Caenorhabditis elegans, Danio rerio, Dictyostelium discoideum, Drosophila melanogaster, Escherichia coli, Gallus gallus, Mus musculus, Rattus norvegicus, Saccharomyces cerevisiae,* and *Schizosaccharomyces pombe (Consortium, 2009)*. This curation effort is evident in the distribution of captured experimentally supported functional annotations (Fig. 1A). For *Chlamydomonas*, a small number (~60) of gene products in v5.5 can be mapped to GO terms curated with experimental evidence from *Chlamydomonas* (Fig. 1B), while for ~30% of *Chlamydomonas* proteins, the most similar *Arabidopsis* protein is

associated with a GO term curated with experimental data (Fig. 1B). The fact remains that functional annotations for ~70% of *Chlamydomonas* genes are either absent or are computational predictions based on information outside the green lineage, which may or may not be appropriately, or correctly, ascribed.

Due to manual curation by the *Chlamydomonas* community at the time of the v3 genome release (Merchant et al., 2007), and the ongoing efforts of many groups in the field, 6554 loci (37%) are associated with biological information. In Phytozome v13 https://phytozome-next.jgi.doe.gov, 6092 loci are now associated with either a manually curated defline (a brief manual annotation usually comprising a few words) and/or a description (a more extensive description consisting of one or several sentences). A total of 5524 loci (31%) are associated with a manually curated gene symbol (i.e., gene name) and 3556 have a curated defline. In addition to these manual curations, Phytozome v13 provides automated curation for 9242, 4175, and 5685 loci based on PFAM, KOG and MapMan ontology, respectively[1]. These numbers are not additive, as many loci classified with one of these annotations are also labelled with another (Fig. 2). Also contained on Phytozome v13 are the predicted cellular localizations of proteins as determined by Predalgo https://mybiosoftware.com/predalgo-1-0-protein-subcellular-localization-prediction-green-algae.html, a tool trained specifically on curated green algal protein localizations (Tardif, 2012). To highlight the biology yet to be discovered in *Chlamydomonas*, ~ 6000 loci are devoid of a curated annotation, ontology or conserved PFAM/KOG domain (Fig. 2).

These statistics underscore the need for genome-scale investigations to provide locus-specific functional inferences at scale. While no single experiment can achieve this, a central concept to functional genomics is to integrate multiple sources of data. As experiments capturing cellular behavior under different conditions accumulate, hypotheses can be formed based on condition-specific expression or phenotypes (Ge, Walhout, & Vidal, 2003; Joyce & Palsson, 2006; Vidal, 2001), although computational and logistical complications have been noted (Palsson & Zengler, 2010). The following sections discuss the nature and extent of functional genomics experiments performed in *Chlamydomonas*, with a view to directing the reader to the accessibility and availability of each resource.

V. FUNCTIONAL GENOMICS OF *CHLAMYDOMONAS*: OVERVIEW

[1] These annotations, and bulk text files for download, are accessible via Phytozome v12, and were assessed in Nov 2020 (Goodstein et al., 2012). N.B. There was a superscript 1 on p. 3 that didn't refer to any footnote so I deleted it.

While defining and characterizing protein function constitutes much of biological research, functional genomics experiments set out to inform some aspect of function on a genome-wide scale. Functional genomics data sets, such as transcriptomics, proteomics, and genome-wide mutant screens, can provide layers of gene-specific data. On the one hand, each experiment provides a global snapshot of cellular behavior under different conditions. On the other hand, by comparing the results of multiple experiments, biological information can be derived by identifying the specific conditions under which a protein is expressed (e.g., when and in what situation the cell requires that transcript or protein) or needed for growth (e.g., which conditions under which the loss of a gene is detrimental), by determining co-expressed proteins (e.g., proteins possibly involved in the same process), and by localizing proteins to specific subcompartments within the cell (e.g., where a protein functions). Within this body of functional genomics resources, we also include phylogenomics techniques that provide a means to contextualize functional knowledge acquired with homologous, non-*Chlamydomonas*, genes and proteins. As with functional genomics experiments, techniques are available for informing on function, such as comparative genomic analyses that reveal whether a gene family is restricted to photosynthetic organisms and, therefore, may function in a photosynthesis-related process. Cumulatively, *Chlamydomonas* has been subjected to hundreds of functional genomics experiments, with thousands of samples. Since a functional genomics investigation provides a (usually quantified) datapoint for each gene in a relatively short time, each type of functional genomics experiment provides a snapshot of some functional data for many, if not all, genes with relatively little time or cost. Clearly such experiments do not individually offer the extensive level of characterization needed for a detailed functional annotation of a given locus, but they can provide the necessary information to generate a conjecture, allowing for further hypothesis-driven experiments to be performed.

VI. TRANSCRIPTOMICS

A transcriptome, whether captured by cDNA sequencing (i.e., RNA-Seq) or by microarray, provides a measure of transcript abundance under defined conditions. The resulting expression data can be used in a number of ways to inform function. Since gene expression is often regulated in response to the environment, a comparative analysis of two or more growth conditions can be used to infer the involvement of genes in that condition. For example, greater transcript abundance of a given gene collected from cells limited for nitrogen compared to cells replete for nitrogen suggests the expression of that locus, and by extension, the encoded protein, is necessary for cell maintenance when restricted for nitrogen. In addition to assessing when a transcript is needed, co-expression analyses can

reveal sets of co-expressed genes, which are more likely to function in the same pathway or process than genes that behave differently (Niehrs & Pollet, 1999). Thus, a functional inference may be made for a gene of unknown function whose transcript abundance alters to similar degrees as another gene of known function. Yet the investigator should be aware of expression-based assumptions: a change in transcript abundance may be post-transcriptional, and changes in transcript abundance do not always correlate with protein abundance. As an example, transcript abundance of some metal-dependent proteins increases during metal deficiency, but protein abundance actually decreases (Blaby-Haas & Merchant, 2011); the increase in transcript abundance is proposed to serve as a feedback loop responding to lower protein activity in the absence of the necessary metal cofactor. In other cases, the altered transcript level may not result directly from the environmental cue but may be a consequence of a downstream or indirect process. These types of changes are seen most often for expression that is responding to a stress caused by the growth condition, such as photooxidative damage during iron limitation (Glaesener, Merchant, & Blaby-Haas, 2013).

The first large-scale interrogations of gene expression in *Chlamydomonas* were performed by microarray. Using an array containing probes representing around 2700 gene sequences derived from Expressed Sequence Tags (ESTs), a comparative analysis was performed to identify genes whose expression profiles differed in response to light intensity and CO_2 levels (Grossman et al., 2003; Im, Zhang, Shrager, Chang, & Grossman, 2003). Other early experiments sought to investigate the influence of other environmental factors, such as sulfur deprivation and reactive oxygen species (ROS) exposure, as well as nutrient deprivation and toxicity (Jamers et al., 2006; Ledford et al., 2004; A. V. Nguyen et al., 2008; Z. Zhang et al., 2004). As access to genomic resources expanded (Asamizu et al., 2004; Eberhard et al., 2006), so did the size of microarrays, as was used to capture a transcriptome-level evaluation of ciliary regeneration (Stolc, Samanta, Tongprasit, & Marshall, 2005).

While the arrays continued to advance coordinately with improved genome releases, and their data was integrated more broadly with other ‘omic scale datasets (Mettler et al., 2014), the technological breakthrough of parallelized second-generation sequencing, and, in particular, the ability to quantitatively sequence cDNA, resulted in RNA-Seq largely replacing microarrays. In addition to being able to detect a higher dynamic range of transcript abundance, reduced cost and cDNA requirements, and availability of standardized analysis tools, RNA-Seq is not limited to a defined set of probes and can capture transcript isoforms as well as novel transcripts. As a consequence, RNA-Seq is routinely exploited to advance

gene model accuracy (Van Verk, Hickman, Pieterse, & Van Wees, 2013; Z. Wang, Gerstein, & Snyder, 2009). Indeed, v5 of the *Chlamydomonas* gene models have benefited from this input (I. K. Blaby et al., 2014). A further advantage is that, since the RNA-Seq read data can be saved digitally as raw sequence, reads can be realigned to the genome and transcripts re-quantified as updated genome drafts and gene models are released (this also allows for data from different investigators to be re-processed using identical computational pipelines, assuming they are made available, allowing for direct comparison and integration of datasets). These advantages have resulted in RNA-Seq becoming a routine technique in many *Chlamydomonas* laboratories. A decade since the first RNA-Seq-based transcriptomes in this alga, there are now hundreds of accessible datasets investigating all areas of *Chlamydomonas* research (examples in Table 1).

The large number of available datasets for a wide breadth of physiological processes is due in part to development of *Chlamydomonas* as a facile experimental system. Carefully controlled and defined manipulations can be made to the growth environment (e.g., precise alterations to temperature, light quality and intensity) and culture media (through chemical supplement or micro/macronutrient dropout). Comparisons of mutants and parental strains (e.g., for regulon analysis with a strain disrupted for a transcription factor) are not complicated by ploidy. While not exhaustive, Table 1 attempts to capture the breadth of availability data by cataloguing *Chlamydomonas* RNA-Seq publications, with particular attention to those datasets that are publicly archived in either the Gene Expression Omnibus (GEO) or the Short Read Archive (SRA) databases, thus enabling access to sequenced reads for future analyses. Some RNA-Seq data can be visualized and accessed in bulk via the *Chlamydomonas* pages of Phytozome v13.

Mining this data can be performed in multiple ways. While each individual study focuses on those genes differentially expressed based on a specific experimental design, deeper investigation can be afforded by collating multiple datasets from different publications or reanalyzing data with different computational techniques. One approach is to identify emergent properties from the hierarchical complexity of the data. For example, a manifold-learning method showed that *FDX7* may have different roles in the day vs. night by treating light- and dark-period transcriptomes of a diurnal transcriptome as two distinct datasets and simultaneously clustering the differentially expressed genes (N. D. Nguyen, Blaby, & Wang, 2019). This approach does not assume a phenotype arises from a single gene. By assessing the behavior of genes under different conditions, cohorts with similar expression profiles can be identified.

