Spring 5-2014

Regulation of Gene Expression in *Karenia brevis*

Skylar C. Rodgers
University of Southern Mississippi

The University of Southern Mississippi

Regulation of Gene Expression in *Karenia brevis*

by

Skylar Rodgers

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science
In the Department of Biological Sciences

May 2014

Approved by

_____

Timothy McLean, Ph.D., Thesis Advisor
Assistant Professor of Biology

_____

Shiao Wang, Ph.D., Chair
Department of Biological Sciences

_____

David R. Davies, Ph.D., Dean
Honors College

Abstract

Karenia brevis is species of dinoflagellate responsible for most of the harmful algal blooms that occur in the Gulf of Mexico. These blooms can be detrimental to the environment and the economy of a coastal region due to the brevetoxins produced by *Karenia brevis*. Currently, the cause of these blooms, as well as the mechanisms of associated toxin production, are unknown. Efforts to characterize *Karenia brevis* at a molecular level are ongoing. However, based on genomic findings, researchers have hypothesized that regulation of gene expression occurs post-transcriptionally. In many organisms, non-coding RNAs, such as natural antisense transcripts (NATs), play crucial roles in gene regulation. The goal of this project was to assess a possible role of NATs in the post-transcriptional regulation of gene expression in *Karenia brevis*. To that end, RNA was extracted from *Karenia brevis* samples at dusk and dawn and sequenced at a core facility using Illumina HiSeq RNA sequencing. The sequences were then processed and assembled into a transcriptome using various softwares. Statistical analyses were performed that corroborated the validity of the transcriptome. Research stopped at this point due to time constraints. Further bioinformatics sequence analysis could yield a better understanding of differential gene expression in *Karenia brevis* and the role of NATs in such expression.

Key Words: harmful algal blooms, natural antisense transcripts, gene regulation

Acknowledgements

I would like to thank my advisor, Dr. Timothy McLean for his guidance and support throughout the completion of this project. Without him, this project would not have been possible. I would also like to thank Scott Anglin for all of his assistance and for his patience in teaching me throughout the course of this project.

Table of Contents

List of Figures

List of Abbreviations

HABs	harmful algal blooms
mRNA	messenger RNA
siRNA	small interfering RNA
miRNA	micro RNA
NATs	natural antisense transcripts
psu	practical salinity unit
rcf	relative centrifugal force

Chapter 1: Introduction

The colloquial term "red tide" refers to the natural phenomenon that occurs when the growth of a colony of marine algae increases exponentially and forms a bloom. These blooms, formally referred to as harmful algal blooms (HABs), can have negative impacts on both the environment and the economy of coastal regions. *Karenia brevis*, a marine dinoflagellate, is responsible for the harmful algal blooms that plague the Gulf of Mexico. The factors that prompt the formation of these blooms are currently unknown. Researchers have tested various environmental factors for correlation to the formation of blooms; however, little data has supported environmental causation of bloom initiation. Cellular processes, therefore, are thought to underlie bloom dynamics. Currently, researchers are studying regulation of gene expression in *Karenia brevis* as a method of better understanding the cellular processes that could impact bloom dynamics. Regulation of gene expression in *Karenia brevis* is believed to occur post-transcriptionally through binding of non-coding RNA to messenger RNA (mRNA). Natural antisense transcripts, a unique type of noncoding RNA, are of particular interest in the post-transcriptional regulation of gene expression in *Karenia brevis*. (McLean, personal communication)

In this study, RNA will be extracted and sequenced from *Karenia brevis* samples to determine whether or not NATs may play a role as post-transcriptional regulators in this organism. Because *Karenia brevis* is a photosynthetic dinoflagellate and is on a circadian cycle, light is thought to influence protein expression. Therefore, the samples will be taken at dawn and dusk to study the differences in the relative concentration of double-stranded RNA at each time of day.

Chapter 2: Literature Review

The term "algae" describes a large group of either unicellular or multicellular organisms that are often photosynthetic and have plant-like characteristics. Some types of algae are capable of forming algal blooms. Blooms occur when the concentration of a particular species of algae increases immensely from the background levels of that species in a particular area. (McLean, 2013) Some algal blooms release toxins to their environment and thus have been labeled "harmful algal blooms" (HABs). (McLean, 2013) Dinoflagellates are responsible for the majority of HAB occurrences. Dinoflagellates are a diverse group of protists that are present in and can adapt to many different environments in both marine and fresh water. Dinoflagellates are often photosynthetic, meaning that they are able to use light as an energy source to gain nutrients; however, studies have revealed that many of these photosynthetic dinoflagellates are able to gain their energy from multiple sources and are actually mixotrophic. Dinoflagellates often have a reddish-brown appearance due to the presence of chlorophylls *a* and *c2*, carotenoid beta-carotene, and several xanthophylls. (Hackett, Anderson, Erdner, Bhattacharya, 2004) The outward cell structure of dinoflagellates can fall into one of two categories: armored or unarmored. Armored cells contain several different polysaccharides, including cellulose, that strengthen the cells. Conversely, unarmored cells have only one layer of plasmalemma around them. Consequently, unarmored dinoflagellates are extremely fragile and easily lysed. (Hackett et al., 2004)

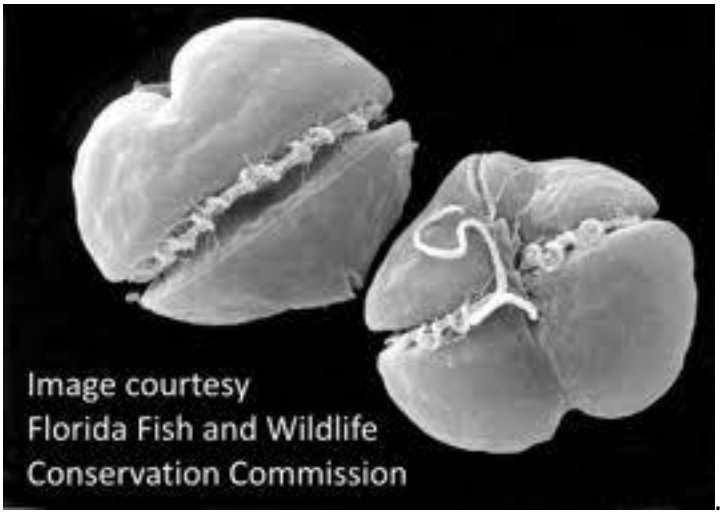

Figure 1: Karenia brevis Cells
Karenia brevis unarmored cells as viewed by an scanning electron microscope.
Florida Fish and Wildlife Conservation Commission

One particular unarmored toxic marine dinoflagellate, *Karenia brevis* (Figure 1), is responsible for the harmful algal blooms that have often plagued the Gulf of Mexico, especially along the Florida and Texas coasts. *Karenia brevis* blooms have historically been called "red tides" due to the red coloration of the water caused by the formation of HABs (Figure 2), although not all HABs appear red. Descriptions of red tides have been documented as early as the 1500s, and have been consistently reported since the 1800s. (Steidinger, 2009) Red tides have historically been responsible for widespread fish kills, animal mortalities, neurotoxic shellfish poisoning (NSP), and respiratory irritation through the release of toxins. (Kirkpatrick et al., 2004) Neurotoxic shellfish poisoning is a serious, although usually non-fatal, illness caused by the ingestion of shellfish that have been heavily exposed to the toxins. (Lekan & Thomas, 2010)

Karenia brevis blooms have a negative impact on the environment, the economy, and human health in coastal regions.

