2009

The exaptation of nitrate/carbon stress-induced smRNAs and their targets from transposable elements in the unicellular green alga Chlamydomonas reinhardtii

Heather Marie Tyra
University of Iowa

THE EXAPTATION OF NITRATE/CARBON STRESS-INDUCED smRNAS AND THEIR TARGETS FROM TRANSPOSABLE ELEMENTS IN THE UNICELLULAR ALGA *CHLAMYDOMONAS REINHARDTII*

by

Heather Marie Tyra

A thesis submitted in partial fulfillment
of the requirements for the Master of
Science degree in Biology
in the Graduate College of
The University of Iowa

May, 2009

Thesis Supervisor: Professor Debashish Bhattacharya

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

MASTER'S THESIS

_____

This is to certify that the Master's thesis of

Heather Marie Tyra

has been approved by the Examining Committee
for the thesis requirement for the Master of Science
degree in Biology at the May, 2009 graduation.

Thesis Committee: _____
Debashish Bhattacharya, Thesis Supervisor

_____
Josep Comeron

_____
John Manak

To my family.

ACKNOWLEDGMENTS

I would like to thank my family first and foremost. Without their love and support, this thesis would not have been possible. Thank you to all of the members of the Bhattacharya lab: Valerie Reeb, Adrian Reyes-Prieto, Jeferson Gross, Ahmed Moustafa, CX Chan, Alaa Soliman, Susanne Ruemelle, Jennifer Bai, and Bill Lanier. I would especially like to thank Bill Lanier and Ric Devor for their guidance and advice both in the lab and in life in general. Thank you to my advisor, Dr. Debashish Bhattacharya, for his patience, support, and guidance, as well. I would also like to thank Phil Ecklund and Eileen Sullivan for all of their help. Thank you to the members of the UI Healthcare system who kept me healthy enough to complete this process. It is only because of the support of the people named above that I was able to complete this thesis.

ABSTRACT

Transposable elements (TEs) are acknowledged sources of genetic change within organisms. The effects of transposition can range from the disruption or creation of a single gene to large-scale genome rearrangements. Transposition events can result in beneficial mutations which allow an organism to adapt to a new environment. In the last three years, several studies have reported that some miRNAs, small RNAs involved in post-transcriptional gene regulation, have evolved from TEs. miRNAs play an important role in the stress responses of many organisms. Interestingly, TEs are derepressed under the same stress conditions that miRNAs are known to ameliorate. The observation that miRNAs are known to evolve from TEs and that TEs are derepressed under stress conditions lead me to question whether TEs play a role in environmental adaptation through the creation of small RNA networks. To test this idea, *Chlamydomonas reinhardtii* cultures were grown under low carbon, nitrate enriched conditions and the small RNA pool was analyzed. I found that these conditions do stimulate the expression of novel small RNAs and that some of these RNAs and their targets are derived from transposition events.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

INTRODUCTION

Transposable elements (TEs) are segments of DNA capable of duplicating and/or reinserting themselves into new locations within a genome through either RNA (Class 1) or DNA (Class 2) intermediates [1-3]. Genomic analyses indicate that these elements are present in all cells to a surprising degree, comprising approximately 45% of the human genome and as much as 60-80% of grass genomes [4]. The precise role of TEs within a genome has been the subject of debate since their discovery in the late 1940's [3, 5]. Ideas have ranged from certainty that there must be some adaptive role for TEs to a complete dismissal of these elements as genomic parasites that have no purpose beyond their own propagation, often at the expense of their host. Growing evidence over the last fifteen years suggests that although TEs are indeed genome parasites, there are numerous examples of how they have been "domesticated" to fulfill sometimes vital roles in the cell [6]. Through the process of transposition, TEs can add exons to or disrupt the coding sequence of existing genes, change the regulatory regions of a gene leading to a change in its expression pattern, or facilitate recombination leading to both local and large-scale genome rearrangements. There are examples of TEs themselves becoming transcribed into functional genes. One striking example of TE exaptation involves the recruitment of the reverse transcriptase from a LINE element to form telomerase, the enzyme critical for DNA replication in eukaryotes [7]. Some TE exaptations have been shown to help the host adjust to environmental stressors. The California population of *Drosophila simulans* experienced a transposition event which changed the expression pattern of the *Cyp6g* gene conferring increased pesticide resistance. This transposition is absent from the African population indicating a selective sweep of this mutation among California individuals. Other examples of TE exaptation include the increased resistance to methylammonium experienced by *Chlamydomonas reinhardtii* in response to TE lesions near the AMT4 gene [8], the formation of polyadenylation signals from Alu insertions in

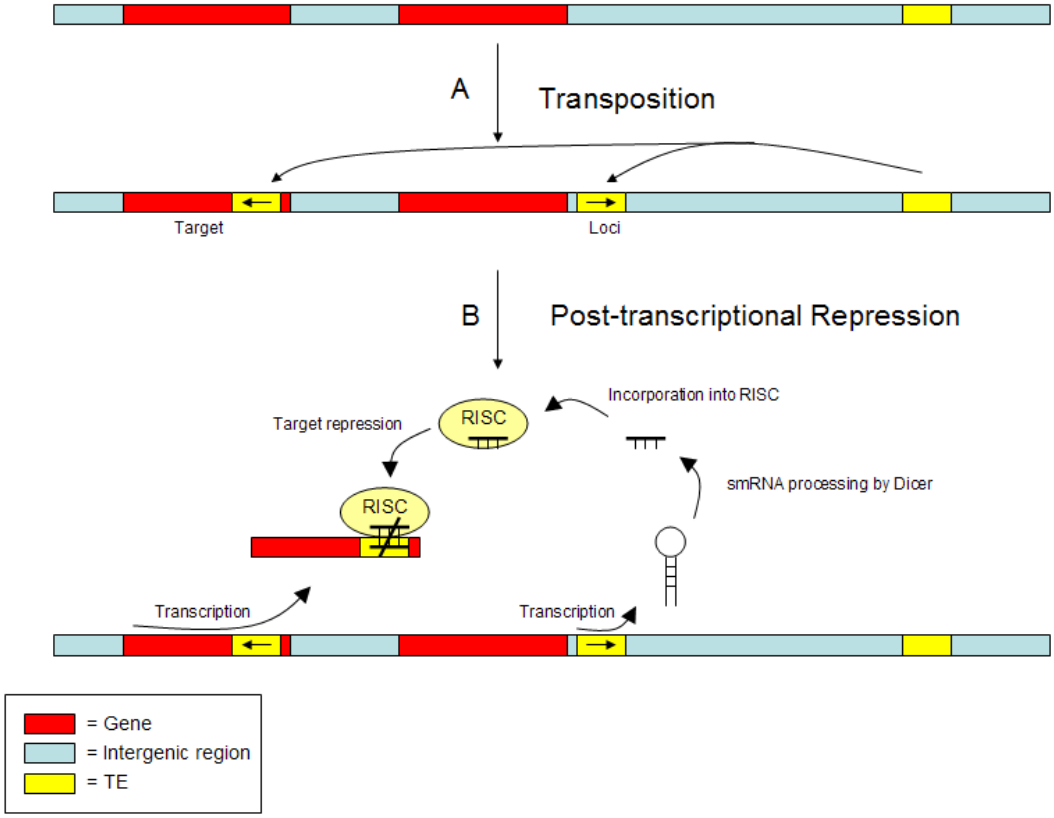
multiple human genes [9], and the reduced susceptibility to Rice Dwarf Virus experienced in an *Oryza sativa* mutant due to an insertion of a Tos17 element into an NAC-domain protein [10]. The examples of beneficial mutations caused by transposition are numerous leading to acceptance of their substantial impact on adaptation.

The mutations caused by TEs are most often deleterious, as is the normal outcome of most types of sequence alteration. Because of this, host genomes have developed elaborate mechanisms designed to control the spread of TEs. RNAi, a small RNA (smRNA)-dependent process involved in gene silencing, is thought to have originally evolved as a genome defense mechanism for controlling the spread of transposable elements and retroviruses [11, 12]. Over time, this mechanism has diversified and can now control endogenous gene expression in addition to TE propagation. The RNAi mechanism includes three classes of smRNAs which control TE propagation and gene expression in different ways: Short interfering RNAs (siRNAs), 21-24mer RNA molecules derived from long double-stranded RNA precursors, that are implicated in cis- and trans-gene silencing, DNA methylation, and TE repression in eukaryotic cells; piwi-interacting RNAs (piRNAs), 28-30mer RNAs, that specifically repress TEs in metazoan germ line cells; and microRNAs (miRNAs), 20-22mer RNAs derived from approximately 150nt long hairpin precursors, that are primarily linked to post-transcriptional gene regulation [13]. Post-transcriptional gene regulation is accomplished after the 20-24mer smRNA is processed from its precursor by the nuclease Dicer and incorporated into a large protein complex called the RISC complex. The smRNA acts as a guide and through standard Watson-Crick base pairing directs the RISC complex to a complementary region within a target mRNA. If the smRNA is almost perfectly complementary to its target (no more than three mismatches are allowed) the RISC complex will cleave the mRNA at the hybridization site leading to mRNA degradation. If the complementarity of the smRNA is limited primarily to the its first 8 bases, the

RISC complex will not cleave the target but remain bound to the target mRNA preventing translation [13, 14].