As noted above, gene co-expression analysis can be a powerful means to inferring gene function, since genes of similar expression profiles may be functionally linked, such as involved in the same pathway or process. Two databases have been developed allowing mining of gene co-expression networks (Aoki, Okamura, Ohta, Kinoshita, & Obayashi, 2016; Romero-Campero, Perez-Hurtado, Lucas-Reina, Romero, & Valverde, 2016), and more recently, a comprehensive study of 58 transcriptomes has illuminated the extent of gene co-expression in *Chlamydomonas*, and provides significant opportunity for further gene-function mining (Salomé & Merchant, 2020).

VII. Proteomics

As valuable as transcriptomes are, their value is limited to inferring transcriptionally and post-transcriptionally regulated mRNA abundance. As the predominant molecular machines in the cell, proteins are prone to additional levels of post-translational regulation. Thus, transcriptome-based expression estimates do not constitute the perfect proxy for protein concentration or enzyme activity, illustrating the benefit of proteomic datasets. The logic for linking proteins with function is identical for proteomes as it is for transcriptomes: proteins enriched in response to an experimental perturbance vs. control cells may perform a role in response to that perturbation. Unlike transcriptomes, proteomic analysis can be performed on biochemically fractionated cells and purified compartments, providing protein localization data. Multiple cellular compartments have been isolated from *Chlamydomonas* and subjected to proteomics to determine the protein composition of cilia, the chloroplast, lipid droplets, nucleus and mitochondria (Table 2). As with transcriptomics, the quality and depth of proteomics experiments in *Chlamydomonas* have increased with revisions to the genome and gene models and new technological developments (Rolland et al., 2009).

Nevertheless, while proteomic data captures a truer estimate of protein levels in the cell, this benefit is offset by a lower dynamic range and an inability to detect all expressed proteins. Consequently, many studies investigating different phenomena in *Chlamydomonas* perform both transcriptomic and proteomic analyses on the same samples (Table 2). Indeed, multiple *Chlamydomonas* studies have additionally captured metabolomics data, providing a systems-level view of the cell that attempts to quantify transcript, protein and metabolite abundances[2]. Integrating multi-disciplinary datasets is a powerful approach for understanding cellular behavior. Not only does the likelihood of a gene exhibiting differential expression increase, but the biological significance of a gene's involvement in a specific

[2] While a valuable tool for gene-function analyses, metabolomics data cannot necessarily be correlated to a specific single genetic locus, and consequently are not discussed in this chapter, but are considered elsewhere in *The Sourcebook* (Volume 2, Chapters ***).

process or response increases if consistent behaviors are observed at the transcript and protein levels. Several *Chlamydomonas* studies have combined proteomes and transcriptomes, often with samples being taken from the same cultures (Tables 1 & 2). Even so, and especially when comparative analyses are performed across datasets resulting from different studies or laboratories, carefully controlled conditions, and the recording of experimental metadata, are crucial.

VII. Phylogenomics

Phylogenomics is a term coined by Eisen and colleagues (Eisen, 1998; Eisen, Kaiser, & Myers, 1997) in the late 1990s and refers to a strategy for improving function predictions based on protein family relationships rather than just sequence similarity. This approach encompasses phylogenetics (i.e., reconstruction of family member relatedness and common ancestry) as a means to weigh the significance of functional information derived from different family members. The underlying assumption is that orthologs (i.e., genes/proteins in different species that evolved through speciation) are more likely to share the same function, while paralogs (i.e., genes/proteins that have diverged from one other due to a duplication event) may have diverged in function. In addition to phylogenetics, phylogenomics encompasses a suite of comparative-genomic techniques that use associations between genes of unknown/uncertain function and genes with known function to provide insight into the function of the former. Sometimes referred to as “guilt-by-association” (Aravind, 2000), a common analysis is the generation of phylogenetic profiles, which are used to identify functional linkages between sets of genes and a phenotype or metabolic capability (Pellegrini, Marcotte, Thompson, Eisenberg, & Yeates, 1999). As an example, the availability of the *Chlamydomonas* genome enabled the identification of genes conserved in sequenced green algae and land plants but absent in non-photosynthetic organisms.

Two major phylogenomics projects undertaken in *Chlamydomonas* have been seminal to our understanding the organism’s photosynthetic and ciliated characteristics. The GreenCut is a phylogenetic inventory, originally comprising 349 genes and since expanded to nearly 600, that are only found in the genomes of organisms capable of photosynthesis (Karpowicz, Prochnik, Grossman, & Merchant, 2011; Merchant et al., 2007). The conservation of these genes in all photosynthetic organisms across nearly a billion years of evolution signifies the paramount importance of their encoded proteins’ functions in the maintenance and efficient operation of photosynthesis. Surprisingly, however, given their apparent importance, 46% of these proteins have only vague functional annotations and

32% are annotated simply as conserved or predicted proteins; hence they are the focus of intense study (Heinnickel & Grossman, 2013; Karpowicz et al., 2011; Merchant et al., 2007; Wittkopp, Saroussi, Yang, & Grossman, 2016). Similarly, the CiliaCut is an assemblage of 195 genes conserved only in the genomes of ciliated organisms; this phylogenetic profile is suggestive of involvement of the proteins encoded by these genes in the biosynthesis, structure or regulation of cilia (Merchant et al., 2007). Both the GreenCut and CiliaCut are discussed at length in chapters X and Y of this publication.

IX. Genome-wide phenotype screens

As with the above approaches, the identification of a gene-specific phenotype can provide an additional layer of biological information for a gene. Since *Chlamydomonas* can be grown in microplates, easily performed screens can involve one mutant and hundreds of growth conditions or thousands of mutants and one or more growth conditions. As underscored by the catalogued collections of *E. coli*, *S. cerevisiae* and *A. thaliana* mutants (Alonso et al., 2003; T. Baba et al., 2006; Winzeler et al., 1999), mapped mutant libraries constitute invaluable resources, especially once they reach genome saturation. When such a library is generated, the entire assemblage (or targeted subgroups) can be subjected to a condition-based selection, identifying genes whose gene products play a role in acclimating to the given selection either because they are essential under the prescribed selection or their loss leads to a fitness advantage. Screens can be conducted under different conditions, such as, for *Chlamydomonas*, heterotrophic vs. photoautotrophic growth regimes, enabling the recorded growth/no growth phenotypes to be attributed to carbon metabolism generally, and photosynthesis specifically. Performing such large-scale experiments with sufficiently large collections of mutants allows for the systematic identification of genotype-to-phenotype relations. Sequenced mutant libraries also allow researchers to cherry-pick specific mutant(s) of interest to their investigations.

To be representative of the genome, such a resource is considered to be approaching genome-saturation once ~85% of loci are disrupted; by definition, genes encoding proteins essential to the cell's viability under the conditions in which the mutants were generated cannot be isolated (Carpenter & Sabatini, 2004). Given the relative ease with which insertional mutants can be produced in *Chlamydomonas* (Chapter **), further facilitated by the haploid status of its genome, this has been a lucrative approach to illuminating key enzymes in a number of pathways and processes in *Chlamydomonas*.

More recently, deep-sequencing technologies have provided opportunities beyond a digital grow/no grow. Second-generation sequencing can be used to quantify the abundance of individual mutants in a population of mutations, obtaining more granular and quantitative strain fitness data of each strain exposed to a given condition (Price et al., 2018). Although not yet available for *Chlamydomonas*, CRISPR-based mutagenesis has enabled genome-scale gene editing in other microbes, overcoming limitations associated with relying on near-random integration of a selection cassette (and the difficulty of identifying disrupted loci) -- the most popular technique in *Chlamydomonas* for generating mutant libraries (as discussed below). Recent progress in applying CRISPR to *Chlamydomonas* (Chapter **) will presumably lead to the development of such a resource in the near future.

Chlamydomonas has a rich history of exploiting screens to obtain mutant strains with particular phenotypes (i.e., forward genetic screens). Since *Chlamydomonas* can grow heterotrophically when provisioned with fixed carbon (i.e., acetate), mutants in genes essential for photosynthesis can be isolated (Levine, 1960b). Taking advantage of this characteristic, early experiments exposed cells to UV and selected acetate-requiring mutants, leading to the discovery of genes encoding components of the photosynthetic electron transport chain and photosystem II (Fork & Urbach, 1965; Levine, 1960a). Ethyl methanesulfonate (EMS) and *N*-methyl-*N′*-nitro-*N*-nitrosoguanidine (MNNG) have also been employed in concert with appropriate screens to identify the genetic basis of arginine biosynthesis (and also generating the Arg- auxotrophic strains) (Loppes, 1968), cell wall biosynthesis (Davies & Plaskitt, 1971) and ciliary assembly (Davies & Plaskitt, 1971; Huang, Rifkin, & Luck, 1977).

The use of mutagens as a tool to generate mutants was later joined by the development of protocols enabling efficient DNA delivery into *Chlamydomonas* and integration of that DNA into the genome by illegitimate recombination in a near-random fashion. The transforming DNA can be as minimal as a drug resistant or auxotrophic marker, of which many have been described in *Chlamydomonas* (Scaife et al., 2015). Since the inserted DNA is of known sequence, selected mutants can be mapped (Cheng et al., 2017; Rachel M Dent, Haglund, Chin, Kobayashi, & Niyogi, 2005; R. M. Dent et al., 2015; Li et al., 2016). Furthermore, if one assumes random integration, and that the genome is 111Mb with an average gene size of ~5kb, a library of 100,000 mutants might be expected to contain several integrations per gene (although integrations in non-coding regions, such as introns and UTRs, may not impair function of the encoded protein). While dependent upon the ability to select for an observable or measurable phenotype, screens based on this approach have identified genes in many pathways. For example, a collection of approaching 50,000

mutants generated by the random insertion of either zeocin or paromycin resistant markers (conferred by *ble* and *aphVIII* respectively) was subjected a series of screens designed to identify genes related to photosynthesis and related metabolism (R. M. Dent et al., 2015). Mutants were isolated displaying a phenotype as a consequence of one of twelve conditional screens, including acetate-dependency (thus null for photosynthesis), deviant growth rates vs. when subjected to high- or low-light, and sensitivity to reactive oxygen species. Using PCR-based methods (David Gonzalez-Ballester et al., 2011; Liu, Mitsukawa, Oosumi, & Whittier, 1995) to enrich for and identify the flanking regions of the lesion, these studies resulted in the identification of 439 mapped mutants displaying known phenotypes (R. M. Dent et al., 2015). A similar study mapped by PCR-based methods (D. Gonzalez-Ballester, de Montaigu, Galvan, & Fernandez, 2005) yielded 26 mutants in 20 genes defective for motility and a further 10 mutants incapable of intraflagellar transport complex assembly (Cheng et al., 2017). Other screens have identified key proteins involved in starch biosynthesis, flagellar biosynthesis, the carbon concentrating mechanism and the circadian system). Another collection of mutants with disrupted cell division, an essential process, overcame the limitation of disrupting essential genes by screening specifically for temperature sensitive mutants (Breker, Lieberman, & Cross, 2018). By screening at both restrictive and permissive temperatures, 350 mutants were isolated, and 260 were determined to have resulted from one or occasionally two mutagenized loci. This and other collections of mapped mutants from such genome-wide screens are available to researchers *en masse* (Table 3), allowing access to these mutants for reverse-genetics based investigations.