Figure 2: Red Tide in Gulf of Mexico
Karenia brevis HAB on the Florida coast.
Image borrowed from www.noaanews.noaa.gov.

The first report of what is now considered a *Karenia brevis* bloom occurred in 1530 when Cabeza de Vaca, a Spanish explorer, described red pigment in the water and fish kills around the coast of Florida in a journal. (Van Dolah et al., 2009) During the 1800s there were about twenty red tide cases documented. One of the worst *Karenia brevis* blooms recorded to date occurred during 1946-1947 and lasted 11 months. This bloom prompted subsequent research into the cause and effect of *Karenia brevis* bloom formation. (Steindinger, 2009) Scientists have studied the growth of *Karenia brevis* at different nutrient levels both in the lab and in nature to determine what nutrients must be present for growth and subsequent HAB formation. *Karenia brevis* is a mixotrophic organism that is capable of utilizing multiple nutrient sources. (McLean, 2013) In marine waters, small amounts of nitrogen, phosphorus, and trace metals, especially iron, may be able to sustain this organism. (Lekan &Thomas, 2010) Laboratory studies have also revealed the importance of trace metals to the growth of *Karenia brevis.* (Steindinger, 2009) The optimum temperature for growth is generally accepted to be around 20-28 degrees Celsius. (Lekan &Thomas, 2010) However, *Karenia brevis* has been grown in laboratory

cultures at as low as 7 degrees Celsius. (Steidinger, 2009) For years, the optimum salinity for growth was considered to be between 24 and 44 psu. However, recently blooms have been documented in areas where the salinity was less than 24 psu. (Lekan & Thomas, 2010)

The toxins released by *Karenia brevis* during harmful algal blooms are called brevetoxins. Brevetoxins have polyether ladder structures that bind to sodium sensitive voltage gated channels in the cell membrane. Their binding causes the channels to remain open and disrupts cellular signaling. Although harmful algal blooms may take days or even weeks to form, brevetoxins can be released and cause toxicity at very low concentrations. (McLean, 2013) Because *Karenia brevis* cells are unarmored, the movement of the ocean easily causes them to split and release harmful toxins. (Lekan & Thomas, 2010) Researchers tested the brevetoxin composition of three samples of *Karenia brevis* at different nutrient levels and at different salinities. They found that the salinity and nutrient level did not cause a particular pattern in brevetoxin composition. Therefore, the study supported the hypothesis that brevetoxin composition is linked more closely to the genome of the cell rather than environmental conditions. (Lekan & Thomas, 2010) However, a more recent study linked an increase in brevetoxin production to phosphorus limitation. (Hardison et al., 2013)

Currently, many researchers are using a molecular approach to understand the dynamics of *Karenia brevis* bloom formation. *Karenia brevis'* genome is composed of 121 chromosomes (Lidie, Ryan, Barbier, Van Dolah, 2005); it is about 30 times larger than the human genome. (VanDolah et al., 2009) The function of the massive amount of DNA found in *Karenia brevis* is not completely understood. By comparison, for many

years, scientist believed that the human genome was composed largely of "junk" DNA. The Human Genome Project revealed that only about 2% of DNA codes for protein. (Robinson, 2009) Subsequent studies, however, have ascribed various regulatory functions to the vast amount of noncoding DNA found in the human genome. A microarray study has shown that the level of *Karenia brevis* transcripts remains practically unchanged between daylight and night. (Van Dolah et al., 2007) Because the genes in dinoflagellates are "on," meaning that they are transcribed all of the time, it is widely believed that gene regulation in these organisms occurs post-transcriptionally. Also, *Karenia brevis* lacks many transcriptional regulators found in other eukaryotes such as histones, promoters, and TATA boxes. (VanDolah et al., 2009) The RNA transcribed in the cell that does not get directly translated into protein is called non-coding RNA. Small interfering RNA (siRNA), micro RNA (miRNA) and natural antisense transcripts (NATs) are non-coding RNA thought to be regulatory non-coding RNAs. (Costa, 2007) siRNA and miRNA are both considered small (generally around 20 base pairs long) non-coding RNAs that suppress gene expression. (Robinson, 2009) siRNA and miRNA, however, suppress gene expression differently. siRNA silences genes through highly specific interactions with mRNA that cause the degradation of its complementary mRNA strand. (Rana, 2007) The interactions between siRNA and mRNA are so specific that it is thought that a particular siRNA is only capable of interacting with and silencing one mRNA. (Robinson, 2009) miRNA regulates gene expression through many different, less specific mechanisms. miRNA, like siRNA, is capable of causing mRNA degradations, but miRNA can also stop translation initiation and protein formation through several different mechanisms that are currently less understood. (Robinson,

2009) NATs are considered long non-coding RNAs that play a role in the regulation of gene expression. They are generally 240-1000 base pairs long and contain non-complementary poly-A tails. mRNA can base pair with the regions of NATs that are complementary; the base pairing of mRNA with NATs forms double stranded RNA that cannot be translated to form proteins. However, this complementarity is not precise as NATs contain poly-A tails. (Brosnan & Voinnet, 2009) Recently, scientists have determined that NATs, as well as miRNA and siRNA, function as important regulators of gene expression in many species of eukaryotes.

Because *Karenia brevis* is a photosynthetic organism and is on a circadian cycle, it is thought that the time of day plays a large role in gene expression. Microarray studies have shown that genes involved in photosynthesis are down regulated at night and up regulated during the day; genes that are present during the night are down regulated during the day. (Lidie et al., 2005)

In this study, *Karenia brevis* samples will be tested at dusk and dawn to determine the relative concentration of double stranded RNA present at each time of day. Theoretically, the double-stranded RNA found in the cells is due to the complementary base pairing between NATs and mRNA. This study will investigate the possible interactions between NATs and mRNA in *Karenia brevis* at different times of day to determine the differences in the level of regulation of gene expression due to NATs at each times of day. Also, this study will investigate whether NATs act as positive or negative regulators in *Karenia brevis*. If NATs have a positive regulatory role, the increase of NATs will coincide with an increase of protein expressed. If NATS have a

negative regulatory role, the presence of NATs bound to mRNA will inhibit protein expression (Figure 3).