That smRNAs control TE propagation and that the RNAi mechanism gave birth to a method of endogenous gene control has been known for years [13, 14]. A direct connection between TEs and endogenous gene control has come to light only very recently, however. Four studies released within the last three years have reported the direct evolution of miRNAs from TEs. Borchert et al were the first to identify Alu-derived miRNAs within the human genome [15]. They were quickly followed by Piriyapongsa et al who found Miniature Inverted Transposable Element (MITE)-derived miRNAs in both human [16] and plant genomes [17]. Finally, Lanier and co-authors reported TE derived miRNAs in the unicellular green alga *Chlamydomonas reinhardtii* [18]. This discovery lead to exciting new hypotheses regarding the evolution of miRNA-mediated gene regulatory networks from TEs [19]. The current model for the creation of TE-derived miRNA networks (Figure 1) suggests that TEs insert in one orientation in a transcriptionally active region of the genome to form the miRNA locus and in the opposite orientation within a gene to form the miRNA target. When the miRNA locus is transcribed, the smRNA is processed and directed into the RNAi machinery where it seeks complementary sequences within mRNA targets leading to post-transcriptional repression [19]. Because the same TE was inserted within the target gene, a complementary region is available and the target is repressed. The derivation of both the target and the locus from the same TE could explain why there is often a striking level of sequence similarity surrounding the miRNA and its target sequence in plants [14].

Figure 1. The evolution of smRNA networks from TEs. The process begins when (A) a TE is reactivated and transposes into both a transcriptionally active region of the genome to form the smRNA loci and a gene to form the future target. The insertions of the TE are in opposite orientations. (B) Post-transcriptional regulation occurs when both the smRNA loci and its target are transcribed. The smRNA locus is processed into a 20-24mer smRNA by a nuclease referred to as Dicer then incorporated into the RISC silencing complex. Here, it acts as a guide, and through standard Watson-Crick base pairing directs the silencing complex to complementary regions within other mRNA transcripts. Because the same TE inserted into the target gene in the opposite orientation, the gene has a region which is complementary to the smRNA and the RISC complex binds the target. From here, depending on the amount of complementarity between the smRNA and its target, it can either cleave the mRNA transcript leading to degradation or can remain bound to the transcript leading to translational repression.


A
Transposition
Target
Loci
B
Post-transcriptional Repression
Incorporation into RISC
Target repression
RISC
RISC
smRNA processing by Dicer
Transcription
Transcription
= Gene
= Intergenic region
= TE

Studying the exaptation of miRNAs and their target sequences from TEs can provide important insight into how complex miRNA regulatory networks are established, particularly those networks involved in response to biotic and abiotic stressors. miRNAs are critical for gene regulation under stress conditions. Oxidative stress is ameliorated in *Arabidopsis* by the downregulation of miR398 which relieves its repression of two Cu/Zn superoxide dismutases. These proteins can then scavenge the superoxide radicals formed during oxidative stress and prevent cellular damage [20]. Salt stress is mitigated by several miRNAs in maize [21]. Many other biotic and abiotic stresses, such as pathogen invasion, nutrient deprivation, heat, and cold are combated in part by miRNA-mediated gene regulation [22, 23]. Interestingly, TEs are upregulated by these same stressors in plants, animals, and fungi [3, 24-29]. Because of this, TE insertions which create miRNA loci and targets could potentially be increased in times of stress providing an organism with an additional way to adapt to their environment. It has long been suggested that TEs may play an adaptive role in abating "genomic stresses" experienced by their hosts [3, 5, 6, 26]. As discussed earlier, many instances of transpositions are known that resulted in beneficial mutations which help the host cope under conditions of stress. Because miRNAs are known to be derived from TEs, the observation that TEs are often derepressed specifically in cases of biotic and abiotic stress provides significant clues to the creation of miRNA-based stress response networks.

Prior to the research reported here, no one had tried to find a link between TEs, miRNAs, and adaptation to stress within a single study. I sought to determine whether exposure to nutrient stress would increase the likelihood of finding novel TE-derived smRNAs in the unicellular, model green alga *Chlamydomonas reinhardtii*. This was accomplished by growing *Chlamydomonas* cultures in a modified f/2 medium which utilized nitrate as its nitrogen source and did not provide acetate as an additional source of carbon. Previous studies of *Chlamydomonas* smRNAs have been conducted by growing strain CC-1690 in standard TAP medium which provides ammonium as its

nitrogen source and acetate as an additional carbon source [30, 31]. The nitrate/acetate-free conditions utilized in this study were more stressful for two reasons. First, ammonium is the preferred nitrogen source for photosynthetic eukaryotes because nitrate must be converted into ammonium through an additional energy-demanding two step reduction process prior to assimilation [32]. It is estimated that 20% of the reducing power generated by photosynthesis is used in this process [33]. The stress caused by the additional cost of reducing nitrate to ammonium was exacerbated by the second reason these conditions were more stressful, the lack of acetate as an additional carbon source. Once nitrate has been reduced to ammonium, the nitrogen is incorporated into organic compounds through the GS/GOGAT pathway. Only after this assimilation step can amino acids, nucleotides, and other nitrogenous compounds be synthesized. In order to complete the nitrogen assimilation process, the photosynthetic organisms must be able to provide sufficient carbon skeletons to incorporate all of the reduced nitrogen [34-36]. At atmospheric levels of CO_2 as were used in this study, *Chlamydomonas* has been shown to excrete nitrogenous compounds into the media because it lacks sufficient carbon resources to incorporate all of the reduced nitrate [33, 35, 37, 38]. Adding acetate to the medium as an additional source of carbon has been shown to mitigate this reduction in nitrogen assimilation capacity [39, 40]. The additional energy and carbon resources provided by acetate supports a more rapid growth rate and a healthier culture because the cells are able to synthesize the nitrogenous compounds needed for rapid reproduction. Because of the use of a nitrate/acetate-free medium in this study, our *Chlamydomonas* cultures experienced an enhanced level of nutrient stress than those previously used to study the smRNA pools in this alga.

It should be noted however that although the growth conditions I used were more stressful, they are experienced by natural populations. Because of this, *Chlamydomonas* can grow successfully under these conditions. Given this approach, my study was designed to detect the possible contributions that TEs make to “environmental”

adaptation in the alga through the creation of novel smRNA networks. The result of this work leads to two important findings: first, novel miRNAs are produced under the stress conditions utilized herein, and second, a surprising number of these novel miRNAs are derived from transposition events.

MATERIALS AND METHODS

Culturing Conditions

Chlamydomonas reinhartii strain CC-1690 was obtained from The *Chlamydomonas* Center, Duke University, Durham, NC. These cells were cultured in freshwater f/2 medium (Bigelow Labs, Boothbay Harbor, ME) containing CaCl2 and $MgSO_4$ supplementation (final concentration 0.1g/L and 0.05g/L, respectively.) This culture medium provided the nitrate/acetate-free stress environment discussed previously. The Ca/Mg supplementation was designed to make the f/2 medium identical to TAP with the exception of the presence of nitrate and lack of acetate. Cells were cultured in this modified f/2 medium at 20^{o}C and with a 12 hour light/dark cycle.

Small RNA Isolation

After two weeks in f/2 medium, the culture was poured into 50mL Falcon conical tubes. The tubes were centrifuged at 3220 x g for 3 minutes in an Eppendorf 5810R centrifuge to pellet the cells. The media was poured off and the cells were transferred to 1.5mL microcentrifuge tubes. A total of 1mL Trizol (Invitrogen, Carlsbad, CA) was added to each tube. The tube was rocked for 10 minutes to lyse the cells. Thereafter, 0.2mL of chloroform was added to each tube. The tubes were shaken vigorously for 15 seconds, and then allowed to rest for 3 minutes prior to centrifugation at 2600 x g for 45 minutes at 4^{o}C. After centrifugation, the clear aqueous layer was extracted and placed in a new 1.5mL tube and 1mL of isopropanol was added. The tubes were vortexed then allowed to settle for 10 minutes prior to a second centrifugation a 2600 x g for 45 minutes at 4^{o}C. The liquid was carefully poured from the tube, and the resulting pellet was washed by adding 1mL of 75% ethanol to the tube, vortexing, then centrifuging at 1600 x g for 15 minutes at 4^{o}C. The ethanol wash was poured from the tube and the pellet was allowed to air dry.

