ORIGINAL RESEARCH PAPER
Volume - 10 | Issue - 08 | August - 2021 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

A STUDY ON ACTIVATION OF THE CARBON CONCENTRATING MECHANISM BY CARBON DI-OXIDE DEPRIVATION OVERLAPS WITH MASSIVE TRANSCRIPTIONAL REARRANGEMENT IN CHLAMYDOMONAS REINHARDTII

Biochemistry

Dr. Anand Shanker Singh Associate Professor, Department Of Chemistry, Chinmaya Degree College, Bhel , Haridwar

Dr . G . Radhika MBBS, MD (Biochemistry), HOD , Department Of Biochemistry, Sri .venkateshwara Medical College And Hospital , Pondicherry

Dr. Ankita Singh* DNB, DGO , MBBS *Corresponding Author

Dr. Debarshi Jana Young Scientist, IPGMER And SSKM Hospital, Kolkata, WB.

ABSTRACT

A CO2-concentrating mechanism (CCM) is essential for the growth of most eukaryotic algae under ambient (392 ppm) and very low (<100 ppm) CO2 concentrations. In this study, we used replicated deep mRNA sequencing and regulatory network reconstruction to capture a remarkable scope of changes in gene expression that occurs when Chlamydomonas reinhardtii cells are shifted from high to very low levels of CO2 (≤100 ppm). CCM induction 30 to 180 min post-CO2 deprivation coincides with statistically significant changes in the expression of an astonishing 38% (5884) of the 15,501 nonoverlapping C. reinhardtii genes. Of these genes, 1088 genes were induced and 3828 genes were downregulated by a log2 factor of 2. The latter indicate a global reduction in photosynthesis, protein synthesis, and energy-related biochemical pathways. The magnitude of transcriptional rearrangement and its major patterns are robust as analyzed by three different statistical methods. De novo DNA motif discovery revealed new putative binding sites for Myeloid oncogene family transcription factors potentially involved in activating low CO2–induced genes. The (CA)n repeat ($9 \leq n \leq 25$) is present in 29% of upregulated genes but almost absent from promoters of downregulated genes. These discoveries open many avenues for new research.

KEYWORDS

Carbon Concentrating Mechanism, Chlamydomonas reinhardtii, mRNA.

INTRODUCTION

In nature, Chlamydomonas reinhardtii and other eukaryotic algae depend on a CO2-concentrating mechanism (CCM) to supply sufficient inorganic carbon (Ci; CO2 or bicarbonate) for photosynthesis-fueled cell growth and proliferation. Mutant cells lacking key components of the CCM molecular machinery or its regulatory system do not grow or grow poorly unless supplied with high concentrations of CO2 (e.g., >10,000 ppm) that are well above the ambient level of ~392 ppm.[1] Because the diffusion rate of CO2 in aqueous environments is ~10,000 times slower than in air, most natural populations of microalgae exist in CO2-limited conditions. This is especially true for dense algal populations growing under abundant sunlight. Under such conditions, CO2 concentrations can become very low (<100 ppm) and cells induce the CCM to maximal levels. CO2 starvation induces the transcription of numerous genes encoding proteins closely associated with the CCM and its activities.[1,2] Indeed, C. reinhardtii and most other eukaryotic algae have developed a finely tuned regulatory system that suppresses expression of CCM-related genes under conditions of replete CO2 (i.e., >0.1% CO2) and activates expression of these genes when CO2 becomes limiting. [1,2] Previous studies using RNA gel blot analyses and microarray analyses have revealed a number of CCM-associated genes and other CO2-responsive genes whose transcription is tied to the physiological changes that accompany cell acclimation to CO2 stress conditions.[3]

Here, we report an extensive global analysis of the massive transcriptional changes evoked by the deprivation of Ci in *C. reinhardtii*. We measured these transcriptional events using replicated deep RNA sequencing (RNA-Seq) on the Illumina platform. The highly reproducible RNA-Seq experiments not only confirm earlier observations based on array analyses quoted above but also extend the list of differentially expressed genes from a few hundred to over 4000.