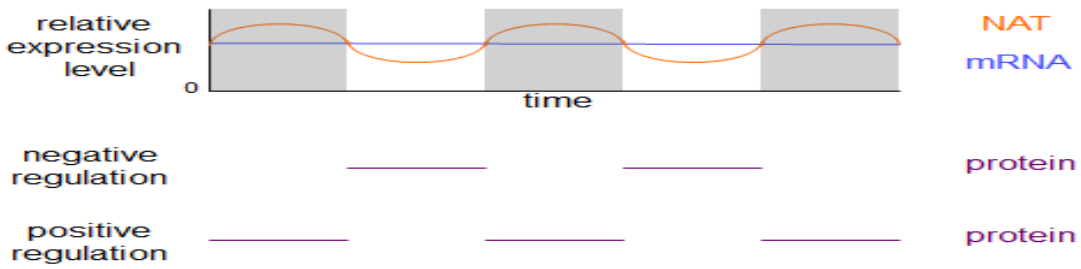

Figure 3: NAT Regulation of Gene Expression
This figure shows one possible relationship between the expression levels of a NAT and its complementary mRNA over successive nights (shaded areas) and days (non-shaded areas). If NATs are positive regulators then as the concentration of NATs increases protein will be expressed (expressed protein depicted by the purple line). If NATS are negative regulators then protein will be expressed when the concentration of NATs decreases.
Figure created by T. McLean and used with his permission.

Chapter 3: Methodology

In order to determine the relative concentration of NATs at different times of day, the total RNA of *Karenia brevis* samples were isolated and sequenced. *Karenia brevis* cultures were grown in rich media in an incubator that is set to a daily light cycle. The light cycle in the incubator closely mimics that of *Karenia brevis'* natural habitat. Each sample used for this experiment was taken from the cultures grown in an incubator. Each culture originated from the same source.

RNA Extraction

To isolate the total RNA, a 200 mL sample of *Karenia brevis* was taken from the incubator at dawn (6 am) and dusk (6 pm). The following protocol was followed for each sample. The cells were harvested by centrifugation for 5 minutes at 1500 rcf, and the supernatant was disposed.

RNA was extracted from the remaining pellet using the RNeasy Blood and Tissue Handbook spin column protocol. 350 microliters of rapid lysis buffer with β-mercaptoethanal was added to the pellet and transferred to a small centrifuge tube. 250 microliters of 100% ethanol was added to the sample and mixed by gently pipetting. The addition of the same volumes of rapid lysis buffer and ethanol was repeated. Next, 700 microliters of the sample was added to a spin column and centrifuged for 1 minute at 8000 rcf. The flow through was discarded. 500 microliters of buffer wash buffer was added to the spin column and centrifuged at 8000 rcf for 1 minute before the flow through was discarded. The same volume of wash buffer was added once more, but the sample was centrifuged at 8000 rcf for 2 minutes. The spin column was transferred to a new centrifuge tube, and 50 microliters of RNase-free water was added to the spin column. The sample was centrifuged at 8000 rcf for 1 minute. The same volume of RNase-free water was added again and the centrifugation was repeated to yield a sample of 100 microliters total volume.

Precipitation of RNA

A cellulose wash was performed on the supernatant to precipitate the double-stranded RNA. Ten milligrams of cellulose and 100% ethanol was added to the supernatant until a final concentration of 20% ethanol is obtained. This solution was shaken at room temperature for one and a half hours. During this time, the double-stranded RNA bound to the cellulose due to a high binding affinity between the two in the presence of a salt. Next, the solution was centrifuged briefly. Upon centrifuging the solution, the cellulose bound RNA precipitated out of solution to form the pellet. The pellet was washed with a 16.5 % solution of salt-tris-EDTA in ethanol and centrifuged

for 30 seconds; the supernatant was discarded. This wash was repeated five times to ensure the purity of the double-stranded RNA. The remaining pellet was then air-dried for approximately 15 minutes. After air-drying, the pellet was re-suspended in salt-tris-EDTA solution and shaken at room temperature for half an hour. This allows for the dissociation of the double-stranded RNA from the cellulose. The solution was centrifuged again, and the supernatant was extracted. The supernatant contains the purified double-stranded RNA. The RNA samples were stored in the freezer with a solution of ethanol and sodium acetate for preservation of the sample.

Illumina HiSeq RNA Sequencing

Both the morning and evening samples had their RNA concentration and purity assessed using an Agilent 2100 Bioanalyzer. Once the samples were proven to be of sound quality, they were sent to The Roy J. Carver Biotechnology Center at the University of Illinois for sequencing. Another assessment of RNA integrity was performed using an Agilent 2100 Bioanalyzer after the samples had been treated with DNAses at the University of Illinois. RNA sequence libraries were prepared using Illumina HiSeq RNA sequencing.

Pre-Assembly

After receiving the Illumina raw sequence reads from The Roy J. Carver Biotechnology Center, pre-assembly processing was performed using various software tools to ensure the quality and validity of the reads. FastQC, a software produced by Babraham Bioinformatics, was utilized to check the quality of the reads. NGS QC Toolkit, a software used to ensure the quality of next-generation sequence data, removed low quality reads. Trimmomatic was used to remove the adaptors used by Illumina for

sequencing. Fast X was used to remove the first ten base pairs of all sequences to improve the quality of the sequence data. Finally, the FastQC software was again utilized to ensure that the quality of the data had increased after processing.

Assembly

Both SOAPdenovo-trans software and IDBA-trans software were used independently for assembly of the transcriptome. After each independent assembly of the transcriptome, the integrity of the transcriptomes was compared using Assemblathon_stats.pr, a javascript statistical analysis program.

Comparison of Gene Expression

Further analysis to determine differences between the concentration of NATs and mRNA at different times of day was unable to be completed due to time constraints. In the future, the results of these analyses will either support or disprove the hypothesis that NATs serve as post-transcriptional regulators of gene expression in *Karenia brevis*. If the hypothesis is supported, this information can be used to further the investigation of the role of NATs in the regulation of gene expression in *Karenia brevis*.

Chapter 4: Results

The integrity of the purified RNA samples had to be assessed before the samples could be sent off for sequencing. The initial integrity check was performed using an Agilent 2100 Bioanalyzer (Figures 4A and 4B).