The dry pellet was resuspended in 25μL of RNase-free water and quantified using a Nanodrop machine. A total of 10μg of resuspended RNA was added to an equal volume of 2X formamide loading dye and 1μL of the miSpike™, the 21nt RNA size marker included with the miRCat™ Small RNA Cloning Kit (Integrated DNA Technologies, Coralville, IA), was also added to the RNA/loading buffer mixture. The RNA was run on a 12.5% denaturing polyacrylamide gel in parallel with a 10-100nt size ladder. The gel was allowed to run at 210 volts until the lower dye band was near the bottom of the gel (approximately 2 hours.) The RNA fragments in the gel were visualized with GelStar® (Lonza Biosciences, Basel, Switzerland) and the 18-24mer RNA band was excised from the gel by cutting slightly above and below the miSpike band. The gel slice was placed in a 1.5mL microcentrifuge tube, crushed, and 200μL of RNase-free water was added. The tube was heated to 70^{o}C for 10 minutes then vortexed. The contents of the tube were transferred to a prepared DTR column (Edge Biosystems, Gaithersburg, MD). The DTR column was centrifuged at 850 x g for three minutes, discarded, and 3μL 10mg/mL glycogen, 25μL of 3M NaOAc (pH 5.2) and 900μL of ice cold 100% EtOH was added to the flow through tube. The tube was placed in a -80^{o}C freezer for 30 minutes and then centrifuged at 16,000 x g for 10 minutes. The supernatant was poured off and the pellet air dried.

Cloning and Sequencing of smRNAs

A set of 5' and 3' cloning linkers were added to each small RNA according to the protocol provided in the miRCat™ Small RNA Cloning Kit Technical Manual. Briefly, an 18bp linker was ligated to the 3' end of the small RNAs, and then the pool was PAGE purified as described previously. The gel was visualized with GelStar® and the 36-42nt RNAs were recovered from the gel as described previously. A 22nt linker was then ligated to the 5' end of the small RNAs. The small RNA pool was reverse-transcribed using the RT primer included with the miRCat™ kit and the SuperScript™III Reverse

Transcriptase kit (Invitrogen, Carlsbad, CA.) The cDNA pool was PCR-amplified using the primers included with the miRCat™ kit and the Bioline PCR kit (London, UK.) The resulting PCR product was cloned using the pGEM® T-Easy Cloning kit (Promega, Madison, WI) per kit instructions. The transformed JM109 competent cells were grown on LB/50μg/mL ampicillin/IPTG/X-GAL plates to facilitate antibiotic and blue/white screening. After incubation at 37°C overnight, large white colonies were picked, added to 0.5mL thin walled PCR tubes with 20μL of RNase-free water, and boiled at 95°C for 5 minutes to lyse the cells. The PCR tubes were spun to collect cellular debris at the bottom of the tube, and then 6.8μL of the supernatant was added to a standard Sanger sequencing reaction mix using the M13 reverse sequence as the primer. The sequencing was performed on an ABI3700 (Applied Biosystems, Foster City, CA) at the Roy J. Carver Center for Comparative Genomics facility in the Department of Biology, University of Iowa.

Analysis of Sequencing Results

Sequence output files were screened for the flanking linker sequences and the small RNA insert was identified. Each insert was compared *via* BLAST searches against previously identified *Chlamydomonas* smRNAs, the NCBI database, and the *Chlamydomonas reinhardtii* genome sequence available from the Joint Genome Institute (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html [41].) Sequences that did not match a previously identified RNA sequence, including smRNAs, tRNAs, and rRNAs, and were found in the *Chlamydomonas* genome were set aside for further study (see below.)

The genome location of each novel sequence was first determined. The sequence and its flanking region were analyzed for miRNA-like hairpin structures using version 2.3 of the Rensselaer Polytechnic Institute RNA mFold server (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1-2.3.cgi) [42]. The fold temperature was set to 20°C. If the sequence was located within an intron, the entire intron was folded. If

the sequence was intergenic, a total of 200nt was folded. Sequences which folded into stable hairpins were designated miRNAs. Sequences which did not fold into hairpins were designated siRNAs. No sequences were designated as piRNAs due to the band size excised from the initial isolation gel (18-24mers) and because piRNAs are apparently limited to metazoan germ line cells.

SmRNA targets were identified through BLAST searches of the *Chlamydomonas reinhardtii* v4 Filtered or Best Transcript database (http://genome.jgi-psf.org/cgi-bin/runAlignment?db =Chlre4&advanced=1) using standard plant target prediction parameters: near perfect sequence complementarity is required along the entire length of the smRNA with no more than three mismatches allowed. Identification of TE-derived smRNAs was accomplished *via* BLAST searches of the *Chlamydomonas reinhardtii* Transposable Repeat Library available through JGI (http://genome.jgi-psf.org/Chlre4/Chlre4.download.ftp.html). In addition, 200nt of each smRNA locus was aligned with each target sequence to look for long regions of identity in order to identify putative transposition events.

Northern Blot Analysis

Northern blots are typically utilized to validate whether a smRNA is a true product of the cell or merely a byproduct of the degradation of a larger RNA. If the 18-24mer is a true product of the smRNA pathway, a band of the proper size should be detected on the blot. The small RNA Northern blot procedure developed by Lau et al was used for this purpose [43]. Briefly, total RNA was separated on a 12.5% denaturing polyacrylamide gel. The first lane of the gel consisted of a ^{32}P γ-ATP labeled 10-100nt size ladder, lanes 2-5 each contained 20μg of total RNA from different RNA preparations, and lane 6 consisted of 150fmole of a DNA oligo with the identical sequence as the smRNA to serve as a positive control. The blot was transferred to an Amersham Hybond™-N$^+$ nylon membrane (GE Healthcare, Chalfont St. Giles, UK)

using a semidry electrotransfer device. The transferred nucleic acids were UV crosslinked to the membrane. The membrane was baked for 30 minutes at 80^{o}C prior to hybridization.

A total of 20pmoles of each Northern probe was ^{32}P γ-ATP end labeled using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA). The membrane was prehybridized with 25mL hyb solution (5X SSC, 20mM NaH_2PO_4, 7% SDS, and 2X Denhardt's solution) and 1mg sheared salmon sperm DNA at 50^{o}C with constant rotation in a Lab-line Hybridization Incubator (Mumbai, India.) After 2 hours, the hyb solution was removed and 25mL of fresh solution was added along with 1mg sheared salmon sperm DNA and 10pmoles of the labeled Northern probe. The blot was hybridized overnight at 50^{o}C with constant rotation. Following hybridization, the blot was washed twice for 10 minutes and two more times for 30 minutes at 50^{o}C with 40mL of non-stringent wash buffer (3X SSC, 25mM NaH_2PO_4, 5% SDS, and 10X Denhardt's solution.) The membrane was washed one final time using 80mL of stringent wash buffer (1X SSC, 1% SDS), then wrapped in plastic. The blot was visualized by overnight exposure of Kodak BioMax XAR film.

RESULTS AND DISCUSSION

Identification and Classification of smRNAs

A total of 192 recombinant plasmids were sequenced in this study. By analyzing these sequences, I was able to identify 128 smRNA inserts. The remaining clones were either empty or contained linker-linker sequences. Small RNA inserts were compared against both known *Chlamydomonas* smRNA sequences, including rRNA and tRNA sequences, and the *Chlamydomonas* genome. The majority of these inserts corresponded to previously reported *Chlamydomonas* smRNAs, but 12 novel smRNAs were identified (Table 1.) The fact that these 12 sequences escaped detection in the previous two large scale 454 pyrosequencing analyses of the smRNA pools derived from TAP grown *Chlamydomonas* [30, 31] indicates that they are very likely produced in response to the non-standard experimental conditions used in my study.

The novel smRNA sequences were mapped to the *Chlamydomonas* genome. 9 of these sequences had only one perfect match within the genome, whereas the remaining appearing from 2-52 times. When the smRNAs and their flanking 200nt were analyzed using the mFold RNA secondary structure prediction software, 8 were capable of forming a stable hairpin (Figure 2.) The ability to form a stable hairpin is a characteristic of miRNAs, and these 8 sequences were designated as such (Table 1.) The inability of three sequences to form stable hairpins indicates that they are siRNAs. Due to ambiguous positions (Ns) in the flanking scaffold information for cre_smRNA7, accurate secondary structure prediction was not possible for this sequence. Because of this, it is not possible to determine whether this smRNA is an siRNA or an miRNA.

