S41-001

Chlamydomonas cDNAs; Assembly and potential role in understanding metabolic processes

J Shrager[1], C-W Chang[1], J Davies[2], J McDermott[5], E Harris[4], C Hauser[4], R Tamse[3], R Surzycki[3], M Gurjal[3], Z Zhang[1] and A R Grossman[1]

[1]*The Carnegie Institution of Washington, 290 Panama Street, Department of Plant Biology, Stanford, CA 94305;*

[2]*Exelixis, 170 Harbor Way, P.O. Box 511, So. San Francisco, CA 94083;*

[3]*Stanford Genome Technology Center, 855 California Avenue, Palo Alto, CA 94304;*

[4]*Duke University, Biology Department, Box 91000, Durham, NC 27708; 5Iowa State University, Dept. of Botany, Bessey Hall, Ames, Iowa, 50011.*

Keywords: genomics, global regulation, differential expression

Introduction:

Chlamydomonas reinhardtii has many characteristics that make it an ideal organism for elucidating the function, biosynthesis and regulation of the photosynthetic apparatus. This unicellular, green alga has haploid genetics and can be grown heterotrophically on acetate as a sole source of carbon; dark-grown cells exhibit normal photosynthesis and chloroplast development. Many in vivo procedures applied to studies of Chlamydomonas are difficult or impossible to perform with more complex systems. In vivo spectrophotometric measurements have yielded information concerning photosynthetic excitation and electron transfer (reviewed in Joliot et al., 1998). The use of electron donors and acceptors, inhibitors of photosynthetic electron flow, chloroplast and cytoplasmic protein synthesis, and $^{35}SO_4^{=}$ for labeling polypeptides, have all been successfully employed for in vivo analyses of photosynthetic activity and chloroplast biogenesis. In vivo pulse labeling of proteins was used to define discrete steps in the biosynthesis of the nuclear-encoded, chloroplast polypeptides cytochrome c6, plastocyanin and the 33 kDa oxygen-evolving complex protein (Howe and Merchant, 1993). Pulse-labeling of vascular plant polypeptides is much more difficult and there may be a considerable lag between the time that the label is administered (to the roots or to a cut edge of the stem) and the time that it is incorporated into chloroplast proteins.

The ability of Chlamydomonas to grow heterotrophically in the dark has permitted the development of several genetic screens for isolating mutants impaired in photosynthesis. Chlamydomonas offers the advantage over vascular plants of rapid phenotypic examination of thousands of colonies on solid medium. In early studies, the characterization of Chlamydomonas mutants elucidated the order of electron carriers in the photosynthetic electron transport chain (Levine, 1969). Furthermore, Chlamydomonas mutants extremely sensitive to light can be readily maintained since the alga can grow well in the dark on acetate (it is difficult to sustain vascular plant photosynthetic mutants that are sensitive to light).

Recently, a number of molecular tools have been tailored for use with Chlamydomonas, and the facility with which these tools can be used (Rochaix, 1995; Lefebvre, 1999; Grossman, 2000; Harris, 2001) has led investigators to bestow the title of ‘green yeast’

(Goodenough, 1992) on this organism. Selectable markers are available for identification of nuclear and chloroplast transformants (Kindle, 1990; Boynton, 1988) as are relatively simple procedures to introduce DNA into cells (Kindle, 1990; Shimogawara, 1998). Furthermore, reporter genes such as GFP (Fuhrmann, 1999) and arylsulfatase (Davies et al., 1992), and antisense suppression of mRNA levels (Schroda, 1999) have been effectively used in Chlamydomonas.

Physiological, genetic and molecular manipulations of Chlamydomonas have become routine and have made this organism ripe for more extensive gene studies. Global analyses of the genome will add considerably to our understanding of photosynthetic function and the ways in which photosynthetic activity is modulated as environmental conditions change. Sequence analysis of cDNAs derived from RNA isolated from Chlamydomonas cells exposed to a number of different environmental conditions is providing the community with a wealth of interesting sequences, which are serving as substrates for functional genomics. cDNA microarray information will provide one aspect of global gene regulation under different environmental conditions. Analysis of high density DNA microarrays with cDNA probes generated from putative 'regulatory mutants' will enable investigators to identify target genes of specific regulatory elements and processes affected in mutant strains: A number of putative regulatory mutants have already been identified in Chlamydomonas that can be immediately used for such studies. Finally, generating a high-density physical map that is linked to the genetic map (Lefebvre, 1998) will make positional cloning of genes altered (e.g. point mutations) in specific mutants easy and rapid. Many of these goals are currently being achieved under the umbrella of the NSF funded Chlamydomonas genome project.