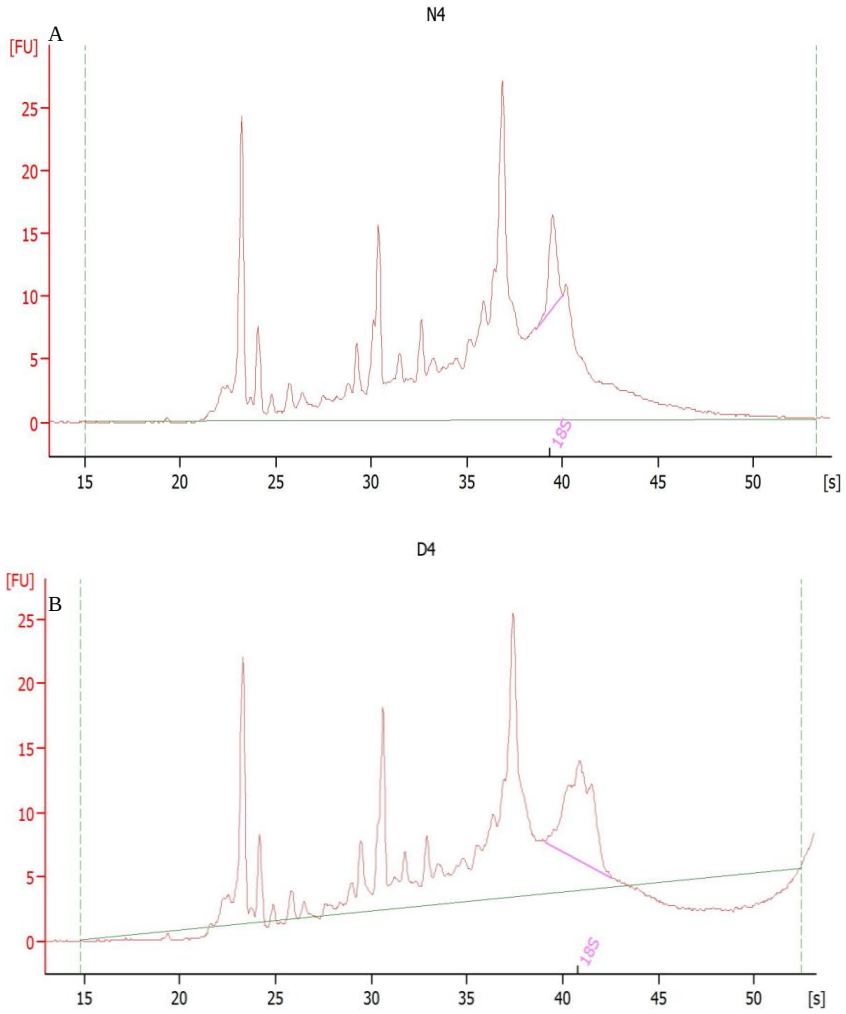

Figure 4- Integrity of RNA Samples
Electropherograms produced by Agilent 2000 bioanalyzer. A) RNA extracted at dawn. B) RNA extracted at dusk.

The presence of two strong peaks around 40 verified that the samples were of sufficient quality to be sequenced. The integrity of the samples was evaluated again at the University of Illinois after being treated with DNases (Figure 5A and 5B). The data from

the Bioanalyzer showed that the DNase treatment had not interfered with the overall integrity of the samples.

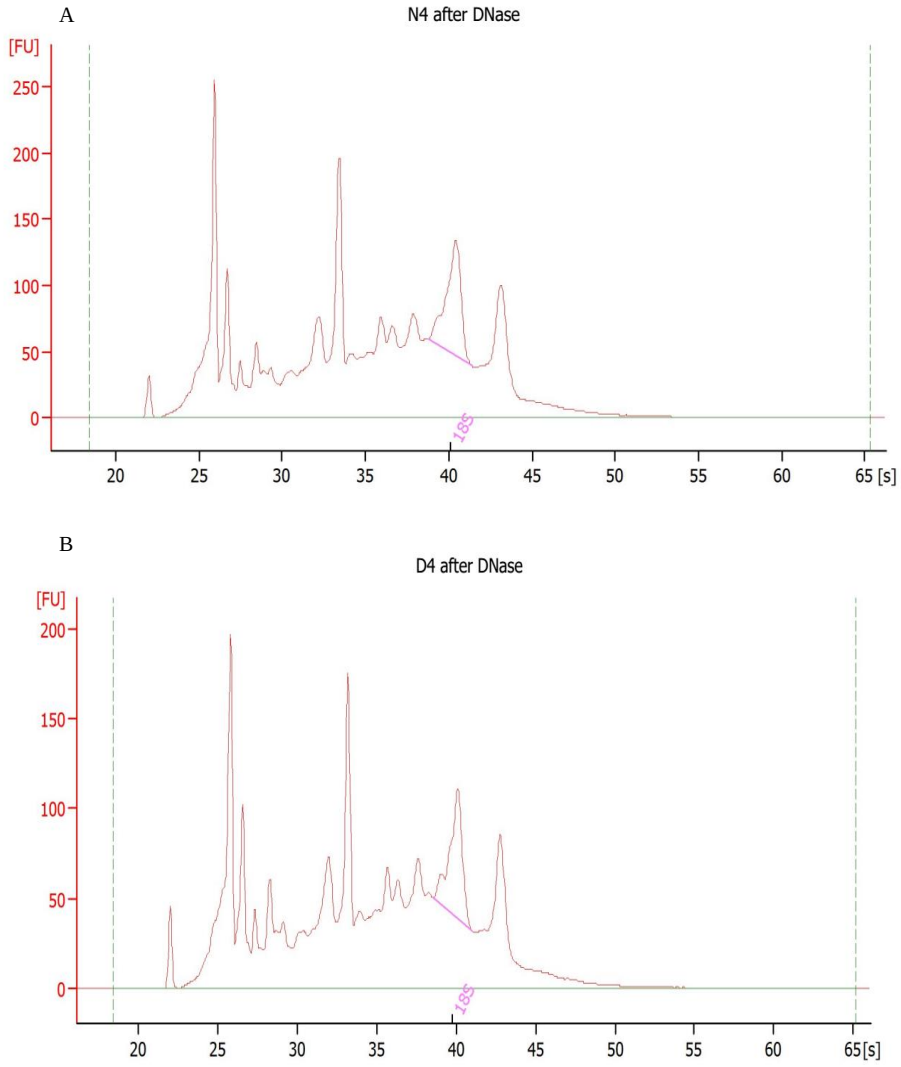

Figure 5: Integrity of RNA Samples After DNase Treatment
Electropherograms produced by Agilent 2000 Bioanlayzer at the Univeristy of Illinois after DNase treatment. A) RNA sample extracted at dawn. B) RNA sample extracted at dusk.

As previously stated, upon receiving the Illumina raw reads, Fast QC software was used to check the quality of the reads. After the initial quality check, the sequences were processed using the aforementioned software tools to improve the overall quality of the sequences.

The mean quality scores for all bases within the sequences before pre-processing were high. However, there was some quality fluctuation noticed in the first ten bases of each read (Figure 6A). Fast X software was used to remove the first ten bases from the sequences. Subsequently, the mean quality of the bases within the sequences improved (Figure 6B).