Total RNA pools were analyzed by Northern blot in order to determine whether the 12 novel smRNAs were stably produced by the cell. Stable expression increases the likelihood that the smRNAs are true products of the cell rather than simply byproducts of RNA degradation. 8 of the 12 smRNAs yielded bands of an appropriate size on the blot

(Figure 3.) Bands of the appropriate size were not detected for the remaining 4 sequences. The expression levels of these smRNAs could be below the minimum threshold required for detection with this method or alternatively, the sequences could be degradation products from longer transcripts.

As a result of the mFold and Northern analyses, I can confirm that six of the novel smRNAs identified in this study, cre_smRNAs 1-4, 6, and 12, are stably expressed miRNAs. miRNAs are exclusively associated with post-transcriptional gene regulation. That these six sequences are stably expressed under the conditions used in this study provides strong evidence that they are involved in stress-induced gene regulation. Cre_smRNA8 is a stably expressed siRNA. Whereas many siRNAs are associated with post-transcriptional gene silencing, others are responsible for DNA methylation, TE repression, and chromatin remodeling. Therefore I cannot be certain which process this smRNA participates in, although its stable expression is an indicator that it does have a functional role in the cell. Whereas it is not certain whether cre_smRNA 7 is an miRNA or an siRNA, it is clearly a stably expressed smRNA. I cannot confirm whether the remaining four sequences, cre_smRNAs 5, 9, 10, or 11, are functional smRNAs or RNA degradation products without additional Northern analyses. In short, I have confirmed the presence of 8 novel smRNAs produced in response to nitrate/acetate-free stress. The next questions I address are what genes do they regulate, and are they TE derived?

Table 1. Sequences and Genomic Positions of Novel Small RNAs

Sequence Name	Sequence	SmRNA Type	Chromosome	Start Position	Stop Position
Cre_smRNA1	AATGCTCGAAATTTCCGCGCCG	miRNA	13	758621	758600
Cre_smRNA2	TGTCTTCAGCCCGCCAATCTTCT	miRNA	10	4578921	4578943
Cre_smRNA3	ATGGGTCCGAACGGGAAGCTTTAT	miRNA	3	2087827	2087804
			3	2100944	2100967
			3	2101951	2101974
Cre_smRNA4	ATGGGTCCGACCGGGAAGCTTTAT	miRNA	3	2087346	2087323
			3	2088469	2088446
			3	2100475	2100498
			3	2102872	2102895
			3	2103780	2103757
			2	6304403	6304380
Cre_smRNA5	CACATAGCCTGGTGCACTC	miRNA	3	5381395	5381413
Cre_smRNA6	TAGAGCTCGAAGAACTTGGGAA	miRNA	6	5471000	5471021
Cre_smRNA7	TCCGGGAGGTGGTCAAAGGGC	unknown	13	6205802	6205782
Cre_smRNA8	TTGTTTACTGCCGTGTCCTGC	siRNA	15	116080	116100
Cre_smRNA9	TGGGTGGGTGCATGGGTCCGA	siRNA	13	525912	525932
Cre_smRNA10	CTCATGGTGTAGTTGGTTAT	miRNA	17	519283	519302
			17	521194	521213
Cre_smRNA11	TTGGTCGCTATCAACCTGACA	siRNA	15	116075	116055
Cre_smRNA12	AGCGCGCACCGGTGTCTCGTCA	miRNA	1	4397800	4397780
			1	4398822	4398802
			1	4492492	4492512
			1	4493514	4493534
			1	8138459	8138439
			1	9700416	9700436
			1	9949903	9949883
			2	7791234	7791254
			2	7952382	7952362
			2	8150202	8150182
			3	2395364	2395344
			3	6317391	6317411
			3	6318413	6318433
			6	580010	580030
			6	4057012	4056992
			6	4335242	4335262
			6	4336264	4336284
			6	6510135	6510155
			7	46151	46131
			7	47173	47153
			7	3101215	3101195
			8	1545802	1545782
			8	1546824	1546804
			8	1547610	1547590
			8	3850696	3850716
			8	4004043	4004063
			8	4085177	4085197
			9	585207	585227
			9	586229	586249
			9	3609384	3609364
			10	2327912	2327932
			12	1892354	1892374

Table 1 Continued

Sequence Name	Sequence	SmRNA Type	Chromosome	Start Position	Stop Position
			12	2527359	2527379
			12	4710123	4710103
			12	6062612	6062592
			12	8296076	8296056
			12	8297098	8297078
			13	2492453	2492473
			15	991705	991685
			16	89839	89859
			16	5334531	5334511
			16	5900028	5900008
			17	3588134	3588154
			17	4398482	4398462
			scaffold_18	145691	145711
			scaffold_18	390033	390013
			scaffold_21	271345	271325
			scaffold_22	125328	125348
			scaffold_36	97194	97214
			scaffold_39	41530	41550
			scaffold_79	10388	10408
			scaffold_79	11410	11430

Figure 2. mFold results for each smRNA. The purpose of this analysis was to determine whether the region surrounding the smRNA could form a hairpin from which an miRNA could be processed. The hairpin must fold in a way which results in the smRNA sequence (in green) forming a duplex with the other arm of the same hairpin. 8 sequences (cre_smRNAs 1-6, 10 and 12) fold into such a stable hairpin. This indicates that they are miRNAs. While the loci of 3 additional sequences (cre_smRNA 8, 9 and 11) are able to produce some hairpin structures, the smRNA does not form a duplex with the other arm of the same hairpin which would make Dicer processing into a functional miRNA impossible. Therefore these sequences are siRNAs. Cre_smRNA7 could not be folded due to the presence of ambiguous positions (Ns) in the flanking genome region so it is not possible to determine whether it is an siRNA or an miRNA. Cre_smRNAs 8 and 11 are included in the same fold because they are located only 5nts away from one another within the genome, another indication that they are siRNAs processed from the same long dsRNA precursor.