X. Use of mutant collections for reverse-genetics

Despite their utility, most libraries built for forward genetic screens are ultimately disposable resources: a handful of mutants with desired phenotypes are maintained, some are successfully mapped to a specific locus, but the remainder (likely tens of thousands of mutant strains) are discarded. Beginning with the publication of a method enabling moderate throughput identification of the disrupted locus (David Gonzalez-Ballester et al., 2011), high-throughput reverse genetics took off in *Chlamydomonas*. Specifically, reverse genetics conceptually differs from forward-genetic screens, since in the latter, investigators hope to identify mutants with a specific phenotype, while in the former researchers seek a phenotype associated with a given gene disruption. The relative ease with which investigators could screen this collection by PCR-based methods enabled a series of genetic mutant-based gene function analyses to be performed (Beel et al., 2012; N. R. Boyle et al., 2012; Catalanotti et al., 2012; Duanmu et al., 2013; Magneschi et al., 2012; Meuser et al., 2012; Yang et al., 2015). Building upon these achievements, several large-scale

Chlamydomonas mutant collections have been generated by insertional mutagenesis, the affected loci mapped, and the collection preserved, allowing for interrogation of the affected region.

A significantly larger collection of known site-of-insertion *Chlamydomonas* mutants has been generated by the Jonikas laboratory. The group has mapped the sites of >80,000 integrants; thus, assuming random integration, the collection is approaching the theoretical point of targeting all loci. The first method to identify the sites of integration exploited a type IIS restriction enzyme, *Mme*I, which cleaves DNA 20 nucleotides downstream of the recognition site (R. Zhang et al., 2014). Since the recognition site, TCCRAC, is incorporated at the extreme of the end of the introduced cassette, 20 nt of the flanking genomic DNA can be captured. Following enrichment, this signature 20bp can be determined by second-generation sequencing methods to identify the genomic insertion site, resulting in roughly 12,000 mapped mutants. As with other protocols to map insertions described above, this technique does not fully capture the genotype of the resultant strains, since only the region immediately flanking the insertion is determined. Notably, the site of insertion can be associated with unmapped deletions or insertions (Rachel M Dent et al., 2005; R. M. Dent et al., 2015; Tam & Lefebvre, 1993; R. Zhang et al., 2014). Multiple integration events, disrupting more than one locus, are also possible. Linking genotype to phenotype requires the implicit assumption that the genotype is known. Therefore, because of the potential for unmapped mutations in these mutants, additional due diligence is required.

In some organisms, rescue of the phenotype can be performed by expressing a wild-type copy of the target gene *in trans*. Since *Chlamydomonas* has a plasmid system for this purpose, complementation can be attempted by introducing a non-replicating expression cassette containing the wild-type gene, which integrates into the genome by illegitimate recombination. This method results in non-uniform expression levels across transformants, sometimes requiring the screening of 100 independent transformants to acquire one or two lines with suitable levels of expression (Ian K Blaby & Blaby-Haas, 2018). Furthermore, the high GC-skew of the genome presents challenges to cloning *Chlamydomonas* genes, although this has been offset to some degree by advances in obtaining synthetic DNA.

To help mitigate these effects, the *Mme*I-based method was subsequently built upon. A paired-end sequencing approach, combined with sequencing the region immediately flanking the insertion site, provides sequence information for a region up to 1.5kb distal to the insertion (Li et al., 2016). If all paired-end reads match the expected genome sequence, a higher level of confidence can be placed on the presence of a single insertion site for the cassette. The added sequence coverage also aided with mapping integrants whereas the 20bp signature DNA did not invariably yield unique sequence (Li et al., 2016). While this

approach does not rule out the possibility of rearrangements, deletions and insertions distal the identified integration site, ~83,000 mutants have been determined using these techniques, and are available as the Chlamydomonas Library Project (CLiP). These mutants plus an additional 439 mutants identified by Dent et al, (R. M. Dent et al., 2015) are available where they are maintained to the community via the Chlamydomonas Resource Center, constituting a reverse-genetics resource (See Vol 1 Chapter ** for a detailed presentation of these libraries). These mutant collections have been analyzed using large numbers of parameters, resulting in the creation of comprehensive datasets for each mutant strain's phenome (Vilarrasa-Blasi et al., 2020). At this depth of experimentation, the same logic can be practiced as with gene co-expression: when multiple mutants carrying deletions for different genes exhibit similar phenotypes when exposed to the same conditions, they may be functionally linked.

Many of these mapped strains are of potential interest to multiple researchers. Even when only those mutants mapping to exons are considered, 14,650 mutants are accessible, affecting 5078 of the 17741 loci in v5.5 (~28%), or conversely, 12,663 genes do not contain an insertion within an exon. If one assumes the sequencing approach devised to map the mutants captured all insertions, including those landing within non-coding regions, 4,876 loci have not been affected at all, suggesting that these genes constitute the essentiome (i.e., genes whose proteins are essential to viability under the conditions in which the library was generated lies).

Of particular relevance are those mutants lying within specific inventories of genes noted earlier, such as the GreenCut, CiliaCut or genes highlighted by transcriptomes to be differentially expressed under certain conditions. For example, of the 596 GreenCut genes mapped to v5.5, 226 exon-specific insertions (or 451 total) are available in the Chlamydomonas Library Project (CLiP) and 108 are accessible in the Niyogi collection (R. M. Dent et al., 2015). Mutants are available for 70 of the 161 CiliaCut genes that can be dependably mapped to v5.5. A further 134 mapped single mutants and 45 double mutants which have been instrumental in studying photosynthesis are available through the ChlamyStation database (See tools section below). Nevertheless, only a minute subset of all available mutants have been sequenced at the whole-genome level, and the methods for mapping the CLiP collection have been found to be inaccurate for some mutants. Due to the possibility of offsite, non-detected, genetic aberrations occurring beyond the sequenced region, caution should be exercised by the researcher before assuming an observed phenotype is directly due to proposed lesion, and newly received strains should be confirmed prior to subjecting them to extensive analysis.

XI. Databases and resources for *Chlamydomonas* genomics and functional genomics

Central to the ability to interpret large, complex, datasets are the computational tools to both warehouse and analyze/visualize those data. Fortunately, multiple web-based resources are available to aid with the curation and dissemination of genomics data for *Chlamydomonas*. Table 3 describes tools that continue to be maintained, as well as detailing websites enabling distribution of strains and plasmids. It should be noted that several of these tools were implemented during previous releases of the *Chlamydomonas* genome, and consequently cater predominantly to earlier genome versions. Therefore, using some tools, and indeed datasets as described in Tables 1 and 2, may necessitate converting gene model generations. This can be automated via the algal annotation tool or on a local machine using the correspondence tables accessible from Phytozome (Table 3). Current information on the genome is presented in the following chapter (***), and library resources are described in chapter ***.

XII. Outlook and future directions

As presented in this chapter, functional genomics datasets provide a means to collect biological information for genes and proteins on a genome-wide scale. Because each dataset typically focuses on a specific biological question and a limited number of growth conditions, mining multiple datasets is needed to build a more complete picture of gene/protein expression and condition-specific phenotypes. When contextualized with biological information acquired through homology searches and phylogenetics, these resources can be used to generate, support, or refute a functional annotation. At the same time, leveraging these data can lead to a hypothesis with respect to gene/protein function, which can be tested at the bench.

Resources are available for acquiring or generating *Chlamydomonas* gene-disruption mutants and/or mutants with the gene of interest overexpressed. Presently, reverse genetics (i.e., the ability to delete or disrupt loci in a targeted manner without affecting any non-targeted region) in *Chlamydomonas* is not as routine as for other microbial reference systems like *S. cerevisiae* (which lacks chloroplasts and cilia), but several techniques are available.

Several studies have employed gene silencing (RNAi) as an approach to knock down the expression of targeted genes (Cerutti, Johnson, Gillham, & Boynton, 1997; E.-J. Kim &

Cerutti, 2009). This has the unique benefit of allowing the phenotypic analysis of essential genes; however, the extent of down-regulation tends to diminish over time, making these mutants short-term resources. Down-regulation is likely often due to the fact that *Chlamydomonas* has an effective mechanism to recognize and silence transgene expression – e.g. expression of the artificial microRNA used for directed knock-down (Neupert et al., 2020). Still, multiple mutants have been made in this way, provisioning researchers with the strains required to assess gene function (Kumar et al., 2017; Oey et al., 2013; S. Schmollinger, Strenkert, & Schroda, 2010).

Zinc-finger nucleases (ZNFs), transcription activator-like effector nucleases (TALENS) and clustered regularly interspaced short palindromic repeat (CRISPR) together with its associated nuclease, Cas, represent progressive generations of genetic tools that can be exploited to induce double-strand breaks in DNA in a targeted manner (Christian et al., 2010; Jinek et al., 2012; Y.-G. Kim, Cha, & Chandrasegaran, 1996). The means of gene targeting differs between each of these tools. Both ZFNs and TALENs are dependent on DNA-binding domains to recruit the nuclease component to the intended locus, and consequently their use necessitates assembly of the DNA constructs, as well as in vivo transcription and translation of the protein machinery, to ensure proper recruitment of the nuclease. Both this complex assembly as well as a propensity for off-target lesions has limited the utilization of either ZFNs or TALENs in *Chlamydomonas*, despite both being demonstrated to viable approaches, with efficiencies of 1% being reported (Gao et al., 2014; Greiner et al., 2017; I. Sizova, Greiner, Awasthi, Kateriya, & Hegemann, 2012, 2013).