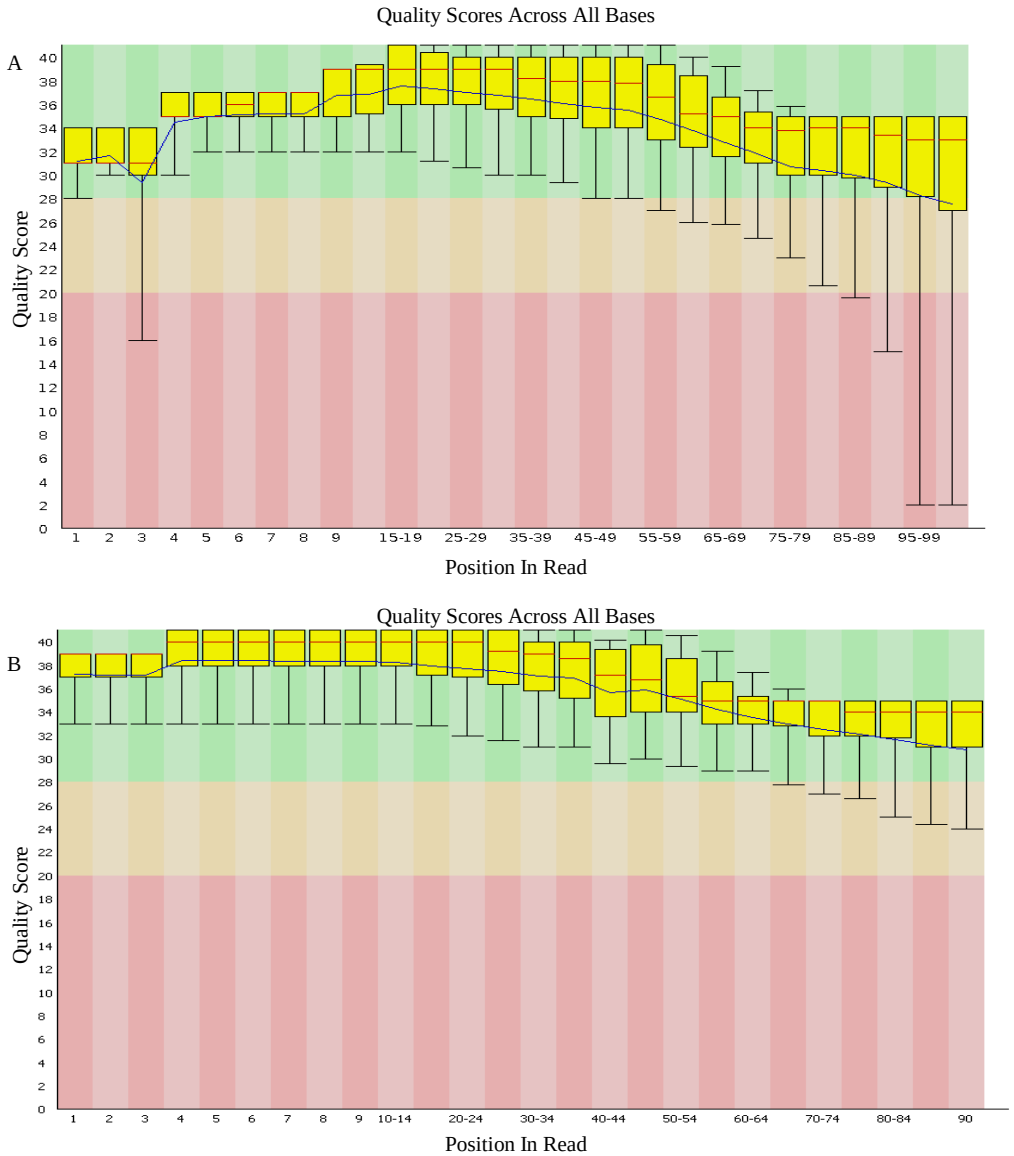

Figure 6: Quality Scores Across All Bases
The red, gold and green zones of the graph represent low, moderate, and high quality scores, respectively. The blue line shows the mean quality of the read, and the red line depicts the median quality score. The yellow boxes depict the interquartile range for each base. The black bars show the 10th to 90th percentile score distribution for each base. A) Quality scores across all bases before pre-processing. B) Quality scores across all bases after pre-processing.

The quality score distribution over all sequences initially was high, with an average quality per read of approximately thirty-six (Figure 7A). After pre-processing the quality score distribution increased, with an average quality per read of approximately thirty-seven (Figure 7B).

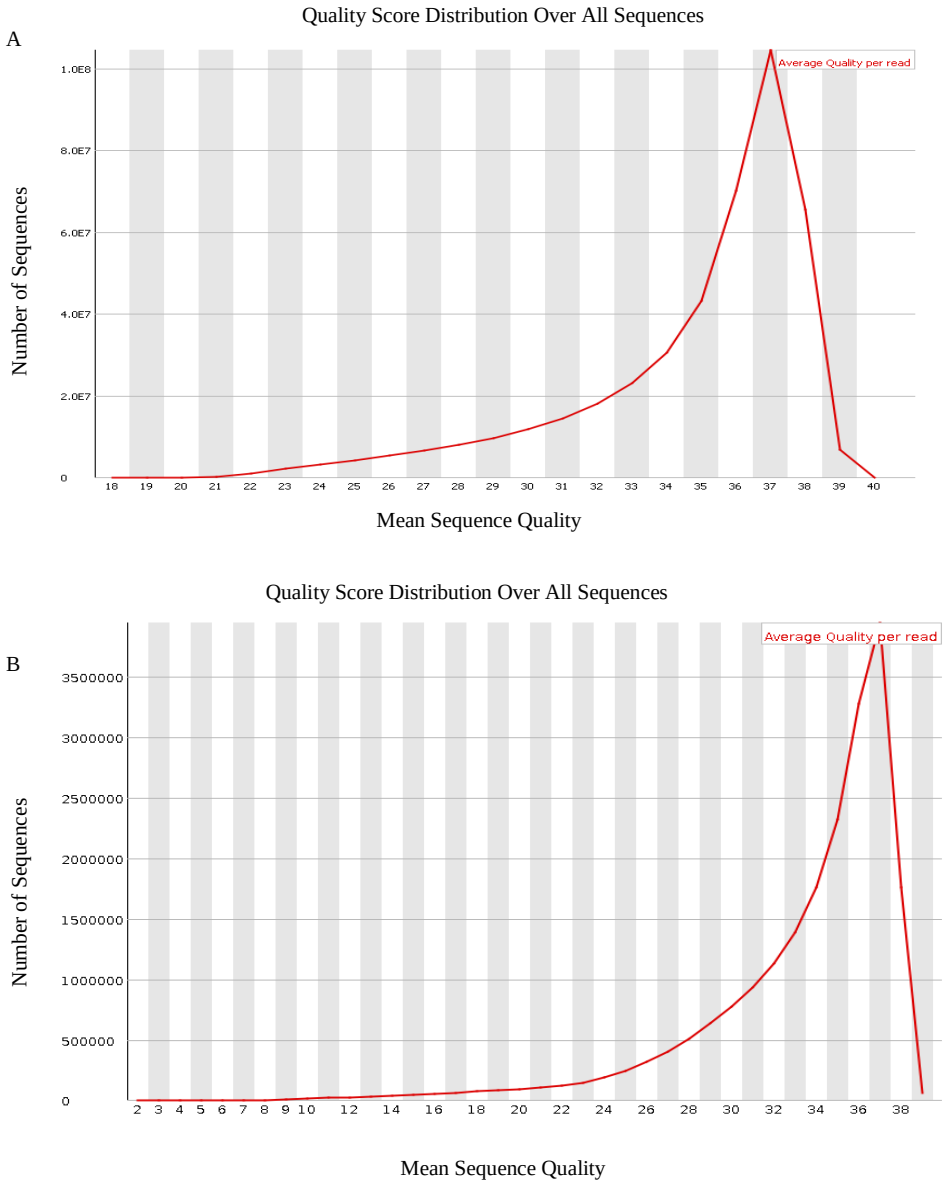

Figure 7: Quality Score Distribution Over All Sequences
The mean quality score over all sequences is depicted by the red line. The tight distribution curve is indicative of high quality. A) Quality score distribution before pre-processing. B) Quality score distribution after pre-processing.