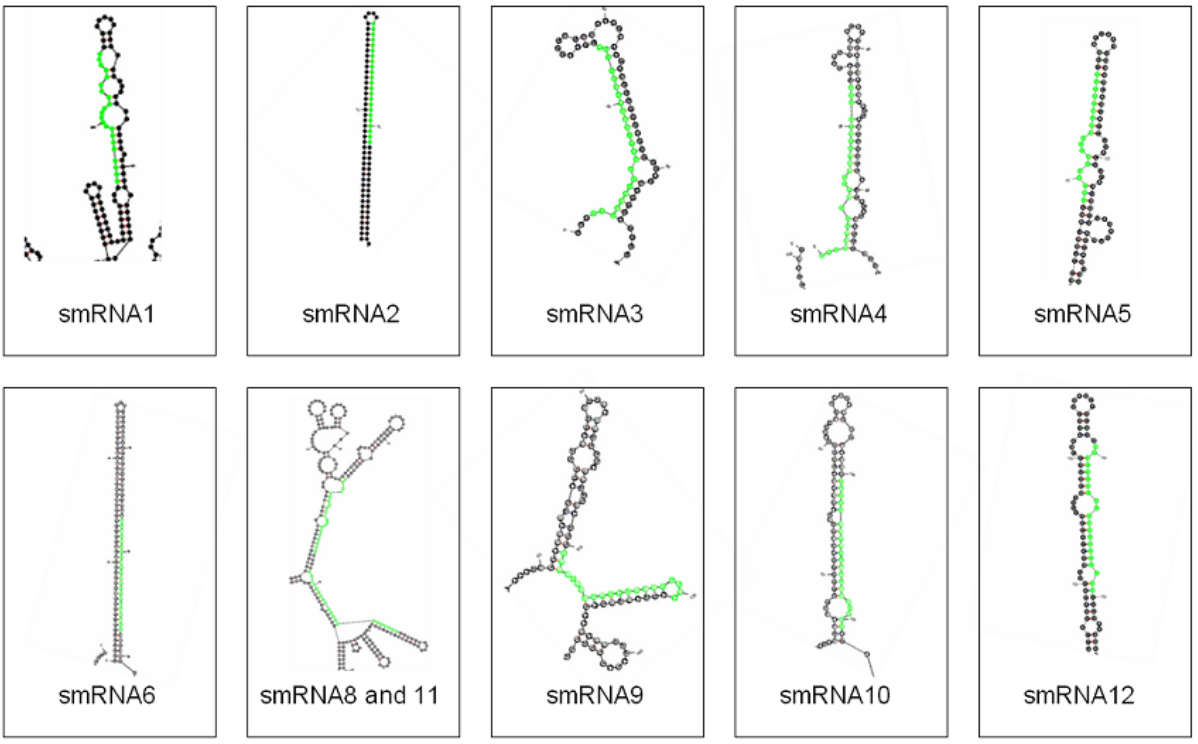
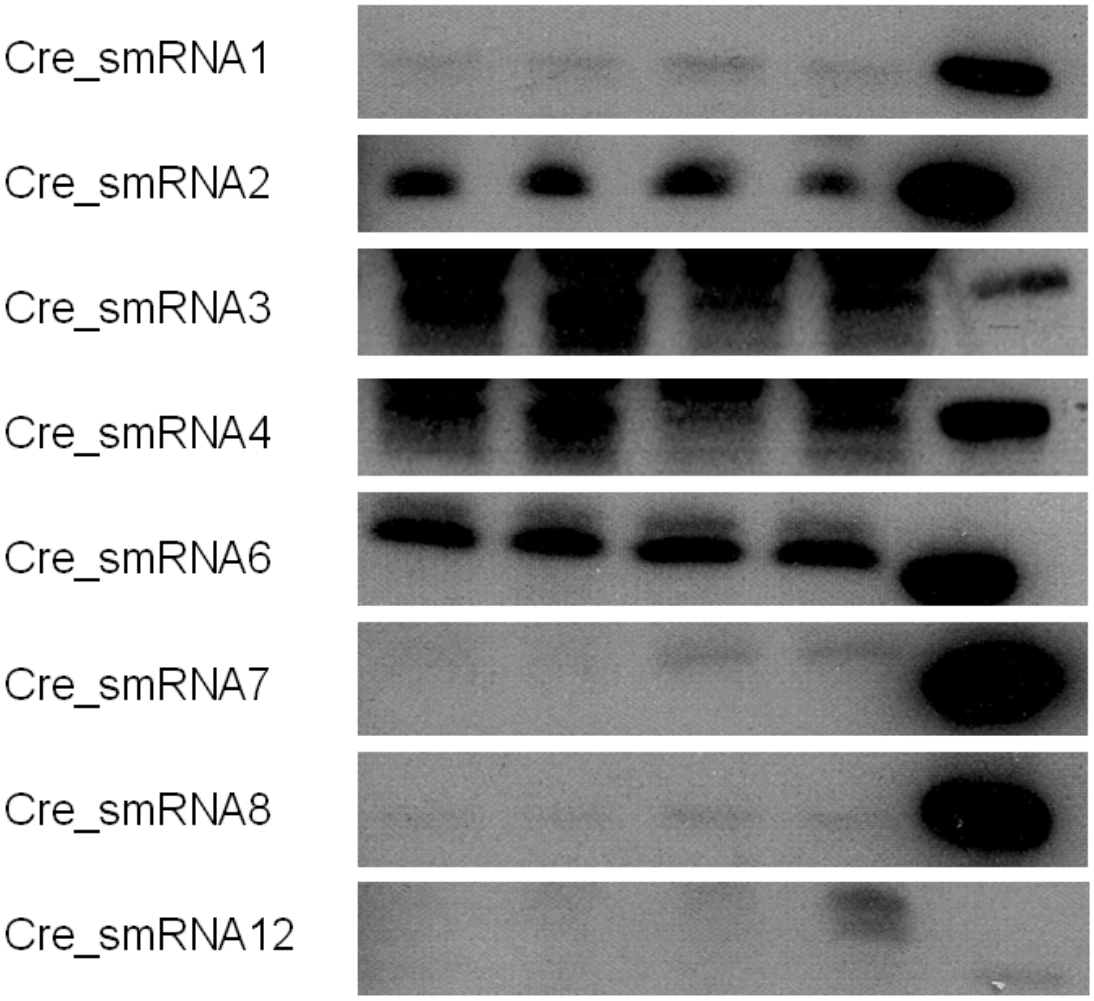

smRNA1
smRNA2
smRNA3
smRNA4
smRNA5
smRNA6
smRNA8 and 11
smRNA9
smRNA10
smRNA12

Figure 3. Positive Northern blot results for 8 smRNAs. The first 4 lanes consist of total RNA isolated from different cell preps grown under my experimental conditions. The 5th lane consists of a DNA oligonucleotide of the same sequence as the smRNA which acted as a positive control. These 8 smRNAs were detected using my approach. The remaining four were not detected by Northern analysis indicating they could either be expressed at extremely low levels or alternatively, are degradation products of larger RNAs.


Cre_smRNA1
Cre_smRNA2
Cre_smRNA3
Cre_smRNA4
Cre_smRNA6
Cre_smRNA7
Cre_smRNA8
Cre_smRNA12

Target Prediction Results

Target prediction was accomplished through BLAST comparison of the 12 smRNAs with the known *Chlamydomonas* transcriptome data to identify near-perfect antisense hits. Potential targets were identified for 11 of the 12 smRNAs (Table 2). A near-perfect antisense hit was not identified for cre-smRNA2. Of the 18 predicted targets identified, 11 are of unknown function. Two predicted targets have a role in transcription. Cre_smRNA5 targets a SMRT/SMRTER nuclear receptor coregulator. This protein interacts with other proteins to repress the transcription of multiple genes simultaneously [44]. Cre_smRNAs 3 and 4 both target an SBP transcription factor. Transcription factors and transcriptional regulators are common targets of miRNAs. In fact, one of the most highly conserved plant miRNAs, miR156, also targets an SBP transcription factor [45]. Other targets include a serine/threonine protein kinase and a mitochondrial substrate carrier responsible for transporting phosphate into the mitochondrial matrix.

The 3 remaining targets are the most intriguing because they each have a role in carbon/nitrogen metabolism (Figure 4). The experimental conditions used in this study changed the available nitrogen source to nitrate and removed acetate as an additional carbon source which would impact carbon/nitrogen metabolism in the cultured cells. The first predicted target, the Hybrid Cluster Protein (HCP), is a target of cre_smRNA12. During nitrate reduction, harmful reactive nitrogen species such as nitrous oxide and hydroxylamine are produced as byproducts [46]. HCP is believed to reduce these harmful byproducts and prevent oxidative damage [47, 48]. HCP is known to be upregulated by nitrite, an intermediate in nitrate reduction [49]. The next predicted target, the cytosolic form of the small subunit of carbamoyl-phosphate synthase (CPSII), is targeted by cre_smRNA7. CPSII catalyzes the first committed step in pyrimidine biosynthesis from glutamine, an important step in the production of nucleotides in the cell and a major part of nitrogen metabolism [50]. The final predicted target, COQ5D, is targeted by cre_smRNAs 3 and 4. This protein catalyzes one of the steps in quinone

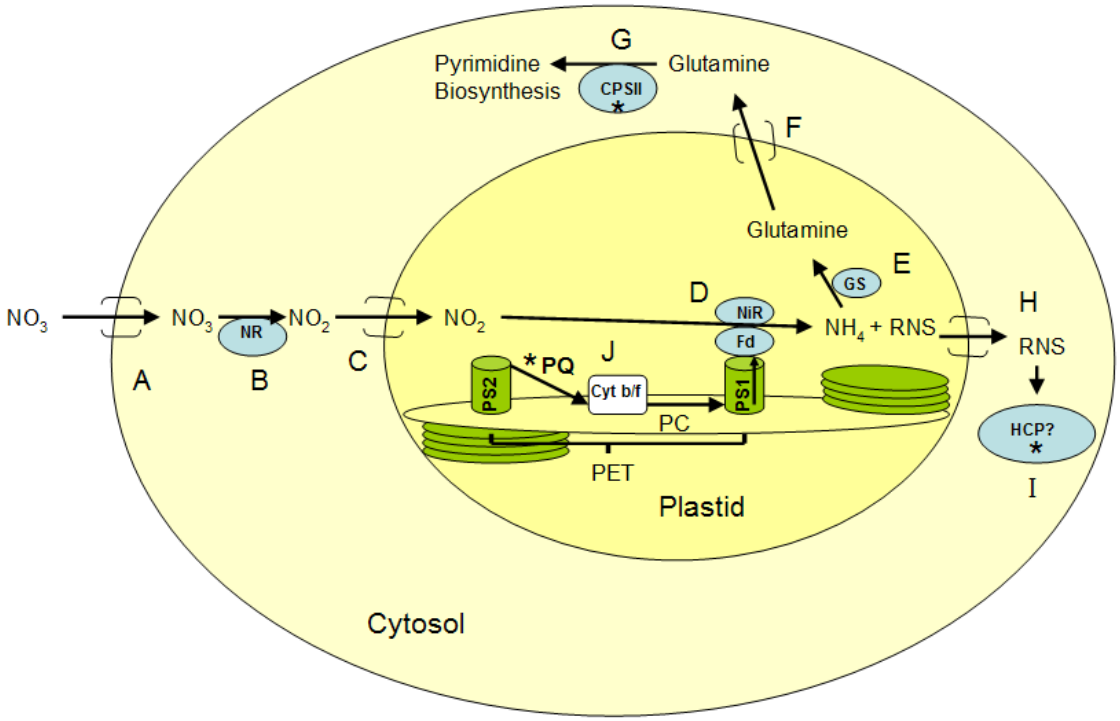
biosynthesis [51]. COQ5D has a plastid targeting signal indicating that it is specifically involved in plastid quinone biosynthesis. Plastid quinones form part of the photosynthetic electron transport (PET) chain which is involved in photosynthetic carbon fixation. Under the low CO_2 conditions which were used in this study, PSII, another component of the PET, can become overexcited and damaged as a result [52]. Plastoquinone (PQ,) one of the plastid quinones, shuttles electrons from PSII to the cytochrome b/f complex in the PET. It is possible that the decreased production of PQ is simply designed to reduce the drain on cellular resources in a nutrient-limiting environment by preventing the production of more PQ than is needed. The same can be said for CPSII and HCP repression. As discussed previously, the level of nitrate assimilation in *Chlamydomonas* is directly related to the available supply of carbon skeletons. In a carbon limited environment such as the one used in this study, the cell must prioritize its use of resources such as amino acids due to the limited influx of nitrogenous compounds. Some genes, such as HCP, are upregulated in response to nitrate, but may not be needed at a normal level due to carbon limitation. By fine-tuning the amount of protein through miRNA-mediated downregulation, the cell can partition its resources more effectively. The miRNA mediated downregulation of HCP, CPSII, and COQ5D could serve some additional adaptive purpose, as well, but more study will be required before this can be said with any certainty.