We report the discovery of an extensive system of head-to-head (HTH; also called divergent) gene pairs, many of them sharing bidirectional or connected promoters. HTH conformation and bidirectional or shared promoters frequently perform the highly accurate coregulation of gene pairs encoding subunits of the same protein complex or two proteins of similar or related functions. Here, we focus on those HTH, coregulated, gene pairs that are most relevant to the CCM. Advanced computational techniques also have allowed an extensive evaluation of potential regulatory elements in promoter regions in CO2-responsive genes and the discovery of new elements shared by several of the most highly stimulated CO2-responsive genes. We also report a previously unrecognized pattern of expression for many genes that suggests a significant, but transient, decrease in gene transcription immediately after a shift to very low CO2 conditions (ASVLCO2). Finally, we employ a vastly expanded pool of transcriptomic data to strengthen earlier observations of metabolic and physiological changes that occur when CO2 becomes limiting in the environment, including significant decreases in transcripts encoding proteins involved in photosynthesis, cytoplasmic, chloroplastic, and mitochondrial protein synthesis, energy use, protein transport, and other Gene Ontology (GO) categories.[4]

METHODS AND MATERIALS

Ci Deprivation

Chlamydomonas reinhardtii wild-type strain CC124 was used for analysis. Briefly, cells were grown in 2 liters of Tris Phosphate medium at 25°C and 3% CO2 to a density of 1 X 106 cells/mL before being transferred to a 3-liter autoclavable glass bioreactor (Applikon Biotechnology) that was connected with EZ control for analysis of temperature, pH, and dissolved oxygen.[5] The bioreactor was illuminated with a light intensity of 200 µmol photons m-2 s-1, and an input gas containing 5% CO2 was introduced. Algal cells were allowed to equilibrate with the new environment for 1 h. Following a sampling of the culture, the input gas for the bioreactor was shifted to 100 ppm CO2, which was monitored in the culture using two CO2 transmitters (Vaisala; models GMT221 and GMT222). Samples were taken at 15, 30, 60, and 180 min following the shift to 100 ppm CO2. During the experiment, pH was maintained at 7.2 using 3 M KOH.

Preliminary Analysis of RNA Samples

To confirm induction of the carbon-concentrating mechanism, preliminary analysis of RNA samples was performed using qRT-PCR. RNA samples were prepared for analysis using the Plexor Two-Step qRT-PCR system (Promega). qRT-PCR analysis was performed using a 7500 Real-Time PCR System (Life Technologies). The genes LCIA (AB168092), LCIB (XM_001698292), and mitochondrial carbonic anhydrase (CAH4, XM_001695951) were chosen for analysis as they have been observed to increase in expression during carbon deprivation.[3,6] CAH2 (X54488) was also selected as a control gene reported as displaying a moderate decrease in expression in response to carbon deprivation. CIA5/CCM1 (AF317732) was used as a positive control as it shows constitutive expression during carbon deprivation. Fluorescently labeled primer pairs were designed for each of the aforementioned genes. Quantitative PCR analysis was performed using a 7500 Real-Time PCR System by measuring the threshold cycle (Ct) of each gene. Using the Ct values of CIA5/CCM1 for each RNA sample as a baseline control, the change in Ct for each gene could be used to calculate the fold change response of each gene throughout the time course.

RNA Sequencing, Mapping, and the Analyses of Gene Expression

From the qRT-PCR data, it was determined that four time points should be analyzed by RNA-Seq. A 15-min ASVLCO2 sample was omitted from RNA-Seq as qRT-PCR analysis of this sample showed limited induction of the aforementioned genes. C. reinhardtii equilibrated at ~5% CO2 was used as the 0 time control, and three time points ASVLCO2 (30, 60, and 180 min) were also analyzed. To provide for biological replicates, RNA samples from two individual bioreactor runs were analyzed. In total, eight RNA samples were submitted for RNA-Seq. Prior to submission, RNA samples were treated with DNase and resuspended in 8.3 mM Tris-HCl and 4.2 mM EDTA. RNA-Seq was performed at the JGI using an Illumina Genome Analyzer II.

Sequencing reads were mapped to the C. reinhardtii version 4 genome (Department of Energy JGI) as well as to the processed Augustus5 exon structure predictions using the tophat and cufflinks software.[7] No more than two mismatches per sequencing read were allowed. Analyses of differential expression including FDR calculations were performed using three independent Bioconductor packages: edgeR, DESeq and baySeq.[8]

Time series analysis of the transcriptional response was performed by k-means cluster analysis.[9] This method partitions fold change patterns into k clusters where each fold change time series belongs to the cluster with the nearest mean. The clusters are iteratively refined. Such analyses have been used to identify temporal expression patterns of a large numbers of genes.