That CRISPR can be programmed with a single guide RNA, circumventing much of the complexity associated with either ZFNs or TALENs, has made this system an indispensable tool for high throughput gene function interrogation in many organisms. Presently, the technology is in its relative infancy in Chlamydomonas. An early investigation suggested toxicity of high levels of Cas expression (Jiang, Brueggeman, Horken, Plucinak, & Weeks, 2014), and, consistent with this, most success has been observed by pre-assembling the RNA protein complex in vitro prior to transformation (Baek et al., 2016; Ferenczi, Pyott, Xipnitou, & Molnar, 2017; Greiner et al., 2017; S. E. Shin et al., 2016). Using this approach, many groups have described success in Chlamydomonas genome editing (Baek et al., 2016; Findinier, Delevoye, & Cohen, 2019; Greiner et al., 2017; Guzmán-Zapata et al., 2019; Jiang et al., 2014; J. Kim, Lee, Baek, & Jin, 2020; S.-E. Shin et al., 2016; Shin et al., 2019), indicating this technique holds promise as a routine reverse genetics approach.

Table 1 Examples of Chlamydomonas RNASeq studies

Experimental variable	Genome version aligned to	Chlamydomonas strain(s)	Data availability[1]	Reference
bilin signalling	4	D66, *hmox1, hmox2, hmox1hmox2*	GSE40031	(Duanmu et al., 2013)
cell cycle	5	not indicated	GSE43004	(Albee et al., 2013)
cell development, strain comparison	5	R3, CJU10	Supplemental Table 2 [2]	(Lopez et al., 2015)
ClpP protease depletion	4	A31, DCH16	GSE56295	(Ramundo et al., 2014)
CO_2 limitation, strain comparison	4	CC125, *cia5*	GSE33927	(Fang et al., 2012)
copper deprivation, strain comparison	3	CC1021, *crr1-2*, *crr1*:*CRR1*	GSE25124	(Castruita et al., 2011)
diurnal cycle	5	CC5152	GSE71469	(Zones, Blaby, Merchant, & Umen, 2015)
diurnal cycle	5	CC5390	GSE112394	(Strenkert et al., 2019)
diurnal cycle	4	dw15.1	PRJNA264777	(Panchy et al., 2014)
haploid to diploid transition	5	JL28	GEO91400	(Joo et al., 2017)
hypoxia	4	CC124	GSE42035	(Hemschemeier et al., 2013)
iron starvation	4	CC4532	GSE44611	(Urzica et al., 2012)
light dark transition	4	CC1609	GSE62690	(Fu et al., 2015)
longterm culturing under constant light	4	4A +	Supplemental Table 1[3]	(Lv et al., 2013)
mercury toxicity	5	CPCC11	GSE70066	(Beauvais-Flück, Slaveykova, & Cosio, 2016)
nickel addition	5	CC3960, CC5073, PST35, CC5071	Supplemental Data [2]	(Blaby-Haas, Castruita, Fitz-Gibbon, Kropat, & Merchant, 2016)
nitrogen deprivation	4	dw15	GSE24367	(Miller et al., 2010)

nitrogen deprivation, additional acetate added	4	CC4348, CC4349, CC4565, CC4566, CC4567	GSE55253	(I. K. Blaby et al., 2013)
nitrogen deprivation, strain comparison	5	CC4348, CC4349	GSE51642	(Goodenough et al., 2014)
nitrogen deprivation, strain comparison	4	CC4532, CC4348, CC4349	SRX038871, GSE51602	(Stefan Schmollinger et al., 2014)
oxidative stress, hydrogen peroxide treatment	5	CC-4532	GSE34826	(I. K. Blaby et al., 2015)
oxidative stress, rose bengal treatment, strain comparison	5	CC4348, *gpx5*	Supplemental Table 1[3]	(Simon et al., 2013)
oxidative stress, rose bengal treatment, strain comparison	4	4A+, sak1	KF985242	(Wakao et al., 2014)
oxidative stress, singlet oxygen acclimation	3	4A+	GSE33548	(Fischer et al., 2012)
salt stress		GY-D55	PRJNA490089	(N. Wang et al., 2018)
solid and liquid media in light and dark	5	137C	SRP132684	(Bogaert et al., 2018)
sulfur nutrition, strain comparison	3	D66, *snrk2.1*	GSE17970	(González-Ballester et al., 2010)
zinc deficiency	4	CC-4532	GSE25622, GSE41096	(Malasarn et al., 2013)
zinc deficiency and resupply	5	CC4532	GSE58786	(Hong-Hermesdorf et al., 2014)

[1] Unless otherwise indicated, accession numbers are provided for Gene Expression Omnibus (GEO) (Edgar, Domrachev, & Lash, 2002) or Sequence Read Archive (SRA) (Leinonen, Sugawara, Shumway, & Collaboration, 2010)
[2] Partially available in supplemental data
[3] Supplemental data includes complete transcriptome aligned to the indicated genome version

Table 2 Examples of Chlamydomonas proteomic studies

Study summary	Strain	Additional 'omics[1]	Proteins identified	Genome version	Data availability[2]	Reference
anaerobic response	CC424		2315	3	SM	(Terashima, Specht, Naumann, & Hippler, 2010)
Comparative analysis of WT and salt tolerant strains	CC503, salt tolerant strain		683 (detected in both strains)	5	SM	(Sithtisarn et al., 2017)
copper, iron, manganese, zinc deprivation	CC4532			4	SM	(Hsieh et al., 2013)
diurnal	CC5390	T, M	6403	5	PXD010794[3]	(Strenkert et al., 2019)
glutathionylation	D66		225	4	SM	(Zaffagnini et al., 2012)
heat stress	CF185		3433	3	SM	(Mühlhaus, Weiss, Hemme, Sommer, & Schroda, 2011)
high CO_2	CC400			4	SM	(M. Baba, Suzuki, & Shiraiwa, 2011)
high light	CC125		444	3	SM	(Förster, Mathesius, & Pogson, 2006)

iron deprivation	CC424		203	2	SM	(Naumann et al., 2007)
light intensity	CC-1690	T (microarray), M	644	4	SM	(Mettler et al., 2014)
nitrogen deprivation	CC4532	T, M		5	SM	(Stefan Schmollinger et al., 2014)
nitrogen deprivation	dw15		259	3	SM	(Moellering & Benning, 2010)
nitrogen deprivation	not indicated		248	4	16644[3]	(H. M. a. B. M. a. C. Nguyen, 2011)
nitrogen deprivation	CC400		2853	5	PXD019491[3]	(Smythers, McConnell, Lewis, Mubarek, & Hicks, 2020)
routine culturing	CC503	M	1069	3	SM	(May et al., 2008)
phosphoproteome	SAG 73.72		328	2	SM	(Wagner et al., 2006)
secretome of mating gametes	CC124, CC125		1216	5	SM	(Luxmi et al., 2018)
thiolation target proteins	CW15		25		MAT	(Michelet et al., 2008)
thioredoxin target proteins	CW15		55	1	MAT	(Lemaire et al., 2004)
TOR kinase inhibition effect on phosphoproteome	CC1690		1432	5	PXD007221[3]	(Werth et al., 2019)
compartment/machinery: thylakoid	CW15		2622	2	SM	(Allmer, Naumann, Markert, Zhang, & Hippler, 2006)
compartment/machinery: stroma	XS1		274	4	SM	(Bienvenut et al., 2011)
compartment/machinery: chloroplast	n/a		2315		SM	(Terashima, Specht, & Hippler, 2011)

compartment/machinery: eyespot	CW15		202	2	SM	(Schmidt et al., 2006)
compartment/machinery: centriole	CW15		61	2	SM	(Keller, Romijn, Zamora, Yates III, & Marshall, 2005)
compartment/machinery: flagellar	137c, *oda1-1*		360	2	SM	(Pazour, Agrin, Leszyk, & Witman, 2005)
compartment/machinery: transition zone of flagellar	*pf18* mt−		115	4	SM	(Diener, Lupetti, & Rosenbaum, 2015)
compartment/machinery: mitochondria	83.82		496	2	SM	(Atteia et al., 2009)
compartment/machinery: nucleus	CC503		672	4	13764[3]	(Winck, Riaño-Pachón, Sommer, Rupprecht, & Mueller-Roeber, 2012)
compartment/machinery: 70S ribosome	CC3395		30	ESTs & plastid genome (Maul et al., 2002)	MAT	(Kenichi Yamaguchi et al., 2003)
compartment/machinery: small subununit of chloroplast ribosome	CC3395		21	ESTs & plastid genome (Maul et al., 2002)	MAT	(K. Yamaguchi et al., 2002)
compartment/machinery: pyrenoid	CC1690	T	368	5	SM	(Mackinder et al., 2016)

[1] T=transcriptome, M=metabolite
[2] SM=Supplemental Material; MAT=Main article table
[3] Accession number for Proteomics Identification Database or proteomeXchange (Martens et al., 2005; Vizcaíno et al., 2014)

Table 3 Summary of data repositories and online tools enabling functional genomics in Chlamydomonas

	Tool	website	genome version	summary	reference
Genome databases	Phytozome 13	https://phytozome-next.jgi.doe.gov	5	Repository of Chlamydomonas genome/gene models allowing bulk-download of data. Gene annotations include user validated annotations, and PFAM and GO predicted annotations, as well as Predalgo predicted protein localizations. Gene co-expression tool utilizing some transcriptomes. Also allows comparative analysis with >200 plants/photosynthetic organisms	(Goodstein et al., 2012)

			https://bioinformatics.psb.ugent.be/plaza/versions/pico-plaza		(Vandepoele et al., 2013)
	Co-expression	picoPlaza		4	Comparative genomics of Chlamydomonas and other photosynthetic organisms
		ChlamyNET	http://viridiplantae.ibvf.csic.es/ChlamyNet	5	Chlamydomonas gene coexpression networks and putative transcription factor binding site predictions (Romero-Campero et al., 2016)
		ALCOdb	http://alcodb.jp	5	Gene coexpression networks for Chlamydomonas and *Cyanidioschyzon merloae* (Aoki et al., 2016)
	Annotation	Algal annotation tool	http://pathways.mcdb.ucla.edu/algal/index.html	3-5	Bulk annotation prediction via Kegg, MapMan, GO, Panther, Metacyc and interconversion of gene model IDs (Lopez, Casero, Cokus, Merchant, & Pellegrini, 2011)
		Predalgo	https://giavapgenomes.ibpc.fr/cgibin/predalgodb.perl?page=main		Protein localization predictions. (Tardif, 2012)