In conclusion, I now know that four of the stably expressed smRNAs induced by our nitrate/acetate-free conditions, cre_smRNAs 3, 4, 7, and 12, target genes known to be involved in carbon/nitrogen metabolism, another strong indicator that they fulfill a stress-mitigating role in the cell. But are these smRNAs TE derived?

Table 2. Small RNA Target Predictions

Sequence Name	Target Protein ID Number	Target Name	Target Function
Cre_smRNA1	155032	Unknown Protein	Unknown
Cre_smRNA2		(no predicated target)	
Cre_smRNA3	417241	SBP	Possible transcription factor
	189872	COQ5D	Ubiquinone/menaquinone biosynthesis
Cre_smRNA4	417241	SBP	Possible transcription factor
	189872	COQ5D	Ubiquinone/menaquinone biosynthesis
Cre_smRNA5	146314	Nuclear receptor coregulator SMRT/SMRTER	Transcription
Cre_smRNA6	146327	Unknown Protein	Unknown
Cre_smRNA7	401382	Carbamoyl-phosphate synthase, small subunit	Nitrogen metabolism
	149657	Unknown Protein	Unknown
	420583	Unknown Protein	Unknown
Cre_smRNA8	382065	Unknown Protein	Unknown
	194753	Hypothetical Protein	Unknown
Cre_smRNA9	395014	Unknown Protein	Unknown
	413366	Unknown Protein	Unknown
	394431	Unknown Protein	Unknown
Cre_smRNA10	184238	Unknown Protein	Unknown
Cre_smRNA11	196712	Mitochondrial substrate carrier	Phosphate transport
Cre_smRNA12	399318	HCP (Prismane)	Electron transport, nitrate/nitrite metabolism
	404883	Unknown Protein	Unknown
	406072	Serine/threonine protein kinase	Serine/threonine protein kinase

Figure 4. The nitrate assimilation process. The assimilation of nitrate begins when (A) nitrate is imported into the cell where it (B) is quickly reduced into nitrite by nitrate reductase (NR.) Nitrite is (C) imported into the chloroplast and (D) further reduced to ammonium by nitrite reductase (NiR) and ferredoxin (Fd). The ammonium is (E) incorporated into glutamine by glutamine synthase (GS.) The glutamine can participate in several possible reactions leading to the formation of nitrogenous compounds such as amino acids and nucleotides in the cell. Some of the glutamine is (F) exported into the cytosol where it (G) can become part of the pyrimidine biosynthesis pathway. Carbamoyl-phosphate synthase II (CPSII,) the target of cre_smRNA7, catalyzes the first committed step in this process. During the nitrite reduction step (D), reactive nitrogen species (RNS) are produced as byproducts. These must be eliminated from the cell in order to prevent oxidative damage. (H) The RNS are exported to cytosol where (I) HCP, the target of cre_smRNA12, is believed to convert these harmful byproducts into less toxic forms for elimination from the cell. The reducing power for nitrate assimilation comes from the photosynthetic electron transport chain (PET.) One component of the PET, Ferredoxin (Fd,) is directly involved in the reduction of nitrite to ammonium. Another important part of the PET is (J) plastoquinone (PQ.) Cre_smRNAs 3 and 4 target COQ5D, a protein which catalyzes a step in PQ biosynthesis. The three proteins targeted by novel smRNAs identified in this study, CPSII, HCP, and PQ, a product of COQ5D, are denoted by an asterisk in this figure.


G
Pyrimidine Biosynthesis
CPSII *
Glutamine
F
Glutamine
E
GS
D
NiR
Fd
NO_3
NO_3
NR
NO_2
NO_2
NH_4 + RNS
H
RNS
A
B
C
* PQ
J
PS2
Cyt b/f
PS1
PC
PET
HCP? *
I
Plastid
Cytosol

Identification of the Possible TE Origins of smRNA Loci and Targets

The possible TE origins of the novel smRNAs identified in this study were investigated through both BLAST comparisons with known *Chlamydomonas* TEs and alignment of the smRNA loci with their target(s) to look for long regions of identity. Two of the stably expressed miRNAs, Cre_smRNA3 and Cre_smRNA4, differ from one another by a single central mismatch and collectively appear nine times within a single 16,000bp region on chromosome 3. These two sequences differ from a *Chlamydomonas* miRNA identified by Molnar et al, cre-miR1146, by the same central mismatch [30]. The two cre-miR1146 loci are located on chromosome 1. Comparison of the cre_smRNA3, cre_smRNA4 and cre-miR1146 flanking regions with known transposable elements indicates that they are derived from the same TE [18]. An alignment of a predicted target, an SBP transcription factor, with the smRNA loci indicates that this same transposable element has inserted itself into this gene in the opposite orientation creating a target for the smRNA (Figure 5A.)

Cre_smRNA12 appears 52 times within the genome. BLAST comparison of this sequence with known TEs indicates that it is derived from the MRC1 retrotransposon (Figure 5B.) This retrotransposon has inserted into multiple locations within the genome, including intergenic regions, intronic regions, and the coding regions of three genes. An alignment of these three genes (HCP, a serine/threonine protein kinase, and a gene of unknown function) with the MRC1 transposon and cre_smRNA12 confirms that the insertion of this TE into these genes in the opposite orientation to the smRNA has created the smRNA target.

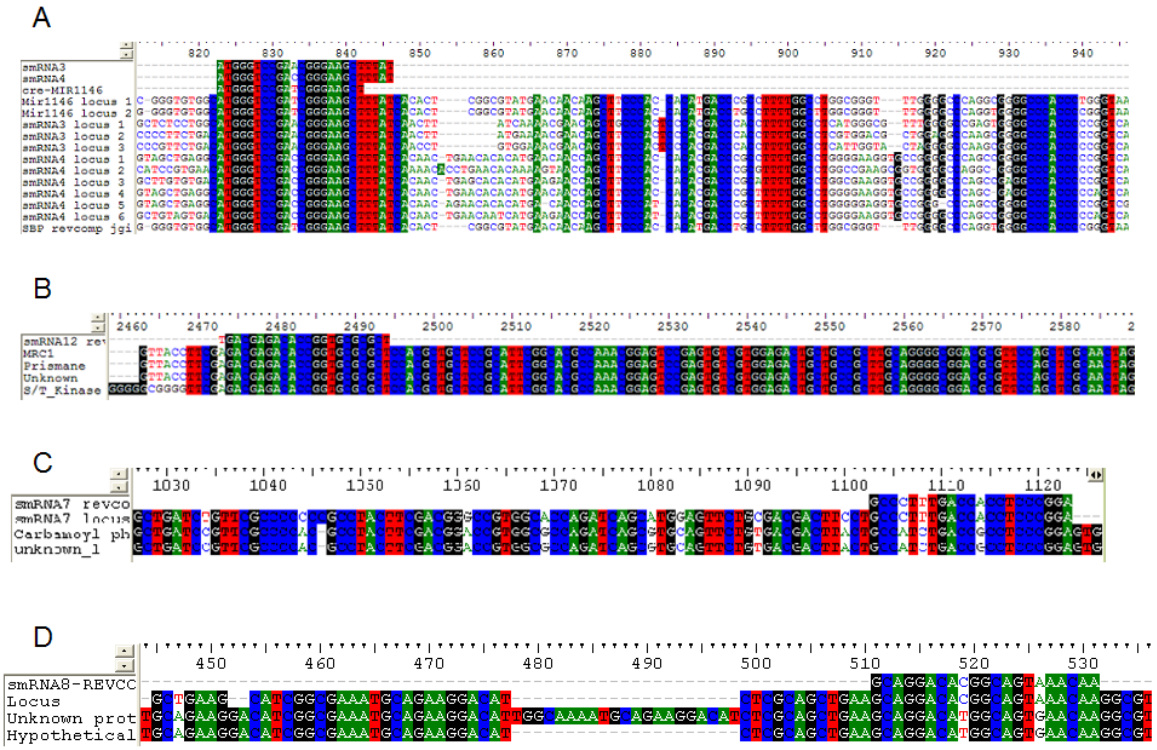
Two sequences, cre-smRNA_8 and cre-smRNA_11, are located only five nucleotides away from each other on chromosome 15. This, in addition to their inability to fold into a stable hairpin, indicates that they could be processed from the same long

siRNA precursor. When 200 flanking bases surrounding these two small RNAs are aligned with the two predicted targets of Cre-smRNA_8, a large inverted region of nearly identical sequence is found. This indicates that a transposition event created the target sequence needed for the RNAi-based repression of these two genes (Figure 5C.)