Gene Ontology Analyses

Complex functional patterns of the differentially regulated genes emerge at the level of photosynthetic categories, low CO2–regulated genes, and plant and diatom genes that have no close relatives in other kingdoms or in prokaryotes other than cyanobacteria. We extended these analyses to GO, a system for the hierarchical annotation of homologous gene and protein sequences in multiple organisms using a common, controlled vocabulary. GO allows the practical, high-throughput interpretation of experiments including RNA-Seq. To avoid the subjectivity inherent in the ad hoc interpretations for less than obvious patterns, a rigorous method was employed to assess the statistical significance of expression patterns, called GSEA.[10] Briefly, GSEA ranks genes by fold changes and calculates enrichment scores for each set. Then, primarily upregulated gene sets are assigned high positive enrichment scores and primarily downregulated sets are assigned low negative scores. For the statistical significance of these enrichment scores, FDR is calculated.[11]

De Novo Motif Discovery of Putative Transcription Factor Binding Sites

Our complex strategy for the discovery and limited confirmation of the transcriptional regulatory network was described earlier. Even in the almost complete absence of chromatin immunoprecipitation, protein binding array, or protein–protein interaction observations for algae, an array of motif discovery algorithms for promoter sequence analysis, each complementing the others, knockout mutants of transcription factors, and RNA-Seq data allowed us to better understand the CCM regulatory network. A key tool is the MEME package for the identification of statistically overrepresented variable sequence motifs. We searched all promoter regions for all motifs represented as positional weight matrices in the commercial version of the TRANSFAC Database using its advanced search tool. Conversely, all identified motifs were queried against the TRANSFAC motifs.[12]

RESULTS AND DISCUSSION

Ci deprivation is a major stress that evokes a dramatic transcriptional response in algae. Using EST-based macroarrays, the Fukuzawa laboratory[3] and Grossman and Weeks laboratories pioneered the transcriptional profiling of Ci deprivation. In initiating our studies, our hypothesis was that revolutionary progress in sequencing technology and statistical methodology would allow us to discover a large number of activated or repressed biological processes and individual genes that may have escaped detection using EST arrays. To test this hypothesis, we performed deep RNA-Seq using the Illumina Genome Analyzer II platform at the Joint Genome Institute (JGI) of the Department of Energy. In total, the eight samples collected at four time points (0, 30, 60, and 180 min after Ci deprivation) produced 98.3 million uniquely mapped sequencing reads (12.3 million 71-base-long reads per sample). When no more than two mismatches were allowed in the anchor regions, ~38% of the reads did not map uniquely or contained more than two base errors due to sequencing errors, genomic variability, alternative splicing, a number of recently duplicated genes, and repetitive DNA elements. Even with this conservative approach, RNA-Seq represents a major advance from micro- and macroarrays: It provides an unprecedentedly high coverage of transcripts, eliminates cross-hybridization effects, does not rely on the commercial availability of arrays, and is more robust against errors in predicted exon structures. The significantly increased performance of RNA-Seq has been shown specifically for *C. reinhardtii*.[13]

The high technological reproducibility of the RNA-Seq measurements performed at the Department of Energy's JGI is shown by the strong correlations of transcript levels between biological replicates (0.958, 0.965, 0.939, and 0.973 for 0, 30, 60, and 180 min time points after carbon deprivation, respectively). These high Pearson correlation coefficients indicate reproducible and multiplicative (linear) biases and that the nonlinear bias is miniscule. Note that linear, multiplicative, and reproducible bias does not alter fold change values by multiplying the transcript levels both in the numerator and the denominator. Such biases include sequencing reads that match imperfectly to the genome or the transcriptome, amplification, and sequencing biases. Additive effects, such as unreal exons, may reduce the extent of differential expression. These effects are due to imperfect gene models, such as those predicted by the Augustus method[7] and alternative splicing. Such additive effects remain our primary concern. Recently duplicated genes pose further challenges in mapping the 71-base-long sequencing reads to the transcriptome because these reads contain erroneous base calls, particularly at their 3' ends. Such gene pairs include major effectors of CO2 concentration, such as four carbonic anhydrases (CAHs), CAH1-CAH2 and CAH4-CAH5, that are recent duplicates. The pair CAH4-CAH5, for example, contains exons that are over 90% identical.[14]

To avoid mappings to the duplicated genes, rigorous procedures (with one or two mismatches in the anchor regions that connect two exons) are necessary. However, this rigor also drastically reduces the coverage of all genes due to both sequencing errors and polymorphisms. Reduced coverage reduces the number of significantly differentially expressed genes. Therefore, we performed the mapping with both one and two allowed mismatches in the anchor region, as implemented in the tophat program. With one allowed mismatch, fewer but more accurate transcript levels were obtained than with two mismatches. For example, our data, as expected from earlier studies, demonstrated that CAH1 is strongly upregulated at 3 h ASVLCO2. However, in our initial analyses allowing two mismatches, CAH2, which had earlier been reported not to respond to CO2, was falsely classified as upregulated. When reanalyzed using only one mismatch per read, the vast majority of reads in the 60- and 180-min time points were mapped to the CAH1 gene, with few being attributable to CAH2. Quantitative RT-PCR (qRT-PCR) confirmed the results of the more rigorous alignments.