Plant transcription factor database	http://planttfdb.gao-lab.org/	5	Predictions for 230 Chlamydomonas transcription factors, and for >160 additional photosynthetic organisms	(Tian, Yang, Meng, Jin, & Gao, 2020)
PlnTFDB	http://plntfdb.bio.uni-potsdam.de/v3.0/index.php?sp_id=CRE4	4	Predictions for 348 Chlamydomonas transcription factors, and for 19 other photosynthetic organisms	(Jin, Zhang, Kong, Gao, & Luo, 2014)
ChlamyCyc 9.0	https://plantcyc.org/content/chlamycyc-9.0	5	Metabolic pathways mapped to genome	(Schläpfer et al., 2017)
Chlamydomonas Resource Center	https://www.chlamycollection.org/		Distributes thousands of catalogued Chlamydomonas strains, in addition to mutant collections (Breker et al., 2018; R. M. Dent et al., 2015; Li et al., 2016). Also maintains plasmids and cDNA libraries	
ChlamyStation	http://chlamystation.free.fr/		IBPC Collection of Chlamydmonas photosynthesis mutants	

Screen data	ChlamyChem	http://chlamychem.utoronto.ca/ChlamyChem/method.php	small molecule screen data repository	(Alfred et al., 2012)

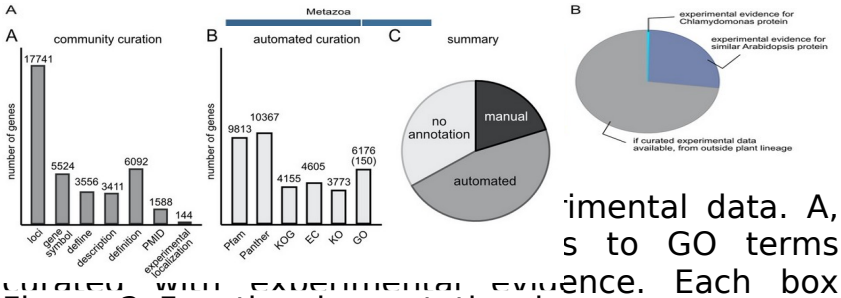

Figure 2 Functional annotation in Chlamydomonas. A, bar chart representing number of loci with associated information as given. Numbers on top of each bar refer to number of loci with given information. B, bar chart representing number of loci whose gene products contain similarity to the given databases. C, pie chart summarizing information in panel A and B. Of the 17,741 loci in v5.5, 20% have been manually curation with a functional annotation. 47% have an automated annotation from one or more of the databases from panel B, and 33% are not associated with an annotation. These 33% are typically either uncharacterized proteins specific to Chlamydomonas, specific to green algae, or belong to a protein family not recognized by one of the databases in panel B.

Acknowledgments

This work has been supported by the DOE Joint Genome Institute (http://jgi.doe.gov) by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy, as well by as well by BR-KP160104 to IKB (PI Francis Alexander, Brookhaven National Laboratory) to IKB. CB-H is supported by the US Department of Energy, Office of Science, Office of Biological, and Environmental Research, as part of the Quantitative Plant Science Initiative SFA at Brookhaven National Laboratory.

References

Albee, A. J., Kwan, A. L., Lin, H., Granas, D., Stormo, G. D., & Dutcher, S. K. (2013). Identification of cilia genes that affect cell-cycle progression using whole-genome transcriptome analysis in Chlamydomonas reinhardtti. *G3: Genes, Genomes, Genetics, 3*(6), 979-991.

Alfred, S. E., Surendra, A., Le, C., Lin, K., Mok, A., Wallace, I. M., et al. (2012). A phenotypic screening platform to identify small molecule modulators of Chlamydomonas reinhardtiigrowth, motility and photosynthesis. *Genome biology, 13*(11), R105.

Allmer, J., Naumann, B., Markert, C., Zhang, M., & Hippler, M. (2006). Mass spectrometric genomic data mining: Novel insights into bioenergetic pathways in Chlamydomonas reinhardtii. *Proteomics, 6*(23), 6207-6220.

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science, 301*(5633), 653-657.
Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410.
Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.
Aoki, Y., Okamura, Y., Ohta, H., Kinoshita, K., & Obayashi, T. (2016). ALCOdb: gene coexpression database for microalgae. *Plant and Cell Physiology, 57*(1), e3-e3.
Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature, 408*(6814), 796-815.
Aravind, L. (2000). Guilt by association: contextual information in genome analysis. *Genome Research, 10*(8), 1074-1077.
Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S., Hirono, M., et al. (2004). Establishment of publicly available cDNA material and information resource of Chlamydomonas reinhardtii (Chlorophyta) to facilitate gene function analysis. *Phycologia, 43*(6), 722-726.
Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nature genetics, 25*(1), 25-29.
Atteia, A., Adrait, A., Brugière, S., Tardif, M., Van Lis, R., Deusch, O., et al. (2009). A proteomic survey of Chlamydomonas reinhardtii mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. *Molecular Biology and Evolution, 26*(7), 1533-1548.
Baba, M., Suzuki, I., & Shiraiwa, Y. (2011). Proteomic analysis of high-CO2-inducible extracellular proteins in the unicellular green alga, Chlamydomonas reinhardtii. *Plant and Cell Physiology, 52*(8), 1302-1314.
Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol, 2*, 2006.0008.
Baek, K., Kim, D. H., Jeong, J., Sim, S. J., Melis, A., Kim, J. S., et al. (2016). DNA-free two-gene knockout in Chlamydomonas reinhardtii via CRISPR-Cas9 ribonucleoproteins. *Sci Rep, 6*, 30620.
Beauvais-Flück, R., Slaveykova, V. I., & Cosio, C. (2016). Transcriptomic and physiological responses of the green microalga Chlamydomonas reinhardtii during short-term exposure to subnanomolar methylmercury concentrations. *Environmental Science & Technology, 50*(13), 7126-7134.
Beel, B., Prager, K., Spexard, M., Sasso, S., Weiss, D., Müller, N., et al. (2012). A flavin binding cryptochrome photoreceptor responds to both blue

and red light in Chlamydomonas reinhardtii. *The Plant Cell, 24*(7), 2992-3008.
Bienvenut, W. V., Espagne, C., Martinez, A., Majeran, W., Valot, B., Zivy, M., et al. (2011). Dynamics of post-translational modifications and protein stability in the stroma of Chlamydomonas reinhardtii chloroplasts. *Proteomics, 11*(9), 1734-1750.
Blaby, I. K., & Blaby-Haas, C. E. (2018). Gene expression analysis by arylsulfatase assays in the green alga Chlamydomonas reinhardtii *Reporter Gene Assays* (pp. 149-161): Springer.
Blaby, I. K., Blaby-Haas, C. E., Pérez-Pérez, M. E., Schmollinger, S., Fitz-Gibbon, S., Lemaire, S. D., et al. (2015). Genome-wide analysis on *Chlamydomonas reinhardtii* reveals impact of hydrogen peroxide on protein stress responses and overlap with other stress transcriptomes. *Plant J*.
Blaby, I. K., Blaby-Haas, C. E., Tourasse, N., Hom, E. F., Lopez, D., Aksoy, M., et al. (2014). The Chlamydomonas genome project: a decade on. *Trends Plant Sci, 19*(10), 672-680.
Blaby, I. K., Glaesener, A. G., Mettler, T., Fitz-Gibbon, S. T., Gallaher, S. D., Liu, B., et al. (2013). Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *Chlamydomonas reinhardtii* starchless mutant. *Plant Cell, 25*(11), 4305-4323.
Blaby-Haas, C. E., Castruita, M., Fitz-Gibbon, S. T., Kropat, J., & Merchant, S. S. (2016). Ni induces the CRR1-dependent regulon revealing overlap and distinction between hypoxia and Cu deficiency responses in Chlamydomonas reinhardtii. *Metallomics, 8*(7), 679-691.
Blaby-Haas, C. E., & Merchant, S. S. (2011). Metal homeostasis: sparing and salvaging metals in chloroplasts. *Encyclopedia of Inorganic and Bioinorganic Chemistry*, 1-13.
Bogaert, K. A., Manoharan-Basil, S. S., Perez, E., Levine, R. D., Remacle, F., & Remacle, C. (2018). Surprisal analysis of genome-wide transcript profiling identifies differentially expressed genes and pathways associated with four growth conditions in the microalga Chlamydomonas. *PloS one, 13*(4), e0195142.
Boyle, N. R., Page, M. D., Liu, B., Blaby, I. K., Casero, D., Kropat, J., et al. (2012). Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in Chlamydomonas. *Journal of Biological Chemistry, 287*(19), 15811-15825.
Breker, M., Lieberman, K., & Cross, F. R. (2018). Comprehensive discovery of cell-cycle-essential pathways in Chlamydomonas reinhardtii. *The Plant Cell, 30*(6), 1178-1198.
Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., et al. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics, 25*(2), 288-289.
Carpenter, A. E., & Sabatini, D. M. (2004). Systematic genome-wide screens of gene function. *Nat Rev Genet, 5*(1), 11-22.

Castruita, M., Casero, D., Karpowicz, S. J., Kropat, J., Vieler, A., Hsieh, S. I., et al. (2011). Systems biology approach in Chlamydomonas reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell, 23*(4), 1273-1292.

Catalanotti, C., Dubini, A., Subramanian, V., Yang, W., Magneschi, L., Mus, F., et al. (2012). Altered fermentative metabolism in Chlamydomonas reinhardtii mutants lacking pyruvate formate lyase and both pyruvate formate lyase and alcohol dehydrogenase. *The Plant Cell, 24*(2), 692-707.

Cerutti, H., Johnson, A. M., Gillham, N. W., & Boynton, J. E. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas. *Plant Cell, 9*(6), 925-945.