This same pattern of large inverted duplications becoming reinserted into an active gene and leading to the creation of smRNA targets was found for 5 of the 12 novel smRNAs: Cre_smRNAs 3, 4, 7, 8, and 12 (Figure 5). Two predicted targets with reverse-orientation insertions are HCP and CPSII, two of the genes with known roles in nitrogen metabolism. Although it was not possible to identify the progenitor TE in all cases, a transposition event is the most likely explanation for the repeated copying and pasting of long sequences into new locations within the genome from 3-52 times.

Using Northern, mFold, and BLAST analyses of the novel smRNAs isolated from my study conditions, I can conclude that 4 of the 12 smRNAs, cre_smRNAs 3, 4, 7, and 12, are stably expressed, have targets that are involved in carbon/nitrogen metabolism, and have both loci and targets derived from transposition events. This was a surprisingly high number given the relative scale of this project. Previous studies of the *Chlamydomonas* smRNA transcriptome derived from non-stress conditions reported a smaller than expected number of repeat-derived smRNAs. For example, Zhao et al found that only 6.11% of the smRNAs identified in their survey mapped to TEs or repetitive elements [31]. That we found 4 additional TE-derived smRNAs among the 12 novel stress-induced smRNAs identified by this study could indicate that TEs do make a large contribution to stress-induced miRNA networks. A large scale 454 pyrosequencing effort of the *Chlamydomonas* smRNA transcriptome derived from nitrate/acetate-free conditions could determine whether this trend continues throughout this smRNA population.

Figure 5. Sequence alignments comparing the smRNA, its locus and its target(s). (A) shows the alignment of cre_smRNA 3 and 4, their 9 loci, the cre_miR1146 loci, and the reverse complement of the SBP gene. Cre_smRNA 3 and 4 were previously identified as being TE derived. Strong similarity is found among these sequences providing strong evidence that the same TE inserted into each of these locations. The insertion of the TE in the opposite orientation in the SBP gene created the target for the smRNAs. (B) This same pattern of reverse orientation TE insertion is found for cre_smRNA12. This alignment shows that cre_smRNA12, the MRC1 transposon, and the reverse complements of HCP, a serine threonine protein kinase, and a gene of unknown function are identical in the region shown. This insertion of the MRC1 transposon in the reverse orientation into these three genes created the targets for cre_smRNA12. Cre_smRNA7 (C) and cre_smRNA8 (D) were also aligned with the reverse complements of their targets and show the same pattern of a reverse-orientation insertion within the target genes creating targets for the smRNAs.


A
B
C
D
smRNA3
smRNA4
cre-MIR1146
Mir1146 locus 1
Mir1146 locus 2
smRNA3 locus 1
smRNA3 locus 2
smRNA3 locus 3
smRNA4 locus 1
smRNA4 locus 2
smRNA4 locus 3
smRNA4 locus 4
smRNA4 locus 5
smRNA4 locus 6
SBP revcomp jgi
smRNA12 re
MRC1
Prismane
Unknown
S/T_Kinase
smRNA7 revco
smRNA7 locus
Carbamoyl ph
unknown_1
smRNA8-REVCC
Locus
Unknown prot
Hypothetical

CONCLUSION

It is widely accepted that an organism must fine-tune its genetic output in response to a changing environment. Over the last 15 years, there has been a growing appreciation of the role that transposition plays in creating genetic changes that an organism may exploit to adapt to new environments. As discussed earlier, TEs are known to make both subtle and profound changes in the genome of an organism, ranging from the disruption or creation of an individual gene to large-scale chromosomal rearrangements. Complex mechanisms exist to control transposition under non-stressed conditions in order to prevent disruptive, unwanted genomic changes. However, it has been noted repeatedly that normally silent transposons become active under conditions of stress [25, 26]. Whether or not this is simply the attempt by a parasitic element to save itself regardless of a possible cost to its host is not of critical importance. That this reactivation leads to mutations that selection can act upon is of high interest, however. Stress induced activation of TEs may provide a source of genetic variation in organisms that can aid in adaptation to a changing environment.

In this study, *Chlamydomonas reinhardtii* was cultured under conditions of nitrate/low carbon stress in order to find examples of TE exaptation that contributed to stress-responsive smRNA networks. Two genes, CPSII and HCP, both of which play a role in nitrogen metabolism, were found to have TE insertions in the 3' termini of their coding regions that formed a target site for smRNA-mediated repression. The repressive smRNA for both of these genes was cloned and its expression was confirmed by Northern analysis indicating that the TE also created a functional smRNA. The novel smRNAs were linked to the experimental conditions used here indicating that they were indeed environmentally induced. Based on these findings, I propose that TE exaptation to form stress-induced smRNA networks has occurred and could provide another avenue whereby TEs aid in the ability of an organism to adapt to changing environments. Based

on this observation, I propose the following modified version of the existing model describing the evolution of smRNA networks from TEs:

> *I propose that under non-stressed conditions, the host organism represses transposition in order to prevent unwanted mutations. When a biotic or abiotic stress is introduced, the repression of transposition is lifted. TE insertions occur at random throughout the genome, although there is evidence that some TEs insert preferentially in gene-rich areas [53]. In some cases, the TE inserts into the coding region or UTR or a gene. In many cases, this leads to disruption of the gene and the insertion creates a null mutation. This mutation may be detrimental to the host and would be removed from the population. However, in some cases the mutation does not disrupt the gene's function and a smRNA target results. At the same time, the TE inserts into a region of the genome that is transcriptionally active at the time of the stress, such as immediately downstream of a gene or in an intron. This creates the smRNA locus. Upon transcription, the TE-derived smRNA is able to repress the target gene because the TE insertion has created a complementary region within the target. Eventually the stress is relieved and the TE is repressed once again, but the smRNA and its target have been formed. Selection and drift can now act on this new network to determine its ultimate fate in the population. Due to the very low rates of successful transposition, these networks will not develop frequently, but as shown by this study and others, these networks are in fact fixed in populations [15-18].*

Barbara McClintock was the first to refer to TEs as "controlling elements" [54]. She recognized early on the potential that these elements had in controlling gene expression and allowing genomes to adapt to challenges presented by the environment [5]. My study presents yet another way in which TEs control gene expression in response

to environmental challenges. The study of the stress-induced evolution of small RNA networks from TEs should provide intriguing insights into how organisms adapt to their environments.

REFERENCES

1. Bowen NJ, Jordan IK: **Transposable elements and the evolution of eukaryotic complexity**. *Current issues in molecular biology* 2002, **4**(3):65-76.

2. Brookfield JF: **Evolutionary genetics: Mobile DNAs as sources of adaptive change?** *Curr Biol* 2004, **14**(9):R344-345.

3. Kidwell MG, Lisch DR: **Perspective: transposable elements, parasitic DNA, and genome evolution**. *Evolution; international journal of organic evolution* 2001, **55**(1):1-24.

4. Feschotte C, Jiang N, Wessler SR: **Plant transposable elements: where genetics meets genomics**. *Nat Rev Genet* 2002, **3**(5):329-341.

5. McClintock B: **The significance of responses of the genome to challenge**. *Science (New York, NY* 1984, **226**(4676):792-801.

6. Miller WJ, McDonald JF, Pinsker W: **Molecular domestication of mobile elements**. *Genetica* 1997, **100**(1-3):261-270.

7. Bowen NJ, Jordan IK: **Exaptation of protein coding sequences from transposable elements**. *Genome dynamics* 2007, **3**:147-162.

8. Kim KS, Kustu S, Inwood W: **Natural history of transposition in the green alga Chlamydomonas reinhardtii: use of the AMT4 locus as an experimental system**. *Genetics* 2006, **173**(4):2005-2019.

9. Chen C, Ara T, Gautheret D: **Using Alu elements as polyadenylation sites: A case of retroposon exaptation**. *Molecular biology and evolution* 2009, **26**(2):327-334.