Procedures and results

cDNA library construction and analysis: One of the primary goals of the Chlamydomonas genome project has been to identify and obtain sequence information from as many genes as possible. To accomplish this we have been constructing, normalizing, and sequencing cDNA libraries. So far, nine cDNA libraries have been generated from cells exposed to a variety of different conditions. The first four libraries constructed were from cells grown in 1) acetate-containing medium in moderate light (~75 μmol photons m^{-2} s^{-1}), 2) acetate-containing medium in the dark, 3) high salt medium at ambient levels of CO_2 and 4) high salt medium in air supplemented with 5% CO_2. An equal number of plasmids from each of these libraries were combined and normalized as a single "Core" Library. Approximately 40,000 clones from these libraries have been sequenced from both the 5' and 3' ends. To enhance the efficiency of gene identification, we pooled the clones identified in the first round of sequencing, subtracted these from the core library and began sequencing this re-normalized library.

In addition to the Core Library, five other libraries have been constructed and normalized (**Table I)**. The "Stress I" and "Stress II" Libraries were constructed from RNA isolated from cells at various times after the cells were shifted to stressful growth conditions. RNA used to construct the Stress I Library was generated from a pooled population of cells that were treated in various ways. The treatments included shifting cells from NO_3^--containing to NH_4^+-containing medium, from NH_4^+-containing to NO_3^--containing medium, from NH_4^+-containing medium to medium lacking nitrogen, from SO_4^{-2}-containing medium to medium lacking sulfur, and from PO_4^{-3}-containing medium to medium lacking phosphorus. The RNA used to construct the Stress II Library also represented a pooled population. RNA was isolated from cells shifted from NH_4^+-containing to NO_3^--containing medium for 24 h, starved for sulfur and placed in an anaerobic environment (a condition that induces hydrogen

production), and treated with sub-lethal concentrations of potentially toxic compounds. These compounds included the oxidizing agent H_2O_2, the osmotically active compound sorbitol, and the heavy metal Cd. RNA from cells deprived of iron was also included in the population of RNA used to construct this library. A "Deflagellation" Library was constructed from RNA isolated from cells that were re-growing flagella after being deflagellated by pH shock, and a "Gametogenesis and Zygote Formation" Library was constructed from RNA from cells of both mating types that had been induced to form gametes. The gamete population was also mixed so that the mating process would be initiated. RNA was isolated at various times during these processes. We are continuing to construct cDNA libraries and have obtained three separate RNA samples from cells deprived of copper, iron and oxygen. We also have RNA from cells in synchronous circadian phases and will isolate RNA from cells adapting to high light. All together, these libraries are providing a rich source of genes that are expressed during normal growth conditions and during adaptive processes.

Table 1. cDNA Libraries

Library	Str.	Condition	Status	Normal-ization	Plates Sequenc-ed
Core	21gr	TAP light, TAP dark, HS+CO_2, HS	Finished	2	228
Stress I	21gr	$NO_3 \rightarrow NH_4$ 0.5, 1, 4 h $NH_4 \rightarrow NO_3$ 0.5, 1, 4 h TAP - N 0.5, 1, 4 h TAP - S 0.5, 1, 4 h TAP - P 4, 12, 24 h	Finished	1	125
Stress II	21gr	NO_3 24 h H_2 production 0, 12, 24 h TAP + H_2O_2 1, 12, 24 h TAP + Sorbitol 1, 2, 6, 24 h TAP + Cd 1, 2, 6, 24 h	Finished	1	50
Stress III	21gr	Fe / Cu deficiency			
Stress IV	21gr	High Light			
Deflagellation	21gr	Deflagellated	Finished	1	0
GAZ	21gr	Gamete and Zygote	Finished	1	0
Polymorphic strain	S1D2	TAP	Finished	1	8

Table I. Status of the libraries and the sequences completed. The strain used for generating the libraries was 21gr and S1D2 (listed under **str**.). 'Plates sequenced' represent clones in 96 well microtitre dishes that have been sequenced.

Normalization of the libraries has helped reduce the relative copy number of genes that are expressed at high and moderate levels. Results of normalization on the abundance of cDNAs for RbcS (small subunit of RuBP carboxylase) in the Core Library and Ats1 (ATP sulfurylase) in the Stress I Library are shown in **Figure 1**. The unnormalized and normalized phage libraries were plated (on the left and right side of the figure, respectively) at a density of approximately 5,000 plaque forming units per plate. Normalization resulted in a considerable reduction in clones for both RbcS and Ats1.