Cheng, X., Liu, G., Ke, W., Zhao, L., Lv, B., Ma, X., et al. (2017). Building a multipurpose insertional mutant library for forward and reverse genetics in Chlamydomonas. *Plant Methods, 13*(1), 36.

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface, 15*(141), 20170387.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., et al. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics, 186*(2), 757-761.

Consortium, R. G. G. o. t. G. O. (2009). The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. *PLoS Comput Biol, 5*(7), e1000431.

Danchin, A., Ouzounis, C., Tokuyasu, T., & Zucker, J. D. (2018). No wisdom in the crowd: genome annotation in the era of big data–current status and future prospects. *Microbial biotechnology, 11*(4), 588-605.

Davies, D. R., & Plaskitt, A. (1971). Genetical and structural analyses of cell-wall formation in Chlamydomonas reinhardi. *Genetics Research, 17*(1), 33-43.

Dent, R. M., Haglund, C. M., Chin, B. L., Kobayashi, M. C., & Niyogi, K. K. (2005). Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of Chlamydomonas reinhardtii. *Plant physiology, 137*(2), 545-556.

Dent, R. M., Sharifi, M. N., Malnoe, A., Haglund, C., Calderon, R. H., Wakao, S., et al. (2015). Large-scale insertional mutagenesis of Chlamydomonas supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants. *Plant J, 82*(2), 337-351.

Diener, D. R., Lupetti, P., & Rosenbaum, J. L. (2015). Proteomic analysis of isolated ciliary transition zones reveals the presence of ESCRT proteins. *Current Biology, 25*(3), 379-384.

Duanmu, D., Casero, D., Dent, R. M., Gallaher, S., Yang, W., Rockwell, N. C., et al. (2013). Retrograde bilin signaling enables Chlamydomonas

greening and phototrophic survival. *Proceedings of the National Academy of Sciences, 110*(9), 3621-3626.
Eberhard, S., Jain, M., Im, C. S., Pollock, S., Shrager, J., Lin, Y., et al. (2006). Generation of an oligonucleotide array for analysis of gene expression in Chlamydomonas reinhardtii. *Current genetics, 49*(2), 106-124.
Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research, 30*(1), 207-210.
Eisen, J. A. (1998). Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome research, 8*(3), 163-167.
Eisen, J. A., Kaiser, D., & Myers, R. M. (1997). Gastrogenomic delights: a movable feast. *Nature medicine, 3*(10), 1076.
Fang, W., Si, Y., Douglass, S., Casero, D., Merchant, S. S., Pellegrini, M., et al. (2012). Transcriptome-wide changes in Chlamydomonas reinhardtii gene expression regulated by carbon dioxide and the CO2-concentrating mechanism regulator CIA5/CCM1. *The Plant Cell, 24*(5), 1876-1893.
Ferenczi, A., Pyott, D. E., Xipnitou, A., & Molnar, A. (2017). Efficient targeted DNA editing and replacement in Chlamydomonas reinhardtii using Cpf1 ribonucleoproteins and single-stranded DNA. *Proceedings of the National Academy of Sciences, 114*(51), 13567-13572.
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), e1008047.
Fischer, B. B., Ledford, H. K., Wakao, S., Huang, S. G., Casero, D., Pellegrini, M., et al. (2012). SINGLET OXYGEN RESISTANT 1 links reactive electrophile signaling to singlet oxygen acclimation in Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences, 109*(20), E1302-E1311.
Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science, 269*(5223), 496-512.
Fork, D. C., & Urbach, W. (1965). Evidence for the Localization of Plastocyanin in the Electron-Transport Chain of Photosynthesis. *Proc Natl Acad Sci U S A, 53*(6), 1307-1315.
Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., et al. (2015). N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell, 161*(4), 879-892.
Förster, B., Mathesius, U., & Pogson, B. J. (2006). Comparative proteomics of high light stress in the model alga Chlamydomonas reinhardtii. *Proteomics, 6*(15), 4309-4320.
Gao, H., Wright, D. A., Li, T., Wang, Y., Horken, K., Weeks, D. P., et al. (2014). TALE activation of endogenous genes in Chlamydomonas reinhardtii. *Algal Research, 5*, 52-60.

Ge, H., Walhout, A. J., & Vidal, M. (2003). Integrating 'omic'information: a bridge between genomics and systems biology. *TRENDS in Genetics, 19*(10), 551-560.
Glaesener, A. G., Merchant, S. S., & Blaby-Haas, C. E. (2013). Iron economy in *Chlamydomonas reinhardtii*. *Front Plant Sci, 4*, 337.
Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science, 274*(5287), 546, 563-547.
Gonzalez-Ballester, D., de Montaigu, A., Galvan, A., & Fernandez, E. (2005). Restriction enzyme site-directed amplification PCR: a tool to identify regions flanking a marker DNA. *Anal Biochem, 340*(2), 330-335.
Gonzalez-Ballester, D., Pootakham, W., Mus, F., Yang, W., Catalanotti, C., Magneschi, L., et al. (2011). Reverse genetics in Chlamydomonas: a platform for isolating insertional mutants. *Plant Methods, 7*(1), 24.
González-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S. S., & Grossman, A. R. (2010). RNA-seq analysis of sulfur-deprived Chlamydomonas cells reveals aspects of acclimation critical for cell survival. *The Plant Cell, 22*(6), 2058-2084.
Goodenough, U., Blaby, I., Casero, D., Gallaher, S. D., Goodson, C., Johnson, S., et al. (2014). The path to triacylglyceride obesity in the sta6 strain of Chlamydomonas reinhardtii. *Eukaryotic cell, 13*(5), 591-613.
Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res, 40*(Database issue), D1178-1186.
Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., Sizova, I., & Hegemann, P. (2017). Targeting of photoreceptor genes in Chlamydomonas reinhardtii via zinc-finger nucleases and CRISPR/Cas9. *The Plant Cell, 29*(10), 2498-2518.
Grossman, A. R., Harris, E. E., Hauser, C., Lefebvre, P. A., Martinez, D., Rokhsar, D., et al. (2003). Chlamydomonas reinhardtii at the crossroads of genomics. *Eukaryot Cell, 2*(6), 1137-1150.
Guzmán-Zapata, D., Sandoval-Vargas, J. M., Macedo-Osorio, K. S., Salgado-Manjarrez, E., Castrejón-Flores, J. L., Oliver-Salvador, M. d. C., et al. (2019). Efficient editing of the nuclear APT reporter gene in Chlamydomonas reinhardtii via expression of a CRISPR-Cas9 module. *International journal of molecular sciences, 20*(5), 1247.
Heinnickel, M. L., & Grossman, A. R. (2013). The GreenCut: re-evaluation of physiological role of previously studied proteins and potential novel protein functions. *Photosynthesis research, 116*(2-3), 427-436.
Hemschemeier, A., Düner, M., Casero, D., Merchant, S. S., Winkler, M., & Happe, T. (2013). Hypoxic survival requires a 2-on-2 hemoglobin in a process involving nitric oxide. *Proceedings of the National Academy of Sciences, 110*(26), 10854-10859.
Hong-Hermesdorf, A., Miethke, M., Gallaher, S. D., Kropat, J., Dodani, S. C., Chan, J., et al. (2014). Subcellular metal imaging identifies dynamic sites of Cu accumulation in Chlamydomonas. *Nat Chem Biol, 10*(12), 1034-1042.

Hsieh, S. I., Castruita, M., Malasarn, D., Urzica, E., Erde, J., Page, M. D., et al. (2013). The proteome of copper, iron, zinc, and manganese micronutrient deficiency in *Chlamydomonas reinhardtii*. *Mol Cell Proteomics, 12*(1), 65-86.

Huang, B., Rifkin, M., & Luck, D. (1977). Temperature-sensitive mutations affecting flagellar assembly and function in Chlamydomonas reinhardtii. *Journal of Cell Biology, 72*(1), 67-85.

Im, C.-S., Zhang, Z., Shrager, J., Chang, C.-W., & Grossman, A. R. (2003). Analysis of light and CO 2 regulation in Chlamydomonas reinhardtii using genome-wide approaches. *Photosynthesis Research, 75*(2), 111-125.

Jamers, A., Van der Ven, K., Moens, L., Robbens, J., Potters, G., Guisez, Y., et al. (2006). Effect of copper exposure on gene expression profiles in Chlamydomonas reinhardtii based on microarray analysis. *Aquatic Toxicology, 80*(3), 249-260.

Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M., & Weeks, D. P. (2014). Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryotic cell, 13*(11), 1465-1469.

Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic acids research, 42*(D1), D1182-D1187.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science, 337*(6096), 816-821.

Joo, S., Nishimura, Y., Cronmiller, E., Hong, R. H., Kariyawasam, T., Wang, M. H., et al. (2017). Gene regulatory networks for the haploid-to-diploid transition of Chlamydomonas reinhardtii. *Plant Physiology, 175*(1), 314-332.

Joyce, A. R., & Palsson, B. Ø. (2006). The model organism as a system: integrating'omics' data sets. *Nature reviews Molecular cell biology, 7*(3), 198-210.

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. *J Biol Chem, 286*(24), 21427-21439.

Keller, L. C., Romijn, E. P., Zamora, I., Yates III, J. R., & Marshall, W. F. (2005). Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Current Biology, 15*(12), 1090-1098.

Kim, E.-J., & Cerutti, H. (2009). Targeted gene silencing by RNA interference in Chlamydomonas *Methods in cell biology* (Vol. 93, pp. 99-110): Elsevier.

Kim, J., Lee, S., Baek, K., & Jin, E. (2020). Site-specific gene knock-out and on-site heterologous gene overexpression in Chlamydomonas reinhardtii via a CRISPR-Cas9-mediated knock-in method. *Frontiers in plant science, 11*, 306.

Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences, 93*(3), 1156-1160.
Kumar, D., Strenkert, D., Patel-King, R. S., Leonard, M. T., Merchant, S. S., Mains, R. E., et al. (2017). A bioactive peptide amidating enzyme is required for ciliogenesis. *Elife, 6*, e25728.
Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921.
Ledford, H., Baroli, I., Shin, J., Fischer, B., Eggen, R., & Niyogi, K. (2004). Comparative profiling of lipid-soluble antioxidants and transcripts reveals two phases of photo-oxidative stress in a xanthophyll-deficient mutant of Chlamydomonas reinhardtii. *Molecular Genetics and Genomics, 272*(4), 470-479.
Leinonen, R., Sugawara, H., Shumway, M., & Collaboration, I. N. S. D. (2010). The sequence read archive. *Nucleic acids research, 39*(suppl_1), D19-D21.
Lemaire, S. D., Guillon, B., Le Maréchal, P., Keryer, E., Miginiac-Maslow, M., & Decottignies, P. (2004). New thioredoxin targets in the unicellular photosynthetic eukaryote Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences, 101*(19), 7475-7480.
Levine, R. P. (1960a). A screening technique for photosynthetic mutants in unicellular algae. *Nature, 188*, 339-340.
Levine, R. P. (1960b). Genetic Control of Photosynthesis in Chlamydomonas Reinhardi. *Proc Natl Acad Sci U S A, 46*(7), 972-978.
Li, X., Zhang, R., Patena, W., Gang, S. S., Blum, S. R., Ivanova, N., et al. (2016). An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in Chlamydomonas reinhardtii. *Plant Cell, 28*(2), 367-387.
Liu, Y. G., Mitsukawa, N., Oosumi, T., & Whittier, R. F. (1995). Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J, 8*(3), 457-463.
Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., et al. (2014). M ercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, cell & environment, 37*(5), 1250-1258.
Lopez, D., Casero, D., Cokus, S. J., Merchant, S. S., & Pellegrini, M. (2011). Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC bioinformatics, 12*(1), 282.
Lopez, D., Hamaji, T., Kropat, J., De Hoff, P., Morselli, M., Rubbi, L., et al. (2015). Dynamic changes in the transcriptome and methylome of Chlamydomonas reinhardtii throughout its life cycle. *Plant Physiology, 169*(4), 2730-2743.
Loppes, R. (1968). Ethyl methanesulfonate: an effective mutagen in Chlamydomonas reinhardi. *Mol Gen Genet, 102*(3), 229-231.

Luxmi, R., Blaby-Haas, C., Kumar, D., Rauniyar, N., King, S. M., Mains, R. E., et al. (2018). Proteases shape the Chlamydomonas secretome: Comparison to classical neuropeptide processing machinery. *Proteomes, 6*(4), 36.

Lv, H., Qu, G., Qi, X., Lu, L., Tian, C., & Ma, Y. (2013). Transcriptome analysis of Chlamydomonas reinhardtii during the process of lipid accumulation. *Genomics, 101*(4), 229-237.

Mackinder, L. C., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., et al. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proceedings of the National Academy of Sciences, 113*(21), 5958-5963.

Magneschi, L., Catalanotti, C., Subramanian, V., Dubini, A., Yang, W., Mus, F., et al. (2012). A mutant in the ADH1 gene of Chlamydomonas reinhardtii elicits metabolic restructuring during anaerobiosis. *Plant physiology, 158*(3), 1293-1305.

Malasarn, D., Kropat, J., Hsieh, S. I., Finazzi, G., Casero, D., Loo, J. A., et al. (2013). Zinc Deficiency Impacts CO_2 Assimilation and Disrupts Copper Homeostasis in *Chlamydomonas reinhardtii*. *J Biol Chem, 288*(15), 10672-10683.

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., et al. (2005). PRIDE: the proteomics identifications database. *Proteomics, 5*(13), 3537-3545.

Matsuo, T., Okamoto, K., Onai, K., Niwa, Y., Shimogawara, K., & Ishiura, M. (2008). A systematic forward genetic analysis identified components of the Chlamydomonas circadian system. *Genes & Development, 22*(7), 918-930.

Maul, J. E., Lilly, J. W., Cui, L., dePamphilis, C. W., Miller, W., Harris, E. H., et al. (2002). The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. *The Plant Cell, 14*(11), 2659-2679.

May, P., Wienkoop, S., Kempa, S., Usadel, B., Christian, N., Rupprecht, J., et al. (2008). Metabolomics-and proteomics-assisted genome annotation and analysis of the draft metabolic network of Chlamydomonas reinhardtii. *Genetics, 179*(1), 157-166.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., et al. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science, 318*(5848), 245-250.

Mettler, T., Mühlhaus, T., Hemme, D., Schöttler, M.-A., Rupprecht, J., Idoine, A., et al. (2014). Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism Chlamydomonas reinhardtii. *The Plant Cell, 26*(6), 2310-2350.

Meuser, J. E., D'Adamo, S., Jinkerson, R. E., Mus, F., Yang, W., Ghirardi, M. L., et al. (2012). Genetic disruption of both Chlamydomonas reinhardtii [FeFe]-hydrogenases: insight into the role of HYDA2 in H2 production.

Biochemical and biophysical research communications, 417(2), 704-709.
Michelet, L., Zaffagnini, M., Vanacker, H., Le Maréchal, P., Marchand, C., Schroda, M., et al. (2008). In vivo targets of S-thiolation in Chlamydomonas reinhardtii. *Journal of Biological Chemistry, 283*(31), 21571-21578.
Miller, R., Wu, G., Deshpande, R. R., Vieler, A., Gärtner, K., Li, X., et al. (2010). Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant physiology, 154*(4), 1737-1752.
Mitchell, D. R., & Sale, W. S. (1999). Characterization of a Chlamydomonas insertional mutant that disrupts flagellar central pair microtubule-associated structures. *The Journal of cell biology, 144*(2), 293-304.
Moellering, E. R., & Benning, C. (2010). RNA interference silencing of a major lipid droplet protein affects lipid droplet size in Chlamydomonas reinhardtii. *Eukaryotic cell, 9*(1), 97-106.
Mouse Genome Sequencing, C., Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, 420*(6915), 520-562.
Mühlhaus, T., Weiss, J., Hemme, D., Sommer, F., & Schroda, M. (2011). Quantitative shotgun proteomics using a uniform 15N-labeled standard to monitor proteome dynamics in time course experiments reveals new insights into the heat stress response of Chlamydomonas reinhardtii. *Molecular & Cellular Proteomics, 10*(9).
Naumann, B., Busch, A., Allmer, J., Ostendorf, E., Zeller, M., Kirchhoff, H., et al. (2007). Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. *Proteomics, 7*(21), 3964-3979.
Neupert, J., Gallaher, S. D., Lu, Y., Strenkert, D., Segal, N. a., Barahimipour, R., et al. (2020). An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga Chlamydomonas. *Nature communications, 11*(1), 1-17.
Nguyen, A. V., Thomas-Hall, S. R., Malnoë, A., Timmins, M., Mussgnug, J. H., Rupprecht, J., et al. (2008). Transcriptome for photobiological hydrogen production induced by sulfur deprivation in the green alga Chlamydomonas reinhardtii. *Eukaryotic cell, 7*(11), 1965-1979.
Nguyen, H. M. a. B. M. a. C. (2011). Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: with focus on proteins involved in lipid metabolism. *Proteomics, 11*(21), 4266--4273.
Nguyen, N. D., Blaby, I. K., & Wang, D. (2019). ManiNetCluster: a novel manifold learning approach to reveal the functional links between gene networks. *BMC genomics, 20*(12), 1-14.
Niehrs, C., & Pollet, N. (1999). Synexpression groups in eukaryotes. *Nature, 402*(6761), 483-487.

Oey, M., Ross, I. L., Stephens, E., Steinbeck, J., Wolf, J., Radzun, K. A., et al. (2013). RNAi knock-down of LHCBM1, 2 and 3 increases photosynthetic H 2 production efficiency of the green alga Chlamydomonas reinhardtii. *PLoS One, 8*(4), e61375.
Palsson, B., & Zengler, K. (2010). The challenges of integrating multi-omic data sets. *Nat Chem Biol, 6*(11), 787-789.
Panchy, N., Wu, G., Newton, L., Tsai, C.-H., Chen, J., Benning, C., et al. (2014). Prevalence, evolution, and cis-regulation of diel transcription in Chlamydomonas reinhardtii. *G3: Genes, Genomes, Genetics, 4*(12), 2461-2471.
Pazour, G. J., Agrin, N., Leszyk, J., & Witman, G. B. (2005). Proteomic analysis of a eukaryotic cilium. *Journal of Cell Biology, 170*(1), 103-113.
Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings of the National Academy of Sciences, 96*(8), 4285-4288.
Pestian, J., Brew, C., Matykiewicz, P., Hovermale, D. J., Johnson, N., Cohen, K. B., et al. (2007). *A shared task involving multi-label classification of clinical free text.* Paper presented at the Biological, translational, and clinical language processing.
Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature, 557*(7706), 503-509.
Promponas, V. J., Iliopoulos, I., & Ouzounis, C. A. (2015). Annotation inconsistencies beyond sequence similarity-based function prediction–phylogeny and genome structure. *Standards in Genomic Sciences, 10*(1), 1-5.
Radivojac, P., Clark, W. T., Oron, T. R., Schnoes, A. M., Wittkop, T., Sokolov, A., et al. (2013). A large-scale evaluation of computational protein function prediction. *Nat Methods, 10*(3), 221-227.
Ramundo, S., Casero, D., Mühlhaus, T., Hemme, D., Sommer, F., Crèvecoeur, M., et al. (2014). Conditional depletion of the Chlamydomonas chloroplast ClpP protease activates nuclear genes involved in autophagy and plastid protein quality control. *The Plant Cell, 26*(5), 2201-2222.
Rhee, S. Y., & Mutwil, M. (2014). Towards revealing the functions of all genes in plants. *Trends Plant Sci, 19*(4), 212-221.
Rolland, N., Atteia, A., Decottignies, P., Garin, J., Hippler, M., Kreimer, G., et al. (2009). Chlamydomonas proteomics. *Current opinion in microbiology, 12*(3), 285-291.
Romero-Campero, F. J., Perez-Hurtado, I., Lucas-Reina, E., Romero, J. M., & Valverde, F. (2016). ChlamyNET: a Chlamydomonas gene co-expression network reveals global properties of the transcriptome and the early setup of key co-expression patterns in the green lineage. *BMC genomics, 17*(1), 227.