The sequence content of all bases was also was measured as an additional component of the initial quality check. The first ten base positions of the sequence demonstrated specific base sequence composition, whereas the remaining bases in the sequences showed random distribution (Figure 8A, 9A). After processing, base composition was randomly distributed in all positions (Figure 8B, 9B).

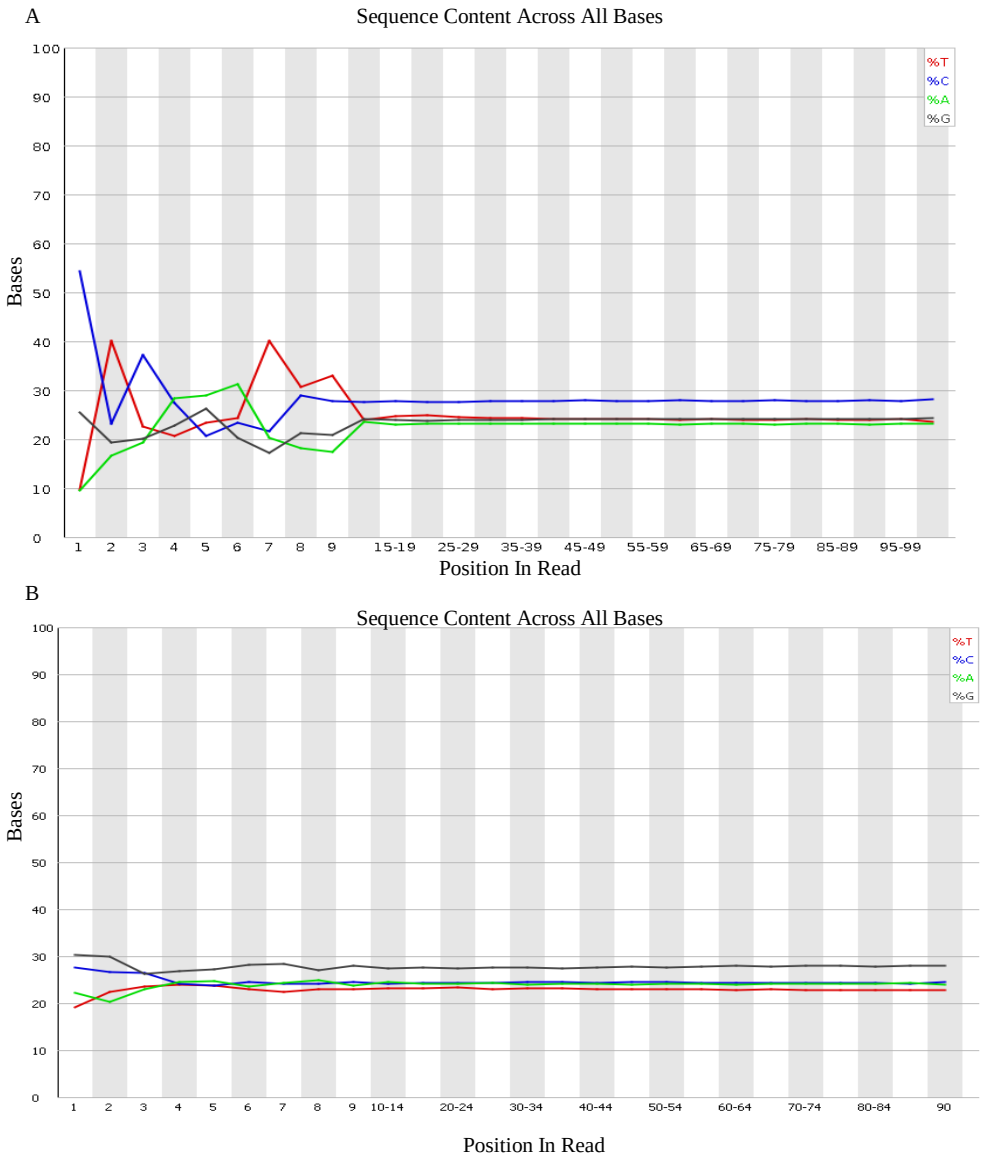

Figure 8: Sequence Content Across All Bases
The base content for all sequences is depicted by color. The concentrations of thymine, cysteine, adenine, and guanine across all sequences are represented by the red, blue, green, and black lines, respectively.