The lists of differentially expressed genes may be influenced by the choice of statistical methodology. Therefore, we analyzed our data using three different computational tools, edgeR, DESeq, and baySeq. Because differential expression of a large number (~16k) of genes is estimated using very few replicates (samples), statistical tools derived from large-sample asymptotic theory do not work. In particular, small sample size affects the correction for overdispersion (greater variability than expected based on Poissonian or other simple models), modeling the empirical distributions, and calculating statistical significance. To solve these issues, edgeR shrinks genewise dispersion estimates toward a constant value using an empirical Bayesian model and performs Fisher's exact test. DESeq uses nonparametric regression models to fit the negative binomial variance as a function of the mean, assuming a locally linear relationship between overdispersion and mean expression levels. baySeq is free of this assumption and uses a fully empirical Bayesian approach to estimate the posterior probabilities. We compared the numbers of overlapping differentially expressed genes reported by the edgeR, DESeq, and baySeq packages at 180 versus 0 min ASVLCO2. Because the exact test implemented in edgeR calculates lower false discovery rate (FDR) q-values than the other two methods, at FDR $\leq$ 0.01, edgeR, DESeq, and baySeq reported 4222, 2364, and 3248 differentially expressed genes, respectively.[15] The lists of differentially expressed genes are more consistent when the FDR threshold is elevated to 0.05, a still conservative level. All three methods reported differential expression for as many as 3141 genes. An additional 702 genes were jointly reported by both edgeR and baySeq, and a further 95 genes were called jointly by edgeR and DESeq. Because of the high overlaps with other methods, and its wider acceptance, below we limit our discussions to the results obtained by the edgeR tool.

Our results reproduced the observed induction of major CCM-associated genes published by Miura et al. (2004)[3] and Yamano et al. (2008)[6] as well as in the companion article[16]. In addition, we report a large number of genes that have not been associated with the CCM previously. Some of the notable similarities and differences in gene sets of our studies and those presented in our companion paper are discussed throughout this section (with special attention to the potential causes of observed differences provided near the end of this section).[16]

Major Transcriptional Changes

Our results greatly extend many aspects of earlier observations of differentially expressed genes following CO2 deprivation. This is indicated by relatively similar lists of induced genes published previously[3,6] and by us. In addition, RNA-Seq and modern statistical methodologies allowed us to discover an unexpectedly high 5884 genes that are differentially regulated at either 30, 60, or 180 min ASVLCO2 relative to the 0 min control [FDR ≤ 0.01 and abs(log2 (fold change)) ≥ 1] or 3828 genes [FDR ≤ 0.001 and abs(log2(fold change))≥2].

Robust temporal expression patterns emerged under our conditions for imposition of CO2 deprivation. We found that the transcriptional response becomes widespread only after 30 min and increases (or decreases) for many, but not all, genes. The relatively slow onset of significant transcript changes is likely coupled to the relatively slow decline in CO2 concentrations employed in our experiments. At 30 min after deprivation, we found only 37 upregulated and five repressed genes relative to the 0 min control (FDR ≤ 0.001 and abs[log2(fold change)≥2; or in absolute, nonlogarithmic scale, a fourfold increase or decrease). At an hour ASVLCO2, 409 genes are upregulated and 1663 genes are repressed. At 3 h ASVLCO2, 981 genes are induced and 1188 genes are repressed. These numbers are approximately doubled at the more typical thresholds (FDR q≤0.01 and abs[log2(fold change)≥1;). To measure transcript levels by a different method, we performed qRT-PCR analyses on the same RNA samples that were submitted for Illumina sequencing. Three different genes induced by CO2 deprivation (Low CO2 Induced A (LCIA), CAH5, and LCIB) displayed similar expression patterns between RNA-Seq and qRT-PCR, while a fourth gene, CAH2, previously reported as not responding or responding negatively to CO2 depletion[17], showed moderate decreases in transcript levels using both RNA-Seq and RT-PCR measurements.