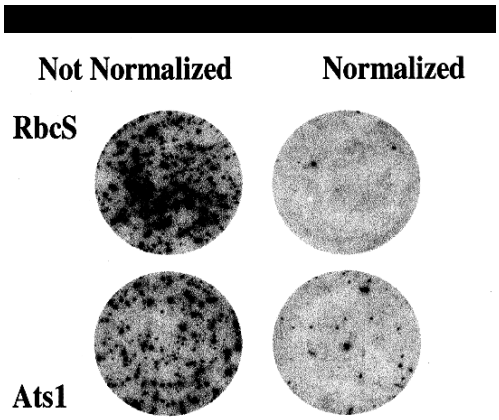

Figure 1. Effect of normalization on the abundance of clones for RbcS and Ats1 in cDNA libraries.

In total, sequences from over 40,000 Chlamydomonas cDNAs (sequenced from both the 5' and 3' ends) have been generated, and most of these sequences have been assembled into unique contigs (each containing multiple sequences that have been aligned) or singletons (unique sequences). Information generated in the latest assemblies suggests that we have identified approximately 7,500 unique cDNAs that were assembled from a total of approximately 45,000 sequencing reads (this includes both 5' and 3' reads). The reads were specific for the Core and Stress I Libraries (no reads were used from the Stress II Library).

We have developed specially designed "intelligent" protocols to perform sequence assembly using the Lisp programming language; this language is primarily applied to research involving Artificial Intelligence. The protocol, part of which is presented in **Figure 2**, is based upon the commonly-used Phrap assembly program, and involves four major phases. In the first phase only the 3' ends of the cDNA sequences are assembled into contigs; the assembly is generated by a repeating quality-control cycle. In the second phase, the 5' ends associated with each 3' end are assembled. At each step of the assembly process, the quality of both the reads and contigs is assessed based upon values reported by Phred and Phrap, and additional phase-specific constraints. For example, when assembling the 3' ends, which are expected to have poly-A tails, we ensure that these tails line up with one another. In the third phase duplicate contigs are removed using semi-automated reasoning about the results of internal cross-match between all contigs. The process of duplicate removal is especially important for microarray construction since it ensures that all of the indexed positions on the array represent unique cDNA clones. The final phase of assembly involves intelligent contig annotation. In this process the contigs are mass blasted (tblastx) and then heuristics are used to determine function in a genome functional taxonomy, developed from Cyanobase, KEGG, and EcoCyc.

The result of our efforts with respect to cDNA sequence generation and assembly is a high quality database for a large fraction of the Chlamydomonas cDNAs that has been partially annotated. Moreover, since we know which cDNA library each read came from, we can perform "in silico" subtractions, statistically sorting the contigs according to those that are specific for each library. For example, we have determined which genes are most highly represented in the Stress I Library and not present in the Core Libraries.

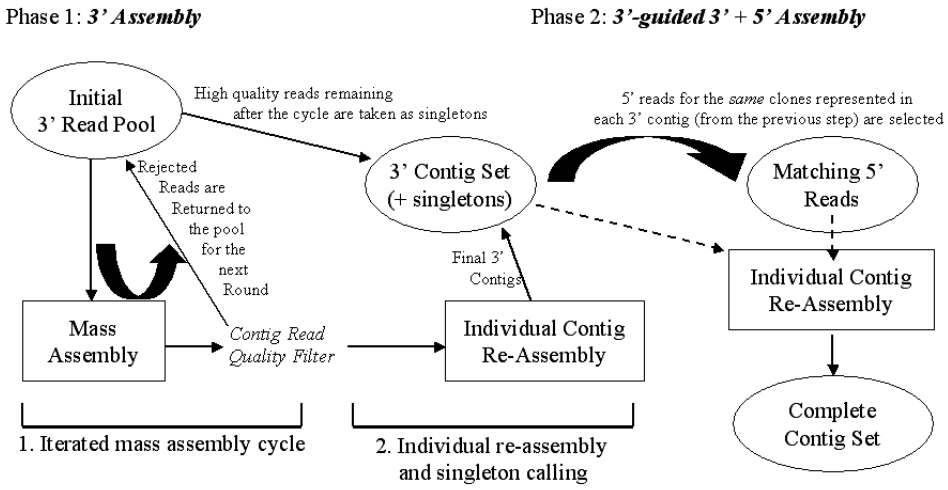

Figure 2. Diagrammatic representation of phases of cDNA assembly.