Rost, B. (2002). Enzyme function less conserved than anticipated. *Journal of molecular biology, 318*(2), 595-608.
Salomé, P. A., & Merchant, S. S. (2020). Co-Expression Networks in the Green Alga Chlamydomonas reinhardtii Empower Gene Discovery and Functional Exploration. *bioRxiv*.
Scaife, M. A., Nguyen, G. T., Rico, J., Lambert, D., Helliwell, K. E., & Smith, A. G. (2015). Establishing Chlamydomonas reinhardtii as an industrial biotechnology host. *The Plant Journal, 82*(3), 532-546.
Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., et al. (2017). Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant physiology, 173*(4), 2041-2059.
Schmidt, M., Geßner, G., Luff, M., Heiland, I., Wagner, V., Kaminski, M., et al. (2006). Proteomic analysis of the eyespot of Chlamydomonas reinhardtii provides novel insights into its components and tactic movements. *The Plant Cell, 18*(8), 1908-1930.
Schmollinger, S., Mühlhaus, T., Boyle, N. R., Blaby, I. K., Casero, D., Mettler, T., et al. (2014). Nitrogen-sparing mechanisms in Chlamydomonas affect the transcriptome, the proteome, and photosynthetic metabolism. *The Plant Cell, 26*(4), 1410-1435.
Schmollinger, S., Strenkert, D., & Schroda, M. (2010). An inducible artificial microRNA system for *Chlamydomonas reinhardtii* confirms a key role for heat shock factor 1 in regulating thermotolerance. *Curr Genet, 56*(4), 383-389.
Schnoes, A. M., Brown, S. D., Dodevski, I., & Babbitt, P. C. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol, 5*(12), e1000605.
Shin, S.-E., Lim, J.-M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., et al. (2016). CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. *Scientific reports, 6*(1), 1-15.
Shin, S. E., Lim, J. M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., et al. (2016). CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci Rep, 6*, 27810.
Shin, Y. S., Jeong, J., Nguyen, T. H. T., Kim, J. Y. H., Jin, E., & Sim, S. J. (2019). Targeted knockout of phospholipase A2 to increase lipid productivity in Chlamydomonas reinhardtii for biodiesel production. *Bioresource technology, 271*, 368-374.
Simon, D. F., Domingos, R. F., Hauser, C., Hutchins, C. M., Zerges, W., & Wilkinson, K. J. (2013). Transcriptome sequencing (RNA-seq) analysis of the effects of metal nanoparticle exposure on the transcriptome of Chlamydomonas reinhardtii. *Applied and environmental microbiology, 79*(16), 4774-4785.
Sithtisarn, S., Yokthongwattana, K., Mahong, B., Roytrakul, S., Paemanee, A., Phaonakrop, N., et al. (2017). Comparative proteomic analysis of Chlamydomonas reinhardtii control and a salinity-tolerant strain revealed a differential protein expression pattern. *Planta, 246*(5), 843-856.

Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., & Hegemann, P. (2012). Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. *The Plant journal : for cell and molecular biology*, 873--882.

Sizova, I., Greiner, A., Awasthi, M., Kateriya, S., & Hegemann, P. (2013). Nuclear gene targeting in C hlamydomonas using engineered zinc-finger nucleases. *The Plant Journal, 73*(5), 873-882.

Smythers, A. L., McConnell, E. W., Lewis, H. C., Mubarek, S. N., & Hicks, L. M. (2020). Photosynthetic Metabolism and Nitrogen Reshuffling Are Regulated by Reversible Cysteine Thiol Oxidation Following Nitrogen Deprivation in Chlamydomonas. *Plants, 9*(6), 784.

Stolc, V., Samanta, M. P., Tongprasit, W., & Marshall, W. F. (2005). Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. *Proceedings of the National Academy of Sciences, 102*(10), 3703-3707.

Strenkert, D., Schmollinger, S., Gallaher, S. D., Salomé, P. A., Purvine, S. O., Nicora, C. D., et al. (2019). Multiomics resolution of molecular events during a day in the life of Chlamydomonas. *Proceedings of the National Academy of Sciences, 116*(6), 2374-2383.

Tam, L. W., & Lefebvre, P. A. (1993). Cloning of flagellar genes in Chlamydomonas reinhardtii by DNA insertional mutagenesis. *Genetics, 135*(2), 375-384.

Tardif, M. a. A. A. a. S. M. a. C. G. (2012). PredAlgo: A New Subcellular Localization Prediction Tool Dedicated to Green Algae. *Molecular Biology and \ldots*.

Terashima, M., Specht, M., & Hippler, M. (2011). The chloroplast proteome: a survey from the Chlamydomonas reinhardtii perspective with a focus on distinctive features. *Current genetics, 57*(3), 151-168.

Terashima, M., Specht, M., Naumann, B., & Hippler, M. (2010). Characterizing the anaerobic response of Chlamydomonas reinhardtii by quantitative proteomics. *Molecular & Cellular Proteomics, 9*(7), 1514-1532.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal, 37*(6), 914-939.

Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., & Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic acids research, 48*(D1), D1104-D1113.

Urzica, E. I., Casero, D., Yamasaki, H., Hsieh, S. I., Adler, L. N., Karpowicz, S. J., et al. (2012). Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. *Plant Cell, 24*(10), 3921-3948.

Van Verk, M. C., Hickman, R., Pieterse, C. M., & Van Wees, S. C. (2013). RNA-Seq: revelation of the messengers. *Trends in plant science, 18*(4), 175-179.

Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., et al. (2013). pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environmental microbiology, 15*(8), 2147-2153.

Vidal, M. (2001). A biological atlas of functional maps. *Cell, 104*(3), 333-339.

Vilarrasa-Blasi, J., Fauser, F., Onishi, M., Ramundo, S., Patena, W., Millican, M., et al. (2020). Systematic characterization of gene function in a photosynthetic organism. *bioRxiv*.

Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature biotechnology, 32*(3), 223-226.

Wagner, V., Geßner, G., Heiland, I., Kaminski, M., Hawat, S., Scheffler, K., et al. (2006). Analysis of the phosphoproteome of Chlamydomonas reinhardtii provides new insights into various cellular pathways. *Eukaryotic Cell, 5*(3), 457-468.

Wakao, S., Chin, B. L., Ledford, H. K., Dent, R. M., Casero, D., Pellegrini, M., et al. (2014). Phosphoprotein SAK1 is a regulator of acclimation to singlet oxygen in Chlamydomonas reinhardtii. *Elife, 3*, e02286.

Wang, N., Qian, Z., Luo, M., Fan, S., Zhang, X., & Zhang, L. (2018). Identification of salt stress responding genes using transcriptome analysis in green alga Chlamydomonas reinhardtii. *International journal of molecular sciences, 19*(11), 3359.

Wang, Y., & Spalding, M. H. (2014). Acclimation to very low CO2: contribution of limiting CO2 inducible proteins, LCIB and LCIA, to inorganic carbon uptake in Chlamydomonas reinhardtii. *Plant Physiology, 166*(4), 2040-2050.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics, 10*(1), 57-63.

Werth, E. G., McConnell, E. W., Couso Lianez, I., Perrine, Z., Crespo, J. L., Umen, J. G., et al. (2019). Investigating the effect of target of rapamycin kinase inhibition on the Chlamydomonas reinhardtii phosphoproteome: from known homologs to new targets. *New Phytologist, 221*(1), 247-260.

Winck, F. V., Riaño-Pachón, D. M., Sommer, F., Rupprecht, J., & Mueller-Roeber, B. (2012). The nuclear proteome of the green alga Chlamydomonas reinhardtii. *Proteomics, 12*(1), 95-100.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science, 285*(5429), 901-906.

Wittkopp, T. M., Saroussi, S., Yang, W., & Grossman, A. R. (2016). The GreenCut: functions and relationships of proteins conserved in green lineage organisms. *Chloroplasts: Current Research and Future Trends*, 241-278.

Yamaguchi, K., Beligni, M. V., Prieto, S., Haynes, P. A., McDonald, W. H., Yates, J. R., et al. (2003). Proteomic characterization of the Chlamydomonas reinhardtii chloroplast ribosome identification of proteins unique to the 70 S ribosome. *Journal of Biological Chemistry, 278*(36), 33774-33785.

Yamaguchi, K., Prieto, S., Beligni, M. V., Haynes, P. A., McDonald, W. H., Yates, J. R., 3rd, et al. (2002). Proteomic characterization of the small subunit of Chlamydomonas reinhardtii chloroplast ribosome: identification of a novel S1 domain-containing protein and unusually large orthologs of bacterial S2, S3, and S5. *Plant Cell, 14*(11), 2957-2974.

Yang, W., Wittkopp, T. M., Li, X., Warakanont, J., Dubini, A., Catalanotti, C., et al. (2015). Critical role of Chlamydomonas reinhardtii ferredoxin-5 in maintaining membrane structure and dark metabolism. *Proceedings of the National Academy of Sciences, 112*(48), 14978-14983.

Zabawinski, C., Van Den Koornhuyse, N., d'Hulst, C., Schlichting, R., Giersch, C., Delrue, B., et al. (2001). Starchless mutants of Chlamydomonas reinhardtii lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase. *Journal of Bacteriology, 183*(3), 1069-1077.

Zaffagnini, M., Bedhomme, M., Groni, H., Marchand, C. H., Puppo, C., Gontero, B., et al. (2012). Glutathionylation in the photosynthetic model organism Chlamydomonas reinhardtii: a proteomic survey. *Molecular & Cellular Proteomics, 11*(2).

Zallot, R., Harrison, K. J., Kolaczkowski, B., & Crécy-Lagard, D. (2016). Functional annotations of paralogs: a blessing and a curse. *Life, 6*(3), 39.

Zhang, R., Patena, W., Armbruster, U., Gang, S. S., Blum, S. R., & Jonikas, M. C. (2014). High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *The Plant Cell, 26*(4), 1398-1409.

Zhang, Z., Shrager, J., Jain, M., Chang, C.-W., Vallon, O., & Grossman, A. R. (2004). Insights into the survival of Chlamydomonas reinhardtii during sulfur starvation based on microarray analysis of gene expression. *Eukaryotic cell, 3*(5), 1331-1348.

Zhao, S., Su, C., Lu, Z., & Wang, F. (2020). Recent advances in biomedical literature mining. *Brief Bioinform*.

Zones, J. M., Blaby, I. K., Merchant, S. S., & Umen, J. G. (2015). High-resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. *The Plant Cell, 27*(10), 2743-2769.