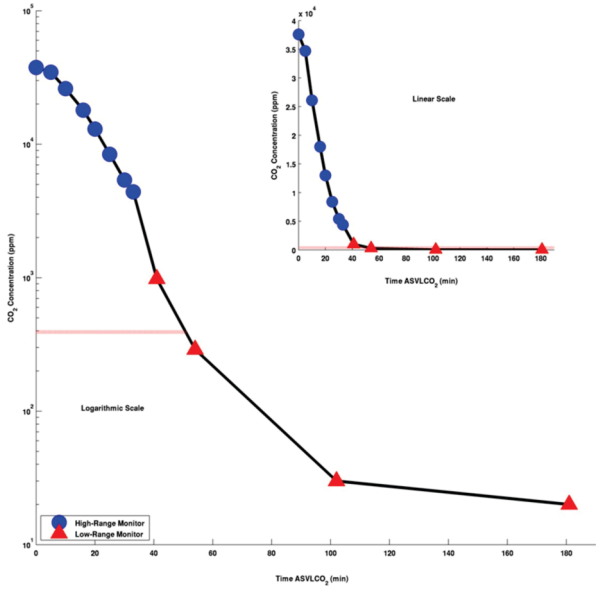

Figure 1 - *Measurements of CO2 Levels Following a Shift of C. reinhardtii Cells from 5% to 100ppm CO2. Two CO2 monitors were used in the fermenter: One was calibrated for high CO2 concentrations (circles), and the other was calibrated to low CO2 concentrations (triangles). CO2 concentrations are plotted on a log10 scale. The horizontal red line represents the 392 ppm concentration of the atmosphere. The reduction in CO2 is represented both in a linear (inset) and logarithmic scale.*

CONCLUSION

The strong correlation of transcript levels in data obtained from biological replicates used for RNA-Seq analyses. Together, these observations confirm the high technological reproducibility of deep RNA-Seq as well as the reproducibility of our biological samples. The magnitude of transcriptional rearrangement and its major patterns are robust as analyzed by three different statistical methods. De novo DNA motif discovery revealed new putative binding sites for Myeloid oncogene family transcription factors potentially involved in activating low CO2–induced genes. The (CA)n repeat (9 ≤ n ≤ 25) is present in 29% of upregulated genes but almost absent from promoters of downregulated genes. These discoveries open many avenues for new research.

REFERENCES

1. Moroney, J.V., and Ynalvez, R.A. (2007). Proposed carbon dioxide concentrating mechanism in Chlamydomonas reinhardtii. Eukaryotic Cell 6, 1251-1259.
2. Duanmu, D., Miller, A.R., Horken, K.M., Weeks, D.P., and Spalding, M.H. (2009a). Knockdown of limiting-CO2-induced gene HLA3 decreases HCO3- transport and photosynthetic Ci affinity in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America 106, 5990-5995.
3. Miura, K., Yamano, T., Yoshioka, S., Kohinata, T., Inoue, Y., Taniguchi, F., . . .Fukuzawa, H. (2004). Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol 135, 1595-1607.
4. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-29.
5. Harris, E.H. (1989). The Chlamydomonas sourcebook : a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).
6. Yamano, T., Miura, K., and Fukuzawa, H. (2008). Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in Chlamydomonas reinhardtii. Plant Physiol 147, 340-354.
7. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, Ii215-Ii225.
8. Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. Bmc Bioinformatics 11.
9. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), pp. 281-297.
10. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., .. . Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545- 15550.
11. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57, 289-300.
12. Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., . . .Wingender, E. (2006). TRANSFAC (R) and its module TRANSCompel (R): transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34, D108-D110.
13. Gonzalez-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S.S., and Grossman, A.R. (2010). RNA-Seq Analysis of Sulfur-Deprived Chlamydomonas Cells Reveals Aspects of Acclimation Critical for Cell Survival. Plant Cell 22, 2058- 2084.
14. Villand, P., Eriksson, M., and Samuelsson, G. (1997). Carbon dioxide and light regulation of promoters controlling the expression of mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii. Biochemical Journal 327, 51-57.
15. Lopez, D., Casero, D., Cokus, S.J., Merchant, S.S., and 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.
16. Fang, W., Si, Y.Q., Douglass, S., Casero, D., Merchant, S.S., Pellegrini, M., . . . Spalding, M.H. (2012). Transcriptome-Wide Changes in Chlamydomonas reinhardtii Gene Expression Regulated by Carbon Dioxide and the CO2-Concentrating Mechanism Regulator CIA5/CCM1. Plant Cell 24, 1876-1893.
17. Moroney, J.V., Ma, Y.B., Frey, W.D., Fusilier, K.A., Pham, T.T., Simms, T.A., . . .Mukherjee, B. (2011). The carbonic anhydrase isoforms of Chlamydomonas reinhardtii: intracellular location, expression, and physiological roles. Photosynthesis Research 109, 133-149.