Table II. Putative polypeptides encoded by genes prevalent in the Stress I library

Contig Number	tblastx Prob.	Description of putative protein product	3' reads (963)
3114	1.00E-43	nucleoredoxin, MGC (Mus musculus)	18
3292	4.00E-38	chlorophyll a/b-binding CP29 F7A7_50 (A. thaliana)	17
2996	6.00E-03	IV early light-inducable Elip (Arabidopsis thaliana)	15
2966	8.00E-03	ferredoxin-binding psi-d1 (Zea mays)	14
2857	2.00E-38	cyclin A1 (Xenopus)	13
2689	1.00E-11	scaffold (Drosophila melanogaster)	12
3643	0	extracellular polypeptide Ecp76 (C. reinhardtii)	12
2708	0	proteasome beta PBD2 PBD2 (A. thaliana)	11
2756	6.00E-25	cysteine proteinase (Nicotiana tabacum)	11
3642	0	putative ribosomal s12 (Fragaria x ananassa)	11
2449	0	proteasome PAA1 PAA1 (Arabidopsis thaliana)	10
2504	2.00E-15	isoflavonoid reductase homologue (A.thaliana)	10
2545	1.00E-33	pherophorin (Volvox carteri)	10
2270	8.00E-10	Sec61 beta-subunit (A.thaliana)	9
1779	3.00E-12	voltage-dependent anion channel vdac2 cds;nuclear mitochondrial (Zea mays)	8
1822	6.00E-27	pherophorin (Volvox carteri)	8
2190	4.00E-27	GTP-binding protein, RAB5A (L.japonicus)	8
2380	0	DEAD box RNA helicase, RH2 (A. thaliana)	8
3768	1.00E-09	early light inducible (Medicago sativa)	8
3722	2.00E-38	cystathionine gamma-synthase CGS (A. thaliana)	7
1877	2.00E-31	glutathione peroxidase-like GPX54Hv (H. vulgare)	6

The library-specific gene products, presented in **Table II,** have a range of different putative activities that are beginning to reveal processes important for acclimation to stress conditions. Genes encoding specific chlorophyll a,b binding polypeptides, early light inducible proteins and a ferredoxin binding protein may be important for restructuring photosynthetic function during stress, while the synthesis of specific proteasome subunits may be critical for modifying protein content of cells. Cyclins are likely involved in controlling cell cycle during suboptimal growth while glutathione peroxidase may provide protection against oxidative stress. This is just a peek at the genes that appear to become activated under specific stress conditions and such analyses are providing clues that can be exploited by individual investigators for unraveling the dynamics of cell metabolism in a changing environment.

Acknowledgements:

We would like to acknowledge NSF MCB 9975765 for supporting the work presented in this manuscript. We also thank people at the Stanford Center for Genome Sequence and Technology for both providing strong technical support and helpful discussions. Finally, we thank other members of the team of investigators including David Stern, Pete Lefebvre and Carolyn Silflow for helping to make our efforts fruitful.

References:

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, Sanford JC. 1988. Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. *Science* 240:1534-1538.

Davies J, Weeks, W, Grossman AR (1992) Expression of the arylsulfatase gene from the ß$_2$ -tubulin promoter in *Chlamydomonas reinhardtii*. *Nucleic Acid Res* 20:2959-2965.

Fuhrmann M, Oertel W, Hegemann P. 1999. A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. *Plant J* 19:353-361.

Goodenough, UW. 1992. Green yeast. *Cell* 70:533-538.

Grossman, A. R. 2000. *Chlamydomonas reinhardtii* and photosynthesis: genetics to genomics. *Curr Opin Plant Biol* 3:132-137.

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

Howe G, Merchant S. 1993. The maturation of thylakoid lumenal proteins in *Chlamydomonas reinhardtii* proceeds post-translationally through an intermediate in vivo. *Proc Natl Acad Sci USA* 19:1862-1866.

Joliot P, Béal D, Delosme R. 1998. In vivo measurements of photosynthetic activity: Methods. In 'The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas'. Edited by Rochaix J-D, Goldschmidt-Clermont M, Merchant S: Kluwer Academic Publishers, pp. 433-449.

Kindle KL, Schnell RA, Fernández E, Lefebvre PA. 1989. Stable nuclear transformation of Chlamydomonas using the Chlamydomonas gene for nitrate reductase. *J Cell Biol* 109:2589-2601.

Kindle KL. 1990. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA* 87:1228-1232.

Lefebvre PA, Silflow CD. 1998. Chlamydomonas: The cell and its genome. *Genetics* 151:9-14.

Rochaix, J-D. 1995. *Chlamydomonas reinhardtii* as the photosynthetic yeast. *Annu Rev Genet* 29:209-230.

Schroda M, Vallon O, Wollman F-A, Beck C. 1999. A chloroplast targeted heat shock protein 70 (HSP70) contributes to the photoprotection and repair of photosystem II during and after photoinhibition. *The Plant Cell* 11:1165-1178.

Shimogawara, K., S. Fujiwara, A. R. Grossman, and H. Usuda. 1998. High efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. *Genetics* 148:1821-1828.