10. Yoshii M, Shimizu T, Yamazaki M, Higashi T, Miyao A, Hirochika H, Omura T: **Disruption of a novel gene for a NAC-domain protein in rice confers resistance to Rice dwarf virus**. *Plant J* 2009, **57**(4):615-625.

11. Cerutti H: **RNA interference: traveling in the cell and gaining functions?** *Trends Genet* 2003, **19**(1):39-46.

12. Cerutti H, Casas-Mollano JA: **On the origin and functions of RNA-mediated silencing: from protists to man**. *Current genetics* 2006, **50**(2):81-99.

13. Axtell MJ, Bowman JL: **Evolution of plant microRNAs and their targets**. *Trends in plant science* 2008, **13**(7):343-349.

14. Chapman EJ, Carrington JC: **Specialization and evolution of endogenous small RNA pathways**. *Nat Rev Genet* 2007, **8**(11):884-896.

15. Borchert GM, Lanier W, Davidson BL: **RNA polymerase III transcribes human microRNAs**. *Nature structural & molecular biology* 2006, **13**(12):1097-1101.

16. Piriyapongsa J, Marino-Ramirez L, Jordan IK: **Origin and evolution of human microRNAs from transposable elements**. *Genetics* 2007, **176**(2):1323-1337.

17. Piriyapongsa J, Jordan IK: **Dual coding of siRNAs and miRNAs by plant transposable elements**. *RNA (New York, NY* 2008, **14**(5):814-821.

18. Lanier W, Tyra HM, Devor EJ, Borchert G, Comeron J, Bhattacharya D: **From So Simple A Beginning: The evolution of miRNAs from Transposable Elements in *Chlamydomonas reinhardtii***. *In review* 2009.

19. Feschotte C: **Transposable elements and the evolution of regulatory networks**. *Nat Rev Genet* 2008, **9**(5):397-405.

20. Jagadeeswaran G, Saini A, Sunkar R: **Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis**. *Planta* 2009.

21. Ding D, Zhang L, Wang H, Liu Z, Zhang Z, Zheng Y: **Differential expression of miRNAs in response to salt stress in maize roots**. *Annals of botany* 2009, **103**(1):29-38.

22. Sunkar R, Chinnusamy V, Zhu J, Zhu JK: **Small RNAs as big players in plant abiotic stress responses and nutrient deprivation**. *Trends in plant science* 2007, **12**(7):301-309.

23. Zhou X, Wang G, Sutoh K, Zhu JK, Zhang W: **Identification of cold-inducible microRNAs in plants by transcriptome analysis**. *Biochimica et biophysica acta* 2008, **1779**(11):780-788.

24. Bouvet GF, Jacobi V, Plourde KV, Bernier L: **Stress-induced mobility of OPHIO1 and OPHIO2, DNA transposons of the Dutch elm disease fungi**. *Fungal Genetics and Biology* 2008, **45**(4):565-578.

25. Daboussi MJ, Capy P: **Transposable elements in filamentous fungi**. *Annual review of microbiology* 2003, **57**:275-299.

26. Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM: **The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses**. *Genetica* 1997, **100**(1-3):241-252.

27. Ikeda K, Nakayashiki H, Takagi M, Tosa Y, Mayama S: **Heat shock, copper sulfate and oxidative stress activate the retrotransposon MAGGY resident in the plant pathogenic fungus Magnaporthe grisea**. *Mol Genet Genomics* 2001, **266**(2):318-325.

28. McGraw JE, Brookfield JF: **The interaction between mobile DNAs and their hosts in a fluctuating environment**. *Journal of theoretical biology* 2006, **243**(1):13-23.

29. Perez-Hormaeche J, Potet F, Beauclair L, Le Masson I, Courtial B, Bouche N, Lucas H: **Invasion of the Arabidopsis genome by the tobacco retrotransposon Tnt1 is controlled by reversible transcriptional gene silencing**. *Plant physiology* 2008, **147**(3):1264-1278.

30. Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC: **miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii**. *Nature* 2007, **447**(7148):1126-1129.

31. Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang XJ, Qi Y: **A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii**. *Genes & development* 2007, **21**(10):1190-1203.

32. Fernandez E, Galvan A: **Nitrate assimilation in Chlamydomonas**. *Eukaryotic cell* 2008, **7**(4):555-559.

33. Azuara MP, Aparicio PJ: **In Vivo Blue-Light Activation of Chlamydomonas reinhardii Nitrate Reductase**. *Plant physiology* 1983, **71**(2):286-290.

34. Aguera E, Ruano D, Cabello P, de la Haba P: **Impact of atmospheric CO2 on growth, photosynthesis and nitrogen metabolism in cucumber (Cucumis sativus L.) plants**. *Journal of plant physiology* 2006, **163**(8):809-817.

35. Rexach J, Fernandez E, Galvan A: **The Chlamydomonas reinhardtii Nar1 gene encodes a chloroplast membrane protein involved in nitrite transport**. *The Plant cell* 2000, **12**(8):1441-1453.

36. Thacker A, Syrett PJ: **Disappearance of nitrate reductase activity from *Chlamydomonas reinhardtii***. *New Phytology* 1972, **71**:435-441.

37. Eriksson M, Villand P, Gardestrom P, Samuelsson G: **Induction and Regulation of Expression of a Low-CO2-Induced Mitochondrial Carbonic Anhydrase in Chlamydomonas reinhardtii**. *Plant physiology* 1998, **116**(2):637-641.

38. Gordillo FJ, Figueroa FL, Niell FX: **Photon- and carbon-use efficiency in Ulva rigida at different CO2 and N levels**. *Planta* 2003, **218**(2):315-322.

39. Quesada A, la G-G, Fernandez E: **Involvement of chloroplast and mitochondria redox valves in nitrate assimilation**. *Trends in plant science* 2000, **5**(11):463-464.

40. Sahu J, Adhikary SP: **Heterotrophic growth and nitrogen fixation in the filamentous blue-green alga Anabaena sp**. *Zeitschrift fur allgemeine Mikrobiologie* 1981, **21**(9):669-676.

41. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L *et al*: **The Chlamydomonas genome reveals the evolution of key animal and plant functions**. *Science (New York, NY* 2007, **318**(5848):245-250.

42. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction**. *Nucleic acids research* 2003, **31**(13):3406-3415.

43. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans**. *Science (New York, NY* 2001, **294**(5543):858-862.

44. Pile LA, Schlag EM, Wassarman DA: **The SIN3/RPD3 deacetylase complex is essential for G(2) phase cell cycle progression and regulation of SMRTER corepressor levels**. *Molecular and cellular biology* 2002, **22**(14):4965-4976.

45. Xie K, Wu C, Xiong L: **Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice**. *Plant physiology* 2006, **142**(1):280-293.

46. Kuznetsova S, Knaff DB, Hirasawa M, Lagoutte B, Setif P: **Mechanism of spinach chloroplast ferredoxin-dependent nitrite reductase: spectroscopic evidence for intermediate states**. *Biochemistry* 2004, **43**(2):510-517.

47. Filenko N, Spiro S, Browning DF, Squire D, Overton TW, Cole J, Constantinidou C: **The NsrR regulon of Escherichia coli K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase**. *Journal of bacteriology* 2007, **189**(12):4410-4417.

48. Singh SP, Kesavan PC: **Post-irradiation modification of O2, N2, and N2O-mediated damage in dry barley seeds by glutathione and cysteine: probable radiation chemical events**. *International journal of radiation biology* 1993, **63**(4):483-491.

49. Filenko NA, Browning DF, Cole JA: **Transcriptional regulation of a hybrid cluster (prismane) protein**. *Biochemical Society transactions* 2005, **33**(Pt 1):195-197.

50. Bagnara AS, Chansiri K: **Sequences upstream and downstream from the glutamine-dependent carbamoyl phosphate synthetase-encoding gene from the protozoan *Babesia bovis***. *Gene* 1996, **172**:173-174.

51. Dibrov E, Robinson KM, Lemire BD: **The COQ5 gene encodes a yeast mitochondrial protein necessary for ubiquinone biosynthesis and the assembly of the respiratory chain**. *The Journal of biological chemistry* 1997, **272**(14):9175-9181.

52. Baroli I, Melis A: **Photoinhibitory damage is modulated by the rate of photosynthesis and by the photosystem II light-harvesting chlorophyll antenna size**. *Planta* 1998, **205**:288-296.

53. Le QH, Melayah D, Bonnivard E, Petit M, Grandbastien MA: **Distribution dynamics of the Tnt1 retrotransposon in tobacco**. *Mol Genet Genomics* 2007, **278**(6):639-651.

54. McClintock B: **Controlling elements and the gene**. *Cold Spring Harbor symposia on quantitative biology* 1956, **21**:197-216.