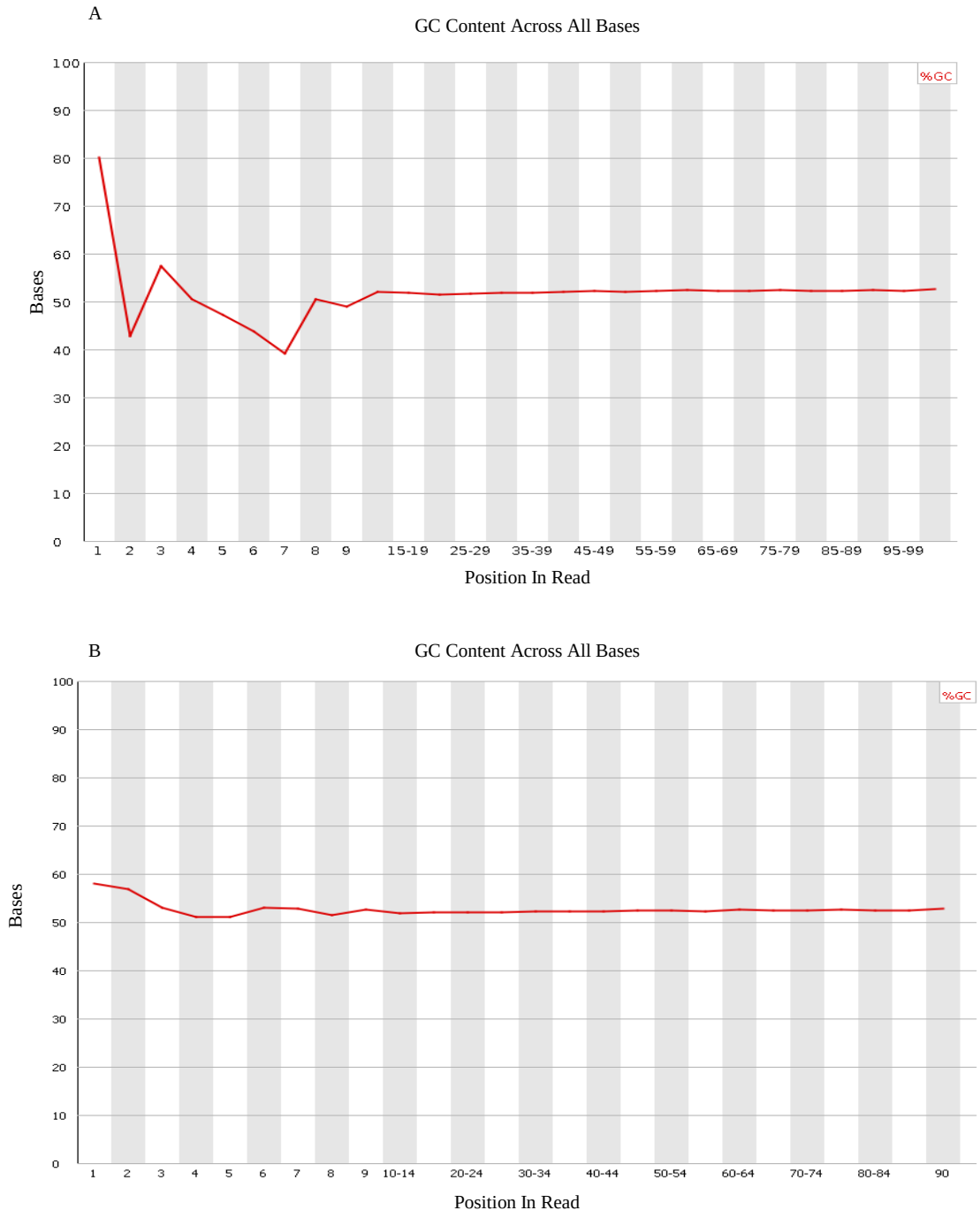

Figure 9: GC Content Across All Bases
The red line represents the GC content across all the bases. A) GC content before pre-processing. B) GC content after pre-processing.

The transcriptomes produced by IDBA and SoapDeNovo assemblers were compared by statistical analysis of their contigs and scaffolds. IDBA produced contigs of larger mean and median size than did SoapDeNovo. However, the N50 contig length was longer in the transcriptome produced by SoapDeNovo than in the transcriptome produced by IDBA (Figure 10).

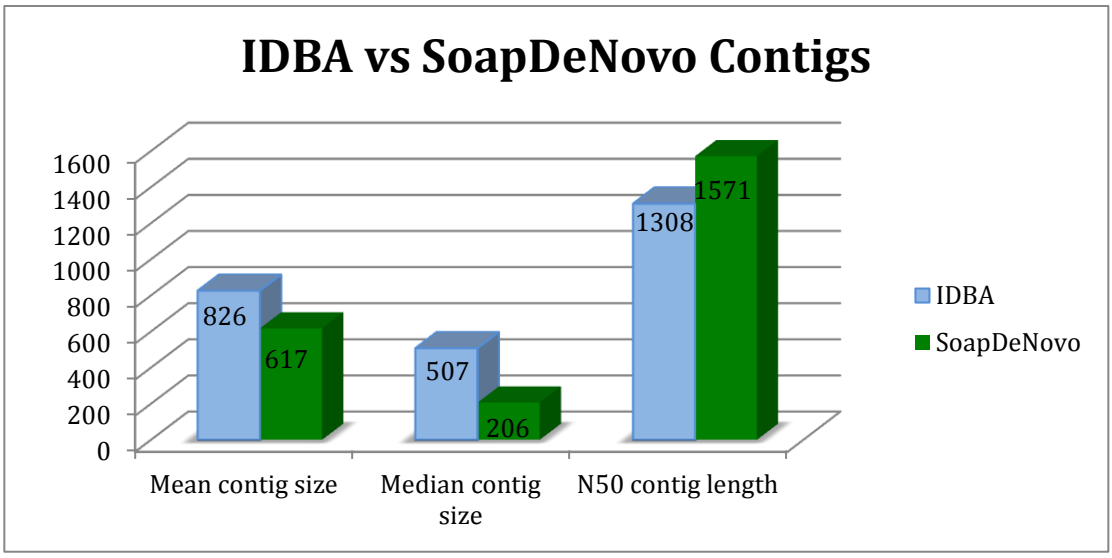

Figure 10: IDBA vs SoapDeNovo Contigs
A comparison of the mean contig size, median contig size, and N50 contig length produced by each assembler.

IDBA also produced scaffolds of larger mean and median size than did SoapDeNovo. The transcriptome produced by SoapDeNovo had the longest N50 scaffold length (Figure 11).

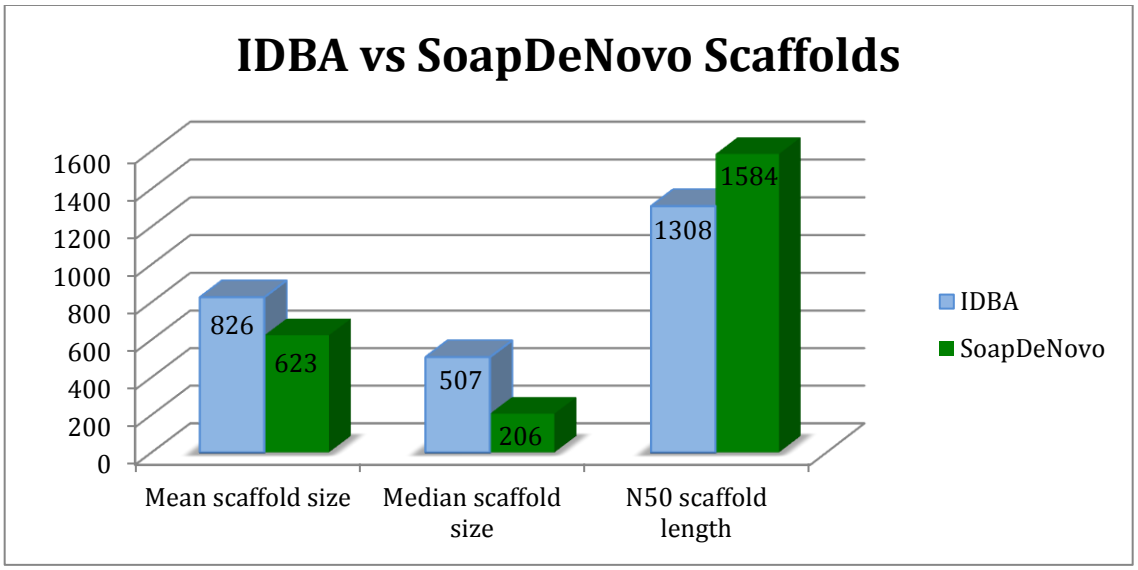

Figure 11: IDBA vs SoapDeNovo Scaffolds
A comparison of the mean scaffold size, median scaffold size, and N50 scaffold length produced by each assembler.

Chapter 5: Discussion

Due to vast size of *Karenia brevis*' genome (Lidie, 2005) and its apparent lack of transcriptional regulation (Van Dolah, 2007), it is believed that regulation of gene expression in *Karenia brevis* occurs post-transcriptionally. NATs, which form long regions of double-stranded RNA, probably play a role in regulation of gene expression in this organism. In order to better understand differential gene expression in *Karenia brevis*, RNA samples were extracted at dusk and dawn and sequenced with the intention of mapping the RNA present in each sample as part of a newly-assembled transcriptome.

After initial extraction of the RNA, the integrity of each sample had to be assessed. The data from the bioanalyzer suggested that both RNA samples were of sufficient quality to be sequenced (Figures 4A, 4B). After the samples were sent to the University of Illinois, they were treated with DNase to futher purify the samples by removing any residual DNA. The integrity of the samples was assessed again, and the

data from the bioanalyzer indicated that the overall quality of the samples had improved (Figures 5A, 5B).

Before the Illumina reads could be used to assemble a transcriptome, quality checks and pre-processing had to occur. Immediately upon receiving the raw reads, Fast QC was used to assess the quality of the reads (Figures 6A, 7A, 8A, 9A). The mean quality across all of the bases was significantly higher than twenty, which is considered an acceptable quality score. However, the mean qualities of the first and last ten bases in the reads were significantly lower than the qualities of the bases in the remaining sequence (Figure 6A). The mean quality score distribution for all sequences was good. The mean quality score across all of the bases within each sequence was calculated, and the distribution of the means was plotted. The mean quality score of the majority of sequences was greater than thirty-two, and the tight distribution curve was indicative of little quality variation. However, some sequences did have low mean quality scores below twenty (Figure 7A). The sequence content across all bases was visualized through an electropherogram. The bases after position ten were randomly distributed throughout all sequences and thus produced uniform lines on the graph. The base composition of positions one through ten in all sequences lacked random distribution. The electropherogram revealed a guanine and cytosine bias and a non-random base composition in the first ten bases (Figure 8A, 9A). This anomaly was probably caused by the priming protocol for sequencing which employs specific adaptors.

In order to improve the quality of the sequences, processing, including the removal of the first ten bases and last ten bases of each read and the removal of any additional low quality reads, occurred. Fast QC was utilized again to check the quality of

the Illumina reads after processing (Figures 6B, 7B, 8B, 9B). After processing, the mean quality score across all bases improved as the mean quality score of all bases was above thirty-two (Figure 6B). The quality score distribution over all sequences increased after processing. The majority of sequences had a mean quality score greater than thirty-five. Additionally, the lowest mean quality scores after processing were above twenty. Therefore, after processing, all sequences were of acceptable quality. The score distribution curve indicated little variation in mean sequence quality in the majority of sequences (Figure 7B). After the first ten bases had been cleaved from all sequences, the base content throughout each read was randomly distributed as expected when working with such a large number of reads (Figure 8B). Additionally, the removal of the first ten bases in each sequence eliminated the guanine and cytosine bias in the reads (Figure 9B).

These sequences, along with additional sequences from an associated experiment, were used to assemble a transcriptome. *De novo* transcriptome assembly software had to be used because there was no reference genome for *Karenia brevis*. IDBA and SoapDeNovo program packages were both used to assemble independent transcriptomes. A statistical analysis was performed on each transcriptome comparing the contigs and scaffolds of each to determine which software produced the most valid transcriptome. Contigs refer to the overlapping reads resulting from the assembly of short sequences. Scaffolds are composed of overlapping contigs. The transcriptome produced by IDBA contained contigs and scaffolds of larger size on average than the transcriptome produced by SoapDeNovo (Figures 10, 11). However, the N50 length for both the contig and the scaffold for the transcriptome produced by SoapDeNovo were larger (Figures 10, 11). N50 is a weighted statistical measurement of median contig and scaffold length. The

N50 refers to a value at which fifty percent of the transcriptome is contained in contigs or scaffolds equal to or larger than the given value. Because the statistical analysis of the assemblers showed little difference between the transcriptomes produced, it was determined that both IDBA and SoapDeNovo produced valid transcriptomes, and either could be used for further analysis.

In the future, further bioinformatics sequence analysis, including mapping and annotation, will be performed in order to better understand the role of NATs as positive or negative regulators of gene expression in *Karenia brevis*.

Chapter 6: Conclusion

To conclude, the role of NATs in the regulation of gene expression in *Karenia brevis* was unable to be determined due to time constraints on the project. However, the bioinformatic pipeline was utilized for successful assembly of the transcriptome, which is a significant accomplishment and will be of great value for future work in the McLean laboratory. The transcriptome that was produced can be utilized for further bioinformatics analysis, including alignment and detection of differentially expressed genes, that could yield a better understanding of the role of NATs in gene expression in *Karenia brevis*.

References

Brosnan, Christopher A., Voinnet, Oliver. 2009. The long and short of noncoding RNAs. Current Opinion in Cell Biology 21:416- 425.

Costa, F. 2007. Noncoding RNAs: lost in translation? Gene 386:1-10.

Hackett, Jermiah D., Anderson, Donald M., Erdner, Deana L., Bhattacharya, Debashish. 2004. Dinoflagellates: a remarkable evolutionary experiment. American Journal of Botany 91(10):1523-1534.

Hardison, Donnie R., Sunda, William G., Shee, Damian, Litaker, Richard W. 2013. Increased Toxicity of Karenia brevis during Phosphate Limited Growth: Ecological and Evolutionary Implications. PLoSONE 8(3).

Kirkpatrick, B., Fleming, L.E., Squicciarini, D., Backer, L.C., Clark, R., Abraham, W., Benson, J., Cheng, Y.S., Johnson, D., Pierce, R. 2004. Literature review of Florida red tide: implications for human health effects. Harmful Algae 3:99–115.

Lekan, Danelle K., Thomas, Carmello R. 2010. The brevetoxin and brevenal composition of three Karenia brevis clones at different salinities and different nutrient conditions. Harmful Algae 9:39–47.

Lidie, K. B., Ryan, J. C., Barbier, M. & Van Dolah, F. M. 2005. Gene expression in Florida red tide dinoflagellate Karenia brevis: analysis of an expressed sequence tag library and development of DNA microarray. Biotechnology 7:481–493.

McLean, Timothy. 2013. “Eco-omics”: A review of the application of genomics, transcriptomics, and proteomics for the study of the ecology of harmful algae. Microbial Ecology 65(4):901-915.

Rana T.M. 2007. Illuminating the silence: Understanding the structure and function of small RNAs. Cell Biology 8:23–36.

Robinson, Victoria L. 2009. Rethinking the central dogma: noncoding RNAs are biologically relevant. Urologic Oncology: Seminars and Original Investigations 27:304–306.

Steidinger, K.A. 2009. Historical perspective on Karenia brevis red tide research in the Gulf of Mexico. Harmful Algae 8 (4):549-561.

VanDolah, F.M. et al. 2009. The Florida red tide dinoflagellate *Karenia brevis*- new insight into cellular and molecular processes underlying bloom dynamics. Harmful Algae 8:562-572.

Van Dolah, F.M. et al. 2007. Microarray analysis of diurnal and circadian regulated genes in the Florida red tide dinoflagellate *Karenia brevis*. Phycology 43(4): 741-752